



UNIVERSITY OF TRENTO

DOCTORAL THESIS

New Analytical Methodologies at the Frontier of Cellular Lipidomics

Author:
Ruggero FERRAZZA

Tutor:
Prof. Graziano GUELLA

Advisor:
Dr. Pietro FRANCESCHI

External Advisor:
Dr. Julian L. GRIFFIN

International Doctoral School in Biomolecular Sciences

29th Cycle

Bioorganic Chemistry Laboratory (Department of Physics)
Centre for Integrative Biology (CIBIO)

Academic Year 2015 – 2016

Declaration of Authorship

I, Ruggero FERRAZZA, declare that this thesis titled, “New Analytical Methodologies at the Frontier of Cellular Lipidomics” and the work presented in it are my own. I confirm that:

- This work was done while in candidature for a research degree at this University;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given;
- I have properly and fully acknowledged all main sources of help;
- Where the work was done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Abstract

New Analytical Methodologies at the Frontier of Cellular Lipidomics

by Ruggero FERRAZZA

Lipids were once thought to only be the building blocks of cell membranes and to serve as energy reserves. With time however, it became increasingly clear that they are actually involved in many more roles. Not surprisingly, the comprehensive characterisation of lipids in cells and tissues has experienced a growing interest worldwide, to the point that the term “lipidomics” was coined. This field is a subset of metabolomics, and the interesting point about these two sciences is that they are closest to the phenotype as compared to their “omics” counterparts (genomics, transcriptomics, ...), because metabolites and lipids are the end products of the -omics cascade.

We have investigated mass spectrometry-based lipidomics from different perspectives: first of all, we have devised a targeted approach in which we have focused on sphingolipids and their perturbations. We started by working on neuronal cell cultures where we inhibited GBA, a key enzyme of the sphingolipid metabolism known to be one of the risk factors for Parkinson’s disease. We found a significant sphingolipid unbalance characterised by an accumulation of glycosyl-ceramides. We then moved on by investigating the effects that LRRK2, an important and complex protein known to be related to autosomal-dominant forms of the disease, has on sphingolipids. We worked on mouse models, and we compared the sphingolipid profiles of wild-type (*Lrrk2*^{+/+}) and knock-out (*Lrrk2*^{-/-}) mice, finding a marked increase in ceramide levels and, more in general, in all lipids downstream of GBA. Such results hint to a possible interaction between LRRK2 and GBA, with LRRK2 playing a role in GBA regulation.

In a second lipidomics investigation, we tried to understand whether or not anti-cancer treatments affect the lipid composition of tumours. Specifically, we concentrated on a common anti-angiogenic drug, whose aim is to starve cancer cells by inhibiting angiogenesis, a process required by the tumours to grow. We considered four different adenocarcinoma cell lines, which were subcutaneously inoculated into mice; the “control” animals received no treatment, whereas the “treated” ones were periodically given the drug. Interestingly, we found the treatment to have significant effects on the cancer lipidome, although the different lines responded unequally to the drug. Such results may reflect the huge heterogeneity of cancers and of individual responses to the treatment.

Finally, we developed an informatics algorithm that deals with labelling experiments. The key point is that mass spectrometry measures isotopic patterns of analytes, which depend on the isotopic distribution of the elements; consequently, if an analyte incorporates the stable isotope employed in a labelling experiment, it will show a modified isotopic pattern. Our algorithm analyses such pattern, estimating the abundance of the incorporated label; we first tested it over carefully planned samples, and then we used it in a biochemical application where we wished to establish whether the rate of *de novo* lipogenesis is influenced by diet. This was accomplished by designing an experiment where mice were given partially deuterated water, while being fed different diets; we were able to ascertain that diet does indeed

affect *de novo* lipogenesis, with the lowest rates occurring on fat-rich diets. We are confident that our tool may find useful applications, considering that stable isotope-based labelling experiments are becoming more and more popular.

Thesis Organisation and Contribution

This PhD Thesis summarises my three years of work in the field of lipidomics, during which time I have followed three different projects. Since there is no common line joining them, I have chosen to dedicate a separate Chapter to each.

1. Chapter 1 is a general introduction: here I briefly describe the main aims of lipidomics and the analytical tools more commonly used, with a focus on the aspects more relevant for this Thesis. In particular, I point out some critical issues of the lipidomics technique that I had to face during the PhD. For the sake of keeping the introduction as general as possible, I do not introduce here the specific biological questions of the following projects, which instead I have placed at the beginning of the related Chapters.
2. In Chapter 2 I describe the first project, a targeted lipidomics investigation where we looked for possible sphingolipids perturbations in either neuronal cell cultures or mouse brains, associated to *Gba* and/or *Lrrk2* genes. Our interest was motivated by the fact that they seem to be involved in the development of Parkinson's disease. The biological part was carefully planned and carried out by Dr. Nicoletta Plotegher and co-workers at the Department of Biochemistry, University of Padova (Italy). On the other hand, my work commenced at the lipid extraction step, and I looked after all the subsequent stages. In particular, I took care of the analytical part, including the choice of the targeted approach, the optimisation of the instrumental parameters and working conditions, and the statistical analysis of the LC-MS data.
3. Chapter 3 is the result of a collaboration with Dr. Stefano Indraccolo and his research group of the Istituto Oncologico Veneto (Padova, Italy). His research focuses on anti-angiogenic treatment of cancer, and our aim here was to understand whether such treatment can trigger lipidomics alterations and, if so, whether they depend on the phenotype of the tumour cells. As with the previous project, I did not conceive the biological experiments, but rather I took care of the analytical part. A great help here came from Steven Murfitt (Julian L. Griffin's group, Department of Biochemistry, University of Cambridge – UK), who kindly carried out the high-resolution LC-MS measurements.
4. Finally, Chapter 4 describes a non-conventional lipidomics approach making use of stable isotopes, and I dealt with this project during my training period abroad at the Department of Biochemistry, University of Cambridge – UK. My advisor was Dr. Julian L. Griffin, and in his group they have been using stable isotopes (^2H and ^{13}C) to label metabolites; I developed a software package aimed at analysing MS isotopic patterns, with the final goal to estimate the amount of label enrichment within target analytes. We also applied it to experimental LC-MS data, kindly provided by Dr. Nyasha Munjoma, Dr. N. Zhang (^{13}C experiments) and Dr. Francis Sanders (^2H experiments).

List of Publications

- Ferrazza, R., Griffin, J. L., Guella, G. & Franceschi, P. IsotopicLabelling: an R package for the analysis of MS isotopic patterns of labelled analytes. *Bioinformatics* (2016). doi: 10.1093/bioinformatics/btw588.
- Ferrazza, R., Cogo, S., Melrose, H., Bubacco, L., Greggio, E., Guella, G., Civiero, L. & Plotegher, N. LRRK2 deficiency impacts ceramide metabolism in brain. *Biochemical and Biophysical Research Communications* **478**, 1141–1146 (2016). doi: 10.1016/j.bbrc.2016.08.082.

Acknowledgements

It is astonishing to think about all the people who made it possible for me to write this Thesis: there are so many to thank, from those who brought scientific insights, ideas and suggestions to those who helped me live through this experience.

A person that undoubtedly belongs to both categories is my tutor, Prof. Graziano Guella, whom I will never cease to thank: in these years, he has personally and patiently shaped my scientific knowledge, allowing me to grow as a research scientist.

Special thanks go to my advisor, Dr. Pietro Franceschi, whose skills in informatics and data analysis, as well as his critical scientific guidance have been priceless for my PhD. Also, I would like to acknowledge my external advisor, Dr. Julian L. Griffin, who hosted me in his laboratories in Cambridge and gave me the opportunity to work on a wonderful research project.

Chapter 2 wouldn't have existed had it not been for Dr. Nicoletta Plotegher, who brought with her a lot of plans and determination. For me, the result of her compelling scientific passion has been a fruitful experience. Here, I would also like to remember her collaborators Prof. Luigi Bubacco, Dr. Susanna Cogo, Dr. Elisa Greggio and Dr. Laura Civiero.

My sincere gratitude also goes to Dr. Stefano Indraccolo and Dr. Matteo Curtarello, with whom another successful collaboration has taken place.

During my PhD, a memorable experience has been my training abroad, at Dr. Griffin's group based in Cambridge. And I would like to thank all the staff that, directly or not, has helped me with my project, including: Steven Murfitt, Dr. Francis Sanders, Dr. Nyasha Munjoma, Dr. Zoe Hall and Dr. Nianshu Zhang.

Of course, I wish to thank all the people working in Prof. Guella's group, in particular I would like to acknowledge the help of my labmate and friend Dr. Andrea Anesi. Also, I am grateful to Adriano Sterni for assisting me whenever some instrumental issue occurred.

Infine, ci tengo a ringraziare i miei genitori Marco e Francesca, mio fratello Lucio e mia sorella Gloria per essermi sempre vicini e per essere sempre riusciti a strapparmi un sorriso, anche nei momenti più impegnativi.

Contents

Declaration of Authorship	i
Abstract	ii
Thesis Organisation and Contribution	iv
List of Publications	v
Acknowledgements	vi
Nomenclature	ix
Glossary	xi
1 An Introduction to Lipidomics	1
1.1 Generalities on Lipids	1
1.2 How to Study Lipids	3
1.3 MS-based Lipidomics	5
1.4 NMR-based Lipidomics	8
1.5 Bioinformatics in MS-based Lipidomics	10
1.6 Lipidomics Applications	11
1.7 Critical Issues	12
1.8 Aims of the Thesis	13
Bibliography of Chapter 1	15
2 A Sphingolipidomics Investigation on the Association Between GBA and LRRK2	19
2.1 Introduction	19
2.2 Primary Cortical Neurons	22
2.2.1 Materials and Methods	22
2.2.2 Results and Discussion	25
2.2.3 Conclusions	29
2.3 <i>Lrrk2</i> ^{+/+} and <i>Lrrk2</i> ^{-/-} Mice	30
2.3.1 Materials and Methods	30
2.3.2 Results and Discussion	31
2.3.3 Conclusions	34
Bibliography of Chapter 2	37
3 The Effects of Anti-VEGF Treatments on the Lipidome of Cancer Cells	41
3.1 Introduction	41
3.2 Materials and Methods	44
3.3 Results and Discussion	46

3.3.1	IGROV-1 Cell Line	46
3.3.2	OC-316 Cell Line	51
3.3.3	OVCAR-3 Cell Line	54
3.3.4	SKOV-3 Cell Line	55
3.4	Conclusions	59
Bibliography of Chapter 3		63
4 <i>IsotopicLabelling: an R Package for the Analysis of Enriched MS</i>		
Isotopic Patterns		66
4.1	Introduction	66
4.2	Development of the <i>IsotopicLabelling</i> R Package	71
4.2.1	Motivation	71
4.2.2	Working Principles	72
4.2.3	The Processing Steps	73
4.2.4	The Example Data Set	78
4.3	Testing the Package	79
4.3.1	LC-MS, Complete ¹³ C Labelling	79
4.3.2	LC-MS, Intermediate ¹³ C Labelling	80
4.3.3	Simulations: Minimum Enrichment Detectable	86
4.4	Biochemical Application	89
4.4.1	² H-labelling	89
4.4.2	Materials and Methods	89
4.4.3	LC-MS Data	90
4.4.4	GC-MS Data	93
4.4.5	Follow-up Experiment	95
4.4.6	Final Considerations	96
Bibliography of Chapter 4		98
A Automatic Interpretation of NMR Spectra of Lipids		102
A.1	The Main Challenges	102
A.2	Block Diagram of the Processing Steps	106
A.3	An Example of the Results	107
A.4	Acyl Chain Composition	108
B Appendix of Chapter 2		109
B.1	Primary Cortical Neurons: Intra-class Profiles	109
B.2	<i>Lrrk2</i> ^{+/+} and <i>Lrrk2</i> ^{-/-} Mice: Intra-class Profiles	110
C Appendix of Chapter 3		112
C.1	Intra-class Distribution for IGROV-1	112
C.2	Intra-class Distribution for OC-316	117
C.3	Intra-class Distribution for OVCAR-3	122
C.4	Intra-class Distribution for SKOV-3	127
D Appendix of Chapter 4		132
D.1	A Practical Script	132
D.2	Biochemical Application: LC-MS Results	133
D.3	Biochemical Application: GC-MS Results	136
D.4	Follow-up Experiment	139

Nomenclature

APCI	Atmospheric Pressure Chemical Ionisation
CBE	Conduritol β -Epoxide
Cer	Ceramide
CI	Chemical Ionisation
CID	Collision-Induced Dissociation
DAG	Diacylglycerol
dc	Direct Current
DLPC	1,2-dilauroyl-sn-glycero-3-phosphocholine
DNL	<i>De Novo</i> Lipogenesis
EI	Electron Impact
ESI	Electrospray Ionisation
FAB	Fast Atom Bombardment
FA	Fatty Acid
FAME	Fatty-Acid Methyl Ester
FFT	Fast Fourier Transform
FID	Flame-Ionisation Detection
FT	Fourier Transform
Gal-Cer	Galactosyl-Ceramide
GBA	Beta-glucocerebrosidase
GC	Gas Chromatography
GD	Gaucher's Disease
Glc-Cer	Glucosyl-Ceramide
GL	Glycerolipid
Gly-Cer	Glycosyl-Ceramide
GP	Glycerophospholipid
HFD	High Fat Diet
HIF	Hypoxia-Inducible Factor
HPLC	High-Performance Liquid Chromatography
HSD	High-Sucrose Diet
LB	Lewy Body
LFD	Low-Fat Diet
LRRK2	Leucine-Rich Repeat Kinase 2
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
<i>m/z</i>	Mass-To-Charge Ratio
NAFLD	Non-Alcoholic Fatty Liver Disease
NA	Not Available
ND	Neurodegenerative Disease
NL	Neutral Loss
NMR	Nuclear Magnetic Resonance
PA	Phosphatidic Acid
PC1	Principal Component 1
PC2	Principal Component 2

PCA	Principal Component Analysis
PC	Phosphatidylcholine
PD	Parkinson's Disease
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PIS	Precursor Ion Scanning
pPC	Plasmanylic and Plasmeylic-PC
pPE	Plasmanylic and Plasmeylic-PE
PS	Phosphatidylserine
PUFA	Polyunsaturated Fatty Acid
QQQ	Triple Quadrupole
RCD	Regular Chow Diet
rf	Radio-Frequency
RT	Retention Time
SCD	Stearoyl-CoA Desaturase
SCID	Severe Combined Immunodeficiency
SEM	Standard Error of the Mean
SER	Standard Error of Regression
SIM	Selected-Ion Monitoring
SM	Sphingomyelin
SNR	Signal-to-noise Ratio
SPE	Solid-Phase Extraction
SP	Sphingolipid
TAG	Triacylglycerol
TIC	Total Ion Current
TOF	Time-Of-Flight
UHPLC	Ultra-High Performance Liquid Chromatography
VEGF	Vascular Endothelial Growth Factor
VLDL	Very Low-Density Lipoproteins
WD	Western Diet
WT	Wild Type
XIC	Extracted-Ion Chromatogram

Glossary

Angiogenesis: The physiological process that generates blood vessels from pre-existing ones. It is of vital importance in growth and development, but it is also involved in the transition of tumours from a benign state to a malignant one;

Conduritol β -Epoxide (CBE): An irreversible inhibitor of the lysosomal enzyme β -glucocerebrosidase (GBA);

Collision-Induced Dissociation (CID): A mass spectrometry technique to induce fragmentation of molecular ions in the gas phase through collisions with an inert gas;

De Novo Lipogenesis (DNL): The enzymatic pathway through which fatty acids are synthesised from acetyl-coenzyme A, which in turn is an intermediate product of metabolism of simple sugars such as glucose;

Electrospray Ionisation (ESI): One of the most common ionisation techniques in mass spectrometry, whereby ions are generated by applying a high voltage to a liquid. As a consequence, it is especially useful when it comes to coupling mass spectrometric detection to liquid chromatography;

Extracted-Ion Chromatogram (XIC): Chromatogram created by plotting the intensity of the signal observed at a chosen m/z value or set of values in a series of mass spectra recorded as a function of retention time;

β -Glucocerebrosidase (GBA): A lysosomal enzyme that breaks down glucosyl-ceramides into glucose and ceramides;

Lewy Bodies (LB): Abnormal protein aggregates that develop inside nerve cells in some neurodegenerative diseases. Their primary structural component is α -synuclein, which can be associated with other proteins such as ubiquitin, neurofilament protein, and α -B-crystallin;

Liquid Chromatography (LC): A technique used to separate a sample mixture into its individual components, based on interactions with two phases, mobile and stationary. In particular, the liquid mobile phase slowly goes through the solid stationary phase, bringing the separated components with it. There are many stationary/mobile phase combinations, and therefore several different types of chromatography;

Leucine-Rich Repeat Kinase 2 (LRRK2): A gene encoding for the LRRK2 protein, a member of the leucine-rich repeat kinase family; the protein is mainly present in the cytoplasm and in the mitochondrial outer membrane. *LRRK2* mutations have been associated with Parkinson's disease;

Mass-To-Charge Ratio (m/z): The mass of a charged chemical species divided by the charge of the species itself;

Neurodegenerative Disease (ND): A range of conditions that primarily affect the neurons in the human brain; such diseases are incurable and debilitating conditions resulting in progressive degeneration and death of nerve cells, thus causing problems with movement or mental functioning. NDs include Parkinson's, Alzheimer's, and Huntington's diseases;

Neutral Loss Scan (NL): An MS/MS-based scan carried out with Triple Quadrupole mass spectrometers, whereby the first and the third mass analysers are synchronised in such a way as to detect the loss of a specified neutral fragment. Such scan mode is useful in targeted analyses;

Parkinson's Disease (PD): One of the most common neurodegenerative diseases, it mainly affects the motor system, although dementia is relatively common in the advanced stages. This disease can arise from a variety of causes, both genetic and environmental;

Precursor Ion Scanning (PIS): A targeted MS/MS-based scan carried out with Triple Quadrupole mass spectrometers, whereby the product ion is selected in the second mass analyser, and the precursor masses are scanned in the first mass analyser. This allows to look for analytes that lose a specified charged species upon fragmentation;

Triple Quadrupole (QQQ): A mass spectrometer having three quadrupoles in series, of which the first and the third act as mass filters, whereas the second one is used to cause fragmentation of the analytes through interaction with a collision gas;

Retention Time (RT): The amount of time a chemical species is retained on a chromatographic column;

Total Ion Current (TIC): The sum of all the separate ion currents carried by the ions of different m/z contributing to a complete mass spectrum;

Unsaturation Index: The average number of unsaturations per carbon atom;

Vascular Endothelial Growth Factor (VEGF): A signal protein produced by cells that stimulates vasculogenesis and angiogenesis. Its normal function is to create new blood vessels during embryonic development and after injury, and when VEGF is overexpressed, it can contribute to disease.

Chapter 1

An Introduction to Lipidomics

1.1 Generalities on Lipids

Many scientists have an understanding on what lipids are, although they have yet to find a universally accepted definition; broadly speaking, lipids are naturally occurring hydrophobic substances which are soluble in organic solvents. According to a more rigorous definition, they are small hydrophobic or amphipathic molecules originating entirely or in part by carbanion-based condensations of thioesters and/or by carbocation-based condensations of isoprene units [1]. This definition encompasses a broad

CATEGORY	ABBREVIATION
Fatty acyls	FA
Glycerolipids	GL
Glycerophospholipids	GP
Sphingolipids	SP
Sterol lipids	ST
Prenol lipids	PR
Saccharolipids	SL
Polyketides	PK

TABLE 1.1: List of the 8 categories into which lipids are grouped.

group of molecules with a wide range of polarity, size, and structure. As a consequence, a lipid classification and nomenclature scheme had to be introduced: lipids are now grouped under eight categories (Table 1.1), each one containing classes, subclasses, subgroups and subsets of lipid molecules. The resulting number of theoretical distinct species is overwhelming and estimated to be close to 200 000 [2]. Although many lipids serve important biological functions, below I will only describe the lipid categories more relevant for the following chapters.

Fatty acyls (FAs) are the building blocks of lipids, consisting of a carboxylic acid with an aliphatic tail that can be saturated or unsaturated; the FA chains usually have an even number of carbon atoms (C14–C26) and *cis*-unsaturated double bonds at definite positions, which can be specified with respect to either the carboxyl group (Δ classification) or the terminal methyl group (ω classification). In humans, FAs can be synthesised through the *de novo* lipogenesis (DNL) process, which mainly produces saturated and mono-unsaturated species [3]. The desaturation process inserts the first double bond in the middle of the chain, and further desaturations occur sequentially toward the terminal group.

Glycerolipids (GLs) have one to three FA chains linked to a glycerol backbone in an ester or ether linkage. Important is the case of triacylglycerols (TAGs), which constitute both the body fat in mammals and the vegetable fats and oils.

Glycerophospholipids (GPs), or simply phospholipids, are similar to GLs, but here a phosphate group is esterified to one of the glycerol hydroxyl groups. Their naming makes use of the stereospecific numbering (*sn*) where the FA substituents are called “radyl” groups; species that lack one radyl group are called “lyso”. GPs are grouped into classes based their polar headgroup at the *sn*-3 position (Figure 1.1), and the most common are phosphatidylcholines (PCs), which usually make up more than 50% of the GPs in eukaryotic membranes [4]. In its turn, each GP class is divided into subclasses on the basis of the *sn*-1 and *sn*-2 substituents. For example, FAs at the *sn*-1 position can be substituted by ether or vinyl ether moieties, giving rise to plasmalyn and plasmenyl GPs, respectively.

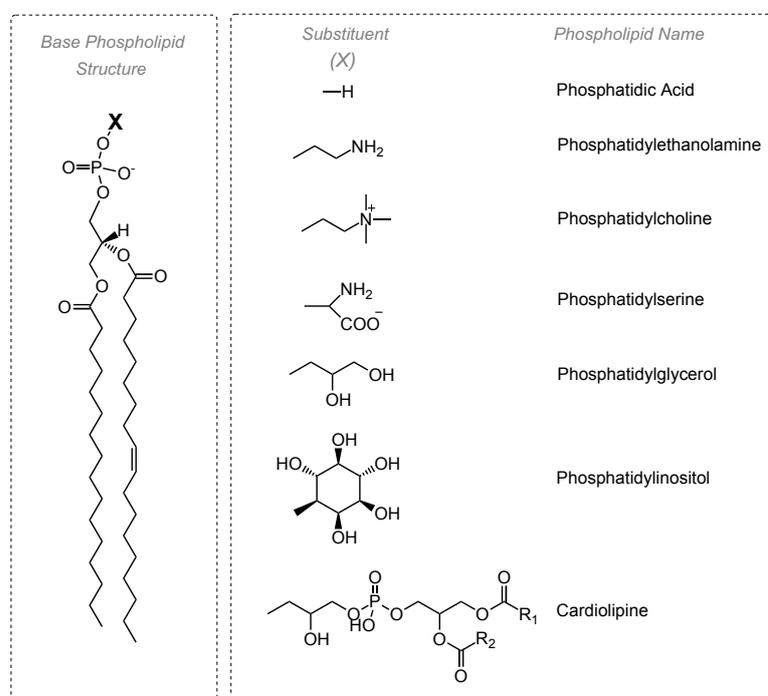


FIGURE 1.1: Scheme of the main glycerophospholipid classes.

Sphingolipids (SPs) are another important lipid category, and their backbone is the aliphatic amino-alcohol sphingosine (Figure 1.2). Sphingolipids are particularly abundant in neural tissues, where they play important roles in signal transduction and cell recognition [5]. Importantly, ceramides (Cer) can be phosphorylated to ceramide-1-phosphate, believed to be a potent pro-inflammatory agent [6]. Also sphingosine and sphingosine-1-phosphate are highly bioactive molecular species [7].

One of the most striking features of lipids is that they can interact with one another and with proteins, giving rise to the important biological membranes cells are made of. Different interactions may take place, depending on the lipids: neutral species interact non-covalently through their chains, whereas polar molecules exploit hydrogen bonds and electrostatic forces. In

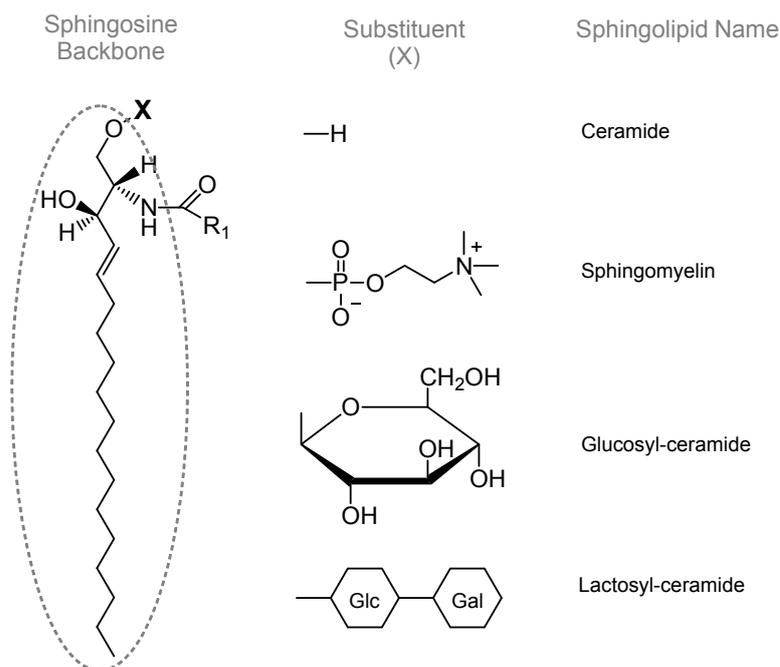


FIGURE 1.2: Scheme of the main sphingolipid classes.

membranes, lipids are relatively mobile, exhibiting lateral diffusion and rotational motions; however, the movement from one leaflet of the lipid bilayer to the other is slow because it requires energy to move the polar headgroups from an aqueous environment into the hydrocarbon domain during the migration. This is the reason why membranes have an asymmetric distribution of polar lipids [4].

For many years lipids have been underestimated, since they were believed to only serve as “bricks” for biological membranes and fuels for bioenergetics. Now it has been finally recognised that, owing to their great diversity, lipids play many crucial biological roles. To name just a few, they are the precursors of important second messengers, they provide the proper environment for membrane-protein function, and their balance is critical for health maintenance. From a dietary point of view, many lipids are essential, meaning that humans cannot synthesise them and have to take them up through the diet. Among these are the polyunsaturated FAs linoleic (C18:2) and linolenic (C18:3). At the same time, lipids can also be deleterious for humans, such as abnormal levels of cholesterol and *trans* FAs [8]. These examples also point out an important aspect related to lipids: despite close analogies among their structures, slightly different lipid species can display divergent biological properties.

1.2 How to Study Lipids

The discovery of lipids’ many biological roles has recently sparked a growing interest in their systematic analysis, to the point that the term “lipidomics” was coined to refer to the large-scale study of lipids and of the factors that interact with them [9]. Lipidomics is a sub-field of metabolomics (the comprehensive investigation of the metabolome), and these two sciences are

important because they are closer to the phenotype than all the other “omics” sciences [10]. Among the many purposes of lipidomics are (i) to link lipids and their pathways to metabolic health, and (ii) to interpret changes in lipid metabolism from a physiological or pathological perspective [11, 12]. One of the greatest challenges of lipidomics comes from the wide range of physical properties that lipids have. As a consequence, it is extremely difficult to extract and analyse the whole lipidome of any biological sample [13].

From an analytical point of view, lipidomics requires three steps:

1. Lipid extraction from the biological samples;
2. Analytical separation through chromatographic methods;
3. Lipid identification and quantification.

The first step aims at extracting as many lipids as possible, at the same time leaving behind all non-lipid components such as proteins, metabolites and inorganic salts. The extraction protocols usually couple organic solvents, where lipids are dissolved, with more polar solvents in order to partition and get rid of polar compounds. A phase separation is triggered between immiscible solvents, and lipids are recovered from the hydrophobic phase. Two common protocols are the Folch [14] and the Bligh & Dyer [15] methods, both relying on methanol, chloroform and water: the upper phase is polar and mainly consists of methanol and water, whereas the bottom phase is rich in chloroform, and lipids dissolve there. Proteins usually accumulate at the interface between the two.

To the extraction there follow lipid separation and analysis, two usually intertwined steps. As a matter of fact, the separation is not always carried out, and recent approaches have now been emerging that analyse the lipid mixture as it is; they are referred to as “shotgun” lipidomics. As for classical methods, the oldest one consists of coupling gas chromatographic (GC) separation with flame-ionisation detection (FID). However, since GC requires the compounds to be thermally stable and to have high enough vapour pressure to volatilise during injection, sample derivatisation has to be performed, whereby FA chains are first cleaved by hydrolysis and then esterified to produce fatty-acid methyl esters (FAMES). With this approach, information about individual lipid classes is completely lost. A better and more common alternative is to use high-performance liquid chromatography (HPLC), which can be easily coupled with different analytical techniques including mass spectrometry (MS) and UV spectroscopy. Two operating modes are possible:

Normal-phase LC: The stationary phase is polar and the mobile phase apolar. Lipids are resolved by class, based on the different polar heads;

Reversed-phase LC: The stationary phase is apolar and the mobile phase is polar. Here the retention times also depend on the composition of the FA chains [16, 17]: species with shorter and more unsaturated chains elute faster than analogous species with longer and more saturated chains [18].

A more recent implementation of HPLC is Ultra-High Performance Liquid Chromatography (UHPLC): it uses smaller stationary phase particles, with a diameter of less than 3 μm . Higher pressures are needed to push the mobile

phase along the column, which results in higher chromatographic efficiency, reduced elution times and increased sensitivity.

Whether or not chromatographic separations are used, in the end lipids have to be analysed; lipidomics now relies on techniques such as MS and Nuclear Magnetic Resonance (NMR) spectroscopy. Whereas MS has been successfully coupled with HPLC in the popular LC-MS technique, NMR is not amenable to such direct coupling. As a consequence, this analysis can be done by either looking at the lipid mixture in its whole complexity, or by collecting chromatographic fractions and studying them “offline” [19].

1.3 MS-based Lipidomics

Mass spectrometry is currently the technique most commonly employed in lipidomics, owing to its high sensitivity, specificity and its ability to be coupled with chromatographic separation techniques [20, 21]. In MS, neutral analyte molecules are ionised, and the resulting charged species are resolved according to their mass-to-charge ratio (m/z) by magnetic and/or electric fields. A mass spectrum is therefore produced, where the ion signal intensity is plotted against m/z .

In general, all mass spectrometers consist of four essential components (Figure 1.3): an inlet for sample introduction, an ion source where the molecular species get charged, a mass analyser where the ions are separated and a detector where the ion signal is measured [22].

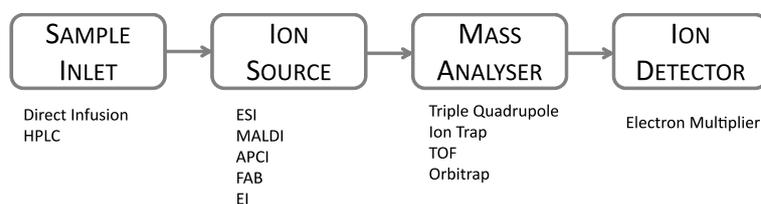


FIGURE 1.3: Block scheme of a mass spectrometer.

Sample Introduction

The sample can be introduced into the mass spectrometer in two ways: with or without a chromatographic step. In the former case, the eluent flow is sent to the mass spectrometer and a series of time-dependent mass spectra is created; in the latter case, the sample is directly injected into the mass spectrometer. Such strategy, referred to as “direct-infusion MS” or “shotgun lipidomics”, is gaining popularity owing to advances in terms of MS instrumentation [23–26]. Its main disadvantage comes from ion suppression issues, where only the most easily ionisable species are detected. In addition, lipids exist in a variety of isomers that cannot be resolved based on their m/z alone.

Ionisation Methods

The ionisation process is required to produce gas-phase ions from neutral molecular species. A number of different methods are available, including

electron impact ionisation (EI), fast atom bombardment (FAB), electrospray ionisation (ESI), chemical ionisation (CI), atmospheric pressure chemical ionisation (APCI) and matrix-assisted laser desorption/ionisation (MALDI) [27]. In my work I only dealt with ESI, and therefore I will describe this technique alone here, which is also the by far most widely employed in lipidomics research.

ESI was developed in the late 1980s by Fenn and co-workers [28]. In ESI, a fine spray of charged droplets is first obtained by injecting the sample solution through a capillary needle at a high electric potential. There follows solvent evaporation, aided by elevated temperatures and by a stream of nitrogen drying gas, that leads to smaller and smaller charged droplets. This causes the surface charge density to increase until a critical point is reached, where ions are ejected in the gaseous phase. These ions are finally taken up by a skimmer cone and accelerated towards the mass analyser [29].

ESI is considered a “soft” ionisation technique, meaning that little or no molecular fragmentation occurs. It is therefore easy to relate the observed m/z values to the mass of the neutral species originating them. ESI can be used in both positive and negative ion modes: in the first case, it is common to observe protonated species, as well as adducts with common cations such as sodium and ammonium (if present in the mobile phase). In negative mode the ions result from the loss of a proton, or from the addition of anions such as chlorine, formate or acetate.

Mass Analysers

Just like ionisation methods, also for the mass analysis there is a range of techniques available on the market; they include quadrupole ion traps, magnetic sector instruments, triple quadrupoles, time-of-flight (TOF), Orbitrap and ion-cyclotron resonance (FT-ICR) mass spectrometers. Below I will just mention the working principles of some of them, mainly TOF, Orbitrap and QQQ mass spectrometers, since I only dealt with them during my PhD.

In the TOF mass analyser ions are accelerated in an electric field to get the same kinetic energy, and are then made to drift along a field-free tube. Different ions are separated on the principle that low-mass ions reach the detector before high-mass ions, and therefore the mass spectrum is a simple recording of the signal as a function of time (t), converted to m/z by the equation

$$m/z = 2 E_{\text{kin}} \left(\frac{t}{l} \right)^2 \quad (1.1)$$

where l is the length of the tube and E_{kin} the kinetic energy [30].

Ion Traps have a doughnut-shaped central electrode and two end-cap electrodes, to which particular direct current (dc) and radio-frequency (rf) potentials are applied, producing a trapping field where ions are confined. By manipulation of the field strength, the analyte ions can be captured and subjected to several analyses including simple m/z scans, MS/MS and MS^n analyses. The Orbitrap is an evolution of the Ion Trap family, characterised by an impressive high mass resolution; it consists of two outer electrodes that have the shape of cups facing each other, and one spindle-like central electrode. A voltage is applied between them, producing a linear electric field that causes harmonic oscillations of the ions trapped inside. With a

correct choice of parameters, the ions remain on a circular spiral inside the trap. At the same time, an axial field triggers harmonic axial oscillations with frequencies that depend on m/z . As a consequence, m/z differences can be indirectly obtained by measuring such frequency: the outer electrodes are used as receiver plates for image current detection of the oscillations, and the obtained current is Fourier-transformed to get a frequency-domain spectrum. The conversion of frequency to m/z ensues. In lipidomics, Orbitrap mass spectrometers have been widely employed [31–33].

Triple quadrupoles (QQQ) are characterised by low mass resolution, and yet they are broadly used in lipidomics research, owing to their abilities to carry out targeted scans where reliable quantitative information can be achieved [34]. These instruments are based on the linear quadrupole, which consists of four parallel metal rods through which dc and rf potentials are applied to produce an oscillating field that achieves mass separation by the oscillatory motions: ions of a fixed m/z pass through the quadrupole rods only with specific values of the dc and rf potentials. A mass spectrum is therefore obtained by changing such potentials over time, while keeping their ratio constant. QQQ instruments employ three quadrupoles, arranged sequentially: while the first (Q1) and third (Q3) can be operated in the mass-selective mode using both dc and rf potentials, the second one (Q2) is operated with only the rf potential to allow all ions above a cut-off m/z to pass through. Q2 is used as a gas collision cell where ion-molecule collisions induce ion fragmentations that can provide chemical information. QQQ instruments can therefore be used in a number of different tandem experiments performed in space:

Product ion scanning: Precursor ions with particular m/z values are selectively transmitted by Q1, fragmented in Q2, and their product ions are analysed by Q3;

Precursor ion scanning (PIS): Q3 is set to transmit only ions of a selected m/z while Q1 scans the precursor ions;

Neutral loss scanning (NL): Q1 and Q3 are scanned together keeping a fixed m/z difference between them;

Multiple reaction monitoring (MRM): Q1 and Q3 are both set to transmit a selected precursor and product ion, respectively. This provides great performances in quantitative analyses.

Untargeted and Targeted Lipidomics

MS-based lipidomics can be either untargeted or targeted: the aim of the former is to detect as many lipids as possible without any prior information, whereas targeted approaches focus on specific lipid classes by exploiting their fragmentation patterns. Targeted lipidomics makes use of tandem-based (MS/MS) experiments to analyse lipids with structural similarities, in order to address specific biological questions. Ion fragmentation is usually achieved by collision-induced dissociation (CID), where the selected precursor ions are first accelerated through an electric potential drop, and then collided with inert gas molecules (either He or N₂), producing fragment ions containing chemical information. Table 1.2 reports some of the more

commonly exploited MS/MS modes in lipidomics, which are very useful for the detection of only the species of interest.

LIPID CLASS	POLARITY	SCAN MODE
PC	+	PIS of m/z 184
PE	+	NL of m/z 141
PS	-	NL of m/z 87
PI	+	NL of m/z 277
PG	+	NL of m/z 189
PA	-	PIS of m/z 153
SM	+	PIS of m/z 184
Ceramides & Hexosyl-Ceramides	+	PIS of m/z 264

TABLE 1.2: Common MS/MS scanning modes in lipidomics.

Despite the ever improving performances of MS instrumentation, there exist analytical challenges that MS is not able to address alone, such as the assignment of double bond configurations [35]. It is for this reason that analytical techniques such as NMR, although not much popular in lipidomics research, should not be overlooked, and in the following section I will describe my brief experience with NMR-based lipidomics.

1.4 NMR-based Lipidomics

NMR spectroscopy is a powerful analytical technique that has been used in many and disparate applications. When compared to MS, NMR boasts a series of advantages:

- ✓ It is intrinsically quantitative, since NMR resonances have areas that are directly proportional to the molar amount of nuclei (^1H , ^{13}C , ...) giving rise to them;
- ✓ The sample preparation step is simple and rapid;
- ✓ Unlike MS, there is no need to extensively optimise instrumental parameters;
- ✓ NMR is non-destructive, hence the sample can be completely recovered after the analysis.

Unfortunately, NMR also has some disadvantages:

- ✗ It has a lower sensitivity when compared to MS;
- ✗ NMR spectra are usually dominated by abundant lipids such as PC and SM, and therefore resonances of low abundant lipids are difficult to identify;
- ✗ Owing to the complexity of biological mixtures, many overlapping resonances may be present;
- ✗ NMR cannot be easily coupled with LC.

Worth of note, the nuclei commonly accessible to NMR (^1H , ^{13}C and ^{31}P) are also important in lipid chemistry. Interesting is the case of phospholipids: since each class only contains a single P atom, the analysis can be easily performed by monitoring the ^{31}P resonances. Crucially, different phospholipid classes can be identified since the chemical environment around ^{31}P depends on the class itself [36].

Figure 1.4 is an ^1H -NMR spectrum of a raw lipid extract (recorded in our labs), and I have reported it just to give an idea of how complex spectra of lipid mixtures can be. To note that also the solvent (methanol- d_4 here) has its own resonances (3.31, 3.35 and 4.8 ppm) that contribute to the final complexity, at the same time masking underlying sample resonances.

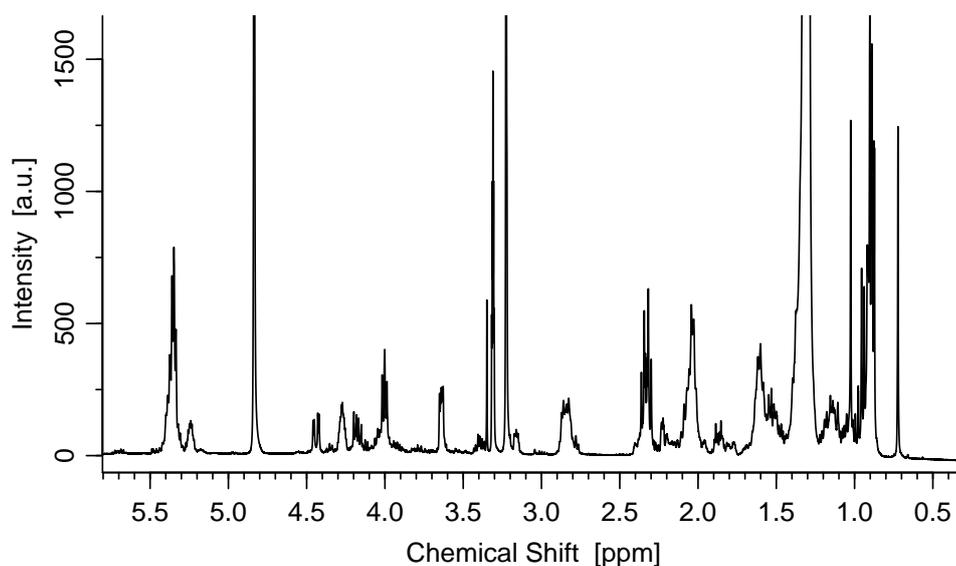


FIGURE 1.4: Example of ^1H -NMR spectrum of a raw lipid extract. It was recorded in deuterated methanol and with a 400 MHz NMR spectrometer.

The signals common to all lipid categories are those related to the FA chains, mainly:

- 0.88 ppm: terminal $-\text{CH}_3$;
- 0.94 ppm: terminal $-\text{CH}_3$ of $\omega 3$ -FAs;
- 1.30 ppm (strong signals): alkyl protons ($-\text{CH}_2-$);
- 1.55 ppm: protons on β position to carbonyl carbon ($\text{CO}-\text{CH}_2-\text{CH}_2-$);
- 2.05 ppm: allylic protons ($-\text{CH}=\text{CH}-\text{CH}_2-$);
- 2.35 ppm: protons on α position to carbonyl carbon ($\text{CO}-\text{CH}_2-$);
- 2.80 ppm: bis-allylic protons ($=\text{CH}-\text{CH}_2-\text{CH}=\text{}$);
- 5.35 ppm: olefinic protons ($-\text{HC}=\text{CH}-$).

These resonances only allow to get general information such as FA chain unsaturation. More detailed information can be obtained by class-specific

resonances which, for the lipid classes more relevant in the present thesis, include¹:

- 0.72(s) ppm: $-\text{CH}_3$ of sterols;
- 3.21(s) ppm: protons of PC and SM headgroups ($-\text{N}-(\text{CH}_3)_3$);
- 4.3(q) and 6.0(dt) ppm: plasmenyl-species ($-\text{CH}=\text{CH}-\text{O}-$);
- 4.04(m) and 3.16(m) ppm: PE headgroup ($-\text{O}-\text{CH}_2-\text{CH}_2-\text{NH}_3$);
- 4.13(m), 4.32(m) and 3.78(dd) ppm:
PS headgroup ($-\text{O}-\text{CH}_2-\text{CH}(\text{NH}_3^+)-\text{COO}^-$);
- 3.85(m), 3.91(m), 3.77(m), 3.62(dd) and 3.54(dd) ppm:
PG headgroup ($-\text{O}-\text{CH}_2-\text{CH}(\text{OH})-\text{CH}_2-\text{OH}$);
- 3.20(dt), 3.38(dd), 3.63(t), 3.76(t), 3.88(dt), 4.18 ppm: PI headgroup.

Based on these premisses, during the very first part of my PhD I tried to figure out whether it was possible to devise an informatics tool to automatically process NMR spectra of lipid mixtures. The aim was to quickly get insights into their composition, without the extensive manual interpretation usually required. Actually, many were the issues that emerged, to the point that I did not succeed in fully implementing such a practical tool. In Appendix A is a more extensive explanation, where I also point out some of the main open challenges.

1.5 Bioinformatics in MS-based Lipidomics

In lipidomics experiments, samples are usually divided into different groups. After the measurements, they have to be compared in order to look for possible group differences: a typical case is to have control and treated samples, these latter having been obtained through the introduction of some kind of treatment or physiological perturbation. However, before data analysis a pre-processing is usually required to get coherent data amenable to comparison; in MS-based experiments this data has to be a single peak table containing common detected features (m/z values) with related amounts (peak intensities or areas), which can be obtained following some crucial steps [37, 38]:

Peak detection and quantification: For each sample the signals are detected and a list of peaks is produced, where the peaks are characterised by m/z and retention time. Peak quantification is achieved by either integration of chromatographic profiles, or measurement of peak intensities;

Peak grouping: Peaks within specified m/z and retention time distances of each other are searched across sample groups, and assumed to represent the same analyte. Time alignments are usually required to correct for retention time shifts;

¹The provided values refer to samples dissolved in deuterated methanol, CD_3OD .

Filling of missing peaks: If some samples miss one or more peaks, they are inserted by assuming to be located at the same position as the identified peaks in related samples. This procedure avoids missing values, problematic for statistical analysis;

Isotope correction: Since each analyte shows more peaks (corresponding to its isotopic pattern), only one of them is kept, usually the base peak;

Compound identification: Compound names are assigned to the peaks, which is achieved by searching appropriate databases.

The pre-processed data can then be used for statistical analysis, and chemometrics methods such as the Principal Component Analysis (PCA) are now popular. The introduction of bioinformatics is a crucial step for solving the challenges related to handling big data sets, but these tools are still emerging in terms of creation of databases and statistical techniques for the analysis and interpretation of lipidomics data. An added issue is the absence of a universal lipid classification scheme, which has led to the birth of different lipid-oriented databases, including LipidBank, LIPIDAT, Lipid Library, Cyberlipids and LIPID Metabolites and Pathway Strategy (LIPID MAPS).

There are many software tools to handle MS-based data: commercial solutions include MarkerLynx (Waters), Lipid Profiler (MDS Sciex), Metabolic Profiler (Bruker BioSpin) and MarkerView (Applied Biosystems), whereas among open-source tools are MZmine, *xcms*, LIMSA, LipidNavigator and TriglyAPCI. During my PhD I mainly worked with *xcms*, a freely available R package [39] developed by Smith and co-workers [40]: it implements useful tools such as a non-linear retention time alignment, matched filtration, peak detection and peak matching. In addition, it is very flexible since it allows data processing coming from several mass spectrometers, the requirement being that the raw data files are exported in the *.NetCDF, *.mzXML or *.mzData formats. If they cannot be directly exported through the proprietary software, this can be achieved with the freely available MSConvert tool (from ProteoWizard).

1.6 Lipidomics Applications

Since its birth, lipidomics has witnessed an increasing number of applications, and here I have summarised just some of them.

Human Diseases and Disease Diagnosis

One of the most important contributions of lipidomics has been in disease diagnosis: lipid homeostasis is essential for health, and lipid alterations have been proven to be central to the pathogenesis of neurodegenerative disorders, metabolic syndromes, cancers and infectious diseases [13, 41]. Altered sulfatide metabolism is an early feature of Alzheimer's disease [42], significant accumulations of TAGs and altered myocardial eicosanoid metabolism have been identified in diabetes [43], and hyperlipidemia is one of the factors related to the occurrence of metabolic syndromes [13]. Lipids are also associated with cancer progression and many of them have been explored as potential biomarkers, such as phospholipids in ovarian, breast and prostate cancers [44–46].

Drug Development

Lipidomics has also found applications in drug development, since in many cases lipid functions are targeted through therapeutic interventions [47]. Accordingly, an increasing number of drugs have been introduced that target lipid metabolism or enzymes involved in the synthesis of cholesterol, fatty acids, phospholipids, sphingolipids and triglycerides [48].

Oxidative Lipidomics

The study and characterisation of oxidised lipids is a novel lipidomics application, made possible by the increasing performances of MS platforms. This new field is particularly challenging, the reason being that oxidised lipids are usually present in very low amounts as compared to their non-oxidised counterparts. In addition, high MS resolution is needed to unequivocally assign molecular formulae. Through oxidative lipidomics it has been found that apoptosis generates oxidised phospholipids, with selective oxidation of CL in the mitochondria and of PS outside them [49]. Also traumatic brain injury was shown to produce oxidised lipids [50].

1.7 Critical Issues

Lipidomics is a promising and powerful discipline, but it has still to face important issues. Below I will point out just some of them, in particular the ones that I found to be more critical during my experience in this field.

Quantitative Lipidomics

In lipidomics, it is very common to have samples sorted in groups, and therefore it is important to apply some normalisation procedure to the data obtained from MS-based measurements, in order to properly take into account the different starting biomasses of the several samples. A possible approach is to normalise the signal intensities or areas within each sample to its total ion current (TIC), but such strategy makes sense only if relative lipid amounts have to be compared; on the contrary, it is of no use when absolute lipid amounts have to be measured. In this more complex case, a general consensus on the protocol to adopt seems not have been achieved yet, the main problem being that in ESI-MS different lipid classes have different response factors. To further worsen things, even FA chain lengths and saturations affect the response factors [51]. Therefore, a rigorous quantification would require internal standards tailored to the analytes of interest: ideally, for each species to quantify there should be added its own standard, represented by the same species but enriched in one or more stable isotopes, in order to have a different m/z but same retention time and response factor. Such rigorous approach is clearly not practical, and a possible alternative is to use external calibration, whereby the standard solutions are not mixed with the experimental samples, but analysed in dedicated runs. A further approach, that we have employed in Chapter 3, is not to convert the signals (areas or intensities) into molar amounts, but directly use them during the statistical analysis, with the precaution to carry out a proper normalisation procedure before doing this.

Another issue regarding quantitative lipidomics comes from fluctuations in extraction yield across samples. Fortunately enough, this problem can be easily overcome by using proper internal standards, added to the samples prior to the extraction, so as to undergo the same extraction steps as the sample analytes. The standard has to be carefully chosen because, obviously, it should not be present in the original samples; for this reason, lipids with odd-numbered or very short FA chains are commonly employed [52]. In our investigations, when required we used the PC species PC 24:0, whose chains are much shorter than the ones commonly found in biological lipid extracts.

Studying the Whole Lipidome

An important goal which is difficult to achieve is to completely characterise the lipidome of biological samples, and there are two main reasons: the first one, related to MS-based detection, has to do with the range of physical and chemical properties of lipids, which is so huge that it is impossible to get a picture of the entire lipidome based on a single MS-based experiment. In addition, as mentioned above different lipid classes have very different MS response factors, and therefore analytes that are difficult to ionise are difficult to detect, too.

The second reason why it is difficult to analyse all lipids within a sample is related to the extraction protocol: again, considering the very different physical properties of different lipid categories, a number of extraction protocols have been devised, each tailored to a particular group of lipids. In other words, there currently is no universal procedure allowing to recover all of the lipid components in a single organic phase, even though the Folch as well as the Bligh & Dyer approaches, which we have broadly used in our investigations, are reckoned to be quite suitable for most lipid classes.

Biological Variability

Another important lipidomics issue that showed up quite often during our investigations is associated to the biological variability: similarly to metabolites, lipids vary between individuals and on a day-to-day basis, with the result of further complicating comparative studies [53–55]. As a consequence, it is usually rather difficult to pinpoint significant differences across sample groups because, even if the treatment under investigation causes a true lipidomics perturbation, this could be masked by the lipid natural variability. The only (but often impractical) solution would be to greatly increase the number of samples per each group [56]. On the other hand, it is worth pointing out that the biological variability is an intrinsic issue, that arises also in all the other “omics” sciences.

1.8 Aims of the Thesis

From what introduced above, it is clear that the field of lipidomics is quite promising, and an increasingly number of scientists all around the world are starting to practice this scientific discipline.

As a PhD student, I was introduced to this field by my tutor, and during the past three years I have worked on three different topics, which are detailed in the coming chapters. Although they are all connected by lipidomics,

these projects are quite different from one another, but I believe that this is a mere reflection of the many facets that lipidomics has. As a consequence, the following are stand-alone chapters and can be read independently of one another.

As already anticipated in the Thesis Organisation Section (page iv), the aim of Chapter 2 was to assess and characterise possible sphingolipidomics perturbations related to pathological conditions. In particular, we were interested in Parkinson's Disease, considering that it is one of the most common neurodegenerative diseases, but unfortunately still poorly understood. To this aim, we worked on both mouse cortical neurons and mouse brain tissues, and we carried out a "targeted lipidomics" in order to focus on sphingolipids; this type of lipidomics exploits the fact that some lipid classes (in our case, sphingolipids) undergo specific fragmentation mechanisms upon collision in mass spectrometry-based investigations, and therefore they can be easily singled out from all the other lipids.

The aim of Chapter 3 was to understand whether anti-angiogenic treatments of cancer cells have some lipidomics repercussions. In terms of methodology, this project was completely different from the one in Chapter 2, because here an "untargeted lipidomics" approach had to be carried out: we were not interested in some particular lipid class, but instead we tried to look at as many lipids as possible at once, and we attempted to identify changes in terms of lipid composition, occurring because of the treatment.

Chapter 4 deals with yet another aspect of lipidomics: the bioinformatics analysis of MS-based data obtained from labelling experiments. Indeed, in biological experiments stable isotope tracers are more and more commonly employed, giving rise to lipids where the labels have been partially taken up. The main aim of this Chapter was to develop an informatics tool able to handle MS-based data of labelled analytes, providing information about the amount of labelling within each investigated chemical species. The details will be given in the following, but the principle behind it is that such information can be retrieved by analysing the MS isotopic patterns.

Bibliography of Chapter 1

1. Fahy, E. *et al.* A comprehensive classification system for lipids. *Journal of Lipid Research* **46**, 839–861 (2005).
2. Orešič, M., Hänninen, V. A. & Vidal-Puig, A. Lipidomics: a new window to biomedical frontiers. *Trends in Biotechnology* **26**, 647–652 (2008).
3. Aarsland, A. & Wolfe, R. R. Hepatic secretion of VLDL fatty acids during stimulated lipogenesis in men. *Journal of Lipid Research* **39**, 1280–1286 (1998).
4. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* **9**, 112–124 (2008).
5. Gault, C. R., Obeid, L. M. & Hannun, Y. A. An overview of sphingolipid metabolism: From synthesis to breakdown. *Advances in Experimental Medicine and Biology* **688**, 1–23 (2010).
6. Pettus, B. J., Chalfant, C. E. & Hannun, Y. Sphingolipids in inflammation: roles and implications. *Current Molecular Medicine* **4**, 405–418 (2004).
7. Kee, T. H., Vit, P. & Melendez, A. J. Sphingosine kinase signalling in immune cells. *Clinical and Experimental Pharmacology & Physiology* **32**, 153–61 (2005).
8. Dhaka, V., Gulia, N., Ahlawat, K. S. & Khatkar, B. S. Trans fats—sources, health risks and alternative approach – A review. *Journal of Food Science and Technology* **48**, 534–541 (2011).
9. Wenk, M. R. The emerging field of lipidomics. *Nature Reviews Drug Discovery* **4**, 594–610 (2005).
10. Navas-Iglesias, N., Carrasco-Pancorbo, A. & Cuadros-Rodríguez, L. From lipid analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part II: analytical lipidomics. *Trends in Analytical Chemistry* **28**, 393–403 (2009).
11. Hu, C. *et al.* Analytical strategies in lipidomics and applications in disease biomarker discovery. *Journal of Chromatography B* **877**, 2836–2846 (2009).
12. Hyötyläinen, T., Bondia-Pons, I. & Orešič, M. Lipidomics in nutrition and food research. *Molecular Nutrition & Food Research* **57**, 1306–1318 (2013).
13. Lam, S. M. & Shui, G. Lipidomics as a principal tool for advancing biomedical research. *Journal of Genetics and Genomics* **40**, 375–390 (2013).
14. Folch, J., Lees, M. & Stanley, G. H. S. A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry* **226**, 497–509 (1957).

15. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911–917 (1959).
16. Lin, J. HPLC separation of acyl lipid classes. *Journal of Liquid Chromatography & Related Technologies* **30**, 2005–2020 (2007).
17. Li, M., Zhou, Z., Nie, H., Bai, Y. & Liu, H. Recent advances of chromatography and mass spectrometry in lipidomics. *Analytical and Bioanalytical Chemistry* **399**, 243–249 (2011).
18. Houjou, T., Yamatani, K., Imagawa, M., Shimizu, T. & Taguchi, R. A shotgun tandem mass spectrometric analysis of phospholipids with normal-phase and/or reverse-phase liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **19**, 654–666 (2005).
19. Willmann, J., Leibfritz, D. & Thiele, H. Hyphenated tools for phospholipidomics. *Journal of Biomolecular Techniques* **19**, 211–216 (2008).
20. Köfeler, H. C., Fauland, A., Rechberger, G. N. & Trötz Müller, M. Mass spectrometry based lipidomics: an overview of technological platforms. *Metabolites* **2**, 19–38 (2012).
21. Milne, S., Ivanova, P., Forrester, J. & Alex Brown, H. Lipidomics: an analysis of cellular lipids by ESI-MS. *Methods* **39**, 92–103 (2006).
22. Watson, J. T. & Sparkman, O. D. *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation* (Wiley, 2007).
23. Han, X., Yang, K. & Gross, R. W. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomics analyses. *Mass Spectrometry Reviews* **31**, 134–178 (2012).
24. Ståhlman, M. *et al.* High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry. *Journal of Chromatography B* **877**, 2664–2672 (2009).
25. Yang, K., Zhao, Z., Gross, R. W. & Han, X. Systematic analysis of choline-containing phospholipids using multi-dimensional mass spectrometry-based shotgun lipidomics. *Journal of Chromatography B* **877**, 2924–2936 (2009).
26. Han, X. L. & Gross, R. W. Shotgun lipidomics: multidimensional MS analysis of cellular lipidomes. *Expert Review of Proteomics* **2**, 253–264 (2005).
27. Awad, H., Khamis, M. M. & El-Aneed, A. Mass spectrometry, review of the basics: ionization. *Applied Spectroscopy Reviews* **50**, 158–175 (2015).
28. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
29. Bruins, A. P. Mechanistic aspects of electrospray ionization. *Journal of Chromatography A* **794**, 345–357 (1998).
30. Schiller, J. *et al.* Matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry in lipid and phospholipid research. *Progress in Lipid Research* **43**, 449–488 (2004).

31. Zubarev, R. A. & Makarov, A. Orbitrap mass spectrometry. *Analytical Chemistry* **85**, 5288–5296 (2013).
32. Perry, R. H., Cooks, R. G. & Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrometry Reviews* **27**, 661–699 (2008).
33. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **40**, 430–443 (2005).
34. Khalil, M. B. *et al.* Lipidomics era: accomplishments and challenges. *Mass Spectrometry Reviews* **29**, 877–929 (2010).
35. Thomas, M. C., Mitchell, T. W. & Blanksby, S. J. OnLine ozonolysis methods for the determination of double bond position in unsaturated lipids. *Methods in Molecular Biology* **579**, 413–441 (2009).
36. Yu, Y., Vidalino, L., Anesi, A., Macchi, P. & Guella, G. A lipidomics investigation of the induced hypoxia stress on HeLa cells by using MS and NMR techniques. *Molecular BioSystems* **10**, 878–890 (2014).
37. Castillo, S., Gopalacharyulu, P., Yetukuri, L. & Orešič, M. Algorithms and tools for the preprocessing of LC–MS metabolomics data. *Chemo-metrics and Intelligent Laboratory Systems* **108**, 23–32 (2011).
38. Wei, X. *et al.* Data preprocessing method for liquid chromatography–mass spectrometry based metabolomics. *Analytical Chemistry* **84**, 7963–7971 (2012).
39. R Development Core Team. *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
40. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using non-linear peak alignment, matching, and identification. *Analytical Chemistry* **78**, 779–787 (2006).
41. Postle, A. D. Phospholipid lipidomics in health and disease. *European Journal of Lipid Science and Technology* **111**, 2–13 (2009).
42. Han, X. Potential mechanisms contributing to sulfatide depletion at the earliest clinically recognizable stage of Alzheimer’s disease: a tale of shotgun lipidomics. *Journal of Neurochemistry* **103**, 171–179 (2007).
43. Bayeva, M., Sawicki, K. T. & Ardehali, H. Taking diabetes to heart - Deregulation of myocardial lipid metabolism in diabetic cardiomyopathy. *Journal of the American Heart Association* **2**, 1–17 (2013).
44. Sutphen, R. *et al.* Lysophospholipids are potential biomarkers of ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention* **13**, 1185–1191 (2004).
45. Min, H. K., Lim, S., Chung, B. C. & Moon, M. H. Shotgun lipidomics for candidate biomarkers of urinary phospholipids in prostate cancer. *Analytical and Bioanalytical Chemistry* **399**, 823–830 (2011).
46. Dória, M. L. *et al.* Lipidomic analysis of phospholipids from human mammary epithelial and breast cancer cell lines. *Journal of Cellular Physiology* **228**, 457–468 (2013).

47. Wenk, M. R. Lipidomics: new tools and applications. *Cell* **143**, 888–895 (2010).
48. Dehairs, J., Derua, R., Rueda-Rincon, N. & Swinnen, J. V. Lipidomics in drug development. *Drug Discovery Today: Technologies* **13**, 33–38 (2015).
49. Tyurin, V. A. *et al.* in *Free Radicals and Antioxidant Protocols* 353–374 (2010).
50. Sparvero, L. J. *et al.* Mass-spectrometry based oxidative lipidomics and lipid imaging: applications in traumatic brain injury. *Journal of Neurochemistry* **115**, 1322–1336 (2010).
51. Koivusalo, M., Haimi, P., Heikinheimo, L., Kostinen, R. & Somerharju, P. Quantitative determination of phospholipid compositions by ESI-MS: effects of acyl chain length, unsaturation, and lipid concentration on instrument response. *Journal of Lipid Research* **42**, 663–672 (2001).
52. Brügger, B. Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry. *Annual Review of Biochemistry* **83**, 79–98 (2014).
53. Pereira, M. A. *et al.* Within-person variation in serum lipids: implications for clinical trials. *International Journal of Epidemiology* **33**, 534–541 (2004).
54. Tsalamandris, C. *et al.* Long-term intraindividual variability of serum lipids in patients with type I and type II diabetes. *Journal of Diabetes and its Complications* **12**, 208–214 (1998).
55. Tolfrey, K., Campbell, I. G. & Jones, A. M. Intra-individual variation of plasma lipids and lipoproteins in prepubescent children. *European Journal of Applied Physiology and Occupational Physiology* **79**, 449–56 (1999).
56. Xiao, Q. *et al.* Sources of variability in metabolite measurements from urinary samples. *PLoS ONE* **9**, e95749 (2014).

Chapter 2

A Sphingolipidomics Investigation on the Association Between GBA and LRRK2

2.1 Introduction

Parkinson's Disease

Parkinson's disease (PD) is a degenerative disorder characterised by the death of dopaminergic neurons in a part of the midbrain called *substantia nigra*. This results in low levels of dopamine, a chemical that spreads messages between the *substantia nigra* and other parts of the brain, making humans able to coordinate muscle movements [1, 2]. As a consequence, the earlier PD symptoms are movement-related (e.g., shaking and difficulty with walking), but there also follow behavioural problems and, in advanced stages, dementia. PD is more common in older people, and the majority of cases take place after the age of 50 [3, 4].

The pathological hallmarks of PD are the so-called Lewy bodies (LBs), neuronal inclusions where α -synuclein accumulates. This is a protein encoded by the synuclein alpha (*SNCA*) gene and its functions are not yet fully understood, even though it is known that it interacts with membranes composed of acidic lipids [5], and it can play roles in vesicular transport and facilitation of dopamine release [6]. α -synuclein oligomers are known to be toxic, since they bind to mitochondrial membranes and lead to their fragmentation followed by neuronal cell death [7, 8]. These oligomers form when α -synuclein misfolds, changing its conformation from α -helix-rich to β -sheet structures [9].

In the past, PD was thought to be solely caused by environmental causes, with epidemiological studies showing associations with factors such as exposure to chemicals and head trauma [10–12]. Later on, genetic studies also revealed heritable causes, and to date a number of mutations in several genes have been linked to PD, although monogenic causes constitute less than 5% of all PD cases, whereas sporadic PD is much more common [13, 14].

There exist some palliative treatments for PD, based on the use of levodopa and dopamine agonists. They can improve quality of life, but eventually they become ineffective. In addition, they produce a complication called dyskinesia, characterised by involuntary writhing movements [15].

Although a common theme for pathogenesis has not yet been identified, it seems that sphingolipid levels are often altered in neurodegenerative diseases (NDs) [16–19]. Ceramides in particular, components of all major sphingolipid species in the brain, were found to have a defective homeostasis in NDs with LBs pathology, including PD [20–22]. Ceramides play an important role in the regulation of plasma membrane dynamics, and in the modulation of membrane protein activity and signalling [23]; they also modulate processes such as autophagy [24] and mitochondrial-mediated apoptosis [25].

Two key genes involved in PD are β -glucocerebrosidase (*GBA*) and leucine-rich repeat kinase 2 (*LRRK2*): *GBA* mutations are the most common risk factor for sporadic PD, and comprise 7% of all sporadic PD cases [26], whereas *LRRK2* mutations are the most common in autosomal dominant forms of the disease.

GBA-associated Parkinsonism

The *GBA* gene encodes GBA, an enzyme abundant in the lysosomes of cells and responsible for the breakdown of glucosyl-ceramides (Glc-Cer) to glucose and Cer (Figure 2.1). Like many lysosomal enzymes, GBA is associated with storage disorders resulting from genetic alterations that ruin its conformation: so far, more than 300 *GBA* mutations have been identified and shown to be linked with Gaucher's disease (GD), one of the most common storage disorders characterised by the accumulation of Glc-Cer within lysosomes [27].

Epidemiological studies have unclosed an association between GD and PD: the former is a recessive disorder, meaning that heterologous *GBA* mutation carriers do not develop the disease; interestingly, they were found 5 to 20-fold more likely to develop PD than non-mutation carriers [22, 26, 28]. However, the relationship between GBA and PD is still unclear, and both gain-of-function and loss-of-function theories have been proposed: according to the former, mutant alleles produce a misfolded and dysfunctional protein, whereas the loss-of-function hypothesis suggests that PD arises as a consequence of GBA deficiency. Also the prion hypothesis has been proposed, where α -synuclein aggregates would have a prion-like mechanism of cell-to-cell transmission [9, 29].

As for the relationship between GBA and α -synuclein, a physical interaction between the two has been observed [30], with the hypothesis that GBA plays a beneficial role in α -synuclein degradation; this mechanism would be disrupted when GBA is mutated [26]. Worthy of note, also mitochondrial disfunctions and oxidative stress related to GBA deficiency have been reported [28].

LRRK2-associated Parkinsonism

LRRK2 is another key gene for PD, for two reasons: (i) some of its mutations are responsible for the late-onset dominant PD forms [31], and (ii) common variations around *LRRK2* are susceptibility factors for the disease [32–34]. The encoding protein is large and complex, with multiple domains [31]. As a consequence, just some of the *LRRK2* roles have been understood, such as its kinase and GTPase activities.

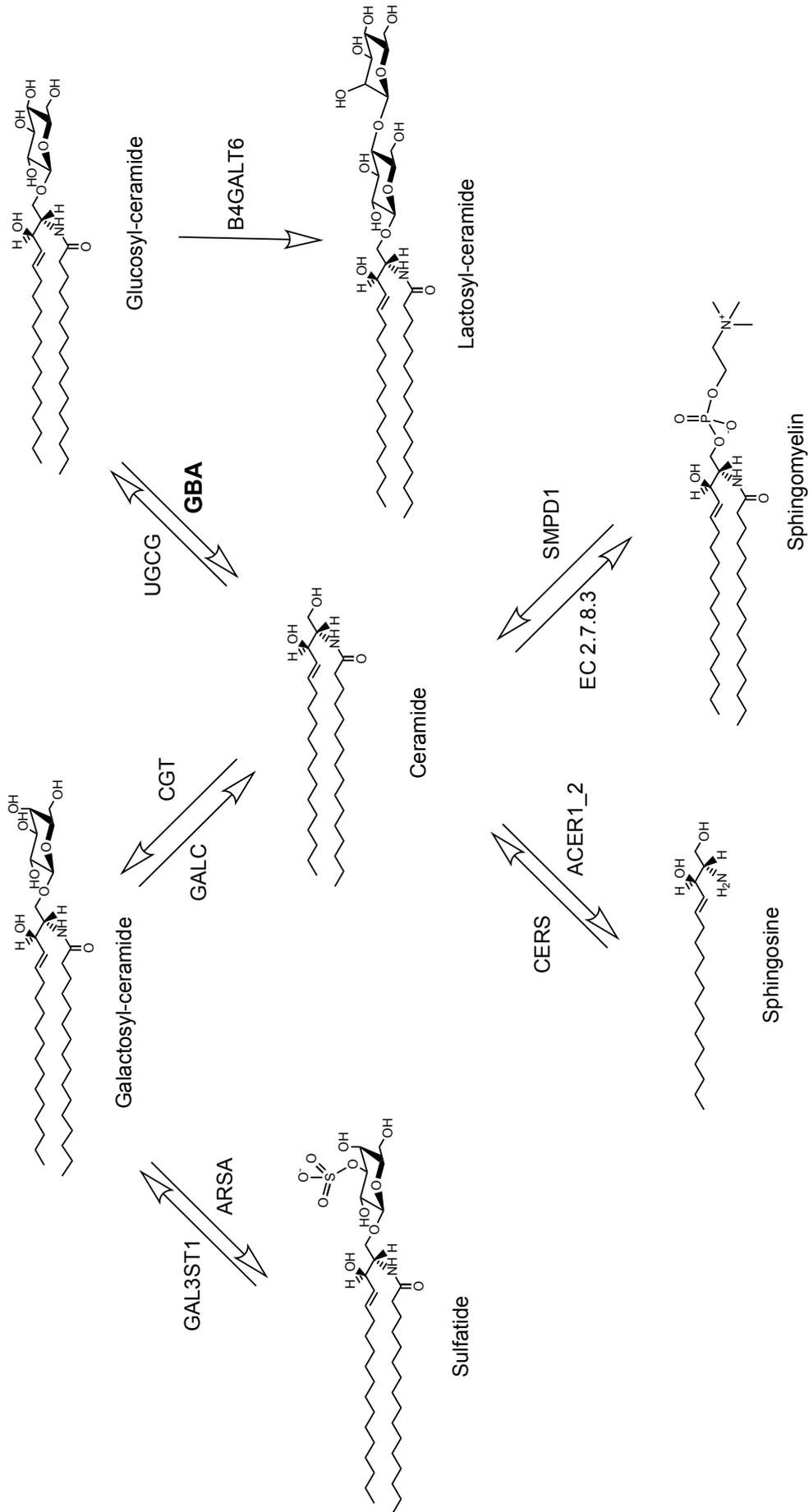


FIGURE 2.1: Simplified sphingolipid metabolic pathway. An important role is played by the GBA enzyme, responsible for the breakdown of Glc-Cer into Cer.

Many *LRRK2* gene mutations have been identified that replace single amino acids in *LRRK2*, affecting its structure and function. Of these, at least 20 mutations have been linked to PD, with the most common being G2019S [32, 35–37]. Of interest, *LRRK2* has been shown to be present in LBs together with α -synuclein, and to play a role in LB formation [38, 39]. In addition, it is known that *LRRK2* mutations impair chaperone-mediated autophagy (CMA), implying that cells accumulate CMA substrates such as α -synuclein [40].

Mouse studies revealed that *Lrrk2*^{-/-} mice have a normal dopaminergic system [41], although *Lrrk2*^{-/-} mouse lungs and kidneys display accumulation of membranous organelles such as lamellar bodies and lysosomes, respectively, in addition to α -synuclein and ubiquitinated proteins [42–46].

Considering the significance of sphingolipids in neurodegenerative diseases, and taking into account the important roles of both GBA and *LRRK2* in PD, we tried to look for possible sphingolipidomics alterations related to either GBA or *LRRK2*: in the first part of our investigation, we focused on GBA-associated alterations by working on cultures of primary cortical neurons, whereas in the second part we focused on *Lrrk2*^{-/-} mouse brains. Ours was a targeted LC-MS lipidomics approach, which we performed with a triple quadrupole (QQQ) mass spectrometer, operated in precursor ion scan (PIS) mode.

2.2 Primary Cortical Neurons

2.2.1 Materials and Methods

Sample Collection and Treatment

Neurons were collected from the cerebral cortex of wild type mice, by dissociating the cells through an enzymatic reaction employing papain [47]. The neuronal cell cultures were then grown in neurobasal medium with 10 % FBS, fungicides and antibiotics. The samples were divided in two groups: CTRL and CBE-TRTD. To the latter, after 3 days of culture the GBA inhibitor conduritol β -epoxide (CBE) was added at a concentration of 50 μ M, and the cells were grown for a further 14 days; the inhibitor was replaced every 3 days, whereas half of the culture medium was replaced once a week. After the treatment, the neurons were harvested and a small amount was used for assessing the GBA activity and the total protein content, obtained through a BCA assay. Overall, 6 CTRL and 6 CBE-TRTD samples were prepared, in three different experiments:

1. *Experiment A*: samples CTRL 1,2 and CBE-TRTD 1,2
2. *Experiment B*: samples CTRL 3,4 and CBE-TRTD 3,4
3. *Experiment C*: samples CTRL 5,6 and CBE-TRTD 5,6

In other words, the same experiment was repeated three times, on every occasion using 2 control and 2 treated samples.

In order to be clear about my role, it is a duty for me to point out that the above-mentioned steps were not performed by myself, but by our collaborators at the Department of Biochemistry, University of Padova; my involvement commenced from the next stage, the lipid extraction procedure.

Lipid Extraction

The collected neurons were lysed using 300 μ L milliQ water, and the lipids were extracted through a Folch extraction procedure [48]: the samples were first immersed in ice and 3 mL chloroform:methanol 2:1 was added, then they were sonicated for 15 min and centrifuged at 4 °C for 10 min to induce phase separation. The lower phase was finally withdrawn and stored at -80 °C until analysis.

NMR Measurements

Before subjecting the samples to the destructive MS-based analyses, their NMR spectra were recorded: the samples were dried using rotary evaporator, and dissolved in 700 μ L of methanol-d4 (CD₃OD, 99.8 % deuteration degree, purchased from Merck, Germany). The measurements were taken using a Bruker-Avance 400 MHz spectrometer operating with a stationary magnetic field of strength 9.4 T and equipped with a 5 mm BBI probe. The 90° proton pulse length was calibrated and established to be 9.4 μ s, with a transmission power of 0 dB. In all of the measurements, the temperature was kept at 300.2 K; the chemical shift scale was referenced to the residual protonated methanol-d4 resonance, set at 3.310 ppm.

LC-MS Measurements

The samples were dried again using a rotary evaporator, and dissolved in 150 μ L HPLC-grade methanol (Sigma-Aldrich Co.).

The LC-MS measurements were performed in positive ion mode using a Triple Quadrupole (QQQ) mass spectrometer (Applied Biosystems, API 3000 QQQ) equipped with an electrospray ion source (ESI), and combined with a Shimadzu High Performance LC system (CBM-20 A, binary pump LC-20AB, Italy) to have chromatographic separation prior to MS detection.

The analytes were chromatographically resolved working in reversed-phase (Kinetex C18 column, 100 Å pore size, 4.6 mm ID, 2.6 μ m particle size, 10 cm length, Phenomenex, Italy). The mobile phase was composed of solvent A, methanol:water (7:3 v/v) with 12 mM ammonium acetate, and solvent B, methanol with 12 mM ammonium acetate. The gradient elution program started with 70 % B, reached 100 % B in 45 min, and was maintained at 100 % B for 20 min, at the steady flow rate of 1.0 mL/min. The sample injection volume was 10 μ L, and each sample was injected twice.

Because of the targeted aim of the study, the mass spectrometer was operated in PIS mode: PIS of m/z 184 was used to detect phosphatidylcholines (PC) and sphingomyelins (SM), whereas PIS of m/z 264 allowed to detect the sphingolipids ceramides (Cer) and glycosyl-ceramides (Gly-Cer) [49], which include Glc-Cer and Gal-Cer. On reversed-phase, the two isomers cannot be chromatographically resolved, implying that only their overall amount can be determined [50].

In Figure 2.2 are the fragmentation mechanisms of some key lipid classes for our targeted investigation (mainly, Cer and related glycosylated forms, SMs and PCs). It is thanks to these peculiar patterns that such lipid species can be easily singled out and investigated through QQQ-based mass spectrometers.

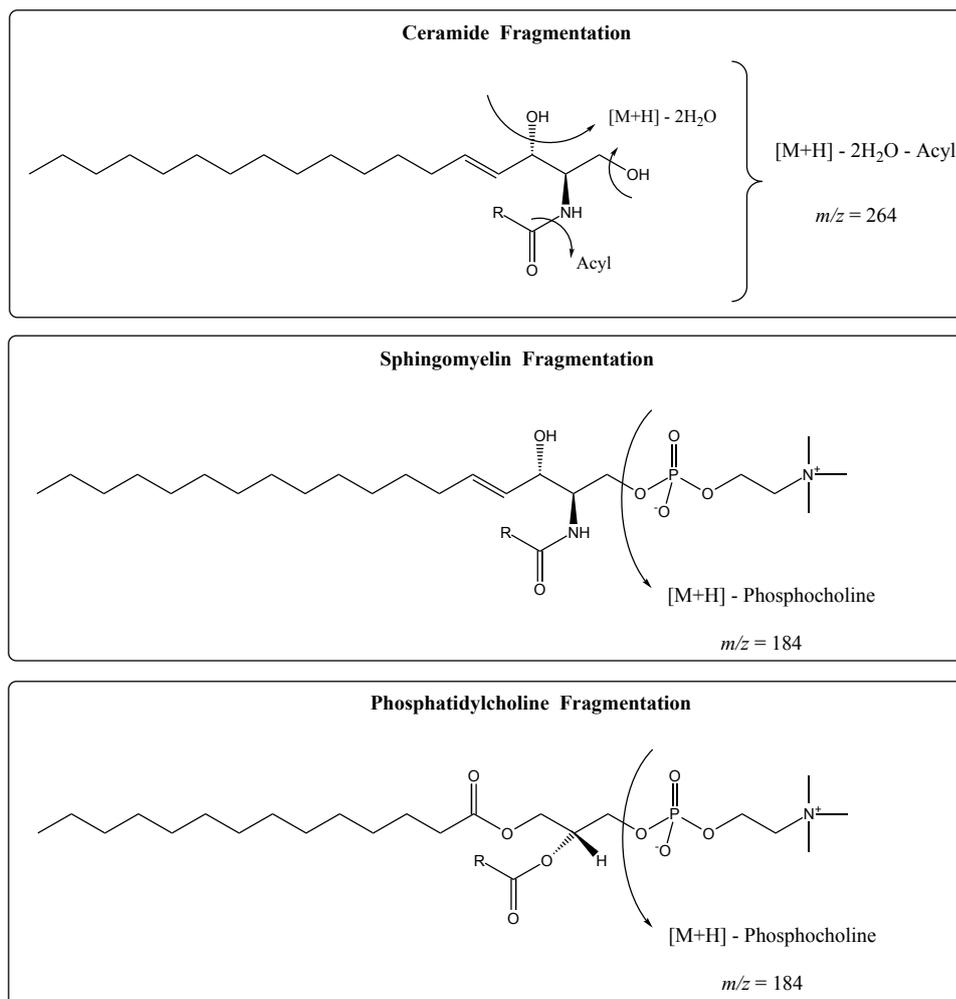


FIGURE 2.2: The main fragmentation patterns exploited in the current targeted MS-based investigation: Cer (as well as Gly-Cer, not shown here) give rise to positive ions characterised by m/z 264, whereas SMs and PCs can be identified through the loss of the phosphocholine headgroup (m/z 184).

To get quantitative information, an external calibration was performed whereby standard solutions were separately injected into the LC-MS system, in order to get the response factors required to convert peak areas into molar quantities. The standards were also used to optimise the instrumental parameters (for PIS of m/z 184, they were NEB=8, CUR=8, TEM=300, IS=5000, CAD=4, DP=50, FP=250, EP=10, CE=40, CXP=20; for PIS of m/z 264, they were NEB=8, CUR=8, TEM=300, IS=5000, CAD=4, DP=50, FP=250, EP=10, CE=50, CXP=20).

Data Analysis

The extracted-ion chromatograms (XIC) relative to the target sphingolipids were integrated with the proprietary software Analyst 1.4.2, and the obtained molar amounts were analysed with the R environment for statistical computing [51].

In order to directly compare sphingolipid levels across samples, their molar amounts were normalised to the total lipid content, set at the arbitrary value of 100. With this approach, percentage molar fractions are obtained.

2.2.2 Results and Discussion

NMR Profiling

The NMR profiles of the samples didn't allow us to identify any significant feature characteristic of either control or treated samples. Figure 2.3 shows the comparison between two of the spectra; although some differences may be noted between them, they are not treatment-specific but are the result of the biological variability in the samples.

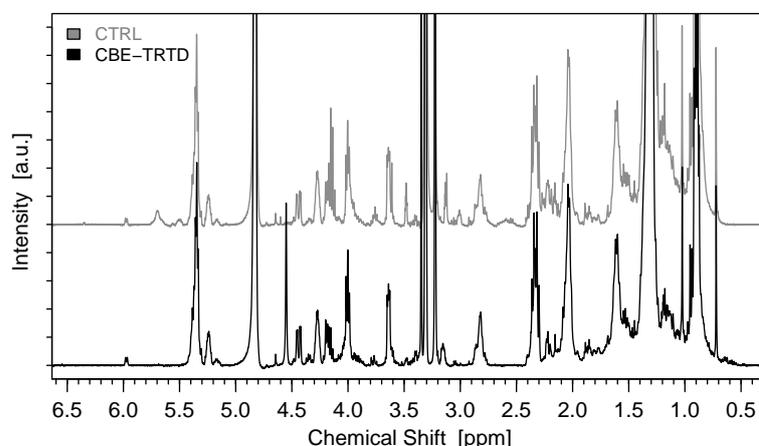


FIGURE 2.3: Comparison between the NMR spectra of CTRL and CBE-TRTD samples.

As I have already pointed out in Chapter 1, the issue with NMR spectra of raw lipid extracts is that they are dominated by the resonances of the most abundant species, and therefore minor components are hidden beneath them. In this case, the sphingolipids we are interested in have no characteristic signals falling in otherwise resonance-free regions. For this reason, we chose the more specific and targeted LC-MS approach based on the PIS scans available in QQQ mass spectrometers.

Relative Distribution Across Sphingolipids

We classified the sphingolipids into the three different groups Cer, Gly-Cer and SM. The sample comparison of the normalised molar amounts is in Figure 2.4, where the samples have been classified depending on the experiment (A, B or C).

In TRTD samples, a significant increase can be seen in Gly-Cer, accompanied by a decrease in SM; interestingly, these two classes are on opposite sides from GBA (see the scheme in Figure 2.1). Specifically, what this analysis shows is that when GBA is inhibited there is a build-up of its direct substrates, and a general depletion of the species downstream to this enzyme.

Next, in order to get a single p-value from these three combined experiments, we considered all 6 pairs of samples and performed a paired

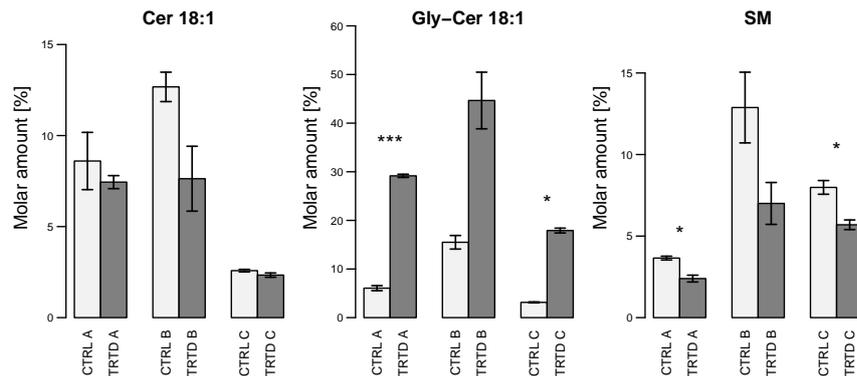


FIGURE 2.4: Distribution of sphingolipids across samples (data normalised to the total lipid content). Each value is the average from two samples, and the error bars are \pm one standard deviation. Significant differences between CTRL and CBE-TRTD can be seen for Gly-Cer and SM (*: $p \leq 0.05$; ***: $p \leq 0.001$).

t-test. This confirmed the increase in Gly-Cer (p-value: 6.4×10^{-4} ***) and the decrease in SM (p-value: 0.035 *).

In Figure 2.5 is the principal component analysis (PCA) we performed with the percentage molar amounts as original variables: interestingly, a dependence on both the experiment (A, B, or C) and the treatment (CTRL or CBE-TRTD) appears. As for the treatment, CBE-TRTD samples have lower PC1 and higher PC2 coordinates than controls. Considering the coordinates of Gly-Cer and SM in the loadings plot, this result is in fair agreement with the above observation that the treatment increases Gly-Cer and decreases SM relative molar amounts.

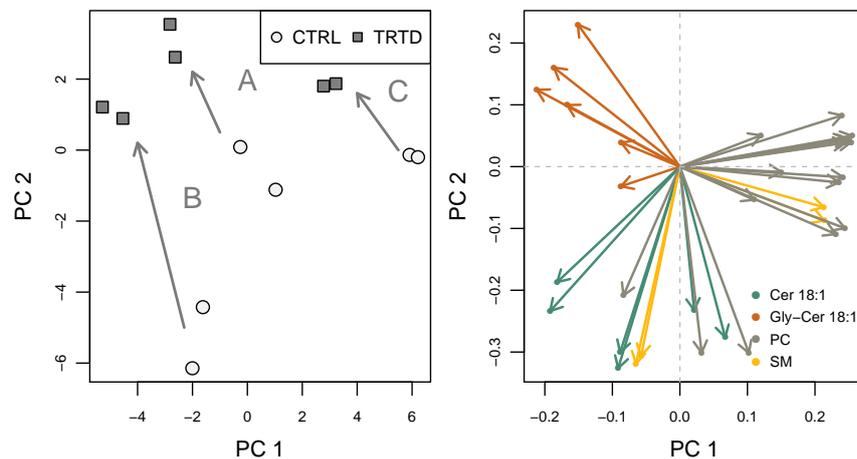


FIGURE 2.5: Scores (left) and loadings (right) plots from the PCA using normalised molar amounts as original variables. The named arrows to the left connect samples from the same experiment.

Gly-Cer/Cer Ratios

The GBA inhibition by CBE is expected to cause important and opposite effects in the molar amounts of its direct substrates (Glc-Cer) and products

(Cer) (see Figure 2.1). Consequently, the Gly-Cer/Cer ratio should be the ideal parameter allowing to look for possible sphingolipid perturbations.

These ratios, computed for each specific couple (*i.e.*, species with the same pair of chains), are in Figure 2.6. It is worth noting that, whereas there is a marked dependence on the treatment, the dependence on the experiment (A, B or C) disappears. The related box-plots (Figure 2.7) show that CBE-TRTD samples are associated with statistically higher ratios than CTRL samples, and this is true for all molecular couples ($p < 0.05$). In addition, we also computed the overall Gly-Cer/Cer ratio (Figure 2.8), which turned out to be significantly higher in CBE-TRTD samples ($p = 0.0016$ **).

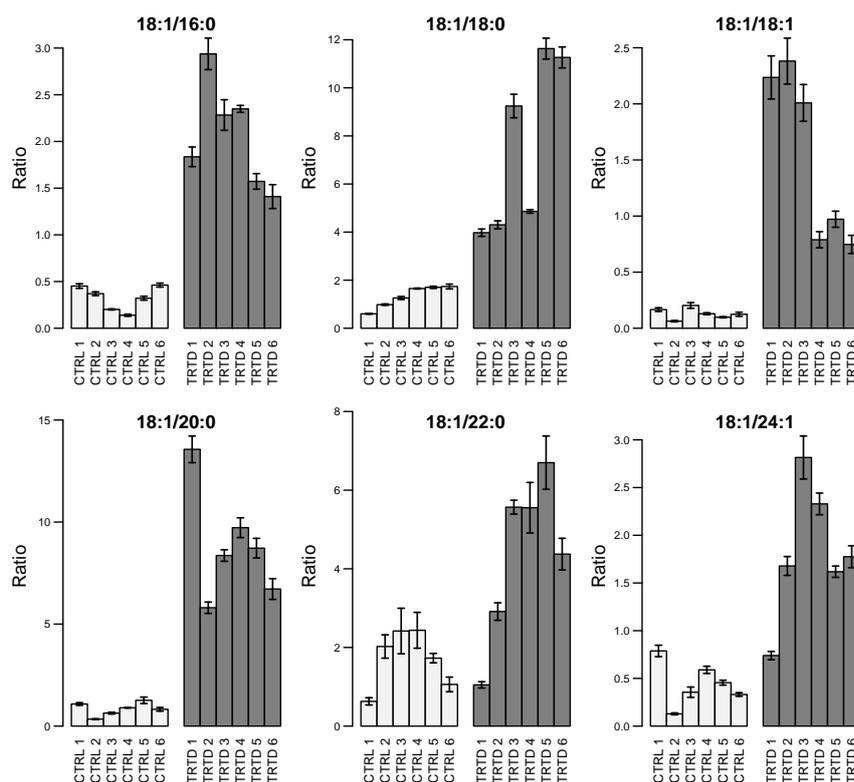


FIGURE 2.6: Individual Gly-Cer/Cer ratios, for CTRL and CBE-TRTD samples. The error bars represent the standard deviation of the technical replicates; the ratios are higher in CBE-TRTD samples.

Our conclusion is that GBA inhibition causes a significant imbalance in Cer and Gly-Cer relative amounts. Besides, taking into consideration the results discussed above, this imbalance appears to be mainly caused by the accumulation of the GBA substrates, Glc-Cer.

The approach of looking at the ratios rather than at the normalised sphingolipid amounts reduces the dependence on the experiment, and we further demonstrated this by running another PCA, this time using the individual Gly-Cer/Cer ratios as original variables (Figure 2.9). The two groups are mostly resolved along PC1, with CBE-TRTD samples being displaced to the right as compared to controls; not surprisingly, all the ratios have positive loadings along PC1.

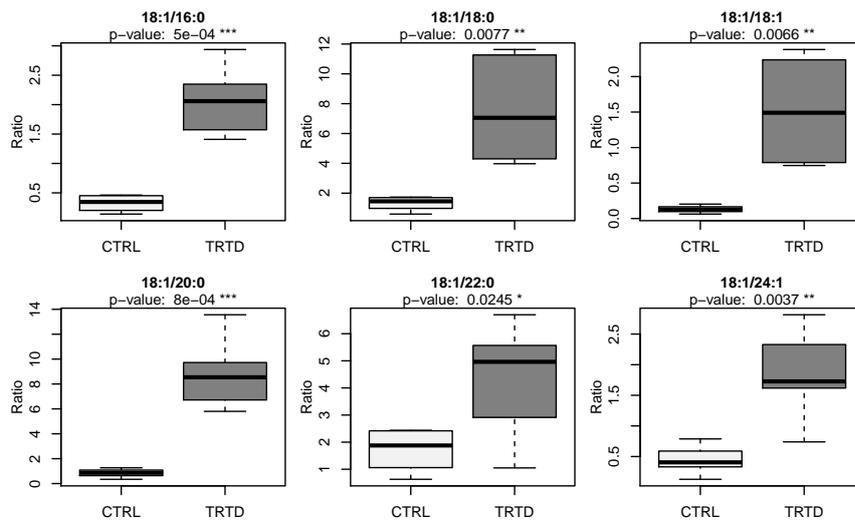


FIGURE 2.7: Box-plots of the individual Gly-Cer/Cer ratios; for all of them the difference between CTRL and CBE-TRTD samples is statistically significant, with the ratios being higher for CBE-TRTD samples.

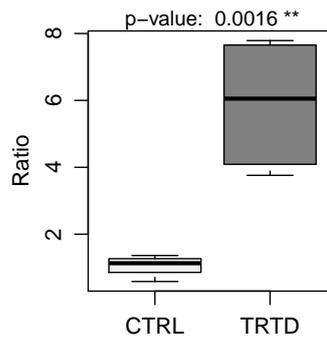


FIGURE 2.8: Overall Gly-Cer/Cer ratio; TRTD samples are associated with a statistically significant higher value.

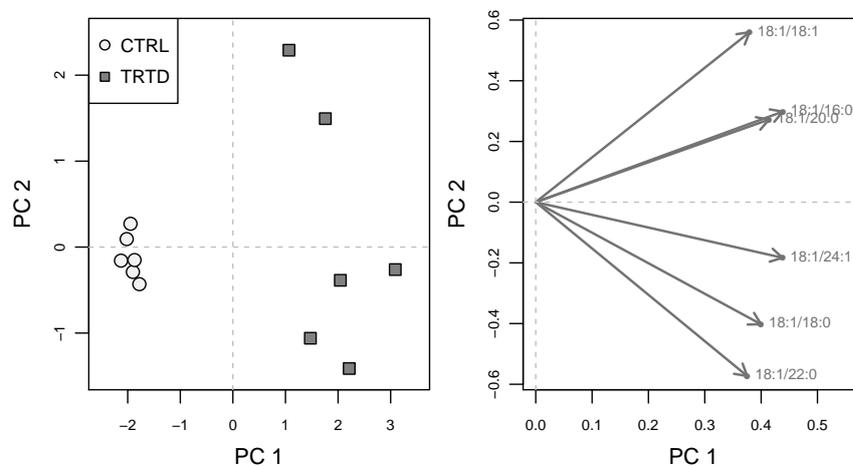


FIGURE 2.9: Scores (left) and loadings (right) plots from the PCA using Gly-Cer/Cer ratios as original variables. The two sample groups are well resolved along PC1, where the original variables have all positive loadings.

Intra-class Distribution

In order to understand whether the treatment is specific for certain acyl chains, we also made a comparison within each sphingolipid class, where we used relative molar amounts (we set the overall molar amount of the investigated class to 100). The resulting distributions are in Figure 2.10, which highlights no remarkable difference between CTRL and TRTD profiles: of the 16 t-tests we performed, only in one case was the p-value lower than the 5% significance level, which is not enough to state that the two profiles are different if we consider that to each t-test there corresponds a 5% probability of getting a significant result just by chance (type I error).

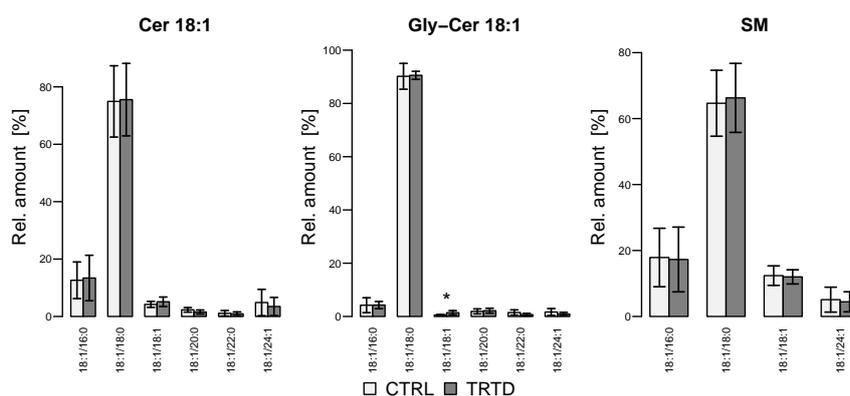


FIGURE 2.10: Comparison between CTRL and CBE-TRTD samples in terms of their intra-class distributions (Cer, Gly-Cer and SM). The profiles are not significantly different.

These results suggest that the overall imbalance observed between Cer and Gly-Cer is not chain-specific, but equally affects different molecular species. As a further confirmation, we computed and compared parameters such as average chain length, average number of unsaturations and unsaturation index¹ (see Appendix B.1), finding no significant difference between sample groups. After all it is worth mentioning that, to the best of our knowledge, there is no report in the literature suggesting substrate specificity for GBA.

2.2.3 Conclusions

With this first, preliminary survey we tried to set up a reliable analytical methodology to identify and characterise possible sphingolipid imbalances. As explained in the Introduction, we were motivated by the fact that sphingolipids seem to play crucial roles in NDs, and their imbalances might therefore be suggestive of ongoing diseased states. It would therefore be important to have an analytical methodology able to selectively monitor these species.

We set out by working with primary cortical neurons, and by inhibiting a key enzyme in the sphingolipid metabolism, GBA, which breaks down the glycosidic bond of Glc-Cer to obtain glucose and Cer (Figure 2.1). Not surprisingly, GBA mutations have been associated to both GD, a disease

¹We define unsaturation index as the average number of unsaturations per carbon atom.

characterised by the accumulation of Glc-Cer in the lysosomes of cells, and PD, where GBA is recognised as a risk factor for sporadic parkinsonism.

We focused on the sphingolipids Gly-Cer, Cer and SM, which we quantified using the PIS modes made available by QQQ instruments; we observed a significant increase, in CBE-TRTD samples, in terms of Gly-Cer molar amounts (Figure 2.4). We then moved on by computing and comparing individual Gly-Cer/Cer ratios (Figure 2.6 and Figure 2.7), finding significant variations upon treatment. The same trend we observed for the overall Gly-Cer/Cer ratio (Figure 2.8). This suggests that such parameter may be used to assess and monitor possible sphingolipid imbalances; in addition, this approach has the big advantage that by using ratios, no internal quantitative reference is required.

Considering the results discussed so far, in principle it should be possible to pinpoint such changes also from the recorded NMR spectra. Unfortunately though, we were not able to put this into practice, owing to (i) the complexity of the spectra themselves (Figure 2.3), coupled with (ii) the low molar abundance of the investigated sphingolipids as compared to the most common phospholipids. On the contrary, the targeted MS-based approach turned out to be more successful owing to its great specificity towards sphingolipids.

Taken together, our results suggest that when GBA activity is compromised, significant sphingolipid perturbations can be seen. In particular, GBA substrate accumulate, since they cannot be processed by this important enzyme. The observed imbalance turned out not to be chain-specific, considering that the intra-class profiles were indistinguishable between groups (Figure 2.10 and Appendix B.1). This means that all Cer and Glc-Cer couples are equally affected, regardless of their FA chain composition.

The GBA treatment with CBE is both time and dose dependent, and we chose to extract the lipids only after 14 days to be sure that GBA was properly inhibited, and also to give cells time to adapt their metabolic networks to GBA inactivity. It is possible that, had we treated the cells for a shorter amount of time, we wouldn't have observed the significant accumulation of GBA substrates discussed above.

The main conclusion of this first part is that our targeted LC-MS approach can be useful to tackle the problem of identifying possible sphingolipid imbalances related to enzyme malfunction or to underlying disease states.

2.3 *Lrrk2*^{+/+} and *Lrrk2*^{-/-} Mice

2.3.1 Materials and Methods

Animal Care and Sample Collection

C57BL/6 LRRK2 wild type (*Lrrk2*^{+/+}) and knock-out (*Lrrk2*^{-/-}) mice were provided by Dr. Heather Melrose. Housing and handling of mice were done in compliance with national guidelines. All animal procedures were approved by the Ethical Committee of the University of Padova and the Italian Ministry of Health (license 46/2012).

Lipid Extraction

Brains were collected from three male *Lrrk2*^{+/+} and three male *Lrrk2*^{-/-} mice at 1 year of age. Lipid extraction from brains was performed through a protocol optimised for ceramides and sphingolipids; briefly, brains were homogenised in ultra-pure water and incubated on ice. 3 mL Folch solution (chloroform:methanol 2:1 v/v) was added to the homogenate and the solution was sonicated for 15 min at 4 °C to complete the lipids extraction. Centrifugation at 16 000 g for 15 min at 4 °C followed to recover the liquid phase.

The extracted samples were dried using a rotary evaporator, dissolved in 150 µL HPLC-grade methanol (Sigma-Aldrich Co.) and stored in closed vials at -20 °C until analysis.

LC-MS Measurements

The LC-MS procedure was the same described above (Section 2.2.1). However, considering the different biological origin of the two sample sets, the species we identified here were dramatically different: through PIS of *m/z* 184 we identified PC, Lyso-PC, plasmenyl-PC (pPC) and SM, whereas with PIS of *m/z* 264 we characterised Cer 18:1, Gly-Cer 18:1, 2-hydroxylated 18:1 glycosyl-ceramides (Gly-Cer 18:1(2-OH)) and sulfatides. Finally, with PIS of *m/z* 266 we detected Cer 18:0 and Gly-Cer 18:0. Again, the reversed-phase chromatography did not allow us to resolve Glc-Cer from Gal-Cer, which is the reason why we refer to them as Gly-Cer.

Also here we used external standard solutions, so as to convert the LC-MS peak areas into molar quantities, and also to optimise the instrumental parameters (the same as before).

Data Analysis

Similarly to above (Section 2.2.1), we handled the LC-MS data with the proprietary software (Analyst 1.4.2), and then we performed the statistical analysis by using R [51].

To compare sphingolipid levels across samples, regardless of their original biomass availability, we normalised their molar amounts to the total lipid content. To investigate intra-class variations, we used relative amounts obtained by normalising the molar amounts in each sphingolipid class to the overall molar amount in that class, set at the arbitrary value of 100. With such data, we also computed and compared parameters such as average chain length, average number of unsaturations and unsaturation index.

2.3.2 Results and Discussion

Relative Distribution Across Sphingolipids

The sphingolipid molar amounts, normalised to the total lipid content, are in Figure 2.11. The comparison between sample groups shows a significant difference in Cer 18:1, more abundant in *Lrrk2*^{-/-} mice. It is also interesting to note a similar behaviour for SM and sulfatides, downstream of GBA as well (see Figure 2.1): both tend to be more abundant in *Lrrk2*^{-/-} mice, although not significantly ($p > 0.05$).

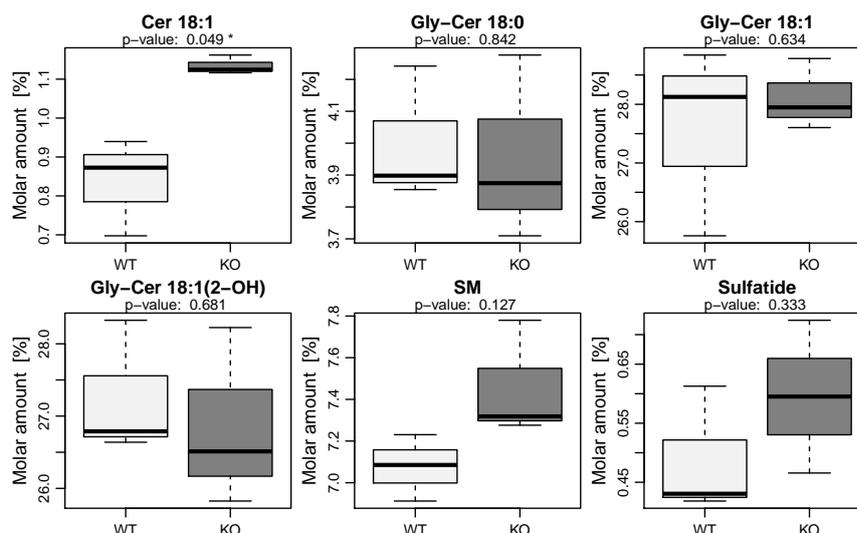


FIGURE 2.11: Inter-class distribution of sphingolipids (data normalised to the total lipid content). *Lrrk2*^{-/-} samples have significantly higher Cer 18:1 amounts as compared to control samples.

These results suggest that the absence of LRRK2 may affect the sphingolipid distribution. To note that the results are quite different from those in the previous investigation (Section 2.2): here, there is an increase in GBA products, as if the absence of LRRK2 increases the GBA activity or its expression.

Gly-Cer/Cer Ratios

We then moved on by computing the 18:1 Gly-Cer/Cer ratios for each couple with the same acyl chains. The results are in Figure 2.12: in most cases, the ratios are lower in KO as compared to WT samples, although the differences are not significant ($p > 0.05$). Similar conclusions hold true for the overall Gly-Cer/Cer ratio (Figure 2.13). The lack of significance is probably caused by the small number of samples ($n = 3$ for each group).

These results confirmed what also observed above, mainly that a possible sphingolipid imbalance may take place when LRRK2 is missing; the most important effect is the increase in GBA products (Cer).

GBA Activity Assays and Protein Expression Levels

At the same time as the LC-MS measurements, our collaborators at the Department of Biology, University of Padova, measured total protein concentration by a BCA assay, and assessed *in vitro* GBA activity. Figure 2.14 shows their results: they unexpectedly found a significant GBA down-regulation in *Lrrk2*^{-/-} brain lysates. In addition, in order to account for the different GBA levels in *Lrrk2*^{+/+} and *Lrrk2*^{-/-} brain lysates, they computed the GBA specific activity by normalising the enzymatic activity to the protein level (measured by western blot). It turned out that GBA specific activity is greater in *Lrrk2*^{-/-} brain lysates as compared to *Lrrk2*^{+/+} samples.

Therefore, the lack of LRRK2 seems to induce a GBA down-regulation, resulting in decreased protein levels. At the same time, though, the GBA

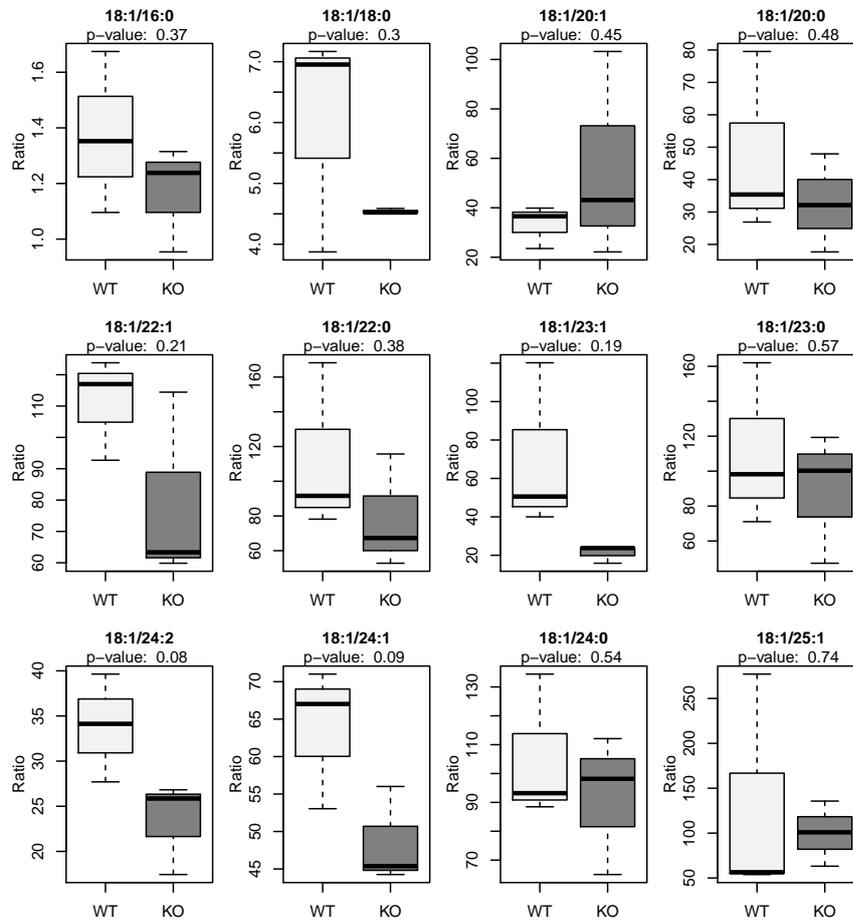


FIGURE 2.12: Individual Gly-Cer/Cer ratios. The group differences are not significant, although in most cases the ratios are lower in KO ($Lrrk2^{-/-}$) as compared to WT ($Lrrk2^{+/+}$) samples.

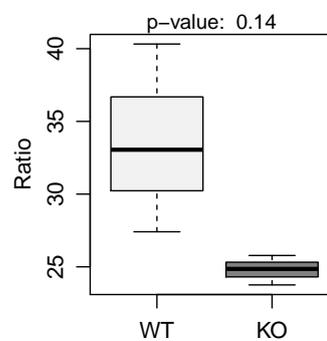


FIGURE 2.13: Overall Gly-Cer/Cer ratio.

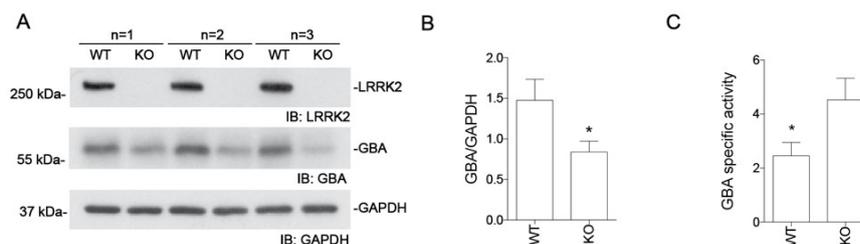


FIGURE 2.14: **A)** Western blot of different brain lysates for GBA and LRRK2 protein levels for *Lrrk2*^{+/+} (WT) and *Lrrk2*^{-/-} (KO) mice. **B)** Relative quantification (n = 9 for each genotype): GBA protein levels are significantly lower in *Lrrk2*^{-/-} mice. **C)** GBA specific activity: it is significantly higher in *Lrrk2*^{-/-} samples (*: p ≤ 0.05).

specific activity increases. By taking into account the LC-MS results, the latter effect seems to be more important, since we observed an increase in GBA products.

The seemingly contrasting results here obtained are difficult to properly interpret: on the one hand, from lipidomics studies we have that GBA products are significantly more abundant in *Lrrk2*^{-/-} samples, whereas on the other hand we have that these same samples are characterised by reduced GBA amount and activity, but an increased specific activity. Of course, we can only provide putative explanations at this stage, and a possible one is the following: from what above, LRRK2 seems to be involved in GBA regulation, and considering that LRRK2 was reported to function as a scaffold for several kinases and in the Wnt signaling by orchestrating their subcellular distribution [52], it may be possible that LRRK2 regulates GBA activity by affecting GBA localisation (for example, its delivery to lysosomes). As a consequence, despite lower GBA amounts, this latter would turn out to be more active if properly delivered to the lysosomes.

Intra-class Distribution

We then looked at the intra-class distributions within each sphingolipid class, to identify possible molecular changes brought about by the treatment: the profiles in Figure 2.15 show no differences in terms of FA chains between *Lrrk2*^{+/+} and *Lrrk2*^{-/-} samples.

Similarly to above, our conclusion is that whereas the lack of LRRK2 has an impact on the sphingolipid relative amounts, the effects are not specific to certain FA chains. In other words, all molecular species belonging to the same class are equally affected by the lack of LRRK2. Again, we further confirmed this observation by computing and comparing the parameters average chain length, average number of unsaturations and unsaturation index (Appendix B.2), for which we found no significant differences.

2.3.3 Conclusions

Our targeted approach was motivated by the fact that sphingolipids have important roles in cell function during disease development and progression. Not surprisingly, many studies have provided evidence that sphingolipid

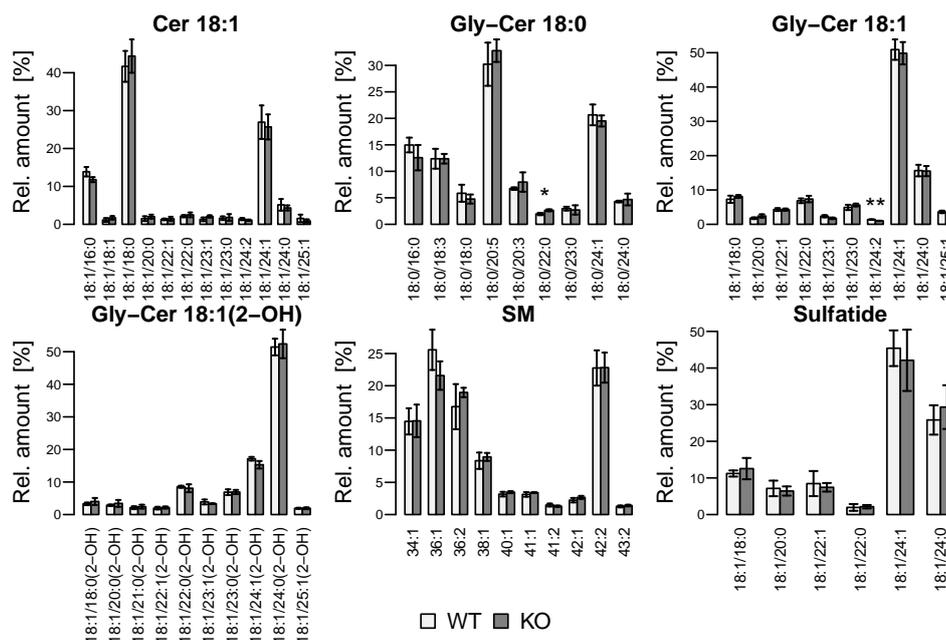


FIGURE 2.15: Intra-class distributions of the sphingolipid classes identified for mouse brain samples. To ease the visualisation, only the molecular species with a relative amount higher than 1% are shown. Overall, there is no difference between *Lrrk2*^{+/+} (WT) and *Lrrk2*^{-/-} (KO) in terms of FA chains.

levels are often altered in NDs [16–19], and Cer were found to have a defective homeostasis in NDs with LB pathology, including PD [16, 20–22].

We carried out a targeted LC-MS analysis using a QQQ mass spectrometer which we operated in PIS mode, to identify possible sphingolipid perturbations arising from the lack of LRRK2 protein in *Lrrk2*^{-/-} mice; to this aim, we compared the sphingolipidomics profiles of brains from *Lrrk2*^{+/+} (WT) and *Lrrk2*^{-/-} (KO) mice. Unfortunately, the number of samples was quite limited ($n = 3$ for each group), which may have worsened our ability to pick out significant group differences. However, this did not preclude us from pinpointing important features.

From the lipidomics analysis, we found *Lrrk2*^{-/-} mouse brains to be associated with a significantly higher amount of Cer 18:1 as compared to *Lrrk2*^{+/+} samples (Figure 2.11); more in general, we observed an increase in all sphingolipids downstream of GBA (Cer, SM and sulfatides). Accordingly, the Gly-Cer/Cer ratios (Figure 2.12 and Figure 2.13) resulted to be lower in *Lrrk2*^{-/-} samples. These sphingolipid perturbations are not chain-specific, since we found no significant differences in the intra-class profiles (Figure 2.15), nor did we observe any variation in the parameters average chain length, average number of unsaturations and unsaturation index (Appendix B.2).

At the same time, the analyses performed on GBA revealed that the enzyme is down-regulated when LRRK2 is missing but, interestingly, its specific activity is increased (Figure 2.14). Taken together, these results and the LC-MS outcomes suggest that the increased GBA specific activity is more important than its decreased expression when it comes to sphingolipid balances. Our results therefore hint to a possible interaction between LRRK2 and GBA, with LRRK2 playing a role in GBA regulation; this conclusion

finds support from recently published work, where it was reported an increase in GBA activity in blood samples from LRRK2 G2019S patients [53]. To complicate things is the fact that in idiopathic PD patients and in GBA mutation carriers, a decrease in GBA level is usually associated with a decrease in GBA activity [54-59]. However, the association of higher GBA activity in LRRK2 G2019S patients and in LRRK2 deficient systems may be caused by a distinct pathogenic mechanism of LRRK2-linked PD that deserves further investigation.

Bibliography of Chapter 2

1. Schultz, W. Multiple dopamine functions at different time courses. *Annual Review of Neuroscience* **30**, 259–288 (2007).
2. Jaber, M., Robinson, S. W., Missale, C. & Caron, M. G. Dopamine receptors and brain function. *Neuropharmacology* **35**, 1503–1519 (1996).
3. Reeve, A., Simcox, E. & Turnbull, D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? *Agein Research Reviews* **14**, 19–30 (2014).
4. Schapira, A. H. & Jenner, P. Etiology and pathogenesis of Parkinson's disease. *Movement Disorders* **26**, 1049–1055 (2011).
5. Jo, E., McLaurin, J., Yip, C. M., St. George-Hyslop, P. & Fraser, P. E. α -synuclein membrane interactions and lipid specificity. *Journal of Biological Chemistry* **275**, 34328–34334 (2000).
6. Norris, E. H., Giasson, B. I. & Lee, V. M. α -synuclein: normal function and role in neurodegenerative diseases. *Current Topics in Developmental Biology* **60**, 17–54 (2004).
7. Nakamura, K. *et al.* Direct membrane association drives mitochondrial fission by the Parkinson Disease-associated protein α -synuclein. *Journal of Biological Chemistry* **286**, 20710–20726 (2011).
8. Stefanovic, A. N. D., Stöckl, M. T., Claessens, M. M. A. E. & Subramaniam, V. α -synuclein oligomers distinctively permeabilize complex model membranes. *The FEBS Journal* **281**, 2838–50 (2014).
9. Olanow, W. C. & Brundin, P. Parkinson's disease and alpha synuclein: is Parkinson's disease a prion-like disorder? *Movement Disorders* **28**, 31–40 (2013).
10. Firestone, J. A. *et al.* Pesticides and risk of Parkinson disease - A population-based case-control study. *Archives of Neurology* **62**, 91–95 (2005).
11. Kamel, F. & Hoppin, J. A. Association of pesticide exposure with neurologic dysfunction and disease. *Environmental Health Perspectives* **112**, 950–958 (2004).
12. Factor, S. A. & Weiner, W. J. Prior history of head trauma in Parkinson's disease. *Movement Disorders* **6**, 225–229 (1991).
13. Lees, A. J., Hardy, J. & Revesz, T. Parkinson's disease. *Lancet* **373**, 2055–2066 (2009).
14. Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harbor Perspectives in Medicine* **2**, a008888–a008888 (2012).
15. Jankovic, J. & Aguilar, L. G. Current approaches to the treatment of Parkinson's disease. *Neuropsychiatric Disease and Treatment* **4**, 743–757 (2008).

16. Bras, J., Singleton, A., Cookson, M. R. & Hardy, J. Emerging pathways in genetic Parkinson's disease: potential role of ceramide metabolism in Lewy body disease. *FEBS Journal* **275**, 5767–5773 (2008).
17. Haughey, N. J. Sphingolipids in neurodegeneration. *NeuroMolecular Medicine* **12**, 301–305 (2010).
18. Haughey, N. J., Bandaru, V. V., Bae, M. & Mattson, M. P. Roles for dysfunctional sphingolipid metabolism in Alzheimer's disease neuropathogenesis. *Biochimica et Biophysica Acta* **1801**, 878–886 (2010).
19. Fabelo, N. *et al.* Severe alterations in lipid composition of frontal cortex lipid rafts from Parkinson's disease and incidental Parkinson's disease. *Molecular Medicine* **17**, 1107–1118 (2011).
20. Abbott, S. K. *et al.* Altered ceramide acyl chain length and ceramide synthase gene expression in Parkinson's disease. *Movement Disorders* **29**, 518–526 (2014).
21. Don, A. S. *et al.* Altered lipid levels provide evidence for myelin dysfunction in multiple system atrophy. *Acta Neuropathologica Communications* **2**, 150 (2014).
22. Mielke, M. M. *et al.* Plasma ceramide and glucosylceramide metabolism is altered in sporadic Parkinson's disease and associated with cognitive impairment: a pilot study. *PLoS One* **8**, e73094 (2013).
23. De Chaves, P. & Sipione, S. Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction. *FEBS Letters* **584**, 1748–1759 (2010).
24. Harvald, E. B., Olsen, A. S. B. & Færgeman, N. J. Autophagy in the light of sphingolipid metabolism. *Apoptosis* **20**, 658–670 (2015).
25. Kogot-Levin, A. & Saada, A. Ceramide and the mitochondrial respiratory chain. *Biochimie* **100**, 88–94 (2014).
26. Sidransky, E. & Lopez, G. The link between the GBA gene and parkinsonism. *The Lancet Neurology* **11**, 986–998 (2012).
27. Stern, G. Niemann–Pick's and Gaucher's diseases. *Parkinsonism and Related Disorders* **20**, S143–S146 (2014).
28. Cleeter, M. W. J. *et al.* Glucocerebrosidase inhibition causes mitochondrial dysfunction and free radical damage. *Neurochemistry International* **62**, 1–7 (2013).
29. Chu, Y. & Kordower, J. H. The prion hypothesis of Parkinson's disease. *Current Neurology and Neuroscience Reports* **15**, 28 (2015).
30. Yap, T. L. *et al.* α -synuclein interacts with glucocerebrosidase providing a molecular link between Parkinson and Gaucher diseases. *The Journal of Biological Chemistry* **286**, 28080–28088 (2011).
31. Ho, F. Y., Rosenbusch, K. E. & Kortholt, A. in *A synopsis of Parkinson's disease* 1–32 (InTech, 2014).
32. Mata, I. F., Wedemeyer, W. J., Farrer, M. J., Taylor, J. P. & Gallo, K. A. LRRK2 in Parkinson's disease: protein domains and functional insights. *Trends in Neurosciences* **29**, 286–93 (2006).

33. Lesage, S. & Brice, A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human Molecular Genetics* **18**, R48–R59 (2009).
34. Heckman, M. G. *et al.* Population-specific frequencies for LRRK2 susceptibility variants in the genetic epidemiology of Parkinson's disease (GEO-PD) consortium. *Movement Disorders* **28**, 1740–1744 (2013).
35. Greggio, E. *et al.* The Parkinson disease-associated leucine-rich repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation. *The Journal of Biological Chemistry* **283**, 16906–16914 (2008).
36. Sheng, Z. *et al.* Ser1292 autophosphorylation is an indicator of LRRK2 kinase activity and contributes to the cellular effects of PD mutations. *Science Translational Medicine* **4**, 164ra161 (2012).
37. Steger, M. *et al.* Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* **5**, e12813 (2016).
38. Zhu, X. *et al.* LRRK2 in Parkinson's disease and dementia with Lewy bodies. *Molecular Neurodegeneration* **1**, 17 (2006).
39. Guerreiro, P. S. *et al.* LRRK2 interactions with α -synuclein in Parkinson's disease brains and in cell models. *Journal of Molecular Medicine* **91**, 513–522 (2013).
40. Orenstein, S. J. *et al.* Interplay of LRRK2 with chaperone-mediated autophagy. *Nature Neuroscience* **16**, 394–406 (2013).
41. Hinkle, K. M. *et al.* LRRK2 knockout mice have an intact dopaminergic system but display alterations in exploratory and motor co-ordination behaviors. *Molecular Neurodegeneration* **7**, 25 (2012).
42. Tong, Y. *et al.* Loss of leucine-rich repeat kinase 2 causes age-dependent bi-phasic alterations of the autophagy pathway. *Molecular Neurodegeneration* **7**, 1–16 (2012).
43. Miklavc, P. *et al.* Surfactant secretion in LRRK2 knock-out rats: changes in lamellar body morphology and rate of exocytosis. *PLoS One* **9**, e84926 (2014).
44. Herzig, M. C. *et al.* LRRK2 protein levels are determined by kinase function and are crucial for kidney and lung homeostasis in mice. *Human Molecular Genetics* **20**, 4209–4223 (2011).
45. Baptista, M. A. *et al.* Loss of leucine-rich repeat kinase 2 (LRRK2) in rats leads to progressive abnormal phenotypes in peripheral organs. *PLoS One* **8**, e80705 (2013).
46. Ness, D. *et al.* Leucine-rich repeat kinase 2 (LRRK2)-deficient rats exhibit renal tubule injury and perturbations in metabolic and immunological homeostasis. *PLoS One* **8**, e66164 (2013).
47. Hilgenberg, L. G. W. & Smith, M. A. Preparation of dissociated mouse cortical neuron cultures. *Journal of Visualized Experiments* **10**, 562 (2007).
48. Folch, J., Lees, M. & Stanley, G. H. S. A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry* **226**, 497–509 (1957).
49. Brügger, B. Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry. *Annual Review of Biochemistry* **83**, 79–98 (2014).

50. Shaner, R. L. *et al.* Quantitative analysis of sphingolipids for lipidomics using triple quadrupole and quadrupole linear ion trap mass spectrometers. *Journal of Lipid Research* **50**, 1692–1707 (2009).
51. R Development Core Team. *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
52. Berwick, D. C. & Harvey, K. LRRK2 functions as a Wnt signaling scaffold, bridging cytosolic proteins and membrane-localized LRP6. *Human Molecular Genetics* **21**, 4966–4979 (2012).
53. Alcalay, R. N. *et al.* Glucocerebrosidase activity in Parkinson’s disease with and without GBA mutations. *Brain* **138**, 2648–2658 (2015).
54. Schöndorf, D. C. *et al.* iPSC-derived neurons from GBA1-associated Parkinson’s disease patients show autophagic defects and impaired calcium homeostasis. *Nature Communications* **5**, 4028 (2014).
55. Gegg, M. E. *et al.* Glucocerebrosidase deficiency in substantia nigra of Parkinson disease brains. *Annals of Neurology* **72**, 455–463 (2012).
56. de la Mata, M. *et al.* Pharmacological chaperones and coenzyme Q10 treatment improves mutant β -glucocerebrosidase activity and mitochondrial function in neuronopathic forms of Gaucher disease. *Scientific Reports* **5**, 10903 (2015).
57. Kim, H. J. *et al.* Leukocyte glucocerebrosidase and β -hexosaminidase activity in sporadic and genetic Parkinson disease. *Parkinsonism and Related Disorders* **23**, 99–101 (2016).
58. Murphy, K. E. *et al.* Reduced glucocerebrosidase is associated with increased α -synuclein in sporadic Parkinson’s disease. *Brain* **137**, 834–848 (2014).
59. Chiasserini, D. *et al.* Selective loss of glucocerebrosidase activity in sporadic Parkinson’s disease and dementia with Lewy bodies. *Molecular Neurodegeneration* **10**, 1–6 (2015).

Chapter 3

The Effects of Anti-VEGF Treatments on the Lipidome of Cancer Cells

3.1 Introduction

The word “cancer” encloses a collection of diseases where some of the body’s cells start to divide uncontrollably and to spread into surrounding tissues. In cancer cells, the balance between growth, proliferation and differentiation is broken through three important steps:

1. **Immortalisation:** cells acquire the ability to divide indefinitely;
2. **Transformation:** cells cease to observe the normal constraints of growth, and develop their own blood supply by way of angiogenesis;
3. **Metastasis:** cancer cells spread from the tissue of origin to other parts of the body.

Cancer accounts for one in eight deaths worldwide, making it the leading cause of death in developed countries and the second leading cause in developing countries, after heart diseases [1]. The risk of contracting cancer increases with age, and in rich countries almost 80 % of newly diagnosed cases involve people older than 55 years of age; this is probably caused by cancer development requiring many steps, which take place over the years [2, 3].

All cancers involve genetic abnormalities in cells. Somatic mutations, the most common, are not hereditary but result from gene damage occurring during a person’s lifetime. These mutations, however, become a heritable trait, and they account for about 5 % of cancer cases.

The genetic changes associated with tumorigenesis can be divided into two categories, depending on whether they result from a gain-of-function or from a loss-of-function: the former involves inappropriate activation of oncogenes, which are stimulatory for growth and cause cancer when hyperactive; conversely, loss-of-function involves inactivation of tumour suppression genes, which inhibit cell growth [4].

Cancer cells are characterised by three abnormal features:

1. **The Warburg effect.** Regardless of the oxygen levels, cancer cells get the energy they need by a high conversion rate of glucose to lactate, rather than metabolising it through oxidative phosphorylation in the mitochondria [5]. These cells are reported to have glycolytic rates as high as 200 times those of normal cells [6].

- 2. A high rate of *de novo* lipogenesis.** DNL is the process through which cells synthesise fatty acids starting from substrates such as glucose and glutamine; in healthy individuals, this biosynthetic pathway usually provides a limited amount of fatty acids (FAs) as compared to those coming from the diet. However, in spite of the extracellular FA availability, cancer cells can synthesise up to 95% of FAs by means of DNL [7–9]. What is still not known is why lipids are not simply taken from the bloodstream, and one putative explanation is that DNL increases the resistance of cancer cells to oxidative stress, since DNL produces saturated and mono-unsaturated FAs, whereas the diet would provide more unsaturated species, vulnerable to peroxidation. Indeed, inhibition of lipogenesis was shown to increase peroxidation products, leading to cells more susceptible to oxidative stress [10]. On the other hand, unsaturated FAs are essential for proper membrane fluidity, which may be the reason why stearoyl-CoA desaturase (SCD), responsible for desaturation during DNL, is usually over-expressed in cancer cells [11]. As a matter of fact, targeting SCD has been suggested for cancer treatment [12].
- 3. A high angiogenesis.** In normal physiology, the process of angiogenesis (the growth of blood vessels from pre-existing vasculature) is highly regulated and useful in processes such as wound healing, skeletal growth, the menstrual cycle and pregnancy. In cancer, angiogenesis is troublesome because cells use it to get a high supply of nutrients and oxygen as they grow in number [13]. Tumour vasculature is also required for removing metabolic products that would be toxic for cells, such as lactic acid [8]. Hypoxia can contribute to angiogenesis: when the tumour outgrows the tissue blood supply, hypoxia-inducible factors (HIFs) are activated, and they promote angiogenesis by inducing expression of the vascular endothelial growth factor (VEGF) [14, 15]. HIFs also promote a metabolic switch to anaerobic energy production, which in its turn induces uptake of FAs with accumulation of lipid droplets [16]. Indeed, cancer cells are known to contain a higher amount of lipid droplets as compared to normal cells, which presumably they use to produce energy through β -oxidation when the environmental conditions become difficult [17, 18].

Many therapeutic strategies have been proposed so far to treat cancer, including surgery, radiation, chemotherapy and immunotherapy. The approach of interest for this Thesis is the targeting of angiogenesis through anti-VEGF therapy [19, 20]: VEGF is a potent pro-angiogenic growth factor expressed by many cancer cells, and it stimulates endothelial-cell proliferation. Unlike tumour vessels, the normal adult vasculature can survive independently from VEGF, and so anti-VEGF therapy is based on the assumption that tumour vessels can be impacted without harming the others [21]. Unfortunately, the benefits of anti-VEGF drugs are still not so clear, especially when used as monotherapy [22]. On the other hand, anti-VEGF combined with chemotherapy proved successful to the point that bevacizumab, a humanised VEGF-specific antibody, became the first anti-VEGF agent to be approved by the Food and Drug Administration for cancer patients [19, 21].

An issue of anti-VEGF drugs is that they do not seem to work on previously treated cancer patients, probably because of an acquired resistance

to antiangiogenic drugs. In addition, it was shown that prolonged treatments can affect the glycolytic phenotype of the cancer cells, leading to more hypoxic and glycolytic tumours [23], as well as to oxidative stress [24]. Pre-clinical studies have also reported increased tumour growth and metastatic formation after withdrawal of anti-VEGF agents [25]; another reported side-effect is the up-regulation of Fatty Acid Binding Proteins (FABP3 and FABP7), involved in FA uptake, therefore resulting in increased lipid droplet levels with triacylglycerol (TAG) accumulation [16], that cannot be degraded under hypoxia via β -oxidation since this process is oxygen-dependent.

The mechanisms by which cancer evades anti-VEGF treatments are not clear, and several guesses have been made, such as (i) escape via different modes of vascularisation, (ii) recruitment of pro-angiogenic cells from the circulation, and (iii) secretion of alternative pro-angiogenic factors [24]. As a consequence, further research is needed to shed some light on these complex processes.

Considering that lipids play crucial roles in cancer development and treatment, in this project we focused our attention on the lipidome composition of cancer cells, and we looked for possible lipidomics changes associated with anti-VEGF treatment.

Ovarian Cancer

Ovarian cancer begins in the ovaries, the female reproductive glands. The ovaries contain 3 types of cells, and each of them can develop into a different type of tumour:

- Epithelial tumours originate from the cells that cover the outer surface of the ovary;
- Germ cell tumours begin from the cells that produce the eggs;
- Stromal tumours start from tissue cells that hold the ovary together and produce the hormones estrogen and progesterone.

In most cases, these tumours are benign and do not spread beyond the ovary; they can be effectively treated by removing the ovary or simply part of it. On the other hand, malignant tumours do spread to other parts of the body. As a consequence, they can be fatal.

Cancerous epithelial tumours are called carcinomas, and about 90 % of them are epithelial ovarian carcinomas. These tumour cells have several features, when viewed under a microscope, that are used to classify epithelial ovarian carcinomas into different types. For example, the serous type is by far the most common, but there are other types like mucinous, endometrioid, and clear cell. If the cells don't look like any of these subtypes, the tumour is called undifferentiated, and these carcinomas tend to grow more quickly than the other types. Epithelial ovarian carcinomas are classified by these subtypes, but they are also given a grade and a stage.

The grade classifies the tumour based on how much it looks like normal tissue: grade 1 carcinomas look more like normal tissue, whereas grade 3 carcinomas look less like normal tissue and usually have a worse prognosis. Grade 2 tumours look and act in between.

The tumour stage describes how far the tumour has spread from where it started in the ovary. Epithelial ovarian cancers tend to spread to the lining

and organs of the pelvis and abdomen first. As it becomes more advanced, it may spread to the lung and liver, but even to the brain, bones, or skin.

There currently are different treatment options for such cancers, including surgery, chemotherapy, hormone therapy and radiation therapy [26, 27]. Outcomes are largely dependent on the subtype of the cancer present. For example, the overall five-year survival rate in the United States is 45 %, but this figure dramatically drops in developing countries.

3.2 Materials and Methods

Sample Collection and Treatment

In this study four different human ovarian adenocarcinoma cell lines were used: they are listed in Table 3.1. Basically, they differ in their glycolytic phenotype (IGROV-1 and SKOV-3 are poorly glycolytic, whereas OC-316 and OVCAR-3 are highly glycolytic cell lines).

CELL LINE	DISEASE
IGROV-1	Ovarian endometrioid adenocarcinoma
OC-316	Ovarian serous adenocarcinoma
OVCAR-3	High grade ovarian serous adenocarcinoma
SKOV-3	Ovarian serous cystadenocarcinoma

TABLE 3.1: Description of the cell lines used in the experiment.

Source: *ExpASY – Bioinformatics Resource Portal* (www.expasy.org).

The cells were grown in RPMI1640 (Euroclone, Pero, Italy) supplemented with 10 % fetal calf serum (FCS; Life Technologies, Gaithersburg, MD), 1 % HEPES (10 mM, Cambrex Bioscience, Verviers, Belgium), 1 % L-Glutamine (2 mM), 1 % sodium pyruvate (1 mM) and 1 % antibiotic-antimycotic mix (Gibco-BRL, Grand Island, NY). Cultures were maintained at 37 °C in a humidified 5 % CO₂ / 95 % air atmosphere.

All procedures involving animal care conformed institutional guidelines that comply national and international policies (EEC Council Directive 86/609, OJ L 358, 12 December, 1987). For tumour establishment, 8-week-old severe combined immunodeficiency (SCID) mice (Charles River, Wilmington, MA) were subcutaneously injected with $0.3\text{--}0.5 \times 10^6$ tumour cells mixed at 4 °C with liquid Matrigel (Becton-Dickinson; Franklin Lakes, NY). About 2 months later, animals developed tumours, and anti-human VEGF mAb (bevacizumab) was administered intraperitoneally at 100 µg/dose twice per week. Control mice received intraperitoneal injections of PBS (Figure 3.1). The treatment lasted 4 weeks, and 48 hours after the last dose the mice were sacrificed, the tumours taken and stored at –80 °C until the analysis. The total number of samples varied depending on the experiment:

1. IGROV-1, 16 mice (11 CTRL and 5 TRTD);
2. OC-316, 15 mice (8 CTRL and 7 TRTD);
3. OVCAR-3, 11 mice (5 CTRL and 6 TRTD);
4. SKOV-3, 11 mice (6 CTRL and 5 TRTD).

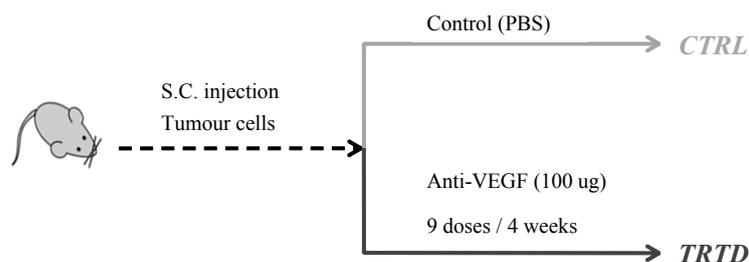


FIGURE 3.1: Scheme of the sample preparation. TRTD mice were given the anti-VEGF drug bevacizumab, whereas CTRL mice were not treated.

Lipid Extraction

Lipids were extracted using a slight modification of the Bligh & Dyer protocol [28]. Briefly, the tumours were placed in a glass centrifuge tube, to which 3.75 mL of MeOH:CHCl₃ 2:1 mixture was added. The samples were thoroughly mixed and mechanically disrupted by means of a homogeniser, while keeping the centrifuge tubes in a bath of water and ice. 20 µL of 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) 0.159 mg/mL standard solution was added to assess possible fluctuations in extraction yield.

The samples were sonicated and vortexed for 15 min, then 1.25 mL CHCl₃ was added, and the samples were sonicated and mixed again for another 15 min. 1.25 mL H₂O was added, and there followed a further sonication/mixing step. Finally, they were centrifuged for 15 min at 2000 rpm to induce phase separation. The organic (bottom) phase was recovered using a Pasteur pipette in a 10 mL round-bottom flask, and the extraction procedure was repeated: 1.88 mL CHCl₃ was added to the tubes, the samples were sonicated, mixed and then centrifuged. The bottom organic phase was recovered again and added to the previously extracted one.

The flasks were dried using a rotary evaporator and the lipids were dissolved in 1 mL MeOH:CHCl₃ 8:2 solution. For the LC-MS analysis, the samples were diluted 1:10 with MeOH.

LC-MS Measurements

The LC-MS measurements were performed in both positive and negative ionisation modes using a Waters Xevo G2 quadrupole time-of-flight (Q-ToF) combined with an Acquity UPLC system (Waters Corporation, Manchester, UK). 10 µL of each sample was injected onto an Acquity UPLC Charged Surface Hybrid (CSH) C18 column (1.7 µm × 2.1 mm × 100 mm, Waters Corporation) held at 55 °C. The flow rate was 0.4 mL/min, and the binary solvent system consisted of solvent A, HPLC-grade acetonitrile:water (60:40) with 10 mM ammonium formate, and solvent B, HPLC-grade acetonitrile:isopropanol (10:90) with 10 mM ammonium formate. The gradient elution program started from 40 % B, reached 99 % B in 18 min, then returned back to the starting condition, remaining there for 2 min. The MS data was collected over the *m/z* range 100–1800 with a scan duration of 0.2 s. The source temperature was set at 120 °C and nitrogen (900 L/h) was used as the desolvation gas. The voltages of the sampling cone, extraction cone and capillary were 30 kV, 3.5 kV and 2 kV respectively, with a collision energy of

6 V for each full scan, and a collision ramp from 20 to 40 V for the fragmentation function. As lock mass, a solution of 2 ng/ μ L acetonitrile:water (50:50) leucine enkephalin (m/z 556.2771) with 0.1 % formic acid was infused into the instrument every 30 s.

Data Analysis

The raw data files were converted to the *.NetCDF format with the software “Databridge” (Waters Corporation) and then handled with the R package *xcms* [29]; the processing procedures included feature detection, peak identification, peak matching, peak grouping across samples, retention time correction and filling in of missing data. The detected peaks were assigned based on their m/z and retention time (RT). The final table containing peak areas was used for data analysis, performed with the R environment for statistical computing [30].

Overall, 10 lipid classes were identified from the positive ion mode scans: ceramides (Cer), diacylglycerols (DAG), glycosyl-ceramides (Gly-Cer), phosphatidylcholines (PC), LysoPC, pPC (*i.e.*, plasmany- and plasmeryl-PC), phosphatidylethanolamines (PE), pPE (*i.e.*, plasmany- and plasmeryl-PE), sphingomyelins (SM) and TAG. The most common adducts were $[M + H]^+$, $[M + Na]^+$ and $[M + NH_4]^+$, depending on the lipid class.

The signal of the DLPC standard was used to account for extraction yield fluctuations, based on the assumption that different yields across samples will result in different DLPC signal intensities. Accordingly, for each sample the analyte signal areas were scaled to the DLPC signal. In addition, in order to study relative differences across samples, for each sample a normalisation to the total signal area, set to 100, was carried out.

As explained in Chapter 1, the most appropriate and meticulous approach would be to employ internal standards in such a way as to relate signal areas to absolute lipid amounts. This approach is however impractical, as well as overly expensive, and we therefore chose to undertake a less ambitious route, whereby we limited ourselves to compare signal areas. After all, our aim is to look for differences of MS features across sample groups, for which no absolute molar amounts are actually required.

3.3 Results and Discussion

3.3.1 IGROV-1 Cell Line

Tumour Dimension and Total Signal

The tumour dimensions are shown in Figure 3.2; within each sample group (CTRL or TRTD), a high biological variability can be noticed. Despite that, the difference in terms of average dimensions is statistically significant, with TRTD tumours being smaller than controls.

In Figure 3.3 is the comparison of the total MS signal across samples. This quantity was obtained by adding, within each sample, the signal areas of the identified lipid species through our MS-based approach. In other words, from the chromatograms we have integrated all the relevant peaks, and added their areas together. The rationale is that the higher the concentration of lipids in the considered sample, the higher this quantity is.

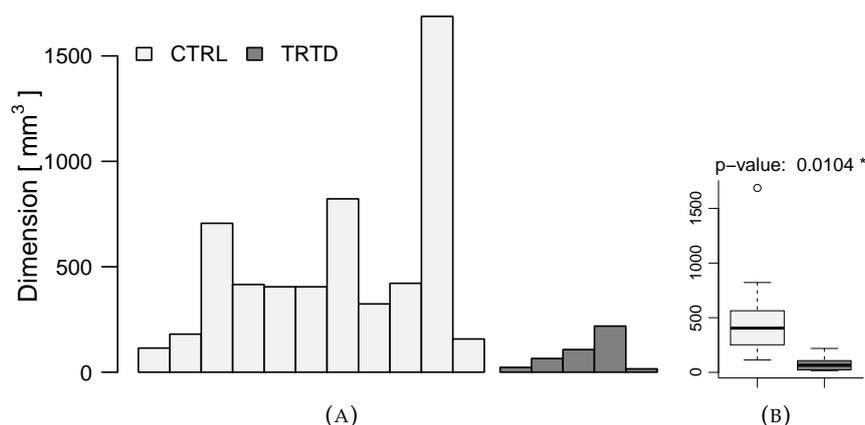


FIGURE 3.2: IGROV-1 – Bar-plot (A) and box-plot (B) of the tumour dimensions. Despite the high biological variability, there is a significant difference (at the 5% level) in terms of tumour dimension, with treated tumours being the smallest.

Surprisingly, no significant group difference can be observed, which therefore suggests the lack of any linear relationship between tumour dimension and overall amount of lipids. This was also confirmed by an analysis of the sample correlation coefficients ($r = 0.30$ for CTRL and $r = 0.82$ for TRTD).

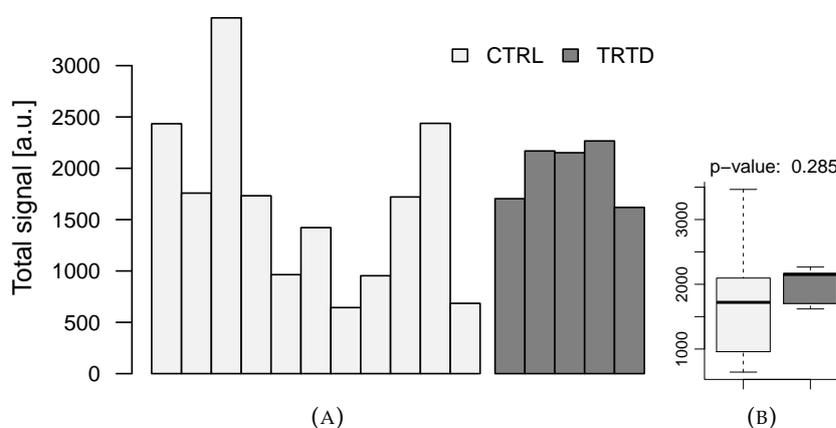


FIGURE 3.3: IGROV-1 – Bar-plot (A) and box-plot (B) of the total signal area. The difference between CTRL and TRTD is not statistically significant (at the 5% level).

Relative Distribution Among Lipid Classes

In Figure 3.4 is a series of box-plots where each identified lipid class is compared between CTRL and TRTD groups. In most cases, the difference is statistically significant ($p \leq 0.05$) and two clear trends can be observed: whereas TAGs are more abundant in TRTD samples, the opposite holds true for almost all the other lipid classes. It is worth stressing that the amounts being compared here are signal areas, and therefore the percentages reported in Figure 3.4 are not to be interpreted as molar abundances.

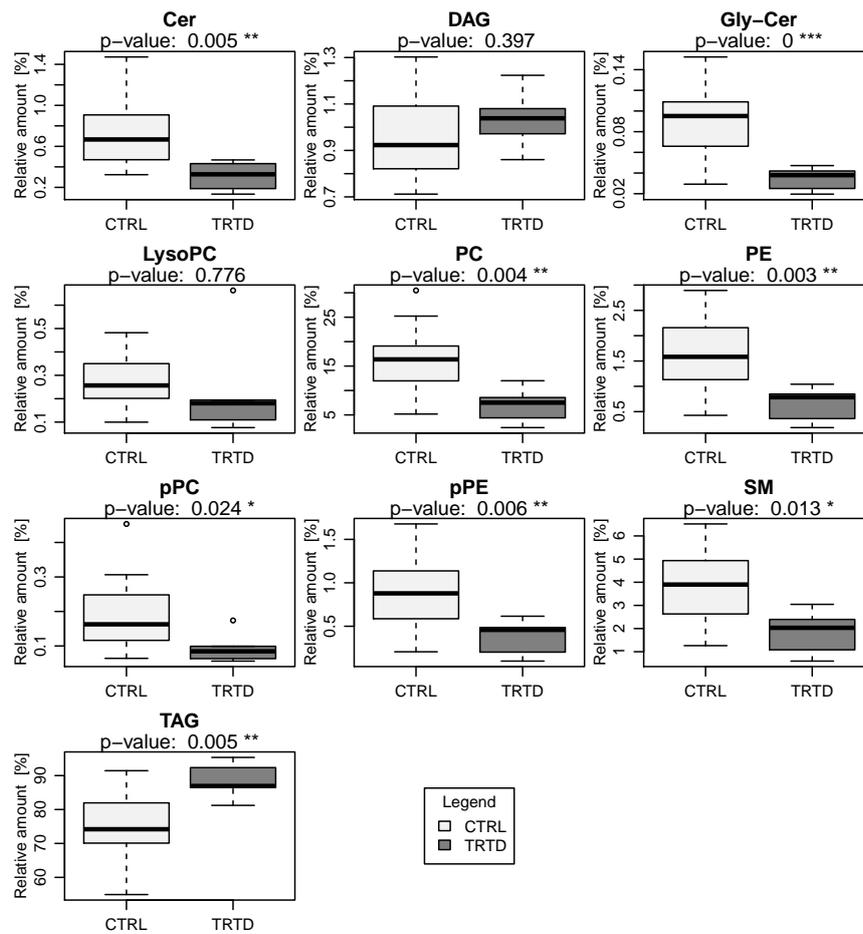


FIGURE 3.4: IGROV-1 – Sample group comparisons in terms of relative amounts of lipid classes. Most of the results are statistically significant (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$).

For a correct understanding of these results, it should be kept in mind that we are comparing relative amounts, implying that the overall sum of lipids in each sample is bound to be constant. As a consequence, if one or more classes increase, the others necessarily have to decrease. In our case, TAGs make up most of the overall signal (roughly 70–90%)¹, and therefore if this class were the only one to truly increase, an apparent decrease would be observed in all the others.

We also performed a principal component analysis (PCA) using such data (269 original variables), and the resulting scores plot is in Figure 3.5. Here, PC1 alone turned out to explain 67% of the total variance. A fair group

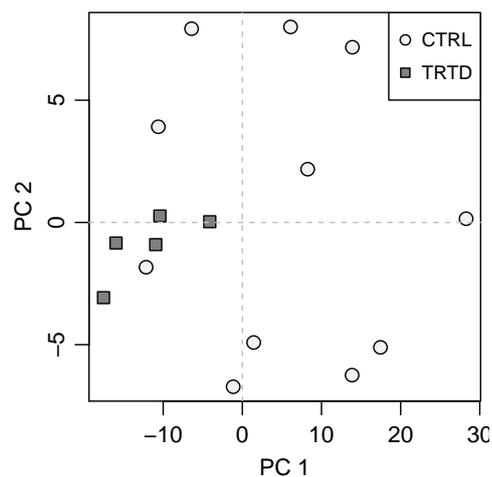


FIGURE 3.5: IGROV-1 – Scores plot from the PCA. The group separation is not clear-cut, although TRTD samples are on the left-hand side along PC1.

separation along PC1 is achieved, where TRTD samples are displaced to the left, while CTRL samples to the right. Not surprisingly, by looking at the loadings (not reported here) it turns out that most of TAG and DAG species have negative values along PC1, whereas the opposite holds for all the other lipid species. This confirms that it is the different relative amount of the lipid classes that contributes the most to the sample group separation; to further verify this, we performed a “restricted” PCA where we used the grouped data (10 original variables, representing the relative amounts of each lipid class), and this is shown in Figure 3.6. Yet again, the sample groups are fairly resolved along PC1; in addition, TAGs and DAGs point in the opposite direction as compared to all the other lipid classes in the loadings plot.

Intra-class Distribution

We then moved on to investigate the intra-class profiles, in order to look for possible intra-class differences caused by the treatment. The molecular distributions are in Appendix C.1, which we obtained by normalising the analyte signals within each lipid class to the overall sum of 100. Table 3.2 summarises the percentage of species for each class where the difference between CTRL and TRTD was statistically significant.

The significance level was set to 5%, meaning that it is expected to find, on average, 5% significant cases just by chance. However, since all the values

¹Of course, we are always referring to relative signal areas and not molar lipid amounts.

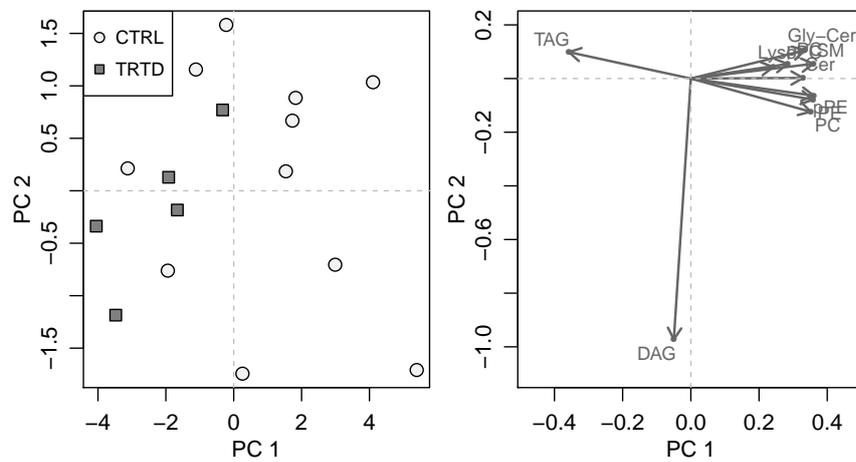


FIGURE 3.6: IGROV-1 – Scores (left) and loadings (right) plots from the PCA on the grouped data (10 original variables representing the relative amount of each lipid class). The sample groups are partially resolved along PC1, where TAG and DAG have negative loadings, in contrast to the positive loadings of all the other lipid classes.

	Cer	DAG	Gly-Cer	LysoPC	PC	PE	pPC	pPE	SM	TAG
↑	0%	45.0%	11.1%	16.7%	4.2%	13.3%	9.1%	11.1%	7.1%	19.4%
↓	5.9%	35.0%	11.1%	16.7%	12.5%	6.7%	9.1%	22.2%	14.3%	56.2%

TABLE 3.2: IGROV-1 – Percentage of species in each class for which the difference between CTRL and TRTD was statistically significant (5% level). The distinction is made between increase (↑) or decrease (↓) from CTRL to TRTD samples.

in the table are higher than such threshold, we can safely state that for each lipid class a true difference between sample groups exists. This is important because it implies that the treatment, in addition to affecting the relative amounts across lipid classes, also changes their intra-class distributions. In order to better characterise the specific effects, we calculated and compared the three different parameters (i) average chain length, (ii) average number of unsaturations, and (iii) unsaturation index. A compact summary of the results is in Table 3.3, where starred arrows indicate significant changes, whereas the obtained box-plots can be found in Appendix C.1.

	Chain length	Unsaturations	Unsaturation index
Cer	--	↑	↑
DAG	↑	↑	↑ *
Gly-Cer	--	↑	↑
LysoPC	↓ **	↑	↑
PC	--	--	--
PE	--	--	--
pPC	↓ *	↓	↓
pPE	↓	--	--
SM	↑	↑ *	↑ *
TAG	↓ **	--	↑ *

TABLE 3.3: IGROV-1 – Summary of the intra-class comparisons. Upward and downward arrows denote higher and lower values, respectively, in TRTD samples. Stars represent the level of significance (* : $p \leq 0.05$, ** : $p \leq 0.01$), whereas missing stars denote non-significant p-values, but still lower than 0.2.

The table provides a picture that is not easily interpreted, although the general trend is for the average chain length to decrease and for the unsaturation degree to increase following the treatment. This picture therefore suggests that lipids in anti-VEGF treated tumours are shorter and more unsaturated than lipids in CTRL tumours, even though the results also depend on the lipid class.

3.3.2 OC-316 Cell Line

Tumour Dimension and Total Signal

In Figure 3.7 are the tumour dimensions, and a significant difference can be seen between sample groups, with treated tumours being the smallest ones. On the other hand, the difference was no more significant when comparing the total MS signal (Figure 3.8), just like for IGROV-1. Again, the lack of a linear relationship between these two quantities was confirmed by an analysis of the sample correlation coefficients ($r = 0.71$ for CTRL, and $r = 0.021$ for TRTD samples).

Relative Distribution Among Lipid Classes

In Figure 3.9 are the inter-class comparisons between sample groups. In contrast to what observed for IGROV-1 cell line, here most of the relative amounts are unaffected by the treatment, and the significant differences only

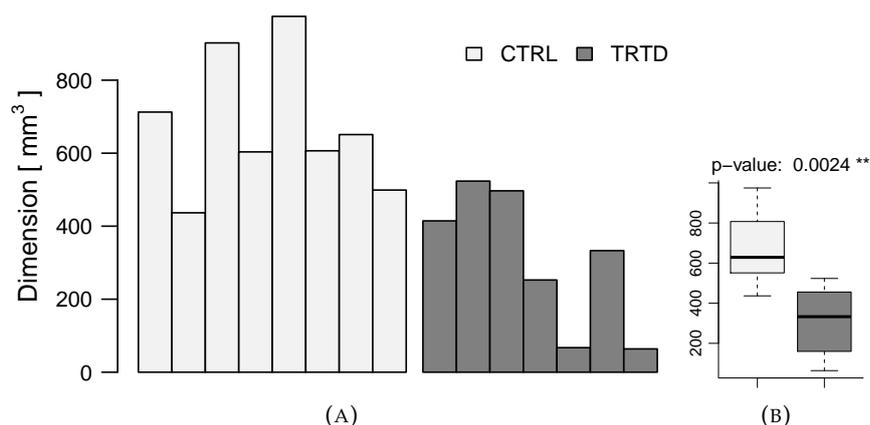


FIGURE 3.7: OC-316 – Bar-plot (A) and box-plot (B) of the tumour dimensions. On average, TRTD tumours are significantly smaller than CTRL tumours.

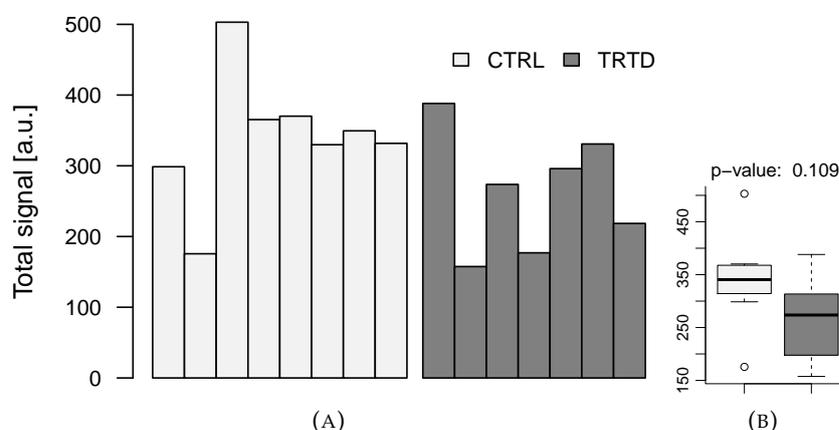


FIGURE 3.8: OC-316 – Bar-plot (A) and box-plot (B) of the total signal areas. The difference between the two groups is not significant at the 5% level.

occur for Gly-Cer and LysoPC, both decreasing in TRTD samples. Also the PCA on the normalised data (335 original variables) failed to separate the two groups, anti-VEGF treated and CTRL tumours, in the scores plot (data not shown).

Intra-class Distribution

The intra-class profiles can be found in Appendix C.2, whereas in Table 3.4 is just a summary of the percentage of significant cases (where the difference between CTRL and TRTD was statistically significant). Interestingly, for all the lipid classes but in TAGs the percentage is well higher than 5%, implying that the treatment deeply affects the intra-class distribution of most classes.

To understand what kind of effects the treatment triggers, we computed and compared average chain length, average number of unsaturations and unsaturation index: Table 3.5 is a compact summary of the results, whereas the details are in Appendix C.2. As for the chain length, in all significant cases (Cer, Gly-Cer, PC, pPC, pPE) lipids are shorter in TRTD samples. The

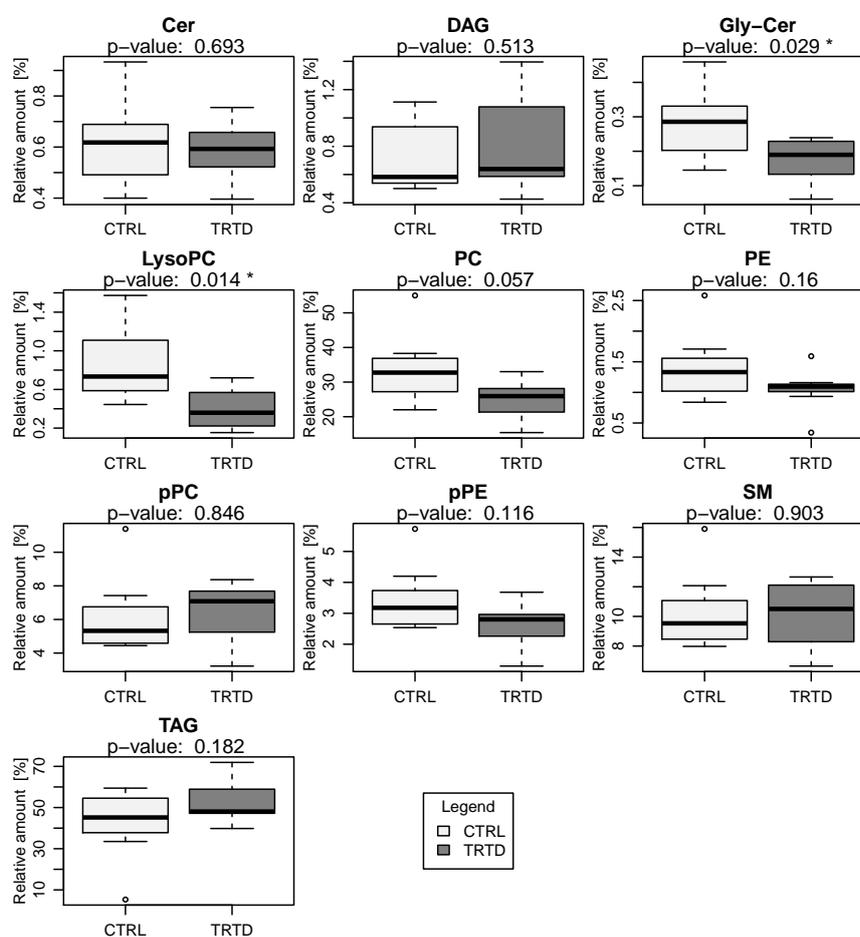


FIGURE 3.9: OC-316 – Inter-class comparisons between sample groups (CTRL VS TRTD).

	Cer	DAG	Gly-Cer	LysoPC	PC	PE	pPC	pPE	SM	TAG
↑	25.0 %	5.0 %	14.3 %	7.1 %	7.3 %	4.8 %	20.6 %	28.0 %	24.0 %	0.0 %
↓	25.0 %	10.0 %	42.9 %	14.3 %	65.5 %	23.8 %	44.1 %	20.0 %	44.0 %	2.6 %

TABLE 3.4: OC-316 – Percentage of chemical species with a statistically significant intra-class difference. The distinction is made between increase (↑) or decrease (↓) from CTRL to TRTD samples.

	Chain length	Unsaturations	Unsaturation index
Cer	↓ *	↓ *	↓
DAG	--	↑	↑ *
Gly-Cer	↓ ***	↑ **	↑ ***
LysoPC	--	--	--
PC	↓ ***	↓ ****	↓ ****
PE	--	--	--
pPC	↓ ***	↓ **	↓ **
pPE	↓ *	--	--
SM	↓	↓	--
TAG	--	--	--

TABLE 3.5: OC-316 – Summary of the intra-class comparison between CTRL and TRTD groups.

situation is more complex as regarding unsaturations and the unsaturation index, since the results vary depending on the lipid class.

3.3.3 OVCAR-3 Cell Line

Tumour Dimension and Total Signal

The comparison of the tumour dimensions is in Figure 3.10, which confirms that the anti-VEGF treatment shrinks the tumours. As previously observed,

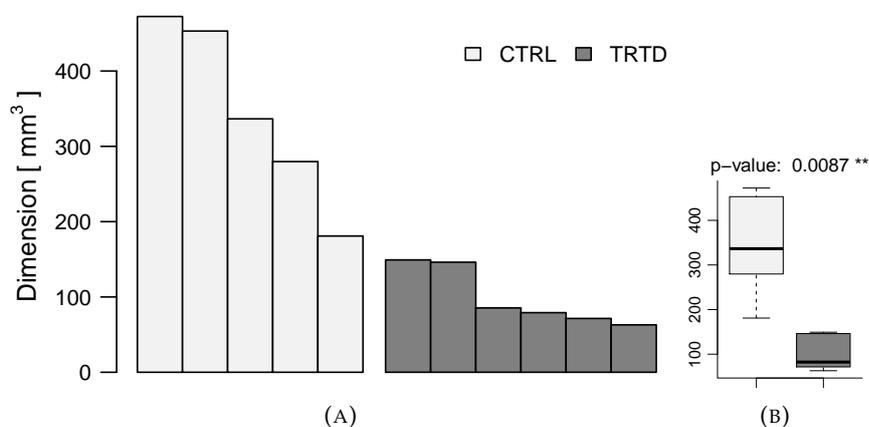


FIGURE 3.10: OVCAR-3 – Bar-plot (A) and box-plot (B) of the tumour dimensions. On average, TRTD tumours are significantly smaller than CTRL tumours.

however, there is no significant group difference in the total MS signal (Figure 3.11); yet again, there is no linear relationship between tumour size and total MS signal ($r = 0.44$ for CTRL, and $r = 0.56$ for TRTD samples).

Relative Distribution Among Lipid Classes

Figure 3.12 shows the inter-class comparisons between sample groups; interestingly, there is a highly significant decrease in Gly-Cer levels, as well as a significant increase in the relative amount of DAG. Worth of note, also in the

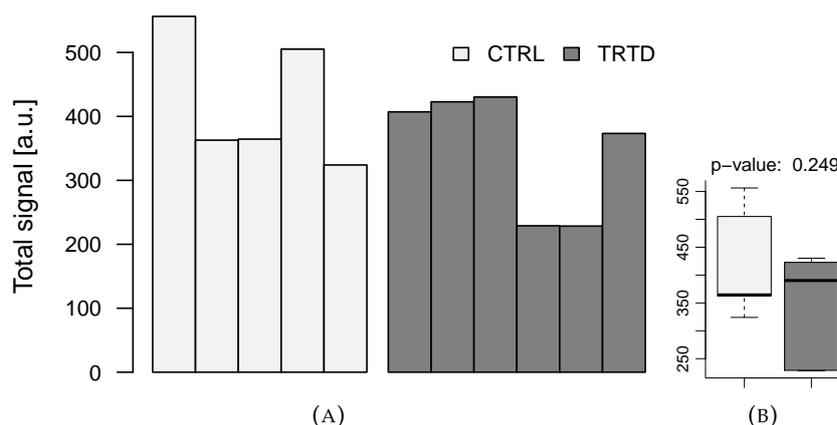


FIGURE 3.11: OVCAR-3 – Bar-plot (A) and box-plot (B) of the total signal areas (after correcting for the DLPC signal). The difference between the two groups is not significant (at the 5 % level).

previous cases the overall Gly-Cer relative amount was significantly lower in TRTD samples. However, a clear and common behaviour for the three cell lines investigated so far cannot be identified.

Intra-class Distribution

The intra-class profiles are detailed in Appendix C.3, whereas here (see Table 3.6) is a compact summary of the percentage of species for which the group difference is statistically significant. Apart from SM and TAG species, our analysis shows that the treatment significantly affects the molecular profiles. The comparison of the parameters average chain length, average

	Cer	DAG	Gly-Cer	LysoPC	PC	PE	pPC	pPE	SM	TAG
↑	4.5 %	0.0 %	30.0 %	28.6 %	4.2 %	7.7 %	7.7 %	25.0 %	0.0 %	1.6 %
↓	4.5 %	11.1 %	10.0 %	14.3 %	10.4 %	15.4 %	7.7 %	16.7 %	0.0 %	3.2 %

TABLE 3.6: OVCAR-3 – Percentage of chemical species with a statistically significant intra-class difference. The distinction is made between increase (↑) or decrease (↓) from CTRL to TRTD samples.

number of unsaturations and unsaturation index is in Table 3.7 and the related box-plots in Appendix C.3. Just like in the previous cases, TRTD samples have lipids which are, on average, shorter than those in CTRL samples. As for the unsaturations, the trend is not clear, although the unsaturation index decreases in most of the species with the notable exception of Gly-Cer, for which it significantly increases (interestingly, a similar behaviour was observed in the previous cases).

3.3.4 SKOV-3 Cell Line

Tumour Dimension and Total Signal

Also in this last case, we show the comparison of both tumour dimensions (Figure 3.13) and total MS signal (Figure 3.14), drawing the same conclusions

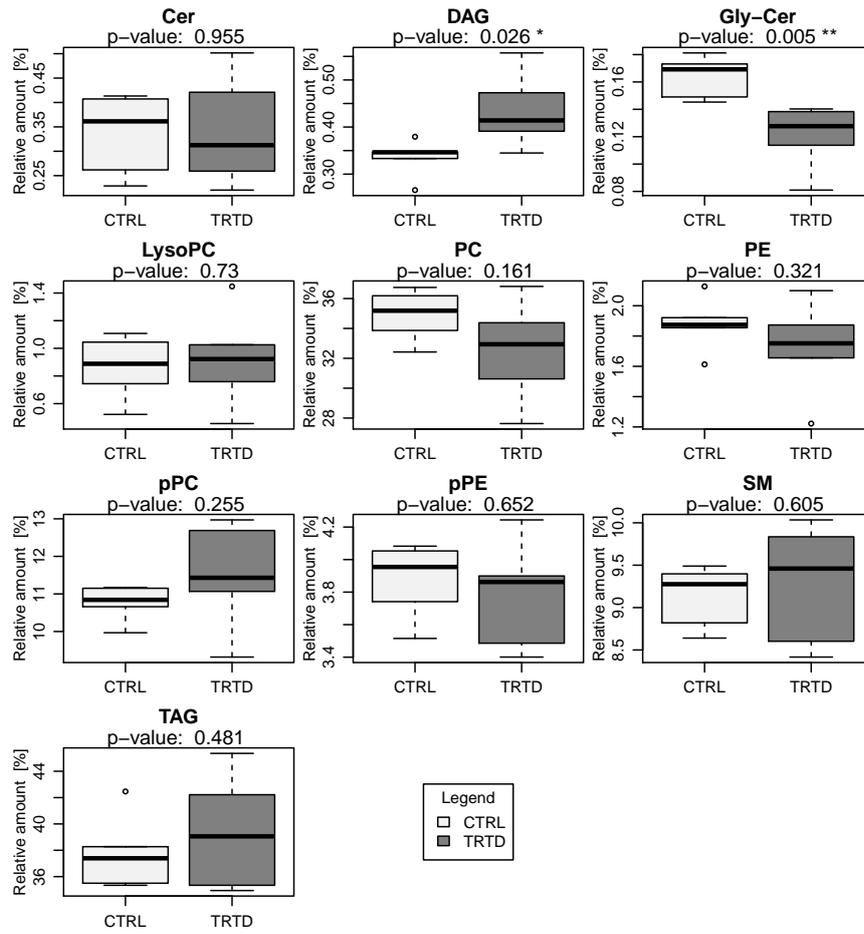


FIGURE 3.12: OVCAR-3 – Inter-class comparisons between sample groups (CTRL VS TRTD).

	Chain length	Unsaturations	Unsaturation index
Cer	--	↑	↑
DAG	--	↓	↓
Gly-Cer	--	↑ *	↑ *
LysoPC	↓ **	--	--
PC	--	↓	↓
PE	--	--	--
pPC	↓	↓ *	↓ *
pPE	↓ ***	↓ **	↓ **
SM	↓ *	↓	--
TAG	--	↓	↓

TABLE 3.7: OVCAR-3 – Summary of the intra-class comparison between CTRL and TRTD groups.

as above: whereas the treatment reduces the tumour size, this is not the case for the total MS signal, related to the total lipid amount. As a consequence, there is no correlation between the two variables ($r = 0.31$ for CTRL, and $r = 0.57$ for TRTD samples).

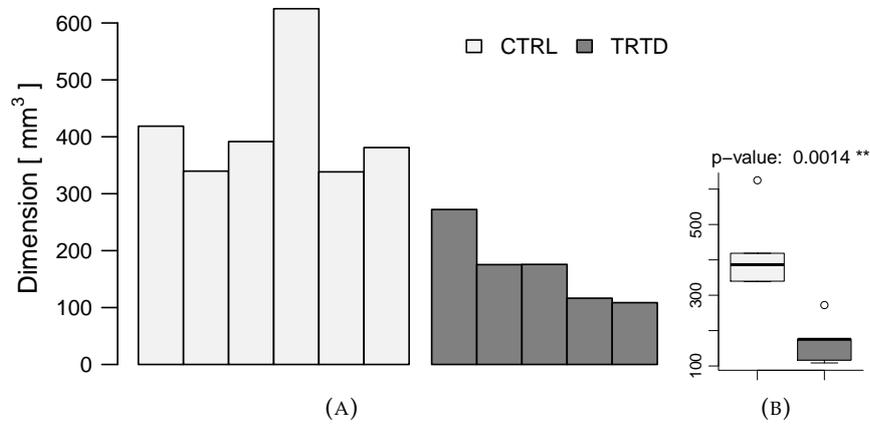


FIGURE 3.13: SKOV-3 – Bar-plot (A) and box-plot (B) of the tumour dimensions. On average, TRTD tumours are significantly smaller than CTRL tumours.

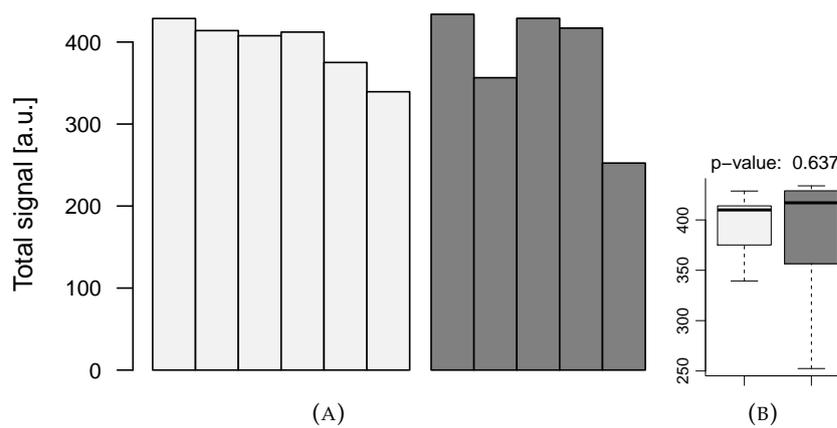


FIGURE 3.14: SKOV-3 – Bar-plot (A) and box-plot (B) of the total signal areas (after correcting for the DLPC signal). The difference between the two groups is not significant (at the 5 % level).

Relative Distribution Among Lipid Classes

As Figure 3.15 shows, the treatment in this case did not have any significant effect on the relative amounts of the investigated lipid classes, as there is not a single significant difference. Therefore, such results represent yet another response to the anti-angiogenic treatment as compared to the previous cell lines.

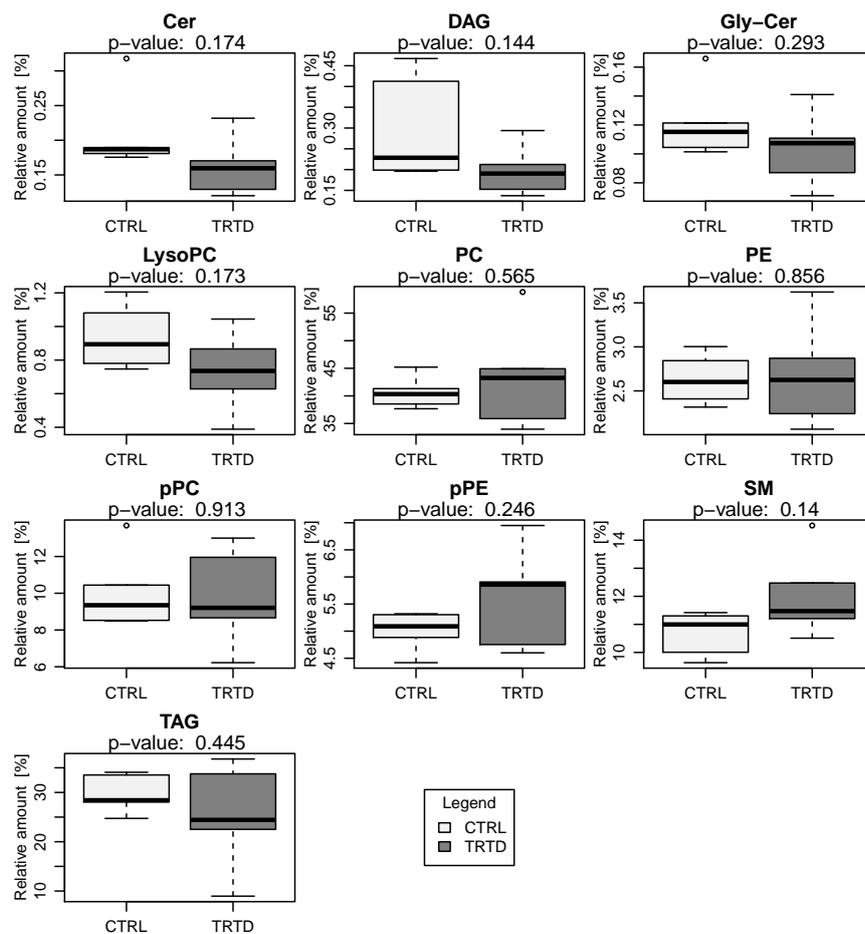


FIGURE 3.15: SKOV-3 – Inter-class comparisons between sample groups (CTRL VS TRTD).

Intra-class Distribution

The intra-class profiles, reported in Appendix C.4, highlight that the treatment does indeed affect the molecular distribution within each class; Table 3.8 displays for how many molecular species were the group differences statistically significant. Interestingly, a feature that seems to be common across the cell lines used is that the intra-class profile of TAG species is not affected by the treatment. On the other hand, we tried to characterise the

	Cer	DAG	Gly-Cer	LysoPC	PC	PE	pPC	pPE	SM	TAG
↑	4.0 %	7.1 %	27.3 %	33.3 %	1.8 %	13.3 %	3.1 %	8.3 %	4.0 %	0.0 %
↓	12.0 %	0.0 %	9.1 %	11.1 %	22.8 %	26.7 %	6.3 %	16.7 %	32.0 %	0.8 %

TABLE 3.8: SKOV-3 – Percentage of chemical species with a statistically significant intra-class difference. The distinction is made between increase (↑) or decrease (↓) from CTRL to TRTD samples.

changes in terms of average chain length, average number of unsaturations and unsaturation index (Appendix C.4 and Table 3.9): once more, TRTD

	Chain length	Unsaturations	Unsaturation index
Cer	↑ *	↑	↑
DAG	--	--	--
Gly-Cer	--	↑ **	↑ ***
LysoPC	↓ *	--	--
PC	↓ **	↓ **	↓ **
PE	--	--	--
pPC	--	↑ *	↑ *
pPE	↓ **	--	--
SM	↓	--	--
TAG	--	--	--

TABLE 3.9: SKOV-3 – Summary of the intra-class comparison between CTRL and TRTD groups.

samples are associated with shorter lipids as compared to CTRL tumours. As for the unsaturations, the general trend is for TRTD lipids to be more unsaturated, although for PCs the opposite holds true, whereas for some other classes the results are not significant.

3.4 Conclusions

The main purpose of the present investigation was to explore the effects of anti-VEGF treatments on the lipidome of cancer cells, and to understand whether such outcomes also depend on the tumour. We were motivated by the fact that (i) lipids are known to play important roles in disease development and progression, and (ii) cancer cells are characterised by abnormalities involving lipids, such as the Warburg effect and a high rate of DNL.

To pursue our aim, we selected four different human ovarian adenocarcinoma cell lines (Table 3.1), and we compared the lipidome of anti-VEGF

treated and untreated tumours in female mice. Quite surprisingly, in all four cases we found significant differences between sample groups, although the effects of the treatment turned out to depend on the cell line itself.

1. **IGROV-1.** In terms of tumour dimensions, we found a significant reduction upon treatment (Figure 3.2). However, the total LC-MS signal did not follow this trend, since the difference was actually not significant (Figure 3.3). This seems to suggest that the overall lipid amount does not increase linearly with the tumour dimension, which was also confirmed by a regression analysis.

When examining the inter-class distributions (Figure 3.4), we observed significant group differences characterised by (i) a higher relative amount of TAGs and DAGs in TRTD tumours as compared to controls, and (ii) the opposite trend for all the other lipid classes. However, it should be considered that we looked at relative amounts, which implies that if some species increase, the others are bound to decrease to keep the overall amount constant. The PCA performed using the relative amounts as original variables (scores plots of Figure 3.5 and Figure 3.6) confirmed the significant difference in terms of lipid class composition.

As for TAGs, they make up most of the total MS signal (70–90%), and their increase upon treatment is worthy of attention. However, we can give only putative biological explanations to this finding: our hypothesis is that such results are indicative of an accumulation of lipid droplets within the tumours. This is interesting because lipid droplets are known to play important roles in cancer progression, and they are usually abundant in diseased states [18, 31–33]. However, it is not clear to us the reason why anti-VEGF treatments would trigger their further accumulation in cells; we speculate that this may be the result of a phenotypic shift of the tumour cells into a more aggressive tumour form, as previously reported [23].

In addition to quantitative variations among lipid classes, we also observed significant differences in their intra-class compositions (Appendix C.1 and Table 3.3): the general trend is for anti-VEGF treated samples to be made of shorter and more unsaturated lipids as compared to CTRL tumours, even though the results are class dependent.

2. **OC-316.** Also here, treated tumours were significantly smaller than controls (Figure 3.7), but the dimension did not correlate well with the total signal intensity (Figure 3.8). Unexpectedly, here we did not find the marked inter-class perturbation observed for IGROV-1. Rather, in this case there only was a significant reduction in Gly-Cer and LysoPC levels (Figure 3.9).

As for the intra-class compositions, we noticed dramatic effects triggered by the treatment (Appendix C.2 and Table 3.5), in that treated lipids turned out to be, on average, significantly shorter than those in CTRL tumours. On the other hand, the trend in the degree of unsaturation turned out not to be the same for all lipid classes: in some cases (DAG, Gly-Cer) it increased whereas in others (Cer, PC, pPC) it decreased as a result of the treatment.

3. **OVCAR-3.** Just like in the previous cases, the anti-VEGF treatment significantly reduced the tumour size (Figure 3.10), although the dimension did not correlate well with the total MS signal (Figure 3.11). In terms of class composition, there was a dramatic drop of Gly-Cer levels, accompanied by an increase in DAG species (Figure 3.12).

As for the intra-class compositions, significant results were observed (Appendix C.3 and Table 3.7), including a pronounced reduction in chain length of many lipid classes. On the other hand, the effects on the unsaturation index were contrasting, since for some classes (DAG, PC, pPC, pPE and TAG) there was a decrease, whereas for some others (Cer and Gly-Cer) an increase.

4. **SKOV-3.** Also with this cell line the treatment induced a significant reduction in tumour dimensions (Figure 3.13), not well related with the total MS signal (Figure 3.14). Unlike for the other cell lines, here the inter-class profiles of CTRL and TRTD samples were quite similar, with no significant differences (Figure 3.15). On the other hand, what did significantly change were the molecular compositions within each class: for many of them, the lipids were shorter in anti-VEGF treated samples. As for the saturations, the effects were contrasting, depending on the lipid class being investigated: (i) no effects for DAG, LysoPC, PE, pPE, SM and TAG, (ii) increase in unsaturation index for Cer, Gly-Cer, pPC, and (iii) decrease in unsaturation index for PC.

The results just described above, although interesting, are quite difficult to properly interpret on a biological basis. Astonishingly, the effects caused by the treatments seem to strongly depend on the cell line under investigation. This is the case, for example, of the different inter-class responses: whereas for IGROV-1 there was a clear increase in the relative amounts of TAGs and DAGs, which we speculated coming from an accumulation of lipid droplets, such behaviour did not repeat itself for the other cell lines. In particular, for OC-316 there only was a decrease in Gly-Cer and LysoPC, for OVCAR-3 there was an increase for DAGs and a decrease in Gly-Cer, whereas for SKOV-3 there was no significant change at all.

In view of such disparate results, it is also worth outlining a lack of any simple correlation between glycolytic phenotype and lipidomics effects of the treatment. If it were so, we would have observed common and pairwise changes, since OC-316 and OVCAR-3 are both highly glycolytic, whereas IGROV-1 and SKOV-3 poorly glycolytic tumours.

Interestingly though, there also are some common effects among the four investigated cell lines. The first one is the tumour size reduction caused by the treatment which, in all cases, did not correlate with the the total MS signal (it did not drop significantly upon treatment). Since such quantity is closely connected with the total lipid amount, these results suggest that tumours are not homogeneous inside in terms of lipid composition. As a matter of fact, intra-tumour heterogeneities have been described and are known to be related to different local microenvironments [34].

A second common feature that can be observed is the general reduction in average chain length of many lipid classes in TRTD samples. Since lipids can be either taken up from the diet or synthesised *de novo*, the observed effect may be due to an altered relative contribution of these two FA sources.

Finally, the third common feature involves Gly-Cer: in all cases, the treatment induces a shift towards more unsaturated species. Considering the pivotal role played by sphingolipids in health and disease, we believe that such perturbation is not to be underestimated, even though we do not know how to biologically explain our findings.

In summary, the picture emerging from our analysis is complex and difficult to properly interpret, not only because there are marked differences among cell lines in terms of the response to the anti-VEGF treatment, but also because a straight cause-effect connection cannot be easily drawn. Taken together our results show that, no matter of the tumour cell line of origin, the treatment does cause significant perturbations in terms of lipidomic profiles, although such effects are tumour-dependent. In a sense, this is not surprising if we consider that anti-angiogenic treatments do not give universal results when used; our results may therefore merely reflect different biological responses to such treatments: here we started with four different sets of samples that, because of the different origin of the tumours, were characterised by significantly different lipidomic profiles. And the results were quite different, as already discussed. It is also worth remembering that the effectiveness of anti-VEGF drugs has been suggested to be curbed by the intrinsic heterogeneity of tumour angiogenesis, and several animal trials have shown that different tumours respond very differently to anti-VEGF therapy [35].

A possible hypothesis that may explain our results is the following: the changes in lipid composition that we see are the consequence of a selection of tumour cells following the use of bevacizumab. Indeed, the initial idea for anti-VEGF therapy was that, in principle, such drugs are less vulnerable to drug resistance, in that their target is not the highly genetically mutable cancer cell population (which can rapidly acquire resistance to any drug treatment), but rather the vascular endothelial cells, which are a genetically stable host cell population. Unfortunately, this is not the case, since anti-VEGF treated tumour often show acquired resistance, which is thought to involve a selection of tumour cell subpopulations capable of surviving in the highly hypoxic tumour environments created by the anti-VEGF therapy [36, 37]. As a consequence, since different tumour cells will have different metabolisms, we speculate that the differences observed in terms of lipid composition between CTRL and TRTD tumours are related to such selection.

Of course, we do recognise that further research is still needed to properly understand and interpret these phenomena, and we believe that lipidomics may be one of the tools needed to help elucidate the mechanisms behind anti-angiogenic treatments. At the same time, we are also well aware that lipidomics alone is unlikely to solve these complex issues, and therefore our investigations will have to be properly integrated with other available methodologies to shed some light into the matter. For example, it may be useful to look for altered expression of some key enzymes involved in the biosynthesis of lipids more strongly affected by the treatments. Another possible approach may be to use electron microscopy to validate our hypothesis that lipid droplets play a role here, too. In summary, we acknowledge that the results discussed here are just preliminary, and further work has to be done if we really want to understand the effects that anti-angiogenic treatments have on cancers.

Bibliography of Chapter 3

1. Ma, X. & Yu, H. Global burden of cancer. *Yale Journal of Biology and Medicine* **79**, 85–94 (2006).
2. Jemal, A. *et al.* Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**, 69–90 (2011).
3. American Cancer Society. *Global Cancer Facts & Figures, 3rd Edition* (American Cancer Society, Atlanta, 2015).
4. Lodish, H., Berk, A. & Zipursky, S. L. *Molecular cell biology – 4th edition* (W. H. Freeman, New York, 2000).
5. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
6. Warburg, O. On the origin of cancer cells. *Science (New York, N.Y.)* **123**, 309–14 (1956).
7. Zhang, F. & Du, G. Dysregulated lipid metabolism in cancer. *World Journal of Biological Chemistry* **3**, 167–174 (2012).
8. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Disease Models & Mechanisms* **6**, 1353–1363 (2013).
9. Santos, C. R. & Schulze, A. Lipid metabolism in cancer. *The FEBS Journal* **279**, 2610–2623 (2012).
10. Rysman, E., Brusselmans, K., Scheys, K. & Timmermans, L. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Research* **70**, 8117–8126 (2010).
11. Ntambi, J. M. Regulation of stearoyl-CoA desaturase by polyunsaturated fatty acids and cholesterol. *Journal of Lipid Research* **40**, 1549–58 (1999).
12. Minville-Walz, M. *et al.* Inhibition of stearoyl-CoA desaturase 1 expression induces CHOP-dependent cell death in human cancer cells. *PLoS One* **5**, e14363 (2010).
13. Chung, A. S. & Ferrara, N. Developmental and pathological angiogenesis. *The Annual Review of Cell and Developmental Biology* **27**, 563–584 (2011).
14. Ackerman, D. & Simon, C. Hypoxia, lipids, and cancer: surviving the harsh tumor microenvironment. *Trends in Cell Biology* **24**, 472–478 (2014).
15. Krock, B. L., Skuli, N. & Celeste Simon, M. Hypoxia-induced angiogenesis: good and evil. *Genes & Cancer* **2**, 1117–1133 (2011).
16. Bensaad, K., Favaro, E., Lewis, C. A. & Peck, B. Fatty acid uptake and lipid storage induced by HIF-1 α contribute to cell growth and survival after hypoxia-reoxygenation. *Cell Reports* **9**, 349–365 (2014).

17. Tirinato, L., Liberale, C., Di Franco, S., Candeloro, P. & Benfante, A. Lipid droplets: a new player in colorectal cancer stem cells unveiled by spectroscopic imaging. *Stem Cells* **33**, 35–44 (2015).
18. Bozza, P. T. & Viola, J. P. B. Lipid droplets in inflammation and cancer. *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)* **82**, 243–250 (2010).
19. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nature Clinical Practice Oncology* **3**, 24–40 (2006).
20. Ferrara, N., Hillan, K. J., Gerber, H. P. & Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nature Reviews Drug Discovery* **3**, 391–400 (2004).
21. Kamba, T. & McDonald, D. M. Mechanisms of adverse effects of anti-VEGF therapy for cancer. *British Journal of Cancer* **96**, 1788–1795 (2007).
22. Quintieri, L., Selmy, M. & Indraccolo, S. Metabolic effects of antiangiogenic drugs in tumors: therapeutic implications. *Biochemical Pharmacology* **89**, 162–170 (2014).
23. Curtarello, M., Zulato, E., Nardo, G. & Indraccolo, S. VEGF-targeted therapy stably modulates the glycolytic phenotype of tumor cells. *Cancer Research* **75**, 120–133 (2015).
24. Fack, F. *et al.* Bevacizumab treatment induces metabolic adaptation toward anaerobic metabolism in glioblastomas. *Acta Neuropathologica* **129**, 115–131 (2015).
25. Sounni, N. E. *et al.* Blocking lipid synthesis overcomes tumor regrowth and metastasis after antiangiogenic therapy withdrawal. *Cell Metabolism* **20**, 280–294 (2014).
26. Pignata, S. *et al.* Chemotherapy in epithelial ovarian cancer. *Cancer Letters* **303**, 73–83 (2011).
27. Kehoe, S. Treatments for gynaecological cancers. *Best Practice & Research Clinical Obstetrics & Gynaecology* **20**, 985–1000 (2006).
28. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911–917 (1959).
29. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using non-linear peak alignment, matching, and identification. *Analytical Chemistry* **78**, 779–787 (2006).
30. R Development Core Team. *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
31. Abramczyk, H. *et al.* The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue. *Analyst* **140**, 2224–2235 (2015).
32. Welte, M. A. Expanding roles for lipid droplets. *Current Biology* **25**, R470–R481 (2015).

33. Stephens, N. A. *et al.* Intramyocellular lipid droplets increase with progression of cachexia in cancer patients. *Journal of Cachexia, Sarcopenia and Muscle* **2**, 111–117 (2011).
34. Junttila, M. R. & de Sauvage, F. J. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* **501**, 346–354 (2013).
35. Bergers, G. & Benjamin, L. E. Angiogenesis: tumorigenesis and the angiogenic switch. *Nature Reviews Cancer* **3**, 401–410 (2003).
36. Bertolini, F. Anti-angiogenesis in cancer; met and unmet goals – An interview with Robert Kerbel. *The International Journal of Developmental Biology* **55**, 395–398 (2011).
37. Kerbel, R. S. Tumor angiogenesis. *New England Journal of Medicine* **358**, 2039–2049 (2008).

Chapter 4

Isotopic Labelling: an R Package for the Analysis of Enriched MS Isotopic Patterns

4.1 Introduction

Biological Labelling Experiments

Metabolomics and lipidomics have been witnessing an increasing interest from biological and biomedical research scientists, especially when it comes to their applications to medical diagnostics. However, because of the complexity of the metabolome/lipidome of cells and tissues, these sciences have to face many issues; among them is the elucidation of metabolic pathways, not readily achievable on account of challenges associated to directly measuring metabolic fluxes. Some limitations can be overcome with the aid of stable isotopes, used to label metabolites in living cells.

In the past, labelling experiments used to employ radioactive tracers, which were detected by highly sensitive scintillation counters. More recently, stable isotopes have replaced radioactive tracers due to improved sensitivities of NMR and MS analytical techniques. Accordingly, isotopic labelling approaches have been developed for a variety of applications including metabolite identification, quantification, and pathway analysis [1–4].

When isotopic labelling is used to study conversion and transport rates of metabolites in cells, it is referred to as fluxomics [5]. Here, the by far most popular label is ^{13}C , given that (i) every bioorganic molecule contains carbon, and (ii) carbon transfer in biochemical reactions follows defined rules. Another common stable isotope is ^2H [6–8], even though caution must be paid with it since ^2H is interchangeable with hydrogen from water.

A typical labelling experiment starts by feeding cell cultures or animals with substrates enriched in the label used. After the growth period, the metabolites are extracted and the samples are analysed. The detection of the label (^2H , ^{13}C , ...) by MS is based on the fact that its mass is different from the one of the isotope most common in nature (^1H , ^{12}C , ...), whereas NMR can directly detect it.

Although MS-based labelling experiments are more common than NMR-based experiments, the potential advantages of this technique over MS should not be underestimated. In particular, (i) it can provide information on specific positions of the label within analytes and (ii) it is not invasive and therefore enables *in vivo* studies [9]. However, in my project I only dealt

with MS-based labelling experiments, and so I will focus on this technique alone in the following.

Isotopes and Mass Spectrometry

Isotopes and MS are historically connected, both starting at the end of the 19th century with the publications by J. J. Thomson, who reported his results on cathode rays [10]. In 1907, he also described a parabola mass spectrograph that he used to measure the m/z of ions coming from different gases; interestingly, while using neon he observed that the main signal, at m/z 20, was always accompanied by a weaker one at m/z 22 [11]. In the following years his assistant, F. W. Aston, was able to describe the isotopic compositions of several elements including Ar, Kr, Hg and Cl [12], and this represented the beginning of the research on isotopes.

Nowadays we know that an isotope is a form of an element whose atomic nucleus contains a specific number of neutrons, in addition to the protons that define the element. Isotopes differ in their nucleon number, the sum of protons and neutrons, which is usually written as a superscript preceding the symbol of the element, as in ^1H for defining the isotope of H having just one nucleon. Elements can be monoisotopic or poly-isotopic depending on whether they exist as one or more naturally occurring stable isotopes, respectively. As a consequence, molecules are mixtures of isotopes, which is the reason why in chemistry atomic and molecular weights are used.

MS identifies ions by their m/z , and therefore it can easily distinguish and quantify the different isotopic variants of a species; to avoid ambiguities, a specific MS terminology has been devised [13, 14]:

Isotopic mass: it is the mass of an isotope, measured in unified atomic mass units (u). 1 u corresponds to 1/12 of the mass of one ^{12}C atom at rest and in its ground state;

Mass defect: it is the difference between the isotopic mass and the sum of the individual masses of the protons and neutrons making up an isotope. Consequently, the exact mass of a molecule depends on its specific isotopic composition [15];

Atomic weight: it is the weighted average obtained by taking into account all naturally occurring isotopes of an element. A similar definition applies to molecular weight;

Monoisotopic mass: it is the exact mass of a molecular species obtained using the mass of the most abundant isotope of each element. For common organic compounds, the most abundant isotopes are also the lightest ones (^1H , ^{12}C , ^{16}O , ^{14}N , ...);

Isotopologues: they are species with the same elemental composition but different isotopic composition, such as $^{12}\text{CH}_4$ and $^{13}\text{CH}_4$;

Isotopomers: they are species of identical isotopic composition, differing only in the position of the isotopes. They cannot be distinguished by MS, no matter of the resolving power of the instrument;

Isotopic molecular ion (M+1, M+2, ...): it is an ion containing one (M+1) or more (M+2, ...) of the less abundant isotopes of the atoms making up the species.

Prediction of MS Isotopic Patterns

Isotopes cause molecules to be present in nature as mixtures of isotopologues, and therefore any analyte observed by MS will display a specific isotopic pattern where each MS peak represents a certain isotopologue. As a consequence, MS isotopic patterns are a sort of elemental fingerprint of the analytes themselves.

The natural isotopic distributions of the elements are well known, and several algorithms have been devised for predicting MS isotopic patterns: the oldest approaches are based on probability theory, whereas the newest ones rely on the Fourier Transform (FT). Here I will just give a quick overview of some of them, including the one that I exploited during my PhD.

The Multinomial Expansion. The most classical and intuitive approach makes use of the multinomial expansion [14]: here, for the generic molecule $A_m B_n C_o$, the first step is to separately consider each of its elements (A, B, C), and to find out its relative isotopic distribution through the expansion of the multinomial [16, 17],

$$(a_1 + a_2 + \dots + a_k)^N = \sum \frac{N!}{n_1! n_2! \dots n_k!} a_1^{n_1} a_2^{n_2} \dots a_k^{n_k} \quad (4.1)$$

where n_i is the number of atoms of isotope "i" of the considered element. The sum is taken over all combinations of n_1, n_2, \dots, n_k for which the overall sum is N, the total number of atoms of the considered element in the molecule. Those on the right-hand side are probability-like terms; basically, each of them represents the probability of occurrence of a certain isotopologue.

Just to give a practical example, for C in $C_{42}H_{82}NO_8P$ the above expression would simplify to

$$(a_{12C} + a_{13C})^{42} = \sum_{i=0}^{42} \frac{42!}{i!(42-i)!} (a_{12C})^i (a_{13C})^{42-i} \quad (4.2)$$

where a_{12C} and a_{13C} are the ^{12}C and ^{13}C natural abundances, respectively. The resulting isotopic distribution is in Figure 4.1, where along the horizontal axis is the number of ^{13}C atoms in the species C_{42} .

After having repeated this procedure for all the elements in the target analyte, the probability of occurrence of a particular variant of the molecule is obtained by multiplying the probabilities for the corresponding element-specific variants [17]. This step can become computationally demanding, especially for big molecules; this is the reason why in the past it was usual to employ the "pruning" technique, whereby variants with an occurrence falling below a specified threshold were not taken into consideration.

Fourier Transform and the Convolution Method. Owing to the mass defect, isobaric isotopologues have slightly different masses, but they can be resolved from one another by MS only if the instrumental resolution is

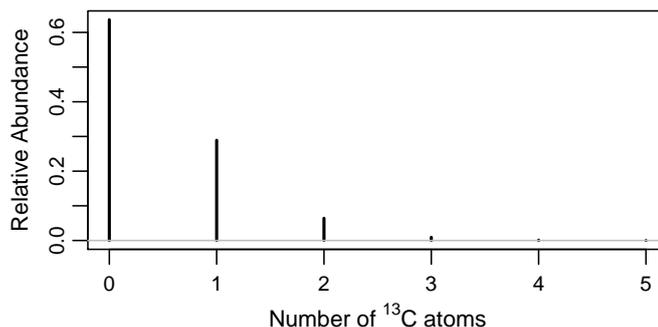


FIGURE 4.1: Expected isotopic distribution for C₄₂ as obtained from a multinomial expansion. Along the x axis is the number of ¹³C atoms in C₄₂, whereas on the y axis is the probability of occurrence. Only the lightest isotopologues are shown, since the occurrence of heavier species is negligible.

high enough. This is not the case for routine mass measurements, implying that such isotopologues are usually grouped together into aggregates with a centre of mass given by a probability-weighted sum of the exact masses of the individual components. The resulting aggregates are separated from each other by about 1 u and so, instead of using their masses, they can be named after the additional neutron content with respect to the monoisotopic variant. For example, the multinomial expansion for the molecule A_mB_nC_o would assume the simple form

$$\left(a_1 I^0 + a_2 I^1 + \dots\right)^m \left(b_1 I^0 + b_2 I^1 + \dots\right)^n \left(c_1 I^0 + c_2 I^1 + \dots\right)^o \quad (4.3)$$

where the superscripts 0, 1, ... specify the additional neutron content. With this approach it is easier to expand and multiply the polynomials, but its very advantage has to do with the FT and the convolution theorem, as pointed out by Alan Rockwood between 1995 and 1996 [18–20]: in summary, he replaced the indicators I^0, I^1, \dots by the Fourier terms $e^{i2\pi m\mu}$, where i is the imaginary number, m is the nominal mass, and μ is a real number between 0 and 1 defining a grid over which the Fourier term is sampled. Equation (4.3) now becomes

$$\left(a_1 e^{i2\pi m a_1 \mu} + a_2 e^{i2\pi m a_2 \mu} + \dots\right)^m \left(b_1 e^{i2\pi m b_1 \mu} + b_2 e^{i2\pi m b_2 \mu} + \dots\right)^n \cdot \left(c_1 e^{i2\pi m c_1 \mu} + c_2 e^{i2\pi m c_2 \mu} + \dots\right)^o \quad (4.4)$$

If it were expanded, each term would represent peak intensity multiplied by a complex exponential containing information on the related nominal mass. However, the above expression can be seen as $f(\mu)$, a function of the parameter μ , and therefore it can be Fourier-transformed into $F(m)$, a function representing a discretely sampled version of the isotopic distribution in the mass domain [19]. To sum up, Rockwood's approach is based on the following steps:

1. Choose a peak shape function, $S(m)$, and inverse-transform it: $s(\mu) = \text{IFT}[S(m)]$;

2. Multiply $s(\mu)$ with $f(\mu)$;
3. Fourier-transform $s(\mu)f(\mu)$ with respect to μ to get to the mass domain:
 $F(m) = \text{FT}[s(\mu)f(\mu)]$.

Importantly, this method is characterised by a reduced computational demand, and therefore no pruning technique is needed. The algorithm could even be used to resolve the isotopic fine structure, in this case exact masses have to be used instead of nominal masses, and the width of the peak shape has to be carefully adjusted, as well as the sampling grid represented by μ .

New Approaches. Since Rockwood, many other algorithms have been devised and fine-tuned to compute the isotopic distributions with ever-increasing efficiencies. In particular, considering that the resolution of modern mass spectrometers is getting higher and higher, the focus has been shifting toward methods able to efficiently resolve the isotopic fine structure. Among these are two-dimensional based FFT methods, that split the calculation into two separate dimensions: a first one is used for getting a low resolution picture, whereas a second FT focuses in the surroundings of the isotopic variants to obtain the details about the fine structure [21, 22].

A very efficient method has been described not long ago, whereby the geometric structure called k -simplex is exploited to describe the distribution of neutrons among the isotopologues of groups of distinct elements [23]. As discussed in the original paper, during a first step each group of elements is considered, and the isotopic patterns are individually computed through an FT applied on the simplex geometric structure. The isotopic pattern of the whole molecule is then obtained as the outer product of all the separate isotopic patterns. The exact masses of each species are similarly obtained, but the products are replaced by sums, and the probabilities by exact isotopic masses. I mentioned this algorithm here because the authors have implemented it in the R package *ecipex*, which I have used during the development of my own R package, *IsotopicLabelling*.

De Novo Lipogenesis (DNL)

One of the biological applications of *IsotopicLabelling* was to investigate the *de novo* lipogenesis (DNL) process, and therefore I would also like to briefly introduce it here before starting with the technical description of the package itself.

DNL is a metabolic pathway whereby fatty acids (FAs) and lipids are synthesised starting from carbohydrates; this process mainly takes place in the liver and in the adipose tissue. In DNL, citrate from the tricarboxylic acid (TCA) cycle is converted to acetyl-CoA, which is polymerised to form FAs: the main DNL product is palmitic acid (16:0), which can be desaturated (SCD enzyme) and/or elongated to palmitoleic acid (16:1), stearic acid (18:0), oleic acid (18:1) and other minor FAs (Figure 4.2).

DNL is known to play important roles in health and disease, and some of its deregulations have been associated with several pathological conditions, including insulin resistance, non-alcoholic fatty liver disease (NAFLD), obesity, viral infections and malignant transformation of normal cells [24–30]. Also diet plays a role, as DNL links carbohydrates and fats, the two pivotal forms of chemical energy for our organism [31]; in healthy humans, DNL is

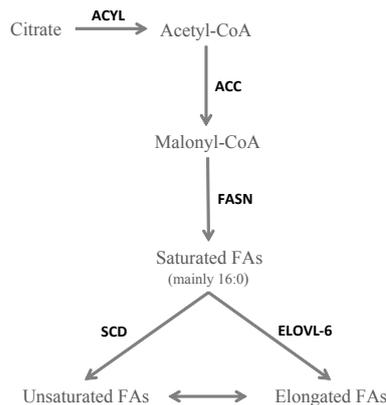


FIGURE 4.2: Exemplified scheme of the DNL metabolic pathway.

a minor source of triacylglycerols (TAGs) as compared to dietary sources. However, some studies suggest that the contribution of DNL becomes significant in individuals on very high carbohydrate diets, which saturate the glycogen reserves and activate lipogenesis [24, 32]. During this process, very low-density lipoproteins (VLDL) are secreted. It was also shown that simple sugars are more effective than complex carbohydrates in stimulating DNL, whereas polyunsaturated fatty acids (PUFAs) can decrease the DNL rate by suppressing gene expression of FASN and SCD enzymes (see Figure 4.2) [33, 34].

Considering the importance of DNL for health, many techniques have been developed to assess its rate, including indirect calorimetry and tracer studies using either ^{13}C or ^2H . In the early 1990s, the method called “mass isotopomer distribution analysis” was developed, based on monitoring the incorporation rate of ^{13}C -labelled acetate into VLDL-palmitate synthesis [35, 36]. Alternatively, the biosynthetic rate can be estimated from the incorporation of deuterated water derived from total plasma water pool into triglycerides [37], although this is technically challenging in that the huge adipose TAG pool size dilutes the newly synthesised TAGs, worsening the sensitivity of the technique.

4.2 Development of the *IsotopicLabelling* R Package

4.2.1 Motivation

When stable isotopes are used to label metabolites, it is important to be able to assess their amount of incorporation within target analytes. Our idea is that such information can be retrieved by proper inspection of MS isotopic patterns, since they strongly depend on the isotopic abundances of the elements. Figure 4.3 is just a simulated example (phosphatidylcholine adduct $[\text{PC } 32:1 + \text{H}]^+$, ^{13}C labelling) to show how the patterns are affected when such distribution is changed: as the ^{13}C abundance increases, there takes place a progressive m/z shift of the most intense signal towards higher m/z values.

The example clearly shows that isotopic patterns convey labelling information; some software packages are already available for dealing with

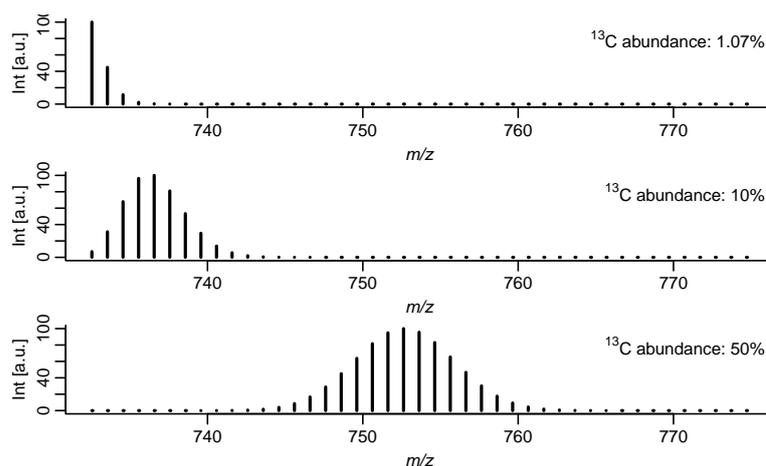


FIGURE 4.3: Simulated MS isotopic patterns for $[\text{PC } 32:1+\text{H}]^+$, assuming different ^{13}C abundances. The underlying assumption here is that the label is uniformly distributed over the C sites of the considered molecule, with the specified average abundance.

labelling-related MS experiments [38–41], but a solution to directly get an estimate of the label abundance is still lacking. This is why we tried to fill this gap by developing *IsotopicLabelling*, an informatics tool that can be used to assess the isotopic abundance of the label in specified metabolites, directly from the results of MS-based experiments. As for the informatics language, we chose the R environment for statistical computing [42], considering that it is widely used by the scientific community. In addition, any R package is freely available and it can be coupled with others, therefore allowing the implementation of more and more complex pipelines.

4.2.2 Working Principles

The MS isotopic pattern of any compound is completely defined once the isotopic distribution of all of its elements is known. In a ^2H - or ^{13}C -based labelling experiment, the only variable is the relative abundance of the label, that we call “X” for convenience; therefore, an isotopic pattern can be seen as a function of X (see Figure 4.4, arrow “a”), since different X values give rise to different patterns.

This line of reasoning can be turned around: since to any relative abundance of the label there corresponds a specific isotopic pattern (for any fixed molecular formula), from an experimental pattern measured by MS it should therefore be possible to estimate the relative isotopic abundance of the labelling element that gives rise to the observed pattern. To put it briefly, this is what the *IsotopicLabelling* R package aims to do, and this is achieved through a fitting procedure whereby X is iteratively changed (Figure 4.4, arrow “c”) and the corresponding patterns are compared to the experimental one (Figure 4.4, arrow “b”), until the “best match” is found.

Uniform Labelling

The package works under the assumption that the target analyte is present in the sample as a homogeneous pool, characterised by a single (unknown)

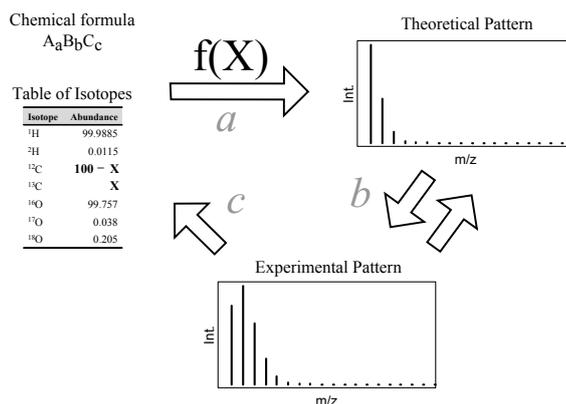


FIGURE 4.4: Exemplified scheme showing how *IsotopicLabelling* works. In this example, a ^{13}C -based experiment is considered.

abundance of the label. This is not the case when, for example, labelled analytes are spiked in the biological sample, or when different biological samples are pooled together, which is a rather common quality-control practice. Therefore, *IsotopicLabelling* cannot handle such cases, since it fits a single parameter.

4.2.3 The Processing Steps

A compact block scheme of the *IsotopicLabelling* processing steps is reported in Figure 4.5, whereas a practical script on how to use it can be found in Appendix D.1.

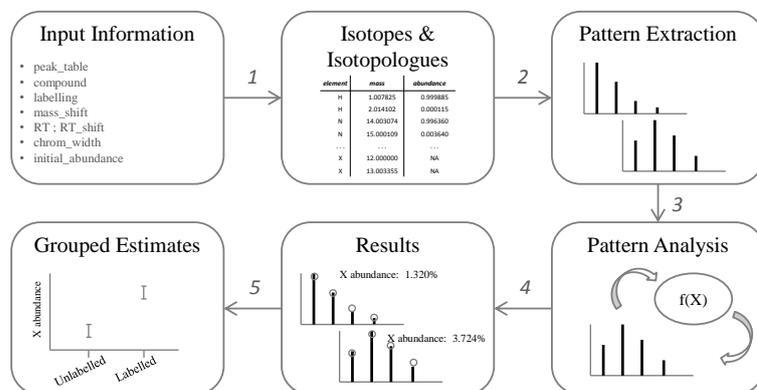


FIGURE 4.5: Block scheme of the main processing steps. See the text for a more extensive explanation.

Input Information

In order to get answers from *IsotopicLabelling*, it is first required to input some important information:

1. **The MS data.** The user should provide an LC-MS or a GC-MS data set relative to labelling experiments. The data should have been already pre-processed, and inputted as an R data frame with its first

two columns representing m/z and retention time (RT) of the identified peaks, and the other columns (one for each sample) containing peak intensities or areas. Considering that a popular R package for processing MS data is *xcms* [43], *IsotopicLabelling* can also accept and handle *xcmsSet* objects;

- 2. The analyte of interest.** The package needs to know the chemical formula of the target analyte, as observed by MS (that is, an ion adduct); this should be specified as a character vector, and a “special” notation is required whereby the atoms with unknown isotopic abundance are denoted by “X”. For example, the PC 32:2 proton adduct, [PC 32:2+H]⁺ with chemical formula C₄₀H₇₇NO₈P, is denoted by X40H77NO8P in ¹³C labelling experiments (since the C relative distribution is unknown), whereas the formula becomes C40X76HNO8P when the label is ²H: here, not all of the 77 hydrogen atoms were replaced by “X”, but only 76 of them, to take into consideration that one H comes from the ESI ionisation process, and it therefore has the natural isotopic distribution;
- 3. The labelling isotope.** The user should specify which kind of labelling experiment was performed; the current implementation of the package only supports ²H and ¹³C labels;
- 4. Chromatographic and MS parameters.** The user should also indicate analyte-dependent experimental parameters, required to identify and extract the isotopic patterns. They include mass accuracy, expected RT range and chromatographic width of the peaks. If known, also an estimate of the label abundance can be specified;
- 5. Charge of the adduct.** Finally, also the charge state of the target adducts has to be provided, and it is needed by the package to calculate the expected m/z values of the isotopologue peaks, as well as their separation among each other. By default, the charge is assumed to be +1.

It is important to point out here that a reliable estimate of the incorporation can be obtained only if the formula of the target metabolite is known in advance, which is the reason why among the required input is the chemical formula of the target, as well as chromatographic information. Indeed, in a completely untargeted scenario the labelling makes the isotopic patterns useless for annotation, and mass accuracy alone cannot be used to infer a reliable elemental composition [44].

Isotopes and Isotopologues

At the beginning, *IsotopicLabelling* takes the user’s input and computes some important parameters, required for the processing. To start with, the package lists all the possible isotopologues generated by the incorporation of the label in the target analyte, determining their m/z : the lightest one is the monoisotopic species (no label incorporation), while the heaviest is the species where all the X atoms have been replaced by the label (²H or ¹³C). At the same time, *IsotopicLabelling* also makes a table containing the natural isotopic abundances of the elements in the target analyte.

To sum up, in this first step an R list is created, containing the following objects:

1. **compound**: the character vector given as input, specifying the target analyte;
2. **isotopes**: a table containing the natural isotopic abundances of the elements in “compound”, as well as the two isotopes of the label X, which are assigned NA (not available) values;
3. **target**: named vector with the exact m/z values of all the possible isotopologues; “M+0” is the monoisotopic mass, whereas in “M+1” there has been a replacement, and so forth. Here the underlying assumption is that the MS resolution is not high enough to resolve the isotopic fine structure, and therefore the replacement of, say, ^1H with ^2H is indistinguishable from ^{12}C with ^{13}C . This is true for most of the instruments currently used in LC-MS measurements;
4. **nX**: the number of X atoms present;
5. **nTOT**: the total number of atoms of the labelling element (either H+X or C+X, depending on the experiment).

Extraction of the Experimental Patterns

The second task of *IsotopicLabelling* is to extract the experimental isotopic patterns of the target analyte from the MS data. The output is a matrix where the first two columns represent exact m/z and RT of each peak, whereas in the others are signal intensities or areas (one column for each sample). Note that each row represents one of the possible isotopologues previously listed.

The extraction of the signals to this matrix is achieved through the following matching procedure:

1. Each possible isotopologue is considered, and all the MS peaks that are within the specified m/z and RT ranges are listed;
2. The RT of the identified peaks are compared across isotopologues, to find groups of peaks (the assumption is that isotopologues of the same analyte have the same RT). Here, the specified chromatographic width is used as a basis for comparison to establish the groupings;
3. If more than one group has been identified, there follows a “selection” procedure: first of all, the groups containing fewer than two isotopologues are discarded, since they are assumed not to represent true isotopic patterns. If more groups are still left, the one with the closest average RT to the expected RT is chosen. At the end, therefore, at most one single pattern is selected for each sample.

Isotopic Pattern Analysis

Once extracted, the patterns are analysed by means of a fitting procedure, where the only parameter to estimate is the abundance of the label. The basic point is that different label abundances will give rise to different patterns, and therefore such fitting aims at estimating the label abundance by choosing

the value originating the theoretical pattern most similar to the experimental one.

The whole process is carried out by a non-linear least squares fitting, and the patterns are computed with the *ecipex* R package [23]. For the comparison, theoretical and experimental patterns are normalised to the most intense signal, which is set to 100; to account for noise and to correct for heteroskedasticity, the signals are given weights proportional to the square root of their intensities.

For the fitting to converge, a good initial guess of the label abundance is required, which *IsotopicLabelling* obtains this way: if the user provides it, such value is directly used to start the regression, otherwise it is computed by looking at the position of the most intense signal in the experimental pattern relative to the monoisotopic peak:

$$\text{Initial estimate} = \frac{m/z_{(\text{max})} - m/z_{(\text{monoisot})}}{nX} \quad (4.5)$$

The principle here is that the higher the label abundance, the more the signals are shifted towards higher masses.

The overall output is an R list containing the following objects:

1. **compound:** the same as input;
2. **best_estimate:** a numeric vector of estimated percentage abundances of the label (one value for each sample analysed);
3. **std_error:** standard errors of regression (one value for each sample);
4. **dev_percent:** percentage deviations between each experimental and related best-fitted isotopic pattern;
5. **x_scale:** the m/z of all possible isotopologues;
6. **y_exp:** a matrix of normalised experimental patterns (one column for each sample);
7. **y_theor:** a matrix of best-fitted patterns (one column for each sample);
8. **residuals:** a matrix of residuals (differences between experimental and fitted patterns);
9. **warnings:** a character vector containing possible warnings from the fitting procedure.

How to Survey the Results

There are a number of ways to look at the results of the processing. If the user is only interested in getting the estimated values, a dedicated function (“summary”) can print a table of percentage estimates with related standard error of regression (SER). At the same time, also different graphical outputs can be produced with the function “plot”:

1. **patterns:** for each sample, the normalised experimental pattern is plotted superimposed to its best-fitted pattern. The estimated label abundance is also given;

2. **residuals:** here, for each sample the residuals are plotted, and this is useful if the aim is to assess whether or not they are randomly distributed;
3. **summary:** a single plot can be produced as well, which summarises the estimated values and related errors: in the x-axis are the different samples, whereas the y-axis represents isotopic abundances.

Additionally, the user can choose whether to only display the plots to the screen, or to export them in a single *.pdf file for later assessment.

IsotopicLabelling also allows to store the results: a detailed table can be saved as a *.csv file containing, for each sample: (i) the estimated label percentage abundance, (ii) the related SER, (iii) the percentage deviation between best-fitted and experimental patterns, and (iv) the outcome message from the fitting.

Dealing with Biological Replicates

The isotopic pattern analysis ends up in as many estimates as the number of samples, each with its own SER. In biological experiments, however, it is common practice to work with replicates and to have samples arranged in different groups. Therefore, it would be convenient to get a single estimate for each group, and to this goal *IsotopicLabelling* provides a dedicated function, “group_labelling”.

The group averages are simply obtained by considering the individual estimates within each group, under the assumption that the samples are representative of the group population. On the other hand, the group standard deviations should take into consideration both the individual SERs associated with each estimate, and the distribution of the estimates within each group (biological variability). As a consequence, the group variances are computed by way of the law of total variance:

$$\frac{1}{m} \sum_{i=1}^m \text{SER}_i^2 + \frac{1}{m-1} \sum_{i=1}^m (a_i - \hat{a})^2 \quad (4.6)$$

where m is the number of samples in the considered group, a_i is the estimate of sample i , and \hat{a} is the group average. Equation (4.6) states that the overall variance has two contributions: the “within-sample” variance (first term) which is the sum of the variances associated to each estimate, and the “between-sample” variance (second term), that takes into consideration the distribution of the estimates in the sample group. Finally, the group standard deviations are computed as the square root of such variances.

Batch Processing

From what described so far, *IsotopicLabelling* can deal with two different scenarios:

- A. **Targeted pre-processing, target analyte:** the users are interested in one particular analyte, and provide just the experimental isotopic pattern of such species. Here, *IsotopicLabelling* does not have to extract any experimental pattern out of LC-MS data, but it directly fits the best pattern in order to get an estimate of the relative abundance of the label.

In this case, the provided experimental pattern could be obtained in a number of independent ways, including the use of proprietary MS software;

- B. **Untargeted pre-processing, target analyte:** the users are interested in one particular analyte, but provide a pre-processed “untargeted” data matrix, containing detected LC-MS features. Here, *IsotopicLabelling* takes care of first extracting and subsequently fitting the patterns. In particular, *xcmsSet* objects (the results of pre-processing with *xcms*) can be used as input, however the data matrix may once again be obtained independently. In this scenario, the chromatographic information is required to match the target analyte with the untargeted list of features.

A further working mode is available, and it regards the processing of a list of target analytes: here, the users provide untargeted pre-processed data and a list of targets, and *IsotopicLabelling* returns the estimated abundances of the label in the several analytes, through the function “*batch_labelling*”. Worthy of note, also in this case the user is required to input chromatographic information, otherwise the patterns extraction would not be feasible.

4.2.4 The Example Data Set

IsotopicLabelling is freely available, and included in it is an example data set that was kindly provided by Dr. Nyasha Munjoma (Julian L. Griffin Group and Nianshu Zhang Group, Department of Biochemistry, University of Cambridge – UK). This contains 8 LC-MS measurements relative to ^{13}C labelling experiments where the first 4 samples are lipid extracts from unlabelled YSBN1 yeast strain cultures (natural ^{13}C abundance), whereas in the last 4 samples the cells were grown in a media supplemented with uniformly-labelled ^{13}C glucose (99 % ^{13}C labelling). Our idea was to use this data set to help the users better understand how to use the package; indeed, also included in the package is a vignette that serves as practical guide, where the example data set is used.

For the sake of completeness, below are further experimental details.

Cell Cultures

For the YSBN1 yeast strain cultures, 13.2 g of yeast nitrogen base (Sigma) and 40 g of Ammonium Sulphate (Sigma) was dissolved in 400 mL of autoclaved distilled water, which was then filter-sterilised, protected from light with aluminium foil and stored at 5 °C prior to use. The yeast nitrogen base was used to make Synthetic Minimum Media (SMM) containing 0.5 % glucose: for isotope labelling experiments, uniformly-labelled ^{13}C (U- ^{13}C) glucose (Cambridge Isotope Laboratories Inc., MA, USA) was employed, whereas for control experiments natural abundance D-(+) glucose (Sigma) was used. The yeast cultures were grown in the described media for 24 h, and the lipids were extracted using a slight modification of the Folch method. Briefly, about 50 mg of cell material was placed in a 2 mL screw top tube, together with about 200 mg of glass beads. 1 mL of chloroform:methanol (2:1) was added, and the yeast cell wall was broken using a TissueLyser (Time: 8 min, Frequency: 240 Hz). The samples were sonicated for 5 min, followed by addition of 400 μL water and vortexing for 60 s. The mixture was centrifuged

for 10 min at 13 200 rpm, and the organic phase was withdrawn and dried overnight in a fume hood. For the mass spectrometric analysis, 500 μL of chloroform:methanol (1:1) was added to the dried lipid extract, and the obtained solution was diluted 20 times using isopropanol:methanol:water (2:1:1).

LC-MS Measurements

The samples were analysed in positive ionisation mode using a Waters Xevo G2 quadrupole time of flight (Q-ToF) combined with an Acquity Ultra Performance Liquid Chromatogram (UPLC) (Waters Corporation, Manchester, UK). 10 μL of the sample was injected onto an Acquity UPLC Charged Surface Hybrid (CSH) C18 column (1.7 μm x 2.1 mm x 100 mm) (Waters Corporation) held at 55 $^{\circ}\text{C}$. The binary solvent system (flow rate 0.4 mL/min) consisted of solvent A containing HPLC grade acetonitrile:water (60:40) with 10 mM ammonium formate and solvent B consisting of LC-MS grade acetonitrile:isopropanol (10:90) and 10 mM ammonium formate. The gradient started from 40 % B, reached 99 % B in 18 min, then returned back to the starting condition, and remained there for the next 2 min. The data was collected over the m/z range 105–1800 with a scan duration of 0.2 s. The source temperature was set at 120 $^{\circ}\text{C}$ and nitrogen was used as the desolvation gas (900 L/h). The voltages of the sampling cone, extraction cone and capillary were 30 kV, 3.5 kV and 2 kV respectively, with a collision energy of 6 V for each single scan. As lock mass, a solution of 2 ng/ μL acetonitrile:water (50:50) leucine enkephaline (m/z 556.2771) with 0.1 % formic acid was infused into the instrument every 30 s.

4.3 Testing the Package

4.3.1 LC-MS, Complete ^{13}C Labelling

After the development of *IsotopicLabelling*, we first tested it on the LC-MS data of the example data set described above: we converted the *.raw files to the *.mzML format using “MSConvert” (from ProteoWizard), and then pre-processed them (peak identification and filtering, peak matching, retention time correction, filling in of missing data, ...) with *xcms*, to get an *xcmsSet* object (which we have included in the package). At this point, we used *IsotopicLabelling* for the pattern analysis, that we performed on the most abundant PC and TAG species.

Table 4.1 shows the estimated ^{13}C abundances with related SEM (standard error of the mean) for some of the species, where we grouped the samples in either controls or labelled species. In all instances, the ^{13}C estimates in control samples are in agreement with the natural ^{13}C occurrence, whose best representative value is 1.07 %, but it can range from 0.99 % to 1.15 % [45]; on the other hand, the estimates for labelled samples are close to the ^{13}C occurrence in the used labelled glucose.

In Figure 4.6 is just a graphical example of two of the extracted patterns (control VS labelled) relative to the analyte $[\text{PC } 32:2+\text{H}]^+$, to show how the patterns change upon labelling: whereas in the control sample the most intense signal corresponds to the monoisotopic species (blue-filled triangle),

	CONTROL		LABELLED	
	Mean	SEM	Mean	SEM
[PC 32:1+H] ⁺	1.069 %	0.006 %	98.963 %	0.005 %
[PC 32:2+H] ⁺	1.079 %	0.004 %	98.943 %	0.005 %
[PC 34:1+H] ⁺	1.075 %	0.008 %	98.982 %	0.007 %
[PC 34:2+H] ⁺	1.080 %	0.003 %	98.940 %	0.004 %
[PC 36:1+H] ⁺	1.04 %	0.02 %	99.05 %	0.01 %
[PC 36:2+H] ⁺	0.99 %	0.02 %	99.02 %	0.02 %
[TAG 48:2+NH ₄] ⁺	1.091 %	0.003 %	98.967 %	0.006 %
[TAG 48:3+NH ₄] ⁺	1.090 %	0.003 %	98.964 %	0.005 %
[TAG 50:2+NH ₄] ⁺	1.09 %	0.02 %	98.970 %	0.006 %
[TAG 50:3+NH ₄] ⁺	1.092 %	0.001 %	98.967 %	0.006 %
[TAG 52:1+NH ₄] ⁺	1.094 %	0.004 %	98.993 %	0.003 %
[TAG 52:2+NH ₄] ⁺	1.100 %	0.004 %	98.970 %	0.008 %
[TAG 52:3+NH ₄] ⁺	1.092 %	0.005 %	98.99 %	0.03 %

TABLE 4.1: Estimated ¹³C abundances on the test data set. For control samples (naturally occurring glucose), the estimated ¹³C abundance is very close to the natural value (1.070 %), whereas for labelled samples it is close to 99.0 %, the nominal ¹³C occurrence in the employed labelled glucose.

in the labelled sample the main signal is shifted upwards by 40 *m/z* units, corresponding to a species made up of just ¹³C, and no ¹²C atoms.

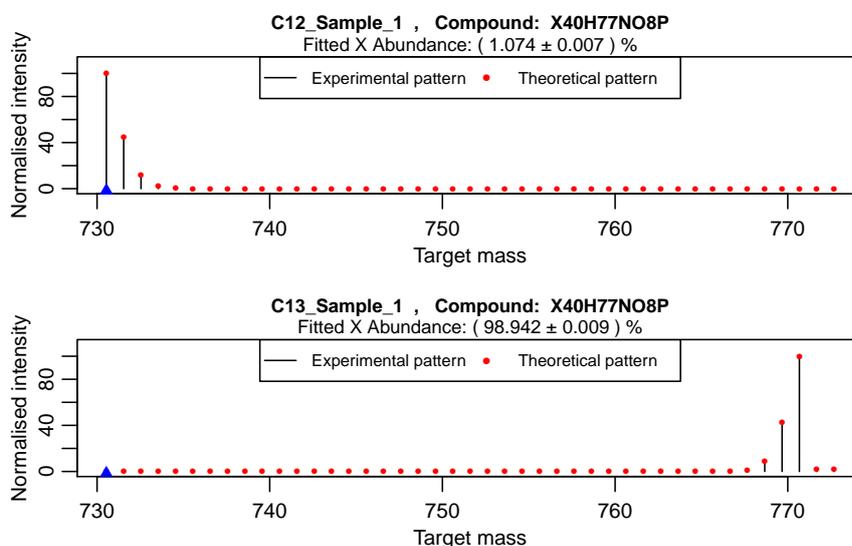


FIGURE 4.6: Two of the extracted patterns for [PC 32:2+H]⁺, with their best-fitted patterns (red dots). The estimated ¹³C abundances are given with the SER.

4.3.2 LC-MS, Intermediate ¹³C Labelling

The complete labelling is a simple case, in that it causes a huge shift in the patterns. To further test *IsotopicLabelling*, we therefore moved on by considering a case of intermediate ¹³C labelling: with the help of Dr. N. Zhang, we prepared another set of 8 samples in a similar way as described above, the

only difference being that in the labelled samples we mixed up 60 % natural glucose and 40 % uniformly-labelled ^{13}C glucose.

The average ^{13}C abundance in the feeding glucose for labelled samples can be easily computed by considering the average molecular weights of glucose,

$$\begin{aligned} \text{MW}(\text{natural Glc}) &= 180.156 \frac{\text{g}}{\text{mol}} \\ \text{MW}(\text{labelled Glc}) &= \left[(0.99 \times 13.003355 + 0.01 \times 12) \times 6 + \right. \\ &\quad \left. + 12 \times 1.007941 + 6 \times 15.99941 \right] \frac{\text{g}}{\text{mol}} = \quad (4.7) \\ &= 186.0517 \frac{\text{g}}{\text{mol}} \end{aligned}$$

Therefore:

- 60 g natural Glc equals 0.333 mol, corresponding to 0.021 mol ^{13}C and 1.977 mol ^{12}C ;
- 40 g labelled Glc equals 0.215 mol, corresponding to 1.277 mol ^{13}C and 0.013 mol ^{12}C .

Putting all together, the expected molar ^{13}C abundance is 39.5 %.

We performed the isotopic pattern analysis as previously done, and the results are in Table 4.2 (to compare with Table 4.1). Whereas control samples gave no unexpected results, for labelled samples the obtained values deserve a closer inspection: first and foremost, the estimates have a higher SEM as compared to unlabelled samples, and secondly they seem to be dependent on the lipid class. In particular, whereas their average is $(36.6 \pm 0.3)\%$ for PCs, for TAGs it is $(41.9 \pm 0.3)\%$; the difference is statistically significant (p-value: 2×10^{-8}), and this finds no easy explanation.

One possible hypothesis comes from looking at the PC and TAG structures, reported in Figure 4.7: it should be considered that both PCs and TAGs have two kinds of carbon atoms, those in the FA chains (R_1 , R_2 and R_3 in the figure) and those making up the polar heads (the numbered atoms in the figure, which for PCs and TAGs are 8 and 3, respectively). The different

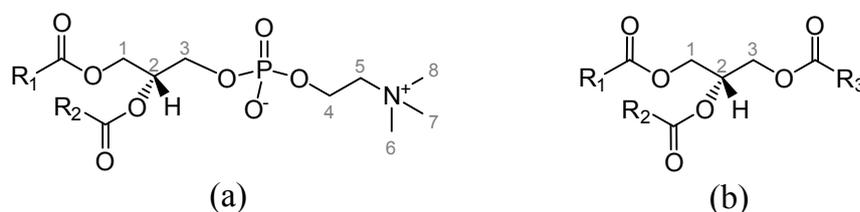


FIGURE 4.7: The structures of (a) a PC, and (b) a TAG. Despite the labelling, we speculate that the numbered C atoms have fixed natural abundance, since they do not come from the glucose metabolism; the labelled carbons only are in the FA chains (R_1 , R_2 and R_3).

	CONTROL		LABELLED	
	Mean	SEM	Mean	SEM
[PC 32:1+H] ⁺	1.03 %	0.07 %	35.9 %	0.4 %
[PC 32:2+H] ⁺	1.08 %	0.01 %	35.9 %	0.3 %
[PC 34:1+H] ⁺	0.96 %	0.06 %	37.4 %	0.5 %
[PC 34:2+H] ⁺	1.07 %	0.01 %	36.7 %	0.3 %
[PC 36:1+H] ⁺	1.2 %	0.1 %	37.1 %	0.6 %
[PC 36:2+H] ⁺	1.20 %	0.08 %	36.7 %	0.4 %
[TAG 48:2+NH ₄] ⁺	NA	NA	NA	NA
[TAG 48:3+NH ₄] ⁺	1.10 %	0.03 %	41.7 %	0.6 %
[TAG 50:2+NH ₄] ⁺	1.09 %	0.04 %	41.4 %	0.5 %
[TAG 50:3+NH ₄] ⁺	1.09 %	0.09 %	41.5 %	0.3 %
[TAG 52:1+NH ₄] ⁺	1.04 %	0.02 %	42.5 %	0.6 %
[TAG 52:2+NH ₄] ⁺	1.11 %	0.05 %	42.6 %	0.4 %
[TAG 52:3+NH ₄] ⁺	1.08 %	0.06 %	41.7 %	0.3 %

TABLE 4.2: Estimated ¹³C abundances for the intermediate ¹³C-labelling study.

building blocks are synthesized in different and complex pathways (as exemplified in Figure 4.8), and this may account for the difference observed in ¹³C abundance between PC and TAG lipids. In particular, the lower values in PCs seem to suggest that some of the polar head carbon atoms are taken up not from the feeding glucose, but from other sources.

To understand whether our hypothesis is correct, we tried to repeat the analysis, but this time we assumed the 5 C atoms of the phosphocholine group to have natural ¹³C abundance, and so we fitted the patterns by changing the average ¹³C occurrence in the remaining C atoms (FA chains plus the glycerol moiety); the new results are in Table 4.3. This time there

	LABELLED	
	Mean	SEM
[PC 32:1+H] ⁺	41.1 %	0.5 %
[PC 32:2+H] ⁺	41.0 %	0.4 %
[PC 34:1+H] ⁺	42.4 %	0.6 %
[PC 34:2+H] ⁺	41.7 %	0.4 %
[PC 36:1+H] ⁺	42.1 %	0.7 %
[PC 36:2+H] ⁺	41.6 %	0.6 %
[TAG 48:2+NH ₄] ⁺	NA	NA
[TAG 48:3+NH ₄] ⁺	41.7 %	0.6 %
[TAG 50:2+NH ₄] ⁺	41.4 %	0.5 %
[TAG 50:3+NH ₄] ⁺	41.5 %	0.3 %
[TAG 52:1+NH ₄] ⁺	42.5 %	0.6 %
[TAG 52:2+NH ₄] ⁺	42.6 %	0.4 %
[TAG 52:3+NH ₄] ⁺	41.7 %	0.3 %

TABLE 4.3: Estimated ¹³C abundances for the intermediate ¹³C-labelling study, assuming some of the C atoms (PC headgroup) to have fixed natural abundance.

is no significant difference between PCs and TAGs (p-value: 0.46), which

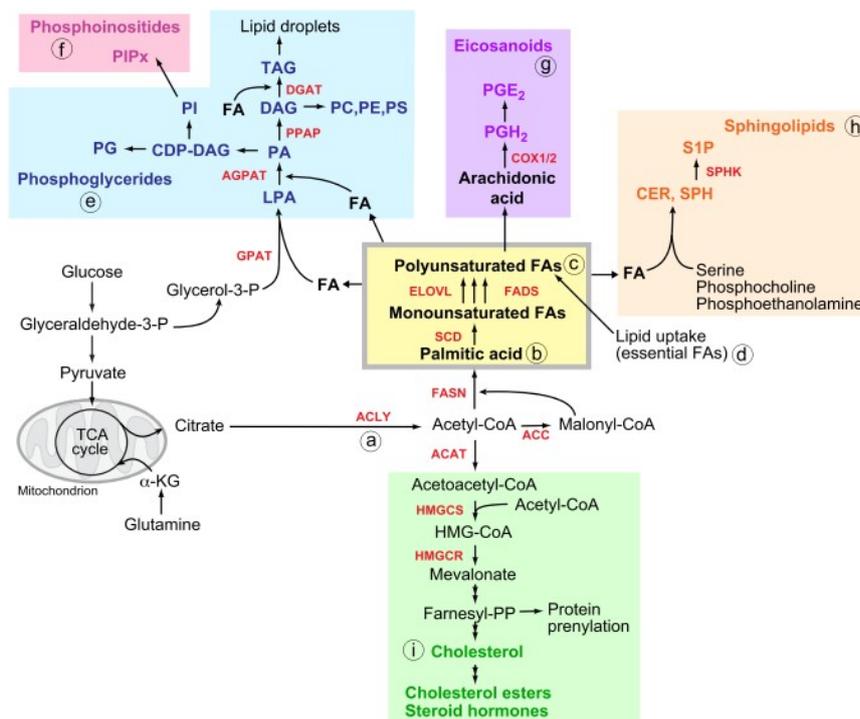


FIGURE 4.8: Scheme of the lipid biosynthesis. As shown here, the C atoms in the FA chains and in the lipids' polar heads are created by two related but distinct mechanisms. Source: [46].

may confirm our hypothesis: in these samples, only the C atoms of the FA chains and of the glycerol moiety are constructed by building blocks coming from the catabolism of glucose, whereas the PC headgroups come from other (unlabelled) sources.

However, it should be noticed that these results are seemingly in contrast to those in Section 4.3.1, where complete labelling was observed also for the C atoms of the PC headgroups. To reconcile these figures, we suggest that what happens is the following: whereas the synthesis of the FA chains immediately starts from the catabolism of glucose, the synthesis of the PC headgroup preferably starts from other available sources, and only when these sources have been completely depleted does the choline synthesis use the glycerol-3-phosphate that comes from the process of glycolysis. If this hypothesis is right, longer experiments should show a gradual increase of labelling and decrease the differences between PCs and TAGs; unfortunately, we have not had the opportunity to test this, but it is an important point that should be considered for future experiments.

Another important issue to consider comes when looking at the patterns themselves, which show peculiar features, and an example for $[PC\ 32:2+H]^+$ is reported in Figure 4.9: whereas our analysis is able to nicely fit the overall shape of the pattern, it cannot account for some details, which are therefore completely missed. Indeed, all of the predicted patterns have a “regular” shape where the isotopologue peaks keep increasing in intensity up to a certain m/z , beyond which value they steadily decrease. This is not the case for our extracted patterns, which have a saw-tooth shape where signals at “even” masses ($M+6, M+8, \dots$) are most intense than those at “odd” masses

(M+5, M+7, ...).

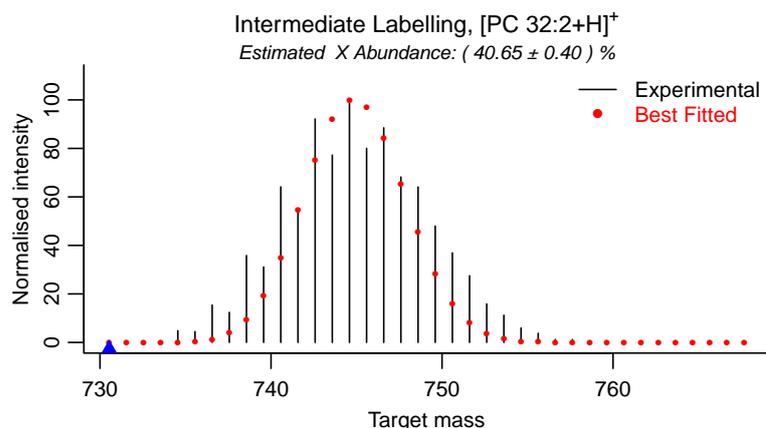


FIGURE 4.9: Example of extracted pattern for $[\text{PC } 32:2+\text{H}]^+$, intermediate labelling. The pattern shows a peculiar saw-tooth shape that is not accounted for by our isotopic pattern analysis.

The discrepancy between fitted and extracted patterns also partly explains why we got higher standard deviations in the estimates (compare Tables 4.1 and 4.3). Another possible reason is related to a loss of sensitivity: whereas in both unlabelled and completely labelled analytes the isotopologue signals are distributed over just a few m/z values (up to 4 signals), here the possible isotopologues span a much broader range, consisting of up to 20 signals (see Figure 4.9). As a consequence, there is a kind of “dilution” where each signal has a lower intensity.

As for the shape of the patterns, the problem with the previous analysis is that it assumes the target analytes to be synthesised from a homogeneous pool of the label. This would hold true only if the glucose molecules were completely broken down before being used to synthesise lipids, which is not the case here. As a matter of fact, the involved pathways are quite complex, as Figure 4.8 points out. In particular, the FA chains are assembled from building blocks of 2 carbon atoms each, since they come from the citrate exiting the TCA cycle following the degradation of glucose to pyruvate. As for the lipids’ polar heads, here things are more complex since each lipid class has its own pathway. In any case, each glyceryl moiety comes from the same glucose molecule which is split to glyceraldehyde-3-phosphate, in its turn converted to glycerol-3-phosphate.

To tackle this specific issue and to understand whether the peculiar shapes can be explained by the incomplete degradation of glucose, we changed our model, assuming the carbon sources to be made up of building blocks of 2 carbon atoms each. As a consequence, there will possibly be three types of blocks:

1. $^{12}\text{C}-^{12}\text{C}$, that we call “ ^{24}X ” in our model;
2. $^{12}\text{C}-^{13}\text{C}$, or “ ^{25}X ”;
3. $^{13}\text{C}-^{13}\text{C}$, or “ ^{26}X ”.

Of course, this is an oversimplification of the true problem, considering what said above. However, it should also be considered that the C atoms in

the FA chains are much higher in number as compared to the polar heads' carbons, and therefore we expect to be able to reproduce the particular features of the patterns by considering just these 2-C building blocks.

Under our new model, for each target analyte instead of having a number of X atoms with unknown distribution between ^{12}X and ^{13}X isotopes, there will now be half that many, with unknown distribution between the three "isotopes" ^{24}X , ^{25}X and ^{26}X . The new problem is therefore to find the relative abundance of these three blocks, which implies that the fitting procedure will now have 2 unknown parameters instead of just one.

As for the polar heads' carbons, during analysis we kept their abundance fixed: we assumed the 5 C atoms of the polar PC head to have natural ^{13}C abundance, whereas we considered the 3 C atoms of the glyceryl moiety to directly come from the glucose metabolism. Therefore, similarly to the 2-C building blocks, we considered them to be 3-C building blocks (that we call "A" for simplicity), but with a fixed distribution between its "isotopes": ^{36}A (58.96%), ^{37}A (1.84%), ^{38}A (1.19%), ^{39}A (38.01%). These figures can be easily obtained by considering how the feeding glucose was experimentally prepared.

As always with fitting procedures, good initial guesses are essential. In our case, we know that we have mixed about 60% natural glucose with 40% labelled glucose, and we therefore expect to roughly have 59% of ^{24}X , 2% of ^{25}X and 39% of ^{26}X . If we consider $[\text{PC } 32:2+\text{H}]^+$ ($\text{C}_{40}\text{H}_{77}\text{NO}_8\text{P}$), upon our model its formula becomes $\text{AC}_5\text{X}_{16}\text{H}_{77}\text{NO}_8\text{P}$, where C are the atoms with fixed natural abundance, A is the 3-C building block of the glyceryl moiety, whereas X plays the role of an element with the 3 stable isotopes ^{24}X , ^{25}X and ^{26}X . It is the X distribution that is being estimated by our analysis.

In Figure 4.10 is the isotopic pattern that would result under these assumptions: interestingly, the saw-tooth shape turns up.

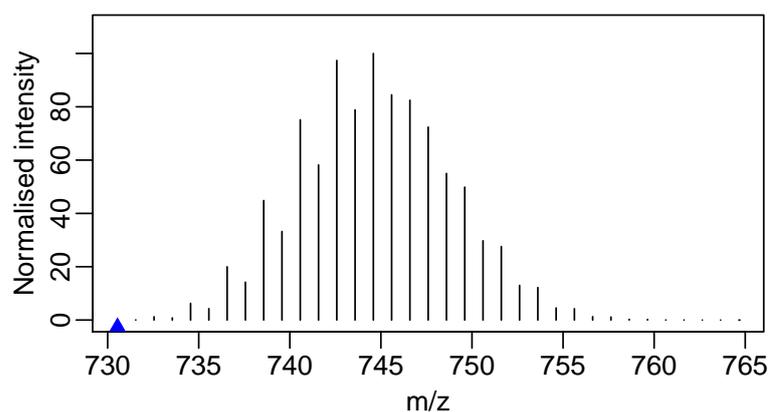


FIGURE 4.10: Simulated isotopic pattern for $[\text{PC } 32:2+\text{H}]^+$ under the assumption that the FA chains are generated from building blocks of 2 carbon units each (relative isotopic distribution: 59% ^{24}X , 2% ^{25}X , 39% ^{26}X), and the glyceryl moiety from a 3-C building block (relative isotopic distribution: 58.96% ^{36}A , 1.84% ^{37}A , 1.19% ^{38}A and 38.01% ^{39}A). Note the saw-tooth shape.

We therefore repeated our analysis, and this time the patterns were strikingly reproduced, as shown in Figure 4.11 (example for $[\text{PC } 32:2+\text{H}]^+$). The obtained estimates are in Table 4.4; interestingly, here no significant

	ABUNDANCE ^{24}X		ABUNDANCE ^{25}X		ABUNDANCE ^{26}X	
	Mean	SEM	Mean	SEM	Mean	SEM
[PC 32:1+H] ⁺	56.4 %	0.5 %	2.6 %	0.5 %	41.0 %	0.4 %
[PC 32:2+H] ⁺	57.3 %	0.2 %	2.6 %	0.2 %	40.1 %	0.2 %
[PC 34:1+H] ⁺	55.3 %	0.9 %	3.6 %	0.9 %	41.1 %	0.6 %
[PC 34:2+H] ⁺	56.8 %	0.2 %	2.7 %	0.2 %	40.5 %	0.3 %
[PC 36:1+H] ⁺	NA	NA	NA	NA	NA	NA
[PC 36:2+H] ⁺	56.9 %	0.5 %	3.0 %	0.6 %	40.1 %	0.5 %
[TAG 48:2+NH ₄] ⁺	56.5 %	0.3 %	2.1 %	0.4 %	41.5 %	0.4 %
[TAG 48:3+NH ₄] ⁺	55.9 %	0.7 %	2.8 %	0.7 %	41.3 %	0.5 %
[TAG 50:2+NH ₄] ⁺	56.1 %	0.2 %	2.4 %	0.3 %	41.5 %	0.3 %
[TAG 50:3+NH ₄] ⁺	56.7 %	0.2 %	2.7 %	0.3 %	40.6 %	0.2 %
[TAG 52:1+NH ₄] ⁺	55.6 %	0.3 %	2.7 %	0.3 %	41.7 %	0.2 %
[TAG 52:2+NH ₄] ⁺	55.6 %	0.3 %	2.8 %	0.4 %	41.6 %	0.2 %
[TAG 52:3+NH ₄] ⁺	56.3 %	0.2 %	3.0 %	0.4 %	40.7 %	0.3 %

TABLE 4.4: Estimated abundances of the “isotopes” ^{24}X , ^{25}X and ^{26}X in the target analytes listed.

difference between PCs and TAGs can be observed. In addition, the estimates are quite in agreement with the expected values, therefore confirming the soundness of our approach.

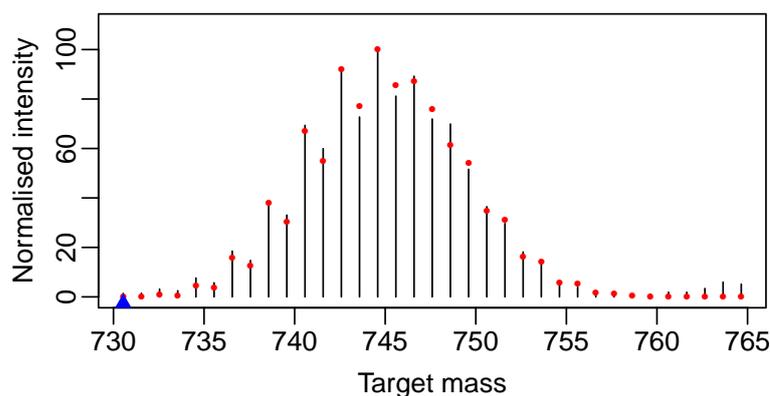


FIGURE 4.11: Experimental (black vertical lines) and fitted (red dots) patterns for [PC 32:2+H]⁺ in one of the labelled samples; the fitting was performed using the 2-C building blocks model.

Although our initial aim was only to test *IsotopicLabelling*, we believe that the case just discussed here is interesting in that it provides an experimental evidence of how lipids are synthesised. In particular, glucose is not completely broken down during glycolysis; instead, it generates 2-C and 3-C building blocks, which are then used during anabolism.

4.3.3 Simulations: Minimum Enrichment Detectable

The developed *IsotopicLabelling* package estimates the ^{13}C or ^2H abundance in target analytes, which is related to the degree of label incorporation. Therefore, it is important to get at least a rough idea as to which is the minimum amount of label incorporation that can be detected. In other words, the problem now is to estimate the minimum label abundances

that labelled analytes should have in order for their patterns to be clearly discernible from those of the related control analytes (no labelling).

In order to achieve this goal, we have to keep into consideration that some parameters may influence the sensitivity of the approach. First of all, the size of the analytes can play an important role, since the relative distribution among isotopologues strongly depends on m/z : the heavier the analyte, the “broader” its isotopic pattern. Another important parameter is expected to be the signal-to-noise ratio (SNR), since the higher it is, the easier it will be to pinpoint minor alterations in the patterns.

To tackle this issue, we tested *IsotopicLabelling* over a set of simulated patterns: as for the analytes, we considered saturated carboxylic acids with acyl chains ranging from C2 to C26 (m/z from 61.02 to 397.40); as for SNR, we took into account 11 values in the range 10 to 10^4 ¹. For each condition, we gradually increased either the ^{13}C or the ^2H abundance, starting from their natural occurrence; for each value, we simulated 50 noisy isotopic patterns and analysed them by means of *IsotopicLabelling*, in order to find an averaged estimated label occurrence with related standard deviation. We repeated the procedure of increasing the label abundance until we found a value for which the 95 % confidence interval did not include the natural ^{13}C or ^2H occurrence.

The overall results are summarised in Figure 4.12 for ^{13}C , and in Figure 4.13 for ^2H labelling experiments. In both, to the left is the minimum label abundance as a function of the (m/z , SNR) pair, whereas to the right are the related standard deviations. What can easily be observed is that the values are strongly dependent on both m/z and SNR: as the SNR increases, so does the sensitivity of the approach, since the noise level decreases and the patterns become less and less noisy. At the same time, the sensitivity also increases as the analyte size increases, a direct consequence of large analytes having “broader” isotopic profiles, for which slight alterations of the patterns are easier to discern.

Taken together, these estimates point out that the sensitivity of the approach depends on the experimental conditions, with the lowest enrichments being detectable when looking at big and abundant metabolites. Worthy of note, the results also depend on the label (^{13}C or ^2H), since (i) the natural occurrences are different, and (ii) the number of sites amenable to labelling differs (usually there are more H than C atoms).

¹We define here SNR as the ratio between the intensity of the monoisotopic peak (the most intense one for poor labelling and small analytes as those considered in this simulation) to the standard deviation of the noise.

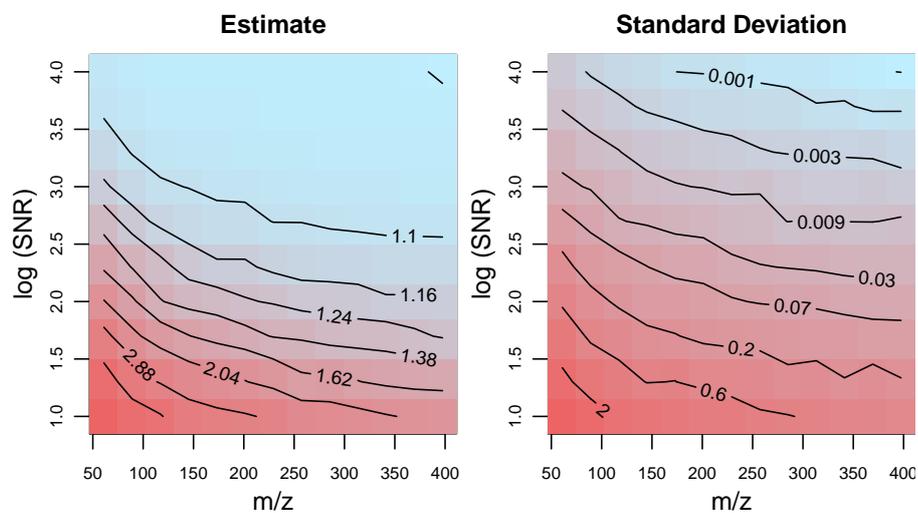


FIGURE 4.12: *Left*: minimum ^{13}C percentage abundance (black level curves) for which the difference between labelled and control patterns is significant. *Right*: related standard deviation. The sensitivity strongly depends on the analyte size and on the SNR.

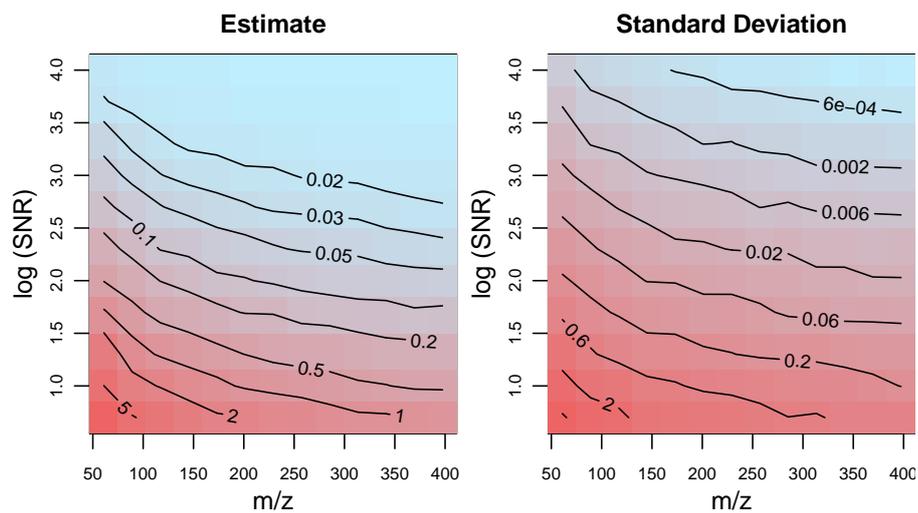


FIGURE 4.13: *Left*: minimum ^2H percentage abundance (black level curves) for which the difference between labelled and control patterns is significant. *Right*: related standard deviation. The sensitivity strongly depends on the analyte size and on the SNR.

4.4 Biochemical Application

4.4.1 ^2H -labelling

After it was tested, *IsotopicLabelling* was applied to biochemical LC-MS and GC-MS data provided by Francis Sanders (Dr. Julian L. Griffin's Lipid Profiling and Signalling Group, MRC Human Nutrition Research, Cambridge, UK), whose scientific interest is related to the DNL process in health and disease. Briefly, he fed different groups of mice with ^2H -enriched water (D_2O) in order to study its uptake in lipids. The rationale is that lipids synthesised by way of DNL will partly incorporate the label, whereas those coming from the diet will not. As a consequence, the higher the observed ^2H abundance, the higher the DNL rate.

^2H -enriched water has been used in many investigations so far, basically with a focus on metabolism and metabolic disorders [8, 35, 47–51]. All such experiments rely on the fact that ^2H from the labelled water can be incorporated during the DNL process; for example, hepatic acetyl-CoA is the product of the TCA cycle, and the precursor for FA synthesis. In the presence of D_2O , the methyl hydrogens of acetyl-CoA become enriched with ^2H thanks to exchange mechanisms taking place during its formation *via* glycolysis and other metabolic pathways [52, 53]. In addition, ^2H can be incorporated from D_2O into C–H bonds of the glycerol moiety of acyl-glycerides [54]. Worthy of note, once in this configuration, ^2H atoms are no more exchangeable with hydrogen atoms from the water pool.

4.4.2 Materials and Methods

Animal Care

All procedures involving the mice were performed in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986, under a project licence held by Dr. Andrew Murray. Five-week-old female *ob/ob* and wild type (WT) mice were purchased from Harlan Laboratory Inc. (UK). Seven animals were housed per cage and were allowed to acclimatise for 7 days. The temperature was maintained at $(20 \pm 4)^\circ\text{C}$ with a 12 h light/dark cycle.

In *ob/ob* mice, the gene responsible for the production of the hormone leptin is mutated. As a consequence, since such hormone is important in the control of appetite, *ob/ob* mice eat excessively and become obese; they are often used as models for type II diabetes.

Animal Diet and Water

The sample groups are shown in Figure 4.14. Seven *ob/ob* and seven WT mice were fed on regular chow diet (RCD, energy composition: 11.5 % fat, 26.9 % protein, 61.6 % carbohydrate); seven *ob/ob* and seven WT mice were fed on a high fat diet (HFD, energy composition: 55 % fat, 29 % protein, and 16 % carbohydrate). The FA composition of the high fat diet was 25 % polyunsaturated FAs, 48 % monounsaturated FAs, and 27 % saturated FAs. Diets were supplied by Teklad Custom Research Diets (Envigo, Huntingdon, UK).

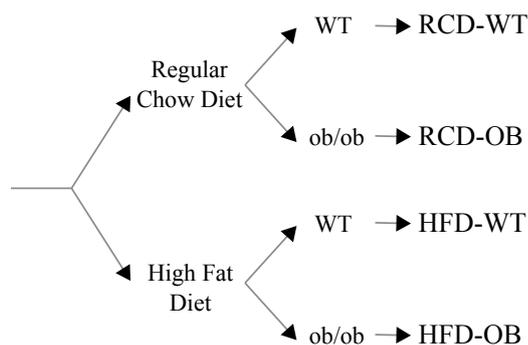


FIGURE 4.14: Grouping of the animals used in the experiment: the 4 groups differ in diet (RCD or HFD) and mouse type (WT or *ob/ob*).

All mice drank water enriched with deuterium adjusted for each group depending on diet and average mouse weight to gain roughly 1% enrichment in the mice's body water. This was maintained for 14 days to allow sufficient enrichment of the body water.

Sample Collection

Animals were killed by CO₂ asphyxiation. The animals were then exsanguinated and dissected to collect hepatic tissue, which was flash frozen in liquid nitrogen; blood was stored at -80°C until extraction.

Sample Extraction and Analysis

The lipids were extracted from both blood and hepatic tissues using a modified version of the Folch method [55], and the LC-MS measurements were taken using a Thermo Exactive Orbitrap mass spectrometer.

The lipid samples were also chromatographically resolved by solid-phase extraction (SPE), and trans-esterified to study the composition of the fatty acid methyl esters (FAMES) by GC-MS. The derivatisation to FAMES was performed only for liver samples. In order to improve the detection sensitivity, the mass scans were performed in selected-ion monitoring (SIM) mode, focusing on the FA chains 14:0, 16:0, 16:1, 18:0, 18:1 and 18:2.

Data Analysis

The LC-MS and GC-MS raw data files were converted into the *.mzML format by means of MsConvert (from ProteoWizard), and then analysed with *IsotopicLabelling*; as for LC-MS, we focused our attention on the abundant PC and TAG species, for which the sensitivity of the approach is higher (see discussion in Section 4.3.3).

4.4.3 LC-MS Data

Considering the way the experiment was planned and carried out, there is low ²H abundance in the feeding water, and therefore we expect the label enrichments to be quite low; Figure 4.15 shows an example of the patterns

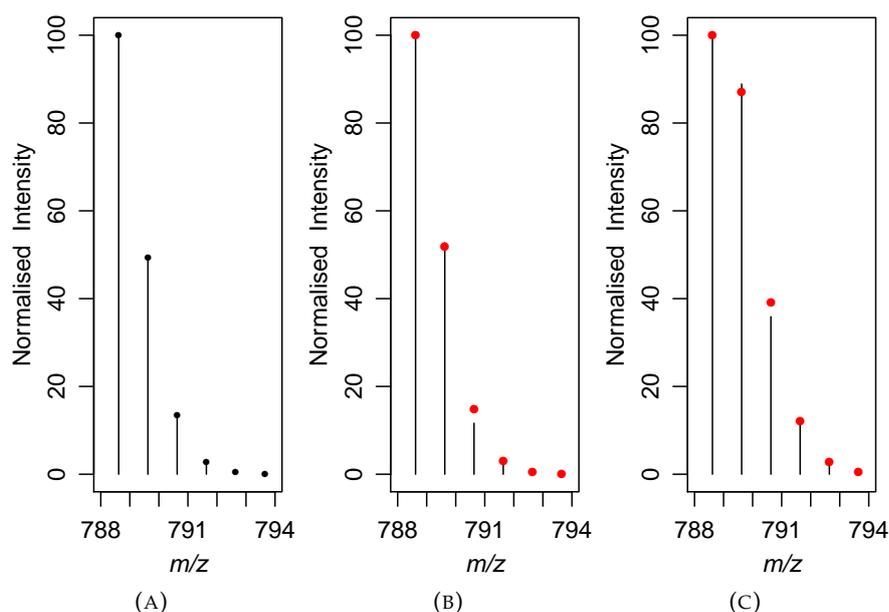


FIGURE 4.15: Comparison between the isotopic patterns of $[\text{PC } 36:1+\text{H}]^+$: (A) natural pattern, (B) experimental pattern of one of the least enriched samples, and (C) experimental pattern of one of the most enriched samples. The red dots are the best-fitted patterns.

we observed for $[\text{PC } 36:1+\text{H}]^+$, whereas Figure 4.16 shows the extracted mass spectra of $[\text{PC } 36:1+\text{H}]^+$ and $[\text{TAG } 54:3+\text{NH}_4]^+$ in one of the samples.

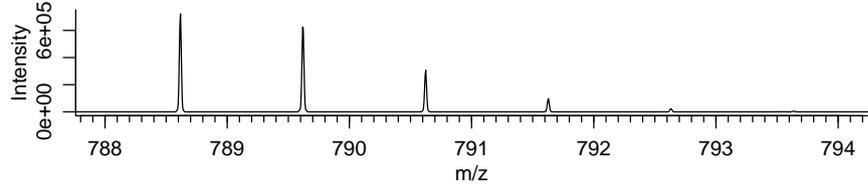
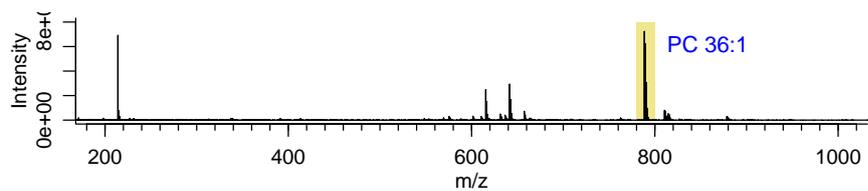
Unlike the previous experimental cases (Section 4.3), here the patterns are only slightly modified following the label incorporation. In spite of this, the analysis by means of *IsotopicLabelling* allowed us to find significant differences across groups: the details can be found in Appendix D.2, whereas in Figure 4.17 are just two representative examples (PC 36:1 and TAG 54:3, liver samples).

What can be seen from our analysis is that the ^2H incorporation depends on both diet and mouse type:

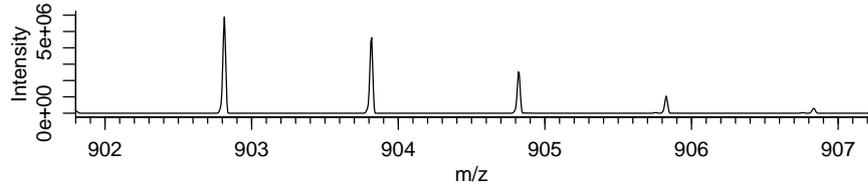
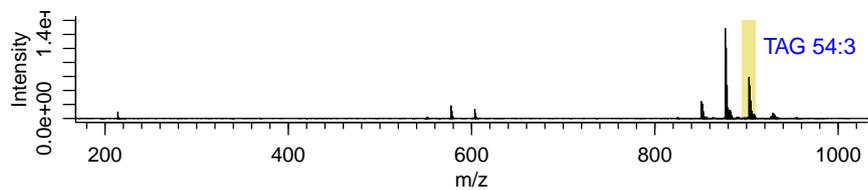
1. **Effect of diet.** RCD is associated with higher ^2H abundances with respect to HFD (for both WT and *ob/ob* mice), therefore suggesting a higher DNL rate in RCD-fed mice;
2. **Effect of mouse type.** The ^2H enrichments are lower in WT mice as compared to *ob/ob* mice, implying that the DNL rate is higher in obese mice.

The p-values from the two-sample t-tests are reported in Appendix D.2. Worthy of note, the above considerations hold true for both liver and blood samples.

We also performed a principal component analysis (PCA) on the data, using the ^2H estimates as original variables (scores plot in Figure 4.18); we obtained a nice separation among sample groups, especially along PC1. Interestingly, the loadings plots (Figure 4.19) point out that all the original variables have negative loadings along PC1, and this is consistent with the fact that HFD-WT, the least enriched samples, are to the left of the scores



(A) Full mass spectrum (top) and isotopic pattern (bottom) of PC 36:1



(B) Full mass spectrum (top) and isotopic pattern (bottom) of TAG 54:3

FIGURE 4.16: Extracted mass spectra for (A) $[\text{PC } 36:1+\text{H}]^+$ (RT: 4.90 min), and (B) $[\text{TAG } 54:3+\text{NH}_4]^+$ (RT: 6.86 min), in the m/z range 200–1000. For each of them, the highlighted area has been enlarged to show the isotopic patterns of the considered lipid species.

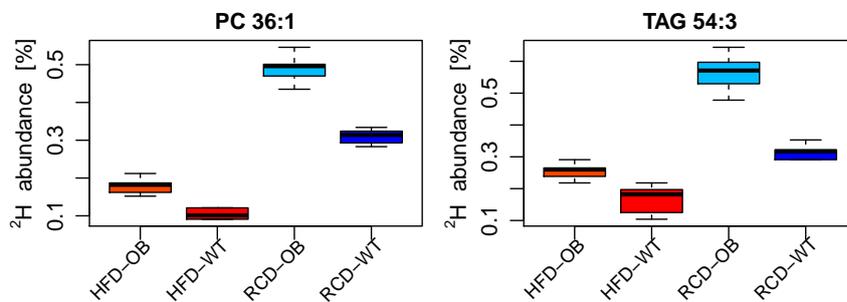


FIGURE 4.17: Comparison between sample groups for PC 36:1 and TAG 54:3, in terms of ^2H abundance in the target analytes. We observed this same trend also for the other target analytes (see Appendix D.2).

plots. Therefore, the separation occurs mainly because of the different ^2H abundances across sample groups.

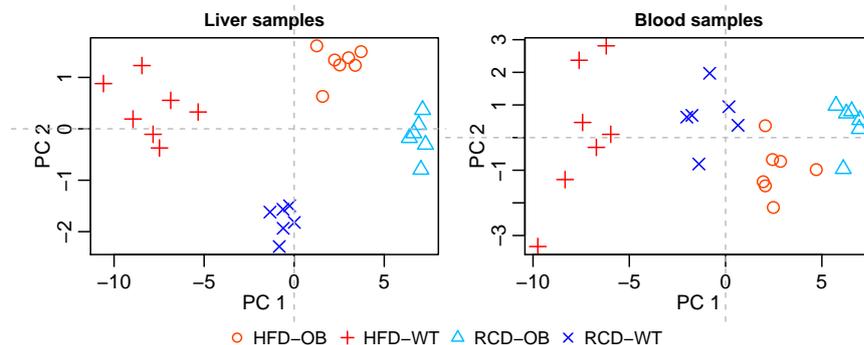


FIGURE 4.18: Scores plot for liver (left) and blood (right) samples. The original variables for the PCA were the estimated ^2H abundances.

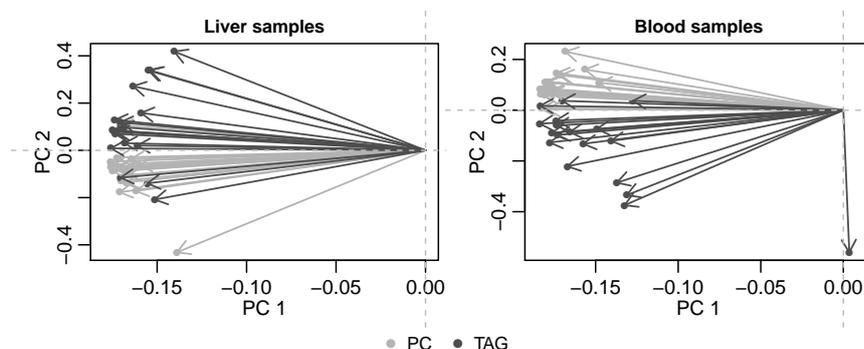


FIGURE 4.19: Loadings plot for liver (left) and blood (right) samples, related to the scores plot of Figure 4.18.

4.4.4 GC-MS Data

We then moved on by analysing the GC-MS data, keeping into consideration the following issues:

1. The samples were derivatised (methyl esterification) prior to analysis, and therefore it is the FAMES that get into the detector;
2. The methyl group which is added through esterification is not labelled, and therefore the elements it contains have fixed natural abundance;
3. The analytes are ionised through electron ionisation (EI), and therefore they are subject to fragmentation. The signal of the molecular ion will have reduced intensity.

In Figure 4.20 are just two of the extracted patterns for FAME C18:0, together with its natural pattern for comparison.

As described above (Section 4.4.2), with GC-MS each sample was split in two: one aliquot was directly derivatised to FAMES, whereas the lipids in the other one were first fractionated into classes (CE, FA, PC, TAG) via

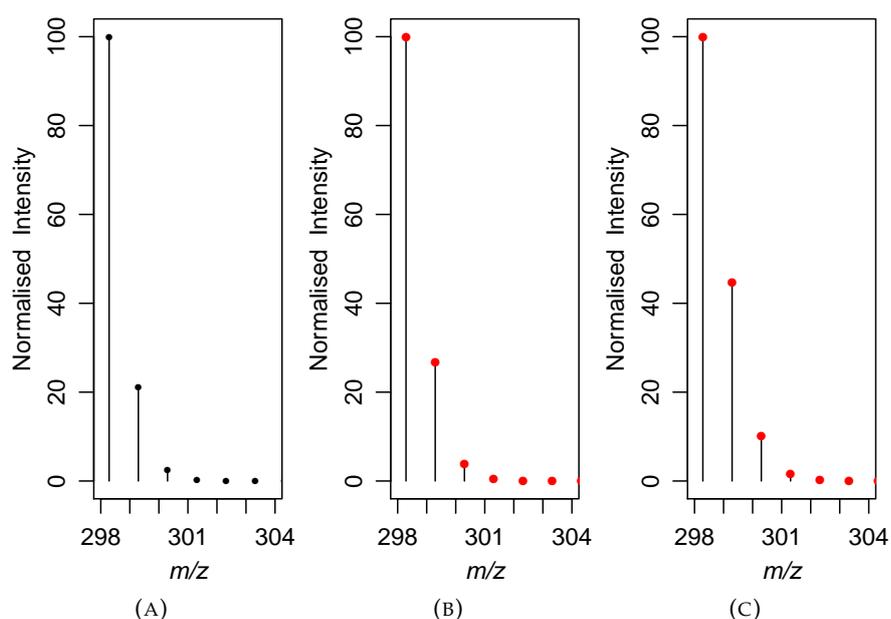


FIGURE 4.20: Isotopic patterns relative to FAME C18:0: (A) natural pattern, (B) experimental pattern of one of the least enriched samples (HFD-WT), and (C) experimental pattern of one of the most enriched samples (RCD-OB). The red dots are the best-fitted patterns.

SPE, and then individually esterified. In Figure 4.21 are the results relative to the total lipid pool, whereas the details of all the other estimates are in Appendix D.3, where also a table detailing the p-values from the two-sample t-tests is reported.

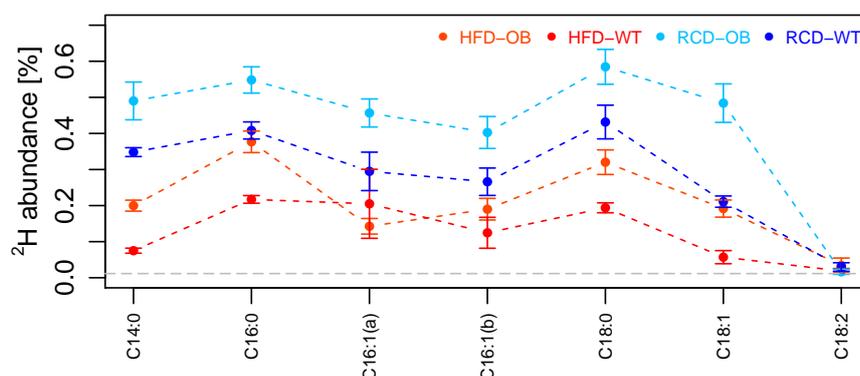


FIGURE 4.21: Estimated ^2H abundances in the FAs of the total lipid pool. The samples are divided into groups, and the horizontal dashed grey line denotes the ^2H natural occurrence (0.0115 %).

Similarly to LC-MS, also here we observed a significant difference between sample groups: both mouse type and diet seem to affect the ^2H occurrence (and therefore the DNL rate), with the lowest ^2H abundances in HFD-WT samples, whereas the highest ones in RCD-OB mice.

It is interesting to note the following: we observed significant differences for all FA species, except FAME C18:2, with an estimated ^2H abundance

very low and close to the natural ^2H occurrence (see both Figure 4.21 and Appendix D.3). This is because C18:2 is an essential FA for mammals, and cannot be synthesised *de novo*: all of it comes only through the diet [56, 57].

To summarise what presented so far, our isotopic pattern analysis allowed us to investigate the DNL process, which turned out to depend on both mouse type (WT, *ob/ob*) and diet (RCD, HFD). Based on the ^2H estimates, we speculate that:

1. The DNL rate is higher for RCD as compared to HFD. This is probably because HFD provides plenty of FAs, which RCD does not; as a consequence, under RCD, DNL is needed to convert some of the carbohydrate fuel into FAs;
2. The DNL rate is higher for *ob/ob* as compared to WT mice. Indeed, in the profoundly obese *ob/ob* mice DNL is probably upregulated, therefore contributing significantly to the building up of the lipid supplies.

4.4.5 Follow-up Experiment

Following the results discussed above, further biological experiments were carried out with slightly different conditions: three diets were here compared, high-sucrose diet (HSD; TD150066), low-fat control diet (LFD; TD08485) and Western Style diet (WD; TD88137). Again, both WT and *ob/ob* mice were used, giving rise to 6 distinct sample groups.

Apart from this modification, all the other experimental conditions were kept the same as those previously described (Section 4.4.2). The aim here was to confirm what observed above, mainly that (i) when considering mouse type, *ob/ob* mice are associated with higher DNL rates than WT mice, and (ii) when considering diet, fat-rich ones are associated with lower DNL rates.

In this case, we only dealt with liver samples and LC-MS data. For the sake of completeness, the new results are reported in Appendix D.4 (plots detailing the estimates for each target analyte, as well as a table of p-values following the two-sample t-tests), whereas in Figure 4.22 is just a representative example (PC 36:1 and TAG 54:3).

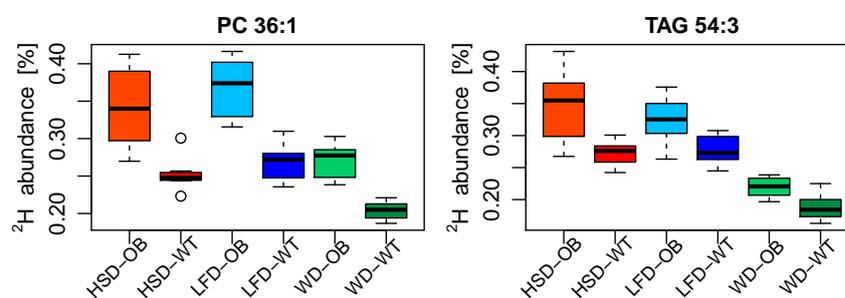


FIGURE 4.22: Representative example with the group comparisons for PC 36:1 and TAG 54:3.

The results shown here can be extended to all the other investigated lipid species. Different shades of the same colour relate are used for the comparison between mouse type (OB: *ob/ob* mice, WT: wild-type mice) with

the same, fixed diet. On the other hand, different colors are used across diets, and the lighter shade is employed for OB mice, whereas the darker one for WT mice. At a glance, it can be immediately noticed how, yet again, obese mice are associated with higher ^2H abundances (and hence higher DNL rates) as compared to wild-type mice. However, the dependence on diet is not so clear as before: as a matter of fact, whereas a clear reduction in ^2H abundance can be observed for Western Style diet (WD) as compared to the other two investigated, no clear difference between high-sucrose and low-fat diets can be seen, which is also confirmed by the t-test analysis shown in Table 4.5. Considering that the low-fat diet serves as “control”, our results imply that a diet characterised by high-sucrose levels do not influence the DNL rate significantly. On the contrary, this rate can be affected by diets rich in fats, such as the Western Style diet.

	EFFECT OF DIET						EFFECT OF MOUSE TYPE		
	HSD VS LFD		HSD VS WD		LFD VS WD		HSD	LFD	WD
	OB/OB	WT	OB/OB	WT	OB/OB	WT			
PC 36:1	3.5×10^{-1}	1.9×10^{-1}	1.6×10^{-2}	2.1×10^{-4}	1.1×10^{-4}	5.7×10^{-5}	5.4×10^{-3}	8.6×10^{-5}	2.0×10^{-5}
TAG 54:3	4.6×10^{-1}	6.4×10^{-1}	1.3×10^{-3}	9.3×10^{-7}	2.0×10^{-5}	9.5×10^{-7}	1.9×10^{-2}	8.1×10^{-3}	4.4×10^{-3}

TABLE 4.5: p-values from the comparison between sample groups; the grey values represent p-values higher than 5%. A table containing information on all the investigated lipid species can be found in Appendix D.4. Notice that the comparison between HSD and LFD results in not significant differences, whereas in the other cases such difference is indeed statistically significant.

To wrap up our most relevant findings, also in this study we found a dependence on both mouse type and diet:

1. **Effect of mouse type.** For each diet, we observed a lower ^2H enrichment in WT mice as compared to *ob/ob* mice, indicative of a higher DNL rate in obese mice;
2. **Effect of diet.** We compared the three different diets pair-wise; whereas the difference between HSD and LFD was not significant, we observed lower ^2H enrichments in WD as compared to both HSD and LFD, suggesting a lower rate of DNL in WD.

4.4.6 Final Considerations

After having successfully tested *IsotopicLabelling*, we exploited it in this biochemical application, where it proved to be reliable. Most importantly, the sensitivity of our approach was high enough to let us tell apart different sample groups, notwithstanding their low ^2H enrichments. This is important because it should always be kept into consideration that heavy water is toxic, and therefore its concentration has to be kept low throughout *in vivo* experiments.

The results obtained here suggest that the DNL rate can be affected by diet, in particular the higher the amount of fats in the dietary regime, the lower the DNL rate. Also, we observed a difference between lean and obese mice, these latter having a higher DNL rate.

A potential complication worth taking into consideration is the toxicity of D_2O itself: in principle, we should not rule out that some metabolic

effects may arise as a consequence of such toxicity, including variations of the DNL rate. On the other hand, even assuming that D₂O does indeed modify the DNL rate, such effect would not overturn our main conclusions, considering that we have treated different groups homogeneously in terms of D₂O administration. As a consequence, the resulting effect would be the same across groups, and therefore different diet and mouse type would still account for the observed group differences.

Of course, we do recognise that it would be helpful to further strengthen our findings with a parallel investigation; a possible solution may be to assess the expression level of key enzymes involved in the DNL process, including FASN and SCD (see Figure 4.2). This is a point that should be taken into consideration for future studies.

Bibliography of Chapter 4

1. Chokkathukalam, A., Kim, D., Barrett, M. P., Breitling, R. & Creek, D. J. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis* **6**, 511–524 (2014).
2. You, L., Zhang, B. & Tang, Y. J. Application of stable isotope-assisted metabolomics for cell metabolism studies. *Metabolites* **4**, 142–165 (2014).
3. Lee, W. N. P. Characterizing phenotype with tracer based metabolomics. *Metabolomics* **2**, 31–39 (2006).
4. Birkemeyer, C., Luedemann, A., Wagner, C., Erban, A. & Kopka, J. Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. *Trends in Biotechnology* **23**, 28–33 (2005).
5. Klein, S. & Heinzle, E. Isotope labeling experiments in metabolomics and fluxomics. *WIREs Systems Biology and Medicine* **4**, 261–272 (2012).
6. Beylot, M., Peroni, O., Diraison, F. & Large, V. In vivo studies of intra-hepatic metabolic pathways. *Diabetes & Metabolites* **23**, 251–257 (1997).
7. Diraison, F., Moulin, P. & Beylot, M. Contribution of hepatic de novo lipogenesis and reesterification of plasma non esterified fatty acids to plasma triglyceride synthesis during non-alcoholic fatty liver disease. *Diabetes & Metabolites* **29**, 478–485 (2003).
8. Strawford, A., Antelo, F., Christiansen, M. & Hellerstein, M. K. Adipose tissue triglyceride turnover, de novo lipogenesis, and cell proliferation in humans measured with $2\text{H}_2\text{O}$. *AJP: Endocrinology and Metabolism* **286**, E577–E588 (2004).
9. Fan, T. W. *et al.* Stable isotope-resolved metabolomics and applications for drug development. *Pharmacology & Therapeutics* **133**, 366–391 (2012).
10. Budzikiewicz, H. & Grigsby, R. D. Mass spectrometry and isotopes: a century of research and discussion. *Mass Spectrometry Reviews* **25**, 146–157 (2006).
11. Thomson, J. J. Rays of positive electricity. *Proceedings of the Royal Society A* **89**, 1–20 (1913).
12. Aston, F. W. The constitution of the elements. *Nature* **104**, 393 (1919).
13. Murray, K. K. *et al.* Definitions of terms relating to mass spectrometry (IUPAC recommendations 2013). *Pure and Applied Chemistry* **85**, 1515–1609 (2013).
14. Gross, J. H. *Mass spectrometry – A textbook* (Springer, 2011).
15. Smith, R. M. *Understanding mass spectra – A basic approach* (Wiley-Interscience, 2004).
16. Yergey, J. A. A general approach to calculating isotopic distributions for mass spectrometry. *International Journal of Mass Spectrometry and Ion Physics* **52**, 337–349 (1983).

17. Valkenborg, D., Mertens, I., Lemi re, F., Witters, E. & Burzykowski, T. The isotopic distribution conundrum. *Mass Spectrometry Reviews* **31**, 96–109 (2012).
18. Rockwood, A. L. Relationship of Fourier transforms to isotope distribution calculations. *Rapid Communication in Mass Spectrometry* **9**, 103–105 (1995).
19. Rockwood, A. L., Van Orden, S. L. & Smith, R. D. Rapid calculation of isotope distributions. *Analytical Chemistry* **67**, 2699–2704 (1995).
20. Rockwood, A. L., Van Orden, S. L. & Smith, R. D. Ultrahigh resolution isotope distribution calculations. *Rapid Communications in Mass Spectrometry* **10**, 54–59 (1996).
21. Fernandez-de Cossio, J. Computation of the isotopic distribution in two dimensions. *Analytical Chemistry* **82**, 6726–6729 (2010).
22. Fernandez-de Cossio, J. Efficient packing Fourier-transform approach for ultrahigh resolution isotopic distribution calculations. *Analytical Chemistry* **82**, 1759–1765 (2010).
23. Ipsen, A. Efficient calculation of exact fine structure isotope patterns via the multidimensional Fourier transform. *Analytical Chemistry* **86**, 5316–5322 (2014).
24. Ameer, F., Scandiuzzi, L., Hasnain, S., Kalbacher, H. & Zaidi, N. De novo lipogenesis in health and disease. *Metabolism* **63**, 895–902 (2014).
25. Menendez, J. A., Vazquez-Martin, A., Ortega, F. J. & Fernandez-Real, J. M. Fatty acid synthase: association with insulin resistance, type 2 diabetes, and cancer. *Clinical Chemistry* **55**, 425–438 (2009).
26. Moore, J. B., Gunn, P. J. & Fielding, B. A. The role of dietary sugars and de novo lipogenesis in non-alcoholic fatty liver disease. *Nutrients* **6**, 5679–5703 (2014).
27. Lambert, J. E., Ramos-Roman, M. A., Browning, J. D. & Parks, E. J. Increased de novo lipogenesis is a distinct characteristic of individuals with nonalcoholic fatty liver disease. *Gastroenterology* **146**, 726–735 (2014).
28. Lam, S. M. & Shui, G. Lipidomics as a principal tool for advancing biomedical research. *Journal of Genetics and Genomics* **40**, 375–390 (2013).
29. Musso, G., Gambino, R. & Cassader, M. Recent insights into hepatic lipid metabolism in non-alcoholic fatty liver disease (NAFLD). *Progress in Lipid Research* **48**, 1–26 (2009).
30. Sanders, F. W. B. & Griffin, J. L. De novo lipogenesis in the liver in health and disease: more than just a shunting yard for glucose. *Biological Reviews of the Cambridge Philosophical Society* **91**, 452–468 (2016).
31. Hellerstein, M. K. No common energy currency: de novo lipogenesis as the road less traveled. *The American Journal of Clinical Nutrition* **74**, 707–708 (2001).
32. Hellerstein, M. K. De novo lipogenesis in humans: metabolic and regulatory aspects. *European Journal of Clinical Nutrition* **53**, S53–S65 (1999).

33. Schwarz, J.-M., Linfoot, P., Dare, D. & Aghajanian, K. Hepatic de novo lipogenesis in normoinsulinemic and hyperinsulinemic subjects consuming high-fat, low-carbohydrate and low-fat, high-carbohydrate isoenergetic diets. *The American Journal of Clinical Nutrition* **77**, 43–50 (2003).
34. Jump, D. B. *et al.* Fatty acid regulation of hepatic gene transcription. *Journal of Nutrition* **135**, 2503–2506 (2005).
35. Hellerstein, M. K. *et al.* Measurement of de novo hepatic lipogenesis in humans using stable isotopes. *The journal of clinical investigation* **87**, 1841–1852 (1991).
36. Hellerstein, M. K., Schwarz, J. M. & Neese, R. A. Regulation of hepatic de novo lipogenesis in humans. *Annual Review of Nutrition* **16**, 523–557 (1996).
37. Murphy, E. J. Stable isotope methods for the in vivo measurement of lipogenesis and triglyceride metabolism. *Journal of Animal Science* **84**, E94–E104 (2006).
38. Millard, P., Letisse, F., Sokol, S. & Portais, J.-C. IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics* **28**, 1294–1296 (2012).
39. Chokkathukalam, A. *et al.* mzMatch-ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics* **29**, 281–283 (2013).
40. Huang, X. *et al.* X¹³CMS: global tracking of isotopic labels in untargeted metabolomics. *Analytical Chemistry* **86**, 1632–1639 (2014).
41. Capellades, J. *et al.* geoRge: a computational tool to detect the presence of stable isotope labeling in LC/MS-based untargeted metabolomics. *Analytical Chemistry* **88**, 621–628 (2016).
42. R Development Core Team. *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).
43. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using non-linear peak alignment, matching, and identification. *Analytical Chemistry* **78**, 779–787 (2006).
44. Kind, T. & Fiehn, O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **7**, 234 (2006).
45. Berglund, M. & Wieser, M. E. Isotopic compositions of the elements 2009 (IUPAC technical report). *Pure and Applied Chemistry* **83**, 397–410 (2011).
46. Baenke, F., Peck, B., Miess, H. & Schulze, A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Disease Models & Mechanisms* **6**, 1353–1363 (2013).
47. Guo, Z. K., Cella, L. K., Baum, C., Ravussin, E. & Schoeller, D. A. De novo lipogenesis in adipose tissue of lean and obese women: application of deuterated water and isotope ratio mass spectrometry. *International Journal of Obesity* **24**, 932–937 (2000).

48. McDevitt, R. M. *et al.* De novo lipogenesis during controlled overfeeding with sucrose or glucose in lean and obese women. *American Journal of Clinical Nutrition* **74**, 737–746 (2001).
49. Turner, S. M. *et al.* Measurement of TG synthesis and turnover in vivo by $^2\text{H}_2\text{O}$ incorporation into the glycerol moiety and application of MIDA. *American Journal of Physiology – Endocrinology And Metabolism* **285**, E790–E803 (2003).
50. Diraison, F., Pachiaudi, C. & Beylot, M. In vivo measurement of plasma cholesterol and fatty acid synthesis with deuterated water: determination of the average number of deuterium atoms incorporated. *Metabolism: Clinical and Experimental* **45**, 817–821 (1996).
51. Leitch, C. A. & Jones, P. J. H. Measurement of human lipogenesis using deuterium incorporation. *Journal of Lipid Research* **34**, 157–163 (1993).
52. Dufner, D. & Previs, S. F. Measuring in vivo metabolism using heavy water. *Current Opinion in Clinical Nutrition and Metabolic Care* **6**, 511–517 (2003).
53. Silva, A. M., Martins, F., Jones, J. G. & Carvalho, R. $^2\text{H}_2\text{O}$ incorporation into hepatic acetyl-CoA and de novo lipogenesis as measured by Krebs cycle-mediated ^2H -enrichment of glutamate and glutamine. *Magnetic Resonance in Medicine* **66**, 1526–1530 (2011).
54. Parks, E. J. Recent advances in liver triacylglycerol and fatty acid metabolism using stable isotope labeling techniques. *The Journal of Lipid Research* **47**, 1651–1660 (2006).
55. Folch, J., Lees, M. & Stanley, G. H. S. A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry* **226**, 497–509 (1957).
56. Rustan, A. C. & Drevon, C. A. in *Encyclopedia of Life Sciences* 1–7 (John Wiley & Sons, Ltd, 2005). <<http://orca.cf.ac.uk/24620/http://doi.wiley.com/10.1038/npg.els.0003894>>.
57. Das, U. Essential fatty acids – A review. *Current Pharmaceutical Biotechnology* **7**, 467–482 (2006).

Appendix A

Automatic Interpretation of NMR Spectra of Lipids

A.1 The Main Challenges

Considering that the manual interpretation of NMR spectra of lipid mixtures is not straightforward, our idea was to understand whether it was possible to devise an informatics tool able to ease this task. Ideally, such a device should automatically interpret NMR spectra, providing the user with an output table detailing the quantitative composition of the lipid mixture. However, soon after we set out developing it, we found ourselves facing a series of issues (critically described below) that made us realise that such ideal software could not actually be developed. What we did instead was to implement a tool that may at least help in the interpretation.

The first critical point to take into account is that lipids, from an NMR point of view, are complex molecules: interpretation of NMR spectra of lipids is therefore a complex issue, especially when lipid mixtures are considered, which may display signal overlaps. This is the case, for example, for the shared structural features of different lipids, such as the FA tails. On the other hand, lipids also possess class-specific domains, which are basically related to their polar heads. As a matter of fact, the approach of exploiting characteristic lipid resonances in order to analyse mixtures by NMR is not new, and a successful ^1H NMR-based method has been recently described for microalgal lipid extracts (Nuzzo, N. *et al.* Composition and Quantitation of Microalgal Lipids by ERETIC ^1H NMR Method. *Marine drugs* **11**,3742–3753 (2013)). We therefore tried to focus on these characteristic resonances, and we first tried to understand whether their signals fall in overlap-free regions: only in this case can such signals be exploited to get qualitative and quantitative information about the lipid composition of the investigated samples. To pursue this aim, we just considered the most common lipid classes occurring in cell cultures and biological tissues, and we tried to superimpose in a synthetic NMR spectrum only their characteristic resonances: the outcome is in Figure A.1, where for each resonance the fine structure is shown. Also highlighted in the figure are three regions common to all lipid classes, arising from the FA chains. Additional details can be found in Table A.1.

What can be immediately noticed from Figure A.1 is that, despite our focused interest on a limited number of species and on their characteristic resonances, the crowding is nonetheless evident: there are just very few analytes whose resonances fall in free regions, such as plasmenyl and sterol species.

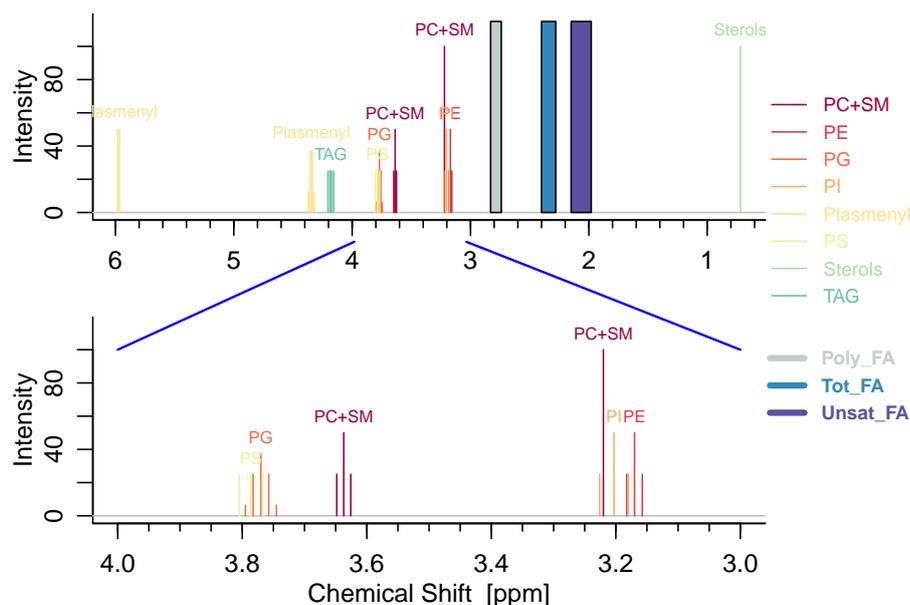


FIGURE A.1: Synthetic NMR spectrum showing characteristic resonances of the most common lipid classes, with a zoom in of the crowded 2–3 ppm chemical shift region. Many are the signal overlaps, making it difficult to unequivocally assign the resonances.

To the issues mentioned so far, there should be added a further complication: we have only shown specific resonances, leaving behind all the others. Real NMR spectra are much more complex, as shown in Figure A.2, where we have reported experimental spectra of some lipid classes; for each of them, the characteristic resonances are highlighted in yellow. The figure clearly reveals that diagnostic signals make up just a small fraction of each spectrum; in addition, in most cases the projection of the yellow regions into the other spectra will fall on other resonances, therefore confirming that a lipid mixture will give rise to a spectrum characterised by many signal overlaps.

Another relevant problem to be mentioned is that each lipid class can occur in a variety of structures depending on the FA chains, and this is better shown in Figure A.3. Importantly, even though it is possible to detect, say, plasmeyl species based on their characteristic resonances, it is impossible to tell which lipid class has those signals (PC, PE, ...).

Finally, an added issue to be always aware of is that proton resonances are not fixed, but their exact chemical shift also depends on parameters such as analyte concentration and ionic strength of the solution. Unfortunately, such dependence is not linear and close resonances may even exchange relative positions.

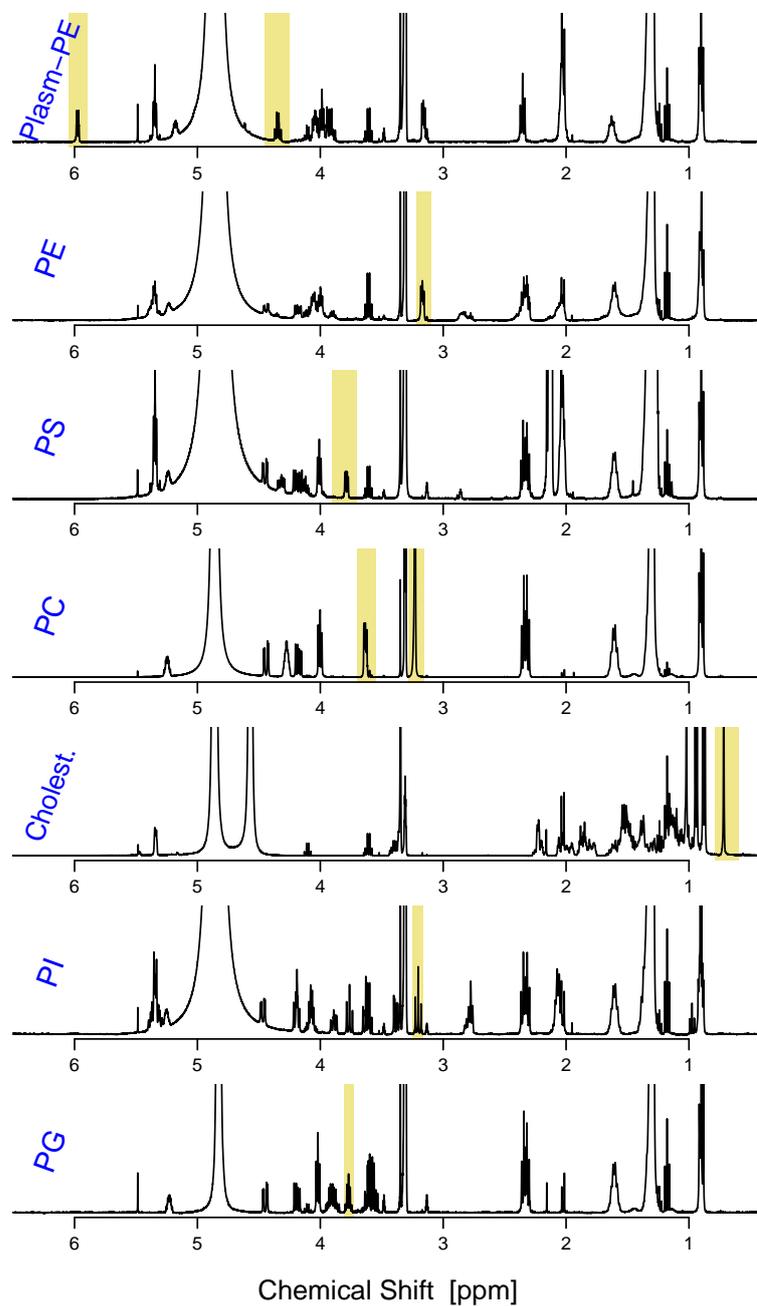


FIGURE A.2: Experimental NMR spectra of the most common lipid classes. Yellow rectangles highlight the characteristic resonances of each class.

SPECIES	SHIFT	MULTIPLLET	PROTONS
Plasmenyl	5.973	d (6.2 Hz)	1
Sphingolipids	5.71	dt (15, 7 Hz)	1
Plasmenyl	4.345	q (6.2 Hz)	1
TAG	4.18	dd (7, 12 Hz)	2
PS	3.83	dd (7.3, 7.6 Hz)	1
PG	3.77	quin (5 Hz)	1
PC + SM	3.637	t (4.5 Hz)	2
PC	3.22	s	9
SM	3.21	s	9
PI	3.203	t (9.2 Hz)	1
PE	3.17	t (5 Hz)	2
Sterols	0.72	s	3
Poly_FA	2.74–2.83	m	2
Tot_FA	2.28–2.40	m	2
Unsat_FA	1.98–2.15	m	4

TABLE A.1: Details of the main lipid resonances.

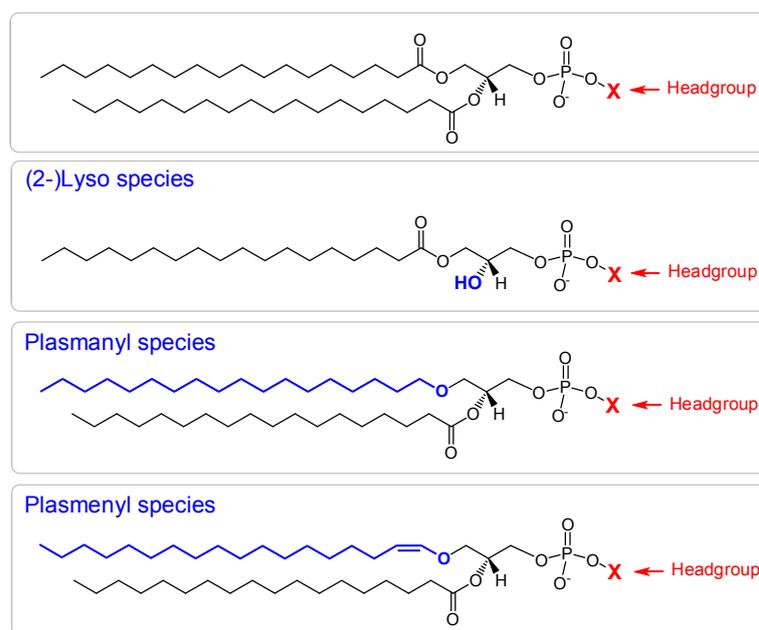


FIGURE A.3: Example showing that within each lipid class a variety of structures can exist.

A.2 Block Diagram of the Processing Steps

Bearing in mind the issues described above, we had to scale back our initial aims, and we tried to devise a tool that may assist the researcher when it comes to inspecting NMR spectra. The basic idea is quite simple, and briefly summarised in Figure A.4.

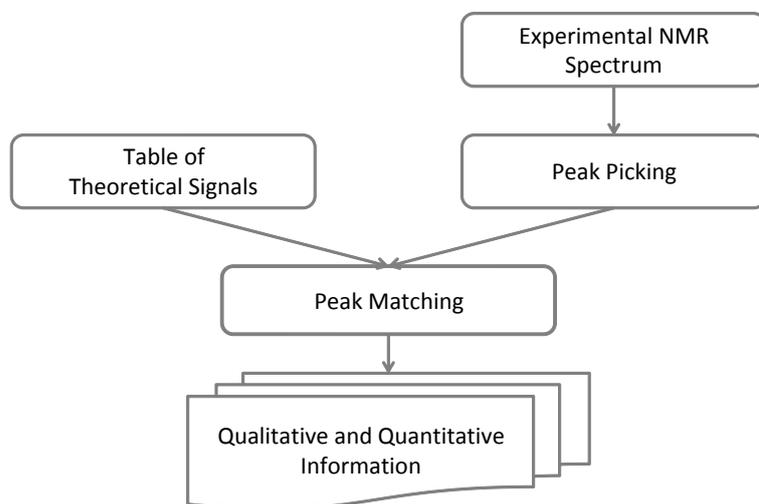


FIGURE A.4: Block diagram of the informatics tool aimed at helping in the interpretation of NMR spectra of lipid mixtures.

- The starting point is to provide a table of theoretical NMR resonances, similar to the one shown in Table A.1: for each lipid class, the user specifies the chemical shift of the diagnostic resonance(s), as well as the details of the fine structure (multiplicity and coupling constants);
- Next, the user inputs the experimental NMR spectrum (already Fourier-transformed, phase- and baseline-corrected, and referenced along the chemical shift scale). A peak-picking procedure ensues, resulting in a list of peaks with related intensities;
- The peak matching process takes place: for each theoretical resonance, the tool detects whether or not it is present in the experimental spectrum, based on both chemical shift (with an accuracy that can be set by the user) and on the fine structure (both coupling constant and relative intensity among the signals are considered);
- After the peak matching, each analyte of the theoretical table is listed, together with the signal intensities found (if any). In order to provide quantitative information, the signals are normalised based on the number of protons giving rise to them;
- Also a set of plots is produced, one for each theoretical signal, where the experimental spectrum in the considered region is shown together with the theoretical resonances and possible matched peaks.

As for the obtained relative intensities, it is important to point out that NMR is intrinsically quantitative, and therefore these quantities directly

reflect the relative molar abundance of the lipid classes, without the need to use any quantitative standard.

A.3 An Example of the Results

We tried to run our tool over some experimental NMR spectra of lipid mixtures, and in Figure A.5 are just a few examples of the graphical output: by comparing the experimental patterns (black noisy lines) with the theoretical expected resonances (red vertical lines), one can immediately understand whether or not the considered class is present, as well as if any overlap with other resonances occurs. For example, what can be said from the results in

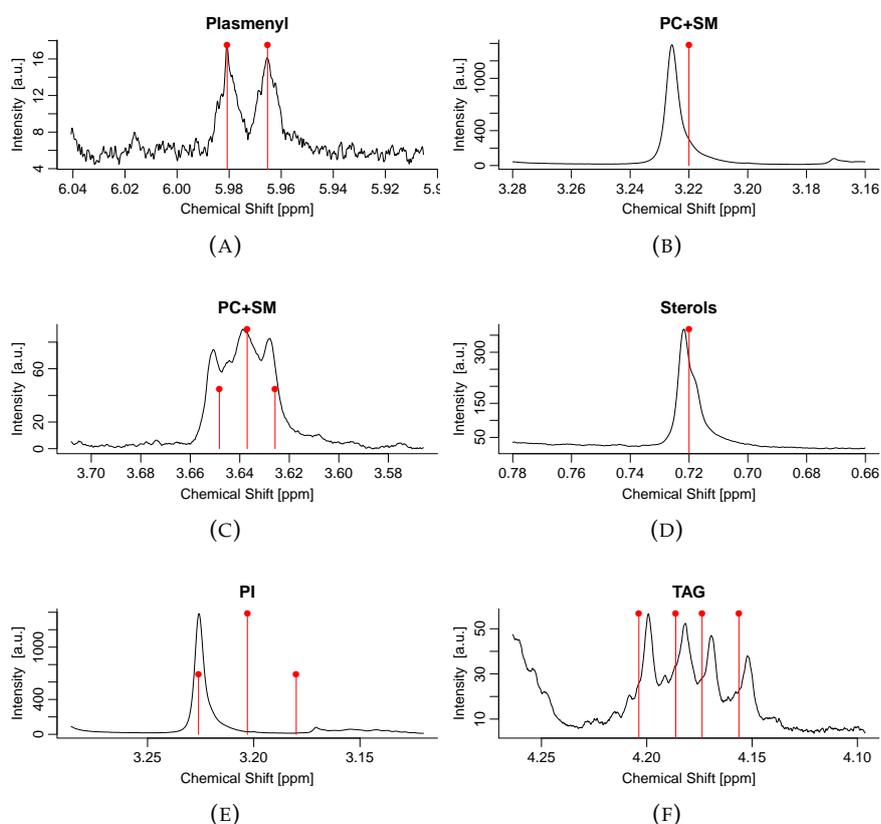


FIGURE A.5: Examples of the graphical outputs produced. The experimental spectrum is dissected and the characteristic regions are plotted together with the theoretical resonances (red vertical lines); it is up to the researcher to understand whether or not the considered class is present.

Figure A.5 is the following:

- (A) Plasmenyl species are present in the sample;
- (B) Also PC+SM are present, although the theoretical and experimental shifts are slightly different;
- (C) The same as above. However, it should be noticed that the pattern is more complex than the expected simple triplet, probably because PC and SM resonances do not perfectly overlap;

- (D) Sterols are clearly present. This is a lucky region, since usually no other lipids have so shielded resonances;
- (E) PI are not present, since the expected triplet cannot be observed;
- (F) Also TAG are present, although there is a slight mismatch between expected and experimental shifts.

A.4 Acyl Chain Composition

Interestingly, in addition to class-specific information, also insights into the FA composition can be obtained, especially by looking at the 2–3 ppm region of the NMR spectrum: the degree of FA unsaturation can be easily obtained, as well as knowledge about the presence of poly-unsaturations. Of course, this will just give an average figure, considering that all acyl chains are taken into account, no matter of the lipid class they belong to.

Unlike the specific resonances, here the signals are very broad and their fine structure is lost, since they are the result of the overlap of many different chains. As a consequence, instead of considering the intensity of the detected peaks, the tool integrates the signals over the specified region (in the input table it is just necessary to provide a chemical shift range, see the last rows of Table [A.1](#)).

Appendix B

Appendix of Chapter 2

B.1 Primary Cortical Neurons: Intra-class Profiles

The following box-plots represent the comparison between sample groups (CTRL versus CBE-TRTD) in terms of the parameters average chain length (Figure B.1), average number of unsaturations (Figure B.2) and unsaturation index (Figure B.3). There are no significant differences, implying that the treatment is not acyl-chain specific.

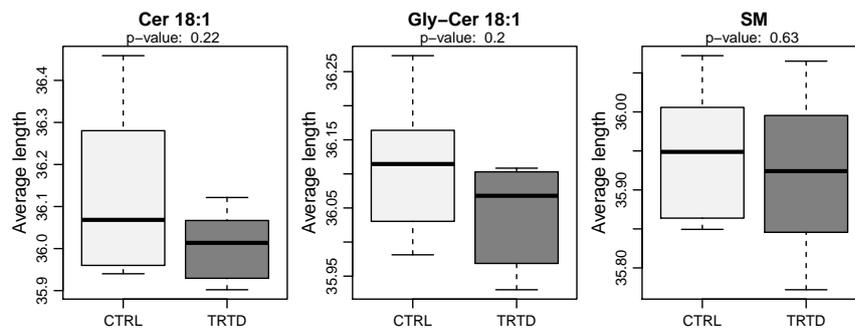


FIGURE B.1: Comparison of the average chain length for each sphingolipid class. There are no significant differences between CTRL and CBE-TRTD samples.

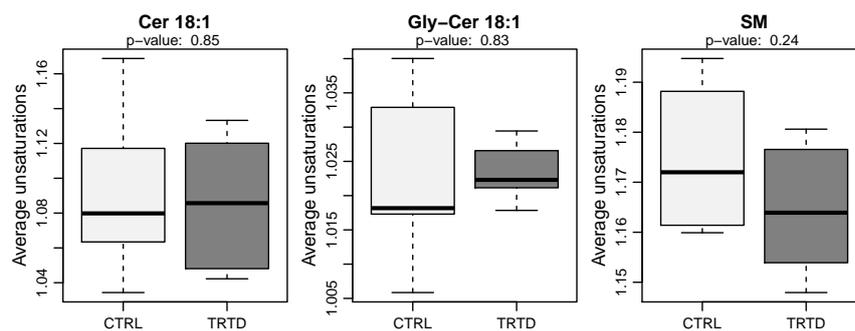


FIGURE B.2: Comparison of the average number of unsaturations for each sphingolipid class. There are no significant differences between CTRL and CBE-TRTD samples.

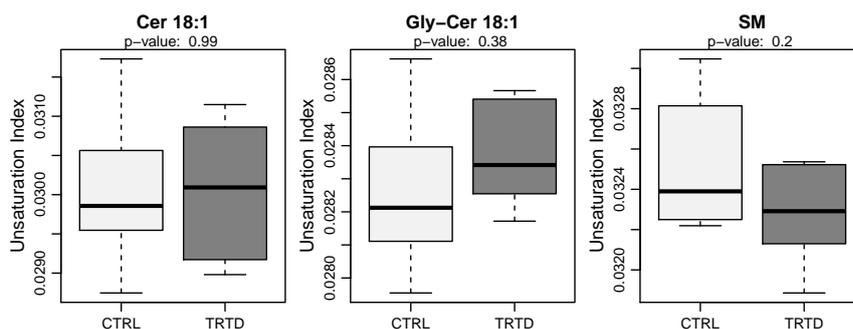


FIGURE B.3: Comparison of the unsaturation index for each sphingolipid class. There are no significant differences between CTRL and CBE-TRTD samples.

B.2 $Lrrk2^{+/+}$ and $Lrrk2^{-/-}$ Mice: Intra-class Profiles

Below are three series of box-plots where the comparison is made between sample groups ($Lrrk2^{+/+}$ versus $Lrrk2^{-/-}$ mice) in terms of the parameters average chain length (Figure B.4), average number of unsaturations (Figure B.5) and unsaturation index (Figure B.6). The lack of any significant difference implies that the observed sphingolipid perturbations are not acyl-chain specific.

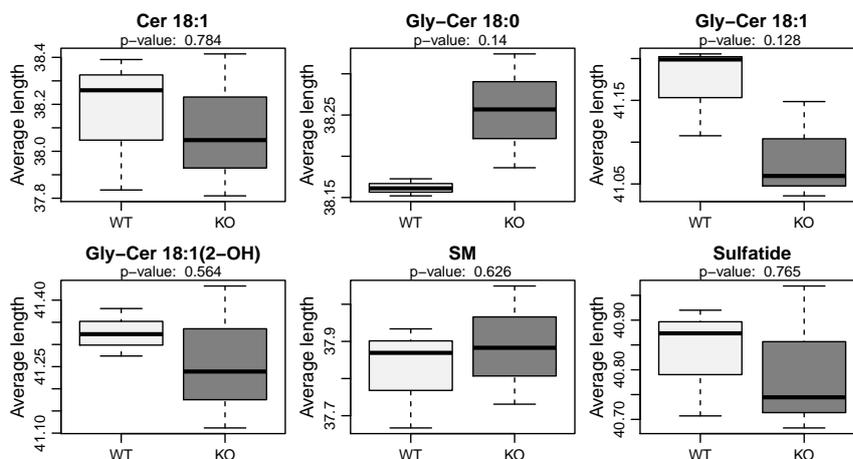


FIGURE B.4: Comparison of the average chain length between sample groups; there are no significant differences.

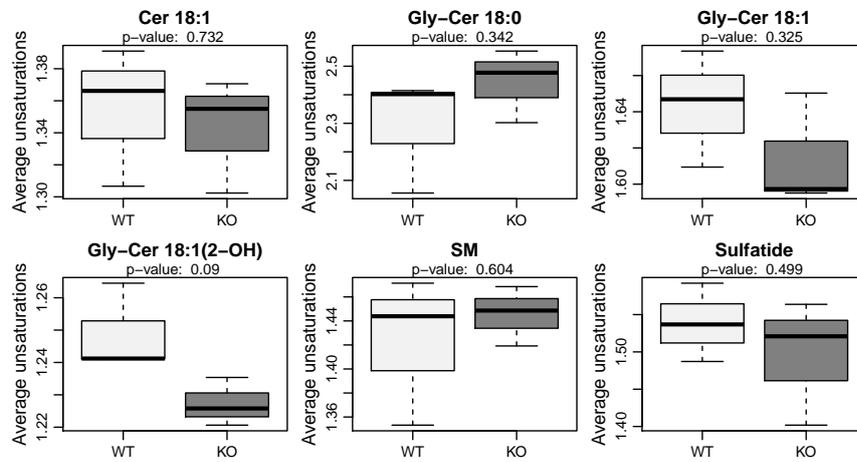


FIGURE B.5: Comparison of the average number of unsaturations between sample groups; there are no significant differences.

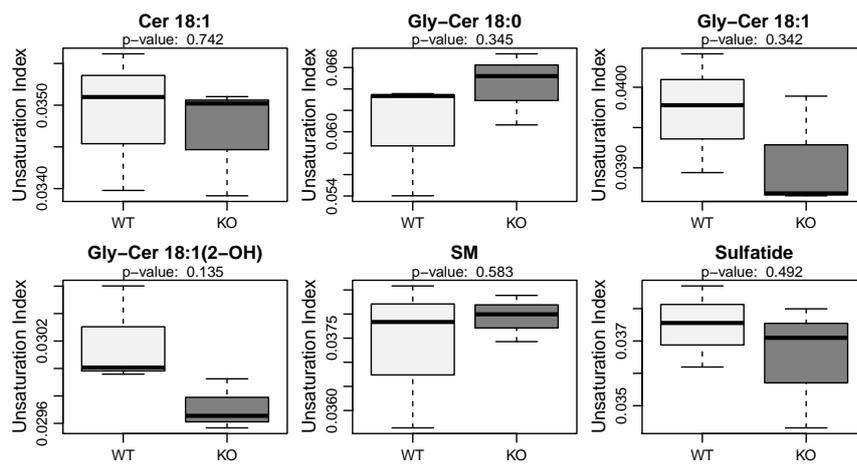


FIGURE B.6: Comparison of the unsaturation index between sample groups; there are no significant differences.

Appendix C

Appendix of Chapter 3

C.1 Intra-class Distribution for IGROV-1

The following barplots (Figure C.1 to Figure C.5) represent the intra-class profiles obtained for each lipid class and sample group. In order to ease the visualisation, only relative amounts higher than 2% are shown. The statistically significant differences are properly pointed out (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$).

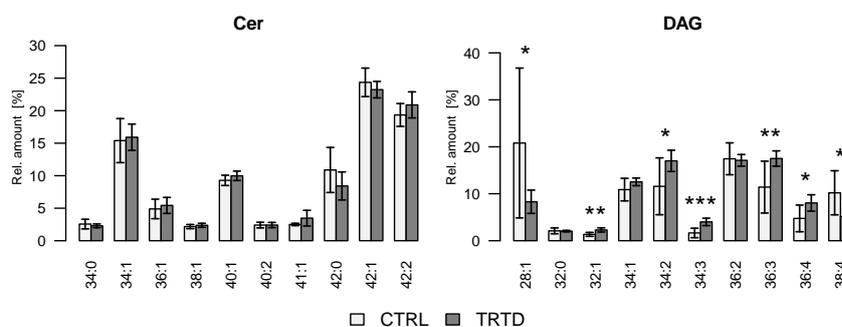


FIGURE C.1: IGROV-1 – Relative distribution of Cer and DAG (threshold: 2%).

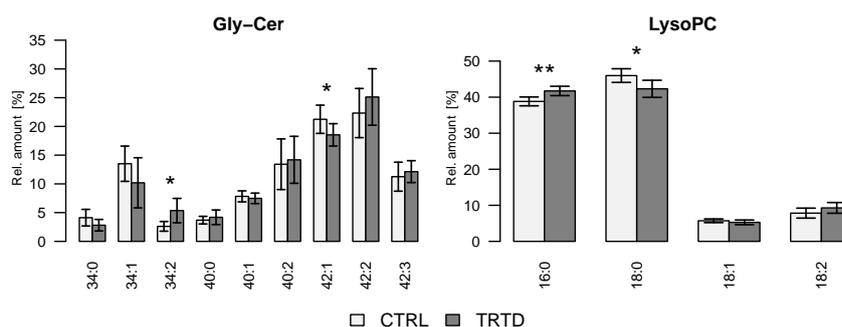


FIGURE C.2: IGROV-1 – Relative distribution of Gly-Cer and LysoPC (threshold: 2%).

Figure C.6 shows the group comparison of the average FA chain length for each of the investigated lipid classes, whereas in Figure C.7 is the average number of unsaturations and in Figure C.8 is the unsaturation index.

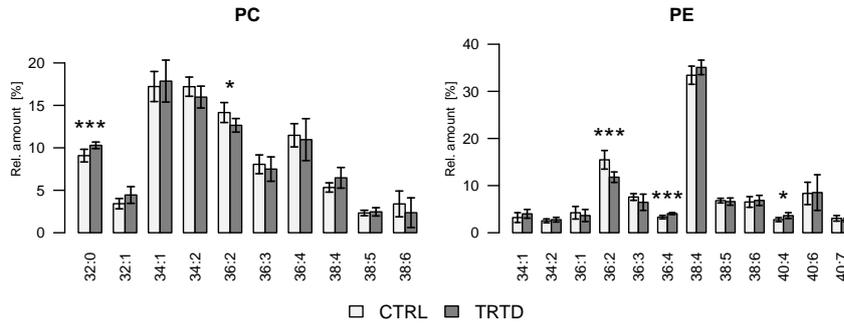


FIGURE C.3: IGROV-1 – Relative distribution of PC and PE (threshold: 2%).

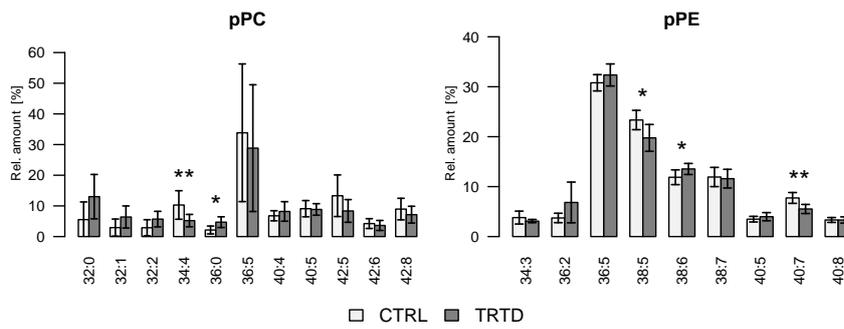


FIGURE C.4: IGROV-1 – Relative distribution of pPC and pPE (threshold: 2%).

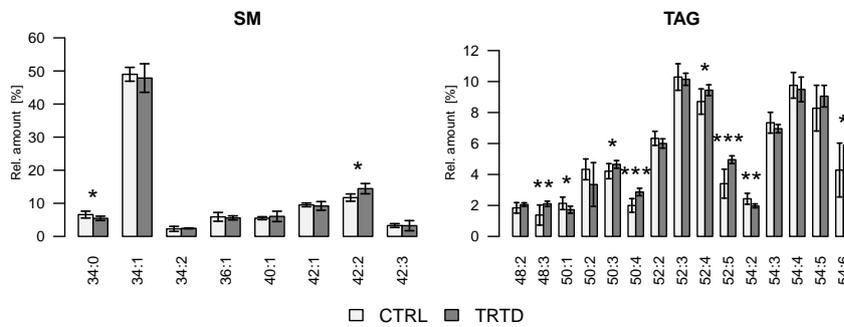


FIGURE C.5: IGROV-1 – Relative distribution of SM and TAG (threshold: 2%).

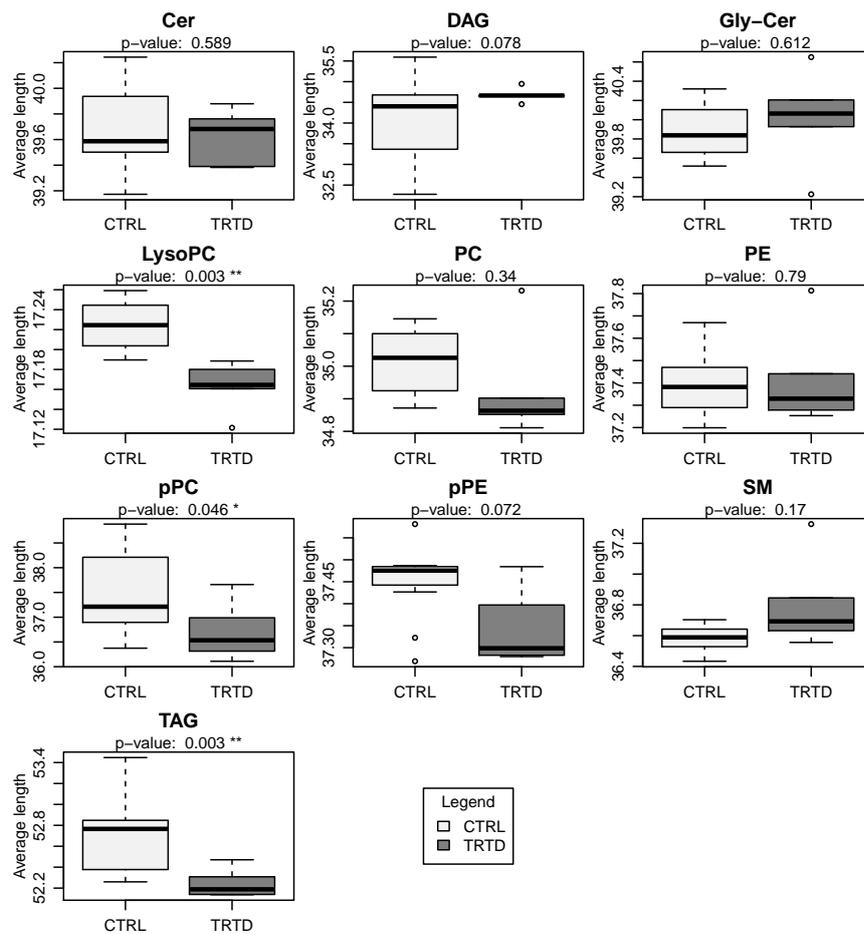


FIGURE C.6: IGROV-1 – Comparison of the average chain length.

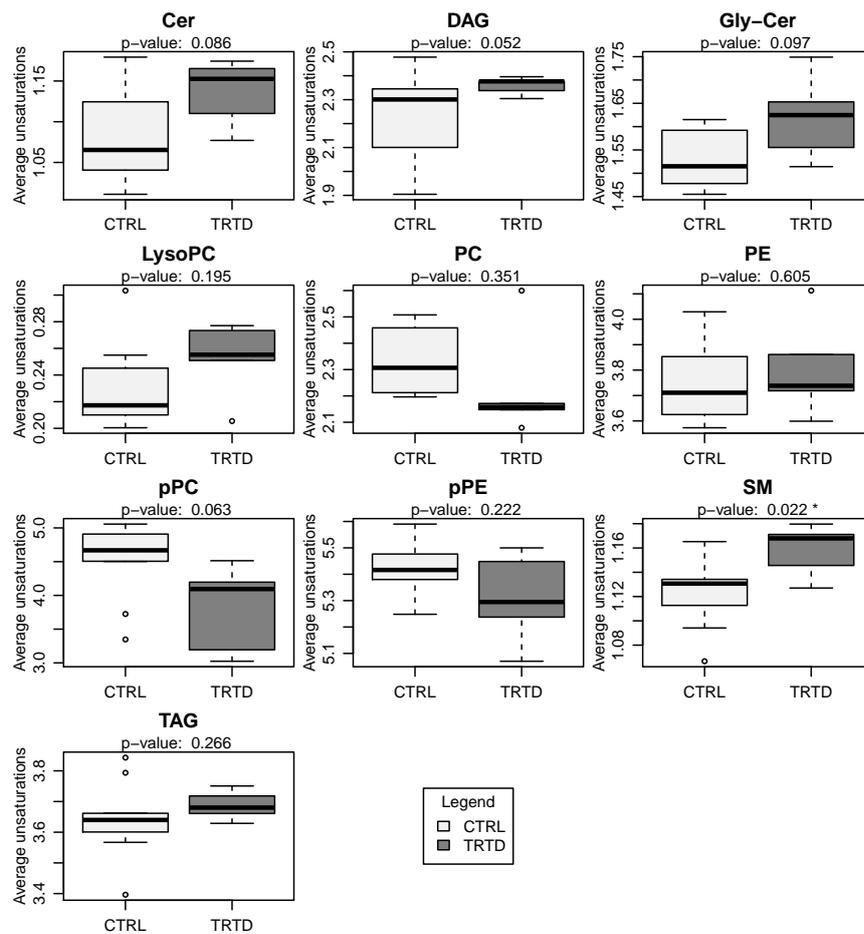


FIGURE C.7: IGROV-1 – Comparison of the average number of unsaturations.

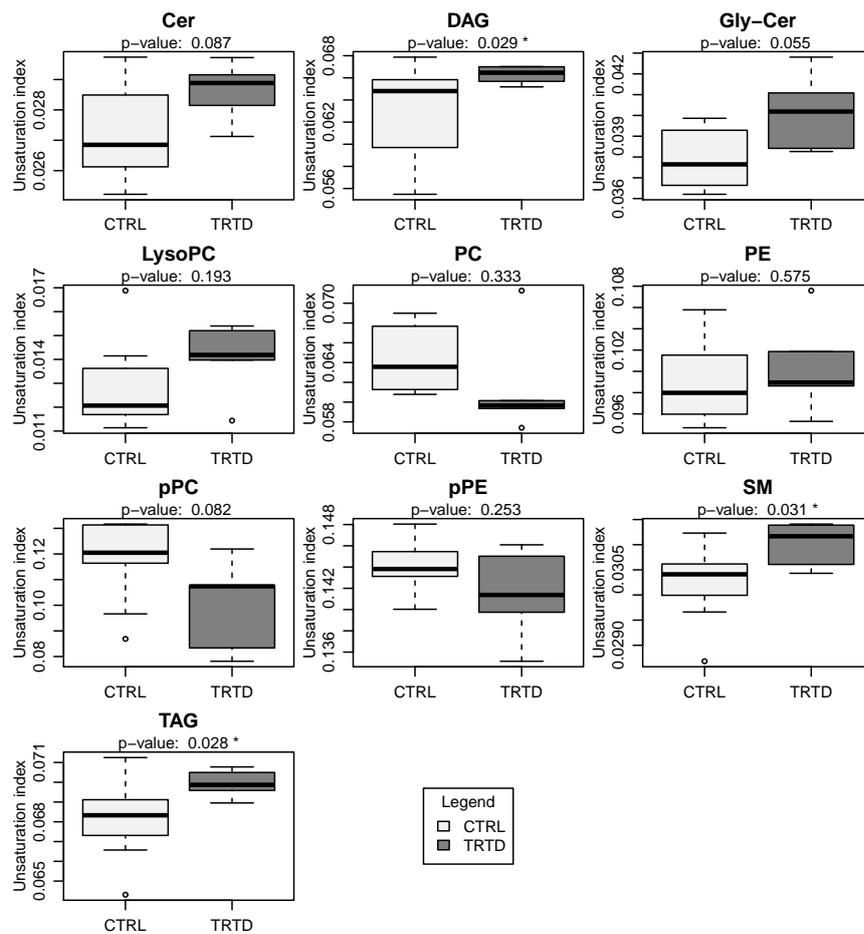


FIGURE C.8: IGROV-1 – Comparison of the unsaturation index.

C.2 Intra-class Distribution for OC-316

In the following bar-plots (Figure C.9 to Figure C.13) are the intra-class profiles for OC-316 cell line, with the comparison between CTRL (light gray) and TRTD (dark gray) samples. Only relative amounts higher than 2% are shown; the stars refer to the level of significance (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$).

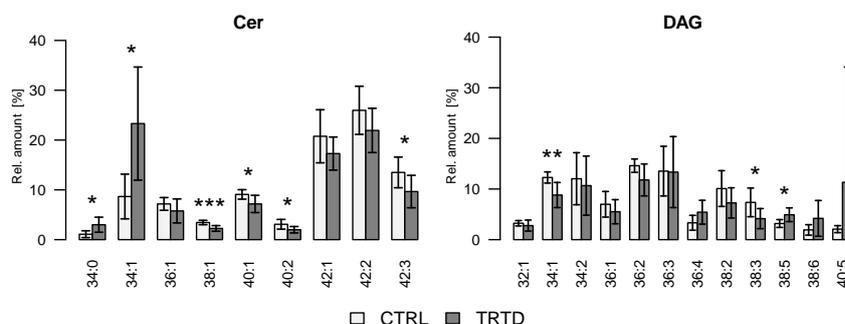


FIGURE C.9: OC-316 – Relative distribution of Cer and DAG (threshold: 2%).

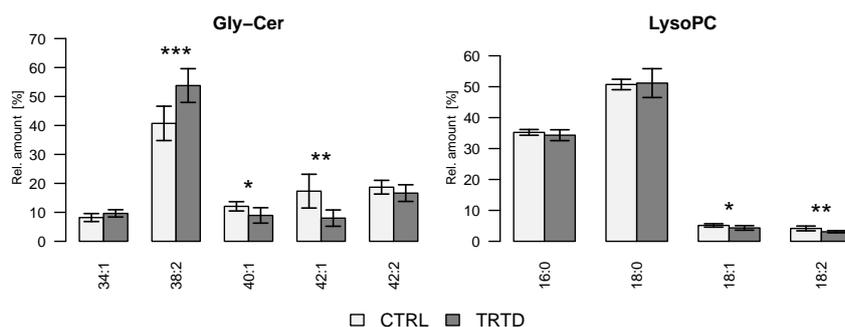


FIGURE C.10: OC-316 – Relative distribution of Gly-Cer and LysoPC (threshold: 2%).

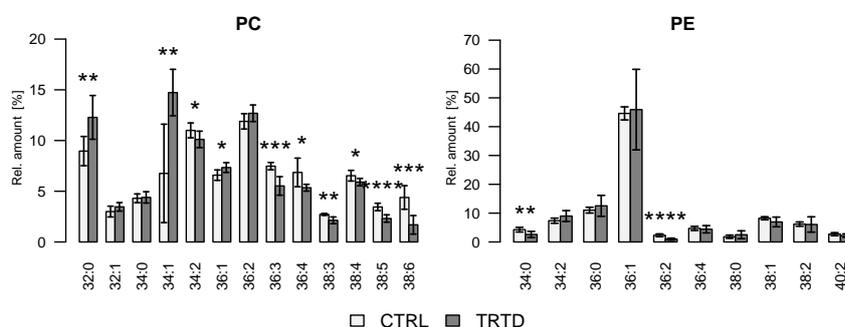


FIGURE C.11: OC-316 – Relative distribution of PC and PE (threshold: 2%).

Figure C.14 shows the group comparison of the average FA chain length for each of the investigated lipid classes, whereas in Figure C.15 is the average number of unsaturations and in Figure C.16 is the unsaturation index.

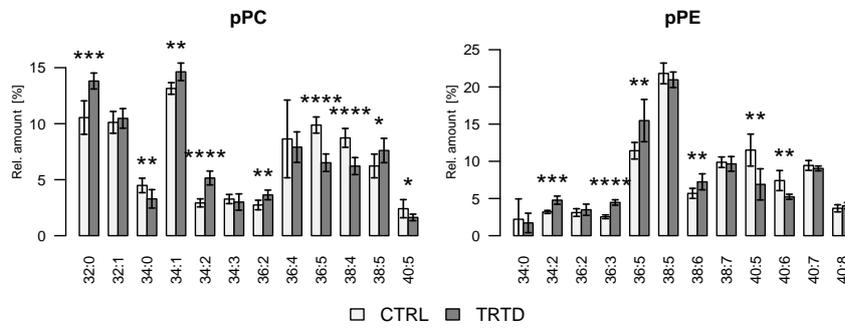


FIGURE C.12: OC-316 – Relative distribution of pPC and pPE (threshold: 2%).

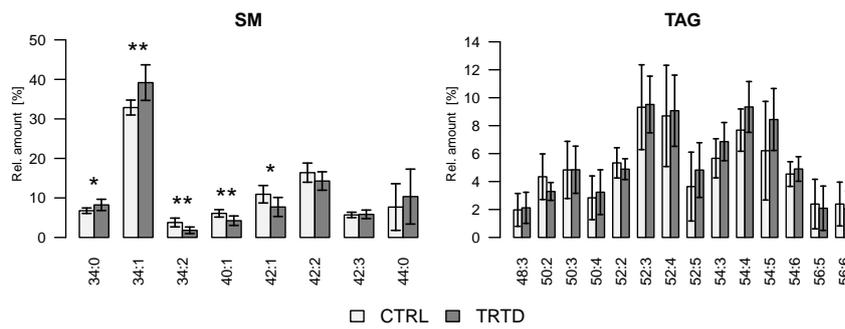


FIGURE C.13: OC-316 – Relative distribution of SM and TAG (threshold: 2%).

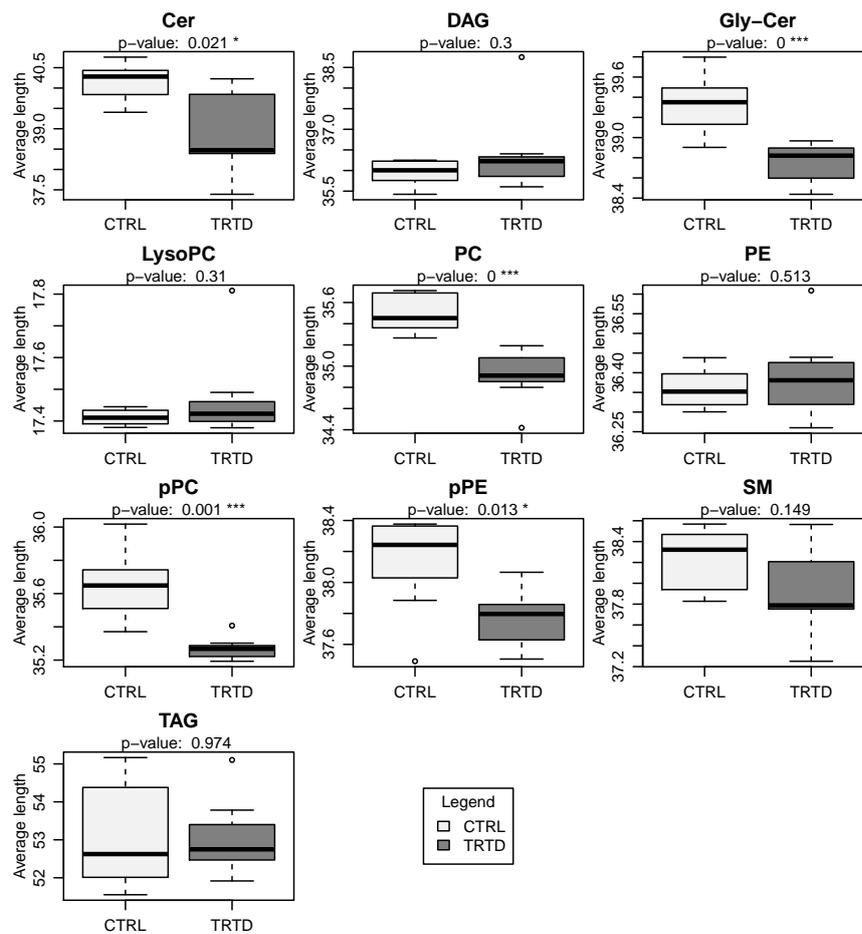


FIGURE C.14: OC-316 – Comparison of the average chain length within each lipid class.

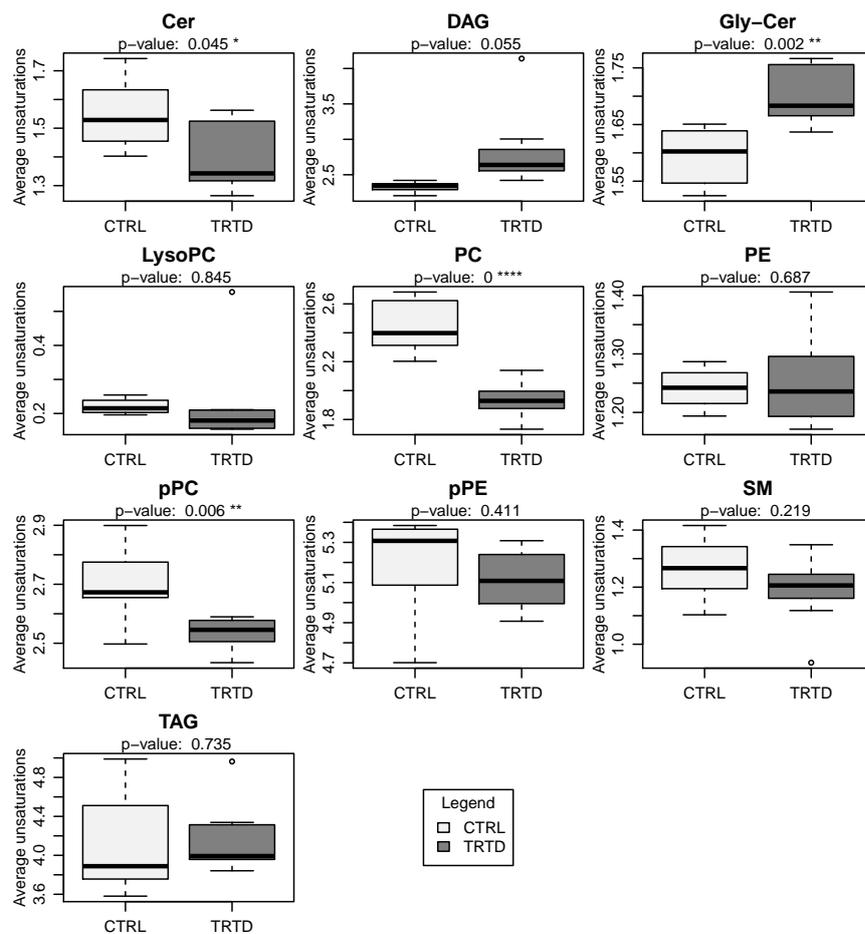


FIGURE C.15: OC-316 – Comparison of the average number of unsaturations within each lipid class.

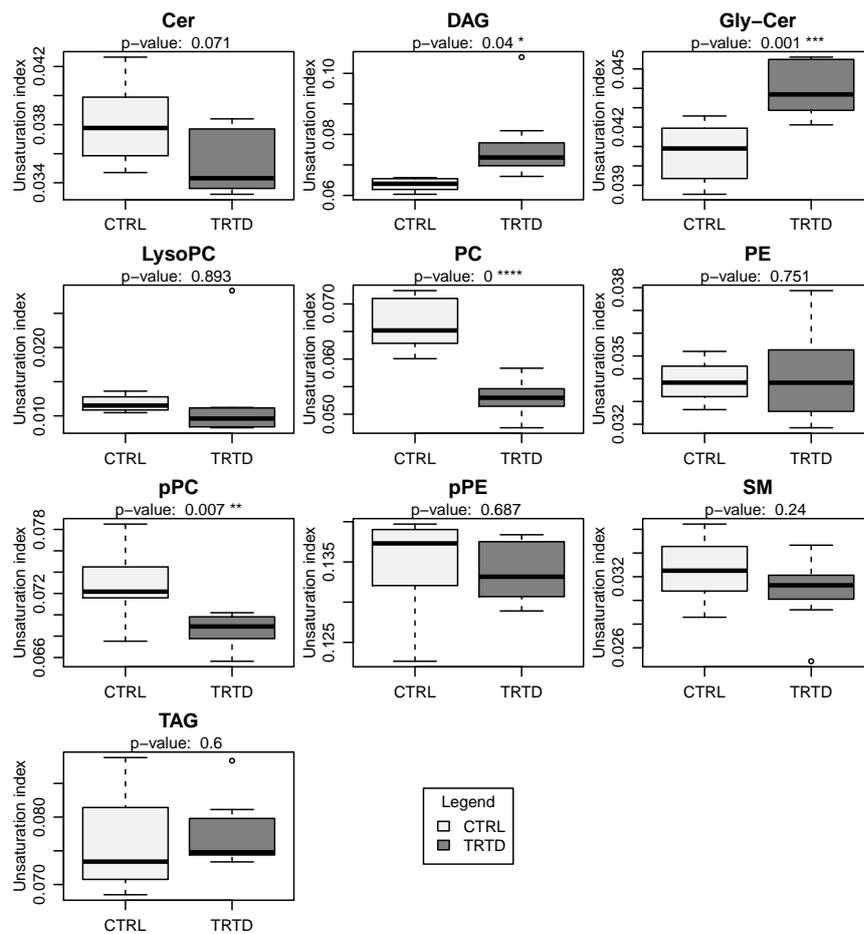


FIGURE C.16: OC-316 – Comparison of the unsaturation index within each lipid class.

C.3 Intra-class Distribution for OVCAR-3

The bar-plots of Figure C.17 to Figure C.21 summarise the intra-class profiles obtained when working with OVCAR-3 cell line. Only relative amounts higher than 2% are shown, and the stars refer to the level of significance (*: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$).

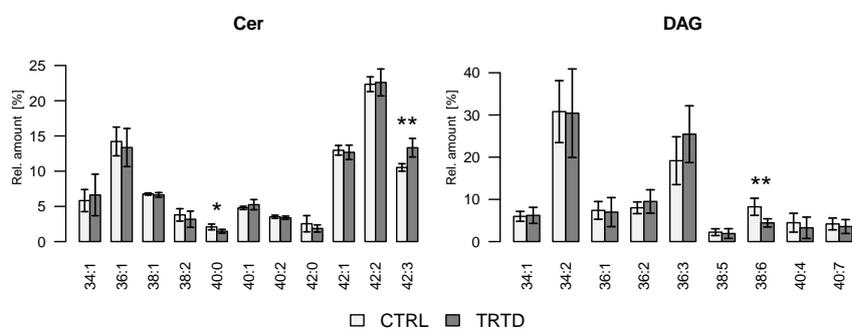


FIGURE C.17: OVCAR-3 – Relative distribution of Cer and DAG (threshold: 2%).

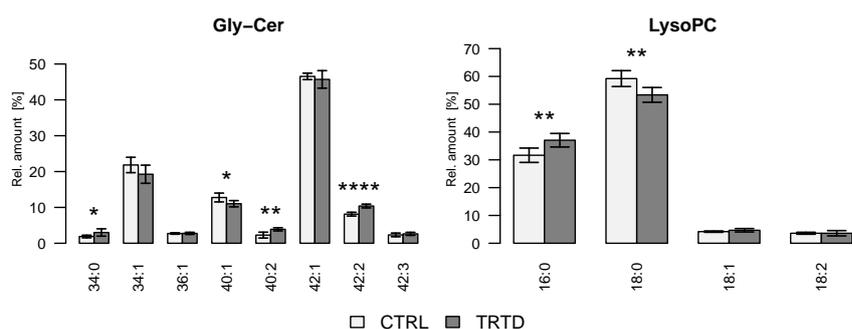


FIGURE C.18: OVCAR-3 – Relative distribution of Gly-Cer and LysoPC (threshold: 2%).

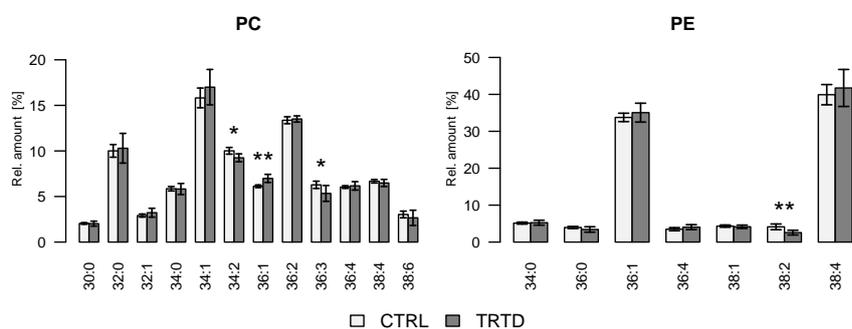


FIGURE C.19: OVCAR-3 – Relative distribution of PC and PE (threshold: 2%).

The comparisons of the parameters average FA chain length, average number of unsaturations and unsaturation index are detailed in Figures C.22, C.23 and C.24, respectively.

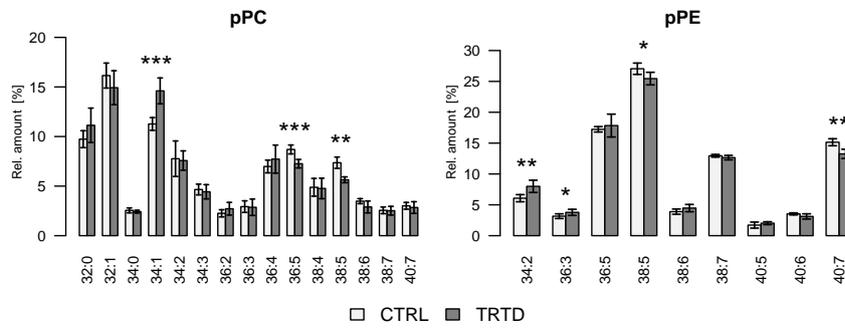


FIGURE C.20: OVCAR-3 – Relative distribution of pPC and pPE (threshold: 2 %).

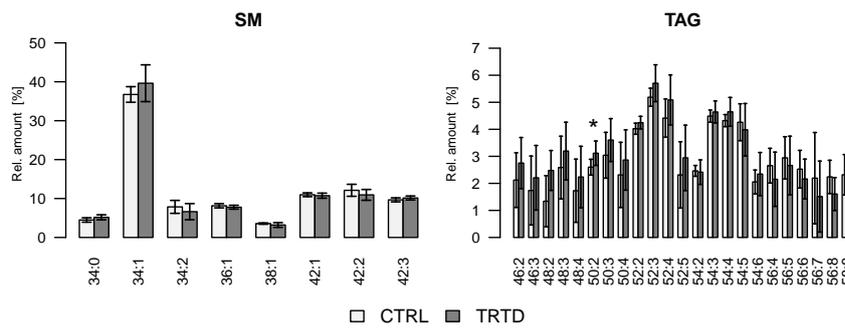


FIGURE C.21: OVCAR-3 – Relative distribution of SM and TAG (threshold: 2 %).

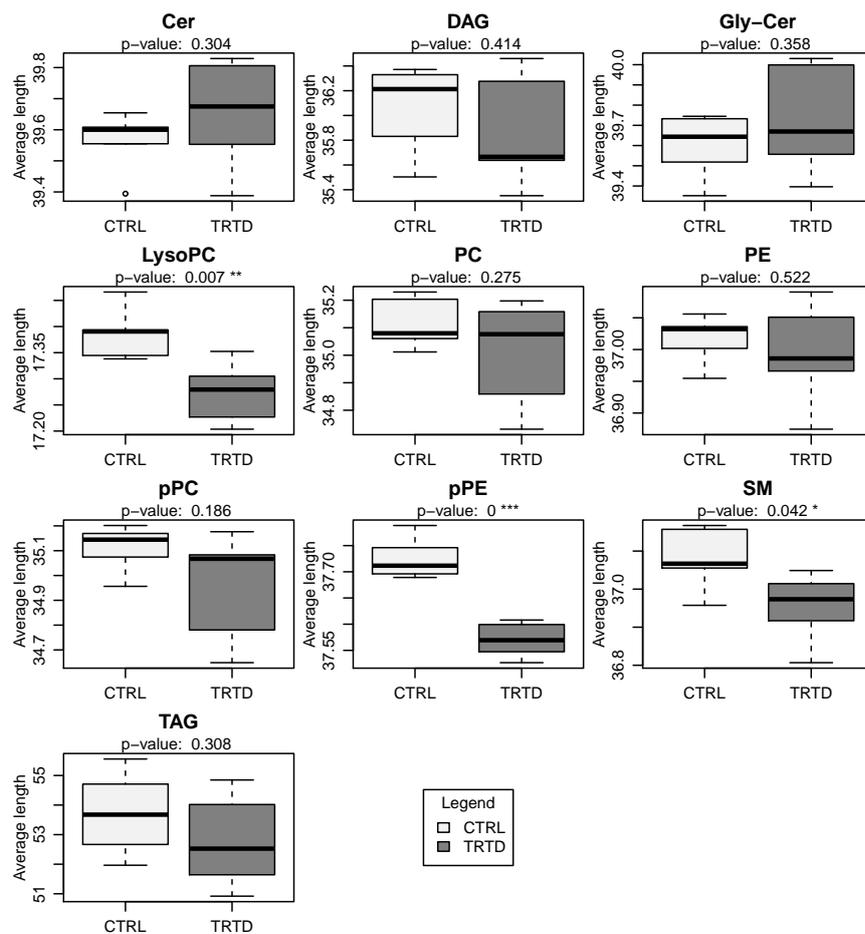


FIGURE C.22: OVCA-3 – Comparison of the average chain length within each lipid class.

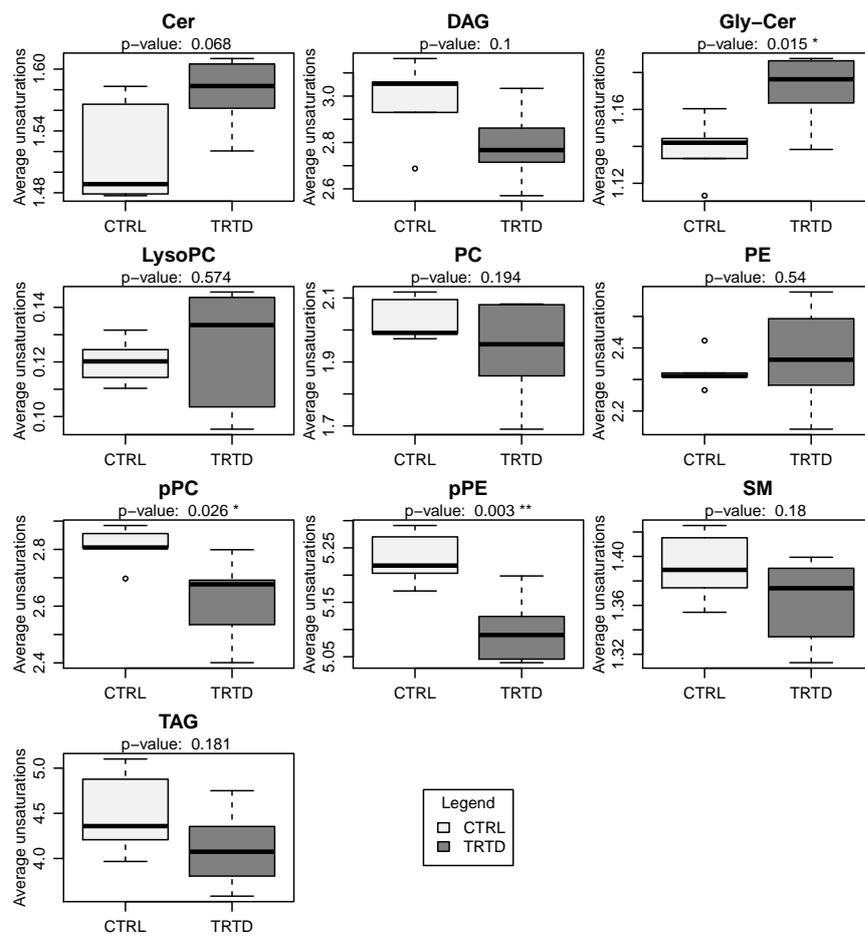


FIGURE C.23: OVCAR-3 – Comparison of the average number of unsaturations within each lipid class.

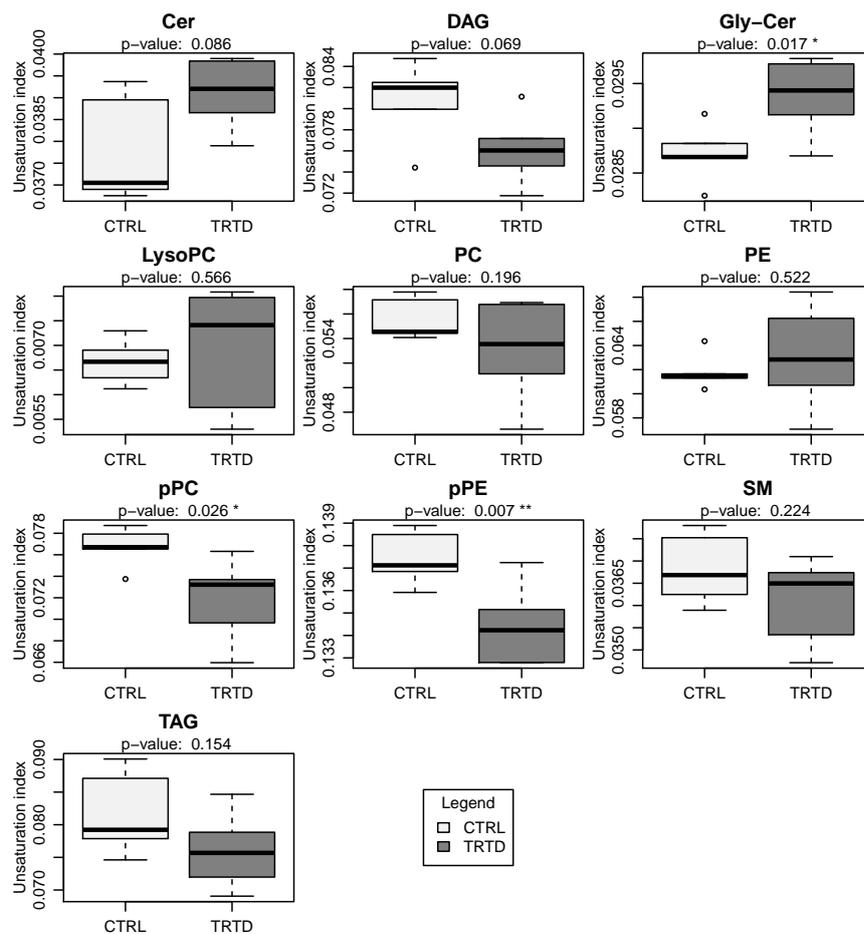


FIGURE C.24: OVCAR-3 – Comparison of the unsaturation index within each lipid class.

C.4 Intra-class Distribution for SKOV-3

The results obtained from SKOV-3 in terms of intra-class profiles are shown in Figure C.25 to Figure C.29.

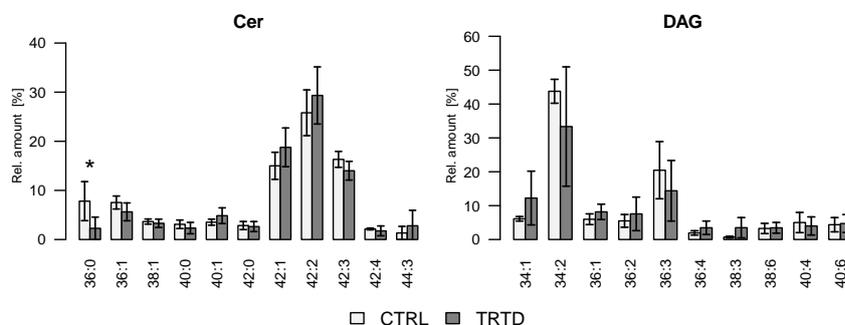


FIGURE C.25: SKOV-3 – Relative distribution of Cer and DAG (threshold: 2%).

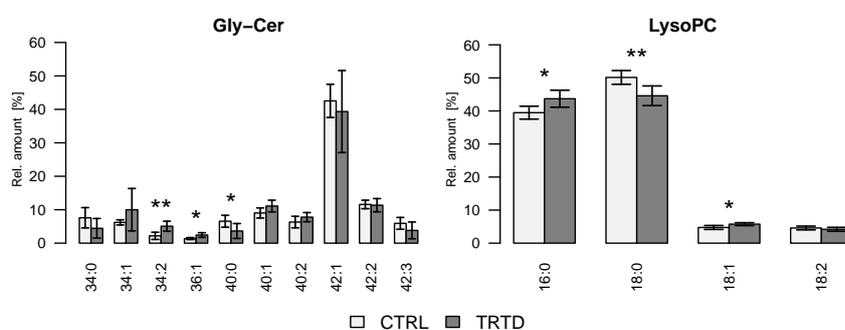


FIGURE C.26: SKOV-3 – Relative distribution of Gly-Cer and LysoPC (threshold: 2%).

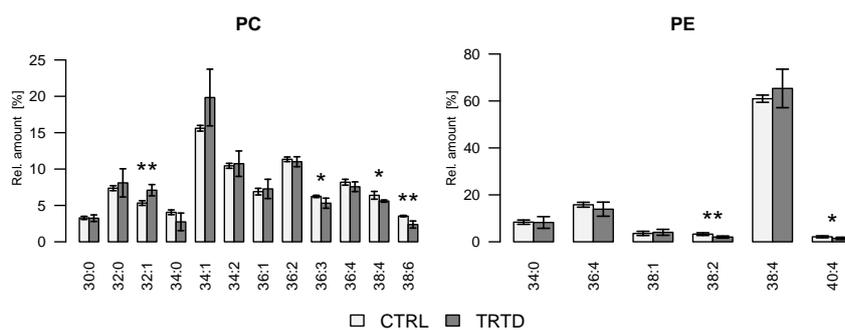


FIGURE C.27: SKOV-3 – Relative distribution of PC and PE (threshold: 2%).

Figures C.30, C.31 and C.32 contain, respectively, the group comparisons of the parameters average FA chain length, average number of unsaturations and unsaturation index.

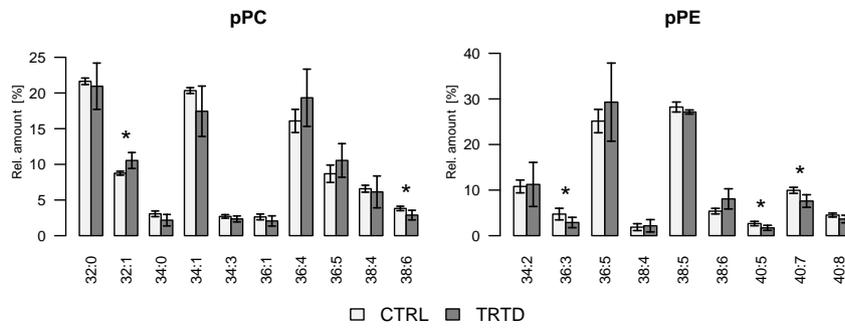


FIGURE C.28: SKOV-3 – Relative distribution of pPC and pPE (threshold: 2 %).

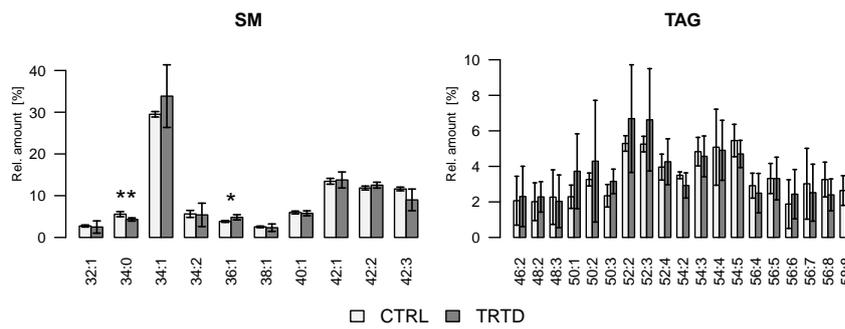


FIGURE C.29: SKOV-3 – Relative distribution of SM and TAG (threshold: 2 %).

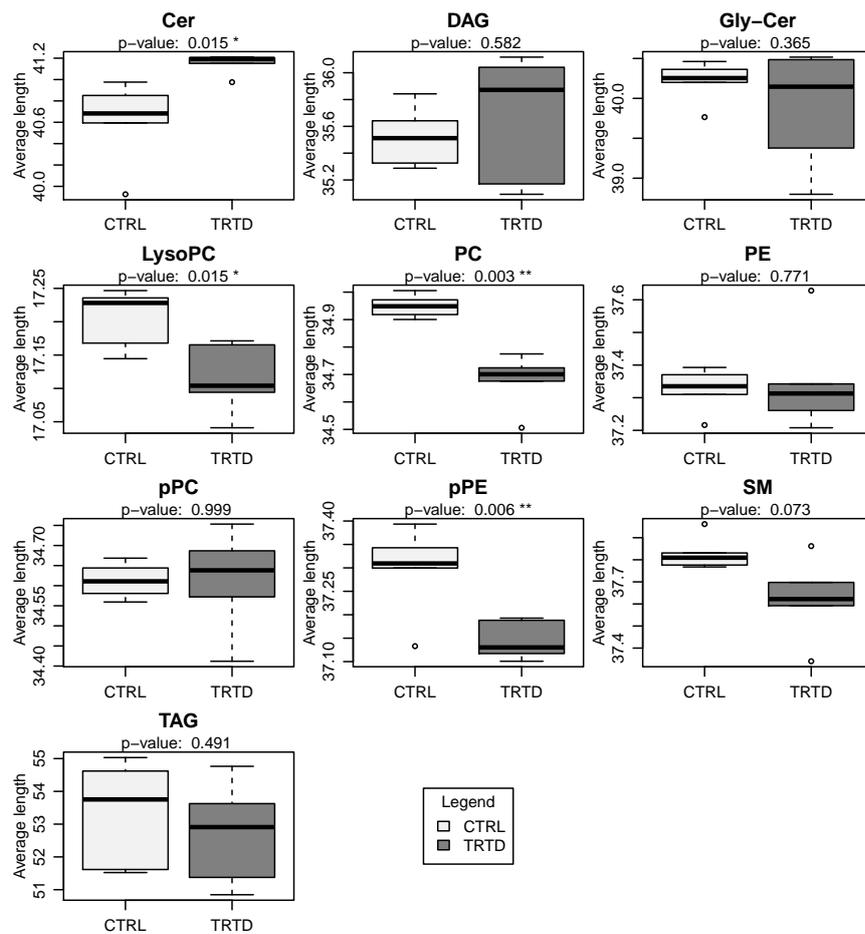


FIGURE C.30: SKOV-3 – Comparison of the average chain length within each lipid class.

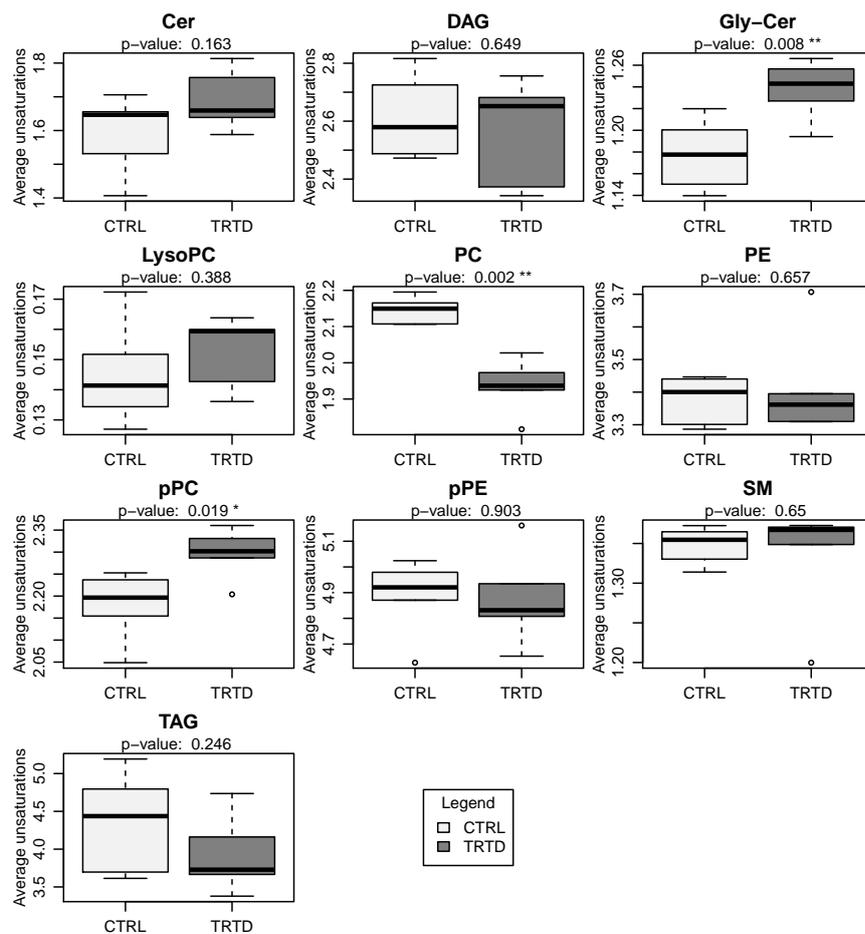


FIGURE C.31: SKOV-3 – Comparison of the average number of unsaturations within each lipid class.

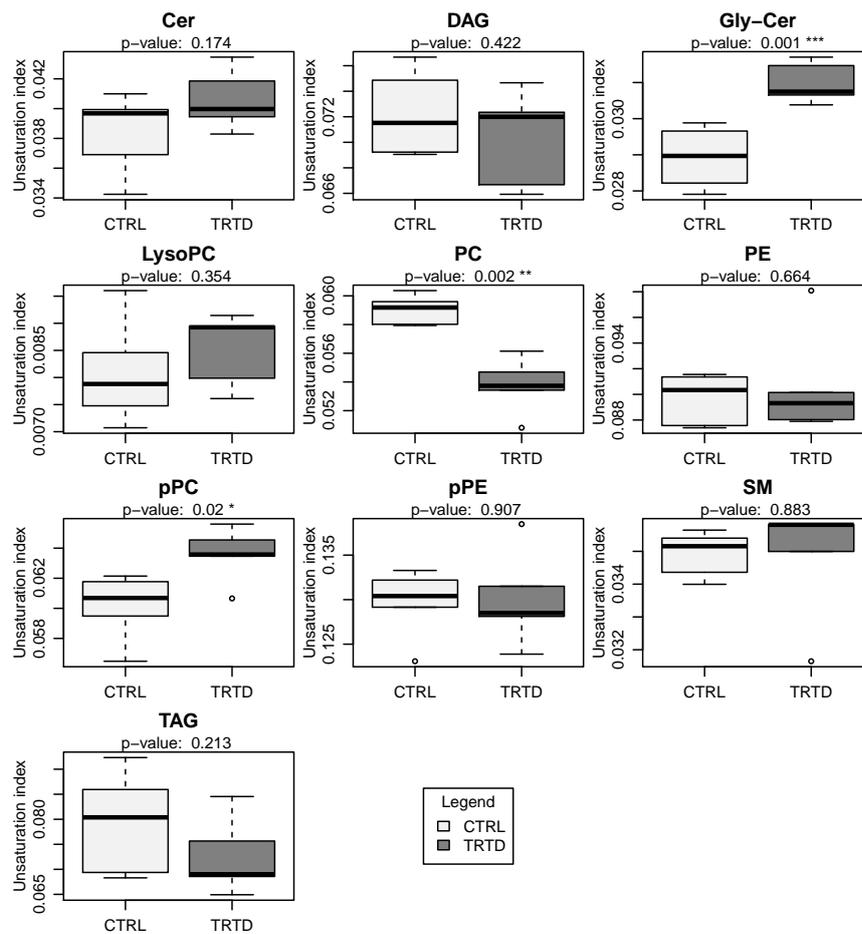


FIGURE C.32: SKOV-3 – Comparison of the unsaturation index within each lipid class.

Appendix D

Appendix of Chapter 4

D.1 A Practical Script

IsotopicLabelling is freely available from the web, where we also included the package reference manual, as well as a vignette explaining how to practically make use of this tool. Below is a concise R script that illustrates the main steps to perform when using *IsotopicLabelling*; here, we load the example data set and look for the patterns of ^{13}C -labelled [PC 32:2+H]⁺ (chemical notation to use: "X40H77NO8P").

```
# Load the package
library("IsotopicLabelling")

# Load the xcmsSet object
data(xcms_obj)

# Convert the object into the required data frame
peak_table <- table_xcms(xcms_obj)

# Process the data
fitted_abundances <- main_labelling(peak_table, compound="X40H77NO8P",
                                   charge=1, labelling="C", mass_shift=0.05, RT=285,
                                   RT_shift=20, chrom_width=7, initial_abundance=NA)

# Quickly look at the results
summary(fitted_abundances)

# Plot the patterns
plot(fitted_abundances, type="patterns", saveplots=T)

# Plot the residuals
plot(fitted_abundances, type="residuals", saveplots=T)

# Plot the overall results
plot(fitted_abundances, type="summary", saveplots=T)

# Save the results to a *.csv file
save_labelling(fitted_abundances)

# Group the samples and obtain grouped estimates
grouped_estimates <- group_labelling(fitted_abundances,
                                     groups=factor(c(rep("C12",4), rep("C13",4))))
```

In this practical example, we exploited the function `main_labelling`, which performs all the processing steps compactly; an alternative would be to use the individual functions it is made of:

```
# From the input, get some useful information
info <- isotopic_information(compound="X40H77NO8P",
                           charge=1, labelling="C")

# Extract the experimental patterns from the data
experimental_patterns <- isotopic_pattern(peak_table, info,
                                         mass_shift=0.05, RT=285, RT_shift=20, chrom_width=7)

# Analyse the patterns
fitted_abundances <- find_abundance(patterns=experimental_patterns,
                                   info=info, initial_abundance=NA, charge=1)
```

Further information can be found at

<https://github.com/RuggeroFerrazza/IsotopicLabelling>.

D.2 Biochemical Application: LC-MS Results

The following plots show the estimated ^2H abundances in the target PC and TAG analytes. The samples have been properly grouped (HFD-OB, HFD-WT, RCD-OB and RCD-WT), and therefore each point represents the group average, whereas the error-bars denote the 95 % confidence intervals. Both blood and liver samples are reported.

Table D.1 details the p-values from the two-sample t-tests; most of them are lower than the 5 % threshold, meaning that the group differences are significant.

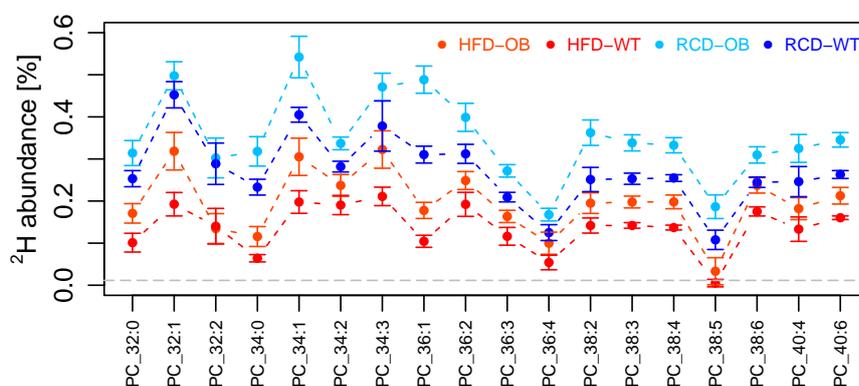


FIGURE D.1: Liver samples, PC species. The horizontal grey line represents the natural ^2H occurrence.

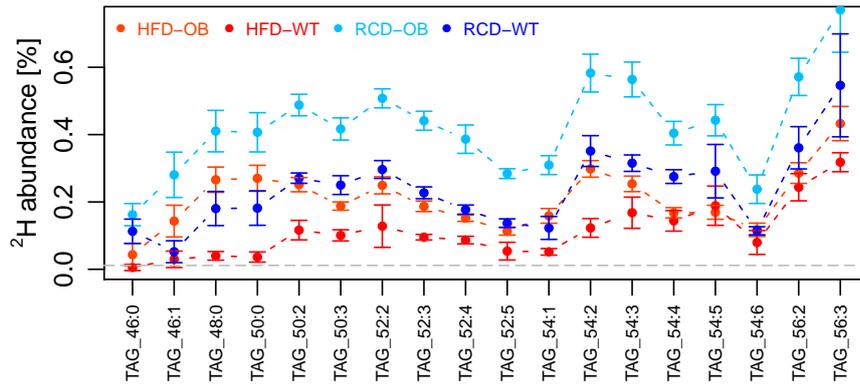


FIGURE D.2: Liver samples, TAG species. The horizontal grey line represents the natural ^2H occurrence.

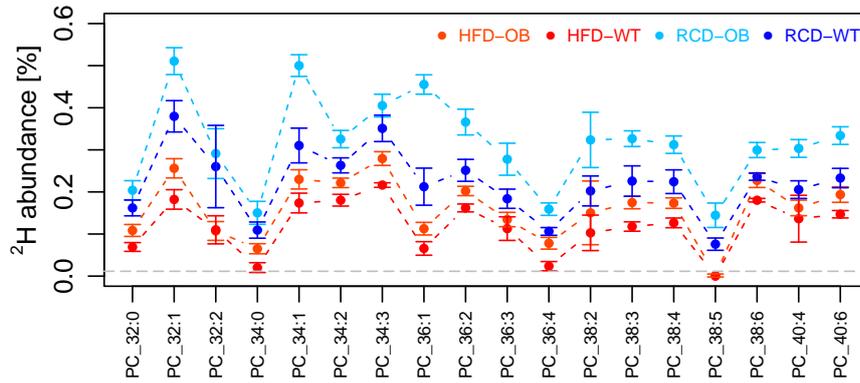


FIGURE D.3: Blood samples, PC species. The horizontal grey line represents the natural ^2H occurrence.

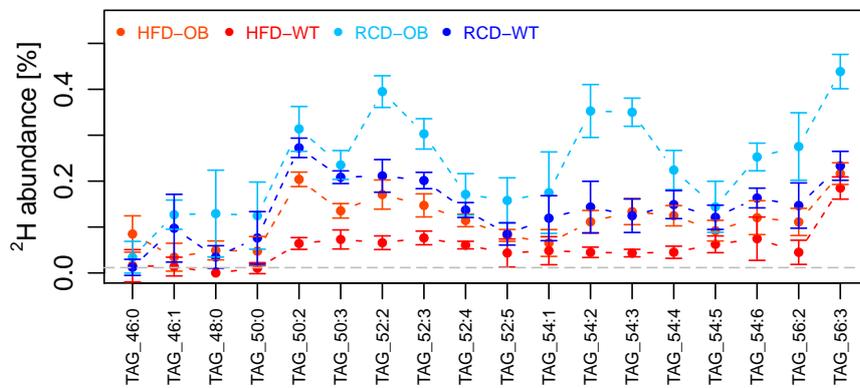


FIGURE D.4: Blood samples, TAG species. The horizontal grey line represents the natural ^2H occurrence.

	LIVER SAMPLES				BLOOD SAMPLES			
	EFFECT OF DIET		EFFECT OF MOUSE TYPE		EFFECT OF DIET		EFFECT OF MOUSE TYPE	
	OB/OB	WT	HFD	RCD	OB/OB	WT	HFD	RCD
PC 32:0	1.2×10^{-6}	1.4×10^{-7}	2.1×10^{-4}	1.8×10^{-3}	5.4×10^{-6}	4.6×10^{-6}	2.2×10^{-4}	4.7×10^{-3}
PC 32:1	7.1×10^{-6}	2.2×10^{-8}	1.7×10^{-4}	3.1×10^{-2}	7.3×10^{-9}	1.9×10^{-6}	1.4×10^{-4}	3.9×10^{-5}
PC 32:2	2.3×10^{-5}	1.7×10^{-4}	8.0×10^{-1}	6.2×10^{-1}	1.2×10^{-4}	9.2×10^{-3}	8.7×10^{-1}	5.1×10^{-1}
PC 34:0	2.2×10^{-7}	1.3×10^{-7}	1.3×10^{-3}	5.2×10^{-4}	1.0×10^{-4}	6.0×10^{-6}	3.2×10^{-5}	1.2×10^{-2}
PC 34:1	1.6×10^{-6}	7.6×10^{-8}	5.3×10^{-4}	2.9×10^{-4}	2.8×10^{-10}	8×10^{-5}	1.3×10^{-3}	4.4×10^{-6}
PC 34:2	1.2×10^{-5}	2.2×10^{-5}	6.2×10^{-3}	2.5×10^{-5}	1.4×10^{-6}	4.2×10^{-6}	1.5×10^{-4}	1.4×10^{-4}
PC 34:3	3.8×10^{-5}	4.1×10^{-4}	4.5×10^{-4}	8.2×10^{-3}	2.1×10^{-6}	7.9×10^{-5}	3.7×10^{-5}	7.4×10^{-3}
PC 36:1	2.7×10^{-9}	4.4×10^{-9}	1.4×10^{-5}	6.2×10^{-7}	2.0×10^{-11}	1.5×10^{-4}	2.3×10^{-4}	1.9×10^{-6}
PC 36:2	2.6×10^{-6}	9.9×10^{-6}	2.5×10^{-3}	3.2×10^{-4}	2.9×10^{-6}	1.3×10^{-4}	2.7×10^{-5}	2.1×10^{-5}
PC 36:3	2.4×10^{-8}	1.1×10^{-5}	1.0×10^{-3}	5.8×10^{-6}	2.4×10^{-5}	5.7×10^{-4}	1.3×10^{-1}	4.6×10^{-4}
PC 36:4	3.2×10^{-4}	3.3×10^{-5}	4.9×10^{-3}	1.1×10^{-3}	4.9×10^{-7}	5.5×10^{-8}	1.2×10^{-5}	2.6×10^{-5}
PC 38:2	3.0×10^{-7}	2.6×10^{-5}	1.3×10^{-3}	3.5×10^{-5}	1.2×10^{-3}	9.9×10^{-4}	2.1×10^{-1}	3.1×10^{-3}
PC 38:3	1.6×10^{-8}	1.5×10^{-7}	1.6×10^{-5}	3.6×10^{-6}	4.7×10^{-9}	3.4×10^{-4}	1.4×10^{-5}	2.6×10^{-4}
PC 38:4	1.4×10^{-8}	1.5×10^{-10}	4.0×10^{-5}	1.0×10^{-5}	7.7×10^{-8}	9.9×10^{-5}	2.5×10^{-5}	8.9×10^{-5}
PC 38:5	1.6×10^{-6}	2.4×10^{-5}	7.9×10^{-2}	2.3×10^{-4}	1.6×10^{-5}	4.1×10^{-5}	3.6×10^{-1}	6.4×10^{-4}
PC 38:6	2.5×10^{-5}	8.3×10^{-7}	4.4×10^{-5}	4.5×10^{-5}	1.1×10^{-5}	8.7×10^{-7}	3.5×10^{-4}	3.9×10^{-5}
PC 40:4	3.7×10^{-6}	1.1×10^{-4}	9.6×10^{-3}	1.9×10^{-3}	4.3×10^{-8}	2.3×10^{-2}	3.1×10^{-1}	4.9×10^{-6}
PC 40:6	4.0×10^{-8}	3.6×10^{-8}	5.1×10^{-4}	2.9×10^{-6}	4.4×10^{-8}	7.2×10^{-5}	4.2×10^{-4}	6.2×10^{-6}
TAG 46:0	4.8×10^{-5}	4.1×10^{-4}	3.3×10^{-2}	2.8×10^{-2}	3.6×10^{-2}	8.3×10^{-1}	7.5×10^{-3}	2.0×10^{-1}
TAG 46:1	1.9×10^{-3}	1.9×10^{-1}	5.7×10^{-4}	5.3×10^{-5}	2.4×10^{-4}	3.2×10^{-2}	2.0×10^{-1}	3.8×10^{-1}
TAG 48:0	6.3×10^{-4}	5.9×10^{-4}	1.7×10^{-6}	2.0×10^{-5}	8.5×10^{-2}	1.5×10^{-2}	1.1×10^{-3}	5.1×10^{-2}
TAG 50:0	6.6×10^{-4}	4.6×10^{-4}	1.1×10^{-6}	1.7×10^{-5}	4.4×10^{-2}	3.1×10^{-2}	2.9×10^{-2}	2.2×10^{-1}
TAG 50:2	2.0×10^{-8}	3.2×10^{-6}	3.0×10^{-6}	2.4×10^{-7}	1.1×10^{-3}	1.5×10^{-8}	3.9×10^{-9}	9.3×10^{-2}
TAG 50:3	3.9×10^{-7}	2.1×10^{-6}	9.7×10^{-7}	1.2×10^{-6}	7.2×10^{-5}	2.9×10^{-7}	1.3×10^{-4}	9.2×10^{-2}
TAG 52:2	1.3×10^{-9}	4.6×10^{-4}	2.8×10^{-3}	3.0×10^{-8}	6.8×10^{-8}	3.5×10^{-5}	6.7×10^{-5}	1.6×10^{-6}
TAG 52:3	7.0×10^{-9}	4.4×10^{-7}	5.9×10^{-7}	3.2×10^{-8}	1.5×10^{-6}	1.0×10^{-7}	1.7×10^{-4}	9.8×10^{-5}
TAG 52:4	5.1×10^{-6}	1.7×10^{-7}	3.8×10^{-7}	6.9×10^{-6}	2.1×10^{-2}	8.7×10^{-6}	1.2×10^{-5}	1.2×10^{-1}
TAG 52:5	7.5×10^{-11}	1.4×10^{-4}	1.1×10^{-3}	1.0×10^{-9}	8.6×10^{-3}	2.2×10^{-2}	2.1×10^{-2}	1.1×10^{-2}
TAG 54:1	3.9×10^{-7}	2.3×10^{-3}	3.9×10^{-6}	6.2×10^{-7}	2.3×10^{-2}	1.3×10^{-2}	3.4×10^{-1}	2.1×10^{-1}
TAG 54:2	2.6×10^{-6}	3.4×10^{-6}	1.7×10^{-7}	7.8×10^{-6}	1.2×10^{-5}	5.7×10^{-3}	2.9×10^{-4}	4.3×10^{-5}
TAG 54:3	6.5×10^{-7}	1.2×10^{-4}	3.3×10^{-3}	3.8×10^{-6}	2.8×10^{-8}	1.8×10^{-3}	1.5×10^{-4}	2.0×10^{-7}
TAG 54:4	2.8×10^{-7}	8.7×10^{-6}	1.1×10^{-1}	2.2×10^{-5}	6.7×10^{-4}	1.0×10^{-4}	2.6×10^{-5}	4.9×10^{-3}
TAG 54:5	7.6×10^{-7}	2.5×10^{-2}	4.5×10^{-1}	2.7×10^{-3}	6.7×10^{-2}	1.1×10^{-3}	2.7×10^{-2}	3.8×10^{-1}
TAG 54:6	1.8×10^{-4}	5.4×10^{-2}	4.6×10^{-2}	2.6×10^{-4}	2.3×10^{-5}	3.2×10^{-3}	8.2×10^{-2}	1.2×10^{-4}
TAG 56:2	1.1×10^{-6}	3.3×10^{-3}	6.3×10^{-2}	6.3×10^{-5}	1.0×10^{-3}	1.8×10^{-3}	1.6×10^{-3}	4.9×10^{-3}
TAG 56:3	3.1×10^{-4}	1.1×10^{-2}	8.8×10^{-4}	1.7×10^{-2}	2.0×10^{-7}	1.2×10^{-2}	4.1×10^{-2}	5.3×10^{-7}

TABLE D.1: p-values from the two-sample t-tests; the grey values are not significant (higher than 0.05).

D.3 Biochemical Application: GC-MS Results

The following plots summarise the estimated ^2H abundances for each sample group (HFD-OB, HFD-WT, RCD-OB and RCD-WT), with the distinction of FAMES based on the lipid class (CE, FA, PC, TAG). The points are group averages, whereas the error-bars denote the 95 % confidence intervals.

Table D.2 contains the p-values from the two-sample t-tests; it can be clearly seen that both mouse type and diet affect the ^2H abundances in the FAs, considering the high number of significant cases. Also worth noting is the lack of any significant difference for FA C18:2, which is the consequence of it being an essential fatty acid, not synthesised through DNL.

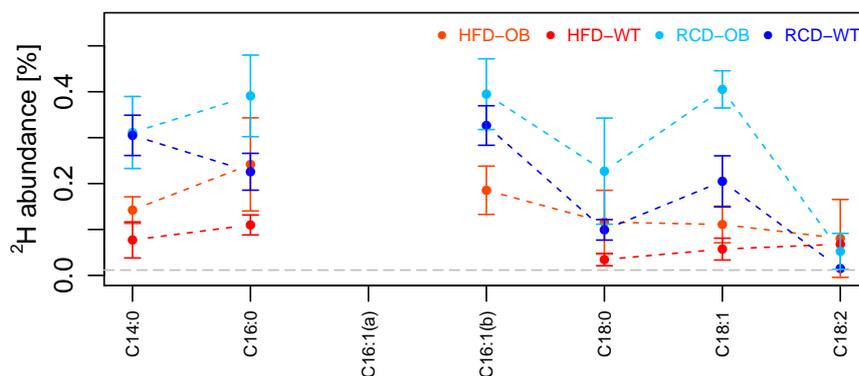


FIGURE D.5: ^2H abundances relative to the FAs of SPE-isolated CEs. Horizontal grey line: natural ^2H occurrence.

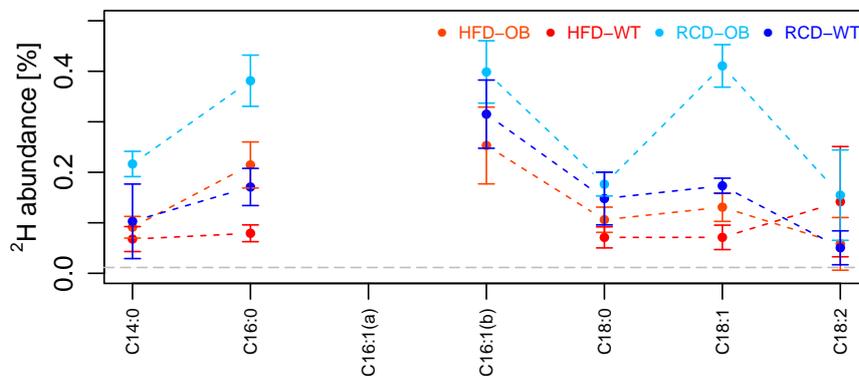


FIGURE D.6: ^2H abundances relative to the FAs of SPE-isolated FAs. Horizontal grey line: natural ^2H occurrence.

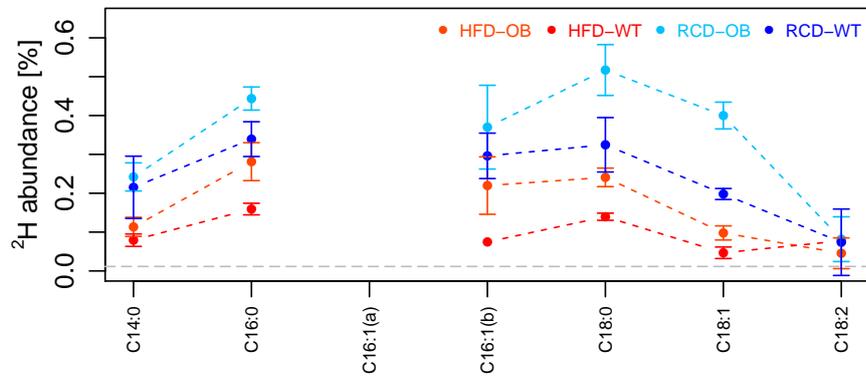


FIGURE D.7: ^2H abundances relative to the FAs of SPE-isolated PCs. Horizontal grey line: natural ^2H occurrence.

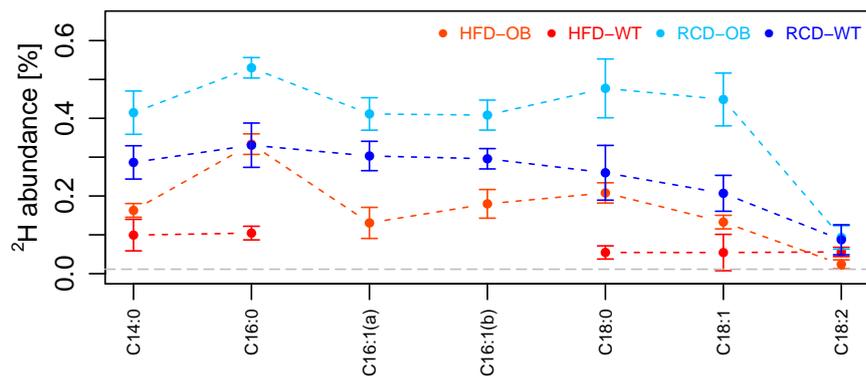


FIGURE D.8: ^2H abundances relative to the FAs of SPE-isolated TAGs. Horizontal grey line: natural ^2H occurrence.

TOTAL FA					SPE – CE				
	EFFECT OF DIET		EFFECT OF MOUSE TYPE			EFFECT OF DIET		EFFECT OF MOUSE TYPE	
	OB/OB	WT	HFD	RCD		OB/OB	WT	HFD	RCD
C14:0	3.6×10^{-6}	3.4×10^{-11}	6.3×10^{-8}	4.4×10^{-4}	C14:0	3.3×10^{-3}	1.6×10^{-3}	1.5×10^{-2}	8.9×10^{-1}
C16:0	1.6×10^{-6}	3.3×10^{-7}	3.6×10^{-6}	1.4×10^{-5}	C16:0	1.9×10^{-2}	2.2×10^{-4}	1.9×10^{-2}	3.0×10^{-3}
C16:1(a)	3.7×10^{-7}	2.2×10^{-1}	3.6×10^{-1}	2.2×10^{-3}	C16:1(a)	NA	NA	NA	NA
C16:1(b)	1.3×10^{-6}	8.9×10^{-5}	1.1×10^{-2}	1.2×10^{-4}	C16:1(b)	6.4×10^{-4}	NA	NA	NA
C18:0	3.6×10^{-7}	1.9×10^{-5}	3.6×10^{-5}	1.4×10^{-4}	C18:0	7.2×10^{-2}	1.9×10^{-4}	2.8×10^{-2}	3.5×10^{-2}
C18:1	1.4×10^{-6}	1.8×10^{-8}	3.5×10^{-7}	6.7×10^{-6}	C18:1	2.6×10^{-8}	4.5×10^{-4}	2.0×10^{-2}	2.8×10^{-5}
C18:2	8.6×10^{-2}	7.1×10^{-2}	1.2×10^{-1}	3.8×10^{-2}	C18:2	6.1×10^{-1}	NA	NA	NA

(A)

SPE – FA					SPE – PC				
	EFFECT OF DIET		EFFECT OF MOUSE TYPE			EFFECT OF DIET		EFFECT OF MOUSE TYPE	
	OB/OB	WT	HFD	RCD		OB/OB	WT	HFD	RCD
C14:0	9.2×10^{-7}	4.0×10^{-1}	1.0×10^{-1}	4.3×10^{-2}	C14:0	2.7×10^{-3}	6.6×10^{-3}	1.6×10^{-2}	4.9×10^{-1}
C16:0	6.7×10^{-5}	6.5×10^{-4}	1.9×10^{-4}	6.7×10^{-6}	C16:0	4.6×10^{-5}	5.9×10^{-5}	6.2×10^{-4}	9.0×10^{-4}
C16:1(a)	NA	NA	NA	NA	C16:1(a)	NA	NA	NA	NA
C16:1(b)	1.2×10^{-2}	NA	NA	NA	C16:1(b)	2.7×10^{-2}	NA	NA	NA
C18:0	2.9×10^{-4}	1.1×10^{-2}	2.2×10^{-2}	2.5×10^{-1}	C18:0	8.9×10^{-5}	9.5×10^{-4}	1.5×10^{-5}	6.8×10^{-4}
C18:1	5.7×10^{-8}	7.9×10^{-4}	6.3×10^{-3}	2.3×10^{-6}	C18:1	2.7×10^{-7}	1.6×10^{-6}	9.8×10^{-4}	1.1×10^{-5}
C18:2	1.5×10^{-1}	2.6×10^{-1}	3.0×10^{-1}	1.2×10^{-1}	C18:2	3.8×10^{-1}	NA	NA	8.8e-1

(C)

SPE – TAG				
	EFFECT OF DIET		EFFECT OF MOUSE TYPE	
	OB/OB	WT	HFD	RCD
C14:0	7.4×10^{-5}	7.5×10^{-5}	1.7×10^{-2}	2.5×10^{-3}
C16:0	1.1×10^{-7}	7.1×10^{-5}	7.7×10^{-9}	6.3×10^{-5}
C16:1(a)	2.9×10^{-6}	NA	NA	6.5×10^{-2}
C16:1(b)	1.0×10^{-6}	NA	NA	4.9×10^{-4}
C18:0	2.5×10^{-4}	4.7×10^{-4}	3.2×10^{-7}	5.8×10^{-4}
C18:1	8.1×10^{-5}	1.4×10^{-4}	5.9×10^{-3}	1.1×10^{-4}
C18:2	5.7×10^{-2}	2.8×10^{-1}	7.9×10^{-2}	8.4×10^{-1}

(E)

TABLE D.2: p-values from the t-tests on the GC-MS data, to study possible effects of diet and mouse type on the isotopic enrichment. The grey values are not significant ($p > 0.05$).

D.4 Follow-up Experiment

The two plots below contain the estimated ^2H abundances for each analyte and sample group (HSD-OB, HSD-WT, LFD-OB, LFD-WT, WD-OB, WD-WT), where the error-bars represent the 95 % confidence intervals.

Table D.3 lists the p-values from the two-sample t-tests; not significant cases are highlighted in grey. It is interesting to observe that no significant difference was found when comparing HSD with LFD, for all experimental settings. This implies that the two dietary regimes result in similar DNL rates.

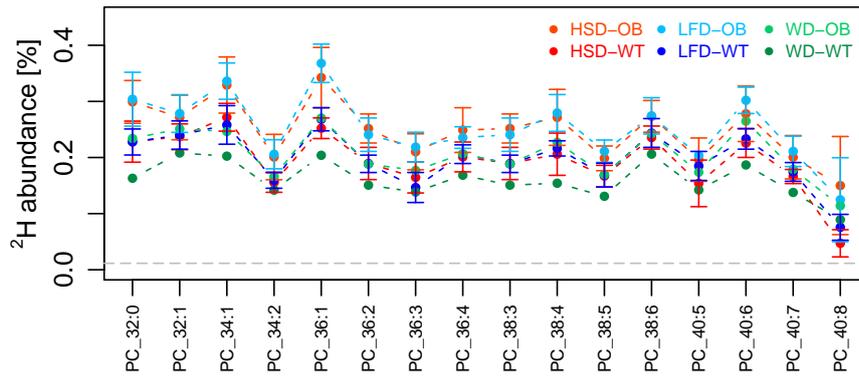


FIGURE D.9: ^2H abundances for PCs (liver samples).

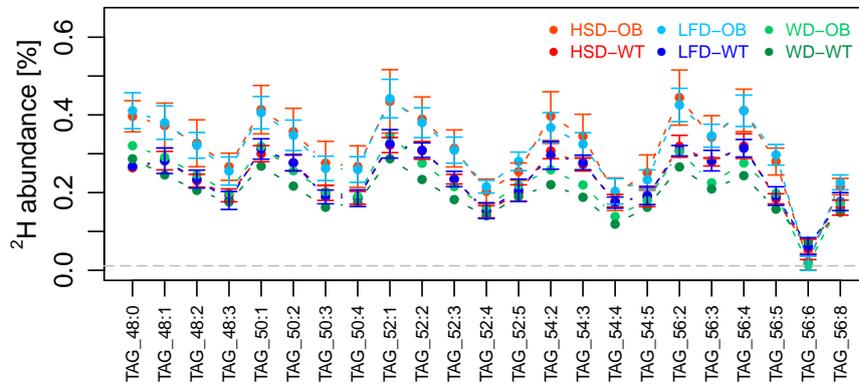


FIGURE D.10: ^2H abundances for TAGs (liver samples).

	EFFECT OF DIET						EFFECT OF MOUSE TYPE		
	HSD VS LFD		HSD VS WD		LFD VS WD		HSD	LFD	WD
	OB/OB	WT	OB/OB	WT	OB/OB	WT			
PC 32:0	8.5×10^{-1}	9.7×10^{-1}	1.2×10^{-2}	6.5×10^{-3}	1.8×10^{-2}	1.7×10^{-3}	6.7×10^{-3}	6.9×10^{-3}	4.3×10^{-3}
PC 32:1	7.5×10^{-1}	8.4×10^{-1}	3.0×10^{-1}	3.0×10^{-2}	1.3×10^{-1}	2.7×10^{-2}	1.0×10^{-1}	5.0×10^{-2}	2.1×10^{-3}
PC 34:1	7.8×10^{-1}	4.6×10^{-1}	5.9×10^{-3}	1.3×10^{-4}	2.2×10^{-4}	6.2×10^{-3}	3.4×10^{-2}	1.6×10^{-3}	8.1×10^{-3}
PC 34:2	8.1×10^{-1}	6.7×10^{-1}	8.9×10^{-2}	1.2×10^{-1}	1.1×10^{-2}	2.5×10^{-2}	3.5×10^{-2}	3.6×10^{-3}	1.3×10^{-2}
PC 36:1	3.5×10^{-1}	1.9×10^{-1}	1.6×10^{-2}	2.1×10^{-4}	1.1×10^{-4}	5.7×10^{-5}	5.4×10^{-3}	8.6×10^{-5}	2.0×10^{-5}
PC 36:2	5.1×10^{-1}	9.6×10^{-1}	6.6×10^{-4}	1.6×10^{-2}	5.4×10^{-3}	3.9×10^{-4}	1.9×10^{-3}	4.0×10^{-3}	4.4×10^{-3}
PC 36:3	6.4×10^{-1}	2.9×10^{-1}	1.0×10^{-1}	7.2×10^{-2}	3.3×10^{-2}	5.3×10^{-1}	2.2×10^{-2}	4.8×10^{-4}	2.6×10^{-2}
PC 36:4	4.8×10^{-1}	7.2×10^{-1}	4.9×10^{-2}	2.5×10^{-2}	3.7×10^{-2}	8.4×10^{-4}	3.4×10^{-2}	1.6×10^{-2}	4.8×10^{-3}
PC 38:3	5.1×10^{-1}	9.6×10^{-1}	6.6×10^{-4}	1.6×10^{-2}	5.4×10^{-3}	3.9×10^{-4}	1.9×10^{-3}	4.0×10^{-3}	4.4×10^{-3}
PC 38:4	7.6×10^{-1}	5.6×10^{-1}	7.7×10^{-2}	1.5×10^{-2}	7.2×10^{-3}	3.2×10^{-5}	2.6×10^{-2}	2.1×10^{-3}	1.7×10^{-5}
PC 38:5	3.2×10^{-1}	8.8×10^{-1}	5.2×10^{-2}	6.5×10^{-3}	7.3×10^{-3}	6.9×10^{-3}	2.2×10^{-2}	3.9×10^{-3}	1.0×10^{-2}
PC 38:6	9.6×10^{-1}	5.7×10^{-1}	9.3×10^{-2}	1.4×10^{-2}	1.1×10^{-1}	1.1×10^{-2}	2.3×10^{-2}	1.0×10^{-1}	2.6×10^{-2}
PC 40:5	7.1×10^{-1}	1.6×10^{-1}	2.2×10^{-1}	5.5×10^{-1}	6.9×10^{-2}	6.6×10^{-3}	9.7×10^{-2}	2.9×10^{-1}	3.1×10^{-3}
PC 40:6	3.1×10^{-1}	5.7×10^{-1}	5.9×10^{-1}	1.8×10^{-2}	3.0×10^{-2}	2.3×10^{-3}	4.7×10^{-2}	1.1×10^{-4}	1.6×10^{-4}
PC 40:7	6.0×10^{-1}	3.7×10^{-1}	2.5×10^{-1}	7.5×10^{-3}	4.0×10^{-2}	2.8×10^{-3}	7.6×10^{-2}	1.9×10^{-2}	1.9×10^{-3}
PC 40:8	6.1×10^{-1}	6.3×10^{-2}	3.6×10^{-1}	5.7×10^{-2}	7.5×10^{-1}	5.0×10^{-1}	2.8×10^{-2}	1.7×10^{-1}	2.0×10^{-1}
TAG 48:0	5.8×10^{-1}	9.1×10^{-1}	4.0×10^{-3}	3.3×10^{-1}	2.5×10^{-3}	5.0×10^{-1}	7.0×10^{-4}	1.6×10^{-3}	6.4×10^{-2}
TAG 48:1	8.1×10^{-1}	9.9×10^{-1}	1.3×10^{-2}	9.7×10^{-3}	1.3×10^{-3}	3.8×10^{-2}	7.7×10^{-3}	8.9×10^{-4}	1.1×10^{-3}
TAG 48:2	8.6×10^{-1}	7.4×10^{-1}	1.6×10^{-2}	9.0×10^{-3}	6.7×10^{-4}	2.2×10^{-2}	7.2×10^{-3}	2.4×10^{-4}	2.3×10^{-4}
TAG 48:3	6.1×10^{-1}	2.9×10^{-1}	2.3×10^{-3}	5.0×10^{-2}	8.0×10^{-3}	7.2×10^{-1}	1.3×10^{-3}	1.3×10^{-3}	3.3×10^{-2}
TAG 50:1	8.1×10^{-1}	4.1×10^{-1}	9.2×10^{-3}	8.8×10^{-3}	1.3×10^{-3}	4.8×10^{-3}	3.6×10^{-3}	7.8×10^{-4}	1.2×10^{-4}
TAG 50:2	7.6×10^{-1}	9.6×10^{-1}	5.6×10^{-3}	9.9×10^{-5}	5.6×10^{-4}	1.2×10^{-4}	1.6×10^{-2}	3.3×10^{-3}	6.7×10^{-4}
TAG 50:3	6.3×10^{-1}	3.8×10^{-1}	1.8×10^{-2}	5.5×10^{-3}	2.2×10^{-3}	2.7×10^{-2}	1.5×10^{-2}	5.3×10^{-4}	1.0×10^{-3}
TAG 50:4	7.8×10^{-1}	9.4×10^{-1}	1.2×10^{-2}	2.5×10^{-1}	1.3×10^{-3}	3.5×10^{-1}	9.5×10^{-3}	8.1×10^{-4}	9.8×10^{-2}
TAG 52:1	8.6×10^{-1}	8.6×10^{-1}	3.9×10^{-2}	7.8×10^{-3}	2.3×10^{-3}	5.1×10^{-2}	1.4×10^{-2}	6.5×10^{-4}	3.4×10^{-5}
TAG 52:2	7.5×10^{-1}	9.1×10^{-1}	2.1×10^{-3}	2.3×10^{-5}	1.7×10^{-4}	4.6×10^{-6}	1.2×10^{-2}	2.8×10^{-3}	1.3×10^{-3}
TAG 52:3	8.6×10^{-1}	9.8×10^{-1}	1.7×10^{-3}	4.2×10^{-5}	1.3×10^{-4}	2.6×10^{-4}	5.9×10^{-3}	8.4×10^{-4}	5.1×10^{-3}
TAG 52:4	4.4×10^{-1}	8.0×10^{-1}	2.3×10^{-2}	3.1×10^{-1}	4.8×10^{-5}	2.6×10^{-1}	5.6×10^{-3}	8.1×10^{-5}	2.1×10^{-2}
TAG 52:5	7.3×10^{-2}	6.7×10^{-1}	1.9×10^{-2}	3.4×10^{-1}	5.3×10^{-4}	2.4×10^{-1}	8.4×10^{-4}	3.7×10^{-4}	1.4×10^{-2}
TAG 54:2	3.5×10^{-1}	5.3×10^{-1}	1.1×10^{-3}	1.9×10^{-6}	1.1×10^{-4}	1.1×10^{-3}	1.2×10^{-2}	6.2×10^{-3}	6.4×10^{-3}
TAG 54:3	4.6×10^{-1}	6.4×10^{-1}	1.3×10^{-3}	9.3×10^{-7}	2.0×10^{-5}	9.5×10^{-7}	1.9×10^{-2}	8.1×10^{-3}	4.4×10^{-3}
TAG 54:4	7.3×10^{-1}	6.0×10^{-1}	1.6×10^{-2}	5.1×10^{-4}	1.9×10^{-3}	9.3×10^{-4}	4.0×10^{-1}	8.5×10^{-2}	1.8×10^{-1}
TAG 54:5	4.1×10^{-1}	7.4×10^{-1}	7.4×10^{-3}	1.6×10^{-1}	2.4×10^{-3}	9.7×10^{-2}	1.4×10^{-2}	1.5×10^{-2}	3.3×10^{-1}
TAG 56:2	5.9×10^{-1}	4.1×10^{-1}	2.9×10^{-3}	2.8×10^{-3}	2.7×10^{-4}	6.9×10^{-4}	3.9×10^{-3}	2.2×10^{-4}	8.4×10^{-4}
TAG 56:3	9.0×10^{-1}	8.0×10^{-1}	1.7×10^{-3}	2.0×10^{-5}	7.8×10^{-6}	1.7×10^{-4}	3.0×10^{-2}	1.9×10^{-3}	1.3×10^{-1}
TAG 56:4	9.7×10^{-1}	7.5×10^{-1}	4.1×10^{-4}	6.6×10^{-4}	2.7×10^{-5}	2.5×10^{-4}	5.4×10^{-3}	4.6×10^{-4}	6.8×10^{-2}
TAG 56:5	3.5×10^{-1}	5.5×10^{-1}	5.9×10^{-4}	9.0×10^{-3}	7.8×10^{-6}	1.6×10^{-2}	2.5×10^{-4}	6.1×10^{-6}	1.2×10^{-3}
TAG 56:6	7.2×10^{-1}	5.7×10^{-1}	3.9×10^{-1}	2.2×10^{-1}	6.3×10^{-1}	4.3×10^{-1}	4.1×10^{-2}	2.6×10^{-3}	7.9×10^{-4}
TAG 56:8	3.7×10^{-1}	2.4×10^{-1}	2.5×10^{-3}	2.7×10^{-1}	1.8×10^{-4}	3.6×10^{-2}	5.9×10^{-4}	1.7×10^{-3}	1.0×10^{-2}

TABLE D.3: p-values from the comparison between sample group couples; the grey values represent p-values higher than 5 %.