



ORBIS-Met: Optimization of R based integrative statistics for metabolome analyses

Sol Culleré, Joaquim

Bioinformatics and Biostatistics Master's Degree

Pharmaceutical bioinformatics

Consultor: Melchor Sánchez Martínez, PhD

Professor in charge: María Jesús Marco Galindo, PhD

External professor: Victòria Ayala Jové, PhD

24/05/2017



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	<i>ORBIS-Met: Optimization of R based integrative statistics for metabolome analyses</i>
Nom de l'autor:	<i>Joaquim Sol Culleré</i>
Nom del consultor/a:	<i>Melchor Sánchez Martínez</i>
Nom del PRA:	<i>María Jesús Marco Galindo</i>
Data de lliurament (mm/aaaa):	<i>24/05/2017</i>
Titulació o programa:	<i>Bioinformatics and Biostatistics Master's Degree</i>
Àrea del Treball Final:	<i>Pharmaceutical bioinformatics</i>
Idioma del treball:	<i>English</i>
Paraules clau	<i>ALS, data analysis, lipidomics</i>

Resum del Treball

Finalitat

Desenvolupar un programa específic per tractar dades òmiques i poder-lo aplicar en la definició d'empremtes lipídòmiques específiques de pacients amb Esclerosi Lateral Amiotròfica (ELA).

Context d'aplicació

La metabolòmica consisteix en l'estudi de molècules petites resultants del metabolisme cel·lular, les quals correlacionen molt bé amb el fenotip i poden reflectir tant processos fisiològics com patològics. Tot i que els recents avenços tècnics ens permeten detectar un gran nombre de metabòlits en mostres biològiques, el tractament d'aquestes dades continua essent un problema. Per això, la millora dels softwares existents i el desenvolupament de nous específicament dissenyats pel tractament de dades òmiques beneficiarà molt la cerca de biomarcadors.

L'ELA es una malaltia neurodegenerativa que no té cura i amb etiologia desconeguda. En aquest sentit, el descobriment de biomarcadors tant d'inici com de progressió de la malaltia poden ajudar a descobrir noves dianes terapèutiques.

Metodologia

S'ha realitzat un anàlisi metabolòmic basat en espectrometria de masses en mostres de plasma y líquid cefaloraquídi de pacients d'ELA i individus control. El tractament de les dades s'ha realitzat utilitzant un software basat en R especialment pensat i dissenyat per l'anàlisi de

metabolòmica.

Resultats

Existeix una empremta lipidòmica en plasma que defineix els pacients d' ELA on sembla que els esfingolípidis poden jugar-hi un paper clau.

Conclusions

La utilització de softwares per al tractament de dades òmiques facilita el descobriment de nous biomarcadors de malalties que permetin monitoritzar la progressió i el tractament, i ajuden en la descripció de la seva etiologia.

Abstract

Aim

Developing a specific software to analyse omics data in order to define an specific lipidomic signature of Amyotrophic Lateral Sclerosis (ALS).

Background

Metabolomics is the study of small molecules or metabolites. Metabolites are the result of cellular metabolism and are easy to correlate with the phenotype, defining both physiological and pathological situations. Although recent technological techniques have allowed determining a wide range of metabolites in biological samples, data analyses is nowadays the main bottle neck. So, the improving and develop of omic specific softwares could help to biomarker search.

ALS is a fatal neurodegenerative disease with an unknown etiology. In this sense, the biomarker discovery of ALS could help to describe new therapeutic targets.

Methods

A mass-spectrometry metabolomic analysis using plasma and cerebrospinal fluid sample of ALS patients and control was performed. Data analysis was done using a R-based software specifically designed for metabolomics analyses.

Results.

There is a specific lipidomic signature in plasma which defines ALS pathology. Sphingolipids metabolism seems to have a crucial role in ALS.

Conclusions

The use of specific designed softwares to analyse omic data could help to biomarker discovery to monitor both disease progression and treatment and contribute to etiology description.

Index

1. Introduction	13
1.1. Work context and justification.	13
1.1.1. General description:	13
1.1.2. Work justification:	13
1.2. Work objectives.....	14
1.3. Approach and methodology	14
1.4. Work plan	15
Weeks are organized in columns and tasks in rows. Blue fill indicates that the row task has been performed in the column week.	15
1.5. Obtained products summary.....	15
1.6. Short description of the rest of the chapters	16
2. Previous concepts	18
2.1. Metabolomics and lipidomics (the search of biomarkers of diseases)	18
2.2. Big data analysis	19
2.3. Amyotrophic lateral sclerosis (ALS): a devastating and traumatic neurodegenerative disease	25
ALS pathophysiology.....	26
2.4. Diagnosis criteria	27
3. Methodology.....	29
3.1. Chemicals	29
3.2. Study population and sample collection	29
3.3. Lipidomic Analysis	29
3.3.1. Preparation of lipid standards	29
3.3.2. Lipid Extraction	30
3.3.3. LC-MS Method.....	31
3.3.4. Data analysis.....	31

4. Results.....	34
4.1. Script	34
4.1.1. Data reading	34
4.1.2. Results displaying	35
4.2. Cases of study.....	41
4.2.1. ALS vs CTL	41
4.2.2. Onset (Bulbar vs Spinal vs Respiratory)	47
4.2.3. Gender.....	52
5. Discussion.....	58
5.1. Developed script	58
5.2. Analysis results.....	59
6. Conclusions	63
6.1. Conclusions	63
6.2. Objective accomplishment.....	63
6.3. Work planning and methodology	63
6.4. Work limitations.....	64
6.5. Future lines of investigation	64
7. Bibliography	65
8. Annexes.....	70
8.1. Annex 1: Interactive shiny markdown script	70
8.2. Annex 2: Tables of metabolites with lower p value.....	81
8.2.1. CTL vs ALS plasma positive	81
The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.....	84
8.2.2. CTL vs ALS plasma negative	85

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time; Yellow font: LC-MS method not prepared to detect this metabolite.	86
8.2.3. CTL vs ALS CSF positive	87
The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.	88
8.2.4. CTL vs ALS CSF negative	89
The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.	90
8.2.5. Onset plasma positive	91
The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Red font: Unexpected retention time.	92
8.2.6. Onset plasma negative	93
The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.	94
8.2.7. Onset CSF positive	95
The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Red font: Unexpected retention time.	96
8.2.8. Onset CSF negative	97
The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.	98

8.2.9. Gender plasma positive	99
Green fill: Confirmed by MS/MS spectra; Red font: Unexpected retention time.	100
8.2.10. Gender plasma negative	101
The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.	102
8.2.11. Gender CSF positive.....	103
The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time; Yellow font: LC-MS method not prepared to detect this metabolite.....	104
8.2.12. Gender CSF negative	105

List of tables

Table 1: Time planning calendar.	15
Table 2. Decision table of the tests that should be applied according to the number of variables and their nature.	23
Table 3. Sample sizes necessary for applying the most used parametric analyses to non-normal data.	25
Table 4. The Escorial criterion and Airlie House revision.....	28
Table 5. List of isotopically labeled lipid standards used for standarization.....	30
Table 6. Example of file format read by the script.	32
Table 7. Methods used for hierarchical clustering in every analysis.	32
Table 8. Plasma lipid species which contribute to discern between control (CTL) and amyotrophic lateral sclerosis (ALS) subjects.	43
Table 9. CSF lipid species which contribute to discern between control (CTL) and amyotrophic lateral sclerosis (ALS) subjects.....	47
Table 10. Plasma lipid species which contribute to discern between disease onsets.	49
Table 11. CFS lipid species which contribute to discern between disease onsets.	51
Table 12. Plasma lipid species which contribute to discern between gender.	54
Table 13. CFS lipid species which contribute to discern between gender.....	57
Table 14: Differential metabolites for CTL vs ALS comparison for plasma samples in positive ionization mode.	81
Table 15: Differential metabolites for CTL vs ALS comparison for plasma samples in negative ionization mode.	85
Table 16: Top 25 metabolites with lower p value obtained from the univariate tests for CTL vs ALS comparison for CSF samples in positive ionization mode.	87
Table 17: Top 25 metabolites with lower p value obtained from the univariate tests for CTL vs ALS comparison for CSF samples in negative ionization mode.....	89
Table 18: Differential metabolites for onset comparison for plasma samples in positive ionization mode.	91
Table 19: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for plasma samples in negative ionization mode.....	93

Table 20: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for CSF samples in positive ionization mode.	95
Table 21: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for CSF samples in negative ionization mode.	97
Table 22: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for plasma samples in positive ionization mode.....	99
Table 23: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for plasma samples in negative ionization mode.....	101
Table 24: Differential metabolites for gender comparison for CSF samples in positive ionization mode.	103
Table 25: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for CSF samples in negative ionization mode.	105

List of Figures

Figure 1. The biological organization of “-omes.”	18
Figure 2. Examples of variable simplification from two dimensions to one.	20
Figure 3 Example of dendrogram representing a hierarchical clustering.	21
Figure 4. Schematic representation of clustering algorithms.	22
Figure 5. File selection panel of the program	34
Figure 6. Window asking for sample organization.	34
Figure 7. Panel asking if mass and retention time are specified in the data	35
Figure 8. Window asking if data normalization is desired.	35
Figure 9. Selection of the kind of test (parametric or non-parametric) to be performed.	35
Figure 10. Scatterplots representing the metabolite p values.	36
Figure 11. Representation of the table of significant metabolites.	36
Figure 12. Boxplot for lipid 765.5472@4.618837.	37
Figure 13. Correlation heatmaps of pairs of metabolites.	37
Figure 14. Example of bar chart of the top 10 metabolites with stronger correlation to the metabolite C32 H47 N3 O S.	38
Figure 15. Scatterplot representing the sample abundances of metabolite 875.7893@10.10477 against the concentrations of the metabolite C36 H77 N3 O5.	38
Figure 16. Principal component analysis of the samples using all the metabolites detected in samples (left panels) or using only those metabolites which show statistically significant difference between groups (right panel).	39
Figure 17. PLS-DA plot example.	39
Figure 18. Examples of the displayed hierarchical clustering.	40
Figure 19. Principal component analyses between plasma samples from CTL and ALS groups.	41
Figure 20. Partial Least Squares Discriminant analysis for plasma samples from CTL and ALS groups.	42
Figure 21. Hierarchical clusterings of plasma samples from CTL and ALS subjects.	43
Figure 22. Principal component analyses between CSF samples from CTL and ALS groups.	45

Figure 23. Partial Least Squares Discriminant analysis for CSF samples from CTL and ALS groups.	46
Figure 24. Hierarchical clusterings of CSF samples from CTL and ALS subjects.	46
Figure 25. Principal component analyses between plasma samples from Bulbar, Spinal and Respiratory groups.	47
Figure 26. Partial Least Squares Discriminant analysis for plasma samples from Bulbar, Spinal and Respiratory groups.	48
Figure 27. Hierarchical clusterings of plasma samples from Bulbar, Spinal and Respiratory onset subjects.	48
Figure 28. Principal component analyses between CSF samples from Bulbar, Spinal and Respiratory groups.	50
Figure 29. Partial Least Squares Discriminant analysis for CSF samples from Bulbar, Spinal and Respiratory groups.	50
Figure 30. Hierarchical clusterings of CSF samples from Bulbar, Spinal and Respiratory onset subjects.	51
Figure 31. Principal component analyses between plasma samples from female (F) and male (M) groups.	53
Figure 32. Partial Least Squares Discriminant analysis for plasma samples from female (F) and male (M) groups.	53
Figure 33. Hierarchical clusterings of plasma samples from female (F) and male (M) subjects.	54
Figure 34. Principal component analyses between CSF samples from female (F) and male (M) groups.	55
Figure 35. Partial Least Squares Discriminant analysis for CSF samples from female (F) and male (M) groups.	56
Figure 36. Hierarchical clusterings of CSF samples from female (F) and male (M) subjects.	56

Glossary

ALS: Amyotrophic Lateral Sclerosis.

BHT: Butylated hydroxytoluene

CE: Cholesteryl Ester

Cer: Ceramide

CSF: Cerebro-Spinal Fluid

CSV: Comma Separated Value

CTL: Control

DG: Diacylglycerol

F: Female

FA: Fatty Acyls

FDR: False Discovery Rate

GL: Glycerolipids

GP: Glycerophospholipids

LMN: Lower Motor Neurone

M: Male

Mb: Metabolite

MG: Monoacylglycerol

PA: Phosphatidic acid

PC: Phosphatidylcholine

PCA: Principal Component Analysis

PE: Phosphatidylethanolamine

PG: Phosphatidylglycerol

PGF: ProstaGlandin F

PI: Phosphatidylinositol

PLS-DA: Partial Least Squares Discriminant Analysis

PS: Phosphatidylserine

RT: Retention Time

SM: Sphingomyelin

SP: Shingolipids

TG: Triacylglycerol

UMN: Upper Motor Neurone

UPLC: Ultrahigh Performance Liquid Chromatograph

1. Introduction

1.1. Work context and justification.

1.1.1. General description:

A population prospective study in debutant or recently diagnosed Amyotrophic Lateral Sclerosis (ALS, OMIM #105400) patients and adequate controls has been performed. In this study we investigate whether there is a specific ALS lipidomic profile of plasma and/or CSF which correlate with clinical phenotype and help to better understand the pathophysiological mechanism of ALS pathogenesis.

1.1.2. Work justification:

ALS is an adult onset neurodegenerative disease characterized by selective loss of both upper (comprising the cortex and the corticospinal tract) and lower (brainstem nuclei and ventral roots of the spinal cord) motor neurons leading to relentlessly worsening paralysis of voluntary muscles until death. The disease has a wide phenotypic range and mean survival of 3 years.

Although its etiology is unknown, it is assumed as multifactorial, with heritance and environment to be determinant for the pathogenesis.¹⁻³ Initially, ALS was classified as sporadic (when a familial linked heritage is missing or etiopathogenesis remains unclear) or familiar (genetically linked, representing up to 15%), but whereas is generally acknowledged that the clinical presentations of sporadic and familial ALS are indistinguishable, there are subtle differences in pathology⁴

Several studies have tried to establish a relationship between pathogenesis and environmental or lifestyle factors, but there is not a general consensus⁵. Furthermore, as recent clinical, physiopathological and genetic advances have revealed, ALS phenotype is heterogeneous and a wide variety of different clinical and pathological subtypes are being discerned. Thus, there is an urgent need of a reliable subclassification in order to reduce the prognosis uncertainty.

Systems biology-associated approaches (such as genomics, transcriptomics, proteomics and metabolomics) allow global characterization of complex global biological systems and they have been previously applied to shed light in the pathogenic mechanisms of neurodegenerative diseases. Unlike genes and proteins, metabolites serve as direct signatures of biochemical activity, being therefore easier to correlate with the phenotype. In this context, metabolomics and its derivative lipidomics have become powerful approaches that have been widely adopted for clinical diagnosis, opening a window to mechanistically investigate how biochemistry relates to clinical phenotype.^{6,7}

Metabolomic and lipidomic profile can be assessed in a variety of tissues and fluids including CSF and plasma which are a rich source of putative biomarkers of various neurological diseases.⁷⁻⁹ Thus, the study of lipidic metabolism deregulation in ALS arises as an interesting research field.

1.2. Work objectives

1.2.1. General objectives:

1. Determine the presence of lipidomic profiles capable of characterize differential phenotypes and the progression of ALS disease.

1.2.2. Specific objectives

- a. Stratify the patients according to age, gender, clinical phenotype and survival.
- b. Identify novel ALS biomarkers based on lipidomic profiles from plasma and CSF samples.
- c. Describe novel underlying mechanism of ALS pathology based on the results obtained in Objective b.

1.3. Approach and methodology

In order to accomplish the proposed objectives, several analyses of plasma and CSF samples of affected ALS individuals have been performed. Individuals confirmed of ALS by El Escorial criteria with different onsets (bulbar, respiratory and spinal) have been recruited in Bellvitge University Hospital within 10-24 months after symptoms appearance, and plasma and CSF samples of these subjects have been collected after an overnight fasting period.

These samples have ben randomized and processed in order to extract the plasma lipidic species in liquid phase with the finality to perform untargeted lipidomic analyses by liquid chromatography coupled to mass spectrometry (LC-MS). An Ultrahigh Performance Liquid Chromatograph coupled to a Quadrupole-Time Of Flight mass spectrometer (UPLC-Q-TOF) has been used for the analyses.

The obtained data has been processed and analysed using an auto scaling algorithm. Multivariate statistics like hierarchical clusterings, principal component analyses (PCA), partial least squares discriminant analyses and correlation analyses were performed, followed by an univariate analyses: Wilcoxon-Mann-Whitney and Kruskal-Wallis(+Dunn post-hoc) tests, obtaining, among other results, a list of differential metabolites capable of describe the samples. These metabolites have been identified by comparing their mass and retention time data against the appropriate databases, and identifications have been confirmed with tandem-mass spectrometry analyses.

1.4. Work plan

1.4.1. Tasks:

1. Collection of Plasma and CSF sample from ALS patients and controls (CTL): 1 week.
2. Collection of clinical information from patients: 1 week.
3. Lipid extraction for injection in UPLC-Q-TOF system: 2 weeks.
4. Chromatographic and mass spectrometry analysis in UPLC-Q-TOF system: 3 weeks.
5. Statistical analysis with specific R-based software for non-targeted lipidomic analysis: 3 weeks
6. Identification of compounds of interests by comparing MS/MS spectra against representative class standards: 4 weeks.

1.4.2. Calendar:

Time planning calendar is represented in Table 1.

Table 1: Time planning calendar.

	Week 0 (27 February - 5 March)	Week 1 (6-12 March)	Week 2 (13-19 March)	Week 3 (20 - 26 March)	Week 4 (27 March - 2 April)	Week 5 (3 - 9 April)	Week 6 (10-16 April)	Week 7 (17-23 April)	Week 8 (24-30 April)	Week 9 (1-7 May)	Week 10 (8- 14 May)	Week 11 (15- 21 May)	Week 12 (22- 24 May)
Sample reception	Blue												
Clinical information obtention	Blue												
Sample processing		Blue	Blue										
Cromatografic analysis				Blue	Blue	Blue							
Statistical analysis					Blue	Blue	Blue	Blue	Blue	Blue	Blue		
Compound identification							Blue	Blue	Blue	Blue	Blue		
Report writing	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue		
Report revision												Blue	Blue

Weeks are organized in columns and tasks in rows. Blue fill indicates that the row task has been performed in the column week.

1.5. Obtained products summary

- CSF and plasma sample data files.** These files contain all the information collected by the UPLC-Q-TOF equipment. They are read and processed with Masshunter Qualitative Analysis and Mass Profiler Professional and several parameters must be introduced in order to align the peaks corresponding to the same metabolites and to obtain a csv file containing the lipidomic profiles of the patients. Files are not presented due to the need of specific softwares to read them.
- Lipidomic profiles from patients.** CSV files containing the sample code, their group and the metabolite relative abundance in positive and negative ionization mode. Only lipid species present in at least 50% of the samples of 1 condition are shown.

- c. **Interactive R script (shiny) used for data analysis.** The script allows exploring the computer to select the file to analyse, to select if samples are organized in rows or in columns and if mass and retention time information is included. It can also perform parametric (t-test/ANOVA) or non-parametric (Wilcoxon/Kruskal-Wallis) tests (as selected), multivariate statistics (correlation analyses, PCA, PLS-DA and hierarchical clustering) and select several parameters. Code is presented and commented and functions are widely explained in results chapter. Script can be found in Annex 1.
- d. **List of lowest p value lipids obtained from univariate analysis.** Excel file containing 25 lipids with lower p value (or all differential lipid species if the number is higher) for every comparison in plasma and CSF samples: ALS vs CTL, according to onset and according to gender. List can be found in Annex 2.
- e. **Results of the multivariate analysis.** Figures obtained with the interactive R script. Those figures are exposed and commented in the results chapter, and discussed in the discussion chapter.

1.6. Short description of the rest of the chapters

- Previous concepts: In this chapter, necessary theoretical concepts to understand the work are described.
 - Metabolomics and lipidomics: The chapter will contain a general explanation of metabolomics and currently used techniques. There will be a focus on how metabolomic studies have contributed to the description of new biomarkers.
 - Big data analysis: This chapter will explain the main multivariate and univariate analyses used to perform data analysis.
 - ALS: In this chapter, mechanisms related to neurodegenerative diseases progression are described, making special mention to ALS.
 - Diagnosis criteria: The current criteria of diagnosis are commented in the text, as well as the upper and lower motor neuron symptoms.
- Methodology
 - Chemicals: Chemicals used for the study will be specified.
 - Study population and sample collection: Population features, such as inclusion criteria, age and gender will be described. Techniques used for the collection of samples and plasma isolation will be also described.
 - Lipidomic analysis: Global protocol will be described. Lipid extraction protocol, LC-MS method and data analysis procedure will be explained.
- Results: Results obtained from univariate and multivariate analyses applied to the following comparisons in plasma and CSF will be specified:
 - ALS vs CSF

- According to ALS onset (bulbar vs respiratory vs spinal)
- According to gender (Male vs Female), only for ALS samples

2. Previous concepts

Once a global view of the work has been explained, the necessary theoretical concepts for the work are introduced:

2.1. Metabolomics and lipidomics (the search of biomarkers of diseases)

Metabolomics is the newest of the omic sciences and consists in the study of small molecules in biological systems. Metabolites can be substrates or products of metabolic pathways derived of cellular functions, such as energy production and storage, signal transduction or apoptosis in live organisms. It includes amino acids, lipids, organic acids, nucleotides and other substances.¹⁰ With this technique, the effects of a disease or a specific diet, among others, can be studied, describing novel potential biomarkers or a metabolic pathway affected by a specific situation. This information could help to generate a new hypothesis which should be confirmed with alternative experiments.

Genomics profile configures a transcriptomic and proteomic pattern that, in turn, resulted in a metabolomics profile, always with the influence of environmental factors (Figure 1). Therefore, metabolites serve as direct signatures of biochemical activity and are easier to correlate with phenotype.⁹ In this context, metabolomics and its derivative, lipidomics, have become a powerful approach that has been widely adopted for clinical diagnosis and opens a window to investigate how mechanistic biochemistry relates to cellular phenotype.

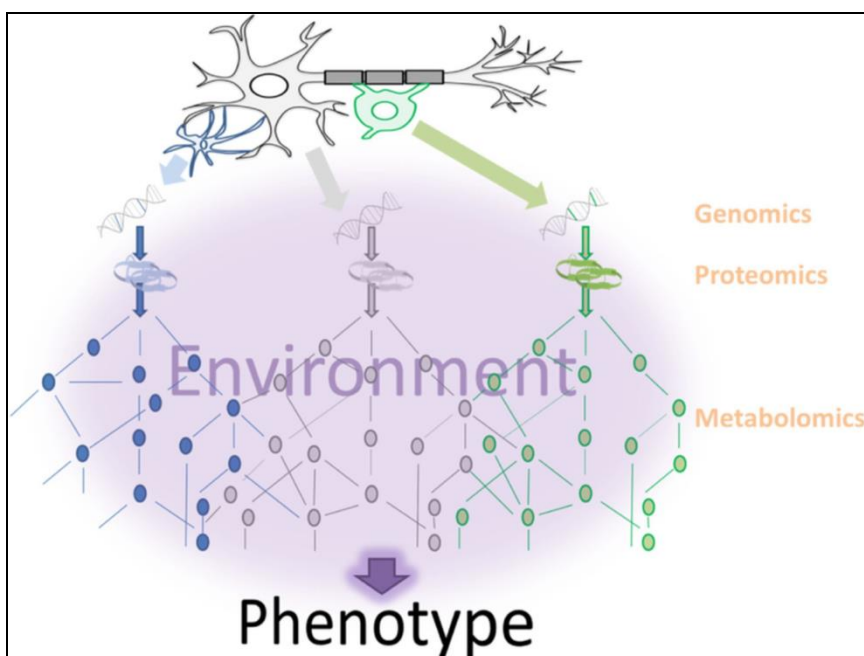


Figure 1. The biological organization of “-omes.” The classical view of cell organization considers the flow of information from the genome to the transcriptome, to the proteome, and then to the metabolome. Because each level of organization depends on the other, a change in 1 network can affect the others. In addition, the environment exerts an important influence in final metabolome. Adapted from Jové et al.⁹

Lipidomics, defined as the complete quantitative and molecular determination of lipid molecules isolated from biological samples, is a particular component of the metabolome. However, the different physicochemical properties of lipid species compared with water-soluble metabolites favour their separate analysis.^{6,11}

Metabolomic studies can be divided into two different types based on commonly used strategies: targeted and untargeted approach. As a hypothesis-generation strategy, untargeted approach is used when there is no initial hypothesis to obtain a global picture of both known and unknown metabolites, with the final goal of obtaining more comprehensive metabolomics data. On the other hand, targeted metabolomics approaches are used when a specified list of known metabolites is measured, typically focusing on one or more related pathways of interest.¹²

Recent advances in mass spectrometry and liquid chromatography (LC/MS) have strongly influenced the evolution of lipidomics. This advanced methodology makes it possible to identify and quantify hundreds of molecular species from the organism's lipidome. The high number of molecules obtained in a metabolomic or lipidomic analysis requires specific statistical methodology. Multivariate statistics simplify the interpretation of the variation between samples that contain thousands of variables, reducing the variation to a two- or three-dimensional model. Multivariate statistics include two major categories: supervised and unsupervised. Unsupervised techniques, principal component analysis (PCA) and hierarchical clustering analysis, are used to establish whether any intrinsic clustering exists within a data set, without a priori knowledge of sample class. In contrast, supervised methods use the class information given for a training set of samples to optimize the separation between two or more sample classes. These later include soft independent modeling of classification analysis and partial least discriminate analysis (PLS-DA), among others.^{13,14}

In this context, metabolomics and its derivative lipidomics (mainly focused on lipid species) arise as powerful approaches that have been widely adopted for clinical diagnosis, opening a window to mechanistically investigate how biochemistry relates to clinical phenotype.

2.2. Big data analysis

As it has been said, untargeted metabolomic studies generate amounts of data that need to be statistically treated. In order to reduce the complexity of the data, a set of variables capable to represent all objects must be obtained.

First of all, irrelevant data needs to be filtered. If a metabolite does not appear in a significant number of samples should be discarded. Several authors have accepted in animals only metabolites that appear in 75% of the samples of at least one condition.^{15,16} In humans, due to the higher variability of the samples, metabolites that appear in 50% of the samples of at least one condition are accepted.¹⁷

In order to be capable to analyse the whole data in a comprehensive way, multivariate statistics, including PCA, PLS and hierarchical clustering methods are used.

PCA is an unsupervised statistical procedure that approximates the sample variation to a low dimensional model in order to simplify and visualize it by giving different loading values to the variables. Then, sample scores are computed by multiplying each variable for its loading and obtaining the sum of the results. PCA produces two main plots depending on which data is displayed: for representing the variable loadings, a loadings plot is displayed where each axis contains the loading values of the corresponding principal component; and for representing the sample classification according to the loadings, a scores plot is displayed, where each axis represents the sample scores. Thus, PCA is used to extract and display the systematic variation in the data, representing more similar samples closer to each other and more different ones further from each other. This provides an overview of all the observations in the data table, allowing us to appreciate groupings, trends and possible outliers.¹³ A simplified example can be seen in Figure 2.

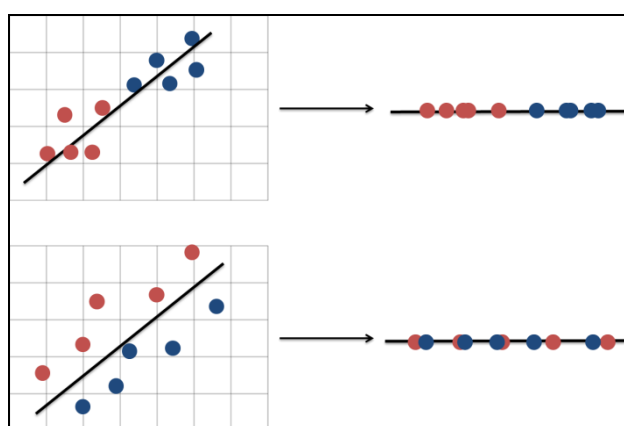


Figure 2. Examples of variable simplification from two dimensions to one. In the left panels, a two dimensional classification of the samples can be seen, with a black line representing the direction of maximum variation of the samples on which they are projected when the representation is done. In right panels, simplified representation can be seen, with the samples projected on the hyperplane. In general, n dimensions (corresponding to the number of metabolites) are analogously reduced to the desired ones (usually 2 or 3). Red circles represent molecules from one class and blue circles from another class. As can be seen, PCA does not have into account the class group but maximizes the variation between all the samples.

PLS is a supervised statistical procedure based on least squares regressions used to search for the fundamental relations between the dependent variables (in metabolomics case, the metabolites) and the independent variable by fitting linear regression models and projecting the variables to a new space. When, as in our study, the independent variables are categorical instead of numerical, a discriminant analysis (PLS-DA) is performed by converting the variables into dummy ones. In this case, linear models between metabolites and classes are searched, and higher loadings are given to the metabolites with a better correlation. Samples are represented according to their scores. Special care must be taken when analysing small number of samples with high number of variables because correlations may be found by chance and the scores plot would give a false suggestion of good separation between groups due to model overtraining. In this case, it should be better used as a suggestion that there are variables likely to be responsible of discrimination rather than a class predictor.¹⁸

Another widely used method for data description is the hierarchical clustering. It allows obtaining groups of samples from most similar to most different ones. Considering the

metabolites as an n dimensional space, distances between samples are computed. Then, samples can be clustered in two ways: i) in divisive clustering, an initial group containing all the samples is considered, and partitions of the cluster are made; ii) in agglomerative clustering, single samples are considered, and clusters are formed until a single group is obtained. The result is a dendrogram (Figure 3) that displays every division.

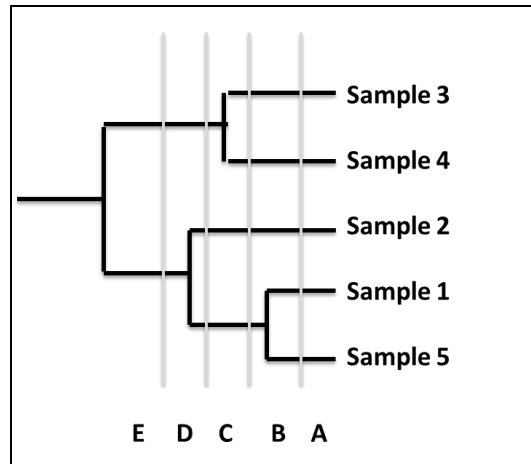


Figure 3 Example of dendrogram representing a hierarchical clustering. Merging lines represent a cluster formation. In region A, the five samples are separate. In region B, a cluster with samples 1,5 is formed. In region C, a cluster with 3,4. In region D a cluster with 2,1,5, and finally, in region E a cluster with all samples. Thus, the main difference between samples is between 3,4-2,1,5. Then, between 2-1,5. The following biggest difference is between 3-4, and the minimum one is between 1-5.

The merging lines of the dendrogram indicate a new cluster formation between the samples with closest data. In Figure 3, first cluster, that is to say, the merging of the two closest samples is formed between 1 and 5. Then samples 3 and 4 are merged, followed by sample 2 with the cluster 1,5. Finally 3,4 with 2,1,5, obtaining a single cluster.

In order to compute the distances between samples, several methods can be used. Same thing happens with the distances between clusters. The ones used by the *dist* and *hclust* R functions are commented in the following lines:

- Distance computing
 - o Euclidean: A straight line between two points: $\sqrt{\sum_{i=1}^n (p_i - q_i)^2}$
 - o Maximum: Maximum metabolite difference: $\max |p_i - q_i|$
 - o Manhattan: Sum of all the metabolite differences: $\sum_{i=1}^n |p_i - q_i|$
 - o Canberra: Weighted version of manhattan distance. $\sum_{i=1}^n \frac{|p_i - q_i|}{|p_i| + |q_i|}$
- Clustering algorithm: Consists on the criteria used to compute the distances between clusters. Merging of two clusters is formed for the closest pair of them. A schematic representation of the methods is represented in Figure 4. For more information, the following references can be consulted.¹⁹⁻²¹.

- Single: The distance between two clusters corresponds to the minimum distance between a pair of their samples. It is a good choice for obviously separated clusters.
- Complete: The distance between two clusters corresponds to the maximum distance between two of their samples. The results tend to be sensitive to outliers.
- Average: The distance between two clusters corresponds to the average distance between all their pairs of samples.
- Median: The distance between two clusters corresponds to the median of the distances between all their pairs of samples. It reduces the outlier effect compared to the average criteria.
- Mcquitty: The distance of a new formed cluster to the other ones is computed as the mean of the distances between the two clusters joined to the other one. For example, if clusters 1 and 2 are to be joined, the distance of this cluster to cluster 3 is the average of the distances from 1 to 3 and 2 to 3.
- Centroid: The distance between two clusters corresponds to the distance between their centroids. A centroid is the arithmetic mean of the position of all the samples in a cluster.
- Ward: With Ward's linkage method, the distance between two clusters is the sum of squared deviations from points to centroids. The goal of Ward's linkage method is to minimize the within-cluster sum of squares. It tends to produce clusters with similar numbers of observations, but it is sensitive to outliers. It is thought to be used with squared Euclidean distances, but it has been reported to behave correctly for non-Euclidean distances.²²

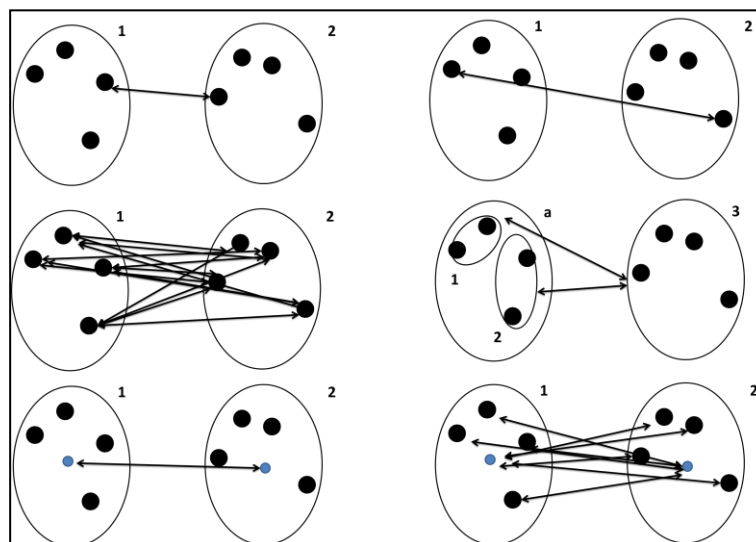


Figure 4. Schematic representation of clustering algorithms. From left to right and from up to down, single, complete, average or median, McQuitty, centroid and Ward methods. 1: Cluster 1; 2: Cluster 2; 3: Cluster 3. Black points correspond to individual samples, and the surrounded ones correspond to a same cluster. Blue points correspond to the cluster centroid. Arrows correspond to the measured distances to calculate the cluster distance.

Once a general overview of the data is obtained, univariate statistics can be applied in order to obtain specific differential metabolites between the comparisons wanted to be studied, that is to say, metabolites that have different concentrations depending on the sample group.

Different univariate tests can be performed according to the number of independent variables and the nature of the data, summarized in Table 2:

Table 2. Decision table of the tests that should be applied according to the number of variables and their nature.

Number of Dependent Variables	Nature of Independent Variables	Nature of Dependent Variable(s)	Test(s)
1	0 IVs (1 population)	interval & normal	one-sample t-test
		ordinal or interval	one-sample median
		categorical (2 categories)	binomial test
		categorical	Chi-square goodness-of-fit
	1 IV with 2 levels (independent groups)	interval & normal	2 independent sample t-test
		ordinal or interval	Wilcoxon-Mann Whitney test
		categorical	Chi-square test
			Fisher's exact test
	1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA
		ordinal or interval	Kruskal Wallis
		categorical	Chi-square test
	1 IV with 2 levels (dependent/matched groups)	interval & normal	paired t-test
		ordinal or interval	Wilcoxon signed ranks test
		categorical	McNemar
	1 IV with 2 or more levels (dependent/matched groups)	interval & normal	one-way repeated measures ANOVA
		ordinal or interval	Friedman test
		categorical	repeated measures logistic regression
2 or more IVs (independent groups)	interval & normal	factorial ANOVA	
	ordinal or interval	ordered logistic regression	
	categorical	factorial logistic regression	
1 interval IV	interval & normal	correlation	

		interval & normal	simple regression linear
		ordinal or interval	non-parametric correlation
		categorical	simple regression logistic
	1 or more interval IVs and/or 1 or more categorical IVs	interval & normal	multiple regression
			analysis of covariance
		categorical	multiple regression logistic discriminant analysis
2+	1 IV with 2 or more levels (independent groups)	interval & normal	one-way MANOVA
	2+	interval & normal	multivariate multiple linear regression
	0	interval & normal	factor analysis
2 sets of 2+	0	interval & normal	canonical correlation

Table must be ridden from left to right in order to determine which test to use, and from right to left to know what a specific test is used for. For example, 2 independent sample t-test would be used for interval and normal data, with 1 independent variable with 2 levels (with independent groups) and 1 dependent variable. Obtained from ²³.

The main concern about these tests is to discern between the using of parametric or non-parametric tests. Parametric tests assume that the data is continuous and follows a normal distribution, and non-parametric tests don't, but assume that groups have the same dispersion. This explains why in Table 2, the main difference in the nature of dependent variables is whether they are normally distributed or not. However, other aspects should be taken into account when choosing between parametric or non-parametric:

Parametric tests have more statistical power than non-parametric, and are more suitable than non-parametric when groups with different dispersion are analyzed. Furthermore, they also behave well with non-normal data for determinate sample sizes (Table 3).²⁴

On the other hand, for skewed normal data which is better represented by the median, for very small sample sizes and for ordinal, ranked data or for data with outliers that cannot be removed, non-parametric tests are more adequate.

These analyses are usually based on a conservative null Hypothesis (in this case, that the lipid abundance does not differ between groups). The results give a confidence value of the rejection of the hypothesis (in this case, that de lipid abundance differs between groups), corresponding to (1-p value)%. Usually, a rejection of the null hypothesis with 95% (p value = 0.05) of confidence is accepted.

Table 3. Sample sizes necessary for applying the most used parametric analyses to non-normal data.

Parametric analyses	Sample size guidelines for nonnormal data
1-sample t test	Greater than 20
2-sample t test	Each group should be greater than 15
One-Way ANOVA	For 2-9 groups, each group should be greater than 15. For 10-12 groups, each group should be greater than 20.

First column contains the parametric analysis to be performed, and the second column contains the conditions to be accomplished.

Although, when not only one lipid is analyzed, false discoveries due to abundance variances that might not be caused by the studied condition are more able to appear. For this motive, a correction of the p value is usually performed. The most used correction methods are Bonferroni correction (dividing the critical p value of 0.05 by the number of lipids analysed, and accepting p values lower than the result), and the Benjamini-Hochberg False Discovery Rate, which is less restrictive and assigns a critical p value depending on the false discovery rate that is accepted and the p values of the lipids. The critical p value is computed by $p_{critical}=(i/m)Q$, where m is the number of lipids analysed, Q is the FDR accepted and $i=1$ for the lipid with lowest p value, $i=2$ for the second and so on. For $p < p_{critical}$, the molecule and all the molecules with lower i value are accepted.^{25,26}

Once univariate analyses are applied and the initial number of metabolites is reduced, multivariate analyses may be applied again in order to know whether the few metabolites obtained are able to describe our data.

2.3. Amyotrophic lateral sclerosis (ALS): a devastating and traumatic neurodegenerative disease

Neurodegenerative diseases are at the critical point of clinical and scientific research interest as their prevalence increases year after year due to increasing life expectancy of world's population. They represent a big challenge for basic and clinical science because of their prevalence, costs, complexity and the lack of specific drugs to treat them. Moreover, many other health and social care disciplines play a vital role in their management, including physiotherapy, speech and language therapy, occupational therapy, podiatry and dietetics. Their consequences affect physical, cognitive and psychological domains and their impact cover almost every social, cultural and economic area.

One of the most devastating neurodegenerative diseases and with worse prognosis is ALS.

Amyotrophic lateral sclerosis (ALS), is an adult-onset, progressive and fatal neurodegenerative disorder involving primarily a degeneration of upper motor neuron in the primary cortex and lower motor neurons in the brainstem and spinal cord.²⁷

Incidence rates for ALS range from 1.2–4.0 per 100,000 person-years in Caucasians.^{28–31} The rate may be lower in some ethnic populations including American Natives,³² and, historically, as much as 50 times higher in Guam, Japan's Kii Peninsula, and western New Guinea. Incidence rates increase with age, peaking between 70 and 80 years, and are higher in men than women.³³ ALS is considered an ageing-related disease.

For the last few years the concept of ALS has changed from the aforementioned definition to a multisystem disorder which shows significant overlaps with frontotemporal dementia.³⁴

Taking into account that sporadic cases of ALS portrays around a 90% of ALS cases, it is becoming clear that ALS is a polygenic disease with a variable penetrance.³⁴

The clinical phenotype can be classified according to the anatomical onset site. Most studies report two typical presentations at disease onset, spinal and bulbar onset, corresponding to about 70% and 30% of the cases.³⁵ However, respiratory onset can also be a presenting symptom in a very small number of ALS patients (less than 3% of cases).³⁶

Spinal onset is the most common form of the disease and symptoms start at the limb. The presenting symptoms are muscle weakness and limb and trunk atrophy. However, most of them exhibit bulbar changes with disease progression.^{37,38}

Bulbar onset first affection is on face, mouth and throat muscles. Symptoms are characterized by impairments in speech (dysarthria) and swallowing musculature (dysphagia). It is the most devastating variant, with fastest decline, shortest survival and significantly reduced quality of life.^{39,40}

Respiratory onset initially affects trunk muscles and presents respiratory muscle impairment, which is also one of the main prognostic factors in ALS. Most deaths from this disease are due to respiratory failure.^{37,41} Thus, a respiratory onset is also devastating, with no significant differences from the survival time in patients with bulbar onset ALS.⁴² In all cases the disease spreads and affects the other regions.⁴³

Despite variations in clinical presentation and disease evolution, the progression of ALS inexorable with a 60% mortality rate three years after diagnosis. However, recent clinical, pathophysiological and genetic discoveries revealed more heterogeneous phenotype in ALS with survival ranges from months to decades, creating prognostic uncertainty. Thus, the existence of both a phenotypical substrate and a genetic susceptibility suggests that ALS should be considered a syndrome instead of a defined disease.⁴⁴

ALS physiopathology

The pathophysiology of ALS is complex and today it is only partially understood. Mechanisms of neurodegeneration include both the motor neurons and non-neuronal cells of the nervous system. Protein aggregation in ALS is one of the major histopathological features observed in post-mortem tissue neurons of patients, both familiar ALS and sporadic ALS.^{45,46} Protein

aggregates, commonly called inclusions or inclusion bodies, are dense, rounded and ubiquitin-positive structures. They can be found predominantly in neurons in various areas of the nervous system such as the frontal and temporal cortex, hippocampus, brainstem, and spinal cord.⁴⁷ In addition, these inclusions have been localized, not only in neurons, but also in glial cells of the spinal cord of patients (Arai et al., 2006). In animal models it has been demonstrated that protein aggregates appear at the beginning of the disease and accumulate progressively until the final phase of the disease.⁴⁸

TDP-43 is the main component of inclusion bodies in both motor neurons of ALS patients and in frontotemporal dementia cortical neurons.^{46,49} This discovery marked a milestone in the investigation and comprehension of these diseases. It is known that positive inclusions to TDP-43 are found in 97% of cases of ALS, both familial and sporadic.^{50,51} The main exceptions are cases caused by mutations in SOD1 or FUS.⁵² As it is previously explained, protein aggregates is a common feature with other NDD pathologies, in which the aggregates constitute one of the most useful diagnostic tools. On the other hand, the association between neurodegeneration and altered protease in the form of aggregate suggests the importance of the relationship between maintenance of the physiological levels of a protein in cytosol and the survival capacity of the neuron, potentially altered in processes such as aging and / or neuroinflammation.

It has been previously described that lipid metabolism is altered in ALS patients and there is an association between indices of dyslipidemia and prognosis.⁵³⁻⁵⁶ Therefore, the study of whole circulating lipidome in ALS samples could be crucial in the discovery of novel biomarkers which could become clinical therapeutic targets.

2.4. Diagnosis criteria

Because of the great clinical variability in presentation and prognosis, the generation of a systematic, consistent description of clinically defined subtypes is not easy. Nevertheless, a classification system for ALS that includes diagnostic criteria and phenotype at the time of diagnosis (clinical presentation) and as the disease progresses (clinical subtype) would be important to help guide treatment, provide an indication of prognosis, and enable analysis in clinical trials of homogenous group for a more personalised approach to therapy, and would be valued by patients and their families.⁵⁷

The **El Escorial criterion** (1994), which was developed by the World Federation of Neurology Research Group on Motor Neuron Diseases to define research-based consensus diagnostic criteria, is the approach with the greatest agreement among experts. According to it, diagnosis of ALS depends on identification of both upper and lower motor neuron signs within body regions defined as bulbar, cervical, thoracic, and lumbar.^{27,58-60} There were four levels of diagnostic certainty, including definite, probable, possible and suspected ALS depending on the dissemination of neuron damage in the regions.⁶¹

To improve diagnostic sensitivity of the El Escorial criteria, in 2000 the “suspected” category was exchanged by a “laboratory-supported probable ALS” category.⁵⁸ Although the resulting **Airlie House criteria** were specific for ALS, sensitivity remained a challenge, particularly in the early stages of the disease. ALS diagnostic criteria from the El Escorial criteria and Airlie House

criteria are summarized on the Table 4. In 2008, a new revision (Awaji-Shima) incorporated a recommendation to use electrophysiological data in the diagnosis.⁵⁹ The conditions for classification are described in Table 4.

Table 4. The Escorial criterion and Airlie House revision.

		Definite ALS*	Probable ALS*	Laboratory-supported probable ALS*	Possible ALS*	Suspected ALS*
El Escorial criteria (1994)		UMN and LMN signs in three regions of the body†.	UMN and LMN signs in at least two regions, with some UMN signs rostral to LMN signs.	..	UMN and LMN signs in only one region, or UMN signs alone in two or more regions, or LMN signs rostral to UMN signs.	LMN signs
Airlie House criteria (2000)		UMN and LMN signs in the bulbar region and at least two spinal regions, or UMN signs in at least two spinal regions and LMN signs in three spinal regions.	UMN and LMN signs in at least two regions, with some UMN signs rostral to LMN signs.	Clinical evidence of UMN and LMN signs in only one region, or UMN signs alone in one region and electrophysiological evidence of LMN signs in at least two regions.	UMN and LMN signs in only one region, or UMN signs alone in two or more regions, or LMN signs rostral to UMN signs.	...

LMN = Lower motor neuron. UMN = Upper motor neuron. .. = Components are not part of the classification. *Neuroimaging and clinical laboratory studies must be done to exclude alternative diagnoses. † = Regions: bulbar, cervical (corresponding to neck, arm, hand, diaphragm, and cervical spinal cord-innervated muscles), thoracic (corresponding to back and abdomen muscles), and lumbar (corresponding to back, abdomen, leg, foot, and lumbosacral spinal cord-innervated muscles). Adapted from Al-Chalabi et al.⁶²

Upper motor neuron signs indicate that the lesion is above the anterior horn cell (i.e. spinal cord, brain stem, motor cortex), and are characterised by increased muscle tone (spasticity) and weakness in the extensors (upper limbs) and flexors (lower limbs), increased reflexes, Clasp-knife spasticity (initial resistance to movement that fades), loss of superficial reflexes, clonus (rhythmic contractions of muscles) and Babinski sign (big toe extension).

Lower motor neuron signs indicate that the lesion is either in the anterior horn cell or distal to the anterior horn cell (i.e. anterior horn cell, root, plexus, peripheral nerve). It is characterised by decreased muscle tone, weakness and atrophy in the muscles, flaccid paralysis, areflexia (loss of muscle stretch reflexes), fasciculations (muscle twitches) and fibrillations (seen only on electromyogram).⁶³

3. Methodology

3.1. Chemicals

Synthetic lipids were obtained from Avanti Polar Lipids Inc. (Alabaster, AL) and Sigma-Aldrich (Madrid, Spain). Fatty acid methyl ester standards were obtained from Larodan Fine Chemicals (Malmö, Sweden) and from Sigma-Aldrich (Madrid, Spain). Methyl tert-butyl ether liquid chromatography coupled to mass-spectrometry (LC-MS), acetonitrile LC-MS, isopropanol LC-MS, potassium chloride, chloroform, ammonium formate, and ammonium hydroxide were purchased from Sigma-Aldrich (Madrid, Spain); methanol was from Carlo Erba (Milano, Italy); acetone was from Riedel-de-Häen (Seelze, Germany); and LC/MS-grade isopropanol and formic acid were from Baker (Phillipsburg, NJ, EUA).

3.2. Study population and sample collection

To explore differential lipids as potential biomarkers underlying signatures that characterize differential phenotypes and disease progression in ALS we obtained 23 patients enrolled at the Hospital Universitary of Bellvitge, Barcelona, between 2015-2016. A group series of 23 cases: 13 male (4 bulbar, 3 spinal, 6 respiratory) and 10 female (3 bulbar, 1 spinal, 6 respiratory) was used.

All individuals were included in the study only if matching ALS diagnosis according to the El Escorial criteria. Patients were enrolled under formal consent between 10-24 months since symptom onset. CSF and blood collections were performed in the morning after fasting the whole night according to standardized operating procedure. We used CSF and plasma (obtained by centrifugation standard procedure and stored at -80 C) of age and sex matched individuals with no neurological or inflammatory conditions (n=10) for control patients.

Butylated hydroxytoluene (BHT) is a lipophilic organic compound used as antioxidant. It is added to the samples at the moment of collection in order to avoid metabolite oxidation during storage and processing of the samples. Its addition at the moment of sample collection could not be performed in all the samples. Thus, prior to lipid extraction, 1 μ L of 100 μ M BHT was added to 99 μ L of non BHT containing sample aliquots.

3.3. Lipidomic Analysis

3.3.1. Preparation of lipid standards

Lipid standards consisting of isotopically labelled lipids were used for external standardization (ie, lipid family assignment) and internal standardization (ie, for adjustment of potential inter- and intra-assay variances). Stock solutions were prepared by dissolving lipid standards in methyl tert-butyl ether at a concentration of 1 mg/mL, and working solutions were diluted to 2.5 μ g/mL in methyl tert-butyl ether. Lipid standards used are specified in Table 5.

Table 5. List of isotopically labeled lipid standards used for standarization.

Compound	Reference
1,3(d5)-dihexadecanoyl-glycerol	110537, Avanti Polar Lipids
1,3(d5)-dihexadecanoyl-2-octadecanoyl-glycerol	110543, Avanti Polar Lipids
1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphate	110920, Avanti Polar Lipids
1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine	110918, Avanti Polar Lipids
1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine	110921, Avanti Polar Lipids
1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1'-rac-glycerol-1',1',2',3',3'-d5)	110899, Avanti Polar Lipids
1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-myo-inositol	110923, Avanti Polar Lipids
1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)-sn-glycero-3-[phospho-L-serine]	110922, Avanti Polar Lipids
26:0-d4 Lyso PC	860389, Avanti Polar Lipids
18:1 Chol (D7) ester	111015, Avanti Polar Lipids
cholest-5-en-3 β -ol(d7)	LM-4100, Avanti Polar Lipids
D-erythro-sphingosine-d7	860657, Avanti Polar Lipids
D-erythro-sphingosine-d7-1-phosphate	860659, Avanti Polar Lipids
N-palmitoyl-d31-D-erythro-sphingosine	868516, Avanti Polar Lipids
N-palmitoyl-d31-D-erythro-sphingosylphosphorylcholine	868584, Avanti Polar Lipids
Octadecanoic acid-2,2-d2	19905-58-9, Sigma Aldrich

3.3.2. Lipid Extraction

Lipidomic analysis was based on a previously validated methodology⁶⁴. In order to precipitate protein fraction, 5 μ L of miliQ water and 20 μ L of methanol were added to 10 μ L of sample. After the addition, samples were vortex-mixed for 2 minutes. For lipid extraction, 250 μ L of methyl tert-butyl ether (containing internal lipid standards) were added, and samples were immersed in a water bath (ATU Ultrasonidos, Valencia, Spain) with an ultrasound frequency of 40 kHz and 100 W power, at 10°C for 30 minutes. Then, 75 μ L of miliQ water were added to the mixture, and organic phase was separated by centrifugation 3000rpm at 10°C for 10

minutes. 170 μ L of the upper phase containing lipid extracts was collected and stored in vials at -20°C in order to perform mass spectrometry analyses. A pool using 30 μ L of all samples was prepared and aliquoted in order to use it as quality control.

3.3.3. LC-MS Method

Lipid extracts were analysed by liquid chromatography coupled to mass-spectrometry (LC-MS) using a liquid chromatograph Agilent UPLC 1290 coupled to mass spectrometer Agilent Q-TOF MS/MS 6520 (Agilent Technologies, Barcelona, Spain). Analysis was based on published method⁶⁵. Sample compartment was refrigerated at 4°C . For each sample, 10 μ L of lipid extract was injected in the system with a 1.8 μm particle 100 \times 2.1 mm id Waters Acquity HSS T3 column (Waters, Mildford, MA) heated to 55°C . Gradient elution with two phases, A and B, phase A composed of 10 mM ammonium acetate in acetonitrile-water (40:60, v/v) and phase B composed of 10 mM ammonium acetate in acetonitrile-isopropanol (10:90, v/v), was applied to the system at a constant flow rate of 400 $\mu\text{L}/\text{min}$. The gradient started at 40% B and reached 100% B in 10 minutes and held for 2 minutes. Finally, the system was switched back to 60% B and equilibrated for 3 minutes. Data from electrospray positive and negative ionized species was obtained in duplicate runs of the samples. System operated in TOF full-scan mode, at 100 to 3000 m/z in an extended dynamic range (2GHz), using N_2 as nebulizer gas (5 L/min, 350°C). The capillary voltage was set 3500 V with a scan rate of 1 scan/s. Continuous infusion using a double spray with masses 121.050873, 922.009798 (positive ion mode) and 119.036320, 966.000725 (negative ion mode) was used for in-run calibration of the mass spectrometer. For MS/MS confirmation the same parameters as MS analyses was used adding collision voltages of 0V, 10V, 20V and 40V. Data was collected with the software *MassHunter Data Acquisition* (Agilent Technologies, Barcelona, Spain).

3.3.4. Data analysis

3.3.4.1. Lipidic profiles obtaining

For the data analysis, molecular features from samples are extracted with *MassHunter Qualitative Analysis* (Agilent Technologies, Barcelona, Spain) and molecular profiles are obtained with *MassHunter Profiler Professional* (Agilent Technologies, Barcelona, Spain). Tolerance conditions are specified in order to discern different molecules: molecules within a $0.1\% \pm 0.25$ minutes retention time window and $30\text{ppm} \pm 2\text{mDA}$ mass window are considered the same one. Only common features (found in at least 50% of the samples of the same condition) were taken into account to correct for individual bias.

Peak intensities are relativized by internal standard peak intensity.

Text file (.txt) containing relative intensities of each metabolite for each sample is obtained and formatted for later statistical analysis with specific R-based scripts.

3.3.4.2. Statistical analysis

Used R-based scripts are prepared to read comma separated value (.csv) files with the format (or transposed) explained in Table 6.

Table 6. Example of file format read by the script.

Group		Mb1	Mb2	Mb3	Mb...
Mass (optional)		Mass Mb1	Mass Mb2	Mass Mb3	...
Retention time (optional)		RT Mb1	RT Mb2	RT Mb3	...
Sample 1	Sample1 group	Concentration Mb1 sample1	Concentration Mb2 sample1	Concentration Mb3 sample1	...
Sample 2	Sample2 group	Concentration Mb1 sample2	Concentration Mb2 sample2	Concentration Mb3 sample1	...
Sample

Non-parametric tests (Mann-Whitney and Kruskal-Wallis+Dunn tests), and multivariate analyses (PCA, PLS-DA and hierarchical clustering) were performed. Methods for hierarchical clustering are specified in Table 7.

Table 7. Methods used for hierarchical clustering in every analysis.

Analysis	Sample	Ionization mode	Used molecules	Distance method	Clustering method
CTL vs ALS	Plasma	+	All features	Euclidean	Ward.D2
			Top 25	Manhattan	Ward.D2
			Significant	Manhattan	Ward.D2
		-	All features	Euclidean	McQuitty
			Top 25	Canberra	McQuitty
			Significant	Manhattan	Ward.D2
	CSF	+	All features	Euclidean	Ward.D2
			Top 25	Manhattan	Ward.D2
			Significant	Manhattan	Ward.D2
		-	All features	Euclidean	Ward.D2
			Top 25	Euclidean	Ward.D2
			Significant	Manhattan	Ward.D2
Onset	Plasma	+	All features	Manhattan	Ward.D2
			Top 25	Manhattan	Ward.D2
			Significant	Euclidean	Ward.D2
		-	All features	Canberra	Complete
			Top 25	Canberra	Complete
			Significant	Canberra	Complete
	CSF	+	All features	Euclidean	Ward.D2
			Top 25	Euclidean	Ward.D2

			Significant	Euclidean	Ward.D2
			All features	Canberra	McQuitty
		-	Top 25	Canberra	McQuitty
			Significant	Canberra	McQuitty
Gender	Plasma	+	All features	Euclidean	Ward.D2
			Top 25	Euclidean	Ward.D2
			Significant	Euclidean	Ward.D2
		-	All features	Euclidean	Ward.D2
			Top 25	Euclidean	Ward.D2
			Significant	Euclidean	Ward.D2
	CSF	+	All features	Euclidean	Ward.D2
			Top 25	Euclidean	Ward.D2
			Significant	Euclidean	Ward.D2
		-	All features	Canberra	Complete
			Top 25	Canberra	Complete
			Significant	Canberra	Complete

Table is ridden from left to right: First the comparison, then the type of sample, ionization mode and used molecules. Distance and clustering methods are specified. +: Positive ionization mode; -: Negative ionization mode.

Statistically significant molecules (with p value < 0.05) were identified by comparing their exact mass and retention time with specific databases^{66,67} in order to obtain potential identities. Those identities were confirmed by MS/MS spectra comparison against representative class standards using lipidmatch.⁶⁸

4. Results

This chapter is divided into two main parts. In the first one, the script obtained is commented and functionalities are shown. In the second part, results obtained with the data analysis are discussed.

4.1. Script

In order to perform data analyses, an R interactive script has been created with two main parts: the first one for data reading, and the second one for results displaying. The script has been implemented as an adaptation of the free software Metaboanalyst^{®69} already used by the research group, and specific group needs not already covered by the public software have been added. The code was fully developed by trial-error method and based on information from specific references.^{70–75}

4.1.1. Data reading

When the script is run, a window for file selection is displayed, where desired file to analyse must be chosen (Figure 5).

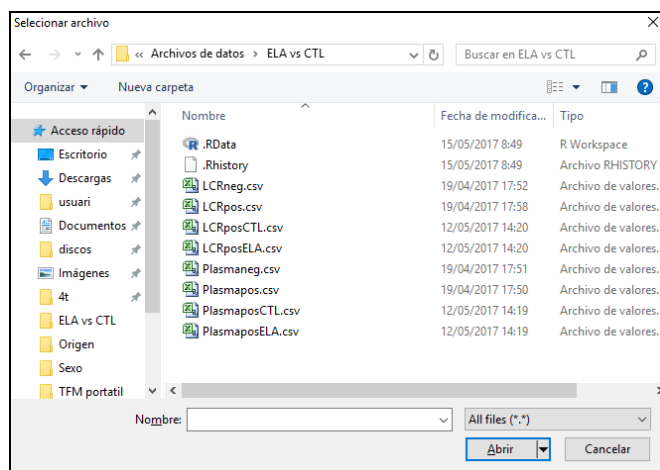


Figure 5. File selection panel of the program

Once file is selected, several questions are asked:

As Figure 6 shows, user is first asked if samples are organised in rows or in columns (user must introduce number 1 or number 2, respectively)

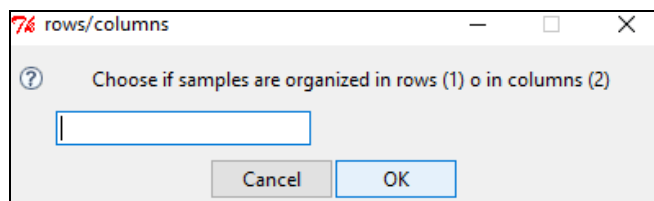


Figure 6. Window asking for sample organization.

Once specified the sample organisation, user is asked if he has included mass and retention time data (Figure 7) and if data normalization is desired to be done (Figure 8).

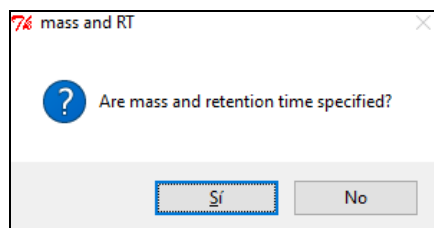


Figure 7. Panel asking if mass and retention time are specified in the data

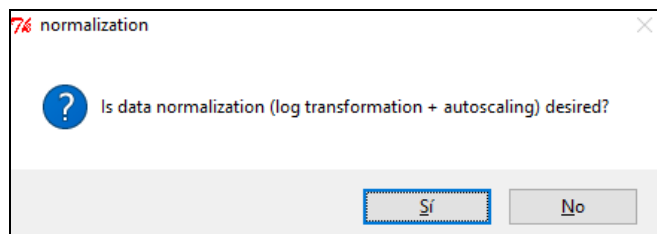


Figure 8. Window asking if data normalization is desired.

Finally, user is asked to choose between non-parametric or parametric analysis (Figure 9).

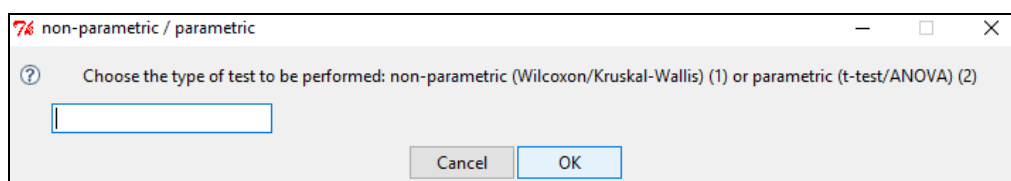


Figure 9. Selection of the kind of test (parametric or non-parametric) to be performed.

Once all the questions are answered, data is read and processed. Then, data analysis results are presented in an interactive interface.

4.1.2. Results displaying

First presented results are the ones from univariate analysis. 3 elements are used:

- Scatterplot representing the $\log_{10}(p)$ value for each metabolite (Figure 10). P threshold is represented with a grey horizontal line and can be modified. P correction method can be also selected, and p or p corrected values can be used for the graph. Compounds with a p value lower than the threshold value are represented over the horizontal line and colored red. Compounds with higher p values are represented under the line and colored blue. The title specifies the p threshold selected, the number of significant lipids and the type of test used.

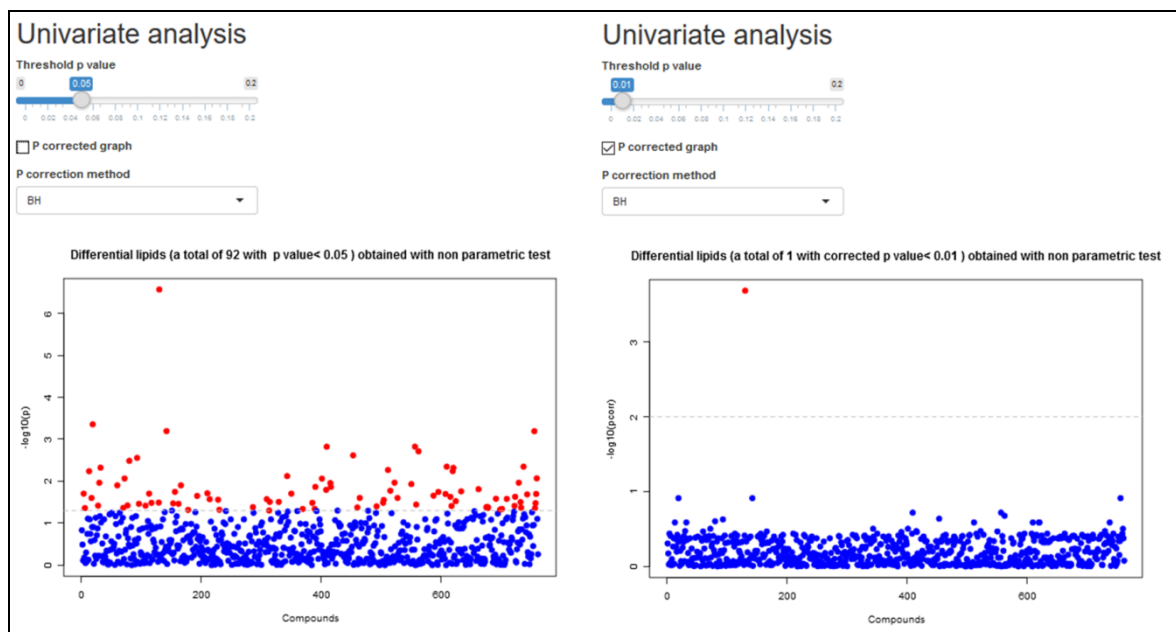


Figure 10. Scatterplots representing the metabolite p values. The metabolites are displayed along the x axis, and $-\log_{10}(p)$ is represented by the y axis. A horizontal grey line represents the threshold p value. Compounds with a p value lower than the threshold value are represented over the horizontal line and colored red. The title specifies the p threshold selected, the number of significant lipids and the type of test used. Compounds with higher p values are represented under the line and colored blue. In left panel, graph for uncorrected p values is displayed. Threshold value is set at $p=0.05$. In right panel, graph for corrected p values (by Benjamini Hochberg FDR) is displayed. Threshold value is set at $p=0.01$.

- A table which includes, among other data, the names of the differential metabolites, their p value, their regulation (up or down) or post-hoc test (in case of more than 2 groups), their mass, retention time and p corrected value. As in the scatterplots, p threshold value and p correction method can be selected, as well as the number of metabolites to be shown in the table. Screenshots of the results can be seen in Figure 11.

Metabolites to show

5

features	estimate.dif	pval	statistic.w	conf.int.1	conf.int.2	regulation (CTL vs ELA)	mass	RT	pcorr
C14 H22 N2 O	2.22	0.00	218.50	2.20	2.27	up	234.172	0.8815	0.00
765.5472@6.618837	-0.84	0.00	30.00	-1.58	-0.33	down	765.5472	6.618837	0.12
839.6116@6.6334314	-1.20	0.00	32.00	-1.81	-0.51	down	839.6116	6.633431	0.12
C20 H5 N3 O20	-2.07	0.00	29.00	-2.08	-0.03	down	606.9473	0.89468	0.12
1972.7578@10.5082855	0.03	0.00	170.00	0.00	2.56	up	1972.758	10.50829	0.19

Figure 11. Representation of the table of significant metabolites. “Metabolites to show” parameter is selected to 5. For every feature is displayed, in columns, the estimate difference, p value, statistic value, the confidence intervals, the regulation, mass, retention time and corrected p values for each metabolite. For unidentified metabolites feature is represented as exactmass@retentiontime.

- Boxplots of the metabolites (Figure 12). Metabolite names can be written in order to obtain their boxplots for the studied groups.

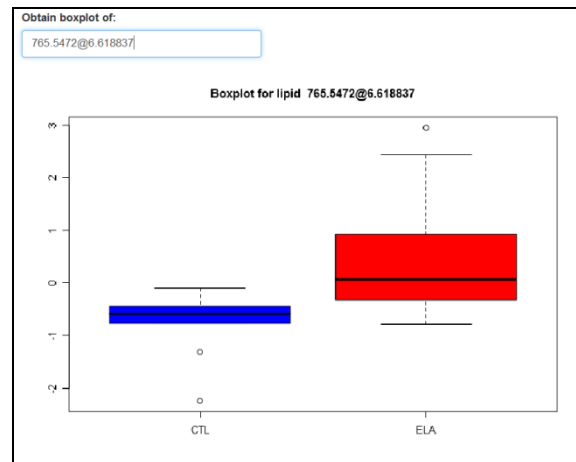


Figure 12. Boxplot for lipid 765.5472@4.618837. In blue, box for control is represented, and in red, box for ELA.

The second results shown are correlation analyses, also presented in 3 elements.

1. A heatmap of all the correlations. It can also be obtained with only significant metabolites (with p value < 0.05) (Figure 13). Pearson, Spearman or Kendall correlations can be applied.

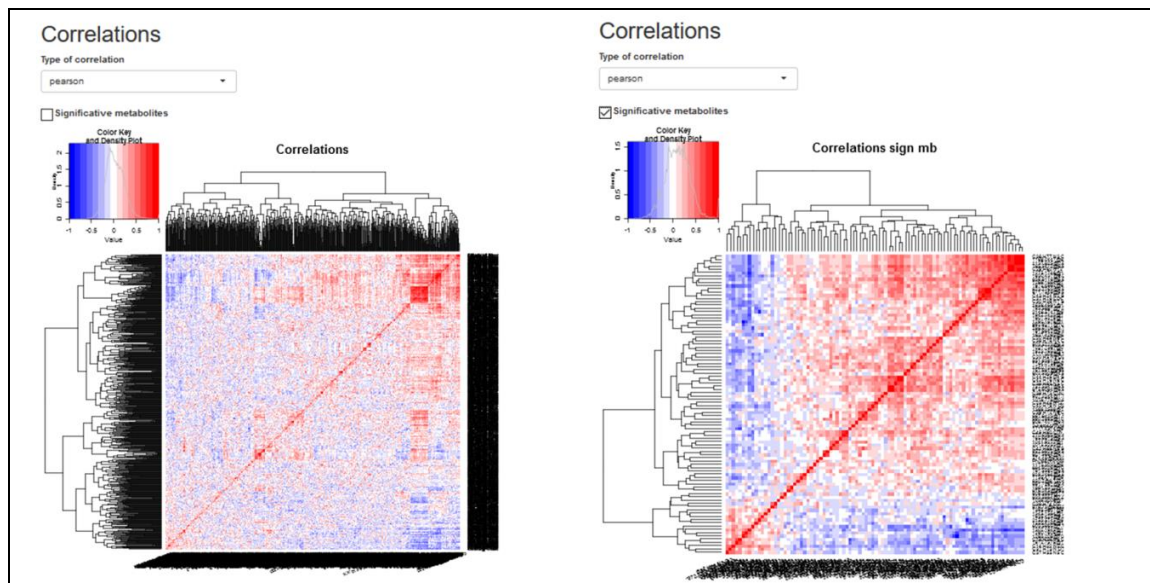


Figure 13. Correlation heatmaps of pairs of metabolites. Negative values are colored in blue, neutral values in white and positive values in red, with a degradation of the color for descending absolute values. In left panel, pearson coefficients of all the metabolites are shown. In right panel, Pearson correlations of significant metabolites are displayed.

2. Bar chart with the top 10 metabolites with higher correlation coefficient with a chosen metabolite (Figure 14). A specific metabolite can be selected from the ones present in the data by means of a drop-down list. Correlation values are represented in the bar chart. Red bars correspond to positively correlated metabolites and blue bars to negatively correlated ones:

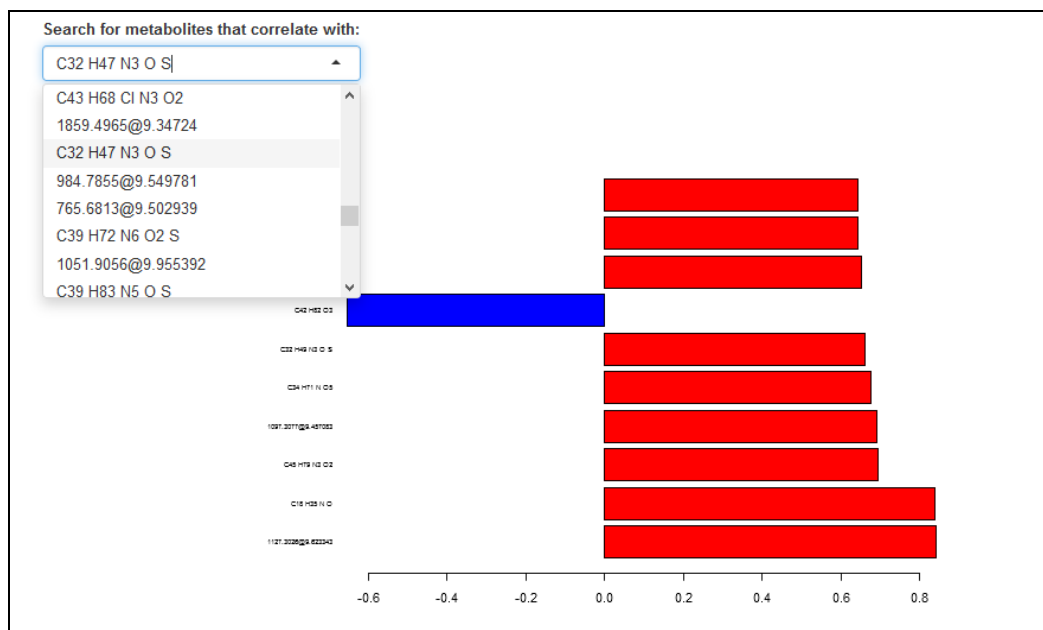


Figure 14. Example of bar chart of the top 10 metabolites with stronger correlation to the metabolite C32 H47 N3 O S. Correlation values are represented in the x axis, and feature names in the y axis. Red bars correspond to positively correlated metabolites and blue bars to negatively correlated ones

3. Scatterplot of the raw abundance in every sample of the previously selected metabolite against a newly selected one (Figure 15). The model line is also plotted.

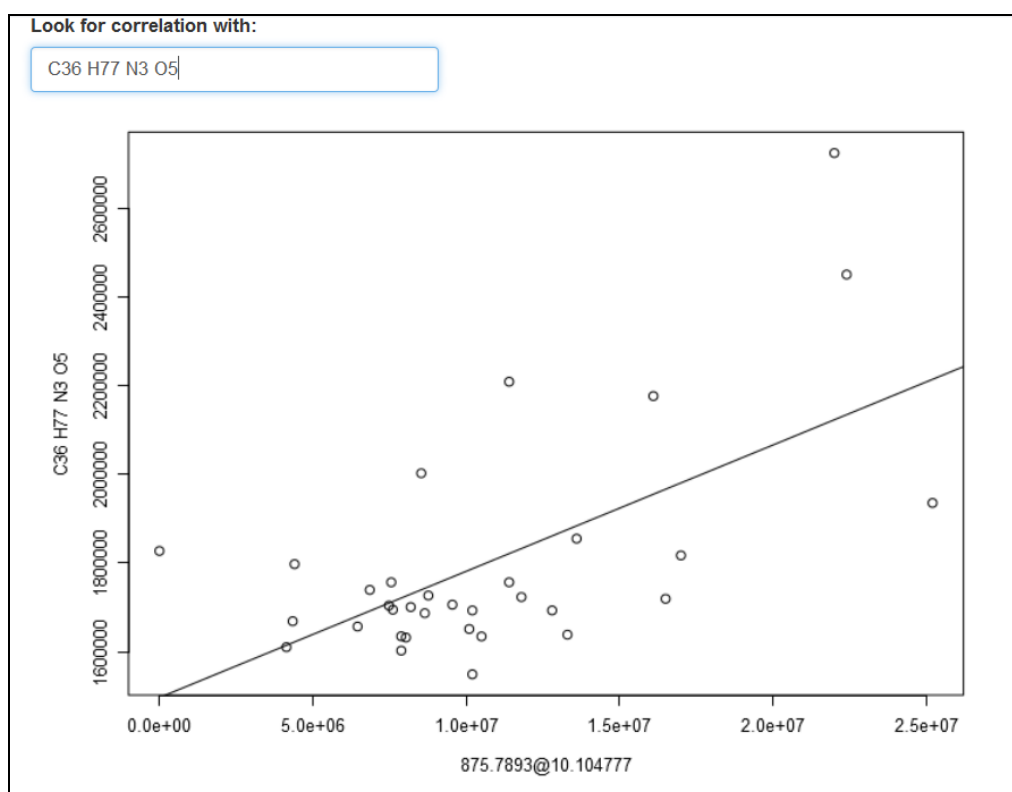


Figure 15. Scatterplot representing the sample abundances of metabolite 875.7893@10.10477 against the concentrations of the metabolite C36 H77 N3 O5. X axis: Raw abundance of lipid 875.7893@10.10477, Y axis: raw abundance of lipid C36 H77 N3 O5. Linear model is represented with a line.

Afterwards, multivariate analyses are presented with 3 more plots:

1. PCA (
2. Figure 16). It can be performed with all the metabolites or only with significant ones. When passing the mouse on a sample, its name and group is displayed in the screen. Loadings are specified in the corresponding axes. Explained variance of the components is also represented below the plots. Plot can be rotated or zoomed in order to have a better vision of the plot.

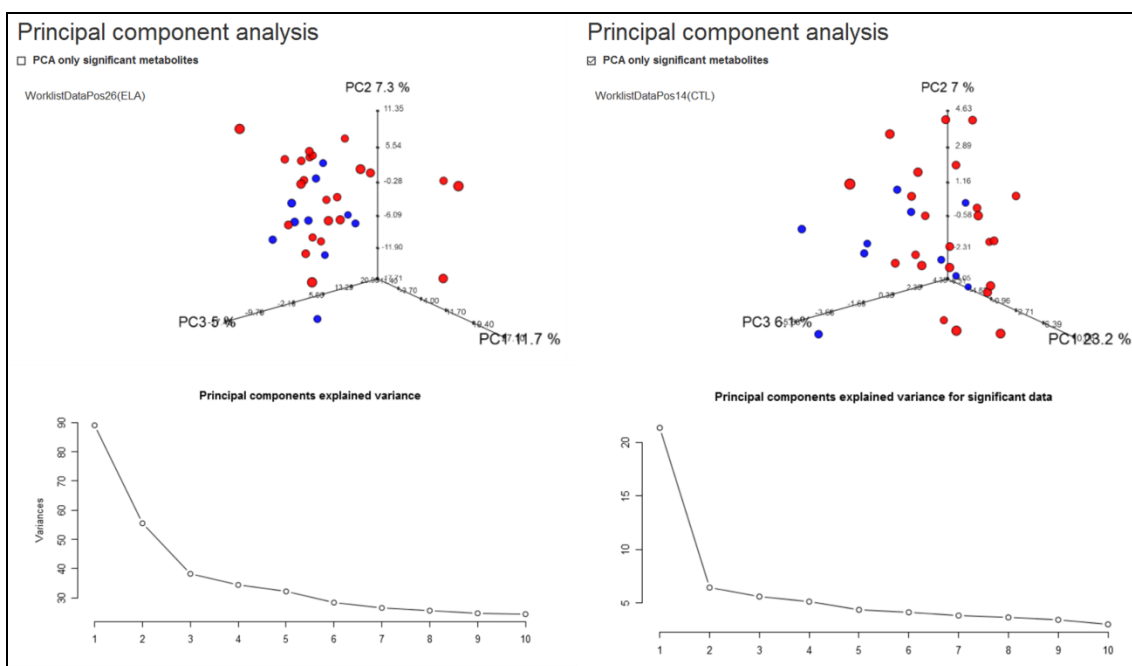


Figure 16. Principal component analysis of the samples using all the metabolites detected in samples (left panels) or using only those metabolites which show statistically significant difference between groups (right panel). Loadings are specified in the corresponding axes. Explained variance of the components is also represented below the plots.

3. PLS-DA. PLS-DA graph of the samples can be obtained (Figure 17). Plot can be rotated or zoomed in order to have a better vision of the plot. When passing the mouse on a sample, its name and group is displayed in the screen.

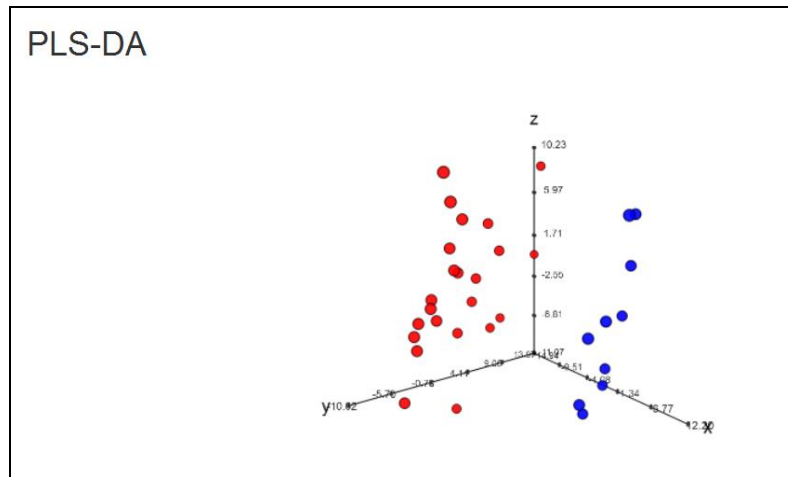


Figure 17. PLS-DA plot example.

4. Hierarchical clustering analyses represented by dendrograms. Dendrograms can be obtained using all the metabolites, only the statistically significant ones or the 25 metabolites with lower p value. Different algorithms for distance computing and clustering can be applied. Heatmaps of the metabolites can be added to the dendrograms (Figure 18).

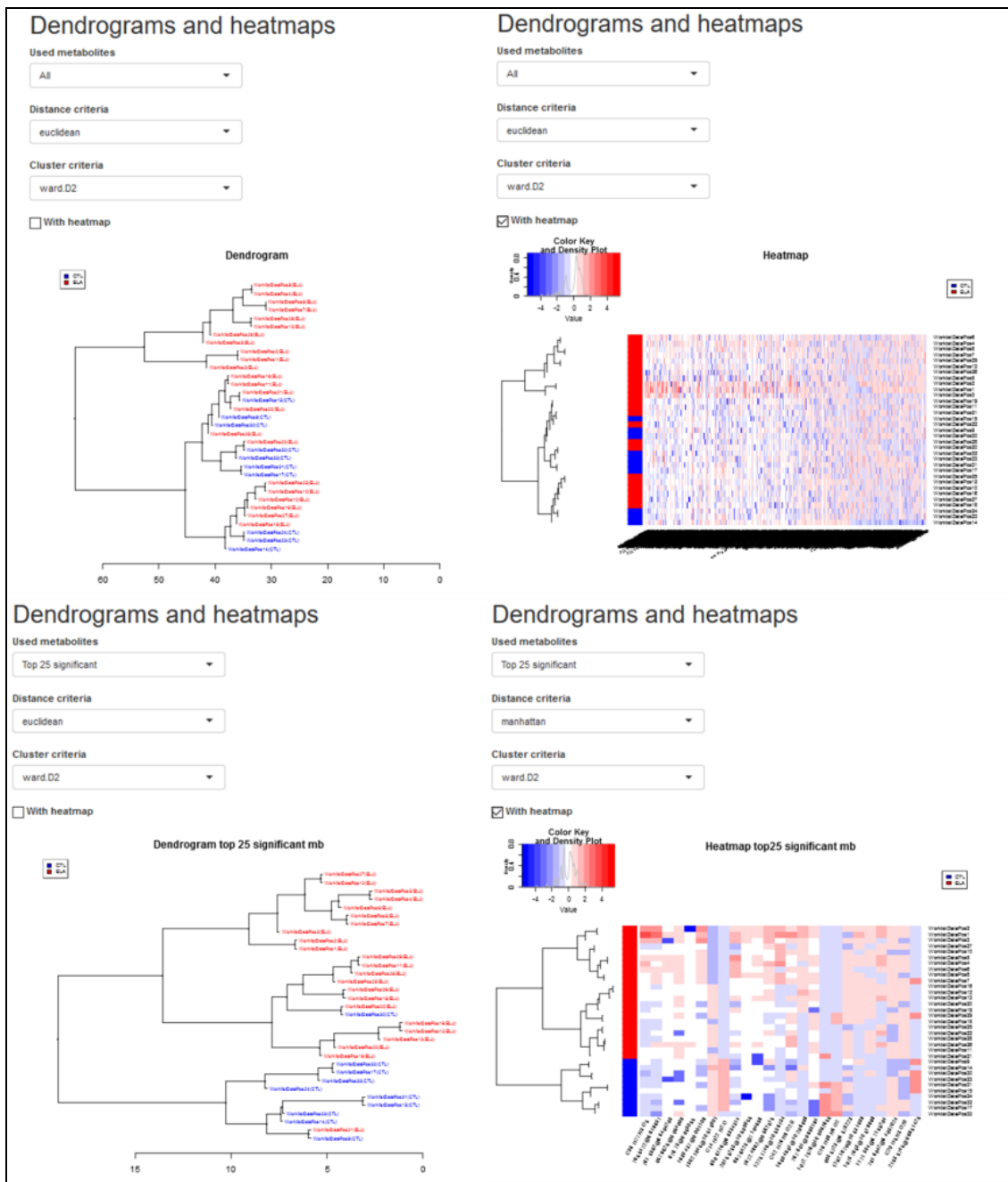


Figure 18. Examples of the displayed hierarchical clustering. Top left panel displays only the dendrogram using all the metabolites detected in samples, with Euclidean distances and ward.D2 clustering method. Top right panel displays the same conditions adding the heatmap. Bottom left panel displays the dendrogram for top 25 metabolites with the same parameters. Bottom right displays the dendrogram with heatmap of the top 25 significant metabolites with Manhattan distances instead of Euclidean ones.

4.2. Cases of study

As commented on the introduction section, ALS is considered one single disease with a wide phenotype, which implies lots of different subcategories. According to this, we decided to apply different approaches to our study. First, all ALS patients were compared to non ALS patients in order to determine the lipids involved in the general pathogenesis and mechanisms of the disease common in the various phenotypes. Secondly, a comparison between different

onsets was made to determine the specific lipids involved in each phenotype. Finally, as commented in the introduction, a lower incidence of the disease is observed in females. Thus, an approach comparing male and female ALS patients was performed in order to determine specific mechanisms involved in the disease progression according to gender.

For all the comparisons, results are presented by means of PCAs, PLS-DAs and heatmaps of positive and negative ionizations. The analyses were first applied to plasma samples and then to CSF ones.

4.2.1. ALS vs CTL

First comparison made was between diseased individuals against non-affected individuals. The main aim of this analysis was to determine general lipids involved in the ALS pathogenesis, without having into account a specific ALS classification.

4.2.1.1. Plasma

In PCAs using the whole lipidome (Figure 19a) we can observe a higher variability in ALS samples than in CTL ones.

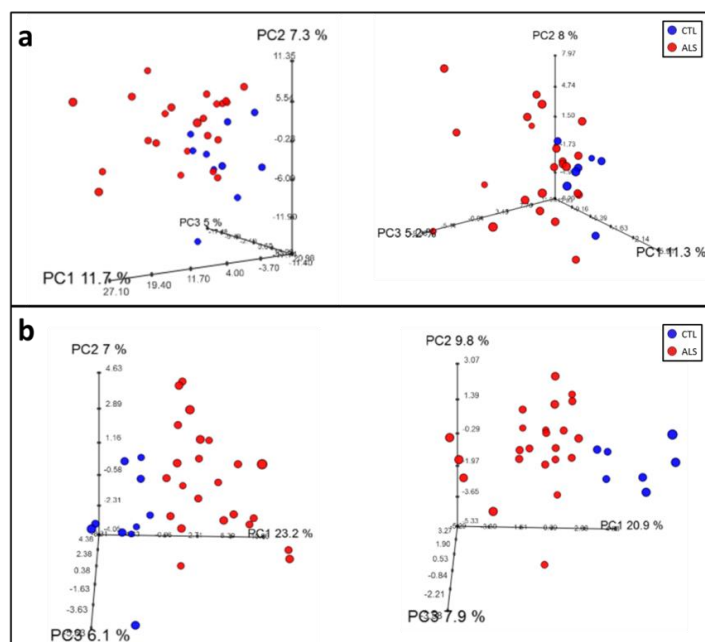


Figure 19. Principal component analyses between plasma samples from CTL and ALS groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from CTL group are represented with blue spheres, and individuals from ALS group with red spheres. Image a. PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. Although there is no clear separation, a tendency is observed in both graphs. b. PCA of samples using only lipids with $p < 0.05$ in Wilcoxon-Mann-Whitney test applied to positive (left panel) and negative (right panel) ion detection modes data. In this case we can observe that a good discrimination is done.

Although a group separation is not very clear, there is a tendency which indicates that several ALS patients are different of control ones according to their lipidome. Furthermore, if we use only the differential metabolites, we can clearly discern both groups (Figure 19b).

PLS-DA (Figure 20) confirms that it is possible to separate both groups using part of their lipidome.

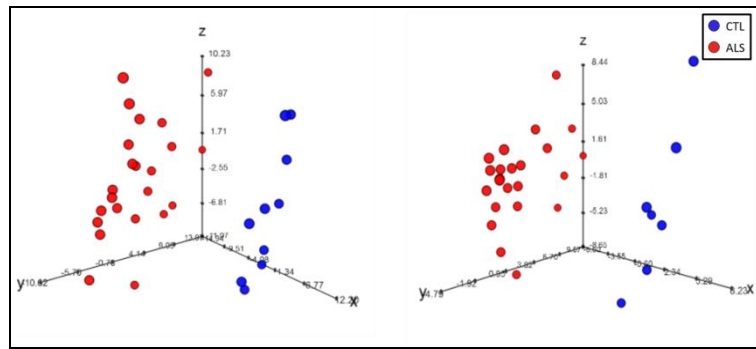


Figure 20. Partial Least Squares Discriminant analysis for plasma samples from CTL and ALS groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. Clear differences can be observed between groups. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from CTL group are represented with blue spheres, and individuals from ALS group with red spheres.

Heatmaps using all the features detected show the same tendency as the PCA analyses (Figure 21): there are samples that clearly differ from both groups and some others that mix together. In the positive case, a big cluster of ALS can be seen, and the other cluster contains the samples of both groups mixed. In negative ionization there are no clear big clusters, but control samples are all close. When using the top 25 species with lower p value, in positive ionization a clear separation can be seen with two main clusters, each one corresponding to one group.

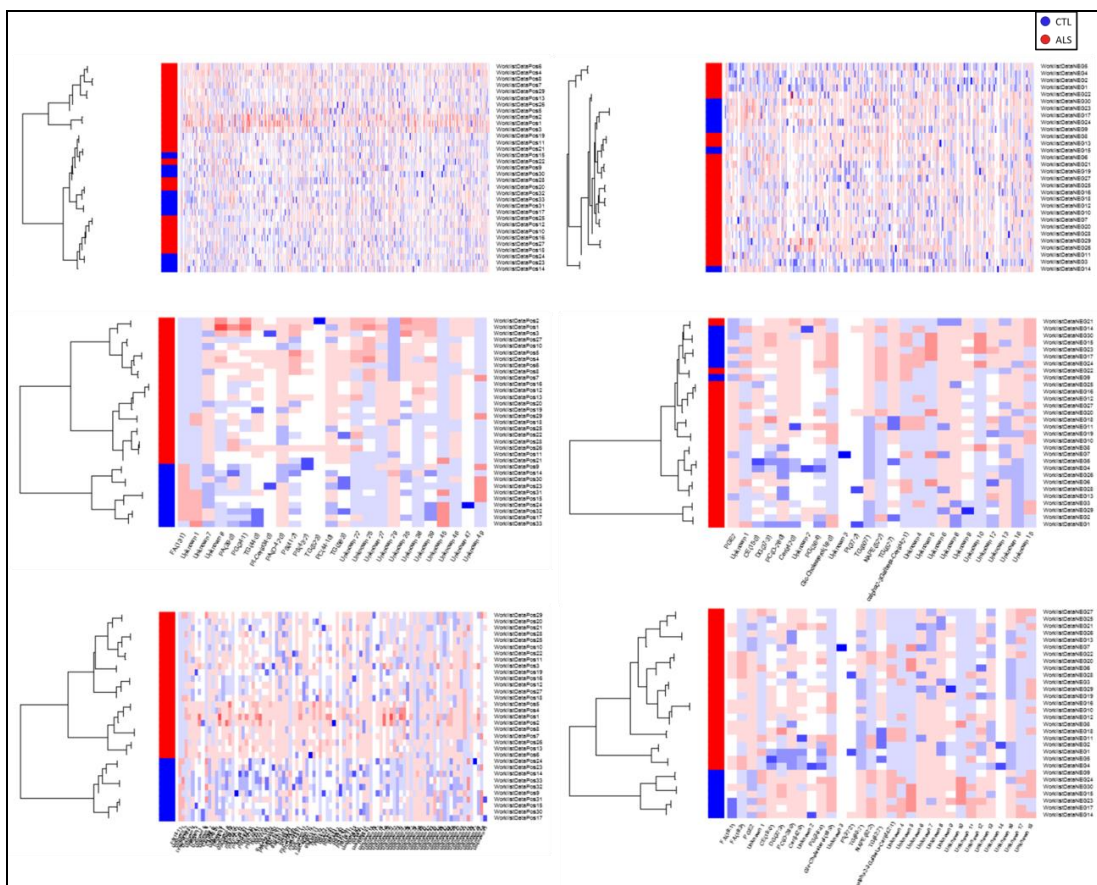


Figure 21. Hierarchical clusterings of plasma samples from CTL and ALS subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (CTL in blue and ALS in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Wilcoxon-Mann-Whitney test p value < 0.05) are displayed.

Thus, the molecules used allow us to describe perfectly our samples. In the case of negative ionization it is not as clear as in positive. However, when using all the differential molecules, perfect clustering for both analyses is seen. The characteristics of lipid species which contribute to clustering both groups are reported in Table 8.

Table 8. Plasma lipid species which contribute to discern between control (CTL) and amyotrophic lateral sclerosis (ALS) subjects.

Family	Metabolite	Regulation (CTL vs ALS)	Ionization mode
FA	FA(13:1)**	up	+
	FA(18:1)*	down	-
	FA(18:3)*	down	-
	PGE2*	down	-
GL	DG(37:3)*	up	-
	DG(38:3)*	up	-

	TG(48:8)*	up	-
	TG(60:7)*	up	-
	TG(62:7)*	up	-
	DG(34:1)*	down	+
	DG(36:4)*	down	+
	DG(38:2)*	down	+
	TG(42:0)*	down	+
	TG(44:0)*	down	+
	TG(44:1)*	down	+
	TG(52:3)*	down	+
	TG(52:4)*	down	+
	TG(54:8)*	down	+
	TG(55:1)*	down	+
	TG(56:5)*	down	+
	TG(56:8)*	down	+
	TG(58:11)*	down	+
	TG(58:3)*	down	+
	TG(63:3)*	down	+
	NAPE(52:2)*	up	-
	PC(36:4)*	up	+
	PC(O-28:0)*	up	-
	PG(36:4)*	up	-
	PI(37:2)*	up	-
	LysoPS(20:3)*	down	+
	PA(39:0)*	down	+
	PA(O-42:0)*	down	+
	PA(P-36:0)*	down	+
	PA(P-37:1)*	down	+
	PA(P-38:0)*	down	+
	PC(44:10)*	down	+
	PE(40:9)*	down	+
	PE(42:1)*	down	+
	PE(P-38:5)*	down	+
	PE(P-40:6)*	down	+
	PG(34:1)*	down	+
	PG(36:1)*	down	+
	PS(41:2)*	down	+
	PS(43:2)*	down	+
GP	PS(P-39:1)*	down	+
	Cer(d38:3)*	up	+
	Cer(t42:0)*	up	-
	NeuAalpha2-3Galbeta-Cer(d42:1)*	up	-
	Cer(t40:0)*	down	+
	LacCer(d32:1)*	down	+
	PI-Cer(d34:0)*	down	+
SP	PI-Cer(d36:0)*	down	+

PI-Cer(t36:0)*	down	+
SM(d39:2)*	down	+
SM(d41:2)*	down	+
SM(d43:1)*	down	+

FA: Fatty Acyls, GL: Glycerolipids, GP: Glycerophospholipids, SP: Shingolipids, *: lipid with p value <0.05, **: lipid with FDR corrected p value < 0.05.

4.2.1.2. CSF

In the case of CSF, PCA analyses shows that samples have neither a clear separation nor a tendency when the whole lipidome is used (Figure 22a). However, when only significant features were used (Figure 22b) a quite good separation was obtained for both positive and negative ionization mode. As it was expected PLS-DA algorithm allows a good separation between groups (Figure 23).

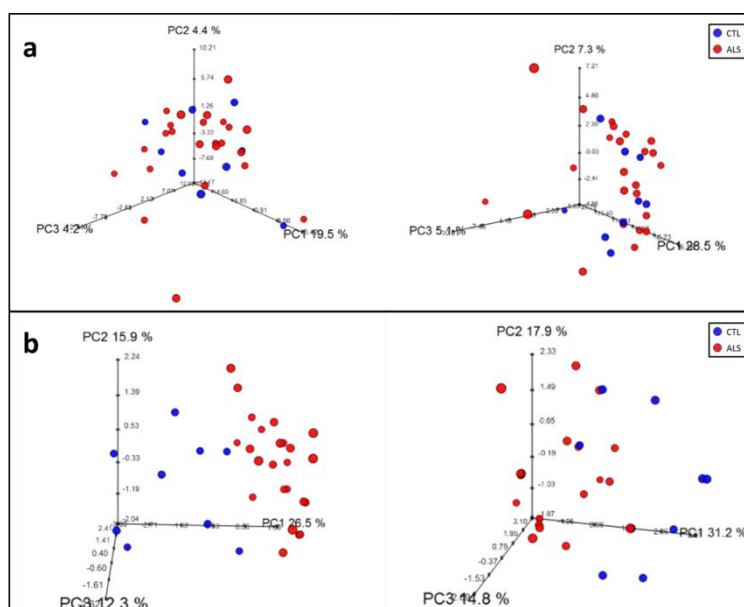


Figure 22. Principal component analyses between CSF samples from CTL and ALS groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from CTL group are represented with blue spheres, and individuals from ALS group with red spheres. Image **a**. PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. **b**. PCA of samples using only lipids with $p < 0.05$ in Wilcoxon-Mann-Whitney test applied to positive (left panel) and negative (right panel) ion detection modes data. In this case we can observe that a good discrimination is done.

As **Figure 24** shows, heatmaps using all features detected in CSF do not show a clear separation between groups. For top 25 molecules and significant ones, a better discrimination can be observed, with two clusters containing the most part of the samples from each group, but the clustering is not as perfect as plasma because some samples are not in the cluster of their group.

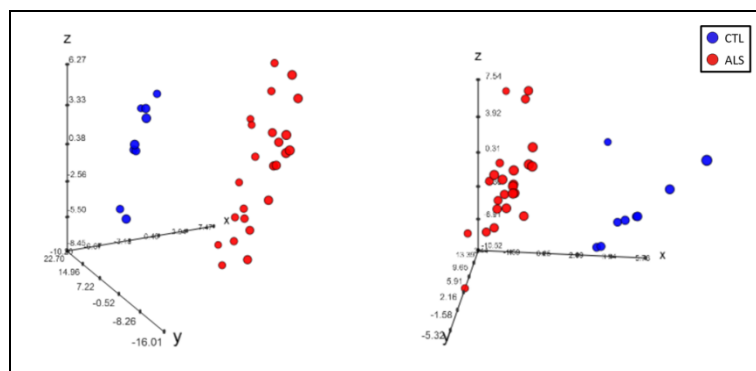


Figure 23. Partial Least Squares Discriminant analysis for CSF samples from CTL and ALS groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from CTL group are represented with blue spheres, and individuals from ALS group with red spheres.

The characteristics of lipid species which contribute to clustering both groups are reported in Table 9.

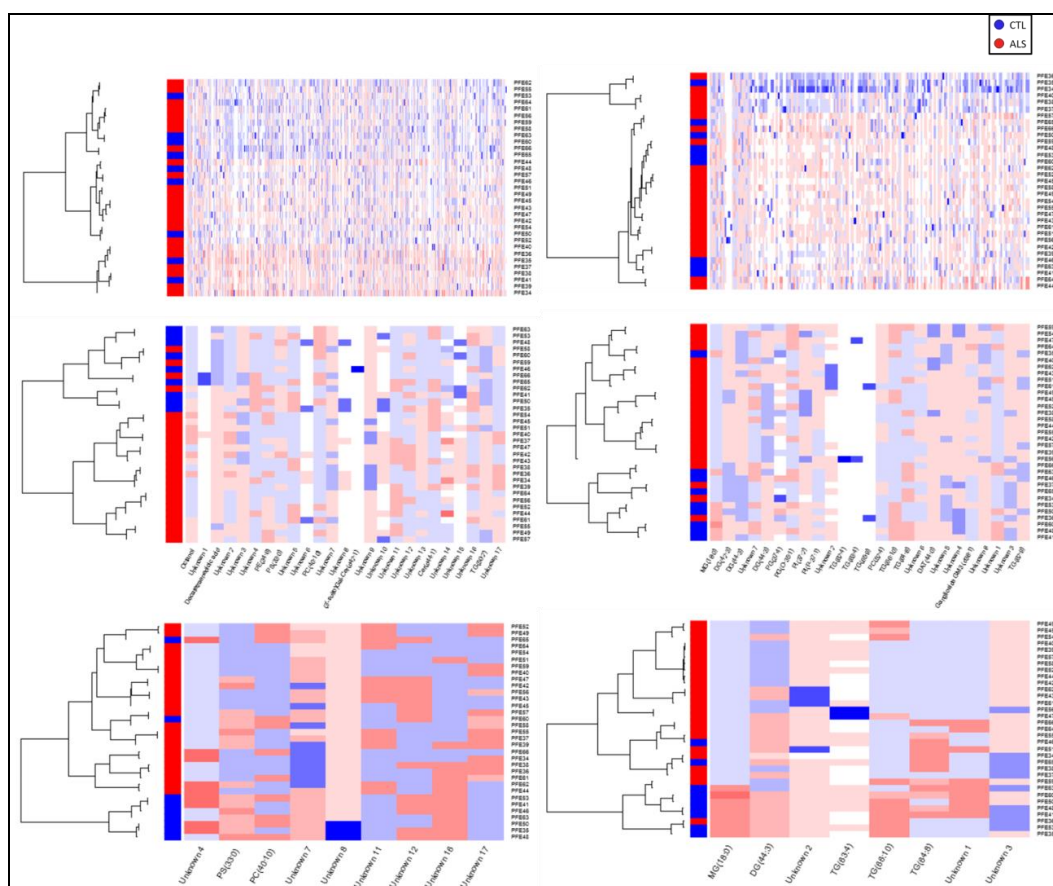


Figure 24. Hierarchical clusterings of CSF samples from CTL and ALS subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (CTL in blue and ALS in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Wilcoxon-Mann-Whitney test p value < 0.05) are displayed.

Table 9. CSF lipid species which contribute to discern between control (CTL) and amyotrophic lateral sclerosis (ALS) subjects.

Family	Metabolite	Regulation (CTL vs ALS)	Ionization mode
GL	DG(44:3)*	up	-
	MG(18:0)**	up	-
	TG(63:4)*	up	-
	TG(64:8)*	up	-
	TG(66:10)*	up	-
GP	PC(40:10)*	up	+
	PS(33:0)*	up	+

GL: Glycerolipids; GP: Glycerophospholipids; *: lipid with p value <0.05; **: lipid with FDR corrected p value < 0.05; +: positive ionization mode; -: negative ionization mode.

4.2.2. Onset (Bulbar vs Spinal vs Respiratory)

4.2.2.1. Plasma

When studying the samples according to the onset, PCA analyses (Figure 25) shows that there is not a good clusterization being the most homogeneous group is the bulbar one. Respiratory samples seem more disperse, and respiratory and spinal samples seem to show differences. When using the significant molecules we can see a good separation in positive samples, and not so good in negative. PLS-DA, however, shows a good clusterization between groups (Figure 26).

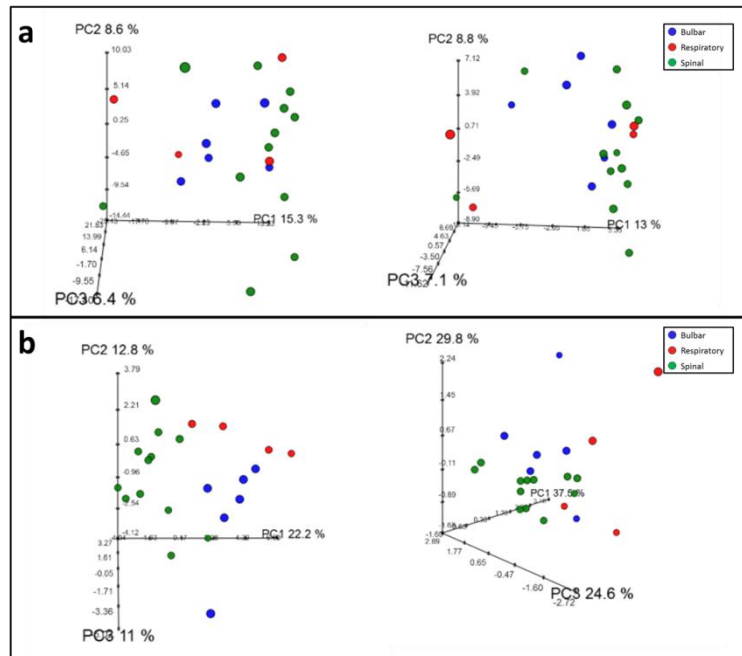


Figure 25. Principal component analyses between plasma samples from Bulbar, Spinal and Respiratory groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from Bulbar group are represented with blue spheres, individuals from Spinal group with green spheres and individuals from Respiratory group with red spheres. **a.** PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. **b.** PCA of samples using only lipids with $p < 0.05$ in Kruskal-Wallis test applied to positive (left panel) and negative (right panel) ion detection modes data.

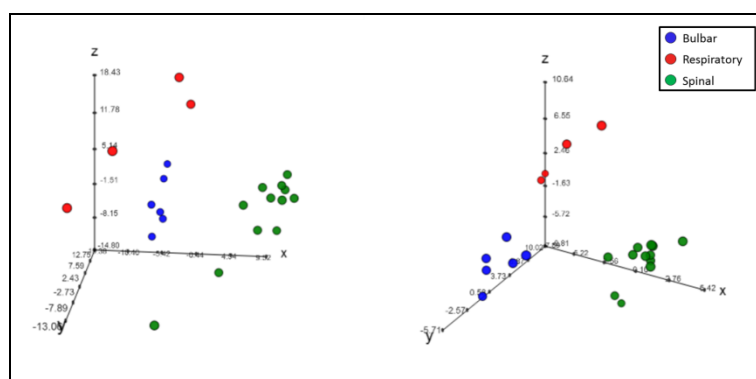


Figure 26. Partial Least Squares Discriminant analysis for plasma samples from Bulbar, Spinal and Respiratory groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from Bulbar group are represented with blue spheres, individuals from Spinal group with green spheres and individuals from Respiratory group with red spheres.

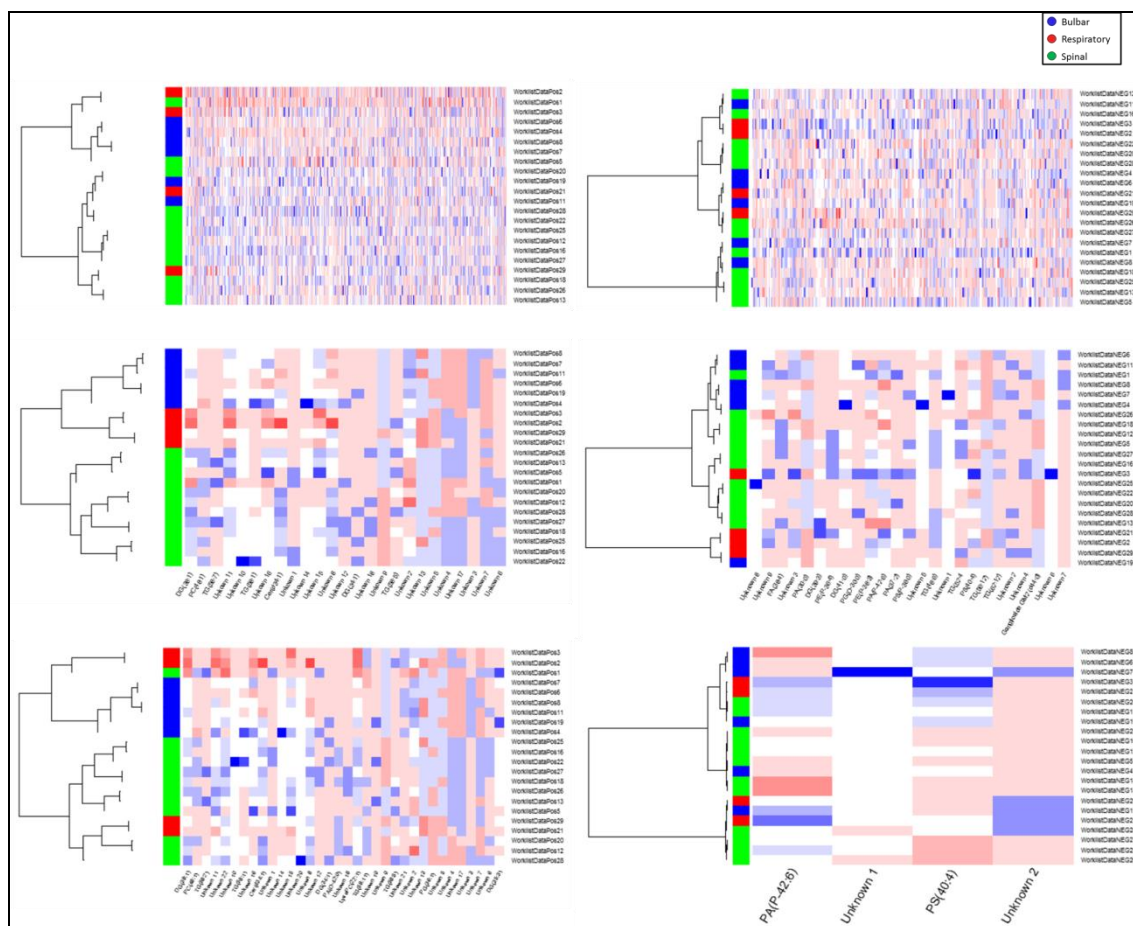


Figure 27. Hierarchical clusterings of plasma samples from Bulbar, Spinal and Respiratory onset subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (Bulbar in blue, Spinal in green and Respiratory in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization

mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Kruskal-Wallis test p value < 0.05) are displayed.

When observing the heatmaps with all the molecules, we can see the same tendency as in the PCA (in the case of positive ionization) (Figure 27). Specifically, two big clusters are formed, one containing mostly spinal samples and other containing mostly bulbar ones, while respiratory samples are disperse. For negative ionization this cannot be seen. When using the top 25 molecules, a perfect clustering can be seen for positive ionization. Furthermore, the two main clusters contain, the first one, only spinal group, and the second one bulbar and respiratory groups (and afterwards splits into a cluster for bulbar and another for respiratory). This is a really interesting fact because respiratory and bulbar onsets have similar survival time. Negative results show a tendency, but the separation is not as clear as in positive. Heatmaps with significant samples are not so good, and with negative ionization only 4 differential molecules were found, which is a too low number for describing the samples. The characteristics of lipid species which contribute to clustering the groups are reported in Table 10.

Table 10. Plasma lipid species which contribute to discern between disease onsets.

Family	Metabolite	Post-hoc	Ionization mode
FA	FA(38:4)		-
	DG(34:1)*	BULBAR - RESPIRATORY (down), RESPIRATORY - SPINAL (up)	+
	DG(38:1)*	BULBAR - SPINAL (up), RESPIRATORY - SPINAL (up)	+
	DG(39:3)		-
	DG(41:0)		-
	TG(48:6)		-
	TG(52:4)		-
	TG(56:12)		-
	TG(58:1)*	BULBAR - RESPIRATORY (down), RESPIRATORY - SPINAL (up)	+
	TG(58:5)*	BULBAR - RESPIRATORY (up) , BULBAR - SPINAL (up)	+
	TG(58:7)*	BULBAR - SPINAL (up), RESPIRATORY - SPINAL (up)	+
GL	TG(62:12)		-
	PA(30:0)		-
	PA(37:2)		-
	PA(P-42:6)*	BULBAR - RESPIRATORY (up) , RESPIRATORY - SPINAL (down)	-
	PC(48:1)*	BULBAR - SPINAL (up), RESPIRATORY - SPINAL (up)	+
	PE(P-36:4)		-
	PE(P-38:3)		-
	PG(O-30:0)		-
	PS(40:4)*	BULBAR - SPINAL (down) , RESPIRATORY - SPINAL (down)	-
GP	PS(P-36:0)		-
	Cer(d34:1)*	BULBAR - SPINAL (up), RESPIRATORY - SPINAL (up)	+
SP	Ganglioside GM2 (d44:0)		-

FA: Fatty Acyls, GL: Glycerolipids, GP: Glycerophospholipids, SP: Shingolipids, *: lipid with p value <0.05, **: lipid with FDR corrected p value < 0.05.

4.2.2.2. CSF

In CSF samples we can see that there are not so clear differences, and samples are very disperse (Figure 28a). However, when we use the significant molecules for the PCA we do obtain separation of the samples (Figure 28b) as well as when PLS-DA is applied (Figure 29).

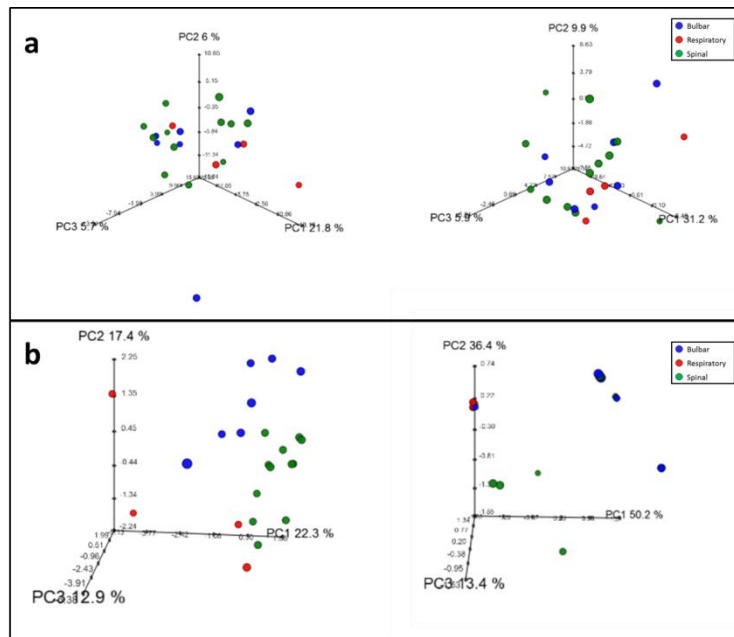


Figure 28. Principal component analyses between CSF samples from Bulbar, Spinal and Respiratory groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from Bulbar group are represented with blue spheres, individuals from Spinal group with green spheres and individuals from Respiratory group with red spheres. **a.** PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. **b.** PCA of samples using only lipids with $p < 0.05$ in Kruskal-Wallis test applied to positive (left panel) and negative (right panel) ion detection modes data.

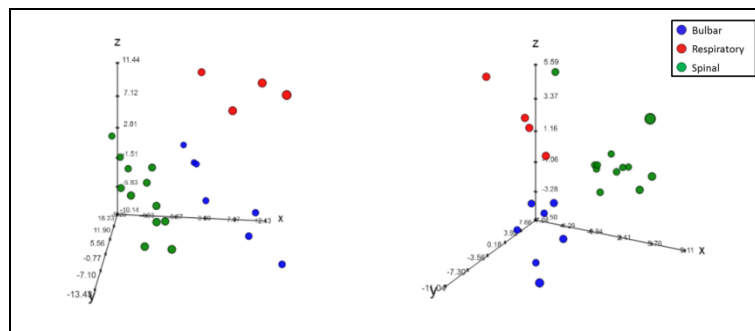


Figure 29. Partial Least Squares Discriminant analysis for CSF samples from Bulbar, Spinal and Respiratory groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from Bulbar group are represented with blue spheres, individuals from Spinal group with green spheres and individuals from Respiratory group with red spheres.

In heatmaps we see also a high dispersion, and none of them seems to be able to obtain a good clustering apart from the one with positive ionization and significant features (Figure 30). However, this clustering is not as good as the obtained for plasma.

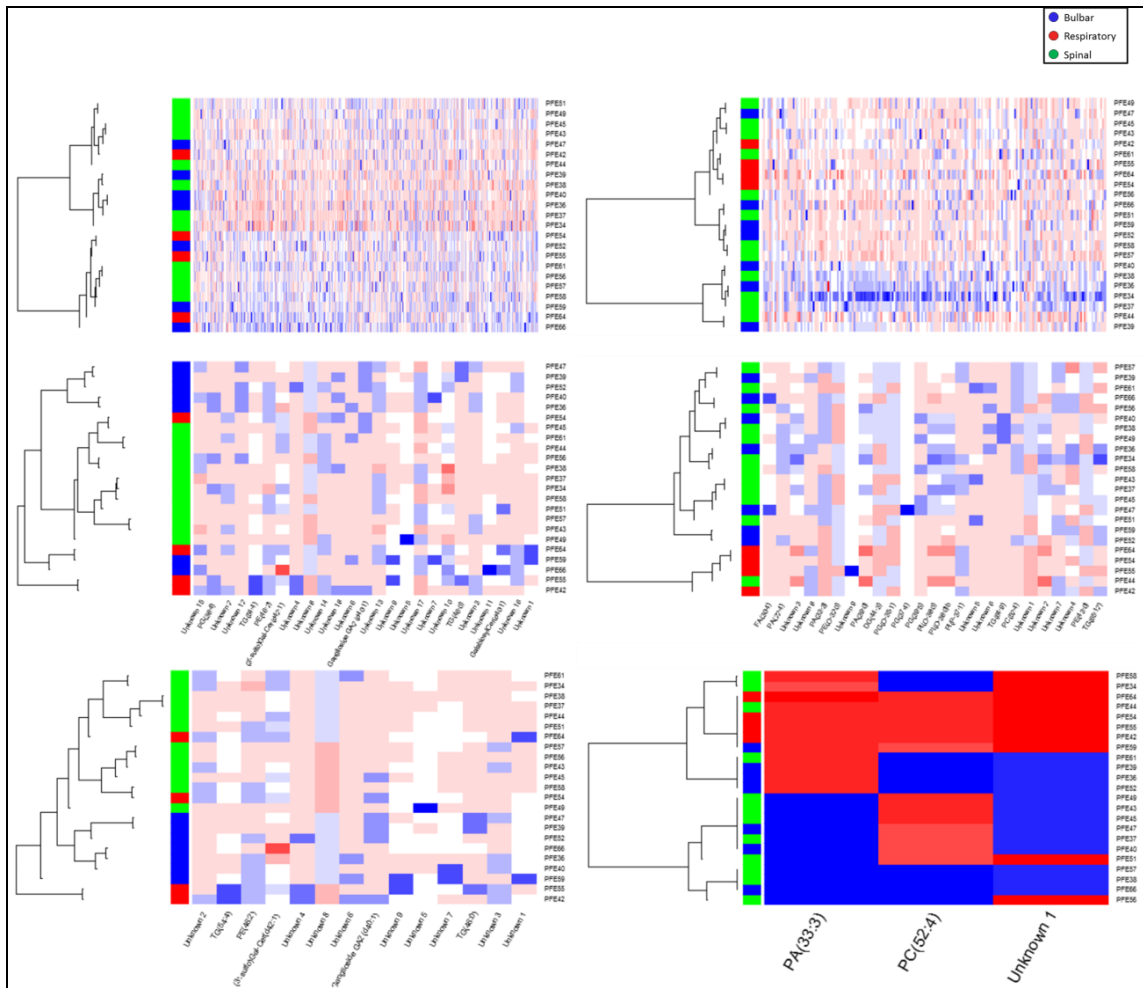


Figure 30. Hierarchical clusterings of CSF samples from Bulbar, Spinal and Respiratory onset subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (Bulbar in blue, Spinal in green and Respiratory in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Kruskal-Wallis test p value < 0.05) are displayed.

The characteristics of lipid species which contribute to clustering the groups are reported in Table 11.

Table 11. CFS lipid species which contribute to discern between disease onsets.

Family	Metabolite	Post-hoc	Ionization mode
FA	FA(30:4)		-
	DG(44:3)		-
	TG(46:0)*	BULBAR - SPINAL (down)	+
	TG(54:4)*	BULBAR - RESPIRATORY (up), RESPIRATORY - SPINAL (down)	+
	TG(64:9)		-
GL	TG(65:12)		-
GP	PA(22:4)		-

	PA(33:3)*	BULBAR - RESPIRATORY (down) , RESPIRATORY - SPINAL (up)	-
	PA(39:0)		-
	PC(52:4)*	BULBAR - RESPIRATORY (down) , RESPIRATORY - SPINAL (up)	-
	PE(43:0)		-
	PE(O-32:0)		-
	PE(46:2)*	RESPIRATORY - SPINAL (down)	+
	PG(37:4)		-
	PG(38:4)		+
	PG(39:5)		-
	PG(O-35:1)		-
	PI(O-38:0)		-
	PI(O-38:0)		-
	PI(P-37:1)		-
	(3'-sulfo)Gal-Cer(d42:1)*	BULBAR - RESPIRATORY (up) , BULBAR - SPINAL (up)	+
	GalabiosylCer(d40:1) Ganglioside		+
SP	GA2 (d40:1)*	BULBAR - SPINAL (down) , RESPIRATORY - SPINAL (down)	+

FA: Fatty Acyls, GL: Glycerolipids, GP: Glycerophospholipids, SP: Shingolipids, *: lipid with p value <0.05, **: lipid with FDR corrected p value < 0.05.

4.2.3. Gender

4.2.3.1. Plasma

In the case of gender comparison, we can see that there are not so much differences, but positive PCA graphs (Figure 31) seems to show a tendency to separation between groups.

. When significant molecules are used a better separation is seen. PLS-DA shows a good separation between groups (Figure 32).

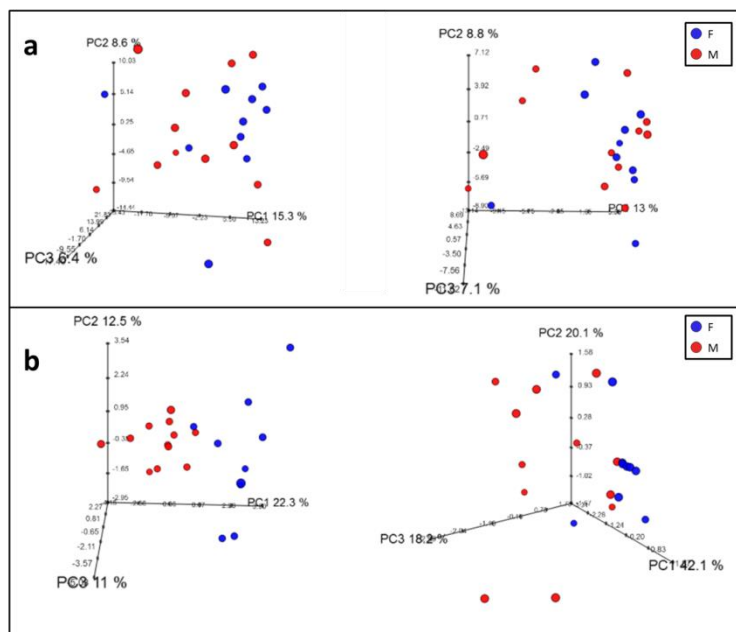


Figure 31. Principal component analyses between plasma samples from female (F) and male (M) groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from F group are represented with blue spheres, and individuals from M group with red spheres. **a.** PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. **b.** PCA of samples using only lipids with $p < 0.05$ in Wilcoxon-Mann-Whitney test applied to positive (left panel) and negative (right panel) ion detection modes data.

Heatmaps using whole lipid species detected do not reflect the tendency observed in the PCA as happened with the comparison ALS vs CTL (Figure 33). Although, when top 25 molecules and significant molecules of positive ionization analysis are used, a perfect clustering is obtained. Negative ionization top 25 molecules also show a pretty good clustering, but not perfect. For negative ionization, only few differential metabolites were obtained and samples cannot be characterized. The characteristics of lipid species which contribute to clustering the groups are reported in Table 12

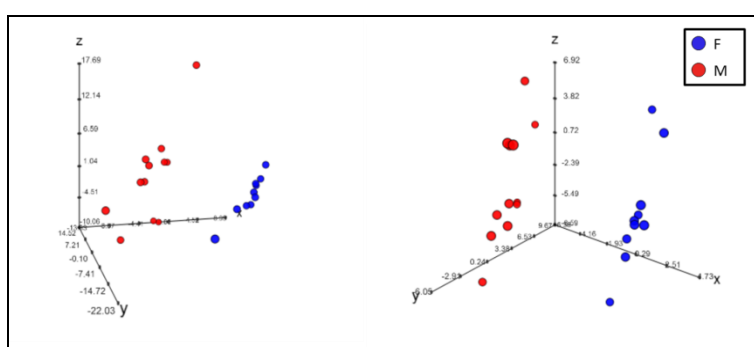


Figure 32. Partial Least Squares Discriminant analysis for plasma samples from female (F) and male (M) groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from F group are represented with blue spheres, and individuals from M group with red spheres.

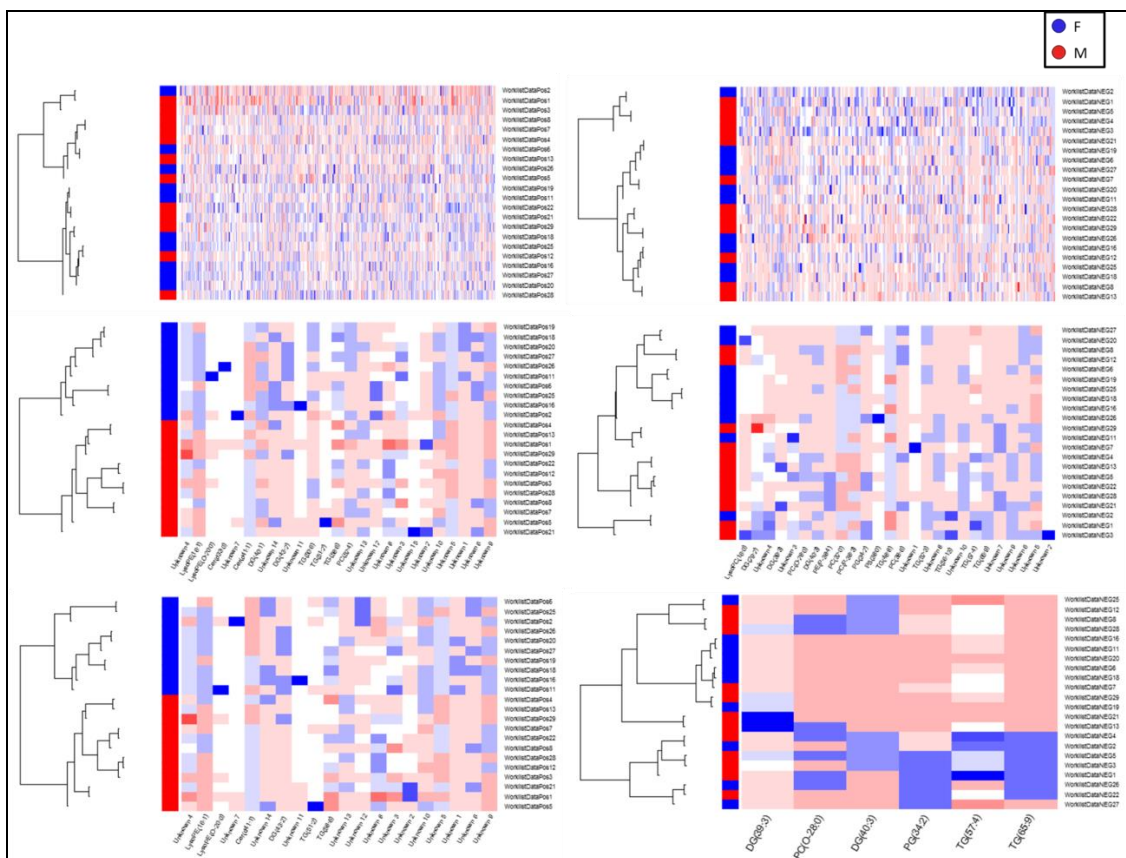


Figure 33. Hierarchical clusterings of plasma samples from female (F) and male (M) subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (F in blue and M in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Wilcoxon-Mann-Whitney test p value < 0.05) are displayed.

Table 12. Plasma lipid species which contribute to discern between gender.

Family	Metabolite	Regulation (CTL vs ALS)	Ionization mode
	DG(29:2)	up	-
	DG(39:3)*	up	-
	DG(40:1)	down	+
	DG(40:3)*	up	-
	DG(43:2)*	down	+
	TG(48:8)	up	-
	TG(50:6)	down	+
	TG(51:2)*	up	+
	TG(52:3)	down	-
	TG(56:13)	up	-
	TG(57:4)*	up	-
GL	TG(58:6)*	down	+

	TG(65:9)*	up	-
	LysoPC(16:0)	down	-
	LysoPE(16:1)*	down	+
	LysoPE(O-20:0)*	down	+
	PC(32:0)	down	-
	PC(38:6)	up	-
	PC(52:4)	down	+
	PC(O-28:0)*	up	-
	PC(P-38:3)	down	-
	PE(P-38:4)	up	-
	PG(34:2)*	up	-
GP	PS(38:0)	down	-
	Cer(d33:0)	down	+
SP	Cer(d41:1)*	up	+

GL: Glycerolipids, GP: Glycerophospholipids, SP: Shingolipids, *: lipid with p value <0.05, **: lipid with FDR corrected p value < 0.05.

4.2.3.2. CSF

For CSF, PCA shows no group separation except for positive ionization with significant molecules, where differences are seen but some samples are mixed (Figure 34a). In negative ionization analysis only 4 significant features were obtained, so PCA with significant molecules does not display differences (Figure 34b).

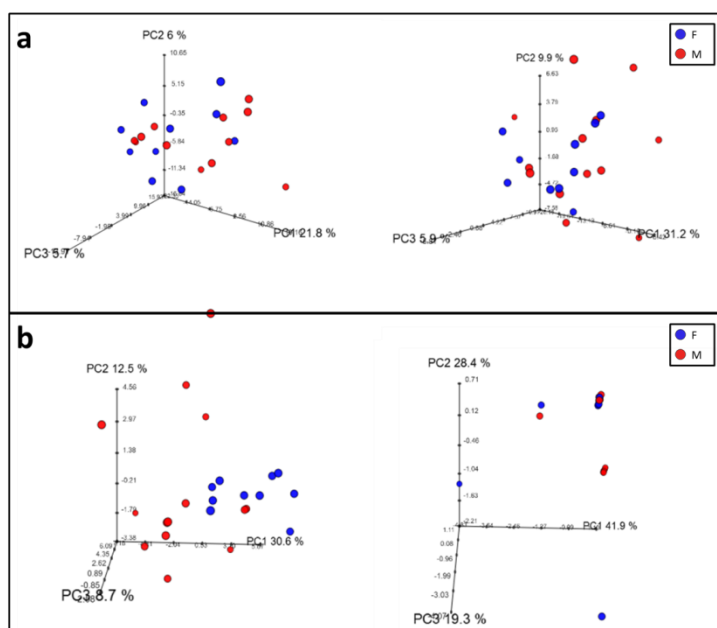


Figure 34. Principal component analyses between CSF samples from female (F) and male (M) groups. X axis: Principal component 1. Y axis: Principal component 2. Z axis: Principal component 3. Individuals from F group are represented with blue spheres, and individuals from M group with red spheres. **a.** PCA from samples using lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ion detection modes. **b.** PCA of samples using only lipids with $p < 0.05$ in Wilcoxon-Mann-Whitney test applied to positive (left panel) and negative (right panel) ion detection modes data.

PLS-DA also shows a good clusterization between groups (Figure 35).

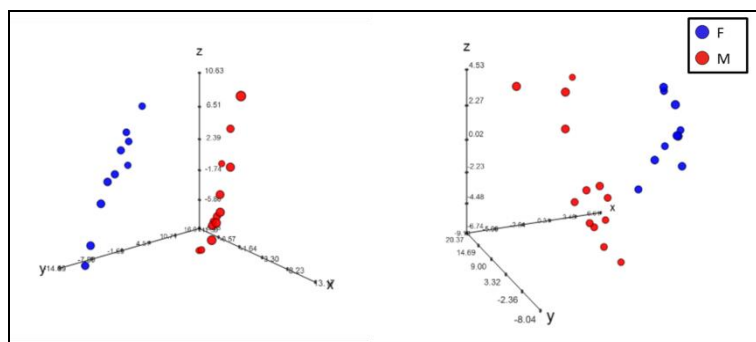


Figure 35. Partial Least Squares Discriminant analysis for CSF samples from female (F) and male (M) groups. Lipids from the whole lipidome obtained with positive (left panel) and negative (right panel) ionization modes are analysed. X axis: PLS component 1. Y axis: PLS component 2. Z axis: PLS component 3. Individuals from F group are represented with blue spheres, and individuals from M group with red spheres.

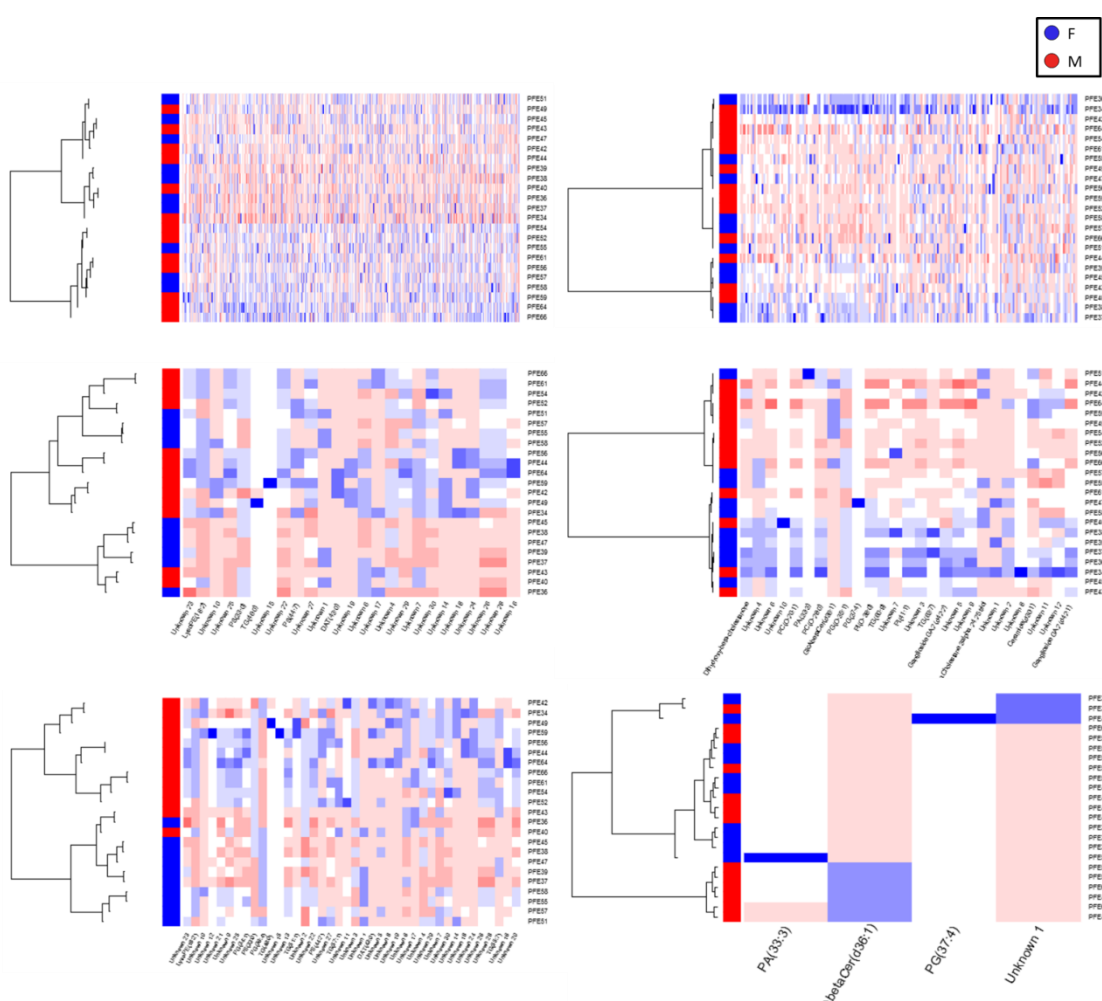


Figure 36. Hierarchical clusterings of CSF samples from female (F) and male (M) subjects. Each line of the graph represents a lipid specie colored by its abundance intensity, normalized to internal standard, log transformed and auto scaled. The scale from blue to red represents this normalized abundance in arbitrary units. Lipids are organized in columns. Samples are organized in rows and ordered according to the hierarchical clustering results. Dendrograms and sample grouping by color (F in blue and M in red) are also displayed. Left panels correspond to positive ionization mode analyses and right panels to negative ionization mode. From top to down, clusterings with all lipid species, clusterings with top 25 significant lipids with lower p value, and clusterings using significant lipids (Wilcoxon-Mann-Whitney test p value < 0.05) are displayed.

When heatmap analyses were applied, only positive significant molecules display an almost perfect clustering (Figure 36). The characteristics of lipid species which contribute to clustering the groups are reported in Table 13.

Table 13. CFS lipid species which contribute to discern between gender.

Family	Metabolite	Regulation (CTL vs ALS)	Ionization mode
	TG(46:0)*	up	+
	TG(51:1)*	up	+
	TG(55:7)*	up	+
	TG(60:7)*	up	+
GL	TG(57:1)*	up	+
	LysoPE(18:2)*	up	+
	PA(33:3)*	down	-
	PG(34:1)*	up	+
	PG(36:4)*	down	+
	PG(37:4)*	down	-
	PS(33:0)*	up	+
GP	PS(44:7)*	up	+
SP	GlcAbetaCer(d36:1)*	up	-

GL: Glycerolipids, GP: Glycerophospholipids, SP: Shingolipids, *: lipid with p value <0.05, **: lipid with FDR corrected p value < 0.05.

5. Discussion

5.1. Developed script

The objective of the present work was to develop an R-based data treatment pipeline to perform a lipidomic study in samples of biological origin. This has been benchmarked in samples from ALS patients, both plasma and CSF. As discussed below, despite it is often considered a single condition, under a clinical point of view ALS diagnosis can be better considered as a single label for many separate disease forms, with separate prognosis, disease evolution and potentially different therapeutic responses. For these reasons, there is a marked need for biomarker development, not only for ALS diagnosis but also for evaluation of prognosis. The developed pipeline has been applied to evaluate if the divergence in origin, clinical phenotype and disease progression of ALS pathogenesis, could also lead to lipidomic differences. As far as it has been explored, this is the first time that a lipidomic study in ALS patients is performed using CSF and plasma simultaneously.

Specific bioinformatics softwares are needed to perform the biostatistics analyses. Most of them are commercial softwares such as Mass Profiler Professional, (Agilent Technologies, Barcelona, Spain). Other open source softwares like Metaboanalyst[®] ⁶⁹ or XCMS⁷⁶ are the main alternative to these software. However, all of them share the same inconvenient: they cannot be modified. Thus, an R-based script has been created in order to supply the specific needs of the scientific community that the present softwares do not cover. The finality of the script is to be continuously updated and amplified according to lipidomic and metabolomics community requirements. For this reason, some of the Metaboanalyst[®] crucial functionalities used by researchers have been adapted and slightly modified in the script and some new ones have been added. Metaboanalyst[®] format for data reading has been conserved in order to be able to analyse previous works with the new script.

The main caveats of Metaboanalyst[®] at its present version that have been modified with the developed interactive script are the following:

1. In univariate analyses, Metaboanalyst[®] only allows to perform FDR p correction and only differential molecules with corrected FDR p value are obtained. The developed software allows choose to correct or not the p value and to perform different corrections.
2. Similarly, in univariate analyses, Kruskal-Wallis tests do not have implemented the post-hoc test. Thus, in the case of the study, post-hoc tests could not have been performed. Dunn post-hoc test has been implemented in the code.
3. Results of univariate analyses do not contain mass and retention time data of the metabolites. This is a major hurdle because in non-targeted schemes (such as liquid chromatography hyphenated with mass spectrometry) they are needed for further identification steps, allowing the annotation of the resulting molecules. With the developed script, mass and retention time data can be obtained if desired. This eases the indexing of molecules by mass and retention time avoiding the need to search and relate this data manually.

4. It was not possible to select which molecules are to be used for correlational analyses. In the presented script a global view of the correlations can be obtained with differential molecules. By implementing this, a global view of related regulation patterns of molecules that significantly differ between classes can be studied.
5. In correlation analyses, scatterplots of specific correlations was not possible to obtain. These scatterplots are interesting, in order to offer a simple view of the relation between pairs of variables for all the samples. The possibility to perform them has been introduced in the code.
6. Similarly to correlational analyses, it was not possible to select subset of data (molecules in this case) to obtain heatmap type graphs. The code has been improved, so now these analyses can be used to know if the significant features are able to describe the samples. Furthermore, more distance and clustering methods can be performed in order to search for the most suitable one.

Thus, although the software lacks other Metaboanalyst® functions, the most used ones by a metabolomics research group have been modified for their specific needs.

5.2. Analysis results

Once the software has been developed with the different functions required, it was applied to samples to perform several comparisons in order to help for searching novel mechanisms of ALS, by potential biomarker discovery. For the analyses, multivariate tests have been performed in order to explore the data, and have been combined with univariate tests in order to determine the specific molecules involved in the cases.

The results allow to observe a tendency for a separation in all the comparisons when using positive ionization mode: in the cases of CTL vs ALS and F vs M, groups did not clearly separate, indicating that the main variability of the samples is not determined by the group differential metabolites –as it often happens with samples of human origin-. Gender, physical condition, age, microbiota, time of day and period of time of sampling, drug or xenobiotic exposure, type of everyday activity are well known factors explaining metabolome and lipidome variations.⁷⁷ Even accounting these variability sources, the tendency observed suggest that these molecules, although they did not report the main variability and other factors could have an influence on this variability, had also an important influence in separating ALS from age and gender-matched individuals. Furthermore, in all cases, a separation between groups was observed when significant metabolites were used, indicating that the studied samples can be described with these metabolites. Similarly, in onset approach, bulbar and respiratory samples also tended to a fair separation when employing the proposed approach.

Even with the low n of individuals examined, in the case of CSF samples no separation or tendency is seen in any PCA with all the molecules for any comparison examined. Furthermore, PCAs with only significant metabolites displayed a good separation for positive analyses, but not that good for negative ones, mainly due to the reduced number of differential metabolites found between CSF negative comparisons.

Then, PLS-DAs were performed in order to determine if a model could help to predict classification by specific molecules. As commented on the introduction, in this case, with a

small samples/variables ratio, it was not expected to use the PLS-DA as a classifier and the only finality was to determine if there were potential molecules that could describe the groups. All PLS-DAs displayed a perfect separation between groups, with a very high accuracy. These analyses allow postulating, in further confirmatory cohort studies, for candidate biomarkers.

The last multivariate analysis performed has been a hierarchical clustering in order to determine if the samples followed a pattern of clustering according to their lipidomic profiles. For cluster determination, all distance and clustering methods allowed by the script have been used and the most suitable one for each case has been chosen. Mostly used algorithms have been Euclidean+Ward.D2 and Manhattan+Ward.D2. Less used ones have been Canberra+McQuitty and Canberra+Complete. As commented on the introduction, Ward.D2 is mainly used with Euclidean distances, but it has been demonstrated to behave well with non-Euclidean ones. The results showed a similar trend to what has been reported with the PCA approach.

For plasma samples, a good separation was found between groups in positive ionization (apart from the gender comparison). This clustering with all the features was not perfect, but having into account that all lipidic species were considered for the obtaining it and that not all the lipidome should be altered, the analyses offer a hopeful result. Furthermore, when using the top 25 significant molecules or the significant ones, a perfect clustering was found in all the cases (except for all significant molecules in onset comparison). In negative ionization the results were not that clear when using all the features, but the grouping was good for significant molecules.

In the case of CSF differences were less marked than those for plasma. Satisfactory clusterings were obtained when using the significant molecules of positive analyses, but not in negative ones or when using all the molecules.

Globally, multivariate results have shown a better classification analysing the samples in positive ion detection mode than in negative, and oppositely as expected, plasma samples have been able to describe the differences between groups. As ALS is a neurodegenerative disease, most differences were expected to be observed in CSF due to its higher interaction with the neurological system. However, due to the invasivity of the CSF extraction procedure compared to the plasma one, results obtained are more interesting in terms of facility of sample extraction.

Univariate analyses have been performed in order to define these specific lipids with capacity to describe the samples. Due to the low number of samples, non-parametric tests have been used: Wilcoxon-Mann Whitney for the two group comparisons (CTL vs ALS and gender) and Kruskal-Wallis test followed by a Dunn post-hoc test for the onset comparison.

As expected observing the multivariate analysis results, a higher number of differential molecules have been found in plasma, and more specifically in positive plasma samples. Only for positive CSF samples for gender comparison a higher number of significant metabolites can be found, but these are not able to describe the samples as well as the fewer plasma ones.

The main identified lipidic families that seem to have influence between the comparisons are cerolipids (GL) and Glycerophospholipids (GPL). Fatty Acyls (FA) and Sphingolipids (SL) also seem to be involved but in a lower degree. However, a high number of SL in plasma is reported allowing CTL vs ALS separation. Also a higher number of Fatty Acyls is seen, and a significant Fatty Acyl with p corrected is found: FA(13:1). Odd chain FA can be the result of uncomplete beta oxidation, or from less typical fatty acid oxidations (such as omega-oxidation or alpha-oxidation),⁷⁸ which have been often implicated as a mechanism of neurological consequences of vitamin inappropriate intake.⁷⁹ Whether this is the case in ALS will require further analyses, but it illustrates that even a single potential biomarker could open a pathogenic, and hopefully, a therapeutical window for this disease.

In this sense, recent advances in mass spectrometry techniques and bioinformatics tools have been critical in the development of “omic” sciences, improving the path to discovery of novel biomarkers and having a deep impact in biomedical research⁸⁰. Further, these developments pave the way for the advance of precision and personalized medicine. However, they require a hand in hand advance with mathematics, bioinformatics and statistical techniques, to help to differentiate the potential biomarkers from background noise, very abundant in samples from biological origin.

Lipidomics, as a branch of metabolomics, is one of the more novel “omic” techniques and it is focused on the identification and quantification of the large variation of lipid molecules in a biological sample. More specifically, plasma lipidome, which is one of with a biomedical importance, can be considered as the result of homeostatic systems that express cellular needs and specific physiological or pathological cell-tissue states.⁸¹ Interestingly, among the more than 4000 metabolites described in a human plasma metabolome, 80% can be defined as lipids,⁸² being therefore a very good sample to delineate, search and confirm physiological and disease’s biomarkers. In addition, 468 metabolites are also described in CSF, being also a good sample to study in biomarker research.⁸³ The importance of this sample is that it is considered a window to the closely protected central nervous system, in which tissue sample extraction is often hindered by ethical reasons.

Multivariate statistics showed that we are able to define the ALS disease using only about 25 lipid species present in plasma samples. Among these lipids we found species belonging to different families such as fatty acids (FA), glycerolipids (GL), sphingolipids (SL) and glycerophospholipids (PL). Globally, we found increased most of these species suggesting a disruption of lipid metabolism in ALS pathogenesis. Previous studies in SOD1 mice, an animal model of familiar ALS, describe changes in amount of PL, SL and TG in spinal cord samples.⁸⁴ Another work demonstrated increased levels of the FA docosahexaenoic acid in spinal cord samples.⁸⁵ These previous studies support our results suggesting that the alteration of lipid metabolism in the nervous system could be reflected in plasma lipidome.

Several researchers have discovered increased values of lipid concentrations in plasma, a result of the so-called hypermetabolic condition.^{53,86} Further, this has been often predicted as a protective condition. The presence of sphingolipids as differential molecules, often implicated in inflammatory pathways could offer interesting opportunities for pathophysiology research.

Concerning CSF analyses, we obtained the best clusterization using only the statistically significant lipids. However, the statistical differential lipids were only 9 in positive and 8 in negative ionization, representing a very small part of whole CSF lipidome. Among identified species we found 5 GL and 2 GP. Interestingly no SL were found differentially expressed in CSF whereas in plasma were the most numerous. As indicated above, SL have an important signalling function in human cells and are especially abundant in central nervous system, but can be also found in plasma. SL are deregulated in several neurological disorders, such as Alzheimer disease.⁶ In summary, according to our results plasma samples would be more appropriate for ALS biomarker searching and SL metabolism could have a crucial role in this pathology. The use of plasma samples instead of CSF is also advantageous for the ease of sample collection.

As it is explained in the introductory section, ALS phenotype is heterogeneous and different clinical and pathological subtypes are being discerned. Thus, once changes in lipid metabolism in ALS pathology were described we focus the analyses on ALS onset. In this case, and in the same line of ALS vs CTL experiment, plasma samples seems to better differentiate experimental groups. Specifically, it was possible to define an onset lipidomic signature using only the 25 lipid with the lower p-value. Among them, we found lipid belonging to FA, GL, GP and SP, but only 1 FA and 2 SP were found. According to these results, disturbance of SL metabolism seems to be a more specific trait of ALS pathogenesis, independently of the onset. As far as we know, this is the first study in which a specific lipidomic signature according to ALS onset is defined.

Finally, and because of the different incidence of ALS pathology between women and men we analysed the lipidomic difference between gender using only ALS patients. Lipidomic analyses revealed that we can better discern gender using plasma samples than CSF. Specifically, there is a specific signature in plasma of ALS according to gender mainly determined by GL and PL. GL account for a high proportion of total lipids in plasma and most of them are dependent on food intake. The differential GL found in our study cannot be attributed to food intake because all the samples were taken after overnight fasting period. Therefore, all the differences should be the result of tissues metabolism.

PL are the main components of cell membranes and are precursor of several signalling molecules implicated in different physiological processes. Moreover, PL and SL contribute to membrane lipid asymmetry.⁶

Globally, these results demonstrate the usefulness of the R script developed, which can be applied by scientific community working with any lipidome or metabolome derived sample in order to perform tailored univariate and multivariate statistic modelling.

6. Conclusions

6.1. Conclusions

1. The use of a handmade software allows to more suitably treat the data from high throughput screening techniques.
2. The use of a handmade software allows to perform a more precise analysis according to each study situation.
3. ALS pathology seems to affect more the lipidomic composition of plasma than CSF.
4. There is a specific lipidomic signature defining ALS pathology in plasma.
5. Sphingolipid metabolism seems to have an important role in the ALS pathology.
6. The different onsets of ALS pathology are more reflected in plasma than in CSF lipid composition.
7. Data analysis development is crucial to advance in biomarker search and in the discovery of new pharmacological targets to treat the diseases.
8. Lipidomic techniques, deeply supported by big data analysis, are useful to better understand pathogenic mechanisms of different diseases

6.2. Objective accomplishment

The general aim was to determine the presence of lipidomic profiles capable of characterize differential phenotypes and the progression of ALS disease.

The specific objectives were to stratify the patients according to age, gender, clinical phenotype and survival, to identify novel ALS biomarkers based on lipidomic profiles from plasma and CSF samples and to describe underlying mechanisms of ALS pathology.

All these objectives are accomplished in the work.

6.3. Work planning and methodology

No remarkable delays or changes have been needed. Planning has been followed in almost all cases, and methodology has been adequate for the work.

6.4. Work limitations

1. Developing a fully functional software requires almost exclusive dedication. New research lines and needs for new analyses keep appearing for every test that is performed. Thus, the software is not as extensive as desired.
2. Due to the rareness of the disease, the number of patient samples was small.
3. Due to the invasivity of CSF extraction, which is not usually performed in sane individuals, number of control samples was also small, and some of the samples were not from fully sane individuals.
4. Due to the devastating conditions of ALS, most of the patients were medicated and the metabolism could also be affected by this medication.
5. Due to external circumstances, most of the clinical data from the patients (apart from the onset and gender) had lots of missing values and correlation analyses between lipidomics data and biochemical or clinical data could not be performed.

6.5. Future lines of investigation

The future lines of investigation are the following ones:

7. To go further in the identification of unknown important compounds that determine the lipidomic signature.
8. To apply the correlation analysis script to the samples in order to know which lipid species of the whole plasma lipidome best correlate with clinical phenotype once missing data is collected.
9. Apply the correlation analysis script to the samples in order to know which lipid species of the whole CSF lipidome best correlate with clinical phenotype once missing data is collected.
10. Apply correlation analysis to compare plasma and CSF to determine which metabolites are correlated between them.
11. To perform pathway analyses and lipid enrichment analyses to study the specific pathways potentially involved in the disease.
12. A correlation analysis between days passed from the first symptom to the date of diagnosis with the lipidic species was performed (not included in the results). In plasma positive samples some strong correlations were found. Those metabolites were not identified, but it is also a future research line.

7. Bibliography

1. Beghi, E. *et al.* The epidemiology and treatment of ALS: focus on the heterogeneity of the disease and critical appraisal of therapeutic trials. *Amyotroph. Lateral Scler.* **12**, 1–10 (2011).
2. Al-Chalabi, A. *et al.* Genetic and epigenetic studies of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **14 Suppl 1**, 44–52 (2013).
3. Al-Chalabi, A. & Hardiman, O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat. Rev. Neurol.* **9**, 617–28 (2013).
4. Ajroud-Driss, S. & Siddique, T. Sporadic and hereditary amyotrophic lateral sclerosis (ALS). *Biochim. Biophys. Acta - Mol. Basis Dis.* **1852**, 679–684 (2015).
5. Trojsi, F., Monsurrò, M. & Tedeschi, G. Exposure to Environmental Toxicants and Pathogenesis of Amyotrophic Lateral Sclerosis: State of the Art and Research Perspectives. *Int. J. Mol. Sci.* **14**, 15286–15311 (2013).
6. Quehenberger, O. & Dennis, E. A. The human plasma lipidome. *N. Engl. J. Med.* **365**, 1812–23 (2011).
7. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–9 (2012).
8. Naudí, A. *et al.* Lipidomics of human brain aging and Alzheimer’s disease pathology. *Int. Rev. Neurobiol.* **122**, 133–89 (2015).
9. Jové, M., Portero-Otín, M., Naudí, A., Ferrer, I. & Pamplona, R. Metabolomics of human brain aging and age-related neurodegenerative diseases. *J. Neuropathol. Exp. Neurol.* **73**, 640–57 (2014).
10. Spratlin, J. L., Serkova, N. J. & Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* **15**, 431–40 (2009).
11. Ortiz, A. & Sanchez-Niño, M. D. The human plasma lipidome. *N. Engl. J. Med.* **366**, 668; author reply 668-9 (2012).
12. Chen, Y., Xu, J., Zhang, R. & Abliz, Z. Methods used to increase the comprehensive coverage of urinary and plasma metabolomes by MS. *Bioanalysis* **8**, 981–97 (2016).
13. Trygg, J., Holmes, E. & Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **6**, 469–479 (2007).
14. Lindon, J. C., Holmes, E., Bollard, M. E., Stanley, E. G. & Nicholson, J. K. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **9**, 1–31 (2004).
15. Jové, M. *et al.* Lipidomic and metabolomic analyses reveal potential plasma biomarkers of early atheromatous plaque formation in hamsters. *Cardiovasc. Res.* **97**, 642–52 (2013).
16. Putluri, N. *et al.* Metabolomic profiling reveals potential markers and bioprocesses altered in bladder cancer progression. *Cancer Res.* **71**, 7376–86 (2011).

17. Jove, M. *et al.* Metabolomics predicts stroke recurrence after transient ischemic attack. *Neurology* **84**, 36–45 (2015).
18. Brereton, R. G. & Lloyd, G. R. Partial least squares discriminant analysis: Taking the magic away. *J. Chemom.* **28**, 213–225 (2014).
19. Greenacre, M. Hierarchical cluster analysis. *Corresp. Anal. Relat. Methods* **2**, 11 (2008).
20. <https://stats.stackexchange.com/questions/195446/choosing-the-right-linkage-method-for-hierarchical-clustering>.
21. Minitab Inc. <http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/item-and-cluster-analyses/linkage-methods/>. (2016).
22. Batagelj, V. Generalized Ward and Related Clustering Problems Ward. *Classif. Relat. Methods Data Anal.* 67–74 (1988).
23. Leeper, J. D. & UCLA: Statistical Consulting Group. Choosing the Correct Statistical Test in SAS, Stata, SPSS and R. Available at: <https://stats.idre.ucla.edu/other/mult-pkg/whatstat/>.
24. Frost, J. <http://blog.minitab.com/blog/adventures-in-statistics-2/choosing-between-a-nonparametric-test-and-a-parametric-test>. (2015).
25. Noble, W. S. How does multiple testing correction work? *Nat. Biotechnol.* **27**, 1135–7 (2009).
26. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368–75 (2003).
27. Brooks, B. R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J. Neurol. Sci.* **124 Suppl**, 96–107 (1994).
28. Logroscino, G. *et al.* Incidence of amyotrophic lateral sclerosis in Europe. *J. Neurol. Neurosurg. Psychiatry* **81**, 385–390 (2010).
29. Marin, B., Gil, J., Preux, P. M., Funalot, B. & Couratier, P. Incidence of amyotrophic lateral sclerosis in the Limousin region of France, 1997–2007. *Amyotroph. Lateral Scler.* **10**, 216–220 (2009).
30. Ragonese, P. *et al.* Accuracy of death certificates for amyotrophic lateral sclerosis varies significantly from north to south of Italy: implications for mortality studies. *Neuroepidemiology* **23**, 73–7 (2004).
31. Sorenson, E. J., Stalker, A. P., Kurland, L. T. & Windebank, A. J. Amyotrophic lateral sclerosis in Olmsted County, Minnesota, 1925 to 1998. *Neurology* **59**, 280–2 (2002).
32. Gordon, P. H. *et al.* Incidence of amyotrophic lateral sclerosis among American Indians and Alaska natives. *JAMA Neurol.* **70**, 476–80 (2013).
33. Alonso, A., Logroscino, G., Jick, S. S. & Hernán, M. A. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. *Eur. J. Neurol.* **16**, 745–51 (2009).

34. Ludolph, A. C., Brettschneider, J. & Weishaupt, J. H. Amyotrophic lateral sclerosis. *Curr. Opin. Neurol.* **25**, 530–535 (2012).
35. Shellikeri, S. *et al.* The neuropathological signature of bulbar-onset ALS: A systematic review. *Neurosci. Biobehav. Rev.* **75**, 378–392 (2017).
36. Gautier, G. *et al.* ALS with respiratory onset: Clinical features and effects of non-invasive ventilation on the prognosis. *Amyotroph. Lateral Scler.* **11**, 379–382 (2010).
37. Haverkamp, L. J., Appel, V. & Appel, S. H. Natural history of amyotrophic lateral sclerosis in a database population. Validation of a scoring system and a model for survival prediction. *Brain* 707–19 (1995).
38. Armon, C. & Moses, D. Linear estimates of rates of disease progression as predictors of survival in patients with ALS entering clinical trials. *J. Neurol. Sci.* S37-41 (1998).
39. Goldstein, L., Atkins, L. & Leigh, P. Correlates of Quality of Life in people with motor neuron disease (MND). *Amyotroph. Lateral Scler. Other Mot. Neuron Disord.* **3**, 123–129 (2002).
40. Mitsumoto, H. & Del Bene, M. Improving the quality of life for people with ALS: the challenge ahead. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **1**, 329–36 (2000).
41. Chio, A. *et al.* Lower serum lipid levels are related to respiratory impairment in patients with ALS. *Neurology* **73**, 1681–1685 (2009).
42. Shoesmith, C. L., Findlater, K., Rowe, A. & Strong, M. J. Prognosis of amyotrophic lateral sclerosis with respiratory onset. *J. Neurol. Neurosurg. Psychiatry* **78**, 629–31 (2007).
43. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. *Lancet* **377**, 942–955 (2011).
44. Beghi, E. *et al.* The epidemiology and treatment of ALS: Focus on the heterogeneity of the disease and critical appraisal of therapeutic trials. *Amyotroph. Lateral Scler.* **12**, 1–10 (2011).
45. Leigh, P. N. *et al.* Ubiquitin deposits in anterior horn cells in motor neurone disease. *Neurosci. Lett.* **93**, 197–203 (1988).
46. Neumann, M. *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science (80-.)*. **314**, 130–133 (2006).
47. Saberi, S., Stauffer, J. E., Schulte, D. J. & Ravits, J. Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. *Neurol. Clin.* **33**, 855–76 (2015).
48. McGoldrick, P., Joyce, P. I., Fisher, E. M. C. & Greensmith, L. Rodent models of amyotrophic lateral sclerosis. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1832**, 1421–1436 (2013).
49. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Commun.* **351**, 602–611 (2006).
50. Mackenzie, I. R. A. *et al.* Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann. Neurol.* **61**, 427–34 (2007).

51. Maekawa, S. *et al.* TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. *Neuropathology* **29**, 672–83 (2009).
52. Scotter, E. L., Chen, H.-J. & Shaw, C. E. TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. *Neurotherapeutics* **12**, 352–363 (2015).
53. Dupuis, L. *et al.* Dyslipidemia is a protective factor in amyotrophic lateral sclerosis. *Neurology* **70**, 1004–1009 (2008).
54. Palamiuc, L. *et al.* A metabolic switch toward lipid use in glycolytic muscle is an early pathologic event in a mouse model of amyotrophic lateral sclerosis. *EMBO Mol. Med.* **7**, 526–46 (2015).
55. Ilieva, E. V *et al.* Oxidative and endoplasmic reticulum stress interplay in sporadic amyotrophic lateral sclerosis. *Brain* **130**, 3111–23 (2007).
56. Cutler, R. G., Pedersen, W. A., Camandola, S., Rothstein, J. D. & Mattson, M. P. Evidence that accumulation of ceramides and cholesterol esters mediates oxidative stress-induced death of motor neurons in amyotrophic lateral sclerosis. *Ann. Neurol.* **52**, 448–57 (2002).
57. Rutter-Locher, Z., Turner, M. R., Leigh, P. N. & Al-Chalabi, A. Analysis of terms used for the diagnosis and classification of amyotrophic lateral sclerosis and motor neuron disease. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **17**, 600–604 (2016).
58. Brooks, B. R., Miller, R. G., Swash, M., Munsat, T. L. & World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **1**, 293–9 (2000).
59. de Carvalho, M. *et al.* Electrodiagnostic criteria for diagnosis of ALS. *Clin. Neurophysiol.* **119**, 497–503 (2008).
60. Carvalho, M. De & Swash, M. Awaji diagnostic algorithm increases sensitivity of El Escorial criteria for ALS diagnosis. *Amyotroph. Lateral Scler.* **10**, 53–57 (2009).
61. Lenglet, T. & Camdessanché, J.-P. Amyotrophic lateral sclerosis or not: Keys for the diagnosis. *Rev. Neurol. (Paris)*. **173**, 280–287 (2017).
62. Al-Chalabi, A. *et al.* Amyotrophic lateral sclerosis: moving towards a new classification system. *Lancet Neurol.* **15**, 1182–1194 (2016).
63. D. Fix, J. *Neuroanatomy*. (Lippincott Williams & Wilkins, 2007).
64. Pizarro, C., Arenzana-Rámila, I., Pérez-del-Notario, N., Pérez-Matute, P. & González-Sáiz, J.-M. Plasma Lipidomic Profiling Method Based on Ultrasound Extraction and Liquid Chromatography Mass Spectrometry. *Anal. Chem.* **85**, 12085–12092 (2013).
65. Castro-Perez, J. M. *et al.* Comprehensive LC-MS E lipidomic analysis using a shotgun approach and its application to biomarker detection and identification in osteoarthritis patients. *J. Proteome Res.* **9**, 2377–89 (2010).
66. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–51 (2005).

67. Fahy, E., Sud, M., Cotter, D. & Subramaniam, S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* **35**, W606–W612 (2007).
68. Koelmel, J. P. *et al.* LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics* (2016).
69. Xia, J. & Wishart, D. S. in *Current Protocols in Bioinformatics* **55**, 14.10.1–14.10.91 (John Wiley & Sons, Inc., 2016).
70. Stack Exchange Inc. stackoverflow.com. (2017).
71. GitHub. github.com. (2017).
72. RStudio Inc. shiny.rstudio.com. (2016).
73. cran.r-project.org.
74. <http://rbloggers.com/>. (2017).
75. Stack Exchange Inc. stats.stackexchange.com. (2017).
76. Huan, T. *et al.* Systems biology guided by XCMS Online metabolomics. *Nat. Methods* **14**, 461–462 (2017).
77. Krug, S. *et al.* The dynamic range of the human metabolome revealed by challenges. *FASEB J.* **26**, 2607–19 (2012).
78. Spiekerkoetter, U. *et al.* Current issues regarding treatment of mitochondrial fatty acid oxidation disorders. *J. Inherit. Metab. Dis.* **33**, 555–61 (2010).
79. Metz, J. Cobalamin deficiency and the pathogenesis of nervous system disease. *Annu. Rev. Nutr.* **12**, 59–79 (1992).
80. Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **88**, 524–45 (2016).
81. Jové, M. *et al.* A Stress-Resistant Lipidomic Signature Confers Extreme Longevity to Humans. *J. Gerontol. A. Biol. Sci. Med. Sci.* **72**, 30–37 (2017).
82. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35**, D521–6 (2007).
83. Wishart, D. S. *et al.* The human cerebrospinal fluid metabolome. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **871**, 164–73 (2008).
84. Henriques, A. *et al.* Amyotrophic lateral sclerosis and denervation alter sphingolipids and up-regulate glucosylceramide synthase. *Hum. Mol. Genet.* **24**, 7390–405 (2015).
85. Cacabelos, D. *et al.* Interplay between TDP-43 and docosahexaenoic acid-related processes in amyotrophic lateral sclerosis. *Neurobiol. Dis.* **88**, 148–60 (2016).
86. Dorst, J. *et al.* Patients with elevated triglyceride and cholesterol serum levels have a prolonged survival in amyotrophic lateral sclerosis. *J. Neurol.* **258**, 613–7 (2011).

8. Annexes

8.1. Annex 1: Interactive shiny markdown script

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, cache=F)
library(knitr)
library(gWidgets)
library(tcltk)
library(htmlwidgets)
```

```{r , message=FALSE, warning=FALSE}
nomcsv<-file.choose()
options(guiToolkit="tcltk")
filescol<-ginput(guiToolkit(),"Choose if samples are organized in rows (1) o in columns (2)",
title="rows/columns", icon="question")
#Si están organizadas en columnas, lo cambiaremos a organización a filas
#De este modo podremos utilizar un mismo código para los análisis.
if (filescol=="2"){
 #Hacemos la matriz transpuesta
 data0<-t(read.csv(nomcsv))
 #Lo pasamos a data frame
 rawdata<-as.data.frame(data0)
 #En hacer la transpuesta no detecta los nombres de los metabolitos como encabezado, así
 que lo indicamos
 names(rawdata)<- data0[1,]
 #Eliminamos la fila que contiene los nombres de los metabolitos
 rawdata<-rawdata[-1,]
 #Eliminamos el archivo data 0
 rm(data0)
 #Se nos ha detectado las variables como factores, así que las transformamos a numéricas
 for (x in 2:ncol(rawdata)){
 rawdata[,x]<-as.numeric(as.character(rawdata[,x]))
 }
 #Si el archivo ya se encuentra organizado por filas hay que hacer menos modificaciones
} else if (filescol == "1"){
 rawdata<-read.csv(nomcsv)
 row.names(rawdata)<-rawdata[,1]
 rawdata<-rawdata[,-1]
} else {message("Error: especificar correctament si l'organització és en files o columnes")}
#Para evitar errores de formato, eliminamos espacios de los grupos y los transformamos en
mayúsculas
rawdata[,1]<-gsub(" ", "", rawdata[,1])
rawdata[,1]<-toupper(rawdata[,1])
#Eliminamos variables constantes (indicamos que si el mínimo es igual al máximo se elimine la
fila)
rawdata<-rawdata[,!apply(rawdata[-c(1,2),], MARGIN = 2, function(x) max(x, na.rm = TRUE) ==
min(x, na.rm = TRUE))]
```

```
massrt<-gconfirm(guiToolkit(),"Are mass and retention time specified?", title="mass and RT",
icon="question")
```

```
if(massrt==TRUE){
 mass<-t(rawdata[1,])
 mass<-mass[-1]
 rt<-t(rawdata[2,])
 rt<-rt[-1]
 rawdata<-rawdata[-c(1,2),]
}else{
 mass<-rep(NA, ncol(rawdata)-1)
 rt<-rep(NA, ncol(rawdata)-1)
}
```

```
#Actualizamos nombres de los grupos (aunque los hayamos unificado, si no hacemos este paso
se detectan los grupos anteriores)
```

```
rawdata[,1]<-factor(rawdata[,1])
rm(filescol, nomcsv)
col=c("blue", "red", "green", "gray", "black", "purple", "coral", "yellow")
col= col[1:length(levels(rawdata[,1]))]
groupcol<-col[(as.numeric(rawdata[,1]))]
```

```
norm<-gconfirm(guiToolkit(),"Is data normalization (log transformation + autoscaling)
desired?", title="normalization", icon="question")
```

```
if(norm==TRUE){
 data<-cbind(rawdata[,1],log(rawdata[,1], base=2))
 data<-as.data.frame(cbind(data[,1],scale(data[,1])))
}else{
 data<-rawdata
}
...

```

```
#Univariate analysis
```

```
...{r}
par<-ginput(guiToolkit(),"Choose the type of test to be performed: non-parametric
(Wilcoxon/Kruskal-Wallis) (1) or parametric (t-test/ANOVA) (2)", title="non-parametric /
parametric", icon="question")
...

```

```
...{r, message=FALSE, results="hide", warning=FALSE}
#Si tenemos 1 o 2 grupos
if(length(levels(data[,1]))<3){
 #Si queremos análisis paramétrico aplicaremos el test por t de student
 if (par=="2"){
 #Aplicaremos la función t test a lo largo de todo nuestros datos
 univar<-as.data.frame(cbind(t(apply(data[,1], 2, function(x)
 #Obtendremos los datos indicados:
 unlist(t.test(x~data[,1])[c("estimate", "p.value", "statistic", "conf.int")])))))
 #Renombramos los datos
 names(univar)<-c(paste("mean for group", levels(data[,1])[1]),paste("mean for group",
 levels(data[,1])[2]), "pval", "statistic.t", "conf.int1", "conf.int2")
 }
}
```

```

#Crearemos una columna con la regulación (según si el promedio del grupo 2 es mayor o
menor, ésta será up o down)
univar$regulation<-c("down", "up")[as.numeric((univar[,1]-univar[,2])>0)+1]
names(univar)[names(univar)=="regulation"]<-
paste("regulation (", levels(data[,1])[1], " vs " ,levels(data[,1])[2],")", sep="")

#Si queremos no paramétrico aplicaremos el test de Mann-Whitney análogamente al
apartado anterior
} else if (par=="1"){
univar<-as.data.frame(cbind(t(apply(data[-1], 2, function(x)
 unlist(wilcox.test(x~data[,1], conf.int=T)[c("estimate", "p.value", "statistic", "conf.int")])))))
names(univar)<-c("estimate.dif", "pval", "statistic.w", "conf.int.1", "conf.int.2")
univar$regulation<-c("down", "up")[as.numeric((univar$estimate.dif)>0)+1]
names(univar)[names(univar)=="regulation"]<-
paste("regulation (", levels(data[,1])[1], " vs " ,levels(data[,1])[2],")", sep="")
}
#Crearemos otra nueva columna con la p ajustada por False Discovery Rate
univar$mass<-mass
univar$RT<-rt
#También creamos una lista únicamente con los metabolitos significativos y otra con los
datos enteros
univarsign<-univar[univar$pval<0.05,]
univarsign<-univarsign[order(univarsign$pval),]
datasign<-as.data.frame(cbind(data[,1],data[colnames(data) %in% rownames(univarsign)]))
#Si tenemos 3 o más grupos realizaremos un ANOVA para el análisis paramétrico o Kruskal-
Wallis si queremos no paramétrico
} else {
if(par==2){
univar<-as.data.frame(cbind(t(apply(data[-1], 2, function(x)
 unlist(summary(aov(x~data[,1])))[c("F value1", "Pr(>F)1")]))))
names(univar)<-c("F value1", "pval")
#Crearemos una nueva columna con la p ajustada por False Discovery Rate

#También creamos una lista únicamente con los metabolitos significativos y otra con los
datos enteros
#A demás, añadiremos una columna con el post-hoc por Tukey.
#Para ello, primero definimos un vector con todas las combinaciones posibles de grupos
combinacions<-combn(levels(datasign[,1]), 2)
#Seguidamente, obtenemos el listado con los valores de p para cada comparación, para los
metabolitos significativos.
tukey<-as.data.frame(cbind(t(apply(data[-1], 2, function(x)
 #Estos valores de p se encuentran en el intervalo indicado,el cual depende del número de
comparaciones posibles
 unlist(TukeyHSD(x=aov(x~data[,1])))[(3*ncol(combinacions)+1):(4*ncol(combinacions))]))))
#Seguidamente, creamos los datos que recogen todas las comparaciones
posthoc<-sapply(1:ncol(combinacions), function(x) c("", paste(combinacions[1,x],"-",
combinacions[2,x], " ,"))[as.numeric((tukey[,x])<0.05)+1])
univar$posthoc<-apply(posthoc, 1, function(x) paste(x, collapse=" "))
univar$mass<-mass
univar$RT<-rt
univarsign<-univar[univar$pval<0.05,]
datasign<-as.data.frame(cbind(data[,1],data[colnames(data) %in% rownames(univarsign)]))

```

```

} else if (par==1){
 univar<-as.data.frame(cbind(t(apply(data[-1], 2, function(x)
 unlist(kruskal.test(x~data[,1]))[-c(2,4,5)]))))
 names(univar)<-c("statistic", "pval")
 univar$pval<-as.numeric(as.character(univar$pval))

 library("dunn.test")
 combinacions<-combn(levels(data[,1]), 2)
 dunn<-as.data.frame(cbind(t(apply(data[-1], 2, function(x)
 unlist(dunn.test(x, data[,1]))[(2*ncol(combinacions)+2):(3*ncol(combinacions)+1)]))))
 posthoc<-sapply(1:ncol(combinacions), function(x) c("", paste(combinacions[1,x],"-",
combinacions[2,x], ","))[as.numeric(as.numeric(as.character(dunn[,x]))<0.05)+1])
 univar$posthoc<-apply(posthoc, 1, function(x) paste(x, collapse=" "))
 univar$mass<-mass
 univar$RT<-rt
 rm(posthoc, dunn, combinacions)
 univarsign<-univar[univar$pval<0.05,]
 datasign<-as.data.frame(cbind(data[,1],data[colnames(data) %in% rownames(univarsign)]))
} else {univar<-NA}
}
...

``{r , echo=FALSE}
library(knitr)
sliderInput("pumbr", "Threshold p value", min = 0, max = 0.2, value = 0.05)

checkboxInput(inputId = "p_corr",
 label = strong("P corrected graph"),
 value = FALSE)
selectInput("corr",
 label = "P correction method",
 choices = c("holm", "hochberg", "hommel", "bonferroni", "BH",
 "BY"),
 selected = "BH")
renderPlot({
 if (input$p_corr==FALSE){
 plot(-log(univar$pval, base=10), col = ifelse(univar$pval
<as.numeric(input$pumbr),'red','blue'), pch = 19, , ylab="-log10(p)", xlab="Compounds",
main=paste("Differential lipids (a total of",sum(univar$pval<input$pumbr), "with p value<",
input$pumbr, ") obtained with", c("non parametric", "parametric")[as.numeric(par)], "test",
collapse=""))
 #Y añadir la línea umbral
 abline(-log(as.numeric(input$pumbr), base=10), 0, lty=2, col="grey")
 } else if (input$p_corr==TRUE){
 pcorregida<-p.adjust(univar$pval, method = input$corr, n = length(univar$pval))
 plot(-log(pcorregida, base=10), col = ifelse(pcorregida <
as.numeric(input$pumbr),'red','blue'), pch = 19, ylab="-log10(pcorr)", xlab="Compounds",
main=paste("Differential lipids (a total of",sum(pcorregida<input$pumbr), "with corrected p
value<", input$pumbr, ") obtained with", c("non parametric", "parametric")[as.numeric(par)],
"test", collapse=""))

```

```

abline(-log(as.numeric(input$pumbr), base=10), 0, lty=2, col="grey")

}

},height = 500, width = 700)

numericInput("rows", "Metabolites to show", 5)
renderTable({
 univarsignn<-univar[univar$pval<input$pumbr,]
 univarsignn<-univarsignn[order(univarsignn$pval),]
 univarsignn$pcorr<-p.adjust(univarsignn$pval, method = input$corr, n = length(univar$pval))
 features<-rownames(univarsignn)
 univarsignn<-cbind(features, univarsignn)
 head(univarsignn, input$rows)
})
```



### ##Boxplots



```

```{r}
textInput("boxplot",
  label = "Obtain boxplot of:",
  value=colnames(rawdata)[2])

renderPlot({
  boxplot(data[,gsub(" ", "",colnames(data))==gsub(" ", "",input$boxplot)]~data[,1],
    main= paste("Boxplot for lipid ", gsub(" ", "",input$boxplot)), col=col)
},height = 500, width = 700)
```

```



### ##Correlations



```

```{r}
densitat <- colorRampPalette(c('blue', 'white', 'red'))

selectInput("correl",
  label = "Type of correlation",
  choices = c("pearson", "kendall", "spearman"),
  selected = "pearson")

checkboxInput(inputId = "corrsign",
  label = strong("significant metabolites"),
  value = FALSE)

renderPlot({
  if (input$corrsign==FALSE){

    corr<-cor(data[-1], method=input$correl)
    gplots::heatmap.2(corr,
      main = "Correlations",
      srtCol = 20,
      trace="none",

```


```

```

 denscol = "grey",
 density.info = "density",
 col=densitat)

} else if (input$corrsign==TRUE){

 corrsign<-cor(datasign[-1], method=input$correl)
 gplots::heatmap.2(corrsign,
 main = "Correlations sign mb",
 srtCol = 20,
 trace="none",
 denscol = "grey",
 density.info = "density",
 col=densitat)
}
},height = 700, width = 700)
```

```{r}
selectInput("cornom",
 label = "Search for metabolites that correlate with:",
 choices = colnames(rawdata)[-1],
 selected = colnames(rawdata)[2])

renderPlot({
 cortot<-cor(rawdata[,colnames(rawdata)==input$cornom],rawdata[,-1])
 cortot<-cortot[colnames(cortot)!=input$cornom]
 par(mar=c(4,20,4,4))
 rawdata2<-rawdata[,-
c(1,as.numeric(labels(colnames(rawdata))[colnames(rawdata)==input$cornom]
))]
 cortop<-colnames(rawdata)[order(abs(cortot), decreasing=TRUE)][1:10]
 corval<-cortot[order(abs(cortot), decreasing=TRUE)][1:10]
 barplot(corval,cex.names=0.5, horiz=TRUE, col = ifelse(corval < 0,'blue','red'),
 las=1,names.arg=cortop)
}, height = 500, width = 900)
```

### Search for specific correlations

```{r}
textInput("coresp",
 label = "Look for correlation with:",
 value=colnames(rawdata)[3])
renderPlot({
 cortot<-cor(rawdata[,colnames(rawdata)==input$cornom],rawdata[,-1])
 cortot<-cortot[colnames(cortot)!=input$cornom]
 rawdata2<-rawdata[,-
c(1,as.numeric(labels(colnames(rawdata))[colnames(rawdata)==input$cornom]
))]
 cortop<-colnames(rawdata2)[order(abs(cortot), decreasing=TRUE)][1:10]
 corval<-cortot[order(abs(cortot), decreasing=TRUE)][1:10]

```

```

par(mar=c(5,5,1,1))
plot(rawdata2[,gsub(" ", "", colnames(rawdata2))==gsub(" ",
"",input$coresp)]~rawdata[,colnames(rawdata)==input$cornom], xlab=input$cornom,
ylab=gsub(" ", "",input$coresp))

```

```

abline(lm(rawdata2[,gsub(" ", "", colnames(rawdata2))==gsub(" ",
"",input$coresp)]~rawdata[,colnames(rawdata)==input$cornom]))
}, height = 500, width = 700)
```

```

#Multivariate analysis

##Principal component analysis

```

```{r, echo=FALSE, warning=FALSE}
pc<-prcomp(data[-1])
loads<-round(pc$sdev^2/sum(pc$sdev^2)*100,1)
pca<-as.data.frame(predict(pc))[1:3]
pcsign<-prcomp(datasign[-1])
loadssign<-round(pcsign$sdev^2/sum(pcsign$sdev^2)*100,1)
pcasign<-as.data.frame(predict(pcsign))[1:3]
options(rgl.useNULL=TRUE)
library(rgl)
library(shinyRGL)
library(car)
library(threejs)

```

```

checkboxInput(inputId = "pcasign",
 label = strong("PCA only significant metabolites"),
 value = FALSE)

```

```

renderScatterplotThree({

```

```

 if (input$pcasign==FALSE){

```

```

 scatterplot3js(x=pca$PC1,y=pca$PC2,z=pca$PC3, axis.col=c("black","black","black"),
surface=FALSE, grid = FALSE, ellipsoid = F, color=groupcol, labels=paste(rownames(data),
("",as.character(data[,1]),""),sep = ""), renderer="canvas",
size=1.2,flip.y=TRUE,axisLabels=c(paste("PC1", loads[1], "%"),paste("PC2", loads[2],
"%"),paste("PC3", loads[3], "%")))})else if (input$pcasign==TRUE){

```

```

 scatterplot3js(x=pcasign$PC1,y=pcasign$PC2,z=pcasign$PC3,
axis.col=c("black","black","black"), surface=FALSE, grid = FALSE, ellipsoid = F, color=groupcol,
labels=paste(rownames(data), ("",as.character(data[,1]),""),sep = ""),
renderer="canvas", size=1.2,flip.y=TRUE,axisLabels=c(paste("PC1", loadssign[1],
"%"),paste("PC2", loadssign[2], "%"),paste("PC3", loadssign[3], "%")))
 }
}

```

```

}}
```

```

```

```{r}

```

```

renderPlot({
 if (input$pcasign==FALSE){

```



```

 plot(pc, type="lines", main="Principal components explained variance")

 }else if (input$pcasign==TRUE){
 plot(pcsign, type="lines", main="Principal components explained variance for significant
data")

 }

},height = 400, width = 700)
...
##PLS-DA

```{r, message=FALSE, warning=FALSE}
library("mixOmics")
plsda <-plsda(data[-1], data[,1]
              , ncomp = 3, scale=F)
plsdacomp<-as.data.frame(plsda$variates)
...

```{r test2gl, webgl=TRUE}

renderScatterplotThree({scatterplot3js(x=plsdacomp$X.comp.1,y=plsdacomp$X.comp.2,z=plsda
acomps$X.comp.3, grid = FALSE, color=groupcol, labels=paste(rownames(data),
"(",as.character(data[,1]),")",sep = ""), renderer="canvas", size=1.2,flip.y=TRUE)
})
...

##Dendrograms and heatmaps

```{r , echo=FALSE, message=FALSE, warning=FALSE}
library("dendextend")

selectInput(inputId = "metabolitos",
            label = "Used metabolites",
            choices = c("significant", "Top 25 significant", "All"),
            selected = "All")

selectInput(inputId = "metododist",
            label = "Distance criteria",
            choices = c("euclidean","maximum","manhattan", "canberra"),
            selected = "euclidean")

selectInput(inputId = "metodoclust",
            label = "Cluster criteria",
            choices = c("ward.D2","single","complete","average","mcquitty","median","centroid"),
            selected = "euclidean")

checkboxInput(inputId = "dendr",
            label = strong("With heatmap"),
            value = FALSE)
renderPlot({
  if (input$metabolitos=="All"){

```

```

d_data <- dist(data[,-1], method=input$metododist)
hc_data <- hclust(d_data, method = input$metodoclust)
dend <- as.dendrogram(hc_data)
dend <- rotate(dend, 1:nrow(data))
labels_colors(dend) <-
  col[as.numeric(data[,1])[order.dendrogram(dend)]
    ]
labels(dend) <- paste(labels(dend),
  ("",as.character(data[,1])[order.dendrogram(dend)],""),
  sep = "")
dend <- hang.dendrogram(dend,hang_height=0.1)
dend <- set(dend, "labels_cex", 0.5)
if (input$dendr==FALSE){
  par(mar = c(3,3,3,7))
  plot(dend,
    main = "Dendrogram",
    horiz = TRUE, nodePar = list(cex = .007))
  legend("topleft", legend = levels(data[,1]), fill = col, cex=0.6)
}
else if (input$dendr==TRUE){
  gplots::heatmap.2(as.matrix(data[,-1]),
    main = "Heatmap",
    srtCol = 30,
    dendrogram = "row",
    Rowv = dend,
    Colv = "NA",
    trace="none",
    margins =c(10,10),
    denscol = "grey",
    density.info = "density",
    RowSideColors = groupcol,
    col=densitat
  )
  legend("topright", legend = levels(data[,1]), fill = col, cex=0.6)
}
}
else if (input$metabolitos=="significant"){
d_datsign <- dist(datsign[,-1], method = input$metododist)
hc_datsign <- hclust(d_datsign, method = input$metodoclust)
dendsign <- as.dendrogram(hc_datsign)
dendsign <- rotate(dendsign, 1:nrow(datsign))
labels_colors(dendsign) <-
  col[as.numeric(datsign[,1])[order.dendrogram(dendsign)]
    ]
labels(dendsign) <- paste(labels(dendsign),
  ("",as.character(datsign[,1])[order.dendrogram(dendsign)],""),
  sep = "")
dendsign <- hang.dendrogram(dendsign,hang_height=0.1)
dendsign <- set(dendsign, "labels_cex", 0.5)
if (input$dendr==FALSE){
  par(mar = c(3,3,3,7))
  plot(dendsign,

```

```

    main = "Dendrogram significant mb",
    horiz = TRUE, nodePar = list(cex = .007))
legend("topleft", legend = levels(datasign[,1]), fill = col, cex=0.6)

}else if (input$dendr==TRUE){
  gplots::heatmap.2(as.matrix(datasign[,-1]),
    main = "Heatmap significant mb",
    srtCol = 60,
    dendrogram = "row",
    Rowv = dendsign,
    Colv = "NA",
    trace="none",
    margins =c(10,10),
    denscol = "grey",
    density.info = "density",
    RowSideColors = groupcol,
    col=densitat
  )

  legend("topright", legend = levels(datasign[,1]), fill = col, cex=0.6)
}

}else if (input$metabolitos=="Top 25 significant"){
  datatop25<-data[,-1]
  datatop25<-as.data.frame(cbind(data[,1], datatop25[,rank(univar$pval)<26]))
  univartop25<-univar[rank(univar$pval)<26,]
  d_datatop25 <- dist(datatop25[,-1], method=input$metododist)
  hc_datatop25 <- hclust(d_datatop25, method = input$metodoclust)
  dendtop25 <- as.dendrogram(hc_datatop25)
  dendtop25 <- rotate(dendtop25, 1:nrow(datatop25))
  labels_colors(dendtop25) <-
    col[as.numeric(datatop25[,1])[order.dendrogram(dendtop25)]]
  labels(dendtop25) <- paste(labels(dendtop25),
    "(",as.character(datatop25[,1])[order.dendrogram(dendtop25)],")",
    sep = "")
  dendtop25 <- hang.dendrogram(dendtop25,hang_height=0.1)
  dendtop25 <- set(dendtop25, "labels_cex", 0.5)
  if (input$dendr==FALSE){
    par(mar = c(3,3,3,7))
    plot(dendtop25,
      main = "Dendrogram top 25 significant mb ",
      horiz = TRUE, nodePar = list(cex = .007))
    legend("topleft", legend = levels(datatop25[,1]), fill = col, cex=0.6)
  }else if (input$dendr==TRUE){
    gplots::heatmap.2(as.matrix(datatop25[,-1]),
      main = "Heatmap top25 significant mb",
      srtCol = 60,
      dendrogram = "row",
      Rowv = dendtop25,
      Colv = "NA",
      trace="none",

```

```
    margins =c(10,10),
    denscol = "grey",
    density.info = "density",
    RowSideColors = groupcol,
    col= densitat
  )
  legend("topright", legend = levels(datasign[,1]), fill = col, cex=0.6)
}
}
},height = 500, width = 700)
```
```

## 8.2. Annex 2: Tables of metabolites with lower p value

### 8.2.1. CTL vs ALS plasma positive

Table 14: Differential metabolites for CTL vs ALS comparison for plasma samples in positive ionization mode.

| Metabolite                                                                           | estimate.dif | pval  | statistic.w | conf.int.1 | conf.int.2 | regulation (CTL vs ALS) | pcorr | mass    | RT    |
|--------------------------------------------------------------------------------------|--------------|-------|-------------|------------|------------|-------------------------|-------|---------|-------|
| FA(13:1)                                                                             | 2.216        | 0.000 | 218.500     | 2.201      | 2.267      | up                      | 0.00  | 234.17  | 0.88  |
| PG(34:1)                                                                             | -0.844       | 0.000 | 30.000      | -1.579     | -0.325     | down                    | 0.34  | 765.55  | 6.62  |
| PS(41:2) / PS(P-39:0) / PS(O-39:1) / PI(P-34:0) / PI(O-34:1) / PC(42:8) / PC(P-40:6) | -1.200       | 0.001 | 32.000      | -1.807     | -0.514     | down                    | 0.49  | 839.61  | 6.63  |
| Unknown                                                                              | -2.068       | 0.001 | 29.000      | -2.083     | -0.026     | down                    | 0.49  | 606.95  | 0.89  |
| Unknown                                                                              | 0.029        | 0.002 | 170.000     | 0.000      | 2.561      | up                      | 1.00  | 1972.76 | 10.51 |
| Unknown                                                                              | -1.126       | 0.002 | 37.000      | -2.185     | -0.256     | down                    | 1.00  | 1672.49 | 9.58  |
| Unknown                                                                              | 2.113        | 0.002 | 175.000     | 0.000      | 2.236      | up                      | 1.00  | 370.30  | 4.90  |
| PA(39:0) / PA(P-39:1) / PE-Cer(d39:2) / SM(d36:2) // GlcCer(d35:2) / TG(41:1)        | -1.176       | 0.002 | 40.000      | -1.788     | -0.483     | down                    | 1.00  | 728.58  | 6.52  |
| Unknown                                                                              | 1.942        | 0.003 | 190.000     | 0.054      | 2.003      | up                      | 1.00  | 1302.20 | 10.52 |
| Unknown                                                                              | -1.013       | 0.003 | 42.000      | -1.663     | -0.228     | down                    | 1.00  | 1630.44 | 9.00  |
| PA(O-42:0) / TG(46:2) / GlcCer(d46:2)                                                | -0.141       | 0.005 | 42.500      | -0.240     | -0.063     | down                    | 1.00  | 791.69  | 6.49  |
| PC(44:10) / PI(38:4)                                                                 | -0.237       | 0.005 | 42.500      | -2.265     | -0.055     | down                    | 1.00  | 903.57  | 6.57  |
| TG(58:3)                                                                             | -0.095       | 0.005 | 42.500      | -2.875     | -0.033     | down                    | 1.00  | 957.86  | 9.95  |
| Unknown                                                                              | -1.954       | 0.005 | 46.000      | -1.974     | 0.000      | down                    | 1.00  | 1973.77 | 10.52 |
| Unknown                                                                              | -1.940       | 0.005 | 49.500      | -1.980     | 0.000      | down                    | 1.00  | 1659.47 | 10.30 |
| Unknown                                                                              | -1.918       | 0.006 | 49.000      | -1.967     | 0.000      | down                    | 1.00  | 1292.19 | 10.53 |
| Unknown                                                                              | -0.439       | 0.006 | 46.000      | -1.007     | -0.147     | down                    | 1.00  | 631.58  | 8.04  |
| Unknown                                                                              | -0.119       | 0.008 | 48.000      | -0.243     | -0.031     | down                    | 1.00  | 2015.53 | 10.59 |

|                                                                                                |        |       |         |        |        |      |      |         |       |
|------------------------------------------------------------------------------------------------|--------|-------|---------|--------|--------|------|------|---------|-------|
| Unknown                                                                                        | 0.000  | 0.009 | 163.500 | 0.000  | 2.408  | up   | 1.00 | 2153.58 | 9.69  |
| TG(52:3) / TG(54:6)                                                                            | -0.065 | 0.009 | 49.000  | -0.121 | -0.015 | down | 1.00 | 878.73  | 9.95  |
| PS(43:2) / PS(P-41:0) / PI(O-36:1) / PI(P-36:0) / PC(44:8) / PC(P-42:6)                        | -0.170 | 0.009 | 47.500  | -0.331 | -0.043 | down | 1.00 | 867.64  | 7.19  |
| TG(44:0) / GlcCer(d40:0) / GalCer(d40:0)                                                       | -0.293 | 0.011 | 49.500  | -2.596 | -0.074 | down | 1.00 | 767.69  | 9.69  |
| Unknown                                                                                        | -1.913 | 0.011 | 56.000  | -1.940 | 0.000  | down | 1.00 | 1171.30 | 6.72  |
| PI-Cer(d34:0) / PC(34:1) / PE(37:1) / PA(41:5) / PS(P-38:2) / PS(O-38:3) / PC(36:4) / PE(39:4) | -0.125 | 0.011 | 49.500  | -0.239 | -0.032 | down | 1.00 | 781.56  | 6.65  |
| Unknown                                                                                        | -0.329 | 0.011 | 51.000  | -2.000 | -0.082 | down | 1.00 | 1275.18 | 10.54 |
| Unknown                                                                                        | 0.000  | 0.012 | 161.500 | 0.000  | 2.329  | up   | 1.00 | 585.44  | 4.80  |
| DG(34:1) / FA(38:5) / Oleyl arachidonate                                                       | -0.781 | 0.013 | 52.000  | -1.465 | -0.120 | down | 1.00 | 576.51  | 6.62  |
| Unknown                                                                                        | 1.935  | 0.013 | 175.500 | 0.013  | 1.975  | up   | 1.00 | 1293.21 | 10.53 |
| LysoPS(20:3)                                                                                   | -0.108 | 0.014 | 56.000  | -1.956 | 0.000  | down | 1.00 | 529.30  | 1.18  |
| Unknown                                                                                        | -1.966 | 0.014 | 54.000  | -1.989 | 0.000  | down | 1.00 | 1173.30 | 8.61  |
| SM(d41:2) / GlcCer(d40:2) / GalCer(d40:2) / TG(46:1) / TG(48:4) / PA(44:0)                     | -0.186 | 0.015 | 53.000  | -0.434 | -0.029 | down | 1.00 | 798.67  | 9.70  |
| Unknown                                                                                        | -0.046 | 0.016 | 53.500  | -0.108 | -0.003 | down | 1.00 | 1881.54 | 10.46 |
| Unknown                                                                                        | -0.192 | 0.017 | 53.500  | -0.384 | -0.049 | down | 1.00 | 574.49  | 9.93  |
| Unknown                                                                                        | -0.093 | 0.017 | 54.000  | -2.523 | -0.010 | down | 1.00 | 973.87  | 10.10 |
| Unknown                                                                                        | -1.299 | 0.018 | 55.000  | -1.778 | -0.085 | down | 1.00 | 2148.58 | 9.69  |
| Unknown                                                                                        | -0.039 | 0.018 | 56.000  | -2.004 | 0.000  | down | 1.00 | 1637.43 | 9.00  |
| Unknown                                                                                        | 0.365  | 0.019 | 175.000 | 0.018  | 0.494  | up   | 1.00 | 1290.19 | 10.51 |
| TG(52:3)                                                                                       | -0.776 | 0.020 | 55.000  | -1.371 | -0.140 | down | 1.00 | 873.78  | 9.96  |
| TG(56:5)                                                                                       | -0.142 | 0.020 | 55.000  | -0.304 | -0.015 | down | 1.00 | 890.80  | 10.10 |
| Unknown                                                                                        | -0.181 | 0.020 | 55.000  | -0.425 | -0.027 | down | 1.00 | 1735.48 | 9.96  |
| Unknown                                                                                        | -0.050 | 0.020 | 58.000  | -1.960 | 0.000  | down | 1.00 | 1021.74 | 0.88  |
| Unknown                                                                                        | -0.255 | 0.020 | 57.500  | -1.957 | 0.000  | down | 1.00 | 1399.26 | 7.16  |
| Unknown                                                                                        | -0.094 | 0.021 | 56.000  | -2.206 | 0.000  | down | 1.00 | 819.25  | 5.86  |
| TG(56:8)                                                                                       | -0.202 | 0.022 | 61.000  | -1.955 | 0.000  | down | 1.00 | 919.82  | 9.97  |
| PG(36:1) / PE(P-39:1) / PC(P-36:1) / PC(O-36:2) / PC(P-38:4) / PE(41:3) / PC(38:3)             | -0.078 | 0.022 | 58.000  | -2.006 | 0.000  | down | 1.00 | 793.59  | 7.06  |

|                                                                                                    |        |       |         |        |        |      |      |         |       |
|----------------------------------------------------------------------------------------------------|--------|-------|---------|--------|--------|------|------|---------|-------|
| PE(42:1) / PC(39:1) / PE(44:4) / PC(41:4)                                                          | -0.051 | 0.023 | 63.500  | -1.983 | 0.000  | down | 1.00 | 851.64  | 7.05  |
| Unknown                                                                                            | -0.013 | 0.023 | 63.500  | -2.016 | 0.000  | down | 1.00 | 1947.74 | 10.54 |
| PE(40:9) / PG(40:7) / PG(P-38:5) / PG(O-38:6) / PI(P-34:1) / PI(O-34:2) / PI(33:2) / PS(36:3) /    | -0.072 | 0.025 | 59.500  | -1.984 | 0.000  | down | 1.00 | 802.53  | 4.81  |
| Unknown                                                                                            | -0.695 | 0.025 | 58.000  | -1.455 | -0.042 | down | 1.00 | 650.56  | 7.77  |
| Unknown                                                                                            | -0.358 | 0.025 | 58.000  | -0.865 | -0.033 | down | 1.00 | 1169.05 | 8.04  |
| Cer(t40:0) / Cer(d40:1)                                                                            | -2.097 | 0.026 | 58.500  | -2.135 | 0.000  | down | 1.00 | 621.61  | 8.49  |
| DG(36:4)                                                                                           | -1.933 | 0.026 | 60.000  | -1.970 | 0.000  | down | 1.00 | 633.53  | 7.53  |
| Unknown                                                                                            | -0.053 | 0.027 | 58.000  | -0.150 | -0.009 | down | 1.00 | 578.57  | 8.42  |
| Unknown                                                                                            | 0.000  | 0.027 | 158.000 | 0.000  | 2.281  | up   | 1.00 | 1635.43 | 9.00  |
| PC(36:4)                                                                                           | 1.900  | 0.028 | 167.000 | 0.000  | 1.941  | up   | 1.00 | 803.57  | 7.78  |
| PA(P-36:0) / PA(O-36:1) / PE-Cer(d36:1) / SM(d33:1)                                                | -0.159 | 0.028 | 59.000  | -0.392 | -0.025 | down | 1.00 | 688.55  | 6.07  |
| DG(38:2) / DG(40:5) / PE-Cer(d36:1) / SM(d33:1) / PA(P-36:0) / PA(O-36:1) / CE(18:2)               | -0.057 | 0.030 | 63.000  | -1.958 | 0.000  | down | 1.00 | 670.56  | 10.37 |
| PA(P-37:1) / PA(O-37:2) / PA(37:0) / DG(44:9) / SM(d34:2) / PE-Cer(d37:2) / GlcCer(d33:2)          | -0.694 | 0.031 | 60.000  | -1.632 | -0.094 | down | 1.00 | 700.55  | 5.89  |
| Unknown                                                                                            | -0.449 | 0.031 | 60.000  | -1.016 | -0.028 | down | 1.00 | 1137.83 | 0.89  |
| Unknown                                                                                            | -0.136 | 0.032 | 60.000  | -2.246 | 0.000  | down | 1.00 | 1540.12 | 6.70  |
| Unknown                                                                                            | -0.055 | 0.032 | 60.000  | -0.122 | 0.000  | down | 1.00 | 816.25  | 5.86  |
| TG(44:1)                                                                                           | -0.171 | 0.033 | 60.000  | -0.450 | -0.009 | down | 1.00 | 765.68  | 9.50  |
| Unknown                                                                                            | 0.030  | 0.033 | 170.000 | 0.004  | 0.063  | up   | 1.00 | 1038.30 | 7.26  |
| Unknown                                                                                            | -0.119 | 0.033 | 60.000  | -0.268 | -0.012 | down | 1.00 | 1041.93 | 10.09 |
| Unknown                                                                                            | -0.024 | 0.033 | 61.500  | -2.030 | 0.000  | down | 1.00 | 1079.78 | 0.89  |
| Cer(d38:3)                                                                                         | 1.902  | 0.034 | 166.500 | 0.000  | 1.975  | up   | 1.00 | 611.52  | 8.07  |
| PS(P-39:1) / PS(39:0) / PC(40:6) / PE(43:6) / PC(P-40:7) / PC(P-38:4) / PC(O-38:5) / LacCer(d32:1) | -0.575 | 0.034 | 61.000  | -1.456 | -0.031 | down | 1.00 | 815.61  | 6.68  |
| Unknown                                                                                            | -0.631 | 0.034 | 61.000  | -1.679 | -0.041 | down | 1.00 | 639.51  | 6.51  |
| Unknown                                                                                            | 0.000  | 0.036 | 157.500 | 0.000  | 2.216  | up   | 1.00 | 1553.43 | 10.18 |
| PI-Cer(t36:0) / PC(36:2) / PE(39:2) / PC(38:5) / PE(41:5) / PA(43:6) / PS(39:4) / PS(P-37:2)       | -0.109 | 0.038 | 61.500  | -0.210 | -0.005 | down | 1.00 | 807.57  | 6.71  |
| LacCer(d32:1) / PC(38:3) / PE(41:3) / PC(40:6) / PS(41:5)                                          | -0.168 | 0.038 | 62.000  | -0.324 | -0.017 | down | 1.00 | 833.59  | 7.08  |
| TG(52:4)                                                                                           | -0.761 | 0.038 | 62.000  | -1.420 | -0.011 | down | 1.00 | 871.75  | 9.80  |

|                                                                                                           |        |       |         |        |        |      |      |         |       |
|-----------------------------------------------------------------------------------------------------------|--------|-------|---------|--------|--------|------|------|---------|-------|
| Unknown                                                                                                   | -0.030 | 0.039 | 67.000  | -1.973 | 0.000  | down | 1.00 | 430.37  | 7.05  |
| TG(63:3)                                                                                                  | -0.100 | 0.039 | 62.500  | -2.106 | 0.000  | down | 1.00 | 992.92  | 10.54 |
| Unknown                                                                                                   | -0.077 | 0.040 | 63.000  | -2.077 | 0.000  | down | 1.00 | 1711.46 | 9.13  |
| TG(42:0)                                                                                                  | -0.220 | 0.041 | 66.000  | -1.945 | 0.000  | down | 1.00 | 739.66  | 9.47  |
| PE(P-38:5) / PE(O-38:6) / PA(37:2) / PE(35:1) / PC(32:1) / PS(33:0) / PS(P-33:1) / PS(O-33:2) / DG(44:11) | -0.164 | 0.041 | 62.500  | -2.826 | -0.001 | down | 1.00 | 731.54  | 6.57  |
| SM(d39:2) / GlcCer(d38:2) / GalCer(d38:2) / TG(44:1) / PA(42:0) / PA(P-42:1) / PA(O-42:2)                 | -0.123 | 0.042 | 65.000  | -1.930 | 0.000  | down | 1.00 | 770.64  | 9.50  |
| TG(54:8)                                                                                                  | -0.103 | 0.042 | 67.000  | -1.921 | 0.000  | down | 1.00 | 891.73  | 9.49  |
| Unknown                                                                                                   | -0.555 | 0.042 | 63.000  | -1.208 | -0.003 | down | 1.00 | 1266.19 | 10.54 |
| PA(P-38:0) / PA(O-38:1) / SM(d35:1) / PE-Cer(d38:1) / PE(O-34:3) / PE(P-34:2) /                           | -0.083 | 0.043 | 64.000  | -2.031 | 0.000  | down | 1.00 | 716.57  | 6.67  |
| PI-Cer(d36:0) / PC(36:1) / PE(39:1) / PC(38:4) / PE(41:4) / PS(O-40:3) / PS(P-40:2) / PS(39:3)            | -0.088 | 0.043 | 63.000  | -0.168 | 0.000  | down | 1.00 | 809.59  | 7.18  |
| PE(P-40:6) / PE(37:2) / PC(34:2) / PA(39:3) / PE-NMe(36:2) / PS(O-36:1) / PS(P-36:0) / PS(35:1) /         | -0.686 | 0.044 | 63.000  | -1.485 | -0.029 | down | 1.00 | 757.55  | 6.69  |
| TG(58:11)                                                                                                 | -0.167 | 0.044 | 65.000  | -1.911 | 0.000  | down | 1.00 | 906.74  | 9.41  |
| Unknown                                                                                                   | -0.080 | 0.044 | 64.000  | -2.079 | 0.000  | down | 1.00 | 548.40  | 0.89  |
| TG(55:1)                                                                                                  | -0.123 | 0.046 | 64.000  | -2.169 | 0.000  | down | 1.00 | 884.82  | 10.23 |
| Unknown                                                                                                   | -0.079 | 0.046 | 64.000  | -2.194 | 0.000  | down | 1.00 | 1807.52 | 10.41 |
| Unknown                                                                                                   | -0.030 | 0.046 | 63.500  | -0.070 | 0.000  | down | 1.00 | 907.22  | 7.79  |
| Unknown                                                                                                   | -0.104 | 0.047 | 64.000  | -2.488 | 0.000  | down | 1.00 | 1164.02 | 8.04  |
| SM(d43:1) / TG(50:3) / TG(52:6) / PS(O-40:0) / Glc-cholesterol(22:1)                                      | -0.114 | 0.048 | 64.000  | -0.219 | 0.000  | down | 1.00 | 850.70  | 9.75  |
| Unknown                                                                                                   | 0.067  | 0.048 | 163.000 | 0.000  | 1.964  | up   | 1.00 | 1345.16 | 10.38 |
| Unknown                                                                                                   | 1.917  | 0.050 | 162.000 | 0.000  | 1.953  | up   | 1.00 | 1395.38 | 9.98  |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.



## 8.2.2. CTL vs ALS plasma negative

Table 15: Differential metabolites for CTL vs ALS comparison for plasma samples in negative ionization mode.

| Metabolite                                                            | estimate.dif | pval     | statistic.w | conf.int.1 | conf.int.2 | regulation (CTL vs ALS) | pcorr    | mass     | RT       |
|-----------------------------------------------------------------------|--------------|----------|-------------|------------|------------|-------------------------|----------|----------|----------|
| Unknown                                                               | 2.628870492  | 0.001165 | 124         | 1.35E-05   | 2.6379372  | up                      | 0.299395 | 1320.069 | 11.11917 |
| CE(15:0) / DG(38:3) / DG(35:0)                                        | 0.427525118  | 0.001809 | 141         | 0.10155039 | 0.88456672 | up                      | 0.463008 | 646.538  | 8.574236 |
| Unknown                                                               | 0.033719377  | 0.003738 | 140         | 0.01193217 | 0.07530576 | up                      | 0.953204 | 1167.268 | 7.868148 |
| PG(36:4) / PA(42:9) / MGDG(36:3) / PI(P-32:0) / PI(O-32:1)            | 0.990969498  | 0.005906 | 135         | 0.30472273 | 1.6168847  | up                      | 1        | 830.5249 | 6.59803  |
| Unknown                                                               | 0.025292926  | 0.00796  | 135         | 0.00743855 | 0.06227919 | up                      | 1        | 1257.057 | 8.564893 |
| PI(37:2)                                                              | 0.080850686  | 0.008094 | 135         | 0.01874655 | 0.15356097 | up                      | 1        | 912.5292 | 6.592938 |
| Unknown                                                               | 0.61740567   | 0.01168  | 131         | 0.1779576  | 1.0973004  | up                      | 1        | 1271.069 | 11.18983 |
| NAPE(52:2)                                                            | 2.114579699  | 0.013156 | 122         | 4.65E-05   | 2.1644233  | up                      | 1        | 981.7657 | 10.72091 |
| Unknown                                                               | 0.022796377  | 0.014407 | 118         | 6.61E-05   | 2.3288969  | up                      | 1        | 1120.898 | 10.73814 |
| Unknown                                                               | 0.511591928  | 0.015988 | 129         | 0.09303163 | 1.44752126 | up                      | 1        | 1510.331 | 8.533118 |
| Unknown                                                               | 0.073001629  | 0.01812  | 128         | 0.00434976 | 2.05724612 | up                      | 1        | 1613.374 | 8.96781  |
| Unknown                                                               | 0.039354563  | 0.018174 | 118.5       | 4.67E-05   | 2.23247716 | up                      | 1        | 604.6761 | 7.1181   |
| TG(60:7) / TG(55:2)                                                   | 1.898479115  | 0.018631 | 127         | 0.00909673 | 1.98260959 | up                      | 1        | 960.8138 | 10.03656 |
| Glc-Cholesterol(18:0) / TG(48:8) / PG(P-38:0) / PG(O-38:1) / PA(44:1) | 1.914612235  | 0.019964 | 122         | 7.42E-05   | 2.08517196 | up                      | 1        | 850.6337 | 9.707544 |
| NeuAalpha2-3Galbeta-Cer(d42:1)                                        | 2.025640584  | 0.019964 | 122         | 1.77E-05   | 2.04822144 | up                      | 1        | 1102.783 | 10.055   |
| DG(37:3) / DG(42:8)                                                   | 0.951313863  | 0.024842 | 126         | 0.20864571 | 1.78414086 | up                      | 1        | 692.5471 | 8.569751 |
| PC(O-28:0)                                                            | 0.03844534   | 0.026391 | 126         | 0.00232764 | 2.31647641 | up                      | 1        | 709.5456 | 8.571888 |
| Unknown                                                               | 0.067332843  | 0.026793 | 121         | 8.66E-05   | 2.01502918 | up                      | 1        | 1811.402 | 9.1275   |
| Unknown                                                               | 0.065102947  | 0.02856  | 125         | 0.00878689 | 0.11473051 | up                      | 1        | 868.1869 | 6.389152 |
| TG(62:7)                                                              | 0.085020993  | 0.030781 | 125         | 0.0084556  | 0.19259314 | up                      | 1        | 1024.797 | 10.04314 |
| Unknown                                                               | 0.042788537  | 0.031694 | 123         | 1.62E-05   | 1.99703526 | up                      | 1        | 1880.416 | 9.24381  |

|                                                    |              |          |       |             |             |      |   |          |          |
|----------------------------------------------------|--------------|----------|-------|-------------|-------------|------|---|----------|----------|
| Unknown                                            | -0.171503078 | 0.032907 | 36.5  | -0.2746473  | -0.04200651 | down | 1 | 791.7042 | 6.494657 |
| Unknown                                            | 0.023741937  | 0.03485  | 114.5 | -7.28E-07   | 2.21547394  | up   | 1 | 1116.904 | 10.73289 |
| PGE2 / PGD2 / EET / HETE                           | -2.207376896 | 0.035405 | 37.5  | -2.22836056 | -2.92E-05   | down | 1 | 380.257  | 4.132522 |
| Cer(t42:0)                                         | 0.554735676  | 0.037345 | 123   | 0.00902412  | 1.18347977  | up   | 1 | 727.6811 | 7.116029 |
| Unknown                                            | 0.024473973  | 0.039765 | 115   | -2.14E-05   | 2.16070994  | up   | 1 | 1342.323 | 8.3359   |
| Unknown                                            | 0.053868354  | 0.041572 | 122   | 1.77E-05    | 2.09160021  | up   | 1 | 1191.285 | 7.86576  |
| Unknown                                            | 0.013956789  | 0.047242 | 112.5 | -2.16E-05   | 2.22275316  | up   | 1 | 2480.923 | 5.1263   |
| FA(18:3) / DHET / deoxy PGF2alpha / deoxy PGF2beta | -0.08675725  | 0.048916 | 43    | -1.97107034 | 2.60E-05    | down | 1 | 338.2397 | 3.166294 |
| FA(18:1)                                           | -0.184871712 | 0.049736 | 40    | -3.25932218 | -0.00109815 | down | 1 | 282.2535 | 3.385161 |
| Unknown                                            | 0.03531833   | 0.049812 | 121   | 0.00093003  | 0.09243306  | up   | 1 | 1658.381 | 8.861469 |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time; Yellow font: LC-MS method not prepared to detect this metabolite.

### 8.2.3. CTL vs ALS CSF positive

Table 16: Top 25 metabolites with lower p value obtained from the univariate tests for CTL vs ALS comparison for CSF samples in positive ionization mode.

| Metabolite                                                                                | estimate.dif | pval     | statistic.w | conf.int.1  | conf.int.2  | regulation (CTL vs ALS) | pcorr | mass     | RT       |
|-------------------------------------------------------------------------------------------|--------------|----------|-------------|-------------|-------------|-------------------------|-------|----------|----------|
| Unknown                                                                                   | -0.054273613 | 0.002958 | 34          | -3.39374636 | -0.01848402 | down                    | 1     | 897.85   | 10.44447 |
| PS(33:0) / PI(28:0) / PG(35:5) / PS(35:3) / PE(37:6) / PC(34:6) / PE(O-38:6) / PE(P-38:5) | 1.874435812  | 0.007472 | 170         | 0.00827775  | 1.94267236  | up                      | 1     | 771.5044 | 0.891945 |
| PC(40:10) / PC(38:7) / PE(41:7) / PI(32:1) / PS(37:1) / PS(39:4) / PG(39:6)               | 2.073488385  | 0.009283 | 162.5       | 6.00E-05    | 2.08842944  | up                      | 1     | 825.5394 | 0.888846 |
| Unknown                                                                                   | 0.059347123  | 0.013138 | 169         | 0.01144949  | 2.14413931  | up                      | 1     | 870.0627 | 0.889885 |
| Unknown                                                                                   | -1.881150954 | 0.016107 | 53          | -1.96637779 | -5.43E-05   | down                    | 1     | 1717.551 | 10.44213 |
| Unknown                                                                                   | 1.923175215  | 0.016232 | 162         | 1.62E-05    | 1.98978898  | up                      | 1     | 1563.499 | 6.551134 |
| Unknown                                                                                   | 1.881842626  | 0.036694 | 155         | 1.03E-05    | 1.96024547  | up                      | 1     | 1207.31  | 9.953177 |
| Unknown                                                                                   | 0.000425253  | 0.036935 | 149         | -1.61E-05   | 2.20188183  | up                      | 1     | 509.5093 | 8.287667 |
| Unknown                                                                                   | -8.67E-05    | 0.04138  | 63          | -2.01838886 | 4.25E-05    | down                    | 1     | 1159.389 | 7.437734 |
| PE(34:6) / PA(36:7) / PE(32:3) / PG(30:2) / PS(30:0)                                      | -0.009165353 | 0.050657 | 64          | -1.92156487 | 7.28E-05    | down                    | 1     | 707.4626 | 0.8895   |
| Unknown                                                                                   | -0.033214181 | 0.052148 | 59.5        | -2.9792123  | 4.97E-05    | down                    | 1     | 928.2253 | 7.111781 |
| Unknown                                                                                   | 0.040991442  | 0.053716 | 156         | -0.00030685 | 2.36564634  | up                      | 1     | 912.1098 | 0.8864   |
| Unknown                                                                                   | -0.052557204 | 0.06007  | 61          | -0.12551554 | 0.00147817  | down                    | 1     | 821.03   | 0.885147 |
| Unknown                                                                                   | 0.023009728  | 0.063176 | 154         | -7.39E-05   | 0.11522627  | up                      | 1     | 504.3925 | 6.442857 |
| Cer(d44:1)                                                                                | 0.017571874  | 0.063632 | 147         | -2.71E-05   | 2.063732    | up                      | 1     | 1355.347 | 10.16825 |
| Unknown                                                                                   | -0.073105659 | 0.068677 | 62.5        | -2.95126098 | 0.00125967  | down                    | 1     | 1446.369 | 8.938281 |
| Unknown                                                                                   | 0.037442958  | 0.068886 | 153.5       | -0.00364165 | 0.06610189  | up                      | 1     | 129.1517 | 0.803629 |
| Unknown                                                                                   | 4.05E-05     | 0.069724 | 145         | -1.41E-05   | 2.10532973  | up                      | 1     | 792.9944 | 0.885    |
| (3'-sulfo)Gal-Cer(d42:1)                                                                  | -0.050048776 | 0.071603 | 63          | -0.11961789 | 0.0056907   | down                    | 1     | 905.6387 | 0.889389 |
| TG(50:2)                                                                                  | -0.086601469 | 0.073193 | 64          | -2.07438734 | 1.05E-05    | down                    | 1     | 1678.515 | 10.28408 |
| Unknown                                                                                   | -0.033161828 | 0.077359 | 67.5        | -1.93336042 | 4.30E-05    | down                    | 1     | 435.336  | 5.044765 |
| Unknown                                                                                   | -0.055932181 | 0.078132 | 67          | -1.94494138 | 4.79E-05    | down                    | 1     | 1314.282 | 8.57353  |

|                        |              |          |      |             |            |      |   |          |          |
|------------------------|--------------|----------|------|-------------|------------|------|---|----------|----------|
| Unknown                | -0.717849125 | 0.078502 | 64   | -1.34024349 | 0.03385745 | down | 1 | 1386.827 | 7.989811 |
| Decatetraenedioic acid | -0.102360017 | 0.082661 | 65   | -2.18953089 | 2.26E-05   | down | 1 | 176.0479 | 0.904704 |
| Octanol                | -0.041044634 | 0.085058 | 68.5 | -1.90400959 | 3.14E-05   | down | 1 | 112.1248 | 5.046353 |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.

## 8.2.4. CTL vs ALS CSF negative

Table 17: Top 25 metabolites with lower p value obtained from the univariate tests for CTL vs ALS comparison for CSF samples in negative ionization mode.

| Metabolite                                                                       | estimate.dif | pval       | statistic.w | conf.int.1  | conf.int.2 | regulation (CTL vs ALS) | pcorr      | mass     | RT       |
|----------------------------------------------------------------------------------|--------------|------------|-------------|-------------|------------|-------------------------|------------|----------|----------|
| MG(18:0)                                                                         | 2.25571805   | 3.92E-05   | 185         | 2.24630437  | 2.30228883 | up                      | 0.00737023 | 418.3287 | 4.799778 |
| TG(63:4)                                                                         | 0.03656828   | 0.01291161 | 170         | 0.0109407   | 0.07039519 | up                      | 1          | 1044.861 | 10.84229 |
| DG(44:3)                                                                         | 0.09818211   | 0.01852435 | 164.5       | 0.00114948  | 2.00604889 | up                      | 1          | 790.6497 | 9.7514   |
| Unknown                                                                          | 0.02301025   | 0.01907429 | 154         | 3.56E-06    | 2.21213401 | up                      | 1          | 1393.318 | 8.575778 |
| TG(66:10)                                                                        | 0.22918467   | 0.02587659 | 156         | 5.72E-05    | 2.04116125 | up                      | 1          | 1074.835 | 10.272   |
| Unknown                                                                          | 0.06601976   | 0.03367331 | 161         | 0.00191731  | 0.13010624 | up                      | 1          | 969.8068 | 10.10119 |
| TG(64:8)                                                                         | 1.05320868   | 0.03685747 | 153         | 4.73E-05    | 2.04202529 | up                      | 1          | 1074.867 | 10.6525  |
| Unknown                                                                          | -0.04959271  | 0.04040496 | 57.5        | -2.13406855 | -3.99E-05  | down                    | 1          | 1908.457 | 9.55688  |
| TG(62:4)                                                                         | 0.02595771   | 0.07160289 | 153         | -0.00295554 | 0.05407932 | up                      | 1          | 1030.848 | 10.71756 |
| Unknown                                                                          | -0.02604247  | 0.08433732 | 65          | -2.3528308  | 3.12E-07   | down                    | 1          | 1122.834 | 10.58983 |
| 9-deoxy-9-methylene-16,16-dimethyl -PGE2 / DAT(44:0)                             | 0.02307929   | 0.08528008 | 151         | -0.00227233 | 0.05767975 | up                      | 1          | 1080.787 | 10.29093 |
| LysoPE(18:1) / PI(P-37:1) / PI(O-37:2) / PI(P-42:6)                              | -0.03206078  | 0.08590116 | 67.5        | -1.96329534 | 4.41E-05   | down                    | 1          | 922.6069 | 9.225841 |
| Ganglioside GM3 (d38:1)                                                          | -0.06784859  | 0.09352415 | 68          | -1.8416055  | 3.42E-05   | down                    | 1          | 1190.763 | 6.806477 |
| Unknown                                                                          | -0.02235016  | 0.09688152 | 67          | -2.11983237 | 3.09E-05   | down                    | 1          | 1104.863 | 10.73064 |
| PI(37:2) / PG(44:7)                                                              | 0.0202967    | 0.1006057  | 149         | -0.00281401 | 0.17770881 | up                      | 1          | 912.5312 | 6.553133 |
| PG(37:4) / TG(46:6) / PA(39:2) / PG(38:2) / PA(P-40:1) / PA(O-40:2) / PG(P-39:1) | 0.40899445   | 0.11139863 | 148         | -0.13127812 | 0.96979894 | up                      | 1          | 802.579  | 9.262784 |
| DG(44:3) / DG(41:0)                                                              | -0.34355478  | 0.11479857 | 70          | -2.06427061 | 4.91E-06   | down                    | 1          | 730.6294 | 9.747137 |
| DG(42:3) / DG(36:0) / CE(16:0)                                                   | -0.02728639  | 0.11592429 | 69          | -2.20590676 | 2.82E-05   | down                    | 1          | 684.5875 | 9.058629 |
| Unknown                                                                          | 0.03430999   | 0.12218081 | 144.5       | -3.12E-05   | 1.98534017 | up                      | 1          | 1077.847 | 10.63428 |
| Arachidonoylmorpholine / PC(52:4)                                                | -0.01535678  | 0.12745462 | 72          | -1.95499618 | 2.25E-05   | down                    | 1          | 1065.83  | 10.52237 |
| PGF2alpha methyl ether /TG(65:9)                                                 | 0.03234938   | 0.12952584 | 146         | -0.01007811 | 0.07777838 | up                      | 1          | 1062.831 | 10.5288  |
| Unknown                                                                          | 0.00369      | 0.13631086 | 141         | -2.24E-05   | 2.02611902 | up                      | 1          | 763.6228 | 9.561785 |

|                                                                         |             |            |       |             |            |      |   |          |          |
|-------------------------------------------------------------------------|-------------|------------|-------|-------------|------------|------|---|----------|----------|
| Unknown                                                                 | 0.02068717  | 0.14277466 | 144.5 | -0.00617496 | 0.11276124 | up   | 1 | 1192.78  | 6.862536 |
| PG(O-35:1) / PG(P-35:0) / PA(41:1) / PA(44:4) / PG(P-40:5) / PG(O-40:6) | 0.00155015  | 0.14878551 | 140   | -7.07E-05   | 2.04325084 | up   | 1 | 808.5749 | 9.245563 |
| 11-deoxy-PGF1b / TG(62:9) / TG(65:12)                                   | -0.03143054 | 0.16017826 | 73    | -2.27480168 | 0.00558801 | down | 1 | 1020.79  | 10.0609  |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.

## 8.2.5. Onset plasma positive

Table 18: Differential metabolites for onset comparison for plasma samples in positive ionization mode.

| Metabolite                                              | statistic  | pval       | posthoc                                       | mass     | RT       |
|---------------------------------------------------------|------------|------------|-----------------------------------------------|----------|----------|
| DG(38:1)                                                | 8.15217391 | 0.01697375 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 306.3244 | 2.495108 |
| Unkown                                                  | 6.52173913 | 0.03835503 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 625.7484 | 7.110135 |
| Unkown                                                  | 6.21958121 | 0.0446103  | RESPIRATORY - SPINAL ,                        | 914.8021 | 9.957381 |
| Unkown                                                  | 7.59683794 | 0.02240617 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 566.6906 | 7.12019  |
| Unkown                                                  | 6          | 0.04978707 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 576.5078 | 6.615459 |
| Unkown                                                  | 6.57509881 | 0.03734526 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 682.7946 | 7.117567 |
| Unkown                                                  | 6.3282226  | 0.04225167 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 816.245  | 5.857839 |
| Unkown                                                  | 10.8537549 | 0.0043968  | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 1094.31  | 9.457305 |
| Unkown                                                  | 7.77860697 | 0.02045959 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 926.8917 | 10.239   |
| Unkown                                                  | 8.05335968 | 0.01783344 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 733.7551 | 6.303918 |
| Cer(d34:1)                                              | 9.26679842 | 0.00972166 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 639.5071 | 6.50908  |
| PA(O-42:0) / TG(46:2) / Gal-Cer(d42:2) / Glc-Cer(d42:2) | 6.13061224 | 0.04663956 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 793.5881 | 7.058999 |
| Unkown                                                  | 11.5017036 | 0.00318007 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 667.6085 | 7.316304 |
| TG(58:1)                                                | 7.46617232 | 0.0239189  | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 604.5144 | 7.554618 |
| LysoPC(22:1)                                            | 6.06521739 | 0.04818976 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 893.2587 | 6.385109 |
| Unkown                                                  | 10.2540297 | 0.00593425 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 1550.05  | 6.63665  |
| Unkown                                                  | 6.35460993 | 0.04169788 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 1555.056 | 6.626369 |
| DG(35:3) / DG(33:0)                                     | 6.2065678  | 0.04490151 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 1848.531 | 10.40632 |
| Unkown                                                  | 7.56594265 | 0.02275498 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 759.1872 | 6.949621 |
| Unkown                                                  | 9.25674274 | 0.00977066 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1811.47  | 9.347261 |

|                                                                                    |            |            |                                               |          |           |
|------------------------------------------------------------------------------------|------------|------------|-----------------------------------------------|----------|-----------|
| Unkown                                                                             | 7.7173913  | 0.0210955  | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 577.4207 | 0.8889714 |
| Unkown                                                                             | 6.73517787 | 0.03447265 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 668.8113 | 7.12244   |
| TG(58:5) / TG(60:8)                                                                | 6.38983532 | 0.0409699  | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 958.778  | 9.487687  |
| Unkown                                                                             | 6.17391304 | 0.04564065 | RESPIRATORY - SPINAL ,                        | 690.9664 | 8.282028  |
| Unkown                                                                             | 10.5311203 | 0.0051665  | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 1742.45  | 9.24021   |
| TG(58:7)                                                                           | 8.08941709 | 0.01751481 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 554.5384 | 9.17709   |
| Unkown                                                                             | 8.79400387 | 0.0123142  | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 1785.461 | 9.242332  |
| Unkown                                                                             | 9.57446809 | 0.00833548 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1379.409 | 9.095188  |
| Unkown                                                                             | 6.86205316 | 0.03235371 | RESPIRATORY - SPINAL ,                        | 1190.35  | 8.7525    |
| PC(48:1)                                                                           | 11.3693878 | 0.00339757 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 443.3666 | 4.091218  |
| PG(36:1) / PE(P-39:1) / PC(P-36:1) / PC(O-36:2) / PC(P-38:4) / PE(41:3) / PC(38:3) | 6.17272727 | 0.04566772 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 1323.174 | 10.52813  |
| Unkown                                                                             | 6.16390041 | 0.04586971 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 977.8031 | 9.615167  |
| TG(58:11)                                                                          | 6.01632653 | 0.0493823  | BULBAR - RESPIRATORY ,                        | 906.7407 | 9.412182  |
| DG(34:1) / FA(38:5) / Oleyl arachidonate                                           | 8.07871948 | 0.01760874 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 791.6942 | 6.493702  |

The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Red font: Unexpected retention time.



## 8.2.6. Onset plasma negative

Table 19: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for plasma samples in negative ionization mode.

| Metabolite                                                                                               | statistic  | pval       | posthoc                                       | mass     | RT       |
|----------------------------------------------------------------------------------------------------------|------------|------------|-----------------------------------------------|----------|----------|
| Unknown                                                                                                  | 6.73517787 | 0.03447265 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 868.1869 | 6.389152 |
| PS(40:4) / Ins-1-P-Cer(d40:1) / LacCer(d34:0) / PC(42:5) / PE(44:10)                                     | 6.63241107 | 0.03629027 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,      | 899.5864 | 6.731    |
| Unknown                                                                                                  | 6.20645346 | 0.04490408 | RESPIRATORY - SPINAL ,                        | 1167.268 | 7.868148 |
| PA(P-42:6) / PA(36:2) / PG(35:2)                                                                         | 6.12648221 | 0.04673597 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 760.5266 | 8.563705 |
| PE(P-36:4) / PE(O-36:5) / PC(O-28:0) / PE(O-31:0) / LysoPC(28:0) / PE-NMe2(28:0) / PC(27:0) / PE(30:0)   | 5.59475806 | 0.06096965 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 723.5184 | 6.999346 |
| TG(62:12) / TG(57:7) / TG(59:9)                                                                          | 5.47977178 | 0.06457772 | RESPIRATORY - SPINAL ,                        | 978.7646 | 10.73168 |
| Ganglioside GM2 (d44:0)                                                                                  | 5.42087542 | 0.06650769 | RESPIRATORY - SPINAL ,                        | 1496.978 | 10.82647 |
| Unknown                                                                                                  | 5.28853755 | 0.07105729 | RESPIRATORY - SPINAL ,                        | 652.7159 | 7.124117 |
| TG(48:6) / PA(41:2) / PG(40:2)                                                                           | 5.05602665 | 0.07981743 | BULBAR - SPINAL ,                             | 830.6051 | 9.058157 |
| PS(P-36:0) / PS(O-36:1) / PC(P-32:1) / PC(O-32:2) / PE(P-35:1) / PE(O-35:1) / PE(P-40:6)                 | 4.98615843 | 0.08265506 | RESPIRATORY - SPINAL ,                        | 775.5554 | 7.395821 |
| TG(56:12)                                                                                                | 4.97954545 | 0.08292881 | BULBAR - SPINAL ,                             | 954.7029 | 7.970666 |
| CE(16:0) / DG(39:3) / DG(36:0)                                                                           | 4.96525424 | 0.08352351 | RESPIRATORY - SPINAL ,                        | 660.5516 | 8.798552 |
| LysoPI(22:0) / PA(30:0)                                                                                  | 4.92537313 | 0.08520573 | RESPIRATORY - SPINAL ,                        | 656.3962 | 6.735429 |
| PA(37:2) / PG(36:2) / DG(44:11) / PA(P-40:3) / PA(O-40:4) / MGDG(36:6) / PA(42:7)                        | 4.92020374 | 0.08542625 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 774.5321 | 6.6865   |
| Unknown                                                                                                  | 4.84723015 | 0.08860074 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 1436.301 | 8.337793 |
| Unknown                                                                                                  | 4.76679842 | 0.09223651 | RESPIRATORY - SPINAL ,                        | 791.7042 | 6.494657 |
| Unknown                                                                                                  | 4.73517787 | 0.09370639 | BULBAR - SPINAL ,                             | 1584.357 | 8.706593 |
| TG(52:4) / TG(55:7)                                                                                      | 4.68057922 | 0.09629974 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,      | 890.7064 | 9.808929 |
| Unknown                                                                                                  | 4.61846962 | 0.09933723 | BULBAR - SPINAL ,                             | 1732.398 | 9.001601 |
| PE(P-38:3) / PE(O-38:4) / PC(O-35:4) / PE(O-35:1) / PE(P-35:0) / PC(P-32:0) / PC(O-32:1) / Cerebroside C | 4.55335968 | 0.10262437 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 753.5488 | 6.606237 |
| Unknown                                                                                                  | 4.53952569 | 0.10333668 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 474.4055 | 7.232344 |
| Unknown                                                                                                  | 4.48814229 | 0.10602598 | RESPIRATORY - SPINAL ,                        | 613.7043 | 7.124972 |

|                                  |            |            |                        |          |          |
|----------------------------------|------------|------------|------------------------|----------|----------|
| DG(41:0) / DG(44:3)              | 4.48418972 | 0.10623572 | BULBAR - SPINAL ,      | 730.6279 | 9.566471 |
| FA(38:4)                         | 4.37903226 | 0.11197091 | BULBAR - SPINAL ,      | 616.5453 | 4.123778 |
| PG(O-30:0) / PA(39:3) / PA(36:0) | 4.36361064 | 0.11283764 | RESPIRATORY - SPINAL , | 740.5173 | 8.567965 |

The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.

## 8.2.7. Onset CSF positive

Table 20: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for CSF samples in positive ionization mode.

| Metabolite                                                               | statistic  | pval       | posthoc                                                         | mass     | RT        |
|--------------------------------------------------------------------------|------------|------------|-----------------------------------------------------------------|----------|-----------|
| Unknown                                                                  | 11.954729  | 0.0025355  | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1964.514 | 9.734933  |
| Unknown                                                                  | 8.38600949 | 0.01510084 | BULBAR - RESPIRATORY , BULBAR - SPINAL , RESPIRATORY - SPINAL , | 829.0213 | 0.8809047 |
| Unknown                                                                  | 7.90505226 | 0.01920612 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1612.449 | 10.42643  |
| Unknown                                                                  | 7.59186704 | 0.02246193 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL ,                   | 926.2596 | 7.111625  |
| (3'-sulfo)Gal-Cer(d42:1)                                                 | 7.28959627 | 0.02612668 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,                        | 905.6387 | 0.8893889 |
| TG(54:4)                                                                 | 7.24063743 | 0.02677414 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL ,                   | 899.7938 | 10.17697  |
| Unknown                                                                  | 7.15398551 | 0.02795965 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,                        | 1303.425 | 9.850638  |
| PE(46:2) / PC(43:2) / TG(56:9) / TG(54:6)                                | 7.10932657 | 0.028591   | RESPIRATORY - SPINAL ,                                          | 900.7339 | 9.869     |
| TG(46:0)                                                                 | 6.93295615 | 0.03122682 | BULBAR - SPINAL ,                                               | 1538.428 | 10.34903  |
| Ganglioside GA2 (d40:1) / GalNAcbeta1-4Galbeta1-4Glcbeta-Cer(d18:1/22:0) | 6.64881252 | 0.03599388 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1130.773 | 7.969731  |
| Unknown                                                                  | 6.61368043 | 0.03663174 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1094.313 | 9.670655  |
| Unknown                                                                  | 6.49570416 | 0.03885758 | BULBAR - SPINAL ,                                               | 1316.368 | 10.06856  |
| Unknown                                                                  | 6.40965583 | 0.04056588 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,                        | 940.1415 | 5.687389  |
| Unknown                                                                  | 6.03643222 | 0.04888835 | RESPIRATORY - SPINAL ,                                          | 1260.36  | 8.360147  |
| GalabiosylCer(d40:1)                                                     | 5.77305516 | 0.05576953 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1872.485 | 10.67313  |
| Unknown                                                                  | 5.76164596 | 0.05608858 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1497.401 | 10.35333  |
| Unknown                                                                  | 5.76164596 | 0.05608858 | BULBAR - SPINAL , RESPIRATORY - SPINAL ,                        | 1867.493 | 10.67259  |
| Unknown                                                                  | 5.52378576 | 0.06317208 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,                        | 848.0531 | 0.8891923 |
| Unknown                                                                  | 5.46470837 | 0.06506593 | BULBAR - RESPIRATORY , BULBAR - SPINAL ,                        | 1245.349 | 9.952369  |
| Unknown                                                                  | 5.36904762 | 0.06825369 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL ,                   | 977.6486 | 0.8875    |
| Unknown                                                                  | 5.3449793  | 0.06908003 | RESPIRATORY - SPINAL ,                                          | 612.6514 | 2.866297  |

|                                           |            |            |                        |          |          |
|-------------------------------------------|------------|------------|------------------------|----------|----------|
| Unknown                                   | 5.33497871 | 0.06942631 | BULBAR - SPINAL ,      | 1941.508 | 10.7195  |
| Unknown                                   | 5.33478513 | 0.06943303 | BULBAR - RESPIRATORY , | 1306.428 | 9.849946 |
| Unknown                                   | 5.27677977 | 0.07147626 | RESPIRATORY - SPINAL , | 981.2467 | 8.340643 |
| PG(38:4) / PI(33:2) / PG(40:7) / PS(38:8) | 5.26285044 | 0.07197581 | RESPIRATORY - SPINAL , | 820.5251 | 0.886931 |

The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Red font: Unexpected retention time.

## 8.2.8. Onset CSF negative

Table 21: Top 25 metabolites with lower p value obtained from the univariate tests for onset comparison for CSF samples in negative ionization mode.

| Metabolite                                                                           | statistic  | pval       | posthoc                                       | mass     | RT       |
|--------------------------------------------------------------------------------------|------------|------------|-----------------------------------------------|----------|----------|
| Arachidonoylmorpholine / PC(52:4)                                                    | 7.87066779 | 0.01953917 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1065.83  | 10.52237 |
| PA(33:3) / PA(30:0)                                                                  | 7.20063879 | 0.027315   | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 656.4273 | 5.202571 |
| Unknown                                                                              | 6.86916236 | 0.03223891 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1077.847 | 10.63428 |
| PG(37:4) / PI(33:2) / PG(40:7) / PA(41:7)                                            | 5.96376812 | 0.05069723 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 820.4986 | 8.709084 |
| Unknown                                                                              | 5.68115942 | 0.05839181 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1223.048 | 10.69381 |
| LysoPE(18:1) / PI(P-37:1) / PI(O-37:2) / PI(P-42:6)                                  | 5.65186074 | 0.0592535  | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 922.6069 | 9.225841 |
| PE(43:0)                                                                             | 4.43785851 | 0.10872546 | BULBAR - SPINAL ,                             | 1691.395 | 9.220823 |
| TG(64:9)                                                                             | 4.41892975 | 0.10975937 | BULBAR - RESPIRATORY , RESPIRATORY - SPINAL , | 1048.818 | 10.24594 |
| TG(65:12) / TG(62:9) / TG(60:7)                                                      | 4.32686335 | 0.11493004 | BULBAR - SPINAL ,                             | 1020.803 | 10.27224 |
| N.arachidonoyl.D.serine / PA(39:0) / PG(O-33:0) / PG(P-38:4) / PG(O-38:5) / PA(42:3) | 4.24534161 | 0.11971147 | RESPIRATORY - SPINAL ,                        | 782.5519 | 8.333594 |
| FA(30:4)                                                                             | 4.22672598 | 0.12083093 | BULBAR - RESPIRATORY ,                        | 480.3792 | 6.488029 |
| Unknown                                                                              | 4.12758799 | 0.12697133 | RESPIRATORY - SPINAL ,                        | 588.3987 | 7.505299 |
| Unknown                                                                              | 3.88535197 | 0.14331991 | BULBAR - SPINAL ,                             | 1686.395 | 9.222379 |
| PG(O-35:1) / PG(P-35:0) / PA(41:1) / PA(44:4) / PG(P-40:5) / PG(O-40:6)              | 3.86667317 | 0.14466471 | RESPIRATORY - SPINAL ,                        | 808.5749 | 9.245563 |
| PI(O-38:0) / PI(40:3) / PG(44:5) / PG(42:3)                                          | 3.6863354  | 0.15831514 | RESPIRATORY - SPINAL ,                        | 916.6062 | 6.861297 |
| PG(39:5) / PI(P-38:4) / PI(O-38:5) / PG(P-42:6) / PI(O-33:0) / PI(32:0) / PG(44:10)  | 3.57492159 | 0.16738465 | RESPIRATORY - SPINAL ,                        | 870.5592 | 8.838483 |
| Unknown                                                                              | 3.55905233 | 0.16871807 |                                               | 969.8068 | 10.10119 |
| Unknown                                                                              | 3.48584315 | 0.17500835 | RESPIRATORY - SPINAL ,                        | 997.8239 | 10.24688 |
| DG(44:3)                                                                             | 3.45177854 | 0.17801468 | RESPIRATORY - SPINAL ,                        | 790.6497 | 9.7514   |
| PE(O-32:0) / PE(O-34:3) / PE(P-34:2)                                                 | 3.43964915 | 0.17909756 |                                               | 699.5315 | 6.950889 |
| Unknown                                                                              | 3.41568047 | 0.18125684 | RESPIRATORY - SPINAL ,                        | 1239.778 | 7.309953 |

|                                             |            |            |                        |          |          |
|---------------------------------------------|------------|------------|------------------------|----------|----------|
| PI(O-38:0) / PI(40:3) / PG(44:5) / PG(42:3) | 3.36749482 | 0.18567686 | RESPIRATORY - SPINAL , | 916.6064 | 5.765081 |
| LysoPA(22:4) / PA(22:4)                     | 3.36004646 | 0.18636965 | RESPIRATORY - SPINAL , | 522.2477 | 5.653857 |
| Unknown                                     | 3.21717599 | 0.20017006 | RESPIRATORY - SPINAL , | 604.6773 | 7.435308 |
| Unknown                                     | 3.19565217 | 0.2023359  | RESPIRATORY - SPINAL , | 747.7172 | 7.411126 |

The table contains, by columns, the possible identifications of the metabolites, the test statistic, the p value, the significant differences between groups by post-hoc Dunn test and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.

## 8.2.9. Gender plasma positive

Table 22: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for plasma samples in positive ionization mode.

| Metabolite                  | estimate.dif | pval        | statistic.w | conf.int.1  | conf.int.2  | regulation (F vs M) | pcorr | mass     | RT        |
|-----------------------------|--------------|-------------|-------------|-------------|-------------|---------------------|-------|----------|-----------|
| Unknown                     | -0.10474798  | 0.004021918 | 16          | -2.52317357 | -0.02679507 | down                | 1     | 1853.511 | 9.8496    |
| Unknown                     | 0.05588241   | 0.011106442 | 99          | 0.0158909   | 0.105390787 | up                  | 1     | 1631.468 | 10.22913  |
| LysoPE(O-20:0) / PC(O-17:0) | -0.09257391  | 0.013769512 | 23          | -0.21128672 | -0.02084729 | down                | 1     | 477.3781 | 4.796853  |
| TG(58:6)                    | -1.02053152  | 0.013769512 | 23          | -1.90534281 | -0.17194244 | down                | 1     | 916.8175 | 10.10578  |
| Unknown                     | -0.81809733  | 0.013769512 | 23          | -1.85568701 | -0.16054856 | down                | 1     | 1466.375 | 9.286162  |
| LysoPE(16:1) / PC(O-13:1)   | -1.9347059   | 0.015063357 | 25          | -1.94667091 | -1.48E-05   | down                | 1     | 473.2456 | 0.8504499 |
| Unknown                     | -0.37819427  | 0.016899509 | 24          | -1.22303326 | -0.05811865 | down                | 1     | 390.2748 | 4.801351  |
| Unknown                     | -1.92881894  | 0.017727572 | 28.5        | -2.04380683 | -2.20E-05   | down                | 1     | 1789.518 | 10.10565  |
| Unknown                     | -0.0626228   | 0.021579993 | 25          | -2.18058097 | -0.00025562 | down                | 1     | 1973.765 | 10.51889  |
| Cer(d41:1)                  | 1.89264316   | 0.024051424 | 91          | 2.69E-05    | 1.994293898 | up                  | 1     | 635.61   | 8.621648  |
| DG(43:2) / CE(20:1)         | -0.16875666  | 0.027964706 | 26.5        | -2.19357058 | -0.01996734 | down                | 1     | 700.6364 | 8.047821  |
| TG(51:2)                    | 0.12117928   | 0.029960751 | 93          | 0.00651009  | 0.257696719 | up                  | 1     | 861.7795 | 10.22231  |
| Unknown                     | -0.06302786  | 0.035827949 | 28          | -0.12749861 | -0.00088897 | down                | 1     | 578.5692 | 8.419089  |
| Unknown                     | 0.87437443   | 0.035827949 | 92          | 0.07458762  | 1.754205083 | up                  | 1     | 1392.356 | 9.146473  |
| Unknown                     | -0.08505161  | 0.037658289 | 30          | -1.94020947 | -5.14E-05   | down                | 1     | 2081.543 | 9.6102    |
| Unknown                     | -1.92280364  | 0.03777625  | 29.5        | -1.95262588 | -3.17E-05   | down                | 1     | 1705.494 | 10.28216  |
| Unknown                     | -0.12395199  | 0.042570433 | 29          | -0.24814116 | -0.00379173 | down                | 1     | 705.6706 | 8.508555  |
| Unknown                     | -0.17999856  | 0.044074094 | 29          | -2.71421034 | -0.01447233 | down                | 1     | 1288.167 | 10.32494  |
| Unknown                     | 0.11462349   | 0.048078354 | 90          | 2.91E-05    | 2.101550925 | up                  | 1     | 1037.792 | 10.0707   |
| Unknown                     | -0.06127034  | 0.049450923 | 30          | -2.17123427 | 6.72E-05    | down                | 1     | 699.6114 | 9.687608  |
| Cer(d33:0)                  | -0.03145355  | 0.050324289 | 30          | -0.06035021 | 0.002886806 | down                | 1     | 547.493  | 8.278584  |

|                                                              |             |             |      |             |             |      |   |          |          |
|--------------------------------------------------------------|-------------|-------------|------|-------------|-------------|------|---|----------|----------|
| Unknown                                                      | -0.04891974 | 0.050324289 | 30   | -0.11493916 | 0.003390185 | down | 1 | 1589.42  | 9.001143 |
| PC(52:4)                                                     | -0.10116351 | 0.051615    | 31   | -1.99495309 | 1.19E-05    | down | 1 | 1009.878 | 10.05042 |
| CE(20:1) / DG(40:1)                                          | -1.89518612 | 0.052526995 | 31.5 | -1.94999713 | 7.76E-05    | down | 1 | 660.6318 | 8.050761 |
| PC(P-37:0) / PE(O-40:1) / PE(P-40:0) / PC(O-37:1) / TG(50:6) | -0.25312982 | 0.05617581  | 31   | -2.11126836 | 1.06E-05    | down | 1 | 804.6872 | 9.309961 |

Green fill: Confirmed by MS/MS spectra; Red font: Unexpected retention time.



## 8.2.10. Gender plasma negative

Table 23: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for plasma samples in negative ionization mode.

| Metabolite                                                                                                                   | estimate.dif | pval     | statistic.<br>w | conf.int.1 | conf.int.2 | regulation (F vs<br>M) | pcor<br>r | mass    | RT     |
|------------------------------------------------------------------------------------------------------------------------------|--------------|----------|-----------------|------------|------------|------------------------|-----------|---------|--------|
| TG(57:4)                                                                                                                     | 0.665521084  | 0.008956 | 99              | 0.1525774  | 1.208724   | up                     | 1         | 984.838 | 10.674 |
| PC(O-28:0)                                                                                                                   | 0.056987424  | 0.011980 | 98              | 0.0061051  | 2.105822   | up                     | 1         | 709.545 | 8.5718 |
| TG(65:9)                                                                                                                     | 0.042557318  | 0.034799 | 92              | 3.63E-05   | 2.102819   | up                     | 1         | 1062.83 | 10.320 |
| CE(16:0) / DG(39:3) / DG(36:0)                                                                                               | 0.158930662  | 0.037742 | 92              | 0.0004535  | 0.752130   | up                     | 1         | 660.551 | 8.7985 |
| DG(40:3) / CE(22:5)                                                                                                          | 0.037990603  | 0.046260 | 90              | 4.34E-05   | 2.030510   | up                     | 1         | 734.587 | 9.0717 |
| PG(34:2) / PI(P-33:1) / PI(O-33:2) / MGDG(34:6) / PG(P-40:6) / PA(40:7) / PG(39:7)                                           | 0.016581534  | 0.049450 | 90              | -3.95E-05  | 2.176875   | up                     | 1         | 806.521 | 8.569  |
| Unknown                                                                                                                      | 0.052416434  | 0.059135 | 89              | 0.0023503  | 0.118829   | up                     | 1         | 868.186 | 6.3891 |
| TG(52:3) / TG(57:8)                                                                                                          | -0.567969318 | 0.063184 | 32              | 2.0344928  | 2.70E-05   | down                   | 1         | 916.757 | 9.9268 |
| LysoPC(16:0) / PC(O-16:0) / LysoPC(P-19:1) /                                                                                 | -0.106098145 | 0.064778 | 31.5            | 0.1764248  | 0.021375   | down                   | 1         | 555.349 | 2.4500 |
| PC(32:0) / PE(35:0) / PC(P-38:4) / PC(O-38:5) / PE(P-38:1) / PC(O-35:2) / PE(O-38:2) / PC(P-35:1) / CerP(d44:1) / PC(O-36:1) | -1.86900675  | 0.066326 | 34              | 1.9942356  | 8.33E-05   | down                   | 1         | 793.582 | 7.0135 |
| PE(P-38:4) / PE(O-38:5) / PE(32:0) / PC(29:0) / PA(34:1) / PE(O-33:0) / PC(O-30:0)                                           | 0.084823307  | 0.068997 | 88              | 0.0078545  | 2.495262   | up                     | 1         | 751.546 | 7.4979 |
| Unknown                                                                                                                      | 0.037365214  | 0.069150 | 88              | 0.0016344  | 0.087484   | up                     | 1         | 1584.35 | 8.7065 |
| PI(P-39:0) / PI(O-39:1) / TG(56:13)                                                                                          | 0.029129575  | 0.074940 | 87.5            | 0.0027553  | 0.097351   | up                     | 1         | 952.654 | 9.8154 |

|                                                                                                          |              |                |      |                |                     |      |              |              |
|----------------------------------------------------------------------------------------------------------|--------------|----------------|------|----------------|---------------------|------|--------------|--------------|
| Unknown                                                                                                  | 0.320619077  | 0.080421<br>13 | 87   | 0.0675672<br>5 | 0.913303<br>24      | up   | 681.688<br>3 | 7.1193<br>52 |
| Unknown                                                                                                  | 0.466846103  | 0.080421<br>13 | 87   | 0.0524277      | -<br>1.278210<br>16 | up   | 633.558<br>7 | 8.0476<br>47 |
| PC(P-38:3) / PC(O-38:4) / PC(P-35:0) / PE(O-38:1) / PE(P-38:0) / PC(O-35:1) /                            | -0.008330419 | 0.083952<br>82 | 35.5 | 1.9542584<br>1 | 1.09E-05            | down | 795.595<br>9 | 8.6651<br>25 |
| Unknown                                                                                                  | 0.035604534  | 0.096548<br>98 | 84.5 | -3.20E-05      | 1.971946<br>81      | up   | 1539.35<br>2 | 8.8299<br>34 |
| Unknown                                                                                                  | 0.064827339  | 0.097881<br>29 | 85   | -6.62E-05      | 2.029212<br>23      | up   | 1371.14<br>4 | 11.065<br>45 |
| PC(38:6) / PS(42:5) / PE(41:6) / PS(O-40:2) / PS(P-40:1) / LacCer(d30:1) / PS(37:0)                      | 0.058812797  | 0.104230<br>33 | 85   | 0.0087981<br>9 | 2.319480<br>75      | up   | 865.581<br>7 | 6.5045<br>18 |
| Unknown                                                                                                  | 0.031187421  | 0.104230<br>33 | 85   | 0.0035402<br>9 | 2.318651<br>02      | up   | 1167.26<br>8 | 7.8681<br>48 |
| Unknown                                                                                                  | 0.074293896  | 0.104924<br>76 | 84.5 | -8.81E-05      | 2.000765<br>66      | up   | 947.877<br>1 | 10.056<br>32 |
| Unknown                                                                                                  | 0.022378909  | 0.108522<br>56 | 84   | -3.49E-05      | 1.991512<br>61      | up   | 1243.28<br>9 | 8.1121<br>18 |
| Unknown                                                                                                  | 0.034708473  | 0.108522<br>56 | 84   | -8.32E-05      | 1.969377<br>07      | up   | 969.855<br>1 | 9.8168<br>96 |
| Glc-Cholesterol(18:0) / TG(48:8) / PG(P-38:0) / PG(O-38:1) / PA(44:1)                                    | 6.40E-05     | 0.116093<br>99 | 78   | -5.25E-05      | 2.352742<br>4       | up   | 850.633<br>7 | 9.7075<br>44 |
| DG(29:2)                                                                                                 | 0.314001164  | 0.122911<br>14 | 84   | 0.1519667<br>4 | 0.865098<br>86      | up   | 582.439<br>9 | 7.5993<br>24 |
| PS(38:0) / PC(34:1) / PE(37:1) / PE(P-40:2) / PE(O-40:3) / PC(P-40:5) / PC(O-40:6) / PE(42:6) / PC(39:6) | -0.091059281 | 0.122911<br>14 | 36   | 0.1977854<br>4 | 0.026149<br>29      | down | 819.593<br>4 | 7.0908<br>18 |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.

## 8.2.11. Gender CSF positive

Table 24: Differential metabolites for gender comparison for CSF samples in positive ionization mode.

| Metabolite                                                                                | estimate.dif | pval       | statistic.w | conf.int.1 | conf.int.2  | regulation (F vs M) | pcorr | mass     | RT        |
|-------------------------------------------------------------------------------------------|--------------|------------|-------------|------------|-------------|---------------------|-------|----------|-----------|
| DAT(43:0) / TG(60:7) / TG(62:10)                                                          | 0.084988361  | 0.00389353 | 112         | 0.02190864 | 0.17348994  | up                  | 1     | 982.7862 | 10.25838  |
| Unknown                                                                                   | 0.119763522  | 0.00457059 | 111         | 0.03996319 | 2.358297    | up                  | 1     | 1561.438 | 9.086374  |
| Unknown                                                                                   | 1.978529906  | 0.00743635 | 103.5       | 3.28E-05   | 2.0131181   | up                  | 1     | 1232.409 | 9.728175  |
| Unknown                                                                                   | 0.916422219  | 0.00802401 | 107         | 0.29937326 | 1.68811164  | up                  | 1     | 1525.482 | 10.13816  |
| PS(44:7) / PI(P-38:3) / PI(O-38:4) / PS(42:4) / PI(37:4)                                  | 0.058677573  | 0.0086003  | 107.5       | 0.01412165 | 2.23544139  | up                  | 1     | 889.5999 | 0.8890356 |
| Unknown                                                                                   | 1.020948064  | 0.01209546 | 105         | 0.22770755 | 1.70880604  | up                  | 1     | 1514.071 | 6.544054  |
| Unknown                                                                                   | 1.896767924  | 0.01376336 | 103         | 6.60E-05   | 1.97223525  | up                  | 1     | 1130.304 | 9.826229  |
| Unknown                                                                                   | 0.064064035  | 0.0138211  | 105         | 0.0105519  | 2.33752024  | up                  | 1     | 1156.374 | 7.437876  |
| TG(46:0)                                                                                  | 0.086296846  | 0.01471069 | 104         | 0.03016378 | 0.13708323  | up                  | 1     | 795.7289 | 10.09959  |
| LysoPE(18:2)                                                                              | 0.054907425  | 0.01598186 | 101.5       | 2.75E-05   | 1.97074451  | up                  | 1     | 459.2704 | 0.8956843 |
| Unknown                                                                                   | 0.098509858  | 0.01668443 | 104         | 0.01713901 | 2.55794333  | up                  | 1     | 1115.323 | 7.825751  |
| Unknown                                                                                   | 0.062659464  | 0.01684693 | 104         | 0.0107411  | 0.15728027  | up                  | 1     | 1541.013 | 6.548581  |
| Unknown                                                                                   | 1.11809259   | 0.0177752  | 103         | 0.1961123  | 1.85270453  | up                  | 1     | 897.5925 | 0.8837297 |
| Unknown                                                                                   | 0.076832753  | 0.01970066 | 101         | 7.35E-05   | 1.97768779  | up                  | 1     | 1523.367 | 9.087706  |
| PS(33:0) / PI(28:0) / PG(35:5) / PS(35:3) / PE(37:6) / PC(34:6) / PE(O-38:6) / PE(P-38:5) | 2.053992726  | 0.02027426 | 96          | 5.44E-05   | 2.07794562  | up                  | 1     | 771.5044 | 0.8919445 |
| Unknown                                                                                   | -0.104041638 | 0.02062696 | 27.5        | -2.2319985 | -0.01198547 | down                | 1     | 979.2687 | 9.490966  |
| Unknown                                                                                   | 0.041615888  | 0.0213624  | 102         | 0.00727754 | 0.07024861  | up                  | 1     | 796.0227 | 0.8869166 |
| Unknown                                                                                   | 0.853596117  | 0.0213624  | 102         | 0.15320335 | 1.69815777  | up                  | 1     | 1696.617 | 10.44787  |
| Unknown                                                                                   | 0.975577097  | 0.0213624  | 102         | 0.19754567 | 1.66190598  | up                  | 1     | 1701.577 | 10.45027  |
| Unknown                                                                                   | 0.07606902   | 0.02431071 | 101.5       | 0.00723616 | 2.21427906  | up                  | 1     | 1530.473 | 10.13604  |
| Unknown                                                                                   | 0.088402064  | 0.02765396 | 101         | 0.00861304 | 0.1915088   | up                  | 1     | 2245.571 | 6.54494   |

|                                                                                   |              |            |       |             |            |      |   |          |           |
|-----------------------------------------------------------------------------------|--------------|------------|-------|-------------|------------|------|---|----------|-----------|
| Unknown                                                                           | 0.039058134  | 0.02955601 | 100.5 | 0.00324149  | 0.08197878 | up   | 1 | 564.6787 | 7.436999  |
| Unknown                                                                           | 1.115801934  | 0.03032867 | 100   | 0.06504651  | 1.85070158 | up   | 1 | 747.7176 | 7.411757  |
| Unknown                                                                           | 0.71782549   | 0.03032867 | 100   | 0.03002448  | 1.50666396 | up   | 1 | 317.2972 | 8.312488  |
| Unknown                                                                           | 0.744719397  | 0.03493495 | 99.5  | 0.02777879  | 1.80394088 | up   | 1 | 856.8275 | 10.44754  |
| PG(36:4) / PG(38:7) / PI(31:2) / PA(44:12) / PA(42:9) / PS(36:8) /                | -1.938807833 | 0.03852357 | 33    | -1.96068945 | -5.22E-05  | down | 1 | 792.4943 | 0.8866    |
| GPGro(36:1) / PG(34:1) / PA(40:6) / PA(38:3) / PS(32:2) / PS(O-33:2) / PS(P-33:1) | 0.633722623  | 0.04215841 | 98    | 0.07693108  | 1.41788007 | up   | 1 | 748.5248 | 6.546432  |
| Unknown                                                                           | 0.099255252  | 0.04215841 | 98    | 0.00612098  | 0.19194443 | up   | 1 | 565.6869 | 7.428514  |
| Unknown                                                                           | 0.755728156  | 0.04215841 | 98    | 0.02125814  | 1.66560211 | up   | 1 | 1169.057 | 8.309758  |
| Unknown                                                                           | 0.066157233  | 0.0436372  | 98    | 0.00095915  | 0.14835909 | up   | 1 | 1112.323 | 7.822035  |
| TG(51:1)                                                                          | 0.043833083  | 0.04379344 | 98    | 0.00149016  | 0.09594821 | up   | 1 | 828.7861 | 10.27988  |
| Unknown                                                                           | 0.035241239  | 0.04379344 | 98    | 0.00362114  | 0.07836066 | up   | 1 | 923.259  | 7.111153  |
| TG(55:7)                                                                          | 0.081002099  | 0.04399191 | 97    | 4.25E-05    | 2.00010757 | up   | 1 | 1819.447 | 9.555727  |
| Unknown                                                                           | 0.240810045  | 0.04469821 | 94    | -5.70E-05   | 2.05845126 | up   | 1 | 935.6064 | 0.8875    |
| Unknown                                                                           | 0.046015406  | 0.04600012 | 97    | 4.77E-05    | 2.04601862 | up   | 1 | 1010.355 | 9.241455  |
| Unknown                                                                           | 8.69E-05     | 0.04921097 | 90    | -7.92E-05   | 2.22787972 | up   | 1 | 854.0708 | 4.073917  |
| Unknown                                                                           | 0.848044403  | 0.04932932 | 97    | 0.00230211  | 1.45809132 | up   | 1 | 724.9761 | 0.8868378 |
| Unknown                                                                           | 0.963545673  | 0.04932932 | 97    | 0.00704275  | 1.78509505 | up   | 1 | 813.0289 | 0.8898107 |
| Unknown                                                                           | 0.608624753  | 0.04932932 | 97    | 0.02384599  | 1.6308166  | up   | 1 | 631.5901 | 8.303732  |
| TG(57:1)                                                                          | 0.082761696  | 0.04961468 | 97    | 0.00552889  | 2.33854335 | up   | 1 | 912.8838 | 10.4403   |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendal's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green fill: Confirmed by MS/MS spectra; Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time; Yellow font: LC-MS method not prepared to detect this metabolite.

## 8.2.12. Gender CSF negative

Table 25: Top 25 metabolites with lower p value obtained from the univariate tests for gender comparison for CSF samples in negative ionization mode.

| Metabolite                                                              | estimate.dif | pval       | statistic.w | conf.int.1  | conf.int.2  | regulation (F vs M) | pcorr | mass     | RT       |
|-------------------------------------------------------------------------|--------------|------------|-------------|-------------|-------------|---------------------|-------|----------|----------|
| PG(37:4) / PI(33:2) / PG(40:7) / PA(41:7)                               | -0.02010261  | 0.02551776 | 29          | -0.05084357 | -0.00221284 | down                | 1     | 820.4986 | 8.709084 |
| GlcAbetaCer(d36:1) / PE(O-32:0)                                         | 0.05564966   | 0.02635755 | 101         | 0.00371212  | 2.22594279  | up                  | 1     | 723.5647 | 9.054241 |
| Unknown                                                                 | -0.038418048 | 0.03755877 | 31          | -0.09780961 | -0.00071907 | down                | 1     | 1279.112 | 10.90932 |
| PA(33:3) / PA(30:0)                                                     | -0.072990582 | 0.04932932 | 33          | -0.15866043 | -0.0028077  | down                | 1     | 656.4266 | 6.386657 |
| PG(O-35:1) / PG(P-35:0) / PA(41:1) / PA(44:4) / PG(P-40:5) / PG(O-40:6) | -0.005860217 | 0.05331186 | 38          | -2.06274237 | 6.68E-05    | down                | 1     | 808.5749 | 9.245563 |
| Unknown                                                                 | -0.02586332  | 0.05738412 | 34.5        | -2.05746245 | 5.88E-05    | down                | 1     | 1299.103 | 11.46137 |
| Unknown                                                                 | -0.691168501 | 0.05745648 | 34          | -1.49553575 | 0.00245343  | down                | 1     | 979.602  | 5.764837 |
| Unknown                                                                 | -0.07593581  | 0.05925667 | 35          | -1.99949699 | 6.34E-06    | down                | 1     | 474.4066 | 7.05164  |
| GlcNAc-beta1-4Man-beta1-4Glc-beta-Cer(d42:2) / Ganglioside GA2 (d42:2)  | -0.631709858 | 0.06659581 | 35          | -1.41368213 | 0.01950911  | down                | 1     | 1174.769 | 7.250891 |
| TG(63:7) / TG(57:4) / TG(62:9)                                          | -0.033076564 | 0.06725525 | 35          | -0.06947213 | 0.00278613  | down                | 1     | 984.8363 | 10.47894 |
| PC(O-28:0)                                                              | -0.457964009 | 0.07685745 | 36          | -1.21743071 | 0.07567426  | down                | 1     | 709.5463 | 8.831676 |
| PI(41:1)                                                                | -0.51849982  | 0.07685745 | 36          | -1.33007459 | 0.12092566  | down                | 1     | 970.6174 | 6.861297 |
| PI(O-38:0) / PI(40:3) / PG(44:5) / PG(42:3)                             | -0.713191013 | 0.07685745 | 36          | -1.50031909 | 0.16193845  | down                | 1     | 916.6064 | 5.765081 |
| Unknown                                                                 | -0.65176767  | 0.07685745 | 36          | -1.42329458 | 0.01782497  | down                | 1     | 1223.048 | 10.69381 |
| GlcNAc-beta1-4Man-beta1-4Glc-beta-Cer(d42:1) / Ganglioside GA2 (d42:1)  | -0.467333931 | 0.08829386 | 37          | -1.54799527 | 0.08977861  | down                | 1     | 1176.784 | 7.353514 |
| PC(O-20:1)                                                              | -0.648764836 | 0.08829386 | 37          | -1.47056323 | 0.11937443  | down                | 1     | 537.4017 | 7.510513 |
| Unknown                                                                 | -0.655721724 | 0.08829386 | 37          | -1.52883959 | 0.18752458  | down                | 1     | 520.4107 | 7.482343 |
| Unknown                                                                 | 0.021947314  | 0.09395724 | 92.5        | -0.00878432 | 0.04763274  | up                  | 1     | 949.209  | 6.63558  |
| Cer(d30:1)                                                              | 0.513927225  | 0.10101165 | 92          | -0.17242445 | 1.32426045  | up                  | 1     | 1390.322 | 8.575784 |
| TG(60:9) / TG(55:1) / TG(54:6)                                          | -0.660210679 | 0.10101165 | 38          | -1.4419596  | 0.06947926  | down                | 1     | 938.7783 | 9.893568 |
| Unknown                                                                 | 0.059253855  | 0.10101165 | 92          | -0.02117889 | 0.19274768  | up                  | 1     | 1316.303 | 8.355529 |
| Unknown                                                                 | -0.506958465 | 0.10101165 | 38          | -1.55236716 | 0.2785165   | down                | 1     | 1251.068 | 10.80881 |

|                                     |              |            |      |             |            |      |   |          |          |
|-------------------------------------|--------------|------------|------|-------------|------------|------|---|----------|----------|
| 5beta.Cholestane.3alpha.24.25.triol | -0.035090533 | 0.10382704 | 38.5 | -0.0938533  | 0.0023002  | down | 1 | 1261.078 | 11.2889  |
| Dihydroxy-beta-cholestanone         | -0.628593876 | 0.11506329 | 39   | -1.51820113 | 0.40033202 | down | 1 | 418.3433 | 6.447216 |
| Unknown                             | -0.034737961 | 0.11506329 | 39   | -0.08868714 | 0.01845705 | down | 1 | 528.4534 | 8.67264  |
| Unknown                             | 0.699612247  | 0.11506329 | 91   | -0.14504232 | 1.52876384 | up   | 1 | 1538.36  | 8.934379 |
| Unknown                             | 0.558771711  | 0.11506329 | 91   | -0.15235573 | 1.36674228 | up   | 1 | 1612.377 | 9.085001 |

The table contains, by columns, the possible identifications of the metabolites, the statistical difference between groups, the p value, the kendall's w statistic, the confidence interval at 95% of confidence, the regulation, the Hochberg corrected p value and the mass and retention times. Green font: Common in PLA and CSF (in the same comparison); Red font: Unexpected retention time.