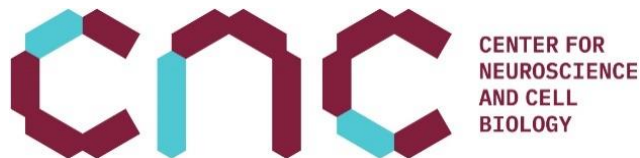1 2 9 0

UNIVERSIDADE Ð COIMBRA

Ana Alexandra Ferreira de Oliveira Martins

AN LC-MS UNTARGETED METABOLOMICS SCREENING IN CSF SAMPLES FROM PATIENTS WITH MILD COGNITIVE IMPAIRMENT

Junho de 2023

INDEX

LIST OF ABBREVIATIONS

ε4 – Epsilon4

2PG – 2-phosphoglyceric acid

7B2 – Neuroendocrine protein

ACh – Acetylcholine

AChE – Acetylcholinesterase

AChEl – Acetylcholinesterase inhibitors

ACN – Acetonitrile

Acyl-CoA – Acylcoenzyme A

AD – Alzheimer's disease

AICD – APP intracellular domain

ApoE – Apolipoprotein E

APP – Amyloid precursor protein

Aβ – β amyloid

BOLD – Blood oxygenation level dependent

BRICA – BRD4-interacting chromatin-remodeling complex-associated protein

C18 – Octadecylsilane bonded silica

CBF – Cerebral blood flow

CBV – Cerebral blood volume

CMGA – Chromogranin-A

CMRglc – Cerebral metabolic rate of glucose

CNS – Central nervous system

CSF – Cerebrospinal fluid

CSTN1 – Calsyntenin-1

CYTC – Cystatin-C

DC – Direct current

DDA – Data-Dependent Acquisition

DHAP – Dihydroxyacetone phosphate

DIA – Data-Independent Acquisition

DSM – Diagnostic and Statistical Manual of Mental Disorders

EGG – Electroencephalography
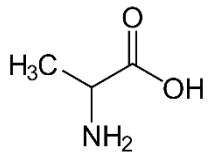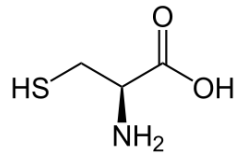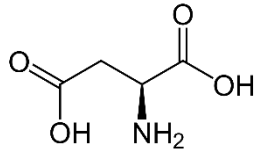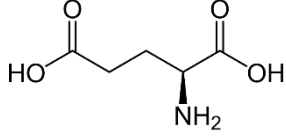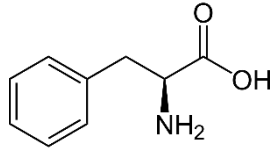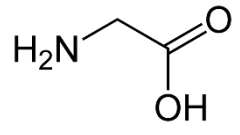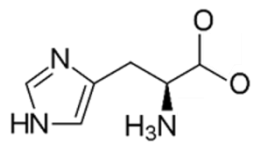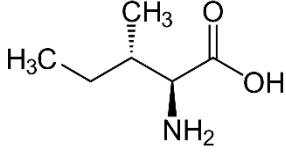
EOAD – Early-onset Alzheimer's disease
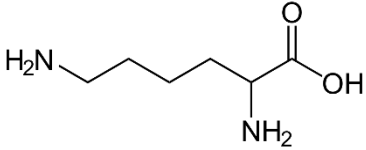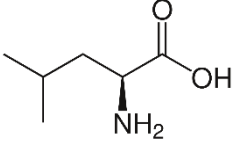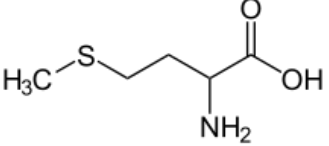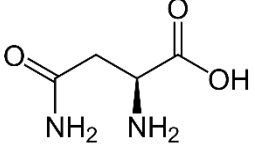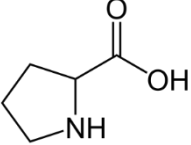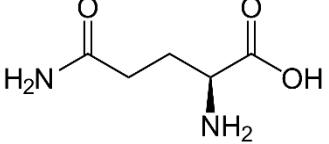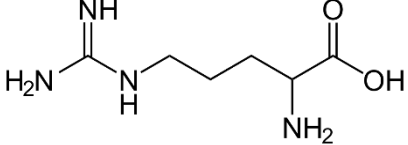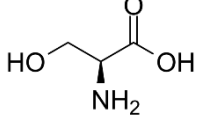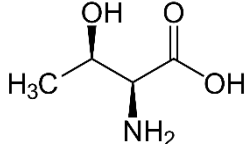
ESI – Electrospray ionization
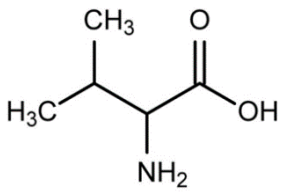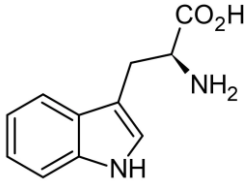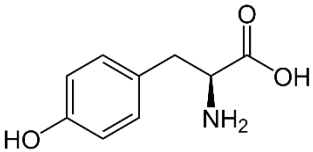
FA – Formic acid

FAD – Familial Alzheimer's disease

FC – Fold change

FDG – Fluorodeoxyglucose

FIBA – Fibrinogen alpha chain

FIBB – Fibrinogen beta chain

fMRI – Functional magnetic resonance imaging

fNIRS – Functional near-infrared spectroscopy

GC – Gas chromatography

H – Hydrogen

HbO – Oxyhemoglobin

HbR – Deoxyhemoglobin

HPLC – High-performance liquid chromatography

IDA – Information-Dependent Acquisition

ITM2B – Integral membrane protein 2B

KI26B – Kinesin-like protein KIF26B

KYN – Kynurenic acid

LC – Liquid chromatography

LOAD – Late-onset Alzheimer's disease

m/z – Mass-to-charge ratio

MALDI – Matrix-assisted laser desorption/ionization

MCI – Mild Cognitive Impairment

MMSE – Mini-Mental State Examination

MoCA – Montreal Cognitive Assessment

MRI – Magnetic resonance imaging

MS – Mass spectrometry

MS/MS – Tandem mass spectrometry

MTA – 5'-methylthioadenosine

NFTs – Neurofibrillary tangles

NIA-AA – National Institute on Aging Alzheimer's Association

NINCDS-ADRDA – National Institute of Neurologic and Communicative Disorders and Stroke-Alzheimer Disease and Related Disorders Association

NMDA – N-methyl-D-aspartate

NMR – Nuclear Magnetic Resonance

OECD – Organization for Economic Cooperation and Development

OSTP – Osteopontin

PCA – Principal Component Analysis

PCS1N – ProSAAS

PEP – Phosphoenolpyruvate

PET – Positron emission tomography

PGCB – Brevican core protein

PLS-DA - Partial Least Square-Discriminant Analysis

PSNE1 – Presenilin 1

PSNE2 – Presenilin 2

PSS – Phosphatidylserine synthase

p-Tau – Phosphorylated Tau

Q – Quadrupole

q2 – Collision cell

Q2 – Cross-validation correlation coefficient

QqQ – Triple quadrupole

QUIN – Quinolinic acid

R2 – Correlation coefficient

SAH – S-adenosylhomocysteine

SAM – S-adenosylmethionine

SD – Standard deviation

SF – Straight filaments

SFC – Supercritical fluid

SM – Sphingolipid

SWATH – Sequential Window Acquisition of All Theorical Fragment-Ion Spectra

TCA – Tricarboxylic acid

TOF – Time-of-flight

t-Tau – Total Tau

VGF – Neurosecretory protein

VIP – Variable importance in projection

Table of the most relevant amino acids

Amino acid	Structure	Letter code	Letter symbol
Alanine		Ala	A
Cysteine		Cys	C
Aspartic acid		Asp	D
Glutamic acid		Glu	E
Phenylalanine		Phe	F
Glycine		Gly	G
Histidine		His	H
Isoleucine		Ile	I

Lysine		Lys	K
Leucine		Leu	L
Methionine		Met	M
Asparagine		Asn	N
Proline		Prp	P
Glutamine		Glu	Q
Arginine		Arg	R
Serine		Ser	S
Threonine		Thr	T

Valine		Val	V
Tryptophan		Trp	W
Tyrosine		Tyr	Y

Resumo

A doença de Alzheimer (DA) é a forma mais comum de demência em todo o mundo e é caracterizada pela deterioração das funções cognitivas. A DA é caracterizada pelas placas senis e pelos emaranhados neurofibrilares. As placas senis são resultado de acumulações extracelulares de compostos na forma anormal da β-amilóide, $A\beta_{42}$. Os emaranhados neurofibrilares são compostos maioritariamente por filamentos helicoidais pareados que consistem em proteína Tau hiperfosforilada. A DA progride de forma lenta e tende a piorar gradualmente ao longo dos anos, sendo dividida em três fases principais da doença, a fase pré-clínica, o défice cognitivo ligeiro (DCL) devido à doença de Alzheimer e a fase de demência, dividida em leve, moderada e severa. O diagnóstico inclui exames neurológicos completos, avaliação do estado cognitivo, comportamental e funcional e estudos laboratoriais e de imagem. As abordagens metabolómicas representam uma importante ferramenta para a descoberta de biomarcadores para um diagnóstico precoce da DA. Este trabalho tem como objetivo fazer um estudo metabolómico não direcionado através de LC-MS/MS em amostras de líquido cefalorraquidiano (LCR) de pacientes com DCL de dois subgrupos, β-amilóide negativo e β-amilóide positivo, de forma a fornecer dados sobre alterações metabolómicas relacionadas com a patologia da DA. Assim, 40 amostras de LCR foram analisadas usando a espectrometria de massa e uma análise univariada e multivariada foi realizada para identificar as características mais interessantes. Destas características, 64 permitiram uma melhor separação entre os dois grupos. Considerando as características mais interessantes, foi possível identificar 34 delas. Uma análise mais abrangente foi realizada, e verificou-se que o metabolismo da tirosina, a biossíntese das hormonas esteroides e o metabolismo da cisteína e da metionina foram as vias identificadas como alteradas nas amostras de LCR. Os metabolitos 3,4-dihidroxi-L-fenilalanina e 19-oxoandrost-4-eno-3,17-diona encontraram-se aumentados no LCR do grupo Aβ negativo, enquanto o metabolito 5'-metiltioadenosina estava aumentado no grupo Aβ positivo. A característica identificada como 5'-metiltioadenosina teve o melhor poder discriminatório entre os dois grupos. No caso dos péptidos, nenhum deles apresentava um bom poder discriminatório. Neste trabalho, identificaram-se as características mais interessantes que podem fornecer informações sobre a fisiopatologia da doença e podem ser utilizadas para diagnóstico. No futuro, estas identificações devem ser validadas e uma comparação com o histórico clínico do paciente deve ser considerado de forma a esclarecer esses resultados.

Abstract

Alzheimer's disease (AD) is the most common form of dementia worldwide and is characterized by the deterioration of cognitive functions. AD is characterized by senile plaques and neurofibrillary tangles. Senile plaques result from extracellular accumulations of compounds in the abnormal form of β-amyloid, Aβ$_{42}$. Neurofibrillary tangles mostly comprise paired helical filaments consisting of hyperphosphorylated Tau protein. AD progresses slowly and tends to worsen gradually over the years, being divided into three main stages of the disease, the preclinical stage, mild cognitive impairment (MCI) due to AD and the dementia stage, divided into mild, moderate and severe. However, patients in the MCI stage may not progress to the dementia stage. Diagnosis includes complete neurological examinations, assessment of cognitive, behavioral, and functional status, and laboratory and imaging studies. Metabolomic approaches represent an important tool for the discovery of biomarkers for an early diagnosis of AD. This work aims to carry out an untargeted metabolomic study through LC-MS/MS in samples of cerebrospinal fluid (CSF) from patients with MCI of two subgroups, negative β-amyloid and positive β-amyloid, in order to provide data on alterations in the metabolome related to AD pathology. Thus, 40 CSF samples were analyzed using mass spectrometry and a univariate and multivariate analysis was performed to identify the most interesting features. Of these, about 64 allowed a better separation between the two groups. Considering the most interesting features, it was possible to identify 34 of them. A more comprehensive analysis was performed, and it was found that tyrosine metabolism, steroid hormone biosynthesis, and cysteine and methionine metabolism were the pathways identified as altered in CSF samples. The 3,4-dihydroxy-L-phenylalanine and 19-oxoandrost-4-ene-3,17-dione metabolites were increased in the CSF of the Aβ negative group, while the 5'-methylthioadenosine metabolite was increased in the Aβ positive group. The feature identified as 5'-methylthioadenosine had the best discriminatory capacity between the two groups. In the case of peptides, none of them had good differentiating ability. In this work, we identified the most interesting features that can provide information about the pathophysiology of the disease and can be used for diagnosis. In the future, these identifications should be validated, and a comparison with the patient's clinical history should be considered in order to clarify these results.

1 – Introduction

1.1 – Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurological disorder where the brain shrinks, brain atrophy occurs, leading to the death of brain cells [1]. According to *Centers for Disease and Healthy Aging*, AD leads to memory loss by affecting an individual's ability to carry out daily activities, for example the ability to carry on a conversation [2].

This disease is called Alzheimer's disease because in 1906, Dr. Alois Alzheimer, a German psychiatrist, found some changes in the brain tissue of a woman who had died of an unusual mental illness, where her symptoms included memory loss, language problems and unpredictable behavior. After she died, Dr. Alois examined her brain, and he found many abnormal clumps (now called amyloid plaques) and tangles of fibers (now called neurofibrillary tangles or Tau protein) [3].

According to data provided by *Alzheimer's Disease International*, in 2019, there were more than 55 million people in the world living with AD. This number is predicted to double every 20 years, reaching 78 million people in 2030 and 139 million in 2050 (Figure 1), with most of this increase occurring in developing countries [4].
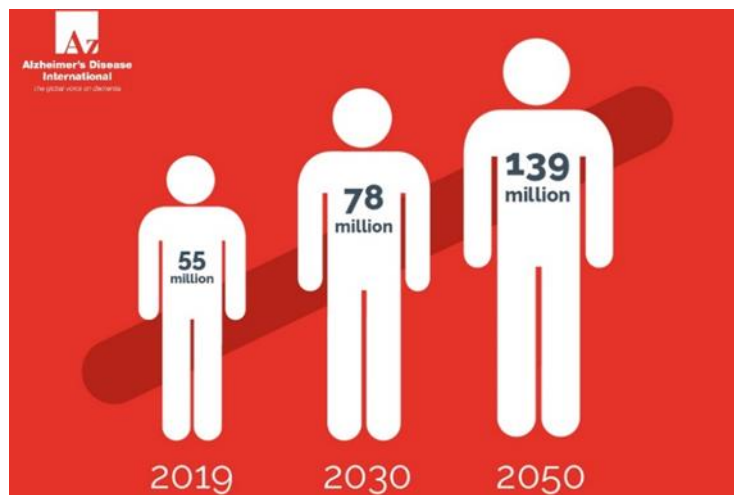
Figure 1 – Dementia statistics around the world: estimated growth in number of people with dementia between 2019 and 2050, adapted from *Alzheimer's Disease International* [4].

Currently, around 60% of people with dementia live in low- and middle-income countries, but by 2050 this will increase to 71% [4]. According to the *Organization for Economic Co-operation and Development* (OECD) report "*Health at a Glance 2017*",

Portugal is the fourth country in the OECD with the most cases of dementia, with about 19,9 cases per 1000 inhabitants, which is higher than the average [5].

According to data from the *Dementia in Europe Yearbook 2019*, in 2018 Portugal had about 193,516 cases of AD, which corresponds to a percentage of 1.88% of the general population. However, it is predicted that in 2050 there will be a doubling of these, reaching 346,905 cases, which is equivalent to 3.82% of the general population [6] (Figure 2 and 3).
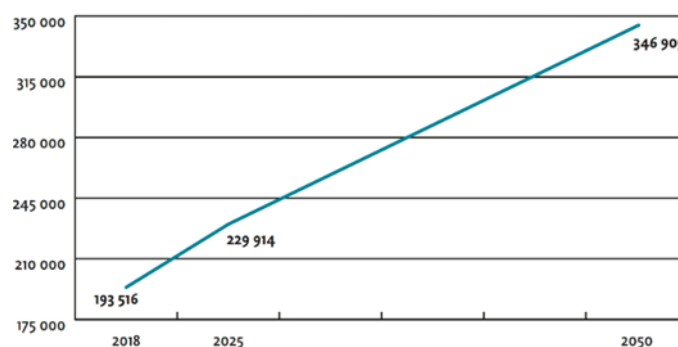
Figure 2 – Dementia statistics in Portugal: the number of people with dementia in Portugal from 2018 to 2050, adapted from *Dementia in Europe Yearbook 2019* [6].
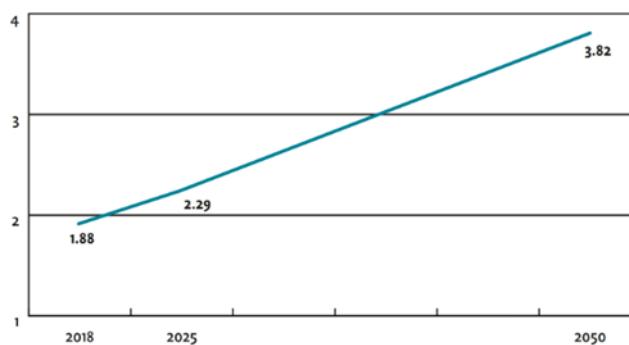
Figure 3 – Dementia statistics in Portugal in percentage: the number of people with dementia in Portugal as % of the total population from 2018 to 2050, adapted from *Dementia in Europe Yearbook 2019* [6].

Despite a small decrease in the Portuguese population over the years, it is expected that by 2050, the number of people with dementia in Portugal will double. One of the most important aspects for this change seems to be the significant increase in the number of people over 70 years old and, in particular, the age group over 85 years old, which will double between 2018 and 2050 [6].

Nowadays, it is very important and relevant to carry out studies on AD in order to improve the diagnosis of the disease and, consequently, the treatments applied.

2

1.1.1 – Alzheimer's etiology

In both developed and developing countries, Alzheimer's disease has a huge impact on individuals.

During the clinical manifestation of dementia disorders, there are some risk factors such as vascular disorders (smoking, high blood pressure, obesity in middle age, diabetes and cerebrovascular lesions) and some psychosocial factors (high schooling, active social activity, physical exercise and mentally stimulating activity) [7].

However, two of the most widely accepted risk factors for developing AD are age and genetic factors. Thus, there are three main steps [8] for an individual to progress from normal cognitive function to a condition in which the full range of clinical symptoms is expressed: an initial injury, a chronic neuroinflammatory response and a change in behavior.

Despite this, the issue of advanced age is one of the risk factors with the greatest impact on cognitive decline and AD. The probability of an individual developing AD increases by about 3% in individuals aged around 65 years and by about 30% at 85 years of age. Although the incidence of AD under 65 years is less certain, there are studies that suggest that this age group is responsible for about 3% of AD cases [9].

As age advances, certain changes occur in the brain that damage neurons and affect other types of brain cells, contributing to the development of Alzheimer's disease. Some of these changes include: atrophy of certain parts of the brain, inflammation, vascular damage, production of free radicals, and even reduced energy production inside cells [10].

Although a large proportion of individuals with AD manifest the same symptoms (memory loss, confusion, problems with daily tasks), this pathology can be divided into two types, depending on age and genetic factors, including early-onset Alzheimer's (EOAD) and late-onset Alzheimer's (LOAD). EOAD usually affects individuals aged between 40 and 50 years, representing only 5% of AD cases. However, LOAD affects people over the age of 65 and accounts for the majority of cases (> 95%) [11].

Late-onset Alzheimer's disease is quite complex as it results from a combination of environmental and genetic influences. The only well-established genetic risk factor for AD is the presence of the epsilon4 (ε4) allele on apolipoprotein E (ApoE). It is important

to note that a single ε4 allele increases the risk of AD 2 to 3 times, and in homozygotes the risk increases up to 12 times [12]. ApoE is the main apolipoprotein in the brain and is extremely important in cholesterol homeostasis.

One of the mechanisms that leads researchers to think that ApoE is involved in the pathogenesis of AD is related to its ability to bind β-amyloid (Aβ), eventually converting A-beta from monomeric and non-toxic forms into oligomers and fibrils. This is visible through the presence of ApoE in senile plaques and intracellular neurofibrillary tangles (NFTs) [12].

In genetic terms, AD is related to familial cases with mendelian inheritance and is defined by rare mutations in one of the three main genes, which are amyloid precursor protein (APP), presenilin 1 (PSEN1) or presenilin 2 (PSEN2). However, PSEN1 is the most common cause of FAD.

APP is a type I membrane protein with a large extracellular domain and a short cytoplasmic region. Thus, there are two cleavage moments that are required to release Aβ from APP, one in the extracellular domain (through a β-secretase cleavage) and another in the transmembrane region (through a γ-secretase cleavage). However, a large part of the mutations affects the activity of secretases, causing an increase in the production of an amyloidogenic alloform of the Aβ peptide, Aβ$_{42}$. This becomes toxic to neurons and can cause Aβ deposition in vessel walls [13].

Thus, although there are no known concrete causes of AD, it is essential to study the risk factors associated with the disease and who has a greater capacity to develop it, in order to prevent this dementia [13].

1.2 – Clinical characteristics of Alzheimer's Disease

Alzheimer's disease is the most frequent cause of dementia, with a median survival duration of between 5 and 8 years after clinical diagnosis [14].
However, during the progression of AD, brain changes that are imperceptible to the patient occur, causing memory problems and, eventually, physical disability. This is called AD continuum progression. [15].

In this continuum, three major phases are present: pre-clinical AD, mild cognitive impairment (MCI) due to Alzheimer's disease and dementia due to Alzheimer's disease, also called Alzheimer's dementia, being divided into mild, moderate and severe dementia [15].

4

The AD continuum is known to start with pre-clinical Alzheimer's disease (when there are no symptoms) and end with severe Alzheimer's dementia (when there are critical symptoms). However, the length of time that occur on each part of the continuum varies, being influenced by age, genetics, biological sex, and other factors [15].

1.2.1 – Preclinical Alzheimer's disease

At this stage individuals, despite not developing symptoms such as memory loss, have some measurable brain changes that indicate the first signs of AD (biomarkers). Some of these biomarkers include abnormal levels of β-amyloid, as shown in positron emission tomography (PET) and cerebrospinal fluid (CSF) analysis, various alterations in Tau protein in CSF and plasma, and even decreased glucose metabolism [15].

1.2.2 – Mild cognitive impairment due to Alzheimer's disease (MCI due to AD)

At this point, patients with MCI due to AD have biomarkers that demonstrate brain changes, such as memory, language and thinking problems, not interfering with the individual's ability to perform daily activities. About 15% of MCI due to AD patients develop dementia after 2 years. However, some individuals with MCI due to AD can return to normal cognition or do not experience further cognitive decline [15].

1.2.3 – Dementia due to Alzheimer's disease (AD)

In this phase, there are symptoms that impair the patient's daily tasks, such as problems with memory, language, thinking or behavior. There are several types of symptoms that change over time, and these reflect the degree of damage to neurons in different parts of the brain. As mentioned before, this stage can be divided into mild, moderate and severe dementia [15]:

• mild Alzheimer's dementia: at this stage patients are able to live mostly independently but will likely need assistance with some activities to maximize independence and stay safe. Handling money and paying bills can be especially challenging, and they may need more time to complete common daily tasks.

• moderate Alzheimer's dementia: this phase is the longest and where individuals experience more memory and language problems. They are more likely to be confused and have more difficulty completing tasks such as bathing and dressing. People may experience personality and behavior changes and also begin to have trouble recognizing loved ones.

- severe Alzheimer's dementia: in this case the ability of individuals to communicate verbally is very low. Due to damage to the areas of the brain involved in movement, individuals become bedridden. Being bedridden makes them vulnerable to physical complications including blood clots, skin infections and sepsis, which triggers inflammation throughout the body that can result in organ failure.

1.3 – Pathology of Alzheimer's Disease

A study in the literature [16] tells us that there are two genetically distinct types of AD, namely:

- familial Alzheimer's Disease (FAD): which is usually characterized by clinical onset before age 60 years and Mendelian inheritance (represents <1% of AD cases).

- sporadic Alzheimer's Disease (SAD): which usually has clinical onset after age 60 years and does not exhibit a consistent pattern of inheritance.

According to Julia TCW and Alison M. Goate [16], AD is characterized neuropathologically by neuronal cell loss, extracellular neuritic plaques composed of β-amyloid (Aβ plaques), and intracellular neurofibrillary tangles composed of hyperphosphorylated Tau protein (NFTs).

A study carried out by Sneham Tiwari and collaborators [17] says that Aβ plaques begin their development in the basal, temporal, and orbitofrontal neocortex regions of the brain and, at a later stage, progress throughout the neocortex, hippocampus, amygdala, diencephalon, and basal ganglia. In more critical situations, these plaques are visible throughout the midbrain, lower brainstem and cerebellar cortex, and their concentration will cause the formation of NFTs, which are found in the locus coeruleus and in the transentorhinal and entorhinal areas of the brain.

It is important to note that there are two main types of Aβ polymers ($A\beta_{40}$ and $A\beta_{42}$), and these have a direct role in plaque formation and induced neurotoxicity. The polymer $A\beta_{40}$ is abundant and less neurotoxic than $A\beta_{42}$, which is less abundant and severely neurotoxic.

When $A\beta_{40}/A\beta_{42}$ aggregation occurs, ion channels are blocked, calcium homeostasis is altered, mitochondrial oxidative stress increases, and energy metabolism and glucose

6

regulation decrease, which contributes to the deterioration of neuronal health and consequently to neuronal cell death [17].

Still in relation to the Aβ plaques, these are protected from the processing of the β-amyloid precursor protein (APP). APPs are proteolytically cleaved by two distinct pathways [16].

In the non-amyloidogenic processing pathway, APP is cleaved within the Aβ domain by α-secretase, with the formation of a large soluble ectodomain (sAPPα) and a C-terminal membrane-associated fragment of 83 residues (C83). Subsequent cleavage of C83 by γ-secretase leads to the formation of P3 and the intracellular APP domain (AICD) [13, 18]

In the amyloidogenic pathway, APP is cleaved at the N-terminus of the Aβ domain by β-secretase or BACE, a membrane-bound protease, resulting in the generation of a soluble ectodomain (sAPPβ) and a 99-residue, retained in the C-terminal membrane fragment (C99). Subsequently, γ-secretase, a membrane-embedded complex with presenilin as a catalytic component, cleaves C99 to release Aβ and AICD peptides (Figure 4). As the γ-secretase cleavage site is promiscuous, it generates Aβ peptides with different C-terminals, including $A\beta_{1-40}$ ($A\beta_{40}$), $A\beta_{1-42}$ ($A\beta_{42}$) and other smaller species [13, 18].
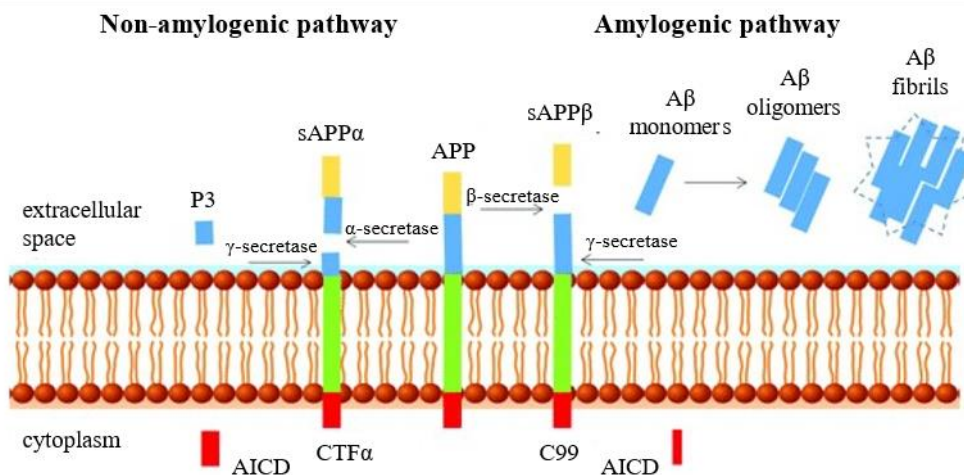
Figure 4 – Proteolytic processing of amyloid precursor protein: proteolytic processing of APP within the non-amyloidogenic and amyloidogenic pathways, adapted from Isabelle Sumner et al [19].

Regarding NFTs, they are abnormal accumulations of a protein called Tau that accumulates inside neurons [20]. Under physiological conditions, Tau protein is a highly soluble and natively unfolded protein that interacts with tubulin and promotes its

assembly into microtubules, facilitating the stabilization of its structure, facilitating the transport of nutrients and molecules from the cell body to the axon and dendrites [20, 21]. However, in AD abnormal chemical changes occur that cause Tau to separate from microtubules and join other Tau molecules, forming threads that will join to form tangles inside neurons. These tangles block the neuron's transport system, which impairs synaptic communication between neurons (Figure 5) [21].

According to some studies in the literature [21-23], there is evidence that indicates additional functions for Tau. One is the fact that tau phosphorylation allows neurons to escape acute apoptotic death by stabilizing ß-catenin. In addition, Tau plays an essential role in balancing microtubule-dependent axonal transport of organelles and biomolecules, modulating kinesin-driven anterograde transport and dynein-driven retrograde transport.

However, AD is clinically characterized by macroscopic and microscopic changes in the brain.
Some of the macroscopic features affect locations involved in thinking and memory function, causing changes in the cerebral cortex (causing a strong reduction in volume) and even an increase in the ventricle [24].
The microscopic changes that are most frequently found in AD include alterations in senile plaque, neurofibrillary tangles, neuronal loss and also synaptic alteration [24].

These changes caused in the brains of Alzheimer's patients are very important for a later diagnosis.
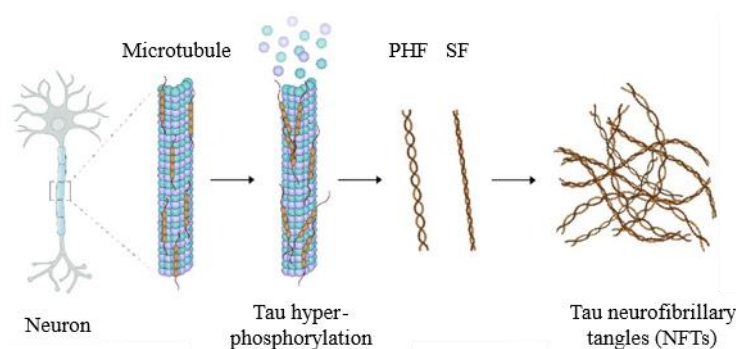
Figure 5 – Formation of neurofibrillary tangles (NFTs): Tau proteins self-assemble to form paired loosely interwoven helical filaments (PHF) and tightly wound straight filaments (SF), forming NFTs, adapted from C. Jie et al [25].

1.4 – Alzheimer's diagnosis

Accurate diagnosis of dementia is essential to provide appropriate treatment and patient counseling, however distinguishing dementia from delirium can be difficult.

Dementia can be grouped into two categories, these being dementia with no prominent motor signs and dementia with prominent motor signs. In this case, this work studies AD that is included in dementias without prominent motor signs [26]. Diagnostic steps include clinical history, neuropsychological tests, mental status assessment, as well as laboratory and imaging studies. According to a study in the literature [26] genetic tests can still be performed, but their use is controversial and raises complex ethical issues.

In 2011, the *National Institute on Aging and Alzheimer's Association* (NIA-AA) formed a set of diagnostic guidelines for the pre-clinical, symptomatic (MCI) or clinical (dementia) phases of AD [27].
Briefly in the pre-clinical phase, some brain changes occur, including the accumulation of β-amyloid and other changes in nerve cells, however no clinical symptoms are evident [28].
In a phase of MCI some problems are visible in terms of thinking and memory, not completely interfering with the independence of patients and patients with MCI may or may not progress to AD [28].
The clinical stage, also called Alzheimer's disease, is the final stage of the disease where symptoms of memory loss, difficulties in carrying on a conversation and visual problems are notorious, impairing a person's ability to act independently [28].

The first step in doing a cognitive assessment involves taking a cognitive screening test, such as the Mini-Mental State Examination (MMSE) [29]. However, there are some studies [30-32] that prove that this test has not been suitable for the detection of MCI and clinical signs of dementia. For this reason, a new screening test was created called The Montreal Cognitive Assessment (MoCA) [33].

The MoCA test is a very important screening tool because it has a high ability to discriminate between normal cognitive function, MCI and early-onset dementia. This test can be performed between 10 and 15 minutes and has a great advantage, which is the sensitivity in the detection of MCI (about 90%) and mild AD (about 100%) [34].

Some studies in the literature [35-37] have evaluated the ability of cognitive screening between MoCA and MMSE for the detection of dementia, concluding that the MoCA test is more useful in relation to the MMSE.

On the other hand, other studies [38, 39] indicate that the MoCA is not superior to the MMSE when evaluating patients with MCI.

This means that there is no consensus among researchers on which tool is most useful for detecting a decline in cognitive function. Nevertheless, when making a diagnosis of AD, it is always important to consider potential risk factors, such as age, sex, educational and occupational level, psychological well-being, physical exercise, social engagement, diet and history of chronic diseases [40].

These tests are useful for monitoring the evolution of cognitive impairment and also for evaluating the effectiveness of anti-dementia drugs or other intervention strategies. The MoCA is useful for mild stages of cognitive impairment (including MCI and mild AD dementia), with sensitivity values of 81% and specificity of 77% for detection, while the MMSE assessment is superior for more advanced stages [41].

1.4.1 – Neuroimaging

As previously mentioned it is important that there are pathological markers to be able to make an early diagnosis [42]. One of the most interesting markers is neuroimaging.

This makes it possible to obtain time-sensitive information, even before the onset of the cascade of neurodegenerative processes, allowing the classification of patients in different stages of the disease. According to Table 1, several imaging methods are currently used to delineate the characteristics of AD-related groups, including magnetic resonance imaging (MRI), positron emission tomography (PET), functional near-infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) [43].

Table 1 – Imaging methods that are used to delineate the characteristics of AD-related groups, including (MRI), (PET), (fNIRS), (fMRI) and (EEG).

Imaging method	Principle	Main findings
Structural MRI	MR with hydrogen	Cerebral atrophy, ventricular enlargement
FDG-PET	Radioluminescence of FDG, amyloid, tau …	Reduced cerebral glucose metabolism
fMRI	MR for hemodynamics	Hyper- and hypo-activation in the task-related regions
fNIRS	NIR light for hemodynamics	Reduction in HbO concentration
EEG	Electrical signal of brain	Altered functional connectivity pattern, slowing, decrease in complexity, alterations in microstate

1.4.1.1 – Magnetic resonance imaging (MRI)

One of the most used imaging techniques to visualize the internal organs of the body and the anatomy of the brain is MRI.

This technique uses strong magnets and low-energy radiofrequency signals to gather information from atomic nuclei inside the body (Figure 6). One atom that is present in all organic compounds is hydrogen (H), so it is well suited for imaging because of its abundance in the human body [44].

No external magnetic field   External magnetic field

External pulse expose: Protons relax:
the phase shift 90°   release signal
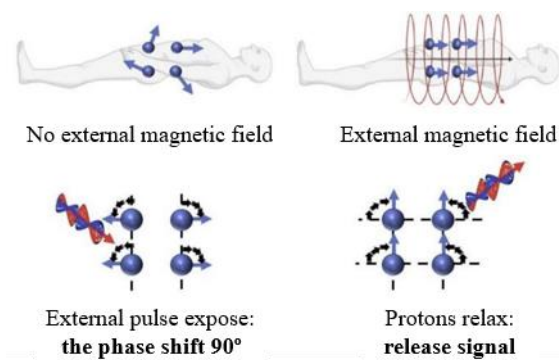
Figure 6 – Magnetic resonance imaging (MRI): when an individual is exposed to a strong external magnetic field, the hydrogen nuclei align parallel or anti-parallel to the external magnetic field. Without a magnetic field, the magnetic moments of the nuclei are randomly distributed, and the net magnetization factor is zero, adapted from JunHyun Kim et al [43].

To obtain a signal, patients need to be exposed to a pulse of radiofrequency energy (resonance frequency), allowing the protons in the H nucleus to reach a high-energy state causing structural changes such as dendritic and neuronal losses in the brain. MRI cannot clearly detect H nucleus inside the atrophied regions, so it is possible to evaluate disease progression [43].

MRI makes it possible to quantitatively measure the morphological changes that occur in the brain structure. On MRI it is possible to see that the first spots of atrophy typically occur in the entorhinal cortex in the medial temporal lobe, followed by the hippocampus and adjacent structures [44]. However, not only volumetric reductions occur throughout the brain or in the hippocampal area.

There are other changes that can be quantitatively measured by MRI such as cortical thickness or ventricular enlargement in correlation with the accumulation of senile plaques and cortical neurofibrillary tangles [45]. Thus, the ability to quantitatively measure all these changes allows for an increase in methodological development in the prediction and classification of AD [46].

1.4.1.2 – Positron emission tomography (PET)

Regarding positron emission (PET), this technique measures metabolic changes through the use of different radioactive tracers, depending on the intended target. PET imaging is a coincident detection of gamma rays that are released during positron annihilation moments to biological molecules such as glucose, peptides and proteins [47].

18F-fluorodeoxyglucose (FDG) PET performs better than structural MRI in predicting conversion to AD and is therefore more useful for early diagnosis. The neurodegeneration biomarker detected by FDG-PET in patients with dementia is brain hypometabolism [48] with changes in FDG uptake in AD patients that correlate with cognitive decline. AD patients show a much greater decline in glucose metabolism and further reductions over 1 year compared to healthy controls.

Figure 7 shows the significant deviations in cortical metabolism in AD patients compared to healthy controls.
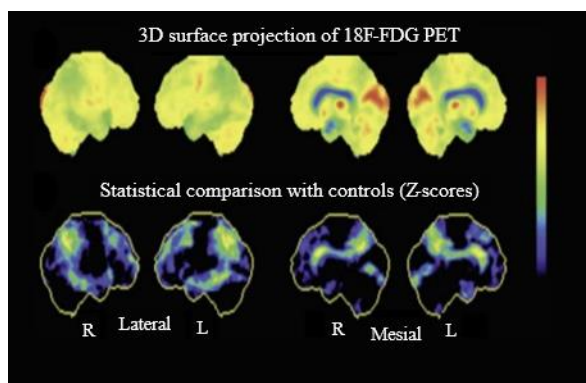
Figure 7 – FDG-PET imaging: 3D surface projection of 18F-FDG PET findings in AD (top row) and a statistical comparison with a healthy control population; green or yellow shows high deviation and black or blue shows no or low deviation (bottom row), adapted from JunHyun Kim et al [43].

1.4.1.3 – Functional magnetic resonance imaging (fMRI)

Regarding functional magnetic resonance imaging (fMRI) methods, they allow the visualization of neural activation indirectly, detecting changes in the hemodynamics of cerebral vessels.

This response is a widely used method to confirm activation in brain regions depending on the blood oxygenation level (BOLD) based on the paramagnetic properties of deoxyhemoglobin in the blood and intact neurovascular coupling [49]. The increase in oxygen consumption in the brain follows a wave of neuronal activity that is associated with an increase in cerebral blood flow (CBF) and blood volume (CBV) to the activated area. Oxygenated blood flow in the activated regions of the cortex increases the MRI signal and is then recorded on fMRI (Figure 8).

Therefore, fMRI images investigate the correlation between the BOLD signal and brain regions comparing the cognitive condition with a control condition [50]. So, fMRI techniques can monitor AD-related brain dysfunction.
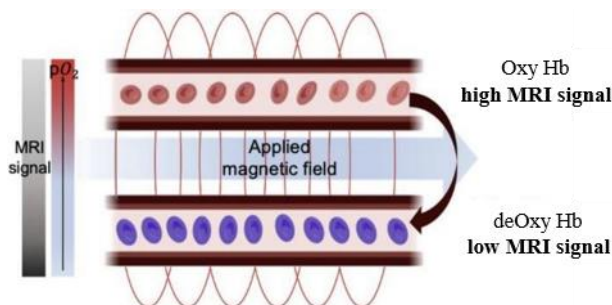
Figure 8 – Functional magnetic resonance imaging (fMRI): elevated oxygen consumption in the brain is associated with increased cerebral blood flow (CBF) and blood volume (CBV), resulting in an amplified MRI signal. Deoxygenated blood exhibits low magnetic susceptibility, causing decay of the MRI, which can be captured by fMRI to effectively track and monitor brain dysfunction associated AD, adapted from JunHyun Kim et al [43].

1.4.1.4 – Functional near-infrared spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) is a non-invasive optical imaging technique that monitors blood vessel hemodynamics, like fMRI, using light from the NIR window (700-900 nm).

As NIR light is less scattered and absorbed by a few biological chromophores, it easily passes through biological tissues reaching the cranium, eventually being absorbed by oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) with different absorption spectra [51].

Changes in oxygenation-deoxygenation states are calculated by the amount of light transmitted through the tissue (Figure 9) using the modified Beer-Lambert law. The sum reflects the total blood volume during brain activity. As more blood flows to activated brain regions, NIRS can be used to detect neural dysfunction when brain shrinkage occurs through neurodegeneration [52].

Synchronization of fNIRS signals is reduced in individuals with mild AD compared to normal aging controls, with loss of regularity inside the brain network with disease progression [53].

It can be concluded that fNIRS is a new modality that has some benefits, for example it is not an invasive method, and it is safe and relatively low-cost.
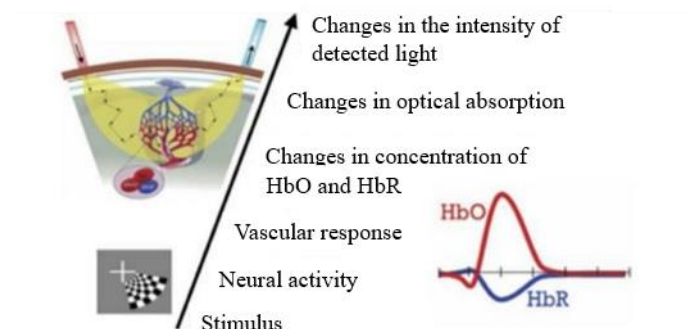
Figure 9 – Functional near-infrared spectroscopy (fNIRS): the changes between oxygenated and deoxygenated states are determined by measuring the amount of light that passes through the tissue. This is done using a modified version of the Lambert-Beer Law and the combined measurements indicate the total volume of blood present during brain activity. By observing changes in blood flow to activated brain areas, NIRS can be used to identify neural dysfunction caused by brain shrinkage due to neurodegeneration, adapted from JunHyun Kim et al [43].

1.4.1.5 – Electroencephalography (EEG)

The electroencephalography (EEG) technique allows recording of the joint electrical activity created by the brain between millions of active neurons using electrodes on the surface of the scalp or intracranially (Figure 10).

In this way, it appears that EEG activity reflects the more or less synchronous activation of a large population of neurons, more precisely, their postsynaptic activity. This means that if a population of neurons is spatially aligned and has synchronous activity, the resulting superimposed electric field will be detected by electrodes on the scalp surface [54].

This technique allows the measured electrical activity to show the functional changes that occur in the cerebral cortex, and which abnormalities in this activity can be used to detect the functional deficits caused by neurodegeneration in AD.
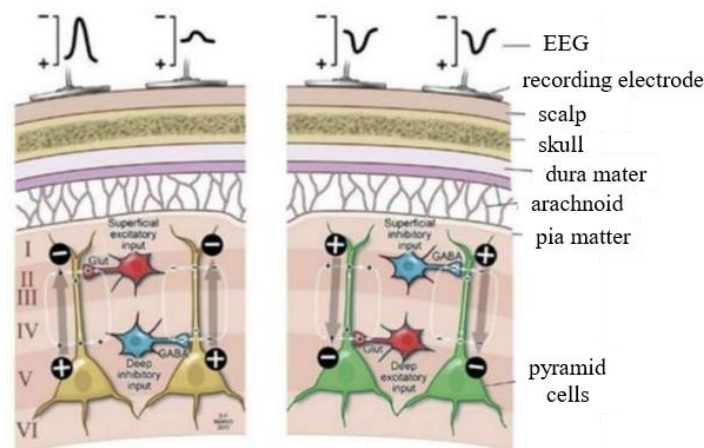
Figure 10 – Electroencephalography (EEG): EEG records the collective electrical activity generated by millions of active neurons in the brain, using electrodes placed either on the surface of the scalp or intracranially. The EEG activity reflects the level of synchronous activation among a substantial population of neurons, along with their postsynaptic activity, adapted from JunHyun Kim et al [43].

An easy way to record the EEG signal in AD patients is in the resting state, as it avoids any discomfort in the patient or the possibilities of not being able to complete the necessary tasks. The most common effects include a gradual loss of complexity, connectivity changes, and changes in microstate complexity [55].

The EEG technique allows diagnosing AD through several frequencies, allowing to discriminate healthy people from AD patients as it can seen in Table 2 [56].

Table 2 – Frequencies of the EEG technique that allow diagnosing Alzheimer's.

Range	Band	Detect
0.1 – 4 Hz	delta	thalamus region and cortex
4 – 8 Hz	teta	cognition, behavior and memory in the hippocampus
8 – 12 Hz	alfa	thalamus region
12 – 30 Hz	beta	motor cortex and anxious thinking
30 – 100 Hz	gama	premotor, parietal and temporal cortical regions

This EEG technique is useful. One of the great advantages is that it provides neural information on a millisecond time scale. However, it has some disadvantages, the main one being related to the distance between the electrodes and the real source of neuronal activity, creating a low-pass filtering in the source signal, and limiting the resolution, making it more difficult to accurately describe the neuronal processes [57].

In conclusion, although the EEG method provides information on possible spatiotemporal distribution patterns of disease progression, subtle EEG abnormalities are difficult to detect in the early stages of AD and are not distinct from markers of other neurodegenerative diseases.

1.5 – Cerebrospinal fluid (CSF) biomarkers

The cerebrospinal fluid (CSF) biomarkers consists of a concentration of $A\beta_{1-42}$ and Tau, total Tau (t-Tau) and phosphorylated Tau (p-Tau), which facilitates the identification of patients with MCI at different risk levels of progression to AD [58-60].

According to a study in the literature [61], when a combination of a decrease in $A\beta$ concentrations in the CSF and an increase in Tau proteins (t-Tau and p-Tau) occurs, it is possible to discriminate patients with AD of controls with a sensitivity and specificity of more than 85%.

However, the combination of these three biomarkers ($A\beta_{1-42}$, t-Tau and p-Tau) is especially sensitive and specific for identifying AD patients among those with subjective memory complaints (defined as people who come to the clinic concerned about their memory function but there is no sign of memory decline). The predictive value of these biomarkers in identifying patients who will progress to AD among a group of patients at

the MCI stage is also quite good, with sensitivity ranging from 81 - 95% and specificity ranging from 72 - 95% [62-64].

A study in the literature [65] carried out in cognitively normal individuals showed that altered baseline levels of $A\beta_{1-42}$ and p-Tau preceded the onset of clinical symptoms of MCI or dementia. This means that the altered levels of biomarkers in the CSF appear to reflect the long pre-clinical AD period, during which preclinical pathological changes occur in the brains of AD patients.

Thus, according to a study in the literature [66], the dementia phase is characterized by the presence of reduced levels of $A\beta_{1-42}$ and increased levels of t-Tau and p-Tau (Figure 11).

These biomarkers help to differentiate AD from important differential diagnoses such as depression and Parkinson's disease, with p-Tau levels also providing substantial assistance in differentiating from other dementias such as frontotemporal lobar dementia and Lewy body dementia [67].
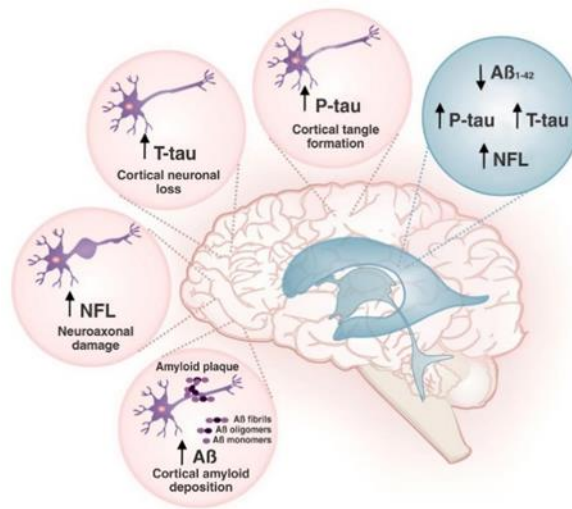
Figure 11 – Cerebrospinal fluid (CSF) biomarkers for Alzheimer's disease: decreased levels of $A\beta$ plaques and increased levels of Tau protein (p-Tau and t-Tau) help distinguish Alzheimer's disease from other diagnoses, adapted from Tatiana Barichello et al [66].

$A\beta_{1-42}$ aggregation sometimes occurs in senile plaques, causing levels of this peptide in AD patients to drop. On the other hand, there are other biomarkers that can explain the increased presence of $A\beta$ in patients, for example BACE1 [68].

According to a study carried out by Lan Li and his collaborators [68], the BACE1 protein, under pathological conditions, cuts APP at a specific point so that the neurotoxic $A\beta_{42}$ is released extracellularly. One of the first pathological changes in AD occurs when there is an abnormal expression or function of BACE1, so it can serve as a biomarker in

early AD. On the other hand, the concentration of Tau protein in CSF is related to the intensity of neuronal degeneration, while the concentration of p-Tau reflects the pathogenesis of NFTs. One of the alternatives to this protein is the measurement of the α-synuclein protein because it induces hyperphosphorylation and aggregation of Tau proteins [69].

It is important to mention that cerebrospinal fluid is considered the most suitable biological fluid for the study of neurodegenerative diseases. CSF is produced in the central nervous system (CNS), in balance with the brain and spinal cord. In normal adults, the volume of fluid is between 125 and 150 mL and its main function is to provide and maintain a suitable chemical environment for neural tissue. The biochemical composition of CSF provides information on the normal or pathological states of brain metabolism, which is very similar to that of ultrafiltered blood plasma containing metabolites that the CNS secretes [69].

CSF collection is an invasive procedure performed by lumbar puncture, where an atraumatic spinal needle is usually inserted between the 3rd and 4th lumbar vertebrae (Figure 12) and at least 12 mL of fluid is withdrawn. The first 1-2 mL is for basic CSF assessment (protein, cell count, erythrocytes) and the last 10 mL for the biobank [69].
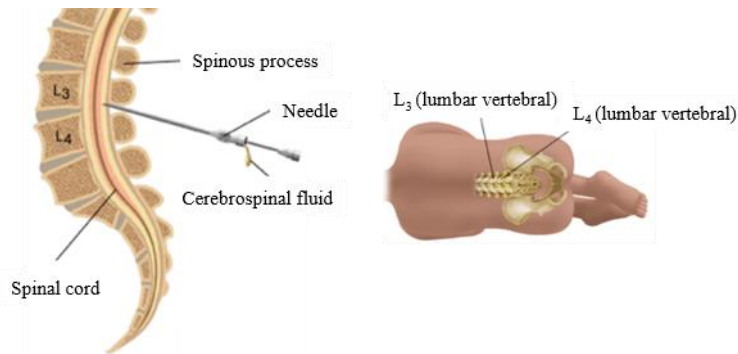
Figure 12 – Lumbar puncture: the patient lies in a hunched position on a table and a needle is inserted into the lower part of the spine to remove the cerebrospinal fluid, adapted from Julia Dixon et al [70].

One of the most likely diagnoses of AD includes evidence of memory decline, deficits in at least one other cognitive domain and progressive worsening of cognitive functions. However, it is important to investigate new biomarkers for the differential diagnosis between the different subtypes of dementia and a more accurate diagnosis to manage it effectively.

The use of cerebrospinal fluid (CSF) biomarkers is shown to facilitate risk identification of patients with MCI at different risk levels of progression to AD [71].

1.5.1 – Mild Cognitive Impairment vs Mild Cognitive Impairment due to AD

According to the literature [72], the mild cognitive impairment (MCI) phase is classified as a cognitive disorder greater than expected for an individual's age and education level.

Although the first symptoms of memory loss and language problems begin to appear at this stage, these do not interfere with the individual's daily activities.

Despite this, some individuals in the cognitive impairment phase remain stable or even return to normal over time. However, more than half of patients in the MCI phase progress to the dementia phase within 5 years [72].

The stage of mild cognitive impairment due to Alzheimer's disease (MCI due to AD), according to the literature [73], is the pre-dementia stage, where symptoms begin to have more impact on the individual's daily life, however it may not interfere with their ability to carry out their day to day tasks.

About 15% of individuals who are in the MCI due to AD stage progress to dementia after 2 years and about one third develop dementia due to Alzheimer's disease inside 5 years [15, 73].

In this way, it is necessary that there are studies that focus on the alterations that occur between these two phases, MCI and MCI due to AD. Although the symptoms presented by the patients are similar, there are some differences.

Patients in the MCI phase have fewer memory problems and are able to relate more easily to other individuals, thus leading a more independent life compared to patients in the MCI due to AD phase. One of the main differences between these two phases is the levels of biomarkers in CSF. Patients in the MCI due to AD phase have reduced levels of $A\beta$ plaques and increased levels of Tau protein (p-Tau and t-Tau) compared to patients in the MCI phase [66].

1.6 – Alzheimer's non-pharmacological and pharmacological therapeutics

The management of AD can become quite complex as it requires a multidisciplinary approach, involving a range of clinicians as a specialist doctor (neurologist, psychiatrist, geriatrician), a general practitioner, nurses and other health professionals (social worker, therapist, psychologist).

For this reason, it is important that treatments target cognitive and functional symptoms. Thus, the treatment is divided into two therapeutic approaches:

19

pharmacological and non-pharmacological. The goal is to help people maintain mental function, manage behavioral symptoms, and delay disease symptoms [74].

1.6.1 – Alzheimer's non-pharmacological therapeutics

The main objective of non-pharmacological intervention, also called behavioral intervention, is to improve or maintain the individual's cognitive function.
In this way, patients can continue to perform activities of daily living or address the behavioral symptoms that often accompany memory impairment (such as depression, wandering, sleep, agitation or aggression).

One of the most relevant approaches is cognitive training, mainly used in the early stages of AD. This approach consists of targeted stimulation of brain functions with a focus on multiple cognitive abilities in order to prevent cognitive decline, promoting self-sufficiency in activities of daily living and a better quality of life [75].

1.6.2 – Alzheimer's pharmacological therapeutics

The main objective of the pharmacological intervention is to delay the progression of neurocognitive and physical decline symptoms. Currently, there are two classes of drugs widely used for the treatment of AD, such as acetylcholinesterase inhibitors (AChEI) (donepezil, galantamine and rivastigmine) and memantine, an antagonist of the N-methyl-D-aspartate (NMDA) receptor [76].

AChEl are widely used for the treatment of mild to moderate AD. Cholinesterase inhibitors reversibly bind to cholinesterase (the enzyme responsible for the degradation of acetylcholine within the synaptic cleft). As one of the main causes of AD is the reduction of acetylcholine (ACh) synthesis, it is important that there is a therapeutic method that increases cholinergic levels in the brain, inhibiting acetylcholinesterase (AChE) increasing cholinergic transmission between neurons [77].

For this, it is necessary to choose which AchEl to use, taking into account the ease of use, tolerability, cost and preference of the doctor and patient. The toxicity of AchEl is dose-related, however this decreases over time or with dose reduction. Thus some adverse effects are associated with the use of AchEls such as nausea, vomiting and diarrhea [78].

When dealing with patients with moderate to severe AD or those who are intolerant to AChEl, the use of memantine is recommended. This is an NMDA receptor

antagonist that is able to normalize the glutamatergic system and reduce glutamate-induced neuronal degradation, since glutamate is the main excitatory neurotransmitter in the brain [76]. Glutamate activates the NMDA receptor and directs calcium ions to the postsynaptic neuron, causing synaptic plasticity, and making learning and memory processes of a higher order [79].

However, intracellular calcium accumulation sometimes occurs, the release of intracellular enzymes, mitochondrial damage and also brain cell damage and death occur. This happens when glutamate transport becomes immobilized during pathological conditions causing extracellular glutamate to accumulate, causing the NMDA receptor to open uncontrolled [79]. The use of memantine has some adverse effects such as gastrointestinal disturbances, confusion, dizziness, drowsiness, headache and agitation. Its administration is still contraindicated in individuals with a history of seizures [76].

1.6.3 – Combination therapy for Alzheimer's disease

The search for combination therapies for AD has been very important, since there is a disappointing history regarding the development of effective treatments.
It is important to note that combination therapy, as the name implies, results in the combination of memantine with any of the AChEl, but according to a study by Taro Kishi and his colleagues [79] there is a subgroup that shows greater improvement in cognitive function, which is donepezil.

A study by Pierre Tariot and his colleagues [80] showed that combination therapy is quite advantageous. They designed a clinical trial using 404 patients, over 50 years old, with moderate to severe AD (who were taking donepezil) from 37 different US centers. All these patients were chosen according to the criteria of the *National Institute of Neurological and Communication Disorders and Stroke-Alzheimer Disease and Related Disorders Association* (NINCDS-ADRDA). During the beginning and end of weeks 4, 8, 12, 18 and 24 all functional, cognitive and global outcomes were collected. At the end of 24 weeks, the combination of memantine and donepezil was found to be effective (P<0.001), allowing the investigators to conclude that memantine is associated with an improvement in cognitive function.

Another study by Susanne Hartmann and Hans Möbius [81] showed that the use of memantine with donepezil for the treatment of AD patients was effective. This study

lasted 24 weeks and donepezil was administered at stable doses (5 - 10 mg/day) together with memantine (20 mg/day). Over the weeks cognitive function improved significantly (P< 0.001).

New drugs have been developed that help to reduce the accumulation of Aβ plaques, since their aggregation initiates a set of events that lead to AD. One of the drugs used in anti-amyloid therapy is aducanumab, and when administered at an early stage of the disease, it manages to limit cognitive decline. This drug is a human monoclonal antibody that interacts with Aβ plaque aggregates, including soluble oligomers and insoluble fibrils, eliminating Aβ plaques. In this way, researchers found that patients treated with aducanumab significantly improved their cognitive abilities [82].
However, there are patients with typical neuropsychological and clinical features of AD who do not have amyloid biomarkers, raising doubts as to whether the Aβ pathway is a pathological moment in AD development or is just a physiological response to neuronal damage [83].

Thus, discovering biomarkers that allow the diagnosis and therapeutic intervention of AD is extremely important. Currently some drugs can delay symptoms, however the existing treatments do not have the ability to stop the progression of the disease. This makes it essential in the future to develop multi-targeted inhibitors (which include anti-amyloid and anti-tau effects, anti-neuroinflammatory and neuroprotective effects, and neurotransmitter modifications) to treat this disease.

1.7 – The potential of metabolomics as a tool for diagnosis

Metabolomics is a scientific study that is based on the identification and quantification of a set of metabolites (metabolome) that are produced or altered by an organism [84].
The human metabolome contains about 2,000 – 40,000 metabolites, which are small molecules (<2,000 Da) resulting from the end products of various biological pathways and processes [85, 86]. In this way, the metabolomics approach has opened new diagnostic possibilities regarding neuropsychiatric disorders, as this approach has the ability to map early biochemical changes in the disease. Thus it is possible to develop new biomarkers that indicate pathological abnormalities before the development of clinical symptoms [87].

The diverse interactions between the tens of thousands of molecules result in very complex biological processes and functions. For this reason, omics data analysis methods (metabolome, genome, proteome and transcriptome) provide information about cellular processes in several areas, such as plant biology, animal science, toxicology, molecular epidemiology and complex diseases.

The inclusion of different omics profiles (Figure 13) allows the collection of more comprehensive information about biological systems and molecular processes [88].
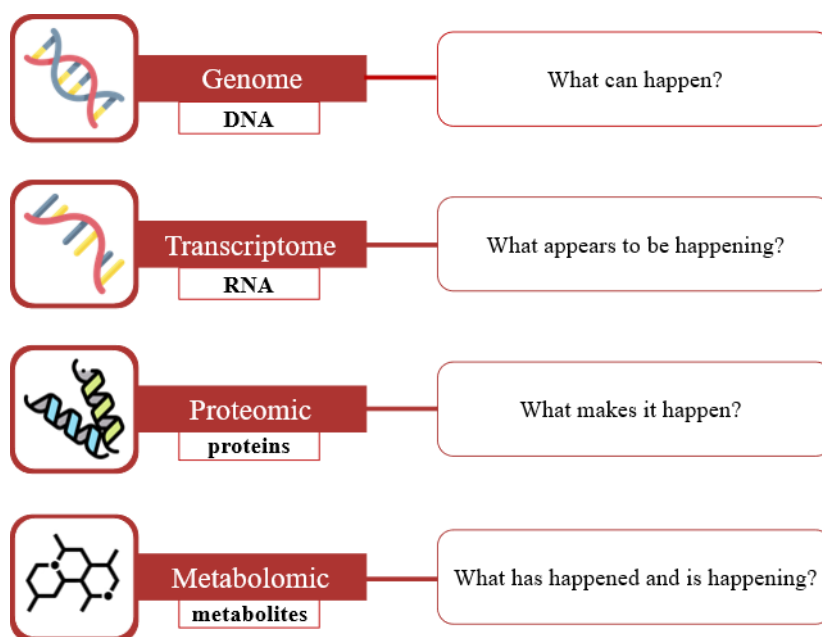
Figure 13 – The "omics cascade" in systems biology: this cascade is defined by different steps, starting from the genome, transcriptome, proteomics and the final step is metabolomics, adapted from Kamil Jurowski et al [89].

Omics approaches are essential tools and when applied and used, they allow understanding an organism's biology and their response to genetic perturbations and environmental stimulus [89].

According to the literature [90] metabolomics is seen as a "direct functional reading of the physiological state of an organism", since the metabolites transmit signals about the genetic structure and the environment. Raoul J. Bino and his collaborators [91] also mention that metabolomics emerged from a genomic methodology that makes it possible to understand the complexity of molecular interactions in biological systems.

Nowadays, metabolomics is applied in several biological studies, since the biochemical response of an organism to a disturbance is characterized by the differential accumulation of individual metabolites [90]. The main objective of metabolomics is to

define the interactions of metabolites quantitatively and qualitatively in biological networks.

When profiling the metabolome, either targeted or untargeted methods are used [92].

Regarding targeted methods, these focus on the exact quantification of a certain set of metabolites in biological samples. Generally, the set of metabolites is predetermined either by the scientific question in question or by the size of the library available in the software [92].

In relation to untargeted methods, they measure and compare the largest possible number of signals in each set of samples, assigning these signals to specific IDs of the metabolites through metabolomic databases. However, despite good progress in the construction of these databases, there is a significant part of the detected signals that are not identified, due to the absence of their spectra in the metabolomic databases. Untargeted approaches are extremely useful in identifying unknown metabolites, especially when they are the biomarkers of a study [92].

Untargeted metabolomics studies provide insights into fundamental biological processes, as they reveal not only that several metabolites are still uncharacterized in terms of structure and function, but also that many of these uncharacterized metabolites undergo changes as a function of health and disease [92].

Usually, untargeted metabolomics studies are hypothesis generators and not hypothesis-driven. However, it is essential to build a workflow to maximize the number of metabolites detected and their quantitative reproducibility (Figure 14).
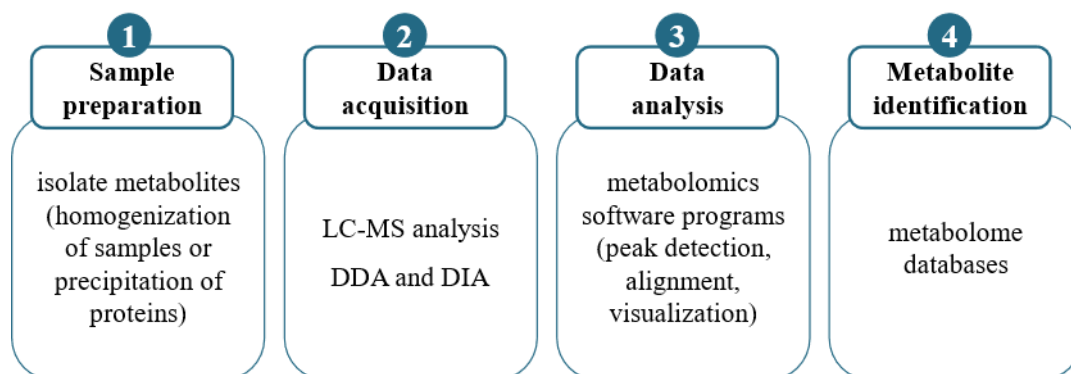
Figure 14 – Workflow of an untargeted metabolomics analysis by LC-MS: this workflow a workflow maximizes the number of metabolites detected and their quantitative reproducibility.

24

Is necessary to use several analytical techniques in metabolomic studies, because the metabolome is quite complex, and the metabolites have a wide range of physicochemical properties.

The most common analytical techniques are spectroscopic, which includes nuclear magnetic resonance (NMR) and spectrometric, which includes mass spectrometry (MS) coupled with separation techniques (liquid chromatography (LC), gas chromatography (GC), supercritical fluid (SFC)) [93].

However mass spectrometry is chosen in most metabolomics studies, because it has high sensitivity and a wide range of covered metabolites. Thus MS in conjunction with a separation technique becomes an essential tool for metabolomic studies [94].

1.8 – Application of metabolomics in Alzheimer's disease

There are currently many individuals with Alzheimer's disease and, as noted earlier in this work, the number of cases is expected to increase over the next few years as the population ages.

The pathophysiology of AD is believed to begin several years before the onset of clinical symptoms. Although several biomarkers and imaging techniques are available to diagnose AD, a definitive diagnosis requires a post-mortem examination of brain tissue [95]. Thus, developing effective biomarkers for an early diagnosis is very important.

Metabolomics is a new approach that aims to improve sensitivity and specificity in the early diagnosis of AD patients. This approach uses multiple platforms to measure the levels of small molecule metabolites in biological samples [95]. As mentioned earlier, these approaches are based on targeted and untargeted analyzes of the CSF using mass spectrometry techniques. These analyzes make it possible to verify the changes that occur in the metabolic pathways of AD patients, including disturbances in methionine metabolism, the tricarboxylic acid (TCA) cycle and lipid metabolism [94].

A study by Eugenia Trushina and collaborators [96] compared individuals with MCI and AD, finding that the number of affected pathways in the CSF increased by 50% in patients in the dementia phase. This is because the number of altered pathways increases with disease severity.

On the other hand, in patients with MCI, the number of affected pathways was higher in plasma, as this fluid reflects changes that occur in organs other than the brain. Some changes can occur in the methionine cycle in CSF, and these are associated with

cardiovascular diseases, vascular dementia, mild cognitive decline and Alzheimer's disease [97].

One of the most important points to understand the mechanisms behind AD is the methionine cycle, because this is essential for cognitive function. One of the main products of the methionine cycle, of the synthesis of nucleic acids and proteins and of the biosynthetic reactions of brain metabolites is S-adenosylmethionine (SAM), which when demethylated forms S-adenosylhomocysteine (SAH).

Normally, when comparing individuals with AD and MCI with healthy individuals, what is observed are low concentrations of SAM levels and high concentrations of SAH levels [97]. Sometimes there is a change in the balance of SAM and SAH and these moments record memory loss and cognitive decline in older populations.

However, there are other metabolites that are also at high levels in patients with AD and MCI, such as choline, methionine and serine [97, 98].

One of the metabolites that acts as a precursor to the neurotransmitter acetylcholine is choline. Choline is involved in betaine and methionine metabolism and in phospholipid biosynthesis, and when it is in normal concentrations it represents good neurotransmitter activity and good metabolism in the brain and CNS. However, higher amounts of choline in AD patients represent neurodegeneration and the breakdown of choline-rich synaptic membranes [97].

Amino acids are very important in the immune response, neurotransmitter function and protein synthesis. However, some patients with MCI and AD undergo some changes in amino acid metabolism. A study by Clara Ibáñez and colleagues [99] showed that several amino acids (valine, serine, histidine and arginine) showed different patterns. Valine and serine showed an increase in the CSF of individuals with AD, the amino acid histidine decreased in individuals with MCI and AD, and arginine showed a small decrease in the CSF of individuals with AD.

However, a study by Yuki Nagata and colleagues [100] showed that there was a reduction in CSF serine levels in AD patients, concluding that serine has a neuroprotective role for the enzyme phosphatidylserine synthase (PSS). This enzyme forms phosphatidylserine through the incorporation of serine into phosphatidylcholine

and phosphatidylethanolamine. The application of PSS indicates an improvement in the cognitive measures of individuals with AD [100].

Tryptophan is also an amino acid of great interest. Tryptophan is a precursor to the synthesis of serotonin, melatonin and niacin. In AD patients, tryptophan levels are reduced, resulting in reduced serotonin synthesis [101]. Changes in the balance of tryptophan metabolism sometimes occur, and these are related to common features of neurodegenerative disorders.

A study by D. Fekkes and colleagues [101] showed that tryptophan, quinolinic acid (QUIN) and kynurenic acid (KYN) had reduced levels in subjects with preclinical AD. The latter two results from tryptophan degradation and are associated with neuroinflammation. However, while QUIN has a neurotoxic role and is present in high amounts in AD patients, KYN has a protective role and is present in smaller amounts [102]. Kynurenic acid acts as an endogenous competitive agonist of the NMDA receptor, reducing its effects. This is because in Alzheimer's disease there is an overstimulation of the NMDA receptor, which can cause neuronal death by excitotoxicity [102].

In this way, it is possible to believe that both tryptophan and its metabolism products (QUIN and KYN) can be potential biomarkers for early AD diagnosis and a focus for new therapeutic interventions.

As previously mentioned, some disturbances at the tricarboxylic acid (TCA) cycle level also occur in patients with AD. It is important to note that the TCA cycle plays an important role in gluconeogenesis, lipogenesis and amino acid interconversion. This cycle is the main source of energy for cells, thus being the final common pathway for the oxidation of proteins, lipids and carbohydrates [87]. The disturbances that occur in the TCA cycle in AD patients cause the intermediates of this cycle, when compared to control groups, to appear in higher concentrations, both in plasma and CSF.

A study carried out by Crystal Sang and colleagues [103] confirms that amino acids and fatty acids fuel the tricarboxylic acid cycle when there is a decrease in glucose levels in individuals with AD. Thus, a decrease in plasma amino acid concentrations is related to their use for TCA cycle replacement. Another way to reset the cycle is through the oxidation of fatty acids through the production of acetyl coenzyme A (acetyl-CoA).

Even so, in AD patients it is also possible to verify an accumulation of acylcarnitines both in the plasma and in the CSF, which is formed by the association of

acylcoenzyme A (acyl-CoA) with carnitine. Carnitine provides energy to cells through reactions that transfer long-chain free fatty acids from the cytoplasm to the mitochondria, facilitating their oxidation. Accumulation of acylcarnitines in AD patients is related to incomplete oxidation of acyl-CoA intermediates leading to a retroconversion to acylcarnitine avoiding the toxic effects of their accumulation in mitochondria [104].

In AD patients, changes in sphingolipid (SM) levels also occur, which are responsible for several processes such as cell division, differentiation and death. The degradation of SM is related to several neurodegenerative disorders in the brain, thus contributing to Alzheimer's disease. However, there are studies that show that during the progression of this disease, SM decrease in both brain tissues and plasma, while others suggest that they increase [103].

A recent study by Therese Koal [105] and colleagues showed that the combination of two metabolites (SM (d18:1/18:0) and SM (18:1/18:1)) showed a significant and positive association with the pathology of AD. Thus, the researchers concluded that the SM metabolite (d18:1/18:0) is a potential biomarker diagnosing AD.

Despite this, there are other potential biomarkers for diagnosing AD, such as uracil and uridine. Uridine is absorbed and phosphorylated in the brain to form nucleotides used for the synthesis of DNA and RNA.

On the one hand, a study in the literature [106] showed that uridine levels increased in the CSF of individuals with AD, resulting in a change in the reversible conversion of uridine to uracil, through the enzyme uridine phosphorylase. This enzyme plays an important role in the detection and initiation of cellular responses to oxidative stress. However, uracil levels decreased in AD patients, which is the main substrate of the enzyme uracil-DNA-glycosylase (which is an enzyme that regulates oxidative stress). This enzyme removes oxidized pyrimidines and is overexpressed under oxidative stress as a protective mechanism against neurodegeneration. This shows that decreased uracil levels may be related to changes in uracil-DNA-glycosylase function leading to neurodegeneration. Even so, uracil acts on the P2X channel in the brain, which controls neurotransmission, neuromodulation, cell proliferation, differentiation and death, causing any alteration to be associated with several neurodegenerative diseases [106].

On the other hand, study [107] shows that the levels of uridine in the CSF of individuals with AD decreased. This decrease may be related to reduction in synaptic plasticity and neuronal deficits.

However, although the results of these two studies [106, 107] are contradictory, there is an explanation that makes them plausible, since the effects of nucleosides depend on brain region, age and sex. This means nucleosides have different roles depending on different brain areas. Thus, it is possible that both uridine and uracil are potential new biomarkers for an early diagnosis of AD.

One of the critical measures of neuronal and synaptic activity is the cerebral metabolic rate of glucose (CMRglc). Thus, some studies have shown that all clinical symptoms of Alzheimer's disease are related to significant decreases in CMRglc. This was visible by PET together with 2-[18F]-fluoro-2-deoxy-D-glucose (FDG) as a tracer [108].

A study in the literature [109] measured glucose levels and the proportions of glycolytic amino acids (serine, alanine and glycine), which are representative of brain glycolytic function. The results indicated an increase in glucose levels in the brain and a decrease in glycolytic flow, associating them with the pathology's severity and the expression of AD symptoms.

Another study [110] analyzed 122 CSF metabolites from patients with and without AD. It was found that only the glycolysis intermediates, dihydroxyacetone phosphate (DHAP) and phosphoenolpyruvate (PEP), showed a significant decrease in AD patients. Thus, the decreases in the levels of glycolytic metabolites in the CSF of Alzheimer's patients translate into inhibition of glycolysis under oxidative stress conditions.

In conclusion, decreases in glucose levels in brain metabolism are associated with the severity of AD pathology, making it essential to study the potential of glycolysis metabolites for the diagnosis of AD.

Therefore, with the identification of all these metabolites, several studies have emerged that allow us to understand whether they can improve the classification of the cerebrospinal fluid profile of AD pathology [106-110]. Some metabolites such as methionine, serine, SAH and choline do not have the ability to improve the CSF profile classification [98].

However, other glycolysis metabolites, such as the combination of PEP-DHAP and PEP-2PG, showed the ability to improve the classification of the CSF profile. Even so, none of these metabolites can achieve the performance of the already validated metabolites, that is, $A\beta_{1-42}$, p-Tau and t-Tau.

This topic aimed to show the potential of several metabolites in order to show their potential performance compared to CSF amyloid and Tau proteins. In conclusion, further studies are needed to identify potential new biomarkers that allow early diagnosis of Alzheimer's disease. For this, it is necessary to pay attention to some criteria, such as the increase in the number of samples, the increase in age differences and an increase in male patient.

2 – Impact of the progression of dementia at a forensic level

Currently, and as mentioned in the previous chapter, there is a rapid increase in the elderly population worldwide. In this way, it has been possible to find some unexpected problems facing patients with dementia [111]. Thus, it is important to understand the impact of dementia at a forensic level, so that it is possible to judge a given individual fairly.

Dementia patients reveal neurocognitive symptoms such as, for example, memory impairment, lack of judgment, and language problems, among others [112]. It is known that these symptoms can change depending on several factors, including the type of dementia, the passage of time, the environment where the individual is, the medications administered, and the care provided. That is, all symptoms can favor (whether separately or in combination with each other) criminal behavior and actions [112].

On the other hand, some of the patients with dementia become more careful and introverted people, as they feel their shortcomings. A study present in the literature [112] indicates that, in the case of Alzheimer's disease, the reduction of cognitive function causes individuals to behave passively, affecting their ability to elaborate and organize a certain crime.

However, a part of the population that suffers from dementia problems shows an increase in aggressive behavior compared to the normal population.

Thus, individuals with dementia who have committed a crime manifest various subtype of dementia, as well as stages of progression. These criminal behaviors are further divided into violent and non-violent crimes. The first is related to alcohol consumption, while the second is related to individuals with dementia who have no history of alcohol consumption [111].

According to the literature [111], the progression of an individual's dementia status is an important factor in relation to the commission of crimes.
On the one hand, crimes associated with alcohol consumption occurred earlier in dementia, because of cognitive dysfunction, hallucinations and delusions.
On the other hand, patients who had no history of alcohol consumption committed a crime at a more advanced stage, where dementia had progressed considerably and when patients had already lost most of their family and socioeconomic support.

According to the literature [112], alcohol is a very common risk factor for violent crimes worldwide. This same study analyzed a group of people with dementia, to increase the understanding of risk factors for crime in people with dementia.

Thus, although the investigators excluded patients with alcohol consumption at the time of the forensic psychiatric examination, approximately 48% of the occurrences indicated an alcohol-related diagnosis.

Still, alcohol dependence decreases cerebral blood flow, leading to deterioration in cognitive function, and provoking more violent behavior, whether in individuals with or without dementia [111].

When faced with a trial of an individual, where the individual is found guilty and the verdict is imprisonment, the Swedish court may accept the submission of a forensic psychiatric examination if the person in question shows signs of dementia.

However, this examination must clarify whether the individual was suffering from a "severe mental disorder" at the time of the crime and at the time of the examination [112].

However, from a forensic perspective, this legal concept of "severe mental disorder" is not related to diagnoses, but to effects of disorders, including psychotic effects; a state of severe depression with suicidal behavior at the time of the crime; severe personality disorder with recurrent episodes of psychotic behavior and reduced psychosocial functioning; mental disorder with increased compulsiveness; severe dementia, severe brain damage, and severe mental retardation [112].

Although the different subtypes of dementia are considered serious by the general health system and the population, they may not be considered as serious from a forensic point of view.

This means that if an individual shows a state of severe psychiatric disorder at the time of committing a given crime, the verdict may implement forensic psychiatric treatment instead of imprisonment [112].

It is important to note that forensic psychiatric examinations are carried out by the *National Council of Legal Medicine*, which is carried out by a team of forensic psychiatrists and psychologists, forensic social researchers, and also a nursing team.

The *Swedish Code of Judicial Procedure Act* [112] allows the court to refuse to prosecute an individual if he has committed a crime while in a state of dementia or another serious mental disorder.

32

In conclusion, an individual with dementia may misinterpret a given situation, acting inappropriately and unfairly. Even so, it is believed that many of the patients who cause criminal acts are not fully aware of what they are doing, often not understanding what is happening.

However, it is extremely important to carry out psychiatric examinations accurately and fairly, so that all people with dementia are judged according to the relevance of their actions, especially when we are dealing with other people's lives.

3 – Analytical techniques

As already mentioned, the study of the metabolome is challenging, as metabolites have a wide range of physicochemical properties.

Because of this, spectroscopic and chromatographic techniques are used to study the compounds of interest [113]. However, the choice of the most appropriate technique depends on the sample matrix, its concentration, the metabolites properties and even the sample amount.

3.1 – Nuclear magnetic resonance spectroscopy (NMR)

Nuclear resonance spectroscopy (NMR) is a highly reproducible, non-destructive, cost-effective technique. However, it is only used in 30% of metabolomics studies.

NMR is based on the absorption and re-emission of energy through some atomic nuclei, which are influenced by an external magnetic field. This technique can be automated, allowing rapid analysis by simultaneously measuring several types of metabolites [113]. Furthermore, it also allows the collection of additional structural information, such as the presence of chemical functional groups and the spatial decomposition of chemical groups in a structure. All this variety of information makes it possible to understand biological processes and biochemical pathways [114].

Despite this, this technique has a major disadvantage, as its sensitivity is significantly lower than mass spectrometry (MS), making NMR unsuitable for analyzing low-abundance metabolites [113]. However, NMR is considered a complementary technique to LC-MS or GC-MS.

3.2 – Mass spectrometry (MS)

Mass spectrometry measures the mass/charge (m/z) ratio of ions in order to identify and quantify the molecules present in a given sample, becoming one of the main techniques for metabolomics studies.
Even so, the separation of the analytes is fundamental for complex matrices, such as biofluids, because a direct injection into the spectrometer without a chromatographic separation makes the analysis of metabolites very fast, with ion suppression and low ionization efficiency [113].

A mass spectrometer is based on a sample introduction system, an ionization source (production of ions resulting from the ionization of sample molecules), a mass analyzer (separation of selected ions according to the m/z ratio) and an ion detector

(collection of ions and signal formation, where the intensity is related to the number of detected ions) (Figure 15). However, both the mass analyzer and the ion detector are under high vacuum making it possible to reduce the background noise and allow the fragmentation pattern not to be influenced [114].

Mass spectrometers can perform tandem MS (MS/MS) using multiple mass analyzers. This allows ions to be selected to induce their fragmentation, obtaining their structural information.
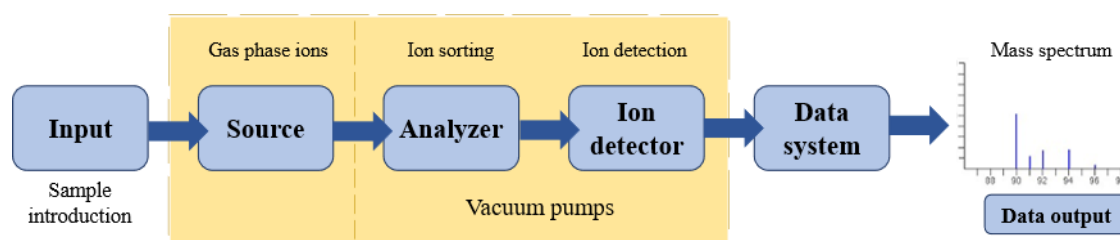
Figure 15 – Mass spectrometry analysis: a mass spectrometer is constituted by an ionization source, a mass analyzer and an ion detector, where the mass analyzer and ion detector are under high vacuum, reducing the background noise and allowing the fragmentation pattern not to be influenced.

As previously mentioned, MS is associated with a separation technique that allows better detection limits and data quality. These separation techniques include gas chromatography (GC) and liquid chromatography (LC).

One of the most used separation techniques in metabolomics studies is GC-MS. It is a robust technique with good separation ability, sensitivity and reproducibility, having almost no problems with matrix effects and ion suppression of coeluent compounds compared to LC-MS. However, GC-MS can only be applied on low molecular weight volatile compounds [113].

Currently, LC-MS has been one of the most used techniques for metabolomics studies. Liquid chromatography is a technique that separates different compounds from a given sample according to the interactions of the compounds with the mobile and stationary phases. The degree of separation of the compounds is related to the affinity of each compound with the mobile phase [115].

LC-MS has several advantages, such as that they operate at lower temperatures (allowing the analysis of thermolabile metabolites) and do not require chemical derivatization (facilitating the steps of sample preparation and identification of metabolites). Even so, it is a technique that allows accurate and precise quantification, as

it presents high selectivity and sensitivity. Thus, it allows to perform the characterization of compounds with similar retention time, but with different m/z ratio or fragmentation patterns [116].

High-performance liquid chromatography (HPLC) is an instrumental technique based on applying more traditional liquid chromatography to theories and instrumentation that were developed for gas chromatography (GC). In HPLC the mobile phase is a solvent or a system of solvents that are pumped through the column in a constant flow under high pressure. The stationary phase is packed inside a chromatographic column in order to maintain high pressure. The main advantage of this technique is its ability to analyze a much wider range of components, such as thermally unstable compounds [117].

A metabolomic study based on LC-MS presents several sources of ionization, however, the most used are matrix-ionization assisted laser desorption/ionization (MALDI) and electrospray (ESI) [118, 119].

The MALDI ion source has established itself as a key instrument for protein sequencing and proteomics research, offering tremendous potential when combined with Electrospray Ionization (ESI) technology. This versatile ion source allows for analysis under different conditions, either at atmospheric pressure or within a high vacuum environment. Notably, in the context of metabolomics research conducted in high vacuum, a specific approach known as MALDI imaging has gained prominence. Through this technique, researchers are able to acquire spatially resolved information on the distribution of metabolites within tissues, thereby enabling a deeper understanding of metabolic processes [118].

The electrospray ionization (ESI) source is used for qualitative and quantitative studies of a huge variety of non-volatile and thermally labile inorganic chemical compounds. This technique is used to produce ions through an electrospray in which a high voltage is applied to a liquid, creating an aerosol [119].

The process only starts when an electric field is applied. This electric field causes an accumulation of ions in the droplet leaving the capillary. While they travel through the space between the needle tip and the cone, the nebulizer gas will initiate the solvent evaporation process, causing the drops to decrease in volume. This causes the droplets to

36

reach the point where the surface tension can no longer sustain the charge and the droplet will break down to become a single charged molecule (Figure 16).

This ion formation process is called the ion evaporation method, providing ions with relatively low m/z values [118, 119].
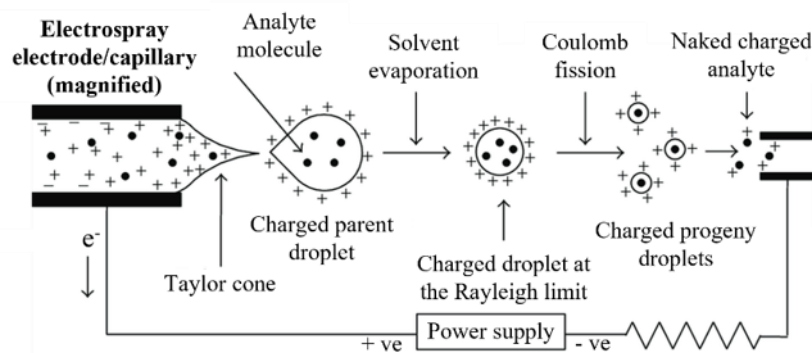
Figure 16 – Electrospray (ESI) mechanism: the formation of a charged drop occurs, given the applied electric field, decreasing the size of the drop until it reaches the point where it can no longer maintain the charge. After that, the drops fragment (Coulomb explosion) causing the division of the drops, where each drop corresponds to a single charged molecule.
Adapted from Shibdas Banerjee et al [120].

Different mass analyzers can be used, depending on mass accuracy, dynamic range and the ability to perform tandem experiments (MS/MS). Another factor is resolution, the higher the resolution, the higher the accuracy of the mass. Two of the most widely used mass analyzers include the quadrupole (Q) and the time-of-flight (TOF).

The quadrupole analyzer (Q) consists of four opposite metal rods, parallel to each other. The voltages applied to these rods will affect the trajectory of the ions traveling through the centralized flight path between the four rods. For certain direct current (DC) and radio frequency (RF) potentials, there are certain ions with a specific m/z ratio that pass through the quadrupole filter and the rest are not transmitted [121, 122].

One of the most widely used tandem mass spectrometers (MS/MS) is the triple quadrupole (QqQ) which consists of two quadrupole mass analyzers (Q1 and Q3) and one collision cell (q2). The first quadrupole (Q1) has the function of selecting specific ions according to their m/z ratio and the ions that are not of interest are discarded. After that, they proceed to the collision cell (q2) which fragments the ions, through bombardment with a neutral gas (nitrogen or argon) and after being fragmented they are

directed to the third quadrupole (Q3). Through this process, an MS/MS spectrum will be produced that allows confirmation of the identity of the analytes [123].

The TOF analyzer uses an electric field that accelerates the ions that are formed through the same electric potential, measuring the time each ion takes to reach the detector. That is, all ions are subjected to the same kinetic energy, traveling the same distance at different speeds depending on their m/z values. If the ions have the same charge, their kinetic energies will be identical, so the speed of each ion will depend only on the mass. This means that the lighter ions reach the detector first, while the heavier ones take longer [123].

Thus, the time-of-flight quadrupole (Q-TOF) is also applied in MS/MS methods. This technique (Q-TOF-MS) combines the benefits of two different mass analyzers by utilizing the high fragmentation efficiency of quadrupole technology with the high analysis speed and high time-of-flight mass resolution capability (Figure 17) [123].
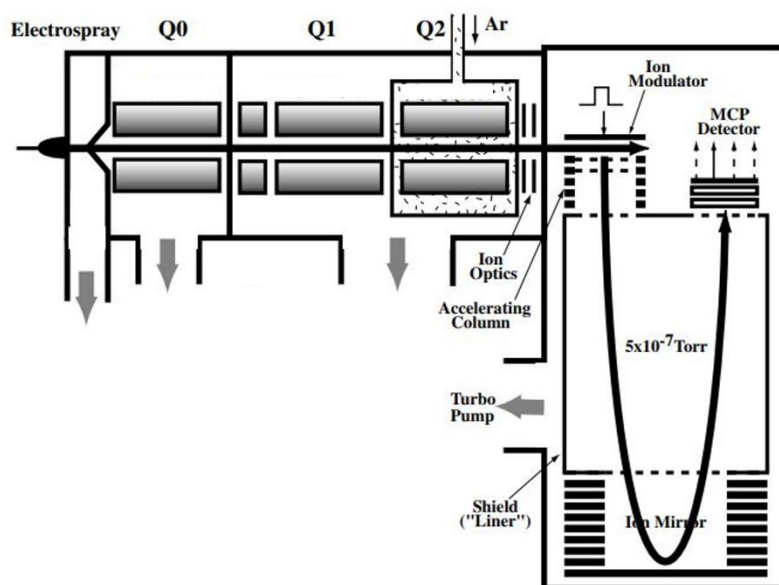
Figure 17 – Schematic Diagram of the Q-TOF-MS: after ionization of the molecules in the ESI source, the molecular ions enter the Q-TOF-MS, starting in Q0 for beam focusing, followed by ion selection in Q1, and fragmentation in Q2. The resulting fragments enter the TOF analyzer, allowing the detection of all fragments, adapted from Chernushevich et al [123].

As previously mentioned, the ions are analyzed according to their m/z ratio and as such need to be selected by the mass analyzer. This selection can be made through different methods, including data dependent acquisition (DDA) and data independent acquisition (DIA), such as SWATH (sequential window acquisition of all theoretical

fragment-ion spectra) (Figure 18). However, in metabolomics studies, the use of data dependent acquisition is more frequent [124].
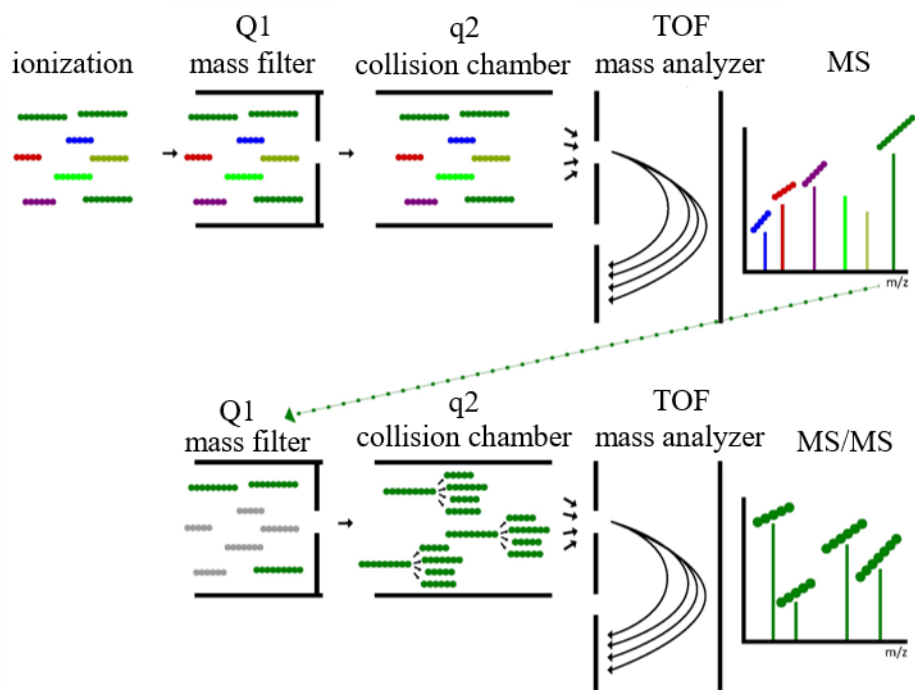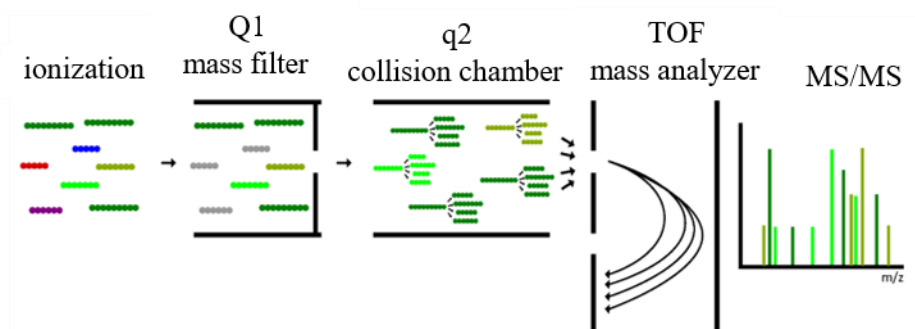
Figure 18 – Data-acquisition approaches for mass spectrometry: DDA obtains fragmentation spectra only for a selected precursor ion based on predefined criteria. DIA the fragmentation spectrum is acquired for all precursor ions regardless of predefined criteria by SWATH, adapted from Santa et al [125].

In DDA, there is a choice of which ions will be selected, through some predefined parameters, including the intensity of signal. Normally, the most intense ions are selected, that is, the most abundant. The DDA method uses a narrow m/z window, producing MS/MS spectra with little interference, as only selected ions are transferred to the q2 collision cell to generate product ions. Even so, this method has some limitations, as some

important precursor ions may not be selected for fragmentation if they do not meet the selection criteria [124, 125].

In DIA, all ions within a given m/z are fragmented and analyzed. Mass spectra are acquired by fragmenting all ions entering the mass spectrometer at a given time (broadband DIA) or by sequentially focusing on a narrow m/z window of precursors, fragmenting all those detected within that window. As previously mentioned, SWATH is one of the DIA techniques [124].

In a SWATH acquisition, the mass specification does not require the initial detection of an MS peak to proceed with the MS/MS analysis, so this technique uses a wider Q1 isolation window covering the entire scanning range, collecting the MS/MS spectra. SWATH technology differs from the traditional quantitative mass spectrometry method in that it directly builds the secondary fragment ion chromatograms, giving each point on the curve sufficient mass spectrometry evidence, increasing quantitative accuracy and reproducibility [124, 125].

However, mass spectrometry is a very applied technique for biological samples. The choice of liquid chromatography as a separation technique reduces the complexity of biological samples, decreasing the effects of the matrix during ionization. In this way, using the Q-TOF LC/MS technique is important for the identification of metabolites, allowing to obtaining qualitative and quantitative information about them.

3.3 – Data analysis

In order to make a more comprehensive detection of metabolites in LC-MS data, several data analysis strategies are necessary. Data analysis in metabolomics studies aims to convert raw data into biological knowledge.

However, these datasets need some treatment, as data pre-processing steps can affect potential metabolite identifications and their subsequent quantification and biological interpretation [126]. As LC-MS data is archived through raw data, it is necessary to use extensive data pre-processing steps for some purposes. One of them is related to reducing the file size, reducing the complexity of the data, and providing them in an adequate format.

Another purpose is related to the alignment of the data to guarantee a correct identification of the metabolites in all analyzed samples [85]. Statistical analysis aims to

detect which peaks have significantly altered intensity between the different biological groups [126].

These two steps, data pre-processing and statistical analysis, are very important, as the MS measurement contains data that may be related to instrument variations or sample heterogeneity, leading to incorrect identification of metabolites.

3.3.1 – LC-MS data analysis

As mentioned before, an untargeted metabolomics approach comprehensively analyzes all measurable analytes in a given sample, including unknown metabolites. After obtaining the raw data, it is necessary to carry out a pre-processing step in order to convert the raw LC-MS data into a list of peaks that are easy to interpret and compare between different runs.

Some pre-processing steps must be performed to correct variations in the raw data, such as outlier screening, baseline correction, peak detection and alignment, and normalization of peak intensities, among others. However, it is not necessary to carry out all these steps, but in comparison studies peak detection and alignment are fundamental [126].

Peak detection consists of converting the raw data into a list of features, m/z and retention time, with the respective responses, i. e. the peak areas, facilitating the removal of noise from the data. Peak detection and alignment can be performed using various software, such as SciexOS, MS-DIAL, MarkerView^{TM} and MZmine. However, in addition to these software, internal standards can be used to correct retention times and/or normalize the data. All these changes in the LC-MS data are fundamental, as they make the data more suitable for future statistical analysis [126].

3.3.2 – Statistical analysis

Statistical analysis is used to extract relevant information about data sets. Normally, univariate and multivariate methods are the most used in statistical analysis. The first consists of analyzing a single variable in question, requiring prior knowledge of the measured variable, while the second consists of analyzing multiple variables. However, multivariate analysis simplifies the data into manageable variables, making it easier to interpret. This analysis can be supervised or unsupervised. In the first case, there is information about the class of the data, while in the second case there is no information about the class of the data [126].

Principal Component Analysis (PCA) is the most used unsupervised approach in LC-MS metabolomics studies. This analysis consists of reducing the size of the data set without significant loss of information, allowing better visualization of the data and the determination of patterns between different samples (control vs. disease) [127]. On the other hand, the partial least squares-discriminant analysis (PLS-DA) is the most used in supervised approaches. This analysis aims to find the metabolic pattern that is causing discrimination between two or more study groups [126].

Appropriate statistical analysis will extract information from the large data set and will allow visualization and interpretation. A multivariate analysis will facilitate the determination of specific patterns in a dataset of biological samples, becoming essential in studies that compare control vs. diseases or different diseases.

4 – Significance of the study

Alzheimer's disease is a progressive neurological disorder that manifests a progressive decline in various cognitive abilities, affecting an individual's behavioral and social skills, making it impossible for them to live their daily lives in a "normal" and independent way.

As mentioned in previous topics, amyloid plaques are observed in the brain and analyzed in the CSF, making them very important for the diagnosis of AD. Despite this, some patients do not have these plaques and have typical features of AD. For this reason, it is extremely important to study new biomarkers for the early diagnosis of AD.

Thus, this work has as main objectives:

• compare the results obtained through two different software, SciexOS and MS-DIAL;

• discover which metabolites differ between patients who are in the MCI phase and those who are in the MCI phase due to AD;

• identify the most significantly altered metabolites;

• discover potential biomarkers for an early diagnosis of Alzheimer's disease, having the ability to overcome the blood-brain barrier so that in the future this type of study can be carried out at the blood level.

5 – Materials and Methods

5.1 – Study group

This study includes 40 patients with mild cognitive impairment aged between 46 and 80 years old. CSF samples were collected at Centro Hospitalar e Universitário de Coimbra (Coimbra, Portugal) under code HUC-43-09 and, more recently, CE-029/2019. Of the analyzed CSF samples, 20 were from patients with Aβ negative (control goup) and 20 from patients with Aβ positive. Their characteristics are summarized in Table 3 and detailed information about the study participants can be found in Table A.1.1 of Appendix A.1.

According to Albert et al [73], the Aβ positive group consists of patients who meet the high probability criteria for MCI due to AD. The Aβ negative group consists of patients with MCI who meet the clinical and cognitive criteria for MCI and the etiology of MCI consistent with the AD pathophysiological process but have negative biomarkers for Aβ deposition and neuronal injury.

Table 3 – Characteristics of study participants.

	β-amyloid negative (Aβ-)	β-amyloid positive (Aβ+)
n	20	20
age (years ± SD)	59.6 ± 5.8	63.8 ± 8.5
female (%)	50	55
male (%)	50	45

5.2 – Protein precipitation

After collecting the CSF samples, 200 μL were used for protein precipitation. We added 800 μL of methanol to each of the tubes and next the samples were vortexed in an IKA™ MS 3 Basic Vortex Mixer and incubated for 1 h at –80˚C. Then, the samples were centrifuged for 20 min at 20,000×g at 4˚C in an Eppendorf® Refrigerated centrifuge Model 5430R. After that, we transferred the supernatant to new tubes and evaporated in an Eppendorf® Concentrator Plus. Finally, we stored the samples at –20˚C until data acquisition by LC-MS.

5.3 – LC-MS data acquisition

5.3.1 – Sample preparation

After evaporating the samples, they were resuspended in 50 µL of a mixture containing the mobile phase, a solution of 2% acetonitrile (ACN), 0.1% formic acid (FA) and an internal standard penicillin V potassium salt and sulfamethazine-D4. Then, samples were sonicated in a Bioblock Scientific Vibracell™ 75041 for 2 min at 40% amplitude in 1 second cycles (1 second on, 1 second off).

After that, the samples were centrifuged for 10 min at 14,000×g in a MiniSpin Plus™ Microcentrifuge and then the supernatant was transferred to HPLC vials.

Finally, 7 µL of each sample were taken to the HPLC vials in order to create 4 pools (Aβ+ female, Aβ+ male, Aβ- female and Aβ- male).

5.3.2 – Data acquisition

The MS analysis was performed on a NanoLC™ 425 system (Eksigent) coupled to an ESI DuoSpray™ ion source (Sciex) operated in positive mode and a Triple TOF™ 6600 System mass spectrometer (Sciex). Metabolite separation occurred on a Triart C18 1/32" capillary column (12 nm, S-3 µm, 150 x 0.3 mm, YMC) at 5 µL/min, with an acetonitrile gradient as described in Table 4.

Table 4 – Acetonitrile gradient used in the LC/MS experiments. The mobile phase B consisted of 0.1% in ACN and mobile phase A of 0.1% FA in water.

Time (min)	Solvent A (%)	Solvent B (%)
0	95	5
20	50	50
25	5	95
30	5	95
31	95	5
40	95	5

Data acquisition was performed in two ways, programmed by the mass spectrometer, including data dependent acquisition (DDA) and data independent acquisition (DIA-SWATH analysis).

For DDA, a full mass spectrum (50-2250 m/z) was obtained with an accumulation time of 250 ms, followed by up to 40 MS/MS scans of 40 ms per spectrum. For SWATH acquisition, the equipment was configured to scan the full spectrum (50-2250 m/z) for 250 ms, followed by 73 product ion windows covering the 50-1500 m/z precursor mass range. Detailed information about SWATH windows can be found in Appendix A.1 in Table A.1.2.

The SWATH-MS/MS spectra were collected from 50-2250 m/z for 23.3 ms, resulting in a cycle time of 2.002 s. Data was acquired using Analyst® TF software (v1.8.1, Sciex).

5.4 – Data processing and statistical analysis

After data acquisition by LC-MS, the data obtained were imported into SciexOs and MS-DIAL in order to perform peak detection and alignment.

The detection of peaks occurs in each of the samples and the alignment is done between samples in order to decide if two certain characteristics (m/z, RT) found in two samples represent the same chemical component or not, that is, if their retention times and m/z values are both within the stated tolerances, they are considered the same feature.

In this way, the set of parameters applied for peak detection, in SciexOs, included a minimum retention time of 0 min and a maximum retention time of 35 min. In the case of MS-DIAL, for peak detection, a minimum peak height of 1000 amplitude and a mass slice width of 0.1 Da were selected. For alignment, the mass tolerance was 0.02 Da and the retention time tolerance 0.60 min.

Relative abundances for each feature were normalized using total area sums in Excel. This normalization technique is used to normalize each sample so that the resulting normalized samples have the same sum of area calculated using all peaks.

For a multivariate statistical analysis, we used MetaboAnalyst 5.0, where Partial Least Squares Discriminant Analysis (PLS-DA) was used to select the most promising features, selecting those with a variable importance in projection (VIP) score above 1.00. For the univariate approach, the Mann-Whitney U test performed in GraphPad Prism 9 was used to determine those that were statistically significant when the p-value was below 0.05.

5.5 – Metabolite identification

The identification of metabolites was performed using four databases, two online, including the Human Metabolome Database (HMDB) (https//hmdb.ca/) and METLIN

(https//metlin.scripps.edu/) and two libraries incorporated in the software used for data analysis, SciexOS and MS-DIAL. The identification was made by comparing the experimental mass measurement with the 4 databases within a mass tolerance window of 30 ppm.

However, in the HMDB database, the fragmentation data of each feature obtained by the DDA were still used to compare with the fragmentation spectrum in the databases through some established parameters, including a parent ion tolerance of 0.05 Da, a m/z tolerance of 30 ppm and a positive ionization mode. The maximum score corresponds to a Fit (%) of 1.00 and RFit (%) of 1.00. The fit score is calculated by comparing the percentage of peaks in the library spectrum to the experimental and RFit is the reverse.

In the case of the MS-DIAL internal library, an online library (http://prime.psc.riken.jp/compms/msdial/main.html) was loaded to identify the metabolites. In the case of the SciexOS software, the Library Search Algorithm method selected in Candidate Search was used. The internal libraries available for all analyzes were MSMLS™ (Mass Spectrometry Metabolite Library, IROA Technologies), with 299 entries, where the threshold for accepting an identification was 0.70 in the library score.

5.6 – Peptidomics

The same DDA and SWATH pools from the sample vials used for the metabolomics approach were also used for this peptidomics approach. Chromatographic separation was performed on a Triart C18 capillary column 1/32" (12 nm, S-3µm, 150 x 0.3 mm, YMC) at 5 µL/min, with an acetonitrile gradient as seen before in Table 4.

Two acquisition methods were also acquired, DDA and SWATH, as in the metabolomics approach. The process performed for both acquisitions was the same used in the metabolomics approach, previously mentioned in subchapter 5.3.2.

5.6.1 – Peptidomics: identification and quantification

Regarding the identification of the peptides, this was done using the ProteinPilot™ software (v5.0, Sciex). The paragon method parameters were: searched the revised Human SwissProt database (downloaded on 30th of January of 2023), no cysteine alkylation, no digestion and considering gel-based ID for the special factors.

Regarding the relative quantification of the peptides, this was performed on each sample (files from swath acquisition) using the SWATH™ processing plug-in in

PeakView^{TM} (v2.0.01, Sciex®). Quantitation results were obtained for peptides with less than 1% FDR and for the sum of up to 5 fragments/peptide. The abundance of each peptide was normalized by the total sum of the areas obtained in the metabolomic analysis.

6 – Results

      The participants included in this study were aged between 46 and 80 years, including 20 participants from the Aβ negative group and another 20 from the Aβ positive group. The Aβ negative group included 10 females and 10 males and the Aβ positive group included 11 females and 9 males. The Aβ negative group has an average age of 59.6 years, while the Aβ positive group has an average age of 63.8, being considered the oldest group. There are no significant differences between the two groups (Figure 19). When dividing the participants according to sex and diagnostic group, the Aβ positive male group is the oldest with a mean age of 61.5, and the youngest is the Aβ negative female group with a mean age of 57.6 years (Figure 20).
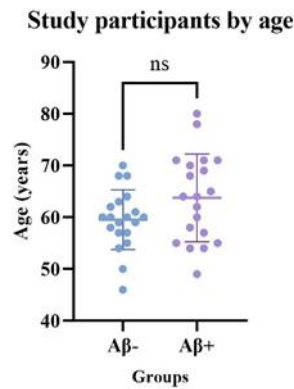
Figure 19 – Study participants by age according to the condition: data represents the mean ± SD regarding the age of the study participants. It represents the distribution, by age, of the 40 study participants divided into two groups according to the Aβ condition. The mean age corresponds to the dark-thickened line near the middle in the box plot. Significance (Student's two-tailed t-test) not significant (ns) between two groups Aβ- (blue) and Aβ+ (purple).
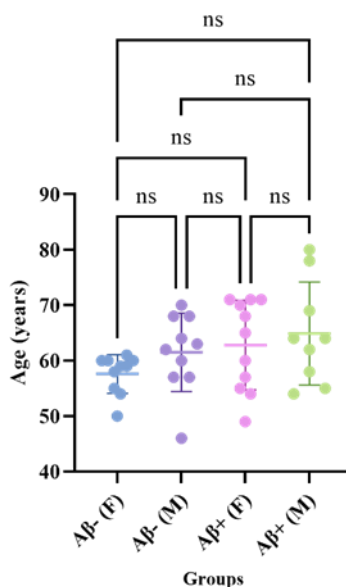
Figure 20 – Study participants by age according to the condition and sex: data represents the mean ± SD regarding the age of the study participants. It represents the
distribution, by age, of the 40 study participants divided into four groups according to the Aβ condition and sex (Aβ negative female (blue), Aβ negative male (purple), Aβ positive female (pink), and Aβ positive male (green)). The mean age corresponds to the dark thickened line near the middle in the box plot, drawn vertically. Significance (One-way ANOVA) not significant (ns) between groups.

Regarding the data obtained by SciexOS, all data were processed according to the parameters described in the previous chapter, and a total of 7956 features were obtained and normalized through the sum of the total area in an Excel spreadsheet. In this way, the number of detected peaks was reduced to 1863 features considering only those that were present in all samples. Thus, the 1863 features were used to perform a principal component analysis (PCA) applying the Pareto scale method (Figure 21), where the values centered on the mean are divided by the square root of the standard deviation.
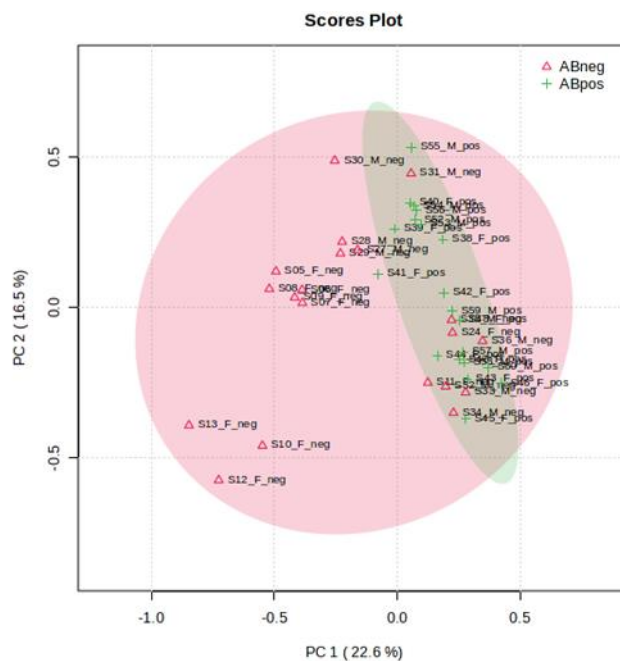
Figure 21 – PCA from SciexOS data: principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (green) groups using the 1863 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

This PCA scores graph shows the separation between the two groups in the first principal component (PC1), observing an overlap of some Aβ negative samples with Aβ positive ones, accounting for 22.6% of the data variance.

The second principal component (PC2) explains 16.5% of the data variance and there is a separation according to the age of the participants included in this study. In the Aβ negative group, we see the formation of three subgroups considering the age and sex of the participants (female samples aged between 59 and 61 years; female samples aged between 50 and 59 years; and male samples aged between 46 and 57) and the same effect is not observed in the Aβ positive group. In this case, nine Aβ negative samples (two women and seven men) showed a different behavior from the other samples of the same group.

Regarding the data obtained by MS-DIAL, once again, all data were processed according to the parameters described in the previous chapter and a total of 2811 features were obtained and normalized through the sum of the total area in an Excel spreadsheet. The number of detected peaks was reduced to 2461 features considering only those that were present in all samples and then a PCA was performed applying the Pareto scale method (Figure 22).
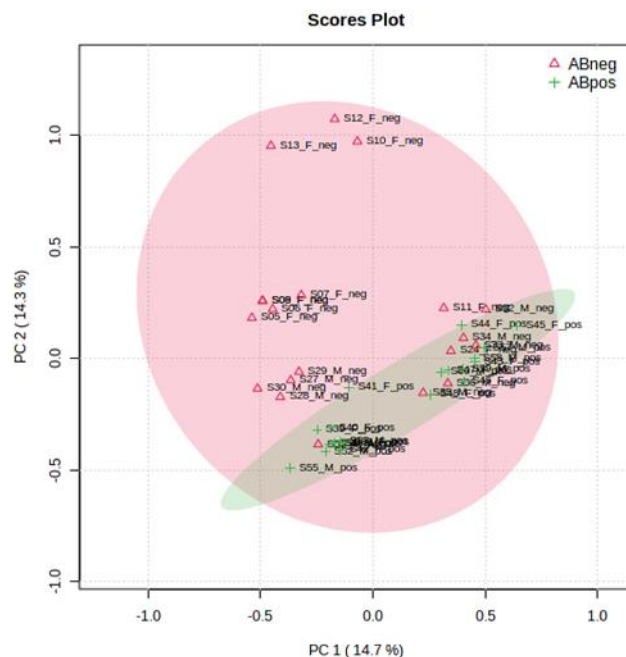
Figure 22 – PCA from MS-DIAL data: principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (green) groups using the 2461 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

In the PC1 it is observed, as in the data obtained by SciexOS, an overlap of some Aβ negative samples with Aβ positive ones, accounting for 14.7% of the data variance. PC2 explains 14.3% of the data variance and indicates a separation according to the age of the participants included in this study. As in the previous data, in the Aβ negative group there is the formation of three subgroups considering the age and sex of the participants (female samples aged between 59 and 61 years; female samples aged between 50 and 59 years; and male samples aged between 46 and 57) and the same effect is not seen in the Aβ positive group. In this case, eight Aβ negative samples (two women and six men) showed a different behavior from the other samples in the same group, showing a similar pattern to the samples from the Aβ positive group.

Then, a partial least squares discriminant analysis (PLS-DA) (supervised analysis) was performed under the same conditions as the PCA in order to improve the separation of the two groups.

In the data obtained by SciexOS, the PLS-DA analysis indicates a better separation in relation to the PCA graph, despite not showing a complete separation, where PC1 explains only 21.9% of the variability and PC2 6.8% (Figure 23).

This PLS-DA score plot had a correlation coefficient (R2) of 0.46 and a cross-validation correlation coefficient (Q2) of 0.26. However, this PLS-DA plot indicates that the two groups have different metabolic characteristics that allow their separation.
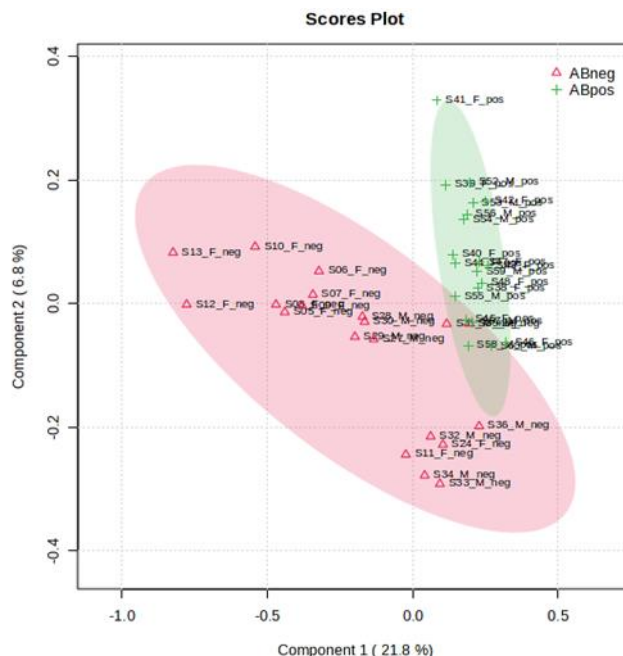
Figure 23 – PLS-DA analysis from SciexOS data: partial least-squares-discriminant (PLS-DA) scores for Aβ negative (red) and Aβ positive (green) groups using the 1863 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

Regarding the MS-DIAL, the PLS-DA analysis indicates a better separation in relation to the PCA plot, but a complete separation is not verified, where PC1 explains only 13.9% of the variability and PC2 4% (Figure 24).

The PLS-DA score chart had a correlation coefficient (R2) of 0.50 and a cross-validation correlation coefficient (Q2) of 0.19. However, as with the SciexOS data, this PLS-DA analysis indicates that the two groups have some different metabolic characteristics allowing their separation.
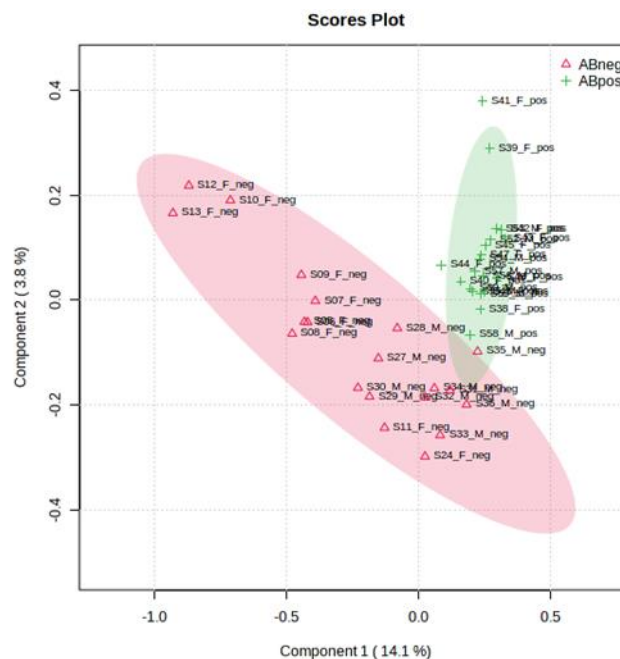
Figure 24 – PLS-DA from MS-DIAL data: partial least-squares-discriminant (PLS-DA) scores for Aβ negative (red) and Aβ positive (green) groups using the 2461 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

After that, an analysis was performed using only the features that had a p-value below 0.05 and a VIP score above 1.00.

In the case of SciexOS, the PCA graph was made using the Pareto scale method, where 144 statistically different features were used (Figure 25). It was verified that the results obtained in this PCA analysis showed a better separation in relation to the PCA that used the initial features.

The results of this PCA analysis showed better separation compared to the PCA that used the 1863 features (Figure 23). The PCA also showed that the variance explained by PC1 increased to 72.1%.

In the data obtained by MS-DIAL, 137 statistically different features were used (Figure 25). Once again, it was verified that the results obtained in this PCA analysis showed a better separation in relation to the PCA that used the initial 2461 features and the variance explained by PC1 increased to 70.9%.
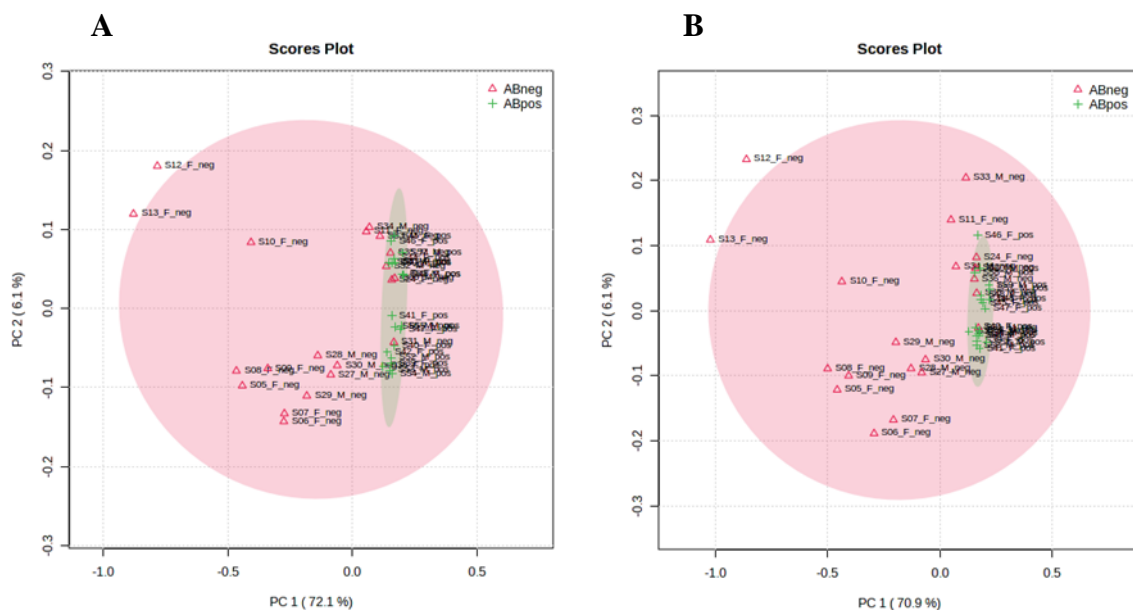
Figure 25 – PCA from SciexOS (A) and MS-DIAL (B) data: principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (green) groups using the features that were statistically different considering a P<0.05 and a VIP>1.00 and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data. **(A)** SciexOs data that used 144 features statistically different and **(B)** MS-DIAL data that used 137 features statistically different.

Then, a sample-sample correlation analysis was carried out between the two groups (Aβ negative and Aβ positive) through an analysis of Sperman's correlation coefficient in MetaboAnalyst 5.0 in two different ways, first without clustering and then with hierarchical clustering. In this way, it was possible to identify samples that did not correlate with samples from the same group. Spearman's correlation coefficient reveals a value between -1 and 1, with the value -1 indicating a perfect negative association between classifications and the value 1 indicating a perfect association of classifications. This coefficient is indicated in the graph through a degree of color intensity.

In both cases (data obtained by SciexOS and by MS-DIAL), in the correlation heatmap without hierarchical clustering, it is verified that in the Aβ negative group there are two female samples (S11_F_neg and S24_F_neg) and six male samples (S31_M_neg, S32_M_neg, S33_M_neg , S34_M_neg, S35_M_neg and S36_M_neg) that do not show the same standard correlation compared to the other samples of the Aβ negative group. These eight samples had already shown a different behavior from the remaining samples

of the same group, in the PCA made with the number of initial features. However, in the Aβ positive group, all samples have a similar correlation with each other (Figure 26).
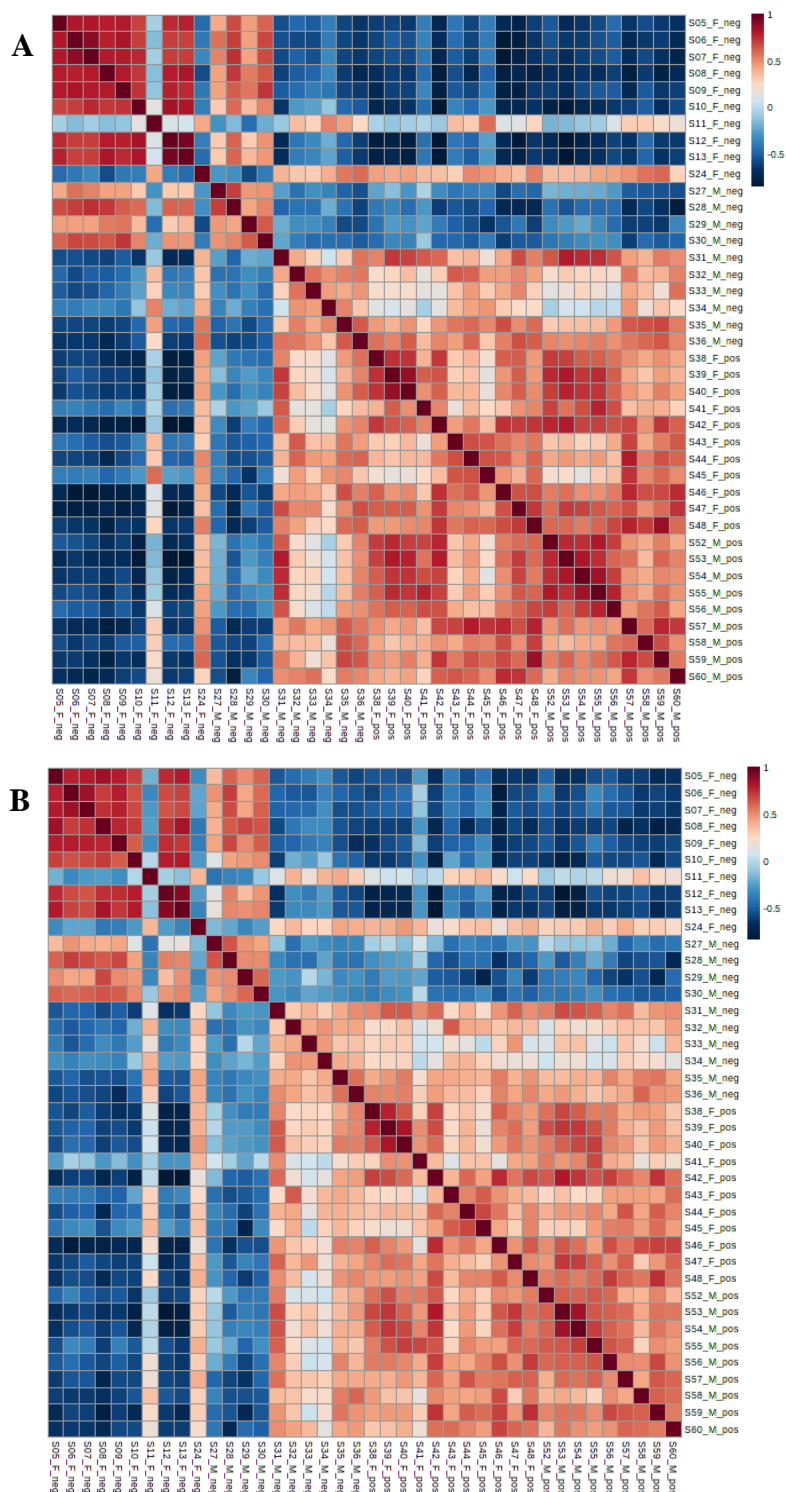
Figure 26 – Correlation heatmap from SciexOS **(A)** and MS-DIAL **(B)** data: correlation heatmap showing the Spearman correlation coefficient analysis of the pairwise comparison between samples. Each colored cell on the map indicates the correlation coefficient, with the scale code shown in the top right corner (red for positive correlations and blue for negative correlations).

In the correlation heatmap with hierarchical clustering, in both cases, the presence of two main groups is observed, one with most of the samples from the Aβ negative group and another with all the samples from the Aβ positive group plus the eight samples from the Aβ negative group (S11_F_neg, S24_F_neg, S31_M_neg, S32_M_neg, S33_M_neg, S34_M_neg, S35_M_neg and S36_M_neg).

In the hierarchical cluster analysis, we can also observe the formation of subgroups within the two main clusters. In the case of SciexOS, in the first main cluster, there is a cluster for male samples and a cluster for female samples together with only one male sample from the Aβ negative group. In the second main cluster there is no separation by sex, one cluster consists of two samples from the Aβ negative group and the other by twenty-seven samples from both groups and both sexes (Figure 27).
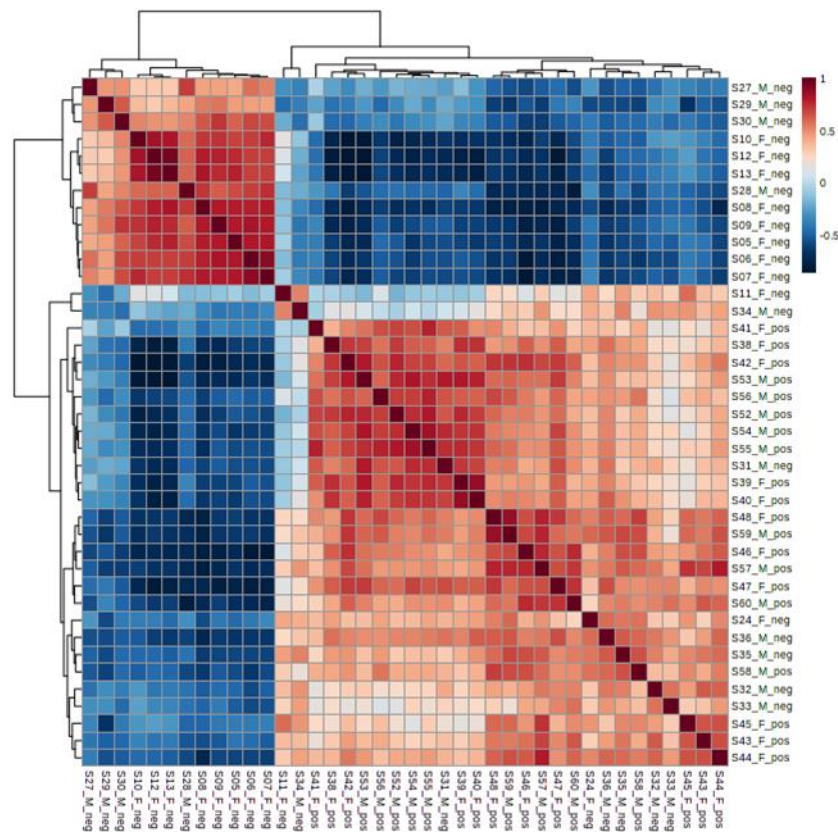
Figure 27 – Correlation heatmap with hierarchical clustering from SciexOS data: correlation heatmap showing the Spearman correlation coefficient analysis of the pairwise comparison between samples. Samples order was determined by hierarchical clustering. Each colored cell on the map indicates the correlation coefficient, with the scale code shown in the top right corner (red for positive correlations and blue for negative correlations).

Regarding the data obtained by MS-DIAL, in the first main cluster, there is a cluster for male samples and a cluster for female samples of the Aβ negative group. In the second main cluster there is also no separation by sex, with a cluster consisting of twenty-one samples from both groups and another cluster of eight samples from both groups (Figure 28).
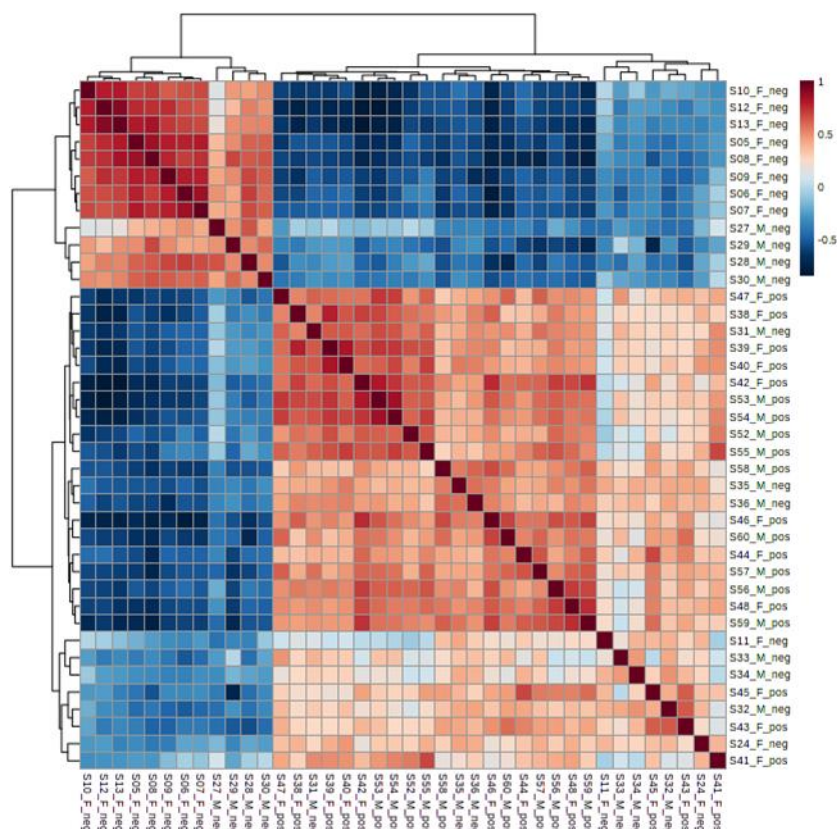
Figure 28 – Correlation heatmap with hierarchical clustering from MS-DIAL data: correlation heatmap showing the Spearman correlation coefficient analysis of the pairwise comparison between samples. Samples order was determined by hierarchical clustering. Each colored cell on the map indicates the correlation coefficient, with the scale code shown in the top right corner (red for positive correlations and blue for negative correlations).

A hierarchical cluster analysis was also carried out between the samples of the two groups and the statistically different features.

In the correlation heatmap, in relation to the data obtained by SciexOS, the presence of two main groups of samples was verified and correlated with these two main groups; there are two main groups of metabolites associated with them. As seen in Figure 29, most metabolites are more expressed in the Aβ negative condition compared to the Aβ positive condition plus the eight samples from the Aβ negative group (S11_F_neg,

S24_F_neg, S31_M_neg, S32_M_neg, S33_M_neg, S34_M_neg, S35_M_neg and S36_M_neg).
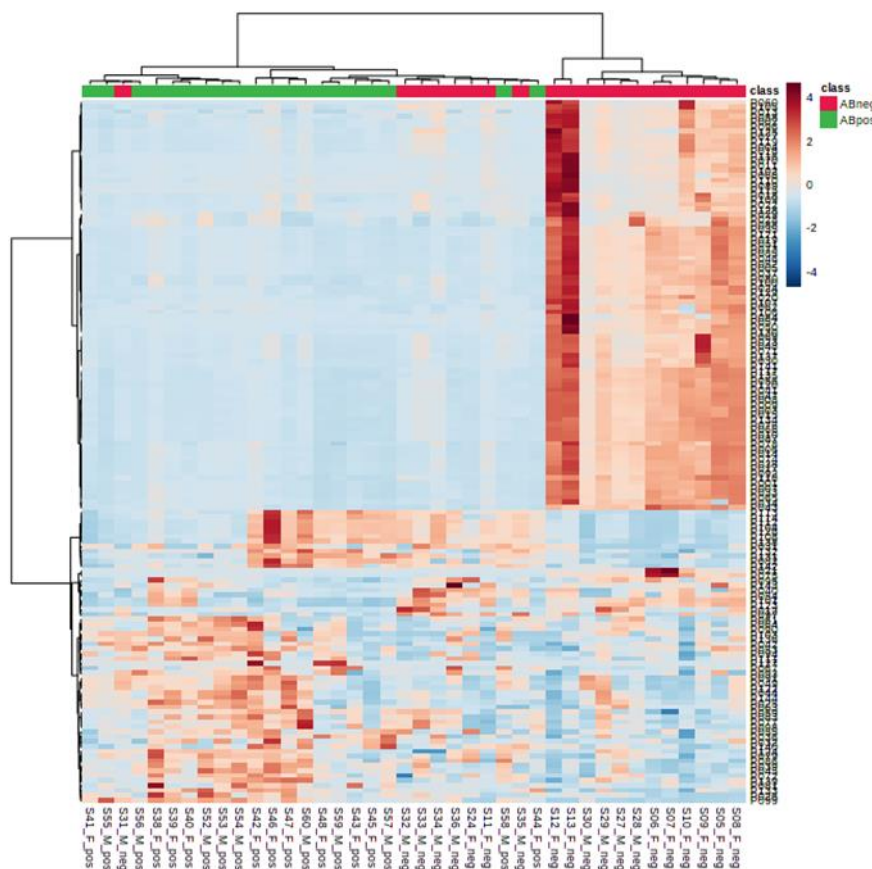
Figure 29 – Heatmap with hierarchical clustering from SciexOS data: heatmap with hierarchical clustering analysis by Euclidian distance using Ward's method generated from the 144 significantly different features between the two groups. The columns represent the individual samples, and the rows indicate differentiating metabolites. On top of the heatmap, Aβ negative samples correspond to the red squares and Aβ positive samples to the green squares. The dendrogram for samples is shown on top of the heatmap and the metabolite dendrogram is on the left side of the heatmap. Each colored cell on the map indicates each feature's relative abundance, with the scale code shown in the top right corner (red for higher abundances and blue for lower abundances).

In the first main cluster, which includes samples from the Aβ positive and Aβ negative groups, we can observe the existence of four groups: the first with samples S31_M_neg, S56_M_pos, S41_F_pos and S55_M_pos; the second with samples S38_F_pos, S39_F_pos, S40_F_pos, S52_M_pos, M53_M_pos and M54_M_pos; the third with S46_F_pos, S42_F_pos, S47_F_pos and S60_M_pos; and the fourth with S59_M_pos, S48_F_pos, S57_M_pod, S43_F_pos, S44_F_pos, S45_F_pos, S56_2_M_pos, S33_M_neg, S32_M_ne, S11_F_neg, S34_M_neg, S36_M_neg, S24_F_neg, S35_M_neg and S58_M_neg.

In the second main cluster, which corresponds to the Aβ negative group, there are two different clusters, one includes only two female samples and the other is divided into male and female samples.

In the correlation heatmap, regarding the MS-DIAL data, the presence of two main groups of samples was again verified (Figure 30). It is also verified in this case that most of the metabolites are more expressed in the Aβ negative group compared to the samples of the Aβ positive group together with eight samples of the Aβ negative group (S11_F_neg, S24_F_neg, S31_M_neg, S32_M_neg, S33_M_neg, S34_M_neg, S35_M_neg and S36_M_neg).
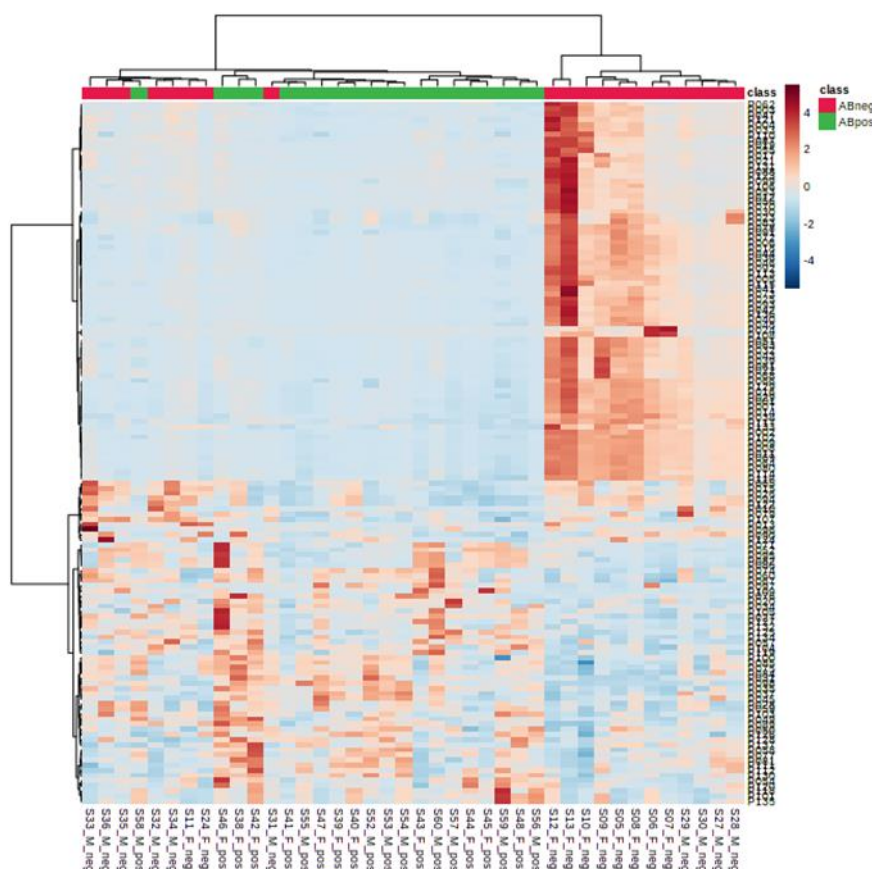
Figure 30 – Heatmap with hierarchical clustering from MS-DIAL data: heatmap with hierarchical clustering analysis by Euclidian distance using Ward's method generated from the 137 significantly different features between the two groups. The columns represent the individual samples, and the rows indicate differentiating metabolites. On top of the heatmap, Aβ negative samples correspond to the red squares and Aβ positive samples to the green squares. The dendrogram for samples is shown on top of the heatmap and the metabolite dendrogram is on the left side of the heatmap. Each colored cell on the map indicates each feature's relative abundance, with the scale code shown in the top right corner (red for higher abundances and blue for lower abundances).

60

In the first main cluster, which includes samples from the Aβ positive and Aβ negative groups, we can observe the existence of four groups: the first with samples S33_M_neg, S32_M_neg, S34_M_neg, S36_M_neg, S11_F_neg and S24_F_neg; the second with samples S46_F_pos, S43_F_pos and S60_M_pos; the third with S38_F_pos, S47_F_pos, S52_M_pos, S54_M_pos, S39_F_pos and S53_M_pos; and the fourth with S42_F_pos, S59_M_pos, S58_M_pos, S48_F_pos, S56_M_pos, S57_M_pos, S44_F_pos, S45_F_pos, S56_2_M_pos, S41_F_pos, S31_M_neg, S55_M_pos, S35_M_neg and S40_F_pos.

In the second main cluster, which corresponds to the Aβ negative group, there are two different clusters, one includes only two female samples and the other is divided into male and female samples, these being the same samples from the analysis made by SciexOS.

The fold change (FC) values describe how much a metabolite varies between two groups. A variance value of 1.5 means a 50% increase between the two groups and a value of 0.67 means there is a 50% decrease.

In the data obtained from SciexOS, of the 144 features significantly different between the two groups, 11 increased significantly (change greater than 1.5) and 94 significantly decreased (change less than 0.67) in the Aβ positive group. In the case of MS-DIAL, of the 137 features significantly different between the two groups, 17 significantly increased and 84 significantly decreased in the Aβ positive group.

Using logarithm to base 2 allows an easier interpretation of the graph, where significantly increased features will have a positive value and significantly decreased features will have a negative value.

In the case of SciexOS analysis, features with 2-fold increase are displayed on axis 1, and in this case, there are four features (216.10_3.52, 796.33_24.14, 851.39_24.94, and 851.38_24.98) that have a 2-fold or more increase in the Aβ positive group.

In the case of MS-DIAL there are five features also with a 2-fold increase (216.10_3.577, 851.39_24.991, 251.01_23.023, 767.21_6.545 and 744.76163_14.52) (Figure 31).
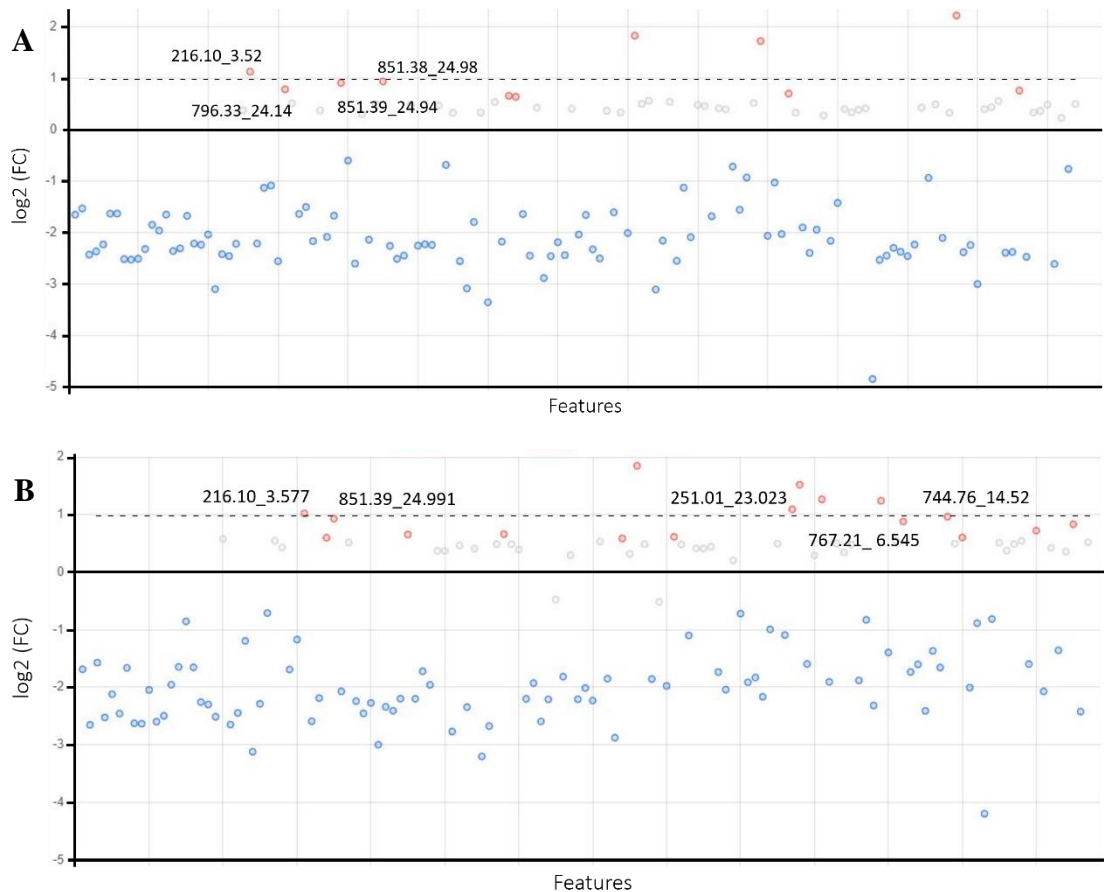
Figure 31 – FC values from SciexOS **(A)** and MS-DIAL **(B)** data: the graph demonstrates the log2 fold change (Aβ + vs. Aβ -) of the means for all the significant features. The red color shows the features significantly increased with a fold change higher than 1.5, and the blue color shows the features significantly down with a fold change lower than 0.67.
(A) SciexOs data that used 144 features statistically different and **(B)** MS-DIAL data that used 137 features statistically different.

Until now, the analysis carried out on the CSF samples was performed considering two groups related to β-amyloid, where the analysis revealed a tendency to separate the two groups considering the sex of the study participants. To better understand this difference between sexes, a new PCA and PLS-DA analysis was made using the statistically different features and the Pareto scale method considering four groups: Aβ negative female, Aβ negative male, Aβ positive female and Aβ positive male.

In both analysis (data obtained by SciexOS and by MS-DIAL), the PCA plot shows a tendency to separate Aβ negative from Aβ positive samples and, in addition, a separation between the two sexes can be observed in the Aβ negative condition (Figure 32). On the other hand, in the Aβ positive group, there is an overlap between the two sexes. Samples S11_F_neg, S24_F_neg, S31_M_neg, S32_M_neg, S33_M_neg,

S34_M_neg, S35_M_neg and S36_M_neg are samples from the Aβ negative group that show a different behavior from those of the same group, as already verified in the previous analysis.
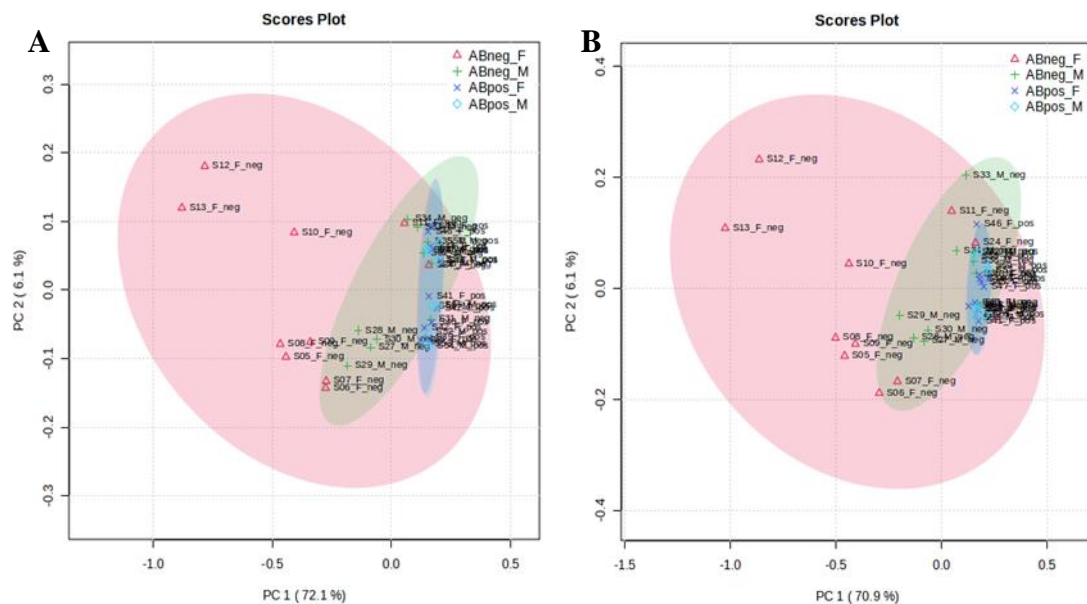
Figure 32 – PCA from SciexOS **(A)** and MS-DIAL **(B)** data: principal component analysis (PCA) scores for Aβ negative female (red), Aβ negative male (green), Aβ positive female (purple) and Aβ positive male (blue) groups using the features that were statistically different considering a P<0.05 and a VIP>1.00 and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data. **(A)** SciexOs data that used 144 features statistically different and **(B)** MS-DIAL data that used 137 features statistically different.

A better separation of the four groups was obtained in the PLS-DA analysis (Figure 33). However, samples S11_F_neg, S24_F_neg, S31_M_neg, S32_M_neg, S33_M_neg, S34_M_neg, S35_M_neg and S36_M_neg are the reason why there continues to be an overlap of the four groups.

In both graphs of the multivariate analysis, the two sexes of the Aβ positive condition did not have any significant difference.

From this analysis of the four groups, it is possible to conclude that there is a difference between the sexes, mainly in the Aβ negative condition.
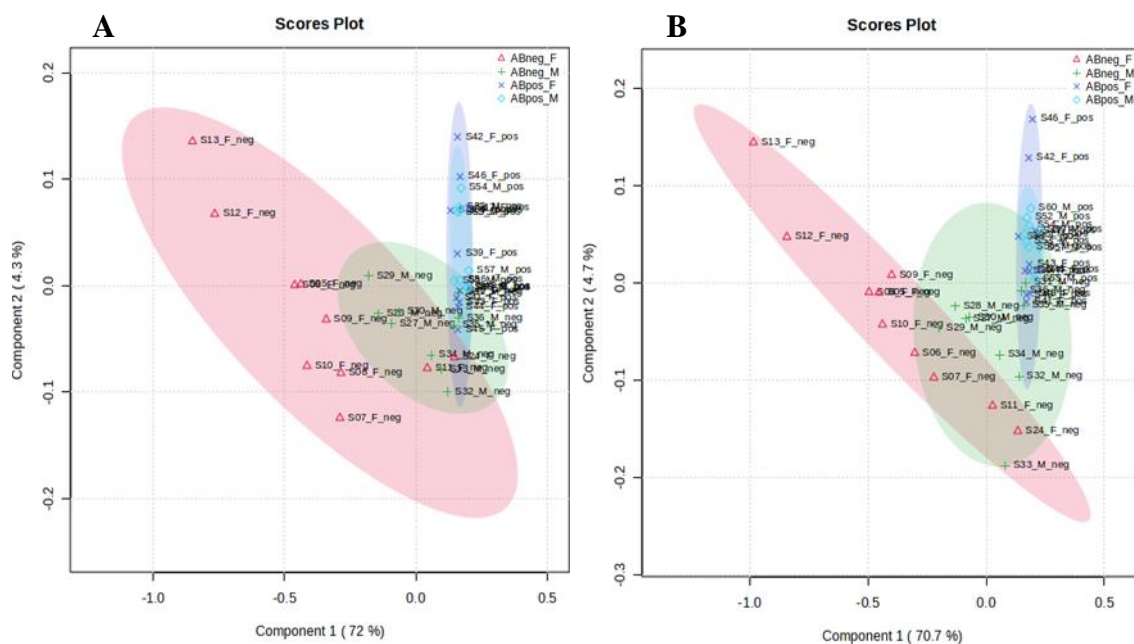
Figure 33 – PLS-DA from SciexOS **(A)** and MS-DIAL **(B)** data: partial least-squares-discriminant (PLS-DA) scores for Aβ negative female (red), Aβ negative male (green), Aβ positive female (purple) and Aβ positive male (blue) groups using the features that were statistically different considering a P<0.05 and a VIP>1.00 and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data. **(A)** SciexOs data that used 144 features statistically different and **(B)** MS-DIAL data that used 137 features statistically different.

During this analysis, as we can see in Figures 28 and 29 there are features (features are presented according to their mass-to-charge ratio and retention time (m/z/RT)) that indicate that there are metabolites that differ between the two conditions. However, in both cases (data obtained by SciexOS and by MS-DIAL), it is observed that the feature P1266 (226.1804/20.28) and the feature P0736 (226.1835/20.28), respectively, have a very high VIP value compared to the remaining features, which leads us to believe that these metabolites allow differentiation only between sexes in the Aβ negative condition (Figure 34).
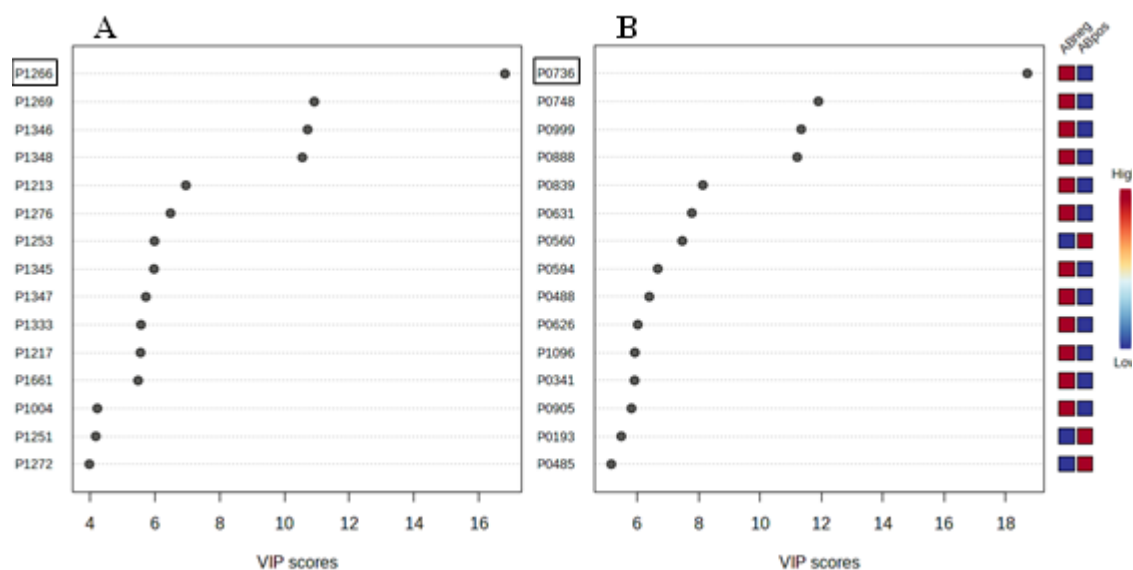
Figure 34 – Top 15 features with the highest VIP scores in SciexOS **(A)** and MS-DIAL **(B)** data: features that have the highest variable importance in projection (VIP) value in an analysis made with the inicial features. **(A)** SciexOs data that used 1863 features statistically different and **(B)** MS-DIAL data that used 2461 features statistically different.

Since features P1266 (226.1804/20.28) and P0736 (226.1835/20.28) have very similar m/z values and retention time, we chose to perform the analysis only for the data obtained by SciexOS and perform the analysis again, without the features correlated to feature P1266 (226.1804/20.28), establishing a cut-off value of 0.75, to visualize again the correlation between the samples and the metabolites, in order to be able to verify which metabolites allow to differentiate the two conditions.

As mentioned before, we excluded all features correlated with feature P1266. So, for further analysis we have 1734 features to perform a PCA applying the Pareto scale method, where values centered on the mean are divided by the square root of the standard deviation.

In PC1 it was observed a total overlap of Aβ negative samples with Aβ positive ones, being responsible for 20.7% of the data variance (Figure 35). However, in this case, subgroups are not formed in the Aβ negative condition, as had happened in the initial analysis.
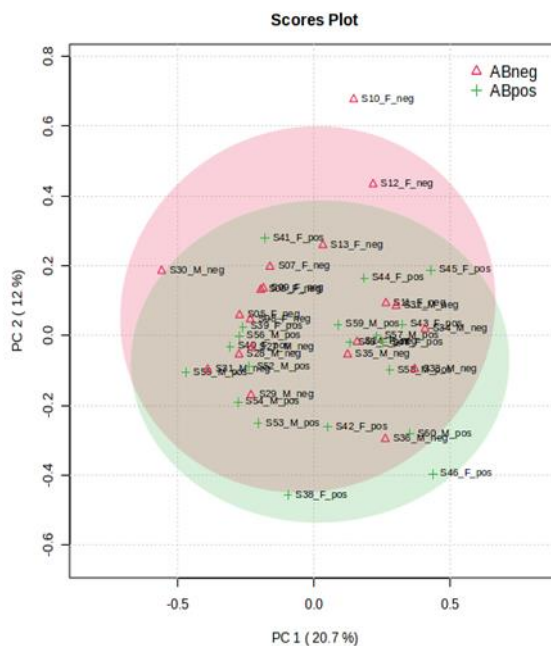
Figure 35 – PCA from data without feature P1266: principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (green) groups using the 1734 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

Then, a PLS-DA analysis was carried out, under the same conditions as the PCA, to improve the separation of the two groups. In this PLS-DA analysis it is possible to observe a better separation in relation to the PCA, although a complete separation is not verified, where PC1 explains only 11% of the variability and PC2 10.9% (Figure 36). The PLS-DA score plot had a correlation coefficient (R2) of 0.96 and a cross-validation correlation coefficient (Q2) of 0.42. However, as in the previous analysis, it is possible to verify that the two groups have some different metabolic characteristics that allow separation.
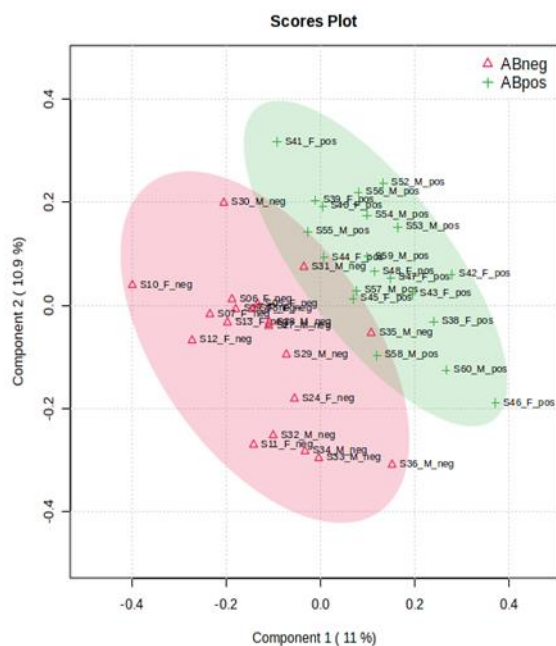
Figure 36 – PLS-DA analysis from data without feature P1266: partial least-squares-discriminant (PLS-DA) scores for Aβ negative (red) and Aβ positive (green) groups using the 1734 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

Then, a heatmap was made with features with a p-value below 0.05 (Figure 37) and another with features that have a VIP score above 1.00 (Figure 38). This was done to select the features that showed which metabolites differed between the Aβ negative condition and the Aβ positive conditions.

From this analysis, it was verified that the heatmap referring to features that have a p-value below 0.05, showed a better separation of the metabolites that allow to distinguish both groups. For this reason, it was decided to select the features that were present in the first cluster of metabolites to carry out a receiver operational characteristic (ROC) analysis; to characterize the predictive value of these metabolites, using MetaboAnalyst 5.0, to discover potential biomarkers.
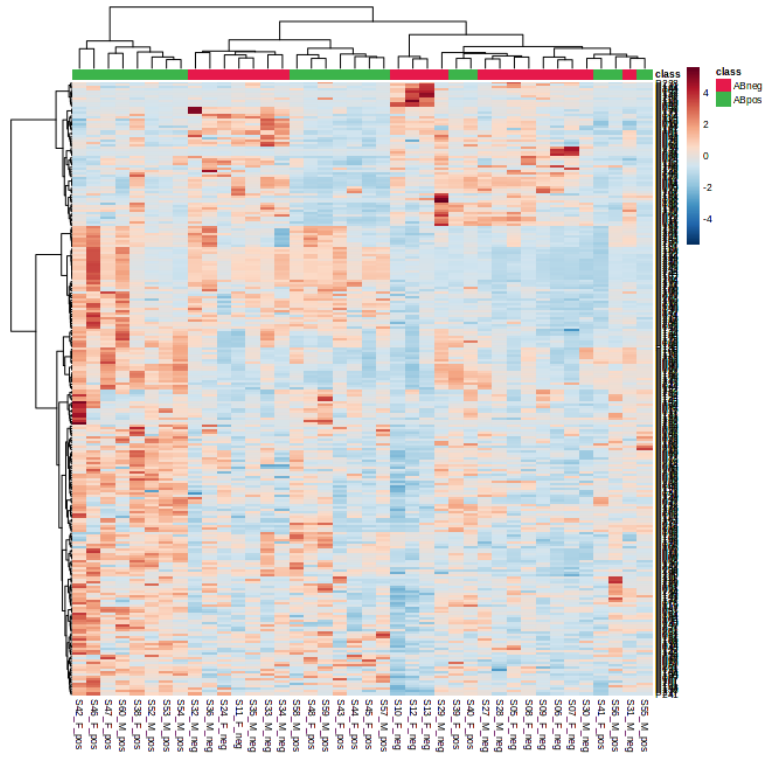
Figure 37 – Heatmap with hierarchical clustering from data with P<0.05: heatmap with hierarchical clustering analysis by Euclidian distance using Ward's method generated from the 263 significantly different features between the two groups.
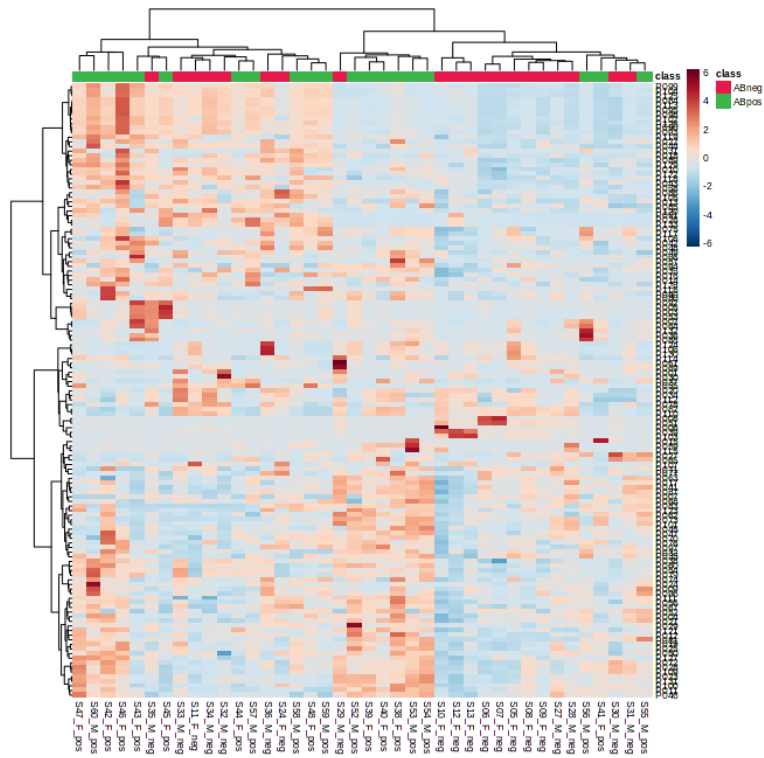
Figure 38 – Heatmap with hierarchical clustering from data with VIP >1.00: heatmap with hierarchical clustering analysis by Euclidian distance using Ward's method generated from the 133 significantly different features between the two groups.

In this way, 62 features were selected for a ROC analysis. However, as we were facing a large number of features, it was decided to use only those with a VIP score above 1.00, so we were using 20 features for further analysis.

In a ROC curve graph, the x-axis indicates the false positive rate (specificity 1), and the y-axis indicates the true positive rate (sensitivity), generating the area under the ROC curve (AUC). AUC is a measure of the accuracy of a diagnostic test, where a value of 1 indicates that the classifier is able to perfectly distinguish between the two groups, a value greater than 0.9 indicates excellent discrimination, a value greater than 0.8 is good discrimination, a value greater than 0.7 is a moderate test, a value between 0.51 and 0.69 is a bad test, and a value of 0.5 shows no discrimination between the two groups [128].

In Figure 39 it is possible to observe the combination of metabolites using logistic regression analysis to improve prediction. The ROC curves created are based on the cross-validation performance of multivariate algorithms using the Linear SVM classifier (support vector machine).
In these ROC curves the ideal is to see a true positive rate of 1, meaning that all "disease" groups are considered positive, and a false positive rate of 0, meaning that none of the "disease" samples were considered controls.
Therefore, an AUC value of 1 or close to this value will be considered the best curve. In this analysis of multivariate ROC curves, the best one used 7 features (Figure 40).

However, it is possible to verify that all models presented an AUC value greater than 0.87, which is why they are good models for discriminating the two groups in studies.
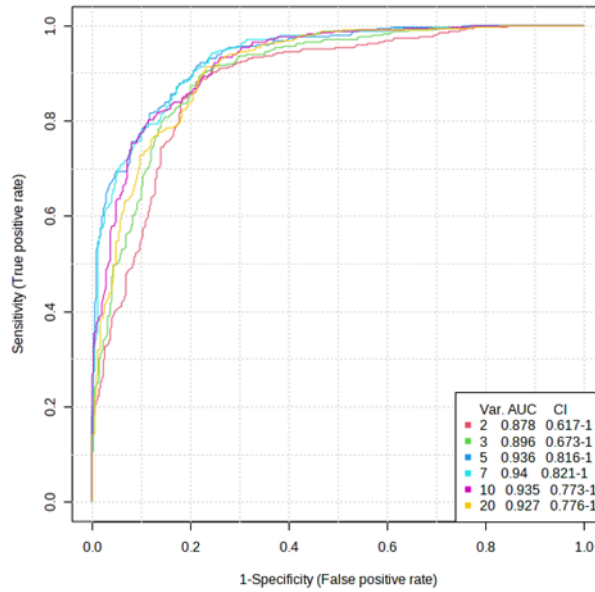
Figure 39 – ROC curves of the significantly different features: multivariate Receiver Operating Characteristic (ROC) curves, considering 2, 3, 5, 7, 10, and 20 features. A Linear Support Vector Machines (SVM) classification method and SVM built-in feature classification method were used in this analysis.
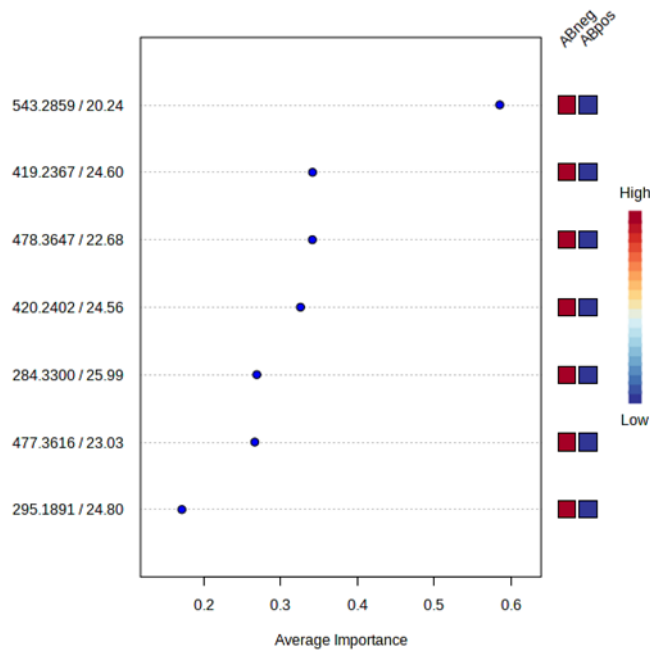
Figure 40 – Features contributions to the model's classification accuracy: features were ranked by their contributions to the model's classification accuracy that uses 7 variables.

Thus, from this analysis, these 7 features were manually quantified, which are present in Figure 41, and a statistical analysis of them was also carried out.
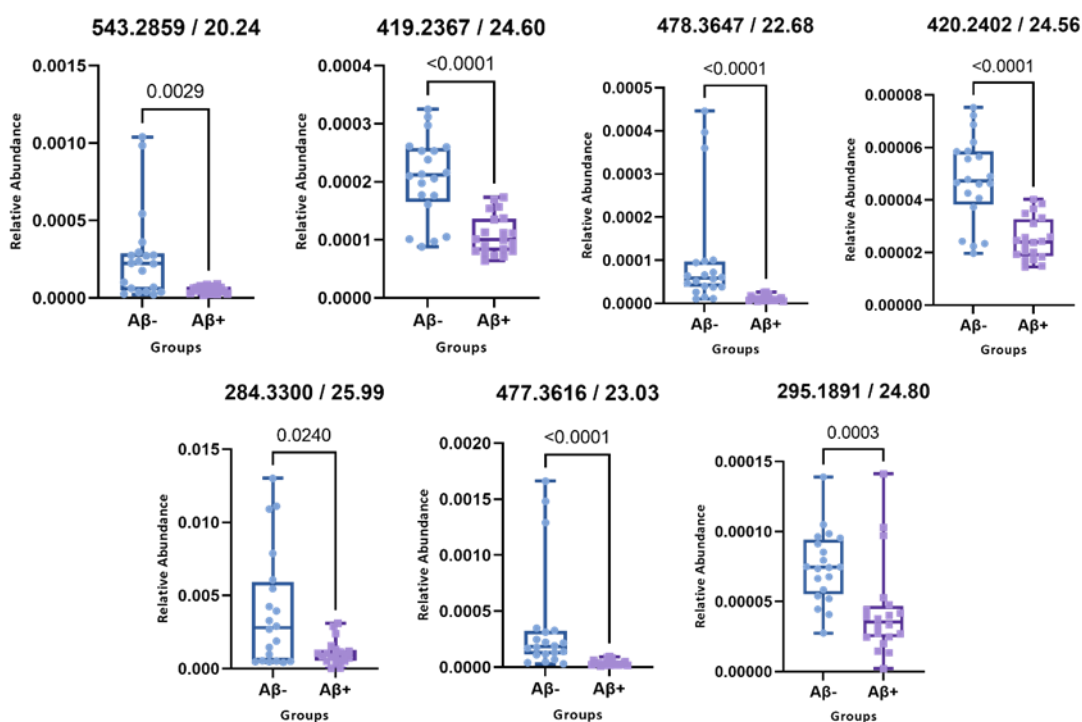
Figure 41 – Boxplots of statistically different features: boxplots of statistically different features, in the control group (blue) and disease group (purple). The plots show the relative abundance of each sample, the median, the minimum and the maximum of each group.

By observing these boxplots, we can verify that all features are statistically different. It can be seen that all features have a higher average in the control group (Aβ negative) than in the disease group (Aβ positive), this means that these metabolites are increased in the samples from the Aβ negative group, as we already had seen in the heatmap represented in Figure 37.

After that, the univariate and multivariate ROC curves were made again, with all newly quantified features. Feature 478.3647/22.68 was the one that presented the best AUC, 0.96, in the univariate analysis, as we can see in Figure 42.
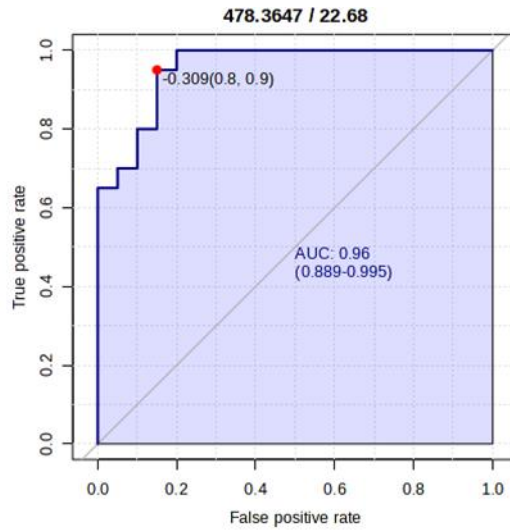
Figure 42 – ROC curve of the feature with the best AUC: univariate receiver operating characteristic (ROC) curve of the feature with the best AUC, 478.3647/22.68.

Then, a multivariate analysis of these 7 features was carried out, as shown in Figure 43. We can see that the multivariate ROC curves showed that the features 543.2859/20.24, 284.3300/25.99, 478.3647/22.68, 419.2367/24.60 and 420.2402/24.56 were the best to characterize the differences between the Aβ negative group and the Aβ positive group. The AUC of this curve was 0.967, which is considered a curve with excellent discrimination.
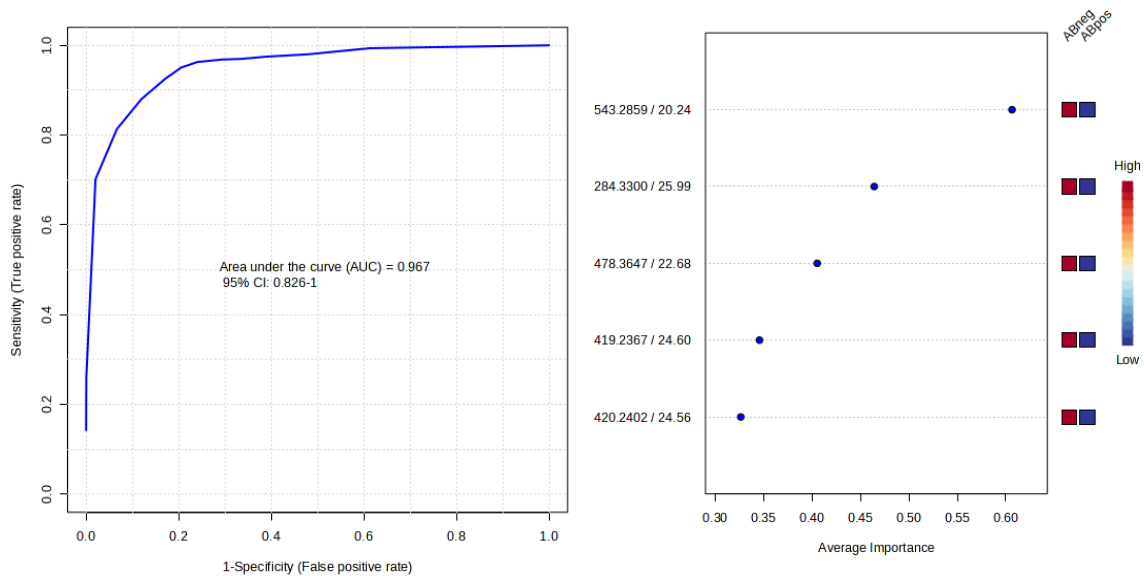
Figure 43 – ROC curve and the 5 features of the model with the best AUC value: the best multivariate Receiver Operating Characteristic (ROC) curve (left), 5 feature model, and the features referent to the model (right). A Linear Support Vector Machines (SVM) classification method and SVM built-in feature classification method were used to perform this analysis.

Regarding the identification of metabolites, 28 features were not identified, 7 features obtained more than one match, 26 features have a match in only one database and 1 feature obtained the same match in all databases.

Feature 298.0980/7.84 was the only that obtained the same correspondence in all databases used in this work and was identified as 5'-methylthioadenosine. The detailed results of the correspondences for each feature considering only the experimental mass/charge of the precursor are provided in supplementary Tables A.2.1 to A.2.4 of Appendix A.2. In Figure 44, it is possible to compare the acquired spectrum with the spectrum present in the SciexOS internal library for the feature 298.0980/7.84.
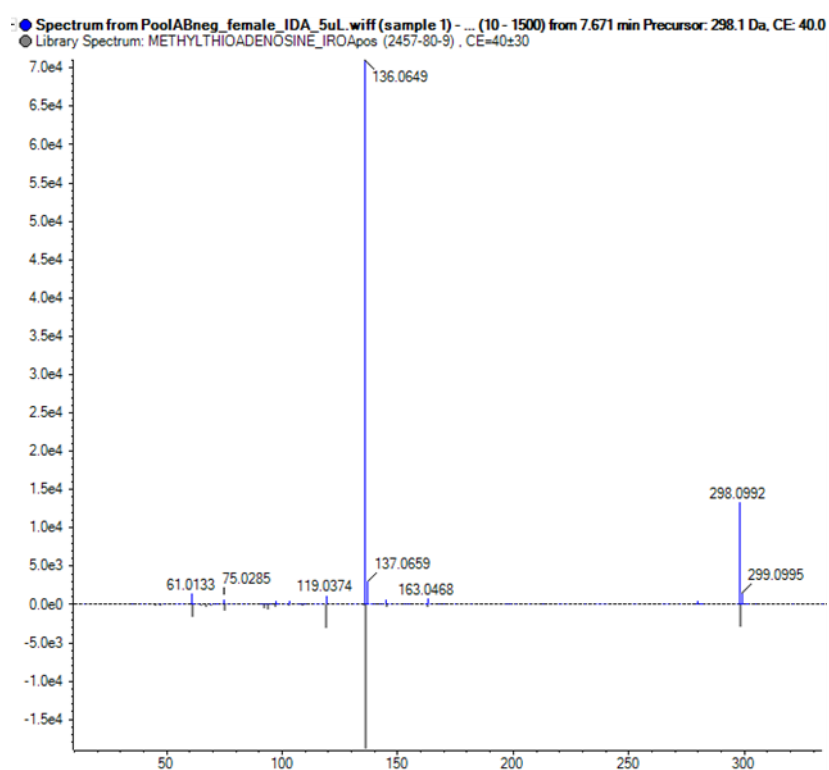
Figure 44 – Mass spectrum obtained for the feature 298.0980/7.84: comparison of the experimental spectrum of the feature 298.0980/7.84 and the spectrum from the SciexOS internal library of the matched metabolite. The experimental MS/MS spectrum (blue stripes) is shown on top, and the library MS/MS spectrum (grey stripes) is shown underneath.

As previously mentioned, there were 7 features (161.0985/24.00, 253.1402/23.83, 313.2325/26.82, 317.1707/24.84, 357.2580/26.82, 467.3562/26. 75 and 489.3351_26.89) representing different matches in the databases. However, the HMDB online database and the MS-DIAL software database were those that presented a greater number of identified metabolites, leading to the belief that there may be a lack of information on MS/MS spectra of some metabolites in the METLIN and SciexOS databases.

For example, (9Z,12E)-15,16-dihydroxyoctadeca-9,12-dienoic acid was correlated with the feature 313.2325/26.82 in the MS-DIAL software database, but in the METLIN database this feature was identified as fruticosonine. When searching for the metabolite (9Z,12E)-15,16-dihydroxyoctadeca-9,12-dienoic acid in the METLIN database, there is no information on the MS/MS spectrum of this metabolite.

However, all incompatibilities are described in supplementary Table A.2.5 of Appendix A.2, where all identified metabolites are described, as well as the associated pathways.

Then, a pathway analysis of the features identified was performed in Metaboanalyst 5.0. For this analysis, the names of the compounds obtained in the databases and the hypergeometric test and the relative-betweenness centrality were used, to identify the most relevant pathways (Figure 45 and Table A.2.6 of Appendix A.2).
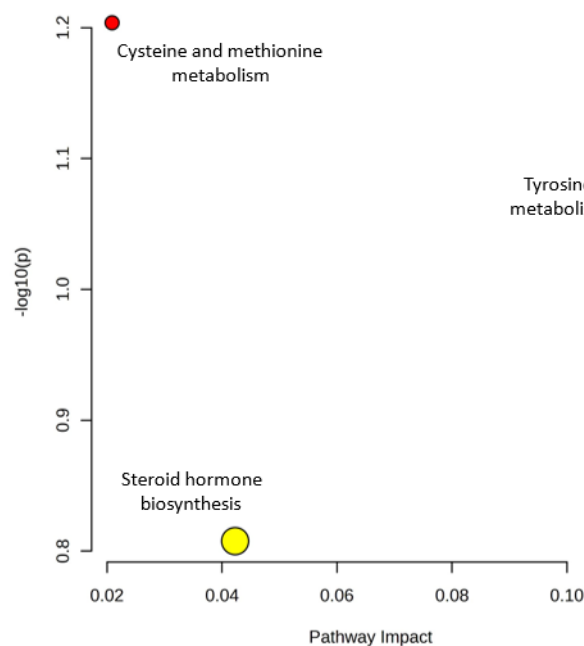
Figure 45 – Pathway analysis results for features identified: pathway analysis results for the features identified in the CSF of Aβ negative and Aβ positive groups. The x-axis represents the pathway impact, and the y-axis is the log of the p-value obtained from the pathway enrichment analysis. The node's color and size are based on its p-value and pathway impact values, respectively, where the pathways are represented: cysteine and methionine metabolism (red), tyrosine metabolism (orange) and the steroid hormones biosynthesis (yellow). The most significantly changed pathways are characterized by a high log(p) and a high impact value.

In this graph, the x-axis represents the pathway impact, and the y-axis is the logarithm of the p-value obtained from the pathway enrichment analysis. The node color and size are based on its p-value and path impact values, respectively.

The most significantly altered pathways are characterized by a high log(p) value and a high impact value (top right region), as in the case of tyrosine metabolism, with a pathway impact of 0.11085 and a p-value of 0.079157, where the metabolite 3,4-dihydroxy-L-phenylalanine (210.1443/14.70) was the only identified metabolite enriched in this pathway.

In the case of steroid hormone biosynthesis, the corresponding feature was 301.1770/25.68, which was identified as 19-oxoandrost-4-ene-3,17-dione, with a pathway impact value of 0.04229 and a p-value of 0.15575.

Finally, the metabolism of cysteine and methionine was the pathway with the least impact, with a value of 0.020809 and a p-value of 0.06256, being enriched with the metabolite 5'-methylthioadenosine (298.0980/7.84).

The metabolites 3,4-dihydroxy-L-phenylalanine and 19-oxoandrost-4-ene-3,17-dione were increased in the CSF of the Aβ negative group (Figure 46) and as previously mentioned, they come from tyrosine metabolism and steroid hormone biosynthesis, respectively.

The metabolite 3,4-dihydroxy-L-phenylalanine, also known as L-DOPA, is a neurotransmitter dopamine precursor produced by the action of the enzyme tyrosine hydroxylase on the amino acid tyrosine. According to the literature [129] Alzheimer's disease is characterized by the aggregation of peptide A that leads to the development of amyloid plaque formation, followed by neurodegenerative alterations. In this way, the metabolite L-DOPA and dopamine can dissolve the A-peptide fibrils and also inhibit the formation of protein tangles. According to the literature, the administration of L-DOPA in patients with AD improves their learning and, consequently, their memory status. However, a study carried out by Oliver Ambrée and his collaborators [130] showed that dopamine levels, in TgCRND8 mice, were reduced in the hippocampus of mice with Alzheimer's. This reduction may be associated with a progression of the pathology that primarily affects the hippocampus and neocortex. However, it was not possible to find studies on L-DOPA levels in the CSF of patients with Alzheimer's disease.

Regarding the metabolite 19-oxoandrost-4-ene-3,17-dione, this was also found to be reduced in the Aβ positive group, being related to the biosynthesis of steroid hormones. According to the literature [131], alterations that occur in the intestinal microbiota (GMB) contribute to the development of neurodegenerative diseases, such as AD. In Alzheimer's patients, there is a decrease in the diversity of the microbiota; for this reason, alterations in the intestinal microbiota can be used for the early detection of Alzheimer's disease. In a study by Jianxiong Xi and colleagues [131], Alzheimer's patients showed changes in their fecal metabolome compared to the control group and also showed changes in the metabolic pathway of steroid hormone biosynthesis, suggesting that the GMB mechanism causes the development of AD. Thus, in this study, the 19-oxoandrost-4-ene-3,17-dione metabolite decreased in AD patients compared to the control group and was shown to be involved in estrogen biosynthesis. This plays a fundamental role in the preservation of neurons and in the repair of neurons damaged by Alzheimer's disease, being used more by healthy individuals than by Alzheimer's patients. Therefore, the decrease in the 19-oxoandrost-4-ene-3,17-dione metabolite may influence AD pathology by estrogen downregulation.

On the other hand, the 5'-methylthioadenosine (MTA) metabolite was increased in the Aβ negative group (Figure 46). This metabolite is produced through S-adenosylmethionine in the metabolism of cysteine and methionine [132]. According to the literature, MTA has regulatory functions such as gene expression, proliferation, differentiation, apoptosis and also acts as a metabolic intermediary in the methionine cycle [133]. A study in the literature [107] demonstrated that patients in the MCI phase had increased levels of MTA in the CSF compared to patients who were in the MCI due to AD phase, as happened in this study.
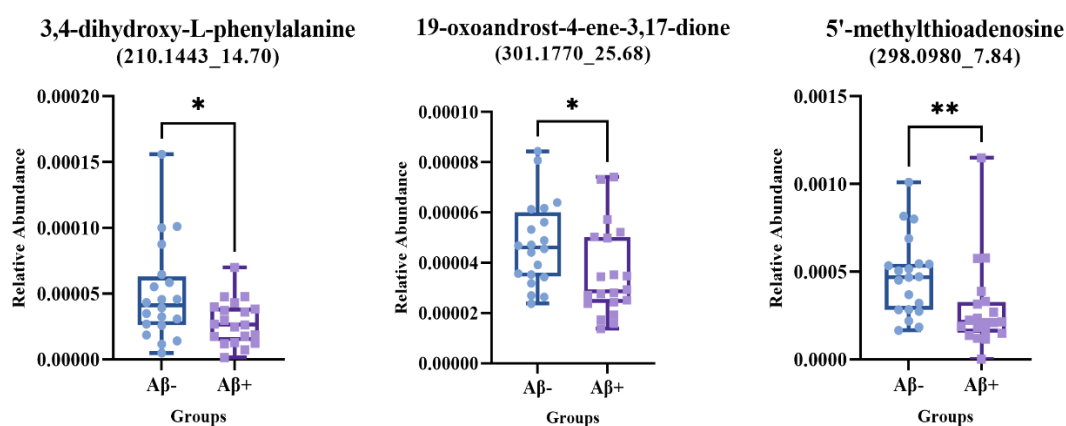
Figure 46 – Boxplot of features identified: boxplots representing three metabolites identified in the CSF of Aβ negative (blue) and Aβ positive (purple) groups. The plots show the relative abundance of each sample, the median, the minimum and the maximum of each group. The level of significant (Mann-Whitney two-tailed test) is represented by *: *P<0.05 and **P<0.01.

Then, ROC curve analysis was performed again for these 3 metabolites, to characterize the predictive value of these metabolites using MetaboAnalyst 5.0.
As previously mentioned, the x-axis indicates the false positive rate and the y-axis indicates the true positive rate, generating the area under the ROC curve (AUC) [128].

Figure 47 shows the ROC curves for these 3 identified metabolites, and the AUC values vary between 0.685 and 0.768. Thus, the 3,4-dihydroxy-L-phenylalanine metabolite was the one that presented the lowest AUC value (0.685) and 5'-methylthioadenosine presented the highest AUC value (0.768), being considered the metabolite with the best discriminatory power for distinguish between Aβ negative and Aβ positive groups. As its value is relatively close to 0.8, we can consider that this metabolite when used in diagnosis, 90% of the time a positive result will be correctly identified and 80% of the time a negative result will be correctly identified.
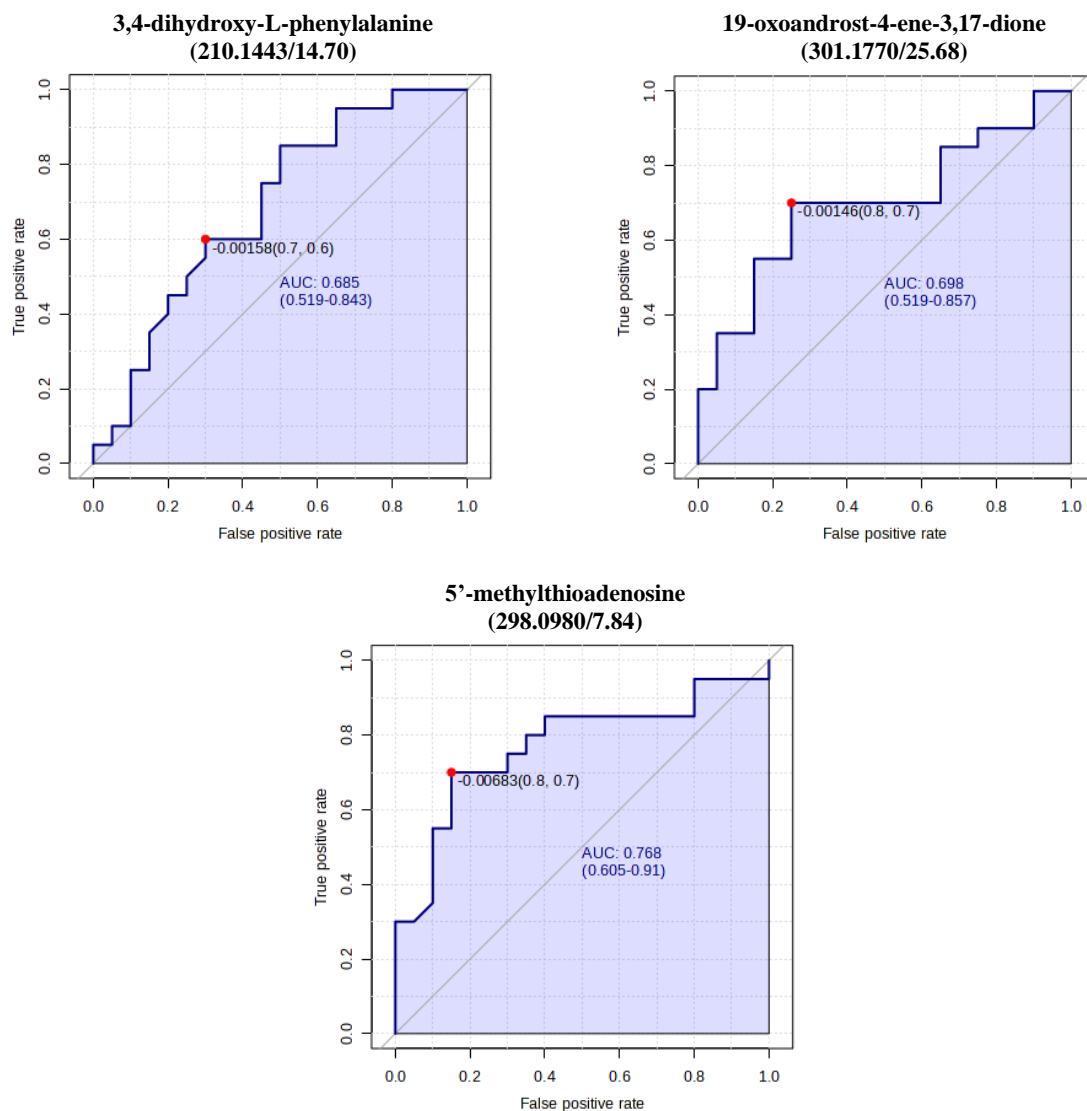
Figure 47 – ROC curve of features identified: receiver operator characteristic (ROC) curve analysis of the 3 features identified between Aβ negative and Aβ positive groups using MetaboAnalyst 5.0.

After that, a PCA analysis was performed again with the 62 statistically different features, considering four groups Aβ negative female, Aβ negative male, Aβ positive female and Aβ positive male, in order to verify if there is a separation of the two groups, considering the sex of the individuals.

The PCA graph shows a tendency to separate the Aβ negative samples from the Aβ positive samples and, in addition, a separation between the two sexes can be observed in the Aβ negative condition (Figure 48). On the other hand, in the Aβ positive group, there is an overlap between the two sexes.
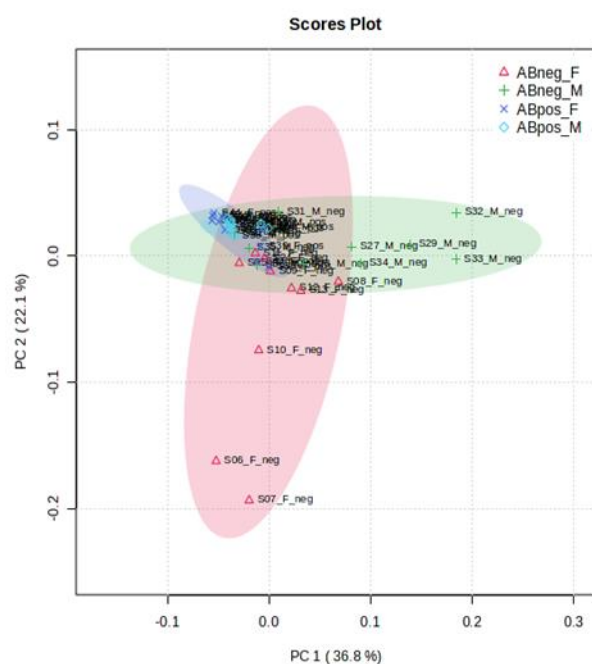
Figure 48 – PCA with 62 statistically different features: principal component analysis (PCA) scores for Aβ negative feminine (red), Aβ negative masculine, (green), Aβ positive feminine (purple) and Aβ positive masculine (blue) groups using the 62 features and the Pareto scaling performed using MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

Subsequently, a peptidomic analysis was conducted to identify the peptides present in the CSF of the individuals included in this study. Peptidomics enables the examination of endogenous fragments present in each sample, facilitating the identification of potential biomarkers in various diseases.

From the search in the ProteinPilot software, 37 different peptides were found, of which 33 were quantified. Of those quantified, 5 belonged to the neurosecretory protein VGF (VGF), 7 to proSAAS (PCS1N), 5 to fibrinogen alpha chain (FIBA), 3 to fibrinogen beta chain (FIBB), 2 to osteopontin (OSTP), 1 to neuroendocrine protein (7B2), 1 to apolipoprotein E (ApoE), 1 to brevican core protein (PGCB), 2 to cystatin-C (CYTC), 1 to BRD4-interacting chromatin-remodeling complex-associated protein (BRICA), 1 to kinesin-like protein KIF26B (KI26B), 1 to calsyntenin-1 (CSTN1), 2 to chromogranin-A (CMGA) and 1 to integral membrane protein 2B (ITM2B) (Table A.2.7 of Appendix A.2).

Then, a PCA analysis was performed with all peptides (Figure 49). It can be observed that the samples from the metabolomics analysis that appeared outside the confidence interval of the respective group do not show this trend in this peptidomics

analysis. In this analysis, the groups appear to be completely overlapping, with no sign of a tendency to separate through multivariate analysis.
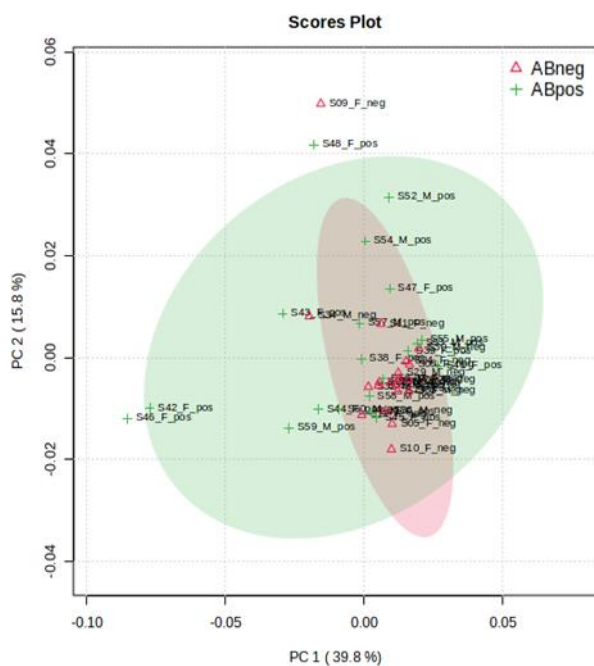
Figure 49 – PCA with all quantified peptides: principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (green) groups using all quantified peptides, normalized with total area sums, performed in MetaboAnalyst 5.0. Each point reflects one individual, and ellipses represent 95% CI of the data.

Since few peptides were quantified, further PCA analyzes with additional filtration were not performed. However, we proceeded to analyze the data in a univariate way, using the Mann-Whitney test. Statistically significant peptides were represented in boxplots, in Figure 50.

From the observation of the boxplots, it is possible to verify that almost all peptides are down regulated in the control group (patients who were in the MCI phase, Aβ negative group) compared to the group of patients who were in the disease phase (MCI due to AD, Aβ positive group). Only the FIBA peptide shows a contrary tendency. The PCS1N (AADHDVGSELPPEGVLGALLR), OSTP (RISHELDSASSEVN), PGCB (ALHPEEDPEGRQGRLLG) and CSTN1 (FVDLSGHNLANPHPFAVVPSTAT) peptides were those that showed the highest statistical significance, with a p-value below 0.01.

The PCS1N peptide was the one with the most identified peptides and the one with the most statistically different peptides. In this case, the peptides

DHDVGSELPPEGVLGA and AADHDVGSELPPEGVLGALLR are related to each other with only a difference of 5 amino acids, which may be related to some enzymatic activity that occurred in the body of the participants in the study. The proSAAS protein is encoded by the PCSK1N gene and is classified as a specialized secretory protein expressed mainly in neural tissues. This protein functions as an anti-amyloid chaperone in AD and is known to play a role in cerebral proteostasis, being found in neurofibrillary tangles and neuritic plaques in patients with Alzheimer's disease [134].

The protein osteopontin (OSTP) can be produced by immune cells found in the brain, contributing to inflammation and being associated with neurological diseases. In the literature, there is a study carried out in the brain of mice modified to develop Alzheimer's disease, where the production of osteopontin by microglia in the brain was tracked. This study found that only a small subset of microglia in the brain produced osteopontin and all belonged to the CD11c+ microglia, with only a few producing the potentially toxic protein. It was further found that the amount of osteopontin in the mice's brains had doubled or tripled as AD progressed. This study also compared the brain tissues of patients with normal cognitive function and patients with Alzheimer's disease, showing that patients with Alzheimer's had an average of three times more osteopontin than healthy individuals. This tendency is verified in the boxplot associated with the OSTP protein (Figure 50), where a higher concentration of this protein is verified in the Aβ positive group [135].

According to a study in the literature [136], the levels of calsyntenin-1 protein (CSTN1) are reduced in the CSF of patients who are in an early stage of Alzheimer's disease. These reduced levels of the protein may reflect the reduced synaptic density in these individuals, as they already show signs of cerebral amyloidosis. This effect is often confused with generalized neurodegeneration in more advanced stages of the disease.

The protein brevican core protein (PGCB) also showed statistical difference, making it an interesting protein to study, however, there is not much information in the literature that correlates it with Alzheimer's disease.
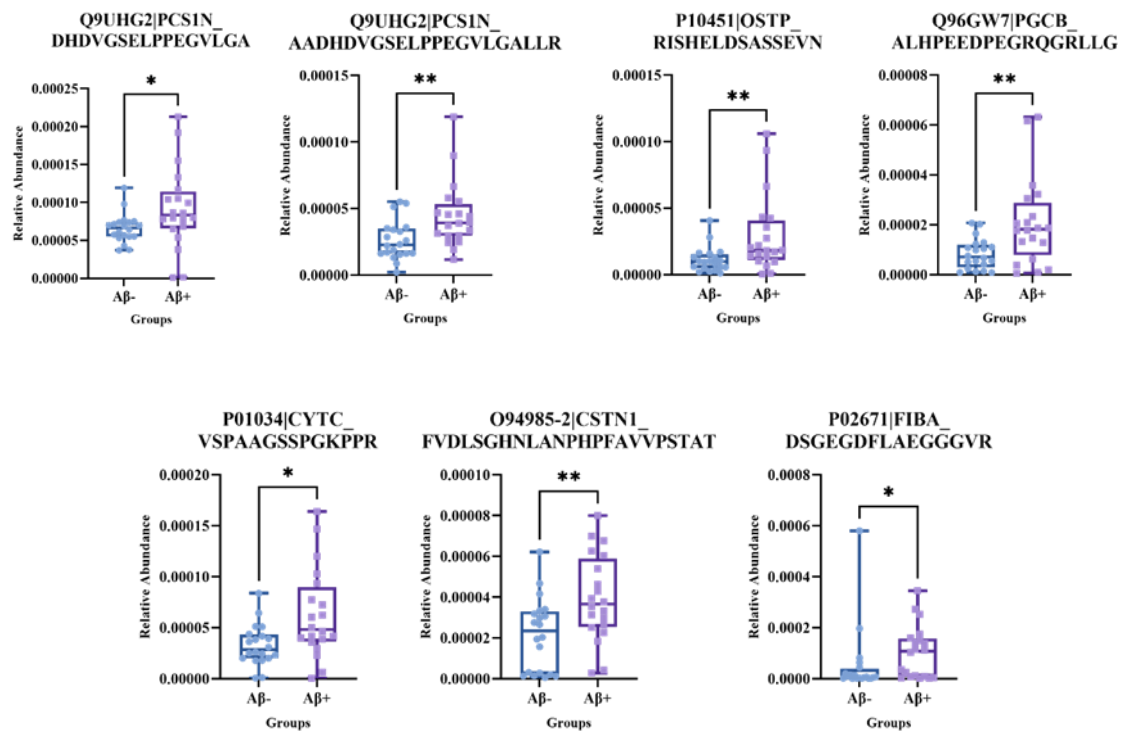
Figure 50 – Boxplot of all the peptides statistically different: boxplot of all the peptides statistically different in Aβ negative (blue) and Aβ positive (purple) group. The plots show the relative abundance of each sample, the median, the minimum and the maximum of each group. The peptides are from proteins: proSAAS (PCS1N); osteopontin (OSTP); brevican core protein (PGCB); cystatin-C (CYTC); calsyntenin-1 (CSTN1) and fibrinogen alpha chain (FIBA). The level of significant is represented by *: * P≤ 0.05; ** P≤ 0.01.

Using this data, we performed a univariate ROC analysis. However, none of the AUCs were above 0.80 and therefore none of these peptides could be used to distinguish between Aβ negative and Aβ positive groups (Supplementary Figure 1 of Appendix A.2).

7 – Conclusion

The main objective of this work was to carry out a non-targeted metabolomics analysis using LC-MS, in CSF samples from β-amyloid positive and β-amyloid negative patients, so that it would be possible to differentiate these two groups by identifying significantly altered metabolites.

For this purpose, 40 samples were used (20 β-amyloid positive samples and 20 β-amyloid negative samples), where the ages varied between 46 and 80 years without statistically significant differences between the groups.

Through LC-MS, 7956 features were detected in the SciexOS software and 2811 features in the MS-DIAL software, and then we selected the features that were present in all samples, 1863 features (in the case of SciexOS) and 2461 features (in the case of MS-DIAL). In this way, the univariate and multivariate analysis allowed highlighting 144 (SciexOS) and 137 (MS-DIAL) features, to better separate the two groups. In the analysis of the features with more statistical significance, 8 samples from the Aβ negative group did not show the same correlation pattern as the other samples from the same group, showing that they were more similar to the samples from the Aβ positive group. These 8 samples could be in a state of progression from Aβ negative to Aβ positive condition at the time of diagnosis and showed similarities with the samples from the Aβ positive group, considering the detected metabolites. For this reason, it would be important to confirm with clinicians whether these patients have progressed from an MCI condition to an MCI due to AD.

In the hierarchical grouping analysis, it was also observed, in the Aβ negative condition, the formation of a group for female samples and another group for male samples, leading to the belief that there were metabolites that differed between sexes in this group.

After this analysis, we found that some features only allowed us to differentiate the sexes present in the Aβ negative condition, as they had a high VIP value overcoming the other features. With this, the features were analyzed, excluding those that were correlated with the feature 226.1804/20.28, thus leaving 62 features, in order to be able to verify the differences between the two groups. In this case, it was verified that, contrary to the previous analysis, all samples were grouped in the respective group, allowing us to

conclude that the two groups have some different metabolic characteristics allowing this separation.

Regarding the identification of metabolites, 34 features were identified. However, feature 298.0980/7.84 was the only one that obtained the same match in all databases (HMDB, METLIN, MS-DIAL and SciexOS), being identified as 5'-methyltheoadenosine, increasing the confidence in this identification.

Of these 34 identified features, there were 7 features that presented different matches. This may be related to the databases lacking information on MS/MS spectra or low-quality spectral results. To obtain a validated identification for these 34 features, it would be necessary to have a resuspended reference standard in the CSF and analyze it under conditions identical to our experimental analysis.

After identifying the features, we carried out pathway analysis, to verify which would be the pathways with the greatest impact in this study.

It was found that the most significantly altered pathway was that of tyrosine metabolism, where the metabolite 3,4-dihydroxy-L-phenylalanine (210.1443/14.70) was found to be enriched, being increased in the CSF of the Aβ negative group.

The metabolite 19-oxoandrost-4-ene-3,17-dione (301.1770/25.68) was found to be enriched in the steroid hormone biosynthesis pathway and was also increased in the CSF of the Aβ negative group.

The 5'-methyltheoadenosine (298.0980/7.84) metabolite was found to be enriched in cysteine and methionine metabolism, which was the least impacting pathway, and unlike the other metabolites, it was increased in the Aβ positive group. In the literature, there are already studies showing that the metabolites 19-oxoandrost-4-ene-3,17-dione and 5'-methyltheoadenosine have an impact on Alzheimer's disease.

The analysis of the ROC curves performed on these 3 identified metabolites, showed that the feature 298.0980/7.84, identified as 5'-methylthioadenosine, was the one that presented the best discriminatory power with an AUC value of 0.768. However, it would be important to verify and confirm the identification of this feature, considering a larger number of samples better to assess the discriminatory power between the two groups. It would also be important to compare this feature with established AD biomarkers, to assess whether this metabolite adds value to existing diagnostic tools.

When the analysis was carried out considering the four groups (Aβ negative female, Aβ negative male, Aβ positive female and Aβ positive male) it was possible to verify that there are differences between sexes in the Aβ negative group; however the same does not occur in the Aβ positive condition. The differences observed between sexes in the Aβ negative group may be related to other factors, such as other pharmacological therapies, since some hypertensive drugs have been identified. Since, in an untargeted metabolomics analysis, all types of metabolites are detected, it is necessary to take into account all the information about the clinical history of the patients, to be easier to exclude the identified metabolites that are unrelated to the Alzheimer's disease.

Regarding the peptidomics analysis, it was found that the samples did not show the same behavior compared to the metabolomics analysis, and in the peptidomics analysis both groups completely overlapped.
However, 7 peptides were quantified and almost all peptides were increased in the CSF of patients in the Aβ negative condition, with the exception of the FIBA peptide, which was increased in the Aβ positive group, possibly being involved in the pathophysiology of AD. However, when analyzing the ROC curves, it was found that no peptide had an AUC value greater than 0.8, concluding that none of them would be a potential biomarker for AD.

In conclusion, this untargeted metabolomics analysis on CSF samples from MCI patients and MCI due to AD patients allowed the detection of 62 statistically significant features and the identification of 34. The identification of the 5'-methyltheoadenosine metabolite that is increased in the Aβ negative group, as already found in the literature, suggests that this metabolite and its respective pathway may be associated with the pathophysiology of AD and can be used as a potential biomarker for early identification of Alzheimer's disease.

Bibliography

1. Breijyeh, Z. and R. Karaman, *Comprehensive Review on Alzheimer's Disease: Causes and Treatment.* Molecules, 2020. **25**(24).

2. Matthews, K.A., et al., *Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015-2060) in adults aged >/=65 years.* Alzheimer's & Dementia, 2019. **15**(1): p. 17-24.

3. Silva, M.V.F., et al., *Alzheimer's disease: risk factors and potentially protective measures.* Journal of Biomedical Science, 2019. **26**(1): p. 33.

4. Alzheimer's Disease International. *Dementia statistics.* Accessed october 10, 2022 [https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/].

5. Campos Fernandes, A., *[A Perspective on the OECD Report "Health at a Glance 2019"].* Acta Médica Portuguesa, 2020. **33**(1): p. 4-6.

6. Georges, J., O. Miller, and C. Bintener, *Estimating the prevalence of dementia in Europe*. 2020.

7. Qiu, C., M. Kivipelto, and E. von Strauss, *Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention.* Dialogues in Clinical Neuroscience, 2009. **11**(2): p. 111-28.

8. Herrup, K., *Reimagining Alzheimer's disease--an age-based hypothesis.* Journal of Neuroscience, 2010. **30**(50): p. 16755-62.

9. Sheppard, O. and M. Coleman, *Alzheimer's Disease: Etiology, Neuropathology and Pathogenesis*, in *Alzheimer's Disease: Drug Discovery*, X. Huang, Editor. 2020, Exon Publications

Copyright: The Authors.: Brisbane (AU).

10. Ferri, C.P., et al., *Global prevalence of dementia: a Delphi consensus study.* The Lancet, 2005. **366**(9503): p. 2112-7.

11. Sakamoto, S., et al., *Differences in cerebral metabolic impairment between early and late onset types of Alzheimer's disease.* Journal of the Neurological Sciences, 2002. **200**(1-2): p. 27-32.

12. Piaceri, I., B. Nacmias, and S. Sorbi, *Genetics of familial and sporadic Alzheimer's disease.* Frontiers in Bioscience (Elite Ed), 2013. **5**(1): p. 167-77.

13. Tcw, J. and A.M. Goate, *Genetics of β-Amyloid Precursor Protein in Alzheimer's Disease.* Cold Spring Harbor Perspectives in Medicine, 2017. **7**(6).

14. DeTure, M.A. and D.W. Dickson, *The neuropathological diagnosis of Alzheimer's disease.* Molecular Neurodegeneration, 2019. **14**(1): p. 32.

15. *2022 Alzheimer's disease facts and figures.* Alzheimer's & Dementia, 2022. **18**(4): p. 700-789.

16. Tcw, J. and A.M. Goate, *Genetics of beta-Amyloid Precursor Protein in Alzheimer's Disease.* Cold Spring Harbor Perspectives in Medicine, 2017. **7**(6).

17. Tiwari, S., et al., *Alzheimer's disease: pathogenesis, diagnostics, and therapeutics.* International Journal of Nanomedicine, 2019. **14**: p. 5541-5554.

18. Nhan, H.S., K. Chiang, and E.H. Koo, *The multifaceted nature of amyloid precursor protein and its proteolytic fragments: friends and foes.* Acta Neuropathologica, 2015. **129**(1): p. 1-19.

19. Sumner, I.L., et al., *Antibody Engineering for Optimized Immunotherapy in Alzheimer's Disease.* Frontiers in Neuroscience, 2018. **12**: p. 254.

20. National Institute on Aging (2017). *What Happens to the Brain in Alzheimer's Disease?* Accessed october 12, 2022 [https://www.nia.nih.gov/health/what-happens-brain-alzheimers-disease].

21. Medeiros, R., D. Baglietto-Vargas, and F.M. LaFerla, *The role of tau in Alzheimer's disease and related disorders.* CNS Neuroscience & Therapeutics, 2011. **17**(5): p. 514-24.

22. Li, H.L., et al., *Phosphorylation of tau antagonizes apoptosis by stabilizing beta-catenin, a mechanism involved in Alzheimer's neurodegeneration.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(9): p. 3591-6.

23. Stokin, G.B., et al., *Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease.* Science, 2005. **307**(5713): p. 1282-8.

24. Iov, T., et al., *Macroscopic and microscopic diagnosis in Alzheimer's disease.* Romanian Journal of Legal Medicine, 2019. **27**(2): p. 89-94.

25. Jie, C., et al., *Tauvid™: The First FDA-Approved PET Tracer for Imaging Tau Pathology in Alzheimer's Disease.* Pharmaceuticals (Basel), 2021. **14**(2).

26. Geldmacher, D.S. and P.J. Whitehouse, Jr., *Differential diagnosis of Alzheimer's disease.* Neurology, 1997. **48**(5 Suppl 6): p. S2-9.

27.   Jack, C.R., Jr., et al., *NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease.* Alzheimer's & Dementia, 2018. **14**(4): p. 535-562.

28.   Jack, C.R., Jr., et al., *Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.* Alzheimer's & Dementia, 2011. **7**(3): p. 257-62.

29.   Pinto, T.C.C., et al., *Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly?* International Psychogeriatrics, 2019. **31**(4): p. 491-504.

30.   Portet, F., et al., *Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease.* Journal of Neurology, Neurosurgery, and Psychiatry, 2006. **77**(6): p. 714-8.

31.   Carnero-Pardo, C., *Should the mini-mental state examination be retired?* Neurologia, 2014. **29**(8): p. 473-81.

32.   Petersen, R.C., *Clinical practice. Mild cognitive impairment.* The New England Journal of Medicine, 2011. **364**(23): p. 2227-34.

33.   Olazaran, J., et al., *Practical application of brief cognitive tests.* Neurologia, 2016. **31**(3): p. 183-94.

34.   Nasreddine, Z.S., et al., *The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.* Journal of the American Geriatrics Society, 2005. **53**(4): p. 695-9.

35.   Freitas, S., et al., *Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease.* Alzheimer Disease and Associated Disorders, 2013. **27**(1): p. 37-43.

36.   Gil, L., et al., *Validation of the Montreal Cognitive Assessment (MoCA) in Spanish as a screening tool for mild cognitive impairment and mild dementia in patients over 65 years old in Bogota, Colombia.* International Journal of Geriatric Psychiatry, 2015. **30**(6): p. 655-62.

37.   Luis, C.A., A.P. Keegan, and M. Mullan, *Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the*

Southeastern US. International Journal of Geriatric Psychiatry, 2009. **24**(2): p. 197-201.

38. Zhou, S., et al., *The influence of education on Chinese version of Montreal cognitive assessment in detecting amnesic mild cognitive impairment among older people in a Beijing rural community.* The Scientific World Journal, 2014. **2014**: p. 689456.

39. Kasai, M., et al., *Screening for very mild subcortical vascular dementia patients aged 75 and above using the montreal cognitive assessment and mini-mental state examination in a community: the kurihara project.* Dementia and Geriatric Cognitive Disorders Extra, 2012. **2**(1): p. 503-15.

40. Jia, X., et al., *A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: a cross-sectional study.* BMC Psychiatry, 2021. **21**(1): p. 485.

41. World Health, O., *The ICD-10 classification of mental and behavioural disorders : diagnostic criteria for research.* 1993, World Health Organization: Geneva.

42. Caselli, R.J., et al., *Alzheimer Disease: Scientific Breakthroughs and Translational Challenges.* Mayo Clinic Proceedings, 2017. **92**(6): p. 978-994.

43. Kim, J., et al., *Neuroimaging Modalities in Alzheimer’s Disease: Diagnosis and Clinical Features.* International Journal of Molecular Sciences, 2022. **23**(11): p. 6079.

44. Grover, V.P., et al., *Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians.* Journal of Clinical and Experimental Hepatology, 2015. **5**(3): p. 246-55.

45. Silbert, L.C., et al., *Changes in premorbid brain volume predict Alzheimer's disease pathology.* Neurology, 2003. **61**(4): p. 487-92.

46. Ledig, C., et al., *Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database.* Scientific Reports, 2018. **8**(1): p. 11258.

47. Kapoor, V., B.M. McCook, and F.S. Torok, *An introduction to PET-CT imaging.* Radiographics, 2004. **24**(2): p. 523-43.

48. Chételat, G., et al., *Amyloid-PET and (18)F-FDG-PET in the diagnostic investigation of Alzheimer's disease and other dementias.* Lancet Neurology, 2020. **19**(11): p. 951-962.

49. Dunn, J.F., et al., *Bold MRI vs. NIR Spectrophotometry*, in *Oxygen Transport to Tissue XX*, A.G. Hudetz and D.F. Bruley, Editors. 1998, Springer US: Boston, MA. p. 103-113.

50. Johnson, K.A., et al., *Brain imaging in Alzheimer disease.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(4): p. a006213.

51. Scarapicchia, V., et al., *Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies.* Front Frontiers in Human Neuroscience, 2017. **11**: p. 419.

52. Mohammadi-Nejad, A.R., et al., *Neonatal brain resting-state functional connectivity imaging modalities.* Photoacoustics, 2018. **10**: p. 1-19.

53. Tang, T.B. and Y.L. Chan, *Functional Connectivity Analysis on Mild Alzheimer's Disease, Mild Cognitive Impairment and Normal Aging using fNIRS.* Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2018. **2018**: p. 17-20.

54. Skrandies, W., *EEG/EP: new techniques.* Brain Topography, 1993. **5**(4): p. 347-50.

55. Cassani, R., et al., *Systematic Review on Resting-State EEG for Alzheimer's Disease Diagnosis and Progression Assessment.* Disease Markers, 2018. **2018**: p. 5174815.

56. Wallstrom, G.L., et al., *Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods.* International Journal of Psychophysiology, 2004. **53**(2): p. 105-119.

57. Tait, L., et al., *EEG microstate complexity for aiding early diagnosis of Alzheimer's disease.* Scientific Reports, 2020. **10**(1): p. 17627.

58. Shaw, L.M., et al., *Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects.* Annals of Neurology, 2009. **65**(4): p. 403-13.

59. van Rossum, I.A., et al., *Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design.* Journal of Alzheimer's Disease, 2010. **20**(3): p. 881-91.

60. Michaud, T.L., et al., *Risk Stratification Using Cerebrospinal Fluid Biomarkers in Patients with Mild Cognitive Impairment: An Exploratory Analysis.* Journal of Alzheimer's Disease, 2015. **47**(3): p. 729-40.

61. Blennow, K., et al., *Cerebrospinal fluid and plasma biomarkers in Alzheimer disease.* Nature Reviews Neurology, 2010. **6**(3): p. 131-44.

62. Hansson, O., et al., *Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study.* Lancet Neurology, 2006. **5**(3): p. 228-34.

63. Parnetti, L., et al., *Performance of aβ1-40, aβ1-42, total tau, and phosphorylated tau as predictors of dementia in a cohort of patients with mild cognitive impairment.* Journal of Alzheimer's Disease, 2012. **29**(1): p. 229-38.

64. Mattsson, N., et al., *CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment.* The Journal of the American Medical Association, 2009. **302**(4): p. 385-93.

65. Moghekar, A., et al., *CSF biomarker changes precede symptom onset of mild cognitive impairment.* Neurology, 2013. **81**(20): p. 1753-8.

66. Barichello, T., V.V. Giridharan, and F. Dal-Pizzol, *A cerebrospinal fluid biosignature for the diagnosis of Alzheimer's disease.* Brazilian Journal of Psychiatry, 2019. **41**(6): p. 467-468.

67. Teunissen, C.E. and E. Willemse, *Cerebrospinal Fluid Biomarkers for Alzheimer's Disease: Emergence of the Solution to an Important Unmet Need.* The electronic Journal of the IFCC, 2013. **24**(3): p. 97-104.

68. Li, L., et al., *BACE1 in the retina: a sensitive biomarker for monitoring early pathological changes in Alzheimer's disease.* Neural Regeneration Research, 2016. **11**(3): p. 447-53.

69. Hall, S., et al., *Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders.* Archives of Neurology, 2012. **69**(11): p. 1445-52.

70. Julia Dixon MD (2023). *Lumbar puncture.* Accessed october 20, 2022 [https://www.saem.org/about-saem/academies-interest-groups-affiliates2/cdem/for-students/online-education/m3-curriculum/group-emergency-department-procedures/lumbar-puncture].

71.   Michaud, T.L., et al., *Using Cerebrospinal Fluid Biomarker Testing to Target Treatment to Patients with Mild Cognitive Impairment: A Cost-Effectiveness Analysis.* PharmacoEconomics - Open, 2018. **2**(3): p. 309-323.

72.   Gauthier, S., et al., *Mild cognitive impairment.* The Lancet, 2006. **367**(9518): p. 1262-70.

73.   Albert, M.S., et al., *The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.* Alzheimer's & Dementia, 2011. **7**(3): p. 270-9.

74.   Loi, S.M., et al., *Alzheimer disease: Non-pharmacological and pharmacological management of cognition and neuropsychiatric symptoms.* Australas Psychiatry, 2018. **26**(4): p. 358-365.

75.   Berg-Weger, M. and D.B. Stewart, *Non-Pharmacologic Interventions for Persons with Dementia.* Missouri Medicine, 2017. **114**(2): p. 116-119.

76.   Tan, E., et al., *Current approaches to the pharmacological treatment of Alzheimer's disease.* Australian Journal for General Practitioners, 2018. **47**: p. 586-592.

77.   Sharma, K., *Cholinesterase inhibitors as Alzheimer's therapeutics (Review).* Molecular Medicine Reports, 2019. **20**(2): p. 1479-1487.

78.   Epperly, T., M.A. Dunay, and J.L. Boice, *Alzheimer Disease: Pharmacologic and Nonpharmacologic Therapies for Cognitive and Functional Symptoms.* American Family Physician, 2017. **95**(12): p. 771-778.

79.   Kishi, T., et al., *Memantine for Alzheimer's Disease: An Updated Systematic Review and Meta-analysis.* Journal of Alzheimer's Disease, 2017. **60**(2): p. 401-425.

80.   Tariot, P.N., et al., *Memantine Treatment in Patients With Moderate to Severe Alzheimer Disease Already Receiving DonepezilA Randomized Controlled Trial.* The Journal of the American Medical Association, 2004. **291**(3): p. 317-324.

81.   Hartmann, S. and H.J. Möbius, *Tolerability of memantine in combination with cholinesterase inhibitors in dementia therapy.* International Clinical Psychopharmacology, 2003. **18**(2): p. 81-5.

82.   Salloway, S.P., et al., *Advancing combination therapy for Alzheimer's disease.* Alzheimer's & Dementia (NY), 2020. **6**(1): p. e12073.

83. Nettiksimmons, J., et al., *Biological heterogeneity in ADNI amnestic mild cognitive impairment.* Alzheimer's & Dementia, 2014. **10**(5): p. 511-521.e1.

84. Ryan, D., et al., *Metabolomics as a tool for diagnosis and monitoring in coeliac disease.* Metabolomics, 2014. **11**(4): p. 980-990.

85. Dunn, W.B., et al., *Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy.* Chemical Society Reviews, 2011. **40**(1): p. 387-426.

86. Wishart, D.S., et al., *HMDB 4.0: the human metabolome database for 2018.* Nucleic Acids Research, 2018. **46**(D1): p. D608-d617.

87. Sethi, S. and E. Brietzke, *Omics-Based Biomarkers: Application of Metabolomics in Neuropsychiatric Disorders.* International Journal of Neuropsychopharmacology, 2015. **19**(3): p. pyv096.

88. Jendoubi, T., *Approaches to Integrating Metabolomics and Multi-Omics Data: A Primer.* Metabolites, 2021. **11**(3).

89. Jurowski, K., et al., *Analytical Techniques in Lipidomics: State of the Art.* 2017.

90. Roessner, U. and J. Bowne, *What is metabolomics all about?* Biotechniques, 2009. **46**(5): p. 363-5.

91. Bino, R.J., et al., *Potential of metabolomics as a functional genomics tool.* Trends in Plant Science, 2004. **9**(9): p. 418-25.

92. Bingol, K., *Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods.* High-Throughput, 2018. **7**(2): p. 9.

93. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: the apogee of the omics trilogy.* Nature Reviews Molecular Cell Biology, 2012. **13**(4): p. 263-9.

94. Wilkins, J.M. and E. Trushina, *Application of Metabolomics in Alzheimer's Disease.* Frontiers in Neurology, 2017. **8**: p. 719.

95. Trushina, E. and M.M. Mielke, *Recent advances in the application of metabolomics to Alzheimer's Disease.* Biochimica et Biophysica Acta, 2014. **1842**(8): p. 1232-9.

96. Trushina, E., et al., *Identification of Altered Metabolic Pathways in Plasma and CSF in Mild Cognitive Impairment and Alzheimer's Disease Using Metabolomics.* PLOS ONE, 2013. **8**(5): p. e63644.

97.   Zhao, Y., et al., *Blood levels of circulating methionine components in Alzheimer's disease and mild cognitive impairment: A systematic review and meta-analysis.* Frontiers in Aging Neuroscience, 2022. **14**: p. 934070.

98.   Dayon, L., et al., *One-carbon metabolism, cognitive impairment and CSF measures of Alzheimer pathology: homocysteine and beyond.* Alzheimer's Research & Therapy, 2017. **9**(1): p. 43.

99.   Ibanez, C., et al., *Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry metabolomics.* Analytical Chemistry, 2012. **84**(20): p. 8532-40.

100.   Nagata, Y., et al., *Comparative analysis of cerebrospinal fluid metabolites in Alzheimer's disease and idiopathic normal pressure hydrocephalus in a Japanese cohort.* Biomarker Research, 2018. **6**: p. 5.

101.   Fekkes, D., et al., *Abnormal amino acid metabolism in patients with early stage Alzheimer dementia.* Journal of Neural Transmission (Vienna), 1998. **105**(2-3): p. 287-94.

102.   Ting, K.K., B.J. Brew, and G.J. Guillemin, *Effect of quinolinic acid on human astrocytes morphology and functions: implications in Alzheimer's disease.* Journal of Neuroinflammation, 2009. **6**(1): p. 36.

103.   Sang, C., et al., *Coenzyme A-Dependent Tricarboxylic Acid Cycle Enzymes Are Decreased in Alzheimer's Disease Consistent With Cerebral Pantothenate Deficiency.* Frontiers in Aging Neuroscience, 2022. **14**: p. 893159.

104.   Pettegrew, J.W., et al., *Clinical and neurochemical effects of acetyl-L-carnitine in Alzheimer's disease.* Neurobiology of Aging, 1995. **16**(1): p. 1-4.

105.   Koal, T., et al., *Sphingomyelin SM(d18:1/18:0) is significantly enhanced in cerebrospinal fluid samples dichotomized by pathological amyloid-β42, tau, and phospho-tau-181 levels.* Journal of Alzheimer's Disease, 2015. **44**(4): p. 1193-201.

106.   Czech, C., et al., *Metabolite profiling of Alzheimer's disease cerebrospinal fluid.* PLOS ONE, 2012. **7**(2): p. e31501.

107.   Ibáñez, C., et al., *A new metabolomic workflow for early detection of Alzheimer's disease.* Journal of Chromatography A, 2013. **1302**: p. 65-71.

108.   Zhang, X., N. Alshakhshir, and L. Zhao, *Glycolytic Metabolism, Brain Resilience, and Alzheimer's Disease.* Frontiers in Neuroscience, 2021. **15**: p. 662242.

94

109. An, Y., et al., *Evidence for brain glucose dysregulation in Alzheimer's disease.* Alzheimer's & Dementia, 2018. **14**(3): p. 318-329.

110. Bergau, N., et al., *Reduction of Glycolysis Intermediate Concentrations in the Cerebrospinal Fluid of Alzheimer's Disease Patients.* Frontiers in Neuroscience, 2019. **13**: p. 871.

111. Kim, J.M., et al., *Criminal manifestations of dementia patients: report from the national forensic hospital.* Dementia and Geriatric Cognitive Disorders Extra, 2011. **1**(1): p. 433-8.

112. Ekström, A., M. Kristiansson, and K.S. Björkstén, *Dementia and cognitive disorder identified at a forensic psychiatric examination - a study from Sweden.* BMC Geriatrics, 2017. **17**(1): p. 219.

113. David, A. and P. Rostkowski, *Chapter 2 - Analytical techniques in metabolomics*, in *Environmental Metabolomics*, D. Álvarez-Muñoz and M. Farré, Editors. 2020, Elsevier. p. 35-64.

114. Segers, K., et al., *Analytical techniques for metabolomic studies: a review.* Bioanalysis, 2019. **11**(24): p. 2297-2318.

115. Xiao, J.F., B. Zhou, and H.W. Ressom, *Metabolite identification and quantitation in LC-MS/MS-based metabolomics.* Trends in Analytical Chemistry, 2012. **32**: p. 1-14.

116. Becker, S., et al., *LC–MS-based metabolomics in the clinical laboratory.* Journal of Chromatography B, 2012. **883-884**: p. 68-75.

117. Wilson, I.D., et al., *HPLC-MS-based methods for the study of metabonomics.* Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 2005. **817**(1): p. 67-76.

118. Duncan, K.D., J. Fyrestam, and I. Lanekoff, *Advances in mass spectrometry based single-cell metabolomics.* Analyst, 2019. **144**(3): p. 782-793.

119. Nordstrom, A., et al., *Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics.* Analytical Chemistry, 2008. **80**(2): p. 421-9.

120. Banerjee, S. and S. Mazumdar, *Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte.* International Journal of Analytical Chemistry, 2012. **2012**: p. 282574.

121.   *Basic Concepts and the Control of Separation*, in *Introduction to Modern Liquid Chromatography*. 2009. p. 19-86.

122.   *Equipment*, in *Introduction to Modern Liquid Chromatography*. 2009. p. 87-145.

123.   Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry.* Journal of Mass Spectrometry, 2001. **36**(8): p. 849-65.

124.   Zhu, X., Y. Chen, and R. Subramanian, *Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry.* Analytical Chemistry, 2014. **86**(2): p. 1202-9.

125.   Santa, C., et al., *Neuroproteomics — LC-MS Quantitative Approaches*, in *Recent Advances in Proteomics Research*, M. Sameh, Editor. 2015, IntechOpen: Rijeka. p. Ch. 3.

126.   Zhou, B., et al., *LC-MS-based metabolomics.* Molecular Omics, 2012. **8**(2): p. 470-81.

127.   Mendes, V.M., M. Coelho, and B. Manadas, *Untargeted Metabolomics Relative Quantification by SWATH Mass Spectrometry Applied to Cerebrospinal Fluid.* Methods in Molecular Biology, 2019. **2044**: p. 321-336.

128.   Carter, J.V., et al., *ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves.* Surgery, 2016. **159**(6): p. 1638-1645.

129.   Nivsarkar, M. and A. Banerjee, *Establishing the probable mechanism of L-DOPA in Alzheimer's disease management.* Acta Poloniae Pharmaceutica, 2009. **66**(5): p. 483-6.

130.   Ambree, O., et al., *Levodopa ameliorates learning and memory deficits in a murine model of Alzheimer's disease.* Neurobiology of Aging, 2009. **30**(8): p. 1192-204.

131.   Xi, J., et al., *Disturbed microbial ecology in Alzheimer's disease: evidence from the gut microbiota and fecal metabolome.* BMC Microbiology, 2021. **21**(1): p. 226.

132.   Moreno, B., et al., *Differential neuroprotective effects of 5'-deoxy-5'-methylthioadenosine.* PLOS ONE, 2014. **9**(3): p. e90671.

133.   Avila, M.A., et al., *Methylthioadenosine.* The International Journal of Biochemistry & Cell Biology, 2004. **36**(11): p. 2125-30.

134.   Shakya, M., T. Yildirim, and I. Lindberg, *Increased expression and retention of the secretory chaperone proSAAS following cell stress.* Cell Stress Chaperones, 2020. **25**(6): p. 929-941.

135.   Qiu, Y., et al., *Definition of the contribution of an Osteopontin-producing CD11c(+) microglial subset to Alzheimer's disease.* Proceedings of the National Academy of Sciences of the United States of America, 2023. **120**(6): p. e2218915120.

136.   Lleo, A., et al., *Changes in Synaptic Proteins Precede Neurodegeneration Markers in Preclinical Alzheimer's Disease Cerebrospinal Fluid.* Molecular & Cellular Proteomics, 2019. **18**(3): p. 546-560.

APPENDIX

A.1 – Material and Methods

Table A.1.1 – Detailed information about the study participants.

Lab ID	Age	Sex	Group	Abbreviation
106008	50	F	Aβ -	S05_F_neg
110203	54	F	Aβ -	S06_F_neg
108382	55	F	Aβ -	S07_F_neg
104968	58	F	Aβ -	S08_F_neg
107018	59	F	Aβ -	S09_F_neg
109456	59	F	Aβ -	S10_F_neg
108259	60	F	Aβ -	S11_F_neg
109500	60	F	Aβ -	S12_F_neg
106389	60	F	Aβ -	S13_F_neg
106628	61	F	Aβ -	S24_F_neg
110073	46	M	Aβ -	S27_M_neg
107080	57	M	Aβ -	S28_M_neg
107461	57	M	Aβ -	S29_M_neg
103293	60	M	Aβ -	S30_M_neg
106583	62	M	Aβ -	S31_M_neg
104251	63	M	Aβ -	S32_M_neg
102434	64	M	Aβ -	S33_M_neg
102721	68	M	Aβ -	S34_M_neg
107282	68	M	Aβ -	S35_M_neg
102037	70	M	Aβ -	S36_M_neg
103698	49	F	Aβ +	S38_F_pos
107702	54	F	Aβ +	S39_F_pos
108726	55	F	Aβ +	S40_F_pos
106022	57	F	Aβ +	S41_F_pos
105728	60	F	Aβ +	S42_F_pos
100855	65	F	Aβ +	S43_F_pos
106746	68	F	Aβ +	S44_F_pos
103423	70	F	Aβ +	S45_F_pos
100112	71	F	Aβ +	S46_F_pos
100596	71	F	Aβ +	S47_F_pos
101960	71	F	Aβ +	S48_F_pos
104726	54	M	Aβ +	S52_M_pos
105438	55	M	Aβ +	S53_M_pos
105080	58	M	Aβ +	S54_M_pos
109901	62	M	Aβ +	S55_M_pos
109089	64	M	Aβ +	S56_M_pos
109522	64	M	Aβ +	S57_M_pos

103176	69	M	Aβ +	S58_M_pos
102357	78	M	Aβ +	S59_M_pos
107362	80	M	Aβ +	S60_M_pos

Table A.1.2 – Window distribution for SWATH method.

	Start Mass (Da)	Stop Mass (Da)
SWATH Exp 1	50	70
SWATH Exp 2	69	90
SWATH Exp 3	89	110
SWATH Exp 4	109	130
SWATH Exp 5	129	150
SWATH Exp 6	149	170
SWATH Exp 7	169	190
SWATH Exp 8	189	210
SWATH Exp 9	209	230
SWATH Exp 10	229	250
SWATH Exp 11	249	270
SWATH Exp 12	269	290
SWATH Exp 13	289	310
SWATH Exp 14	309	330
SWATH Exp 15	329	350
SWATH Exp 16	349	370
SWATH Exp 17	369	390
SWATH Exp 18	389	410
SWATH Exp 19	409	430
SWATH Exp 20	429	450
SWATH Exp 21	449	470
SWATH Exp 22	469	490
SWATH Exp 23	489	510
SWATH Exp 24	509	530
SWATH Exp 25	529	550
SWATH Exp 26	549	570
SWATH Exp 27	569	590
SWATH Exp 28	589	610
SWATH Exp 29	609	630
SWATH Exp 30	629	650
SWATH Exp 31	649	670
SWATH Exp 32	669	690
SWATH Exp 33	689	710
SWATH Exp 34	709	730

SWATH Exp 35	729	750
SWATH Exp 36	749	770
SWATH Exp 37	769	790
SWATH Exp 38	789	810
SWATH Exp 39	809	830
SWATH Exp 40	829	850
SWATH Exp 41	849	870
SWATH Exp 42	869	890
SWATH Exp 43	889	910
SWATH Exp 44	909	930
SWATH Exp 45	929	950
SWATH Exp 46	949	970
SWATH Exp 47	969	990
SWATH Exp 48	989	1010
SWATH Exp 49	1009	1030
SWATH Exp 50	1029	1050
SWATH Exp 51	1049	1070
SWATH Exp 52	1069	1090
SWATH Exp 53	1089	1110
SWATH Exp 54	1109	1130
SWATH Exp 55	1129	1150
SWATH Exp 56	1149	1170
SWATH Exp 57	1169	1190
SWATH Exp 58	1189	1210
SWATH Exp 59	1209	1230
SWATH Exp 60	1229	1250
SWATH Exp 61	1249	1270
SWATH Exp 62	1269	1290
SWATH Exp 63	1289	1310
SWATH Exp 64	1309	1330
SWATH Exp 65	1329	1350
SWATH Exp 66	1349	1370
SWATH Exp 67	1369	1390
SWATH Exp 68	1389	1410
SWATH Exp 69	1409	1430
SWATH Exp 70	1429	1450
SWATH Exp 71	1449	1470
SWATH Exp 72	1469	1490
SWATH Exp 73	1489	1500

A.2 – Results and discussion

Table A.2.1 – List of matches for each of the 62 statistically different features in the HMDB database based on mass/charge of the precursors only.

m/z	Retention Time (RT)	HMDB ID matches	Monoisotopic Mass	Adduct	Delta (ppm)
158.1527	22.33	HMDB0257395; HMDB0253703; HMDB0245344	157.1467	M+H	8
161.0213	5.68	No results for this query mass	−	−	−
161.0985	24.00	HMDB0246567	160.0900	M+H	8
		HMDB0032426; HMDB0031572	160.0922		6
		HMDB0038790; HMDB0031567; HMDB0035207	160.0888		15
197.0781	14.29	No results for this query mass	−	−	−
210.1443	14.70	HMDB0254752; HMDB0246283; HMDB0248857; HMDB0245501	209.1416	M+H	22
244.2234	23.24	No results for this query mass	−	−	−
247.0907	19.68	HMDB0247212	224.1085	M+Na	28
		HMDB0246998	224.1022		3
		HMDB0340720; HMDB0340719; HMDB0340718; HMDB0340717; HMDB0340716; HMDB0340715; HMDB0032552; HMDB0036203; HMDB0030703; HMDB0032663; HMDB0040629; HMDB0030680; HMDB0040890	224.1049		14
		HMDB0041008	224.1062		19
		HMDB0030379	224.0950		26
251.1603	24.94	HMDB0251058	250.1470	M+H	24
		HMDB0243865	250.1603		29
		HMDB0240642; HMDB0304066; HMDB0255460; HMDB0061839; HMDB0302702; HMDB0030917; HMDB0035148; HMDB0037529; HMDB0039635; HMDB0035760; HMDB0036036; HMDB0039644; HMDB0036563; HMDB0035798;	250.1569		15

		HMDB0035117; HMDB0035137; HMDB0035358; HMDB0302605; HMDB0039156; HMDB0040754; HMDB0036550; HMDB0031901; HMDB0037559; HMDB0036664; HMDB0034721; HMDB0036037; HMDB0037064; HMDB0301855; HMDB0015371			
253.1402	23.83	HMDB0340846; HMDB0340845; HMDB0340844; HMDB0340843; HMDB0340842; HMDB0340841; HMDB0340840; HMDB0340839; HMDB0059661; HMDB0036047; HMDB0033197	252.1362	M+H	13
264.1908	24.32	HMDB0260305; HMDB0259401; HMDB0244949; HMDB0013892; HMDB0015646; HMDB0029567; HMDB0014339	263.1885	M+H	19
266.1702	16.79	HMDB0251062	265.1597	M+H	12
		HMDB0061122; HMDB0015520	265.1678		18
		HMDB0015689; HMDB0014514	265.1579		19
266.1786	17.32	No results for this query mass	−	−	−
273.1660	16.79	No results for this query mass	−	−	−
273.1687	13.83	HMDB0241062; HMDB0241060; HMDB0241057; HMDB0241059; HMDB0241063; HMDB0241061; HMDB0241056; HMDB0340797; HMDB0340796; HMDB0340795; HMDB0340794; HMDB0340793	331.2358	M+H	3
277.1053	14.98	HMDB0254414; HMDB0244962	276.0899	M+H	29
		HMDB0028818; HMDB0250751; HMDB0042036; HMDB0011737	276.0958		8
		HMDB0304112; HMDB0304054	276.1042		22
		HMDB0030145; HMDB0030791; HMDB0035814; HMDB0030287; HMDB0041520	276.0998		6
		HMDB0249010; HMDB0253452	276.0939		15

		HMDB0247442	280.1576		17
281.1696	25.71	HMDB0340908; HMDB0340907; HMDB0340906; HMDB0340905; HMDB0340904; HMDB0242339; HMDB0030148; HMDB0041229; HMDB0036979; HMDB0040755	280.1675	M+H	18
284.33	25.99	No results for this query mass	−	−	−
295.1891	27.80	HMDB0251767; HMDB0040646; HMDB0035754; HMDB0035722; HMDB0034779; HMDB0005783	294.1831	M+H	4
		HMDB0033432; HMDB0030282	294.1732		29
298.098	7.84	HMDB0304350; HMDB0258835; HMDB0254957	297.0961	M+H	18
		HMDB0037297	297.0849		20
		HMDB0257557; HMDB0001173	297.0896		4
301.177	25.68	HMDB0260292	300.1620	M+H	26
		HMDB0258027	300.1725		9
		HMDB0254504	300.1660		12
		HMDB0249556	300.1626		23
		HMDB0248401; HMDB0256730; HMDB0304035; HMDB0060088; HMDB0038684; HMDB0039569; HMDB0011195; HMDB0006772; HMDB0006768; HMDB0000010; HMDB0015031	300.1725		9
313.2325	26.82	HMDB0242602; HMDB0340915; HMDB0340914; HMDB0062281; HMDB0304442; HMDB0062434; HMDB0062637; HMDB0040900; HMDB0010221; HMDB0010208; HMDB0010201; HMDB0006940; HMDB0004706; HMDB0003871	312.2301	M+H	15
		HMDB0302976	312.2312		19
317.1707	24.84	HMDB0256709	316.1688	M+H	17
		HMDB0252371	316.1587		15
		HMDB0247634; HMDB0303451; HMDB0029566	316.1675		13
		HMDB0249442; HMDB0014517	316.1569		21

		HMDB0256709	316.1688		14
		HMDB0252371	316.1587		17
317.1715	24.00	HMDB0247634; HMDB0303451; HMDB0029566	316.1675	M+H	10
		HMDB0249442; HMDB0014517	316.1569		23
335.2769	26.82	No results for this query mass	−	−	−
341.1913	24.42	HMDB0257041; HMDB0253577; HMDB0249355; HMDB0245907; HMDB0040799; HMDB0001091	340.1787	M+H	16
		HMDB0252882	340.1859		6
		HMDB0244346	340.1886		13
		HMDB0014876	340.1899		17
351.1541	24.35	HMDB0258286	350.1531	M+H	18
		HMDB0251232	350.1518		14
		HMDB0249307	350.1379		25
		HMDB0256819; HMDB0032003	350.1552		24
357.258	26.82	HMDB0341190	356.2424	M+H	23
		HMDB0259370; HMDB0041882	356.2464		12
		HMDB0244518; HMDB0245719; HMDB0302208; HMDB0004239; HMDB0002689; HMDB0002685	356.2563		16
378.1209	20.21	No results for this query mass	−	−	−
379.1469	25.57	HMDB0341149	378.1362	M+H	9
		HMDB0035037; HMDB0029304; HMDB0301749; HMDB0039137	378.1315		22
		HMDB0030656; HMDB0040350	378.1467		19
387.1765	25.15	HMDB0258919; HMDB0256720	386.1630	M+H	16
		HMDB0255928	386.1754		16
		HMDB0249294	386.1664		7
		HMDB0303928; HMDB0303467; HMDB0030810; HMDB0035404; HMDB0035405	386.1729		10
387.1775	24.46	HMDB0258919; HMDB0256720	386.1630	M+H	19
		HMDB0255928	386.1754		13
		HMDB0249294	386.1664		10
		HMDB0303928; HMDB0303467; HMDB0030810; HMDB0035404; HMDB0035405	386.1729		7

104

		HMDB0257135	390.1400		7
		HMDB0251865	390.1387		11
		HMDB0250075	390.1492		16
391.1501	26.09	HMDB0304725; HMDB0242527; HMDB0030564; HMDB0031422; HMDB0030565; HMDB0036294	390.1315	M+H	29
		HMDB0247827	390.1526		25
		HMDB0035876	390.1467		10
		HMDB0037260	390.1526		25
397.2529	24.42	HMDB0259404	396.2438	M+H	5
		HMDB0249570	396.2413		11
401.2856	26.09	No results for this query mass	−	−	−
410.1617	24.56	HMDB0257957; HMDB0015108	409.1489	M+H	13
		HMDB0256532	409.1557		3
		HMDB0259210	409.1665		29
		HMDB0061159	409.1656		27
		HMDB0060470; HMDB0060454; HMDB0303673	409.1584		10
		HMDB0059662; HMDB0011596	409.1597		13
419.2367	24.6	HMDB0256094	418.2369	M+H	18
		HMDB0253027	418.2329		8
420.2402	24.56	HMDB0254922; HMDB0248012	419.2209	M+H	29
		HMDB0247141	419.242		22
461.2836	26.93	HMDB0297030; HMDB0297029; HMDB0297028; HMDB0297027; HMDB0297022; HMDB0297021; HMDB0297020; HMDB0297019; HMDB0297018; HMDB0297017; HMDB0297016; HMDB0297015; HMDB0297014; HMDB0297013; HMDB0297012; HMDB0297011; HMDB0297010; HMDB0297009; HMDB0297008; HMDB0297007; HMDB0256357; HMDB0038352; HMDB0030909	460.2825	M+H	13
		HMDB0254731	460.2699		14
467.3562	26.75	HMDB0002972; HMDB0245416	466.3447	M+H	9
		HMDB0012082	466.3536		10

		HMDB0247191	476.3443		21
477.3616	23.03	HMDB0030271	476.3614	M+H	15
		HMDB0034079; HMDB0039443	476.3502		9
		HMDB0032879	460.3341		7
478.3647	22.68	HMDB0303309	460.3400	M+NH4	19
		HMDB0038500; HMDB0038501; HMDB0032690	460.3189		25
481.3194	18.47	HMDB0253340; HMDB0013267	229.1678	2M + Na	11
489.3351	26.89	HMDB0255626	489.3279	M	15
		HMDB0242438; HMDB0242406	489.3454		21
491.3411	26.89	HMDB0039090	490.3447	M+H	22
505.3092	26.89	HMDB0013021	504.2948	M+H	14
		HMDB0033709	504.3087		13
507.1326	24.80	HMDB0254498	506.1177	M+H	15
		HMDB0259909	506.1213		8
511.3784	26.75	No results for this query mass	−	−	−
512.3825	26.75	No results for this query mass	−	−	−
533.3639	26.86	No results for this query mass	−	−	−
535.3661	26.89	HMDB0297130; HMDB0297128; HMDB0297129: HMDB0297127	534.3556	M+H	6
		HMDB0035899	534.3709		23
536.2870	20.24	HMDB0259747; HMDB0240271	535.2795	M+H	0
541.2894	20.28	No results for this query mass	−	−	−
543.2859	20.24	No results for this query mass	−	−	−
549.3352	26.86	No results for this query mass	−	−	−
555.4047	26.72	HMDB0297254; HMDB0297253; HMDB0297252; HMDB0297251; HMDB0297250; HMDB0297249; HMDB0297248; HMDB0297247; HMDB0297126; HMDB0297125: HMDB0297124; HMDB0297123; HMDB0294496; HMDB0294495; HMDB0294494; HMDB0294493	554.3819	M+H	28
578.3958	26.82	No results for this query mass	−	−	−
599.4132	26.72	No results for this query mass	−	−	−
599.4284	26.72	HMDB0294484; HMDB0294483; HMDB0294482; HMDB0294481;	598.4081	M+H	22

		HMDB0294464; HMDB0294463; HMDB0294462; HMDB0294461			
		HMDB0260191; HMDB0039139; HMDB0039594	598.4233		4
		HMDB0030100	598.4386		29
643.4563	26.69	No results for this query mass	−	−	−
701.3865	14.25	HMDB0267486; HMDB0267485; HMDB0266866; HMDB0266865; HMDB0262921; HMDB0262920	700.3951	M+H	23
795. 3352	25.22	No results for this query mass	−	−	−
805.404	24.46	HMDB0274840; HMDB0274839	804.419	M+H	28
		HMDB0034945; HMDB0034949	804.378		23

Table A.2.2 – List of matches for each of the 62 statistically different features in the METLIN database based on mass/charge of the precursors only.

m/z	Retetion Time (RT)	METLIN ID matches	Monoisotopic Mass	Adduct	Delta (ppm)
158.1527	22.33	No results for this query mass	−	−	−
161.0213	5.68	346616	160.0095	[M+H]+	27
161.0985	24.00	No results for this query mass	−	−	−
197.0781	14.29	488670	196.0653	[M+H]+	27
210.1443	14.70	464755	209.1328	[M+H]+	19
244.2234	23.24	267513	243.2198	[M+H]+	−
247.0907	19.68	No results for this query mass	−	−	−
251.1603	24.94	485265	250.1463	[M+H]+	26
253.1402	23.83	273346	252.1263	[M+H]+	26
264.1908	24.32	No results for this query mass	−	−	−
266.1702	16.79	445689	265.1551	[M+H]+	29
266.1786	17.32	964257	265.1642	[M+H]+	26
273.166	16.79	273847	272.1525	[M+H]+	22
273.1687	13.83	267750	272.1541	[M+H]+	26
277.1053	14.98	No results for this query mass	−	−	−
281.1696	25.71	No results for this query mass	−	−	−
284.33	25.99	43833	283.3239	[M+H]+	4
295.1891	24.80	No results for this query mass	−	−	−
298.098	7.84	3425	297.0896	[M+H]+	3

301.177	**25.68**	**41865**	**300.1725**	**[M+H]+**	**9**
313.2325	26.82	67501	312.2202	[M+H]+	16
317.1707	24.84	527031	316.1549	[M+H]+	26
317.1715	24.00	2856	316.1569	[M+H]+	23
335.2769	26.82	No results for this query mass	−	−	−
341.1913	24.42	2234	340.1787	[M+H]+	15
351.1541	24.35	72303	350.1413	[M+H]+	15
357.258	26.82	−	−	−	−
378.1209	20.21	No results for this query mass	356.2501	[M+H]+	−
379.1469	25.57	67733	378.1315	[M+H]+	21
387.1765	25.15	68834	386.1632	[M+H]+	15
387.1775	24.46	68834	386.1632	[M+H]+	18
391.1501	26.09	53240	−	−	−
397.2529	24.42	36160	396.2512	[M+H]+	13
401.2856	26.93	907894	400.2686	[M+H]+	24
410.1617	24.56	66912	409.1584	[M+H]+	9
		302656	409.1426	[M+H]+	28
419.2367	24.60	No results for this query mass	−	−	−
420.2402	24.56	267275	419.2209	[M+H]+	28
461.2836	26.93	433700	460.2625	[M+H]+	29
		41882	460.2825	[M+H]+	13
467.3562	26.75	42409	466.3447	[M+H]+	8
		3030	466.3447	[M+H]+	8
477.3616	23.03	634222	476.3403	[M+H]+	29
478.3647	22.68	No results for this query mass	−	−	−
481.3194	18.47	No results for this query mass	−	−	−
489.3351	26.89	No results for this query mass	−	−	−
491.3411	26.89	367039	490.3195	[M+H]+	29
		68419	490.3308	[M+H]+	6
505.3092	26.89	741838	504.2891	[M+H]+	25
507.1326	24.80	512868	506.1129	[M+H]+	24
511.3784	26.75	346153	510.3669	[M+H]+	8
512.3825	26.75	739540	511.3621	[M+H]+	25
533.3639	26.86	951964	532.3472	[M+H]+	17
535.3661	26.89	634086	534.343	[M+H]+	29
536.287	20.24	129233	535.2642	[M+H]+	28
541.2894	20.28	42027	540.2674	[M+H]+	27
543.2859	20.24	No results for this query mass	−	−	−
549.3352	26.86	No results for this query mass	−	−	−
555.4047	26.72	300126	554.3971	[M+H]+	0
578.3958	26.82	No results for this query mass	−	−	−
599.4132	26.72	No results for this query mass	−	−	−

108

599.4284	26.72	898775	598.4085	[M+H]+	21
643.4563	26.69	264170	642.4541	[M+H]+	7
701.3865	14.25	No results for this query mass	–	–	–
795.3352	25.22	718426	794.3445	[M+H]+	20
805.404	24.46	No results for this query mass	–	–	–

Table A.2.3 – List of matches for each of the 62 statistically different features in the MS-DIAL database.

m/z	Retetion Time (RT)	MS-DIAL ID matches	Adduct
158.1527	22.33	No results for this query mass	[M+H]+
161.0213	5.68	Salicyclic acid	[M+H]+
161.0985	24.00	L-alanyl-L-alanine	[M+H]+
197.0781	14.29	SUBERATE	[M+H]+
210.1443	14.70	No results for this query mass	[M+H]+
244.2234	23.24	No results for this query mass	[M+H]+
247.0907	19.68	No results for this query mass	[M+Na]+
251.1603	24.94	Normianserin	[M+H]+
253.1402	23.83	(8aR,12S,12aR)-12-hydroxy-4-methyl-4,5,6,7,8,8a,12,12a-octahydro-1H-3-benzoxecine-2,9-dione	[M+H]+
264.1908	24.32	Dendrobine	[M+Na]+
266.1702	16.79	No results for this query mass	[M+H]+
266.1786	17.32	7-methyl-3-methylidene-6-(3-oxobutyl)-4,7,8,8a-tetrahydro-3aH-cyclohepta[b]furan-2-one	[M+H]+
273.166	16.79	(E)-5-hydroxy-3-isobutyl-6-(3-methylbenzylidene)-1,6-dihydropyrazin-2(3H)-one	[M+H]+
273.1687	13.83	1,3-bis(4,4-dimethyl-4,5-dihydrooxazol-2-yl)benzene	[M+H]+
277.1053	14.98	Phomalone	[M+H]+
281.1696	25.71	No results for this query mass	[M+H]+
284.33	25.99	9,10-alpha-Epoxy-eremanthin	[M+H]+
295.1891	24.80	[6]-Gingerol	[M+H]+
298.098	7.84	5'-S-Methylthioadenosine	[M+H]+
301.177	25.68	No results for this query mass	[M+H]+
313.2325	26.82	(9Z,12E)-15,16-dihydroxyoctadeca-9,12-dienoic acid	[M+H]+
317.1707	24.84	(S)-[6]gingerol	[M+H]+
317.1715	24.00	NCGC00380965-01_C19H24O4_11a-Hydroxy-4,4,9-trimethyl-9-vinyl-1,2,3,4,9,10,11,11a-octahydrodibenzo[c,e]oxepine-5,7-dione	[M+H]+
335.2769	26.82	No results for this query mass	[M+H]+
341.1913	24.42	No results for this query mass	[M+H]+
351.1541	24.35	No results for this query mass	[M+H]+
357.258	26.82	Chenodeoxycholic acid; CE0; RUDATBOHQWOJDD-BSWAIDMHSA-N	[M+H]+

378.1209	20.21	3,4-dihydro-papaveraldine	[M+H]+
379.1469	25.57	No results for this query mass	[M+H]+
387.1765	25.15	No results for this query mass	[M+H]+
387.1775	24.46	No results for this query mass	[M+H]+
391.1501	26.09	No results for this query mass	[M+H]+
397.2529	24.42	Ginkgolic Acid C17-1	[M+H]+
401.2856	26.93	NCGC00385238-01_C25H36O4_	[M+H]+
410.1617	24.56	No results for this query mass	[M+H]+
419.2367	24.60	4-[(2R,3R)-3-[(3,4-dimethoxyphenyl)methyl]-4-methoxy-2-(methoxymethyl)butyl]-1,2-dimethoxybenzene	[M+H]+
420.2402	24.56	No results for this query mass	[M+H]+
461.2836	26.93	MMV676602	[M+H]+
467.3562	26.75	(6R)-3-hydroxy-2-methyl-6-((3R,5S,7R,8R,9S,10S,12S,13R,14S,17R)-3,7,12-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)heptanoic acid"	[M+H]+
477.3616	23.03	No results for this query mass	[M+H]+
478.3647	22.68	No results for this query mass	[M+NH4]+
481.3194	18.47	No results for this query mass	[2M+Na]+
489.3351	26.89	(E)-(3S,10R,13R)-10,13-dimethyl-17-(6-methylheptan-2-yl)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tetradecahydro-1H-cyclopenta[a]phenanthren-3-yl 3-chlorobut-2-enoate	[M]+
491.3411	26.89	No results for this query mass	[M+H]+
505.3092	26.89	Polyporenic acid C	[M+H]+
507.1326	24.80	No results for this query mass	[M+H]+
511.3784	26.75	Dehydroeburicoic acid monoacetate	[M+H]+
512.3825	26.75	No results for this query mass	[M+H]+
533.3639	26.86	Alisol A,24-acetate	[M+H]+
535.3661	26.89	No results for this query mass	[M+H]+
536.287	20.24	No results for this query mass	[M+H]+
541.2894	20.28	NCGC00385727-01_C31H40O8_(1R,5R,6R,13R,14R,16S)-6-(3-Furyl)-16-(2-methoxy-2-oxoethyl)-1,5,15,15-tetramethyl-8,17-dioxo-7-oxatetracyclo[11.3.1.0~2,11~.0~5,10~]heptadec-10-en-14-yl 2-methylpropanoate	[M+H]+
543.2859	20.24	(3S,4S,5R)-4-[(2R,3R)-2,3-dihydroxy-3-[(2S,3R,5R,10R,13R,14S,17S)-2,3,14-trihydroxy-10,13-dimethyl-6-oxo-2,3,4,5,9,11,12,15,16,17-decahydro-1H-cyclopenta[a]phenanthren-17-yl]butyl]-3,5-dimethyloxolan-2-one	[M]+
549.3352	26.86	Dehydropachymic acid	[M+K]+
555.4047	26.72	No results for this query mass	[M+H]+
578.3958	26.82	No results for this query mass	[M+H]+
599.4132	26.72	No results for this query mass	[M+NH4]+
599.4284	26.72	No results for this query mass	[M+H]+
643.4563	26.69	No results for this query mass	[M+H]+
701.3865	14.25	Cimiracemoside D	[M+H]+
795.3352	25.22	No results for this query mass	[M+H]+

110

m/z	Retetion Time (RT)		Adduct
805.404	24.46	No results for this query mass	[M+H]+

Table A.2.4 – List of matches for each of the 62 statistically different features in the SciexOS database.

m/z	Retetion Time (RT)	SciexOS ID matches	Adduct
158.1527	22.33	No results for this query mass	[M+H]+
161.0213	5.68	No results for this query mass	[M+H]+
161.0985	24.00	No results for this query mass	[M+H]+
197.0781	14.29	No results for this query mass	[M+H]+
210.1443	14.70	No results for this query mass	[M+H]+
244.2234	23.24	No results for this query mass	[M+H]+
247.0907	19.68	GAMMA,GAMMA-DIMETHYLALLYL PYROPHOSPHATE	[M+Na]+
251.1603	24.94	No results for this query mass	[M+H]+
253.1402	23.83	No results for this query mass	[M+H]+
264.1908	24.32	No results for this query mass	[M+Na]+
266.1702	16.79	No results for this query mass	[M+H]+
266.1786	17.32	No results for this query mass	[M+H]+
273.166	16.79	No results for this query mass	[M+H]+
273.1687	13.83	No results for this query mass	[M+H]+
277.1053	14.98	No results for this query mass	[M+H]+
281.1696	25.71	No results for this query mass	[M+H]+
284.33	25.99	No results for this query mass	[M+H]+
295.1891	24.80	No results for this query mass	[M+H]+
298.098	7.84	METHYLTHIOADENOSINE	[M+H]+
301.177	25.68	No results for this query mass	[M+H]+
313.2325	26.82	No results for this query mass	[M+H]+
317.1707	24.84	No results for this query mass	[M+H]+
317.1715	24.00	No results for this query mass	[M+H]+
335.2769	26.82	No results for this query mass	[M+H]+
341.1913	24.42	No results for this query mass	[M+H]+
351.1541	24.35	No results for this query mass	[M+H]+
357.258	26.82	No results for this query mass	[M+H]+
378.1209	20.21	No results for this query mass	[M+H]+

379.1469	**25.57**	**No results for this query mass**	**[M+H]+**
387.1765	25.15	No results for this query mass	[M+H]+
387.1775	24.46	No results for this query mass	[M+H]+
391.1501	26.09	No results for this query mass	[M+H]+
397.2529	24.42	No results for this query mass	[M+H]+
401.2856	26.93	No results for this query mass	[M+H]+
410.1617	24.56	No results for this query mass	[M+H]+
419.2367	24.60	No results for this query mass	[M+H]+
420.2402	24.56	No results for this query mass	[M+H]+
461.2836	26.93	No results for this query mass	[M+H]+
467.3562	26.75	No results for this query mass	[M+H]+
477.3616	23.03	No results for this query mass	[M+H]+
478.3647	22.68	No results for this query mass	[M+NH4]+
481.3194	18.47	No results for this query mass	[2M+Na]+
489.3351	26.89	CITICOLINE	[M]+
491.3411	26.89	No results for this query mass	[M+H]+
505.3092	26.89	No results for this query mass	[M+H]+
507.1326	24.80	No results for this query mass	[M+H]+
511.3784	26.75	No results for this query mass	[M+H]+
512.3825	26.75	No results for this query mass	[M+H]+
533.3639	26.86	No results for this query mass	[M+H]+
535.3661	26.89	No results for this query mass	[M+H]+
536.287	20.24	No results for this query mass	[M+H]+
541.2894	20.28	No results for this query mass	[M+H]+
543.2859	20.24	No results for this query mass	[M]+
549.3352	26.86	No results for this query mass	[M+K]+
555.4047	26.72	No results for this query mass	[M+H]+
578.3958	26.82	No results for this query mass	[M+H]+
599.4132	26.72	No results for this query mass	[M+NH4]+
599.4284	26.72	No results for this query mass	[M+H]+
643.4563	26.69	No results for this query mass	[M+H]+
701.3865	14.25	No results for this query mass	[M+H]+
795.3352	25.22	No results for this query mass	[M+H]+
805.404	24.46	No results for this query mass	[M+H]+

Table A.2.5 – List of features identified with a maximum score in HMDB, METLIN, MS-DIAL and SciexOS.

m/z	Retetion Time (RT)	Compound	Pathway
161.0213	5.68	Salicyclic acid	–
161.0985	24.00	Multiple matches ((±)-2,2'-Iminobispropanoic acid or Salicyclic acid)	–
197.0781	14.29	Suberate	–
210.1443	14.70	L-Dopa	Tyrosine metabolism
251.1603	24.94	Normianserin	–
253.1402	23.83	Multiple matches (Triamterene or (8aR,12S,12aR)-12-hydroxy-4-methyl-4,5,6,7,8,8a,12,12a-octahydro-1H-3-benzoxecine-2,9-dione)	–
266.1786	17.32	7-methyl-3-methylidene-6-(3-oxobutyl)-4,7,8,8a-tetrahydro-3aH-cyclohepta[b]furan-2-one	–
273.1660	16.79	(E)-5-hydroxy-3-isobutyl-6-(3-methylbenzylidene)-1,6-dihydropyrazin-2(3H)-one	–
273.1687	13.83	1,3-bis(4,4-dimethyl-4,5-dihydrooxazol-2-yl)benzene	–
277.1053	14.98	Phomalone	–
281.1696	25.71	Phentolamine	–
284.3300	25.99	9,10-alpha-Epoxy-eremanthin	–
295.1891	24.80	Esmolol	–
298.0980	7.84	5'-Methylthioadenosine	Cysteine and methionine metabolism
301.1770	25.68	19-oxoandrost-4-ene-3,17-dione	Steroid hormone biosynthesis
313.2325	26.82	Multiple matches (Fruticosonine or (9Z,12E)-15,16-dihydroxyoctadeca-9,12-dienoic acid)	–
317.1707	24.84	Multiple matches (Piplartine or (S)-gingerol)	–
317.1715	24.00	NCGC00380965-01_C19H24O4_11a-Hydroxy-4,4,9-trimethyl-9-vinyl-1,2,3,4,9,10,11,11a-octahydrodibenzo[c,e]oxepine-5,7-dione	–
341.1913	24.42	3-Hydroxyquinine	–
357.2580	26.82	Multiple matches (PGD2-d4 or Chenodeoxycholic acid)	–
378.1209	20.21	3,4-dihydro-papaveraldine	–

397.2529	24.42	Ginkgolic Acid C17-1	−
401.2856	26.93	NCGC00385238-01_C25H36O4_	−
419.2367	24.60	4-[(2R,3R)-3-[(3,4-dimethoxyphenyl)methyl]-4-methoxy-2-(methoxymethyl)butyl]-1,2-dimethoxybenzene	−
461.2836	26.93	MMV676602	−
467.3562	26.75	Multiple matches (2,3-epoxyphylloquinone or (6R)-3-hydroxy-2-methyl-6-((3R,5S,7R,8R,9S,10S,12S,13R,14S,17R)-3,7,12-trihydroxy-10,13-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-17-yl)heptanoic acid")	−
489.3351	26.89	Multiple matches ((E)-(3S,10R,13R)-10,13-dimethyl-17-(6-methylheptan-2-yl)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tetradecahydro-1H-cyclopenta[a]phenanthren-3-yl 3-chlorobut-2-enoate or citicoline)	−
505.3092	26.89	Polyporenic acid C	−
511.3784	26.75	Dehydroeburicoic acid monoacetate	−
533.3639	26.86	Alisol A,24-acetate	−
541.2894	20.28	NCGC00385727-01_C31H40O8_(1R,5R,6R,13R,14R,16S)-6-(3-Furyl)-16-(2-methoxy-2-oxoethyl)-1,5,15,15-tetramethyl-8,17-dioxo-7-oxatetracyclo[11.3.1.0~2,11~.0~5,10~]heptadec-10-en-14-yl 2-methylpropanoate	−
543.2859	20.24	(3S,4S,5R)-4-[(2R,3R)-2,3-dihydroxy-3-[(2S,3R,5R,10R,13R,14S,17S)-2,3,14-trihydroxy-10,13-dimethyl-6-oxo-2,3,4,5,9,11,12,15,16,17-decahydro-1H-cyclopenta[a]phenanthren-17-yl]butyl]-3,5-dimethyloxolan-2-one	−
549.3352	26.86	Dehydropachymic acid	−
701.3865	14.25	Cimiracemoside D	−

Table A.2.6 – List of the features identified in the CSF sample of the Aβ negative and Aβ positive groups belonging to the significant pathways from the Pathway Enrichment analysis performed in the MetaboAnalyst 5.0.

| Pathway name	p-value	Impact	Metabolites included in the pathway
Cysteine and methionine metabolism	0.06256	0.02089	5'-methylthioadenosine (298.0980/7.84)
Tyrosine metabolism	0.079157	0.11085	3,4-dihydroxy-L-phenylalanine (210.1443/14.70)
Steroid hormone biosynthesis	0.15575	0.04229	19-oxoandrost-4-ene-3,17-dione (301.1770/25.68)

Table A.2.7 – List of the peptides quantified in the CSF samples of the Aβ negative and Aβ positive groups.

Protein	Peptide	m/z	RT
sp|O15240|VGF_HUMAN	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEVRGA	659.505	9.2445
sp|O15240|VGF_HUMAN	APPGRPEAQPPPLSSEHKEPVAGDAVPGPKDGSAPEV	734.372	9.8850
sp|O15240|VGF_HUMAN	NAPPEPVPPPRAAPAPTHV	639.011	10.033
sp|O15240|VGF_HUMAN	APPEPVPPPR	528.795	8.5135
sp|O15240|VGF_HUMAN	ERAPLPPPAPS	566.311	9.8097
sp|Q9UHG2|PCS1N_HUMAN	DHDVGSELPPEGVLGA	796.383	12.951
sp|Q9UHG2|PCS1N_HUMAN	AADHDVGSELPPEGVLGALLR	706.039	15.857
sp|Q9UHG2|PCS1N_HUMAN	DHDVGSELPPEGVLGALLRV	691.704	17.671
sp|Q9UHG2|PCS1N_HUMAN	SPPLAETGAPR	548.293	8.7322
sp|Q9UHG2|PCS1N_HUMAN	ADHDVGSELPPEGVLGA	831.902	11.786
sp|Q9UHG2|PCS1N_HUMAN	DHDVGSELPPEGVLGALLR	658.681	16.283
sp|P02671|FIBA_HUMAN	SGEGDFLAEGGGVR	675.817	10.901
sp|P02671|FIBA_HUMAN	EGDFLAEGGGVR	603.791	10.071
sp|P02675|FIBB_HUMAN	[PGQ]-QGVNDNEE[NaX]GFFSA	709.777	14.116
sp|P02675|FIBB_HUMAN	[PGQ]-QGVNDNEEGFFSA	698.786	14.112
sp|P10451|OSTP_HUMAN	RISHELDSASSEVN	515.249	7.5818

sp|P05408|7B2_HUMAN	SVNPYLQGQRLDNVVAKKSVPHFSDEDKDPE	703.154	10.646
sp|P02649|APOE_HUMAN	KVEQAVETEPEPELR-[-1R]	799.401	10.851
sp|Q96GW7|PGCB_HUMAN	ALHPEEDPEGRQGRLLG	625.322	9.7000
sp|P01034|CYTC_HUMAN	VSPAAGSSPGKPPR	436.573	6.4420
sp|P01034|CYTC_HUMAN	VSPAAGSSP[Oxi]GKPPR	441.905	5.5581
sp|Q9NZM4|BICRA_HUMAN	PRPPPPPPP	476.274	8.3755
sp|Q2KJY2|KI26B_HUMAN	HQAKVSLQMATS	650.837	26.108
sp|O94985-2|CSTN1_HUMAN	FVDLSGHNLANPHPFAVVPSTAT	797.741	13.583
sp|P10645|CMGA_HUMAN	HSGFEDELSEVLENQSSQAELKEAVEEPSSKDVME	977.448	13.664
sp|Q9Y287|ITM2B_HUMAN	FENKFAVET	542.769	10.678
sp|Q9UHG2|PCS1N_HUMAN	AADHDVGSELPPEGVLGA	867.420	13.741
sp|P02671|FIBA_HUMAN	ADSGEGDFLAEGGGVR	768.849	11.826
sp|P02671|FIBA_HUMAN	DSGEGDFLAEGGGVR	733.331	13.425
sp|P02671|FIBA_HUMAN	KPVPDLVPGNFK	655.877	11.884
sp|P02675|FIBB_HUMAN	[PGQ]-QGVNDNEEGFFSAR	776.836	12.699
sp|P10451|OSTP_HUMAN	AQDLNAPSDWDSRGKDSYETSQLDDQSAETHSHKQS	673.133	10.118
sp|P10645|CMGA_HUMAN	HSGFEDELSEVLENQSSQAELKEAVEEPSSKDVM[Oxi]E	981.446	13.329

117
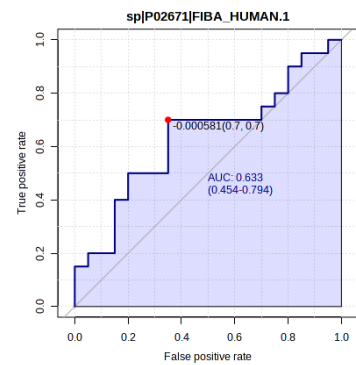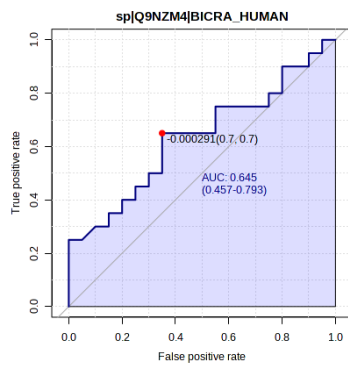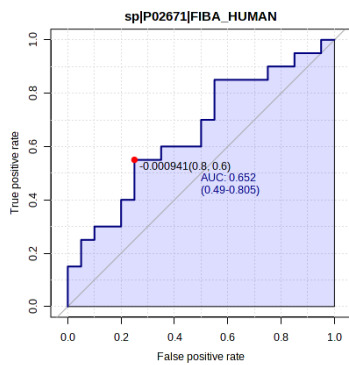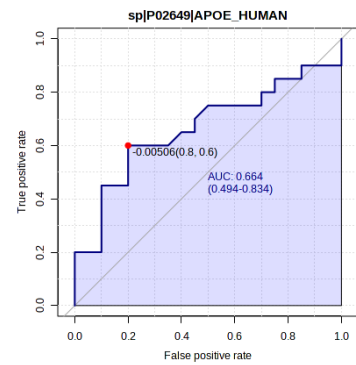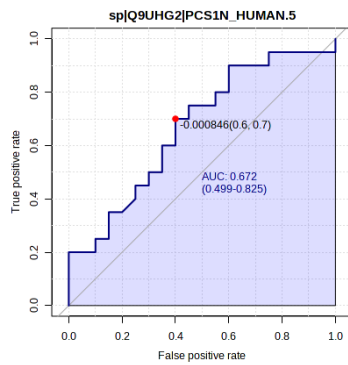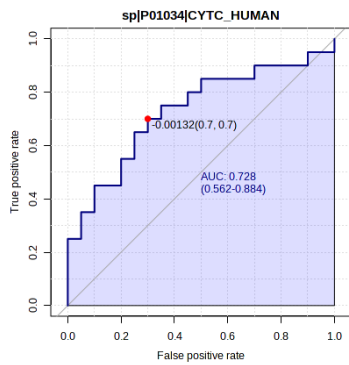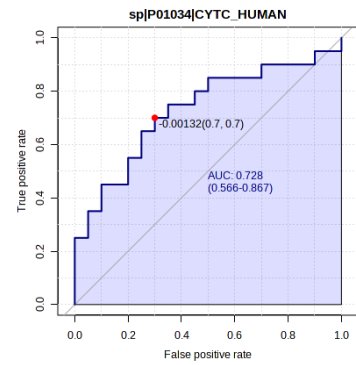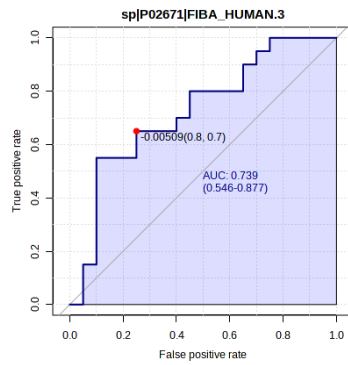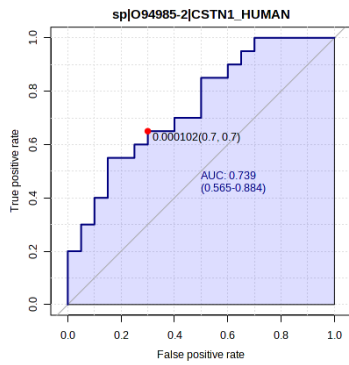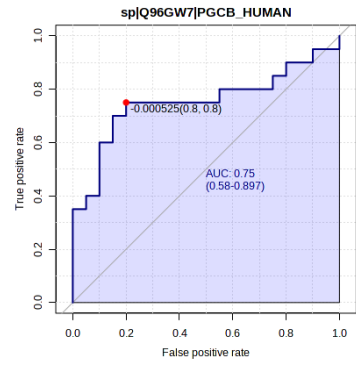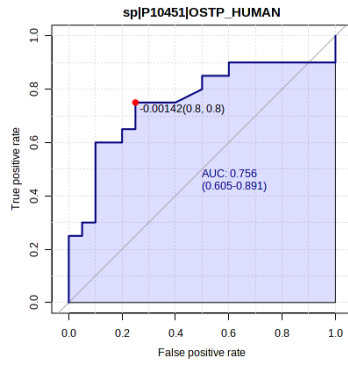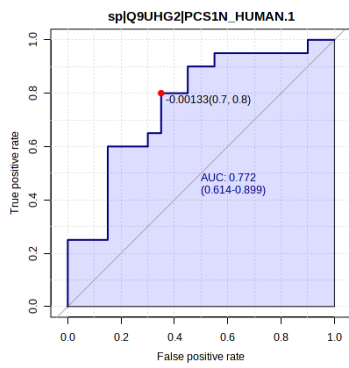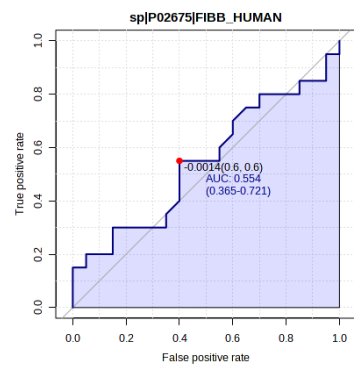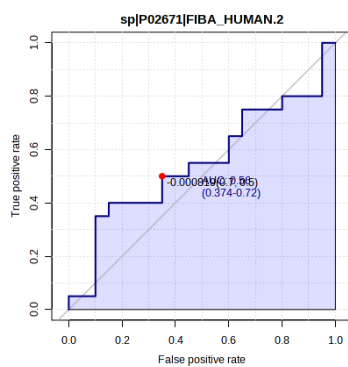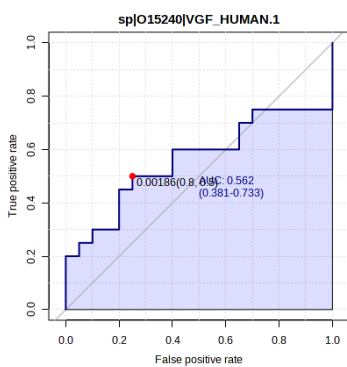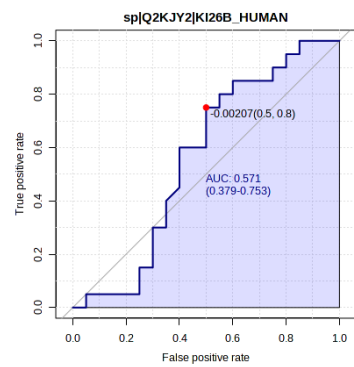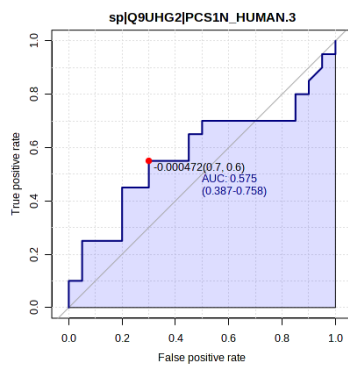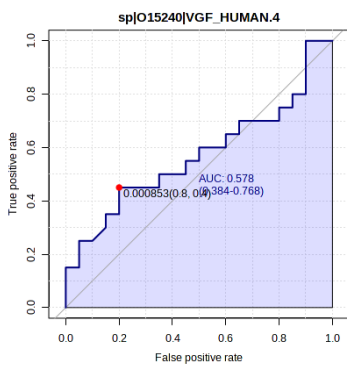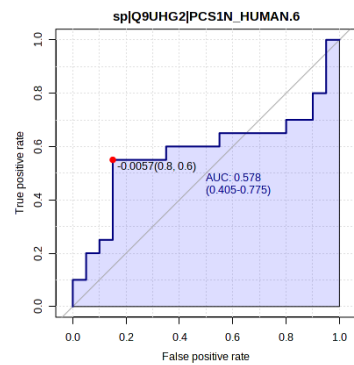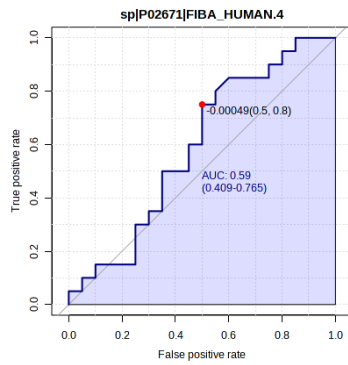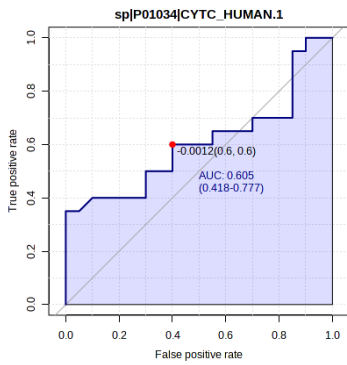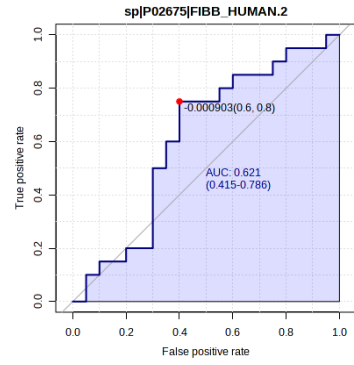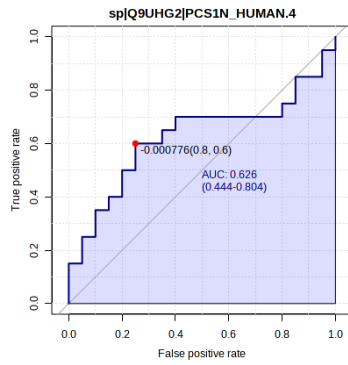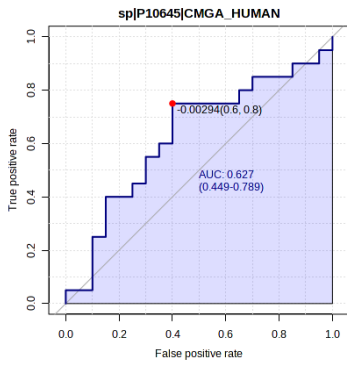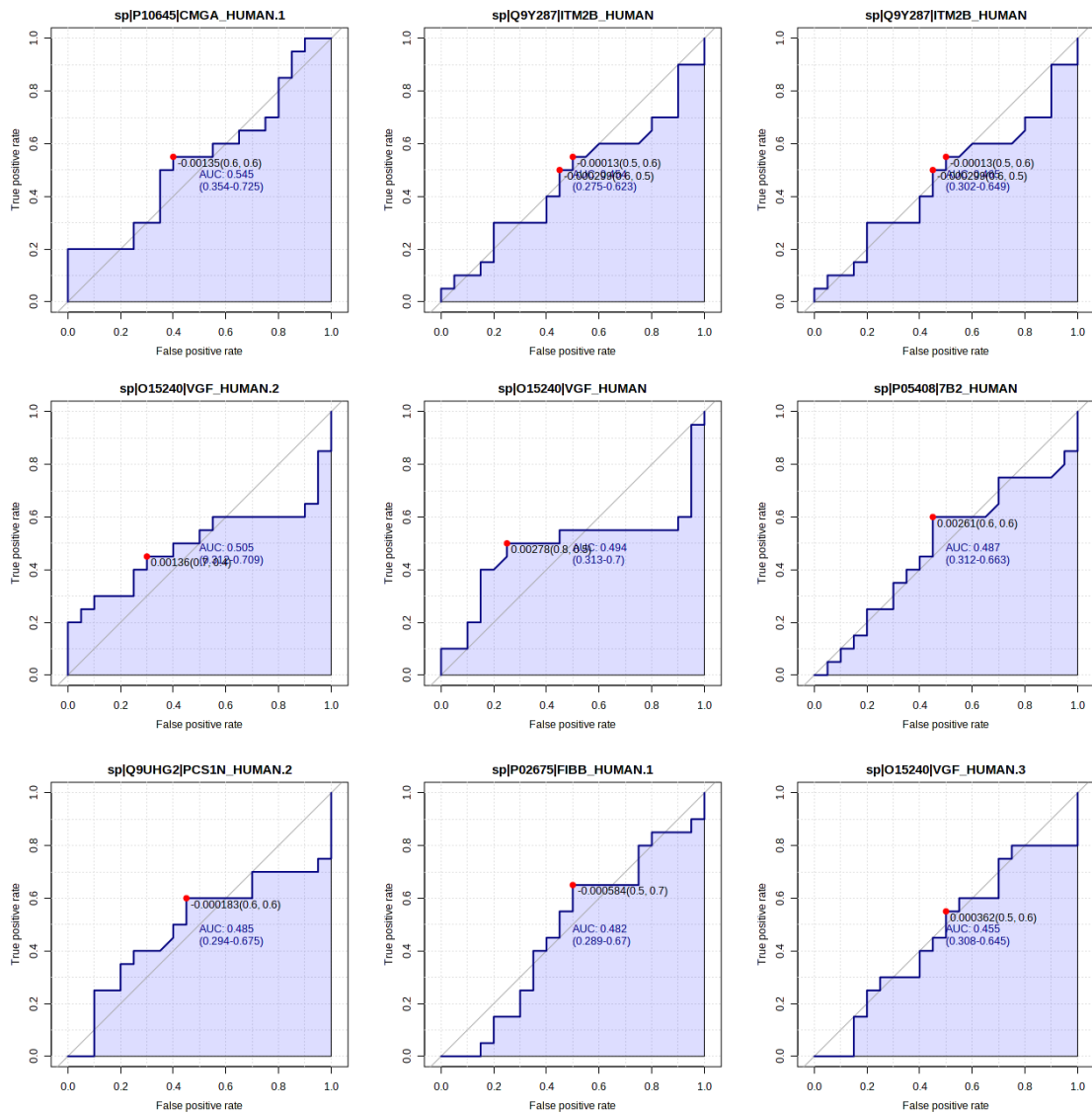
Supplementary Figure 1 – ROC curve of peptides: receiver operator characteristic (ROC) curve analysis of peptides between Aβ negative and Aβ positive groups using MetaboAnalyst 5.0.