Universidade de Aveiro
Ano 2022

Maria Manuel de Almeida Leal Santiago

Análise metabolómica não direcionada de amostras de LCR em pacientes com a Doença de Alzheimer usando LC-MS

An LC-MS untargeted metabolomics screening in CSF samples from patients with Alzheimer's Disease

Universidade de Aveiro
Ano 2022

Maria Manuel de Almeida Leal Santiago

Análise metabolómica não direcionada de amostras de LCR em pacientes com a Doença de Alzheimer usando LC-MS

An LC-MS untargeted metabolomics screening in CSF samples from patients with Alzheimer's Disease

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, realizada sobre a orientação científica do Doutor Bruno Manadas, Investigador Principal do Centro de Neurociências e Biologia Celular da Universidade de Coimbra, e do Doutor Pedro Domingues, Professor Associado com Agregação do Departamento de Química da Universidade de Aveiro

O trabalho foi financiado pela Fundação para a Ciência e Tecnologia (FCT), Portugal, com os projetos PTDC/MED-NEU/27946/2017, POCI-01-0145-FEDER-007440 (UIDB/04539/2020 and UIDP/04539/2020), POCI-01-0145-FEDER-016428 e pela Rede Nacional de Espectrometria de Massa (RNEM), sob o contracto POCI-01-0145-FEDER-402-022125 (Ref.ROTEIRO/0028/2013).

Dedico este trabalho à minha família pelo incansável apoio.

o júri

presidente

Prof. Doutor Brian James Goodfellow
Professor Auxiliar do Departamento de Química da Universidade de Aveiro

vogais

Prof. Doutora Ana Maria Pissarra Coelho Gil
Professora Associada com Agregação do Departamento de Química e membro do CICECO da Universidade de Aveiro

Doutor Bruno José Fernandes Oliveira Manadas
Investigador Principal do Centro de Neurociências e Biologia Celular da Universidade de Coimbra

agradecimentos

Gostaria de agradecer em primeiro lugar ao Dr. Bruno Manadas pela oportunidade de realizar esta tese e de toda a disponibilidade ao longo deste trabalho, assim como pelo incentivo e motivação. Agradecer também a todo o grupo do Life Sciences Mass Spectrometry por todo o conhecimento partilhado, em especial à Vera, pela disponibilidade, ajuda, ensinamentos e paciência.

Agradecer também ao Professor Pedro Domingues, pela orientação desta tese, e à Universidade de Aveiro, em particular ao Departamento de Química.

Agradecer também ao Dr. Alexandre Mendonça da Faculdade de Medicina da Universidade de Lisboa e equipa, e à Dra. Isabel Santana da Faculdade de Medicina da Universidade de Coimbra e equipa, pelo acesso às amostras utilizadas neste estudo.

Às minhas colegas de mestrado, Mariana, Diana e Erika, por todo o apoio nesta fase, e pela partilha de experiências e entreajuda. Agradecer também aos meus amigos Bia, Mariana, Sara, Daniel, Isabel, Nuno, e Adriana, por toda a motivação e apoio dado durante esta fase, assim como, por todos os bons momentos que me proporcionam.

Por fim, à minha família, por serem um exemplo e por me encorajarem a nunca desistir das minhas capacidades e do meu trabalho.

palavras-chave

Doença de Alzheimer, beta-amilóide, líquido cefalorraquidiano, metabolómica, diagnóstico

resumo

A doença de Alzheimer (DA) é a forma mais comum de demência no mundo e é caracterizada pela deterioração das funções cognitivas, o comprometimento das atividades da vida diária, e anormalidades neuropsicológicas. As marcas mais distintivas da DA são as placas senis e os emaranhados neurofibrilares. As placas senis são acumulações extracelulares principalmente compostas pela forma anormal do péptido β-amilóide, a Aβ-42. Os emaranhados neurofibrilares são compostos maioritariamente por filamentos helicoidais emparelhados de proteína Tau hiperfosforilada. A DA tende a progredir lentamente e gradualmente pior ao longo dos anos, existem cinco fases associadas com a doença: DA pré-clínica, Défice Cognitivo Ligeiro (DCL) devido à DA,e demência ligeira, moderada e severa devido à DA. No caso do DCL nem todos os pacientes progridem para a DA. O diagnóstico inclui exames físicos e neurológicos completos; avaliação do estado cognitivo, comportamental e funcional; e estudos laboratoriais e de imagiologia. Apesar das placas senis serem uma parte essencial do diagnóstico desta doença, há pacientes diagnosticados com DA que não apresentam biomarcadores positivos para o β-amilóide. As abordagens metabolómicas representam uma importante ferramenta para a descoberta de biomarcadores alternativos para diagnóstico da DA e para o desenvolvimento de um diagnóstico mais sensível e específico. O objetivo deste trabalho foi realizar um estudo metabolómico não-direcionado em amostras de líquido cefalorraquidiano (LCR) de pacientes com DCL categorizados em dois subgrupos, β-amilóide positivo e β-amilóide negativo, para fornecer dados sobre as alterações metabolómicas relacionadas com a patologia da DA. Desta forma, 40 amostras de LCR foram analisadas usando espetrometria de massa de alta resolução LC-MS, e foi realizada uma análise univariada e multivariada para identificar os elementos mais interessantes. Dos 4804 elementos detetados, foram identificadas 119 que permitiram uma melhor separação entre os dois grupos em estudo. Considerando os elementos mais interessantes, foi possível identificar 35 com uma identificação putativa e realizar uma análise mais compreensiva dos resultados. O metabolismo do triptofano, a biossíntese da arginina, o metabolismo dos glicerolípidos, o metabolismo da metionina e cisteína, o metabolismo glicerofosfolípido, e o metabolismo da arginina e prolina, foram as vias identificadas como estando alteradas no LCR das amostras analisadas. Os metabolitos 3-hidroxi-L-quinurenina, 5'-metiltioadenosina, e triptamina, estavam aumentados no LCR do grupo Aβ negativo, enquanto o L-histidinol estava diminuído. O elemento identificado com múltiplas correspondências (ornitina ou L-asparagina ou cinamaldeído) teve o melhor poder discriminatório entre os dois grupos. Neste trabalho, foi possível identificar elementos interessantes nas amostras de LCR de pacientes Aβ negativo e Aβ positivo que podem fornecer informação acerca da fisiopatologia da doença e que podem ser consideradas para diagnóstico. No futuro, estas identificações devem ser validadas, e uma comparação com o histórico clínico dos pacientes deve ser realizada para clarificar estes resultados.

keywords Alzheimer's disease, beta-amyloid, cerebrospinal fluid, metabolomics, diagnosis

abstract

Alzheimer's disease (AD) is the most common form of dementia in the world and is characterized by the deterioration of cognitive functions, the impairment of activities of daily living, and neuropsychological abnormalities. The most distinctive hallmarks of AD are senile plaques and neurofibrillary tangles. Senile plaques are extracellular accumulations principally composed of the abnormal form of β-amyloid, the Aβ-42. The neurofibrillary tangles are composed mostly of paired helical filaments consisting of hyperphosphorylated Tau protein. AD tends to progress slowly and gradually worsen over the years, there are five stages associated with this disease preclinical AD, Mild Cognitive Impairment (MCI) due to AD, mild dementia due to AD, moderate dementia due to AD and severe dementia due to AD. In the case of the MCI, not all patients progress to AD. The diagnosis includes complete physical and neurologic examinations; evaluation of cognitive, behavioral, and functional status; and laboratory and imaging studies. Although the senile plaques are an essential part of the diagnosis of this disease, there are patients diagnosed with AD that do not present positive biomarkers for β-amyloid. Metabolomics approaches represent an important tool for the discovery of alternative biomarkers for AD diagnosis and the development of a more sensitive and specific diagnosis. This study aims to do an LC-MS/MS untargeted metabolomic study on cerebrospinal fluid (CSF) samples from patients with MCI categorized into two sub-groups, the β-amyloid positive and β-amyloid negative, to provide data on metabolomic changes related with AD pathology. Therefore, 40 samples of CSF were analyzed using LC-MS high-resolution mass spectrometry, and univariate and multivariate analysis was performed to identify the more interesting features. Of the 4804 features detected, there were 119 features that allowed a better separation between the two groups in the study. Considering the most interesting features, it was possible to identify 35 features with a putative identification, and perform a more comprehensive analysis of the results. The tryptophan metabolism, arginine biosynthesis, glycerolipid metabolism, cysteine and methionine metabolism, glycerophospholipid metabolism, and arginine and proline metabolism, were the pathways identified as being altered in the CSF samples analyzed. The metabolites 3-hydroxy-L-kynurenine, 5'-methylthioadenosine, and tryptamine were increased in the CSF of the Aβ negative group, while the L-histidinol was decreased. The feature identified has the multiple matches (ornithine or L-asparagine or cinnamaldehyde) had the best discriminatory power between the two groups. In this work, it was possible to identify interesting features in the CSF samples of patients Aβ negative and Aβ positive that can provide information about the Alzheimer's disease pathophysiology and be considered for diagnosis. In the future, these identifications must be validated and a comparison with the patient's clinical history to clarify these results.

INDEX

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

^{1}H-MRS – Proton-magnetic resonance spectroscopy
2PG – 2-phosphoglyceric acid
3-HK – 3-hydroxy-L-kyurenine
3PG – 3-phosphoglyceric acid
ACIs – Acetylcholibesterase inhibitors
ACN – Acetonitrile
AD - Alzheimer's disease
AICD – APP intracellular domain
ApoE – Apolipoprotein E
APP – Amyloid precursor protein
Aβ – β amyloid
C18 – Octadecylsilane bonded silica
CE – Capillary electrophoresis
CID – Collision-induced dissociation
CNS – Central nervous system
CSF – Cerebrospinal fluid
DC – Direct-current
DDA – Data-Dependent Acquisition
DHAP – Dihydroxyacetone phosphate
DIA – Data-Independent Acquisition
DLB – Dementia with Lewy bodies
DSM – Diagnostic and Statistical Manual of Mental Disorders
EOAD – Early-onset Alzheimer's disease
ESI – Electrospray ionization
FA – Formic acid
FAD – Familial Alzheimer's disease
FC – Fold change
FDG – Fluorodeoxyglucose
FT-ICR – Fourier transform ion cyclotron
FTIR – Fourier transform infrared
GC – Gas chromatography
HILIC – Hydrophilic interaction chromatography
HPLC – High-performance liquid chromatography
ICD-10 – International Classification of Diseases 10
IDA – Information-Dependent Acquisition
IP-MS – Immunoprecipitation-mass spectrometry
LC – Liquid chromatography
LOAD – Late-onset Alzheimer's disease
LP – Lumbar puncture
m/z – Mass-to-charge ratio
MALDI – Matrix-assisted laser desorption/ionization
MAP – Microtubule-associated protein
MCI – Mild Cognitive Impairment
MMSE – Mini-Mental State Examination
MoCA – Montreal Cognitive Assessment
MRI – Magnetic resonance imaging
MRM – Multiple reaction monitoring
MS – Mass spectrometry
MS/MS – Tandem mass spectrometry
MTA – 5'-methylthioadenosine
NAA – *N*-acetyl aspartate
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 Co-operation and Development

PC – Phosphatidylcholine

PCA – Principal Component Analysis

PCC – Posterior cingulate cortex

PD - Parkinson's disease

PEP - Phosphoenolpyruvate

PET – Positron emission tomography

PFC – Prefrontal cortex

PHFs – Paired helical filaments

PiB – Pittsburgh Compound-B

PLS-DA - Partial Least Square-Discriminant Analysis

PSNE1 – Presenilin 1

PSNE2 – Presenilin 2

PSS – Phosphatidylserine synthase

Pyr – Pyruvate

Q – Quadrupole

QqQ - Triple quadrupole

QqTOF– Quadrupole time-of-flight

RF – Radiofrequency

RP-LC – Reversed-phase liquid chromatography

SAH – S-adenosylhomocysteine

SAM – S-adenosylmethionine

SD – Standard deviation

SF – Straight filaments

SM – Sphingomyelin

SOMs – Self-Organizing Maps

SPECT – Single-proton emission computed tomography

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

TCA – Tricarboxylic acid

WHO – World Health Organization

CHAPTER 1 | Introduction

1. Introduction

1.1 Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease that results in the deterioration of cognitive functions, impaired activities of daily living, and neuropsychological abnormalities. It was first described in 1906 by Alois Alzheimer, who gave his name to this dementing disease[1]. AD is the most common form of dementia, accounting for 60% to 70% of cases worldwide. According to the World Health Organization (WHO), around 50 million people suffer from dementia worldwide, of which nearly 60% of whom live in low- and middle-income countries[2].

Portugal is the fourth OECD country with more people with dementia per 1000 habitants[3]. The aging of the Portuguese population will lead to an increase in the total number of people with dementia, from 193,516 in 2018 to 346,905 in 2050, representing 1.88% and 3.82% of the overall population, respectively[4]. AD has a significant impact on the economy, being one of the most expensive disorders in the world. The total global social costs were estimated at US$ 818 billion in 2015. The global social costs of AD depend on where the patient lives, the severity of the disease, and the stage of the disease[2].

Understanding the cause of this disorder is important for an early and accurate diagnosis and effective treatment. Effective therapeutic approaches will only work with the development of sensitive and specific diagnostic tools.

1.1.1 Alzheimer's Aetiology

Since cases of AD have been reported, many risk factors associated with dementia in general and AD have been identified. Diabetes, hypertension, smoking, obesity, and dyslipidemia increase the risk as well as a history of traumatic brain injury, vascular disease, and vasculopathies. A higher level of education, physical activity, as well as the Mediterranean diet has been shown to reduce the risk of developing AD[5]. In addition to those previously mentioned, age and genetic factors account for a large part of the risk of developing AD[5,6].

Older age is one of the most important risk factors for cognitive decline and AD, suggesting that biological processes associated with aging may be involved in the pathogenesis of the disease. The prevalence of AD increases by about 19% in individuals aged 75 to 84, and from 30 to 35%, or even up to 50%, in people over 85 years of age[7]. There are reductions in gray and white matter in a cognitively normal brain, associated enlargement of the cerebral ventricles, and loss of synapses and dendrites in selected areas accompanied by senile plaques

and neurofibrillary tangles. AD could be an accelerated form of normal aging since the pathological changes identified in AD are similar, although more severe, to those present in normal aging[6].

Alzheimer's disease can be divided into two types depending on the age of onset and genetic factors. The vast majority of cases of AD (>95%) are late-onset (over 65 years), while early-onset (30 to 50 years) only accounts for less than 5% of the AD cases[6].

Sporadic AD (SAD), or late-onset AD (LOAD), is characterized by a late-onset and due to a complex interaction between genetic and environmental factors. The inheritance of the ε4 allele of apolipoprotein E (apoE) is the only well-established genetic risk factor for this type of AD. A single ε4 allele increases the risk of AD 2-3 times, and the risk increases up to 12 times in homozygotes[8]. ApoE is a protein found in the central nervous system (CNS) and secreted mainly by astrocytes and microglia. This apolipoprotein has an important function in the transport of lipids in the CNS to maintain lipid homeostasis, maintain synaptic connections, repair injured neurons, and scavenge toxins[9]. The main effect of apoE isoforms on AD risk is via the effect of apoE on β-amyloid (Aβ) metabolism, influencing the time to onset of Aβ deposition in the brain parenchyma and vasculature[10]. Additionally, apoE4 is associated with several other factors that may contribute to AD pathology, including low glucose uptake, mitochondria abnormalities, and cytoskeletal dysfunction[9].

The development of autosomal dominant familial AD, called early-onset AD (EOAD) or familial AD (FAD), is determined by the presence of mutations in one of three genes – amyloid precursor protein (APP), presenilin 1 (PSEN1), or presenilin 2 (PSEN2) – or duplication of APP[11,12]. PSEN1 is the most common cause of familial AD. Presenilins are central components of atypical aspartyl protease complexes responsible for the γ-secretase cleavage of APP. Mutation of the gene presenilin increases the ratio of Aβ-42 to Aβ-40, which appears to result from a change in function manifested by reduced γ-secretase activity[11]. In the case of mutations in the APP gene, including APP duplication and missense mutations, these directly implicate Aβ generation as a causative factor in AD pathology. APP is a type I membrane protein with a large extracellular domain and a short cytoplasmatic region that goes through two cleavage events to release Aβ from APP. Most mutations interfere with secretase activity, leading to increased production of an amyloidogenic alloform of the Aβ peptide, Aβ-42, which is toxic to neurons and can result in Aβ deposition in vessel walls[12].

Alzheimer's disease is a multifactorial neurodegenerative disorder with different factors contributing to its development, such as advanced age and genetic mutations. Although the exact cause of AD is still unknown, knowing the risk factors associated with this disease and who

is most likely to develop it represents an advantage in preventing and managing this dementing illness.

1.1.2 Clinical characteristics and pathology of Alzheimer's Disease

The symptoms of AD progress slowly over several years. An early sign of the disease is usually difficulty in remembering recent events and getting lost in familiar places. As the disease progresses, memory impairment will worsen, and other symptoms will develop. Gait disturbance, motor and occasional sensory abnormalities, and seizures are likely to occur at a later stage[2,13]. There are three stages associated with AD: preclinical AD, mild cognitive impairment due to AD, and dementia due to AD. The length of each stage is influenced by age, genetics, gender, and other factors[14-16].

AD begins long before symptoms become apparent. Preclinical AD spans from the first neuropathological brain lesions to the onset of the first clinical symptoms of AD, such as memory loss[14]. The next stage is the Mild Cognitive Impairment (MCI) due to AD, which refers to the symptomatic predementia phase of AD. At this stage, there is a more severe cognitive decline than expected for normal aging. In some cases, individuals with MCI can return to their normal cognitive function with the use of medication; therefore, it is necessary to identify those who are more likely to develop AD. The patient has the highest level of certainty that over time will progress to AD dementia if the patient meets the Core Clinical Criteria for MCI and, in addition, has positive biomarkers for Aβ deposits and neuronal injury. On the other hand, if the patient has negative biomarkers for both Aβ deposition and neuronal injury, the patient is considered to have the lowest probability of progressing to dementia due to AD[15].

Dementia can affect the patient in different ways, and the progression of the disease depends on the impact of the disease itself and the person's personality and state of health. The Alzheimer's dementia phase can be divided into mild, moderate, and severe dementia, reflecting the severity of the symptoms (Table I)[16].

Table I. Common symptoms experienced in the different stages of dementia. Adapted from Porth et al[16].

Stage of dementia	Symptoms
Mild dementia	• Memory loss of recent events • Changes in personality • Difficulty organizing and expressing thoughts • Getting lost or misplacing belongings

Moderate dementia	• Experience even greater memory loss • Undergo significant changes in personality and behavior (restlessness and agitation, unfounded suspicions, and aggressive physical behavior) • Show increasingly poor judgment and deepening confusion
Severe dementia	• Lose the ability to communicate coherently • Require daily assistance with personal care (eating, dressing, and all other daily self-care tasks) • Experience a decline in physical abilities (urinary and fecal incontinence, difficulty swallowing, unable to walk without assistance and to hold up his or her head without support) • Seizures

The cardinal features of Alzheimer's disease are senile plaques, neurofibrillary tangles (NFTs), and neuronal loss, dystrophic neuritis, and gliosis[17-19]. Senile or neuritic plaques are extracellular lesions that result from the accumulation of Aβ-40 and Aβ-42. Furthermore, these neuritic plaques contain dense bodies believed to be the remnants of lysosomes, mitochondria and paired helical filaments (PHFs)[18]. NFTs are intracellular lesions found in cell bodies or dendrites and composed mostly of PHFs of hyperphosphorylated Tau protein[19]. The senile plaques and neurofibrillary tangles can be found in both intellectually normal individuals and patients with dementia, differing from the area in which they are found. In patients with Alzheimer's disease, the amyloid plaques are found in the frontal and temporoparietal cortical areas[20], and NFTs are located primarily in the large pyramidal neurons of the Ammon's horn (hippocampus) and the cerebral neocortex[19].

As previously noted, the amyloid precursor protein, one of the most abundant proteins in the CNS, is proteolytically cleaved by two distinct pathways to release Aβ peptides, as shown in Figure 1. In the non-amyloidogenic processing pathway, APP is cleaved in the Aβ domain by α-secretase with the formation of a large soluble ectodomain (sAPPα) and an 83-residue membrane-associated C-terminal fragment (C83). Subsequent cleavage of C83 by γ-secretase leads to the formation of P3 peptide, which corresponds to residues $A\beta_{17-40}$ and $A\beta_{17-42}$, and the APP intracellular domain (AICD). In the amyloidogenic pathway, APP is cleaved at the N-terminal of the Aβ domain by β-secretase, or BACE, resulting in the generation of a soluble ectodomain (sAPPβ) and a 99-residue, membrane retained C-terminal fragment (C99). Then, the γ-secretase cleaves C99 to release Aβ peptides and AICD[12]. Aβ species are released as monomers which

progressively aggregate into dimers, trimers and oligomers, and fibrils to finally originate the amyloid plaques. The Aβ oligomers interact with neurons and glial cells, leading to oxidative stress, activation of inflammatory cascades, deregulation of Tau protein phosphorylation, and induction of neuronal apoptosis[17,18].

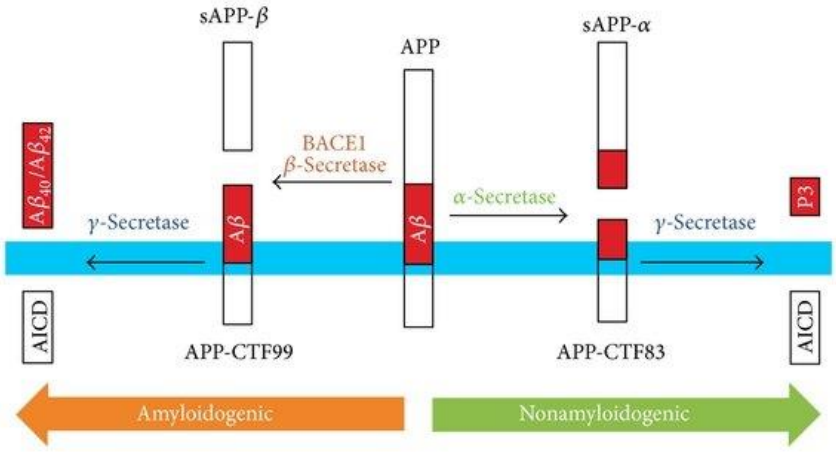

Figure 1. A diagram of the amyloid precursor protein (APP) processing pathway. The transmembrane protein APP (membrane indicated in blue) can be processed by two pathways the nonamyloidogenic pathway and the amyloidogenic pathway. Adapted from Wang *et al*[21].

The second hallmark characteristic of AD, as previously mentioned, is neurofibrillary tangles. Paired helical and straight filaments of NFTs were identified as aggregates of Tau protein. The Tau protein belongs to the microtubule-associated proteins family, can be found in most tissues, and is highly expressed in the central and peripheral nervous systems. This protein interacts with tubulin to stabilize the microtubules which are essential components for maintaining the neuronal structure, axonal transport, and synaptic plasticity. In the interaction between Tau and tubulin, Tau promotes its polymerization and inhibits the fast depolymerization of tubulin[22,23]. Under pathological conditions, including the pathophysiology of AD, Tau can be abnormally hyperphosphorylated, which impairs its capability to bind to tubulin and leads to increased self-assembly of the Tau protein. The increased aggregation creates the neurofibrillary tangles, Figure 2, which deposit in the brain of AD patients. There is also a destabilization of the microtubule structure and impairment of axonal transport and synaptic metabolism. These events result in the collapse of the cytoskeleton, loss of cellular viability, and, ultimately, neuronal death[23].

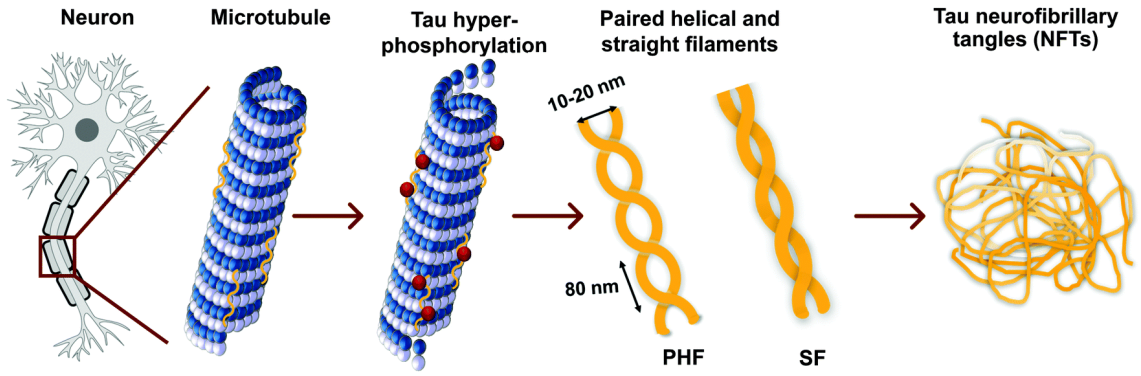

Figure 2. Tau hyperphosphorylation leads to the microtubule's destabilization. It then dissociates from microtubules and aggregates into oligomers, paired helical filaments (PHFs) and straight filaments (SF), and ultimately neurofibrillary tangles (NFTs). Adapted from Verwilst *et al*[22].

Overall, the clinical characteristics of Alzheimer's disease are the result of macroscopic and microscopic alterations in the brain. The macroscopic alterations include the atrophy in the neocortex and hippocampus, places involved in thinking and memory function, along with ventricular enlargement and smaller total brain volume. The senile plaques and neurofibrillary tangles are the most common and distinctive hallmark lesions at a microscopic level, leading to cellular viability loss and neuronal apoptosis. All these changes in the brain of AD patients are important tools in diagnosing this type of dementia.

1.1.3 Alzheimer's Diagnosis

Alzheimer's Disease is only fully diagnosed by post-mortem analysis of the brain. An accurate diagnosis of AD is important to ensure an earlier diagnosis and an appropriate treatment. The diagnosis is difficult because of the frequent presence, in older adults, of co-morbidities that can contribute to cognitive impairment. The diagnosis is based on the patient's clinical history; complete physical and neurologic examination; evaluation of cognitive, behavioral, and functional status; and laboratory and imaging studies[16].

The first step of the diagnosis is the diagnosis of dementia. The criteria used are the Diagnostic and Statistical Manual of Mental Disorders (DSM)[24], the National Institute of Neurologic and Communicative Disorders and Stroke Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA)[25] criteria, and the International Classification of Diseases (ICD)[26]. According to these guidelines, dementia is defined as the development of multiple cognitive deficits that necessarily include recent memory impairment and at least one more cognitive impairment (apraxia, agnosia, aphasia, and loss of executive function). Regarding MCI due to AD, the diagnosis follows the National Institute on Aging-Alzheimer's Association (NIA-AA)[15] recommendations. The detection and characterization of cognitive impairment require a

cognitive assessment, using, for example, the Mini-Mental State Examination (MMSE)[27] or the Montreal Cognitive Assessment (MoCA)[28]. These brief cognitive tests are designed to identify cognitive deficits at the community level and in primary health care in specific consultations. In addition, they are useful in controlling the evolution of cognitive impairment and in assessing the effectiveness of anti-dementia drugs or other intervention strategies. The MoCA is useful for the mild stages of cognitive impairment (including MCI and mild dementia due to AD), with sensitivity values of 81% and specificity of 77% for detection, while the MMSE assessment is superior for more advanced stages[28].

An essential part of the diagnosis of dementia is neuroimaging, which allows the *in vivo* examination of the central nervous system's anatomy and function. The neuroimaging methods help in the early diagnosis of Alzheimer's and differentiate AD from other neurodegenerative disorders. Several molecular imaging techniques can be used, such as structural resonance imaging (MRI), proton-magnetic resonance spectroscopy (^{1}H-MRS), and positron emission tomography (PET). The findings of these brain imaging techniques are summarized in Table II[20,29-32].

Table II. Clinically neuroimaging findings in patients with Alzheimer's disease[20,29-32].

Imaging techniques	Major findings in Alzheimer's disease
MRI	Gray matter atrophy begins in the medial temporal lobe and progresses to the temporal neocortex, parietal cortex, and frontal cortex
^{1}H-MRS	Decreased *N*-acetyl aspartate and glutamate levels Increased *myo*-inositol and cholinesterase levels
FDG-PET	Decreased fluorodeoxyglucose uptake in a specific pattern involving the temporoparietal cortex
PiB-PET	Abnormal tracer retention in the cortex
SPECT	Hypoperfusion in the temporoparietal cortex, posterior cingulate cortex, and media temporal regions

MRI is the most widely used approach for cerebral imaging *in vivo* for the assessment of dementia, as it provides detailed information about brain structure, thus allowing the characterization of regional brain atrophy and the identification of causes for the cognitive impairment, such as white matter lesions, cerebral infarcts, and brain tumors. The most consistent findings in MRI studies of patients with AD are atrophy in the medial temporal lobe

(hippocampus, amygdala, entorhinal cortex, and parahippocampal gyrus) and ventricular enlargement[29]. MRI allows for other neurodegenerative diseases to be excluded, such as dementia with Lewy bodies (DLB), frontotemporal lobe degeneration, and Parkinson's disease (PD) dementia. On the other hand, the ^{1}H-MRS provides a quantitative analysis of chemical changes in brain tissue. Different metabolites can be evaluated using this technique, like the amino acid *N*-acetyl aspartate (NAA), which reflects the functional status of neuronal mitochondria; and the glutamate and glutamine metabolites, whose metabolism is related to both neurons and glial cells. A reduction in the NAA has been reported in superior temporal and occipital areas of AD patients, suggesting a loss of neuronal function in AD, and a decrease in glutamate was found in the posterior cingulate cortex (PCC) in AD patients, implying a dysfunction and loss of glutamatergic neurons[30].

Along with MRI and MRS, PET is also a valuable tool in the diagnosis of Alzheimer's disease[20,31]. PET is an imaging technique which uses radioactive substances to visualize and measure changes in metabolic processes, chemical composition, and absorption in the brain. PET with 2-[^{18}F]-fluoro-2-deoxy-D-glucose (FDG) is used to evaluate cerebral glucose metabolism. A reduction of glucose metabolism is detected in the hippocampus, followed by the parietotemporal and PCC cortices in AD patients. Reductions in the cerebral metabolic rate for glucose begins before the onset of clinical symptoms by many years and correlate with dementia severity and pathological diagnosis of AD[31]. Another radioactive substance used with PET is the compound N-methyl-[11C]2-(4'-methylaminophenyl)-6-hydroxybenzothiazole, also named the Pittsburgh Compound-B (PiB). This compound can cross the blood-brain barrier as well as binds to fibrillar Aβ with high affinity. The preferential retention of PiB is in the frontal and temporoparietal cortical areas, where amyloid deposits in AD patients exist, as shown in Figure 3. An accumulation of the PiB is also observed in the medial parietal cortex, lateral temporal cortex, striatum, and posterior cingulate cortex[20].

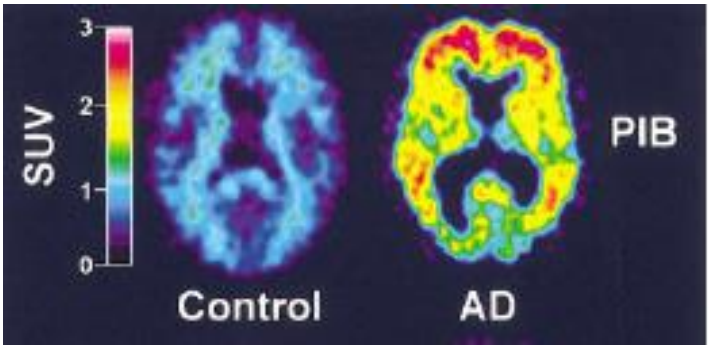

Figure 3. PiB-PET scan in healthy control (left) and an Alzheimer's disease patient (right). There is a lack of PiB retention in the entire gray matter in the control subjects, while in the AD patient there is a high PiB retention in the frontal and temporoparietal cortices. The PiB standardized uptake value (SUV) indicates increased retention of PiB in areas known to have amyloid deposition in AD patients, and therefore the mean PiB SUV is significantly greater in these patients. Adapted from Klunk *et al*[20].

A cheaper and more available alternative to PET is SPECT which comprises a radioisotope scan accomplished with single-photon emission CT (SPECT). This technique allows regional cerebral blood flow measurement, which reflects the neuronal synaptic activity. The structural and functional integrity of neurons is decreased in neurodegenerative diseases, such as AD and PD. In AD, a pattern of hypoperfusion is found in the temporoparietal, PCC, and medial temporal regions[32].

The use of biomarkers in cerebrospinal fluid (CSF) and the imaging techniques and cognitive assessments improve clinical diagnostic accuracy, especially in the earlier diagnosis of Alzheimer's disease[13,14]. The biomarkers that are used more commonly for diagnosing AD in the clinic are Aβ-42 and Tau (total and phosphorylated form) levels in CSF. Each of them is well-validated enough to be incorporated into AD diagnosis criteria and used in therapeutic trials. Both CSF Aβ-42 and amyloid PET imaging are biomarkers of Aβ plaque deposition, while the others are related to neurodegeneration[13]. The combination of low CSF Aβ-42 levels and high t-Tau and p-Tau levels is often called the "Alzheimer's disease CSF profile", with a sensitivity from 81 to 95% and a specificity from 72 to 95%[33]. These biomarkers help to differentiate AD from important differential diagnoses, such as depression and Parkinson's disease, with p-Tau levels also providing substantial aid in the differentiation of other dementias, such as frontotemporal lobar dementia and Lewy body dementia[34].

Table III. Levels of typical AD biomarkers in CSF samples. Adapted from Teunissen et al[33].

[CSF] (pg/mL)	Reference range	AD
Aβ-42	>550	365-535

t-Tau	<375	419-860
p-Tau	<52	63-112

The lower CSF levels of Aβ-42 in AD patients are the result of the aggregation of this peptide in senile plaques[12,17]. Other potential biomarkers can be evaluated to explain the increase of Aβ presence in AD patients like BACE1. The protein BACE1, also known as β-secretase, under pathological conditions cuts the APP at a specific point so that neurotoxic Aβ-42 is released extracellularly. An abnormal expression of β-secretase is one of the earliest processes of AD's pathogenesis, being a possible biomarker for the earlier stage of AD[35]. In the case of the evaluation of protein Tau, the concentration of t-Tau in CSF correlates with the intensity of neuronal degeneration, while the concentration of p-Tau reflects the pathogenesis of NFTs[23]. The measurement of protein α-synuclein can be an alternative to Tau protein since the α-synuclein induces hyperphosphorylation and aggregation of Tau proteins[34].

The cerebrospinal fluid represents the most suitable biological fluid to study neurodegenerative diseases. CSF is produced in the CNS, in equilibrium with the brain and spinal cord, and is predominantly secreted by the choroid plexus. In normal adults, this fluid volume is between 125 and 150 mL and has the important function of providing and maintaining an appropriate chemical environment for the neural tissue. CSF indirectly reflects the biochemical processes occurring in the brain. Therefore, its composition can be anticipated to provide information about states of normal or pathological metabolism of the brain. The biochemical composition of CSF resembles that of ultrafiltered blood plasma, but it also contains metabolites that the CNS secretes. The collection of CSF is an invasive procedure made by a lumbar puncture. An atraumatic spinal needle is usually inserted between the 3rd and 4th lumbar vertebrae (Figure 4) and at least 12 mL of fluid is taken. The first 1–2 mL for basic CSF assessment (protein, cell counts, erythrocytes, and others) and the last 10 mL for biobanking. There are some risks associated with this procedure, and headaches can occur within 24 hours after the puncture in less than 10% of the cases. More rarely, bleeding into the spinal canal or around the brain and infection introduced by the needle can occur[36].

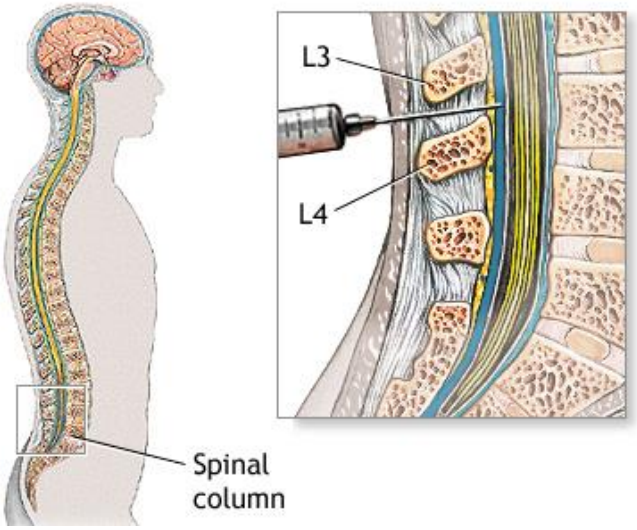

Figure 4. An illustration of the lumbar puncture procedure. The atraumatic spinal needle passes through the spinal column. Adapted from MedlinePlus Website[37].

Although CSF analysis provides important information for the diagnosis of neurological disorders, its analysis is not always easily feasible for large-scale screening. In contrast to CSF, blood-based biomarkers are minimally invasive, less costly, and could increase diagnostic accuracy. The most promising MS-based blood biomarkers for AD seem to be the determination of Aβ peptides by IP-MS (immunoprecipitation-mass spectrometry)[38] and the genotyping of apoE by MRM (multiple reaction monitoring)[39]. However, it is important to note that brain-derived biomarkers are typically present at relatively low concentrations in the blood because the blood-brain barrier prevents the free passage of molecules between the CNS and the blood[40].

To summarize, a probable AD diagnosis includes clear evidence of the decline in memory and learning, a deficit in at least one other cognitive domain, and a progressive worsening of cognitive functions. This diagnosis is supported by biomarkers of brain amyloid-β deposition and neuronal degeneration or injury. Nevertheless, new biomarkers are needed for differential diagnosis between different subtypes of dementia and a more accurate diagnosis to manage it effectively.

1.1.4 Alzheimer's Pharmacological and Non-Pharmacological therapeutics

Currently, there is no cure for Alzheimer's disease or medication that stops or slows its progression. The treatment for AD is divided into two approaches pharmacological and non-pharmacological therapeutic. These approaches aim to help people maintain mental function, manage behavioral symptoms, and slow down the symptoms of the disease.

The goal of non-pharmacological therapeutic is to maintain and improve cognitive function, motor skills, ability to perform activities of daily living, and overall quality of life.

Different non-pharmacological approaches can be chosen considering the age and gender of the patient, and the stage of the disease. An important approach for dementia patients is cognitive training, which is mainly used in the initial and middle stages of AD and consists of targeted stimulation of brain functions with a focus on multiple cognitive abilities. For example, solving a mathematical problem without a calculator. The stimulation of cognitive function will have an important role in preventing cognitive decline, promoting self-sufficiency in daily living activities, and helping maintain quality of life[41].

In the case of pharmacological therapeutic, there are only two currently approved treatments for AD by the authorities of most countries and include the cholinesterase inhibitors and the *N*-methyl-D-aspartate (NMDA) receptor antagonist memantine[16,42,43]. Acetylcholinesterase inhibitors (ACIs) – donepezil, rivastigmine, and galantamine – have been shown to be effective in slowing the progression of the disease in the mild and moderate stages by potentiating the action of available acetylcholine (ACh)[16]. ACh is the neurotransmitter used by cholinergic neurons that are selectively degenerated in AD, with loss of cholinergic function being closely related to cognitive decline in patients with this type of dementia. The therapeutic strategy is to increase the cholinergic levels in the brain by inhibiting the biological activity of acetylcholinesterase, the enzyme that catalyzes the hydrolysis of ACh to generate choline and acetate ions. The ACIs will limit the degradation of ACh and consequently increase the function of neuronal cells. The current ACIs can cause several side effects, like convulsions, nausea, facial flushing, and diarrhea. Due to this, ACIs should be taken at low doses initially and then titrated up to the maximum tolerated dose[42].

Through its action as an NMDA receptor antagonist, memantine is indicated for the treatment of patients with moderate to severe AD[16,43]. Memantine blocks the effect of pathologically elevated levels of glutamate that may otherwise lead to neuronal dysfunction[43]. Glutamate is the principal excitatory neurotransmitter in the brain, particularly in the cortical and hippocampal regions. The interaction of this neurotransmitter with specific receptors is responsible for many functions, including memory and cognition. In normal conditions, glutamate activates the NMDA receptor and calcium ions flow into the post-synaptic neuron, triggering a cascade that produces synaptic plasticity, such as long-term potentiation, and thereby facilitates the higher-order processes of learning and memory. The glutamate transport mechanism becomes immobilized during pathological conditions, causing extracellular glutamate to accumulate. This glutamate excess then drives the uncontrolled opening of the NMDA receptor, producing an increase in intracellular calcium. The excess of intracellular calcium leads to a series of processes mediated by calcium, including the release of intracellular

enzymes, free radical formation, mitochondrial injury, and eventually brain cell injury and death[16]. Adverse effects of memantine include risk of dizziness, somnolence, and weight increase[43].

Combination therapy between memantine and a cholinesterase inhibitor is common and provides benefits in treating AD[43]. It was concluded that combination therapy with memantine and any ACIs shows a positive trend when compared with ACIs monotherapy. Among ACIs, donepezil is the best ACIs to be combined with memantine since this subgroup shows the greatest improvement in cognitive function[43]. Another important part of pharmacological therapeutic is the use of antipsychotics and mood stabilizers to treat neuropsychiatric symptoms, like agitation, aggression, and sleep disturbances[16].

An approach for new drug development is the reduction of Aβ or the prevention of Aβ aggregation. A meta-analysis that considered the results from 14 randomized control trials targeting β-amyloid, concluded that in these trials the reduction in amyloid levels does not slow the cognition decline within the follow-up period[44]. An example of an anti-amyloid clinical trial that have failed in recent years is the elenbecestat (E2609). Elenbecestat is a BACE-1 inhibitor used in amyloid-PET-positive patients with MCI due to AD and mild to moderate dementia due to AD. A phase 2b study showed a decrease in CSF Aβ levels but no significant improvements in cognition function. In the case of AD patients that were submitted to anti-amyloid immunotherapy, Aβ plaques were cleared from their brains also without cognitive improvements[45]. It is also noted that some patients with typical clinical, neuropsychological and neuroimaging features of AD do not present amyloid biomarkers[46]. These results raise the question of whether the Aβ pathway is, in fact, an important pathological event for the development of Alzheimer's dementia or is instead a physiological response to neuronal damage. Therefore, it is important to discover other biomarkers for the diagnosis of AD and therapeutic intervention.

The existing treatment for AD cannot stop the progression of this disease. The medication available only helps to slow down the symptoms temporarily. On that account, future approaches should focus on developing multi-target inhibitors to treat this multifactorial disease. The multi-target inhibitors should include anti-amyloid and anti-tau effects, neurotransmitter modifications, anti-neuroinflammatory, and neuroprotective effects.

1.2 The potential of metabolomics as a tool for diagnosis

The metabolome is the global collection of all metabolites that are produced during metabolism in a biological system. The complexity and size of the metabolome are dependent on the organism[47], with the human metabolome containing 2,000 to 40,000 or even more

metabolites, with a large fraction still to be identified[48]. The metabolites are small molecules (typically <1500 Da) that are end products of multiple biological pathways and processes, the quantification and examination of these metabolites are the focus of metabolomic research[47,49]. Metabolomics measures the downstream products of the "omics cascade", which provides information about perturbations in biological systems that is not accessible through other "omics" agents. Consequently, metabolomics is the most informative of the functional status of cells (Figure 5)[47,49].

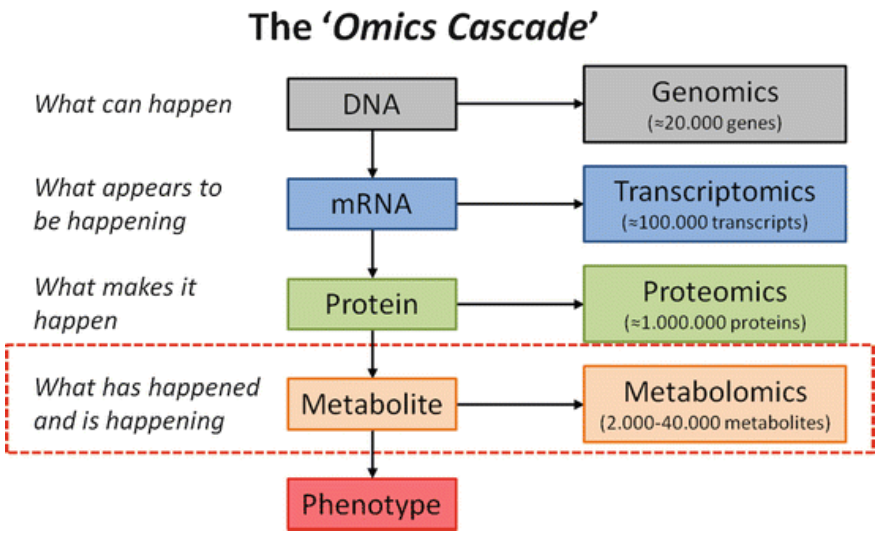

Figure 5. The "omics cascade" in systems biology. Adapted from Cortes *et al*[49].

As one of the omics technologies, metabolomics has applications in several fields, including medical science, synthetic biology, and predictive modeling of plant, animal, and microbial systems. Unlike genomics, transcriptomics, and proteomics, metabolomics provides a greater understanding of global system biology. System biology has the objective of studying all the complex interactions between the biological components of a system and understand and predict how environmental and genetic changes influence biological function at the systems level. The role of metabolomics in systems biology is to define quantitatively and qualitatively the interactions of metabolites in biological networks[47].

Metabolomics research can be divided into targeted and untargeted approaches, also called metabolic profiling and metabolic fingerprinting. The first approach refers to a method that measures a specified list of metabolites, typically focusing on one or more related pathways of interest. In the second one, the objective is to simultaneously measure as many metabolites as possible from biological samples without any bias[47,50].

Untargeted metabolomics has wide applicability across countless biological questions, given its sensitivity, high throughput, and minimal sample requirements. Metabolic

fingerprinting is frequently used in a comparative analysis of two subject groups, such as healthy vs. disease or disease A vs. disease B, making it a promising tool in studies focused on disease diagnosis and prognosis. Although untargeted metabolomics experiments are often hypothesis-driven, the construction of an experimental workflow that maximizes the number of metabolites detected and their quantitative reproducibility is an important step (Figure 6)[50].

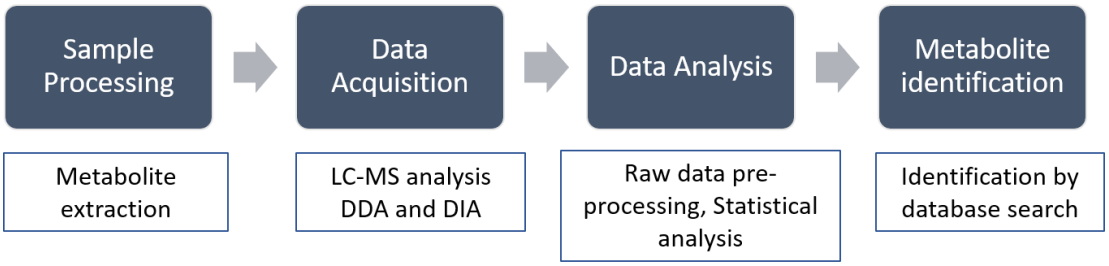

Figure 6. The untargeted workflow for LC-MS-based metabolomics.

In metabolomics studies, the main techniques used are Nuclear Magnetic Resonance (NMR) spectroscopy or mass spectrometry (MS), although other techniques such as Fourier transform infrared (FTIR) and Raman spectroscopy can also be used[47]. MS is usually coupled with a separation technique, namely gas chromatography (GC) or liquid chromatography (LC), to increase the capacity to identify more compounds and with higher sensitivity. Even though NMR allows the analysis of samples without or with reduced sample preparation, its sensitivity is low, making it difficult to analyze metabolites present in lower concentrations[49].

Metabolomic studies facilitate the understanding of biochemical fluxes and the discovery of metabolites that are indicative of abnormal biological or environmental perturbations[47]. Untargeted analysis of these biochemical transformations at a global level serves as a phenotypic screen, which can be used in clinical diagnosis to identify therapeutic targets of diseases and explore the mechanisms of fundamental biological processes[50]. Mass spectrometry has a high versatility, that has the potential of covering a large part of the metabolome. MS, along with a separation technique, represents an advantageous tool in the metabolomic study[47,50].

1.3 Application of metabolomics in Alzheimer's disease

A vast number of people have Alzheimer's disease, and it is expected that this number will continue to increase in the next years. As previously discussed, AD begins long before symptoms become apparent, and the early symptoms of AD can be the same as various

neuropathological disorders[14]. Although there are CSF biomarkers and imaging techniques available for the diagnosis of AD, the definitive diagnosis only occurs after death through an examination of brain tissue[16]. Due to these reasons, it becomes imperative to discover more accurate biomarkers for early diagnosis.

In the last decade, metabolomics approaches have been developed to improve sensitivity and specificity in diagnosing early Alzheimer's disease patients. These approaches focus on a targeted and untargeted analysis of CSF using mass-spectrometry techniques. Using these approaches, it is possible to establish panels of metabolic biomarkers and understand better the mechanism behind this disease by identifying changes in the metabolic pathways of mild cognitive and AD patients.

The metabolomic studies have identified metabolites and subsequent metabolic pathways altered in the CSF of AD patients. The metabolic pathways perturbed in AD patients are related to energy metabolism and mitochondrial function and include lipid biosynthesis and metabolism, amino acid metabolism, Krebs cycle, and homocysteine-methionine cycle[51-54]. When comparing AD patients with MCI, the total of altered pathways increases by 50% in CSF since the number of affected pathways increases with the disease severity. It is also noted that the number of altered pathways is greater in the plasma because this fluid reflects changes in organs other than the brain that are associated with AD[51].

The homocysteine-methionine cycle produces the universal methyl group donor S-adenosylmethionine (SAM), therefore having a critical role in many biochemical processes and an important contribution to human health. Dysfunction in the methionine cycle has been associated with cardiovascular disease, mild cognitive decline, vascular dementia, and Alzheimer's disease. SAM is a key product of this pathway involved in nucleic acid and protein metabolism and synthesis, as well as biosynthetic reactions of key brain metabolites. The demethylation of SAM results in the formation of S-adenosylhomocysteine (SAH). In both AD and MCI individuals, the SAH is in high concentrations and the SAM in low concentrations compared with healthy individuals[52,55]. The altered SAM/SAH ratio and/or the hyperhomocysteinemia is a hallmark of memory loss and cognitive decline in elderly populations[55]. Other metabolites related to the homocysteine-methionine, such as methionine, choline, and serine, were also elevated in patients with cognitive impairment[56-58]. Choline acts as a precursor of the neurotransmitter acetylcholine and participates in betaine metabolism, methionine metabolism, and phospholipid metabolism. Normal concentrations of choline indicate an adequate neurotransmitter activity and metabolism in the brain and CNS, the increased amounts of choline may be the reflection of neurodegeneration and the breakdown of

synaptic membranes enriched with this metabolite[56,57]. The homocysteine-methionine cycle is important for cognitive function and consequently to understand the mechanisms behind AD.

As previously mentioned, amino acids metabolism is also perturbed in AD patients. Amino acids have important roles in the immune response, neurotransmitter function, protein synthesis, among others. Alterations in amino acid pathways were observed in the plasma and CSF of both MCI and AD patients[51]. For example, arginine, alanine, and threonine were found to be decreased in AD patients[53,57], while methionine, serine, valine, and cysteine were increased in these patients[56,57,59]. Contradictory to these results, Nagata *et al* reported decreased values of serine and methionine in AD individuals[53]. The serine reduction in the CSF of patients with AD may reflect the neuroprotective role of the enzyme phosphatidylserine synthase (PSS). PSS generates phosphatidylserine by incorporating serine in phosphatidylcholine and phosphatidylethanolamine, and its administration was found to improve cognitive measures of AD individuals[53].

Another amino acid of interest is tryptophan, a precursor of serotonin, melatonin, and niacin synthesis. Tryptophan and related metabolites, like kynurenine and kynurenic acid, were decreased in preclinical AD patients[54,60]. Alterations in the balance of tryptophan metabolism are related to common features of neurogenerative disorders[60]. The neurotransmitter systems that involve acetylcholine, noradrenalin, dopamine, and serotonin are primarily affected in AD with the subsequent loss of associated neurons. The synaptic loss and malfunction occur prior to the development of Aβ plaques and neurofibrillary tangles. These alterations in the synapses are directly associated with synaptic strength and plasticity[51]. The downstream products of tryptophan degradation, quinolinic acid and kynurenic acid, have been associated with neuroinflammation in CNS diseases. The quinolinic acid was significantly higher in AD patients[61], while the kynurenic acid was decreased[54,62]. The quinolinic acid is considered neurotoxic and the kynurenic is considered to have a neuroprotective role[61,62]. Kynurenic acts as an antagonist against the NMDA receptor. In AD neuropathology, the overstimulation of the NMDA receptor can lead to neuronal death by excitotoxicity, but the presence of kynurenic acid can reduce this effect[62]. These findings of tryptophan suggest an alteration in his pathway in AD patients, with its catabolites appearing to be closely related to AD pathology. Therefore, tryptophan and its downstream products have the potential to be a biomarker for AD diagnosis and a possible target for new therapeutic interventions.

The TCA cycle (Tricarboxylic Acid cycle) also called the Krebs cycle and Citric acid cycle, is the central energy metabolism process in the mitochondria. The TCA cycle is the final common pathway for lipid, carbohydrate, and protein oxidation and plays a central role in lipogenesis,

gluconeogenesis, and interconversion of amino acids (Figure 7). The TCA cycle was found to be affected in MCI and AD patients[51]. In addition, intermediates from the TCA cycle had higher concentrations in AD patients, in both plasma and CSF, when compared to control subjects. Citrate had a statistically significant increase in both plasma and CSF of AD patients, while α-ketoglutarate had only a statistically significant increase in CSF. Alternative sources to fuel the TCA cycle, such as amino acids and fatty acids, can be activated when the brain senses a decrease of glucose in AD patients. The lower concentrations of amino acids found in AD patients' plasma suggest that readily available amino acids could have been used to replenish the TCA cycle[61].

In addition to the pool of free amino acids, fatty acid oxidation can fuel the TCA cycle by the production of acetyl-CoA. In AD, a trend of accumulation of acylcarnitines with a chain length between C6 and C12 in both CSF and plasma is noted. This increase could be related to incomplete oxidation of acyl-CoA intermediates resulting in their retro conversion to acylcarnitine for the transport and release to the plasma to avoid the toxic effects of their accumulation in mitochondria[61]. The acylcarnitine is formed by the conjugation of acyl-CoA with carnitine for the transport of long-chain fatty acids across the inner mitochondrial membrane for β-oxidation. Carnitine has a critical role in energy provision and balance via fatty acid and carbohydrate utilization. In patients without Alzheimer's disease, the carnitine concentration is high, while in patients with MCI due to AD the values are low[57]. Koal *et al* found in CSF samples with AD-like pathology that acylcarnitine (C3), sphingomyelins [SM (d18:1/18:0) and SM (d18:1/18:1)] and glycerophospholipids (PC aa C32:0, PC aa C34:1, PC aa C36:1, PC aa C38:4, and PC aa C38:6) were significantly increased[63]. Neuronal membranes contain sphingolipids (ceramides, sphingomyelins, and glycosphingolipids), being the sphingomyelins the major components of the myelin sheath. The sphingolipids play a major role in several signal transduction processes in the brain. The catabolism of these lipids is directly linked to neurodegenerative disorders in the brain. In AD pathology, pathways of sphingolipid and sphingomyelin had been identified to contribute to this dementia. This study concluded that SM (d18:1/18:0) could become a novel robust biomarker for AD to distinguish AD-like pathology in unknown samples with 70% accuracy[63].

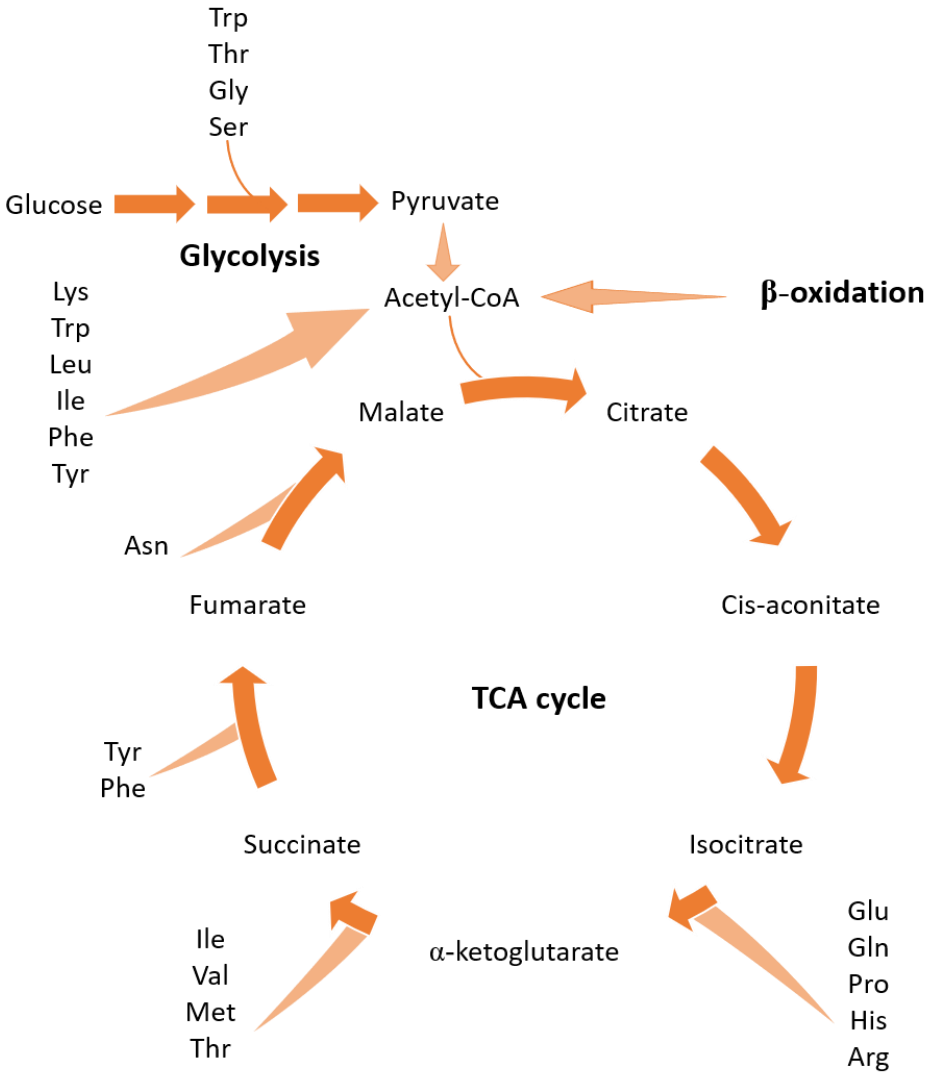

Figure 7. A scheme of the TCA cycle and its anaplerotic pathways (amino acid metabolism, glycolysis, and β-oxidation). Abbreviations Arg, Arginine; Asn, Asparagine; Gln, Glutamine; Glu, Glutamic acid; Gly, Glycine; His, Histidine; Ile, Isoleucine; Leu, Leucine; Lys, Lysine; Met, Methionine; Phe, Phenylalanine; Pro, Proline; Ser, Serine; Thr, Threonine; Trp, Tryptophan; Tyr, Tyrosine; Val, Valine.

Uracil and uridine were also identified as possible biomarkers for AD diagnosis. Uridine was found to be continuously increased from the control to AD patients. This pyrimidine is taken up by the brain and phosphorylated to form nucleotides used for DNA and RNA synthesis. The increased concentrations in AD patients suggest that these patients do not correctly metabolize uridine[60]. However, it was observed by Czech *et al* decreased levels of uridine in AD patients[59]. The decreased concentration could reflect the reduced synaptic plasticity and neuronal deficits found in AD pathology[59]. These contradictory results can be explained by considering that nucleoside levels also depend on age, gender, and brain region, suggesting that nucleosides have different functional roles in different areas. Furthermore, the specific effects of nucleosides

depend on the expression of their receptors and transporters in neuronal and glial cells[59,60]. Uridine phosphorylase catalyzes the reversible conversion of uridine to uracil. This enzyme has been suggested to have functions in initiating a cellular response to oxidative stress.

Contrary to uridine, uracil was decreased in AD patients. These decreased levels can be explained by the alteration of the protection role of the enzyme uracil-DNA glycosylase. Uracil-DNA glycosylase is a key regulatory enzyme of oxidative stress in the brain that removes oxidized pyrimidines and is overexpressed under oxidative stress as a mechanism of protection from neurodegeneration. In addition, uracil acts via channel P2X in the brain. The P2X channel controls neurotransmission, neuromodulation, cell proliferation, differentiation and death, and its alteration has been associated with different neurodegenerative diseases[60]. Both uracil and uridine can be good candidates for biomarkers for diagnosing Alzheimer's disease.

Alzheimer's disease is associated with reduced glucose metabolism in the CNS, as detected in central imaging using FDG-PET[31]. The significantly reduced concentrations of dihydroxyacetone phosphate (DHAP), 2-phosphoglyceric acid (2PG), 3-phosphoglyceric acid (3PG), phosphoenolpyruvate (PEP) and pyruvate (Pyr), indicate a reduced glycolysis activity reflected in these patients. The reduced levels of glycolytic metabolites in the CSF of AD patients can be explained by the inhibition of glycolysis under conditions of oxidative stress. An increased level of proteins involved in the regulation of glucose metabolism in AD patients, like the increase of the enzyme involved in the catabolism of glyceraldehyde-3-phosphate to 2,3-biphosphoglycerate, the glyceraldehyde-3-phosphate dehydrogenase, supports the results of aberrant glucose metabolism in AD patients[64]. Since the reduction in cerebral metabolism glucose is an early event in AD[14,15], it would be important to study with more detail the potential of the glycolysis metabolites for the diagnosis of AD, particularly in MCI and preclinical patients.

Table IV. Summary of altered metabolites in Alzheimer's disease and Mild Cognitive Impairment due to AD patients.

Metabolite	Sample	AD	MCI due to AD	References
2PG	CSF	Decreased	-	[64]
3PG	CSF	Decreased	-	[64]
Acylcarnitines	CSF and plasma	Increased	-	[61]
Alanine	CSF	Decreased	Decreased	[53]
Arginine	CSF	Decreased	Decreased	[57]
Carnitines	CSF	Increased	Decreased	[57]

Choline	CSF	Increased	Increased	[56,57]
Cysteine	CSF	Increased	Increased	[59]
DHAP	CSF	Decreased	-	[64]
Kynurenic acid	CSF	Decreased	-	[51,54,62]
Methionine	CSF	Increased*	Increased*	[56,58]
PC aa C32:0, PC aa C34:1, PC aa C36:1, PC aa C38:4, and PC aa C38:6	CSF	Increased	-	[63]
PEP	CSF	Decreased	-	[64]
Pyr	CSF	Decreased	-	[64]
Quinolinic acid	CSF	Increased	-	[61]
SAH	CSF	Increased	Increased	[52,55]
SAM	CSF	Decreased	Decreased	[52,55]
Serine	CSF	Increased*	Increased*	[56,57]
SM (d18:1/18:0) and SM (d18:1/18:1)	CSF	Increased	-	[63]
TCA cycle intermediates	CSF and plasma	Increased	Increased	[51,61]
Threonine	CSF	Decreased	Decreased	[53]
Tryptophan	CSF and plasma	Decreased	Decreased	[54,60]
Uracil	CSF	Decreased	-	[60]
Uridine	CSF	Increased**	-	[60]
Valine	CSF	Increased	Increased	[57]

*Results contradicted by Nagata *et al*[53].

**Results contradicted by Czech *et al*[59].

Abbreviations: 2PG, 2-phosphoglyceric acid; 3PG, 3-phosphoglyceric acid; AD, Alzheimer's disease; CSF, cerebrospinal fluid; MCI, Mild Cognitive Impairment; PC aa, Phosphatidylcholine with diacyl residue; PEP, Phosphoenolpyruvate; Pyr, pyruvate; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; SM, Sphingomyelin; TCA, Tricarboxylic acid.

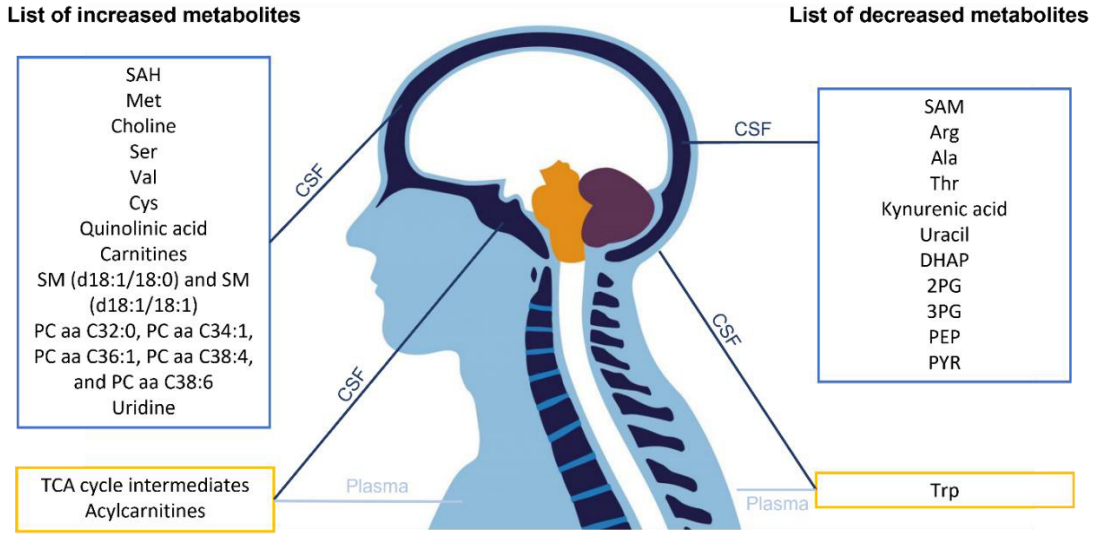

Figure 8. A scheme of the altered metabolites in Alzheimer's disease patients.

After the identification of metabolites of interest, many studies conducted prediction analysis to understand if these metabolites had the potential to improve the classification of the CSF profile of AD pathology[56,58,63,64]. In the case of metabolites that comprehend the one-carbon metabolism in the cytosol, such as methionine, serine, SAH, and choline, they were unable to improve the classification of CSF profile[56]. The unknown metabolites identified by Motsinger-Reif *et al* were able to produce initial prediction levels like those produced by the CSF-validated biomarkers for discriminating between clinically diagnosed AD and cognitively normal patients[58]. In addition, it was concluded that these metabolites have the potential to be applicable in mild stages where there is a need for accurate biomarkers. In the case of glycolysis metabolites, PEP achieved the best performance as a sole diagnostic instrument, with the combination of PEP and DHAP, and PEP and 2PG having an increase in performance. Nevertheless, none of these glycolysis metabolites reach the performance of the validated metabolites (Aβ1-42, p-Tau, and t-Tau)[64]. As already mentioned, SM (18:1/18:0) had an accuracy of 70% in distinguishing samples with AD-like pathology; however, it cannot be concluded that SM (18:1/18:0) can replace the established makers[63]. Further research needs to be done to identify more possible biomarkers that could improve the diagnosis of AD pathology.

Most of the studies stated to have the limitation of having a small number of samples, age differences between groups, and mainly male patients. Future studies will need to assess these limitations to have more clarified results. The results mentioned in this chapter show the potential of different metabolites to provide a discrimination potential of the same magnitude as the CSF amyloid and tau proteins. A further validation phase of these metabolites and others

included in metabolic pathways altered would be helpful to identify AD patients with greater precision, mainly in earlier phases.

1.4 Significance of the study

Alzheimer's disease is a multifactorial neurodegenerative disorder that manifests a progressive decline in several cognitive abilities. AD severely affects the daily living of patients and their caregivers since the diagnosis is not fully accurate and there is no cure available to treat this disease.

The amyloid plaques have been described as one of the hallmarks of AD, being its observation in the brain and its analysis in the CSF used for the diagnosis. However, they are not present in some patients with typical clinical, neuropsychological, and neuroimaging features of AD. Therefore, the discovery of other biomarkers for the diagnosis of AD, mainly in the earlier phases, where the medication can be more useful, is important.

In this work, the main aims are to do:

- An LC-MS/MS untargeted metabolomic study on CSF samples from patients with MCI categorized into two sub-groups: the β-amyloid positive (Aβ+) and the β-amyloid negative (Aβ-) group;
- Differentiation of samples into the two groups (Aβ- vs. Aβ+) through multivariate data analysis;
- The identification of the more significant altered metabolites and possibly its role in the disease, whether an amyloid response was triggered or not.

CHAPTER 2 | Analytical methodology

2. Analytical methodology

2.1 Mass spectrometry

Mass spectrometry has become a technique of choice for metabolomics studies because of its high sensibility and resolution, allowing the detection and quantification of a wide range of metabolites. MS is based on detecting ions after their separation according to their total atomic mass (m) and electrical charge (z) by electric or magnetic fields. Data generated by the mass spectrometer is represented by the relative abundance versus the ratio of the mass of the ion and its charge (as *m/z*)[47].

The main components of a mass spectrometer include the ionization source, which is used to produce ions in the gas phase resulting from the ionization of analyte molecules from the sample; the mass analyzer, where the ions will be selected and separated according to the *m/z* ratio; and a detector, in which the separated ions are collected and characterized by producing a signal whose intensity is related to the number of detected ions (Figure 9)[65,66]. The analyzer and the detector are always under a high vacuum to not influence the fragmentation pattern and not increase the background noise. In addition, the mass spectrometer may contain several analyzers, which will allow the tandem mass spectrometry (MS/MS) experimenters to be performed[65]. With tandem mass spectrometers, ions can be selected to induce their fragmentation and obtain structural information for the selected species[66].

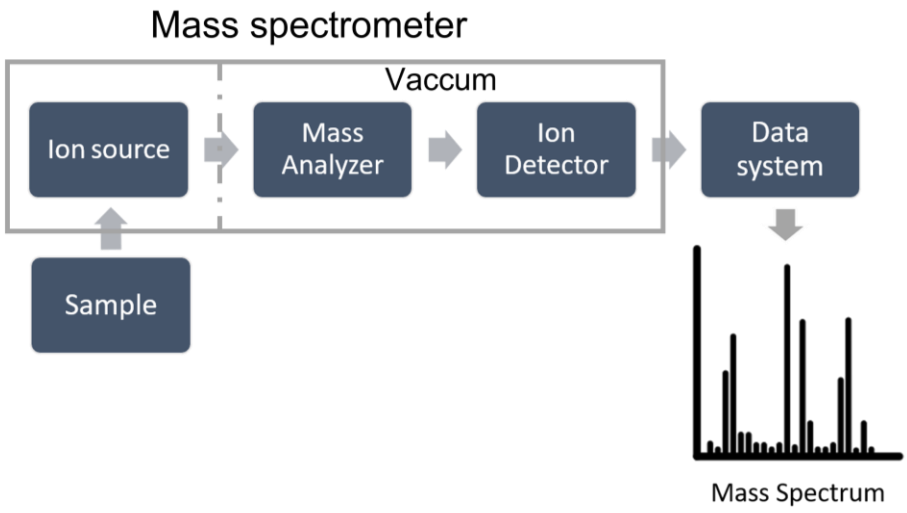

Figure 9. Example of a MS analysis, highlighting MS components and also the sample entry and the data system.

MS is generally coupled with a separation technique that will result in better detection limits and data quality. The separation is mainly divided into gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE)[47]. The separation of the compounds occurs due to the interaction between the compounds and the stationary phase, resulting in differences in the retention time[65]. GC-MS is applied to the analysis of metabolites which are volatile and thermostable. Most metabolites do not have sufficient volatility, and therefore, the derivatization is applied to increase the range of metabolites analyzed. LC is used for compounds that are not suitable for GC and are not volatile[65,66]. CE-MS is less frequently used than the others and allows the separation of polar and charged metabolites[47].

2.1.1 LC-MS based metabolomics

Liquid chromatography is a physical separation method accomplished by the metabolite equilibration between a liquid mobile phase and a solid (or liquid) stationary phase. The mobile phase traverses an LC column packed with particles on which a stationary phase is present. Liquid chromatography coupled with mass spectrometry (LC-MS) provides a powerful tool for accurate and precise quantification because of its high selectivity and high sensitivity[47,67,68]. This analytical technique allows the characterization of compounds with similar retention time, but with different *m/z* or fragmentation pattern[47,67].

In high-performance/pressure liquid chromatography (HPLC), the mobile phase is a solvent or system of solvents pumped through the column at a constant flow rate under high pressure, while the stationary phase is packed inside a chromatographic column in a way to maintain this high pressure. The reversed-phase liquid chromatography (RP-LC) is the most common variation of LC in metabolomics and the most common of all the methods used in HPLC. RP-LC consists of a silica matrix covered with non-polar materials, such as C18, that is used with a polar mobile phase to provide chromatographic separation. The mobile phase is frequently a mixture of water and either methanol or acetonitrile. Since the mobile phase is more polar than the stationary phase, more polar compounds will elute first, followed by the less polar ones[47,68]. Although the RP-LC is ideal for non-polar compounds, it is not applicable for polar compounds like sugar and some amino acids. The Hydrophilic Interaction Chromatography (HILIC) allows the separation of more polar compounds that are unretained under reversed-phase conditions. The HILIC consists of a silica column with gradient elution's running from high organic to high aqueous[47].

Different ionization sources can be used for the analysis of organic compounds and, in LC-MS based metabolomics, the most used are electrospray ionization (ESI) mainly for on-line LC-MS and matrix-assisted laser desorption/ionization (MALDI) for off-line LC-MS. ESI and MALDI

are both soft ionization techniques since they induce little or no fragmentation. While in MALDI, analytes are embedded in a solid matrix and gaseous ions are formed by exposure to a laser pulse, the ESI converts solution-phase analytes into gas-phase ions. Additionally, MALDI is performed at high vacuum and ESI at atmospheric pressure[47,65]. Both techniques have common advantages, like the simplicity of operation and the high sensitivity, requiring few picomoles of the analyte per analysis[66].

ESI is the method of choice in LC-MS-based metabolomics studies because its soft ionization capability produces many ions through charge exchange in solution and often forms intact molecular ions which support initial identification[67]. ESI is a process by which ions of an analyte are formed in the liquid state and then transferred to the gas phase into the mass spectrometer. The process is initiated when the analytical solution is pumped through a capillary and an electric field is applied. The electric field at the tip of the capillary causes the ions to accumulate in the drop emerging from the capillary. As the droplets traverse the space between the needle tip and the cone, the nebulizing gas begins the process of evaporation of the solvent. In this process, the droplets shrink until it reaches the point that the surface tension can no longer sustain the charge; consequently, a Coulomb explosion occurs, and the droplet is ripped apart until becoming a single charged molecule (Figure 10)[66].

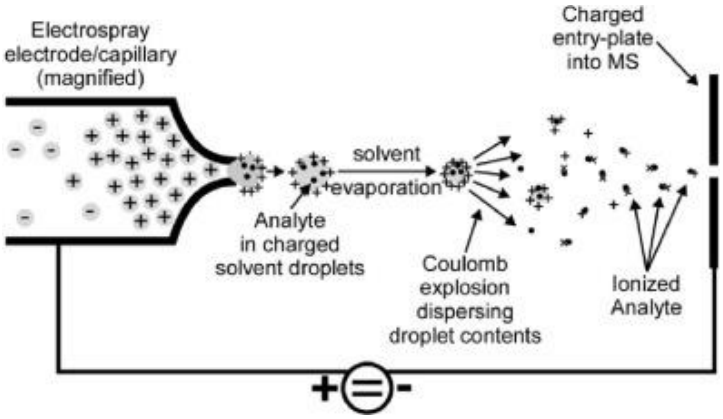

Figure 10. Schematic of ESI process, illustrating the formation of a charged droplet due to the electrical field applied. When the droplet size decreases to a point where it can no longer hold the charge, a fragmentation of the droplets occurs, the Coulomb explosion. This will force the droplets to divide into smaller droplets until each droplet corresponds to a single charged molecule. Adapted from Grebe *et al*[69].

After sample ionization, the ions enter into the mass spectrometer and then are separated by the mass analyzer according to their *m/z* ratio. The different mass analyzers can vary in mass resolution and accuracy, dynamic range, and capability to perform tandem-MS

experiments. High resolution correlates with higher mass accuracy, and this will improve the identification capabilities of the mass spectrometry equipment[66]. The more common analyzers are the quadrupole (Q), ion trap, time-of-flight (TOF), cyclotron resonance of Fourier/ Fourier transform ion cyclotron (FT-ICR), and orbitrap[65].

A quadrupole consists of four parallel rods, where direct-current (DC) and radiofrequency (RF) potentials are applied to the opposite rod pairs. Ions are consequently alternately attracted and repelled by the rod voltages as they pass through these quadrupolar fields along the central axis of the rods. If specific DC and RF are chosen, ions with a specific m/z ratio will have stable trajectories and pass the analyzer to the detector, while the other ions will not be transmitted[66]. The TOF analyzer relies on the fact that after the ions being subjected to the same kinetic energy, they will travel the same distance at different speeds according to their *m/z* values, consequently, the ions will take different times to travel across the flight tube. The ions with a lighter mass will have a shorter flight time while the heavier ions will take longer to fly towards the detector[66,70].

In MS/MS mass spectrometry, two stages of MS are involved, and a step of fragmentation occurs between the two stages. The fragmentation of ions in MS/MS experiments requires an input of energy; this is achieved by converting the kinetic energy of a collision between the selected ion and an inert collision gas into vibrational energy. This collision-induced dissociation (CID) occurs in a collision cell between the two mass analyzers[66]. The triple quadrupole (QqQ) is the most widely used MS/MS instrument and consists in three quadrupoles in sequence. The first quadrupole (Q1) operates as a mass filter to select specific ions according to their m/z ratio. The ions then enter the second quadrupole (q2), used as a collision cell, where they are bombarded by a neutral gas, such as N_2 or Ar, forcing them to fragment. After leaving q2, the generated product ions are transmitted into the third quadrupole (Q3)[66].

The hybrid MS/MS instrument that is frequently used with the ESI technique is the quadrupole time-of-flight (QqTOF). The QqTOF combines the quadrupole technologies with a TOF mass analyzer, in this hybrid system the TOF system replaces the Q3. The QqTOF instruments has the benefits of the two different mass analyzers, the high dynamic range of the quadrupole technology and the high mass resolution and accuracy of TOF (Figure 11)[70].

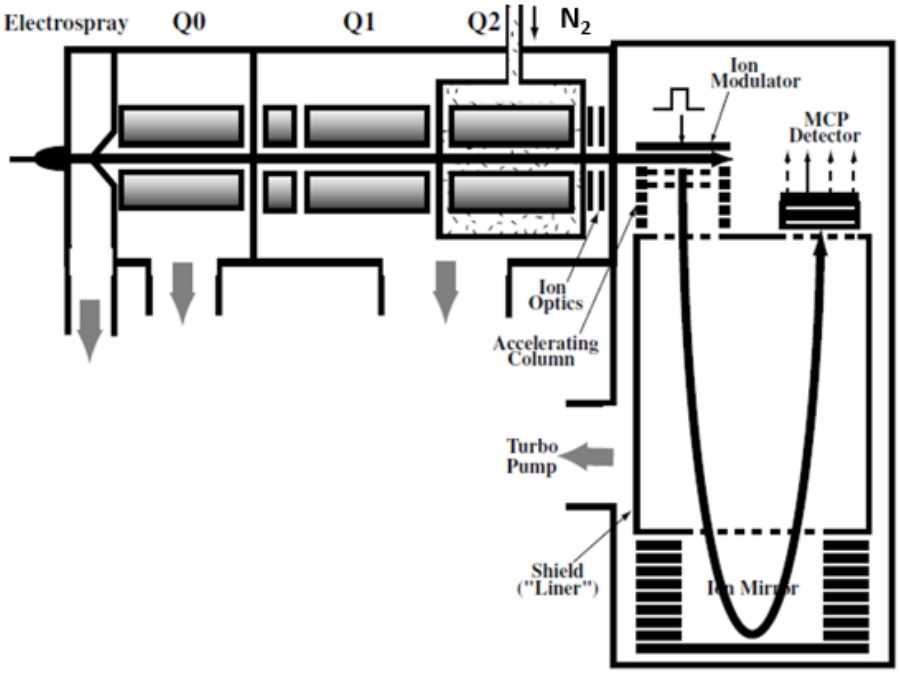

Figure 11. Schematic diagram of Q-TOF-MS. After ionization of the molecules in the ESI source, the molecular ions enter into the Q-TOF-MS, starting in Q0 for beam focusing, followed by ion selection in Q1, and fragmentation in Q2. The resulting fragments enter the TOF analyzer, allowing the detection of all fragments. Adapted from Chernushevich *et al*[70].

Before the TOF component analyzes the product ions, the molecular ions can be selected using different methods. Nowadays, the main approaches are the Data-Dependent Acquisition (DDA), the Data-Independent Acquisition (DIA), such as MS^E and SWATH (Sequential Window Acquisition of All Theoretical Fragment-Ion Spectra) (Figure 12) for SCIEX instruments. These different methods vary in how ions are selected and measured in the MS/MS scan[71].

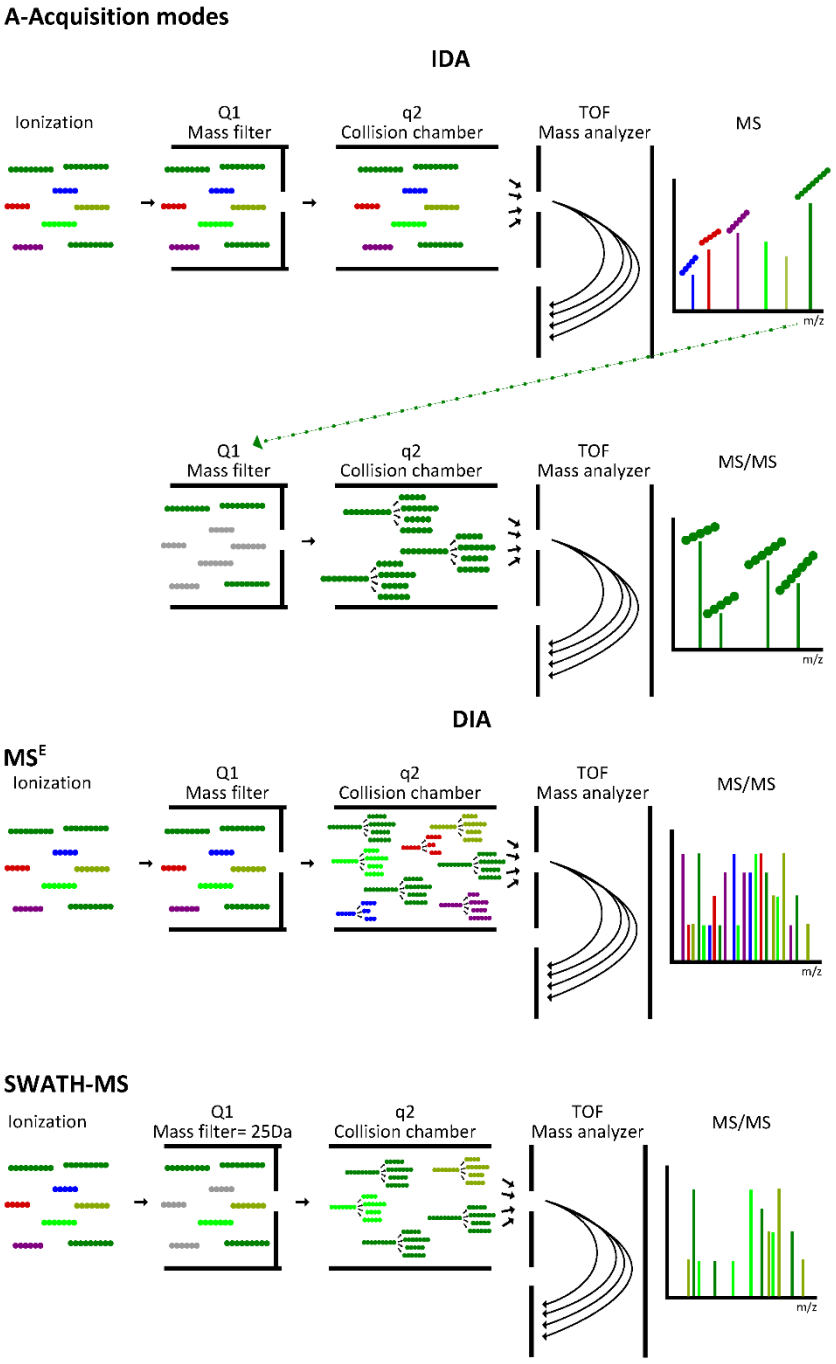

Figure 12. An illustration of data-acquisition approaches for mass spectrometry. The Information-Dependent Acquisition (IDA) obtains the fragmentation spectra only for a selected precursor ion based on pre-defined criteria. In the Data-Independent Acquisition (DIA) the fragmentation spectrum is acquired for all precursor ions independently of the pre-defined criteria, either by acquiring the entire mass range simultaneously (MS^E) or by covering the mass range in sequential smaller windows of defined size (SWATH). Adapted from Santa *et al*[72].

Mass spectrometry metabolomic studies mainly use Data-Dependent Acquisition (DDA or Information-Dependent Acquisition – IDA), where the software automatically selects the molecular ions using some pre-defined parameters. The pre-defined criteria more common is the intensity of the signal, usually selecting the most intense ions, therefore the most abundant ions. DDA can produce MS/MS spectra with minimal interference since only ions within the selected narrow *m/z* window are transferred to the collision cell Q2 to generate product ions[71]. However, this method has some limitations, important precursor ions may not be selected for fragmentation if they do not meet the defined selection criteria or if they are closely eluting isomers[71,73].

Data-Independent Acquisition (DIA), unlike DDA, submits all the ions within an *m/z* to fragmentation instead of selecting a precursor ion. DIA is independent of prior information of the sample, making all precursor ions formed in the ion source to be sent to the collision cell q2 for fragmentation, resulting in a nonselective MS/MS spectrum. One DIA technique is SWATH; this technique allows the comprehensive detection and quantification of the compounds present in a sample in a single run. SWATH, the Q1 is stepped continuously with an *m/z* window of 20 or 25 Da, covering the whole *m/z* range of interest. Then, the ions are generated under high collision energy[71]. When comparing SWATH with the MS^E the first one presents an MS/MS spectrum with more quality, but when compared with DDA it presents less selectivity in the precursors that are allowed to pass through Q1[73].

Mass spectrometry is a fast-evolving analytical technique that has become an instrument of choice for the analysis of biological samples. The use of a separation technique, such as liquid chromatography, can reduce the complexity of biological samples and decrease matrix effects during ionization. QqTOF system has a fast acquisition speed, high resolution, and good mass accuracy, being a good choice for identifying metabolites. LC-QqTOF mass spectrometry can provide qualitative and quantitative information on the metabolism.

2.2 Data analysis

Combining various data analysis strategies is necessary to allow a more comprehensive detection of metabolites in LC-MS data. The purpose of data analysis in metabolomic studies is to convert raw data into biological knowledge. MS metabolomics datasets need careful treatment since the data pre-processing steps can affect the potential for metabolite identification and subsequent quantification and biological interpretation[47].

LC-MS data is stored as raw data files, requiring extensive data pre-processing steps performed with two goals. The first one is to reduce the file size by reducing data complexity and providing the data in a suitable format, and the second one is to provide alignment of data to

ensure the correct identification of the metabolites in all samples analysed[47]. After the pre-processing step, statistical analysis is performed to detect those peaks whose intensity levels are significantly altered between different biological groups[67].

An MS measurement contains data that can be originated from different sources, such as variations in instruments or due to heterogeneity of the sample. These diverse sources can result in incorrect identification of metabolites. Therefore, pre-processing and statistical analysis are fundamental steps in the data analysis of metabolomic studies.

2.2.1 LC-MS data analysis

As previously discussed, the untargeted approach attempts the comprehensive analysis of all measurable analytes in a sample, including unknown metabolites. This approach requires the analysis of data sets with a lot of information, and because of that, previous processing of these data sets is an essential step in the data analysis[50].

After obtaining the raw data, a pre-processing step is performed to convert the raw LC-MS data into a peak list which can be easily interpreted and compared across runs. In order to correct the variations and obtain meaningful information from the raw data, multiple pre-processing steps must be performed, including outlier screening, filtering, baseline correction, peak detection, retention time alignment, and normalization of peak intensities, among others. Not all of these steps have to be made, and it can be chosen which ones are more suitable[67]. In comparative studies, peak detention and alignment are fundamental steps. The first one, peak detection, converts the raw data into a list of features (*m/z*; retention time) with the respective responses (peak areas). This conversion offers the advantage that part of the noise in the continuous data is removed. The second one aligns each feature between the samples, considering the retention time and *m/z* shifts between samples[67,73].

Software tools are available to provide these correction methods. XCMS, MZmine, MetAlign, MS-DIAL, MarkerView™ are softwares that can be used to perform peak detection and alignment. In addition to the softwares, the use of internal standards can also correct the retention times and can be used for data normalization. All these transformations in LC-MS data are needed to make the data more suitable for subsequent statistical analysis[67,73].

2.2.2 Statistical analysis

After the LC-MS data has been processed, metabolomics measurements result in signal intensities across a wide range of metabolites. Statistical analysis is applied to extract relevant information from these datasets. The statistical analysis methods can be categorized as univariate and multivariate. The univariate approach implicates the analysis of a single variable

in question and requires prior knowledge of the measured variable. In contrast, the multivariate approach implicates the analysis of multiple variables[67].

In untargeted metabolomics studies, the data generated involves a large measurement of metabolites. The multivariate analysis simplifies the data into manageable variables, facilitating an interpretation of the complex metabolomics datasets. This analysis can be unsupervised, where there is no information about the class of the data, or supervised, where information about the class of the data is known[67]. The most used unsupervised technique in LC-MS-based metabolomic study is Principal Component Analysis (PCA), this technique allows a dimensionality reduction of the dataset without a significant loss of information. PCA can also allow better visualization of the data and the determination of patterns between distinct samples, for example, control vs disease[73]. Other unsupervised techniques include hierarchical clustering and Self-Organizing Maps (SOMs)[67,73]. The Partial Least Square-Discriminant Analysis (PLS-DA) is the most used in the supervised analysis. PLS-DA aims to find the metabolite pattern that is causing the discrimination between two or more study groups[67].

Appropriate statistical techniques are needed to analyze the multiple variables that result from the metabolomic study. The analysis will extract comprehensive information from the large dataset and allow subsequent visualization and interpretation. Multivariate analysis facilitates the determination of specific patterns in datasets from biological samples, being useful in the studies that compare controls vs. disease or different diseases.

CHAPTER 3 | Materials and methods

3. Materials and Methods

3.1 Study group

For this study, 40 Mild Cognitive Impairment patients were included with ages between 46 and 80 years old. CSF samples were collected at the Centro Hospitalar e Universitário de Coimbra (Coimbra, Portugal) under the code HUC-43-09 and, more recently, using CE-029/2019. Of the CSF samples analyzed, 20 were from patients with Aβ positive and 20 from patients with Aβ negative (control group). Their characteristics are summarized in Table V and detailed information about study participants can be found in table A.1.1 of Appendix A.1.

The Aβ positive group consists of patients that meet the criteria of a high likelihood of MCI due to AD, according to Albert *et al*[15]. The Aβ negative group consists of patients with MCI that meet the clinical and cognitive criteria for MCI and the etiology of MCI consistent with the AD pathophysiological process but have negative biomarkers for both Aβ deposition and neuronal injury.

Table V. 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
Male (%)	50	45
Female (%)	50	55

3.2 Protein precipitation

From the CSF samples collected, 200 µL were used for protein precipitation. To each tube, 800 µL of methanol were added, and the samples were vortexed in an IKA™ MS 3 Basic Vortex Mixer and incubated for 1 h at –80˚C. After the incubation, samples were centrifuged for 20 min at 20,000 × g at 4˚C in an Eppendorf® Refrigerated centrifuge Model 5430R. The supernatant was transferred to new tubes and evaporated in an Eppendorf® Concentrator Plus. In the end, samples were stored at –20˚C until the LC-MS data acquisition step.

3.3 LC-MS data acquisition

3.3.1 Sample preparation

The evaporated samples were resuspended in 50 µL of a mixture that included the mobile phase, a solution of 2% acetonitrile (ACN) and 0.1% formic acid (FA), and the internal standards penicillin V potassium salt and sulfamethazine-D4. After the resuspension, they were sonicated in a Bioblock Scientific Vibracell™ 75041 for 2 min at 40% amplitude in 1-second cycles (1 second on, 1 second off). Samples were centrifuged for 10 min at 14,000 × g in a MiniSpin Plus™ Microcentrifuge and the supernatant was transferred to HPLC vials. From each sample, 7 µL were transferred to HPLC vials to create 4 pools (Aβ+ female, Aβ+ male, Aβ- female, and Aβ- male).

Data Acquisition

The MS analysis was performed on a NanoLC™ 425 System (Eksigent) coupled with an ESI DuoSpray™ (Sciex) ionization source operated in positive mode and a Triple TOF™ 6600 System mass spectrometer (Sciex). Metabolites were separated in a Triart C18 Capillary column 1/32" (12 nm, S-3 µm, 150 x 0.3 mm, YMC) at 5 µL/min, with an acetonitrile gradient as described in table VI.

Table VI. 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

The mass spectrometer was programmed for two different forms of data acquisition Data-Dependent Acquisition (DDA or IDA) and Data-Independent Acquisition (DIA-SWATH analysis). For DDA, a full mass spectra (50-2250 *m/z*) was acquired with an accumulation time of 250 ms, followed by up to 40 MS/MS scans for 40 ms per spectrum. For SWATH experiments, the equipment was set to scan the full spectra (50-2250 *m/z*) for 250 ms, followed by 73 product ion windows covering the precursor mass range of 50-1500 *m/z*. Detailed information about the 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).

3.4 Data processing and statistical analysis

The data files obtained from the LC-MS system were imported to MasterView™ (v1.3.1, Sciex) for peak detection and to MarkerView™ (v2.0, Sciex) for alignment. Peak detection processes each mass spectra of a sample by orders of increasing number. The set of parameters applied for peak detection were the following minimum retention time of 0 min, minimum spectral peak width of 0.02 Da, minimum retention time peak width of 2 min, and a threshold of 100 counts per second. The alignment step was used to decide if two features (m/z, RT) found in two samples represent the same chemical component or not. If their retention times and m/z are both within the stated tolerances, they were assumed to be the same feature. The retention time tolerance applied was 0.60 min and the mass tolerance was 40 ppm. Relative abundances for each feature were normalized using the total area sums in MarkerView™. This normalization technique is used to normalize each sample so that the resulting normalized samples have the same area sum calculated using all peaks.

Multivariate statistical analysis was performed in MetaboAnalyst 5.0 where Partial Least Square-Discriminant Analysis (PLS-DA) was used to select the most promising features, selecting the ones with VIP score above 1.00. For the univariate approach, the Mann-Whitney U test, performed by IBM SPSS statistics 28, was used to determine the statistically significant ones when the *p*-value was below 0.05. The target quantification of the fragments for the most interesting features was performed in Multiquant™ (v2.1.1, Sciex).

The Kruskal-Wallis test was performed in R software when considering four groups according to the diagnosis and the gender (Aβ negative female, Aβ negative male, Aβ positive female, and Aβ positive male) to identify which features where statistically significant different. When significance was observed, a post-hoc non-parametric Dunn's test of Multiple Comparisons with Benjamini-Hochberg *p*-value adjustment was performed in R programming software and used to see which pairs of groups differ significantly.

3.5 Metabolite identification

Metabolite identification was accomplished using two online databases, Human Metabolome Database (HMDB) (https//hmdb.ca/) and METLIN (https//metlin.scripps.edu/), and on in-house library. The in-house library was created using the kit Mass Spectrometry Metabolite Library of Standards (MSMLS) from the IROA Technologies LLC. The identification was divided into two levels the annotation level and the putative ID level, these levels were assigned based on an adapted version of the criteria described by Schrimpe-Rutledge *et al*[74].

The annotation level was performed by comparing the experimental mass measurement to the three databases within a mass tolerance window of 30 ppm. This level was divided into four sub-levels 1) confirmed in the three databases with a lower mass error; 2) confirmed in only two databases with a lower mass error; 3) confirmed in just one database with a lower mass error; and 4) confirmed in just one database with a lower mass error but with more than one hit per feature. The putative ID level, that has the highest confidence identification, was performed by using the fragmentation data of each feature obtained by DDA to compare to the fragmentation spectrum in the databases with the following parameters a parent ion tolerance of 0.05 Da, an m/z tolerance of 30 ppm and a positive ionization mode. This level was also divided into four sub-levels 1) confirmed in the three databases with a maximum score; 2) confirmed in only two databases with a maximum score; 3) confirmed in just one database with a maximum score; 4) confirmed in just one database with a maximum score but with more than one hit per feature. 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 one and RFit is the reverse. The Figure 13 illustrates the workflow used for the metabolite identification.

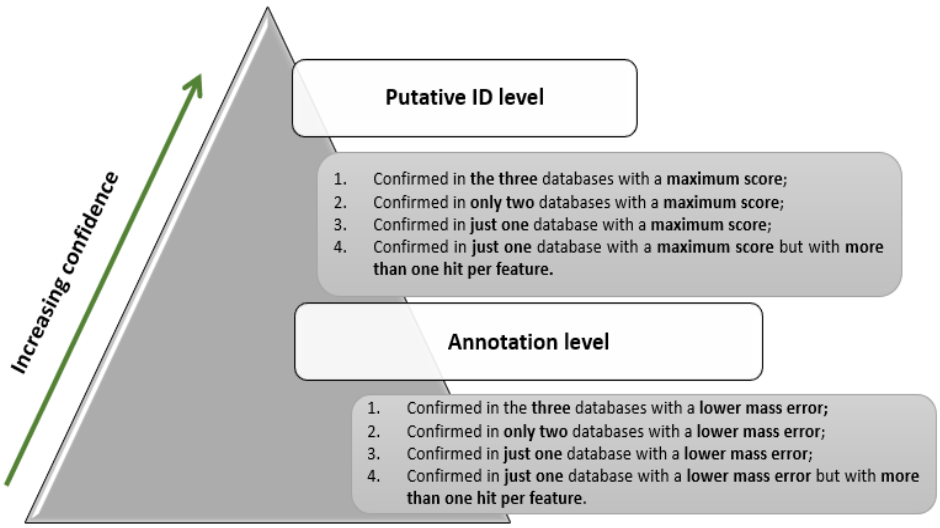

Figure 13. Workflow used for metabolite identification. The Annotation level corresponds to the match using only the experimental mass/charge of the precursor, and the putative identification (ID) corresponds to the match with the fragmentation data. The Putative ID level is the level with higher confidence. The three databases used were the HMDB, the METLIN, and the in-house library.

CHAPTER 4 | Results and discussion

4. Results and discussion

Participants' ages ranged from 46 to 80 years with 20 participants identified as Aβ negative and the other 20 identified as Aβ positive. In the Aβ negative group, there were 10 females and 10 males, and in the Aβ positive group, there were 11 females and 9 males. The Aβ positive group is the older group with a mean age of 63.8, while the Aβ negative has a mean age of 59.6. There are no significant differences between the two groups as shown in Figure 14. When dividing the participants according to their gender and diagnostic group, the Aβ positive male group is the oldest with a mean age of 64.9, and the youngest is the Aβ negative female group with a mean age of 57.6, as observed in Figure 15.

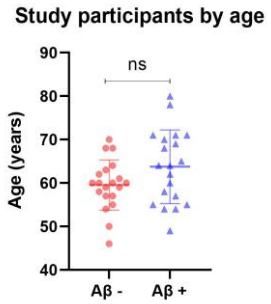

Figure 14. Data represent the mean ± SD regarding the age of the study participants. It represents the distribution, by age, of the 40 study participants divided into two groups according to the Aβ condition. The mean age corresponds to the dark-thickened line near the middle in the box plot, drawn vertically. Significance (Student's two-tailed t-test)) not significant (ns) between two groups Aβ negative (red) and Aβ positive female (blue).

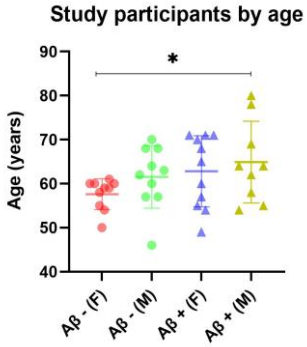

Figure 15. Data represent the mean ± SD regarding the age of the study participants. It represents the distribution, by age, of the 40 study participants divided into four groups according to the Aβ condition and gender (Aβ negative female (red), Aβ negative male (green), Aβ positive female (blue), and Aβ positive male (yellow)). The mean age corresponds to the dark thickened line near the middle in the box plot, drawn vertically. Significance (Student's two-tailed t-test)) *p<0.05 between two groups Aβ negative female and Aβ positive male.

Data files were processed for peak detection and alignment, using the parameters described in the previous chapter, and a total of 4804 features were detected and normalized using the total area sums. The number of peaks detected was reduced to 1445 features by considering only the ones present in all samples (20 Aβ positive and 20 Aβ negative). Using the 1445 features, a principal component analysis (PCA) was performed applying the Pareto scaling method, where the mean-centered values are divided by the square root of the standard deviation (Figure 16).

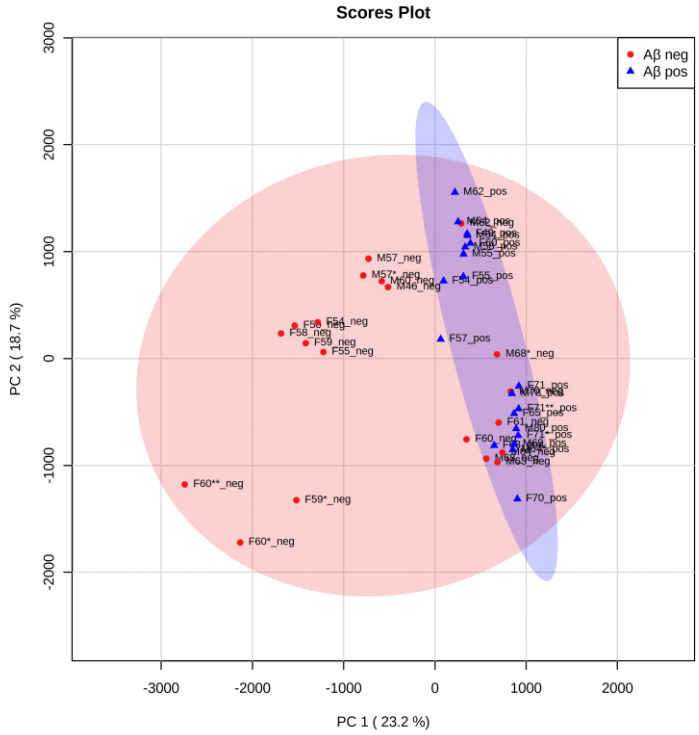

Figure 16. Principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (blue) groups using the 1445 features and the Pareto scaling performed using MetaboAnalyst.

The PCA scores plot shows in the first principal component (PC1) the separation between the two groups with an overlap of some Aβ negative samples with the Aβ positives, and it accounts for 23.2% of the variance. The second principal component (PC2) explains 18.7% of data variance and it is observed a separation according to the age of the study participants. In the Aβ negative group, it is noted the formation of three sub-groups considering the age and the gender of the participants (female samples with ages between 59 and 60; female samples with ages between 50 and 59; and male samples with ages between 46 and 60), the same effect is not observed in Aβ positive group. Eight Aβ negative samples (two females and six males) showed a different behavior from the other samples of the same group being grouped in the Aβ positive one.

A supervised analysis, partial least-squares-discriminant analysis (PLS-DA), was then performed in the same conditions as the PCA to improve the separation of the two groups (Figure 17). The PLS-DA shows a clearer separation when compared with the PCA plot but not a complete separation, PC1 explaining only 22.8% of variability and PC2 7.1%. The PLS-DA score plot had a correlation coefficient (R2) of 0.44, and a cross-validation correlation coefficient (Q2) of 0.27, these values show a moderate predictive relevance[75]. It was expected a moderate predictive ability, since in this supervised analysis it was considered the features present in all samples. Nevertheless, this plot indicates that the two groups have some different metabolic characteristics that allow this separation.

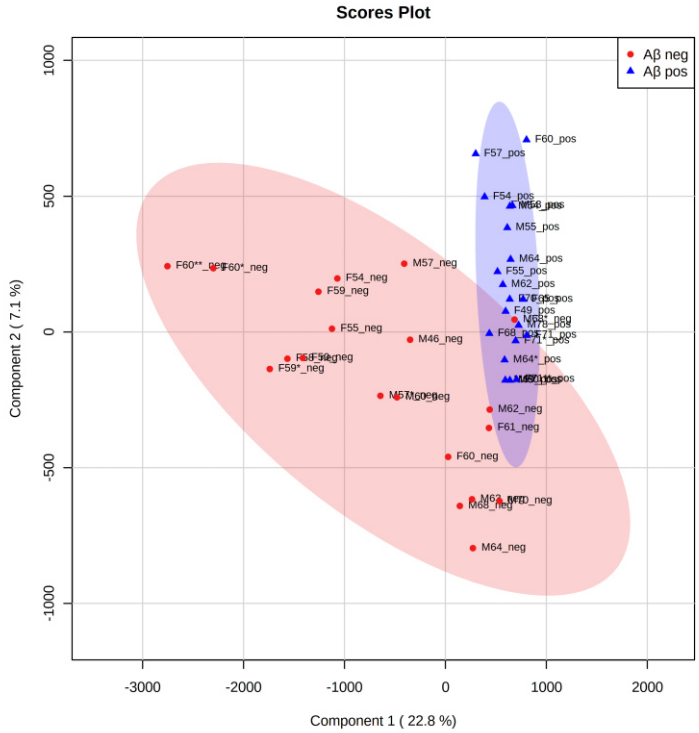

Figure 17. Partial least-squares-discriminant analysis (PLS-DA) scores for Aβ negative (red) and Aβ positive (blue) groups using the 1445 features and the Pareto scaling performed using MetaboAnalyst.

For further analysis, only the features with a *p*-value below to 0.05 and a VIP score higher than 1.00 were used. The PCA of the 119 features statistically different was performed using the Pareto scaling method (Figure 18). The results of the PCA analyses evidenced a better separation when compared with the PCA that used the 1445 features (Figure 17). The PCA also demonstrated that the variance explained by the PC1 increased to 61.9% and almost no change according to PC2.

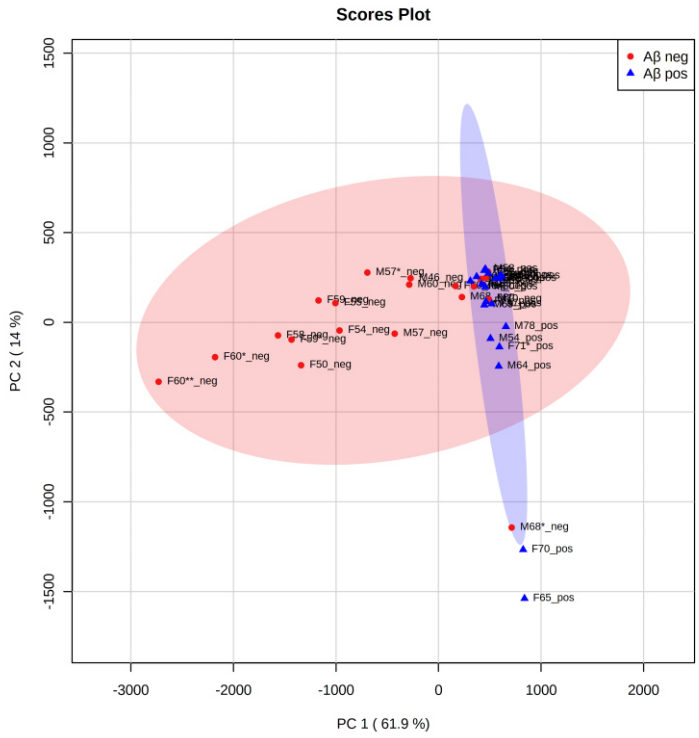

Figure 18. Principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (blue) groups using the 119 features that were statistically different considering a p<0.05 and a VIP>1.00 and the Pareto scaling performed using MetaboAnalyst.

Sample-sample correlation analysis among the Aβ negative and Aβ positive groups was conducted by Spearman correlation coefficient analysis within MetaboAnalyst, first without clustering (Figure 19) and then with hierarchical clustering (Figure 20). This analysis allowed the identification of samples that did not correlate with samples of the same group. The Spearman coefficient, shown in the plot by a degree of color intensity, can take a value between 1 to -1, where a value of 1 means a perfect association of rank and a value of -1 means a perfect negative association between ranks.

In Figure 19, it is observed that in the Aβ negative group there are two female samples (F60_neg and F61_neg) and six male samples (M62_neg, M63_neg, M64_neg, M68_neg, M68*_neg, and M70_neg) that do not show the same correlation pattern as the other samples of the Aβ negative group. These eight samples had already shown in the PCA of the 1445 features (Figure 16) a different behavior from the others of the same group. On the other hand, the Aβ positive group, all the samples present a similar correlation between them.

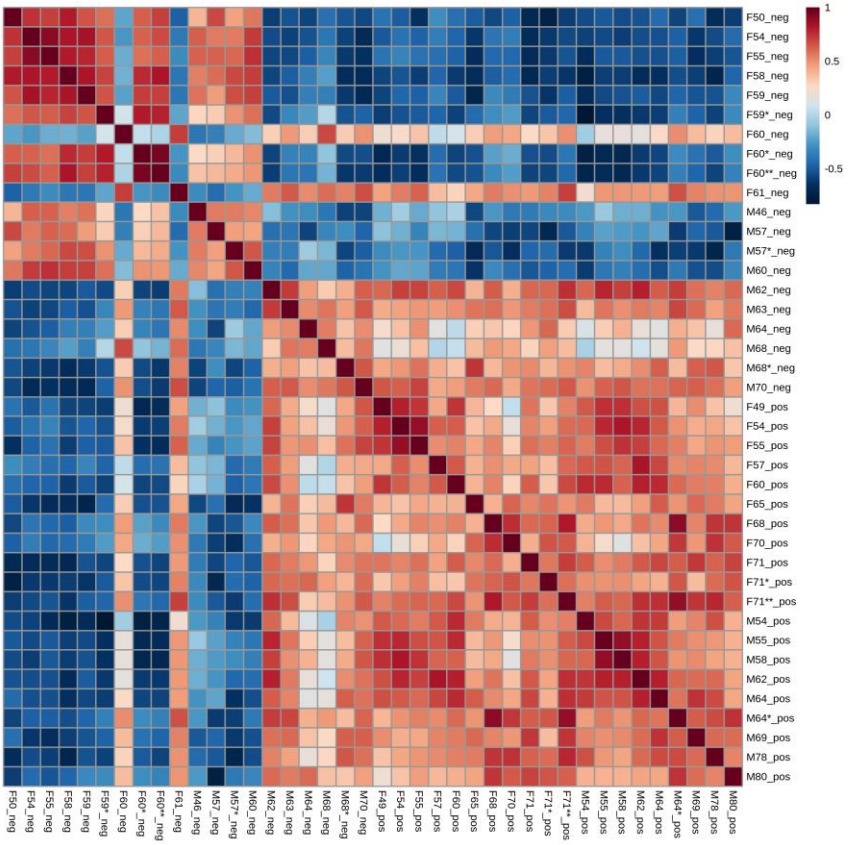

Figure 19. Correlation heatmap showing the Spearman correlation coefficient analysis of the pairwise comparison between samples. Each colored cell on the map indicates the correlation coefficient, with the scale code shown in the top right corner (red for positive correlations and blue for negative correlations).

The correlation heatmap with hierarchical clustering is shown in Figure 20 and two main clusters can be observed, one with most of the samples of the Aβ negative group and a second one with all samples of the Aβ positive group plus the eight samples of the Aβ negative group (F60_neg, F61_neg, M62_neg, M63_neg, M64_neg, M68_neg, M68*_neg, and M70_neg). The hierarchical cluster analysis also highlighted sub-groups within the two main clusters. In the first main cluster, it is observed a cluster for the female samples and a cluster for the male samples of the Aβ negative group; while in the second main cluster, there is no separation according to gender, one cluster consists of eighteen samples, and the other consists in only ten samples from both groups and both genders.

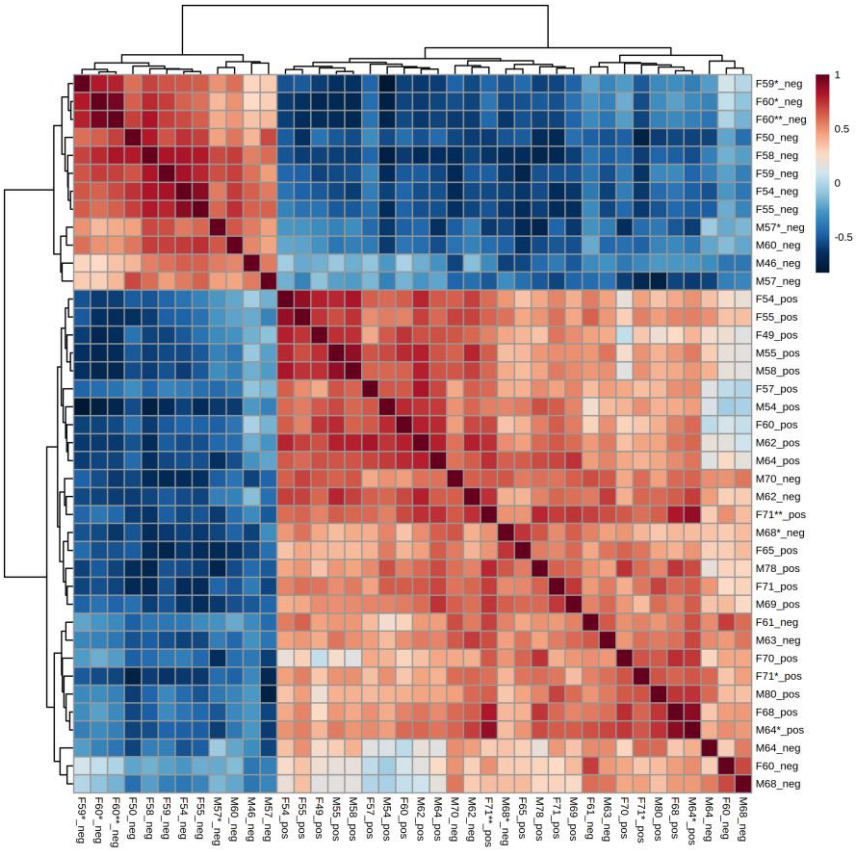

Figure 20. Correlation heatmap showing the Spearman correlation coefficient analysis of the pairwise comparison between samples. Samples order was determined by hierarchical clustering. Each colored cell on the map indicates the correlation coefficient, with the scale code shown in the top right corner (red for positive correlations and blue for negative correlations).

Hierarchical clustering analysis was also performed between the samples of the two groups and the 119 statistically different features (Figure 21). As already shown in the previous correlation heatmap (Figure 20) there are two main clusters of samples, and correlated with these two main groups there are two main clusters of metabolites associated with them. As observed in Figure 21, most of the metabolites are higher expressed in the Aβ negative group when compared with the Aβ positive samples and the eight samples of the Aβ negative group (F60_neg, F61_neg, M62_neg, M63_neg, M64_neg, M68_neg, M68*_neg, and M70_neg). In the first main cluster, that includes samples from the Aβ positive and the Aβ negative groups, we can observe the existence of four groups: the first one with the samples F70_pos, M68*_neg, F65_pos, and F49_pos; the second with the samples F60_pos, F71*_pos, M58_pos, M54_pos, M55_pos, F49_pos, F54_pos, and F55_pos; the third with M64_neg, M70_neg, F60_neg, M68_neg, F61_neg, and M63_neg; and the fourth with F71_pos, M63_neg, F57_pos, M62_neg, M62_pos, F68_pos, F71**_pos, M64*_pos, M69_pos, and M78_pos. In the second main cluster, that correspond to the Aβ negative group, there are two different clusters, one of

them only includes the two female samples with 60 years old, and a second one that can be divided in male and female samples.

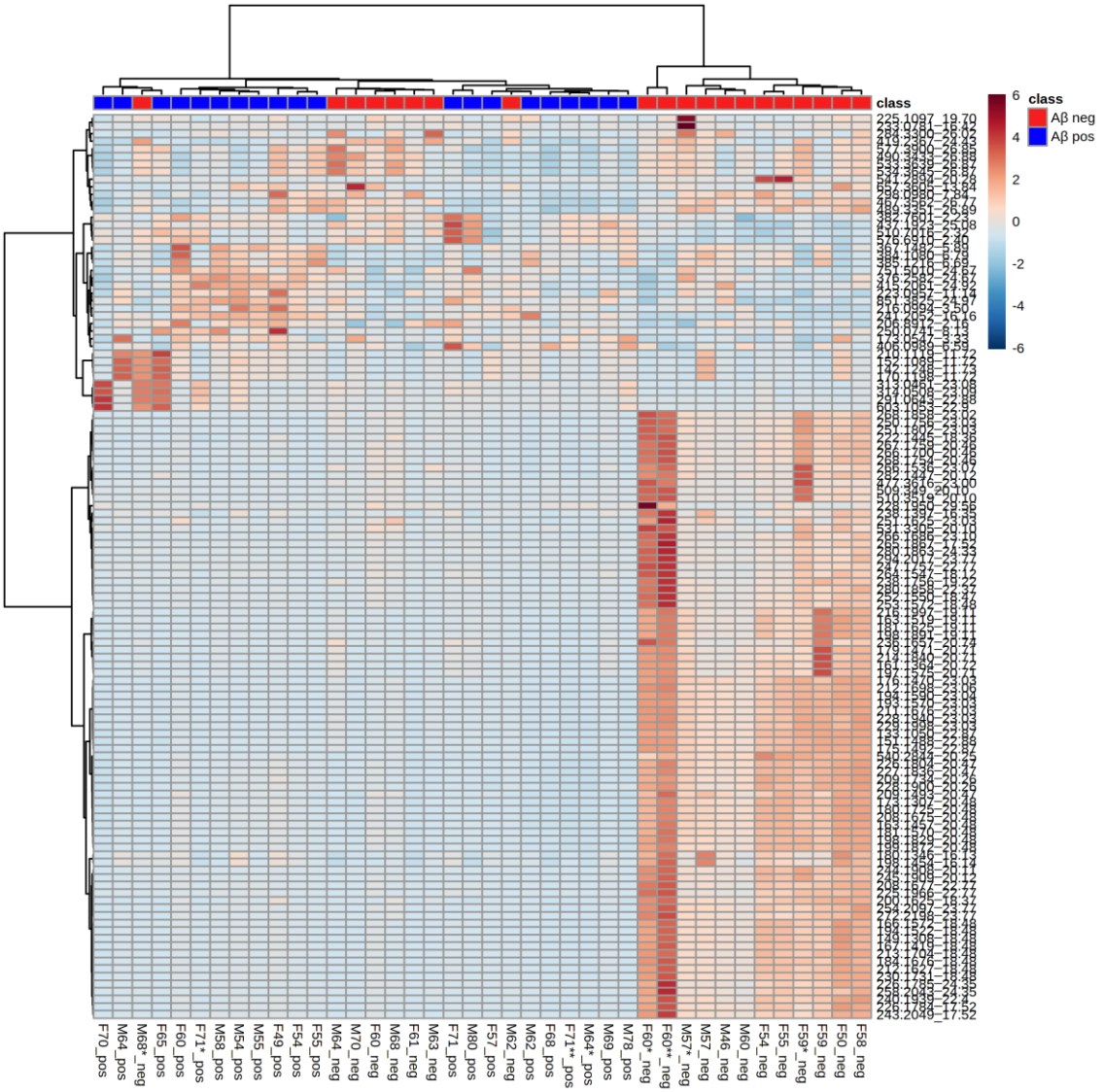

Figure 21. Heatmap with hierarchical clustering analysis by Euclidian distance using Ward's method generated from the 119 significantly different features between the two groups. The columns represent the individual samples, and the rows indicate differentiating metabolites. On top of the heatmap, Aβ negative samples correspond to the red squares and Aβ positive samples to the blue 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).

The fold change (FC) is a measure that describes how much a metabolite changes between two groups. A value of fold change of 1.5 means there is an increase of 50% between the two groups and a value of 0.67 means there is a decrease of 50%. Of the 119 features significantly different between the two groups, 13 features were significantly up (fold change

higher than 1.5) and 92 significantly down (fold change lower than 0.67) in the Aβ positive group, as shown in Figure 22. The use of logarithm to base 2 allows an easier interpretation of the plot, where features significantly increased will have a positive value, and features significantly decreased will have a negative value. Features with a 2-fold increase are displayed at an axis of 1, in this case, there are six features (241.2052_16.16, 250.0741_8.13 291.0643_22.88, 313.0461_23.08, 314.0508_23.09, and 603.1053_22.9) that have a 2-fold or more increase in the Aβ positive group. The features 226.1785_24.35, 258.2043_24.35, 477.3616_23.00 and 541.2894_20.28 have an 8-fold decrease, displayed at the axis value of -3.

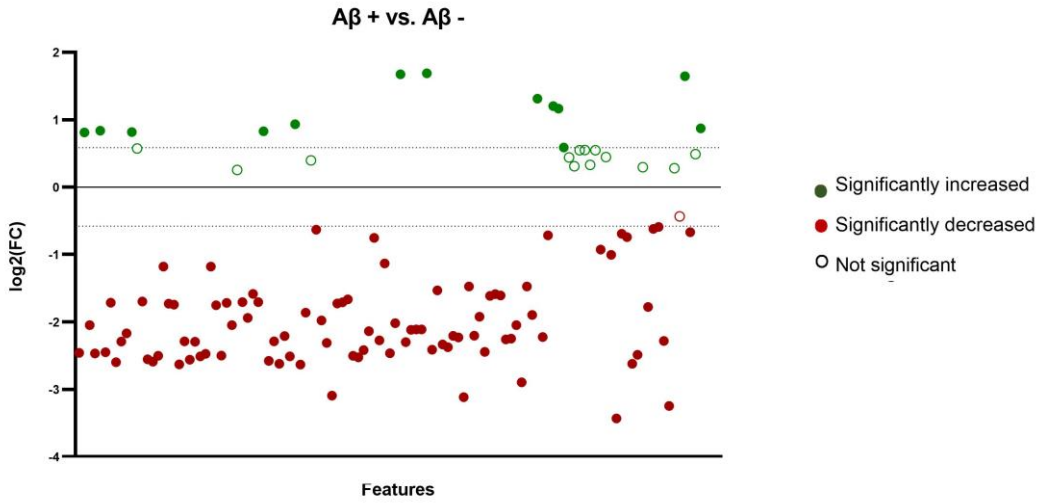

Figure 22. The graph demonstrates the log2 fold change (Aβ + vs. Aβ -) of the means for all the 119 features. The green color shows the features significantly increased with a fold change higher than 1.5, and the red color shows the features significantly down with a fold change lower than 0.67. The dotted lines represent log2(1.5) and log2(0.67), respectively.

Metabolite identification was performed on the online databases and the in-house library, using the 119 statistically different features, as explained in Chapter 3. The annotation level is the less confident classification, and the putative ID represents the highest confidence classification. The number of matches in all the levels and sub-levels assigned for the annotation and identification of the metabolites is represented in Figure 23.

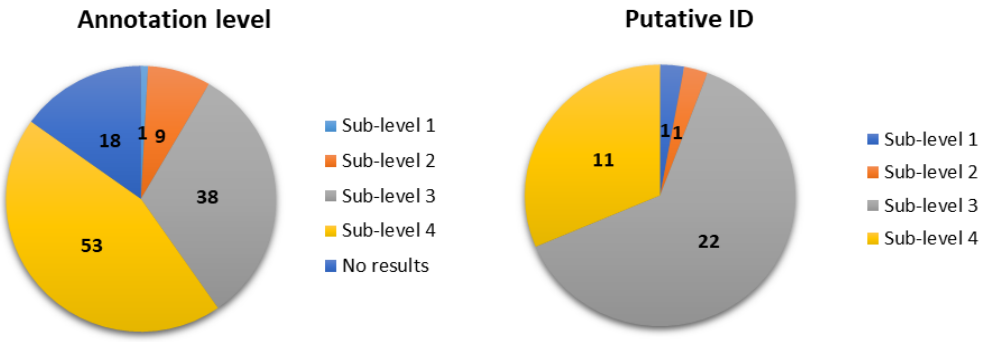

Figure 23. Metabolite identification of 119 statistically different features using the HMDB, METLIN, and the in-house library. The pie chart on the left represents the number of features matched in the Annotation level and the pie chart on the right represents the number of features matched in the Putative identification level. Sub-level 1 confirmed in the three databases; Sub-level 2 confirmed in only two databases; Sub-level 3 confirmed in just one database; Sub-level 4 confirmed in one database but with more than one hit per feature.

As observed in Figure 23, at the metabolite annotation level, 18 features got no identification, 53 features got more than one match, 38 had a match in just one database and 9 features had a match in both online databases. The feature 298.0980_7.84 was the only one that got the same match in all the databases used and matched with the 5'-Methylthioadenosine metabolite. Detailed results of the matches for each feature considering only the experimental mass/charge of the precursor are provided in the supplementary tables A.2.1 and A.2.2 of the Appendix A.2. For the putative ID the fragmentation data of each feature was used; however, 39 features did not have the information about the MS/MS spectrum and for that, they were excluded. Of the 80 features that remained, 24 got no results and 21 did not present the maximum score of FIT. Eleven features had more than one match with a maximum score and 22 were confirmed in just one database with a maximum score. The feature 268.1858_22.37 got the same match in both online databases, being identified as Metoprolol. As well as in the metabolite annotation, the feature 298.0980_7.84 was the only one that got the same match in the three databases, 5'-methylthioadenosine. In Figure 24, it is shown the comparison between the acquired spectrum and the in-house library for the feature 298.0980_7.84 identified as 5'-methylthioadenosine. The table with the features that got a putative ID is described in the supplementary table A.2.3 of Appendix A.2.

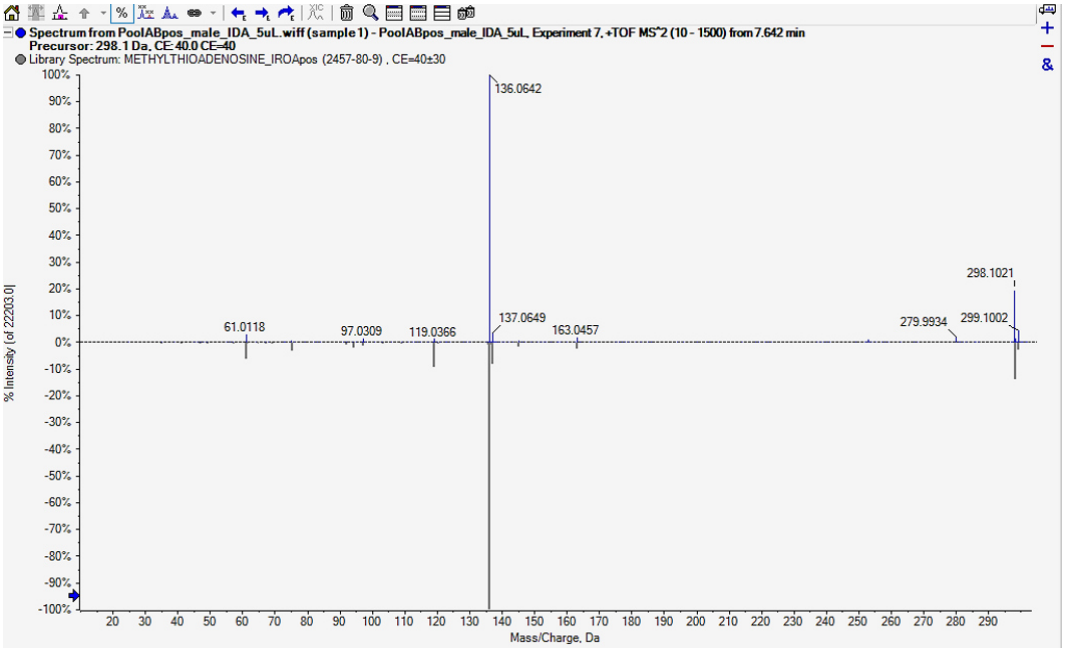

Figure 24. Comparison of the experimental spectrum of the feature 298.0980_7.84 and the spectrum from the in-house library of the matched metabolite. The experimental MS/MS spectrum (blue stripes) is shown on top, and the library MS/MS spectrum (grey stripes) is shown underneath.

There were four features (142.1248_11.73, 265.1867_17.52, 384.1080_6.79, and 419.2367_24.43) that got different matches with maximum scores in the HMDB and METLIN databases; these different matches may have resulted from a lack of the information from the MS/MS spectrum of some metabolites in the METLIN database. For example, the 2-(4-allyl-2,6-dimethoxyphenoxy)-1-(3,4,5-trimethoxyphenyl)-1-propanol was matched with the feature 419.2367_24.43 on the HMDB database but in the METLIN this feature was matched with the tripeptide Threonyl-tryptophyl-leucyine, and if we search for the metabolite 2-(4-allyl-2,6-dimethoxyphenoxy)-1-(3,4,5-trimethoxyphenyl)-1-propanol on METLIN it is indicated that the database has no information about the MS/MS spectrum of this compound. Nevertheless, these mismatches are identified in supplementary table A.2.3 of Appendix A.2, where all the metabolites identified within the putative ID level are described, as well as the associated pathways.

Most of the features identified with a putative ID were significantly decreased in the Aβ positive group, as it is shown in Figure 25, this tendency was already noted when all the 119 features were considered. The decreased compounds are more correlated with the Aβ negative group and responsible for the separation between the two main clusters observed in the heatmap in Figure 21.

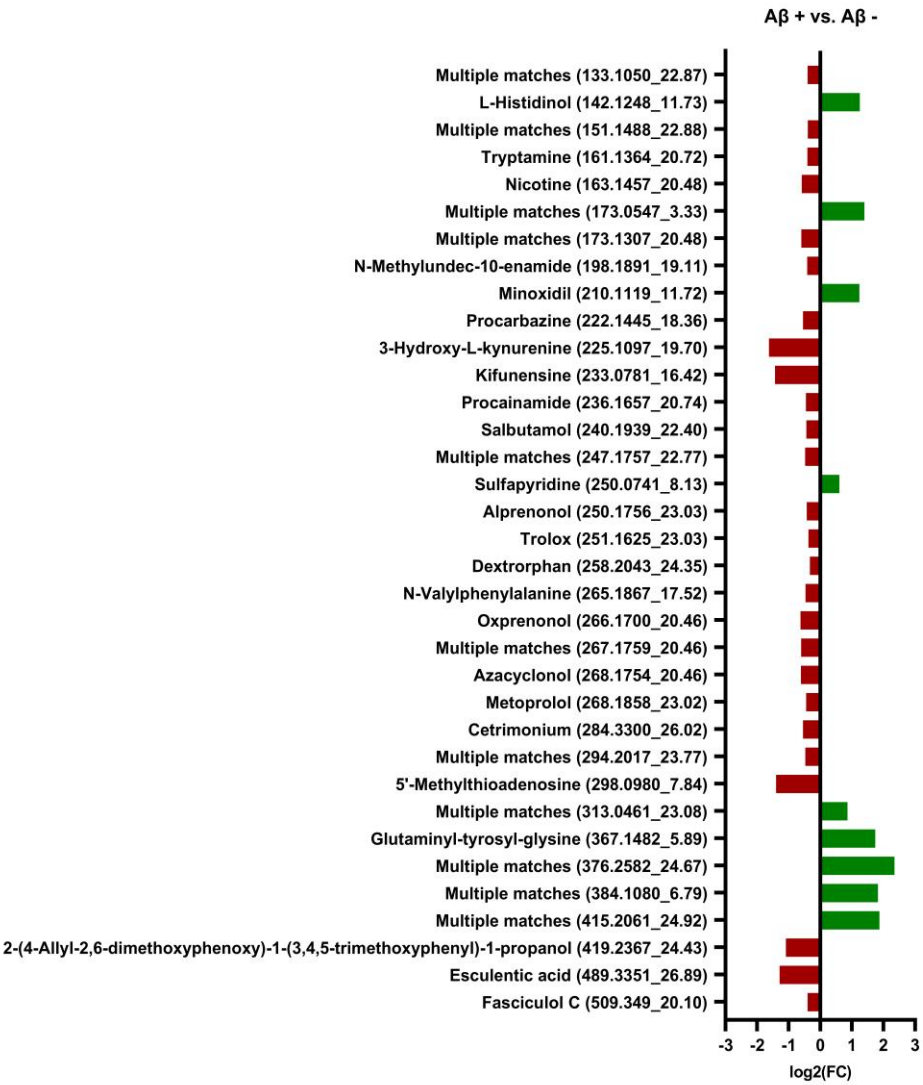

Figure 25. The graph shows the log2 fold change (Aβ + vs. Aβ -) of the means for the 35 features identified with a putative ID in the CSF of Aβ negative and Aβ positive groups. The green color shows the features significantly increased, and the red color shows the features significantly decreased.

The pathway analysis of the features identified with a putative ID was conducted using MetaboAnalyst to identify the most relevant pathways. The metabolites' human metabolome database (HMDB) IDs were up-loaded and the Hypergeometric Test and the relative-betweenness Centrality were used in the analysis. Within the 35 features identified with a Putative ID the ones from an endogenous source were used for the pathway analysis, and this includes the metabolites tryptamine (161.1364_20.72), 3-hydroxy-L-kynurenine (225.1097_19.70), ornithine (133.1050_22.87), glycerol-3-phosphate (173.0547_3.33), 5'-methylthioadenosine (298.0980_7.84), capric acid (173.1307_20.48), and L-asparagine (133.1050_22.87). The pathway analysis results are shown in Figure 26 and the information is described in Table A.2.4 of Appendix A.2.

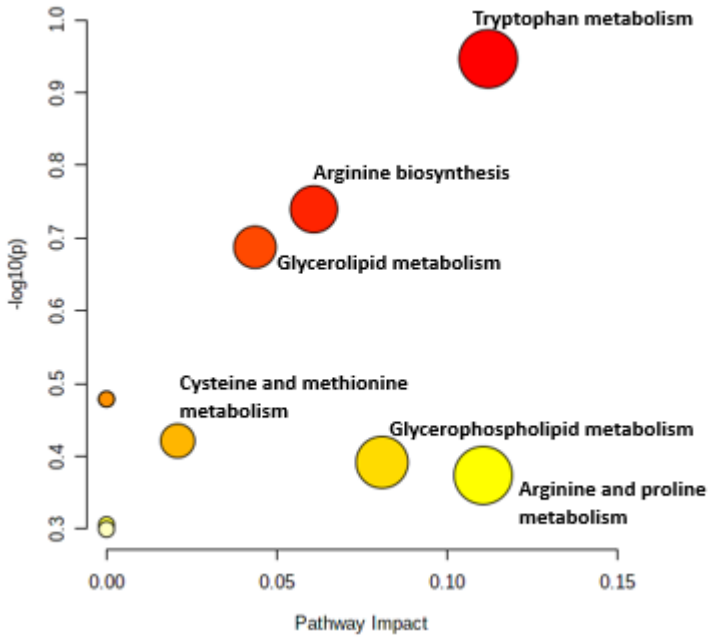

Figure 26. Pathway analysis results for the endogenous features identified with a putative ID in the CSF of Aβ negative and Aβ positive groups. The x-axis represents the pathway impact, and the y-axis is the log of the *p*-value obtained from the pathway enrichment analysis. The node color and size are based on its *p*-value and pathway impact values, respectively. The most significantly changed pathways are characterized by a high log (*p*) and a high impact value (top right region).

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 color and size are based on its *p*-value and pathway impact values, respectively. The most significantly changed pathways are characterized by a high log (*p*) and a high impact value (top right region) as it is in the case of the tryptophan metabolism, with a pathway impact of 0.1121 and a *p*-value of 0.11308. The metabolites tryptamine (161.1364_20.72) and 3-hydroxy-L-kynurenine (225.1097_19.70) were the two identified metabolites enriched in this pathway. In the arginine biosynthesis and the arginine and proline metabolism, the correspondent feature was the 133.1050_22.87, this feature was identified with multiple matches, including the metabolite ornithine enriched in these two metabolic pathways. The metabolite glycerol-3-phosphate was one of the two possible identifications for feature 173.0547_3.33 and was enriched in the glycerolipid metabolism and the glycerophospholipid metabolism. The cysteine and methionine metabolism was the pathway with a lower impact enriched with the metabolite 5'-methylthioadenosine (298.0980_7.84).

Tryptamine and 3-hydroxy-L-kyurenine (3-HK) were both increased in the CSF of the Aβ negative group (Figure 27). These two metabolites are from the tryptophan metabolism, a metabolomic pathway altered in patients with neurodegenerative disorders, including

Alzheimer's disease[51,60,61]. The 3-HK is a downstream product of tryptophan metabolism, synthesized from kynurenine by kynurenine 3-monooxygenase. 3-HK is a neurotoxic compound that can generate toxic free radicals, causing neurodegeneration and apoptosis of neurons[76]. In CSF, decreased concentrations of 3-HK in patients with AD were already reported by Toghi *et al*[77]. In plasma samples of AD patients, Gulaj *et al* found no significant differences between AD patients and control patients[76], while van der Velpen *et al* described a decrease in tryptophan pathway intermediates in AD patients, including 3-HK[61]. Tryptamine is an important neuromodulator of serotonin, and results in decarboxylation of tryptophan by the enzyme tryptophan decarboxylase. Serotonin is predominantly present in the nervous system and gut, and it is an important neuromodulator in the hippocampus and others brain structures critical for memory[78]. There are no reports of the identification of tryptamine in the CSF of patients with AD; however, the metabolite tryptophan was found to decrease in patients with AD[54,61], which points to a potential positive correlation between tryptamine and tryptophan.

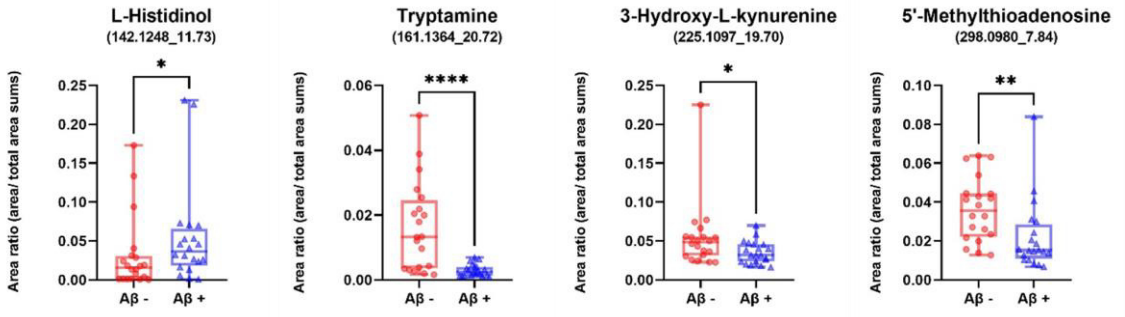

Figure 27. Box plots representing four metabolites identified with a putative ID from an endogenous source in the CSF of Aβ negative and Aβ positive groups. Significance (Mann-Whitney two-tailed U test) *p<0.05, **p<0.01, ****p<0.0001 between the Aβ negative (red) group and Aβ positive (blue) group.

As observed in Figure 27, the 5'-methylthioadenosine (MTA) was increased in the CSF samples of Aβ negative patients, this metabolite is produced from S-adenosylmethionine in the cysteine and methionine metabolism. Its increase was already reported by Ibáñez *et al* in patients with MCI when compared with MCI due to AD, AD, and control patients[60]. MTA plays relevant regulatory functions in the cells as gene expression, proliferation, differentiation, apoptosis, and a role as a metabolic intermediate in the methionine cycle[79]. The dysfunction in the methionine cycle has been associated with AD, and an alteration of the SAM/SAH is a hallmark of memory loss and cognitive decline[55].

L-Histidinol was also identified in the CSF of the study participants and was increased in Aβ positive group (Figure 27). L-Histidinol is the second to the last step in histidine biosynthesis and results from the transamination of the histidinol phosphate. Its presence was

not reported in AD patients, but the metabolite histidine was identified and analyzed in the CSF of MCI and AD patients. The histidine residues are present in Aβ peptides and are suggested to influence the toxicity of this peptide[80]. The histidine was found to decrease in patients with MCI due to AD in two different researches that compare AD patients with MCI patients that remained stable (MCI-S) for a period of two years and patients with MCI that progressed to AD (MCI-AD) in the same two-year period[57,60]. In our results an increase of L-Histidinol was noted in the Aβ positive group, the group that corresponds to patients diagnosed with MCI due to AD. In plasma and CSF samples of patients with AD, an increase of histidine was detected[51,57].

Considering only the identified features from an endogenous source, an unsupervised analysis, PCA, and a supervised analysis, PLS-DA, were performed using the Pareto scaling method, Figures 28 and 29. These features include multiple matches (133.1050_22.87), L-histidinol (142.1248_11.73), tryptamine (161.1354_20.72), multiple matches (173.0547_3.33), multiple matches (173.1307_20.48), 3-hydroxy-L-kynurenine (225.1097_19.70), and 5'-methylthioadenosine (298.0980_7.84). In the PCA scores plot of Figure 28, an overlap between the two groups is observed, with a 41.2% explained variance in PC1 and 24.2% in PC2, and it is also observed a tendency to gather the samples from the same group together. In the PLS-DA, Figure 29, the separation observed was better, with the samples from the Aβ negative group being more dispersed than the Aβ positive group.

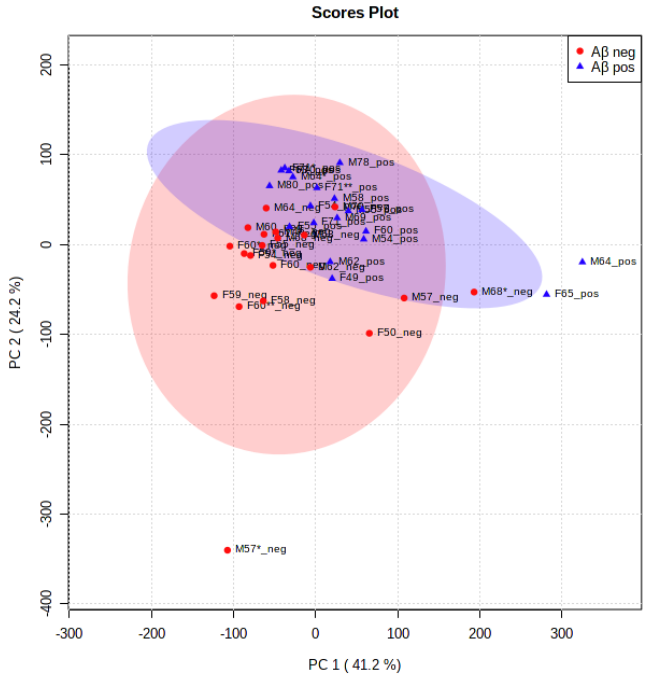

Figure 28. Principal component analysis (PCA) scores for Aβ negative (red) and Aβ positive (blue) groups using the 7 features from an endogenous source and the Pareto scaling performed using MetaboAnalyst.

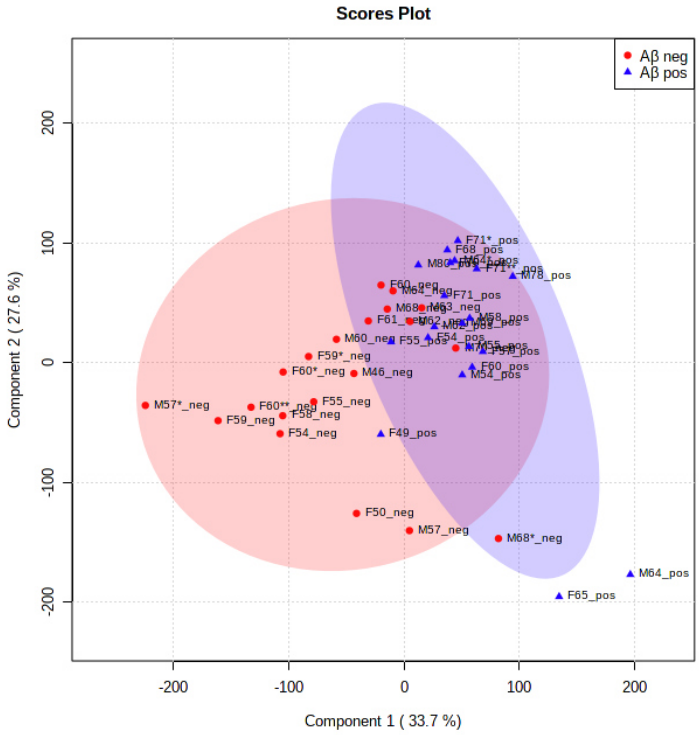

Figure 29. Partial least-squares-discriminant analysis (PLS-DA) scores for Aβ negative (red) and Aβ positive (blue) groups using the 7 features from an endogenous source and the Pareto scaling performed using MetaboAnalyst.

For the 7 features identified with a putative ID from an endogenous source, a Receiver-operating characteristic (ROC) analysis was performed to characterize the predictive value of these metabolites using MetaboAnalyst. In a ROC diagram plot, the x-axis indicates the false positive rate (1-specificity), and the y-axis indicates the true positive rate (sensitivity), generating the area under the ROC curve (AUC). The AUC is a measure of the accuracy of a diagnostic test a value of 1 indicates that the classifier is able to perfectly distinguish between the two groups, a value superior to 0.9 indicates excellent discrimination, a value superior to 0.8 is a good discrimination, a value superior to 0.7 is a fair test, a value between 0.51 and 0.69 is a poor test, and an AUC value of 0.5 does not show discrimination between the two groups[81].

Figure 30 shows the ROC curves for each endogenous metabolite identified, and it is observed that the AUC values range from 0.695 to 0.906, with sensitivities between 70 to 100% and specificities between 60 to 80%. Among all the endogenous metabolites, the metabolite 3-hydroxi-L-kynurenine had the lower AUC value, 0.695, and the multiple matches (ornithine or L-asparagine or cinnamaldehyde) had the highest AUC value, 0.906, which means that this metabolite had the best discriminatory power to distinguish between the Aβ negative

and Aβ positive groups. If this metabolite is considered for diagnosis, 90% of the time a positive result will be identified correctly and 80% of the time a negative result will be identified correctly. The pathways related to ornithine and L-asparagine, as the arginine and proline metabolism and the aminoacyl-tRNA biosynthesis, respectively, had already been found affected in the CSF of AD patients[51,53,54,57]. In case of the cinnamaldehyde, there is no report of its detection in AD patients.

The metabolite tryptamine had an AUC value of 0.853 with a sensitivity of 100% and a specificity of 70%, which means that compared to multiple matches (ornithine or L-asparagine or cinnamaldehyde) the proportion of positive cases that are identified correctly is higher.

A combination of more than one endogenous metabolite was developed using logistic regression analysis to improve the prediction. The ROC curves created are based on the cross-validation performance of multivariate algorithms using the Linear SVM (support vector machine) classifier. Figure 31 represents the ROC curves for all models created and the metabolites ranked by their selection importance in the six-feature panel. In all models, the AUC values were superior to 0.800, and the model that achieved the highest AUC was the signature with 6 metabolites. It is noted that the AUC values increased with the number of features used in the signatures. The feature ranking method shows which metabolites had contributed to the model, in this case, the 3-hydroxy-L-kynurenine was the one that had higher importance in the six-feature panel, and the multiple matches (capric acid or 1,2,3,4-tetrahydro-beta-carboline) the lower importance.

Comparing the AUC that each metabolite individually got and the AUC with the combination of different metabolites, the metabolite identified as multiple matches (ornithine or L-asparagine or cinnamaldehyde) was the best discriminator (AUC of 0.906) of the two groups.

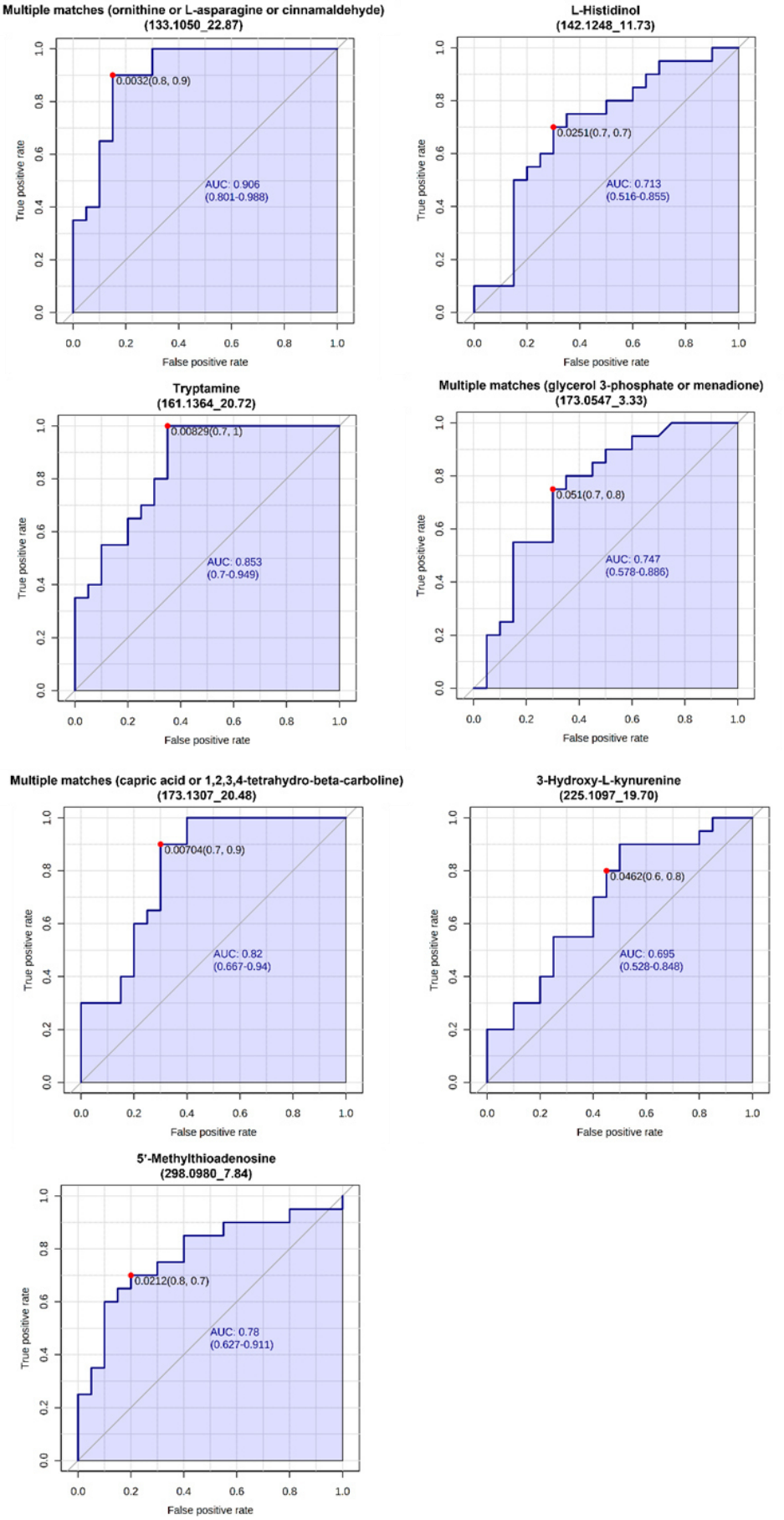

Figure 30. Receiver operator characteristic (ROC) curve analysis of the 7 features from an endogenous source between Aβ negative and Aβ positive groups using MetaboAnalyst.

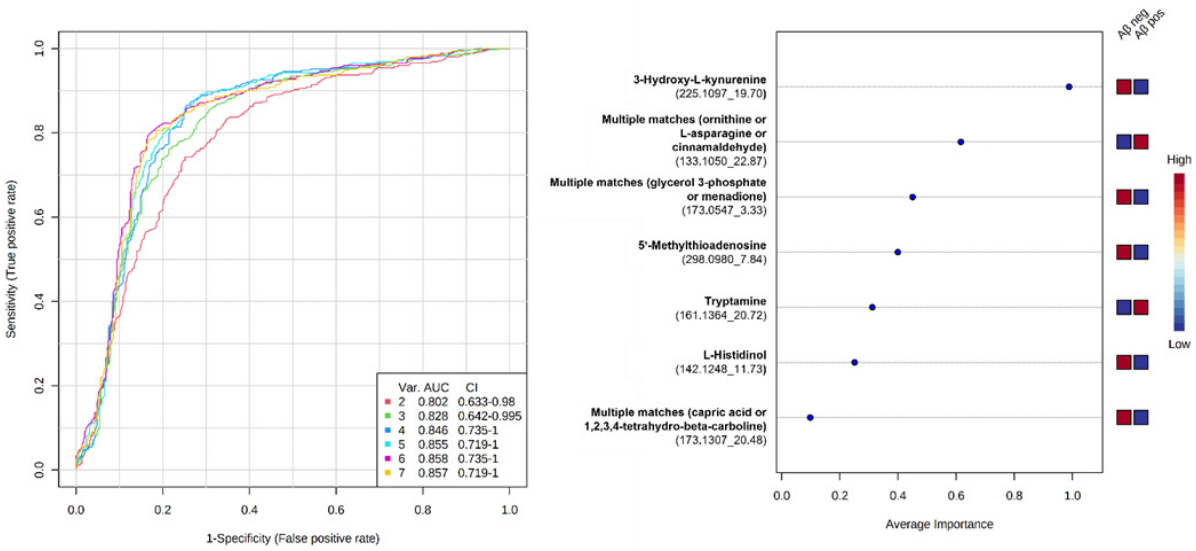

Figure 31. ROC curve for the predictive model of the 7 features from an endogenous source between Aβ negative and Aβ positive groups using MetaboAnalyst. On the right side, a combination metabolites model was calculated from the logistic regression analysis; on the left side, features were ranked by their contributions to the model's classification accuracy that uses 6 variables.

For each of the 35 features identified with a putative ID, the fragmentation spectra were acquired, and some fragments of the 35 precursors were chosen to confirm the quantification of the respective precursor ion in the biological samples, as exemplified in Figure 32.

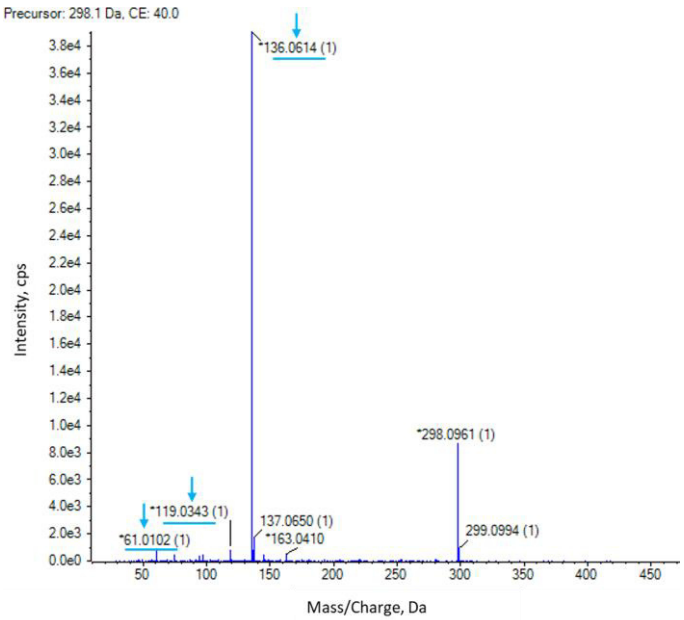

Figure 32. Fragmentation spectrum of the precursor ion with m/z of 298.0980 and RT of 7.84, the feature identified as 5'-methylthioadenosine. The blue arrows represent the fragments selected for quantification.

The peak integration of the precursor ions and their fragments was performed using the Multiquant™ software. Of the 35 features considered, the confirmation of the quantification of the precursor was only possible in 12 features**,** these features are represented in Table VII with the respective fold-change values and *p*-values. For features 298.0980_7.84 and 250.0741_8.13, the peak area of the ions in the two conditions (Aβ positive and Aβ negative) were graphically represented in Figures 33 and 34, respectively.

By analyzing the Table VII, it can be concluded that in all the features the tendencies were the same when comparing precursors and fragments. In the sulfapyridine, glutaminyl-tyrosyl-glicine, multiple matches (384.1080_6.79), and multiple matches (415.2061_24.92), there were an improvement of the *p*-values, since it values decrease in the fragments. In the case of the 5'-methylthioadenosine there was a greater decrease in the fold change and a decrease in the *p*-values in the fragments.

Table VII. The 12 features where it was possible to confirm the quantification of the precursors performed using the Multiquant™ software. The green arrows represent an increase in the Aβ positive group, and the red arrows represent a decrease in the Aβ positive group when compared with the Aβ negative group.

Compound	Precursor		Fragment 1		Fragment 2		Fragment 3	
	FC(Aβ+/Aβ-)	p-value	FC(Aβ+/Aβ-)	p-value	FC(Aβ+/Aβ-)	p-value	FC(Aβ+/Aβ-)	p-value
Multiple matches (173.1307_20.48)	↓ 0.31	<0.001	↓ 0.40	0.002	↓ 0.32	0.002	↓ 0.42	0.02
N-Methylundec-10-enamide (198.1891_19.11)	↓ 0.19	<0.001	↓ 0.79	0.102	↓ 0.29	<0.001	↓ 0.23	<0.001
Salbutamol (240.1939_22.40)	↓ 0.21	<0.001	↓ 0.25	<0.001	↓ 0.38	<0.001	↓ 0.26	0.001
Sulfapyridine (250.0741_8.13)	↑ 3.15	0.009	↑ 3.78	0.002	↑ 3.49	0.004	↑ 3.51	0.003
Dextrorphan (258.2043_24.35)	↓ 0.12	<0.001	↓ 0.09	<0.001	↓ 0.10	<0.001	↓ 0.13	<0.001
Oxprenolol (266.1700_20.46)	↓ 0.33	<0.001	↓ 0.29	<0.001	↓ 0.36	<0.001	↓ 0.06	<0.001
Metoprolol (268.1858_23.02)	↓ 0.21	<0.001	↓ 0.16	<0.001	↓ 0.32	<0.001	-	-
5'-Methylthioadenosine (298.0980_7.84)	↓ 0.61	0.002	↓ 0.50	<0.001	↓ 0.58	0.01	↓ 0.62	<0.001
Glutaminyl-tyrosyl-glycine (367.1482_5.89)	↑ 1.49	0.017	↑ 1.36	0.009	↑ 1.39	0.006	↑ 1.35	0.005
Multiple matches (384.1080_6.79)	↑ 1.46	0.038	↑ 1.24	0.023	↑ 1.27	0.017	↑ 1.32	0.005
Multiple matches (415.2061_24.92)	↑ 1.45	0.012	↑ 1.46	0.002	↑ 1.45	0.001	↑ 1.41	0.003
Fasciculol C (509.3490_20.10)	↓ 0.18	<0.001	↓ 0.15	<0.001	↓ 0.17	<0.001	↓ 0.15	<0.001

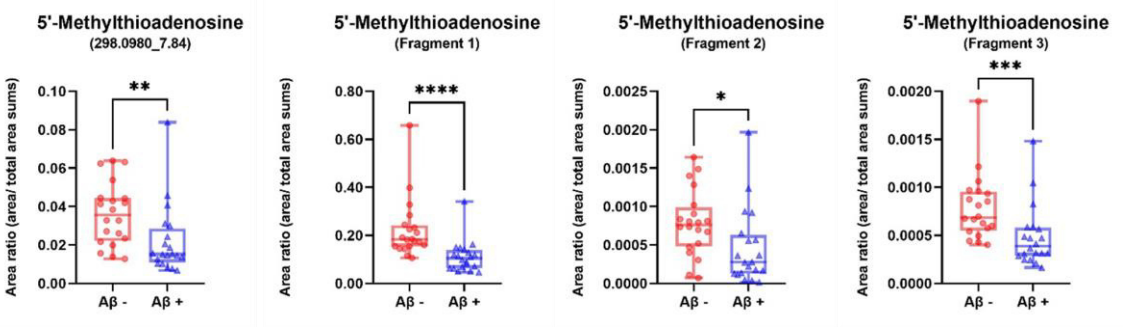

Figure 33. Area ratio for the feature with m/z value of 298.0980 and RT of 7.84 minutes and its selected fragment ions in the two different groups. Significance (Mann-Whitney two-tailed U test) **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001 between the Aβ negative (red) group and Aβ positive (blue) group.

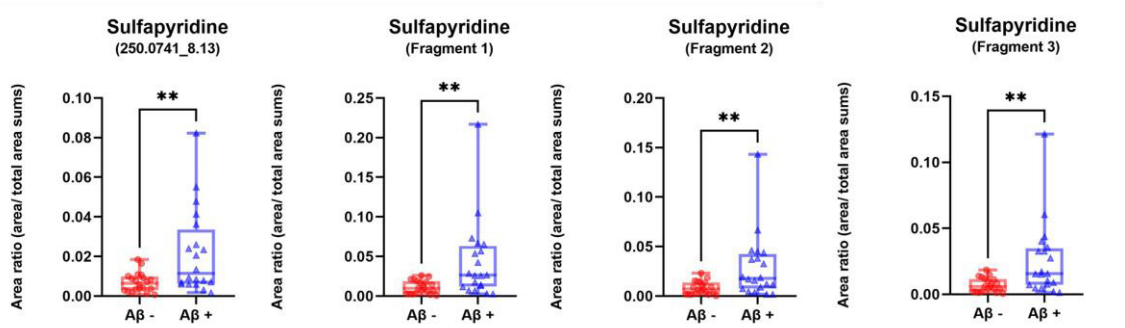

Figure 34. Area ratio for the feature with m/z value of 250.0741 and RT of 8.13 minutes and its selected fragment ions in the two different groups. Significance (Mann-Whitney two-tailed U test) ***p*<0.01, between the Aβ negative (red) group and Aβ positive (blue) group.

Until now, the analysis performed in the CSF samples was done considering two groups related to the β-amyloid, where the analysis revealed a tendency to separate the two groups if the gender of the study participants was considered. In order to understand better this difference between gender, a new PCA (Figure 35) and PLS-DA (Figure 36) analyses were performed using the 119 statistically different features and the Pareto scaling method considering four groups: Aβ negative female, Aβ negative male, Aβ positive female, and Aβ positive male.

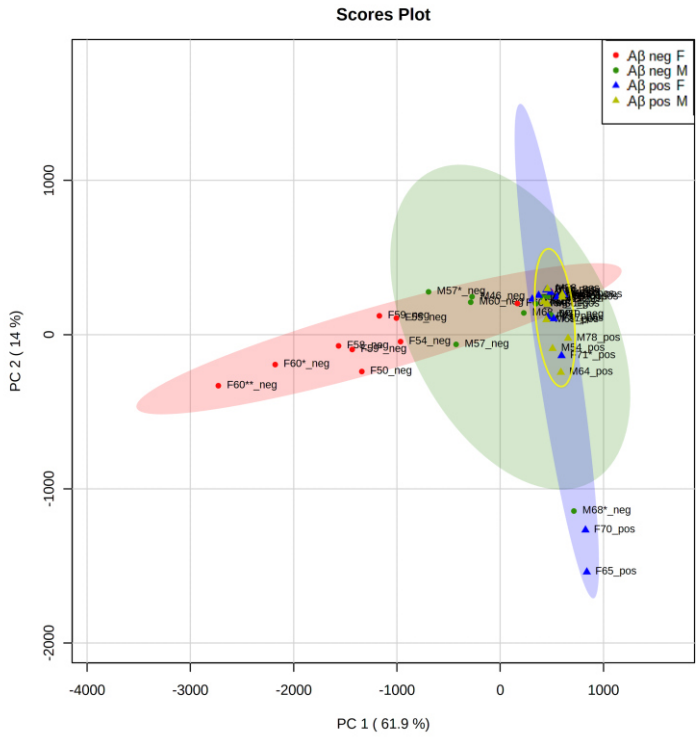

Figure 35. Principal component analysis (PCA) scores for the four conditions considering the Aβ condition and the gender of the study participants Aβ negative female (red), Aβ negative male (green), Aβ positive female (blue), and Aβ positive male (yellow), using the 119 features statistically different with a p<0.05 and a VIP>1.00 and performing Pareto scaling in MetaboAnalyst.

The PCA scores plot of Figure 35 shows a tendency to separate the Aβ negative samples from the Aβ positive ones, and in addition, there is a separation between the two genders in the Aβ negative group. In the case of Aβ positive condition, there is an overlap between the two genders, males and females. Samples F60_neg, F61_neg, M62_neg, M63_neg, M64_neg, M68_neg, M68*_neg, and M70_neg are the samples from Aβ negative group that display a different tendency from the ones in the same group, as already demonstrated in the previous analysis (Figures 16-20).

A better separation of the four groups was obtained in PLS-DA analysis, as observed in Figure 36. The samples F60_neg, F61_neg, M62_neg, M63_neg, M64_neg, M68_neg, M68*_neg, and M70_neg, are the reason why an overlap of the four groups was continuously observed. In both plots of the multivariate analysis, the two genders of the Aβ positive condition did not have any significant difference between them, this was then also confirmed by the analysis of the boxplots of the identified features with a putative ID (Figures 37 and 38). From the analysis of the four groups, it is possible to conclude that there is a difference between genders, particularly in the Aβ negative condition.

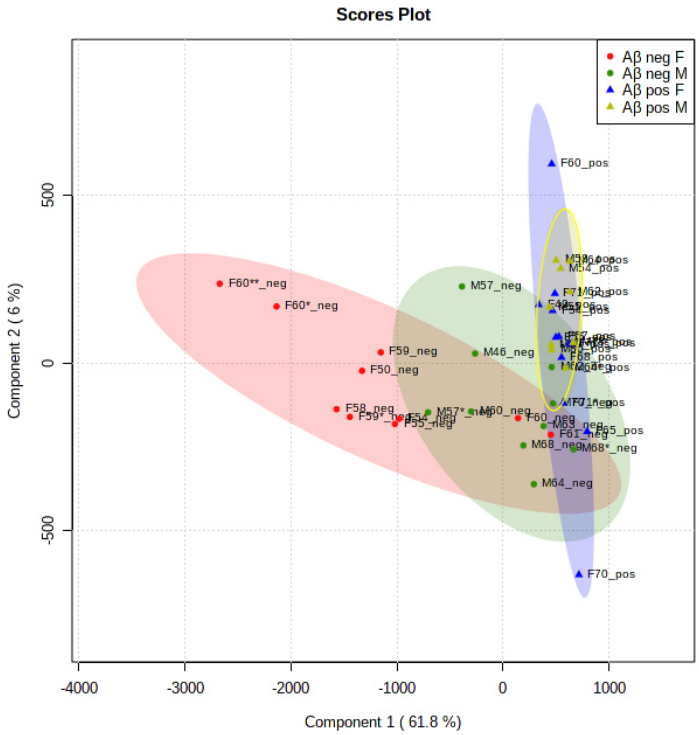

Figure 36. Partial least-squares-discriminant analysis (PLS-DA) scores for the four conditions considering the Aβ condition and the gender of the study participants Aβ negative female (red), Aβ negative male (green), Aβ positive female (blue), and Aβ positive male (yellow), using the 119 features statistically different with a p<0.05 and a VIP>1.00 and performing with Pareto scaling in MetaboAnalyst 5.0.

The quantification of the 35 metabolites identified with a putative ID was considered for the metabolic profile of the four groups. The Kruskal-Wallis test found 29 metabolites significantly different in at least one of the comparisons. Two different profiles were identified, one with a higher ratio in the samples of the Aβ negative male group (Figure 37), and a second one with a higher ratio in the samples of the Aβ negative female group (Figure 38). The first group had three metabolites while the second had eighteen metabolites. The metabolites identified increased in the male group of Aβ negative and had no correlation with an endogenous source or with sources related to any type of medication. The identified metabolites that were increased in the female group of Aβ negative include metabolites from an endogenous source, such as the tryptamine and the ornithine, and metabolites from an exogenous source, such as the salbutamol, alprenolol, and the procainamide.

In both metabolic profiles, Figure 37 and 38, as well as in the multivariate analysis, Figure 35 and 36, a significant statistical difference between the two genders of the Aβ positive group is never observed, concluding that there is no difference in this group between genders.

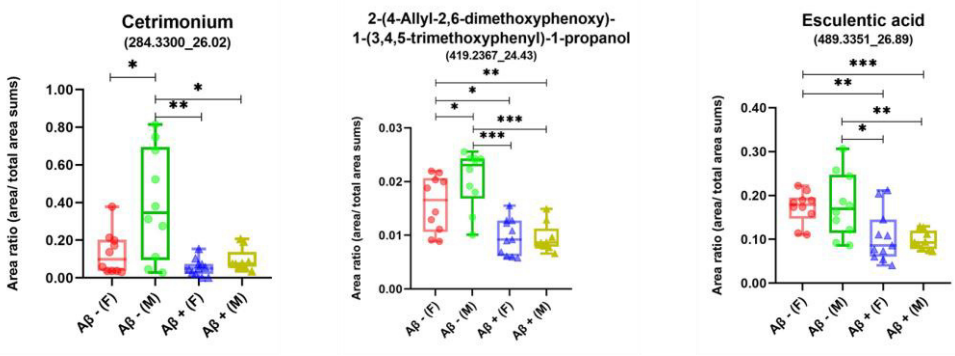

Figure 37. Box plots representing the metabolites identified with a putative ID in the CSF of the four groups, that had a higher area ratio in the Aβ negative male group. Significance (Mann-Whitney two-tailed U test) **p*<0.05, ***p*<0.01, *****p*<0.0001 between four groups Aβ negative female (red), Aβ negative male (green), Aβ positive female (blue), and Aβ positive male (yellow) groups.

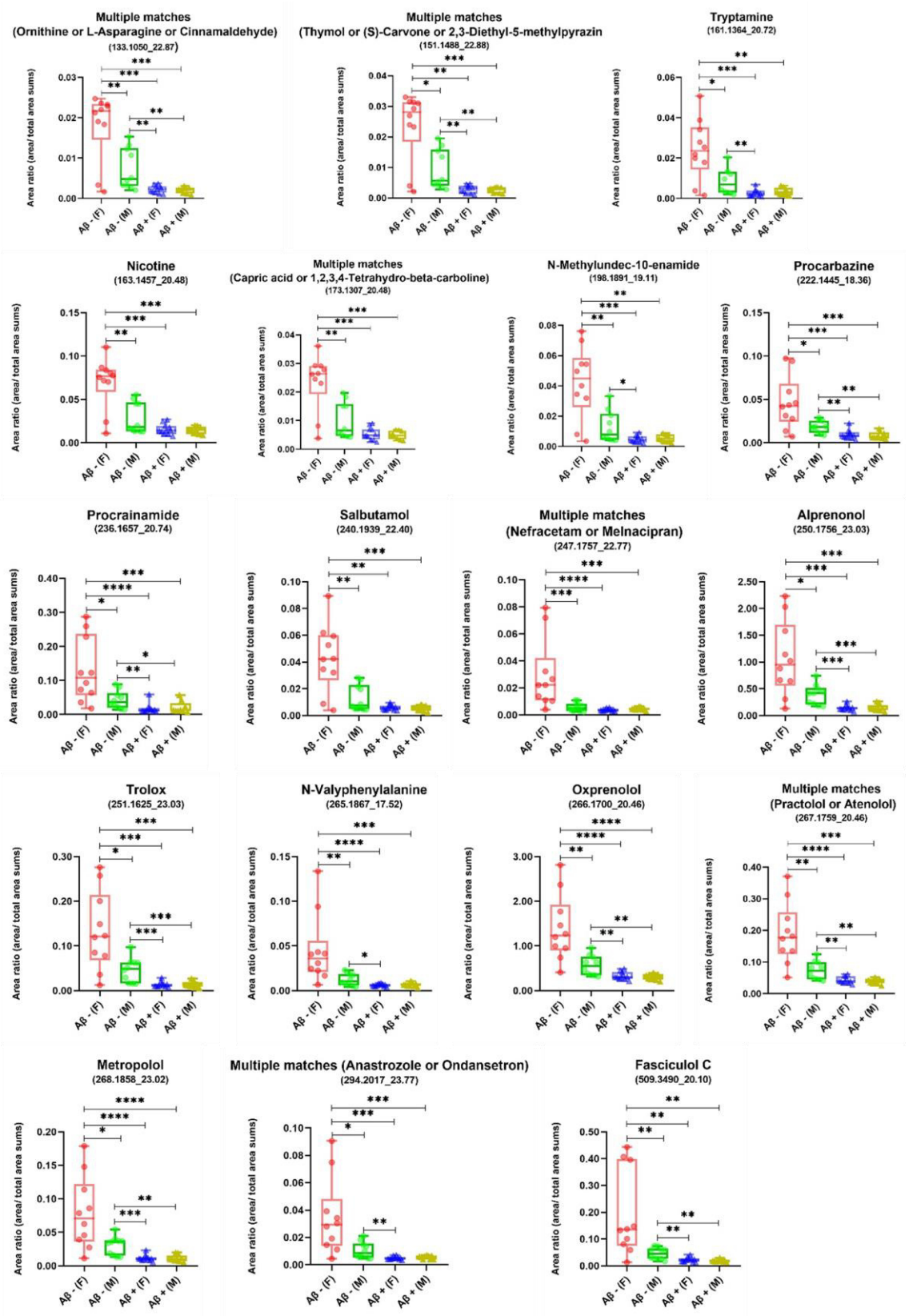

Figure 38. Box plots representing the metabolites identified with a putative ID from an endogenous and exogenous source in the CSF of the four groups that had a higher area ratio in the Aβ negative female group. Significance (Mann-Whitney two-tailed U test) **p*<0.05, ***p*<0.01, *****p*<0.0001 between four groups Aβ negative female (red), Aβ negative male (green), Aβ positive female (blue), and Aβ positive male (yellow) groups.

A drug-associated metabolite enrichment analysis was conducted to provide a possible understanding of the differences noticed in the different groups. All the features identified with a putative ID were submitted to MetaboAnalyst, and a total of 38 pathways were identified (Table A.2.5 of Appendix A.2). The top 25 drug-associated pathways are shown in Figure 39. The most significant pathways and with a higher enrichment ratio are related to the treatment of cardiovascular diseases. Most of the metabolites set in each pathway are increased in the Aβ negative female group and may explain why this group differentiates from the one with the same condition but with a different gender. Alprenolol, atenolol, metoprolol, and oxprenolol are related to the treatment of hypertension and varied heart conditions, while practolol and procainamide are related to the treatment of cardiac arrhythmias.

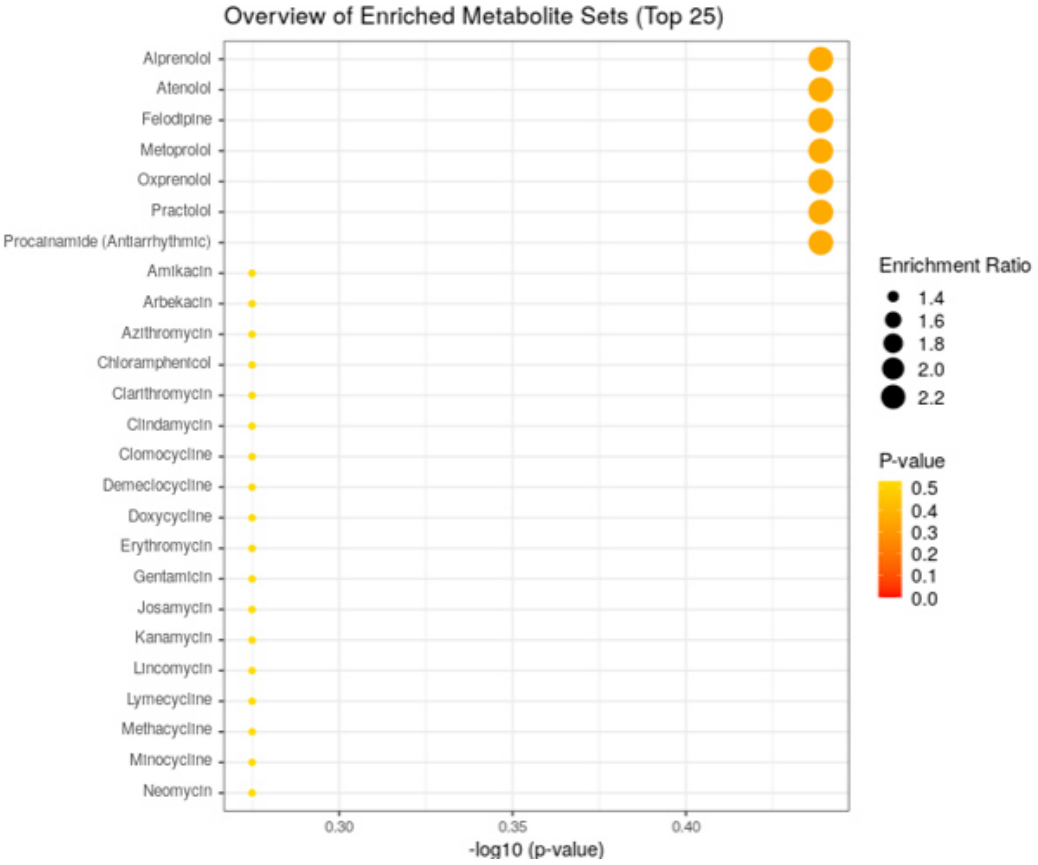

Figure 39. The top 25 enriched drug-associated pathways of 35 features were identified with a putative ID. The node color and size are based on its p-value and enrichment ratio, respectively. The most significantly changed pathways are characterized by a high log (*p*) and a high impact value.

CHAPTER 5 | Conclusion

5. Conclusion

The main focus of this work was to do an untargeted metabolomic analysis on CSF samples of β-amyloid positive and β-amyloid negative patients using LC-MS to differentiate the two groups to identify the more significant altered metabolites. With this aim, 40 samples were analyzed, 20 β-amyloid positives and 20 β-amyloid negatives. The ages ranged from 46 to 80 years, with no significant differences between the two groups.

From the LC-MS data, 4804 features were detected, and after filtering, 1445 features were used for data analysis. The univariate and multivariate analysis allowed to highlight 119 features that can be used to have a better separation of the two groups. In the analysis of the 119 features, eight samples of the Aβ negative group did not show the same correlation pattern that the other samples of the same group and were clustered with the Aβ positive group. The eight samples diagnosed as Aβ negative group but grouped with the Aβ positive group could be in a state of progression from Aβ negative to Aβ positive at the time of the diagnosis, leading to their clustering with the Aβ positive group considering the detected metabolites. It would be necessary to confirm with the clinicians if these patients progressed from an MCI condition to an MCI due to AD one. The hierarchical cluster analysis also observed the formation of a cluster for the female samples and a cluster for the male samples of the Aβ negative group suggesting that there were metabolites that differed between genders in this group.

In the metabolite identification, 35 features were identified with a putative identification, the level with higher confidence. The feature 298.0980_7.84, identified as the 5'-methylthioadenosine, was the only feature that had the same match in all the databases (HMDB, METLIN, and the in-house library), which increases the confidence in this identification. Of the 35 features, four features had different matches in the HMDB and METLIN databases. False-positive or false-negative identifications can be the result of incomplete databases, mainly in the information of MS/MS spectrum, or the result of poor spectra quality. To have a validated identification for these 35 features, it would be necessary to have a reference standard resuspended in CSF and to analyze it under identical conditions as our experimental analysis.

After the identification of the features, it was possible to do a more comprehensive analysis of the results, largely because of the interest features from an endogenous source identified in the samples. The pathway that was most significantly changed was the tryptophan metabolism. In this pathway, the two enriched metabolites were the tryptamine and the 3-hydroxy-L-kynurenine, both increased in the CSF of the Aβ negative group. The 5'-

methylthioadenosine metabolite was enriched in a pathway with a lower impact, the cysteine and methionine metabolism, and was also increased in the Aβ negative group. These pathways and related metabolites, mainly the metabolite 3-HK and the MTA, had already been reported to be altered in patients with AD, MCI due to AD, and MCI, identifying them as possible biomarkers. The metabolite identified to be decreased in the Aβ negative group, the L-histidinol, has not yet been reported. Even though some of these endogenous metabolites had already been reported in the literature, it is important to confirm if its increase is related to the disease or if it is related to an alteration caused by specific drug therapy.

In the ROC analysis performed for the endogenous metabolites, it was possible to conclude that the feature identified as multiple matches (ornithine or L-asparagine or cinnamaldehyde) had the best discriminatory power with an AUC value of 0.906, followed by the six-feature panel with an AUC value of 0.856. It would be important to verify and confirm the identification of the feature that got the best AUC value and to consider a higher number of samples to evaluate better the discriminatory power between the two groups. Lastly, it would be interesting to compare it with the established biomarkers of AD to evaluate if this metabolite adds value to the existing diagnostic tools.

The analysis considering four groups (Aβ negative female, Aβ negative male, Aβ positive female, and Aβ positive male) demonstrated that there were no differences between genders in the Aβ positive group. In the case of the Aβ negative group, the differences observed were the result of some exogenous metabolites related to hypertension therapy. These results require further confirmation by consulting the patient's clinical history.

In the future, the validation of identification will allow to clarify our results and a higher number of samples, which would help distinguish the two groups better. It would also be advantageous to have complete information about the drugs and the pre-existing conditions of the patients under study. In the untargeted metabolomic analysis, all types of metabolites (endogenous and exogenous) are detected, creating the need to have as much information as possible about the patient's clinical history to facilitate a later exclusion of the identified metabolites are not related to the disease.

In conclusion, the untargeted metabolomic analysis of the CSF samples of patients Aβ negative and Aβ positive allowed the detection of 119 statistically significant features and to identify 35 of them with a Putative ID. The identification of the increased metabolites 3-hydroxy-L-kynurenine and 5'-methylthioadenosine in the Aβ negative group, both already reported in the literature, suggest that these metabolites and their respective pathways could be associated with the Alzheimer's disease pathophysiology. The feature 133.1050_22.87, identified as multiple matches (ornithine or L-asparagine or cinnamaldehyde), revealed

excellent discrimination between the two groups. The pathways associated with the ornithine and the L-asparagine are already identified as being altered in the CSF of AD patients, which highpoint the potential of these metabolites for diagnostic tools.

BIBLIOGRAPHY

1 Alzheimer, A. Uber einen eigenartigen schweren Erkrankungsprozess der Hirninde. *Neurologisches Centralblatt* **25**, 1134 (1906).

2 World Health Organization, *Dementia*, accessed 21 September 2020, <https://www.who.int/news-room/fact-sheets/detail/dementia>

3 OECD. *Health at a Glance 2019*. (2019).

4 Europe, A. *Dementia in Europe Yearbook 2019: Estimating the prevalence of dementia in Europe*. (2019).

5 Mayeux, R. & Stern, Y. Epidemiology of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine* **2**, doi:10.1101/cshperspect.a006239 (2012).

6 Herrup, K. Reimagining Alzheimer's Disease—An Age-Based Hypothesis. *The Journal of Neuroscience* **30**, 16755-16762, doi:10.1523/jneurosci.4521-10.2010 (2010).

7 Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P.R., Rimmer, E. & Scazufca, M. Global prevalence of dementia: a Delphi consensus study. *Lancet (London, England)* **366**, 2112-2117, doi:10.1016/s0140-6736(05)67889-0 (2005).

8 Saunders, A. M., Strittmatter, W. J., Schmechel, D., St. George-Hyslop, P. H, Pericak-Vance, M. A., Joo, S. H., Rosi, B. L., Gusella, J. F., Crapper-MacLachlan, D. R., Alberts, M. J., Hulette, C., Crain, B., Goldgaber, D. & Roses, A. D. Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467-1467, doi:10.1212/wnl.43.8.1467 (1993).

9 Mahley, R. W., Weisgraber, K. H. & Huang, Y. Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proceedings of the National Academy of Sciences* **103**, 5644-5651, doi:10.1073/pnas.0600549103 (2006).

10 Näslund, J., Thyberg, J., Tjernberg, L. O., Wernstedt, C., Karlström, A. R., Bogdanovic, N., Gandy, S. E., Lannfelt, L. & Terenius, L., Nordstedt, C. Characterization of stable complexes involving apolipoprotein E and the amyloid beta peptide in Alzheimer's disease brain. *Neuron* **15**, 219-228, doi:10.1016/0896-6273(95)90079-9 (1995).

11 Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., Johnson-Wood, K., Lee, M., Seubert, P., Davis, A., Kholodenko, D., Motter, R., Sherrington, R., Perry, B., Yao, H., Strome, R., Lieberburg, I., Rommens, J., Kim, S., Schenk, D., Fraser, P., St George Hyslop, P. & Selkoe, D. J. Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. *Nature medicine* **3**, 67-72, doi:10.1038/nm0197-67 (1997).

12 Tcw, J. & Goate, A. M. Genetics of β-Amyloid Precursor Protein in Alzheimer's Disease. *Cold Spring Harbor Perspectives in Medicine* **7**, doi:10.1101/cshperspect.a024539 (2017).

13 McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S. & Phelps, C. H. The diagnosis of dementia 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: The Journal of the Alzheimer's Association* **7**, 263-269, doi:10.1016/j.jalz.2011.03.005 (2011).

14 Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J. F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., Gouw, A. A., Habert, M. O., Holtzman, D. M., Kivipelto, M., Lista, S., Molinuevo, J. L., O'Bryant, S. E., Rabinovici, G. D., Rowe, C., Salloway, S., Schneider, L. S., Sperling, R., Teichmann, M., Carrillo, M. C., Cummings, J. & Jack, C. R., Jr. Preclinical Alzheimer's disease: Definition,

natural history, and diagnostic criteria. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **12**, 292-323, doi:10.1016/j.jalz.2016.02.002 (2016).

15 Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B. & Phelps, C. H. 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: The Journal of the Alzheimer's Association* **7**, 270-279, doi:10.1016/j.jalz.2011.03.008 (2011).

16 Porth, C. & Matfin, G. *Pathophysiology: Concepts of Altered Health States*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2009).

17 Korczyn, A. D. The amyloid cascade hypothesis. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **4**, 176-178, doi:10.1016/j.jalz.2007.11.008 (2008).

18 Roberts, G. W., Gentleman, S. M., Lynch, A., Murray, L., Landon, M. & Graham, D. I. Beta amyloid protein deposition in the brain after severe head injury: implications for the pathogenesis of Alzheimer's disease. *Journal of neurology, Neurosurgery, and Psychiatry* **57**, 419-425, doi:10.1136/jnnp.57.4.419 (1994).

19 Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica* **82**, 239-259, doi:10.1007/BF00308809 (1991).

20 Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G. F., Estrada, S., Ausén, B., Debnath, M. L., Barletta, J., Price, J. C., Sandell, J., Lopresti, B. J., Wall, A., Koivisto, P., Antoni, G., Mathis, C. A. & Långström, B. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Annals of Neurology* **55**, 306-319, doi:10.1002/ana.20009 (2004).

21 Wang, H., Megill, A., He, K., Kirkwood, A. & Lee, H.-K. Consequences of Inhibiting Amyloid Precursor Protein Processing Enzymes on Synaptic Function and Plasticity. *Neural Plasticity* **2012**, 272374, doi:10.1155/2012/272374 (2012).

22 Verwilst, P., Kim, H. S., Kim, S., Kang, C. & Kim, J. S. Shedding light on tau protein aggregation: the progress in developing highly selective fluorophores. *Chemical Society Reviews* **47**, 2249-2265, doi:10.1039/C7CS00706J (2018).

23 Gong, C. X., Lidsky, T., Wegiel, J., Zuck, L., Grundke-Iqbal, I. & Iqbal, K. Phosphorylation of microtubule-associated protein tau is regulated by protein phosphatase 2A in mammalian brain. Implications for neurofibrillary degeneration in Alzheimer's disease. *Journal of Biological Chemistry* **275**, 5535-5544, doi:10.1074/jbc.275.8.5535 (2000).

24 Association, A. P. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. (American Psychiatric Publishing, 2013).

25 McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. & Stadlan, E. *M.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944, doi:10.1212/wnl.34.7.939 (1984).

26 World Health Organization, *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. (World Health Organization, 1993).

27 Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* **12**, 189-198, doi:10.1016/0022-3956(75)90026-6 (1975).

28 Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L. & Chertkow, H. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society* **53**, 695-699, doi:https://doi.org/10.1111/j.1532-5415.2005.53221.x (2005).

29 Thompson, P. M., Hayashi, K. M., De Zubicaray, G. I., Janke, A. L., Rose, S. E., Semple, J., Hong, M. S., Herman, D. H., Gravano, D., Doddrell, D. M. & Toga, A. W. Mapping

hippocampal and ventricular change in Alzheimer disease. *NeuroImage* **22**, 1754-1766, doi:10.1016/j.neuroimage.2004.03.040 (2004).

30 Su, L., Blamire, A. M., Watson, R., He, J., Hayes, L. & O'Brien, J. T. Whole-brain patterns of 1H-magnetic resonance spectroscopy imaging in Alzheimer's disease and dementia with Lewy bodies. *Translational Psychiatry* **6**, e877-e877, doi:10.1038/tp.2016.140 (2016).

31 Mosconi, L., Mistur, R., Switalski, R., Tsui, W. H., Glodzik, L., Li, Y., Pirraglia, E., De Santi, S., Reisberg, B., Wisniewski, T. & de Leon, M. J. FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease. *European Journal of Nuclear Medicine and Molecular Imaging* **36**, 811-822, doi:10.1007/s00259-008-1039-z (2009).

32 Johnson, K. A., Moran, E. K., Becker, J. A., Blacker, D., Fischman, A. J. & Albert, M. S. Single photon emission computed tomography perfusion differences in mild cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry* **78**, 240-247, doi:10.1136/jnnp.2006.096800 (2007).

33 Teunissen, C. & Willemse, E. Cerebrospinal Fluid Biomarkers for Alzheimer's Disease: Emergence of the Solution to an Important Unmet Need. *The Journal of the international Federation of Clinical Chemistry and Laboratory Medicine* **24**, 97-104 (2013).

34 Hall, S., Öhrfelt, A., Constantinescu, R., Andreasson, U., Surova, Y., Bostrom, F., Nilsson, C., Widner, H., Decraemer, H., Nägga, K., Minthon, L., Londos, E., Vanmechelen, E., Holmberg, B., Zetterberg, H., Blennow, K. & Hansson, O. Accuracy of a Panel of 5 Cerebrospinal Fluid Biomarkers in the Differential Diagnosis of Patients With Dementia and/or Parkinsonian Disorders. *Archives of Neurology* **69**, 1445-1452, doi:10.1001/archneurol.2012.1654 (2012).

35 Li, L., Luo, J., Chen, D., Tong, J. B., Zeng, L. P., Cao, Y. Q., Xiang, J., Luo, X. G., Shi, J. M., Wang, H. & Huang, J. F. BACE1 in the retina: a sensitive biomarker for monitoring early pathological changes in Alzheimer's disease. *Neural Regeneration Research* **11**, 447-453, doi:10.4103/1673-5374.179057 (2016).

36 Wright, B. L., Lai, J. T. & Sinclair, A. J. Cerebrospinal fluid and lumbar puncture: a practical review. *Journal of neurology* **259**, 1530-1545, doi:10.1007/s00415-012-6413-x (2012).

37 Johns Creek (GA): Ebix, I. *A.D.A.M. Medical Encyclopedia*, accessed 8 December 2020, <https://medlineplus.gov/ency/imagepages/9242.htm>

38 Nakamura, A., Kaneko, N., Villemagne, V. L., Kato, T., Doecke, J., Doré, V., Fowler, C., Li, Q. X., Martins, R., Rowe, C., Tomita, T., Matsuzaki, K., Ishii, K., Ishii, K., Arahata, Y., Iwamoto, S., Ito, K., Tanaka, K., Masters, C. L. & Yanagisawa, K. High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* **554**, 249-254, doi:10.1038/nature25456 (2018).

39 Hirtz, C., Vialaret, J., Nouadje, G., Schraen, S., Benlian, P., Mary, S., Philibert, P., Tiers, L., Bros, P., Delaby, C., Gabelle, A. & Lehmann, S. Development of new quantitative mass spectrometry and semi-automatic isofocusing methods for the determination of Apolipoprotein E typing. *Clinica Chimica Acta* **454**, 33-38, doi:https://doi.org/10.1016/j.cca.2015.12.020 (2016).

40 Oeckl, P. & Otto, M. A Review on MS-Based Blood Biomarkers for Alzheimer's Disease. *Neurology and therapy* **8**, 113-127, doi:10.1007/s40120-019-00165-4 (2019).

41 Fertalova, T. & Ducaiova, I. in *Redirecting Alzheimer Strategy - Tracing Memory Loss to Self Pathology* (IntechOpen, 2019).

42 Sharma, K. Cholinesterase inhibitors as Alzheimer's therapeutics (Review). *Molecular Medicine Reports* **20**, 1479-1487, doi:10.3892/mmr.2019.10374 (2019).

43 Kishi, T., Matsunaga, S., Oya, K., Nomura, I., Ikuta, T. & Iwata, N. Memantine for Alzheimer's Disease: An Updated Systematic Review and Meta-analysis. *Jounal of Alzheimer's Disease* **60**, 401-425, doi:10.3233/jad-170424 (2017).

44 Ackley, S. F., Zimmerman, S. C., Brenowitz, W. D., Tchetgen Tchetgen, E. J., Gold, A. L., Manly, J. J., Mayeda, E. R., Filshtein, T. J., Power, M. C., Elahi, F. M., Brickman, A. M. & Glymour, M. M. Effect of reductions in amyloid levels on cognitive change in randomized trials: instrumental variable meta-analysis. *BMJ* **372**, n156, doi:10.1136/bmj.n156 (2021).

45 Huang, L. K., Chao, S. P. & Hu, C. J. Clinical trials of new drugs for Alzheimer disease. *Journal of Biomedical Science* **27**, 18, doi:10.1186/s12929-019-0609-7 (2020).

46 Nettiksimmons, J., DeCarli, C., Landau, S., Beckett, L. & Alzheimer's Disease Neuroimaging, I. Biological heterogeneity in ADNI amnestic mild cognitive impairment. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **10**, 511-521.e511, doi:10.1016/j.jalz.2013.09.003 (2014).

47 Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews* **40**, 387-426, doi:10.1039/B906712B (2011).

48 Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C. & Scalbert, A. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research* **46**, D608-D617, doi:10.1093/nar/gkx1089 (2017).

49 Cortes, M., García-Cañaveras, J. C., Pareja, E. & Lahoz, A. in *Biomarkers in Disease: Methods, Discoveries and Applications* (eds Vinood B. Patel & Victor R. Preedy) 99-128 (Springer Netherlands, 2017).

50 Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **13**, 263-269, doi:10.1038/nrm3314 (2012).

51 Trushina, E., Dutta, T., Persson, X. M., Mielke, M. M. & Petersen, R. C. Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's disease using metabolomics. *PloS One* **8**, e63644, doi:10.1371/journal.pone.0063644 (2013).

52 Hajjar, I., Liu, C., Jones, D. P. & Uppal, K. Untargeted metabolomics reveal dysregulations in sugar, methionine, and tyrosine pathways in the prodromal state of AD. *Alzheimer's & Dementia (Amsterdam, Netherlands)* **12**, e12064, doi:10.1002/dad2.12064 (2020).

53 Nagata, Y., Hirayama, A., Ikeda, S., Shirahata, A., Shoji, F., Maruyama, M., Kayano, M., Bundo, M., Hattori, K., Yoshida, S., Goto, Y., Urakami, K., Soga, T., Ozaki, K. & Niida, S. Comparative analysis of cerebrospinal fluid metabolites in Alzheimer's disease and idiopathic normal pressure hydrocephalus in a Japanese cohort. *Biomarker Research* **6**, doi:10.1186/s40364-018-0119-x (2018).

54 Hao, L., Wang, J. X., Page, D., Asthana, S., Zetterberg, H., Carlsson, C., Okonkwo, O. C. & Li, L. J. Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease. *Scientific Reports* **8**, doi:10.1038/s41598-018-27031-x (2018).

55 Guiraud, S. P., Montoliu, I., Da Silva, L., Dayon, L., Galindo, A. N., Corthésy, J., Kussmann, M. & Martin, F. P. High-throughput and simultaneous quantitative analysis of homocysteine-methionine cycle metabolites and co-factors in blood plasma and cerebrospinal fluid by isotope dilution LC-MS/MS. *Analytical and Bioanalytical Chemistry* **409**, 295-305, doi:10.1007/s00216-016-0003-1 (2017).

56 Dayon, L., Guiraud, S. P., Corthesy, J., Da Silva, L., Migliavacca, E., Tautvydaite, D., Oikonomidi, A., Moullet, B., Henry, H., Metairon, S., Marquis, J., Descombes, P., Collino, S., Martin, F. P. J., Montoliu, I., Kussmann, M., Wojcik, J., Bowman, G. L. & Popp, J. One-carbon metabolism, cognitive impairment and CSF measures of Alzheimer pathology: homocysteine and beyond. *Alzheimers Research & Therapy* **9**, doi:10.1186/s13195-017-0270-x (2017).

57 Ibanez, C., Simo, C., Martin-Alvarez, P. J., Kivipelto, M., Winblad, B., Cedazo-Minguez, A. & Cifuentes, A. Toward a Predictive Model of Alzheimer's Disease Progression Using Capillary Electrophoresis-Mass Spectrometry Metabolomics. *Analytical Chemistry* **84**, 8532-8540, doi:10.1021/ac301243k (2012).

58 Motsinger-Reif, A. A., Zhu, H., Kling, M. A., Matson, W., Sharma, S., Fiehn, O., Reif, D. M., Appleby, D. H., Doraiswamy, P. M., Trojanowski, J. Q., Kaddurah-Daouk, R. & Arnold, S. E. Comparing metabolomic and pathologic biomarkers alone and in combination for discriminating Alzheimer's disease from normal cognitive aging. *Acta Neuropathologica Communications* **1**, 28, doi:10.1186/2051-5960-1-28 (2013).

59 Czech, C., Berndt, P., Busch, K., Schmitz, O., Wiemer, J., Most, V., Hampel, H., Kastler, J. & Senn, H. Metabolite profiling of Alzheimer's disease cerebrospinal fluid. *PloS One* **7**, e31501, doi:10.1371/journal.pone.0031501 (2012).

60 Ibáñez, C., Simó, C., Barupal, D. K., Fiehn, O., Kivipelto, M., Cedazo-Mínguez, A. & Cifuentes, A. A new metabolomic workflow for early detection of Alzheimer's disease. *Journal of Chromatography. A* **1302**, 65-71, doi:10.1016/j.chroma.2013.06.005 (2013).

61 van der Velpen, V., Teav, T., Gallart-Ayala, H., Mehl, F., Konz, I., Clark, C., Oikonomidi, A., Peyratout, G., Henry, H., Delorenzi, M., Ivanisevic, J. & Popp, J. Systemic and central nervous system metabolic alterations in Alzheimer's disease. *Alzheimer's Research & Therapy* **11**, 93, doi:10.1186/s13195-019-0551-7 (2019).

62 Takayama, T., Mizuno, H., Toyo'oka, T., Akatsu, H., Inoue, K. & Todoroki, K. Isotope Corrected Chiral and Achiral Nontargeted Metabolomics: An Approach for High Accuracy and Precision Metabolomics Based on Derivatization and Its Application to Cerebrospinal Fluid of Patients with Alzheimer's Disease. *Analytical Chemistry* **91**, 4396-4404, doi:10.1021/acs.analchem.8b04852 (2019).

63 Koal, T., Klavins, K., Seppi, D., Kemmler, G. & Humpel, C. Sphingomyelin SM(d18:1/18:0) is Significantly Enhanced in Cerebrospinal Fluid Samples Dichotomized by Pathological Amyloid-beta(42), Tau, and Phospho-Tau-181 Levels. *Journal of Alzheimers Disease* **44**, 1193-1201, doi:10.3233/jad-142319 (2015).

64 Bergau, N., Maul, S., Rujescu, D., Simm, A. & Navarrete Santos, A. Reduction of Glycolysis Intermediate Concentrations in the Cerebrospinal Fluid of Alzheimer's Disease Patients. *Frontiers in Neuroscience* **13**, 871, doi:10.3389/fnins.2019.00871 (2019).

65 de Hoffmann, E. & Stroobant, V. *Mass Spectrometry: Principles and Applications*. 3 edn, (Wiley, 2007).

66 Hocart, C. H. in *Comprehensive Natural Products II* (eds Hung-Wen Liu & Lew Mander) 327-388 (Elsevier, 2010).

67 Zhou, B., Xiao, J. F., Tuli, L. & Ressom, H. W. LC-MS-based metabolomics. *Molecular bioSystems* **8**, 470-481, doi:10.1039/C1MB05350G (2012).

68 Snyder, L. R., Kirkland, J. J. & Dolan, J. W. *Introduction to Modern Liquid Chromatography*. (Wiley, 2011).

69 Grebe, S. K. G. & Singh, R. J. Clinical peptide and protein quantification by mass spectrometry (MS). *TrAC Trends in Analytical Chemistry* **84**, 131-143, doi:https://doi.org/10.1016/j.trac.2016.01.026 (2016).

70 Chernushevich, I. V., Loboda, A. V. & Thomson, B. A. An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **36**, 849-865, doi:https://doi.org/10.1002/jms.207 (2001).

71 Zhu, X., Chen, Y. & Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MSAll Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography–Quadrupole Time-of-Flight Mass Spectrometry. *Analytical Chemistry* **86**, 1202-1209, doi:10.1021/ac403385y (2014).

72 Santa, C., Anjo, S., Mendes, V. M. & Manadas, B. *Neuroproteomics - LC-MS Quantitative Approaches*. (IntechOpen, 2015).

73 Mendes, V. M., Coelho, M. & Manadas, B. Untargeted Metabolomics Relative Quantification by SWATH Mass Spectrometry Applied to Cerebrospinal Fluid. *Methods in Molecular Biology (Clifton, N.J.)* **2044**, 321-336, doi:10.1007/978-1-4939-9706-0_20 (2019).

74 Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry* **27**, 1897-1905, doi:10.1007/s13361-016-1469-y (2016).

75 Hair, J. F., Ringle, C. M. & Sarstedt, M. Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long Range Planning* **46**, 1-12, doi:https://doi.org/10.1016/j.lrp.2013.01.001 (2013).

76 Gulaj, E., Pawlak, K., Bien, B. & Pawlak, D. Kynurenine and its metabolites in Alzheimer's disease patients. *Advances in medical sciences* **55**, 204-211, doi:10.2478/v10039-010-0023-6 (2010).

77 Tohgi, H., Abe, T., Takahashi, S., Kimura, M., Takahashi, J. & Kikuchi, T. Concentrations of serotonin and its related substances in the cerebrospinal fluid in patients with Alzheimer type dementia. *Neuroscience Letters* **141**, 9-12, doi:10.1016/0304-3940(92)90322-x (1992).

78 Roth, W., Zadeh, K., Vekariya, R., Ge, Y. & Mohamadzadeh, M. Tryptophan Metabolism and Gut-Brain Homeostasis. *International Journal of Molecular Sciences* **22**, 2973 (2021).

79 Avila, M. A., García-Trevijano, E. R., Lu, S. C., Corrales, F. J. & Mato, J. M. Methylthioadenosine. *The International Journal of Biochemistry & Cell Biology* **36**, 2125-2130, doi:10.1016/j.biocel.2003.11.016 (2004).

80 Arispe, N., Diaz, J. C. & Flora, M. Efficiency of Histidine-Associating Compounds for Blocking the Alzheimer's Aβ Channel Activity and Cytotoxicity. *Biophysical Journal* **95**, 4879-4889, doi:https://doi.org/10.1529/biophysj.108.135517 (2008).

81 Carter, J. V., Pan, J., Rai, S. N. & Galandiuk, S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery* **159**, 1638-1645, doi:https://doi.org/10.1016/j.surg.2015.12.029 (2016).

APPENDIX

A.1 Material and Methods

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

Lab ID	Age	Gender	Group	Abbreviation
106008	50	F	Aβ -	F50_neg
110203	54	F	Aβ -	F54_neg
108382	55	F	Aβ -	F55_neg
104968	58	F	Aβ -	F58_neg
107018	59	F	Aβ -	F59_neg
109456	59	F	Aβ -	F59*_neg
108259	60	F	Aβ -	F60_neg
109500	60	F	Aβ -	F60*_neg
106389	60	F	Aβ -	F60**_neg
106628	61	F	Aβ -	F61_neg
110073	46	M	Aβ -	M46_neg
107080	57	M	Aβ -	M57_neg
107461	57	M	Aβ -	M57*_neg
103293	60	M	Aβ -	M60_neg
106583	62	M	Aβ -	M62_neg
104251	63	M	Aβ -	M63_neg
102434	64	M	Aβ -	M64_neg
102721	68	M	Aβ -	M68_neg
107282	68	M	Aβ -	M68*_neg
102037	70	M	Aβ -	M70_neg
103698	49	F	Aβ +	F49_pos
107702	54	F	Aβ +	F54_pos
108726	55	F	Aβ +	F55_pos
106022	57	F	Aβ +	F57_pos
105728	60	F	Aβ +	F60_pos
100855	65	F	Aβ +	F65_pos
106746	68	F	Aβ +	F68_pos
103423	70	F	Aβ +	F70_pos
100112	71	F	Aβ +	F71_pos
100596	71	F	Aβ +	F71*_pos
101960	71	F	Aβ +	F71**_pos
104726	54	M	Aβ +	M54_pos
105438	55	M	Aβ +	M55_pos
105080	58	M	Aβ +	M58_pos
109901	62	M	Aβ +	M62_pos
109089	64	M	Aβ +	M64_pos
109522	64	M	Aβ +	M64*_pos
103176	69	M	Aβ +	M69_pos

102357	78	M	Aβ +	M78_pos
107362	80	M	Aβ +	M80_pos

*, ** are used to distinguish samples with same gender, age, and diagnosis.

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 119 statistically different features in the METLIN database based on mass/charge of the precursors only.

m/z	Retention time	METLIN ID matches	Monoisotopic Mass	Adduct	Delta (ppm)
133.1050	**22.87**	88347; 88348	132.0973	M+H	3
142.1248	**11.73**	1164; 3268; 43983; 64426; 64441; 68640; 86891; 89208; 94387; 94388	141.1154	M+H	15
149.1308	**18.48**	71487; 263602	148.1252	M+H	11
151.1488	**22.88**	87306; 91026; 97465	150.1409	M+H	4
152.1089	**11.72**	1493; 1846; 1961; 1962; 2190; 43546; 44096; 44808; 45122; 58167; 68162; 69213; 94213; 94214; 94218; 94229; 94383; 94578; 94580; 94585	151.0997	M+H	12
161.1364	**20.72**	No results for this query mass	-	-	-
163.1457	**20.48**	85265; 91785	162.1409	M+H	14
163.1519	**19.11**	85265; 91786	162.1409	M+H	23
166.1572	**18.48**	No results for this query mass	-	-	-
167.1419	**18.48**	75346; 86655; 86834; 87612; 87770; 87899; 88513; 89356; 89356; 91339; 91829	166.1358	M+H	6
170.1198	**11.72**	44245; 71886; 93042	169.1103	M+H	13
173.0547	**3.33**	6084; 69672; 89846	172.0484	M+H	5
173.1307	**20.48**	94809	172.1252	M+H	10
175.1492	**22.87**	91720; 96269; 96270; 265058	174.1409	M+H	6
176.1470	**23.03**	68166	175.1361	M+H	20
179.1471	**20.71**	2155; 70082; 87484; 87773; 87902; 88540; 90410; 91718; 92224; 92863; 92930	178.1358	M+H	22
180.1346	**16.13**	1262; 2013; 44229; 45381; 68067; 94230; 94231; 94232; 94238; 94239; 94241; 96184	179.1310	M+H	20
180.1725	**20.48**	263887	179.1674	M+H	12

181.1570	**20.48**	36515; 46052; 46113; 46117; 46443; 46445; 46447; 46448; 46449; 46453; 46455; 46456; 46457; 75281; 87473; 88509; 92864; 93921; 94148	180.1514	M+H	9
181.1625	**19.11**	36515; 46052; 46113; 46117; 46443; 46445; 46447; 46448; 46449; 46453; 46455; 46456; 46457; 75281; 87473; 88509; 92864; 93921; 94149	180.1514	M+H	21
184.1676	**18.48**	68068	183.1623	M+H	10
193.1570	**23.03**	69412; 69413; 70270; 75353; 86525; 86526; 86527; 86528; 86529; 86530; 87904; 88498; 88514; 89355; 90125; 90330; 90346; 90945; 91003; 91200; 91205; 91300; 91721; 91832; 92107; 95987	192.1514	M+H	8
194.1522	**18.48**	88262	193.1467	M+H	8
194.1590	**23.04**	88262	193.1467	M+H	26
197.1575	**20.71**	34856; 34895; 34896; 34897; 34898; 34899; 34900; 34901; 34902; 34903; 34904; 41094; 41103; 41117; 41143; 41144; 41145; 44507; 45778; 46188; 46193; 46204; 46208; 46210; 46436; 67014; 69426; 73905; 73906; 73907; 73908; 73909; 73910; 73911; 73912; 73913; 73914; 74123; 74124; 74125; 74126; 74127; 74128; 74129; 74130; 74131; 84985; 87239; 87307; 87361; 87868; 88250; 88369; 88810; 90291; 91226; 92250; 92995; 93005; 94758; 94906; 95861; 95880; 97269; 97286; 97287; 97351; 97387; 97390; 97391; 97392; 97397	196.1463	M+H	19
198.1454	**16.14**	No results for this query mass			
198.1829	**20.48**	65467; 263847; 263903; 263970	197.1780	M+H	11
198.1891	**19.11**	65467; 263847; 263903; 263970	197.1780	M+H	19
199.1872	**20.48**	515	198.1844	M+H	22
200.1625	**18.37**	No results for this query mass	-	-	-
206.8912	**2.16**	No results for this query mass	-	-	-
208.1675	**20.48**	2014; 67979	207.1623	M+H	10
208.1677	**22.77**	2014; 67979	207.1623	M+H	9
209.1493	**20.47**	69898; 86412; 87987; 90126; 90350; 90542; 91723; 91835;	208.1463	M+H	20

		91838; 92591; 92819; 93202; 94764; 96000			
209.1734	**20.26**	No results for this query mass	-	-	-
210.1119	**11.72**	44243; 65601; 69378; 69921	209.1052	M+H	2
211.1676	**23.03**	3923	210.1593	M+H	4
212.1627	**18.48**	2399; 68413	211.1572	M+H	8
212.1698	**23.06**	2399; 68413	211.1572	M+H	24
213.1704	**18.48**	No results for this query mass	-	-	-
214.1840	**20.71**	984916	213.1729	M+H	17
216.0994	**3.50**	44582; 62986	215.0938	M+H	7
216.1997	**19.11**	35924	215.1885	M+H	18
222.1445	**18.36**	2020; 3539; 44112; 67642; 68148; 985189	221.1416	M+H	19
223.0957	**11.14**	43594; 43823; 44543; 58086; 68332; 69854; 72951; 74901; 86105; 86289; 87847; 89684; 92099; 92511; 93234; 95782	222.0892	M+H	3
225.1097	**19.70**	68482; 87137; 87144; 87752; 88524; 88604; 91351; 95107; 95343	224.1049	M+H	10
225.1966	**22.77**	34504	224.1889	M+H	1
226.1784	**17.52**	263684; 263706	225.1729	M+H	7
226.1785	**24.35**	263684; 263706	225.1729	M+H	7
226.1804	**20.47**	263684; 263706	225.1729	M+H	1
227.1836	**20.47**	87369	226.1722	M+H	18
228.1900	**20.26**	No results for this query mass	-	-	-
228.1940	**23.03**	No results for this query mass	-	-	-
228.1950	**29.56**	No results for this query mass	-	-	-
229.1998	**23.03**	No results for this query mass	-	-	-
230.1731	**18.48**	No results for this query mass	-	-	-
233.0781	**16.42**	45180	232.0695	M+H	5
236.1657	**20.74**	69848	235.1572	M+H	5
238.1397	**16.35**	2336; 3540; 67703; 67841; 68213; 71291; 85031; 96325	237.1365	M+H	17

238.1756	**19.22**	263707	237.1729	M+H	19
240.1939	**22.4**	263740	239.1885	M+H	7
241.2052	**16.16**	No results for this query mass	-	-	-
243.2049	**17.52**	No results for this query mass	-	-	-
244.1908	**20.11**	No results for this query mass	-	-	-
245.1909	**20.12**	36476; 87985; 97008	244.1827	M+H	9
247.1757	**22.77**	62968; 93558	246.1654	M+H	12
250.0741	**8.13**	No results for this query mass	-	-	-
250.1756	**23.03**	3007; 44522; 67813; 71708; 985162	249.1729	M+H	18
251.1625	**23.03**	3023; 3025; 53360; 53506; 64102; 64130; 67581; 67900; 67921; 70284; 87285; 88017; 90162; 90455; 90474; 90484; 90658; 91007; 91214; 91215; 91631; 91639; 91706; 92042; 92395; 92416; 93752; 94199; 95219; 984914; 985113; 985215; 985293; 985294; 985299; 985308; 985316; 985361; 985483	250.1569	M+H	6
251.1802	**23.03**	997; 69205; 86905	250.1681	M+H	19
252.1550	**18.47**	72673	251.1521	M+H	17
253.1572	**18.48**	85359	252.1474	M+H	9
254.2097	**23.77**	263687; 263710; 263791	253.2042	M+H	6
258.2043	**24.35**	No results for this query mass	-	-	-
264.1547	**18.12**	1130; 1131; 88349	263.1521	M+H	17
265.1867	**17.52**	2807; 64452; 68548; 68564; 68569; 71537	264.1838	M+H	16
266.1536	**23.07**	1609; 2475; 93473; 95250	265.1467	M+H	1
266.1686	**23.10**	3996; 67414; 73025	265.1579	M+H	12
266.1700	**20.46**	3996; 67414; 73025	265.1579	M+H	18
267.1759	**20.46**	86754	266.1671	M+H	5
268.1754	**20.46**	68411; 729211	267.1735	M+H	20
268.1858	**23.02**	1250; 68411; 729211	267.1735	M+H	18
272.2198	**23.77**	No results for this query mass	-	-	-

280.1858	**22.37**	67996; 91470	279.1834	M+H	17
280.1863	**24.33**	67996; 91470	279.1834	M+H	15
282.1447	**20.12**	68776	281.1376	M+H	0
284.3300	**26.02**	43833	283.3239	M+H	4
291.0643	**22.88**	95551	290.0579	M+H	3
294.2017	**23.77**	939; 44179; 66272; 91472; 263716	293.1991	M+H	15
298.0980	**7.84**	425	297.0896	M+H	3
313.0461	**23.08**	85252; 86148	312.0481	M+H	29
314.0508	**23.09**	72752	313.0417	M+H	5
367.1482	**5.89**	86646; 88999	366.1427	M+H	4
376.2582	**24.67**	65463; 65464; 67123; 84994	375.2410	M+H	26
382.7601	**2.23**	No results for this query mass	-	-	-
384.1080	**6.79**	50393	383.1005	M+H	0
385.1216	**6.69**	2042; 47770; 47796; 47797; 48013; 48337; 49962; 50033; 50709; 50810; 51133; 51135; 51264; 51488; 51514; 51555; 53199; 86327; 90302; 90762; 91426; 93561; 94174; 94191; 94674	384.1209	M+H	17
406.0989	**6.59**	1707; 1708	405.0882	M+H	8
415.2061	**24.92**	72052	414.1943	M+H	10
419.2367	**24.43**	68545; 70751; 70776	418.2355	M+H	14
437.1923	**25.08**	47390; 47764; 47768; 47800; 47824; 48001; 48440; 48441; 49224; 49326; 52005; 52929; 52991; 95632	436.1886	M+H	8
467.3562	**26.77**	3030; 41562; 42409	466.3447	M+H	9
477.3616	**23.00**	42369; 83937; 89701; 94016	476.3502	M+H	8
489.3351	**26.89**	91884	488.3290	M+H	2
490.3433	**26.88**	85382	489.3302	M+H	11
509.349	**20.10**	73164; 73423; 84919	508.3400	M+H	3
510.3519	**20.10**	24068; 40033; 40053; 40054; 40057; 40158; 40290; 40291; 62269; 62296; 62785; 76587; 77678	509.3481	M+H	6

510.7016	**2.32**	No results for this query mass	-	-	-
531.3305	**20.10**	44214	530.3244	M+H	2
533.3639	**26.87**	No results for this query mass	-	-	-
534.3645	**26.87**	62280; 62306; 77690	533.3481	M+H	17
540.2844	**20.25**	44099	539.2803	M+H	5
541.2894	**20.28**	42027; 42030	540.2674	M+H	27
576.691	**2.40**	No results for this query mass	-	-	-
577.3900	**26.85**	81212; 81234; 81274; 81325	576.3791	M+H	6
603.1053	**22.9**	47200	602.0908	M+H	12
657.3605	**13.84**	No results for this query mass	-	-	-
751.501	**24.67**	43580	750.4918	M+H	2
851.3825	**24.97**	No results for this query mass	-	-	-

Table A.2.2. List of matches for each of the 119 statistically different features in the HMDB database based on mass/charge of the precursors only.

m/z	Retention time	HMDB ID matches	Monoisotopic Mass	Adduct	Delta (ppm)
133.1050	**22.87**	HMDB0032303; HMDB0032304	132.0973	M+H	3
142.1248	**11.73**	HMDB0030324; HMDB0039839; HMDB0030325; HMDB0039840; HMDB0033364; HMDB0030338	141.1154	M+H	15
149.1308	**18.48**	HMDB0059834; HMDB0061809; HMDB0061810	148.1252	M+H	11
151.1488	**22.88**	HMDB0030944; HMDB0035792; HMDB0061811	150.1409	M+H	4
152.1089	**11.72**	HMDB0039658; HMDB0039835; HMDB0060765; HMDB0001387; HMDB0039659; HMDB0040037; HMDB0039674; HMDB0003633; HMDB0039663; HMDB0040044; HMDB0001942	151.0997	M+H	13
161.1364	**20.72**	HMDB0061813; HMDB0061814	160.1252	M+H	24
163.1457	**20.48**	HMDB0013806; HMDB0036759; HMDB0061815	162.1409	M+H	15
163.1519	**19.11**	HMDB0013806; HMDB0036759; HMDB0061815	162.1409	M+H	23
166.1572	**18.48**	No results for this query mass	-	-	-
167.1419	**18.48**	HMDB0030002; HMDB0032540; HMDB0030252; HMDB0033546; HMDB0031565; HMDB0036815; HMDB0031359; HMDB0036190	166.1358	M+H	7
170.1198	**11.72**	HMDB0038321	169.1103	M+H	13
173.0547	**3.33**	HMDB0034322; HMDB0001212	172.0484	M+H	6
173.1307	**20.48**	HMDB0040284	172.1252	M+H	10
175.1492	**22.87**	HMDB0061824; HMDB0036683; HMDB0059826; HMDB0059696	174.1409	M+H	6
176.1470	**23.03**	No results for this query mass	-	-	-
179.1471	**20.71**	HMDB0014956; HMDB0032579; HMDB0038107; HMDB0038194; HMDB0031194; HMDB0035055; HMDB0031731; HMDB0037272; HMDB0031568; HMDB0036681	178.1358	M+H	23

180.1346	**16.13**	HMDB0041930; HMDB0014523; HMDB0039677; HMDB0041932; HMDB0039683; HMDB0039676; HMDB0039686; HMDB0039675; HMDB0039684	179.1310	M+H	20
180.1725	**20.48**	HMDB0015177; HMDB0014621	179.1674	M+H	12
181.1570	**20.48**	HMDB0031181; HMDB0039590; HMDB0032531; HMDB0039339; HMDB0038108	180.1514	M+H	9
181.1625	**19.11**	HMDB0031181; HMDB0039590; HMDB0032531; HMDB0039339; HMDB0038108	180.1514	M+H	21
184.1676	**18.48**	No results for this query mass	-	-	-
193.1570	**23.03**	HMDB0036684; HMDB0029820; HMDB0029824; HMDB0034671; HMDB0036818; HMDB0029821; HMDB0029825; HMDB0034959; HMDB0041615; HMDB0029823; HMDB0033545; HMDB0036144; HMDB0037139; HMDB0029822; HMDB0031733; HMDB0036022	192.1514	M+H	9
194.1522	**18.48**	No results for this query mass	-	-	-
194.1590	**23.04**	No results for this query mass	-	-	-
197.1575	**20.71**	HMDB0038270; HMDB0040227; HMDB0030867; HMDB0041494; HMDB0032051; HMDB0034903; HMDB0038280; HMDB0040397; HMDB0030945; HMDB0061816; HMDB0032193; HMDB0036048; HMDB0037301; HMDB0040206; HMDB0041469; HMDB0031692; HMDB0032901; HMDB0039522; HMDB0040725; HMDB0031030; HMDB0032331; HMDB0036599	196.1463	M+H	20
198.1454	**16.14**	No results for this query mass	-	-	-
198.1829	**20.48**	No results for this query mass	-	-	-
198.1891	**19.11**	No results for this query mass	-	-	-
199.1872	**20.48**	HMDB0015301	198.1844	M+H	23

200.1625	**18.37**	HMDB0062669	199.1572	M+H	10
206.8912	**2.16**	No results for this query mass	-	-	-
208.1675	**20.48**	HMDB0060611	207.1623	M+H	10
208.1677	**22.77**	HMDB0060611	207.1623	M+H	9
209.1493	**20.47**	HMDB0029668; HMDB0036686; HMDB0038529; HMDB0031869; HMDB0036824; HMDB0040233; HMDB0035228; HMDB0038047; HMDB0034672; HMDB0037776	208.1463	M+H	21
209.1734	**20.26**	No results for this query mass	-	-	-
210.1119	**11.72**	No results for this query mass	-	-	-
211.1676	**23.03**	HMDB0014631	210.1593	M+H	5
212.1627	**18.48**	No results for this query mass	-	-	-
212.1698	**23.06**	No results for this query mass	-	-	-
213.1704	**18.48**	No results for this query mass	-	-	-
214.1840	**20.71**	No results for this query mass	-	-	-
216.0994	**3.50**	HMDB0041830	215.0938	M+H	8
216.1997	**19.11**	No results for this query mass	-	-	-
222.1445	**18.36**	HMDB0030305; HMDB0060850	221.1416	M+H	20
223.0957	**11.14**	HMDB0041379; HMDB0033776; HMDB0029470; HMDB0034056; HMDB0037681; HMDB0013247; HMDB0031668; HMDB0038563; HMDB0029226; HMDB0094660; HMDB0002056; HMDB0030725; HMDB0037130	222.0892	M+H	4
225.1097	**19.70**	HMDB0036203; HMDB0030680; HMDB0040629; HMDB0030703; HMDB0032663; HMDB0040890; HMDB0032552	224.1049	M+H	11
225.1966	**22.77**	HMDB0030290	224.1889	M+H	2
226.1784	**17.52**	No results for this query mass	-	-	-

226.1785	**24.35**	No results for this query mass	-	-	-
226.1804	**20.47**	No results for this query mass	-	-	-
227.1836	**20.47**	HMDB0031041	226.1722	M+H	18
228.1900	**20.26**	No results for this query mass	-	-	-
228.1940	**23.03**	No results for this query mass	-	-	-
228.1950	**29.56**	No results for this query mass	-	-	-
229.1998	**23.03**	No results for this query mass	-	-	-
230.1731	**18.48**	HMDB0013267	229.1678	M+H	9
233.0781	**16.42**	HMDB0030813; HMDB0033330; HMDB0041325; HMDB0038911	232.0736	M+H	12
236.1657	**20.74**	HMDB0060851	235.1572	M+H	5
238.1397	**16.35**	No results for this query mass	-	-	-
238.1756	**19.22**	No results for this query mass	-	-	-
240.1939	**22.4**	No results for this query mass	-	-	-
241.2052	**16.16**	No results for this query mass	-	-	-
243.2049	**17.52**	No results for this query mass	-	-	-
244.1908	**20.11**	HMDB0013286	243.1834	M+H	0
245.1909	**20.12**	HMDB0031867; HMDB0039588	244.1827	M+H	4
247.1757	**22.77**	HMDB0038947	246.1654	M+H	12
250.0741	**8.13**	No results for this query mass	-	-	-
250.1756	**23.03**	HMDB0015004; HMDB0060852	249.1729	M+H	18

251.1625	**23.03**	HMDB0015371; HMDB0039644; HMDB0035117; HMDB0035760; HMDB0036550; HMDB0037529; HMDB0037559; HMDB0030917; HMDB0040754; HMDB0035137; HMDB0035798; HMDB0036563; HMDB0039635; HMDB0034721; HMDB0035358; HMDB0036037; HMDB0037064; HMDB0039156; HMDB0031901; HMDB0061839; HMDB0035148; HMDB0036036; HMDB0036664	250.1569	M+H	7
251.1802	**23.03**	HMDB0015124; HMDB0060655	250.1681	M+H	19
252.1550	**18.47**	HMDB0060956	251.1422	M+H	22
253.1572	**18.48**	HMDB0014451	252.1474	M+H	10
254.2097	**23.77**	No results for this query mass	-	-	-
258.2043	**24.35**	HMDB0013272	257.1991	M+H	8
264.1547	**18.12**	HMDB0032305	263.1521	M+H	18
265.1867	**17.52**	HMDB0032090; HMDB0061935	264.1725	M+H	26
266.1536	**23.07**	HMDB0014257; HMDB0038838; HMDB0060840; HMDB0040785	265.1467	M+H	1
266.1686	**23.10**	HMDB0014514; HMDB0015639	265.1579	M+H	13
266.1700	**20.46**	HMDB0014514; HMDB0015639	265.1579	M+H	18
267.1759	**20.46**	HMDB0030146	266.1671	M+H	6
268.1754	**20.46**	HMDB0030347	268.1701	M+H	20
268.1858	**23.02**	HMDB0001932	267.1834	M+H	18
272.2198	**23.77**	HMDB0013317	271.2147	M+H	8
280.1858	**22.37**	HMDB0015237	279.1754	M+H	11
280.1863	**24.33**	HMDB0015237	279.1754	M+H	13
282.1447	**20.12**	HMDB0030312	281.1416	M+H	15
284.3300	**26.02**	No results for this query mass	-	-	-
291.0643	**22.88**	HMDB0041121	290.0579	M+H	3

294.2017	**23.77**	HMDB0029846; HMDB0036328	293.1991	M+H	16
298.0980	**7.84**	HMDB0001173	297.0896	M+H	4
313.0461	**23.08**	No results for this query mass	-	-	-
314.0508	**23.09**	No results for this query mass	-	-	-
367.1482	**5.89**	HMDB0033108; HMDB0012492	366.1427	M+H	5
376.2582	**24.67**	No results for this query mass	-	-	-
382.7601	**2.23**	No results for this query mass	-	-	-
384.1080	**6.79**	HMDB0000912	383.1077	M+H	18
385.1216	**6.69**	HMDB0036281; HMDB0039634; HMDB0030621; HMDB0038874; HMDB0040934; HMDB0035479; HMDB0039617; HMDB0034914; HMDB0038950; HMDB0029541	384.1209	M+H	17
406.0989	**6.59**	HMDB0014827	405.0995	M+H	19
415.2061	**24.92**	HMDB0014838; HMDB0039380; HMDB0030404; HMDB0037039; HMDB0041407	414.2042	M+H	13
419.2367	**24.43**	No results for this query mass	-	-	-
437.1923	**25.08**	HMDB0033806	436.1886	M+H	8
467.3562	**26.77**	No results for this query mass	-	-	-
477.3616	**23.00**	HMDB0034079; HMDB0039443	476.3502	M+H	9
489.3351	**26.89**	HMDB0036884	488.3290	M+H	2
490.3433	**26.88**	No results for this query mass	-	-	-
509.349	**20.10**	No results for this query mass	-	-	-
510.3519	**20.10**	HMDB0011481; HMDB0011511	509.3481	M+H	7
510.7016	**2.32**	No results for this query mass	-	-	-
531.3305	**20.1**	HMDB0030392; HMDB0060527; HMDB0060526; HMDB0060525	530.3104	M+H	24
533.3639	**26.87**	No results for this query mass	-	-	-
534.3645	**26.87**	HMDB0011492; HMDB0011522	533.3481	M+H	17

540.2844	**20.25**	No results for this query mass	-	-	-
541.2894	**20.28**	No results for this query mass	-	-	-
576.691	**2.4**	No results for this query mass	-	-	-
577.39	**26.85**	HMDB0029310	576.3662	M+H	29
603.1053	**22.9**	No results for this query mass	-	-	-
657.3605	**13.84**	No results for this query mass	-	-	-
751.501	**24.67**	No results for this query mass	-	-	-
851.3825	**24.97**	No results for this query mass	-	-	-

Table A.2.3. List of features identified with a maximum score in the putative ID level in HMDB, METLIN, and in-house library.

m/z	Retention time	Compound	Pathway	Database ID	KEGG
133.1050	**22.87**	Multiple matches (ornithine or L-asparagine or cinnamaldehyde)	Biosynthesis of amino acids (first two); Aminoacyl-tRNA biosynthesis (second one)	HMDB0000214 HMDB0000168 HMDB0003441	C00077 C00152 C00903
142.1248	**11.73**	L-Histidinol	Histidine metabolism	HMDB0003431	C00860
142.1248	**11.73**	Multiple matches (tropine or pelletierine)	Tropane, piperidine, and pyridine alkaloid biosynthesis; Biosynthesis of alkaloids derived from ornithine, lysine, and nicotinic acid	Metlin ID 1164 Metlin ID 43983	C00729 C06182
151.1488	**22.88**	Multiple matches (thymol or (S)-carvone or 2,3-diethyl-5-methylpyrazine)	Monoterpenoid biosynthesis (second one)	HMDB0001878 HMDB0004487 HMDB0041345	C09908 C11383 Not found
161.1364	**20.72**	Tryptamine	Tryptophan metabolism	HMDB0000303	C00398
163.1457	**20.48**	Nicotine	Tropane, piperidine and pyridine alkaloid biosynthesis	HMDB0001934	C00745
173.0547	**3.33**	Multiple matches (glycerol 3-phosphate or menadione)	Glycerolipid and glycerophospholipid metabolism (first one); Vitamin digestion and absorption (second one)	HMDB0000126 HMDB0001892	C00093 C05377
173.1307	**20.48**	Multiple matches (capric acid or 1,2,3,4-tetrahydro-beta-carboline)	Fatty acid biosynthesis (first one)	HMDB0000511 HMDB0012488	C01571 Not found
198.1891	**19.11**	N-Methylundec-10-enamide	-	Metlin ID 65467	Not found
210.1119	**11.72**	Minoxidil	-	HMDB0014494	Not found

222.1445	**18.36**	Procarbazine	-	HMDB0015299	C07402
225.1097	**19.70**	3-Hydroxy-L-kynurenine	Tryptophan metabolism	HMDB0000732	C03227
233.0781	**16.42**	Kifunensine	-	Metlin ID 45180	Not found
236.1657	**20.74**	Procainamide	-	HMDB0015169	C07401
240.1939	**22.40**	Salbutamol	-	HMDB0001937	C11770
247.1757	**22.77**	Multiple matches (nefiracetam or milnacipran)	-	Not found HMDB0015602	Not found Not found
250.0741	**8.13**	Sulfapyridine	-	HMDB0015028	Not found
250.1756	**23.03**	Alprenonol	-	Metlin ID 44522	Not found
251.1625	**23.03**	Trolox	-	HMDB0038804	Not found
258.2043	**24.35**	Dextrorphan	-	HMDB0060552	Not found
265.1867	**17.52**	N-Valylphenylalanine	-	HMDB0255244	Not found
265.1867	**17.52**	Tetracaine	-	Metlin ID 2807	C07526
266.1700	**20.46**	Oxprenolol	-	Metlin ID 34545	Not found
267.1759	**20.46**	Multiple matches (Practolol or Atenolol)	-	Metlin ID 44683 Metlin ID 1128	C11696 Not found
268.1754	**20.46**	Azacyclonol	-	Metlin ID 2771	Not found
268.1858	**23.02**	Metoprolol	-	HMDB0001932 Metlin ID 1250	C07202
284.3300	**26.02**	Cetrimonium	-	Metlin ID 43833	Not found
294.2017	**23.77**	Multiple matches (anastrozole or ondansetron)	-	HMDB0015348 HMDB0005035	C08159 C07325
298.0980	**7.84**	5'-Methylthioadenosine	Cysteine and methionine metabolism	HMDB0001173 Metlin ID 3425	C00170
313.0461	**23.08**	Multiple matches (2',5,6-trimethoxyflavone or cabreuvin)	-	HMDB0037331 HMDB0030679	Not found Not found
367.1482	**5.89**	Glutaminyl-tyrosyl-glycine	-	Metlin ID 17118	Not found

376.2582	**24.67**	Multiple matches (R-4-benzyl-3-((R)-3-hydroxy-2,2-dimethyloctanoyl)-5,5-dimethyloxazolidin-2-one or Lysyl-lysyl-threonine or S-4-benzyl-3-((S)-3-hydroxy-2,2-dimethyloctanoyl)-5,5-dimethyloxazolidin-2-one)	-	Metlin ID 65464 Metlin ID 19689 Metlin ID 65463	Not found Not found Not found
384.1080	**6.79**	Multiple matches (felodipine or prazosin)	Felodipine metabolism pathway (first one)	HMDB0015158 HMDB0014600	Not found C07368
384.1080	**6.79**	Methionyl-methionyl-cysteine	-	Metlin ID 16843	Not found
415.2061	**24.92**	Multiple matches (Lys His Met or Met His Lys or Met Lys His or His Lys Met)	-	Metlin ID 17238 Metlin ID 18880 Metlin ID 21819 Metlin ID 16754	Not found Not found Not found Not found
419.2367	**24.43**	2-(4-Allyl-2,6-dimethoxyphenoxy)-1-(3,4,5-trimethoxyphenyl)-1-propanol	-	HMDB0030828	Not found
419.2367	**24.43**	Threonyl-tryptophyl-leucyine	-	Metlin ID 20859	Not found
489.3351	**26.89**	Esculentic acid	-	HMDB0035782	Not found
509.3490	**20.10**	Fasciculol C	-	HMDB0035853	Not found

Table A.2.4. List of the features identified with a putative ID in the CSF sample of the Aβ positive and Aβ negative groups belonging to the significant pathways from the Pathway Enrichment analysis performed in the MetaboAnalyst.

Pathway Name	p-value	Impact	Metabolites included in the pathway
Tryptophan metabolism	0.11308	0.1121	Tryptamine (161.1364_20.72) and 3-Hydroxy-L-kynurenine (225.1097_19.70)
Arginine biosynthesis	0.18207	0.0609	Ornithine (133.1050_22.87)
Glycerolipid metabolism	0.20534	0.0436	Glycerol-3-phosphate (173.0547_3.33)
Glutathione metabolism	0.33224	0.0000	Ornithine (133.1050_22.87)
Cysteine and methionine metabolism	0.37918	0.0209	5'-methylthioadenosine (298.0980_7.84)
Glycerophospholipid metabolism	0.40582	0.0809	Glycerol-3-phosphate (173.0547_3.33)
Arginine and proline metabolism	0.42297	0.1106	Ornithine (133.1050_22.87)
Fatty acid biosynthesis	0.49446	0.0000	Capric acid (173.1307_20.48)
Aminoacyl-tRNA biosynthesis	0.50186	0.0000	L-Asparagine (133.1050_22.87)

Table A.2.5. List of the features identified with a putative ID in the CSF sample of the Aβ positive and Aβ negative groups belonging to the drug-associated pathways from the Enrichment analysis performed in the MetaboAnalyst.

Drug-associated pathway	p-value	Metabolite included in the pathway
Alprenolol	0.364	Alprenolol (250.1756_23.03)
Atenolol	0.364	Atenolol (267.1759_20.46)
Felodipine	0.364	Felodipine (384.1080_6.79)
Metoprolol	0.364	Metoprolol (268.1858_23.02)
Oxprenolol	0.364	Oxprenolol (266.1700_20.46)
Practolol	0.364	Practolol_267.1759_20.46)
Procainamide (Antiarrhythmic)	0.364	Procainamide (236.1657_20.74)
Amikacin	0.531	L-Asparagine (133.1050_22.87)
Arbekacin	0.531	L-Asparagine (133.1050_22.87)
Azithromycin	0.531	L-Asparagine (133.1050_22.87)
Chloramphenicol	0.531	L-Asparagine (133.1050_22.87)
Clarithromycin	0.531	L-Asparagine (133.1050_22.87)

Clindamycin	0.531	L-Asparagine (133.1050_22.87)
Clomocycline	0.531	L-Asparagine (133.1050_22.87)
Demeclocycline	0.531	L-Asparagine (133.1050_22.87)
Doxycycline	0.531	L-Asparagine (133.1050_22.87)
Erythromycin	0.531	L-Asparagine (133.1050_22.87)
Gentamicin	0.531	L-Asparagine (133.1050_22.87)
Josamycin	0.531	L-Asparagine (133.1050_22.87)
Kanamycin	0.531	L-Asparagine (133.1050_22.87)
Lincomycin	0.531	L-Asparagine (133.1050_22.87)
Lymecycline	0.531	L-Asparagine (133.1050_22.87)
Methacycline	0.531	L-Asparagine (133.1050_22.87)
Minocycline	0.531	L-Asparagine (133.1050_22.87)
Neomycin	0.531	L-Asparagine (133.1050_22.87)
Netilmicin	0.531	L-Asparagine (133.1050_22.87)
Oxytetracycline	0.531	L-Asparagine (133.1050_22.87)
Paromomycin	0.531	L-Asparagine (133.1050_22.87)
Rolitetracycline	0.531	L-Asparagine (133.1050_22.87)
Roxithromycin	0.531	L-Asparagine (133.1050_22.87)
Spectinomycin	0.531	L-Asparagine (133.1050_22.87)
Streptomycin	0.531	L-Asparagine (133.1050_22.87)
Telithromycin	0.531	L-Asparagine (133.1050_22.87)
Tetracycline	0.531	L-Asparagine (133.1050_22.87)
Tigecycline	0.531	L-Asparagine (133.1050_22.87)
Tobramycin	0.531	L-Asparagine (133.1050_22.87)
Troleandomycin	0.531	L-Asparagine (133.1050_22.87)
Nicotine	0.747	Nicotine (163.1457_20.48)