UNIVERSITAT POLITÈCNICA DE VALÈNCIA

INSTITUTO INTERUNIVERSITARIO DE INVESTIGACIÓN DE RECONOCIMIENTO MOLECULAR Y DESARROLLO TECNOLÓGICO

NMR-based metabolomics for the identification of biomarkers of disease

PhD. THESIS

Submitted by

Marina Dolores Botello Marabotto

PhD. Supervisors:

Dra. M. Carmen Martínez Bisbal

Dra. Andrea Bernardos Bau

Prof. Ramón Martínez Máñez

Valencia, September 2024

M. CARMEN MARTINEZ BISBAL, PhD in Chemistry and Associate Professor at the *Universitat de València*, ANDREA BERNARDOS BAU, PhD in Chemistry and Distinguished Researcher at the *Universitat Politècnica de València* and RAMÓN MARTÍNEZ MÁÑEZ, PhD in Chemistry and Professor at the *Universitat Politècnica de València*.

CERTIFY:

That the work *"NMR-based metabolomics for the identification of biomarkers of disease"* has been developed by *Marina Dolores Botello Marabotto* under their supervision in the Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico (IDM) of the *Universitat Politècnica de València*, forming part of the Unidad Mixta en Nanomedicina y Sensores de Universitat Politècnica de València Instituto de Investigación Sanitaria LaFe , as a Thesis Project in order to obtain the degree of PhD in Biotechnology at the *Universitat Politècnica de València.*

Valencia, 2024

Dra. M. Carmen Martínez Bisbal   Dra. Andrea Bernados Bau

Directora   Directora

Prof. Ramón Martínez Máñez

Director

A mi familia

"No dejes que el miedo al tiempo que tomará realizar algo se interponga en tu camino. El tiempo pasará de todos modos; podrías también poner ese tiempo en el mejor uso posible".

-**Earl Nightingale**

"Todo se cura con agua salda: sudor, lágrimas o el mar".

- **Karen Blixen**

Agradecimientos

Acknowledgments

Es muy bonito, al acabar una etapa, mirar hacia atrás y ver que tienes muchas razones para sentirte agradecida, y muchas personas a las que dar las gracias y expresar tu cariño. Me siento muy feliz y afortunada, porque la lista es larga, y no todos los días se tiene la oportunidad de expresarlo.

Me gustaría empezar dando las gracias a mis directores de tesis, Ramón, Andrea y Mari Carmen. En primer lugar, a Ramón, gracias por darme la oportunidad de realizar esta tesis, y por todas las correcciones y consejos. A Andrea, porque, aunque lo de que acabaras siendo mi directora nos pilló un poco por sorpresa a las dos, te has implicado conmigo desde el principio. Muchas gracias por tu energía y optimismo, y por no tener problema en volver a ponerte los guantes, cuando necesitaba ayuda con las síntesis. Y a Mari Carmen, muchas gracias por tu sentido del humor, por toda la confianza que has depositado en mí, y por estar siempre disponible... y también por tener siempre chocolate en tu despacho jajaja.

A la gente del departamento de química. Gracias a todos por la ayuda prestada y el cariño que me habéis mostrado en mis etapas en el 2.6. A Giovanni, gracias por apuntarte conmigo a mi primer torneo, aunque fuera una patata jugando, nos lo pasamos bien jajaja. A Marcia, Andy y a Eva.G por su sabiduría y buen rollo, Guille, Rocío y Eva.C por su energía y buen humor, a Jessie, Guille.F y Edu, por su amabilidad y dulzura. También Laura, a Jordi, a Pablo, a David.B... También gracias a la gente del 0.2 por prestarme su ayuda cuando la he necesitado, a Paula, a Javi. H, a Carol, y a Sandra. Gracias también a la gente de la CPI, Miguel, Serena, Angy,.. a Yoel, que ha sido un poco trotamundos como yo. A la gente del CIPF, a Javi, por apuntarse sin dudarlo a formar

parte del pack de yogures, y a David, a los dos por estar siempre dispuestos a echar unas risas. También a Xente, por su entusiasmo contagioso, a Juanjo, Andrea.E, y a Alba.G por su ayuda. A las personas que se han ido incorporando al CIPF y con las que he tenido la suerte de coincidir en un momento u otro: Alba, Paula, Jenni, Sandra, María… y a Isa, gracias a ti he empezado a ganar partidas de padel jajaja. También me gustaría darle las gracias por sus labores de gestión a Eva B. y a Tania por su alegría y buenos consejos. Y a Elena, Toni, Vicente y Estela, por su ayuda.

I would also like to thank Professor Craig Wheelock and all the amazing people from his lab at the Karolinska Institutet, where I had the opportunity to stay for three months. Thank you all for your kindness and for making me feel so welcome. Especialmente gracias a Toni, por toda su ayuda, y sus enseñanzas durante la estancia y Tanize, por sus ganas de ayudar y sus palabras de aliento.

A mis supervivientes, habéis hecho que esta etapa valiera la pena a muchos niveles más allá del académico. Gracias por todas las conversaciones, los planes, las risas, los viajes... En resumen, gracias por vuestra amistad. A Elena, por todos los buenos ratos, y por toda tu ayuda en estos últimos momentos de preparación de la tesis. A Blanca por ser tan risueña, por tu amabilidad y por siempre estar dispuesta ayudar. A mis chicas de la Fe: a Nieves, que llegó la primera a inaugurar el equipo de la nueva Fe, cuando yo acababa de aterrizar en el grupo y lo estaba pasando... vamos a decir que regular. Gracias por siempre intentar que todo el mundo esté a gusto. A Isa, por tu energía incansable y por siempre estar dispuesta para cualquier plan, desde una noche de fiesta… hasta una mudanza jajaj. A Andrea, gracias por tu entusiasmo y tus ganas, sigue así y llegarás donde te dé la gana. Y a Alba, quien me iba a decir que iba a encontrar a mi gemela en el mundo en Valencia. Por nuestros ratitos de charlas y risas en la cena y por las sesiones de cine. Gracias por aguantarme cuando estoy de mal humor e intentar alegrarme cuando notas que estoy triste o estresada. En fin... que os quiero chicas, y que os voy a echar de menos.

También me gustaría darle las gracias a Clara, porque, aunque su tiempo en La Fe ha sido breve ha aportado muchas cosas buenas. Gracias por estar siempre dispuesta a ayudar. A Pasqual, por su entusiasmo... y por haberme ayudado a ejercitar la paciencia jajaja. Y también a Ana, por su alegría, y por sus ganas de ayudar. I would also like to thank my metabolomics partner, Molud. Thank you for being always so kind and willing to share your culture with me. It has been great getting to know you. Y a María Ovejero, fue genial tener una compañera con la que compartir proyectos. Muchas gracias por toda tu ayuda.

Y por supuesto quiero darle las gracias a mucha gente que ha estado a mi lado a lo largo de estos años, y, aunque no directamente, han contribuido también a que esta tesis sea posible.

A mis ovejas negras, las de siempre y para siempre. A María, porque siempre estas para recogernos cuando estamos hechas cachitos. Nunca me cansaré de darte las gracias, ya lo sabes, mi vida ahora sería muy diferente si no me hubieras rescatado de los recreos de biblioteca. A Luisa, tan auténtica, siempre directa, siempre honesta, por querernos con tanta fiereza. Mi eterna compañera de pupitre, nuestras yo de hace ya más de diez años se sentirían orgullosas. A Ana, mi Zape, con su capacidad de adaptarse a cualquier situación y sus palabras profundas, siempre acertadas. Y por último a mi Aracelika, mi medio limón. Gracias por siempre creer en mí y darme ánimos para que yo también lo haga, y por ser un ejemplo vital. Créeme cuando te digo que hay cosas que si son para siempre. Gracias chicas, porque, aunque no estéis, estáis. Qué bonito haber crecido juntas, y que bonito seguir viéndonos crecer. Gracias por las palabras de aliento, y por las reuniones por videollamada que son mejores que una sesión de terapia. Porque, aunque cada vez sea más difícil, siempre conseguimos sacar un ratito para nosotras. Os quiero mucho, chicas.

A mi familia, no sería nada sin vosotros. A mi madre, por ser la persona con más paciencia del mundo. Muchas gracias por escucharme e intentar entenderme siempre. A mi padre, por estar dispuesto a recorrer el mundo entero por mí. A los dos, por estar tan orgullosos de mí. Y a mi hermano Rodrigo, por ser mi primer amigo, y mi primer ejemplo, y porque sé que siempre voy a poder contar contigo. Gracias a los tres por haberme demostrado vuestro amor incondicional en cada paso que he dado. No podría haber hecho esto sin vosotros. Os quiero mucho.

A todos mis titos y primos, por alegrarse tanto de mis logros. Y a mis abuelos. A mi abuela Mari Loli por mostrarme su cariño y su ilusión a través de todos los medios posibles, y a mi abuela Maria Jesús y mis abuelos Julio y Modesto, que sé que habrían estado muy orgullosos. Muchas gracias también a toda mi familia de Valencia, en especial a mis titos Luisi y Pedro, y a mis primas Marina y Alba, por haberme acogido con tanto cariño cuando llegué y me sentía un poco sola y perdida, y a mi tío Vicente por siempre ofrecerme su casa. Gracias por haberme ayudado cada vez que lo he necesitado a lo largo de estos años.

Y a Nacho. Gracias por todo tu apoyo en la realización de esta tesis, no solo técnico jajaja, si no sobre todo emocional. Gracias por ser el mejor compañero, por quererme tan bonito. Por darme fuerzas, cuando yo no las tengo, y ayudarme a arrancar cuando me siento atascada. Te quiero mucho. Gran parte de esto también es tuyo. Ahora todo recto hacia nuestro futuro. Se acabó ver series cada uno en su cama, sincronizando llamadas. Ahora lo haremos juntos, en nuestro sofá.

Resumen

La presente tesis doctoral titulada *"NMR-based metabolomics for the identification of biomarkers of disease"* se centra en explorar y mostrar el potencial de la metabolómica mediante espectroscopía de resonancia magnética nuclear (RMN) como herramienta para la detección de nuevos biomarcadores de enfermedad que permitan un diagnóstico temprano y no invasivo, así como un seguimiento del paciente. La metabolómica mediante espectroscopía de RMN se aplica específicamente en esta tesis al estudio de cuatro enfermedades: demencia tipo Alzheimer, glaucoma, aterosclerosis y vulnerabilidad de placa, y fibrosis pulmonar desarrollada tras neumonía por COVID-19.

En la introducción se presenta una descripción del proceso de análisis metabolómico para la identificación de biomarcadores de enfermedad, así como de las principales plataformas empleadas en metabolómica. Posteriormente se profundiza en el proceso seguido para análisis metabolómico mediante espectroscopía de RMN, así como en las principales herramientas estadísticas utilizadas en esta tesis. Finalmente, se describen las principales características fisiopatológicas de las enfermedades estudiadas en esta tesis, así como las formas de diagnóstico actuales y la necesidad de identificar nuevos marcadores de enfermedad.

A continuación, se describe el objetivo principal de esta tesis, así como los objetivos específicos que se abordan en los capítulos experimentales.

En el primer capítulo, se identifican biomarcadores de la enfermedad de Alzheimer (*Alzheimer disease*, AD) y del progreso de pacientes con deterioro cognitivo leve (*mild cognitive impairment*, MCI) a enfermedad de Alzheimer mediante metabolómica de RMN en muestras de suero. Para ello, en primer lugar, se producen modelos de estadística multivariante de análisis discriminante de mínimos cuadrados parciales

(*Partial least squares-discriminant analysis*, PLS-DA) para la discriminación entre pacientes de AD y MCI y controles con cognición normal (*healthy controls,* HC). Los modelos de clasificación obtenidos para discriminar AD frente a MCI y HC mostraron excelentes parámetros de clasificación. Sin embargo, la discriminación entre MCI y HC no alcanzó significación, poniendo de manifiesto cómo el desarrollo de AD altera el metabolismo celular. Posteriormente, para la identificación de biomarcadores del proceso de evolución de MCI a AD, el grupo MCI se subdividió atendiendo a, si en un periodo de 1 a 3 años tras la adquisición de la muestra, su estado cognitivo había deteriorado a AD. Se observó que la lisina, el piruvato, la colina, la fenilalanina y las lipoproteínas, tenían unos valores de concentración relativa en suero diferente dependiendo si el paciente evolucionaba a AD o se mantenía estable. Además, se observaron diferencias en las rutas metabólicas entre estos grupos de pacientes.

En el segundo capítulo de la tesis, se estudiaron las diferencias metabolómicas entre pacientes MCI y control de un grupo de pacientes diferente, por medio del análisis de los metabolitos presentes en muestras de plasma con espectroscopía de RMN, así como la presencia de marcadores de peroxidación lipídica, que fueron analizados por medio de espectrometría de masas acoplada a cromatografía líquida (UPLC-MS/MS). El proceso de oxidación celular se ha visto estrechamente relacionado con el desarrollo de la enfermedad de Alzheimer y, de hecho, en el trabajo presentado se observa como la adición de marcadores de peroxidación lipídica al conjunto de datos de metabolómica de RMN para su estudio mediante estadística multivariante mejora los modelos de clasificación para la identificación de biomarcadores de MCI. En este trabajo se realizaron modelos de clasificación PLS-DA utilizando los metabolitos identificados mediante espectroscopía de RMN, los marcadores de peroxidación lipídica detectados con UPLC-MS/MS, y la combinación de ambos, obteniéndose el resultado con una mayor significación estadística cuando se combinan ambas técnicas. Los metabolitos identificados como potenciales biomarcadores de MCI en plasma son

isoleucina, valina, 3-hidroxibutirato, 2-hidroxi-3-metilvalerato, glutamato, glutatión, cisteína, malonato, n-nitroso dimetilamina, taurina, prolina e isoprostanos. Especialmente, por su relación con la patología, se destacan como potenciales biomarcadores la isoleucina, la valina, el glutamato, la taurina, la prolina, la cisteína y los isoprostanos.

En el tercer capítulo de la tesis, se analizaron lágrimas de pacientes de glaucoma primario de ángulo abierto (POAG) y controles por medio de espectroscopía de RMN, para la identificación de biomarcadores de la patología en un medio mínimamente invasivo, como son las lágrimas. Los modelos de clasificación obtenidos permitieron clasificar las muestras de POAG y controles con altos valores de sensibilidad y especificidad. Del análisis de los metabolitos participantes en el modelo de clasificación se determinaron como potenciales biomarcadores de POAG en lágrima la taurina, glicina, urea, glucosa, ácidos grasos insaturados, fenilalanina, fenilacetato, leucina, compuestos n-acetilados, el ácido fórmico y la uridina. De especial relevancia son la fenilalanina, glucosa, leucina, glicina y taurina por su relación con la patología.

A continuación, en el cuarto capítulo se presentan los resultados obtenidos del análisis metabolómico mediante espectroscopía de RMN de muestras de placas de ateroma y suero de pacientes con estenosis de carótida. Se obtuvieron muestras de pacientes sintomáticos (que habían presentado algún signo de accidente cerebrovascular) y asintomáticos, a fin de determinar biomarcadores de vulnerabilidad de placa. De los análisis realizados en placa, por medio de espectroscopía de ángulo mágico de alta resolución (*high resolution magic angle spinning*, HRMAS), se determinaron como potenciales biomarcadores de vulnerabilidad el mio-inositol, glutamato, y ácidos grasos insaturados, teniendo especial relevancia el glutamato por su relación con la patología. La determinación de biomarcadores en suero resultaba más interesante a nivel clínico, y del análisis estadístico realizado se obtuvieron como potenciales marcadores de vulnerabilidad de placa la treonina, la histamina y los ácidos grasos insaturados.

En el quinto y último capítulo experimental se analizaron por medio de espectroscopía de RMN muestras de suero de pacientes que habían estado ingresados en el hospital por neumonía causada por el virus SARS-CoV-2, a fin de identificar biomarcadores para predecir el desarrollo de fibrosis pulmonar, una de las secuelas más graves desarrolladas tras la COVID-19. Los modelos de clasificación PLS-DA generados fueron capaces de discriminar que pacientes habrían desarrollado fibrosis pulmonar un año después del alta hospitalaria (y la toma de muestras) con altos valores de sensibilidad y especificidad. Los metabolitos propuestos como potenciales biomarcadores de desarrollo de fibrosis, por su participación en el modelo y la presencia de diferencias significativas en su concentración relativa entre ambos grupos son la glucosa, valina y ácidos grasos.

Finalmente, se presenta la discusión general y las principales conclusiones generales de la tesis, así como las conclusiones derivadas del trabajo experimental aquí presentado. Esperamos que estos resultados abran la puerta al uso de la metabolómica mediante espectroscopía de RMN para la identificación de biomarcadores tempranos y no invasivos, y de este modo resolver una de las principales necesidades de la medicina actual.

Abstract

The present doctoral thesis entitled "NMR-based metabolomics for the identification of biomarkers of disease" focuses on exploring and demonstrating the potential of nuclear magnetic resonance (NMR) spectroscopy-based metabolomics as a tool for identifying new disease biomarkers that allow for early and non-invasive diagnosis, as well as patient monitoring. NMR spectroscopy-based metabolomics is here specifically applied to the study of four diseases: Alzheimer's disease, glaucoma, atherosclerosis and plaque vulnerability, and pulmonary fibrosis developed after COVID-19 pneumonia.

The general introduction provides a description of the metabolomic analysis process for the identification of disease biomarkers, as well as the main platforms used in metabolomics. Subsequently, it focuses on the process to follow for an NMR spectroscopy metabolomic analysis, as well as on the main statistical tools used in this thesis. Finally, the main pathophysiological characteristics of the diseases studied in this thesis are described, as well as the current diagnostic methods and the need to identify new biomarkers of disease.

Next, the main objective of this thesis is described, as well as the specific objectives addressed in the experimental chapters.

In the first chapter, biomarkers of Alzheimer's disease (AD) and the progression of patients with mild cognitive impairment (MCI) to Alzheimer's are identified using NMR spectroscopy-based metabolomics in serum samples. Classification models were produced for discriminating between AD and MCI patients and normal cognition controls (HC). The classification models obtained for discriminating AD from MCI and

HC showed excellent classification parameters. However, discrimination between MCI and HC was not as significant, highlighting how the development of AD alters cellular metabolism. Subsequently, for the identification of biomarkers of the MCI to AD progression process, the MCI group was subdivided based on whether, within a period of 1 to 3 years after sample acquisition, their cognitive status had deteriorated to AD. It was observed that lysine, pyruvate, choline, phenylalanine, and lipoproteins had different relative concentration values in serum depending on whether the patient progressed to AD or remained stable. Furthermore, differences in metabolic pathways were observed between these groups of patients.

In the second chapter of the thesis, metabolic differences between MCI patients and controls from a different sample group were studied by analyzing the metabolites present in plasma samples by NMR spectroscopy, as well as the presence of markers of lipid peroxidation, which were analyzed using UPLC-MS/MS. The process of cellular oxidation has been closely related to the development of Alzheimer's disease, and indeed, in the presented work, it is observed that adding biomarkers of lipid peroxidation to NMR spectroscopy metabolomics data improves multivariate statistical classification models for identifying MCI biomarkers. In this work, PLS-DA classification models were performed using the metabolites identified by NMR spectroscopy, lipid peroxidation markers detected with ultra-performance liquid chromatography-mass spectroscopy (UPLC-MS/MS), and the combination of both, yielding greater statistical significance when both techniques are combined. The identified potential MCI biomarkers in plasma are isoleucine, valine, 3-hydroxybutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, taurine, proline, and isoprostanes. Especially noteworthy are isoleucine, valine, glutamate, taurine, proline, cysteine, and isoprostanes due to their relationship with the pathology.

In the third chapter of the thesis, tears from primary open-angle glaucoma (POAG) patients and controls were analyzed using NMR spectroscopy to identify biomarkers of the pathology in a minimally invasive medium, such as tears. The classification model obtained allowed the classification of POAG and control samples with high values of sensitivity and specificity. From the analysis of the metabolites participating in the classification model, taurine, glycine, urea, glucose, unsaturated fatty acids, phenylalanine, phenylacetate, leucine, N-acetylated compounds, formic acid, and uridine were determined as potential tear biomarkers of POAG. Phenylalanine, glucose, leucine, glycine, and taurine are particularly relevant due to their relationship to the pathology.

Next, in the fourth chapter, the results obtained from metabolomic analysis by NMR spectroscopy of atheroma plaque samples and serum from patients with carotid stenosis are presented. Samples were obtained from symptomatic patients (who had shown signs of a cerebrovascular incident) and asymptomatic ones to determine plaque vulnerability biomarkers. From the plaque analysis performed using high resolution-magic angle spinning (HRMAS), myo-inositol, glutamate, and unsaturated fatty acids were determined as potential vulnerability plaque biomarkers, with glutamate being particularly relevant due to its relation to the pathology. Determining biomarkers in serum was more clinically relevant, and potential plaque vulnerability markers were obtained from statistical analysis as threonine, histamine, and unsaturated fatty acids.

In the fifth and final experimental chapter, serum samples from patients hospitalized for pneumonia caused by the SARS-CoV-2 virus were analyzed by NMR spectroscopy to identify biomarkers to predict the development of pulmonary fibrosis, one of the most severe sequelae after COVID-19. The PLS-DA classification models generated were able to discriminate which patients developed pulmonary fibrosis one year after hospital

discharge (and sample collection) with high values of sensitivity and specificity. The proposed metabolites as potential fibrosis development biomarkers, due to their participation in the model and significant differences in their relative concentration between both groups are glucose, valine, and fatty acids.

Finally, the general discussion and the main general conclusions of the thesis are presented, as well as the conclusions derived from the experimental work presented here. We hope that these results pave the way for the use of NMR spectroscopy metabolomics to identify early and non-invasive biomarkers, thus addressing one of the main needs of current medicine.

Resum

Esta tesi doctoral titulada "*NMR-based metabolomics for the identification of biomarkers of disease*" es centra en explorar i mostrar el potencial de la metabolòmica d'espectroscopia de ressonància magnètica nuclear (RMN) com a eina per a la detecció de nous biomarcadors de malalties que faciliten un diagnòstic precoç i no invasiu, així com un seguiment del pacient. L'espectroscopia de RMN s'aplica ací específicament a l'estudi de quatre malalties: enfermetat d'Alzheimer, glaucoma, aterosclerosi i vulnerabilitat de placa, i fibrosi pulmonar desenvolupada després de la pneumònia per COVID-19.

En la introducció es presenta una descripció del procés d'anàlisi metabolòmic per a la identificació de biomarcadors de malalties, així com de les principals plataformes emprades en metabolòmica. Posteriorment es profunditza en el funcionament de l'espectroscopia de RMN per a l'anàlisi metabolòmic, així com en les principals eines estadístiques utilitzades en aquesta tesi. Finalment, es descriuen les principals característiques fisiopatològiques de les malalties estudiades en aquesta tesi, així com les formes de diagnòstic actuals i la necessitat d'identificar nous marcadors de malalties.

A continuació, es descriu l'objectiu principal d'aquesta tesi, així com els objectius específics que s'aborden en els capítols experimentals.

En el primer capítol, s'identifiquen biomarcadors d'enfermetat d'Alzheimer (AD, de l'anglés *Alzheimer Disease*) i del progrés de pacients amb deteriorament cognitiu lleu MCI (de l'anglés *mild cognitive impairment*) a Alzheimer mitjançant metabolòmica per espectroscopia de RMN en mostres de sèrum. Per a això, en primer lloc, es produeixen models PLS-DA per a la discriminació entre pacients d'AD i MCI i controls amb cognició

normal (HC, de l'anglés *healthy controls*). Els models de classificació obtinguts per a discriminar AD davant MCI i HC van mostrar excel·lents paràmetres de classificació. No obstant això, la discriminació entre MCI i HC no va ser tan significativa, posant de manifest com el desenvolupament d'AD altera el metabolisme cel·lular. Posteriorment, per a la identificació de biomarcadors del procés d'evolució de MCI a AD, el grup MCI es va subdividir atenent a si en un període de 1 a 3 anys després de l'adquisició de la mostra, el seu estat cognitiu havia deteriorat a AD. Es va observar que la lisina, el piruvat, la colina, la fenilalanina i les lipoproteïnes, tenien uns valors de concentració diferent depenent si el pacient evolucionava a AD o es mantenia estable. A més, es van observar diferències en les rutes metabòliques entre estos grups de pacients.

En el segon capítol de la tesi, es van estudiar les diferències metabolòmiques entre pacients MCI i control d'un grup de pacients diferent, mitjançant l'anàlisi dels metabolits presents en mostres de plasma en espectroscopia de RMN, així com la presència de marcadors de peroxidació lipídica, que van ser analitzats mitjançant UPLC-MS/MS. El procés d'oxidació cel·lular s'ha vist estretament relacionat amb el desenvolupament de la malaltia d'Alzheimer, i de fet, en el treball presentat s'observa com afegir marcadors de peroxidació lipídica a la metabolòmica de RMN millora els models de classificació d'estadística multivariant per a la identificació de biomarcadors de MCI. En aquest treball es van realitzar models de classificació PLS-DA utilitzant els metabolits identificats per RMN, els marcadors de peroxidació lipídica detectats amb UPLC-MS/MS, i la combinació de tots dos, obtenint el resultat amb una major significació estadística quan es combinen ambdues tècniques. Els metabolits identificats com a potencials biomarcadors de MCI en plasma són isoleucina, valina, 3-hidroxibutirato, 2-hidroxi-3-metilvalerato, glutamat, glutatió, cisteïna, malonat, n-nitrosodimetilamina, taurina, prolina i isoprostans. Especialment, per la seua relació amb la patologia, es destaquen com a potencials biomarcadors la isoleucina, la valina, el glutamat, la taurina, la prolina, la cisteïna i els isoprostans.

En el tercer capítol de la tesi, es van analitzar llàgrimes de pacients de glaucoma primari d'angle obert (POAG) i controls mitjançant espectroscopia de RMN, per a la identificació de biomarcadors de la patologia en un mitjà mínimament invasiu, com són les llàgrimes. Els models de classificació obtinguts van permetre classificar les mostres de POAG i controls amb elevats valors de sensibilitat i especificitat i de l'anàlisi dels metabolits participants en el model de classificació es van determinar com a potencials biomarcadors de POAG en llàgrima la taurina, glicina, urea, glucosa, àcids grassos insaturats, fenilalanina, fenilacetat, leucina, compostos n-acetilats, l'àcid fòrmic i la uridina. De especial rellevància són la fenilalanina, glucosa, leucina, glicina i taurina per la seva relació amb la patologia.

A continuació, en el quart capítol es presenten els resultats obtinguts de l'anàlisi metabolómico mitjançant espectroscopía de RMN de mostres de plaques d'ateroma i sèrum de pacients amb estenosis de caròtide. Es van obtindre mostres de pacients simptomàtics (que havien presentat algun signe d'accident cerebrovascular) i asimptomàtics, a fi de determinar biomarcadores de vulnerabilitat de placa. De les anàlisis realitzades en placa, per mitjà de HRMAS, es van determinar com a potencials biomarcadores de vulnerabilitat el mio-inositol, glutamat, i àcids grassos insaturats, tenint especial rellevància el glutamat per la seua relació amb la patologia. La determinació de biomarcadores en sèrum resultava més interessant a nivell clínic, i de l'anàlisi estadística realitzada es van obtindre com a potencials marcadors de vulnerabilitat de placa la treonina, la histamina i els àcids grassos insaturats.

En el quint i últim capítol experimental es van analitzar per mitjà de espectroscopía de RMN mostres de sèrum de pacients que havien estat ingressats a l'hospital per pneumònia causada pel virus SARS-CoV-2, a fi d'identificar biomarcadores per a predir el desenvolupament de fibrosi pulmonar, una de les seqüeles més greus

desenvolupades després de la COVID-19. Els models de classificació PLS-DA generats van ser capaces de discriminar que pacients haurien desenvolupat fibrosi pulmonar un any després de l'alta hospitalària (i la presa de mostres) amb alts valors de sensibilitat i especificitat. Els metabòlits proposats com a potencials biomarcadores de desenvolupament de fibrosi, per la seua participació en el model i la presència de diferències significatives en la seua concentració relativa entre tots dos grups són glucosa, valina i àcids grassos.

Finalment, es presenta la discussió general i les principals conclusions generals de la tesi, així com les conclusions derivades del treball experimental ací presentat. Esperem que estos resultats òbriguen la porta a l'ús de la metabolómica mitjançant espectroscópia de RMN per a la identificació de biomarcadores primerencs i no invasius, i d'esta manera resoldre una de les principals necessitats de la medicina actual.

Publications

Results of this PhD thesis and other contributions have resulted in the following scientific publications

- Luis Pla; Félix Sancenón; M. Carmen Martínez-Bisbal; Ricardo Prat-Acín; Inmaculada Galeano-Senabre; **Marina Botello-Marabotto**; Sarai Palanca-Suela; Elena Aznar; Sara Santiago-Felipe; Ramón Martínez-Máñez. A gated material as immunosensor for in-tissue detection of IDH1-R132H mutation in gliomas. *Sensors and Actuators B: Chemical.* Volume 345 130406 (2021)

- Luis Pla; Félix Sancenón; M. Carmen Martínez-Bisbal; Celia Bañuls; Nuria Estañ; **Marina Botello-Marabotto**; Elena Aznar; Guillermo Sáez; Sara Santiago-Felipe; Ramón Martínez-Máñez. *Nanoscale*. Volume 13 (18), pages 8648-8657 (2021)

- **Marina Botello-Marabotto**; M. Carmen Martínez-Bisbal; Miguel Calero; Andrea Bernardos; Ana B. Pastor; Miguel Medina; Ramón Martínez-Máñez. Non-invasive boimarkers for mild cognitive impairment and Alzheimer's disease. *Neurobiology of disease*. Volume 187 106312 (2023)

- **Marina Botello-Marabotto**; M.Carmen Martinez-Bisbal; M. Dolores Pinazo-Durán; Ramón Martínez-Máñez. Tear metabolomics for the diagnosis of primary open-angle glaucoma. *Talanta.* Volume 273 125826 (2024)

- **Marina Botello-Marabotto**; Emma Plana; M.Carmen Martinez-Bisbal; Pilar medina; Andrea Bernardos; Ramón Martínez-Máñez; Manuel Miralles. Metabolomic study for the identification of symptomatic carotid plaque biomarkers. *Submitted*

- María Ovejero-Sánchez; **Marina Botello-Marabotto**; Carmen Peña-Bautista; Consuelo Cháfer-Pericás; M.Carmen Martínez-Bisbal; Andrea Bernardos; Miguel Baquero; Ramón Martínez-Máñez. Highly sensitive and specific approach to identify metabolites and pathways in early Alzheimer's disease. *Submitted*.

- **Marina Botello-Marabotto**; Julia Tarrasó; Alba Mulet; Lucía Fernández-Presa; Jaime Signes-Costa; Andrea Bernardos; M. Martínez-Bisbal; Ramón Martínez-Máñez. Early biomarkers of fibrosis in COVID-19 patients one year after hospital discharge. *Submitted*

Abbreviations and Acronyms

ACG	Angle closure glaucoma
AD	Alzheimer's Disease
ARDS	Acute respiratory distress syndrome
AUC	Area under the curve
Aß	Amyloid-ß
BMRB	Biological magnetic resonance data bank
CDR	Clinical dementia rating
CDT	Clock drawing test
CDU	Carotid duplex ultrasound
COSY	Correlation spectroscopy
CoV	Coronavirus
CPMG	Carr-Purcell-Meiboom-Gill
CSF	Cerebrospinal fluid
CT	Computed tomography
CTA	Computed tomography angiography
CV	Cross validation
CVD	Cardiovascular diseases
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ECM	Extracelular matrix
FID	Free induction decay
GA	Genetic algorithm
GC	Gas-chromathogrphy
HMDB	Human metabolome database
HRMAS	High resolution magic angle spinning
IHD	Ischemic heart disease
IOP	Intraocular pressure
IPF	Idiopathic pulmonary fibrosis

IS	Internal standard
KEGG	Kyoto encyclopedia of genes and genomes
LC	Liquid chromathography
LDL	Low density lipoprotein
MCI	Mild cognitive impairment
MMSE	Mini-mental state examination
MRA	MRI angiography
MRI	Magnetic resonance imaging
MS	Magnetic spectroscopy
NFT	Neurofibrillary tangles
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
OAG	Open angle glaucoma
PC	Principal component
PCA	Principal component analysis
PLS	Partial least square
PLSDA	Partial least square discriminant analysis
POAG	Primary open angle glaucoma
PPM	Parts per million
QC	Quality control
RGC	Retinal ganglion cells
RNS	Reactive nitrogen species
ROC	Receiver-Operating Characteristic
ROS	Ractive oxygen species
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SMPDB	Small molecule pathway database
TOCSY	Total correlation spectroscopy
TSP	3-trimethylsilyl propionate
UPLC	Ultraperformance liquid chromatography
VIP	Variable importance in projection
VSMC	Vascular smooth muscle cells

WHO   World health organization

General Introduction

1. Metabolomics as a source of biomarkers of disease

Biomarkers are defined by the World Health Organization (WHO) as "almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological. The measured response may be functional and physiological, biochemical at the cellular level, or a molecular interaction".[1]

For the diagnosis of disease, it has always been important to find reliable biomarkers to identify health problems and to provide adequate treatment when available. Precision medicine has emerged as the optimization of the current approach for a better classification of patients according to their specific conditions to improve diagnosis, prognosis, and response to treatment.

In the last decades, the development of the omic sciences has provided great amounts of information that has contributed to understanding biological processes associated with health and disease.[2] Genomics, transcriptomics, proteomics, and metabolomics are the main omic sciences, all of them being interconnected and showing different aspects of a living system.[3]

Where the genome could be considered the instruction book for all the processes occurring in a given organism, the transcriptome reflects the information that is actually being executed for a given tissue or cell type at a specific time. The transcripts provide the code for the generation of proteins, which ultimately catalyze the chemical reactions in which the metabolites are involved or produced (Figure 1).[2]
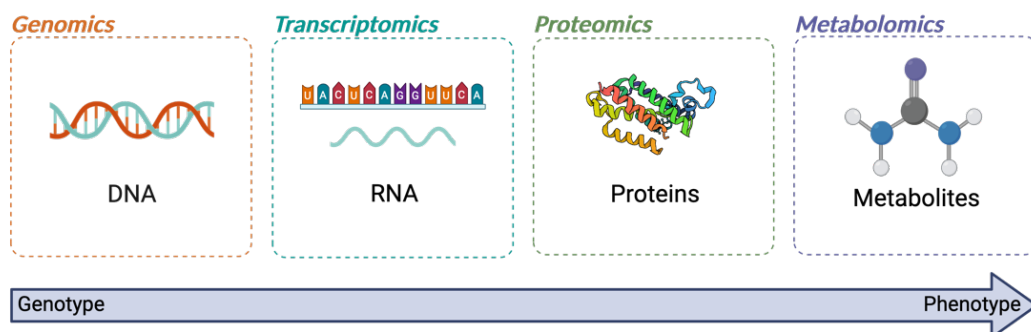
Figure 1. Schematic representation of the main "omic" sciences. Omic sciences are consecutively displayed from genotype to phenotype, being metabolomics the one that best can represent the macroscopic phenotype.

Metabolites are small molecules, of usually less than 1500 Da,[4] including amino acids, fatty acids, or sugars, among other organic compounds which act as intermediaries or end products of the cellular metabolism.[5]

The complete set of metabolites found in a specific biological sample is defined as the metabolome. The metabolome reflects the interplay between the alterations or changes produced in the expression of genes and protein activity and the influence of external factors, such as lifestyle, environment, stress, drug treatment, or different health conditions, being the omic science closest to phenotype.[6]

Metabolomics is the omic science determining the relative or absolute abundance of metabolites in a biological sample. For the relative quantitation, the normalized signal from the analytical instrument given by a specific metabolite is considered a measure of that metabolite in a specific sample, which can be compared to that cohort or batch. However, as the concentration value does not have a specific measuring unit,

comparing samples or groups from different studies is difficult.[7] For absolute quantification, calibration and internal standards (IS) are required.[7]

By measuring metabolic perturbations compared to the reference state (samples from healthy donors or controls), information of biological processes related to phenotype can be inferred. The use of metabolomics is then a good strategy to seek potential biomarkers of disease as well as to find new therapeutic targets.

1.1 Metabolomic Approaches

Metabolomic studies are usually classified as untargeted or targeted.[8] Targeted metabolomics are used when a defined set of metabolites is studied. Absolute quantification is easier in targeted analysis, and it is usually used when there is a pre-existing hypothesis about changes in a specific metabolic pathway under a given condition that needs to be validated.[9]

On the other hand, untargeted metabolomics aims to identify and measure as many metabolites as possible in a biological sample. It is usually used to determine the differences in the metabolome of groups of samples found under any differential condition (e.g., health and disease, treatment and placebo...), seeking for a specific pattern of metabolites that would allow to classify the groups. In this case more complex information is found, and usually, relative, more that absolute quantification is performed.[9] This approach is more associated with the discovery of biomarkers.[10–12]

1.2 The metabolomic workflow

The first step in an untargeted metabolomic study is to define the objectives. According to this, the type of sample to be analyzed and the size of the study is established, as

well as the analytical platform. There are two main techniques used for metabolomic analysis, i.e. nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).

MS is performed coupled with different chromatographic methods, including gas chromatography (GC-MS) and liquid chromatography (LC-MS). Taking this into consideration, the pre-processing of the sample and the data acquisition parameters are defined, according to the requirements of each platform, NMR or MS. After acquisition, the data needs to be preprocessed, and quality control (QC) is checked. For obtaining meaningful data from the metabolomic spectrum, the signals produced by the metabolites need to be translated or exported into a standard format. A numerical matrix is generated with the information obtained from each meaningful signal in the spectrum for every sample. Then, statistical data analysis together with metabolite identification is performed to obtain meaningful results. Currently, there are several databases available for the study of the metabolome. One of the most relevant for the study of human metabolites is the Human Metabolome Database (HMDB).[13] The HMDB contains information about all of the human metabolites identified up to date (more than 200.000) in different biofluids from urine or serum to tears, as well as tissues. Detailed information about each metabolite including its chemical structure, biological roles, physiological concentrations, among others, as well as reference MS/MS (tandem mass spectrometry), GC-MS and NMR spectra.[13]

Finally, the biological meaning of the obtained results and its relationship with the pathology, or condition under study, is established. Biological data bases such as Kyoto Encyclopedia of Genes and Genomes (KEGG)[14] or small molecule pathway database (SMPDB)[15] present wide information about metabolic pathways, which links the chemical reactions occurring between metabolites in a given biological process. A basic workflow of an untargeted metabolomic study is shown in figure 2.
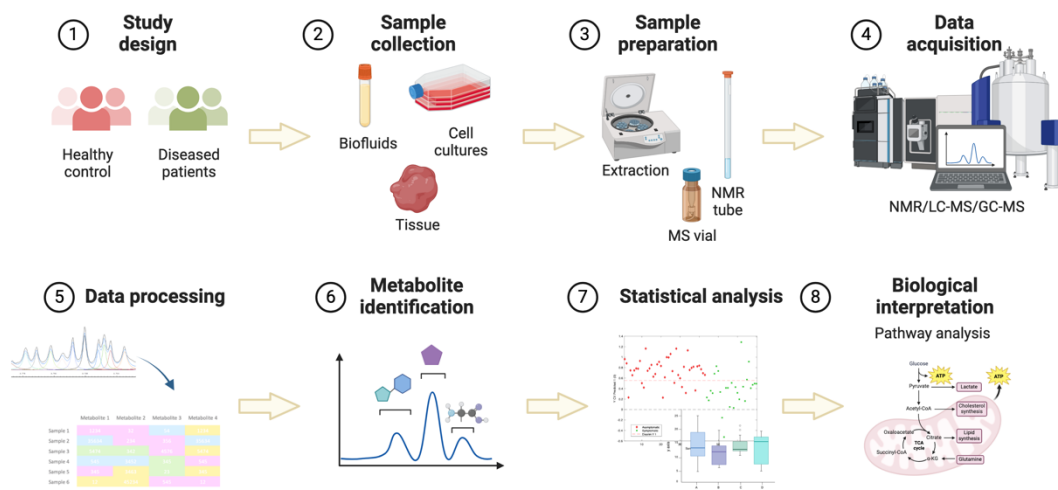
Figure 2. General workflow of a metabolomic study. The main steps of a metabolomic study for the identification of biomarkers of disease are shown.

1.3 Analytical platforms for metabolomics

The development of metabolomics in the last decade has been possible due to the advances in analytical technology, which has allowed the performance of high-throughput analysis.[16] Compound identification can be difficult due to the complexity of the biological matrixes, and none of the techniques allows a complete qualitative and quantitative description of the full metabolome. Still, through the application of these analytical tools it is possible to identify a wide range of metabolites present in biological samples such as serum[17] and plasma,[11] urine[10,18,19] or tears[20] among other biofluids, as well as tissues,[21–25] or cell cultures.[26]

The analytical platform of choice for a given metabolomic study depends on several factors, as both NMR spectroscopy and MS present different advantages and limitations. In fact, the use of both platforms can be complementary to obtain an increased coverage of the metabolome.[27]

As has been previously mentioned, separation techniques, mainly GC or LC, are used before MS. The chromatographic method is generally chosen depending on the matrix or the nature of the compounds of interest. In GC compounds are separated depending on their volatility,[28] whereas in LC, in most of the cases metabolites are separated according to their polarity (differentiating between polar and non-polar compounds).[29] Also, the advancements in analytical instrumentation have allowed the emergence of ultraperformance LC (UPLC), which improves the sensitivity, resolution and speed of LC.[30] Briefly, after separation, metabolites are sent into the MS equipment, which is composed of an ionizer, a mass analyzer, and a detector. Then, the MS identification of compounds is based on the ionization patterns of molecules. The mass over charge (m/z) of ions, as well as their relative quantities is used to identify the metabolites.[31]

For NMR analysis, the most common preparation consists of the addiction of a phosphate buffer for the regulation of the pH of the sample, and a deuterate solvent. The sample usually doesn't require any further preparation, while concentration or lyophilization can be done previously. Then, the prepared samples are subjected to NMR spectroscopy, where metabolites resonate at distinct frequencies based on their chemical environments. By measuring these resonances and processing the resulting spectra, metabolites can be identified and quantified.

MS has higher sensitivity than NMR, being able to identify compounds whose concentration is in the order of nanomolar (nM) or picomolar (pM) for some equipment. NMR can detect metabolites that are 10 or 100 times more concentrated. That also means that a higher number of compounds can be identified with MS.[32]

On the other hand, NMR spectroscopy is more reproducible than MS and does not require any preparation step prior to the analysis of the sample.[33] Furthermore, NMR is a non-destructive technique, so the samples can be further analyzed after NMR spectroscopy if needed. In contrast, MS is a destructive technique, but less volume of

sample is required for the analysis.[33] Advantages and disadvantages of NMR and MS for metabolomic studies are summarized in figure 3.

	Nuclear Magnetic Resonance (NMR)	Mass spectrometry (MS)
Sensitivity	Low (nanomolar)	High (picomolar)
Reproducibility	Very High	Average
Sample preparation	Minimal	More complex preparation reriquered
Samples recovery	Yes (non-destructive)	No (sample processing required)
Experimental time	Short time (1D) Long time (2D)	Longer times when coupled with chromatography
Experimental cost	Lower	Higher
Number of detectable metabolites	Lower (30-100 metabolites)	Higher (thousand of metabolites)
Tissue Extraction	Not-required (HRMAS*)	Reriquered
Targeted Analysis	Not optimal	Better than NMR

Figure 3. Comparison between NMR and MS for metabolomic studies. The figure shows the main differences between both techniques.[33]

In this PhD thesis, the technique used is NMR spectroscopy. Theoretical fundaments of the technique and the most relevant features are further exposed in the following sections.

1.4 Nuclear magnetic resonance spectroscopy

NMR spectroscopy is a fundamental tool in chemistry, having a wide range of applications, among which we can find structure elucidation, purity assessment, or compound quantification. Beyond chemistry, other disciplines such as food, pharmacy,

biology, medicine, and materials also use this technique to support their studies. Moreover, this technique is a powerful tool for metabolomic studies.[34]

1.4.1 NMR spectroscopy principles

Spectroscopy, as a broad discipline, focuses on how matter interacts with electromagnetic radiation. Atoms and molecules possess distinct energy levels associated with various electronic, vibrational, or rotational states. This interaction involves the absorption and emission of photons by atoms. Spectroscopic techniques vary based on the frequencies employed. For instance, NMR spectroscopy utilizes radio frequencies typically ranging from 10 to 800 MHz.[35]

NMR is the study of the magnetic properties (and energies) of nuclei. The absorption and emission of electromagnetic radiation can be observed when the nuclei are placed in a strong external magnetic field. The principles of NMR spectroscopy are based on the concept of nuclear spin.[36] The spin is the nuclear angular momentum that makes nuclei act like magnets. The nuclear spin is the resultant of the spin and orbital angular moment of the neutrons and protons that form the nucleus. The neutron and the proton each have a spin quantum number 1/2. A nucleus with an odd mass number A has a half-integral value (1/2, 3/2, etc.). A nucleus with even mass number A has an integral value of I.[37] Nuclear spin determines the energy levels that a nucleus can take under a magnetic field. Depending on the nature of the nucleus the interaction of the atom with an external magnetic field (B_0) would be different. NMR principally studies nuclei with a spin of 1/2 (such us, ^1H, ^{15}N, ^{13}C, ^{19}F, ^{31}P).[38] A spin of 1/2 means that under an applied magnetic field it has two energy levels (Figure 4), depending on the spin's orientation:

1) $+ \frac{1}{2}$: a low energy state (α), that is produced when the spins are parallel with the applied magnetic field

2) $- \frac{1}{2}$: A high energy state (ß), that is produced when the spins are antiparallel to B_0

As in any other physic system, the nuclei tend to be in the ground energy state. In the equilibrium, there would be nuclei associated to one spin state or the other, causing the macroscopic magnetization in the sample.[38] The application of a radiofrequency pulse perpendicular to the magnetic field produce the transition of the nuclei between both stages. When the pulse finishes the nuclei come back to its basal state (relaxation), releasing the excess energy, that can be detected as a free induction decay (FID) (Figure 4).
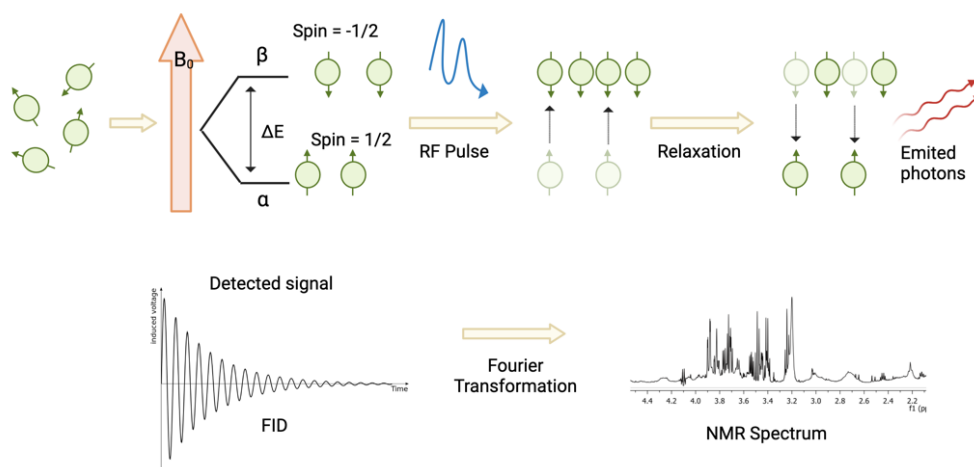
Figure 4. Schematic representation of the fundaments of ^{1}H-NMR spectroscopy. Nuclei are represented as green circles. The arrows indicate the orientation of the spin.

Then, the application of the Fourier transform allows to convert this signal into a spectrum, transforming the time domain into the frequency domain. The resultant spectrum is formed by the x-axis, with the frequency (ω), and a y-axis with the intensity of the signals (Figure 4).[36]

The NMR signal (frequency ω) produced by nuclei is influenced by the chemical environment of the nuclei. All nuclei are surrounded by an electron cloud, that shields it from the external electromagnetic field B_0, generating a small magnetic field B_1, generally in opposition to B_0. The shape of the cloud depends on the electronegativity

of the neighboring atoms, so it is specific for each chemical compound.[38] The frequency shift caused by the shielding of the electron clouds is called chemical shift (δ). The chemical shift is normalized to the magnetic field, being measured in part per million (ppm). The extent of this chemical shift is dependent on the electronic environment around the nucleus, which is influenced by the electronegativity of neighboring atoms and the overall molecular structure.

For the identification of metabolites in a spectrum, spin-spin coupling (J) is also important that is produced by two or more adjacent protons in a molecule. This produces a splitting in the signal called multiplicity.[38] The multiplicity is produced following rule n+1, where n is the number of nuclei of the same type in the neighboring environments (Figure 5).

These two characteristics allow the identification of the molecules present in a sample. In complex matrixes, such as biological samples, the signals overlap due to the presence of hundreds of metabolites, making the identification of metabolites challenging. Furthermore, the intensity of the signals or the area under the peak can be related to the concentration of a given metabolite in the samples.[39]
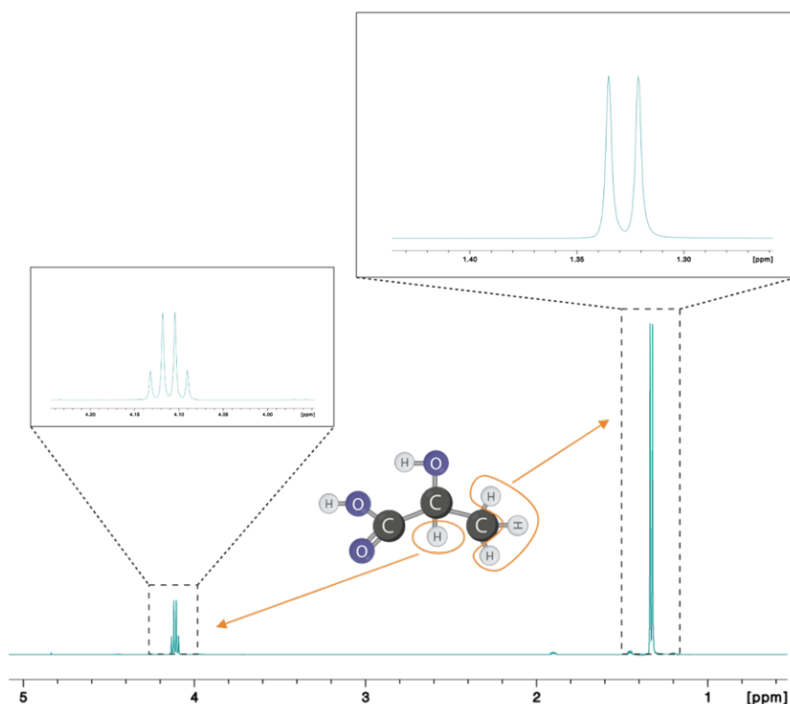
Figure 5. ¹H-NMR spectra of lactate. Lactate molecule and its NMR spectrum are displayed. The orange arrows indicate which ¹H group is responsible for the generation of the signal. Lactate spectrum was downloaded from Biological Magnetic Resonance Data Bank (BMRB).

Once acquired, spectra need to be pre-processed and quality controlled. The preprocessing of NMR spectra usually consists on the application of a windowing function (usually exponential line broadening), baseline correction and phase adjustment, as well as chemical shift calibration. The chemical shift can be referenced to the shift of an already known metabolite, or through the addition of reference compounds, such as 3-trimethylsilyl propionate (TSP) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

1.4.2 ^{1}H-NMR experiments

Proton NMR (^{1}H-NMR) is the most extended for metabolomic studies, due to the high abundance of ^{1}H in biomolecules, which allows its identification and quantification. As has been previously mentioned, radiofrequency pulses are required to obtain NMR spectra. A pulse sequence is a series of pulses designed to acquire specific types of information from the sample. The most commonly used pulse sequences for ^{1}H-NMR experiments in metabolomics are monodimensional versions of Nuclear Overhauser Effect Spectroscopy (NOESY)[40] and Carr-Purcell-Meiboom-Gill (CPMG) sequences.[41]

The most prominent signal in the spectra of samples with biomedical interest is the solvent (water), given that the samples are in an aqueous environment. To observe the signals of the metabolites, it is necessary to eliminate as much as possible the water signal. There are different ways to selectively suppress this signal. The 1D NOESY sequence allows an efficient suppression of the water signal with a resulting flat baseline close to the residual water signal. It is a highly reproducible method for the general acquisition of spectra from complex samples containing metabolites (small and large biomolecules).[40]

Usually, the samples of biomedical interest contain molecules of high weight and with restrictions in relaxation that are in the origin of wide and low resolved signals, that compromise the baseline of the spectra. When a better identification of low molecular weight molecules is required, like in serum and plasma samples, the CPMG sequence is principally applied.[41] It is based on the different relaxation properties of the nuclei depending of the size of the molecules allowing the elimination of wide signals.[41]

As it has been mentioned previously, one of the limitations in metabolite identification of ^{1}H-NMR is the overlapping of signals due to the complexity of biological samples. However, this problem can be in part solved with 2D experiments of correlation homonuclear (^{1}H-^{1}H) or heteronuclear (^{1}H-^{13}C for instance) which can be acquired

thanks to the possibility of observing multiple nuclei during a single NMR experiment.[42] In these experiments, the use of two dimensions enables the separation of peaks, that in 1D experiments would be overlapping, providing further information on the chemical structure of the molecule.[42] Most used 2D NMR experiments include homo-nuclear experiments (^{1}H-^{1}H), such as ^{1}H J-Resolved, COSY (COrrelation SpectroscopY) [42] and TOCSY (TOtal Correlation Spectroscopy), and heteronuclear experiments like HSQC (Heteronuclear Single Quantum Coherence) that can be acquired correlating signals of different nuclei as ^{1}H-^{13}C, ^{1}H-^{15}N or ^{1}H-^{31}P.[43,44]

1.4.3 HRMAS NMR spectroscopy

Metabolomics by NMR spectroscopy can also be applied in intact semisolids or semifluids, such as tissue biopsies or culture cells thanks to the application of high-resolution magic angle spinning (HRMAS) NMR.[45] This technique allows the acquisition of high-resolution spectra of such semisolid samples such as intact tissue to evaluate the metabolic composition.[46]

In tissue there are limitations in the mobility of molecules, that are not found in liquids, which increase the proton dipolar interactions. This restriction together with the heterogeneity in magnetic susceptibility, results in broad lines of the NMR spectra.[46] To reduce the line broadening, the samples are mechanically rotated at speeds faster than the spectral broadening around an axis inclined 54º44' (the magic angle), which decreases the effect of the interactions. Using HRMAS NMR spectroscopy the resolution of the tissue spectra is close to the resolution found in ^{1}H-NMR of tissue extracts.[46]

1.5 Statistical analysis

As it has been previously mentioned, after the data acquisition and pre-processing of the spectra, the signals are integrated, and a complex matrix of observations (samples) and variables (peaks) is formed. To identify the relevant information and to interpret the data, the application of statistical methods is required.

The first steps are the normalization and the scaling of the data. Normalization allows to reduce the differences between samples produced by the nature of the sample (for example, urine samples could be more concentered or diluted prior to the analysis). The most common strategies for normalization are the normalization to the total area of the spectrum or to the area of a determined peak. The scaling is used to minimize the impact that bigger or more intense signals (which have higher values due to the characteristics of the analytical technique) have in the analysis. Scaling allows to prevent the presence of signals with high intensity from overshadowing or concealing the presence of other metabolites with lower signal intensity.[47] In the works presented in this thesis, the scaling used is autoscaling, in which every variable is mean centered and divided by the standard deviation. Once the matrix of data is prepared, different statistical analyses can be performed.

Univariate statistics, such as t-test or Mann-Withney U test (non- parametric version of t-test) allows to compare the specific value of a metabolite among groups. However, the current consensus suggests that the entire metabolic profile is more sensitive in identifying and characterizing disease compared to the study of single metabolites.[48] Bearing this in mind, the use of techniques able to provide a holistic picture of the samples is highly appealing and may find applications in the development of tools for diagnosis. In this context, given the high number of variables obtained from a metabolomic analysis, and the interconnections between the different metabolites that can provide redundant information, more potent statistics methods (multivariate

statistics) are generally used for the identification of potential biomarkers, and the generation of predictive models.

1.5.1 Multivariate analysis

The application of multivariate statistics allows the reduction of the dimensionality of the data, considering all the metabolomic features simultaneously, and identifying relationship patterns among them. The multivariate analysis can be unsupervised, which do not require information about type of sample (e.g., health and diseased), and supervised methods which are "directed" to discriminate the samples according to class, and therefore highlighting the metabolomic differences between the groups.

1.5.1.1 Unsupervised data analysis (PCA)

Principal component analysis (PCA) is the most widely used unsupervised multivariate tool employed in omic analyses.[49] It is commonly used to determine whether the samples in a batch are grouped based on the characteristics under analysis, in our case the metabolome.[49]

When PCA is applied, the information in a data set is reorganized in new uncorrelated variables called "Principal Components" (PCs), which are calculated as linear combinations of the original variables representing most of the information in the data.[50] With this method, the first component explains the maximum variance in the data and the subsequent components explain consecutive reduced amounts of variance.[50] After performing a PCA, every observation in a dataset is described by a set of scores according to the principal components. These scores indicate the position of

the data point in the new coordinate system created by those principal components (Figure 6).[50]
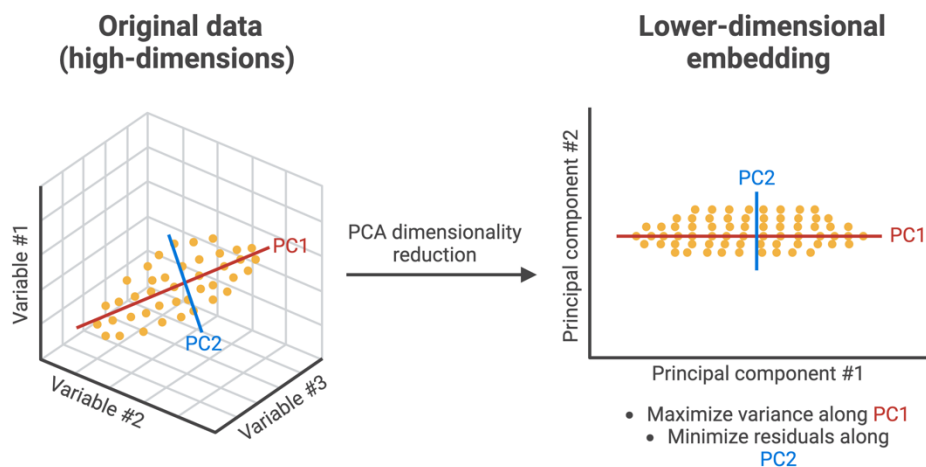
Figure 6. Schematic representation of PCA dimensionality reduction. Dimensionality is reduced and a new system of variables that maintain most of the information is created, in such a way that similar samples are clustered together. (Image from Nashed, M. (2023) BioRender).

In a metabolomic analysis, several factors can influence the final composition of the samples, and therefore commonly it is not possible to observe a structure on the data after PCA, or the structure formed does not represent the groups under study. To find patterns associated to groups a supervised analysis can be performed.

1.5.1.2 Supervised data analysis (PLSDA)

Supervised methods use the information about the groups present in the dataset to discover variable patterns associated to such groups.[51] There are several supervised methods for the identification of discriminant variables between groups, such as support vector machine,[52] artificial neural networks[53] or partial least squares discriminant analysis (PLS-DA).[54] PLS-DA has become the most popular algorithm used for metabolomic studies,[54] and it is the algorithm used in the works presented in this thesis.

16

Partial least square (PLS) is a statistical tool that allows the generation of predictive models to find the relation between two matrices, a matrix of independent variables (predictors) and a secondary matrix of dependent variables (response). PLS creates new variables, in a similar way to PCA, called latent variables but instead of looking for maximum variance, latent variables (LV) are used for the generation of a regression model between predictors and response variables.[55]

PLS-DA is a variation of PLS used when the response variable is categorical (i.e., health and disease). Therefore, it is useful for the generation of discriminative predictive models. The feature coefficients (loadings) represent the relative contribution of each predictor variable to the discrimination.[55]

Once the predictive model is generated, its accuracy and predictive capacity (sensitivity and specificity) are studied. Sensitivity is defined as the rate of true positives divided by all the positive samples (true positive and false negative). Specificity is defined as the rate of true negatives divided by all the negative samples (true negatives and false positives).[56] An interesting method for the representation of the sensitivity and specificity of a model, and to determine its "goodness" is the Receiver-Operating Characteristic (ROC) curve. In the ROC curve, sensitivity is displayed on the y-axis and 1-specificity on the x-axis for different threshold points in the testing of samples (Figure 7). As a result, the area under the curve (AUC) is a combined measure of sensitivity and specificity useful for the determination of the prediction capacity of the model.[56]
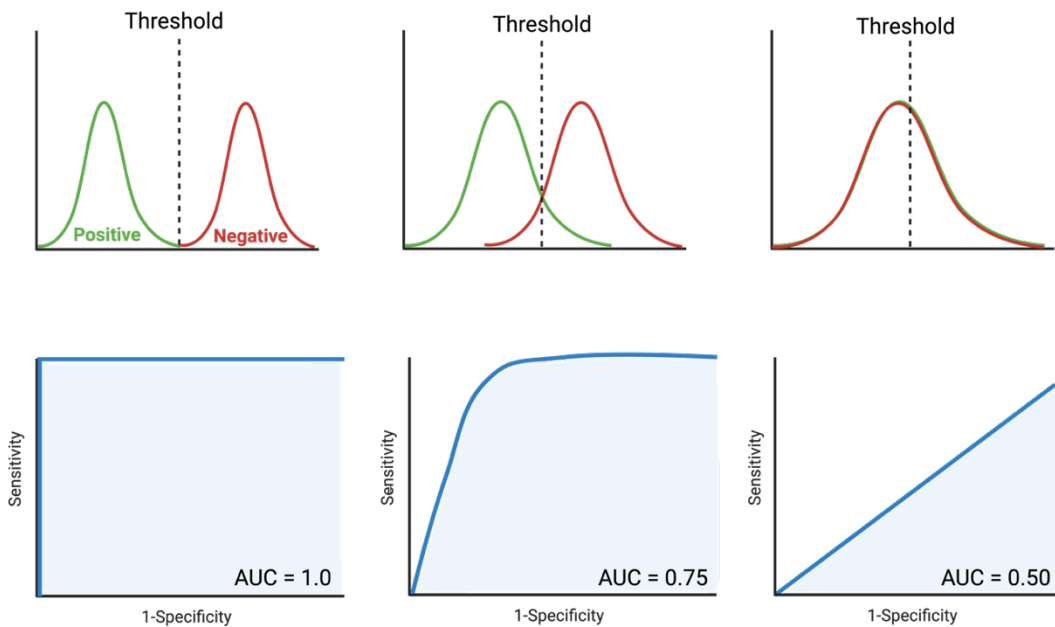
Figure 7. Schematic representation of ROC curves and distribution of positive and negative samples under a classification model. With an ideal classifier, positive and negative samples can always be distinguished (left curve). Most of the time, there is an overlap in the classification of positive and negative samples (middle curve). When the classifier is not able to separate samples by class, the classification is produced by chance (right curve).

An important step when evaluating a classifier is to check for overfitting.[57] Predictive models are trained prior to prediction, and with the use of too many latent variables, the model could be too suited for the current set of data but would not work for a different group of data. To ensure that the model is not overfitted, different strategies can be followed. In an ideal scenario, samples can be divided into a calibration group for training and a validation group for testing. However, if the available cohort is not big enough, cross validation (CV) and permutation test can be used to determine whether the model is overfitted.[54]

CV determines the predictive capacity of a classifier or model using an iterative approach. At each round of CV, the dataset is split into different subsets (folds). The

model is trained on different combinations of these subsets, and a remaining subset (validation set) is used to evaluate the performance of the model. This process is repeated until all the samples have been used for validation. The average of these results provides an estimate of the performance and reliability of the predictive model.[54] After CV, permutation test can be performed to determine the significance of the obtained predictive results. In the permutation test, the group labels are randomly assigned to the samples, and the CV is performed. This process is repeated hundreds to thousands of times, to generate a distribution of tests to assess the statistical significance. The null hypothesis would be that there is not difference between the results obtained when group labels are randomly assigned and the actual class of the samples. If the p-value is lower than the established threshold (usually 0.05) the null hypothesis would be rejected, and the model could be considered significant.[58]

Variable selection methods

Variable selection is a crucial step for the generation of predictive models in metabolomic studies. Different strategies can be followed to select the most representative variables of a PLS-DA model and refine the prediction results. In the works presented in this thesis, mainly two methods have been used for variable selection: variable importance in projection (VIP) scores and genetic algorithms (GA).

Variable Importance in Projection (VIP)

VIP scores indicate the importance of each variable in the prediction capacity of the PLS-DA model. It is calculated according to the weighted contribution of each variable to the model's predictive performance across its latent variables. When a variable has a higher VIP value, it means that it contributes more to the discrimination between classes generated by the model.[59] The refining of the model, by the only inclusion of variables with VIP scores above a given value, (e.g., VIP > 1) often improves CV performance and significance (while decreasing overfitting).

Genetic Algorithms (GA)

Genetic algorithms (GA) is an optimization technique designed for the selection of variables. It is based on the concepts of mutation, natural selection, and evolution.[60] Variables are represented as genes in a chromosome, each chromosome being an initial random distribution of a subset of variables, that will be used for the prediction model. Fitness of the model is assessed, and in an iterative process, mutation and recombination (change of variables) and natural selection (progressive selection of the best performing model) is produced for a predefined number of generations, or until a certain fitness threshold is achieved. In the end, the group of variables that give the best prediction in the PLS-DA model is selected.[60]

Based on the use of NMR-based metabolomics as a source of biomarkers of disease, different studies have been developed during this PhD thesis. We have sought an improvement in the diagnosis of four diseases: Alzheimer's disease, primary open angle glaucoma, atherosclerosis and plaque vulnerability, and pulmonary fibrosis development after COVID-19 infection. The main pathological characteristics and current diagnosis procedures of these diseases are further discussed in the following sections.

2. Alzheimer's disease

Dementia is a general term used to define a group of pathological symptoms produced in several age-related brain diseases. It is characterized by a progressive cognitive decline which causes an impairment with everyday activities, that is not produced with normal aging. Main dementia symptoms include memory and language problems, problem-solving disfunction, and difficulties in other cognitive skills, as well as behavioral problems.[61] With the increased aging of the population, the development of dementia and its impact in society and economy is also rising. In 2019, about 55

million people were affected by dementia, a figure expected to rise to 139 million by 2050.[62]

Dementia has different potential causes, Alzheimer's disease (AD) being the most prevalent cause. AD is a progressive and irreversible neurodegenerative disease, responsible of 60-80% of dementia cases.[63] It is a complex multifactorial disorder caused by the interaction of genetic susceptibility and environmental factors, being aging the principal risk factor, with 10% individuals older than 65 having AD.[64]

2.1 Etiology of the disease

Alzheimer's disease was first described in 1906 by the German psychiatrist Alois Alzheimer. In 1901, Alzheimer observed a patient named Auguste D. with a progressive loss of cognitive capacity. When she died, Alzheimer observed in a histological analysis of her brain the formation of plaques in the cortex, which today we know is amyloid-β (Aβ) accumulation, and the formation of neurofibrils, which we now understand are caused by the accumulation of hyperphosphorylated Tau protein.[65] More than 100 years after, the formation of Aß plaques and neurofibrillary tangles (NFT) are still the main pathological hallmarks of Alzheimer's disease.[66]

Amyloid-ß is produced from the cleavage of amyloid protein precursor (APP), which is a transmembrane protein related to membrane-limited organelles. The accumulation of Aß in the cerebral cortex is related to an increased neurotoxicity.[67]

On the other hand, Tau is a protein which associates with the microtubules to increase its instability and allow to form a stable network of microtubules and hold them together. Due to the accumulation of Aß, Tau protein is hyperphosphorylated, and oligomerized, which results in the formation of the NFT (Figure 8). Neurofibrillary tangles, together with Aß-plaques generate an inflammatory environment in the brain and produce an impairment in synaptic signaling.[67] These changes lead to clinical manifestations that go from memory lapses to severe and debilitating loss of memory and cognitive function.[66]
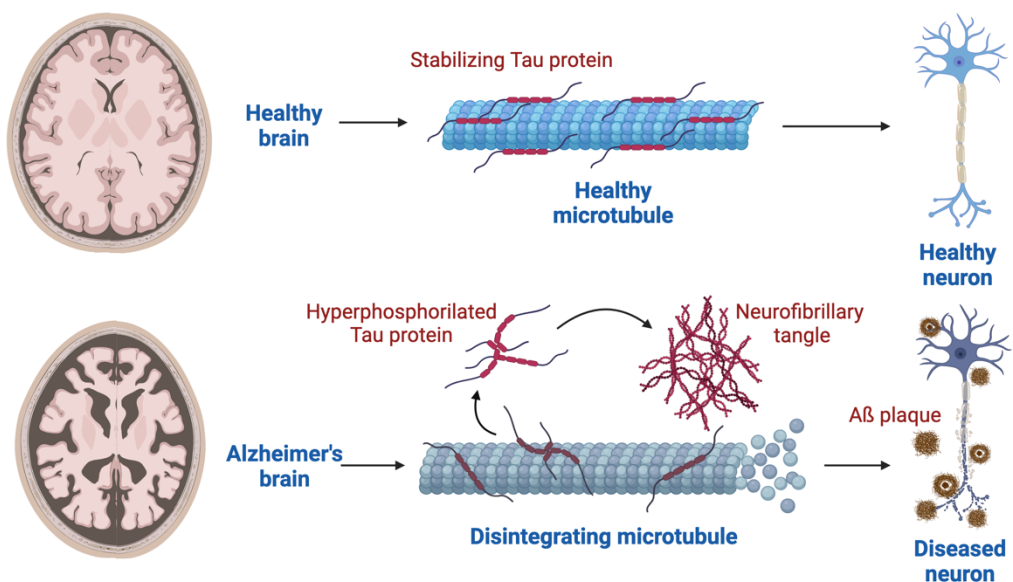
Figure 8. Formation of NFT tangles and Aß plaques in AD brain. The top scheme represents healthy brains with the presence of Tau protein stabilizing microtubules, resulting in healthy neurons. The bottom scheme shows an AD brain, with NFT and Aß plaques resulting in a diseased neuron with an impairment in the synapsis ability.

Although AD is not a hereditary disease *per se*, several genetic factors have been associated with the development of the disease. There are more than 80 risk and protective genes for AD, most of them with small effects when evaluated individually. The apolipoprotein E (APOE) gene is the most important risk gene for AD.[68] One of the three major allelic variants (*APOE*ε4*) is related to an increased risk and lower age of

AD onset. The effect that it has on the onset of the disease might be related to the inhibition of Aß clearance and promotion of Aß aggregation.[69]

2.2 The course of Alzheimer's disease

AD is a chronic irreversible illness that can last between 8-10 years. However, it is preceded by preclinical and prodromal stages, which can begin up to 20 years before the onset of clinical symptoms.[70] AD patients experience a gradual onset of the disease, with a slow, but progressive, decline in most of their cortical functions.[71]

There are three principal stages in AD, with a progressive loss of cognitive capacities. In the first preclinical stage of AD, there are no clinical symptoms, but the aforementioned pathological changes in the brain, including the accumulation of Aß and hyperphosphorylated Tau protein, start occurring. Then, Mild Cognitive Impairment (MCI) is used to define those patients who do not have dementia but have some deficits in cognition. It could be defined as an intermediate stage between normal aging and dementia. People with MCI might have memory lapses and changes in cognition that can be measured.[66] Finally, the progression to dementia can be divided into different substages (mild, moderate and severe dementia) according to the sequential loss of capacities.[68] In this point it is also worth mentioning that not all MCI patients will progress to dementia, as MCI might not be caused by AD. In some cases, cognitive decline in MCI patients can be reverted or remain cognitively stable.[72] Progressive deterioration in people with AD might vary according to the elements affected in cognition. First, word recall and orientation in time are affected. Then a progressive decline in language, attention and concentration, and finally, impairment in constructional praxis, orientation in place and immediate memory deterioration are observed in severe AD dementia.[73] Changes in personality are as well observed (Figure 9).[74]
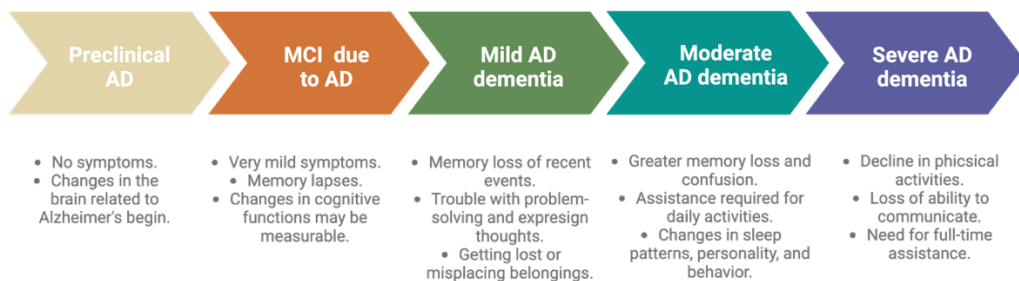
Figure 9. Schematic representation of progressive cognitive impairments in AD. The main stages in the course of Alzheimer's disease and the limitations produced in each stage are represented.[75]

2.3 Diagnosis criteria

The diagnosis of AD involves the combination of different methods, including cognitive testing, physical examination, neuroimaging, and Aß and Tau proteins determination.

First, cognitive tests are performed to determine the stage of dementia of the patients. Among the most popular cognitive tests are the Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), and Clock Drawing Test (CDT). The MMSE is a brief test consisting of 30 items with a total score that ranges from normal (30 points) to severe impairment (0 points). The questions are grouped into seven categories: orientation in time, orientation in place, registration, attention and concentration, recall, language, and drawing. The sub scores obtained in each category provide information about AD progression.[73] In the CDR test, six functional domains are examined (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care), with scores ranging from 0 to 3, representing normal aging (0) to severe dementia (3).[76] In the Clock Drawing Test, the level of dementia is established based on the evaluation of the patient's drawing of a clock (Figure 10).[77]
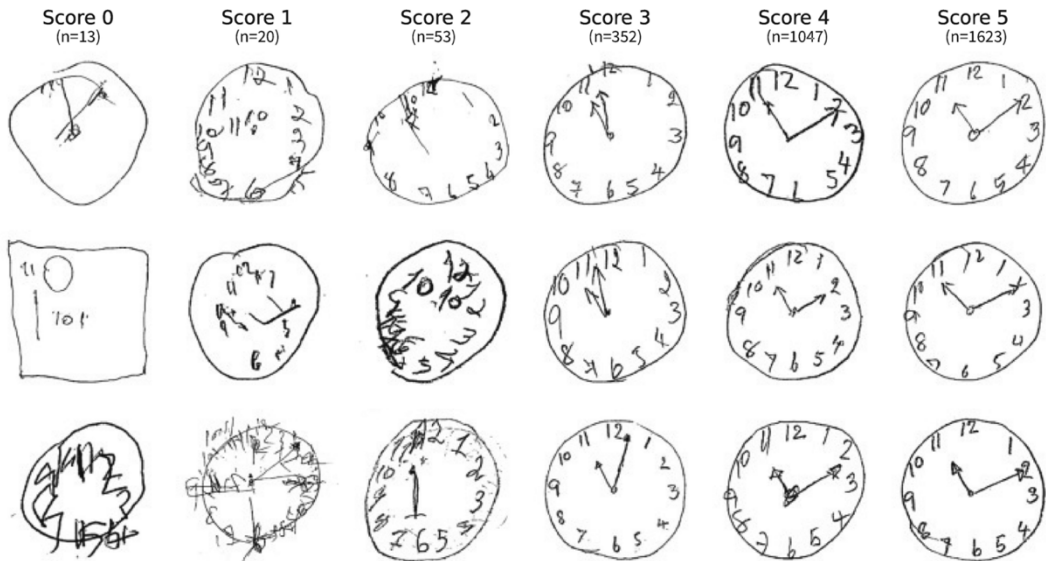
Figure 10. Example of clock drawing images. Scores range from 0 to 5, according to the level of dementia. The number of images in each score is shown below the score label (Figure from *Raksasat et al., 2013.*[77])

In addition to clinical diagnosis, molecular biomarkers such as Aß, total Tau protein and phosphorylated Tau protein can be measured in cerebrospinal fluid (CSF), and recently also in blood.[78] However, the lumbar puncture for CSF is an invasive procedure, and the measurement in blood is not validated for clinical use yet.[68] Neuroimaging techniques are also used for the diagnosis of AD. Position emission tomography (PET) allows to study not only the concentration, but also the spatial distribution of Aß and Tau protein. However, the costs and infrastructural requirements limit its use. On the other hand, magnetic resonance imaging (MRI) is also useful for the diagnosis of AD, as allows the identification of structural changes in the brain and vascular pathology, but it is not specific for AD.[68] Altogether, these diagnosis methods do not capture the full complexity of AD pathophysiology (as they only account for the presence of Aß, Tau, and NFT), ignoring processes also occurring in AD, such as inflammation and synaptic loss. Furthermore, the clinical symptomatology in most patients is produced by co-existing pathologies like cerebrovascular diseases, and these techniques do not account

for the presence of mixed pathologies.[68] The availability of molecular, non-invasive biomarkers would enable more effective diagnosis, patient stratification, treatment selection, and follow-up.

2.4 The need for new biomarkers. Metabolism and AD

As it has been previously mentioned, the clinical diagnosis of AD usually occurs decades after the initiation of the pathological processes. Currently, there are more than 187 clinical trials conducted searching for an effective treatment of AD, most of them being disease-modifying therapies.[79] However, therapy that begins when clinical symptoms appear is not effective, as it may be too late to reverse the damage already done. Efforts are being focused in finding biomarkers to determine which individuals would develop AD, time before pathological irreversible changes occur, and to develop new therapeutical interventions to slow down or to stop its progression.

The identification of biomarkers of impaired metabolism in AD has previously been explored. AD has a wide effect on brain metabolism, affecting different pathways and functions.

Oxidative stress is a well-known characteristic of AD, contributing to its pathogenesis and progression.[80] Oxidative stress is produced due to the generation of reactive oxygen species (ROS) and reactive nitrogen species (RNS), which react with biomolecules producing oxidative damage that can be measured by the concentration of biomarkers of oxidative damage.[80] Increased concentration in biomarkers of lipids (i.e. peroxidized lipids and isoprostanes),[81] proteins (i.e. 3-Nitrotyrosine),[82] and nucleic acids (i.e. 8-Hydroxy-deoxyguanosine and 8-Hydroxyguanine)[83] in patients with AD have been reported. This oxidative environment damages the synaptic functions related to learning and memory observed in AD patients.

Furthermore, alterations in the energetic metabolism have been observed in the development of AD.[80] For example, glucose metabolism becomes dysfunctional in AD brain due to mitochondrial impairment. Glucose hypometabolism has been observed in the brains of AD patients, suggesting that hypometabolism could be occurring before the onset of clinical symptoms.

Altered lipid metabolism has also been related to AD. As previously mentioned, *APOE* is the principal risk gene for the development of AD, which encodes an apolipoprotein. Furthermore, high cholesterol levels in the brain have been associated with AD. Cholesterol is involved in the maintenance of neuronal functions in the brain, such as neurotransmitter release and synaptic plasticity.[84,85] However, an excess of cholesterol in the brain can enhance the release of Aß.[86]

Further analysis of previous metabolomic studies performed for the identification of biomarkers is exposed in Chapter 1. In this thesis, two studies (in Chapters 1 and 2) have been carried out, (i) one studies the potential of metabolomics for the identification of early biomarkers of MCI and AD, and potential indicators of the progression of MCI to AD, (ii) the other studies oxidative stress biomarkers of MCI patients, obtained by UPLC-MS in combination with NMR-based metabolomics.

3. Primary Open Angle Glaucoma

Glaucoma is a term used to englobe a group of optic neuropathies, characterized by the progressive loss of retinal ganglion cells (RGC) together with supportive glia and vasculature. This produces morphological changes within the retinal nerve fiber layer and the optic nerve head.[87] The disease is usually bilateral, but asymmetry can be seen, depending on the etiology of the disease.[88] Glaucoma affects more than 70 million people worldwide and it is estimated that 111.8 million by 2040 will develop the disease. It is the leading cause of irreversible blindness and the second cause of blindness worldwide, affecting 3.5% of individuals aged between 40 and 80, a percentage increased to 10% for people older than 90.[89]

3.1 Etiology of the disease

There are two main types of glaucoma, according to its underlying anatomy and pathophysiology: open angle glaucoma (OAG) and angle closure (ACG). Both can then be subdivided into primary and secondary, depending on if it has an undelaying known cause (secondary), or not (primary).[90]

The anterior segment of the eye has its own circulatory system, which maintains the crystalline nerves and the cornea, both of which lack a blood supply. The ciliary body, a structure located behind the iris, produces the aqueous humor which circulates through the pupil into the anterior chamber where it nourishes the avascular structures of the eye, and drains out through the trabecular meshwork, which is a structure in the iridocorneal angle, where the iris and the cornea meet.[91] Primary angle closure glaucoma (PACG) is produced by disorders of the iris, lenses, and retro-lenticular structures which narrow the angle between the iris and the cornea, which leads to an increase in the intraocular pressure (IOP) as the aqueous humor cannot be correctly

drained. On the other hand, in primary open angle glaucoma (POAG) the angle between the cornea and the iris remains open, but the aqueous humor is not correctly drained due to an increased resistance through the trabecular meshwork (Figure 11).[92]
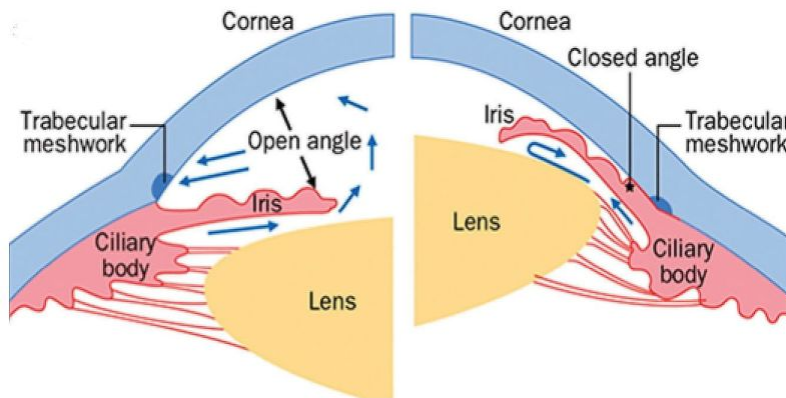
Figure 11. Schematic diagram of open angle and angle closure glaucoma. The anterior eye in open angle (left) and angle closure (right) glaucoma is shown. It can be appreciated how the displacement of the iris and lens obstructs the fluid drainage through the trabecular meshwork. (Figure from *Krizaj*, 2019.[92])

This inefficient drainage of the aqueous humor elevates the IOP which is related to the progressive apoptosis of the retinal ganglion cells (RGC).[93] The retinal ganglion cells are the neurons that receive the signals from the photoreceptors, process them, and transmit them to the brain through their axons which form the optic nerve.[94] Structural changes in the eye and impaired transport in the optic nerve are produced when there is an increase in IOP, low perfusion pressure (the pressure difference between the blood pressure and the intraocular pressure), or low CSF pressure.[94] The loss of RGC leads to a progressive impairment of the visual field, beginning from the periphery, and progressing until only a central or peripheral island of intact vision remains.[94]

POAG is the most prevalent form of glaucoma, and it has been associated with an increased IOP. However, although high IOP is a risk factor for the development of glaucoma, most patients with high IOP will not develop POAG, and conversely, not all

POAG patients have high IOP.[90] This suggests that factors other than IOP are important in the development and evolution of POAG, especially in those patients with statistically normal IOP levels (21 mmHg or less), known as normal tension glaucoma.[90] Furthermore, the mechanisms leading to the increase in intraocular pressure in POAG are not fully understood.[94]

The pathogenesis of POAG remains unclear, and the combination of different factors and systemic conditions is believed to be involved in the onset and progression of the disease. Among the potential risk factors of the disease are autoimmune dysfunction,[95] oxidative stress,[96] systemic and ocular vascular factors,[97] as well as sex, age, and ethnicity (Figure 12).[98]
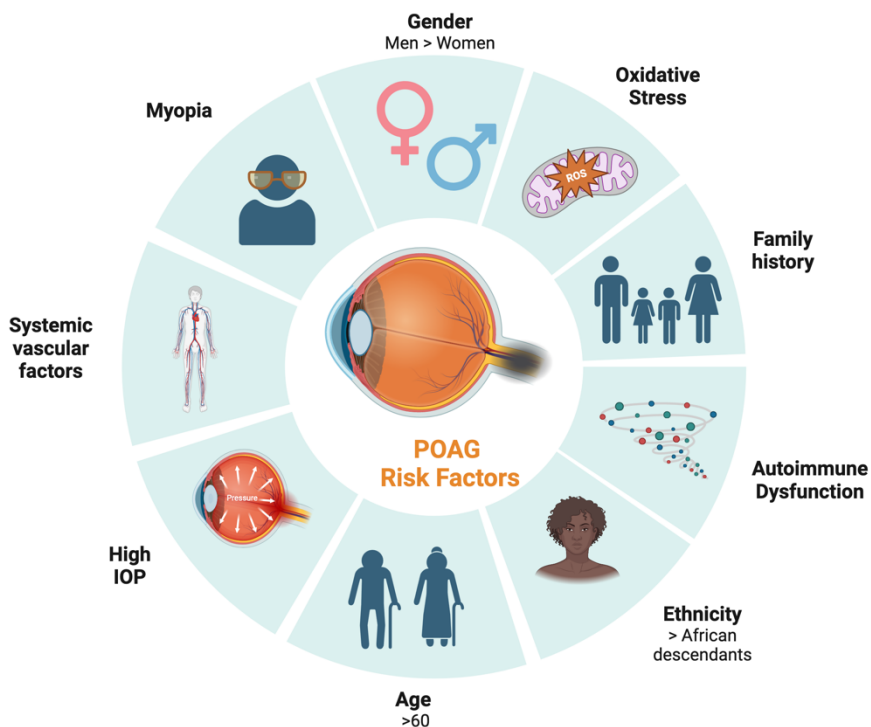
Figure 12. Risk factors associated with POAG. Schematic representation of the factors that have been associated with the development of POAG.

3.2 Current Diagnosis

For the diagnosis of glaucoma, different factors are clinically evaluated:

- First, the visual field is examined to determine the degree of functional impairment produced by the loss of optic nerve fibers, to determine any areas of visual loss.[94]

- Then tonometry is used to determine intraocular pressure.[94] The corneal thickness and curvature are measured at the same time as it can affect the accuracy of the IOP measurements.[94]

- Optic nerve is examined with imaging techniques such as optical coherence tomography. The optic nerve head appearance and the optic nerve fiber thickness change in glaucoma, with a thinning of the optic nerve.[94]

- Finally, gonioscopy is used to examine the chamber drainage angle at the time of initial diagnosis to assess if it is open or closed.[94]

3.3 The need of new biomarkers. Metabolism and POAG

Glaucoma symptomatology appears years after the initiation of the pathophysiological changes and as many as half of the RCG axons can be lost before the first pathological changes are detectable.[99] In early glaucoma when symptoms are absent, diagnosis may be missed. There is a need for the identification of non-invasive biomarkers for the early diagnosis and follow up of glaucoma.[100] Furthermore, the early diagnosis of the disease is extremely important, as currently the only way to prevent blindness is by early treatment reducing the IOP,[99] and high IOP is not even a necessary or a sufficient condition for the development of POAG. Therefore, the identification of molecular biomarkers of glaucoma could not only allow an earlier diagnosis, but also lead to a better understanding of the pathophysiology of the disease, and to the development of new targets for treatment.

Some alterations in the metabolism have previously been associated with the development of glaucoma. Oxidative stress has an important impact in the pathology of glaucoma, as observed in AD. A significant decrease in the antioxidant capacity of POAG patients has been reported, as well as an increase in markers of lipid peroxidation such as malondialdehyde. It has been proved that ROS increases the oxidative DNA damage in the trabecular meshwork of POAG patients.[101] Glucose metabolism has also been observed to be affected in POAG patients. Previous studies have shown alterations in the metabolism and transport of glucose and pyruvate.[102]

When searching for molecular biomarkers of glaucoma, different biofluids can be explored. The process of glaucoma neurodegeneration occurs in the optic nerve and RGC in the inner retina. However, the sampling of this tissue would be highly invasive and not feasible. Serum, vitreous fluid, aqueous humor, and tear film are available and barely invasive biofluids (especially tear film) for the study of biomarkers of glaucoma.[100]

The tear film is a thin layer of fluid that covers the surface of the eye. Although tear film is more distant from the glaucoma neurodegeneration sites, previous studies have found biomarkers associated with glaucoma in this biofluid. For example, endothelin-1, which promotes the decrease in aqueous humor drainage, was found to be increased in tear samples of POAG patients,[103] and factors related to the protection of RGCs were decreased in POAG patients tear film.[100] The presence of POAG biomarkers in tears has been proposed to come mainly from the aqueous humor after scleral perlocation.[104] Previous metabolomic studies of POAG, as well as the identification of new POAG biomarkers through NMR-based metabolomics, are further assessed in this thesis in Chapter 3.

4. Atherosclerosis and plaque vulnerability

Cardiovascular diseases (CVDs), principally stroke and ischemic heart disease (IHD) are the prevalent causes of death worldwide. In 2021 approximately 20.5 million people died from CVDs, accounting for approximately 30% of all global deaths.[105] The main underlying cause of stroke and IHD is the development of atherosclerosis.[106]

Atherosclerosis is a progressive inflammatory disease of the big arteries characterized by the accumulation of fibrous elements, lipids, and different cell types in the subendothelial layer of the arteries, forming what is known as atherosclerotic plaque.[107,108] It is a complex disease produced by a combination of environmental and genetic factors,[106] being hypertension, dyslipidemia (elevated circulating lipids), smoking, obesity, diabetes mellitus and insulin resistance, gender, and age, the main risk factors for the development of the disease.[109]

Arteries are formed mainly by three layers. From outside to inside these layers are *tunica adventitia*, which is formed by epithelial cells; *tunica media*, formed by vascular smooth muscle cells (VSMC); and *tunica intima*, which is in contact with the blood and is formed by a single layer of endothelial cells (Figure 13).[110]
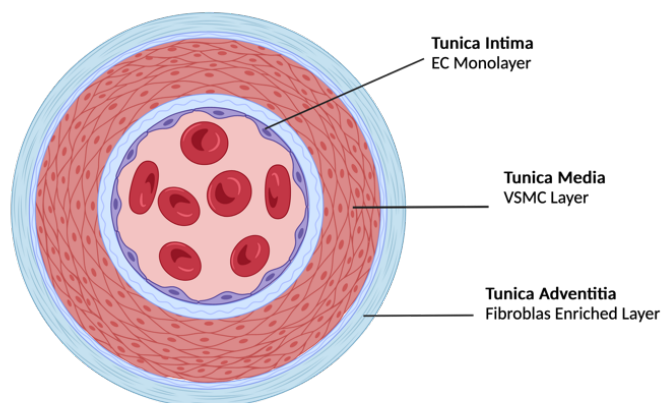
Figure 13. Schematic representation of the cross-section of an artery. The three layers are shown with their main cellular type.

4.1 Etiology of the disease

The atherosclerosis plaque formation begins with the accumulation of low-density lipoprotein (LDL) in the arterial *intima* due to an impairment in the endothelial barrier. In the sub-endothelial, the LDL particles are oxidized by lipoxygenases or ROS.[104] The oxidation of the LDL starts a pro-inflammatory response with the activation of endothelial cells and recruitment of monocytes, which differentiate into macrophages, and lymphocytes. The oxidized LDL is internalized by the macrophages which are then transformed into foam cells. Due to the inflammatory environment, VSMCs migrate to the *intima* and secrete extracellular matrix, which covers the lesion forming a fibrous cap. Eventually, foam cells die releasing the LDL, forming the necrotic lipid reach core. Atherosclerotic plaques continue developing and can evolve to a stable situation or become vulnerable and rupture (Figure 14).
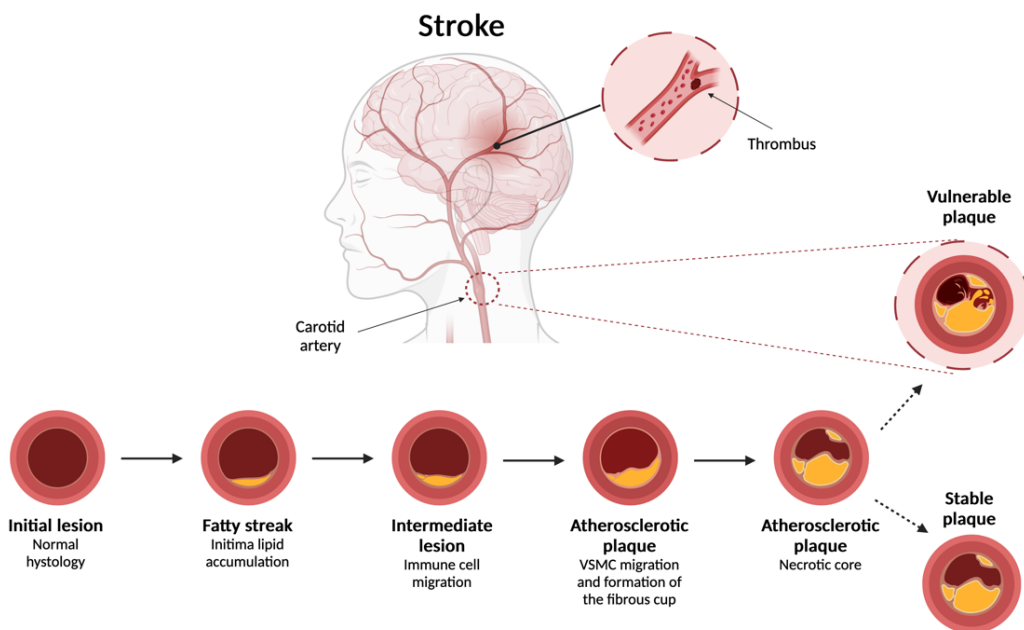
Figure 14. Atherosclerosis progression. A schematic representation of the build-up of the atherosclerotic plaque is shown. The plaque can evolve to a stable situation or become vulnerable and rupture. In the example represented, the rupture of the plaque in the carotid artery leads to the generation of a thrombus that can block the brain vessels, leading to an ischemic stroke.

The rupture of the plaque exposes the collagen-rich core, and the pro-thrombotic factors accumulated inside the lesion to the bloodstream, which leads to the activation of the coagulation cascade. This results in the formation of a thrombus, necessary for the healing of the plaque. However, a large thrombus can completely block the artery and impede the blood flow.[111] The most important clinical complications of atherosclerosis are acute myocardial infarction and stroke, mainly produced when a thrombus occludes the blood flow in the coronary artery or in the carotid, respectively, due to the rupture of the plaque.

4.2 Plaque vulnerability

Vulnerable plaques are defined as non-obstructive silent carotid or coronary lesions that suddenly become obstructive and symptomatic. The vulnerability of the plaque and formation of thrombus rather than the degree of stenosis is responsible for clinical symptomatology in most cases.[112]

There are three principal causes for the formation of a thrombus: plaque rupture, plaque erosion, and the calcification of the nodules. The rupture of the plaque is the most prevalent cause, being responsible for 70-75% of the events.[112] Vulnerable plaques are characterized by a thinning on the fibrous cap previous to the rupture of the plaque, also known as thin cap fibroatheroma (TCFA). It is characterized by a thin fibrous cap, with a large necrotic core with increased free/esterified cholesterol ratio, increased plaque inflammation, vascular remodelling, increased vasa-vasorum neovascularization, and intraplaque haemorrhage.[112]

The molecular processes that are underneath plaque vulnerability are not yet clear, although it is believed that immune cells and inflammatory mediators have an important role in the thrombogenesis of the plaque. Macrophages and T-lymphocytes are capable of degrading the extracellular matrix by phagocytosis and secretion of

proteases, weakening the already thin fibrous cap.[112] Deficient resolution of inflammation results in more unstable and vulnerable plaques.[113]

4.3 Current Diagnosis Strategies.

Atherosclerotic plaques develop silently for decades, without the expression of any symptom until the moment of the event.[114] In a situation of imminent risk, the clinical protocol indicates vascular interventions to restore the blood flow. Due to the absence of symptoms, in most cases, the intervention criteria are based on the evaluation of the degree of stenosis. However, there is increasing concern about the need to include parameters of plaque instability for the determination of vascular risk.

The diagnosis of vulnerable plaques is challenging. Different imaging techniques are starting to be used, in an attempt to identify features associated with plaque vulnerability, such us computed tomography angiography (CTA), MRI angiography (MRA), and carotid duplex ultrasound (CDU), among others.[115]

- CDU is widely used in clinical practice. It is a non-invasive technique that uses two dimensional ultrasounds to create images of the arteries. It allows the identification of factors of plaque vulnerability, such as irregularities of the plaque surface. Nevertheless, it is not useful to determine the internal components of the plaque and it is insensitive to plaque surface features, like fibrous cap thickness. Furthermore, it is highly affected by calcified sound and shadow, respiratory movement and operator dependence.[116]

- CTA can provide information about location, shape and size of the atherosclerotic lesions. It is based in irrading X-ray to the tissue, as differences in thickness and density of tissue leads to different X-ray attenuation, combined with the injection of a contrast agent in the patient, which allows

obtaining two dimensional or three dimensional images of the vessels, after processing of the images. However, this technique is not effective in identifying vulnerable plaques, as it mainly provides information about the location, degree, and extent of stenosis. Furthermore, its use is not very extended in clinical practice. [116]

- MRA allows to obtain information about plaque composition, surface characteristics, and the presence of intraplaque hemorrhage or a lipid-rich core, characteristics associated with plaque vulnerability.[116]

4.4 The need for new biomarkers. Metabolism and atherosclerosis.

A better understanding of plaque vulnerability has been achieved thanks to the use of the aforementioned imaging techniques. However, the accurate prediction of plaque rupture is still far from being accomplished, in part due to the limitations of these imaging techniques and, in many cases, they are not implemented in the clinical routine. The identification of new molecular biomarkers, especially in biofluids, which could be obtained in a medical routine exam would allow a non-invasive, simple method to improve the diagnosis and intervention criteria, as well as allow the development of preventive treatments.

Furthermore, a deeper understanding of the molecular events occurring locally in the plaque would also allow the identification of new targets for therapy for the stabilization of vulnerable plaques.

Metabolic dysregulation in the different cell types involved in atherosclerosis plaque formation and progression have been previously described. An imbalance in lipid metabolism, glucose utilization, and amino acid metabolism has been described. Increased glycolysis in endothelial cells, VSMCs, macrophages, and T-cells is described

in unstable plaques. Increased glucose uptake and glycolysis lead to VSMCs dedifferentiation, migration, and proliferation. Furthermore, elevated lactate has been previously observed in unstable plaques compared to stable plaques.[117]

NMR-based metabolomics could be a reliable and useful tool for the study of biomarkers of plaque vulnerability, as the impact of the cellular metabolism in the instability of the plaque as well as in the associated inflammatory process has previously been highlighted. In Chapter 4, further insights into the metabolomics of atherosclerotic plaque are described, as well as the generation of a predictive model for plaque vulnerability.

5. COVID-19 and Pulmonary Fibrosis

In December 2019, several cases of pneumonia without known origin appeared in Wuhan City. Isolation of the virus from the patients allowed to identify it as a new coronavirus (CoV), and the disease caused by it was named COVID-19 by the WHO.[118] In March of 2020, the World Health Organization declared a state of pandemic due to the rapid global expansion of pneumonia (COVID-19) caused by the virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).[119] SARS-CoV-2 rapid spreading is due to its transmission from person to person through aerosol emission of droplets that are inhaled, getting the virus in contact with the nose, mouth, and eyes.[120] Currently, more than 700 million people have been affected by COVID-19 worldwide (Figure 15). The number of deaths produced as a consequence of SARS-CoV-2 infection is above 7 million.[121]
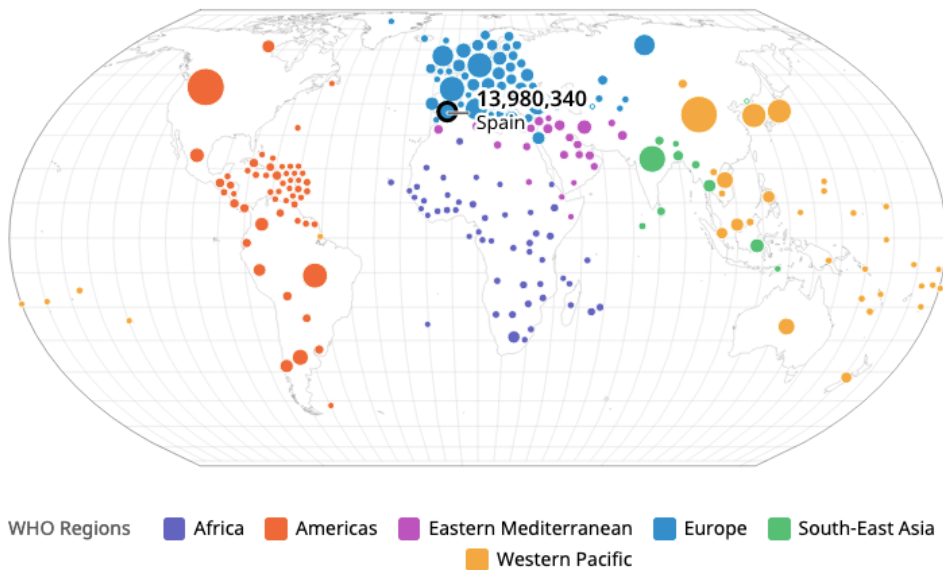
Figure 15. Total cumulative cases of COVID-19 reported to WHO. Almost 14 million COVID-19 cases have been reported in Spain at the moment of elaboration of this thesis.[121]

CoV causes intestinal and respiratory infections in animals and humans. COVID-19 is mainly characterized by causing problems in the respiratory tract. However, the symptomatology presented by COVID-19 patients is wide, going from completely asymptomatic to mild or severe COVID-19. The principal presented symptoms are fever, cough, respiratory difficulties, or muscle pain, but also other symptoms such as diarrhea or vomiting,[118] as well as loss of smell or taste.[119] The presence of comorbidities, such as hypertension, diabetes, obesity, and cardiovascular diseases are risk factors associated with more severe symptomatology.[122] Also, COVID-19 is mainly asymptomatic in people younger than 14, increasing its severity with age.[123] However, the factors related to the severity of the symptoms, and sequelae depend on the personal characteristics of the individual, apart from the aforementioned risk factors, being some persons more susceptible to the development of a severe condition and serious sequelae.

5.1 COVID-19 pneumonia and post-COVID Pulmonary Fibrosis.

The main complication of SARS-CoV-2 infection is the development of pneumonia, with approximately 15% of COVID-19 patients developing it. Symptomatology in these cases includes dyspnea, low arterial oxygen saturation, and high respiratory rate, among others. Around 5% of the cases progress to critical illness, presenting acute respiratory distress syndrome (ARDS), shock, or multiple organ failure, requiring admission to the intensive care unit and mechanical ventilation.[124] The severity of these cases is produced due to an overwhelming immune response. The cytokine release syndrome (CRS) observed in some severe COVID-19 patients, results from an excessive release of cytokines, which causes severe inflammation leading to ARDS.[125] In some cases, inflammation is not successfully resolved, leading to a state of chronic inflammation, which can result in the development of pulmonary fibrosis. Fibrosis occurs due to dysregulation of the wound-healing mechanism. Myofibroblasts are differentiated from fibroblasts to restore damaged tissue. These cells actively synthesize extracellular matrix (ECM) components during the restoration of lung tissue and are eliminated through apoptosis when enough ECM has been produced. However, during chronic inflammation, these myofibroblasts evade apoptosis, leading to the overproduction of ECM and pulmonary fibrosis.[126] Pulmonary fibrosis stands out as one of the most concerning COVID-19 sequelae. Around 10-15% of COVID-19 severe patients develop medium and long-term consequences.[127] Furthermore, patients who had mild COVID-19 have also shown fibrotic changes after the infection, meaning that the virus itself could have a profibrotic effect.[128]

5.2 Current diagnosis

During the COVID-19 pandemic, pulmonary function testing demonstrated limitations in diffusion capacity in hospitalized patients. These tests measure lung volume and capacity, rates of flow, and gas exchange. The second tool used for determining pulmonary fibrosis is computed tomography (CT) scans. In CT scans, pulmonary fibrosis

can be observed as interstitial changes characterized by the appearance of ground glass opacities and irregular lines (Figure 16).[129] The long-term prognosis of pulmonary fibrosis after COVID-19 can vary greatly among individuals. Some can experience mild symptoms and gradually improve over time, while others may develop progressive and chronic lung disease.[130] Currently, two drugs are approved for the treatment of pulmonary fibrosis, *pirfenidone* and *nintedanib*. However, the efficacy of the drugs is limited, and they can merely stop the progression of the disease, far from cure it.[131] The identification of early biomarkers of pulmonary fibrosis in COVID-19 is important to start the treatment and to preserve lung function for as long as possible.
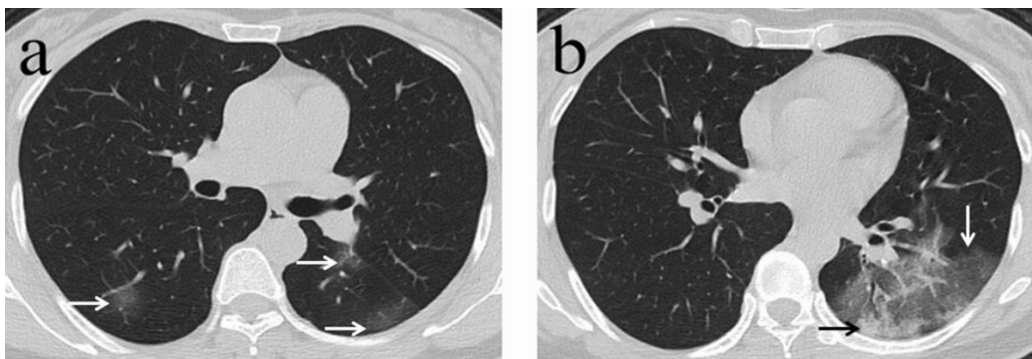
Figure 16. CT of lung from a COVID-19 patient. CT of a 45-year-old infected with SARS-Cov-2 woman showing peripheral patchy ground glass opacities in both lower lobes (white arrows). CT scan also demonstrated consolidation in the left lower lobe (black arrow). (Figure from Qiang Lei, *et.al.*, 2021.[132])

5. 3 The need for new biomarkers.

The development of fibrosis post-COVID-19 is a current clinical concern. Biomarkers are required to determine which patients are more likely to develop pulmonary fibrosis, and help to make treatment decisions, as long-term consequences are still not well studied. Research has been conducted in an attempt to identify biomarkers for the prognosis of pulmonary fibrosis in COVID-19 patients.

Metabolism appears to have an important role in infectious disease development depending on the individual, leading to more severe or milder symptoms.[133] In this sense, metabolomics has proved to be useful for the identification of potential causes underneath the severity of the outcome of COVID-19 patients, as well as for the discrimination of patients according to disease severity, and to determine the effects of treatment and vaccination in different patients.[134] Furthermore, previous metabolomics studies have been performed for the identification of biomarkers of fibrosis in idiopathic pulmonary fibrosis (IPF). Mitochondrial dysfunction and metabolic reprogramming and regulation seem to be involved in the aberrant activation of signaling pathways in IPF.[135]

NMR-based metabolomics is proposed in this thesis as a tool for the identification of biomarkers for the prognosis of pulmonary fibrosis development and progression in COVID-19 patients, as further described in Chapter 5.

6. Bibliography

1. Slikker, W. Biomarkers and their impact on precision medicine. *Exp. Biol. Med.* **243**, 211-212 (2018).

2. Boja, E. S., Kinsinger, C. R., Rodriguez, H. & Srinivas, P. Integration of omics sciences to advance biology and medicine. *Clin. Proteomics* **11**, 45 (2014).

3. Pinu, F. R. *et al.* Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. *Metabolites* **9**, 76 (2019).

4. Qiu, S. *et al.* Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 132 (2023).

5. Muthubharathi, B. C., Gowripriya, T. & Balamurugan, K. Metabolomics: small molecules that matter more. *Mol. Omi.* **17**, 210–229 (2021).

6.   Zampieri, M. & Sauer, U. Metabolomics-driven understanding of genotype-phenotype relations in model organisms. *Curr. Opin. Syst. Biol.* **6**, 28–36 (2017).

7.   Rankin, N. J. *et al.* The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis* **237**, 287–300 (2014).

8.   Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

9.   Wang, J. H., Byun, J. & Pennathur, S. Analytical approaches to metabolomics and applications to systems biology. *Semin. Nephrol.* **30**, 500–511 (2010).

10.   Nizioł, J. *et al.* Untargeted urinary metabolomics for bladder cancer biomarker screening with ultrahigh-resolution mass spectrometry. *Sci. Rep.* **13**, 9802 (2023).

11.   Dang, V. T., Huang, A. & Werstuck, G. H. Untargeted Metabolomics in the Discovery of Novel Biomarkers and Therapeutic Targets for Atherosclerotic Cardiovascular Diseases. *Cardiovasc. Hematol. Disord. Targets* **18**, 166–175 (2018).

12.   Oto, J. *et al.* LC–MS metabolomics of urine reveals distinct profiles for non-muscle-invasive and muscle-invasive bladder cancer. *World J. Urol.* **40**, 2387–2398 (2022).

13.   Wishart, D. S. *et al.* HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).

14.   Jin, Z., Sato, Y., Kawashima, M. & Kanehisa, M. KEGG tools for classification and analysis of viral proteins. *Protein Sci.* **32**, e4820 (2023).

15.   Frolkis, A. *et al.* SMPDB: The small molecule pathway database. *Nucleic Acids Res.* **38**, 480–487 (2009).

16.   Pinu, F. R., Goldansaz, S. A. & Jaine, J. Translational metabolomics: Current challenges and future opportunities. *Metabolites* **9**, 108 (2019).

17.   Botello-Marabotto, M. *et al.* Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease. *Neurobiol. Dis.* **187**, 106312 (2023).

18.   Yilmaz, A. *et al.* Targeted metabolic profiling of urine highlights a potential biomarker panel for the diagnosis of alzheimer's disease and mild cognitive

impairment: A pilot study. *Metabolites* **10**, 357 (2020).

19. Loras, A. *et al.* Integrative metabolomic and transcriptomic analysis for the study of bladder cancer. *Cancers (Basel).* **11**, 686 (2019).

20. Botello-Marabotto, M., Martínez-Bisbal, M. C., Pinazo-Durán, M. D. & Martínez-Máñez, R. Tear metabolomics for the diagnosis of primary open-angle glaucoma. *Talanta* **273**, 125826 (2024).

21. Fuss, T. L. & Cheng, L. L. Evaluation of cancer metabolomics using ex vivo high resolution magic angle spinning (HRMAS) magnetic resonance spectroscopy (MRS). *Metabolites* **6**, 11 (2016).

22. Zheng, H. *et al.* Tissue-Specific Metabolomics Analysis Identifies the Liver as a Major Organ of Metabolic Disorders in Amyloid Precursor Protein/Presenilin 1 Mice of Alzheimer's Disease. *J. Proteome Res.* **18**, 1218–1227 (2019).

23. Martínez-Granados, B. *et al.* Metabolite identification in human liver needle biopsies by high-resolution magic angle spinning 1H NMR spectroscopy. *NMR Biomed.* **19**, 90–100 (2006).

24. Martínez-Bisbal, M. C., Martínez-granados, B. & Rovira, V. Magnetic resonance spectroscopy and imaging on fresh human brain tumor biopsies at microscopic resolution. *Anal. Bioanal. Chem.* **40,** 6771–6780 (2015).

25. Martínez-Bisbal, M. C. *et al.* [1]H and ^{13}C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo ^{1}H MRS study of human high grade gliomas. *NMR Biomed.* **17**, 191–205 (2004).

26. Kostidis, S., Addie, R. D., Morreau, H., Mayboroda, O. A. & Giera, M. Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Anal. Chim. Acta* **980**, 1–24 (2017).

27. Bhinderwala, F., Wase, N., Dirusso, C. & Powers, R. Combining Mass Spectrometry and NMR Improves Metabolite Detection and Annotation Graphical Abstract HHS Public Access. *J Proteome Res* **17**, 4017–4022 (2018).

28. Fiehn, O. Metabolomics by Gas Chromatography-Mass Spectrometry: the combination of targeted and untargeted profiling. *Curr Protoc Mol Biol.* **114,** 30.4.1-30.4.32 (2016).

29. Feng, J. *et al.* Simultaneous Analysis of the Metabolome and Lipidome Using Polarity Partition Two-Dimensional Liquid Chromatography-Mass Spectrometry. *Anal. Chem.* **93**, 15192–15199 (2021).

30.   Tamošiūnas, V. & Padarauskas, A. Comparison of LC and UPLC coupled to MS-MS for the determination of sulfonamides in egg and honey. *Chromatographia* **67**, 783–788 (2008).

31.   Faull, K. F. *et al.* An Introduction to the Basic Principles and Concepts of Mass Spectrometry. *Compr. Anal. Chem.* **52**, 1–46 (2008).

32.   Emwas, A. H. *et al.* NMR spectroscopy for metabolomics research. *Metabolites* **9**, 123 (2019).

33.   Emwas, A. H. M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Meth. Mol. Biol.* **1277**, 161-193 (2015).

34.   Nagana Gowda, G. A. & Raftery, D. NMR-Based Metabolomics. *Adv. Exp. Med. Biol.* **1280**, 19–37 (2021).

35.   De Graaf, R. A. *In vivo* NMR spectroscopy: principles and techniques (John Wiley, 2007).

36.   Tognarelli, J. M. *et al.* Magnetic Resonance Spectroscopy: Principles and Techniques: Lessons for Clinicians. *OURNAL Clin. Exp. Hepatol.* **5**, 320–328 (2015).

37.   Levine, I. N. Physical chemistry. (2009).

38.   Claridge, T. D. W. High-Resolution NMR Techniques in Organic Chemistry: Third Edition. *Elsevier* (2016).

39.   Hays, P. A. & Thompson, R. A. A processing method enabling the use of peak height for accurate and precise proton NMR quantitation. *Magn. Reson. Chem.* **47**, 819–824 (2009).

40.   Mckay, R. T. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **38 A**, 197–220 (2011).

41.   Wang, Y. *et al.* Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues. *Anal. Biochem.* **323**, 26–32 (2003).

42.   Hatada, K. & Kitayama, T. Two-dimensional NMR Spectroscopy In: NMR Spectrosc. Polym*. Springer* 143-168 (2004).

43. Hansen, A. L. *et al.* 2D NMR-Based Metabolomics with HSQC/TOCSY NOAH Supersequences. *Anal. Chem.* **93**, 6112–6119 (2021).

44. Esteve, V., Celda, B. & Martínez-Bisbal, M. C. Use of 1H and 31P HRMAS to evaluate the relationship between quantitative alterations in metabolite concentrations and tissue features in human brain tumor biopsies. *Anal. Bioanal. Chem.* **403**, 2611–2625 (2012).

45. Martínez-Bisbal, M. C., Esteve, V., Martínez-Granados, B. & Celda, B. Magnetic resonance microscopy contribution to interpret high-resolution magic angle spinning metabolomic data of human tumor tissue. *J. Biomed. Biotechnol.* **2011**, 763684 (2011).

46. Zheng, L., Herzfeld, J., Fishbein, K. W. & Griffin, R. G. Two-Dimensional Solid-State 1H-NMR and Proton Exchange. *J. Am. Chem. Soc.* **115**, 6254–6261 (1993).

47. van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006).

48. Spur, E. M., Decelle, E. A. & Cheng, L. L. Metabolomic Imaging of Prostate Cancer with Magnetic Resonance Spectroscopy and Mass Spectrometry. *Eur. J. Nucl. Med. Mol. Imaging* **40**, 60-71 (2013).

49. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **1**, 92-107 (2013).

50. Davies, A. M. C. & Fearn, T. Back to basics: the principles of principal component analysis. *Spectrosopy Eur.* (2004)

51. Debik, J., Sangermani, M., Wang, F., Madssen, T. S. & Giskeødegård, G. F. Multivariate analysis of NMR-based metabolomic data. *NMR Biomed.* **35**, e4638 (2022).

52. Mahadevan, S., Shah, S. L., Marrie, T. J. & Slupsky, C. M. Analysis of metabolomic data using support vector machines. *Anal. Chem.* **80**, 7562–7570 (2008).

53. Date, Y. & Kikuchi, J. Application of a Deep Neural Network to Metabolomics Studies and Its Performance in Determining Important Variables. *Anal. Chem.* **90**, 1805–1810 (2018).

54. Westerhuis, J. A. *et al.* Assessment of PLSDA cross validation. *Metabolomics* **4**, 81–89 (2008).

55. Fordellone, M., Bellincontro, A. & Mencarelli, F. Partial least squares discriminant analysis: A dimensionality reduction method to classify hyperspectral data. *Stat. Appl.* **31**, 181–200 (2019).

56. Kumar, R. & Indrayan, A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr.* **48**, 277–287 (2011).

57. Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **44**, 1–12 (2004).

58. Pérez-Guaita, D., Kuligowski, J., Garrigues, S., Quintás, G. & Wood, B. R. Assessment of the statistical significance of classifications in infrared spectroscopy based diagnostic models. *Analyst* **140**, 2422–2427 (2015).

59. Banerjee, P. *et al.* Identification of key contributory factors responsible for vascular dysfunction in idiopathic recurrent spontaneous miscarriage. *PLoS One* **8**, e80940 (2013).

60. Ramadan, Z., Jacobs, D., Grigorov, M. & Kochhar, S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. *Talanta* **68**, 1683–1691 (2006).

61. Duong, S., Patel, T. & Chang, F. Dementia: What pharmacists need to know. *Can. Pharm. J.* **150**, 118–129 (2017).

62. Long, S., Benoist, C., Weidner, W. World Alzheimer Report 2023: Reducing dementia risk: never too early, never too late. *England: Alzheimer's Disease International*. (2023).

63. Abubakar, M. B. *et al.* Alzheimer's Disease: An Update and Insights Into Pathophysiology. *Front. Aging Neurosci.* **14**, 742408 (2022).

64. Hou, Y. *et al.* Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019).

65. Šerý, O., Povová, J., Míšek, I., Pešák, L. & Janout, V. Molecular mechanisms of neuropathological changes in Alzheimer's disease: A review. *Folia Neuropathol.* **51**, 1–9 (2013).

66. Tahami Monfared, A. A., Zhang, Á. Q., Byrnes, M. J., White, L. A. & Zhang, Q., Alzheimer's Disease: Epidemiology and Clinical Progression. *Neurol. Ther.* **11**, 553-569 (2022).

67. Tiwari, S., Atluri, V., Kaushik, A., Yndart, A. & Nair, M. Alzheimer's disease: pathogenesis, diagnostics, and therapeutics. *Int. J. Nanomedicine* **14**, 5541 (2019).

68. van der Flier, W. M., de Vugt, M. E., Smets, E. M. A., Blom, M. & Teunissen, C. E. Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. *Nat. Aging* **3**, 494–505 (2023).

69. Yamazaki, Y., Zhao, N., Caulfield, T. R., Liu, C. C. & Bu, G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat. Rev. Neurol.* **15**, 501–518 (2019).

70. Masters, C. L. *et al.* Alzheimer's disease. *Nat. Rev. Dis. Prim.* **1**, 1–18 (2015).

71. Tarawneh, R. & Holtzman, D. M. The Clinical Problem of Symptomatic Alzheimer Disease and Mild Cognitive Impairment. *Cold Spring Harb. Perspect. Med.* **2**, a006148. (2012).

72. Morris, J. C. *et al.* Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease. *Arch. Neurol.* **58**, 397–405 (2001).

73. Henneges, C., Reed, C., Chen, Y. F., Dell'Agnello, G. & Lebrec, J. Describing the Sequence of Cognitive Decline in Alzheimer's Disease Patients: Results from an Observational Study. *J. Alzheimer's Dis.* **52**, 1065-80 (2016).

74. Robins Wahlin, T. B. & Byrne, G. J. Personality changes in Alzheimer's disease: a systematic review. *Int. J. Geriatr. Psychiatry* **26**, 1019–1029 (2011).

75. Alzheimer's stages: How the disease progresses - Mayo Clinic. Available at: https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-stages/art-20048448. (Accessed: 5th April 2024)

76. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).

77. Raksasat, R. *et al.* Attentive pairwise interaction network for AI-assisted clock drawing test assessment of early visuospatial deficits. *Sci. Reports* **13**, 18113 (2023).

78. Leuzy, A. *et al.* Blood-based biomarkers for Alzheimer's disease. *EMBO Mol. Med.* **14**, e14408 (2022).

79. Cummings, J. *et al.* Alzheimer's disease drug development pipeline: 2023. *Alzheimer's Dement. (New York, N. Y.)* **9**, e12385 (2023).

80. Butterfield, D. A. & Halliwell, B. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nat. Rev. Neurosci.* **20**, 148–160 (2019).

81. Milne, G. L., Dai, Q. & Roberts, L. J. The isoprostanes--25 years later. *Biochim. Biophys. Acta* **1851**, 433–445 (2015).

82. Smith, M. A., Richey Harris, P. L., Sayre, L. M., Beckman, J. S. & Perry, G. Widespread peroxynitrite-mediated damage in Alzheimer's disease. *J. Neurosci.* **17**, 2653–2657 (1997).

83. Halliwell, B. & Gutteridge, J. M. C. Free Radicals in Biology and Medicine. *Oxford academic* (2015).

84. Koudinov, A. R. & Koudinova, N. V. Essential role for cholesterol in synaptic plasticity and neuronal degeneration. *FASEB J.* **15**, 1858–1860 (2001).

85. Linetti, A. *et al.* Cholesterol reduction impairs exocytosis of synaptic vesicles. *J. Cell Sci.* **123**, 595–605 (2010).

86. Di Paolo, G. & Kim, T. W. Linking lipids to Alzheimer's disease: cholesterol and beyond. *Nat. Rev. Neurosci.* **12**, 284–296 (2011).

87. Tuulonen, A. *et al.* The Finnish evidence-based guideline for open-angle glaucoma. *Acta Ophthalmol. Scand.* **81**, 3–18 (2003).

88. Ibinson, J. W. & Ferguson, L. H. Open Angle Glaucoma. In: Essence of Anesthesia Practique. *StatPearls.* 161 (2011).

89. Allison, K., Patel, D. & Alabi, O. Epidemiology of Glaucoma: The Past, Present, and Predictions for the Future. *Cureus* **12**, e11686. (2020).

90. Harasymowycz, P. *et al.* Medical Management of Glaucoma in the 21st Century from a Canadian Perspective. *J. Ophthalmol.***2016,** 6509809 (2016).

91. Kwon, Y. H., Fingert, J. H., Kuehn, M. H. & Alward, W. L. M. Primary Open-Angle Glaucoma. *N. Engl. J. Med.* **360**, 1113-24 (2009).

92. Križaj, D. What is glaucoma? *Webvision Organ. Retin. Vis. Syst.* 31241881 (2019).

93. Wagner, I. V, Stewart, M. W. & Dorairaj, S. K. Updates on the Diagnosis and Management of Glaucoma. *Mayo Clin. Proc. Innov. Qual. Outcomes* **6**, 618–635 (2022).

94. Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T. & Pfeiffer, N. The Diagnosis and Treatment of Glaucoma. *Dtsch. Arztebl. Int.* **117**, 225-234 (2020).

95. Wakefield, D. & Wildner, G. Is glaucoma an autoimmune disease? *Clin. Transl. Immunol.* **9**, e1180 (2020).

96. Welge-Lüßen, U. & Birke, K. Oxidativer stress im trabekelwerk beim POWG. *Klin. Monbl. Augenheilkd.* **227**, 99–107 (2010).

97. Yanagi, M. *et al.* Vascular risk factors in glaucoma: a review. *Wiley Online Libr. Clinical Exp. Ophthalmol.* **39**, 252–258 (2011).

98. Quigley, H.A., Broman A.T. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol.* **90**, 262-7 (2006).

99. Von Thun Und Hohenstein-Blaul, N., Kunst, S., Pfeiffer, N. & Grus, F. H. Biomarkers for glaucoma: from the lab to the clinic. *Eye* **31**, 225–231 (2017).

100. Beykin, G. & Goldberg, J. L. Molecular Biomarkers for Glaucoma. *Curr. Ophthalmol. Rep.* **7**, 171–176 (2019).

101. Pinazo-Durán, M. D., Zanón-Moreno, V., Gallego-Pinazo, R. & García-Medina, J. J. Oxidative stress and mitochondrial failure in the pathogenesis of glaucoma neurodegeneration. *Prog. Brain Res.* **220**, 127–153 (2015).

102. Harder, J. M. *et al.* Disturbed glucose and pyruvate metabolism in glaucoma with neuroprotection by pyruvate or rapamycin. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33619–33627 (2020).

103. Pavlenko, T. A. *et al.* Endothelins and dopamine levels in tears for assessment of neurovascular disorders in glaucoma. *Vestn. Oftalmol.* **134**, 41–46 (2018).

104. Mastropasqua, R. *et al.* Trans-conjunctival aqueous humor outflow in glaucomatous patients treated with prostaglandin analogues: an in vivo confocal microscopy study. *Graefe's Arch. Clin. Exp. Ophthalmol.* **252**, 1469–1476 (2014).

105. Di Cesare, M. *et al.* The Heart of the World. *Glob. Heart* **19**, 11 (2024).

106. Le Borgne, M., Caligiuri, G. & Nicoletti, A. Once Upon a Time: The Adaptive Immune Response in Atherosclerosis—a Fairy Tale No More. *Mol. Med.* **21**, S13–S18 (2015).

107. Lusis, A. J. Atherosclerosis. *Nature* **407**, 233–241 (2000).

108. Hansson, G. K. & Hermansson, A. The immune system in atherosclerosis. *Nat. Immunol.* **12**, 204–212 (2011).

109. Fruchart, J. C., Nierman, M. C., Stroes, E. S. G., Kastelein, J. J. P. & Duriez, P. New Risk Factors for Atherosclerosis and Patient Risk Assessment. *Circulation* **109**, III-15-III-19 (2004).

110. Bax, M. *et al.* Arterial dissections: Common features and new perspectives. *Front. Cardiovasc. Med.* **9**, 1055862 (2022).

111. Hansson, G. K. Inflammation, Atherosclerosis, and Coronary Artery Disease. *N. Engl. J. Med.* **325**, 1685–1695 (2005).

112. Moreno, P. R. Vulnerable Plaque: Definition, Diagnosis, and Treatment. *Cardiol. Clin.* **28**, 1–30 (2010).

113. Hansson, G. K., Libby, P. & Tabas, I. Inflammation and plaque vulnerability. *J. Intern. Med.* **278**, 483–493 (2015).

114. Fender, A. C. & Dobrev, D. Contemporary plaque imaging for risk stratification of coronary artery disease: Are we getting there? *Int. J. Cardiol. Hear. Vasc.* **31**, 100678 (2020).

115. Hafiane, A. Vulnerable Plaque, Characteristics, Detection, and Potential Therapies. *J. Cardiovasc. Dev. Dis.* **6**, 1–24 (2019).

116. Zhang, L., li, X., Lyu, Q. & Shi, G. Imaging diagnosis and research progress of carotid plaque vulnerability. *J. Clin. Ultrasound* **50**, 905–912 (2022).

117. Stroope, C. *et al.* Dysregulated cellular metabolism in atherosclerosis: mediators and therapeutic opportunities. *Nat. Metab.* **6**, 617–638 (2024).

118. Sun, J. *et al.* COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. *Trends Mol. Med.* **26**, 483–495 (2020).

119. Wang, Y. *et al.* Differential COVID-19 Symptoms Given Pandemic Locations, Time, and Comorbidities During the Early Pandemic. *Front. Med.* **9**, 770031 (2022).

120. Zhou, L., Ayeh, S. K., Chidambaram, V. & Karakousis, P. C. Modes of transmission of SARS-CoV-2 and evidence for preventive behavioral interventions. *BMC Infect. Dis.* **21**, 496 (2021).

121. COVID-19 cases | WHO COVID-19 dashboard. Available at: https://data.who.int/dashboards/covid19/cases?n=c

122. Chatterjee, S. *et al.* Association of COVID-19 with Comorbidities: An Update. *ACS*

Pharmacol. Transl. Sci. **6**, 334–354 (2023).

123. Xie, C. *et al.* Association of Early Inflammation with Age and Asymptomatic Disease in COVID-19. *J. Inflamm. Res.* **14**, 1207 (2021).

124. Papadopoulou, G. *et al.* Molecular and Clinical Prognostic Biomarkers of COVID-19 Severity and Persistence. *Pathogens* **11**, 311 (2022).

125. Sievers, B. L., Cheng, M. T. K., Csiba, K., Meng, B. & Gupta, R. K. SARS-CoV-2 and innate immunity: the good, the bad, and the "goldilocks". *Cell. Mol. Immunol.* **21**, 171–183 (2023).

126. Savin, I. A., Zenkova, M. A. & Sen'kova, A. V. Pulmonary Fibrosis as a Result of Acute Lung Inflammation: Molecular Mechanisms, Relevant In Vivo Models, Prognostic and Therapeutic Approaches. *Int. J. Mol. Sci.* **23**, 14959 (2022).

127. Cojocaru, E. *et al.* Perspectives on Post-COVID-19 Pulmonary Fibrosis Treatment. *J. Pers. Med.* **14**, 51 (2024).

128. Mulet, A. *et al.* Biomarkers of Fibrosis in Patients with COVID-19 One Year After Hospital Discharge: A Prospective Cohort Study. *Am. J. Respir. Cell Mol. Biol.* **69**, 321–327 (2023).

129. Duong-Quy, S. *et al.* Post-COVID-19 Pulmonary Fibrosis: Facts—Challenges and Futures: A Narrative Review. *Pulm. Ther.* **9**, 295–307 (2023).

130. Alrajhi, N. N. Post-COVID-19 pulmonary fibrosis: An ongoing concern. *Ann. Thorac. Med.* **18**, 173-181 (2023).

131. Li, Y. *et al.* Exploring therapeutic targets for molecular therapy of idiopathic pulmonary fibrosis. **107**, 368504241247402 (2024).

132. Lei, Q. *et al.* Correlation between CT findings and outcomes in 46 patients with coronavirus disease 2019. *Sci. Reports* **11**, 1103 (2021).

133. Ayres, J. S. A metabolic handbook for the COVID-19 pandemic. *Nat. Metab. Q* **2**, 572–585 (2020).

134. de Moraes Pontes, J. G., dos Santos, R. Vander & Tasic, L. NMR-Metabolomics in COVID-19 Research. *Adv. Exp. Med. Biol.* **1412**, 197–209 (2023).

135. Roque, W. & Romero, F. Cellular metabolomics of pulmonary fibrosis, from amino acids to lipids. *Am. J. Physiol. - Cell Physiol.* **320**, C689 (20

Objectives

There is a need for the identification of new biomarkers of disease, that in the context of personalized medicine, would allow a better classification of patients according to their specific conditions for an early diagnosis, prognosis, and response to treatment. NMR-based metabolomics stands out as a potent resource for the identification of non-invasive, early biomarkers of diseases. This PhD thesis, therefore, aims to explore the potential of NMR-based metabolomics for the identification of biomarkers in four highly relevant diseases

The main objectives of this project are:

- To elaborate predictive models for AD and MCI diagnosis, and to identify new biomarkers of MCI progression in serum.
- To develop predictive models for the diagnosis of MCI based on the combination of NMR-based metabolomics and lipid peroxidation metabolites obtained by UPLC-MS/MS in plasma.
- To obtain a predictive model for the diagnosis of POAG, and to identify biomarkers of POAG in tear samples.
- To generate predictive models of plaque vulnerability and to identify biomarkers of plaque vulnerability in serum and atheroma plaque tissue samples.
- To obtain biomarkers of pulmonary fibrosis in serum of post COVID-19 pneumonia patients to predict fibrotic changes in CT suggestive of pulmonary fibrosis sequelae one year after hospital discharge.

Chapter 1 | Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease

Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease

Marina Botello-Marabotto[1,2,4], M. Carmen Martínez-Bisbal[1,2,3,4,5*], Miguel Calero[6,7,8], Andrea Bernardos[1,4,5,9], Ana B. Pastor[7], Miguel Medina[6,7], Ramón Martínez-Máñez[1,2,4,5,9]

1. Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), Universitat Politècnica de València, Universitat de València, Valencia, Spain

2. Unidad Mixta de Investigación en Nanomedicina y Sensores. Instituto de Investigación Sanitaria La Fe (IISLAFE), Universitat Politècnica de València. Valencia, Spain

3. Departamento de Química-Física, Universitat de València, Valencia, Spain

4. CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain

5. Unidad Mixta UPV-CIPF de Investigación en Mecanismos de Enfermedades y Nanomedicina, Universitat Politècnica de València, Centro de Investigación Príncipe Felipe, Valencia, Spain

6. Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

7. CIEN Foundation, Queen Sofia Foundation Alzheimer Research Center, Madrid, Spain

8. Instituto de Salud Carlos III, Madrid, Spain

9. Departamento de Química, Universitat Politècnica de València, Valencia, Spain

*Corresponding Author: M. Carmen Martínez-Bisbal. E-mail address:
carmen.martinez-bisbal@uv.es

Graphical abstract

Keywords

My contribution

I performed the deconvolution of the signals from the spectra, the statistical analysis, discussion, and wrote the original draft.

1. Abstract
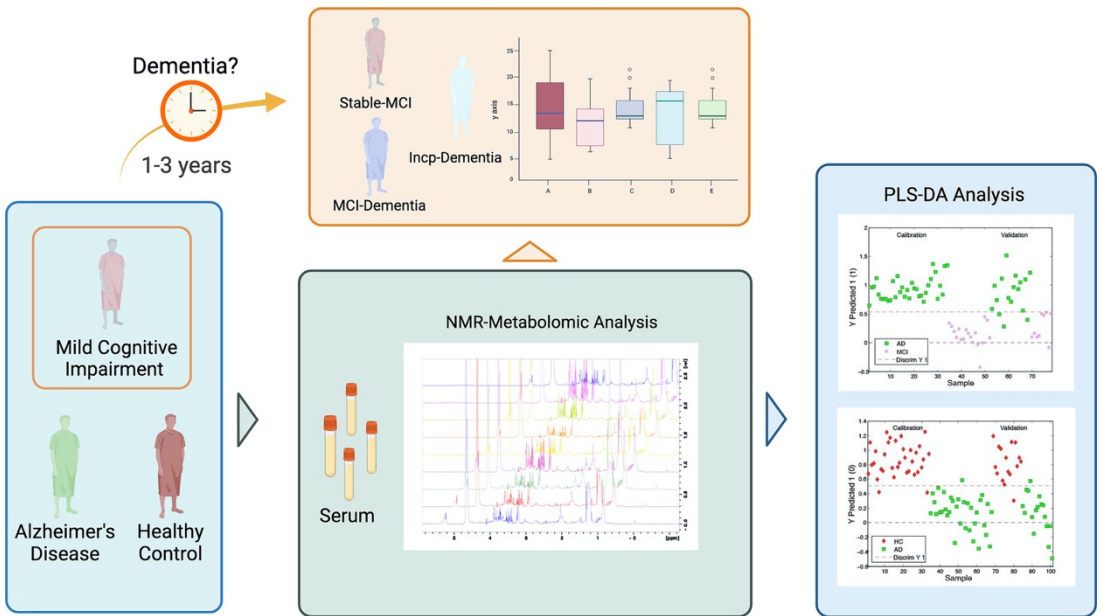
Alzheimer's disease is the most common type of dementia in the elderly. It is a progressive degenerative disorder that may begin to develop up to 15 years before clinical symptoms appear. The identification of early biomarkers is crucial to enable a prompt diagnosis and to start effective interventions. In this work, we conducted a metabolomic study using proton Nuclear Magnetic Resonance (^{1}H-NMR) spectroscopy in serum samples from patients with neuropathologically confirmed Alzheimer's disease (AD, n=51), mild cognitive impairment (MCI, n=27), and cognitively healthy controls (HC, n=50) to search for metabolites that could be used as biomarkers. Patients and controls underwent yearly clinical follow-ups for up to six years. MCI group included samples from three subgroups of subjects with different disease progression rates. The first subgroup included subjects that remained clinically stable at the MCI stage during the period of study (stable MCI, S-MCI, n= 9). The second subgroup accounted for subjects which were diagnosed with MCI at the moment of blood extraction but progressed to clinical dementia in subsequent years (MCI-to-dementia, MCI-D, n=14). The last subgroup was composed of subjects that had been diagnosed as dementia for the first time at the moment of sample collection (incipient dementia, Incp-D, n=4). Partial Least Square Discriminant Analysis (PLS-DA) models were developed. Three models were obtained, one to discriminate between AD and HC samples with high sensitivity (93.75%) and specificity (94.75%), another model to discriminate between AD and MCI samples (100% sensitivity and 82.35 % specificity), and a last model to discriminate HC and MCI with lower sensitivity and specificity (67% and 50%). Differences within the MCI group were further studied in an attempt to determine those MCI subjects that could develop AD-type dementia in the future. The relative concentration of metabolites and metabolic pathways were studied. Alterations in the pathways of alanine, aspartate and glutamate metabolism,

pantothenate and CoA biosynthesis, and beta-alanine metabolism, were found when HC and MCI- D patients were compared. In contrast, no pathway was found disturbed in the comparison of S-MCI with HC groups. These results highlight the potential of ¹H-NMR metabolomics to support the diagnosis of dementia in a less invasive way, and set a starting point for the study of potential biomarkers to identify MCI or HC subjects at risk of developing AD in the future.

2. Introduction

Alzheimer's disease (AD) is the most common type of dementia in older adults.[1] In 2020, over 55 million people worldwide were estimated to have dementia, a figure expected to rise to 78 million by 2030.[2] AD is a progressive neurodegenerative disorder characterized by the accumulation of extracellular plaques of ß-amyloid peptides and intracellular aggregation of tau protein, with the concomitant neuronal and synaptic loss,[3] resulting in the development of cognitive dysfunction and dementia. The pathophysiological alterations usually start between 10 and 15 years before clinical onset.[4] The definitive diagnosis of Alzheimer's disease can only be made post-mortem.[5] However, dementia of the Alzheimer's type is the clinical term to refer to dementia patients in which the Alzheimer's diagnosis is achieved based on the clinical symptoms *pre-mortem*. Age is the greatest risk factor for AD, and the incidence doubles every five years after the age of 65.[3] Mild cognitive impairment (MCI) is a pathological condition characterized by the manifestation of a premature cognitive decline.[6] Dementia-related MCI could be described as an intermediate state between normal aging and AD-type dementia. The prevalence of MCI in the population ranges between 15% and 20% in adults older than 60 years. MCI population is of great interest since their annual rate of progression to dementia is in the range of 8% to 15%.[7] In this context, there is an increasing interest in identifying the subjects with MCI that will progress to dementia. Anticipating the progression of these MCI patients to dementia would be of interest to

apply therapies in the early stages of the disease when these interventions may be more effective.

Nowadays, there is no treatment to restore the cognitive decline of AD patients. Very recently, anti-amyloid immunotherapies have been approved by the FDA, showing a rather modest clinical benefit by slightly slowing down disease progression at early phases of the disease, albeit amid some safety concerns.[8–10] AD patients are identified according to their cognitive state, neuroimaging studies and biomarkers profile. Current clinical biomarkers for AD are variations in the cerebrospinal fluid (CSF) levels of tau proteins (total tau protein and phosphorylated tau protein) and β-amyloid 1-42 peptide (Aß42). AD patients show an increase in tau protein and a decrease in Aß42 levels compared to healthy subjects.[11] Unfortunately, CSF extraction requires invasive procedures and specialised staff, and it is not exempt from risks in aged patients who usually present concomitant pathologies. Considering these factors, the identification of new biomarkers is urgently needed to implement non-invasive diagnostic techniques for AD, to stratify populations for clinical trials, and to find new therapeutic targets to prevent cognitive decline. AD has lately been related to a metabolic disease. In fact, it has been shown that a metabolic dysfunction of the brain could be a potential driver of AD.[12] These metabolic changes in the brain might be translated early on to other organs and biofluids. The observation of the global biochemical changes produced could provide information to reveal biomarkers related to AD and to further deepen our understanding of the molecular mechanism underlying cognitive decline and AD.

Accordingly, the study of the metabolic profile seems relevant in this context. Metabolomics is the study of the footprint of all metabolic pathways and chemical processes occurring in a living system, that is, the study of those metabolites present in a biological sample.[13] Metabolomics is closer to the phenotype than any other -omics discipline, and informs about what has happened, and not about the possibility of something happening as in other -omics.[14] The analysis of differential metabolites in a biological sample isolated from a healthy subject and a patient gives information about the biochemical processes occurring underneath the disease and gives an insight into

new therapeutic approaches and the identification of biomarkers for the diagnosis of the disease. To perform metabolomic studies, nuclear magnetic resonance (NMR) spectroscopy together with mass spectrometry (MS) are the most predominant techniques. NMR spectroscopy allows the untargeted qualitative and quantitative analysis of a wide variety of biological samples (blood derivatives, urine, CSF, tissue, culture cells, and others).[15–18] It is a non-destructive technique, thus enabling the performance of complementary assays in the same sample. The requirements to perform NMR spectroscopy metabolomic studies include straightforward sample processing procedures and small sample quantities. Moreover, there is a wide set of experiments for metabolite identification, and metabolite quantification is also possible with this technique. Remarkably, it is also a robust and very reproducible technique.[19]

Metabolomics, either using NMR spectroscopy or MS, has been extensively applied to the study of AD. Different biofluids have been studied to find AD biomarkers of disease or progression. CSF is in contact with nervous tissue, and it is supposed to be closer to the pathology, so it would be the best candidate to provide information on AD. Different metabolomic studies have been performed in CSF.[4,20,21] Ibáñez *et al.*, 2012 applied capillary electrophoresis-mass spectrometry (CE-MS) for the generation of models of AD progression, reaching 83% of accuracy in the discrimination of patients undergoing AD compared to MCI-AD patients and non-AD subjects. Metabolites such as choline, dimethylarginine, arginine, valine, proline, serine, histidine, creatine, carnitine, and suberylglycine were identified as potential AD progression biomarkers.[20] Jääskeläinen *et al.* compared the ability of classic CSF biomarkers (amyloid-ß 42, phosphorylated tau protein, and total tau) and metabolic profiles obtained by NMR spectroscopy for the classification of AD and healthy controls. The authors concluded that classic CSF biomarkers were better for classifying cognitive healthy controls (HC) vs. AD patients (AUC = 0.89), but metabolic subclasses by NMR spectroscopy were more effective for classifying MCI vs. AD samples (AUC = 0.68).[21] Vignoli *et al.,* 2020 used NMR spectroscopy to obtain the metabolomic profile of CSF samples, and generated

discriminant models with 86.1% accuracy in discriminating between AD and HC and 70% accuracy for classifying AD vs. MCI. They found that acetate, valine, and 3-hydroxyisovalerate were altered in AD.[4] Despite the interesting results obtained from CSF samples, this is an invasive technique that implies a lumbar puncture. Therefore, different biofluids that require less invasive procedures for sample collection, such as urine or blood, have been explored to determine AD biomarkers. Recently, studies in urine using NMR spectroscopy and UHPLC-MS together with metabolic quantitative trait loci (mQTL) were used to calculate models able to classify correctly 82.96% of cases MCI converting to AD, and 77.78% of stable MCI vs. controls. However, urine is separated from the brain not only by the blood-brain barrier, but also by glomerular filtration.[22] Blood fractions (serum, plasma) seem a compromise between less invasiveness and relation to the pathology. Olazarán *et al*., 2015 used UPLC-MS to determine metabolomic biomarkers for the diagnosis of AD using plasma as a biological sample in a set of 251 AD, HC and MCI subjects.[23] In this study a panel of seven metabolites (glutamic acid, alanine, aspartic acid, 22:6n-3 DHA, deoxycholic acid, PE(36:4), SM(39:1)) was determined for the discrimination of AD from HC samples (AUC = 0.918) and MCI from HC (AUC = 0.826). Figuera *et al.,* 2019 used NMR spectroscopy on serum to generate discriminative models able to classify AD and MCI samples from HC with AUC values of 0.61 and 0.71, respectively. In the study, the authors determined that the threonine-linked metabolic pathways were important in the pathological process.[24] Graham *et al*. combined NMR and LC-MS to identify biomarkers able to discriminate AD and MCI from HC, obtaining models with sensitivity and specificity values ranging from 0.75-0.85 and 0.69-0.81, respectively. The authors concluded that the lipid metabolism was the most perturbed biochemical pathway in MCI and AD.[25]

Aimed with this background, herein, we present a metabolomic study by NMR spectroscopy of serum samples from HC, MCI and AD patients to determine biomarkers of disease and early biomarkers of progression.

3. Material and Methods

3.1 Patient selection

Cognitively healthy, non-demented participants and subjects diagnosed with MCI (n=77) were recruited from the Vallecas project, a single-center, multidisciplinary, observational, longitudinal study of a cohort of 1,213 volunteers, aged 69–86 years and home-dwelling at baseline, recruited between 2011 and 2013 in Madrid, Spain, which is carried out in the Queen Sofia Foundation, funded by CIEN Foundation and Queen Sofia Foundation.[23] Participants of the Vallecas project were cognitively healthy volunteers at baseline attending Queen Sofia Foundation Alzheimer Research Center. The AD group (n=51) consisted of clinically diagnosed patients with moderate to severe AD that were institutionalized at the Queen Sofia Foundation Healthcare Center. Written informed consent was obtained from all participants or representatives according to the Declaration of Helsinki. Approval was obtained from the Research Ethics Committee of the Instituto de Salud Carlos III (CEI PEI 46_2011-v2015; CEI PI 78_2019).

Healthy controls (n=50) and subjects included in the MCI group (n=27) were followed up for six years to determine their clinical evolution and observe their possible progression to dementia and AD (Figure 1). According to their clinical progression, subjects in the MCI group were classified into 3 subsets: i) first dementia diagnosis (incipient D, 4 subjects with a first diagnosis of dementia at the time of sample collection); ii) MCI diagnosis at sample collection were diagnosed with MCI but who, in the following one to three years, progressed to dementia (MCI-to-dementia, 14 patients), and MCI patients that in the follow-up, remained stable in MCI condition (stable MCI, 9 patients). Clinical and demographic data are provided in table 1. The Figure 1 represents schematically the study.

Table 1. Sociodemographic and clinical data of participants in this study.

	Healthy controls (HC)	Mild Cognitive Impairment (MCI)			Alzheimer's disease (AD)
		Stable MCI (S-MCI)	MCI progressing to dementia (MCI-D)	Incipient Dementia (Incp D)	
Age, Mean (range)	77.2 (72-87)	80.2 (74-91)	78.9 (71-85)	80.2 (76-86)	82.35 (58-93)
Number (%)	50 (39.1)	14 (10.9)	9 (7)	4 (3.1)	51(39.8)
Sex, male/female	17/33	4/10	5/4	10/17	9/42
APOE4+ (%)	10 (20%)	2 (13%)	3 (37.5%)	2 (50%)	22 (43.1%)
MMSE* / SMMSE**	28.2 (24-30)*	25.9 (22-28)*	24.8 (19-30)*	21 (14-25)*	10.84 (0-30)**
CDR***	0 (0-0.5)	(0.5)	(0.5)	(1)	2.75 (1-3)

* MMSE = Mini Mental State Examination (0-30)
** SMMSE = Severe MMSE (0-30) (validated for Spanish-speaking population)[26]
*** CDR = Clinical Dementia Rating (0-3)
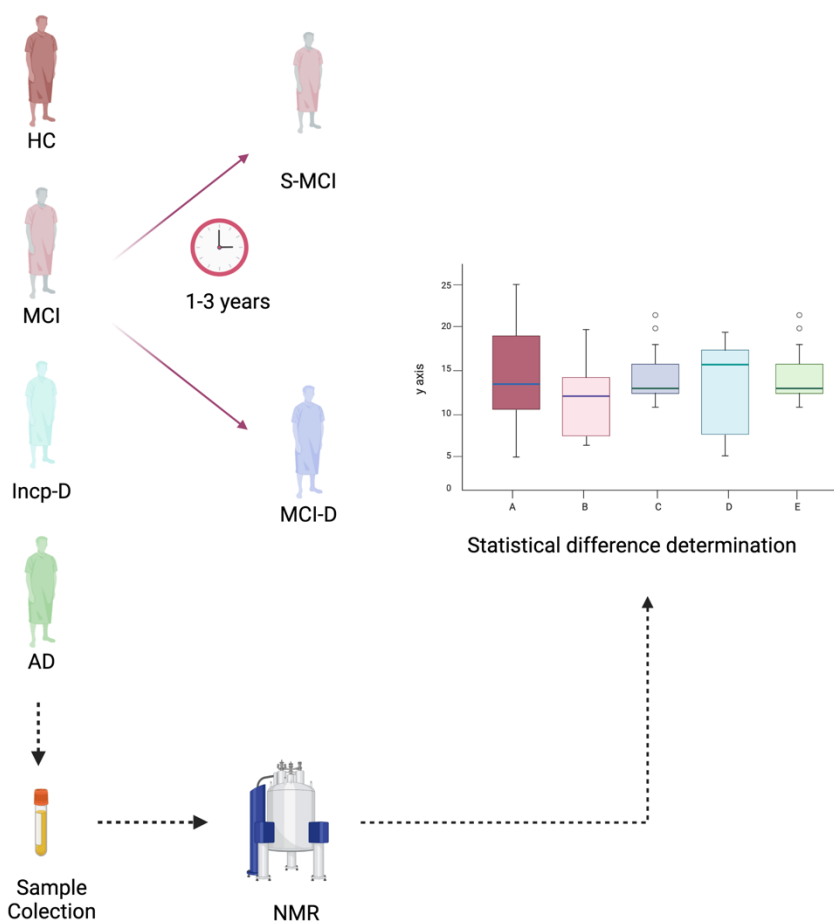
Figure 1. Schematic representation of the analysed groups. Serum samples were collected from patients with different clinical diagnosis (HC, MCI, and AD). Within the MCI group, the clinical stage at sample collection and the progression in the follow up were diverse. Some MCI patients were in clinical condition close to dementia at sample collection, and were classified as Incp-D. In the follow-up time, some of the MCI patients progressed to a more evolved state of dementia (MCI-D) whereas for other MCI patients the clinical diagnosis remained the same (S-MCI). Taken this into consideration, statistical analyses were performed to these sets of NMR spectra to determine metabolomic differences between S-MCI and MCI-D, together with the other clinical groups: HC, Incp-D and AD.

3.2 Cognitive assessment

Participants from the Vallecas project cohort were cognitively assessed by the well-known Mini Mental Scale Examination (MMSE) and the Clinical Dementia Rating (CDR) scales. To overcome the floor effect effect observed with the standard MMSE scale on moderate-to-severe AD population,[27] the AD group was cognitively assessed with the Severe MMSE (SMMSE).[28] The fact that scoring in both scales give a range of values between 0 and 30 can be confusing, but scores from both scales should not be directly compared.[29] Thus, the CDR scale has also been included, which can be used with all groups studied (CDR=0, cognitively unimpaired; CDR=0,5, MCI; CDR=1 mild dementia; CDR=2, moderate dementia; CDR=3, severe dementia).

AD-type dementia diagnosis was established according to the National Institute on Neurological Disorders and Stroke, and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) guidelines.[30] Forty three out of fifty one AD subjects had donated their brains, and the AD diagnosis was neuropathologically confirmed *post mortem*. Participants with MCI were defined using criteria described by Petersen *et al.,* 1999.[31]

3.3 Sample preparation

Serum samples from healthy controls and MCI subjects from the Vallecas Project cohort as well as AD samples from the CIEN Foundation Brain Tissue Bank (BT-CIEN), were all collected in fasting conditions by venous puncture. After clot removal, aliquots of 500 µL were preserved at -80 ºC until the analysis.

Samples were prepared following the protocol described by Beckonert *et al.*[32] Briefly, before sample preparation, samples were thawed. Immediately, 400 µL of serum were introduced in 5 mm NMR tubes, and 200 µL of phosphate buffer (pH 7.4) were added. Phosphate buffer contained deuterated water (20% v/v) and sodium 2,2-dimethyl-2-

silapentane-5-sulphonate (DSS) 1 mM as internal standard for chemical shift referencing.

3.4 NMR spectra acquisition and processing

Once the samples were prepared, NMR spectra were recorded in a Bruker Avance DRX 600 MHz spectrometer (Bruker GmbH, Rheinstetten, Germany) at *U26 NMR: Biomedical Applications II platform* from *Nanbiosis (Research Infrastructures & Services of CIBER-BBN)*. 1D ^{1}H-NMR spectra were acquired for each sample using Carr-Purcell-Meiboom-Gill (cpmg) pulse sequence with water signal suppression and a total spin echo of 32 ms for each sample (interpulse delay between 180º pulses was 0.001 s, and the number of loops was 16). This pulse sequence reduces the contribution of signals from high molecular weighted molecules to the spectra, such as proteins or other macromolecules, owing to their short times of transverse relaxation (T_2). The temperature of the probe was set at 300 ºK (27 ºC). Together with the acquisition of ^{1}H cpmg spectra, 2D homonuclear (^{1}H-^{1}H TOCSY) and heteronuclear spectra were acquired (^{1}H-^{13}C HSQC) in a reduced set of samples to unequivocally identify and assign the signals in the spectra.

Once acquired, the 1D and 2D spectra were Fourier transformed and processed with *TopSpin 4.0.0* (Bruker BioSpin Corporation). For processing the 1D spectra an exponential line-broadening function of 0.5 Hz was applied followed by Fourier transformation. Phasing, baseline correction and chemical shift referencing to the trimethylsilyl signal of DSS at 0.0 ppm was also performed. For the processing of the 2D spectra the phase was corrected for rows and columns, and the chemical shift referenced to the trimethylsilyl signal of DSS at 0.0,0.0 ppm. The main signals in the spectra were assigned according to the data in the bibliography [33,34] and the Human Metabolome Data Base (HMDB).[35]

After processing, meaningful signals in the cpmg spectra underwent deconvolution using *AMIX 4.0.2* software (Bruker BioSpin Corporation). The residual signals after

water suppression in the area between 4.5 and 5.0 ppm, and those regions with chemical shifts lower than 0.5 ppm and higher than 8.5 ppm were excluded from the analysis. A total of 130 signals were selected in the 1D spectra and included for deconvolution (Figure S1). Then, a mixed Gaussian/Lorentzian variable function was applied for deconvolution of these signals. After deconvolution, integrals were obtained for all cpmg spectra and were normalized to the sum of all integrals in each sample, resulting finally in a data set of 130 normalized integrals for each sample.

3.5 Multivariate statistical analysis

To determine differences between the serum metabolomic profiles of, HC, MCI and AD patients, multivariate statistical analyses were performed. For this purpose, the normalized data were fed into the software *PLS_Toolbox Solo 8.9* (Eigenvector Research, Inc., Manson, WA, USA).

Partial least squares-discriminant analyses (PLS-DA) were performed to generate predictive models able to classify the samples according to the clinical diagnosis, and using the information from the cpmg spectra deconvolution, i.e., based on its metabolic profile. Models were generated to discriminate HC vs. AD, and MCI vs. AD.

Before PLS-DA analyses the data were split in discovery (66% of the data was included for training and calculating the models) and validation subsets (33% of the data was used to apply the calculated model and to see the performance of the model). The split in 70% and 30% has been empirically proven to provide accurate models and results.[36] Cross validation was used to determine the appropriated number of principal components.

The performance of multivariate statistic calculations is generally improved when the number of variables and samples is equilibrated.[37] With this purpose, variable selection strategies are usually included previous to these analyses.[38] Accordingly, in this study first, applying the knowledge on the signals in the spectra, only one representative peak from each metabolite was selected, thus reducing the number of variables from 130 to

48. Afterwards, variable selection in the calibration sets was performed selecting those variables with values of Variable Importance in Projection (VIP) higher than 0.8.

After variable selection in the discovery sets, two PLS-DA models were obtained (to discriminate HC vs. AD, and MCI vs. AD, respectively). Cross validation (using venetian blinds) was used to determine the optimum number of latent variables for the model. After applying the models to the validation sets, sensitivity, specificity and the area under the ROC curve (AUC) were calculated to determine the goodness of the models to discriminate between each set of samples. Once validated, to determine the robustness of the model and to test for over-fitting, permutation tests (200 iterations) were performed and pairwise Wilcoxon signed rank text (Wilcoxon test), pairwise signed rank test (Rank test) and randomization t-test (Rand t-test) probabilities were obtained in the self-prediction and in the cross-validated residuals.

3.6 Univariate statistical analysis

Mean comparison of the identified metabolites in HC vs. AD and MCI vs. AD was performed. For comparison of pairs t-test and Mann Whitney U test were used, depending on the results obtained in the normality test (Kolmogorov-Smirnov (Dementia) or Shapiro Wilks (HC and MCI)). *IBM SPSS Statistics* 25 version was used for univariate statistics. MCI-to-dementia, stable MCI and incipient D as defined previously were considered in the study of MCI set. ANOVA test and the post hoc Scheffé test were used for comparison between the different MCI subsets, and AD and HC groups. Boxplots of the above-mentioned groups of metabolites showing sequential changes according to the progression in the disease were performed using *R* (*RStudio 1.2.5001*).

3.7 Analysis of altered metabolic pathways in AD and MCI

To determine the potential metabolic pathways involved in the pathological processes, *Metaboanalyst* [39] was used. A concentration table made with the relative concentration of each metabolite as columns and samples as rows as used. The HMDB ID of each

metabolite was used to include them in the pathway analysis, so those metabolites whose HMDB ID was not available were not included, such as fatty acids or unknown metabolites. The global test enrichment analysis selected for the topological analysis was Relative-betweenness centrality and the Homo sapiens library provided by metaboanalyst was used as reference metabolome. The enrichment method selected was global test. Once the analysis was obtained, the pathways with p value < 0.05 and impact factor > 0 were chosen as representative pathways.

4. Results and Discussion

4.1 Metabolic profile of serum samples

The main signals in the spectra were assigned to enable the identification of potential biomarkers of AD-type dementia and progression in the discriminant models. Figure 2 shows the serum spectrum of one of the samples with the assignment of the main peaks. For a better observation of the signals in the figure, spectrum has been split in two parts, aliphatic (Figure 2.a) and aromatic part (Figure 2.b). Reference spectra of MCI (Figure 2.c) and HC (Figure 2.d) are also shown for comparison.

The trimethylsilyl peak of DSS can be observed at 0.00 ppm as a singlet. All the assigned resonances are shown in table S1 with the detail in the compound and in the functional group, and for each of them, the chemical shift, the multiplicity and J coupling. With this information, 27 compounds were assigned. Within the assigned compounds, there were 11 amino acids identified, 5 organic acids and 2 sugars, among other molecules, such as fatty acids or alcohols.
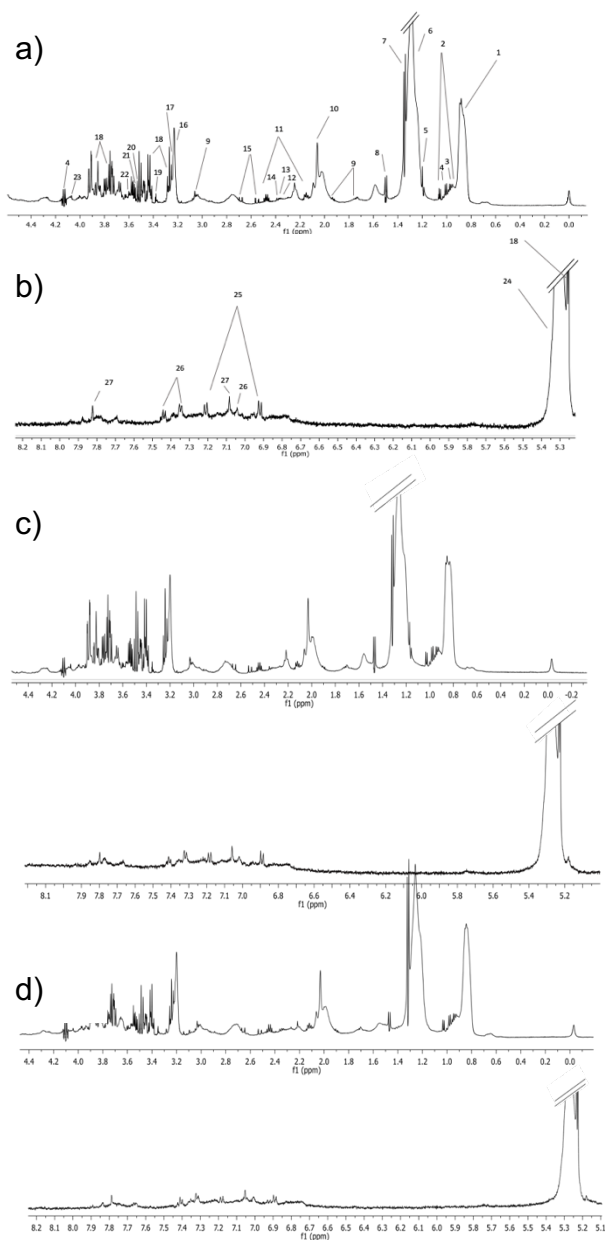
Figure 2. **Serum ¹H-NMR spectrum of an AD patient** a) Aliphatic region of the spectrum (δH 0-4.4 ppm). b) Aromatic region of the spectrum (δH 5.3-8.2 ppm). Due to a lower signal intensity in the aromatic region, the intensity was amplified 10 times in regard to the aliphatic region. The spectrum area corresponding the water suppression signal is not shown. c) Serum ¹H-NMR spectrum of an MCI patient d) Serum ¹H-NMR spectrum of a HC.

4.2 Multivariate analysis of the serum metabolomic profiles

A PLS-DA analysis was performed to generate a predictive model able to classify and differentiate between HC and AD serum samples. Three principal components were selected. From the first model generated with the calibration subset, variables with VIP > 0.8 were selected, resulting in a total of 21 variables. The R^2 value was 0.7 and the Q^2 was 0.59. This model was then applied to the validation set and 93.75 % of sensitivity and 94.18% of specificity were obtained, with an AUC value of 0.9816 (Figure 3.a). All the permutation tests performed proved the robustness of the model ($p < 0.05$). The variables participating in this model (and potential biomarkers of the disease) are shown in the table 2, having a greater impact in the model n-acetylglucosamine, CH_3- mixed lipoproteins, and n-acetylated compounds, pyruvate, lysine, threonine and glycine.

Following the same procedure, a PLS-DA was performed to generate a predictive model to discriminate between dementia of the AD type against MCI. The discrimination of these two sets of patients is of the most importance in this clinical context for early diagnosis. The model was generated as described before. In this case the model had a total of 18 variables and was made with 2 principal components with an R^2 value of 0.7 and a Q^2 value of 0.61 and was able to determine with a 100% of sensitivity and 82.35 % of specificity (for the validation set) between AD and MCI with an AUC value of 0.9281 (Figure 3.b). All the permutation tests performed proved the robustness of the model ($p < 0.05$). The variables participating in this model are shown in table 2. The metabolites with a greater impact in the model are CH_3- mixed lipoproteins, n-acetylglucosamine, n-acetylated compound, lysine, pyruvate and CH_2- mixed lipoproteins. Finally, a third model for the discrimination of HC and MCI was performed. The model was made with 2 principal components. The classification and prediction results obtained by PLS-DA did not offer a good performance, yielding a specificity of the 50% and 67% of sensitivity (Figure 3.c.). The AUC value was 0.556. The Q^2 value

was -0.1 and the R2 0.17. As reflected by the AUC and Q2 and R2 values, this model does not have capability for prediction.
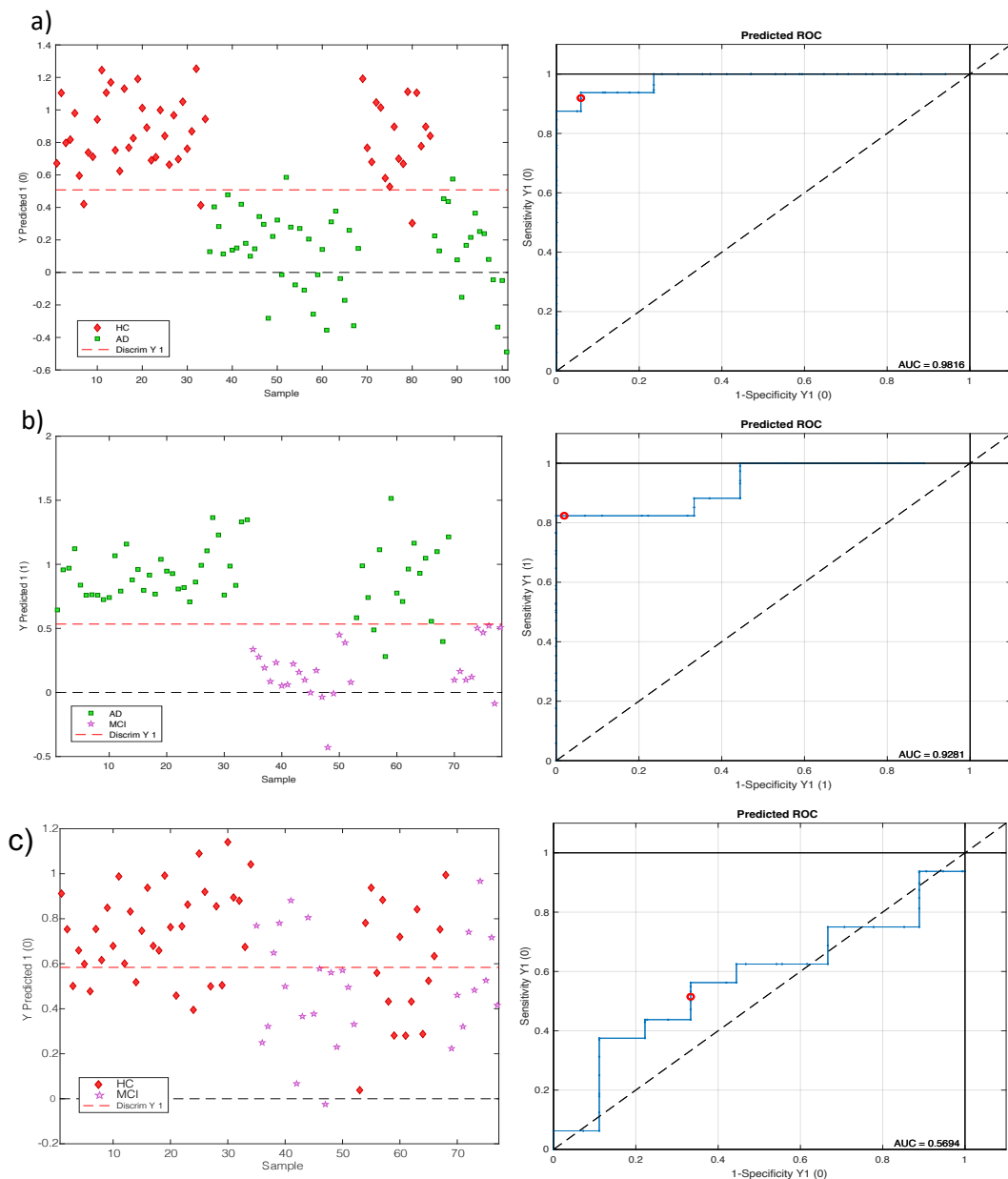
Figure 3. PLS-DA scores and ROC curve for the classification of (a) HC vs. AD, (b) AD vs. MCI and (c) HC vs MCI. In the left side of the panel the prediction plot is shown, divided in the scores obtained in the calibration and validation subsets. In the right side of the panel, ROC curve is shown for each model, with an AUC value of (a) 0.9816, (b) 0.9281 and (c) 0.5694 for discrimination of HC vs. AD, AD vs. MCI and HC vs MCI

The metabolites participating in this model are shown in table 2. Given the clinical relevance of finding differences between these two groups, further analysis will be performed seeking for potential metabolic biomarkers of the differences between HC and MCI.

Table 2. Variables participating in the models sorted by VIP value.

AD vs HC		AD vs MCI		HC vs MCI	
Variable	**VIP S.**	**Variable**	**VIP S.**	**Variable**	**VIP S.**
N-acetylglucosamine	1.54	CH_2 m.l. *(1.20 ppm)	1.67	Leucine	1.57
CH_2 m.l. (1.20 ppm)	1.52	N-acetylglucosamine	1.56	Choline	1.39
N-acetyled comp.	1.34	N-acetyled comp.	1.31	Valine	1.27
Pyruvate	1.30	Lysine	1.27	Pyruvate	1.25
Lysine	1.10	Pyruvate	1.15	Creatinine	1.22
Threonine	1.09	CH_3 m.l. (0.83 ppm)	1.13	N-acetyled comp.	1.02
Glycine	1.08	Phenylalanine	0.99	Lysine	1.02
Ethanol	0.93	Ethanol	0.99	Arginine	1.01
CH_3 m.l. (0.85 ppm)	0.91	Tyrosine	0.87	Glutamine	1.00
Choline	0.90	Unk (3.86)	0.87	Glycerol	0.94
CH_3 m.l. (0.83 ppm)	0.90	Citrate	0.82	Alanine	0.84
Valine	0.90	Choline	0.79	Acetyl Choline	0.83
CH_3 m.l. (0.87 ppm)	0.89	Glycine	0.68	Isoleucine	0.83
Creatine	0.87	Lactate	0.66	Threonine	0.77
Phenylalanine	0.84	Creatine	0.65	CH_3 m.l. (0.81 ppm)	0.71
Glycerol	0.81	Unk (7.01 ppm)	0.63	CH_3 m.l. (0.83ppm)	0.70
Unk (7.01ppm)	0.73	Acetone	0.44	Ethanol	0.66
Acetylcholine	0.68	Glycerol	0.41	Lactate	0.66
Methanol	0.67				
CH_3 m.l. (0.88 ppm)	0.61				
Acetone	0.55				

* m.l. means mixed lipoproteins

Previous NMR metabolomic studies in serum have been performed to obtain discriminative models for the identification of potential biomarkers of AD and MCI. Figuera *et al.,* 2019, generated models able to classify MCI and HC samples with an AUC value of 0.6, similar to the results we obtained. They found that threonine, 2-hydroxybutyrate, glutamine, L-tyrosine, trimethylamine, isobutyrate and propylene glycol were important in the discriminative model. They as well generated models able to classify HC and AD samples with an AUC of 0.71. The metabolites that they found important for the discrimination of samples were threonine, aspartate, creatine, N,N-Dimethylglycine, L-alanine, acetic acid and acetoacetic acid.[24] Yilmaz *et al.,* 2020, generated discriminative models for the identification of potential biomarkers of MCI and AD by NMR and LC-MS in plasma samples. They found that acetic acid and lyso-phosphatidilcholine (C16:1), ceramide (C18:2), sphingomiosine (C24:1) and sphingomiosine (C24:0) were the most important metabolites for the discrimination between HC and MCI.[25] Among the set of metabolites found in the work here presented, glutamine was also found important in the discrimination between MCI and HC, and threonine and creatine were found important in the discrimination between AD and HC, as had been previously reported in the bibliography. Regarding to the discrimination between MCI and AD, the metabolites reported were obtained by LC-MS in the model found in Yilmaz *et al.,* 2020. It seems controversial the scarce coincidence of biomarkers found in the diverse studies. Nevertheless, even though NMR is a highly reproducible technique, there are several issues to address when comparing the results obtained from different metabolomic studies, including the way the samples have been collected and prepared, the patient inclusion and classification criteria, or the different ways for data processing as the integration, scaling and normalizing methods for NMR data.

4.3 Mean comparison of metabolites

The relative concentration of metabolites in HC, AD and MCI was compared. The 48 signals, used for the multivariate statistics were here analysed, including identified metabolites, unknown peaks, and different peaks from fatty acids and lipoproteins. Significant differences were identified in most of the metabolites analysed in both comparison HC vs. AD and AD vs. MCI, highlighting the wide impact that the disease has in the cellular metabolism (Table 3).

Creatine, ethanol, threonine, glycine, methanol, lysine, n-acetylglucosamine, alanine, CH_2 mixed lipoproteins, valine, CH_3 mixed lipoproteins, phenylalanine, acetylcholine, choline, pyruvate, acetone, glycerol, isoleucine, N-acetyled compounds and two unassigned signals (7.01 ppm and 3.86 ppm) showed statistical differences in its relative mean concentrations between AD and HC. In the comparison of AD vs. MCI, statistical differences ($p < 0.05$) were found in lactate, creatine, ethanol, glycine, lysine, n-acetylglucosamine, CH_3 mixed lipoproteins, CH_2 mixed lipoproteins, phenylalanine, pyruvate and choline. At this point, it deserves to be noted that most of the variables in the PLS-DA models showed significant differences in their means for both comparisons, HC vs. AD and MCI vs. AD (with exception of valine in both comparisons, as well as the unknown compound at 7.01 ppm in MCI vs. HC)

For the comparison between HC and MCI we found significant differences in the relative concentration of lactate and threonine. The levels of lactate in cerebrospinal fluid have been associated to disease severity in other neurological diseases.[40] However, other studies, also performed in CSF observed the same behaviour as here is presented in serum, showing an increase in the concentration of lactate between HC and AD that is not significant, whereas a higher and statistically significant increase is observed in the serum samples of MCI patients when compared to HC and AD.[41] It is interesting to observe that this association between higher lactate levels and earlier stages of dementia is also found in serum, which can be obtained following minimally

invasive procedures compared to those needed to obtain CSF. Zebhauser *et al.* suggested that the increase of lactate levels in CSF in MCI could be produced by the activation of microglia.[41]

Table 3. Mean comparison of relative concentration of metabolites with significant differences between HC and AD and/or MCI and AD.

Metabolites	[Metab]rel x 10^3			p value		
	HC	MCI	AD	HC vs AD	MCI vs AD	HC vs MCI
Unk (7.01)	1.030	0.959	0.763	0.007	0.069	0.578
Lactate	1.181	1.460	1.249	0.740	0.038	0.025
Creatine	3.221	3.350	4.256	0.000	0.008	0.669
Unk (3.86)	11.850	13.511	7.235	0.002	0.000	0.370
Threonine	1.838	2.031	2.108	0.002	0.797	0.023
Glycine	3.524	3.625	4.221	0.001	0.025	0.685
Methanol	1.483	1.505	2.003	0.011	0.079	0.468
Lysine	4.856	4.499	6.101	0.000	0.000	0.211
N-acetylglucosamine	35.144	34.781	45.484	0.000	0.000	0.741
Alanine	4.860	5.169	5.767	0.012	0.224	0.179
CH_2 m.l. (0.852)	44.679	41.141	52.991	0.000	0.000	0.795
Phenylalanine	0.487	0.476	0.744	0.000	0.001	0.866
Acetylcholine	26.860	22.151	21.353	0.006	0.805	0.660
Choline	33.989	32.328	28.069	0.000	0.000	0.146
Pyruvate	2.429	2.521	3.360	0.000	0.000	0.551
Acetone	7.827	7.803	8.973	0.038	0.098	0.940
Isoleucine	4.577	4.560	4.261	0.048	0.655	0.781
Ethanol	3.527	3.680	3.429	0.000	0.000	0.423
Glycerol	4.631	4.883	2.594	0.000	0.000	0.417
N-acetyled compound	12.271	12.608	15.963	0.000	0.000	0.527
CH_3 m.l. (1.204)	45.991	46.874	25.969	0.000	0.000	0.814
CH_2 m.l. (0.865)	16.242	14.679	23.375	0.004	0.000	0.240
CH_2 m.l. (0.832)	60.757	53.507	65.815	0.248	0.031	0.130
CH_2 m.l. (0.813)	37.802	33.615	35.517	0.034	0.152	0.823

* m.l. means mixed lipoproteins

4.4 Metabolic changes between MCI patients with different progression rates

Attention should be drawn to MCI that is a diverse group, and a clinically relevant issue would be to be able to discriminate, previous to the manifestation of overt clinical symptoms, whether a patient with MCI in the near future will progress to dementia or, on the contrary, if this patient will remain in MCI condition. For that purpose, the MCI patient's progression along the follow up period (between 1 and 3 years) was considered. According to their clinical progression, the MCI patients were classified in MCI-to-dementia (MCI-D), stable MCI (S-MCI) and incipient dementia (Incp-D).

When the MCI patients were projected onto the HC vs AD model, they were not classified according to the progression of cognitive impairment (Figure S2), that is why a different strategy was followed seeking for differences that could be found within the MCI group and could be related to the progression of the disease.

With that purpose of identifying metabolic changes between the different clinical evolution in MCI subgroups in comparison to HC and AD patients, boxplots of the different metabolites were represented, ordered according to the clinical severity of each subset of patients, HC, S-MCI, MCI-D, Incp-D and AD-type dementia (Figure 1).
A progressive increase in the concentration of phenylalanine, lysine, pyruvate, and CH_2 mixed lipoproteins, and a progressive decrease in the concentration of choline, was observed (Figure 4).
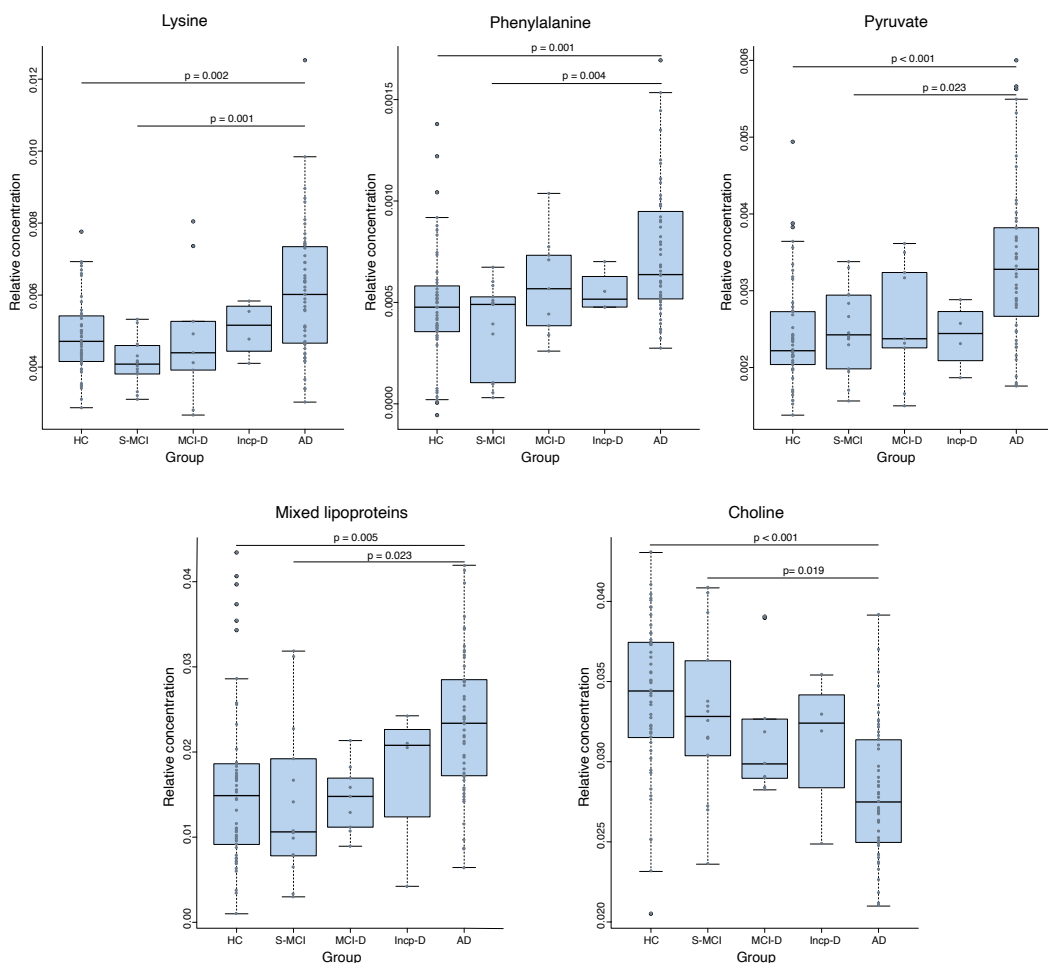
Figure 4. Relative concentration of metabolites with a progressive trend between the groups of study. The median (horizontal bar) of lysine, phenylalanine, pyruvate, CH_2 mixed lipoproteins and choline relative concentration for the different groups (HC, S-MCI, MCI-D, Incp-D and AD) is depicted. The p value of groups with means significantly different is shown.

Statistical value of these differences was assessed by ANOVA test, and afterwards by Scheffé test to determine the difference between each group. The Scheffé test is appropriated to be used when the size of the groups is small or the samples size between groups is unbalanced as it is in our case. Differences between S-MCI and AD patients were found ($p < 0.05$), but these differences were not significant between MCI-

D and AD patients. These results show that at the time of sample collection the concentration of these metabolites in MCI patients that later evolved to dementia (MCI-D) had already values closer to AD than those from MCI patients who did not evolve (S-MCI). In that way, these metabolites could be postulated as potential predictive biomarkers. An interval value of these metabolites could be determined to discriminate whether MCI samples would evolve to AD or other advanced states of dementia before the clinical symptoms were developed.

In a parallel analysis, the metabolic pathways that could be altered in the development of dementia were explored. The different groups of patients, including the MCI subgroups, were compared to find metabolic pathways affected in the different conditions of the disease in the MCI and dementia context.

There were 14 metabolic pathways altered between dementia and HC (not shown), 8 of them were as well altered between AD and MCI-D / Incp-D, whereas 9 pathways were altered between AD and S-MCI (Figure 5). Following a logic similar to the one presented before, a higher number of metabolic pathways were altered in the comparison between AD and S-MCI than when AD samples were compared with those samples that were closer, or later evolve, to dementia (MCI-D / Incp-D) (Figure 5). Remarkably, four pathways did not show significant differences between AD and MCI-D / Incp-D but showed significant differences between AD and S-MCI: phenylalanine, tyrosine and tryptophan biosynthesis, phenylalanine metabolism, tyrosine metabolism, and TCA cycle.
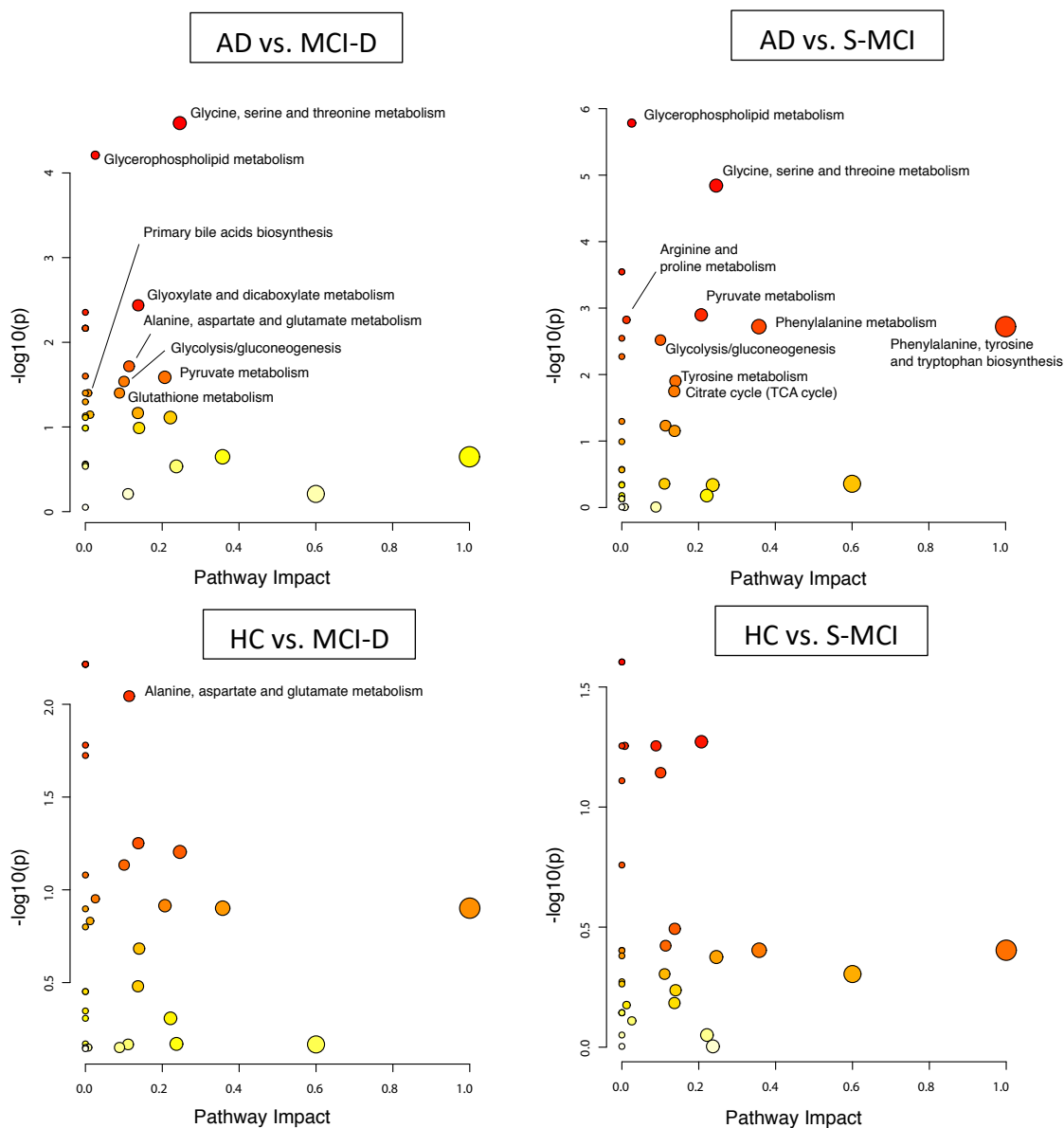
Figure 5. Metabolic pathways altered in AD and MCI subsets. Only the significant metabolic pathways are labelled (p-value < 0.05, impact > 0). Colour and size of the circles indicate the p-value and impact index, respectively: $-\log_{10}(p)$ is represented from higher values (red) to lower values (yellow) and the pathway impact is reflected in the size of the circles from smaller circles (lower impact) to bigger circles (higher impact).

These results suggest that these four pathways and the involved metabolites should be further studied in an attempt to determine which MCI patients are at higher risk of developing AD. Phenylalanine, tyrosine, acetoacetate, citrate and pyruvate are metabolites involved in the routes above mentioned. Phenylalanine and pyruvate, as described before, have shown a progressive increase in its concentration from HC to AD passing though the different MCI subgroups (Figure 4), and both metabolites have been demonstrated to be important in the discrimination between dementia and HC and between dementia and MCI according to the PLS-DA models (Table 2). The disturbance of some of these pathways and metabolites, and its related biological processes has been previously reported.

Current knowledge about Alzheimer's disease points out the alterations in the metabolism glucose,[42] highlighting the dysfunction of glycolysis,[43] as well as successive processes involved in the energetic metabolism, such as the decreased functioning of the pyruvate dehydrogenase complex.[44] In this work, an increment in the concentration of pyruvate already in MCI patients has been observed, being higher the increment in MCI-D patients than in S-MCI. In this context, it is also remarkable the differences found in the TCA cycle between dementia patients and HC and dementia patients and stable MCI, but not between AD and MCI-D / Incp-D. The TCA cycle is the main pathway of glucose oxidation in the brain, and a diminution in isocitrate dehydrogenase and α-ketoglutarate dehydrogenase complex has been previously described.[45] Phenylalanine metabolism has also been previously reported to be modified in dementia serum samples,[46] and brain tissue.[47]

On the other hand, the comparison of HC and stable MCI pathways did not show any relevant difference, whereas the alanine, aspartate and glutamate metabolism, was altered in the comparison of HC group and MCI-to-dementia/incipient dementia (Figure 5). Glutamine, citrate, pyruvate, and alanine are the metabolites involved in this pathway. The implications of pyruvate and associated pathways has already been

addressed in this work. However, other studies have involved some of these metabolites with dementia. For example, alterations in the levels of circulating glutamine [48] and impairments in the glutamate/glutamine cycle [49] has been previously related to the development of dementia. This impairment of the glutamate/glutamine cycle are related with changes in mood, behaviour and memory loss among others, all of them are alterations observed in AD patients.[49] All together these results show that there are changes in the MCI group that could be further analysed in order to have an earlier predictive diagnosis of those patients that would evolve to more advanced stages of dementia. To confer clinical applicability potential to the analysis performed by NMR spectroscopy, threshold values and metabolic signatures should be defined and stablished for a molecular diagnosis and prognosis of AD. To do that, a study with a big enough cohort should be developed ensuring that the clinical criteria and analysis process are standardized.

5. Conclusions

In this study the differences between serum samples from HC, AD patients and MCI patients with different levels and progressions have been analyzed by ^{1}H-NMR spectroscopy. Two predictive models have been generated to discriminate between AD and HC samples and AD and MCI samples, with high levels of sensitivity and specificity (93.75% and 94.75% for discrimination of AD and HC, and 100 and 82.35% for AD and MCI), according to the metabolic information in the NMR spectra. These models could be of use as a non-invasive tool to support dementia diagnosis. Furthermore, significant differences between AD and HC have been found in the relative concentration of most of the analyzed metabolites, highlighting the impact that the cellular metabolism has in dementia. Moreover, significant differences were also found in 12 metabolites when comparing MCI and dementia serum samples. Furthermore, differences within the MCI group in agreement with the clinical evolution have been found, which would allow to find biomarkers that could help to determine which MCI patients would progress to

dementia. An increase in phenylalanine, lysine and pyruvate and a progressive decrease in the concentration of choline, can be observed in the progression to dementia. On the other hand, alanine, aspartate and glutamate metabolism; pantothenate and CoA biosynthesis; and beta-alanine metabolism have been found altered when comparing HC and MCI-D whereas no pathway was altered between HC and S-MCI, which has allowed us to determine some differences in the metabolism of the different kind of patients inside the MCI group. What is described here could be a starting point to explore how the development of MCI to dementia could be effectively predicted by the study of serum using NMR spectroscopy. Future projects include the follow-up of the patients here studied and the increment of the number of patients to confirm the results.

6. Bibliography

1.   Olajide, O. A. & Sarker, S. D. Alzheimer's disease: natural products as inhibitors of neuroinflammation. *Inflammopharmacology* **28**, 1439–1455 (2020).

2.   Gauthier, S., Rosa-Neto, P., Morais, J. A. & Claire, W. World Alzheimer Report 2021: Journey through the diagnosis of dementia. *Alzheimer's Dis. Int.* (2021).

3.   Lucey, B. P. It's complicated: The relationship between sleep and Alzheimer's disease in humans. *Neurobiol. Dis.* **144**, 105031 (2020).

4.   Vignoli, A. *et al.* Fingerprinting Alzheimer's Disease by 1H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J. Proteome Res.* **19**, 1696–1705 (2020).

5.   Wurtman, R. Biomarkers in the diagnosis and management of Alzheimer's disease. *Metabolism.* **64**, S47–S50 (2015).

6.   Gauthier, S. *et al.* Mild cognitive impairment. *Lancet* **367**, 1262–1270 (2006).

7.   Duara, R., Loewenstein, D. A., Wright, C., Crocco, E. & Varon, D. Mild Cognitive Impairment. *Dementia* 77–95 (2013).

8.   Larkin, H. D. Lecanemab Gains FDA Approval for Early Alzheimer Disease. *Jama* **329**, 363 (2023).

9.   Reardon, S. FDA approves Alzheimer's drug amid safety concerns. *Nature* **613**, 227–228 (2023).

10.   Couzin-Frankel, J. Alzheimer's drug approval gets a mixed reception: FDA allows use of antibody despite ongoing debates over its benefits and dangers. *Science (80-.).* **379**, 126–127 (2023).

11.   Mattsson, N. *et al.* CSF Biomarkers and Incipient Alzheimer Disease in patients with mild cognitive impairment. *JAMA* **302**, 385–393 (2009).

12.   Zheng, H. *et al.* Tissue-Specific Metabolomics Analysis Identifies the Liver as a Major Organ of Metabolic Disorders in Amyloid Precursor Protein/Presenilin 1 Mice of Alzheimer's Disease. *J. Proteome Res.* **18**, 1218–1227 (2019).

13.   Peng, B., Li, H. & Peng, X. X. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* **6**, 628–637 (2015).

14.   D'Alessandro, A., Giardina, B., Federica, G., Timperio, A. M. & Zolla, L. Clinical metabolomics: The next stage of clinical biochemistry. *Blood Transfus.* **10**, 19–24 (2012).

15.   Duarte, I. F., Diaz, S. O. & Gil, A. M. NMR metabolomics of human blood and urine in disease research. *Journal of Pharmaceutical and Biomedical Analysis* **93**, 17–26 (2014).

16.   Fuss, T. L. & Cheng, L. L. Evaluation of cancer metabolomics using ex vivo high resolution magic angle spinning (HRMAS) magnetic resonance spectroscopy (MRS). *Metabolites* **6**, 11 (2016).

17.   Kast, R. E. *et al.* Emerging technology: Applications of Raman spectroscopy for

prostate cancer. *Cancer Metastasis Rev.* **33**, 673–693 (2014).

18. Čuperlović-Culf, M., Barnett, D. A., Culf, A. S. & Chute, I. Cell culture metabolomics: Applications and future directions. *Drug Discov. Today* **15**, 610–621 (2010).

19. Ward, J. L. *et al.* An inter-laboratory comparison demonstrates that [1H]-NMR metabolite fingerprinting is a robust technique for collaborative plant metabolomic data collection. *Metabolomics* **6**, 263–273 (2010).

20. Ibáñez, C. *et al.* Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry metabolomics. *Anal. Chem.* **84**, 8532–8540 (2012).

21. Jääskeläinen, O. *et al.* Metabolic profiles help discriminate mild cognitive impairment from dementia stage in Alzheimer's disease. *J. Alzheimer's Dis.* **74**, 277–286 (2020).

22. Kurbatova, N. *et al.* Urinary metabolic phenotyping for Alzheimer's disease. *Sci. Rep.* **10**, 1–17 (2020).

23. Olazarán, J. *et al.* A blood-based, 7-metabolite signature for the early diagnosis of Alzheimer's disease. *J. Alzheimer's Dis.* **45**, 1157–1173 (2015).

24. Figueira, J., Adolfsson, R., Adolfsson, A. N., Nyberg, L. & Öhman, A. Serum Metabolite Markers of Dementia Through Quantitative NMR Analysis: The Importance of Threonine-Linked Metabolic Pathways. *J. Alzheimer's Dis.* **69**, 763–774 (2019).

25. Yilmaz, A. *et al.* A Community-Based Study Identifying Metabolic Biomarkers of Mild Cognitive Impairment and Alzheimer's Disease Using Artificial Intelligence and Machine Learning. *J. Alzheimer's Dis.* **78**, 1381–1392 (2020).

26. Peavy, Guerry, M. *et al.* Neuropsychological assessment of severely demented elderly: The severe cognitive impairment profile. *Arch Neurol* **53**, 367–372

(1996).

27.   Harrell, L. E., Marson, D., Chatterjee, A. & Parrish, J. A. The severe mini-mental state examination: A new neuropsychologic instrument for the bedside assessment of severely impaired patients with Alzheimer disease. *Alzheimer Dis. Assoc. Disord.* **14**, 168–175 (2000).

28.   Mougias, A. A., Christidi, F., Kiosterakis, G., Messinis, L. & Politis, A. Dealing with severe dementia in clinical practice: A validity and reliability study of Severe Mini-Mental State Examination in Greek population. *Int. J. Geriatr. Psychiatry* **33**, 1236–1242 (2018).

29.   McKhann, G. M. *et al.* 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. *Alzheimers Dement* **7**, 263–269 (2012).

30.   Petersen, R. C. *et al.* Mild cognitive impairment: Clinical characterization and outcome. *Arch. Neurol.* **56**, 303–308 (1999).

31.   Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).

32.   Govindaraju, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **13**, 129–153 (2000).

33.   Martínez-Bisbal, M. C. *et al.* 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed.* **17**, 191–205 (2004).

34.   Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

35.   Gholamy, A., Kreinovich, V. & Kosheleva, O. Why 70/30 or 80/20 Relation

Between Training and Testing Sets: A Pedagogical Explanation. *Dep. Tech. Reports* **1209**, 1–6 (2018).

36.   Adler, N. & Yazhemsky, E. Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *Eur. J. Oper. Res.* **202**, 273–284 (2010).

37.   Mehmood, T., Liland, K. H., Snipen, L. & Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118**, 62–69 (2012).

38.   Chong, J. *et al.* MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

39.   Albanese, M. *et al.* Cerebrospinal fluid lactate is associated with multiple sclerosis disease progression. *J. Neuroinflammation* **13**, 36 (2016).

40.   Zebhauser, P. T. *et al.* Cerebrospinal fluid lactate levels along the Alzheimer's disease continuum and associations with blood-brain barrier integrity, age, cognition, and biomarkers. *Alzheimer's Res. Ther.* **14**, 61 (2022).

41.   Yan, X., Hu, Y., Wang, B., Wang, S. & Zhang, X. Metabolic Dysregulation Contributes to the Progression of Alzheimer's Disease. *Front. Neurosci.* **14,** 530219 (2020).

42.   Hipkiss, A. R. Aging, Alzheimer's disease and dysfunctional glycolysis; Similar effects of too much and too little. *Aging Dis.* **10**, 1328–1331 (2019).

43.   Sorbi, S., Bird, E. D. & Blass, J. P. Decreased pyruvate dehydrogenase complex activity in Huntington and Alzheimer brain. *Ann. Neurol.* **13**, 72–78 (1983).

44.   Bubber, P., Haroutunian, V., Fisch, G., Blass, J. P. & Gibson, G. E. Mitochondrial abnormalities in Alzheimer brain: Mechanistic implications. *Ann. Neurol.* **57**, 695–703 (2005).

45.   Sun, C. *et al.* Serum metabolomic profiling in patients with Alzheimer disease and amnestic mild cognitive impairment by GC/MS. *Biomed. Chromatogr.* **34**

e4875 (2020).

46.   Liu, P. *et al.* Phenylalanine metabolism is dysregulated in human hippocampus with Alzheimer's Disease related pathological changes. *J. Alzheimer's Dis.* **83**, 609–622 (2021).

47.   Adams, C. D. Circulating glutamine and Alzheimer's Disease: A mendelian randomization study. *Clin. Interv. Aging* **15**, 185–193 (2020).

48.   Robinson, S. R. Neuronal expression of glutamine synthetase in Alzheimer's disease indicates a profound impairment of metabolic interactions with astrocytes. *Neurochem. Int.* **36**, 471–482 (2000).

49.   Díaz-Orueta, U. *et al.* Generalization of Results from Cognitive Stimulation Programs to Real Life Is it Possible? *Alzheimer's Dis. Res. J.* **2**, 183–306 (2010).

7. Supporting Information

Table S1. [1]H Chemical shifts on 1D human serum spectra. Assignation of metabolites.

Number assignment	Metabolite	Group	Chemical Shift (ppm)	Multiplicity and J coupling (Hz)
1	Mixed lipoproteins	CH_3	0.83	
2	Isoleucine	δCH_3	0.92	t, J = 7.41
3	Leucine	δCH_3	0.93	t, J = 5.89
4	Valine	γCH_3	0.98	d, J = 7.2
3	Isoleucine	γCH_3	1.00	d, J = 7.00
5	Ethanol	CH_2	1.16	t, J = 7.08
6	Mixed lipoproteins	CH_2	1.26	
7	Lactate	CH_3	1.32	d, J = 7.00
8	Alanine	βCH_3	1.48	d, J = 7.40
9	Lysine	γCH_2	1.70	m
9	Lysine	$3CH_2$	1.91	m
10	N-acetylglucosamine	CH_3	2.03	s
11	Glutamine	βCH_2	2.13	m
12	Acetone	CH_3	2.22	s
13	Acetoacetate	CH_3	2.27	s
14	Pyruvate	CH_3	2.36	s
11	Glutamine	βCH_2 u	2.44	m
15	Citrate	CH_2	2.52	d, J = 15.80
15	Citrate	CH_2	2.65	d, J = 15.80
9	Lysine	ϵCH_2	3.05	m
16	Choline/ glycerophosphocholine	$-N^+-(CH_3)_3$	3.19	s
17	Acetylcholine	$-N^+-(CH_3)_3$	3.21	s
18	Glucose	2CH	3.25	dd, J = 9.22, 8.06

Number assignment	Metabolite	Group	Chemical Shift (ppm)	Multiplicity and J coupling (Hz)
19	Methanol	CH	3.35	s
18	Glucose	C4H	3.42	m
18	Glucose	C3H	3.49	m
20	Glycine	αCH	3.55	s
21	Threonine	αCH	3.59	d, J = 4.88
4	Valine	αCH	3.60	d, J = 4.33
22	Glycerol	1,3 CH$_2$OH	3.64	dd, J = 11.70, 4.30
5	Ethanol	CH$_2$OH	3.67	q, J = 7.07
18	Glucose	C6H u	3.73	m
9	Lysine	αCH	3.74	m
18	Glucose	C6H u	3.77	m
8	Alanine	αCH	3.78	m
18	Glucose	C6H d	3.83	m
18	Glucose	C6H d	3.90	dd, J = 12.29, 2.11
23	Creatine	CH$_2$	4.04	s
7	Lactate	CH	4.11	c, J = 7.00
18	Glucose	C1H	5.22	d, J = 3.88
24	Unsaturated Fatty Acids	CH	5.28	
25	Tyrosine	CH 3,5	6.88	m
26	Phenylalanine	CH 3,5	7.41	m
27	Histidine	2CH	7.79	s
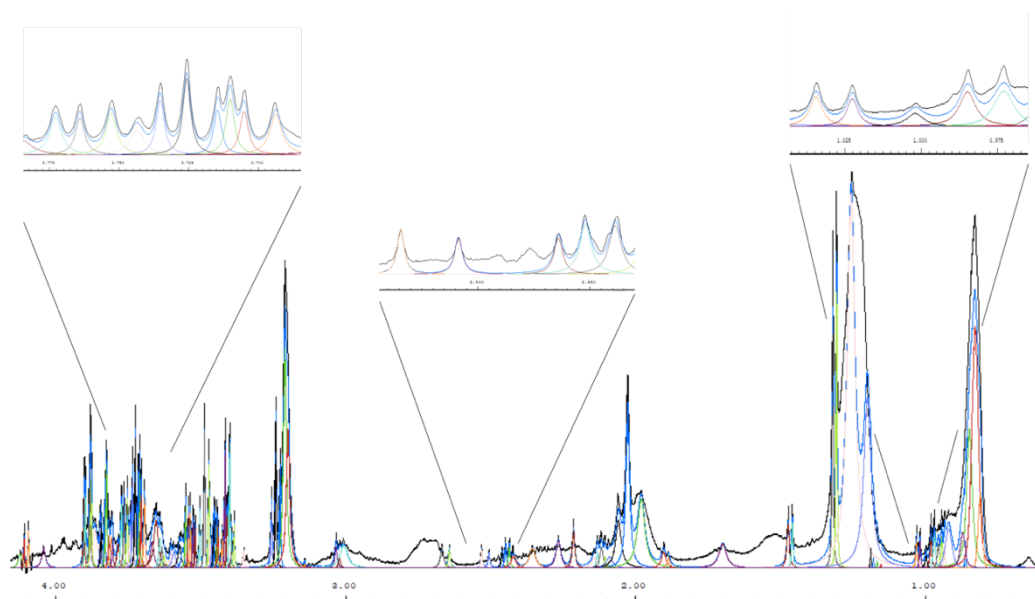
Figure S1. Deconvolution of the aliphatic region of the spectra preformed using AMIX 4.0.2 software. Three regions are shown in detail (0.95-1.05 ppm, 2.43-2.60 ppm, 3.68 -3.78 ppm)
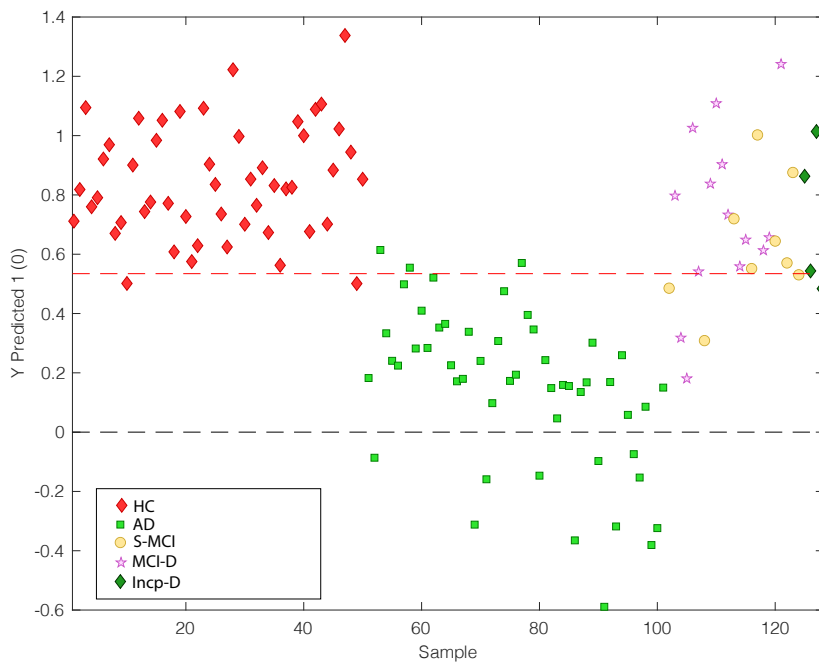
Figure S2. Projection of the MCI subgroups into the HC vs AD model

Chapter 2 | Highly Sensitive and Specific
Approach to Identify Metabolites and Pathways in
Early Alzheimer Disease

Highly Sensitive and Specific Approach to Identify Metabolites and Pathways in Early Alzheimer Disease

María Ovejero-Sánchez[1,‡], Marina Botello-Marabotto[1,2,3,‡], Carmen Peña-Bautista[4], Consuelo Cháfer-Pericás[4,*], M.Carmen Martínez-Bisbal[1,2,3,5,*], Andrea Bernardos[1,3,6,7], Miguel Baquero[4], Ramón Martínez-Máñez[1,2,3,6,7]

1. Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico, Universitat Politècnica de València, Universitat de València, Valencia, Spain.

2. Unidad Mixta de Investigación en Nanomedicina y Sensores, Universitat Politècnica de València-Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

3. CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III

4. Grupo de Investigación en Enfermedad de Alzheimer, Instituto de Investigación Sanitaria La Fe, Valencia, Spain.

5. Departamento de Química Física, Facultad de Química, Universitat de València, Burjassot, Spain.

6. Unidad Mixta UPV-CIPF de Investigación en Mecanismos de Enfermedades y Nanomedicina, Universitat Politècnica de València, Centro de Investigación Príncipe Felipe, Valencia, Spain.

7. Departamento de Química, Universitat Politècnica de València, Valencia, Spain.

‡ These two authors (M.O.-S.;M.B.-M.) contributed equaly

* Corresponding authors: Consuelo Cháfer-Pericás (m.consuelo.chafer@uv.es) and M. Carmen Martínez-Bisbal (carmen.martinez-bisbal@uv.es
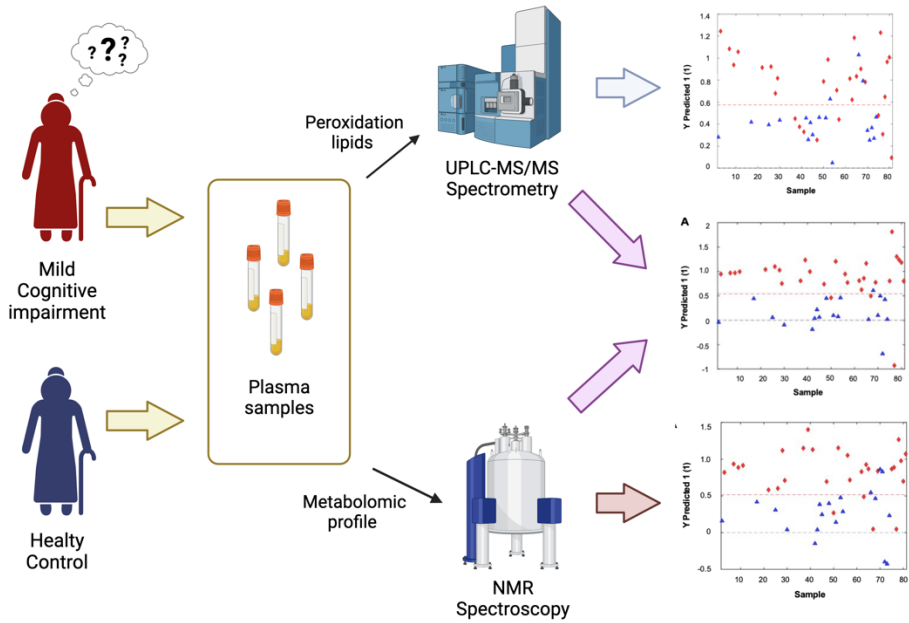
Submitted

Graphical Abstract

My contribution

I performed the preparation of the samples, the deconvolution of the signals from the spectra, the discussion, and participated in the writing of the original draft.

1.Abstract

Alzheimer's Disease (AD) is a neurodegenerative disorder and the most common cause of dementia. It is characterized by the accumulation of beta-amyloid plaques and tau tangles. Clinical diagnosis is based on the evaluation of cognitive impairment, neuroimaging techniques and cerebrospinal fluid-specific analysis. Unequivocal diagnosis of AD is often complex and requires the use of invasive techniques. Preclinical and early stages of AD (Mild Cognitive Impairment (MCI-AD)) already show functional and structural changes. This work aimed to identify metabolites by minimally invasive techniques that discriminate between patients with early stages of the disease (MCI-AD) and healthy controls to establish a simple method for the identification and monitoring of early AD patients.

Forty-eight participants (29 MCI-AD patients and 19 healthy controls) from the Neurology Unit of the University and Polytechnic Hospital La Fe (Valencia, Spain) are included in this study. Plasma samples were collected. From them, lipid peroxidation metabolites were determined using Ultra-Performance Liquid Chromatography coupled to tandem Mass Spectrometry (UPLC-MS/MS), and plasma metabolomic profiles were obtained by proton Nuclear Magnetic Resonance (^{1}H-NMR). A Partial Least Square Discriminant Analysis (PLS-DA) was performed to identify metabolite differences between participant groups.

A predictive model including both metabolites determined by ^{1}H-NMR and lipid peroxidation metabolites was obtained. This model shows high specificity and sensitivity, discriminating between MCI-AD patients and healthy controls. Pathways analysis reveals that taurine and hypotaurine metabolism, bile acid biosynthesis, and arginine and proline metabolism are altered in early AD.

Altogether, these results demonstrate that ^{1}H-NMR and UPLC-MS/MS techniques could represent a useful approach for detecting the early stages of AD with clinical potential.

2.Introduction

Alzheimer's disease (AD) is the most common type of dementia. Dementia affects more than 55 million people worldwide, and in 70% of the cases AD is the underlying cause.[1] AD is a neurodegenerative disorder characterized by cognitive decline and progressive loss of functional capacities. Mild cognitive impairment (MCI-AD) can be defined as a premature cognitive decline.[2] MCI patients can stay with this degree of cognitive loss through life, or progress to more advanced stages of dementia. The annual progression rate of MCI to AD is in the range of 8% to 15%.[3] The early identification of subjects with MCI is of great interest as it would allow the application of some therapies in the future that may be effective, slowing down the progress of dementia. To date, there are no effective treatments to restore the cognitive decline in dementia patients and recently the FDA approved anti-amyloid immunotherapies, have revealed a somewhat limited clinical advantage as they marginally impede the progression of the disease in its early stages, despite certain safety considerations.[4–6]

Currently, the diagnosis of AD is based on neuroimaging studies, neuropsychological evaluation (e.g. Mini-Mental State Examination (MMSE) [7] or clinical dementia rating (CDR) [8]), and levels of certain biomarkers (ß-amyloid, Tau, and phosphorylated Tau (p-Tau) in the cerebrospinal fluid (CSF)). Some of these pathophysiological alterations may start between 10 and 15 years before the clinical onset.[9] In this context, the identification of new and early AD biomarkers, which could be determined in a less invasive procedure than CSF extraction, would be of great value for both treatment and diagnosis.

In some cases, the alteration of the metabolism in the brain could be associated to AD,[10] and the use of metabolomics to investigate AD potential biomarkers has already been assessed.[11–14] In a previous work performed by our group with serum samples from MCI, AD patients, and healthy controls, we found that phenylalanine, lysine, and

pyruvate concentrations increased along the dementia progression, whereas choline concentration decreased.[15]

On the other hand, oxidative stress plays an important role in the development of AD, being involved in inflammation, neurotoxicity and cell death, and the presence of some oxidation products has been studied as early biomarkers of the disease.[16,17] Specifically, lipid peroxidation products showed promising results as diagnostic biomarkers of AD in different stages,[18–20] reflecting brain structural alterations.[21]

According to these previous results, the combination of lipid peroxidation and metabolic disturbances might improve the discovery of new biomarkers for AD. In this work, we aimed to identify potential biomarkers of early AD from circulating metabolites in plasma. For this, a combined strategy applying the two most extended techniques for metabolomics analysis (nuclear magnetic resonance spectroscopy (NMR spectroscopy), and ultra-performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS)) is carried out in MCI-AD patients and healthy controls.

3. Materials and Methods

3.1 Participant selection and sample collection

Participants (n=48) from the Neurology Unit of the University and Polytechnic Hospital La Fe (Valencia, Spain), aged between 50-75 years, were included in this study. They were classified as MCI-AD (n=29) and control (n=19) groups according to cerebrospinal fluid (CSF) and imaging biomarkers and neuropsychological assessment following the National Institute on Aging-Alzheimer's Association (NIA-AA) recommendations.[22–24] MCI-AD patients were positive for CSF biomarkers (ß-amyloid, Tau, and p-Tau) or neuroimaging (amyloid PET), while controls were negative. According to the neuropsychological evaluation (clinical dementia rating (CDR), and Repeatable Battery

for the Assessment of Neuropsychological Status-Delayed Memory (RBANS.DM)), control group showed standard score, whereas MCI-AD patients showed cognitive complaints without or with minor daily living activities impairment. Informed consent was obtained from all participants, and the study protocol (project reference number 2020-462-1) was approved by the Ethics Committee (CEIC) from Health Research Institute La Fe (Valencia, Spain).

Blood samples were collected from all participants in tubes containing ethylenediaminetetraacetic acid (EDTA), and centrifuged for 15 min at 1160 g. Butylated hydroxytoluene (BHT) (0.25% (w/v) in ethanol) was added to the plasma samples to avoid further oxidation and samples were stored at −80 °C until the analysis.

3.2 Lipid peroxidation metabolites determination by means of UPLC-MS/MS

A panel of 18 lipid peroxidation metabolites (15(R)-15-F2t-IsoP; PGE2; 2,3-dinor-15-epi-15-F2t-IsoP; 15-keto-15-E2t-IsoP; 15-keto-15-F2t-IsoP; 15-E2t-IsoP; 5-F2t-IsoP; 15-F2t-IsoP; PGF2α; 4(RS)-4-F4t-NeuroP; 1a,1b-dihomo-PGF2α; 10-epi-10-F4t-NeuroP; 14(RS)-14-F4t-NeuroP; Ent-7(RS)-7-F2t-dihomo-IsoP; 17-F2t-dihomo-IsoP; 17-epi-17-F2t-dihomo-IsoP; 17(RS)-10-epi-SC-Δ15-11-dihomo-IsoF; 7(RS)-ST-Δ8-11-dihomo-IsoF) and four total parameters (isoprostanes, neuroprostanes, isofurans and neurofurans) were analysed by UPLC-MS/MS after sample treatment as described by Peña Bautista et al.[18] Briefly, 400 μL of plasma with internal standard (IS) (PGF2α-D4 and D4-10-epi-10-F4t-NeuroP) were subjected to basic hydrolysis with potassium hydroxide solution (15% w/v), protein precipitation with hydrochloric acid and purification by solid phase extraction after adjusting the pH to 7.

3.3 Metabolites identification by means of NMR

Plasma samples (200 µl) were lyophilized using a speed vacuum concentrator (mi Vac) from Genevac LTD (Ipswich, United Kingdom). The samples were then stored at -80ºC until their analysis. Finally, the 81 lyophilized samples were resuspended in 200 µl of D_2O. 110 µl of this solution were mixed with 55 µl of phosphate buffer (pH 7.4) in deuterated water and sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) 1 mM as internal standard for chemical shift referencing. The total volume (165 µl) of the solution was introduced into a 3 mm NMR tube.

After sample preparation, NMR spectra were acquired using a Bruker Avance DRX 600 MHz spectrometer (Bruker GmbH, Rheinstetten, Germany) equipped with a 5 mm TCI (^1H, ^{15}N ^{13}C) cryoprobe, a BCU05 temperature unit, and a cooled SampleJet for automated sample analysis at the NMR facility of the Príncipe Felipe Research Center Foundation (CIPF). Samples were measured at 300 ºK (27 ºC). For each sample, a 1D ^1H-NMR spectrum was acquired using the Carr-Purcell-Meiboom-Gill (cpmg) pulse sequence (cpmgpr1d from the Bruker pulse sequence library). This pulse sequence effectively minimizes contributions from high molecular weight molecules, such as proteins or other macromolecules, due to their short transverse relaxation times (T2). The relaxation delay was 0.5 ms applied 160 times resulting in an echo time of 160 ms. The acquired 1D spectra were then processed using TopSpin 4.0.0 (Bruker BioSpin Corporation). Processing included the application of an exponential line-broadening function of 0.5 Hz followed by a Fourier transform. In addition, phasing, baseline correction, and chemical shift referencing to the trimethylsilyl signal of DSS at 0.0 ppm were performed. Spectral assignments were made using the Human Metabolome Data Base (HMDB),[25] and relevant data from the literature.[26,27] After processing, meaningful signals within the cpmg spectra underwent deconvolution using AMIX 4.0.2 software (Bruker BioSpin Corporation). Signals in the region between 4.5 and 5.0 ppm, as well as those with chemical shifts lower than 0.5 ppm or higher than 8.5 ppm, were excluded from the analysis. A total of 292 signals from the 1D spectra were selected and included

for deconvolution. A mixed Gaussian/Lorentzian variable function was applied for the deconvolution process, followed by obtaining integrals for all cpmg spectra. Each integral was normalized by the sum of integrals in each spectrum.

3.4 Multivariate statistical analysis

Differences between MCI-AD and control patients were determined by multivariate statistical analysis. *Solo* 9.2 software (2023, Eigenvector Research, Inc., Manson, WA USA 98831; available at http://www.eigenvector.com) was used for statistical analysis of normalized data. Partial Least Squares Discriminant Analysis (PLS-DA) was used to generate predictive models that discriminated within MCI-AD and control groups. Data from metabolomic profiles and/or lipid peroxidation metabolites were included in several analyses. For these analyses, the data were randomly divided into two groups: training (2/3 of the data were included to calculate the model) and validation (1/3 of the data was used to validate the calculated model).[28] In addition, the number of Principal Components (PCs) was determined using cross-validation (using venetian blinds). Taking into account that reducing the number of variables could improve classification accuracy,[29,30] only one signal was selected from NMR data for those metabolites with more than one signal assigned, reducing finally the number of variables to 158 NMR signals. Furthermore, variables with a Variable Importance in Projection (VIP) value > 0.65 were selected for lipid peroxidation metabolites and for metabolomic profiles along with lipid peroxidation metabolites analysis, and variables with a VIP>1 were selected for metabolomic profiles analysis. Once selected, PLS-DA analyses were performed using NMR, UPLC-MS/MS, or both data sets, and the optimal number of PCs was calculated by cross-validation. Sensitivity, specificity, and Area Under the ROC Curve (AUC) were then calculated in validation sets to determine the goodness of the model to discriminate between the data sets.

3.5 Univariate statistical analysis

Differences between metabolites included in the best-fit models that discriminate between MCI-AD and control groups (obtained in PLS-DA analyses) were assessed for statistical significance. Before performing the comparison, normality distribution was tested using Saphiro Wilk tests. Student's two tailed t-tests were performed for variables with a normal distribution and Mann Whitney U tests for non-normal distribution. Different analyses and boxplots were obtained using jamovi for Mac version 2.2.5. The jamovi project (2021). Retrieved from https://www.jamovi.org. Statistical significance was concluded for values of p≤0.05 (**p < 0.01, *p < 0.05).

3.6 Metabolite set enrichment analysis

Potential metabolic pathways that could be involved in the pathological processes were analyzed using *Metaboanalyst.*[31] For the Metabolite Set Enrichment Analysis (MSEA), a table with the relative concentration of each metabolite (column) in each patient (row) was constructed. The HMDB ID of each metabolite was provided to perform the analysis. Metabolites without this ID were not included. During MSEA, relative-betweenness centrality was selected for topology analysis and *Homo sapiens* (KEGG) was selected as the reference metabolome. Pathways with p-values ≤0.05 and impact factors >0 were selected as representative pathways. Then, a pathway analysis (targeted) module integrating enrichment analysis and pathway topology analysis of metabolic pathways was also performed using the same table as in the previous study.

4. Results

4.1 Demographic and clinical characteristics of the participants

A total of 48 plasma samples from two groups of participants (29 MCI-AD patients and 19 controls) were studied. Table 1 describes the clinical and demographic characteristics of the participants. As expected, neuropsychological tests (CDR, RBANS.DM) and CSF biomarker levels (ß-amyloid, Tau, p-Tau) showed statistically significant differences between groups. No differences were found for variables other than age and education level. Control participants tended to be younger and with higher education level.

Table 1. Demographic and clinical characteristics of the two groups of participants (MCI-AD or controls).

		MCI-AD	controls	p-value
Number (n, %)		29 (35.8)	19 (23.46)	
Age (years, median, IQR)		70 (65,73)	63 (61,66)	0.000
Sex (female (n, %))		22 (76%)	9 (47%)	0.127
Educational level	**Primary**	18 (62%)	2 (11%)	
	Secondary	9 (31%)	7 (37%)	0.000
	Universitary	2 (7%)	10 (53%)	
Drugs (n, %)	**Statins**	16 (67%)	9 (47%)	0.440
	Fibrates	4 (20%)	2 (12%)	0.362
	Benzodiazepines	4 (16%)	1 (5%)	0.191
	Antidepressants	1 (4%)	0 (0%)	0.541
	Antiepileptics	0 (0%)	0 (0%)	0.498
	Anticoagulants	0 (0%)	0 (0%)	0.999
	Antihypertensives	12 (48%)	4 (21%)	0.157
	Corticoids	0 (0%)	0 (0%)	0.498
	Anti-inflammatories	0 (0%)	0 (0%)	0.999
Comorbidities (n, %)	**Dyslipidemia**	16 (64%)	8 (42%)	0.347
	Diabetes	2 (8%)	1 (5%)	0.870
	Hypertension	12 (48%)	6 (32%)	0.547
	Heart Disease	0 (0%)	1 (5%)	0.543
	Cerebrovascular	0 (0%)	0 (0%)	0.999
Smoke (n, %)	**Yes, or ex-smoker**	3 (12%)	7 (39%)	0.118
Alcohol consumption (n, %)		2 (8%)	2 (11%)	0.941
Depression (n, %)		7 (29%)	2 (11%)	0.111
Anxiety (n, %)		4 (17%)	1 (5%)	0.337
Amyloid β42 (pg ml^{-1} (median, IQR))		593 (472,694)	1197 (1074,1458)	0.000

t-Tau (pg ml^{-1} (median, IQR))	679 (435,1142)	196 (130,306)	0.000
p-Tau181 (pg ml^{-1} (median, IQR))	96 (65,120)	48 (36,69)	0.000
CDR (median, IQR)	0.5 (0.5,0.5)	0 (0,0)	0.000
MMSE (median, IQR)	24 (21,26)	29 (28,30)	0.000
RBANS.DM (median, IQR)	44 (40,55)	100 (95,102)	0.000

4.2 Plasma metabolomic profiles

From the NMR data, a PLS-DA analysis was carried out. An initial one principal component model was constructed using the calibrated subset, and variables with VIP>1 were selected for the next analysis. Finally, a three principal components model including 21 metabolites was constructed (Table 2). This model was then applied to the validation subset, yielding 77.8% sensitivity, and 57.1% specificity, with an AUC value of 0.7937 (Figure 1a). The variables included in this model are listed in table 2 according to their VIP score. The variables that showed a higher influence were valine, threonine, tryptophan, isoleucine, proline, malonate, kynurenine, 2-hydroxy-3-methylvalerate, and taurine. Subsequently, the relative levels of metabolites included in this model were compared (Figure 1b, Table 3). Statistically significant differences were observed for isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, and taurine (Figure 1b, Table 3). Patients with MCI-AD had lower levels of isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, and taurine, and higher levels of glutathione, cysteine, malonate, and N-nitrosodimethylamine in plasma than control group (Figure 1b, Table 3). Analysis of plasma levels of these metabolites may be useful for early detection of the disease and the identification of essential pathways with clinical potential.

Table 2. Variables participating in the model calculated from the data by ^{1}H-NMR spectroscopy and sorted by VIP value. Variables with VIP>1 are shown in *italic*.

Variable	VIP Score
Valine	*1.31*
Threonine	*1.19*
Tryptophan	*1.17*
Isoleucine	*1.11*
Proline	*1.10*
Malonate	*1.08*
Kynurenine	*1.07*
2-Hydroxy-3-methylvalerate	*1.03*
Taurine	*1.00*
Glutathione	0.98
Mannose	0.97
Glutamate	0.97
Cysteine	0.96
N-Nitrosodimethylamine	0.94
Acetate	0.92
3-Hydroxyisobutyrate	0.91
4-Hydroxybenzoate	0.90
Arginine	0.82
Ribose	0.81
Citrate	0.80
τ-Methylhistidine	0.77
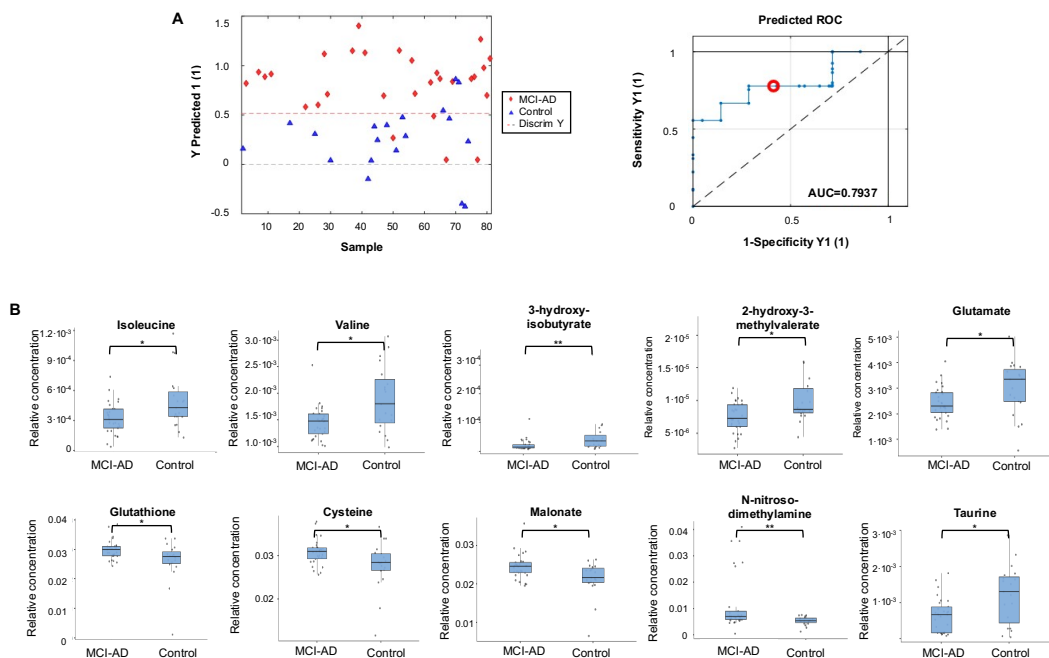
Figure 1. Analyses of plasma metabolomic profiles. (A) Multivariate analysis. PLS-DA scores and ROC curve for the classification of MCI-AD patients and controls. The prediction plot divided into the scores obtained in the calibration and validation subsets is shown in the left panel. The ROC curve for the model with an AUC value of 0.7937 for discriminating MCI-AD from controls is shown in the right panel, (B) Univariate analysis. Relative concentration of metabolites with a significant difference between MCI-AD and control groups. The median (horizontal bar) of the relative concentration of isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, and taurine for MCI-AD and control groups is shown ($**p < 0.01$, $*p < 0.05$).

Table 3. Comparison of relative metabolite concentrations determined by ^{1}H-NMR spectroscopy with significant differences between MCI-AD and control groups.

Metabolite	Relative metabolite concentration		p-value
	MCI-AD	control	
Isoleucine	$3.22 \cdot 10^{-4}$	$4.99 \cdot 10^{-4}$	0.022
Valine	$1.45 \cdot 10^{-3}$	$1.90 \cdot 10^{-3}$	0.013
3-hydroxyisobutyrate	$2.25 \cdot 10^{-5}$	$5.36 \cdot 10^{-5}$	0.007
2-hydroxy-3-methylvalerate	$7.54 \cdot 10^{-6}$	$1.03 \cdot 10^{-5}$	0.021
Glutamate	$2.44 \cdot 10^{-3}$	$3.00 \cdot 10^{-3}$	0.021
Glutathione	$3.00 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	0.015
Cysteine	$3.08 \cdot 10^{-2}$	$2.77 \cdot 10^{-2}$	0.015
Malonate	$2.44 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	0.026
N-nitrosodimethylamine	$1.17 \cdot 10^{-2}$	$5.38 \cdot 10^{-3}$	0.005
Taurine	$6.59 \cdot 10^{-4}$	$1.22 \cdot 10^{-3}$	0.017

4.3 Plasma lipid peroxidation metabolites

From the UPLC-MS/MS data, a PLS-DA analysis was performed (29 MCI-AD patients and 19 controls). From the first model built on the calibration subset with two principal components, variables with VIP>0.65 were selected, resulting in a total of 17 variables (Table 4). This model was then applied to the validation subset and 55.6% of sensitivity and 71.4% of specificity were obtained with an AUC value of 0.6667 (Figure 2a). The variables involved in this model are shown in table 4 and those with a higher impact are: total isoprostanes, 8-iso-PGE2, 10-epi-10-F4t-NeuroP, 8-iso-15R-PGF2a, total neurofurans, 1a,1b-dihomo-PGF2α, and total neuroprostanes. The relative levels of the lipid peroxidation metabolites involved in this model were then compared (Figure 2b, Table 5). Significant differences were observed for isoprostanes and neurofurans, with higher levels in MCI-AD patients (Figure 2b, Table 5), highlighting the importance of lipid peroxidation in the pathobiology of AD. Isoprostanes and neurofurans may represent potential biomarkers of early stages of this disease.

Table 4. Variables participating in the model calculated with the lipid peroxidation metabolites from UPLC-MS/MS sorted by VIP value. Data with VIP>1 is shown in *italics*.

Variable	VIP
Isoprostanes	*1.63*
8-iso-PGE2	*1.38*
10-epi-10-F4t-NeuroP	*1.28*
8-iso-15R-PGF2α	*1.15*
Neurofurans	*1.09*
1a,1b-dihomo-PGF2α	*1.09*
Neuroprostanes	*1.04*
8-iso-PGF2a	0.97
Ent-7(RS)-7-F2t-dihomo-IsoP	0.95
PGF2a	0.84
17(RS)-10-epi-SC-Δ15-11-dihomo-IsoF	0.76
17-F2t-dihomo-IsoP	0.76
8-iso-15-keto-PGE2	0.74
PGE2	0.71
8-iso-15-keto-PGF2a	0.67
7(RS)-ST-Δ8-11-dihomo-IsoF	0.65
14(RS)-14-F4t-NeuroP	0.64

Table 5. Comparison of relative lipid peroxidation metabolite concentrations determined by UPLC-MS/MS with significant differences between MCI-AD and control groups.

Metabolite	Relative metabolite concentration		p-value (test)
	MCI-AD patients	controls	
Isoprostanes	$3.94 \cdot 10^{-1}$	$2.75 \cdot 10^{-1}$	0.005
Neurofurans	$1.73 \cdot 10^{-1}$	$7.41 \cdot 10^{-2}$	0.019
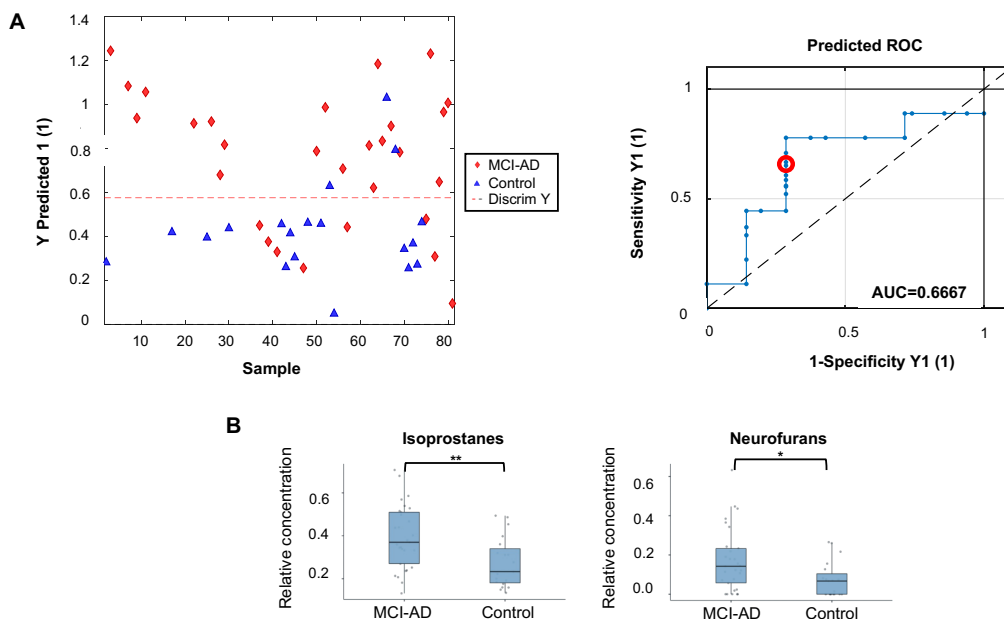
Figure 2. Analyses of lipid peroxidation metabolites: (A) Multivariate analysis. PLS-DA scores and ROC curve for the classification of MCI-AD patients and controls. The prediction plot divided into the scores obtained in the calibration and validation subsets is shown in the left panel. The ROC curve for the model with an AUC value of 0.6667 for discriminating MCI-AD from controls is shown in the right panel, (B) Univariate analysis. Relative concentration of metabolites with a significant difference between MCI-AD and control groups. The median (horizontal bar) of the relative concentration of total isoprostanes and neurofurans for MCI-AD and control groups is shown (**$p < 0.01$, *$p < 0.05$).

4.4 Integration of plasma metabolomic profiles and lipid peroxidation metabolites

From the combination of both datasets (lipid peroxidation metabolites and plasma metabolomic profiles), all MCI-AD patients (n=29) and controls (n=19) were included in the analysis. An initial single principal component model was constructed using the calibrated subset, and variables with VIP>0.65 were selected for further analysis. Finally, a five principal component model including 20 metabolites was constructed (Table 6). This model was then applied to the validation subset and 77.8% sensitivity, and 85.7% specificity were obtained with an AUC value of 0.8730 (Figure 3a). The addition of lipid peroxidation metabolites appears to improve the ability of the model to discriminate between MCI-AD and control groups. The variables included in this model are listed in table 6 according to their VIP score. Variables that showed a higher influence were 8-iso-PGE2, valine, tryptophan, kynurenine, citrate, and glutathione. Subsequently, the relative levels of metabolites included in this model were compared (Figure 3b, Table 7). Significant differences were observed for isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, taurine, proline, and isoprostanes (Figure 3b, Table 7). Patients with MCI-AD had lower levels of isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, taurine, and proline, and higher levels of glutathione, cysteine, malonate, N-nitrosodimethylamine, and isoprostanes than healthy controls (Figure 3b, Table 7).
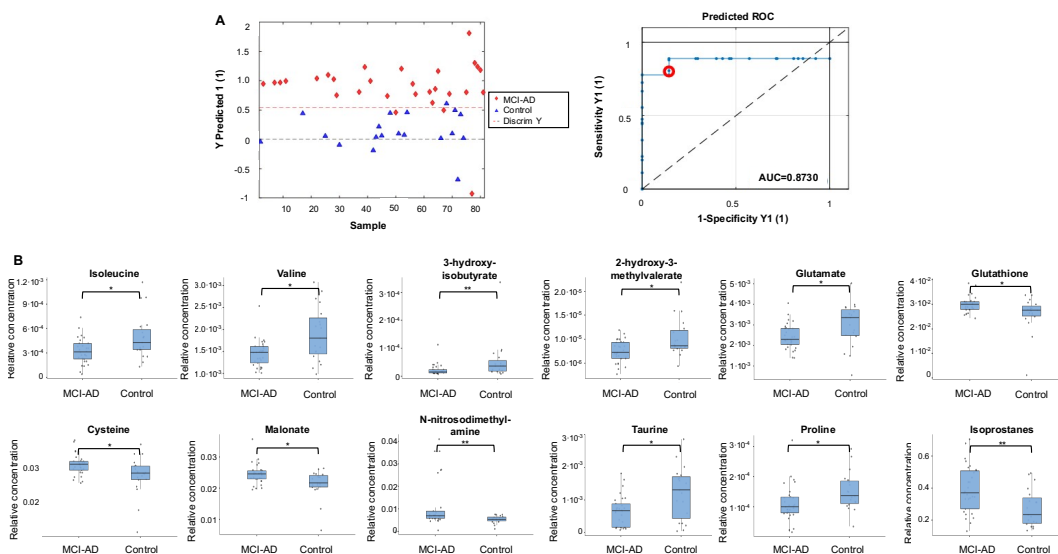
Figure 3. Analyses of plasma metabolomic profiles together with lipid peroxidation metabolites: (A) Multivariate analysis. PLS-DA scores and ROC curve for the classification of MCI-AD patients and controls. The prediction plot divided into the scores obtained in the calibration and validation subsets is shown in the left panel. The ROC curve for the model with an AUC value of 0.8730 for discriminating MCI-AD from controls is shown in the right panel, (B) Univariate analysis. Relative concentration of metabolites with a significant difference between MCI-AD and control groups. The median (horizontal bar) of the relative concentration of total isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, taurine, proline, and isoprostanes for MCI-AD and control groups is shown (**$p < 0.01$, *$p < 0.05$).

Table 6. Variables participating in the model calculated with metabolites from ^{1}H-NMR spectroscopy and lipids peroxidation determined by UPLC-MS/MS and sorted by VIP value. Variables in *italic* show VIP>1.

Variable	VIP
8-iso-PGE2	*1.46*
Valine	*1.20*
Tryptophan	*1.16*
Kynurenine	*1.10*
Citrate	*1.05*
Glutathione	*1.03*
2-hydroxy-3-methylvalerate	0.98
Isoleucine	0.98
Cysteine	0.97
Isoprostanes	0.96
Malonate	0.96
Arginine	0.96
3-hydroxyisobutyrate	0.96
Taurine	0.92
Glutamate	0.91
Proline	0.91
N-nitrosodimethylamine	0.88
Glutamine	0.83
Threonine	0.79
Ribose	0.76

Table 7. Comparison of relative metabolite concentrations with significant differences between MCI-AD and control groups.

Metabolite	Relative metabolite concentration		p-value
	MCI-AD patients	controls	
Isoleucine	$3.22 \cdot 10^{-4}$	$4.99 \cdot 10^{-4}$	0.022
Valine	$1.45 \cdot 10^{-3}$	$1.90 \cdot 10^{-3}$	0.013
3-hydroxyisobutyrate	$2.25 \cdot 10^{-5}$	$5.36 \cdot 10^{-5}$	0.007
2-hydroxy-3-methylvalerate	$7.54 \cdot 10^{-6}$	$1.03 \cdot 10^{-5}$	0.021
Glutamate	$2.44 \cdot 10^{-3}$	$3.00 \cdot 10^{-3}$	0.021
Glutathione	$3.00 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	0.015
Cysteine	$3.08 \cdot 10^{-2}$	$2.77 \cdot 10^{-2}$	0.019
Malonate	$2.44 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	0.026
N-Nitrosodimethylamine	$1.17 \cdot 10^{-2}$	$5.38 \cdot 10^{-3}$	0.005
Taurine	$6.94 \cdot 10^{-4}$	$1.22 \cdot 10^{-3}$	0.017
Proline	$1.13 \cdot 10^{-4}$	$1.52 \cdot 10^{-4}$	0.013
Isoprostanes	$3.94 \cdot 10^{-1}$	$2.75 \cdot 10^{-1}$	0.005

4.5 Impaired pathways in early AD

From the different metabolites' levels found between MCI-AD and control groups. (isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, taurine, isoprostanes) (Table 3-5), the corresponding metabolic pathways were evaluated. For this purpose, metabolite set enrichment analyses were performed. In the case of the predictive model obtained from plasma metabolomic profiles, 22 metabolic pathways were altered; however, only four pathways showed significant differences between MCI-AD and control groups (Figure 4A). On the other hand, the predictive model analysis integrating plasma metabolomic profiles and lipid peroxidation metabolites revealed an alteration of 18 metabolic pathways, 3 of which showed statistically significant differences between MCI-AD and control groups. In both cases, primary bile acid biosynthesis, taurine and hypotaurine metabolism, and D-amino acid metabolism appear to be involved in the early pathogenesis of AD.
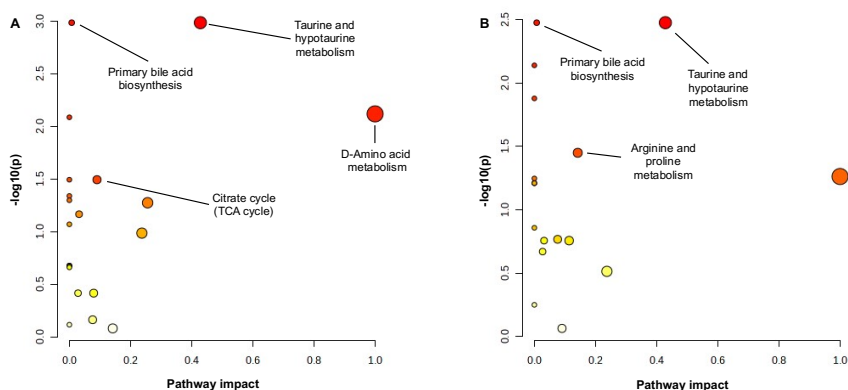
Figure 4. Metabolic pathways altered in MCI-AD and control groups. Pathways altered in the predictive model derived from plasma metabolomic profiles (A) and from plasma metabolomic profiles together with lipid peroxidation metabolites (B). Significant metabolic pathways are labelled (p<0.05, impact >0). The colour and the size of the circles indicate the p-value and impact index, respectively: -log$_{10}$(p) is represented from higher values (red) to lower values (yellow) and the pathway impact is reflected in the size of the circles from smaller circles (lower impact) to larger circles (higher impact).

5. Discussion

AD is the most prominent form of dementia, and it has a great impact on society. The identification of biomarkers using minimally invasive procedures for early diagnosis of the disease would be of great value for clinical practice. In this work, three different classification models for the diagnosis of MCI-AD have been elaborated based on the presence of metabolomic and lipid peroxidation biomarkers in plasma. The first model generated based on metabolomic biomarkers obtained by NMR was able to discriminate with a 77.8% of sensitivity and a 57.1% of specificity between control and MCI-AD patients. The metabolites isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, and taurine showed lower levels in MCI-AD while glutathione, cysteine, malonate, and N-nitrosodimethylamine were higher in comparison with the control group. In a second model, the concentration of lipid peroxidation metabolites obtained by UPLC-MS/MS was used, with the model showing 55.6% of sensitivity and 71.4% of specificity. In this case, significant differences were observed for isoprostanes and neurofurans, with higher levels in MCI-AD patients.

Finally, a combined model using both data sets improved the classification capacity of the previous models, with 77.8% sensitivity, and 85.7% specificity. The metabolites included in this combined model, showing significative different levels between groups were isoleucine, valine, 3-hydroxyisobutyrate, 2-hydroxy-3-methylvalerate, glutamate, glutathione, cysteine, malonate, N-nitrosodimethylamine, taurine, proline, and isoprostanes. It was also observed that combining UPLC-MS/MS and NMR improves the classification potential, indicating that NMR and UPLC-MS/MS are two complementary, not mutually exclusive, techniques for the study of metabolites as potential biomarkers of disease.

Alzheimer's disease could be defined as a complex metabolic disorder,[32] and in fact various observations substantiate the theory that AD is a systemic disorder characterized by disruptions in glucose metabolism, mitochondrial dysfunction, and aberrant metabolism of branched-chain amino acids (BCAAs).[33] BCAAs are involved in

a variety of processes in the brain, from neurotransmitters metabolism to protein synthesis and energy production.[34] A decrease in the relative concentration of BCAAs has been described in AD patients.[33,35], and lower valine plasma levels have been related to cognitive decline.[36] In our work, a decrease in the relative plasma concentration of valine and isoleucine was found in MCI-AD patients, suggesting that BCAA metabolism may be altered in the early stages of AD.

Glutamate is the major neurotransmitter in the brain, and it does not cross the blood-brain barrier, so it must be produced in the brain from BCAA[33]. Alterations in circulating glutamine levels[37] and impairments in the glutamate/glutamine cycle[38] have previously been associated with the development of dementia. This impairment of the glutamate/glutamine cycle is related to changes in mood, behavior, and memory loss, among other symptoms, all of which are observed in AD patients.[38] However, whether these metabolic changes are a cause, or a consequence of AD remains to be elucidated. Moreover, recent studies have suggested that changes in the gastrointestinal function in AD may lead to reduced absorption of dietary BCAAs.[35] A decrease in the relative plasma concentration of glutamate was observed in MCI-AD patients. Other authors have previously described changes in glutamine levels in AD.[15]

We have observed that several amino acids, such as taurine, proline, and cysteine, also suffer alterations in their concentration in MCI-AD; a decrease in taurine and proline and an increase in cysteine for MCI-AD patients. Taurine is involved in several functions in the central nervous system (CNS) and has a neuroprotective effect.[39] A decrease in circulating taurine levels has been associated with aging and age-related diseases.[40] In fact, taurine supplementation has been shown to have a potential therapeutic effect against AD in mice by restoring the glutamate system.[41] However, the mechanism involved in the decrease of taurine levels in AD is still unknown. Taurine can be produced from the metabolism of cysteine: L-cysteine is oxidized by cysteine dioxygenase (CDO), which generates cysteine sulfinate which is decarboxylated by cysteine sulfinic acid decarboxylase (CSD), which converts cysteine sulfinate to hypotaurine, which is converted to taurine by hypotaurine dehydrogenase.[42] In this

work, we observed an increase in cysteine levels. Impairment of an enzyme on the taurine and hypotaurine metabolism pathway could explain the differences observed in the concentrations of these metabolites; however, further studies would be needed to confirm this hypothesis.

Oxidative stress also plays an important role in neurodegenerative diseases including AD.[18,43] In fact, reactive oxygen and nitrogen species can induce lipid peroxidation of cell membrane lipids, thereby disrupting normal cell physiology.[18,43] Lipid peroxidation has been suggested to be implicated in AD since the brain has a high lipid composition and a high oxygen consumption.[18,43] In this work we have found an increase in the concentration of isoprostanes and neurofurans in MCI-AD patients. The association between isoprostanes and AD has been described previously.[21,44] Isoprostanes and neuroprostanes are the product of araquidonic acid and docosahexaenoic acid (DHA) oxidation, and they are known as biomarkers of oxidative stress and inflammation.[43,45] In fact, DHA is highly present in the grey matter of the brain and neuroprostanes and isoprostanes may reflect brain atrophy and pathophysiological mechanisms of AD.[43,45] Taken together, these results suggest that metabolomics could be a useful technique for the early diagnosis of AD-MCI and a reliable source for the search for therapeutic targets, as well as for a better understanding of the molecular mechanism underlying the disease.

6. Bibliography

1.   Nichols, E. & Vos, T. The estimation of the global prevalence of dementia from 1990-2019 and forecasted prevalence through 2050: An analysis for the Global Burden of Disease (GBD) study 2019. *Alzheimer's Dement.* **17**, 1–2 (2021).

2.   Petersen, R. C. Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* **256**, 183–194 (2004).

3.   Duara, R., Loewenstein, D. A., Wright, C., Crocco, E. & Varon, D. Mild Cognitive

Impairment. *Dementia* 77–95 (2013).

4.   Couzin-Frankel, J. Alzheimer's drug approval gets a mixed reception: FDA allows use of antibody despite ongoing debates over its benefits and dangers. *Science (80-.).* **379**, 126–127 (2023).

5.   Larkin, H. D. Lecanemab Gains FDA Approval for Early Alzheimer Disease. *Jama* **329**, 363 (2023).

6.   Reardon, S. FDA approves Alzheimer's drug amid safety concerns. *Nature* **613**, 227–228 (2023).

7.   Mougias, A. A., Christidi, F., Kiosterakis, G., Messinis, L. & Politis, A. Dealing with severe dementia in clinical practice: A validity and reliability study of Severe Mini-Mental State Examination in Greek population. *Int. J. Geriatr. Psychiatry* **33**, 1236–1242 (2018).

8.   Morris, J. C. The clinical dementia rating (cdr): Current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).

9.   Vignoli, A. *et al.* Fingerprinting Alzheimer's Disease by 1H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J. Proteome Res.* **19**, 1696–1705 (2020).

10.   Zheng, H. *et al.* Tissue-Specific Metabolomics Analysis Identifies the Liver as a Major Organ of Metabolic Disorders in Amyloid Precursor Protein/Presenilin 1 Mice of Alzheimer's Disease. *J. Proteome Res.* **18**, 1218–1227 (2019).

11.   Yilmaz, A. *et al.* Diagnostic Biomarkers of Alzheimer's Disease as Identified in Saliva using 1H NMR-Based Metabolomics. *J. Alzheimer's Dis.* **58**, 355–359 (2017).

12.   Yilmaz, A. *et al.* Targeted metabolic profiling of urine highlights a potential biomarker panel for the diagnosis of alzheimer's disease and mild cognitive impairment: A pilot study. *Metabolites* **10**, 357 (2020).

13. Figueira, J., Adolfsson, R., Adolfsson, A. N., Nyberg, L. & Öhman, A. Serum Metabolite Markers of Dementia Through Quantitative NMR Analysis: The Importance of Threonine-Linked Metabolic Pathways. *J. Alzheimer's Dis.* **69**, 763–774 (2019).

14. Kurbatova, N. *et al.* Urinary metabolic phenotyping for Alzheimer's disease. *Sci. Rep.* **10**, 21745 (2020).

15. Botello-Marabotto, M. *et al.* Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease. *Neurobiol. Dis.* **187**, 106312 (2023).

16. Fulda, S., Gorman, A. M., Hori, O. & Samali, A. Cellular stress responses: Cell survival and cell death. *Int. J. Cell Biol.* **2010**, 214074 (2010).

17. Marrocco, I., Altieri, F. & Peluso, I. Measurement and Clinical Significance of Biomarkers of Oxidative Stress in Humans. *Oxid. Med. Cell. Longev.* **2017**, 6501046 (2017).

18. Peña-Bautista, C. *et al.* Plasma lipid peroxidation biomarkers for early and non-invasive Alzheimer Disease detection. *Free Radic. Biol. Med.* **124**, 388–394 (2018).

19. Ferré-González, L., Peña-Bautista, C., Baquero, M. & Cháfer-Pericás, C. Assessment of Lipid Peroxidation in Alzheimer's Disease Differential Diagnosis and Prognosis. *Antioxidants* **11**, 551 (2022).

20. Peña-Bautista, C. *et al.* Lipid peroxidation assessment in preclinical alzheimer disease diagnosis. *Antioxidants* **10**, 1043 (2021).

21. Peña-Bautista, C., López-Cuevas, R., Cuevas, A., Baquero, M. & Cháfer-Pericás, C. Lipid peroxidation biomarkers correlation with medial temporal atrophy in early Alzheimer Disease. *Neurochem. Int.* **129**, 104519 (2019).

22. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's

Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7**, 270–279 (2011).

23.   McKhann, G. M. *et al.* 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 Dement.* **7**, 263–269 (2011).

24.   Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **14**, 535–562 (2018).

25.   Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

26.   Govindaraju, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **13**, 129–153 (2000).

27.   Martínez-Bisbal, M. C. *et al.* 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed.* **17**, 191–205 (2004).

28.   Gholamy, A., Kreinovich, V. & Kosheleva, O. Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation. *Dep. Tech. Reports* **1209**, 1–6 (2018).

29.   Martín-Gómez, A. *et al.* Guidelines to build PLS-DA chemometric classification models using a GC-IMS method: Dry-cured ham as a case of study. *Talanta Open* **7**, 100175 (2023).

30.   Mehmood, T., Liland, K. H., Snipen, L. & Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118**, 62–69 (2012).

31.   Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, 388–396 (2021).

32. Zhou, S. *et al.* Alzheimer's disease, a metabolic disorder: Clinical advances and basic model studies (Review). *Exp. Ther. Med.* **27**, 63 (2023).

33. Polis, B. & Samson, A. O. Role of the metabolism of branched-chain amino acids in the development of Alzheimer's disease and other metabolic disorders. *Neural Regen. Res.* **15**, 1460–1470 (2020).

34. Fernstrom, J. D. Branched-chain amino acids and brain function. *J. Nutr.* **135**, 1539S-1546S (2005).

35. Qian, X. H. *et al.* Investigating the causal association between branched-chain amino acids and Alzheimer's disease: A bidirectional Mendelian randomized study. *Front. Nutr.* **10**, 1103303 (2023).

36. Xiong, Y. lan, Therriault, J., Ren, S. jiang, Jing, X. jun & Zhang, H. The associations of serum valine with mild cognitive impairment and Alzheimer's disease. *Aging Clin. Exp. Res.* **34**, 1807–1817 (2022).

37. Adams, C. D. Circulating glutamine and Alzheimer's Disease: A mendelian randomization study. *Clin. Interv. Aging* **15**, 185–193 (2020).

38. Robinson, S. R. Neuronal expression of glutamine synthetase in Alzheimer's disease indicates a profound impairment of metabolic interactions with astrocytes. *Neurochem. Int.* **36**, 471–482 (2000).

39. Menzie, J., Pan, C., Prentice, H. & Wu, J. Y. Taurine and central nervous system disorders. *Amino Acids* **46**, 31–46 (2014).

40. Singh, P., Gollapalli, K., Mangiola, S., Schranner, D. & Yusuf, M. A. Taurine deficiency as a driver of aging. *Science..* **380**, 6649, (1010-1011) (2023).

41. Oh, S. J. *et al.* Evaluation of the neuroprotective effect of taurine in Alzheimer's disease using functional molecular imaging. *Sci. Rep.* **10**, 15551 (2020).

42. Froger, N. *et al.* Taurine: The comeback of a neutraceutical in the prevention of retinal degenerations. *Prog. Retin. Eye Res.* **41**, 44–63 (2014).

43.   Peña-Bautista, C., Baquero, M., Vento, M. & Cháfer-Pericás, C. Free radicals in Alzheimer's disease: Lipid peroxidation biomarkers. *Clin. Chim. Acta* **491**, 85–90 (2019).

44.   Peña-bautista, C. *et al.* Clinical utility of plasma lipid peroxidation biomarkers in alzheimer's disease differential diagnosis. *Antioxidants* **9**, 649 (2020).

45.   Peña-Bautista, C. *et al.* Isoprostanoids levels in cerebrospinal fluid do not reflect alzheimer's disease. *Antioxidants* **9**, 407 (2020).

Chapter 3 | Tear Metabolomics for the Diagnosis of Primary Open-Angle Glaucoma

Tear Metabolomics for the Diagnosis of Primary Open-Angle Glaucoma

Marina Botello-Marabotto,[1,2,3] M. Carmen Martínez-Bisbal,[1,2,3,4] *
M. Dolores Pinazo-Durán,[5,6,7] # Ramón Martínez-Máñez[1,2,3,8,9] #

1. Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), Universitat Politècnica de València – Universitat de València, Valencia, Spain.

2. Unidad Mixta de Investigación en Nanomedicina y Sensores. Instituto de Investigación Sanitaria La Fe (IISLAFE) - Universitat Politècnica de València. Valencia, Spain

3. CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Spain

4. Departamento de Química Física, Universitat de València, Valencia, Spain

5. Ophthalmic Research Unit "Santiago Grisolia"/FISABIO, Valencia, Spain

6. Cellular and Molecular Ophthalmobiology Research Group at the University of Valencia, Valencia, Spain

7. Spanish Net of Inflammatory Research (REI-RICORS) Institute of Health Carlos III, Madrid, Spain

8. Departamento de Química, Universitat Politècnica de València, Valencia, Spain

9. Unidad Mixta UPV-CIPF de Investigación en Mecanismos de Enfermedades y Nanomedicina, València, Universitat Politècnica de València, Centro de Investigación Príncipe Felipe, Valencia, Spain.

# These two authors (M.D.P.-D.; R.M.-M.) are sharing last place as the study coordinators.

 *Corresponding author: M. Carmen Martínez-Bisbal. E-mail address: carmen.martinez-bisbal@uv.es

Graphical abstract
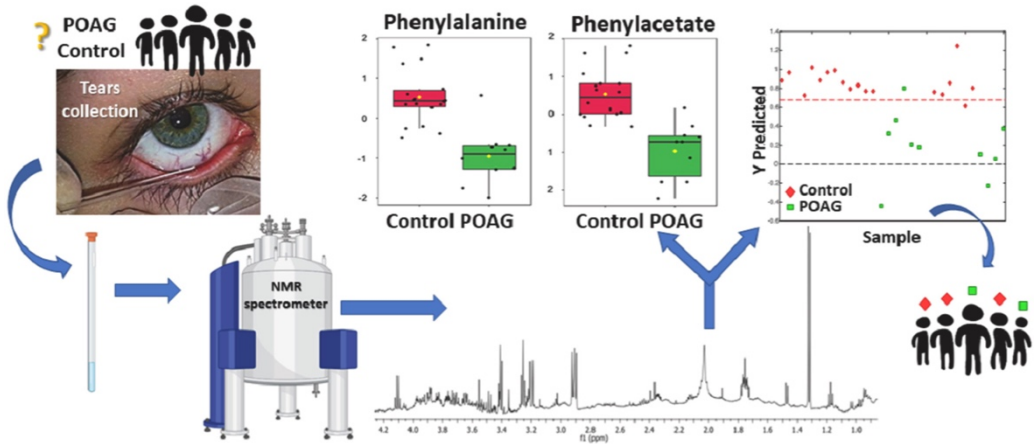
Keywords

My contribution

I participated in the preparation of the samples, performed the deconvolution of the signals from the spectra, the statistical analysis, discussion, and wrote the original draft.

1. Abstract

Primary Open-Angle Glaucoma (POAG) is the most prevalent glaucoma type, and the leading cause of irreversible visual impairment and blindness worldwide. Identification of early POAG biomarkers is of enormous value, as there is not an effective treatment for the glaucomatous optic nerve degeneration (OND). In this pilot study, a metabolomic analysis, by using proton (^{1}H) nuclear magnetic resonance (NMR) spectroscopy was conducted in tears, in order to determine the changes of specific metabolites in the initial glaucoma eyes and to discover potential diagnostic biomarkers. A classification model, based on the metabolomic fingerprint in tears was generated as a non-invasive tool to support the preclinical and clinical POAG diagnosis. ^{1}H-NMR spectra were acquired from 30 tear samples corresponding to the POAG group (n=11) and the control group (n=19). Data were analysed by multivariate statistics (partial least squares-discriminant analysis: PLS-DA) to determine a model capable of differentiating between groups. The whole data set was split into calibration (65%) / validation (35%), to test the performance and the ability for glaucoma discrimination. The calculated PLS-DA model showed an area under the curve (AUC) of 1, as well as a sensitivity of 100% and a specificity of 83.3% to distinguish POAG group versus control group tear data. This model included 11 metabolites, potential biomarkers of the disease. When comparing the study groups, a decrease in the tear concentration of phenylalanine, phenylacetate, leucine, n-acetylated compounds, formic acid, and uridine, was found. Moreover, an increase in the tear concentration of taurine, glycine, urea, glucose, and unsaturated fatty acids was observed in the POAG group. These results highlight the potential of tear metabolomics by ^{1}H-NMR spectroscopy as a non-invasive approach to support early POAG diagnosis and in order to prevent visual loss.

2. Introduction

Glaucoma is the second leading cause of blindness worldwide after cataracts, with the differential characteristic of becoming irreversible due to the progressive damage to the retinal ganglion cells (RGCs) and optic nerve fibres (ONFs). This injury leads to optic atrophy, peripheral vision decline, and loss of vision-related quality-of-life, also constituting an important matter of global socio-economic burden.[1,2] Glaucoma affects more than 70 million people worldwide and it is estimated that 111.8 million by 2040 will develop the disease.[3,4]

The most common glaucoma type is primary open angle glaucoma (POAG), which is characterized by mechanical insult due to the elevated intraocular pressure (IOP), morphological alteration of the optic nerve head, and functional landmarks, as the progressive visual field loss, constituting a neurodegenerative process, globally known as the glaucomatous OND.[2,5] Unfortunately, POAG is asymptomatic in the early stages. The only therapeutic action is the prompt initiation of elevated IOP treatment, and no neuroprotective treatments are currently available in clinical practice.[5] However, the medical-laser-surgical hypotensive glaucoma therapy does not prevent disease progression or visual impairment.[6] In spite of the advances in epidemiological and experimental studies, more research is needed on the molecular mechanisms responsible for the glaucoma development and progression.[7] Reliable biomarkers for early glaucoma diagnosis are yet to be discovered.[8]

A singular set of metabolites, named metabolome, is the result of the combination of genetic and environmental factors, among others, that is found in a biological sample. Genomics, transcriptomics, proteomics and metabolomics constitute the biological omics cascade.[9,10] The metabolome gives information more directly related to the phenotype than any other omics science.[9,10] Mainly, two sophisticated techniques, mass spectrometry (MS),[11,12] and nuclear magnetic resonance (NMR) spectroscopy, [12,13]

lead the determination of metabolomic profiles. NMR spectroscopy is a robust and reproducible method that allows to determine the metabolic composition of biofluids (blood derivatives, cerebrospinal, urine, tears, saliva, sweat, synovial, etc.) usually through very simple preparation procedures and using small amount of sample, being these two features the most relevant in the clinical context.[14,15]Both techniques need computationally intensive statistical tools to refine interpretation.[11–15]

When POAG diagnosis and prognosis is approached, a key issue to address is the selection of a suitable biologic sample to provide information on the pathology.[16–18] Most studies agree in the use of blood for biomedical glaucoma research, but also aqueous humor, vitreous body, and tear samples have been used, as recently reviewed by Tezel.[19] Our research group has conducted extensive research on the pathophysiology of ocular diseases, mainly ocular surface disorders, glaucoma, and diabetic retinopathy, by using tear samples.[20–24] Another key point in glaucoma research is to select the most suitable participants for the study, according to an accurate diagnosis, and to make an appropriate classification of the disease stage.[2,5,7,8]

Some metabolites and metabolic pathways associated to pathological processes have been reported to be altered in glaucoma, mainly regarding carbohydrates,[25] amino acids,[25,26] and fatty acids,[27] along with inflammation and neurodegeneration pathways. However, it has not yet been possible to identify a panel of reliable biomarkers that can be obtained non-invasively for translation as a diagnostic and prognostic tool to the clinical practice.

Aimed for this context, in this work we intended to develop a non-invasive method to support the diagnosis and prognosis of POAG patients at the initial stage of the disease, based on the tear metabolomic fingerprinting obtained by ^{1}H-NMR spectroscopy. The secondary objective is to search for potential biomarkers of the disease, to help

increase knowledge about the molecular processes underlying the clinically asymptomatic initial steps of the glaucomatous OND.

3. Materials and Methods

3.1 Focused topic and study characteristics

There is no cure for glaucoma. There is growing interest in identifying and validating clinical, imaging, biochemical, and molecular-genetic biomarkers that may help early detection of POAG. To progress in knowledge on the clinical and molecular basis of POAG, a collaborative multicenter analytical case-control pilot study was planned for 50 male and female participants aged 40-80. This work was conducted in accordance with the tenets of the Declaration of Helsinki (Edimburgh 2000), reviewed and approved by the Institutional Boards (code: 131/18; code P14_23_01_19). All clinical requirements to maintain the data privacy from the study participants were specifically met. All volunteers were informed and signed the consent to participate.

3.2 Eligibility requirements for the study participants

Ophthalmic specialists caring for glaucoma patients carried out a pre-selection by personal interview, according to the inclusion/exclusion criteria listed in table 1. Socio-demographics, personal and family characteristics, lifestyle, and treatments were recorded in a Microsoft Excel spreadsheet, as DEMO. A systematized ocular examination was done in the potential participants that got an appointment for the eye clinic. Best-corrected visual acuity (BCVA) was obtained from each eye calculating the logarithm of the minimum angle of resolution (LogMAR). The IOP was measured by Goldman applanation tonometry (Haag-Streit AT 900; Haag-Streit Köniz, Switzerland). Morphological determination (indirect gonioscopy) through a slit-lamp (IMAGEnet, Topcon, Barcelona, Spain) with the Goldmann 3-mirror lens was carried out to identify

an open anterior chamber angle; ocular fundus exploration with a 78D lens was performed through a slit-lamp; examination by optical coherence tomography (OCT) (Cirrus spectral-domain OCT, Carl Zeiss Meditec, Inc., Madrid, Spain) of the anterior and posterior eye segments, and functional probes by means of the visual field (VF) performance, using the 24-2 Swedish interactive threshold algorithm (Humphrey field analyzer, Carl Zeiss Meditec, Inc., Madrid, Spain), were also carried out. For the evaluation of participants, standard definitions of IOP, central corneal thickness (CCT), cup-to-disc (C/D) ratio, retinal nerve fiber layer (RNFL) thickness, RGCs density and VF median deviation (MD) were applied. In this context, normal IOP was considered as <21 mmHg, and any IOP above this threshold was defined as ocular hypertension (OHT). The CCT was determined by OCT and the normal values were estimated at 533 μm.

Table 1. Inclusion and exclusion criteria for the study participants

POAG group	Control group
INCLUSION	
Diagnosis of POAG	Healthy non-glaucomatous individuals
Aged >40 and < 80 years	Aged <40 and < 80 years
Initial glaucoma stage	-
Precise data at the clinical records	Precise data at the clinical records
Psychic and physical status that permits the participation in the study	Psychic and physical status that permits the participation in the study
EXCLUSION	
Other Glaucoma type	-
Aged <40 and > 80 years	Aged <40 and > 80 years
Other glaucoma stage	-
Other eye diseases or recent ophthalmic laser/surgery.	Other eye diseases or recent ophthalmic laser/surgery
Other systemic diseases/treatments/surgery	Other systemic diseases/treatments/surgery
Missing data or incomplete clinical history	Missing data or incomplete clinical history
No able to participate	No able to participate

Participants were classified as POAG group if they met one of these criteria: 1) patients previously diagnosed and confirmed in the clinical history as initial glaucomatous OND, under hypotensive eye drop therapy (Latanoprost, Timolol, and/or Brinzolamide); 2)

naïve POAG cases with corrected IOP/CCT higher than 21 mmHg, with an initial glaucomatous OND including specific optic nerve head alterations such as neuroretinal rim thinning, peripapillary nerve fiber loss, asymmetry of cupping between the patient eyes, and parapapillary atrophy, etc. Glaucoma damage was staged into the adequate category for better managing the disease. In this concern, automated VF is the hallmark for testing the visual function in glaucoma patients. We have used static automated perimetry (SAP) for the POAG diagnosis by the Humphrey Swedish Interactive Thresholding Algorithm (SITA) 24-2 Fast, with fixation monitoring and gaze-tracking (Humphrey visual field analyzer; Carl Zeiss Meditec, Madrid, Spain). In this study population, glaucomatous defects have been detected using the above techniques, with the reliability indices of the European Glaucoma Society (EGS), mean deviation (MD) and pattern standard deviation (PSD). The MD corresponds to the mean elevation or depression in the VF, as compared to the reference normal VF. According to the MD values, it was classified as mild, moderate, or severe VF damage (>6.00dB, -6.01 a 12.00 dB, and <12.01 dB, respectively).[28] The PSD corresponds to the irregularity measurements, in each of 6 regions of the VF, by adding the absolute value of the difference between the threshold value for each point, and the average VF sensitivity at each point. Therefore, according to the Hodapp *et al.*[28] approach, minimum criteria for considering the initial glaucoma stage is: 1) a glaucoma hemifield test outside normal limits (in at least two VF); 2) a cluster of three or more non-edge points in a location typical for glaucoma, all of which are depressed on the pattern deviation plot at a p < 5% level, and one of which is depressed at a p < 1% level (on two consecutive VF); 3) a corrected pattern standard deviation that occurs in less than 5% of normal VF (on two consecutive VF performances). Within the VF indexes, it has to be contemplated that the MD is the average elevation/depression of visual sensitivity in the overall VF, compared with that of the normal age-corrected reference VF. Therefore, the classification of VF defects for early glaucoma includes: 1) VF mean deviation less than -6 dB; 2) Less than 25% of the points are depressed below the 5%

level and less than 10 points are depressed below the 1% level on the pattern deviation plot; and 3) All point in the central 5º must have a sensitivity of at least 15 dB. Participants were classified as control group when the IOP was lower than 21 mm Hg with normal visual fields, optic disc, and RNFL in absence of other ocular or systemic disease (as in the inclusion/exclusion criteria).

All data were recorded into a Microsoft Excel spreadsheet, as OPHTHAL, which was reviewed by the glaucoma specialist. At baseline, a total of 50 individuals were selected by a nonrandom consecutive sampling procedure, to better confirm the health and ocular condition of the suitable participants and were distributed into two groups: patients with POAG diagnosis (n = 23) and individuals without glaucoma, as a control group (n = 27). The final sample size of our pilot study participants was 30 (11 POAG patients and 19 control individuals). Changes in the potential number of participants were due to the volunteer decision, clinical issues, and/or sampling contingences. Recruitment characteristics and operative procedures are depicted in Figure S1.

3.3. Sample collection.

Reflex tears were collected through capillarity by using a microhematocrit tube from each eye of the study participants, by a gentle rubbing of the inferior meniscus and external canthus of each eye, without instilling anaesthetics as described elsewhere.[20–24] A tear volume ranging from 6 to 25 μL was collected from each participant. Each sample was transferred into micro Eppendorf tubes, appropriately labelled and stored at −80 °C until processing. A total of 30 samples were collected from POAG patients (n=11), and the control group (n=19), as previously described.[20-24] The figure 1 shows the sampling technique for collecting reflex tears from the study participants.

Figure 1. Reflex tear collection from the inferior lacrimal meniscus by capillarity.

3.4 Sample preparation for ^{1}H-NMR spectroscopy study

Before sample preparation, tears were thawed. To prepare each sample, 20 μL of tear fluid were introduced in NMR tubes with a diameter of 1,7 mm. The remaining volume was completed up to 60 μl with phosphate buffer (which additionally contained deuterated water, and the internal standard sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) 1 mM), following previously published procedures.[29] In three cases the volume of tear available was less than 20 μl (6, 15 and 16 μL respectively). In those cases, extra quantity of phosphate buffer was added to complete the volume of 60 μL. Following the same procedure, three of the pharmacological active principles of the topical glaucoma therapy were considered for better classifying our participants, as those administered in monotherapeutic regimen or in association [Lumigan® (Bimatoprost), Azopt® (Brinzolamide) and Azarga® (Timolol/Brinzolamide)] and were prepared for NMR spectroscopy acquisition, to rule out that any of the signals included in the analyses were directly produced by the presence of the drugs.

3.5 NMR spectra acquisition and processing

The samples underwent NMR spectroscopy study (Príncipe Felipe Research Center Foundation (CIPF) NMR facility). NMR spectra were acquired using a 600 MHz spectrometer (Bruker AVII-600, Bruker Biospin, Germany) equipped with a 5 mm TCI cryoprobe (^{1}H, ^{15}N ^{13}C), a temperature unit BCU05 and a refrigerated SampleJet. The temperature of the probe was set at 300 ºK (27 ºC). ^{1}H-NMR monodimensional spectra were acquired for each sample with noesy pulse sequence and presaturation of the water signal during the relaxation time and mixing time. 200 scans were programmed with a spectral width of 30 ppm. Following the same procedures ^{1}H-NMR noesy spectra were as well acquired from the topical drugs used for ocular hypertension treatment. The chemical shift of the signals from the drugs spectra were taken into consideration in later analyses. Once acquired, the spectra were transformed and pre-processed. For the pre-processing of the spectra an exponential line-broadening function of 0.5 Hz was applied followed by Fourier transformation with TopSpin 3.6.2. Phasing, baseline correction and chemical shift calibration to DSS resonance at 0.0 ppm was done with the program MestReNova version 6.0.2 (Mestrelab Research SL, Santiago de Compostela, Spain). Areas from the different peaks in the 1D ^{1}H spectra were integrated. The integrated areas were used to determine the differences in metabolite concentration between tears from CG and POAGG. The peaks were assigned according to their chemical shift and the information in different databases, such as the human metabolome database (HMDB) [30] and the biological magnetic resonance data bank (BMRB).[31]

3.6 Statistical Analysis

3.6.1. General Statistical Proceedings

Statistics for the clinical data were performed by the *IBM SPSS 28.0* program (IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp). The Shapiro-Wilk test (subgroups) and the Kolmogorov-Smirnov test (main groups) were used to verify the normal distribution of the quantitative variables, whereas the qualitative variables were described by absolute and relative frequencies. Quantitative variables were described using the mean and standard deviation (normal distribution) or median and interquartile range (non-normal distribution). Differences between quantitative variables were analysed using the student's t test for independent samples and ANOVA (normal variable) or the Mann-Whitney and Kruskal Wallis U test (non-normal variable). Differences between groups were considered statistically significant when the p-value was less than or equal to 0.05.

To determine the metabolomic differences between tears from the POAG and the control groups, statistical analysis was performed. The integration data from each peak was normalized to the sum of all signals and auto scaled. These data were fed into the software *PLS_Toolbox Solo 8.9* (Eigenvector Research, Inc., Manson, WA, USA) to perform multivariate statistical analysis. A Principal Component Analysis (PCA) was employed to determine the presence of outsiders, to remove them from further analysis.

Partial Least Squares-Discriminant Analysis (PLS-DA) was used to generate a predictive model, able to classify the samples based on the relative concentration of metabolites in each group. The data were split into calibration (2/3 of the samples were used for the generation of the model) and validation (1/3 of the samples was used to prove the discriminative capacity of the generated model in an independent data set) subsets.

The variables included in the model were iteratively selected based on its Variable Importance in Projection (VIP) value until reaching the optimization of the model. Cross Validation (venetian blinds) was used to select the optimal number of latent variables for the model. In order to determine the goodness of the model to discriminate between different sets of samples, the area under the ROC curve value (AUC), the sensitivity and the specificity were calculated. After validation, the robustness and over-fitting of the model were tested through permutation test (100 iterations, Rand-t-test, Wilcoxon and Sign test). p-value < 0.05 was considered significant.

Univariate analysis of the metabolites participating in the model was performed by using *Metaboanalyst.*[32] The mean of normalized intensity for each metabolite was calculated for the two groups, POAG and control groups. The t-test was used to determine significant differences between metabolites in POAG and control tear samples after testing the normality of the variables. False Discovery Rates (FDR) adjusted p-values were as well obtained and considered for statistical significance.

3.6.2 Analysis of altered metabolic pathways

Metaboanalyst [32] was used to explore the potential metabolic pathways involved in the pathological processes. HMDB ID of each metabolite was used to include them in the pathway analysis. The global test enrichment analysis selected for the topological analysis was relative-betweenness centrality. The pathways with p-value < 0.05 and impact factor > 0 were selected as representative pathways.

4. Results

4.1 Patient characteristics

The final number of participants in this pilot study was 30 (11 POAG patients and 19 control individuals), as depicted in figure S1. The breakdown of participants was 20 (40%), that failed to complete the study for a variety of reasons, including loss of interest and volunteer drop out, experimenter error, clinical findings, and alterations in sampling, transportation, or laboratory processing of the biological samples.

Mean age was 69 ± 8 years (71 ± 9 years in the POAG group and 68 ± 7 years in the control group). Gender distribution was 64% women / 36% men in the POAG group and 55% women / 45% men in the control group. The mean age and gender distribution were not significantly different between POAG and control groups. All participants were Caucasian.

4.2 Systemic and ophthalmologic clinical characteristics

Comorbidities were recognized non-IOP risk factors for the POAG course, and the following were taken into consideration in the study participants: hypertension blood pressure, cardiovascular disease, diabetes mellitus, myopia, and/or obesity (increased body mass index). The participants with the above disorders were excluded from the study, according to the criteria established in table 1 (Figure S1).

The POAG patients had IOP elevation, verified through the augmented optic disc excavation, optic nerve damage, and altered VF. The participants in POAG group were under glaucoma treatment (hypotensive eye drops as monotherapy: 52% Bimatoprost (Lumigan ®), and 8% Brinzolamide (Azopt ®), as well as fixed combination: 40%

Timolol/Brinzolamide (Azarga ®)). The CG was constituted by those healthy individuals without any of the above glaucoma milestones.

Ophthalmological examination showed that mean BCVA LogMAR was 0.00 for the RE (right eye) and 0.10 for the LE (left eye) in the CG, versus 0.20 (RE) and 0.20 (LE) in the POAG group. Moreover, the mean IOP, and the mean CCT were significantly lower in the control group respect to the POAG group ($p < 0.001$; $p < 0.001$). The most important OCT and VF parameters were the mean cup-to-disc ratio, RNFL thickness, and RGCs density, as well as the VF MD and PSD. Overall, the SAP data analyses showed a significant decrease in MD values ($p < 0.001$), and a significant increase in the PSD values ($p < 0.001$) in the POAG patients respect to the controls. In fact, the POAG patients displayed mild VF damage. No moderate or severe VF damage was detected in our glaucomatous population. All the above parameters were significantly different between both study groups. The ophthalmological parameters of the study for both groups and the p-values are shown in table 2.

Table 2. Ophthalmic characteristics of the study participants.

Parameters	POAG group	Control group	p value
BCVA RE	0.2	0.0	<0.05
BCVA LE	0.2	0.1	<0.05
IOP RE (mm Hg)	20 ± 2	14 ± 1	<0.001
IOP LE (mm Hg)	19 ± 2	15 ± 1	<0.001
CCT RE (m)	527 ± 13	575 ± 12	<0.01
CCT LE (m)	532 ± 12	568 ± 14	<0.01
Average C/D ratio RE	0.6 ± 0.2	0.1 ± 0.01	<0.001
Average C/D ratio LE	0.5 ± 0.3	0.1 ± 0.01	<0.001
Average RNFL thickness RE (m)	70 ± 10	94 ± 11	<0.05
Average RNFL thickness LE (m)	72 ± 8	89 ± 10	<0.05
RCCs density RE	65 ± 8	94 ± 12	<0.001
RGCs density LE	68 ± 9	90 ± 12	<0.001
VF MD RE	-3.2 ± 1.6	$-1dB \pm 1$	<0.001
VF MD LE	-2.5 ± 1.2	$-1dB \pm 1$	<0.001

BCVA: best corrected visual acuity; logMAR: logarithm of the minimum angle of resolution; VF MD: visual field mean deviation; RE: right eye, LE: left eye

4.3 ^{1}H-NMR spectroscopy study

Finally, 30 tear samples corresponding to both eyes of the study participants (POAGG n=11; CG n=19) were analysed by ^{1}H-NMR metabolomics.

4.3.1. ^{1}H NMR profile of tears obtained from the study participants.

^{1}H-NMR noesy spectra were acquired for all tear samples and drugs (Lumigan ®, Azopt ® and Azarga ®). The spectra showed an acceptable signal to noise ratio, despite the low concentration of the metabolites in the samples, enough to enable the assignment and the relative quantification of the signals. The main signals in tears were assigned and are shown in the figure 2. The spectrum is shown divided into two parts, aliphatic (Figure 2a) and aromatic (Figure 2b), to enable a better observation of the signals. A total of 40 metabolites were assigned according to their chemical shifts, multiplicity and J coupling (Table 3). The spectra of the drugs were also obtained and are shown in the figure S2, to assess that the signals in the discriminant models between POAG and control groups are not due to the treatments.
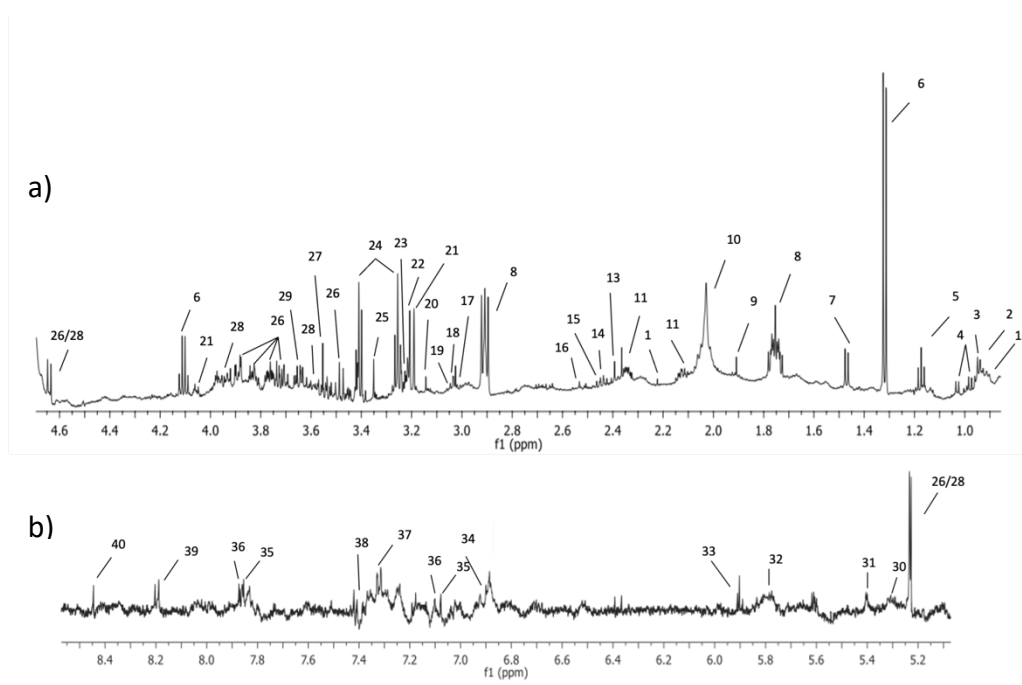
Figure 2. ¹H-NMR spectrum of one tear sample. The residual water signal (4.7-5.1 ppm) is not shown a) Aliphatic region of the spectrum (0.8-4.6 ppm) b) Aromatic region of the spectrum (5.2-8.5 ppm). The intensity of peaks in the aromatic region (5.0-8.7 ppm) has been scaled (5x) respect to the aliphatic region for a more appropiated display. 1. Fatty Acids (–CH$_3$), 2. Isoleucine, 3. Leucine, 4. Valine, 5. Ethanol, 6. Lactate, 7. Alanine, 8. DSS, 9. Acetate, 10. N-acetyled compounds, 11. Glutamate, 12. Acetone, 13. Pyruvate, 14. Pyroglutamic Acid, 15. Glutamine, 16. Citrate, 17.Creatine, 18. Creatinine, 19. Lysine, 20. Dimethyl Sulfone, 21. Choline, 22. Carnitine, 23. Arginine, 24. Taurine, 25. Methanol, 26. Glucose, 27. Glycine, 28. Glucose 6-phosphate, 29. Glycerol, 30. Unsaturated Fatty Acids, UFA (-CH=CH-), 31. Sucrose, 32. Urea, 33. Uridine, 34. Tyrosine, 35. Histidine, 36. Histamine, 37. Phenylacetate, 38. Phenylalanine, 39. Hypoxanthine, 40. Formic acid.

Table 3. Chemical shift, multiplicity and J coupling of the signals from metabolites identified in the tear samples.

	Metabolite	Chemical Shift (ppm) and J coupling (Hz) *
1	Fatty Acids (–CH_3),	0.8-0.9
2	Isoleucine	0. 93 (t, J = 7.0), 1 (d, J = 6.5), 1.31 (m)
3	Leucine	0.97 (d), 0.98 (d), 1.72 (m)
4	Valine	1.01 (d, J = 7.2), 1.06 (d, J = 7.2)
5	Ethanol	1.16 (t, J = 7.08), 3.67 (q, J = 7.07)
6	Lactate	1.35 (d, J = 7.0), 4.14 (c, J = 7.0)
7	Alanine	1.47(d, J = 7.2), 3,77 (q, J = 7.2)
8	DSS	0.00 (s), 1.75 (m), 2.92 (t)
9	Acetate	1.91 (s)
10	N-acetyled compounds	2.0-2.1
11	Glutamate	2.11 (td, J = 6.8, 6.2), 2.15 (dt, J = 15.4, 6.8), 3.75 (t, J = 6.2)
12	Acetone	2.22 (s)
13	Pyruvate	2.36 (s)
14	Pyroglutamic acid	2.39 (m), 2.50 (m), 4.17 (dd, J = 9.02, 5.83)
15	Glutamine	2.42 (dt, J = 14.4, 6.8), 3.76 (t, J = 6.2)
16	Citrate	2.52 (d, J = 15.4), 2.66 (d, J = 15.4)
17	Creatine	3.02 (s), 3.92 (s)
18	Creatinine	3.03 (s), 4.05 (s)
19	Lysine	1.46 (m), 1.71 (m), 1.89 (m), 3.02 (t), 3.74 (t, J = 6.09)
20	Dimethyl sulfone	3.14 (s)
21	Choline	3.19 (s), 3.51 (dd, J = 5.81, 4.16), 4.05 (ddd)
22	Carnitine	2.13 (s), 2.48 (dd, J = ND), 2.61 (dd, J = ND), 3.18 (s), 3.61 (d, J = ND), 3.82 (dd, J = ND), 5.57 (q)
23	Arginine	1.68 (m), 1.90 (m), 3.23 (t, J = 6.93), 3.76 (t, J = 6.11)
24	Taurine	3.25 (t, J = 6.57), 3.42 (t, J = 6.62)
25	Methanol	3.34 (s)
26	Glucose	3.74 (d, J = 5.4), 3.81 (dt, J = 8.4,5.4), 5.22 (d, J = 1.6)
27	Glycine	3.55 (s)
28	Glucose 6-phosphate	3.27 (dd, J = 9.21, 7.99), 3.71 (t, J = 9.54), 4.64 (d, J = 7.99), 5.22 (d, J = 3.75)
29	Glycerol	3.58 (dd), 3.67 (dd, J = 11.7, 4.3) y 3.90 (m)
30	UFA	5.25-5.35
31	Sucrose	3.46 (t, J = 9.30), 3.55 (dd, J = 3.98, 3.89), t (3.75),
32	Urea	5.78 (s)
33	Uridine	3.801 (dd, J = 12.77, 4.4), 4.21 (m), 4.22 (dd),
34	Tyrosine	6.91 (m), 7.21 (m)
35	Histidine	3.16 (dd, J = 15.55, 7.7), 3.23 (dd, J = 16.10, 4.9), 3.98 (dd, J =7.73, 4.98), 7.09 (d, 0.58), 7.90 (d, J = 1.13)
36	Histamine	3.03 (m), 3.29 (t, J = 7.11), 7.14 (s), 7.99 (s)
37	Phenylacetate	3.53 (s), 7.29 (m), 7.36 (m)
38	Phenylalanine	7.35 (m), 7.39 (m), 7.44 (m)
39	Hypoxanthine	8.17 (s), 8.20 (s)
40	Formic acid	8.44 (s)

* multiplicity is indicated as s (singlet), d (doublet), t (triplet), q (quadruplet), m (multiplet)

4.3.2 Multivariate data analysis of the tear metabolomic profiles

In order to identify the presence of outliers or significant clustering of POAG and control tear samples, an unsupervised PCA analysis was performed with the relative concentrations of the different metabolites. Some clustering related to the presence or absence of the disease could already be appreciated in the undirected classification analysis, based on the scores of the samples in the principal components one and two (Figure 3a). Two samples, one POAG and one control, were located outside the 95% confidence level at the PCA diagram and were removed from further studies.

The remaining samples (from 18 controls and 10 POAG patients) were randomly divided into two groups (calibration and validation), to perform a PLS-DA analysis. The model was generated with the calibration subset, through the iterative selection of the most representative variants (VIP > 1). When applied to the validation subset, the ROC curve of the model showed excellent classification capacity with an AUC value of 1 (Figure 3b). This model predicted with high sensibility (100%) and specificity (83.3%) if the tear samples were from control or POAG (Figure 3c). Wilcoxon permutation test provided a $p < 0.05$, which confirmed the robustness of the model to discriminate between both groups. From the 40 identified metabolites, the multivariate model included 11 of them to discriminate POAG group from control group: phenylalanine, phenylacetate, leucine, taurine, glycine, urea, glucose, n-acetyled compounds, UFA, formic acid and uridine, as well as an unassigned species at 8.35 ppm (unk 8.35 ppm). For this final model, the most relevant metabolites with a VIP > 1 were phenylalanine, phenylacetate, leucine, taurine, glycine, urea and glucose (Figure 3d).
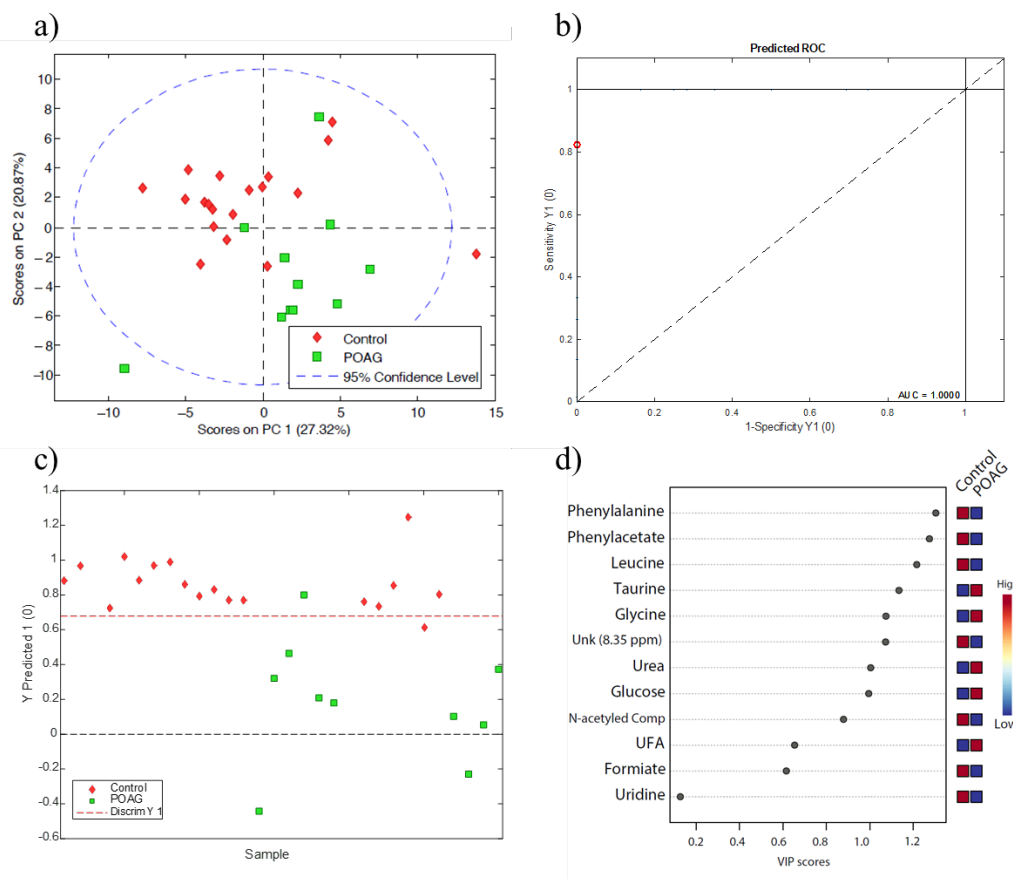
Figure 3. Multivariate statistical analysis. Data from control group are presented in red diamonds and data from POAG group are presented in green squares in a) and c). a) Principal component analysis (PCA) for the whole set of samples to determine outliers b) ROC curve of the model c) Partial least squares-discriminant analysis. (PLS-DA) score plot. The classification of calibration (left) and validation (right) samples is shown d) Variables participating in the prediction model. The VIP values of the metabolites and its relative changes in control and POAG group is displayed. (red = increaed in POAG group vs. control group, blue = decreased in POAG group vs. control group).

4.3.3 Mean comparison of metabolites included in the discriminative model

Afterwards, the relative concentration of the metabolites with discriminative capacity, as shown in the multivariate classification model was studied (Figure 4, Table 4). All metabolites, with the exception of formic acid and uridine, showed significant difference with an FDR < 0.05. The analysis showed a decrease in the relative concentration of phenylalanine, phenylacetate, leucine, n-acetyled compounds, formic acid, uridine and unk (8.35 ppm) in POAG tear samples compared to the tears from control group. On the other hand, an increase in the concentration of taurine, glycine, urea, glucose and UFA in tears of POAG group was observed.

Table 4. Relative mean concentrations of discriminative metabolites between control and POAG groups, and statistical significance of normalized data.

	[Metab]*		p-value	FDR**
	control group	POAG group		
Phenylacetate	50.6	19.6	9.12E-06	7.26E-05
Phenylalanine	51.9	20.9	1.21E-05	7.26E-05
Unknown (8.35 ppm)	0.8	0.2	2.05E-05	7.43E-05
Taurine	8.5	10.9	2.48E-05	7.43E-05
Leucine	48	25.4	3.68E-05	8.83E-05
Glycine	13	13.5	5.82E-05	1.16E-04
Glucose	0.3	0.9	5.52E-04	9.46E-04
Urea	26.6	41.4	1.69E-03	2.29E-03
N-acetyled compounds	114.2	67.6	1.72E-03	2.29E-03
UFA (-CH=CH-)	0.3	0.5	0.024	0.028
Formic acid	0.3	0.1	0.049	0.054
Uridine	14.8	7.9	0.22	0.22

* Relative mean concentration in control group and POAG group are shown x10^3
**FDR: False discovery Rate
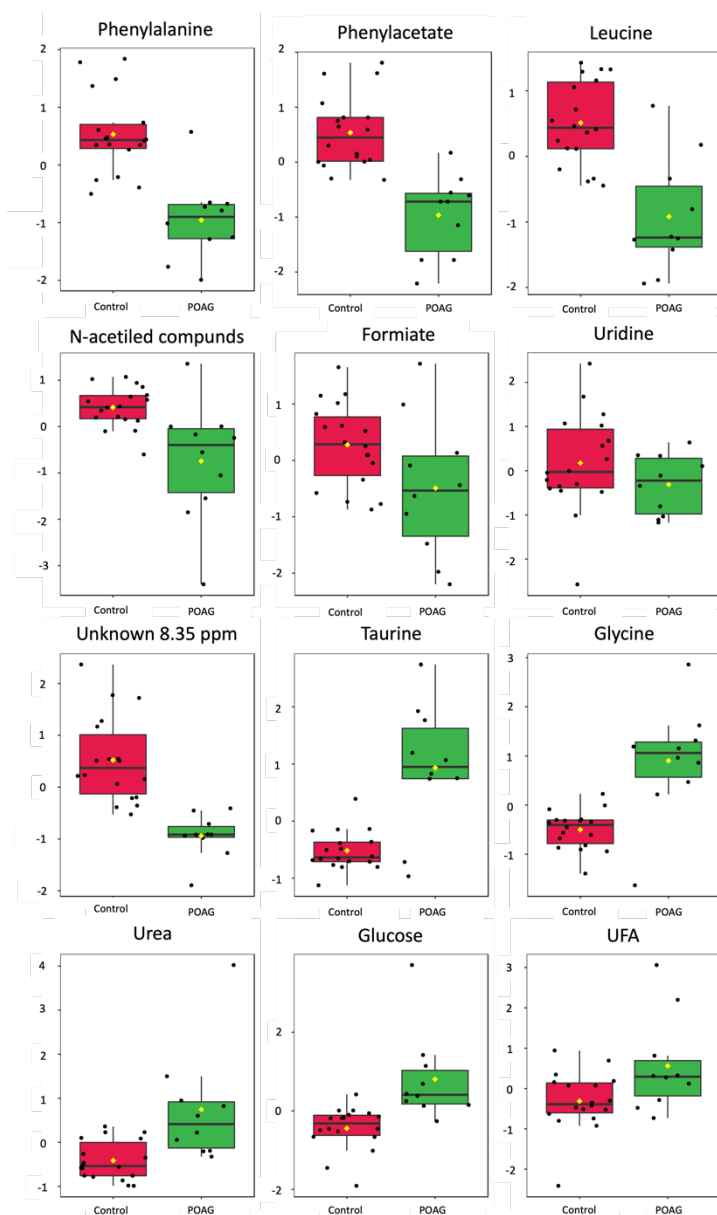
Figure 4. Comparison of the mean concentration of the metabolites participating in the discriminative model. Box plots representing the mean concentration of normalized metabolites in control and POAG tears samples are shown. The yellow diamonds represent the mean and the horizontal line the median of each group for a determined feature. The normalized concentrations of each sample are shown as black dots.
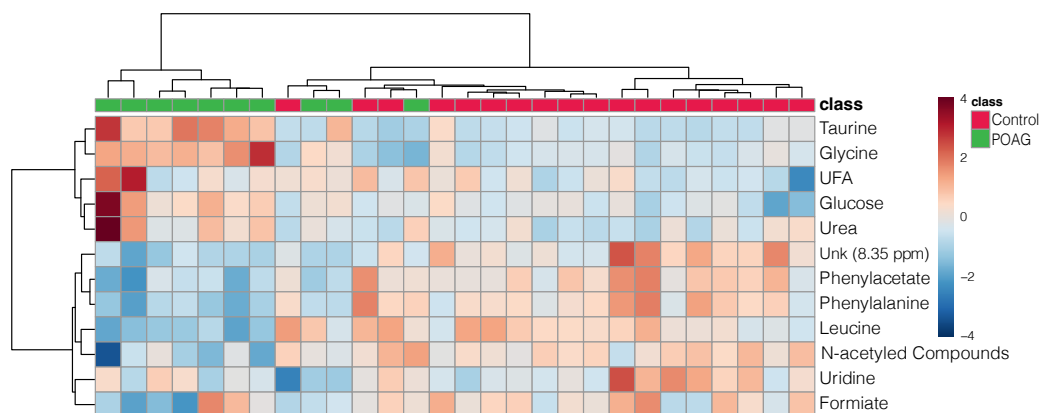
Figure 5. Heat map with the concentration of the metabolites participating in the discriminative model. Metabolites are clustered according to its relative increase (red) /decrease (blue).

4.3.4 Metabolic Pathways Analysis

Quantitative pathway topological analysis of the metabolites included in the discriminative model, revealed significant alterations in phenylalanine metabolism, taurine and hypo-taurine metabolism, glyoxylate and dicarboxylate metabolism, glycine, serine and threonine metabolism, glutathione metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, primary bile acid biosynthesis and glycolysis/gluconeogenesis (Figure 6, Table 5).
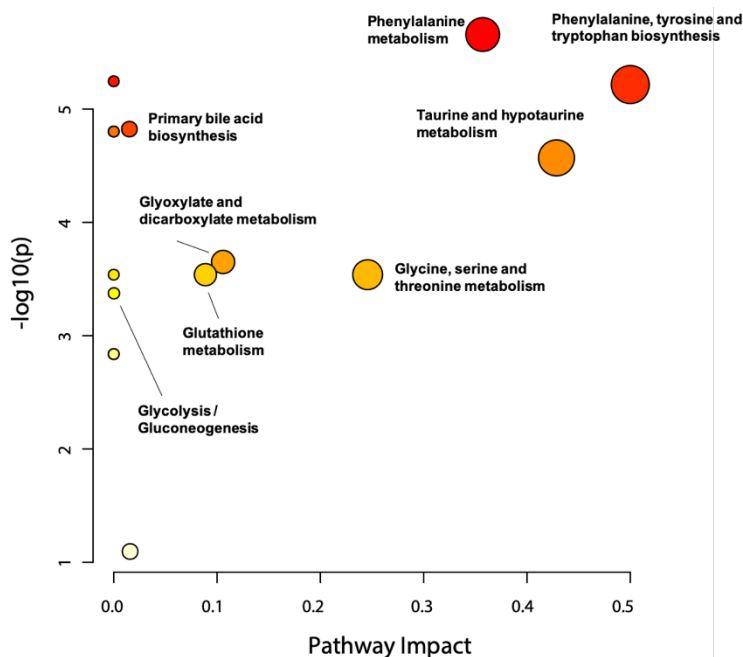
Figure 6. Significant metabolic pathways altered in POAG group compared to control group. Each circle represents an identified metabolic pathway. The size of the circle is proportional to the pathway impact value (PIV), and the color is proportional to the statistical significance [-log$_{10}$(p)] from highest (red) to lowest (white).

Table 5. Significant metabolic pathways and pathway impact values obtained from integrating enrichment analysis and pathway topology analysis.

Metabolic Pathway	Total Cmp*	Hits	p - value	FDR**	Impact
Phenylalanine metabolism	10	2	2.20E-06	3.04E-05	0.36
Phenylalanine, tyrosine and tryptophan biosynthesis	4	1	6.07E-06	3.04E-05	0.50
Primary bile acid biosynthesis	46	2	1.50E-05	3.94E-05	0.02
Taurine and hypotaurine metabolism	8	1	2.70E-05	5.78E-05	0.43
Glyoxylate and dicarboxylate metabolism	32	2	2.24E-04	3.95E-04	0.11
Glycine, serine and threonine metabolism	33	1	2.89E-04	3.95E-04	0.25
Glutathione metabolism	28	1	2.89E-04	3.95E-04	0.09
Glycolysis / Gluconeogenesis	26	1	4.23E-04	5.28E-04	0.00021

*Cmp: Compounds

** FDR: False Discovery Rate

5. Discussion

In this pilot study, we investigated the metabolites present in tear samples from POAG patients and controls by ^{1}H-NMR spectroscopy to generate, through multivariate statistics, a predictive model able to identify patients at initial glaucoma stage, at risk of OND and vision loss. We found that the generated model showed excellent discriminant ability with an AUC of 1, and 100% of sensibility and 83.3% of specificity in the classification of POAG group and control group. From the statistical model generated, a list of potential biomarkers of the disease, and associated metabolic pathways were obtained. Moreover, changes in the concentration of biomarkers involved in the discriminant model were confirmed by the significant differences shown in the univariate analysis. A descent in the relative concentration of phenylalanine, phenylacetate, leucine, n-acetyled compounds, formic acid, uridine and unk (8.35 ppm), and an increment in the concentration of taurine, glycine, urea, glucose and UFA in tears was found in the POAG group as compared to the control group.

The discovery of molecular biomarkers in tears could reveal essential information regarding POAG pathophysiology, as well as help to manage medical-laser-surgical hypotensive therapy.[2–4,7,8,21,23,28] In this context, early diagnosis and treatment are pivotal to avoid glaucoma progression, optic atrophy, and blindness.[28,33] It is essential to consider that: i) a significant number of patients present at first medical appointment with elevated IOP and moderate-to-severe VF loss, due to the silent period of undetected OHT, and ii) some glaucoma patients suffer higher rates of progression than others. Therefore, some findings have been reached in the present work that could be transferred to ophthalmic practice, which are discussed below.

First, tear samples were used for the metabolomic study in glaucoma, in agreement with Agnifili *et al*.[34] and Wu *et al*.[35] and in contrast with other authors usually using aqueous humor or plasma samples for this purpose.[25,26,36–38] As widely known, the tear

film is located on the outermost part of the ocular surface, in direct contact with the environment. Under normal conditions, tear volumes range from 4 to 12 µL per eye.[39] The tear film is a complex and interesting biological fluid that contains water and a wide variety of electrolytes, lipids, proteins, glycoproteins and small molecules from different sources.[39–41] After a period of controversy, it has been established that metabolites associated with glaucoma present in the tear film mainly come from the aqueous humor, through the uveoscleral outflow pathway, after precise scleral percolation.[39–41] Trying to counteract the small tear volumes that can be collected for analytical issues, a technique for obtaining reflex tears by gentle rubbing of the inferior lacrimal meniscus and palpebral lateral canthus, has been used here, to relatively easy obtaining 20-30   L of tears from both eyes by capillarity, as described elsewhere.[20–24]

Next, we analysed the metabolites in tears from POAG patients at initial stage of disease, according to Hodapp *et al*.[28] but at risk of glaucoma OND and blindness. Epidemiological and experimental studies have established that the early detection of OHT, and the prompt IOP reduction significantly diminish the risk of glaucoma progression.[2–5,33] This stage is very important, because the elevated IOP leads to progressive damage and death of the RGCs and optic fibre loss, glaucoma hallmarks that manifests themselves in the structural/functional ophthalmological examination of these patients. However, there are no specific and complete standard references for accurately establishing the early glaucoma diagnosis.[2–5,28,33] Bearing this in mind, the POAG group was accurately selected for the main purpose of this study, that was to characterize the metabolomic fingerprint in tears of POAG patients at initial stage of disease in order to identify potential biomarkers for better eye and vision care.

Furthermore, ^{1}H-NMR spectroscopy was used to generate, through multivariate statistics, a predictive model able to identify POAG patients at initial glaucoma stage. Biomedical and biotechnological advances in metabolomics have provided information on a considerable number of metabolites to better understand the metabolic changes that occur in glaucoma. Previous reports on the identification of metabolomic

biomarkers of glaucoma by MS have been performed using blood and aqueous humor samples.[26,4]

[1–46] However, other researchers used ^{1}H-NMR spectroscopy, as in this study, to perform glaucoma metabolomics, either using blood or aqueous humor samples, with optimum results.[33,47]

Data are quite different among the studies. In spite of describing the biological sample and the analytical platform, the statistical processing as well as the changes of the statistically significant metabolites identified in a differential profile, important variations arise that make it difficult to identify potential biomarkers of glaucoma. Leruez *et al.*[37] described amino acids, carbohydrates and polyamine families, among others, altered in POAG samples and associated to mitochondrial dysfunction, senescence and polyamines deficiency. However, they did not explore the diagnosis potential of the metabolites identified. Myer *et al.*[36] pointed out disturbances in the concentration of several amino acids, such as arginine, cysteine, threonine and lysine, and carbohydrates in the aqueous humor of POAG patients compared to the CG, and generated cross-validated PLS-DA models with Q2 values of 0.15. Better significance for the classification was reached by Buisset *et al.*[37] who obtained a model able to classify samples from the POAG patients with an AUC of 0.89.- Moreover they pointed out to taurine and spermine deficiency in aqueous humor from POAG patients.

From the statistical model generated in the present work, a list of potential biomarkers of the disease, and associated metabolic pathways were obtained. An increase in the concentration of taurine, glycine, urea, glucose and UFA and a decrease in the concentration of phenylalanine, phenylacetate, leucine, n-acetyled compounds, formic acid, uridine and unk (8.35 ppm) was observed in tears from the POAG group as compared to the control group. Moreover, according to these results, the phenylalanine metabolism and phenylaniline, tyrosine and tryptophan biosynthesis pathways were altered in POAG tears with a significant decrease in the concentration of phenylalanine and phenylacetate. Phenylalanine had already been pointed as a

potential biomarker of glaucoma in previous metabolomic studies, both in aqueous humor, and in tears,[41,43] in which the authors observed a decrease in the concentration of phenylalanine in POAG patients in a similar manner to us. Previous studies have as well reported mutations in the synthesis pathway of phenylalanine related to POAG development.[48] The significant change in the glycine, leucine and serine pathway, with the decrease in the relative concentration of leucine [49] and the relative increase in the concentration of the amino acid glycine [41,43] were also previously reported in aqueous humor. Furthermore, the increase in glucose levels observed in tears from the POAG patients presented herein is consistent with the pathological processes associated to glaucoma, as it has been reported that patients with POAG present glucose hypometabolism. [50] Also, the increase in serum glucose levels has been associated with elevated IOP,[51] which is strongly associated with glaucoma. An increase in the relative concentration of taurine has been detected in the present work, while previous reports noticed a decreased concentration in aqueous humor [37] and tears.[41] Taurine has a neuroprotective effect against inflammation, oxidative stress, and osmotic stress. Nevertheless, taurine could as well be produced as a counteracting mechanism against oxidative stress. In fact, in agreement with our current results, taurine was observed to increase in other previous studies conducted with aqueous humor of POAG patients,[47] and in a canine glaucoma model.[52]

Tear metabolomic signature of POAG patients by ^{1}H-NMR spectroscopy, has been described in this work. The small quantity of sample available (in the order of 20 μL) and the inherent low concentration of metabolites in tear samples have been overcomed by using a high field NMR spectrometer equipped with a cryoprobe. This experimental setting has provided a non-invasive way to provide samples and data related to ocular disease. A statistical model for the diagnosis of POAG has been developed using samples obtained in a non-invasive way. Moreover, a group of potential biomarkers of the disease has been obtained from the statistic model. Future improving of this study includes the consideration of a higher number of samples, and

the use of samples from patients naïve to treatment. Despite the analysis of the topical drugs assessed that the classification of the samples and the potential biomarkers obtained in this study were not directly derived from the treatment, the changes observed in the POAG samples spectra might be produced by the metabolism of the therapy. Furthermore, unlike blood derivatives that circulate through the general bloodstream, tears that come into contact with the eye can serve as a valuable source of biomarkers for glaucoma and can be collected in a non-invasive manner in contrast to aqueous humor. Previous works have studied the metabolic profile of POAG in tears by different methods of MS.[35,41] However, unlike NMR,[12–14,32] MS platform requires more elaborated preparation procedures for the analysis of each sample.[11,12,15,32]

Overall, data presented herein provided a model with a very good performance for disease prediction, from samples that have been obtained both by a non-invasive collecting technique (tears) and through an analytical method that requires a simplest sample preparation procedure (NMR) in contrast with previous reports that use aqueous humor and MS. Moreover, these data were also used to generate and validate the model with a spare group of samples. In conclusion, a model able to classify with great values of specificity and sensitivity POAG and control groups in an independent set of samples has been here obtained. Phenylalanine, phenylacetate, leucine, formic acid, n-acetyl compounds, uridine, taurine, glycine, urea, glucose, UFA and an unknown metabolite (8.35 ppm) could be considered potential biomarkers of patients at the initial glaucoma stage, that can be obtained in a non-invasive, relatively affordable way to improve eye and vision care. In this sense, it seems to us that pathological changes occurring in the eyes can be reflected in the whole ocular constituents, including the ocular surface, either via local and systemic circulation, and/or via simple diffusion through the cornea and sclera. Nonetheless, by analysing tear film samples for target metabolites, we can design a promising window for optimizing POAG diagnosis and preventing blindness. This research could be a starting point for developing a non-invasive diagnostic system for POAG. In fact, we manage for the first time a new

approach on the identification of POAG-related metabolites in tears, to improve the personalized diagnosis of the disease, that allows to early identify patients at highest risk of POAG or POAG progression. Future directions necessarily include the increment of the number of samples to confirm the results here shown.

6. Bibliography

1.   Bourne, R. R. A. *et al.* Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study. *Lancet Glob. Heal.* **9**, e144–e160 (2021).

2.   Greco, A. *et al.* Emerging Concepts in Glaucoma and Review of the Literature. *Am. J. Med.* **129**, 1000.e7-1000.e13 (2016).

3.   Tham, Y. C. *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* **121**, 2081–2090 (2014).

4.   Kapetanakis, V. V. *et al.* Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis. *Br. J. Ophthalmol.* **100**, 86–93 (2016).

5.   Gauthier, A. C. & Liu, J. Neurodegeneration and neuroprotection in glaucoma. *Yale J. Biol. Med.* **89**, 73–79 (2016).

6.   Conlon, R., Saheb, H. & Ahmed, I. I. K. Glaucoma treatment trends: a review. *Can. J. Ophthalmol.* **52**, 114–124 (2017).

7.   Beykin, G. & Goldberg, J. L. Molecular Biomarkers for Glaucoma. *Curr. Ophthalmol. Rep.* **7**, 171–176 (2019).

8.   Bua, S. & Supuran, C. T. Diagnostic markers for glaucoma: a patent and literature

review (2013-2019). *Expert Opin. Ther. Pat.* **29**, 829–839 (2019).

9.   Johnson C, Ivanisevic J & Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).

10. Jones, O. A. H. Illuminating the dark metabolome to advance the molecular characterisation of biological systems. *Metabolomics* **14**, 101 (2018).

11. Kim, S. J., Song, H. E., Lee, H. Y. & Yoo, H. J. Mass Spectrometry-based Metabolomics in Translational Research. *Adv. Exp. Med. Biol.* **1310**, 509–531 (2021).

12. Emwas, A. H. M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol. Biol.* **1277**, 161–193 (2015).

13. Bizzarri, D. *et al.* 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. *eBioMedicine* **75**, 103764 (2022).

14. Nagana Gowda, G. A. & Raftery, D. NMR-Based Metabolomics. *Adv. Exp. Med. Biol.* **1280**, 19–37 (2021).

15. Alves, S., Paris, A. & Rathahao-Paris, E. Mass spectrometry-based metabolomics for an in-depth questioning of human health. *Adv. Clin. Chem.* **99**, 147–191 (2020).

16. Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A deep profiler's guide to cytometry. *Trends Immunol.* **33**, 323–332 (2012).

17. Pieragostino, D. *et al.* Differential protein expression in tears of patients with primary open angle and pseudoexfoliative glaucoma. *Mol. Biosyst.* **8**, 1017–1028 (2012).

18. Izzotti, A., Longobardi, M., Cartiglia, C. & Saccà, S. C. Proteome alterations in primary open angle glaucoma aqueous humor. *J. Proteome Res.* **9**, 4831–4838

(2010).

19.   Tezel, G. Multiplex protein analysis for the study of glaucoma. *Expert Rev. Proteomics* **18**, 911–924 (2021).

20.   Pinazo-Durán, M. D. *et al.* Effects of a nutraceutical formulation based on the combination of antioxidants and ω-3 essential fatty acids in the expression of inflammation and immune response mediators in tears from patients with dry eye disorders. *Clin. Interv. Aging* **8**, 139–148 (2013).

21.   Galbis-Estrada, C. *et al.* Patients undergoing long-term treatment with antihypertensive eye drops responded positively with respect to their ocular surface disorder to oral supplementation with antioxidants and essential fatty acids. *Clin. Interv. Aging* **8**, 711–719 (2013).

22.   Galbis-Estrada, C. *et al.* Differential effects of dry eye disorders on metabolomic profile by 1H nuclear magnetic resonance spectroscopy. *Biomed Res. Int.* **2014**, 542549 (2014).

23.   Benitez-del-Castillo, J., Cantu-Dibildox, J., Sanz-González, S. M., Zanón-Moreno, V. & Pinazo-Duran, M. D. Cytokine expression in tears of patients with glaucoma or dry eye disease: A prospective, observational cohort study. *Eur. J. Ophthalmol.* **29**, 437–443 (2018).

24.   Benítez Del Castillo, J. M. *et al.* Tear 1H Nuclear Magnetic Resonance-Based Metabolomics Application to the Molecular Diagnosis of Aqueous Tear Deficiency and Meibomian Gland Dysfunction. *Ophthalmic Res.* **64**, 297–309 (2021).

25.   Lereuz, S. *et al.* A Metabolomics Profiling of Glaucoma Points to Mitochondrial Dysfunction, Senescence, and Polyamines Deficiency. *Biochem. Mol. Biol.* **59**, 4355–4361 (2018).

26.   Zhang, Z. *et al.* Relationship between plasma amino acid and carnitine levels and

primary angle-closure glaucoma based on mass spectrometry metabolomics. *Exp. Eye Res.* **227**, 109366 (2023).

27. Rong, S. *et al.* Long-chain unsaturated fatty acids as possible important metabolites for primary angle-closure glaucoma based on targeted metabolomic analysis. *Biomed. Chromatogr.* **31**, e3963 (2017).

28. Hodapp, E., Parrish, R. & Anderson, D. R. Clinical Decisions In Glaucoma. (1993). Available at: https://api.semanticscholar.org/CorpusID:56639770.

29. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc. 2007 211* **2**, 2692–2703 (2007).

30. Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

31. Hoch, J. C. *et al.* Biological Magnetic Resonance Data Bank. *Nucleic Acids Res.* **51**, D368–D376 (2023).

32. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

33. Weinreb, R. N., Aung, T. & Medeiros, F. A. The pathophysiology and treatment of glaucoma: a review. *JAMA* **311**, 1901–1911 (2014).

34. Agnifili, L. *et al.* Molecular biomarkers in primary open-angle glaucoma: from noninvasive to invasive. *Prog. Brain Res.* **221**, 1–32 (2015).

35. Wu, J. *et al.* Glaucoma Characterization by Machine Learning of Tear Metabolic Fingerprinting. *Small Methods* **6**, 2200264 (2022).

36. Myer, C. *et al.* Differentiation of soluble aqueous humor metabolites in primary open angle glaucoma and controls. *Exp. Eye Res.* **194**, 108024 (2020).

37. Buisset, A. *et al.* Metabolomic Profiling of Aqueous Humor in Glaucoma Points

to Taurine and Spermine Deficiency: Findings from the Eye-D Study. *J. Proteome Res.* **18**, 1307–1315 (2019).

38. Chen, X., Chen, Y., Wang, L. & Sun, X. Metabolomics of the aqueous humor in patients with primary congenital glaucoma. *Mol. Vis.* **25**, 489–501 (2019).

39. Tiffany, J. M. The Normal Tear Film. in *Surgery for the Dry Eye: Scientific Evidence and Guidelines for the Clinical Management of Dry Eye Associated Ocular Surface Disease* (eds. Geerling, G. & Brewitt, H.) **41**, 0 (S.Karger AG, 2008).

40. Mastropasqua, R. *et al.* Trans-conjunctival aqueous humor outflow in glaucomatous patients treated with prostaglandin analogues: an in vivo confocal microscopy study. *Graefe's Arch. Clin. Exp. Ophthalmol.* **252**, 1469–1476 (2014).

41. Rossi, C. *et al.* Multi-omics approach for studying tears in treatment-naïve glaucoma patients. *Int. J. Mol. Sci.* **20**, 4029 (2019).

42. Barbosa-Breda, J., Himmelreich, U., Ghesquière, B., Rocha-Sousa, A. & Stalmans, I. Clinical Metabolomics and Glaucoma. *Ophthalmic Res.* **59**, 1–6 (2017).

43. Wang, Y., Hou, X. W., Liang, G. & Pan, C. W. Metabolomics in glaucoma: A systematic review. *Investig. Ophthalmol. Vis. Sci.* **62**, 9 (2021).

44. Umeno, A. *et al.* Comprehensive measurements of hydroxylinoleate and hydroxyarachidonate isomers in blood samples from primary open-angle glaucoma patients and controls. *Sci. Rep.* **9**, 2171 (2019).

45. Javadiyan, S. *et al.* Elevation of serum asymmetrical and symmetrical dimethylarginine in patients with advanced glaucoma. *Invest. Ophthalmol. Vis. Sci.* **53**, 1923–1927 (2012).

46. Zeleznik, O. A. *et al.* Plasma metabolite profile for primary open-angle glaucoma in three US cohorts and the UK Biobank. *Nat. Commun.* **14**, 2860 (2023).

47.   Barbosa Breda, J. *et al.* Metabolomic profiling of aqueous humor from glaucoma patients - The metabolomics in surgical ophthalmological patients (MISO) study. *Exp. Eye Res.* **201**, 108268 (2020).

48.   Bailey, J. N. C. *et al.* Hypothesis-independent pathway analysis implicates GABA and Acetyl-CoA metabolism in primary open-angle glaucoma and normal-pressure glaucoma. *Hum. Genet.* **133**, 1319–1330 (2014).

49.   Rossi, C. *et al.* Multi-omics approach for studying tears in treatment-naïve glaucoma patients. *Int. J. Mol. Sci.* **20**, 4029 (2019).

50.   Murai, H. *et al.* Cerebral glucose metabolism in the striate cortex positively correlates with fractional anisotropy values of the optic radiation in patients with glaucoma. *Clin. Exp. Ophthalmol.* **43**, 711–719 (2015).

51.   Qian, C. *et al.* Machine learning identifying peripheral circulating metabolites associated with intraocular pressure alterations. *Br. J. Ophthalmol.* 1–6 (2022). doi:10.1136/bjophthalmol-2021-320584

52.   Boillot, T. *et al.* Determination of morphological, biometric and biochemical susceptibilities in healthy eurasier dogs with suspected inherited glaucoma. *PLoS One* **9**, e111873 (2014).

7. Supporting Information
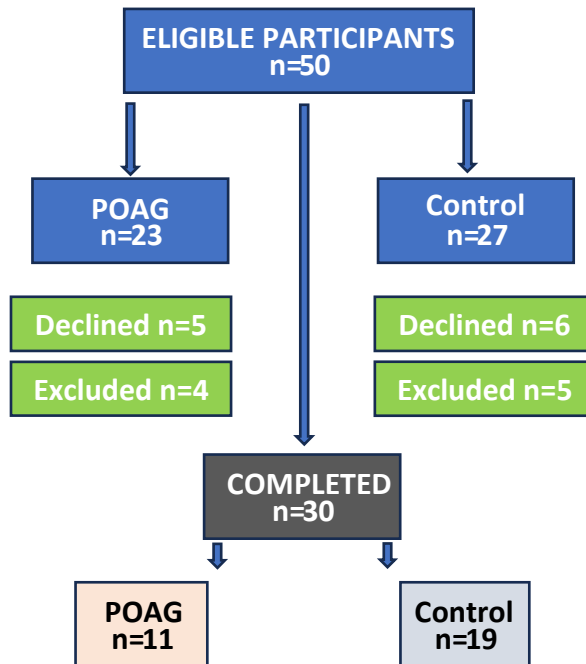
Figure S1. Flowchart of the participants recruitment and the final sample size of the pilot study.
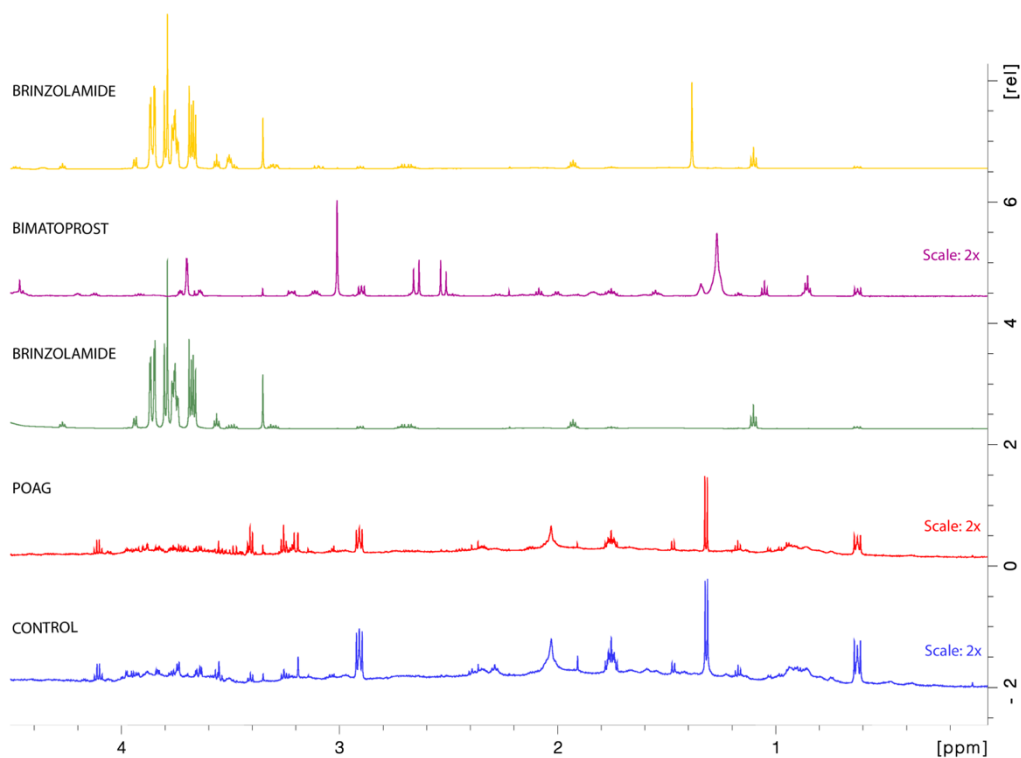
Figure S2. Aliphatic region of the spectra from bimatoprost and two brinzolamide comertial drugs, one POAG and one control sample. NMR of the topical drugs commonly used for the treatment of ocular hypertension in POAG patients (yellow, purple and green) compared to a POAG (red) and a control (blue) tears samples. The spectra where acquiered to ensure that the statistical differences observed between control and POAG where not directly produced by the clinical treatment.

Chapter 4 | Metabolomic Study for the Identification of Symptomatic Carotid Plaque Biomarkers

Metabolomic Study for the Identification of Symptomatic Carotid Plaque Biomarkers

Marina Botello-Marabotto, [1,2,3] Emma Plana, [4,*] M.Carmen Martínez-Bisbal, [1,2,3,5,*] , Pilar Medina, [4] Andrea Bernardos, [2,3,6] Ramón Martínez-Máñez, [1,2,3,6,7,#] Manuel Miralles, [4,8,9,#]

1. Unidad Mixta de Investigación en Nanomedicina y Sensores. Instituto de Investigación Sanitaria La Fe (IISLAFE), Universitat Politècnica de València. Valencia, Spain
2. Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), UPV-UV, Valencia, Spain
3. CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III
4. Grupo acreditado de Hemostasia, Trombosis, Arteriosclerosis y Biología Vascular. Instituto de Investigación Sanitaria La Fe (IISLAFE), Valencia, Spain
5. Departamento de Química Física, Universitat de València, Valencia, Spain
6. Unidad Mixta UPV-CIPF de Investigación en Mecanismos de Enfermedades y Nanomedicina, Universitat Politècnica de València, Centro de Investigación Príncipe Felipe, Valencia, Spain
7. Departamento de Química, Universitat Politècnica de València, Valencia, Spain
8. Servicio de Angiología y Cirugía Vascular. Hospital Universitario y Politécnico La Fe. Valencia.
9. Departamento de Cirugía. Facultad de Medicina. Universidad de Valencia.

\# These two authors (R.M.-M.; M.M.) are sharing last place as the study coordinators.

* Corresponding authors: Emma Plana. E-mail address: emma_plana@iislafe.es and M. Carmen Martínez-Bisbal. E-mail address: carmen.martinez-bisbal@uv.es

Submitted

Graphical Abstract
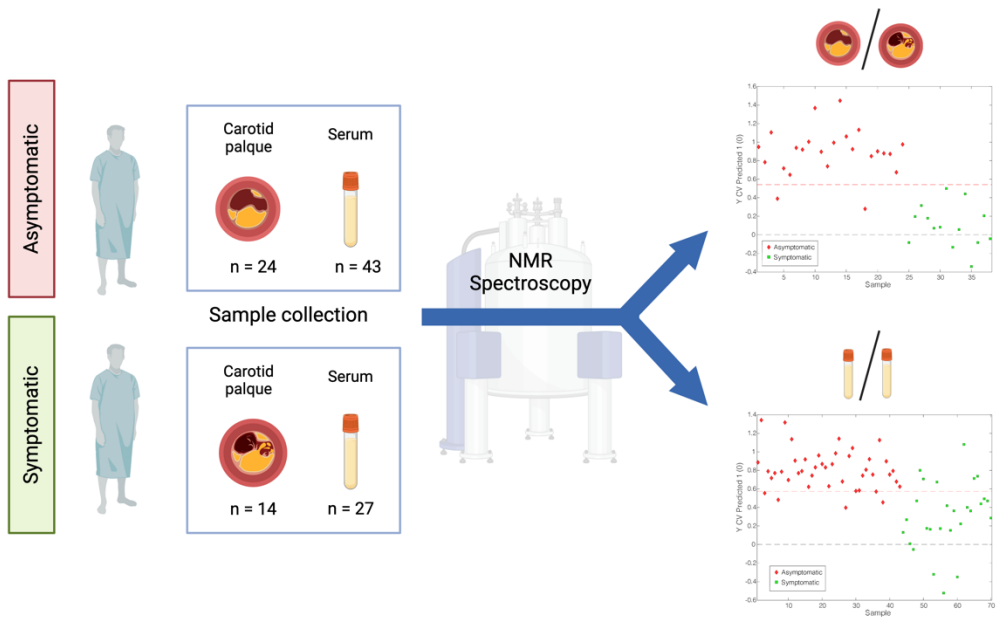
Keywords

Atherosclerosis; Vulnerable plaque; Metabolomics; ^{1}H-NMR Spectroscopy; Biomarkers

My contribution

I participated in the preparation of the samples, the acquisition of the spectra, performed the deconvolution of the signals from the spectra, the statistical analysis, discussion, and wrote the original draft.

1. Abstract

Carotid artery stenosis is mainly produced due to the progressive accumulation of atherosclerotic plaque in the vascular wall. The atherosclerotic plaque is characterized by the accumulation of lipids, low density proteins, expression of chemokines and adhesion molecules, and migration of monocytes and lymphocytes into the plaque. Its rupture can produce stroke, but embolic propensity depends principally on the composition and vulnerability of plaque rather than the severity of stenosis. It is important, then, to ascertain which patients with carotid arterial stenosis have a greater risk of developing neurological symptomatology. Here, we present a metabolomic study by using nuclear magnetic resonance (NMR) spectroscopy in atheroma plaque and serum samples from patients with recently symptomatic and asymptomatic carotid stenosis to search for metabolites that could be used as biomarkers associated with plaque vulnerability and subsequent risk of rupture. Thirty-eight atheromatous plaque samples (24 asymptomatic patients and 14 symptomatic) and 70 serum samples (43 asymptomatic and 27 symptomatic) were studied by NMR spectroscopy. The data were analysed using multivariate statistics (PLS-DA) to determine a model to discriminate between symptomatic and asymptomatic samples (atheroma plaques and sera). The calculated PLS-DA models showed a 100% sensitivity and a 96.6% specificity for the cross validation to discriminate between symptomatic and asymptomatic plaques, and 88,37% sensitivity and 77,78% specificity when serum samples were analysed. According to the results of univariate and multivariate analysis, histamine, phenylalanine, and tyrosine are postulated as potential biomarkers of plaque rupture in serum, as well as glutamate, which is found in higher concentrations in symptomatic plaques.

2. Introduction

Stroke and myocardial infarction are still among the leading causes of death worldwide,[1] being the development of atherosclerosis (AE) the underlying cause in most cases.[2] Based on its histological appearance and physiopathology, AE can be considered a chronic inflammatory disease of the circulatory system.[3] It usually begins with impairment of the molecular activity of the intima layer or endothelial dysfunction, allowing the entry of macromolecules such as low-density lipoproteins (LDL) into the subendothelial layer, particularly in turbulent blood flow areas. This leads to an inflammatory reaction, with LDL becoming oxidized (ox-LDL). These particles develop chemoattraction and recruitment of monocytes and lymphocytes from the bloodstream coming inside the vessel wall. Ox-LDL particles are subsequently phagocyted and internalized by macrophages and transformed into foam cells, which are highly stimulated to produce inflammatory cytokines. Ultimately, this inflammatory environment, expression of cytokines, and growth factors result in vascular smooth muscle cell (VSMC) migration from the *media* towards the *intima*, changing their phenotype and producing extracellular matrix proteins, including collagen, as well as new waves of cytokines and growing factors. This process leads to the formation of a fibrous cap surrounding a lipid core of merging LDL particles released from apoptotic foam cells and the constitution of the atheroma plaque. The evolution of these plaques can lead to either a stable or unstable situation, primarily depending on the integrity and thickness of the fibrous cap.[4]

The most clinically significant consequence of atherosclerosis plaques is the impairment of blood flow to the downstream tissues and organs, mainly due to the diameter reduction or stenosis of the vessel. Clot formation at this point may block blood flow, leading to a stroke if the occlusion is produced in the carotid artery or to myocardial infarction if it is produced in the coronary artery. The formation of thrombus is produced by the rupture or erosion of the plaque and the release of pro-thrombotic factors into the bloodstream. Previous to the rupture of the plaque, the

fibrous cap may become thin and unstable, which is known as vulnerable plaque.[5] The reasons why the plaque can become vulnerable, and rupture remain unclear.

In this context, the identification of biomarkers of plaque vulnerability would be of utmost clinical value as it could help to determine for which patients an intervention would be required to prevent clinical symptomatology.

Metabolomics has been explored as a tool for the search of biomarkers in different clinical scenarios such as cancer, inborn errors of metabolism, diabetes, infection, glaucoma, or Alzheimer's disease, among many others.[6–12] Actually, metabolomics is the omic that encloses more information to determine the cause of the pathologic problem. That is because a concrete metabolome is produced due to the combination of genetic and environmental factors, among others, so it provides broader information about the molecular features of the phenotype than any other omics science.[13] There are two main techniques to ascertain the metabolome: mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy is a robust and reproducible technique that allows the determination of the metabolic composition of any sample (blood derivatives, cerebrospinal fluid, culture cells and media, urine, tears, tissue, etc.) without destroying it. In addition, the acquisition of NMR spectra requires simple preparation procedures and only a small number of samples.[14] Furthermore, NMR spectroscopy allows the study of intact tissue samples through high-resolution magic angle spinning (HRMAS) without the need for extracts.[15] This opens the possibility to study samples of lower size/weight[15,16] and also enables the performance of quantitative studies.[17]

In the context of atherosclerosis diagnosis and research, previous metabolomic studies for the identification of biomarkers have been performed using NMR spectroscopy. For instance, MS. P. Wurtz *et al*.[18] performed an NMR spectroscopy study for the identification of subclinical biomarkers of atherosclerosis in plasma and found docosahexaenoic acid, glutamine, and tyrosine as potential biomarkers of atherosclerosis progression. P.A. Vorkas *et al.*[19] studied the metabolome composition

of atheroma plaque by ultra-high performance liquid chromatography mass spectrometry (UHPLC–MS), and they found changes in the levels of acylcarnitine and phosphatidylethanolamine-ceramides. L. Tomas *et al.*[20] studied by MS the presence of biomarkers of vulnerability in carotid plaque samples. These authors found that the metabolite profile of the carotid plaques correlated with the histological analysis of the plaques in the assessment of their stability. In high-risk plaques, there was increased glycolysis and amino acid metabolism and decreased fatty acid oxidation. However, previous studies do not offer a leap in the understanding of the causes leading to the development of a high-risk plaque, and a set of biomarkers that could be explored as a prognostic tool for plaque instability associated with neurological symptomatology has not yet been identified.

Aimed for this context, we have performed a metabolomic analysis of atheroma plaque and serum from patients with symptomatic and asymptomatic carotid stenosis, using NMR spectroscopy for the search of new molecules and circulating markers related to plaque vulnerability and risk of stroke in patients with carotid stenosis. These metabolites could be used as biomarkers to determine plaque vulnerability and risk of rupture.

3. Materials and Methods

3.1 Patients selection

This is a case-control study, where subjects were classified as recently symptomatic patients (case) and asymptomatic patients (control). Symptomatic patients were considered when transient ischaemic attack, *amaurosis fugax*, or stroke with ischaemic origin occurred at most 3 weeks before surgery. Asymptomatic patients were subjects with severe carotid stenosis (>70%) without neurological symptomatology submitted for carotid endarterectomy on the basis of international guidelines. Inclusion criteria

were age between 18-90 and absence of malignancies in an advanced state, and human immunodeficient virus. Patients were recruited by the Angiology and Vascular Surgery Service from the Hospital Universitario y Politécnico La Fe. The present study was performed according to the declaration of Helsinki and was approved by the ethical committee of the Medical Research Institute Hospital La Fe (references 2019-094-1 and 2020-189-1_PI20/01171). Samples and data from subjects included in this study were managed and provided by Biobank La Fe (PT17/0015/0043) after approval by the scientific and ethical committees. All participants agreed to donate samples to biobank and gave written informed consent.

3.2 Sample collection

Blood samples were obtained during surgery in a clot activator tube, centrifuged at 1811 g for 30 min at 4 °C within 3h after collection to separate the serum fraction, which was distributed in 300-500μl aliquots and stored at -80 °C until used.

Atheromatous plaque samples were obtained during endarterectomy procedure according to standardized surgical protocols in the Angiology and Vascular surgery service from Hospital Universitario y Politécnico La Fe. Samples were rinsed with phosphate buffered saline (PBS), snap frozen in liquid nitrogen, and stored at -80 °C until use. A fragment of 0.3 cm from the most occluded zone was used for RMN analysis.

3.3 Sample preparation

The 38 collected atherosclerotic plaques (14 from symptomatic patients and 24 from asymptomatic ones) were analysed by HRMAS spectroscopy. For this purpose, tissue samples (10 ± 5 mg) were introduced in a 4 mm disposable micro-rotor with 10 l of

D_2O, with sodium 3-(trimethylsilyl)-1-propanesulfonic acid d_6 sodium salt (DSS-d_6) 5 mM as an internal standard for chemical shift referencing.

In a parallel substudy, 70 serum samples (35 of them paired with the atheroma plaque samples studied by HRMAS) were analysed by NMR spectroscopy. The samples were prepared following the protocol described by Beckonert *et al.*[21] Briefly, 250 µL of the serum sample were introduced in 5 mm NMR tubes and 250 µL of phosphate buffer was added. Phosphate buffer contained deuterated water (20% v/v) and DSS 1 mM as an internal standard for chemical shift referencing.

3.4 NMR spectra acquisition and processing

The atheroma plaque experiments were carried out in a Bruker Avance II 500 MHz equipped with an HRMAS probe (Bruker BioSpin). Noesy sequence was used for the acquisition of the 1D ^{1}H-NMR spectra of the plaques. 256 scans and a spectral width of 14 ppm were acquired. NMR spectra from the serum samples were acquired in the same 500 MHz spectrometer equipped with a TXI probe. Noesy experiments were acquired, with 128 repetitions and a spectral width of 30 ppm. In both types of samples, 2D homo (^{1}H-^{1}H) and heteronuclear (^{1}H-^{13}C) spectra were acquired to guarantee the assignment of the metabolites in the spectra. The temperature of the probes for both tissues and sera was set to 278 K (5 °C).

Once obtained, the 1D spectra were transformed and post-processed with the *TopSpin 4.0.7* software (Bruker BioSpin Corporation). The phase and baseline of the spectra were corrected, and the chemical shift was calibrated according to the DSS CH_3 signal at 0.0 ppm. The Human Metabolome Database (HMDB),[22] Chenomx NMR Profiler,[23] and previously published bibliography[15] were used for metabolite identification.

Meaningful signals underwent deconvolution using *AMIX 4.0.2* software (Bruker BioSpin Corporation), excluding the water suppression signal (4.5-5.0 ppm) and the

signals at ppm lower than 0.5 ppm and higher than 8.5 ppm. Then, a mixed Gaussian/Lorentzian variable function was applied for the deconvolution of these signals. Later on, integrals were obtained for all spectra in atheroma plaque tissue and serum samples. For serum samples, relative quantification was based on the integration of the signal divided by the sum of the integrals of all the signals in the spectrum. Normalization was done for the atheroma plaque samples according to their weight.

3.5 Multivariate statistical analysis

Multivariate statistical analysis was performed using the *PLS_Toolbox Solo 8.9* software (Eigenvector Research, Inc., Manson, WA, USA). The same procedures were followed with the data from the tissue and from the serum samples.

Before the model calculation, a variable reduction was performed in order to simplify the data set. The variables included in the model were selected using Genetic Algorithms (GA). GA is a common machine learning technique based on the process of natural selection that has been proven to be helpful in feature selection processes.[24] Fifty maximum generations, PLS regression with 10 latent variables, and random cross-validation (CV) with autoscaling and normalization by sum of the data were applied for variable selection.

After variable selection, partial least squares-discriminant analysis (PLS-DA) was used to generate a predictive model with the obtained variables. CV (Venetian blinds) was used to select the finest number of latent variables for the model. In order to determine the goodness of the model to discriminate between different sets of samples, a receiver operating characteristic (ROC) was generated, and the area under the curve (AUC), sensitivity and specificity were calculated. The robustness and over-fitting of the model were tested through permutation tests (100 iterations).

3.6 Univariate statistical analysis

Univariate statistical analysis was performed on the data from atheroma plaque and serum. The Shapiro-Wilks test was applied to check the normality of the variables included in the study. Subsequently, t-test or U de Mann Whitney test were used, when appropriate, for the comparison of the means of the relative concentrations of the analysed metabolites. *IBM SPSS Statistics 25 version* was used for univariate statistics.

3.7 Metabolite set enrichment analysis

To investigate which metabolic pathways could be involved in plaque vulnerability, *Metaboanalyst*[25] was used. HMDB ID of each metabolite was used to include them in the enrichment analysis. The lipid motifs and unknown metabolites were excluded from the analysis. The metabolite set library chosen for the analysis was the one based on the KEGG human metabolic pathways. Pathways with a p.value < 0.1 and impact > 0 were considered to be significant.

4. Results and Discussion

Plaques from symptomatic and asymptomatic patients, as previously defined, have been analysed by NMR spectroscopy to identify differences in the metabolome that might be associated with plaque vulnerability. Serum samples obtained from patients with and without clinical symptomatology were also analysed. A total of 70 serum samples and 38 atheroma plaque samples were included, 35 of which were paired

plaque and serum samples from the same patient. Both groups were balanced in terms of age, sex, and main cardiovascular risk factors (Table 1).

Table 1. Clinical data of patients participating in the study.

	Plaque		Serum	
	Symptomatic (n=14)	Asymptomatic (n=24)	Symptomatic (n=27)	Asymptomatic (n=43)
Age, Mean (range)	70.4 (60-83)	69.4 (50-81)	68.6 (44-83)	68.8 (50-81)
Sex, male/female	13/1	21/3	24/3	38/5
Hypertension (%)	92.85%	70.83%	80.76%	76.74%
Diabetes Mellitus (%)	50%	50%	50%	39.53%
Dyslipidemia (%)	64.28%	62.5%	61.54%	69.76%

4.1 Metabolic profiling of atherosclerotic plaque tissues and serum samples

The signals in the spectra were assigned to the corresponding metabolites for atheroma plaques (Figure 1a) and serum samples (Figure 1b). In the tissue spectra, due to the importance of the broad and intense signals coming from molecules such as lipids or lipoproteins, the identification of smaller signals produced by metabolites was challenging. Nevertheless, we were able to identify a total of 25 metabolites; 6 of them were amino acids, and 5 were fatty acids, among others (Table 2). In serum samples, a total of 26 metabolites were identified (Table 2), 10 of them were amino acids, and 6 were fatty acids. Remarkably, 17 of the 34 metabolites identified were found both in tissue and serum samples.
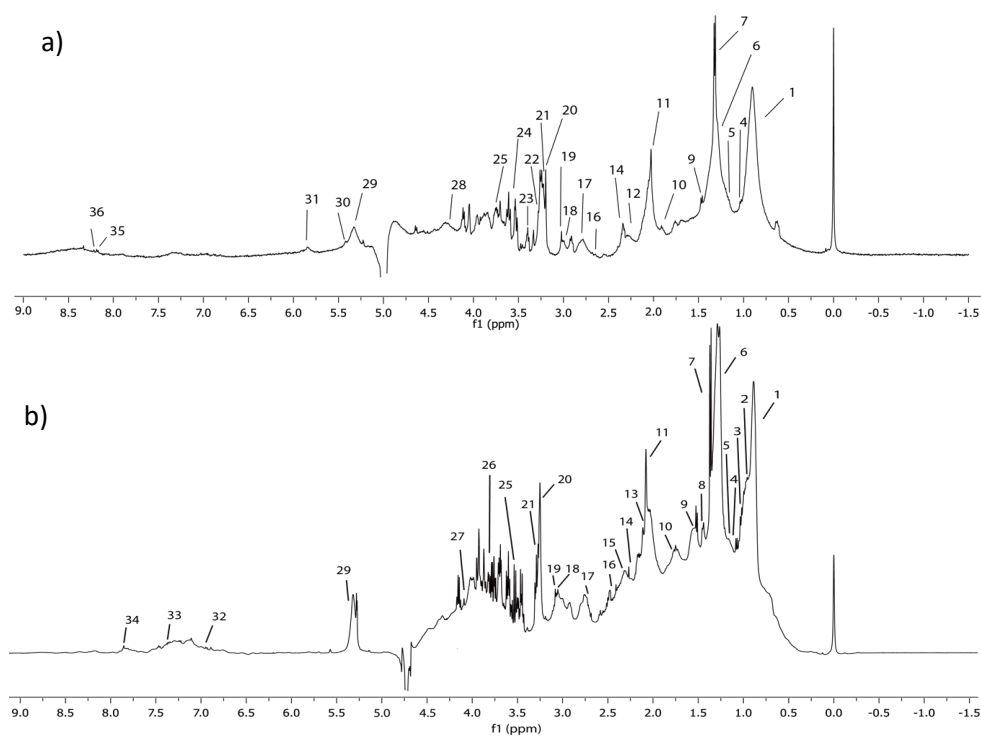
Figure 1. Annotations of the metabolites present in the samples. a) Representative spectrum of tissue samples from atherosclerotic plaques. b) Representative spectrum of serum samples from patients with atherosclerosis. 1. Fatty Acids (-CH_2-CH_3), 2. Isoleucine, 3. Leucine, 4. Valine, 5. Ethanol, 6. -(CH_2)n-, 7.Lactate, 8. Threonine, 9. Alanine, 10. Acetate, 11. N-acetylglucosamine, 12. Glutamate, 13. Acetoacetate, 14. Pyruvate, 15. Glutamine, 16. Citrate, 17. -CH=CH-CH_2-CH=CH-, 18. Lysine, 19. Creatine, 20. Choline, 21. Acetylcholine, 22. Glucose-6-phosphate, 23. Proline, 24. Myo-inositol, 25. Glucose, 26. Glycerol, 27. Creatinine, 28. Triacylglycerol (-CH_2-OCO-), 29. -CH=CH-, 30. Glucose-1-phosphate, 31. Uracil, 32. Tyrosine, 33. Phenylalanine, 34. Histamine, 35. Hypoxanthine, 36. Inosine.

4.2 Identification of potential biomarkers of vulnerability in plaque and serum

Partial Least Square discriminant analysis (PLS-DA) was performed in both serum and tissue data. Cross-validated models were generated to discriminate between plaques from symptomatic and asymptomatic patients.

For the atheromatous plaque tissue analysis, a total of 120 signals (variables) were integrated from the NMR spectra. GA was applied to the data matrix, and the best model was obtained with a total of 23 variables. The resultant PLS-DA, using 3 principal components, was able to correctly classify stable and vulnerable plaques with 100% sensitivity and 91.6% specificity. The ROC curve rendered an AUC value of 0.985, showing the excellent classification capacity of the model (Figure 2a). The model was statistically significant ($p < 0.05$) for all the permutation tests. The metabolites included in the model were threonine, alanine, glutamate, citrate, myo-inositol, glucose, creatine, lactate, uracil, hypoxanthine, inosine, lipid fragments $-(CH_2)_n-$, -CH=CH- and -CH=CH-CH$_2$-CH=CH-, and two unknown compounds (Table 3). All the metabolites in vulnerable plaques were increased. Moreover, myo-inositol, glutamate, and the -CH=CH- lipid motif showed significant differences between plaques from symptomatic and asymptomatic patients according to the univariate analysis (Table 3).

Table 2. Annotation of the metabolites present in serum and plaque samples. The sample type, the group responsible for the NMR signal, chemical shift, multiplicity, and J coupling are displayed. To simplify, only one signal per metabolite is shown in the table.

Number assignment	Metabolite	Sample type	Group	Chemical Shift (ppm)	Multiplicity and J coupling (Hz) *
1	Fatty Acids (-CH_2-CH_3)	Plaque/Serum	CH_3	0.83	-
2	Isoleucine	Serum	CH_3	0.93	t, J = 7.41
3	Leucine	Serum	CH_3	0.95	t, J = 5.89
4	Valine	Plaque/Serum	γXH_3	0.98	d, J = 7.2
5	Ethanol	Plaque/Serum	CH_2	1.16	t, J = 7.08
6	-$(CH_2)_n$-	Plaque/Serum	CH_2	1.26	-
7	Lactate	Plaque/Serum	CH_3	1.32	d, J = 7.00
8	Threonine	Serum	γCH_3	1.33	d, J =6.58
9	Alanine	Plaque/Serum	βXH_3	1.48	d, J = 7.40
10	Acetate	Serum	CH_3	1.91	s
11	N-acetylglucosamine	Plaque/Serum	CH_3	2.03	s
12	Glutamate	Plaque	βCH_2	2.13	m
13	Acetoacetate	Serum	CH_3	2.27	s
14	Pyruvate	Plaque/Serum	CH_3	2.36	s
15	Glutamine	Serum	γCH_2	2.46	dt, J = 14.4, 6.8
16	Citrate	Plaque/Serum	CH_2	2.52	d, J = 15.80
17	-CH=CH-CH_2-CH=CH-	Plaque/Serum	CH_2	2.79	-
18	Lysine	Plaque/Serum	CH_2	3.09	t, J = 6.50
19	Creatine	Plaque/Serum	CH_3	3.021	s
20	Choline	Plaque/Serum	-N^+-$(CH_3)_3$	3.19	s
21	Acetylcholine	Plaque/Serum	-N^+-$(CH_3)_3$	3.21	s
22	Glucose-6-phosphate	Plaque	CH_2	3.25	dd, J = 9.21, 7.99
23	Proline	Plaque/Serum	CH_2	3.33	m
24	Myo-inositol	Plaque/Serum	C1HC3H	3.59	dd, J = 3.4, 2.5
25	Glucose	Plaque/Serum	C6H u	3.60	m
26	Glycerol	Serum	1,3 CH_2OH	3.65	m
27	Creatinine	Serum	CH_2	4.09	s
28	Tryacilglicerols	Plaque	-CH_2-OCO-	4.30	-
29	-CH=CH-	Plaque/Serum	CH	5.28	-
30	Glucose-1-phosphate	Plaque	C1H	5.42	s
31	Uracil	Plaque	C6H	5.84	-
32	Tyrosine	Serum	CH 2,6	7.19	m
33	Phenylalanine	Serum	CH 2,6	7.27	m
34	Histamine	Serum	CH	7.79	s
35	Hypoxanthine	Plaque	CH	8.17	s
36	Inosine	Plaque	CH	8.22	s

* The multiplicity of the signals is indicated as s: singlet; d: doublet, t: triplet, m: multiplet, dd: doublet of doublets

Table 3. Metabolites participating in the discriminative models of asymptomatic *vs.* symptomatic plaques and sera.

	Metabolites	VIPS*	p-value	Trend
Plaque	*Myo-inositol*	**1.989**	**0.011**	↑**
	Glutamate	**1.768**	**0.006**	↑
	-CH=CH-	**1.511**	**0.055**	↑
	Glucose	1.063	0.224	↑
	Threonine	0.988	0.341	↑
	Hypoxanthine	0.947	0.715	↑
	$-(CH_2)_n-$	0.861	0.286	↑
	$-CH=CH-CH_2-CH=CH-$	0.828	0.686	↑
	Lactate	0.747	0.274	↑
	Uracil	0.737	0.304	↑
	Citrate	0.675	0.171	↑
	2-aminobutyrate	0.558	0.107	↑
	Glucose-1-phosphate	0.533	0.135	↑
	Inosine	0.526	0.421	↑
	Creatine	0.517	0.201	↑
Serum	*Threonine*	**1.887**	**0.000**	↑
	-CH=CH-	**1.431**	**0.015**	↑
	Histamine	**1.179**	**0.056**	↑
	Acetoacetate	0.973	0.606	↓
	$-CH_2-CH_3-$	0.868	0.277	↓
	Methanol	0.829	0.540	↓
	Isoleucine	0.701	0.198	↓
	Valine	0.700	0.672	↑
	Lysine	0.642	0.876	↓
	Alanine	0.511	0.258	↑
	Lactate	0.493	0.613	↓

*VIP = Variable importance in projection, represents the importance of the metabolite in the discriminative capacity of the model
** The arrow direction indicates the change in concentration of symptomatic compared to asymptomatic patients. Metabolites with significant differences are shown in bold

Previous studies have shown that an increase in plasma glutamate is related to a higher risk of cardiovascular diseases (CVD). Plasma glutamate levels were associated with an increased risk of stroke in a study performed in 980 participants under follow up and found that baseline glutamate was associated with 43% and 81% increased risk of CVD and stroke, respectively.[26] These results were validated by a different study where type 2 diabetes mellitus patients showed a positive correlation between glutamate plasma levels and the risk of stroke.[27]

In the analysis of serum samples, a total of 130 signals from the spectra were integrated. Representative signals from each metabolite were included as variables for the PLS-DA model, with a total of 32 variables. After applying GA, 12 variables were selected for the discriminative model. The final model was able to classify the serum of asymptomatic from symptomatic patients with an 88.37% sensitivity and a 77.78% specificity. The AUC value was 0.89 (Figure 2b). The discriminant metabolites were threonine, histamine, acetoacetate, methanol, isoleucine, valine, lysine, alanine, lactate, and the lipid motifs -CH_2-CH_3- and -CH=CH-. The model was also significant for all the permutation tests. Moreover, according to the univariate analysis, threonine, -CH=CH- and histamine showed a notable difference, with an increase in the serum of symptomatic patients for the three metabolites (Table 3).

As it has been observed in this study, an increase in the concentration of histamine has previously been related to the occurrence of cardiovascular events.[28] Blood histamine levels are increased in the inflammatory process produced in unstable plaques.[29] Histamine is an inflammatory mediator produced from L-histidine released from mast cells. The accumulation of activated mast cells and histamine in the atherosclerotic lesion has been associated with the vulnerability of the plaque.[30] Blood vessels dilatation increased vascular permeability, and the migration of leukocytes into the atheroma plaque are among the effects of high levels of histamine in blood. Furthermore, histamine can trigger the production of the glycoprotein tissue factor (TF) by endothelial cells and smooth muscle cells.[31] TF, usually encrypted in the endothelial cells, is a key protein that triggers coagulation through the extrinsic pathway once exposed to the blood flow, leading to thrombin formation,[32] and contributes to the development of thrombosis within the atherosclerotic arteries.[31]
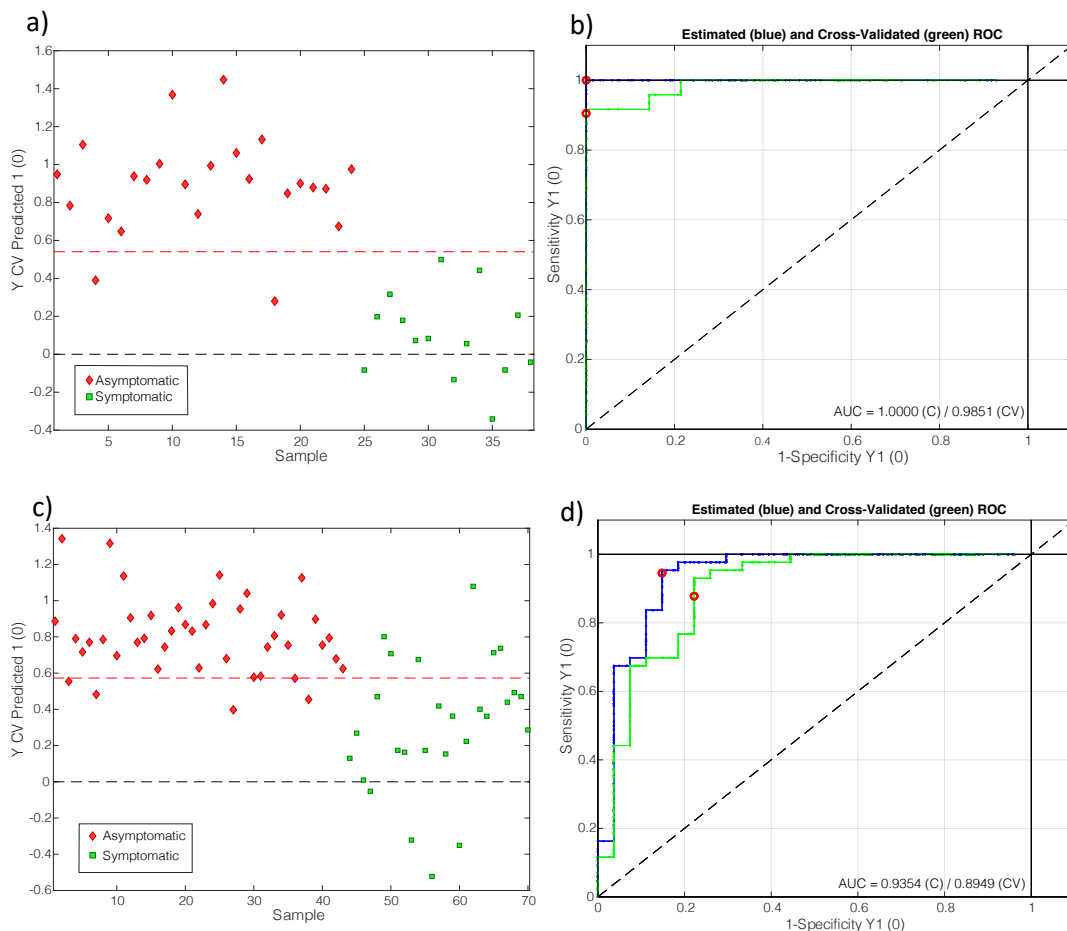
Figure 2. PLS-DA model and ROC curve with AUC values obtained based on the variables selected by the application of genetic algorithms. a) PLS-DA samples classification of the plaque samples. b) ROC curve for the discriminative model of plaque samples. Calibration and CV results are shown. An AUC value of 0.98 for the CV is reached. c) PLS-DA samples classification of the serum samples. d) ROC curve of the classification of the estimated (blue) and calibrated (green) of the serum samples. An AUC value of 0.89 is obtained. Asymptomatic samples are represented by red diamonds and symptomatic samples by green squares.

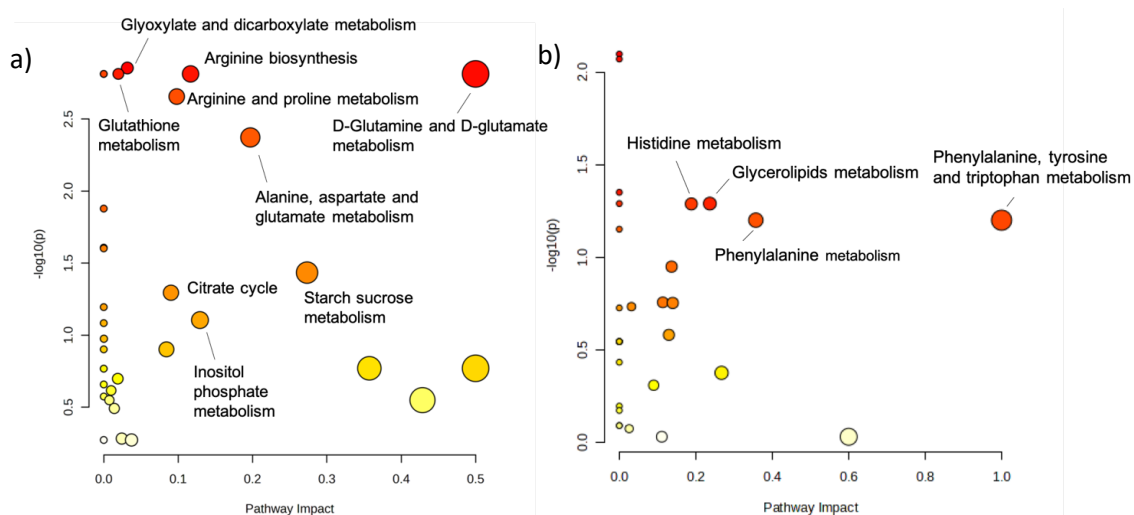## 4.3 Metabolic pathways involved in plaque vulnerability

Metaboanalyst 4.0 was used to generate a metabolite set enrichment analysis for the identification of biological patterns associated with the metabolomic data obtained from the serum and plaque samples. First, the metabolites are individually analyzed to determine if they are significant, and later, the existence of meaningful patterns is determined. It can detect subtle but consistent differences among related compounds that conventional methods may not detect. We defined as notable differences those with a p.value < 0.1 and an impact > 0.

In atheroma plaque samples, nine metabolic pathways were potentially altered between symptomatic and asymptomatic patients. Six pathways were statistically significant with a p.value < 0.05: glyoxylate and dicarboxylate metabolism, D-glutamine and D-glutamate metabolism, arginine biosynthesis, arginine, and proline metabolism, alanine, aspartate and glutamate metabolism, starch and sucrose metabolism, and three pathways notably altered (p. value < 0.1: citrate cycle (TCA cycle), valine, leucine, and isoleucine biosynthesis, and inositol phosphate metabolism (Figure 3a).

Our results are consistent with the fact that both groups are composed of individuals with atherosclerosis, which may lead to discrete changes in the metabolic profiles. The p.value was set to 0.1 due to the limited number of samples but also because subtle differences were expected to be observed as it is not anticipated that the metabolism of cells in vulnerable and stable plaques will produce substantial changes, and even smaller changes are expected to be reflected in the bloodstream.

As detailed in table 4, glutamate is a common metabolite in the pathways. As it has been previously referred, increased glutamate levels are positively correlated with cardiovascular risk. Also, circulating glutamate has been associated with subclinical atherosclerosis independently of other risk factors.[33] In fact, glutamate-mediated excitotoxicity has been previously proposed as a potential therapeutic target for

ischemic stroke.[34] Besides, citrate and citric acid metabolism has also previously been related to increased risk in different CVD.[35,36]

Figure 3. **Metabolic Pathways related to plaque vulnerability based on** a) the metabolites found in the plaque. b) the metabolites found in the serum.

In serum samples, there were four pathways notably altered between symptomatic and asymptomatic patients: histidine metabolism, glycerolipids metabolism, phenylalanine metabolism, and phenylalanine, tyrosine, and tryptophan metabolism (Figure 3b). The alteration in the metabolism of amino acids has previously been reported to be involved in the development of atherosclerosis.[37] More specifically, the levels of phenylalanine and tyrosine have been associated with an increased risk of coronary artery disease and stroke.[38] Accordingly, differences in metabolic pathways where phenylalanine and tyrosine are participating occur between symptomatic and asymptomatic patients (Table 5). The relevance of histamine and histidine metabolism has already been discussed.

Table 4. Metabolic pathways involved in plaque vulnerability, based on the differences in the metabolites found in the plaques.

Metabolic pathway	Metabolites	P-value	Impact*
Glyoxylate and dicarboxylate metabolism	L-Glutamate, Citrate	0.0014	0.03
D-Glutamine and D-glutamate metabolism	L-Glutamate	0.0015	0.50
Arginine biosynthesis	L-Glutamate	0.0015	0.12
Glutathione metabolism	L-Glutamate	0.0015	0.02
Arginine and proline metabolism	Creatine, L-Glutamate	0.002	0.10
Alanine, aspartate and glutamate metabolism	L-Alanine, L-Glutamate, Citrate	0.004	0.20
Starch and sucrose metabolism	Glucose-1-P, Glucose-6-P	0.036	0.27
Citrate cycle (TCA** cycle)	Citrate	0.05	0.09
Inositol phosphate metabolism	Glucose-6-P, Myo-inositol	0.078	0.13

*Impact: Pathway impact value calculated from pathway topology analysis
**Tricarboxylic Acid Cycle

Table 5. Metabolic pathways involved in plaque vulnerability, based on the differences in the metabolites found in serum.

Metabolic pathway	Metabolites	P-value	*Impact
Glycerolipid metabolism	Glycerol	0.051	0.24
Histidine metabolism	Histamine	0.052	0.19
Phenylalanine, tyrosine, and tryptophan biosynthesis	Phenylalanine, Tyrosine	0.063	1.0
Phenylalanine metabolism	Phenylalanine, Tyrosine	0.063	0.036

*Impact: Pathway impact value calculated from pathway topology analysis

5. Conclusions

We have performed a preliminary study in plaques and serum samples from patients with carotid stenosis by NMR. We have sought biomarkers of plaque vulnerability through the generation of statistical models able to discriminate symptomatic and asymptomatic patients with high sensitivity (100% and 88.37%) and specificity (91.6% and 77.78%) in plaque and serum, respectively. Increased concentration of myo-inositol, glutamate, and fatty acids was found in vulnerable plaques, whereas an

increase in histamine, threonine, and fatty acids was observed in serum samples. Overall, the data presented herein provided valuable information on potential biomarkers of plaque vulnerability, both in plaque and serum samples. Although the presence of biomarkers in plaque is not of high value from a diagnosis perspective (as it implies the extraction of the plaque), the molecular mechanism involved in plaque vulnerability is still unknown, and the metabolites and metabolic pathways described in this work could be a starting point to understand better the reasons why some plaques become vulnerable and other remain stable. Also, locally dysregulated metabolites could highlight new therapeutic targets. On the other hand, the analysis of metabolites present in serum of symptomatic and asymptomatic patients could be studied as biomarkers of risk of plaque rupture. As previously mentioned, histamine, phenylalanine, and tyrosine, which are described in this work as potential biomarkers of plaque rupture, have already been associated with plaque vulnerability. A limitation of the study is the limited number of samples; thus, further studies with an increased number of patients would be necessary to validate our results. Moreover, our samples were obtained when the ischemic event had already occurred; thus, causality cannot be established with complete confidence. A longitudinal and a follow-up study should be performed to understand the applicability of our results better.

6. Bibliography

1.   Mattiuzzi, C. & Lippi, G. Worldwide disease epidemiology in the older persons. *Eur. Geriatr. Med.* **11**, 147–153 (2020).

2.   Banerjee, C. & Chimowitz, M. I. Stroke Caused by Atherosclerosis of the Major Intracranial Arteries. *Circ. Res.* **120**, 502–513 (2017).

3.   Hansson, G. K. Inflammation, Atherosclerosis, and Coronary Artery Disease. *N.*

Engl. J. Med. **325**, 1685–1695 (2005).

4. Hansson, G. K. & Hermansson, A. The immune system in atherosclerosis. *Nat. Immunol.* **12**, 204–212 (2011).

5. Virmani, R., Burke, A. P., Farb, A. & Kolodgie, F. D. Pathology of the unstable plaque. *Prog. Cardiovasc. Dis.* **44**, 349–356 (2002).

6. Gowda, G. A. N. *et al.* Metabolomics-based methods for early disease diagnostics. *Expert Rev. Mol. Diagn.* **8**, 617–633 (2008).

7. Tounta, V., Liu, Y., Cheyne, A. & Larrouy-Maumus, G. Metabolomics in infectious diseases and drug discovery. *Mol. Omi.* **17**, 376–393 (2021).

8. Botello-Marabotto, M. *et al.* Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease. *Neurobiol. Dis.* **187**, 106312 (2023).

9. Loras, A. *et al.* Integrative metabolomic and transcriptomic analysis for the study of bladder cancer. *Cancers (Basel).* **11**, 686 (2019).

10. Loras, A. *et al.* Urinary Metabolic Signatures Detect Recurrences in Non-Muscle Invasive Bladder Cancer. *Cancers (Basel).* **11**, 914 (2019).

11. Oto, J. *et al.* LC–MS metabolomics of urine reveals distinct profiles for non-muscle-invasive and muscle-invasive bladder cancer. *World J. Urol.* **40**, 2387–2398 (2022).

12. Oto, J. *et al.* Urine metabolomic analysis in clear cell and papillary renal cell carcinoma: A pilot study. *J. Proteomics* **218**, 103723 (2020).

13. D'Alessandro, A., Giardina, B., Federica, G., Timperio, A. M. & Zolla, L. Clinical metabolomics: The next stage of clinical biochemistry. *Blood Transfus.* **10**, 19–24 (2012).

14. Ward, J. L. *et al.* An inter-laboratory comparison demonstrates that [1H]-NMR metabolite fingerprinting is a robust technique for collaborative plant

metabolomic data collection. *Metabolomics* **6**, 263–273 (2010).

15.   Martínez-Bisbal, M. C. *et al.* 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed.* **17**, 191–205 (2004).

16.   Martínez-Granados, B. *et al.* Metabolite identification in human liver needle biopsies by high-resolution magic angle spinning 1H NMR spectroscopy. *NMR Biomed.* **19**, 90–100 (2006).

17.   Martínez-Bisbal, M. C. *et al.* Determination of metabolite concentrations in human brain tumour biopsy samples using HR-MAS and ERETIC measurements. *NMR Biomed.* **22**, 199–206 (2009).

18.   Würtz, P. *et al.* High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. *Eur. Heart J.* **33**, 2307–2316 (2012).

19.   Vorkas, P. A. *et al.* Metabolic phenotyping of atherosclerotic plaques reveals latent associations between free cholesterol and ceramide metabolism in atherogenesis. *J. Proteome Res.* **14**, 1389–1399 (2015).

20.   Tomas, L. *et al.* Altered metabolism distinguishes high-risk from stable carotid atherosclerotic plaques. *Eur. Heart J.* **39**, 2301–2310 (2018).

21.   Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).

22.   Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

23.   Chenomx Inc | Metabolite Discovery and Measurement. Available at: https://www.chenomx.com/. (Accessed: 20th March 2024)

24.   Cavill, R. *et al.* Genetic algorithms for simultaneous variable and sample

selection in metabonomics. *Bioinformatics* **25**, 112–118 (2009).

25. Chong, J. *et al.* MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

26. Zheng, Y. *et al.* Metabolites of Glutamate Metabolism Are Associated With Incident Cardiovascular Events in the PREDIMED PREvención con DIeta MEDiterránea (PREDIMED) Trial. *J. Am. Heart Assoc.* **5**, e003755 (2016).

27. Li, R. T. *et al.* Relationship between plasma glutamate and cardiovascular disease risk in Chinese patients with type 2 diabetes mellitus by gender. *Front. Endocrinol. (Lausanne).* **14**, 1095550 (2023).

28. Rosic, M. *et al.* Histamine blood concentration in ischemic heart disease patients. *J. Biomed. Biotechnol.* **2011**, 315709 (2011).

29. Clejan, S. *et al.* Blood histamine is associated with coronary artery disease, cardiac events and severity of inflammation and atherosclerosis. *J. Cell. Mol. Med.* **6**, 583–592 (2002).

30. Yamada, S., Wang, K. Y., Tanimoto, A. & Sasaguri, Y. Novel function of histamine signaling in hyperlipidemia-induced atherosclerosis: Histamine H1 receptors protect and H2 receptors accelerate atherosclerosis. *Pathol. Int.* **65**, 67–80 (2015).

31. Lindstedt, K. A., Mäyränpää, M. I. & Kovanen, P. T. Mast cells in vulnerable atherosclerotic plaques - A view to a kill. *J. Cell. Mol. Med.* **11**, 739–758 (2007).

32. Steffel, J., Akhmedov, A., Greutert, H., Lüscher, T. F. & Tanner, F. C. Histamine induces tissue factor expression: Implications for acute coronary syndromes. *Circulation* **112**, 341–349 (2005).

33. Lehn-Stefan, A. *et al.* Elevated Circulating Glutamate Is Associated with Subclinical Atherosclerosis Independently of Established Risk Markers: A Cross-Sectional Study. *J. Clin. Endocrinol. Metab.* **106**, E982–E989 (2021).

34. Shen, Z. *et al.* Glutamate excitotoxicity: Potential therapeutic target for ischemic stroke. *Biomed. Pharmacother.* **151**, 113125 (2022).

35. Santos, J. L. *et al.* Circulating citric acid cycle metabolites and risk of cardiovascular disease in the PREDIMED study. *Nutr. Metab. Cardiovasc. Dis.* **33**, 835–843 (2023).

36. Stryeck, S. *et al.* Serum Concentrations of Citrate, Tyrosine, 2- and 3-Hydroxybutyrate are Associated with Increased 3-Month Mortality in Acute Heart Failure Patients. *Sci. Rep.* **9**, 6743 (2019).

37. Nitz, K., Lacy, M. & Atzler, D. Amino acids and their metabolism in atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **39**, 319–330 (2019).

38. Jauhiainen, R. *et al.* The Association of 9 Amino Acids with Cardiovascular Events in Finnish Men in a 12-Year Follow-up Study. *J. Clin. Endocrinol. Metab.* **106**, 3448–3454 (2021).

Chapter 5 | Early biomarkers of fibrosis in covid-19 patients one year after hospital discharge

Early biomarkers of fibrosis in covid-19 patients one year after hospital discharge

Marina Botello-Marabotto, [1,2,3] , Julia Tarrasó [4], Alba Mulet [4], Lucía Fernández-Presa [4],

Jaime Signes-Costa [4], Andrea Bernardos, [2,3,5] M.Carmen Martínez-Bisbal*, [1,2,3,6,7]

Ramón Martínez-Máñez, [1,2,3,5,7]

1. Unidad Mixta de Investigación en Nanomedicina y Sensores. Instituto de Investigación Sanitaria La Fe (IISLAFE), Universitat Politècnica de València. Valencia, Spain

2. Instituto Interuniversitario de Investigación de Reconocimiento Molecular y Desarrollo Tecnológico (IDM), UPV-UV, Valencia, Spain

3. CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III

4. Pulmonary Department. Hospital Clinico. INCLIVA. Universitat de València. Spain.

5. Departamento de Química, Universitat Politècnica de València, Valencia, Spain

6. Departamento de Química Física, Universitat de València, Valencia, Spain

7. Unidad Mixta UPV-CIPF de Investigación en Mecanismos de Enfermedades y Nanomedicina, Universitat Politècnica de València, Centro de Investigación Príncipe Felipe, Valencia, Spain

* Corresponding Author: M. Carmen Martínez-Bisbal. E-mail address: carmen.martinez-bisbal@uv.es

Submitted

205

Graphical abstract
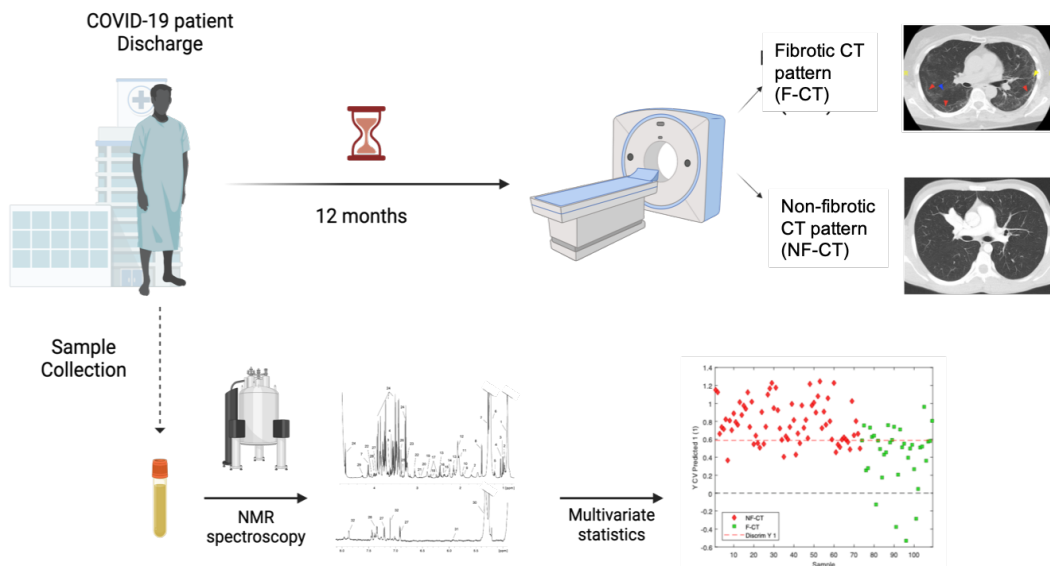
Keywords

COVID-19, Pulmonary Fibrosis, Metabolomics, ^{1}H-NMR Spectroscopy, Biomarkers

My contribution

I performed the deconvolution of the signals from the spectra, assignation of the signals from the spectra, the statistical analysis, discussion, and wrote the original draft.

1. Abstract

COVID-19 global pandemic has affected more than 600 million people up to date. The symptomatology and severity of COVID-19 are very broad, and there are still concerns about the long-term sequelae that it can have on discharged patients. The development of pulmonary fibrotic sequelae after SARS-CoV-2 infection is especially worrying. Metabolomic studies performed in serum of COVID-19 patients have shown that there is a multisystemic effect of the infection. In this work, we have performed a metabolomic study in the serum samples of 109 COVID-19 patients 2 months after hospital discharge to determine if there is a metabolomic signature that could predict the development of pulmonary fibrotic sequelae. We have observed that based on the nuclear magnetic resonance (NMR) analysis of the serum samples, it is possible to distinguish with 80.82% of sensitivity, 72.22 % of specificity and an AUC value of 0.83 which patients would have radiological signs of pulmonary fibrotic pattern one year after sample collection. According to the metabolites participating in the discriminative model and the univariate statistics, glucose, valine, and fatty acids (=CH-CH_2-CH=) are suggested as potential biomarkers of the development of pulmonary fibrotic sequalae after COVID-19.

2. Introduction

The global pandemic caused by SARS-CoV-2 has affected more than 600 million people up to date, and 10-20% of COVID-19 patients developed pneumonia.[1] SARS-CoV-2 produced a broad variety of symptomatology and severity depending on several factors, many of which remain to be clarified.[2] Risk factors and comorbidities such as advanced age coupled with reduced lung function, hypertension, pre-existing diabetes,

cardiovascular diseases, and obesity have heightened the susceptibility to severe COVID-19 infection.[3]

Furthermore, concerns arise about long term effects of COVID-19 survivors, independently of their initial severity.[1] Several studies have shown that 6 to 12 months after recovery from COVID-19 a substantial number of survivors had pulmonary sequelae, with a mean of 32% of survivors showing pulmonary fibrotic patterns in chest computed tomography (CT) studies.[4] Pulmonary fibrosis is produced from recurrent epithelial mild injuries and inadequate healing processes, leading to the over activation of fibroblasts and the excessive accumulation of extracellular matrix. This results in a mechanically stretched environment, producing an alteration in the respiratory function which can lead to dyspnea.[5]

Pulmonary fibrosis sequelae are not only produced in COVID-19 patients affected with acute respiratory distress syndrome, but also in patients with mild disease. This means that SARS-CoV-2 could have pro-fibrotic effects itself.[6] There is a need to find potential biomarkers of the development of pulmonary fibrosis to determine which COVID-19 patients are more susceptible to have sequelae.

Metabolomics has been proved to be a good source of biomarkers of different conditions and diseases, such as Alzheimer's disease,[7] cancer,[8] glaucoma,[9] or infectious diseases[10] among others. Metabolomics can be defined as the analysis of the product of all the metabolic processes occurring in a living system, and implies the study of small molecules called metabolites. It reflects the interplay between the expression of genes in an individual and the effect of environmental changes, being the *omic* discipline that can be more directly related to phenotype.[11]

Different metabolomic studies have been performed up to date to identify biomarkers of COVID-19 infection and progression,[12] and some articles have been published where metabolomic biomarkers of COVID-19 complications have been found. *Yang et al.*

performed a metabolomic and proteomic study by mass spectrometry to identify biomarkers of pulmonary fibrosis and fibrosis progression. They found an association of cγR-mediated phagocytosis, PPAR signaling, TRP-inflammatory pathways, and the urea cycle with the progression of pulmonary fibrosis.[13] However, they did not identify a clear panel of metabolites that would allow to determine in advance which patients would be more prompt to develop pulmonary fibrosis.

Encouraged by this context, in this work we aimed to identify metabolomic biomarkers in acute COVID-19 patients after discharge by NMR spectroscopy, able to predict long term fibrotic sequelae, as would be assessed by CT routine studies, one year later after discharge. The early identification of patients prone to fibrotic sequelae after acute COVID-19 would be of a valuable information for the management of these patients in the follow-up.

3. Material and methods

3.1 Study Design

This is a prospective observational study in which an NMR metabolomic study was performed on serum samples from 109 COVID-19 pneumonia patients taken 2 months after hospital discharge (between March and June 2020). In order to find potential biomarkers of pulmonary fibrosis, the metabolite composition was associated with the presence of CT fibrotic lesions after one year (Figure 1)*.*
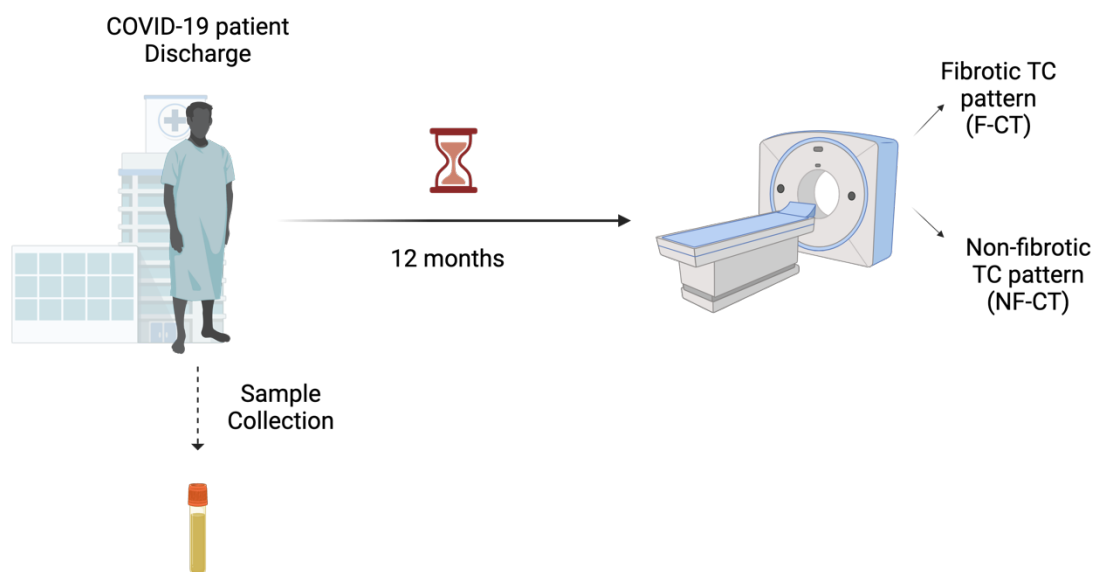
Figure 1. Schematic representation of sample collection and diagnosis of fibrosis after 12 months. Serum samples were collected form COVID-19 patients two months after being discharged. 12 months after discharge, CT scan was performed and the presence pulmonary fibrosis was determined. Discharged patients were then classified in fibrotic CT patients (F-CT) and non-fibrotic CT patients (NF-CT) for the identification of potential biomarkers of future fibrosis development in the serum samples.

3.2 Patient selection

This study is based on a subcohort of the COVID-FIBROTIC[14,15] study (Study of the Appearance of Lung Fibrotic Changes Associated with SARS-CoV-2 Infection; www.clinicaltrials.gov identifier: NCT04409275; June 1, 2020). The COVID-FIBROTIC study is a prospective, observational, multicenter study of patients admitted for bilateral COVID-19 pneumonia in 12 hospitals in Spain. All patients aged over 18 with a life expectancy > 1 year discharged from respiratory services between May 1 and July 31 (2020) were invited to participate. Diagnosis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was based on centers of disease control (CDC) criteria, with

all patients confirmed by reverse transcription polymerase chain reaction (PCR). Diagnosis of COVID-19 pneumonia was established in accordance with World Health Organization (WHO).[16] All patients were informed of the study procedures and gave their consent The study (version 3·0; May 12, 2020) was approved by the Ethics Research Committee from Hospital Clinico, INCLIVA (Valencia, Spain) (2020/149) and by local committees wherever needed. Patients with unilateral COVID-19 pneumonia, a previous diagnosis of interstitial lung disease (ILD) or chronic obstructive pulmonary disease (COPD) and/or difficulties in attending the centers[1] for follow-up visits were excluded. Our substudy cohort consisted of patients from five of these centers[1]*. Center selection was based on the availability and logistics required to send biobank samples.

3.3 Hospital Procedures

All participants were scheduled for the first visit 2 months after hospital discharge (between May 1 and July 31; 2020). Baseline data (demographic, comorbidities, clinical course) were retrieved from electronic medical records. At first visit, we performed blood sampling, pulmonary function tests (PFR) and simple chest radiography (X-ray). Computed tomography (CT) was indicated in those patients who presented respiratory functional alterations and/or persistent images in the control X-ray. After one year, all patients underwent PFR and only patients with unresolved CT at 2 months underwent repeat CT scans.

PFR were performed in the respiratory function testing laboratory in all participating centers and included plethysmography' lung volumes determination (Total Lung Capacity [TLC] and Residual Volume [VR]), spirometry (Forced Vital Capacity [FVC] and

---

[1] *Hospital Clínico de Valencia, H. General de Valencia, H. Virgen del Rocío (Sevilla), H. Virgen de la Arrixaca (Murcia), H. Arcos del Mar Menor (Murcia).

Forced Expiratory Volume in the first second [FEV1]) and diffusion capacity [DLCO] (*Masterscreen, Jaeger, Germany*). All procedures were performed according to American Thoracic Society (ATS) and European Respiratory Society (ERS) guidelines.[17–20] The functional alterations were defined as DLCO <80% and/or FVC <80%. Chest X-rays (CXRs) were performed using standardized techniques. CT scans (*SOMATOM, Siemens, Germany; AQUILION, Toshiba, Japan; OPTIMA, General Electric, USA*) were obtained with subjects in the supine position during breath holding (slices 1mm). CT images were evaluated following the Fleischner society glossary of terms.[21] Fibrotic pattern on CT was defined by the presence of at least one of the following findings: traction bronchiectasis, reticular pattern and parenchymal bands (Figure 2).[22,23]

Two peripheral blood aliquots were obtained under fasting conditions by venipuncture in all patients. The serum was obtained after blood centrifugation at 4000 rpm at 4ºC for 20 minutes and 500 μL aliquots were stored at -80ºC in a INCLIVA biobank until analysis.
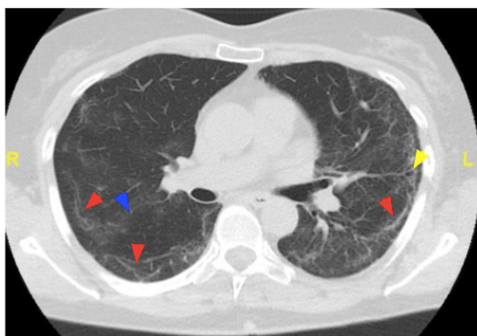
a) b)
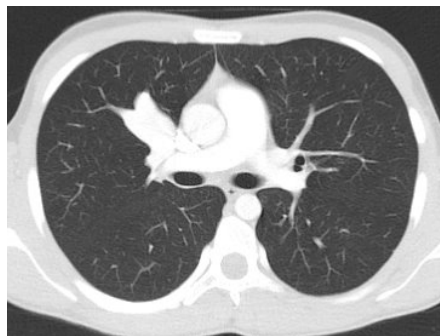
Figure 2. Lung CT images. a) CT findings of discharged COVID-19 patients after 12 months. The image shows the persistence of areas of ground glass (blue arrowheads). However, the most striking finding is the appearance of subpleural parenchymal bands, predominantly in the left hemithorax (red arrowheads), as well as some traction bronchiectasis (yellow arrowhead).
b) Normal CT.

3.4 Sample preparation

Following the protocol outlined by Beckonert *et al.*, 2007.[24] Firstly, samples were thawed. Subsequently, 400 µL of serum and 200 µL of phosphate buffer (pH 7.4) were introduced into 5 mm NMR tubes. The phosphate buffer was prepared with deuterated water (20% v/v) and the internal standard sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) at 1 mM concentration.

3.5 NMR spectra acquisition and processing

Once the samples were prepared, NMR spectra were acquired in a Bruker 600 MHz spectrometer equipped with a cryo-probe (NMR service, Principe Felipe Research Center (CIPF), Valencia. Spain). For each sample, 1D ^{1}H-NMR spectra were acquired using the Carr-Purcell-Meiboom-Gill (cpmg) pulse sequence, incorporating water signal suppression and a total spin echo of 32 ms (interpulse delay between 180º pulses was 0.001 s, and the number of loops was 16). This pulse sequence minimizes the contribution of signals from high molecular weighted molecules, such as proteins or other macromolecules, due to their short times of transverse relaxation (T2). The probe temperature was set at 300 K (27 ºC).

Following the acquisition, the spectra underwent Fourier transformation and processing using *TopSpin 4.0.0* (Bruker BioSpin Corporation). For processing the 1D spectra, a 0.5 Hz exponential line-broadening function was applied, followed by Fourier transformation. Phasing, baseline correction, and chemical shift referencing the trimethylsilyl signal of DSS at 0.0 ppm were also performed. Signals in the spectra were assigned based on literature data,[25,26] the Human Metabolome Database (HMDB)[27] and *Chenomx NMR Profiler*.[28]

After processing, significant signals in the cpmg spectra underwent deconvolution using *AMIX 4.0.2* software (Bruker BioSpin Corporation). Residual signals after water suppression in the 4.5 to 5.0 ppm range, and regions with chemical shifts below 0.5 ppm and above 8.5 ppm, were excluded from the analysis. A total of 196 signals were selected in the 1D spectra and included for deconvolution. Subsequently, a mixed Gaussian/Lorentzian variable function was applied for the deconvolution of these signals. After deconvolution, integrals were obtained for all cpmg spectra and the data were normalized to the sum of all integrals in each sample.

3.6 Multivariate statistical analysis

Multivariate statistical analysis was performed using the software *PLS_Toolbox Solo 9.2* (Eigenvector Research, Inc., Manson, WA, USA). The variables included in the model were selected using the genetic algorithms (GA) module. The initial population size selected was 256, and 100 maximum generations were set. PLS regression with a maximum of 10 latent variables and random cross-validation with 10 splits and 5 iterations were set for the variable selection. After applying GA, a total of 23 variables were selected to generate the PLS-DA model.

Partial least squares-discriminant analysis (PLS-DA) was employed to construct a predictive model using the selected variables. Cross-validation, specifically Venetian blinds, was applied to identify the optimal number of latent variables for the model. To assess the model's ability to differentiate between distinct sample sets, the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, and specificity were obtained. The model's robustness and susceptibility to overfitting were evaluated using permutation tests conducted through 100 iterations.

3.7 Univariate statistical analysis

Qualitative variables were described using frequencies and percentages while quantitative variables by means and their standard deviation. A t-test was employed for comparison in quantitative variables and Fisher's exact test for the comparison between groups of categorical variables. Normality test (Kolmogorov-Smirnov) was applied to check the normality of the metabolomic variables included in the study. Subsequently, based on the normality results, either a t-test or Mann-Whitney U test was employed to compare the means of the relative concentrations of the analyzed metabolites. Univariate statistics were conducted using GraphPad, and R (*RStudio 1.2.5001*). The ggplot2 package as used for the generation of plots.

3.8 Metabolic pathways analysis

To identify potential metabolic pathways involved in pathological processes, *Metaboanalyst*[29] was used. A concentration table was generated with data from metabolites as columns and samples as rows, one signal from each metabolite was chosen for the generation of the table. Metabolites were included in the pathway analysis using their corresponding HMDB ID. Metabolites lacking an HMDB ID, such as fatty acids or unknown metabolites, were excluded. For the topological analysis, Relative-betweenness centrality was employed, and the Homo sapiens library from *Metaboanalyst* served as the reference metabolome. The enrichment method chosen was the global test. Subsequently, pathways with a p-value < 0.05 and an impact factor > 0 were selected as representative pathways.

4. Results

4.1 Demographic and clinical characteristics of the participants

Between 1 May, and 31 July 2020, 109 serum samples were collected. Mean age was 57.68 (14.03) years and 65.13% were male. For further analysis, patients were classified in two groups according to the presence of fibrotic findings on CT one year after infection: Group 1, absence of fibrotic sequelae (73/109; 66.9%); Group 2, presence of fibrotic sequelae (36/109; 33%). There were significant between-group differences in age ($p=0.0005$) and severity of disease ($p<0.0001$) but not in comorbidities (hypertension, diabetes, cardiopathy, previous respiratory diseases) smoking status or body mass index (BMI) (Table 1).

At 12 months, 85 patients had completed the entire study protocol and up to 32.94% had diffusion impairment. There were no functional differences between-group in spirometry or diffusion. There were several losses in plethysmography volume determination, so this variable was not considered for the analysis. According to the study protocol, chest CT at 12 months was performed only in patients with unresolved CT scans at 2 months (41/109; 37.61%). The non-resolved CT was more frequent in Group 2 ($p<0.0001$). The fibrotic pattern (traction bronchiectasis and/or parenchymal bands and/or reticular pattern) was present in 33.02% of the total cohort (Table 1).

Table 1. Clinical Characteristics of enrolled patients.

	Total n = 109	Group 1 n = 73	Group 2 n = 36	p Value
Age, years	57.68 (14.03)	54.45 (14.52)	64.25 (10.38)	0.0005
Male sex	71 (65.13%)	46 (63.01%)	25 (69.44%)	0.53
BMI, kg/m^2	27.43 (4.00)	27.10 (4.15)	28.10 (3.66)	0.26
Never-smoker	74 (67.89%)	52 (71.23%)	22 (61.11%)	0.38
Comorbidities				
Pulmonary disease*	21 (19.27%)	16 (21.92%)	5 (13.89%)	0.44
Hypertension	38 (34.86%)	25 (34.25%)	13 (36.11%)	0.99
Diabetes	22 (20.18%)	15 (20.55%)	7 (19.44%)	0.99
Cardiovascular disease	10 (9.17%)	7 (9.59 %)	3 (8.33%)	0.99
Clinical severe disease	36 (33%)	14 (19.17%)	22 (61%)	<0.0001
12 months Pulmonary function**				
FVC, % pred	105.11 (15.93)	106.25 (17.72)	102.66 (11.04)	0.34
FEV1, % pred	104.68 (14.17)	103.84 (15.63)	106.43 (10.41)	0.43
D_{LCO}, % pred	86.34 (15.98)	88.24 (16.70)	82.46 (13.90)	0.12
D_{LCO} <80%, pred	28 (32.94%)	16 (27.58%)	12 (44.44%)	0.14
12 months CT findings				
CT non - resolved	41 (37.61%)	5 (6.84%)	36 (100%)	<0.0001
GGO	27 (24.77%)	5 (6.84%)	22 (61.11%)	<0.0001
Reticular pattern	14 (12.84%)	0 (0%)	14 (38.89%)	<0.0001
Traction Bronchiectasis	17 (15.59%)	0 (0%)	17 (47.22%)	<0.0001
Parenchymal bands	22 (20.18%)	0 (0%)	22 (61.11%)	<0.0001
Fibrotic pattern***	36 (33.02%)	0 (0%)	36 (100 %)	<0.0001

Data are n (%) or mean (SD). BMI = Body Mass Index. FVC = forced vital capacity. FEV1 = forced expiratory volume in one second. D_{LCO} = diffusing capacity for carbon monoxide. TLC = total lung capacity. CT= Computed Tomography. GGO= ground glass opacity
*Pulmonary disease: asthma, obstructive sleep apnea
** Pulmonary function: the values were obtained for 85 samples, distribuided in 58 for group 1 and 27 for group 2
***Fibrotic pattern: defined as the presence of traction bronchiectasis, reticular pattern and/or parenchymal bands

4.2 Metabolic profiling of serum samples

^{1}H-NMR cpmg spectra were acquired for all the sera samples. The main signals in the spectra were assigned for the later identification of potential biomarkers of fibrosis sequela in COVID-19 discharged patients. Figure 3 shows ^{1}H-NMR spectrum of one

serum sample. Spectrum has been split in two parts, the aliphatic region (Figure 3.a) and the aromatic region (Figure 3.b), for a better visualization of the signals. The assigned metabolites are shown in table 2 with details about chemical shift, as well as the multiplicity and J coupling. A total of 31 compounds were assigned, with 4 signals from fatty acids and 13 amino acids, among others.
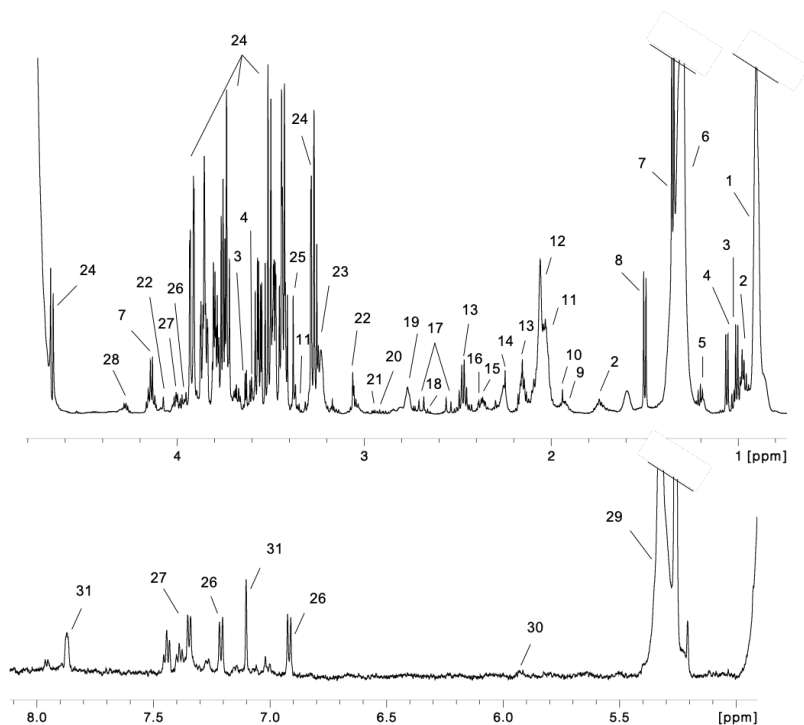
Figure 3. ^{1}H-NMR spectrum of one serum sample. The region with remaining water signal after water suppression (4.7-5.1 ppm) is not shown a) Aliphatic region of the spectrum (0.8-4.6 ppm) b) Aromatic region of the spectrum (5.2-8.5 ppm). The intensity of peaks in the aromatic region (5.0-8.7 ppm) has been scaled (10x) respect to the aliphatic region for a more appropriated display. 1. Fatty Acids (–CH$_3$), 2. Leucine, 3. Valine, 4. Isoleucine, 5. Ethanol, 6. Fatty Acids (–CH$_2$–), 7. Lactate, 8. Alanine, 9.4-Aminobutyrate, 10. Acetate, 11. Proline, 12. N-acetyled compounds, 13. Glutamine, 14. Acetone, 15. Glutamate, 16. Pyruvate, 17. Citrate, 18. Methionine, 19. Fatty Acids (=CH-CH$_2$-CH=), 20. Trimethylamine, 21. Dimethylglycine, 22. Creatinine, 23. Carnitine, 24. Glucose, 25. Methanol, 26. Tyrosine, 27. Phenylalanine, 28. Threonine, 29. Fatty Acids (-CH=CH-), 30. Urea, 31. Histidine.

Table 2. Chemical shift, multiplicity and J coupling of the signals from metabolites identified in the sera samples.

Number	Metabolite	Chemical Shift (ppm) and J coupling (Hz) *
1	Fatty Acids ($-CH_3$)	0.8-0.9
2	Leucine	0.97 (d), 0.98 (d), 1.72 (m)
3	Valine	1.01 (d, J = 7.2), 1.06 (d, J = 7.2)
4	Isoleucine	0. 93 (t, J = 7.0), 1 (d, J = 6.5), 1.31 (m)
5	Ethanol	1.16 (t, J = 7.08), 3.67 (q, J = 7.07)
6	Fatty Acids ($-CH_2$)	1.21-1.31
7	Lactate	1.35 (d, J = 7.0), 4.14 (c, J = 7.0)
8	Alanine	1.47 (d, J = 7.2), 3,77 (q, J = 7.2)
9	4-Aminobutyrate	1.90 (m), 2.28 (t, J = 7.36), 3.0 (t, J = 7.58)
10	Acetate	1.91 (s)
11	Proline	1.98 (m), 2.2(m), 3.4 (ddd, J = 12.6,9.4,3.2), 3.4 (ddd, J = 9.4,4.3,1.0), 3.74 (t, J = 9.3)
12	N-acetyled compounds	2.0-2.1
13	Glutamine	2.14 (m), 2.42 (m), 3.76 (t, J = 6.2)
14	Acetone	2.22 (s)
15	Glutamate	2.11 (td, J = 6.8, 6.2), 2.15 (dt, J = 15.4, 6.8), 3.75 (t, J = 6.2)
16	Pyruvate	2.36 (s)
17	Citrate	2.52 (d, J = 15.4), 2.66 (d, J = 15.4)
18	Methionine	2.11 (dtd, J = 14.4, 6.6, 6.2), 2.12 (s), 2.19 (dtd, J = 14.4, 6.6, 6.2), 2.64 (t, J = 6.6), 3.85 (t, J = 6.2)
19	Fatty Acids (=CH-CH_2-CH=)	2.76
20	Trimethylamine	2.88 (s)
21	Dimethylglycine	2.89 (s), 3.71 (s)
22	Creatinine	3.03 (s), 4.05 (s)
23	Carnitine	2.13 (s), 2.48 (dd, J = ND), 2.61 (dd, J = ND), 3.18 (s), 3.61 (d, J = ND), 3.82 (dd, J = ND), 5.57 (q)
24	Glucose	3.74 (d, J = 5.4), 3.81 (dt, J = 8.4,5.4), 5.22 (d, J = 1.6)
25	Methanol	3.38 (s)
26	Tyrosine	6.91 (m), 7.21 (m)
27	Phenylalanine	7.35 (m), 7.39 (m), 7.44 (m)
28	Threonine	1.31 (d, J = 6.58), 3.57 (d, J = 4.86), 4.24 (m)
29	Fatty Acids (-CH=CH-)	5.26
30	Urea	5.82 (s)
31	Histidine	3.16 (dd, J = 15.55, 7.7), 3.23 (dd, J = 16.10, 4.9), 3.98 (dd, J =7.73, 4.98), 7.09 (d, 0.58), 7.90 (d, J = 1.13)

* the multiplicity of the signals is indicated as s: singlet; d: doublet, t: triplet, m: multiplet, dd: doublet of doublets

4.3 Statistical analysis of serum samples

Aiming to identify potential biomarkers, partial least square discriminant analysis (PLS-DA) were performed with the data obtained from the integration of the spectrum signals. The generated models were able to classify the serum samples according to the future development of pulmonary fibrotic sequelae with an 80.82% of sensitivity and 72.22% of specificity (Figure 4.a). The AUC value for the cross validation (CV) was 0.83 (Figure 4.b). Five latent variables were used for model generation. The results were significant ($p < 0.05$) for all test permutations.

a)   b)
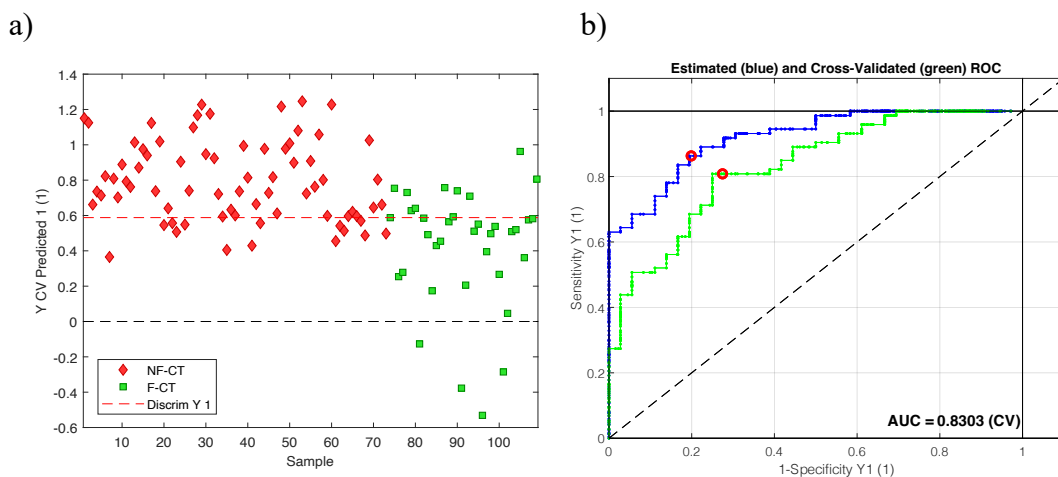
Figure 4. PLS-DA model and ROC curve with AUC values. a) samples distribution based on latent variable (LV) 1 and LV 5. Red diamonds show samples from NF-CT and green squares show samples from F-CT patients. b) ROC curve for the CV of the model generated.

The metabolites involved in the classification models were fatty acids ($-CH_3$ moiety), valine, proline, glutamine, glutamate, citrate, glucose, n-acetyled compound, fatty acids ($=CH-CH_2-CH=$ moiety) and phenylalanine (Table 3). Three of them showed significant differences in relative concentration, with a concentration increase of fatty acids ($=CH-CH_2-CH=$ moiety) and a decrease concentration in glucose and valine (Figure 5).

Table 3. Metabolites participating in the discriminative model

Metabolites	VIPS*	p-value	Trend
=CH-CH$_2$-CH=	**1.81**	**0.004**	**↑****
Phenylalanine	1.11	0.245	↓
Valine	**1.02**	**0.016**	↓
-CH$_3$	0.95	0.090	↑
Glutamine	0.85	0.668	↓
N-acetyled compound	0.83	0.064	↑
Glutamate	0.79	0.854	↓
Proline	0.74	0.298	↑
Citrate	0.71	0.864	↓
Glucose	**0.64**	**0.000**	↓

*VIPS = Variable importance in projection score, represents the importance of the metabolite in the discriminative capacity of the model.

** The arrow direction indicates the change in concentration of metabolites in patients without fibrotic sequelae *vs.* patients with fibrotic sequelae. Metabolites with significant differences are shown in bold.
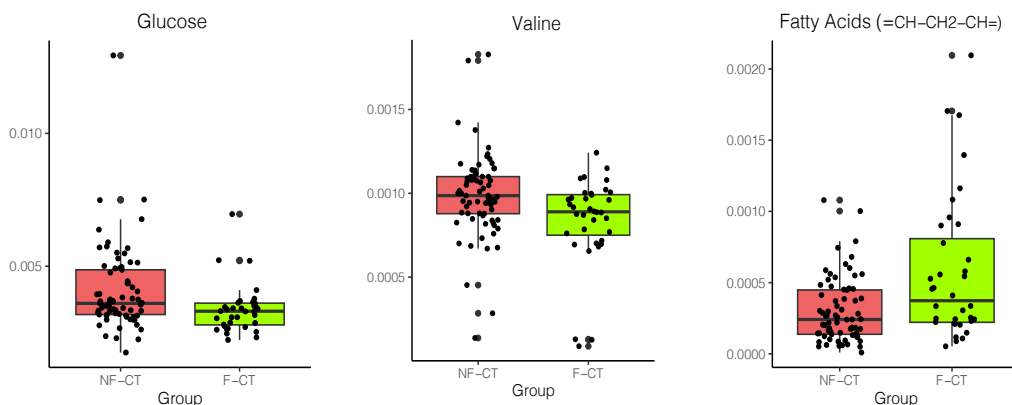
Figure 5. Comparison of the relative concentration of the metabolites in the model showing significant differences between NF-CT and F-CT groups. Boxplots representing the mean concentration of metabolites in serum samples are shown. The horizontal black line inside each box represents the median of each group for a determined feature. The relative concentration of each sample is shown as a black dot.

4.4 Metabolic Pathways Analysis

To identify possible metabolic routes, all the metabolites participating in the discriminatory model were analyzed in *Metaboanalyst*. Starch and sucrose metabolism and galactose metabolism pathways were identified to be significantly altered in patients who showed fibrotic sequelae 12 months after hospital discharge (Figure 6). Glucose is the identified metabolite that participates in both metabolic pathways. The differences in the relative concentration of glucose in both groups has already been highlighted (Figure 5).
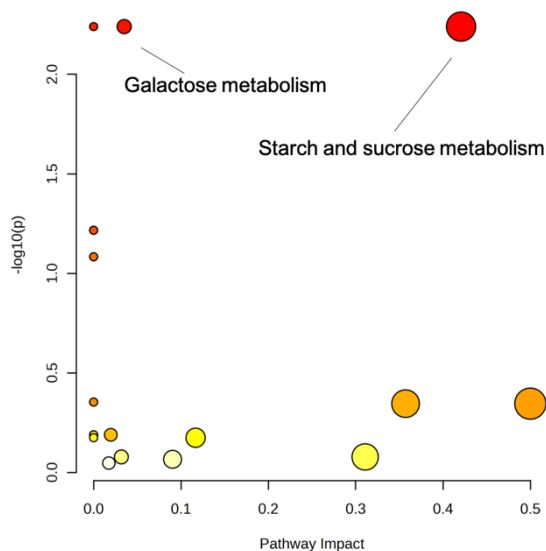
Figure 5. Metabolic pathways altered in COVID-19 patients who developed fibrotic sequelae. Only significant pathways are labeled (p-value < 0.05, impact > 0). Color and size of the circles indicate the p-value and impact index, respectively: $-\log_{10}(p)$ is represented from higher values (red) to lower values (yellow), and the pathway impact is reflected in the size of the circles from smaller circles (lower impact) to bigger circles (higher impact).

5. Discussion

COVID-19 has affected millions of people worldwide up to date and the long-lasting sequelae are a clinical concern still under study. In this study, we have generated a model capable of predicting which patients will develop fibrotic sequelae one year after hospital discharge with 80.82% sensitivity, 70.22% specificity, and 0.83 AUC value.

We found DLCO impairment in 32.94% of the whole cohort. These results are in agreement with other previously published studies. Wu X *et al*. found that, after one year, despite the improvement in mean DLCO values over the following months, up to 33% of individuals still had an impaired DLCO, regardless of the severity of the disease.[30]

We identified the fibrotic pattern in 33.02% of total patients. This percentage is consistent with other studies which found CT abnormalities on scan 1 year after in 24% of the patients.[31] Another study reported radiological fibrotic sequelae one year after discharge in 25% of patients.[32]

Valine, proline, glutamine, glutamate, citrate, glucose, n-acetyl compounds, phenylalanine and general fatty acids (-CH$_3$) and unsaturated fatty acids (=CH-CH$_2$-CH=) were the metabolites participating in the model. Among these metabolites, univariate statistics showed a significant increase in the amount of unsaturated fatty acids (=CH-CH$_2$-CH=) chain, and a decrease in the concentration of glucose and valine.

Starch and sucrose metabolism, as well as galactose metabolism, are highlighted as statistically significant metabolic pathways affected in patients who developed fibrotic sequelae, being glucose the identified metabolite that participates in both pathways. These pathways involve the breakdown of polysaccharides into glucose. To our knowledge, this is the first time these pathways have been linked to the development of pulmonary fibrosis. The association of these pathways in pulmonary fibrosis could be related to a higher demand for glucose in patients who develop fibrotic sequelae. The observed decrease in glucose concentration could indicate an increase in glycolysis,

which has been related to the development of pulmonary fibrosis. Fibrosis occurs due to the recurrent injury of epithelial cells and the activation of fibroblasts in the lungs, that leads to increased glycolysis.[33] Lung myofibroblasts are reprogrammed into a secretory phenotype, producing excessive extracellular matrix, leading to the characteristic tissue stiffening and scarring observed in fibrotic lungs.[34]

An impairment in lipid metabolism has previously been related to the development of IPF. We observed an increase in fatty acids (-CH_3) chain concentration, and a significative increase in the concentration of unsaturated fatty acids (=CH-CH_2-CH=) chain. Similar to changes in glucose metabolism, the reprogramming of cells in the development of a pro-fibrotic phenotype, involves changes in the lipid metabolism, including alteration in the synthesis, storage and oxidation of lipids.[35] Fatty acids have important roles in several cellular processes are associated with the pathogenesis and progression of pulmonary fibrosis.[35] Furthermore, previous research on pulmonary fibrosis has shown an increase in total fatty acids serum levels in IPF patients,[36] as well as the accumulation of lipid droplets in fibrotic lung tissue.[37]

We observe a decrease in the relative concentration of valine in patient with fibrotic sequelae, which contrast with previous studies that describe an increase in the exhaled concentration of valine, as well as other amino acids, such as proline, alanine, and leucine/isoleucine in IPF patients.[38] Branched chain amino acids, including valine, are key substrates for energy metabolism and protein synthesis, and play significant roles in signaling and energy pathways, being, for example, the main activators of the mammalian target of rapamycin (mTOR), a pro-growth enzyme that is hyperactivated in fibroblast in IPF.[39] Further research would be needed to determine the specific mechanisms by which valine contributes to the development of fibrotic sequelae in COVID-19 patients.

Furthermore, glutamine is as well involved in the discriminative model for the identification of patients with fibrotic sequelae. Glutamine metabolism has been

proved to be important in the regeneration of alveolar cells in IPF and in the alveoli repair after epithelial injury, crucial pathway to stop fibrosis progression. Glutamine is involved in the TGF-β-induced myofibroblast activation.[40] The impairment in glutamine metabolism might as well be involved in the development of fibrosis sequelae in COVID-19 patients. Further studies to verify the misfunction of enzymes involved in glutamine metabolism should be performed.[41] Glutamine metabolism is also required for the synthesis of proline, which also participates in the discriminative model. In fact, flux analysis has shown that proline is the major cellular destination for glutamine metabolism in myofibroblasts. Proline is the second most abundant amino acid in collagen, one of the main proteins in the extracellular matrix.[33]

Altogether, our results show that it is possible to identify changes in the metabolic signature of COVID-19 patients that will develop radiological fibrotic pattern. This suggests the potential of NMR metabolomics for predicting the onset of fibrotic sequelae. However, further research would need to be performed to elucidate the cellular mechanisms that lead to the development of this sequelae and are responsible for the differences observed in the metabolome of COVID-19 patients with and without fibrotic signs one year after hospital discharge.

6. Bibliography

1. Mulet, A. *et al.* Biomarkers of Fibrosis in Patients with COVID-19 One Year After Hospital Discharge: A Prospective Cohort Study. *Am. J. Respir. Cell Mol. Biol.* **69**, 321–327 (2023).

2. Weaver, A. K., Head, J. R., Gould, C. F., Carlton, E. J. & Remais, J. V. Environmental Factors Influencing COVID-19 Incidence and Severity. *Annu. Rev. Public Health* **43**, 271–291 (2022).

3. Chatterjee, S. *et al.* Association of COVID-19 with Comorbidities: An Update. *ACS Pharmacol. Transl. Sci.* **6**, 334–354 (2023).

4. Lee, J. H., Yim, J. J. & Park, J. Pulmonary function and chest computed tomography abnormalities 6–12 months after recovery from COVID-19: a systematic review and meta-analysis. *Respir. Res.* **23**, 233 (2022).

5. Yang, J., Pan, X., Wang, L. & Yu, G. Alveolar cells under mechanical stressed niche: critical contributors to pulmonary fibrosis. *Mol. Med.* **26**, 95 (2020).

6. Myall, K. J. *et al.* Persistent post–COVID-19 interstitial lung disease: An observational study of corticosteroid treatment. *Ann. Am. Thorac. Soc.* **18**, 799–806 (2021).

7. Botello-Marabotto, M. *et al.* Non-invasive biomarkers for mild cognitive impairment and Alzheimer's disease. *Neurobiol. Dis.* **187**, 106312 (2023).

8. Loras, A. *et al.* Integrative metabolomic and transcriptomic analysis for the study of bladder cancer. *Cancers (Basel).* **11**, 686 (2019).

9. Botello-Marabotto, M., Martínez-Bisbal, M. C., Pinazo-Durán, M. D. & Martínez-Máñez, R. Tear metabolomics for the diagnosis of primary open-angle glaucoma. *Talanta* **273**, 125826 (2024).

10. Araújo, R. *et al.* Infection Biomarkers Based on Metabolomics. *Metabolites* **12**, 92 (2022).

11. Santos-Rebouças, C. B., Cordovil Cotrin, J. & dos Santos Junior, G. C. Exploring the interplay between metabolomics and genetics in Parkinson's disease: Insights from ongoing research and future avenues. *Mech. Ageing Dev.* **216**, 111875 (2023).

12. Bruzzone, C., Conde, R., Embade, N., Mato, J. M. & Millet, O. Metabolomics as a powerful tool for diagnostic, pronostic and drug intervention analysis in COVID-19. *Front. Mol. Biosci.* **10**, 1111482 (2023).

13. Yang, J. *et al.* Proteomics and metabonomics analyses of Covid-19 complications in patients with pulmonary fibrosis. *Sci. Rep.* **11**, 14601 (2021).

14. Tarraso, J. *et al.* Lung function and radiological findings 1 year after COVID-19: a prospective follow-up. *Respir. Res.* **23**, 242 (2022).

15. Safont, B. *et al.* Lung Function, Radiological Findings and Biomarkers of

Fibrogenesis in a Cohort of COVID-19 Patients Six Months After Hospital Discharge. *Arch. Bronconeumol.* **58**, 142–149 (2022).

16. Clinical Spectrum | COVID-19 Treatment Guidelines. Available at: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/. (Accessed: 20th May 2024).

17. Graham, B. L. *et al.* 2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung. *Eur. Respir. J.* **49**, 1600016 (2017).

18. Graham, B. L. *et al.* Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *Am. J. Respir. Crit. Care Med.* **200**, E70–E88 (2019).

19. Quanjer, P. H. *et al.* Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur. Respir. J.* **40**, 1324–1343 (2012).

20. Stanojevic, S. *et al.* Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. *Eur. Respir. J.* **50**, 1700010 (2017).

21. Hansell, D. M. *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* **246**, 697–722 (2008).

22. Antonio, G. E. *et al.* Thin-section CT in patients with severe acute respiratory syndrome following hospital discharge: preliminary experience. *Radiology* **228**, 810–815 (2003).

23. Westcott, J. L. & Cole, S. R. Traction bronchiectasis in end-stage pulmonary fibrosis. *Radiology* **161**, 665–669 (1986).

24. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc. 2007 211* **2**, 2692–2703 (2007).

25. Govindaraju, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **13**, 129–153 (2000).

26. Martínez-Bisbal, M. C. *et al.* 1H and 13C HR-MAS spectroscopy of intact biopsy

samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed.* **17**, 191–205 (2004).

27. Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

28. Chenomx Inc | Metabolite Discovery and Measurement. Available at: https://www.chenomx.com/. (Accessed: 20th March 2024)

29. Pang, Z. *et al.* MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* **49**, 388–396 (2021).

30. Wu, X. *et al.* 3-month, 6-month, 9-month, and 12-month respiratory outcomes in patients following COVID-19-related hospitalisation: a prospective study. *Lancet Respir. Med.* **9**, 747–754 (2021).

31. Han, X. *et al.* Six-month follow-up chest CT findings after severe COVID-19 pneumonia. *Radiology* **299**, E177–E186 (2021).

32. Pan F, Yang L, Liang B, *et al*. Chest CT Patterns from Diagnosis to 1 Year of Follow-up in COVID-19*. Radiology*. **302**, 709-719 (2021).

33. Hamanaka, R. B., Okhan, G., Mutlu, M. & Mutlu, G. M. Metabolic requirements of pulmonary fibrosis: role of fibroblast metabolism. *FEBS J.* **288**, 6331–6352 (2021).

34. Ortiz-zapater, E., Signes-costa, J., Montero, P. & Roger, I. Lung Fibrosis and Fibrosis in the Lungs: Is It All about Myofibroblasts? *Biomedicines* **10**, 1423 (2022).

35. Geng, J., Liu, Y., Dai, H. & Wang, C. Fatty Acid Metabolism and Idiopathic Pulmonary Fibrosis. *Front. Physiol.* **12**, 794629 (2022).

36. Iannello, S. *et al.* Low fasting serum triglyceride and high free fatty acid levels in pulmonary fibrosis: a previously unreported finding. *MedGenMed* **4**, 5 (2002).

37. Shi, X. *et al.* The novel molecular mechanism of pulmonary fibrosis: insight into lipid metabolism from reanalysis of single-cell RNA-seq databases. *Lipids in Health and Disease* **23**, 98 (2024).

38. Gaugg, M. T. *et al.* Molecular breath analysis supports altered amino acid metabolism in idiopathic pulmonary fibrosis. *Respirology* **24**, 437–444 (2019).

39. Summer, R. *et al.* Circulating metabolic profile in idiopathic pulmonary fibrosis: data from the IPF-PRO Registry. *Respir. Res.* **25**, 58 (2024).

40. Choudhury, M. *et al.* Targeting Pulmonary Fibrosis by SLC1A5-Dependent Glutamine Transport Blockade. *Am. J. Respir. Cell Mol. Biol.* **69**, 441–455 (2023).

41. Wang, S. *et al.* Glutamine Metabolism Is Required for Alveolar Regeneration during Lung Injury. *Biomolecules* **12**, 728 (2022).

General Discussion

The research performed in this thesis has aimed to explore the potential of NMR-based metabolomics as a tool for the identification of early non-invasive biomarkers of disease. NMR is used in this thesis for metabolomic research because it is a highly reproducible and non-destructive technique. To date, several studies have validated the usefulness of NMR as a diagnosis support.

Metabolomics is the omic science closest to phenotype, as it focuses in the end products of all the metabolic processes occurring in a living organism, showing the interplay between genetic and environmental factors.[1] There is a need to identify early biomarkers of diseases in the context of personalized medicine for a better clusterization of patients and for the identification of more straightforward treatment strategies. Furthermore, the importance and impact of cellular metabolism in disease development and in the different outcomes of patients with the same conditions is gaining relevance.[2] This work has studied four diseases to identify new biomarkers of diagnosis and progression through NMR-based metabolomics.

In the first chapter, following the proposed objective, serum samples from patients with AD and MCI were analyzed by NMR-based metabolomics. The PLSDA models generated were able to discriminate with high sensitivity and specificity AD patients from controls (93.75% and 94.75%) and AD from MCI (100% and 82.35%). When the metabolites in the model were analyzed, we found significant differences for most of them, highlighting the impact that the development of AD has on cellular metabolism. A third model for the classification between MCI and HC was performed. However, in this case, the model was not discriminative (67% sensitivity and 50% specificity). This shows that patients with MCI are closer in phenotype to HC than to AD patients. The analysis of the evolution of MCI patients to AD in a frame time of one to three years resulted in the identification of five metabolites (lysine, pyruvate, phenylalanine, mixed lipoproteins and choline) as potential biomarkers of the progression of MCI patients to AD. To date, there is no precise biomarker able to predict the evolution of MCI patients

to AD, as MCI patients can remain stable, as previously mentioned in the general introduction. Furthermore, the metabolic pathways involved in the progression of MCI to AD were studied. Overall, in this chapter, we validated already known information about the impact of AD on cellular metabolism while providing new non-invasive metabolites for the follow-up of MCI patients by NMR-based metabolomics. However, further studies would need to be conducted to validate the aforementioned metabolites as biomarkers of MCI to AD transition. It would be necessary to increase the size of the study, as when dividing the MCI group in our study, the size of the subgroups was very limited. Moreover, it would be interesting to carry out a longitudinal study to corroborate that in MCI patients, the value of the metabolites increases or decreases, as described here, when they progress to more advanced stages of dementia.

In Chapter 2, MCI patients of the AD type were further studied. We observed that the combination of NMR-based metabolomics with peroxidation lipids obtained by UPLC-MS/MS improves the generation of classification models for the discrimination between MCI and HC. As mentioned in the introduction, comparing metabolomic results among different studies is difficult. Furthermore, the study presented in Chapter 1 was made with serum samples, and for Chapter 2, the sample used was plasma. Still, in both projects, we can observe that the metabolism of amino acids, such as valine and isoleucine, is involved in the discriminative classification of MCI patients from HC. Furthermore, from this second study, we can corroborate that oxidative stress is a crucial player in the development of cognitive impairment, as the inclusion of isoprostanes in the discriminative model improves its classification ability.

Attending to the third objective of this thesis, in Chapter 3, biomarkers of primary open glaucoma were found in tears. Through NMR-based metabolomics, discriminative models of POAG patients with high sensitivity and specificity values (100% and 83.3%) were built. A decrease in the relative concentration of phenylalanine, phenylacetate,

leucine, n-acetyled compounds, formic acid, and uridine, and an increase in the concentration of taurine, glycine, urea, glucose, and UFA in tears of POAG patients was observed, being these metabolites postulated as biomarkers for the diagnosis of POAG. However, this work is a preliminary study, as the number of samples is limited (n=30), and increasing the study size would be crucial for validating these metabolites as biomarkers of POAG.

Chapter 4 aimed to identify potential biomarkers of plaque vulnerability that could be used as a supportive tool for intervention decisions. Serum (70) and tissue (atheroma plaque, 38) samples from patients with recently symptomatic and asymptomatic carotid stenosis were analyzed by NMR-based metabolomics. The discrimination model obtained for atheroma plaque samples could classify stable (asymptomatic patients) from unstable plaques (symptomatic patients) with high sensitivity and specificity (100% and 96.4%). In addition, attending to the univariate statistics and the biological contest, glutamate is proposed as a relevant biomarker of plaque vulnerability. Furthermore, circulant biomarkers were identified in serum by generating another PLS-DA model, able to classify unstable from stable plaques with 88.37% sensitivity and 77.78% specificity. Threonine, histamine, and unsaturated fatty acids (-CH=CH-) are described as potential biomarkers of plaque vulnerability. In this Chapter, we have discovered potential new biomarkers of plaque instability in serum and tissue samples. However, as discussed in Chapter 4, it is not possible, from the study performed, to determine whether the biomarkers described here are found as a cause or as a consequence of plaque rupture. To validate the use of these biomarkers for rupture prediction, a longitudinal study following patients previously to a cardiovascular event and after it should be performed.

In the last Chapter, the identification of biomarkers of post-COVID pulmonary fibrosis is assessed. Serum samples from 109 COVID-19 patients obtained after hospital discharge were analyzed by NMR-based metabolomics. The diagnosis one year after

discharge was used for the classification model. The obtained model was able to classify with 85.56% sensitivity and 83.33 % specificity the patients that showed pulmonary fibrosis one year after the acquisition from those who didn't show pulmonary sequelae. (=CH-CH2-CH) chain, found in higher concentrations in patients who later would develop fibrosis, and glucose and valine, found in lower concentrations, are proposed as potential biomarkers of fibrosis development. It is remarkable that in this project, the metabolic changes were already observed one year before the evaluation of pulmonary fibrosis by CT, showing the importance of cellular metabolism in all the processes occurring in a living system that can lead to the development of diseases. Nevertheless, although these results are promising, further studies would need to be performed in order to validate the use of glucose, valine and fatty acids concentration as biomarkers of pulmonary fibrosis development.

In summary, the research presented in this thesis highlights the potential of NMR-based metabolomics as a valuable tool to support diagnosis of disease, and for the identification of new biomarkers. However, it is important to note that while the findings presented here are promising, they are still preliminary, as in most cases, the number of samples studied is limited, and longitudinal studies would be required for validation. Still, the use of metabolomics allows us to get an insight into the effect that the development of diseases has on metabolism, as well as how differences in the metabolism can result in different illness outcomes.

References

1. Qiu, S. *et al.* Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 1–37 (2023).
2. Deberardinis, R. J. & Thompson, C. B. Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **148**, 1132 (2012).

Conclusion

This PhD thesis has contributed to the development of new strategies for an early and non-invasive disease diagnosis by the use of NMR-based metabolomics, covering a need for the identification of new biomarkers of diseases. Specifically, the conclusions acquired from this work are:

1. Metabolites and metabolic pathways that could potentially be biomarkers for AD diagnosis and MCI progression have been identified by the NMR spectroscopy analysis of serum samples.

2. The combination of UPLC-MS/MS and ^{1}H-NMR allows the identification of metabolic differences between patients with MCI and healthy controls.

3. NMR-based metabolomics of tear samples shows promising results in the identification of potential biomarkers of POAG.

4. NMR-based metabolomics in atheroma plaque tissue and serum samples provides insights into potential biomarkers associated with plaque vulnerability and subsequent risk of rupture in carotid artery stenosis patients.

5. Metabolomic analysis of serum samples from discharged COVID-19 patients reveals potential biomarkers for predicting the development of pulmonary fibrosis sequelae.