



**PRINCIPE FELIPE**  
CENTRO DE INVESTIGACION



VNIVERSITAT  
D VALÈNCIA

Departament de Bioquímica i Biologia Molecular  
Programa de Doctorado: Bioquímica y Biomedicina

**APLICACIONES DE LA RMN A LA IDENTIFICACIÓN  
DE NUEVOS BIOMARCADORES DE UTILIDAD  
CLÍNICA EN ONCOLOGÍA**

TESIS DOCTORAL presentada por:

Leonor Puchades Carrasco

Dirigida por:

Dr. Antonio Pineda Lucena

Valencia, 2013





PRINCIPE FELIPE  
CENTRO DE INVESTIGACION

**Antonio Pineda Lucena**, Doctor en Ciencias Químicas, Investigador Jefe del Laboratorio de Bioquímica Estructural y Director del Programa de Terapias Avanzadas del Centro de Investigación Príncipe Felipe,

**INFORMA:**

Que Dña. Leonor Puchades Carrasco, Licenciada en Farmacia por la Universidad de Valencia, ha realizado bajo mi dirección el trabajo de investigación titulado: **“APLICACIONES DE LA RMN A LA IDENTIFICACIÓN DE NUEVOS BIOMARCADORES DE UTILIDAD CLÍNICA EN ONCOLOGÍA”**. Dicho trabajo ha sido revisado por mí, estando conforme con su presentación para ser juzgado por el Tribunal que en su día se designe, para optar al grado de Doctor.

Valencia, 10 de Diciembre de 2013

Dr. Antonio Pineda Lucena  
DNI: 30517484A



## **AGRADECIMIENTOS**

Me gustaría expresar mis agradecimientos a todas las personas que de alguna manera han contribuido a la realización de esta Tesis.

En primer lugar al Dr. Antonio Pineda, mi director de Tesis, por el entusiasmo que me ha transmitido por nuestro trabajo y, en especial, por el interés que ha puesto en cada detalle de este estudio y las oportunidades que me ha brindado para que todo saliera lo mejor posible.

Al Dr. Felipe Prósper por creer en la metabolómica, porque nos abrió la puerta al mundo del Mieloma Múltiple, en el que hemos podido colaborar con los Dres. Lahuerta, Lecumberri, Mateos, Martínez-López y San-Miguel. Ha sido un placer trabajar con todos ellos y poder contar con aportaciones que han enriquecido enormemente este trabajo.

Al Dr. Carlos Camps y la Dra. Eloisa Jantus, por todo lo que me enseñaron durante mi estancia, y en cada reunión que hemos mantenido. Por su disponibilidad y por el entusiasmo con el que empujan este proyecto. Muchas gracias también a la Dra. Lucas por sus consejos desde la Universidad de Valencia.

A Jacobo Martínez y Carol Abril, de la Red Valenciana de Biobancos, por su implicación en la selección de muestras para este proyecto y en la aprobación por los Comités Éticos. Muchas gracias a todos los pacientes que donaron sus muestras voluntariamente para investigación, sin ellos no hubiera sido posible este proyecto.

A todos mis compañeros del CIPF, en especial a mis compañeros del Laboratorio porque, siempre que lo he necesitado, me habéis prestado vuestra ayuda. De todos vosotros he aprendido un montón de cosas en estos años, empezando por aquellas ligaciones que nunca salían, la purificación de proteínas, la RMN, la estadística, y todos esos programas imposibles.

A Paco García, por todas sus sugerencias, por la paciencia con la que ha escuchado todos mis planteamientos y por el ánimo que me ha dado en esta última etapa.

A todos nuestros colaboradores que, aún sin participar de manera directa en este trabajo, me han mostrado su apoyo de forma constante. Gracias a los Dres. López-Guerrero, Rubio, Guillén y Nagore del IVO, al Dr. Dominguez del IVI, al Dr.

Cano y Clara Pérez del Hospital General, y a los miembros de TBomics, los Dres. Mande, Banerjee, Patel y Varela.

A los Crisálidos, que me ayudaron tanto a tener otra perspectiva distinta en ese año que seguro fue el más crítico de toda esta etapa. Me lo pasé muy bien en las clases y me ayudasteis a desconectar cada semana, empezando por la puesta a punto con Vero.

A mis amigos, que a pesar de estar tan lejos de este mundo profesionalmente, me han ayudado con todo su cariño. A Payo, Carolina, Lorena, a los London Connection, a Hernani por estar siempre dispuesto.

Para finalizar, me gustaría dar las gracias a mi familia. A mis tíos y primos, a Vicente y Dada por su apoyo. A mi madre, Mar, Ale, Tere, Max, Marina y Óscar, porque sé que los tengo siempre y para todo, desde un diseño gráfico, a una comida, un cine, o escucharme una charla. Por interesarse e ilusionarse sinceramente en todo lo que hago. Y el agradecimiento más especial para mi padre, por todo lo que aprendí de él, por todo su cariño y porque sigue siendo mi mejor ejemplo.

Gracias a todos aquellos que se han preocupado, interesado, y aportado entusiasmo durante este tiempo. Espero que todos vosotros os sintáis parte de este trabajo.

## ÍNDICE

<b>ABREVIATURAS.....</b>	<b>11</b>
<b>I. INTRODUCCIÓN.....</b>	<b>15</b>
<b>1- Cáncer.....</b>	<b>17</b>
1.1- Clasificación del cáncer.....	18
1.2- Estadificación del cáncer.....	18
<b>2- Biomarcadores en cáncer.....</b>	<b>21</b>
2.1- Tipos de biomarcadores.....	21
2.2- Utilidad de los biomarcadores.....	22
2.3- Características de un biomarcador.....	23
2.4- Biomarcadores utilizados en oncología.....	24
2.5- Impacto de las nuevas tecnologías en la identificación de biomarcadores..	25
<b>3- Metabolómica en la identificación de biomarcadores oncológicos.....</b>	<b>29</b>
3.1- Metaboloma Humano.....	30
3.2- Posibilidades de la metabolómica.....	30
3.3- Plataformas tecnológicas en metabolómica.....	32
<b>4- Metabolómica por RMN.....</b>	<b>33</b>
4.1- Principios básicos de RMN.....	33
4.2- Esquema de trabajo de un estudio de metabolómica por RMN.....	36
a) Diseño del estudio	
b) Preparación de las muestras	
c) Adquisición de los espectros	
d) Pre-procesado de los datos	
e) Asignación de los espectros	
f) Procesado de los datos	
g) Análisis estadístico multivariante	
h) Análisis de los modelos estadísticos	
i) Cuantificación de los metabolitos y análisis estadístico univariante	
j) Regresión logística	

k) Valoración de la capacidad predictiva	
l) Interpretación biológica de los resultados	
<b>5- Aplicaciones de la metabolómica por RMN en oncología.....</b>	<b>53</b>
5.1- Mieloma Múltiple.....	53
5.2- Cáncer de pulmón.....	54
<b>6- Objetivos.....</b>	<b>56</b>
<b>II. METODOLOGÍA.....</b>	<b>57</b>
<b>1- Diseño de los estudios.....</b>	<b>59</b>
1.1- Mieloma Múltiple.....	59
1.2- Cáncer de pulmón.....	59
<b>2- Recogida de muestras.....</b>	<b>60</b>
<b>3- Preparación de muestras para RMN.....</b>	<b>61</b>
<b>4- Adquisición de espectros de RMN.....</b>	<b>61</b>
<b>5- Pre-procesado de los espectros de RMN.....</b>	<b>65</b>
<b>6- Asignación de los espectros.....</b>	<b>65</b>
<b>7- Procesado de datos.....</b>	<b>65</b>
<b>8- Análisis estadístico multivariante.....</b>	<b>66</b>
<b>9- Cuantificación de los metabolitos relevantes.....</b>	<b>66</b>
<b>10- Análisis estadístico univariante.....</b>	<b>66</b>
<b>III. RESULTADOS Y DESARROLLO ARGUMENTAL.....</b>	<b>69</b>
<b>1- Mieloma Múltiple.....</b>	<b>71</b>
1.1- Análisis y asignación de los espectros.....	71
1.2- Análisis no supervisado.....	72
a) Identificación de <i>outliers</i>	
b) Identificación de otras agrupaciones	
1.3- Análisis supervisado.....	81
a) Pacientes con MM en el momento del diagnóstico vs voluntarios sanos	



b) Pacientes con MM en el momento del diagnóstico vs pacientes con MM tras la remisión completa	
c) Pacientes con MM tras la remisión completa vs voluntarios sanos	
1.4- Integración de los metabolitos más relevantes.....	86
1.5- Interpretación biológica.....	95
a) Pacientes con MM en el momento del diagnóstico vs voluntarios sanos	
b) Pacientes con MM en el momento del diagnóstico vs pacientes con MM tras la remisión completa	
c) Pacientes con MM tras la remisión completa vs voluntarios sanos	
1.6- Relevancia de los resultados.....	97
<b>2- Cáncer de pulmón.....</b>	<b>100</b>
2.1- Análisis y asignación de los espectros.....	100
2.2- Análisis no supervisado.....	102
a) Identificación de <i>outliers</i>	
b) Análisis de homogeneidad	
2.3- Análisis supervisado.....	108
2.4- Integración de los metabolitos más relevantes.....	114
2.5- Modelo de regresión logística.....	120
2.6- Valoración de la capacidad predictiva de los metabolitos identificados.....	122
2.7- Valoración de la capacidad predictiva del modelo de regresión.....	124
2.8- Interpretación biológica.....	126
<b>IV. CONCLUSIONES FINALES.....</b>	<b>133</b>
<b>V. BIBLIOGRAFÍA.....</b>	<b>137</b>
<b>VI. ANEXOS.....</b>	<b>149</b>
<b>1- Comunicaciones a congresos nacionales e internacionales.....</b>	<b>151</b>
<b>2- Publicaciones derivadas de la Tesis Doctoral.....</b>	<b>153</b>



## ABREVIATURAS

- **1D:** Espectro monodimensional
- **<sup>1</sup>H-RMN:** Resonancia magnética nuclear de protón
- **2D:** Espectro bidimensional
- **ADN:** Ácido desoxirribonucleico.
- **AFP:** Alfa-fetoproteína
- **APC:** Poliposis adenomatosa
- **AUC:** Area under the curve
- **BRCA-1:** Gen del cáncer de mama-1
- **BRCA-2:** Gen del cáncer de mama-2
- **CA125:** Antígeno cancerígeno 125
- **CA15-3:** Antígeno cancerígeno 15-3
- **CA19-9:** Antígeno cancerígeno 19-9
- **CEA:** Antígeno carcinoembrionario
- **CHGUV:** Consorcio Hospital General Universitario de Valencia.
- **CMP:** Cáncer microcítico de pulmón.
- **COSY:** 2D <sup>1</sup>H-<sup>1</sup>H correlation spectroscopy
- **CP:** Cáncer de pulmón
- **CPMG:** Carr Purcell Meiboom and Gill
- **CPNM:** Cáncer de pulmón no microcítico.
- **CTCs:** Células tumorales circulantes
- **Ctr:** Centrado
- **CV-ANOVA:** Analysis Of Variance testing of Cross-Validated predictive residuals.
- **EGFR:** Epidermal growth factor receptor. Receptor del factor de crecimiento epidérmico
- **EM:** Espectrometría de masas
- **FDG-PET:** Tomografía de emisión de positrones basada en fluorodeoxiglucosa
- **FID:** Free induction decay
- **GECP:** Grupo español de cáncer de pulmón
- **hCG:** Gonadotropina coriónica humana
- **HDT-ASCT:** High-dose therapy followed by autologous stem cell transplantation.

- **HER2:** Human epidermal growth factor receptor 2
- **HIF:** Factor inducible por hipoxia
- **HMDB:** Human metabolome database
- **HSPs:** Proteínas de choque térmico
- **HSQC:** 2D  $^1\text{H}$ ,  $^{13}\text{C}$  heteronuclear single quantum correlation
- **J:** Constante de acoplamiento
- **Ib:** Line broadening
- **LLA:** Leucemia linfocítica aguda
- **LMA:** Leucemia mieloide aguda
- **LMC:** Leucemia mieloide crónica
- **MGUS:** Gammapatía monoclonal de significación indeterminada
- **MM:** Mieloma múltiple
- **NOESY:** Nuclear Overhauser effect spectroscopy
- **OPLS-DA:** Análisis discriminante por mínimos cuadrados con corrección ortogonal
- **OR:** Odds ratio
- **Par:** Pareto
- **PBS:** Phosphate buffered saline
- **PCA:** Análisis de los componentes principales
- **PCs:** Componentes principales
- **PDC:** Complejo piruvato deshidrogenasa
- **PDK:** Piruvato deshidrogenasa quinasa
- **PLS-DA:** Análisis discriminante por mínimos cuadrados
- **PSA:** Antígeno prostático específico
- **PTE/CT:** positron emission tomography / computerized tomography
- **RC:** Remisión completa
- **RMN:** Resonancia magnética nuclear.
- **ROC:** Receiver operating curve
- **S/N:** Signal to noise ratio
- **SEOM:** Sociedad Española de Oncología Médica.
- **SUS-plot:** Shared and Unique Structures plot
- **TC:** Tomografía computarizada
- **Tg:** Tiroglobulina
- **TGF $\beta$ :** Factor de crecimiento transformante  $\beta$
- **TOCSY:** Total correlation spectroscopy

- **TSP:** Trimethylsilyl propanoic acid
- **UV:** Varianza unitaria
- **VIP:** Variable importance in the projection



## **I. INTRODUCCIÓN**

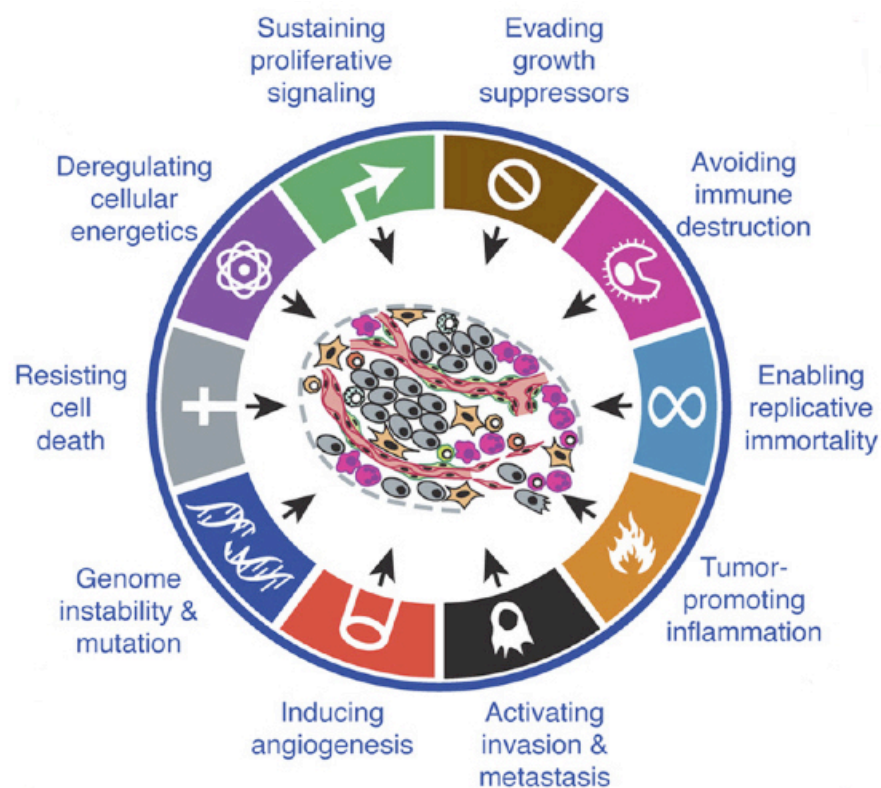




## 1- CÁNCER

El cáncer es una patología heterogénea que agrupa enfermedades muy diferentes en su etiología y en sus manifestaciones clínicas. Su principal característica es que las células de un tejido comienzan a dividirse sin control, proliferando de manera anormal y adquiriendo la capacidad de invadir otros tejidos.

El proceso de carcinogénesis incluye múltiples cambios genéticos que resultan en la transformación de una célula normal en una célula tumoral. Hanahan y Weinberg (Hanahan et al., 2000) intentaron definir y agrupar todos estos cambios en seis características esenciales que deben ser adquiridas por cualquier célula tumoral durante este proceso: i) división celular en ausencia de señales externas que estimulen el crecimiento, ii) crecimiento en presencia de señales inhibitoras de crecimiento, iii) evasión de apoptosis, iv) angiogénesis sostenida, v) potencial infinito de replicación e vi) invasión y metástasis. En una revisión posterior (Hanahan & Weinberg, 2011b), añadieron otras dos capacidades relacionadas con la interacción entre el tumor y el microambiente: la reprogramación del metabolismo energético y la evasión de la respuesta inmune (**Figura 1**).



**Figura 1:** Características adquiridas durante el proceso de transformación tumoral (Hanahan & Weinberg, 2011b)

### 1.1- Clasificación del cáncer

Se han identificado numerosos tipos distintos de cáncer, que pueden clasificarse en función del órgano donde se origina. De manera general, según el tipo celular del que derivan, se dividen en cánceres de origen epitelial (carcinomas) y cánceres de origen no epitelial (sarcomas, cánceres hematológicos y tumores neuroectodérmicos):

**Carcinomas:** Son los más abundantes. Derivan de tejidos epiteliales y representan aproximadamente el 80% de los tumores. Pueden subdividirse en adenocarcinomas (derivados de células de origen glandular) y carcinomas epidermoides o escamosos (células no glandulares).

**Sarcomas:** Son poco frecuentes. Derivan del tejido conectivo y fundamentalmente de distintos tipos celulares mesenquimales (fibroblastos, adipocitos, osteoblastos y miocitos).

**Cánceres hematológicos:** Incluyen linfomas (caracterizados por un aumento en el número de linfocitos en los ganglios linfáticos), leucemias (caracterizadas por un aumento incontrolado de leucocitos en la médula ósea) y mieloma (caracterizado por una producción excesiva de fragmentos de inmunoglobulinas por parte de los linfocitos B).

**Tumores neuroectodérmicos:** Comprenden alrededor del 1.3% de todos los cánceres diagnosticados. Derivan de células que forman parte del sistema nervioso central y periférico. En este grupo se encuentran los gliomas, glioblastomas, schwannomas y meduloblastomas.

### 1.2- Estadificación del cáncer

Los sistemas que se emplean para la estadificación del cáncer están internacionalmente aceptados y son homogéneos. En particular, el sistema TNM, aceptado por las principales organizaciones internacionales para estandarizar las publicaciones y las recomendaciones terapéuticas, está basado en la determinación de tres parámetros, que son reflejo de las vías de extensión de los tumores sólidos:

**T:** Extensión del tumor primario, atendiendo al tamaño y a la invasión de las estructuras vecinas.

**N:** Extensión tumoral a los ganglios linfáticos regionales. Sólo se incluyen en esta clasificación los ganglios linfáticos del área de drenaje del tumor primario.

La afectación de ganglios linfáticos a distancia se considera enfermedad metastásica.

**M:** Analiza la presencia o no de metástasis a distancia.

Tal y como se indica en la **Tabla 1**, adaptada de (Greene, F. L., Page D.L, Fleming I.D, Fritz A.G, Balch C.M, Haller D.G, 2002), estos parámetros presentan diversos subíndices que indican la magnitud de la afectación:

**Tabla 1:** Clasificación TNM

<b>T</b>	<b>N</b>	<b>M</b>
<b>Tx:</b> No se puede determinar la existencia o el tamaño del tumor primario.	<b>Nx:</b> No se puede determinar la existencia de adenopatías regionales metastásicas.	<b>Mx:</b> No se puede determinar la existencia de metástasis a distancia.
<b>T0:</b> No hay pruebas de la existencia de tumor primario.	<b>N0:</b> No hay demostración de adenopatías metastásicas.	<b>M0:</b> No hay pruebas de la existencia de metástasis a distancia.
<b>Tis:</b> Carcinoma <i>in situ</i> (cáncer inicial no diseminado a tejidos vecinos).	<b>N1-3:</b> Invasión ganglionar regional por orden creciente de número y tamaño.	<b>M1:</b> Existen metástasis a distancia, debiendo especificar su localización.
<b>T1-4:</b> Tumor de tamaño y capacidad invasora creciente.		

Mediante este sistema de clasificación se obtiene una descripción muy precisa de la extensión de la enfermedad. La importancia de este sistema de clasificación radica también en que proporciona una importante información pronóstica y terapéutica de los tumores, ya que estas categorías se pueden agrupar en 4 estadios clínicos, incluyendo en cada uno de ellos categorías TNM homogéneas en cuanto al pronóstico y la supervivencia:

**Estadio I:** Pequeños, localizados y habitualmente curables.

**Estadios II y III:** Localmente avanzados y/o con afectación de los ganglios linfáticos locales.

**Estadio IV:** Metastásicos y en la mayoría de los casos inoperables.

Existe gran variabilidad en el comportamiento, pronóstico y tratamiento de los distintos tipos de cáncer e incluso entre distintas etapas evolutivas del mismo tipo de cáncer. La elección del tratamiento en las enfermedades oncológicas depende de la extensión de la enfermedad, la localización del tumor primario, la clasificación anatomopatológica y la presencia o no de comorbilidades en el paciente.

En este sentido, las tecnologías “ómicas” poseen un gran potencial a la hora de profundizar en los mecanismos moleculares ligados a la etiología de las enfermedades oncológicas, siendo de gran utilidad también en el desarrollo de estrategias orientadas a la identificación de biomarcadores que permitan la detección temprana del cáncer, su diagnóstico, así como la evaluación individualizada del pronóstico y la respuesta a tratamientos.

## **2- BIOMARCADORES EN CÁNCER**

El cáncer es actualmente la principal causa de muerte en el mundo. La Sociedad Española de Oncología Médica (SEOM) estima que en el año 2015 se diagnosticarán 220.000 casos nuevos en España (“GLOBOCAN: Country Fast Stat”).

En la última década, se han realizado importantes avances en el conocimiento de las bases moleculares del cáncer, lo que ha impulsado el desarrollo de terapias dirigidas a dianas específicas. Este tipo de terapias presentan una gran selectividad, ya que actúan directamente sobre la actividad de dianas biológicas específicas (p.ej., erlotinib, inhibiendo la actividad de EGFR; trastuzumab, la de HER2; bevacizumab, la angiogénesis tumoral). Sin embargo, a pesar de los grandes avances que se han conseguido en el tratamiento de la enfermedad en los últimos años, la mayoría de los fármacos antineoplásicos no son efectivos en todos los pacientes y muchos presentan además efectos adversos graves (Spratlin et al., 2009).

Por otro lado, los métodos tradicionales de diagnóstico del cáncer, principalmente técnicas de imagen y análisis morfológico de biopsias del tumor, presentan importantes limitaciones a la hora de evaluar el pronóstico del paciente y predecir la respuesta individual al tratamiento.

Todos estos factores hacen que el cáncer sea una enfermedad en la que existe una necesidad, no cubierta con los métodos actualmente existentes, de identificar nuevos biomarcadores que faciliten la detección de la enfermedad en etapas tempranas, antes de que se disemine a otros órganos y se haga más difícil su curación. Por otro lado, la posibilidad de seleccionar subgrupos de pacientes con una mayor probabilidad de beneficiarse de un determinado tratamiento facilitaría la elección de la terapia apropiada para cada paciente, minimizando así el riesgo de reacciones adversas y traduciéndose finalmente en la obtención de terapias más seguras y eficaces.

### **2.1- Tipos de biomarcadores**

Un biomarcador es una característica biológica que puede ser medida de manera objetiva y que proporciona información sobre procesos biológicos normales o patológicos, lo que permite discriminar entre ambos procesos, caracterizar la situación patológica en la que se encuentra un individuo, evaluar la progresión de la

enfermedad o cuantificar su respuesta frente a un tratamiento farmacológico (Genoma España, 2010).

En oncología, los biomarcadores utilizados abarcan desde la medida de perfiles de expresión de genes, mutaciones en la secuencia de ADN, patrones de metilación del ADN, perfiles metabólicos y proteicos, hasta la evaluación de procesos como la apoptosis, la proliferación o la angiogénesis. Estos marcadores pueden ser producidos por el mismo tumor o por otros tejidos en respuesta a la presencia del tumor, por lo que pueden encontrarse en distintos biofluidos (suero, plasma, líquido cefalorraquídeo, orina, saliva, etc.) y/o tejidos.

Los biomarcadores pueden clasificarse en cuatro tipos principales, en función de su uso en la práctica clínica:

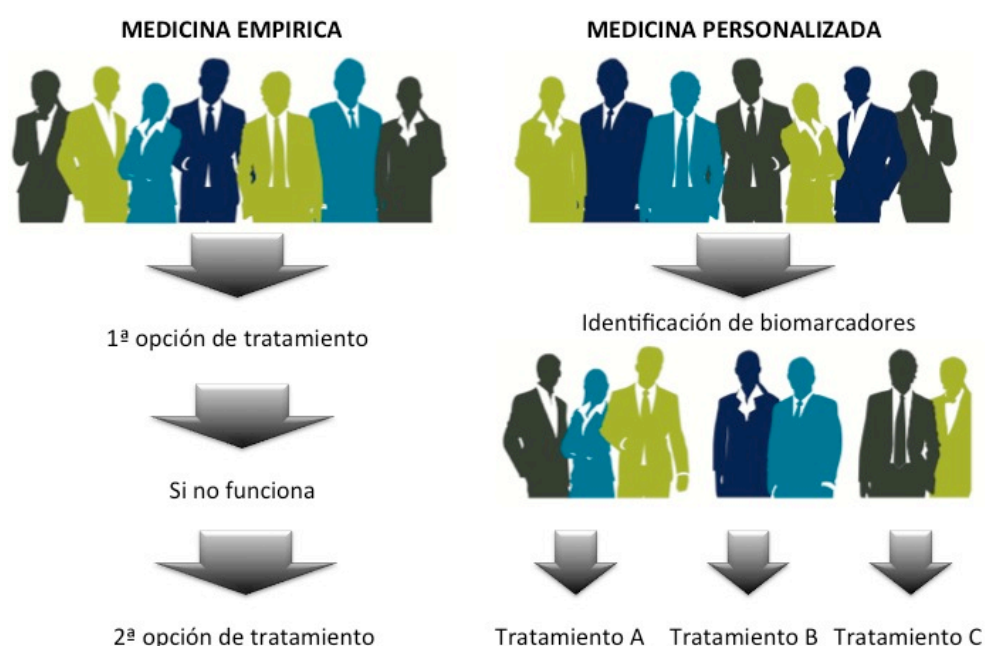
- Biomarcadores con valor diagnóstico: Permiten detectar la presencia de una determinada patología y, por tanto, diferenciar individuos sanos y enfermos. Este tipo de biomarcadores también pueden ser de gran utilidad en el seguimiento de pacientes, permitiendo identificar posibles recidivas de la enfermedad.
- Biomarcadores con capacidad predictiva: Permiten predecir la posibilidad de desarrollar una patología antes de que se produzcan los síntomas, es decir, miden la predisposición del individuo a desarrollar un patología.
- Biomarcadores con utilidad pronóstica: Proporcionan información sobre el posible curso de la patología, aportando información útil para la selección del tratamiento en función del riesgo.
- Biomarcadores de respuesta: Aportan información, antes de comenzar un tratamiento, sobre la respuesta a una intervención farmacológica o la posibilidad de desarrollar efectos adversos.

## **2.2- Utilidad de los biomarcadores**

La identificación de nuevos biomarcadores con valor diagnóstico en etapas tempranas de la enfermedad aportaría un gran valor en el caso de las enfermedades oncológicas, ya que ayudaría a aumentar significativamente el porcentaje de pacientes que se puede beneficiar de las terapias administradas, que suelen ser mucho más efectivas en etapas iniciales de la enfermedad.

Por otro lado, los biomarcadores con valor pronóstico y predictivo juegan un papel muy importante en el desarrollo de terapias personalizadas, facilitando la

estratificación de pacientes. Mediante esta estrategia se busca, de manera retrospectiva, la identificación de biomarcadores característicos de un grupo de pacientes definido que responde mejor a una determinada terapia. Una vez identificados, esta información es utilizada, de manera prospectiva, para seleccionar la terapia más adecuada para cada paciente, en función de sus características. De esta forma, al poder seleccionar los tratamientos de una manera más personalizada, se administra el tratamiento más apropiado a cada paciente desde el primer momento, evitando así la administración de terapias inefectivas que, en el caso de las enfermedades oncológicas, suelen producir además numerosos efectos adversos (**Figura 2**, adaptada de (Research Advocacy Network, 2010)).



**Figura 2:** Ventajas de la utilización de biomarcadores en la Medicina Personalizada

### 2.3- Características de un biomarcador

Las características que debe reunir un biomarcador para poder ser utilizado en la práctica clínica son fundamentalmente las siguientes:

- Poder ser detectado en un biofluido y/o tejido periférico fácilmente accesible de manera no invasiva o mínimamente invasiva.
- Presentar niveles significativamente diferentes en individuos sanos y pacientes.
- Ser fácilmente cuantificable en un ensayo fiable y estable.
- Estar relacionado de forma específica con el efecto fisiológico estudiado, preferiblemente de forma cuantitativa.

- Tener una alta sensibilidad y especificidad.

Por otro lado, en la **Tabla 2** (adaptada de (Elmer et al., 2011)) se indican las propiedades deseables de cada tipo de biomarcador, en función de su aplicación en oncología.

**Tabla 2:** Características deseables en un biomarcador oncológico.

	No-invasividad	Bajo coste	Fácil determinación	Precisión	Capacidad discriminación
Cribado	+++	+++	+++	+++	+++
Predisposición	+++	+++	+++	+++	+++
Detección temprana	++	++	++	+++	+++
Respuesta a tratamiento	+++	++	++	+++	+++
Diana terapéutica	NA	+	NA	+++	NA

NA: no aplica

#### 2.4- Biomarcadores utilizados en oncología

A pesar de que en los últimos años se ha avanzado enormemente en el conocimiento de la biología del cáncer y en la identificación de biomarcadores, su uso en oncología está limitado a un reducido número de casos y tumores, ya que su inclusión en la práctica clínica rutinaria debe ir necesariamente ligado a un incremento de la supervivencia de los pacientes y/o a una mejora de la calidad de vida. En la **Tabla 3** se recogen algunos de los biomarcadores tumorales (Bhatt et al., 2010) actualmente utilizados en la práctica clínica, los cuales incluyen hormonas, enzimas, antígenos, glicoproteínas, mutaciones genéticas, etc.

**Tabla 3:** Biomarcadores oncológicos de valor diagnóstico, pronóstico y/o terapéutico.

Biomarcador	Tumor	Aplicación	Muestra
<i>Biomarcadores basados en la presencia de antígenos</i>			
PSA	Cáncer de próstata	D/P	Suero
AFP	Carcinoma hepatocelular	D/P	Suero
CA125	Cáncer de ovario, trompas de Falopio	D/P	Suero
CA15-3	Cáncer de mama	D/P	Suero, nódulos linfáticos, médula ósea
CA19-9	Cáncer de páncreas Cáncer de vejiga	D/P	Suero Orina



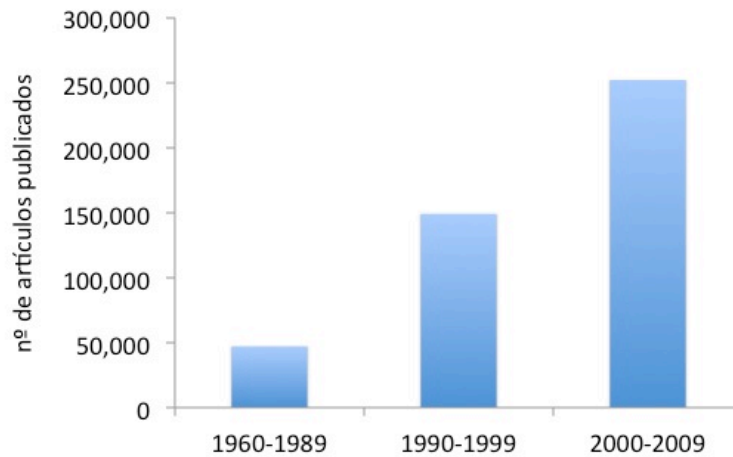
<b>Biomarcador</b>	<b>Tumor</b>	<b>Aplicación</b>	<b>Muestra</b>
BRCA-1, BRCA-2	Cáncer de mama	D/P	Tumor
CEA	Cáncer colorrectal	D/P	Suero
hCG	Tumor de células germinales	D/P	Suero
Tg	Cáncer papilar y folicular tiroideo	D/P	Suero
HSPs: Hsp70, Hsp27	Cáncer de próstata, gástrico, osteosarcomas, carcinoma de útero, cervical y vejiga	D/P	Suero
TGFβ	Tumores malignos	D/P	Suero
<i>Biomarcadores metabólicos</i>			
Glucosa	En general, todos los cánceres	D/P/T	Imagen
<i>Biomarcadores genéticos</i>			
Traslocaciones genéticas y otros productos de fusión génica	LMA, LLA, LMC, SMD y linfoma de Burkitt	D	Médula ósea y sangre periférica
Gen APC	Adenocarcinoma, carcinoma células escamosas del estómago, cáncer de páncreas, tiroides y ovario	D/P	Sangre, tumor
<i>Biomarcadores celulares</i>			
CTCs	Cáncer de mama metastásico, etc.	D/P	Sangre
CSCs	LMA, melanoma, tumor cerebral, cáncer de mama, cáncer de próstata	D/P/T	Tumor

D: diagnóstico; P: pronóstico; T: terapéutico

## 2.5- Impacto de las nuevas tecnologías en la identificación de biomarcadores

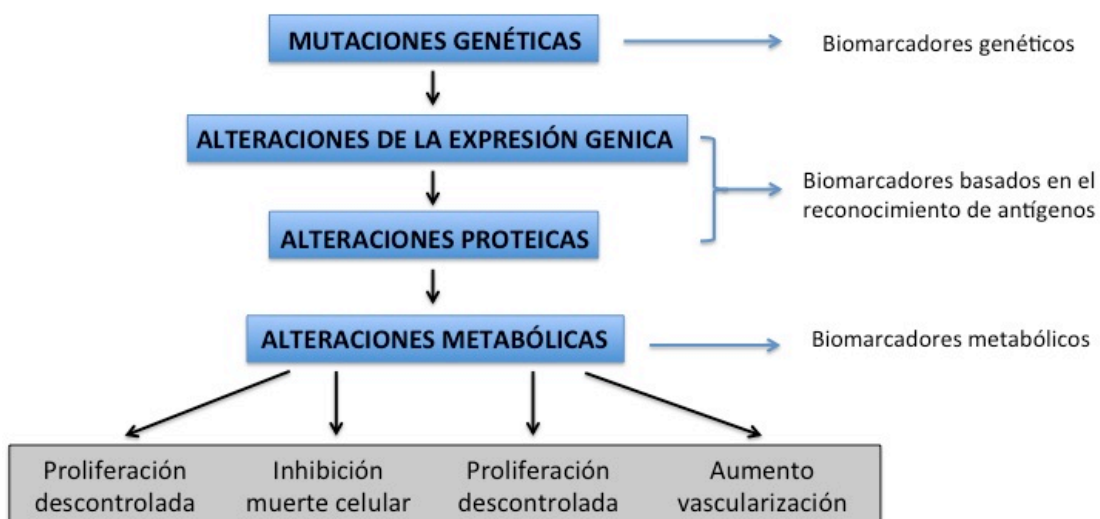
El avance en la identificación de nuevos biomarcadores está íntimamente ligado al desarrollo de nuevas plataformas analíticas. A nivel técnico, el desarrollo en los últimos años de las tecnologías “ómicas” (genómica, proteómica, metabolómica) ha permitido la identificación de una nueva generación de biomarcadores moleculares basados en la Biología de Sistemas. La Biología de Sistemas trata de integrar el estudio sistemático de los flujos e interacciones en sistemas biológicos con el fin de

proporcionar una mejor comprensión del organismo. Estas nuevas aproximaciones han abierto nuevas vías de conocimiento para el descubrimiento de biomarcadores característicos de distintas patologías, lo que viene reflejado por un importante aumento en el número de trabajos dirigidos a la identificación de nuevos marcadores (**Figura 3**, adaptada de (Research Advocacy Network, 2010)).



**Figura 3:** Número de artículos científicos publicados relacionados con biomarcadores

Tanto la genómica como la proteómica aportan gran información a la identificación de biomarcadores característicos del genotipo de cada individuo. En este contexto, la metabolómica aporta una información complementaria a través del estudio del metaboloma, que es el producto final de todos los procesos de transcripción y transducción de señales. El perfil metabólico de la muestra aporta información sobre el fenotipo, influenciado por el genotipo, pero también por otros factores como la dieta, el ejercicio, el estado de salud, la microflora intestinal, factores ambientales, la edad, etc. (**Figura 4**, adaptada de (Bhatt et al., 2010)).

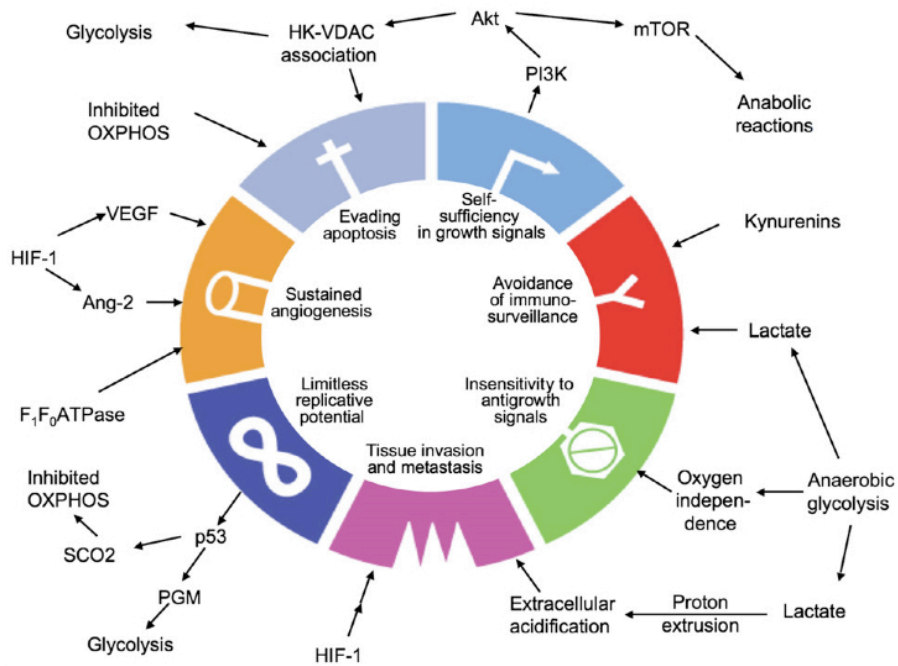


**Figura 4:** Oportunidades para la identificación de biomarcadores en el proceso de carcinogénesis.

Todas estas plataformas tecnológicas son muy prometedoras en la identificación de nuevos biomarcadores que podrían mejorar significativamente el tratamiento de los pacientes oncológicos a través la detección temprana de la enfermedad y de la individualización de las terapias administradas.

### 3- METABOLÓMICA EN LA IDENTIFICACIÓN DE BIOMARCADORES ONCOLÓGICOS

Como se ha mencionado anteriormente, y se indica en la **Figura 5** (adaptada de (Kroemer & Pouyssegur, 2008), existen 8 características comunes adquiridas por todas las células tumorales durante el proceso de carcinogénesis (Hanahan et al. 2000; Hanahan & R. a Weinberg 2011; Vizán et al. 2008).



**Figura 5:** Cambios producidos durante el proceso de carcinogénesis

Globalmente, todas estas alteraciones son derivadas de la adaptación a la limitada disponibilidad de nutrientes y oxígeno, así como a la necesidad de satisfacer la elevada demanda energética y biosíntesis debida a la proliferación celular descontrolada. Algunas de estas características representan la base bioquímica para el desarrollo de nuevas técnicas ampliamente conocidas y utilizadas en la actualidad para el diagnóstico clínico de enfermedades oncológicas como, por ejemplo, el FDG-PET (Rajendran et al., 2004).

Uno de los cambios más importantes derivado de este proceso es la alteración del metabolismo celular, lo que provoca cambios significativos en los niveles de los metabolitos. El primer trabajo que describió una relación entre la enfermedad oncológica y el perfil metabólico fue realizado por Warburg (Warburg, 1956), que observó importantes diferencias en el balance energético entre células normales y cancerígenas en condiciones aerobias, derivadas de un aumento de la glucólisis en las células tumorales. Desde entonces, se han identificado distintas alteraciones

metabólicas asociadas al crecimiento de las células cancerosas:

- ↑ Glucólisis (Warburg, 1956)
- ↑ Acidificación del microambiente (Muñoz-Pinedo et al., 2012)
- ↑ Captación y metabolismo de la glutamina (DeBerardinis & Cheng, 2010)
- ↑ Síntesis lipídica (principalmente síntesis *de novo*) (Baron et al., 2004; Mashima et al., 2009)
- ↑ Síntesis de nucleótidos y dependencia de folato

Los metabolitos presentes en una muestra biológica, así como sus niveles, son un reflejo directo del metabolismo celular, que se verá alterado frente a distintas situaciones como enfermedades, hábitos alimenticios, tratamientos farmacológicos, intervenciones quirúrgicas, etc. En este sentido, la metabolómica estudia todos los compuestos de bajo peso molecular (metabolitos) presentes en una muestra biológica en un determinado momento en el tiempo, empleando para ello plataformas tecnológicas avanzadas en combinación con técnicas de análisis estadístico multivariante. Este tipo de aproximaciones permiten medir en un solo experimento, los niveles de gran cantidad de metabolitos (sustratos, productos y cofactores de reacciones enzimáticas, fármacos y productos de su metabolismo, etc.). Por otro lado, el desarrollo de nuevas herramientas bioinformáticas aplicadas al análisis de los datos metabólicos, unido a un mejor conocimiento del metaboloma humano y el desarrollo de bases de datos específicas, ha contribuido al desarrollo de nuevas aproximaciones en este área, basadas en la metabolómica.

Por este motivo se considera que la metabolómica es una técnica complementaria a las demás aproximaciones “ómicas” (genómica y proteómica) que posee un enorme potencial en el desarrollo de técnicas de diagnóstico temprano y monitorización de tratamientos, lo que viene reflejado por un aumento progresivo en el número de estudios en los últimos años (Duarte & Gil, 2012).

Los resultados de estos estudios han aportado evidencias directas de las alteraciones metabólicas descritas anteriormente, así como de otras, a través del estudio de los perfiles metabólicos de células y tejidos tumorales. Además, el estudio directo de tejidos tumorales ha hecho posible la identificación de marcadores metabólicos para la detección y caracterización de distintos tipos de cáncer (Bathen et al., 2010). Resulta importante destacar que el estudio de los perfiles metabólicos de distintos biofluidos (suero, plasma, orina, saliva, etc.) ofrece

la posibilidad de evaluar de manera dinámica tanto los efectos de la enfermedad como los de distintos tratamientos en el metabolismo del paciente.

### **3.1- El metaboloma humano**

El metaboloma humano se define como el conjunto de todos los metabolitos que se encuentran en el cuerpo humano (Wishart et al., 2013). Los metabolitos incluyen moléculas de bajo peso molecular (inferior a 1 kDa): péptidos, lípidos, carbohidratos, aminoácidos, vitaminas, minerales, medicamentos, toxinas, ácidos orgánicos, aditivos alimentarios, etc. Desde 2005, a través del *Human Metabolome Project* ("The Human Metabolome Project website"), se han desarrollado numerosos estudios con el fin de contribuir al conocimiento del metaboloma humano. Toda la información resultante es depositada periódicamente para su acceso público en la *Human Metabolome Database* (HMDB) (Wishart et al., 2007, 2009, 2013). Esta base de datos incluye actualmente información sobre más de 40.000 metabolitos detectados o a los que el cuerpo humano puede estar expuesto. En este sentido, debe indicarse que el número e identidad de los metabolitos puede diferir mucho en función de la muestra que se analice.

Los biofluidos que se emplean con más frecuencia en los estudios de metabolómica dirigidos al desarrollo de nuevos métodos diagnósticos son suero, plasma y orina, debido principalmente a la posibilidad de obtenerlos de manera mínimamente invasiva. En este sentido, dada su relevancia clínica, se han realizado importantes esfuerzos dirigidos a la caracterización del metaboloma de estos biofluidos humanos (Psychogios et al., 2011).

### **3.2- Posibilidades de la metabolómica**

En el ámbito clínico, el objetivo de los estudios de metabolómica es analizar las fluctuaciones en los niveles de los metabolitos presentes en una muestra biológica y buscar correlaciones sobre la existencia o no de una patología, la respuesta a un tratamiento farmacológico, etc.

Este tipo de aproximaciones están resultando de gran utilidad en:

1- La identificación de biomarcadores para el diagnóstico temprano de enfermedades. Existen diversos estudios relacionados con enfermedades cardiovasculares (Brindle et al., 2002; Kirschenlohr et al., 2006), diabetes (Fiehn et al., 2010; Manna et al., 2011), alteraciones metabólicas en neonatos (Arn, 2007;

Wikoff et al., 2007), daño hepático inducido por alcohol (Manna et al., 2011) y cáncer (Chan et al., 2009; Gu et al., 2012; Mamas et al., 2011).

2- La caracterización de perfiles metabólicos y biomarcadores específicos de la respuesta a un determinado tratamiento, a través de los estudios de farmacometabolómica (Clayton et al., 2009; Jeremy K Nicholson et al., 2011). La finalidad de este tipo de estudios es poder identificar *a priori*, a través del estudio del perfil metabólico, aquellos pacientes que van a responder favorablemente a una terapia, cuáles son los que poseen una mayor probabilidad de desarrollar efectos adversos, etc. Esta información resulta de gran utilidad para la estratificación de pacientes y puede contribuir significativamente al desarrollo de terapias más eficaces y seguras.

3- El desarrollo de nuevas terapias dirigidas a partir de la información que ofrece la metabolómica en el proceso de desarrollo de fármacos. La identificación de las enzimas responsables de las alteraciones específicas observadas en el metaboloma de las células tumorales permite el desarrollo de terapias dirigidas, mucho más selectivas y eficaces (D'Alessandro & Zolla, 2012; Jeremy K Nicholson et al., 2011).

En relación con el análisis de los datos, existen principalmente dos enfoques:

- **Análisis dirigido de metabolitos:** Se emplea en los casos en los que se posee información previa sobre rutas metabólicas implicadas en el desarrollo de la patología o el proceso estudiado. En este caso, el análisis metabólico va dirigido a la identificación y cuantificación de un grupo de metabolitos específicos, implicados en un mismo proceso o ruta metabólica. A partir de ahí, el análisis de los datos se puede hacer de forma individualizada o en conjunto.
- **Estudio del perfil metabólico:** Se utiliza para estudiar el metaboloma de forma holística, sin ser necesario un conocimiento previo sobre las rutas o metabolitos implicados en los cambios, y permite la identificación de un gran número de metabolitos (100-1000s). El análisis de los resultados se realiza mediante estadística multivariante, y no persigue necesariamente la cuantificación de los metabolitos identificados. Esta aproximación permite explorar de forma simultánea diversas rutas metabólicas, pudiendo asociar un perfil metabólico a una condición biológica determinada.

### 3.3- Plataformas tecnológicas en metabolómica

Uno de los elementos claves en los estudios de metabolómica es la elección de una plataforma tecnológica para la realización de las medidas, lo que dependerá fundamentalmente del tipo de muestra que se pretende analizar y los objetivos planteados en el estudio.

Las dos plataformas tecnológicas más utilizadas en metabolómica son la espectroscopía de Resonancia Magnética Nuclear (RMN) y la Espectrometría de Masas (EM). Los principios físicos en los que se basa cada una de ellas son diferentes, y ambas presentan distintas ventajas e inconvenientes que las hacen complementarias. La **Tabla 4** recoge las características principales de cada una de ellas.

**Tabla 4:** Principales ventajas e inconvenientes de la EM y RMN en metabolómica

	EM	RMN
VENTAJAS	<p>Elevada sensibilidad</p> <p>Permite mayor selectividad: En función de la preparación de la muestra, se pueden dirigir el análisis a distintos tipos de compuestos (análisis dirigido)</p>	<p>No destructiva</p> <p>No selectiva: Detecta todos los metabolitos presentes en la muestra que contengan algún hidrogeno en su estructura</p> <p>Mínima preparación de la muestra</p> <p>Altamente reproducible</p> <p>Gran utilidad en la identificación de compuestos</p> <p>Permite análisis directo de tejidos (HR-MAS)</p> <p>Asignación de los compuestos a través de la información disponible en bases de datos de acceso público</p>
INCONVENIENTES	<p>Requiere preparación y separación previa de la muestra</p> <p>Baja reproducibilidad</p> <p>Necesidad de patrones internos para la identificación y cuantificación de compuestos</p> <p>Técnica destructiva</p>	<p>Baja resolución en muestras muy complejas debido a la superposición de las señales</p> <p>Baja sensibilidad: Límite de detección en el rango micromolar</p>



## 4- METABOLÓMICA POR RMN

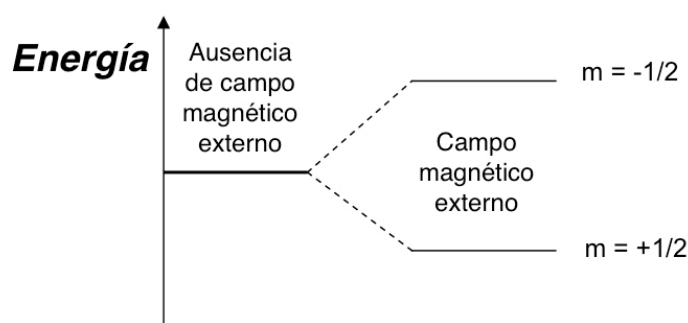
La RMN es una de las plataformas tecnológicas más empleadas en metabolómica dada su versatilidad en el análisis de muestras biológicas.

### 4.1- Principios básicos de RMN

Ciertos núcleos atómicos, entre los que se incluye el protón ( $^1\text{H}$ ), presentan un momento magnético, determinado por su número de espín ( $I$ ), que los hace sensibles a la presencia de campos magnéticos externos. Se trata de núcleos magnéticamente activos, capaces de absorber radiaciones de una determinada frecuencia en presencia de un campo magnético externo. La RMN de protón ( $^1\text{H}$ -RMN), un núcleo presente en la naturaleza con una abundancia de más del 99.9%, se basa en las propiedades magnéticas de este núcleo.

La  $^1\text{H}$ -RMN está basada en el distinto comportamiento de los núcleos activos de hidrógeno contenidos en las moléculas orgánicas cuando son expuestos a una radiación electromagnética en presencia de un campo magnético externo. Aunque esta técnica encuentra un gran campo de aplicación en Química Orgánica para caracterizar la estructura de pequeñas moléculas, su uso está cada vez más extendido en el área de la Biomedicina.

El modelo mecano-cuántico de la RMN establece que cada núcleo posee  $2I+1$  estados de espín posibles ( $m$ ). En ausencia de un campo magnético, los estados de espín posibles presentan igual energía y se distribuyen de manera aleatoria. En presencia de un campo magnético externo ( $B_0$ ), se produce un alineamiento bien a favor o en contra del campo magnético, por lo que se crea una diferencia de energía entre los posibles estados de espín, siendo el estado de menor energía el más poblado (**Figura 6**).



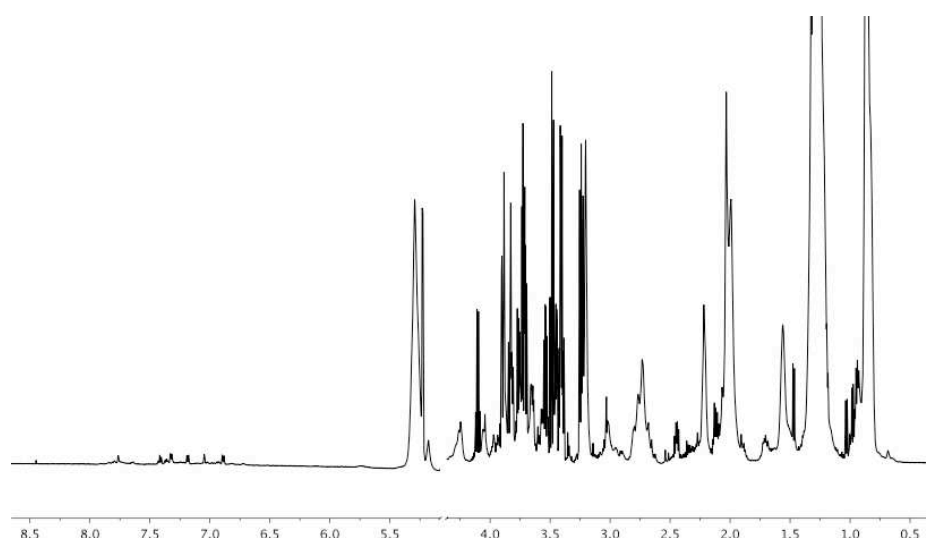
**Figura 6:** Diferencia de energía entre los dos estados de espín para el núcleo de hidrógeno. Los protones ( $^1\text{H}$ ) presentan un espín de  $1/2$ , dando lugar a  $(2 \times 1/2) + 1 = 2$  niveles de energía:  $+1/2$  (más estable y de menor energía) y  $-1/2$  (menos estable y de mayor energía) (Edwards & Ph, 1946).

La diferencia de energía entre estos dos niveles ( $\Delta E$ ) es la que determina la frecuencia de la radiación necesaria para la excitación del núcleo a un estado de espín de mayor energía. Esta frecuencia, característica para cada núcleo, se conoce como frecuencia de *Larmor*. En presencia de un campo magnético externo, la aplicación de un pulso de radiofrecuencia perpendicular a la dirección del campo magnético principal provoca la aparición de una magnetización transversal. La vuelta al estado de equilibrio, o relajación, se caracteriza por la desaparición de la magnetización transversal, dando lugar a lo que se conoce como *Free Induction Decay* (FID).

La transformada de *Fourier* permite convertir la FID en un espectro de RMN, con picos de distintas formas y tamaños, posicionados a lo largo del eje de las frecuencias, en función de su desplazamiento químico ( $\delta$ ). El desplazamiento químico al que aparecen las distintas señales en el espectro varía en función del entorno químico en que se encuentra cada  $^1\text{H}$ , de manera que la posición en el espectro no es aleatorio. Las unidades del desplazamiento químico se expresan en partes por millón (ppm) y equivalen a  $10^6$  veces la relación entre la separación de las señales y la frecuencia del campo magnético externo, ambas expresadas en Hz. El área de cada pico en el espectro es proporcional al número de protones que está contribuyendo a esa señal, por lo que los valores de la intensidad de las señales pueden emplearse con fines cuantitativos.

Los experimentos de RMN se pueden adquirir de manera consecutiva, de tal modo que la segunda y siguientes FIDs se van acumulando, aumentando así la relación señal/ruido (S/N) en el espectro de RMN y por tanto la sensibilidad. Después de cada adquisición, los núcleos no vuelven al equilibrio de manera instantánea, sino en un tiempo de relajación que es característico para cada núcleo. Por esta razón, es importante introducir tiempos de relajación suficientemente largos entre adquisiciones para asegurar que todos los núcleos de las moléculas presentes en la muestra han vuelto al equilibrio antes de la siguiente adquisición.

Los espectros  $^1\text{H}$  RMN de muestras biológicas son complejos ya que contienen señales correspondientes a múltiples compuestos. Esta característica queda reflejada en la **Figura 7**, que muestra un espectro representativo de una muestra de suero humano.



**Figura 7:** Espectro  $^1\text{H}$ -RMN CPMG de una muestra de suero de un individuo sano

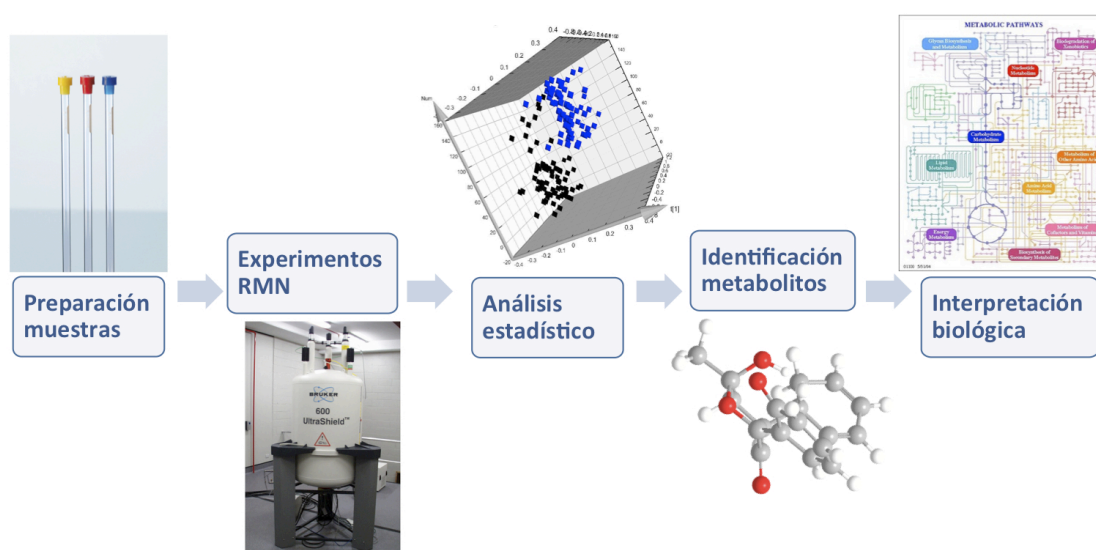
Los espectros de suero o plasma humano presentan una serie de señales anchas características derivadas de las lipoproteínas presentes en la muestra. Este tipo de señales se superpone sobre muchas otras señales también presentes en el espectro, y dificulta la identificación y cuantificación de otros metabolitos menos abundantes en la muestra que resuenan en esa misma región del espectro. Este tipo de problemas en RMN pueden abordarse mediante la utilización de distintas secuencias de pulso que, a través de la utilización de filtros, facilitan la eliminación de las contribuciones espectrales de especies moleculares de gran tamaño. Más adelante, en este mismo apartado, se detallará el tipo de secuencias utilizadas habitualmente para el análisis de muestras biológicas en los estudios de metabolómica.

Finalmente, y como se ha comentado con anterioridad, una de las limitaciones que presenta la RMN en metabolómica es su baja sensibilidad. La sensibilidad y resolución máxima que pueden conseguirse en el espectro RMN viene determinada por la intensidad, homogeneidad y estabilidad del campo magnético generado por el espectrómetro ( $B_0$ ). Cuanto mayor es la intensidad del campo magnético generado por el equipo, mayor es la relación S/N. En las últimas décadas, los avances en el campo de la RMN se han traducido en espectrómetros cada vez más sensibles, debido principalmente a la obtención de campos cada vez más intensos con la introducción de imanes superconductores y al uso de criosondas (Keun et al., 2002), mejorando significativamente el límite de detección de los instrumentos. Estos avances técnicos, unidos al desarrollo de secuencias de pulso específicas

para este tipo de aplicaciones, se ha traducido en un aumento significativo en el número de estudios de metabolómica.

#### 4.2- Esquema de trabajo en un estudio de metabolómica por RMN

En general, y desde el punto de vista metodológico, el esquema de trabajo que se sigue en cualquier estudio de metabolómica por RMN encaminado a la identificación de nuevos biomarcadores es el que se muestra en la **Figura 8** (adaptada de (Beckonert et al., 2007)) y se detalla en los siguientes apartados.



**Figura 8:** Procedimiento de trabajo típico en el estudio de perfiles metabólicos mediante RMN.

##### a) Diseño del estudio

En este tipo de proyectos es fundamental contar con un diseño experimental que se adapte adecuadamente a la enfermedad que se pretende estudiar y a los objetivos que se quieren alcanzar. Entre los aspectos a tener en cuenta se encuentran: i) el número de sujetos a incluir en el estudio y la distribución de edades, sexo, procedencia étnica y otras variables en cada uno de ellos, ii) el tipo de muestra a analizar, iii) los procedimientos para la recogida de muestras, iv) el procesamiento de las mismas, v) los parámetros relativos tanto al análisis por RMN como al análisis estadístico de los resultados obtenidos, y vi) los datos clínicos asociados a las muestras necesarios para la interpretación biológica de esos resultados.

En los estudios de metabolómica por RMN las diferencias entre muestras debidas a inestabilidades del instrumento son mínimas, ya que se trata de una

técnica muy robusta. La mayor fuente de variabilidad en este tipo de estudios suele proceder de la heterogeneidad dentro de cada grupo de individuos. Esta variabilidad no suele ser un problema importante en estudios con animales. Sin embargo, en estudios clínicos, factores como el estilo de vida, la alimentación, el hábito tabáquico, el consumo de medicamentos, la presencia de otras enfermedades crónicas, el ejercicio, etc. pueden dar lugar a una gran variabilidad en los grupos de muestras, por lo que es importante definir antes de comenzar el análisis los criterios de inclusión/exclusión de cada uno de los grupos del estudio. También es importante identificar previamente otros parámetros que podrían contribuir a las variaciones, y que deben ser controladas durante el estudio a fin de asegurar la calidad de los datos obtenidos, como pueden ser: la manipulación y conservación de las muestras, el momento de recogida, la calibración y el mantenimiento de los equipos, la preparación de las muestras, la recogida de datos clínicos, etc.

### b) Preparación de las muestras

El procedimiento de preparación de la muestra para su análisis mediante RMN debe reunir los siguientes criterios: reproducibilidad, facilidad, reducción de variaciones entre muestras (p.ej., variaciones en el pH), optimización de la calidad espectral en lo que respecta a resolución y sensibilidad, ausencia de artefactos analíticamente detectables, etc.

Las muestras biológicas habitualmente utilizadas en los estudios de metabolómica por RMN incluyen biofluidos (p.ej., suero, plasma, líquido cefalorraquídeo, orina, etc.) y tejidos (John C. Lindon et al., 2000). Cada uno de estos tipos de muestras presenta propiedades físico-químicas diferentes (**Tabla 5**), por lo que la preparación de la muestra deberá adecuarse en cada caso.

**Tabla 5:** Características físico-químicas de los biofluidos más habituales en los estudios de metabolómica.

	% agua	Viscosidad	% proteínas	% lípidos	Solapamiento
Orina	+++	-	-	-	-
Plasma/Suero	+++	++	+++	+++	++
Líquido cefalorraquídeo	+++	+	+	+	++
Saliva	+++	++	++	++	+
Líquido seminal	+	+++	+	+	+++
Líquido	+	++	+	+	++

	% agua	Viscosidad	% proteínas	% lípidos	Solapamiento
prostático					
Líquido amniótico	+++	+	+	+	+

En los estudios metabolómicos relacionados con procesos oncológicos, los biofluidos más estudiados son fundamentalmente la orina y el suero (o plasma) sanguíneo. Ambos biofluidos presentan la ventaja de que su obtención es mínimamente invasiva. Además, la determinación de perfiles metabólicos de biofluidos mediante RMN de orina y suero (o plasma) no requiere procedimientos sofisticados en lo que concierne a la preparación de la muestra. Por otro lado, el desplazamiento químico al que aparecen las señales en el espectro de RMN puede variar en función del disolvente, pH, etc. que se utilice en la preparación. Este efecto es especialmente acusado en algunos metabolitos de pequeño peso molecular, como el citrato o la histidina. La adición de agua o buffer es suficiente, en la gran mayoría de los casos, para reducir posibles variaciones debidas al pH o a la viscosidad (Beckonert et al., 2007). Todo esto hace que este tipo de medidas suelen estar exentas de posibles errores debidos a procedimientos de extracción variables.

Otros procedimientos de preparación de muestras menos establecidos, y que implican mayor variabilidad, incluyen la liofilización y posterior reconstitución de la muestra en el caso de muestras muy diluidas (p.ej., líquido amniótico) y la extracción previa de los metabolitos presentes en la muestra, obteniendo una fase acuosa y otra orgánica, en el análisis de tejidos. En este contexto, el empleo de biofluidos como líquido biliar, fluido ovárico, secreciones prostáticas, etc. ha quedado limitado a casos en los que la enfermedad está directamente relacionada con la afectación del órgano del que son reflejo directo.

### **c) Adquisición de los espectros**

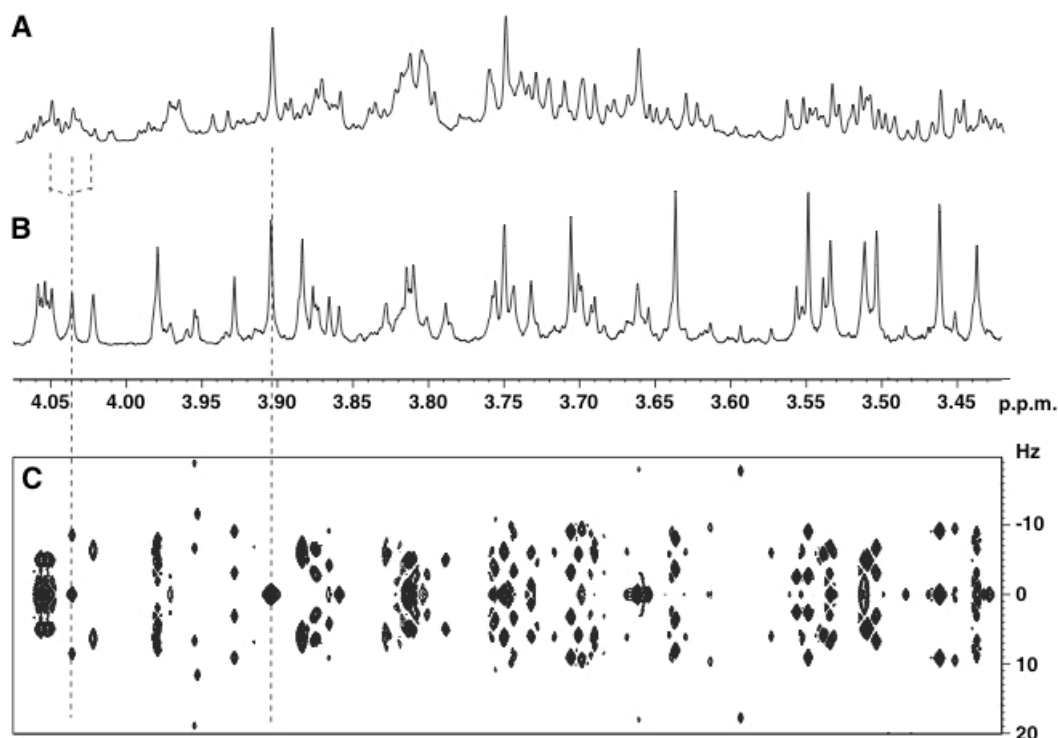
Si se compara con otras aplicaciones de la RMN, la metabolómica no requiere de un alto nivel de sofisticación en cuanto a las secuencias de pulsos que se emplean. Sin embargo, la aplicación de esta técnica está condicionada por el hecho de que en todos los experimentos, basados en la detección de  $^1\text{H}$  realizados sobre biofluidos, se intenta caracterizar moléculas que se encuentran a bajas concentraciones (micromolar) sobre un fondo de protones de la molécula de agua que se encuentran a una concentración de 110M. Por otro lado, los espectros de

RMN de muestras biológicas suelen presentar señales anchas correspondientes a moléculas de mayor peso molecular, fundamentalmente lipoproteínas y proteínas, que se solapan con señales de otros metabolitos de menor peso molecular, dificultando así su detección y cuantificación. En general, las secuencias de pulsos generalmente empleadas en los experimentos de metabolómica van encaminadas a superar estos dos aspectos.

Normalmente, para cada muestra incluida en el estudio, se adquiere un conjunto de experimentos 1D RMN (Beckonert et al., 2007):

- **1D NOESY (presaturación):** Utilizando esta secuencia de pulsos, conocida como “noesy-presat” (Kumar et al., 1980), la señal del agua es saturada a través de una irradiación de frecuencia selectiva, mejorando significativamente la supresión de la señal correspondiente a esta molécula. Esta técnica proporciona un método simple, altamente reproducible y robusto para la adquisición de espectros de RMN de alta calidad en soluciones acuosas.
- **CPMG (editado basado en las propiedades de relajación):** Las moléculas de alto peso molecular, como proteínas y lipoproteínas, son más rígidas y experimentan una relajación transversal muy rápida. Este fenómeno se utiliza en metabolómica para la eliminación de señales pertenecientes a especies de elevado peso molecular, mejorando así la resolución de las señales correspondientes a metabolitos de bajo peso molecular. En este experimento, lo que se hace es introducir un período de relajación de entre 40-100 ms antes de la adquisición. La secuencia más empleada en metabolómica es la denominada CPMG (“Carr Purcell Meiboom and Gill” (Meiboom & Gill, 1958)). Esta técnica es especialmente útil para muestras biológicas, como el suero o plasma sanguíneo, que tengan un alto contenido en proteínas.
- **Difusión (editado basado en las propiedades de difusión):** Este tipo de experimentos de RMN permite la atenuación de las señales pertenecientes a moléculas de pequeño tamaño gracias a que presentan coeficientes de difusión molecular rápidos, al contrario que el de las moléculas de elevado tamaño, con coeficientes de difusión mucho mayores (Lenz & Wilson, 2007).
- **J-resuelto (experimento homonuclear 2D):** Se trata del experimento 2D homonuclear más simple en RMN. En este tipo de experimento, la información relacionada con los acoplamientos de tipo escalar es separada de los desplazamientos químicos en una segunda dimensión (Viant, 2003). El acoplamiento escalar provoca el desdoblamiento de las señales debido al

acoplamiento entre protones magnéticos equivalentes, entre 1 y 3 enlaces de distancia, definido por la constante de acoplamiento,  $J$ . Una característica muy interesante de este experimento es el cálculo del denominado espectro de desplazamientos químicos que se obtiene mediante una proyección de la matriz 2D completa, denominada proyección *skyline*. De este modo, la complejidad espectral se reduce de manera considerable (**Figura 9**).



**Figura 9:** Región espectral correspondiente un extracto de *Arabidopsis thaliana* en agua:cloroformo: A) Espectro 1D convencional; B) Proyección tipo *skyline* de un experimento 2D J-res; C) Experimento 2D J-res (Viant, 2003)

En la **Tabla 6** se recogen las principales ventajas e inconvenientes de cada uno de los experimentos descritos anteriormente.

**Tabla 6:** Principales características de los experimentos de RMN habitualmente utilizados en los estudios de metabolómica.

Método	Finalidad	Ventajas	Inconvenientes
Presaturación (1D-noesy)	Detecta señales de compuestos de bajo y alto/medio peso molecular.	Útil en la cuantificación.	Espectros complejos debido a la presencia de metabolitos de bajo y alto peso molecular en el mismo espectro, unido a los acoplamientos de señales ( $J$ ).



Método	Finalidad	Ventajas	Inconvenientes
Editado basado en las propiedades de relajación (CPMG)	Muestra únicamente las señales de las especies de bajo peso molecular, a través de la supresión de las señales de moléculas de mayor tamaño.	Espectro fácilmente interpretables debido a la simplificación de señales. Muy utilizados en muestras de suero/plasma.	No presenta ventajas significativas en el caso de biofluidos como orina, donde predominan los compuestos de bajo peso molecular.
Editado basado en las propiedades de difusión (Difusión)	Muestra únicamente señales correspondientes a moléculas de peso molecular relativamente alto (p. ej., lípidos).	Muy utilizados en muestras de suero/plasma.	No presenta ventajas significativas en el caso de biofluidos como orina, donde no existen diferencias apreciables en los coeficientes de difusión de las distintas moléculas.
Experimento homonuclear 2D (J-res)	Separa la información relacionada con los acoplamientos de tipo escalar en una segunda dimensión.	Facilita enormemente la interpretación y asignación de las señales.	En comparación con los otros espectros 1D, su interpretación es más complicada.

Como se ha indicado anteriormente, los experimentos 1D de muestras biológicas se caracterizan por presentar un elevado solapamiento de señales debido a la presencia de múltiples especies moleculares distintas en una misma muestra. Por esta razón, en los estudios de metabolómica suelen adquirirse una serie de experimentos 2D con muestras representativas de cada grupo del estudio. Este tipo de experimentos permiten dar solución a las dificultades asociadas al solapamiento de señales en los espectros 1D, ya que las frecuencias de resonancia se resuelven en dos dimensiones: F1 y F2, lo que resulta de gran utilidad en la asignación de metabolitos. Los experimentos más utilizados en metabolómica como apoyo para la asignación de las señales son: 2D  $^1\text{H}$ - $^1\text{H}$  *Correlation Spectroscopy* (COSY), *Total Correlation Spectroscopy* (TOCSY), *Nuclear Overhauser Effect Spectroscopy* (NOESY) y 2D  $^1\text{H}$ ,  $^{13}\text{C}$  *Heteronuclear Single Quantum Correlation* (HSQC).

#### d) Pre-procesado de los datos

Una vez adquiridos los experimentos de RMN, se realiza un pre-procesado de los datos que permita obtener una representación interpretable de los datos. Este procesado incluye: i) la transformada de *Fourier*, ii) la corrección de la fase y la

línea base del espectro, iii) la calibración de los valores de desplazamiento químico respecto a una señal de referencia para conseguir un alineamiento adecuado de los espectros de las distintas muestras, corrigiendo así posibles variaciones debidas a ligeras modificaciones del pH, fuerza iónica, etc. entre muestras. En la actualidad, es posible realizar todas estas transformaciones de forma semi-automática. El objetivo final de esta etapa es transformar los datos obtenidos mediante la adquisición de los espectros de RMN en información interpretable estadísticamente. Este proceso facilitará el análisis posterior y hará que éste sea más fiable.

El pre-procesado de los datos continúa con la exclusión de regiones espectrales con información que no es reproducible entre muestras o que no corresponde a señales de metabolitos. La región espectral que se incluye normalmente los estudios de RMN es la que abarca desde 0.15 a 10 ppm. Dentro de esta ventana espectral, se excluye la región correspondiente a la señal del agua (4.6 - 4.9 ppm, aproximadamente) o, en el caso de extractos orgánicos, cualquier otro disolvente que se haya utilizado en la preparación de la muestra. Otra señal que suele eliminarse es la correspondiente a la urea (entre 5.4 – 6.0 ppm, aproximadamente), muy aparente en los espectros de orina, ya que se trata de una señal que no puede ser cuantificada con exactitud al encontrarse sus protones en continuo intercambio con el agua, por lo que su inclusión en el análisis estadístico podría introducir cierta variabilidad, no necesariamente significativa, entre muestras.

#### **e) Asignación de los espectros**

Tras el pre-procesado de los espectros, se procede a la asignación de las señales más representativas identificadas en los espectros de RMN. Para ello, y partiendo de los desplazamientos químicos obtenidos en los experimentos 1D y 2D, se emplean bases de datos de acceso público con datos disponibles para miles de metabolitos identificados, o potencialmente identificables, en muestras biológicas. Las bases de datos más utilizadas en metabolómica son la *Human Metabolome Database* (HMDB) (Wishart et al., 2009) y la del *Biological Magnetic Resonance Data Bank* (BMRB) (“BMRB - Biological Magnetic Resonance Bank”).

#### **f) Procesado de los datos**

Este paso consiste en la subdivisión, e integración del espectro, en pequeñas regiones denominadas *buckets*, un proceso conocido como *bucketting*. En la gran mayoría de los casos se emplea para la integración un tamaño de *bucket* constante

de 0.04 ppm. El mayor inconveniente de este procedimiento es la ausencia de flexibilidad en los límites, de ahí la importancia, como se indicó en el apartado anterior, de alinear correctamente los espectros de las distintas muestras empleando un valor apropiado como referencia para los desplazamientos químicos dentro del espectro. Si un pico cruza el límite entre dos *buckets*, este desplazamiento dará lugar a cambios importantes en el posterior análisis de los datos. En este contexto, es posible también emplear tamaños de *bucket* menores (p.ej., 0.01 ppm e incluso inferiores) en aquellos casos en los que se pretende evaluar diferencias en regiones muy próximas en el espectro. Tras la integración de los distintos *buckets* en cada espectro, se obtiene una relación de valores proporcional a la intensidad de los distintos picos en cada muestra. Este proceso de *bucketting* también se realiza de forma automática mediante la utilización de herramientas informáticas específicas.

En función del tipo de muestra analizada, será necesario normalizar el área total de cada espectro a un determinado valor de referencia antes de poder comparar los espectros entre sí. La normalización se hace necesaria en muestras donde la concentración total de la muestra varía entre individuos debido a variaciones en el volumen, tamaño de la muestra, etc. Es muy común en el caso de muestras de orina y extracción de tejidos biológicos, donde el factor de normalización de cada espectro dependerá del factor de dilución de la muestra correspondiente.

#### **g) Análisis estadístico multivariante**

Los métodos de análisis estadístico multivariante proporcionan un medio para evaluar la información contenida en los espectros de RMN de mezclas complejas. La ejecución de los pasos indicados anteriormente genera una matriz de datos ( $X$ ) que contiene tantas filas como muestras ( $n$ ) y tantas columnas ( $k$ ) como *buckets* se hayan empleado para segmentar el espectro de RMN.

El análisis estadístico multivariante posterior de los datos requiere generalmente la realización de una transformación matemática previa, entre las que se incluye el centrado y el escalado de los datos (van den Berg et al., 2006). El centrado convierte todas las intensidades en fluctuaciones alrededor de cero, en lugar de en torno a la media de intensidades de cada *bucket*. Mediante esta transformación lo que se consigue es focalizar los resultados en las diferencias entre muestras, y no en la variabilidad global. A diferencia de la normalización entre muestras, únicamente necesaria en el caso de muestras con distintos factores de dilución, el

escalado es un proceso de normalización entre variables que suele ser habitual en todos los estudios de metabolómica. Los espectros de muestras biológicas presentan normalmente unos pocos metabolitos con concentraciones muy elevadas, y muchos otros con señales mucho menos intensas. Los métodos de escalado están dirigidos a ajustar las diferencias en las fluctuaciones entre los distintos metabolitos. En función del método de escalado empleado, se consigue enfatizar diferentes aspectos de los datos experimentales, cada uno con ventajas e inconvenientes, por lo que habitualmente se evalúan distintas aproximaciones antes de seleccionar el modelo definitivo para el análisis de los datos.

En general, los métodos de centrado/escalado más utilizados en metabolómica son:

- Centrado (Ctr): Elimina únicamente el valor de la media de cada *bucket* en las variables correspondientes.
- Centrado + Varianza Unitaria (UV): Emplea la desviación estándar de cada variable como factor de escalado. Mediante esta transformación se consigue que cada variable tenga el mismo peso en la totalidad del espectro, independientemente de la intensidad de la señal en el espectro original. El inconveniente de esta aproximación es que al tener el mismo peso todas las variables es posible introducir variabilidad entre las muestras debida a errores instrumentales y/o metodológicos, que nada tienen que ver con la variabilidad entre los grupos de muestras objeto del estudio.
- Centrado + Pareto (Par): Emplea la raíz cuadrada de la desviación estándar de cada variable como factor de escalado. Este método es ampliamente utilizado para el tratamiento de datos espectroscópicos, ya que continúa manteniendo una distribución en el peso de las variables similar a la de los datos originales.

El análisis estadístico multivariante se realiza sobre estos datos transformados. En función del objetivo del estudio (p.ej., visualización de los datos, selección de variables, clasificación de grupos, modelos predictivos, etc.) se utilizarán distintas aproximaciones. Generalmente, en primer lugar se emplean métodos de análisis no supervisado, que permiten evaluar la variabilidad de los datos sin tener en cuenta su clasificación dentro de ningún grupo. Seguidamente, los modelos supervisados, que utilizan la información sobre la clasificación de las muestras para maximizar la separación entre grupos, aportan información de gran utilidad para la identificación

de potenciales biomarcadores característicos de cada grupo.

La variabilidad entre muestras es muy difícil de detectar mediante la observación directa de los espectros de RMN y hace necesaria la introducción de técnicas estadísticas que permitan evaluar esas diferencias en relación a las variables que se conocen de cada muestra, entre las que se encuentran desde los datos clínicos del individuo hasta los parámetros instrumentales de adquisición del experimento de RMN. El análisis de los componentes principales (PCA) es un método no supervisado, es decir, no requiere conocimiento previo de las muestras, que resulta de gran utilidad para la identificación de patrones y/o tendencias dentro de los grupos de muestras. Consiste en la reducción de la dimensionalidad de los datos de manera que las múltiples variaciones quedan agrupadas en distintos componentes (primero, segundo, etc.), de acuerdo a su importancia en la interpretación de la variable observada (Halouska & Powers, 2006). Como resultado de este análisis, los datos extraídos de una matriz ( $X$ ) con  $k$ -dimensiones, donde  $k$  corresponde al número de variables, son reducidos a un espacio definido por unos pocos componentes principales (PCs), capaces de explicar la mayor variabilidad entre las muestras ( $n$ ). Una particularidad de este método es que los PCs son ortogonales entre ellos, de manera que la representación gráfica de los datos en el espacio bidimensional definido por los dos primeros PCs ( $t_1$  y  $t_2$ ) proporciona una proyección de los datos multidimensionales conocido como gráfico de puntuaciones de las muestras en función de los componentes principales (*score plot*), en el que pueden apreciarse las agrupaciones y tendencias entre las muestras ( $n$ ). De manera análoga, en el gráfico de las puntuaciones de las variables en función de los componentes principales (*loading plot*),  $p_1$  y  $p_2$  definen la relación entre las  $k$  variables que integran la matriz de datos original ( $X$ ) con las mismas direcciones que las del *score plot*, lo que permite identificar fácilmente, de forma gráfica, cuáles son las variables (*buckets*) responsables de las tendencias observadas entre las distintas muestras. El PCA es el método de elección para obtener una visión global de la variabilidad entre muestras, y su correspondiente *score plot* es el gráfico generalmente utilizado para la identificación de posibles tendencias o agrupaciones que se den dentro de los grupos de muestras y que sean debidas a otras variables, distintas de la variable de estudio (p. ej., edad, sexo, estilo de vida, etc.). Este tipo de análisis es también de gran utilidad para la identificación de *outliers*, muestras que se desvían significativamente del perfil que muestran, de forma global, el resto de muestras incluidas dentro de los grupos del estudio. Otro gráfico muy utilizado

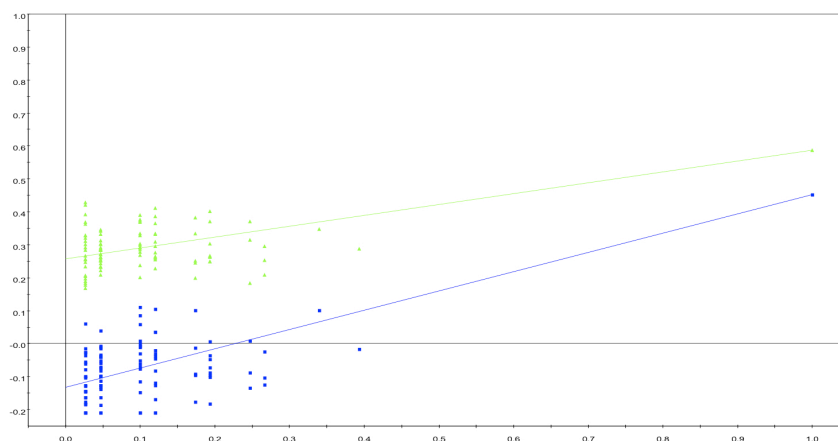
para la identificación de *outliers* dentro de los grupos incluidos en el estudio es el gráfico  $T^2$  de *Hotelling*. En este tipo de gráficos se pueden identificar todas las muestras que se encuentran fuera del intervalo de confianza del 95% en el modelo no supervisado.

Por otro lado, métodos como el análisis discriminante por mínimos cuadrados (PLS-DA) y su modificación con corrección ortogonal (OPLS-DA) utilizan el conocimiento previo que se tiene de las muestras para construir modelos estadísticos robustos que permitan, posteriormente, predecir la clase a la que pertenece una muestra sin necesidad de estar previamente clasificada. Se trata de métodos de regresión iterativa que modelan la relación existente entre un grupo de variables independientes y una variable dependiente. En estos modelos, se maximiza la covarianza entre dos grupos de muestras, buscando explicar la separación entre ellos a través de un número definido de variables latentes (J.C. Lindon et al., 2001). La representación gráfica de estos modelos, de manera similar al PCA, revela el grado de discriminación entre los grupos.

En el OPLS-DA, la variación sistemática encontrada en el conjunto original de los datos se estructura en dos niveles: la variación que está correlacionada con la pertenencia a una u otra clase en el modelo, y aquella que no está relacionada (ortogonal) (Rantalainen et al., 2007; Trygg & Wold, 2002). Los vectores del *score* y *loading plot* se dividen a su vez en los vectores predictivos: *tp* y *pp*, y ortogonales: *to* y *po*. El objetivo de este tipo de estudios es intentar reducir la influencia de las variables ortogonales, no relacionadas con la clasificación, para poder obtener modelos más precisos y fácilmente interpretables.

En este tipo de modelos supervisados, basados en un conocimiento previo de las propiedades de las muestras, es muy importante evaluar la calidad de los análisis realizados y la fiabilidad de los modelos obtenidos (Westerhuis et al., 2008). Existen distintas aproximaciones para la validación de los modelos de clasificación, una de ellas es el test de permutación, que consiste en el cálculo de nuevos modelos de clasificación obtenidos tras una asignación aleatoria de las variables de clasificación entre las distintas muestras del modelo. A partir de los datos obtenidos, se calcula la bondad de ajuste ( $R^2$ ) y la capacidad predictiva ( $Q^2$ ) de los modelos generados de forma aleatoria, siendo el resultado esperado que la capacidad predictiva de los modelos aleatorios sea significativamente inferior a la obtenida para el modelo original. Para el análisis de los resultados del test de permutación,

se representan los valores de  $R^2$  y  $Q^2$  obtenidos en los modelos de permutación y los valores de  $R^2$  y  $Q^2$  del modelo original, y se calcula la recta de regresión (**Figura 10**). El modelo se considera correctamente validado cuando los valores de la ordenada en el origen para la recta de regresión entre los distintos valores de  $R^2$  se encuentra entre 0.3-0.4, y el de la recta de regresión de  $Q^2$  es inferior a 0.05.



**Figura 10:** Ejemplo representativo de un test de permutación de un modelo PLS-DA. En el eje horizontal se representa el valor del coeficiente de correlación entre el valor del modelo original y los valores de cada permutación. Mientras que el eje vertical representa los valores obtenidos para  $R^2(Y)$  (verde) y  $Q^2(Y)$  (azul).

Otra aproximación consiste en la denominada validación cruzada, en la que se vuelve a construir el mismo modelo, pero esta vez dejando fuera un porcentaje determinado del conjunto de muestras original. Posteriormente se calcula el error de la predicción del modelo para el grupo de muestras inicialmente excluido del modelo. Este procedimiento se repite hasta que todas las muestras han sido excluidas del modelo al menos una vez. Los parámetros obtenidos, utilizados para evaluar la fiabilidad del modelo, son los estadísticos  $R^2$  y  $Q^2$ .  $R^2$  es una medida del grado de ajuste del modelo a los datos originales y  $Q^2$ , obtenido como resultado de la validación cruzada, da idea de la capacidad predictiva del modelo. De manera análoga, en aquellos casos en los que el número de muestras disponibles para el estudio lo permite, es posible realizar un test de validación más robusto a través de la predicción sobre un conjunto de muestras de validación, obtenidas y medidas de manera independiente. En general, un modelo de clasificación en metabolómica se considera significativo cuando se obtienen valores de  $Q^2$  superiores a 0.5, siempre y cuando la diferencia entre  $R^2$  y  $Q^2$  sea de al menos 0.2.

Finalmente, otro parámetro muy utilizado para evaluar la probabilidad de que la discriminación identificada en el modelo generado sea casual, es el p-valor

calculado en el test CV-ANOVA (*ANalysis Of VAriance testing of Cross-Validated predictive residuals*), que considera que un modelo es estadísticamente significativo cuando el valor de  $p$  es inferior a 0.05.

#### **h) Análisis de los modelos estadísticos**

El objetivo último del análisis estadístico supervisado es el de evaluar la capacidad de discriminación entre los grupos de estudio e identificar las variables responsables de esas diferencias. En el caso de estudios de metabolómica por RMN, el análisis supervisado conduce a la identificación de un número reducido de *buckets* que contribuyen mayoritariamente a las diferencias existentes entre los grupos de muestras incluidos en el estudio. Estos *buckets* corresponden a regiones específicas del espectro de RMN y, a su vez, a señales de determinados metabolitos.

En este sentido, existen una serie de representaciones de los datos obtenidos que resultan muy útiles para identificar las variables más relevantes en el modelo:

- **Loading plot:** Define la relación entre las variables independientes que integran la matriz de datos original ( $X$ ) con la variable de clasificación definida en cada modelo.
- **VIP list (Variable Importance in the Projection):** Muestra cuáles son las variables (*buckets*) que más peso tienen en el modelo estadístico, asignando a cada *bucket* un valor de VIP que es proporcional a la importancia de esa variable en el modelo.
- **Contribution plot:** Diagrama de barras en el que se representan todas las variables incluidas en el modelo, en relación a su peso relativo en la discriminación entre las clases analizadas en el modelo.
- **S-plot:** Combina la covarianza encontrada para cada variable con su correlación respecto a la variable dependiente. La relevancia de una variable en la separación entre grupos será mayor cuanto mayor covarianza y correlación con la variable dependiente presente.
- **SUS-plot (Shared and Unique Structures):** Se utiliza para el análisis de las diferencias entre dos modelos supervisados. Detecta cuáles son las diferencias entre las distintas clases de cada modelo que son únicas y cuáles están presentes en ambos modelos.

El análisis de la información derivada de estas representaciones, unido a la



información aportada por la asignación de los metabolitos en los espectros de RMN, permite la identificación de los metabolitos más representativos en cada modelo estadístico, que son los responsables de las diferencias entre grupos.

### **i) Cuantificación de los metabolitos y análisis estadístico univariante**

Una vez identificadas las señales más relevantes en la discriminación entre grupos y los metabolitos responsables de las diferencias, se realiza de nuevo una integración de los espectros de RMN, pero esta vez dirigida específicamente a las señales identificadas. A partir de las intensidades obtenidas para cada señal se comprueba, mediante la prueba de *Kolmogorov-Smirnov*, si la mayoría de las variables siguen una distribución normal. En caso positivo, y para confirmar la importancia estadística univariante de las variables identificadas, se utiliza la prueba paramétrica *Student* para la comparación de dos medias independientes. En los casos en los que la mayoría de las variables no sigan una distribución normal, se emplea una prueba no paramétrica para la comparación de dos medias independientes, la prueba de *Mann-Whitney*. El objetivo de este análisis es poder descartar cualquier variable que, de manera individual, no contribuya de manera significativa a la discriminación entre grupos.

### **j) Regresión logística**

Entre las regiones identificadas como relevantes para la discriminación, a partir del modelo multivariante, pueden existir: i) regiones del espectro que, presentando desplazamientos químicos distintos, correspondan al mismo metabolito, y ii) regiones que correspondan a metabolitos distintos pero cuyos niveles en suero varían de manera conjunta (p.ej., metabolitos implicados en las mismas rutas metabólicas). Por tanto, en los casos en que las regiones identificadas como relevantes son numerosas, y con el fin de simplificar el modelo de predicción, el análisis de regresión logística permite identificar entre todas las variables seleccionadas, cuáles aportan información esencial y cuáles están aportando información redundante.

El modelo de regresión logística explica la variabilidad de una variable dicotómica que sólo puede tomar los valores comprendidos entre 0 y 1 (p.ej., 0=control y 1=paciente), a través de una expresión matemática que contempla la existencia de un número de variables independientes o explicativas y que se

representa matemáticamente según la fórmula para la función de probabilidad:

$$f(p) = \frac{e^Z}{1+e^Z}$$

siendo Z una expresión matemática del sumatorio de los productos de cada una de las variables independientes ( $X_i$ ) por un coeficiente ( $\beta_i$ ), más una constante ( $\beta_0$ ), que se calcula según la fórmula:

$$Z = \beta_0 + \beta_1 \times \text{metabolito}_1 + \beta_2 \times \text{metabolito}_2 + \dots + \beta_i \times \text{metabolito}_i$$

El modelo matemático asigna un valor a los distintos coeficientes  $\beta_i$  con el fin de ajustar la función de probabilidad de presentar el evento al valor real que presenta la variable resultado en la muestra de estudio.

Siendo  $\beta$  el elemento más importante en la regresión logística, habitualmente para la interpretación de los resultados en el campo de la Biomedicina se utiliza el valor de la *odds ratio* (OR) como medida de la relación entre la variable de estudio y la intensidad de la variable explicativa. El valor de OR se calcula a partir del coeficiente de la regresión ( $\beta$ ):

$$OR_i = \text{EXP}(\beta_i)$$

Por tanto, OR podrá tomar valores desde 0 a infinito. En función del valor de  $\beta$ :

$\beta = 0$        $OR = 1$       indica que no existe relación entre las dos variables

$\beta > 0$        $OR > 1$       indica que existe una relación positiva entre las dos variables

$\beta < 0$        $OR < 1$       indica que existe una relación negativa entre las dos variables

Dado el rango asimétrico de valores que puede tomar el OR (de cero a infinito, con el uno como punto de bifurcación) y a fin de poder comparar en una misma escala los valores de OR entre las distintas variables, se utiliza para la interpretación el valor de OR en el caso de las variables con  $OR > 1$  y  $1/OR$  en las variables para las que  $OR < 1$ . De esta manera, una OR de 0.1 será equivalente a 10, 0.01 será equivalente a 100, y así sucesivamente. Estos valores serán equivalentes en cuanto a la magnitud de la relación, pero opuestos en cuanto al sentido de la relación.

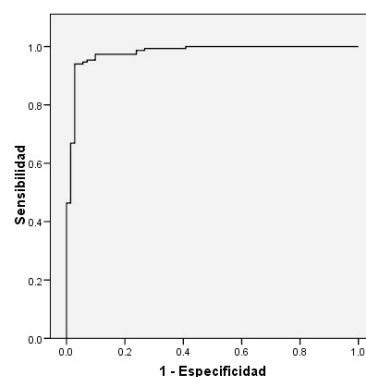
El análisis de regresión logística se puede realizar de manera univariante, teniendo en cuenta una única variable explicativa, o multivariante, cuando introducimos más de una variable explicativa en el análisis. En este último caso, a través de la regresión logística multivariante, se obtendrá un modelo de regresión logística en el que únicamente se verán representadas las variables explicativas esenciales con sus correspondientes OR ajustados, calculados en presencia de las demás variables de la ecuación.

### k) Valoración de la capacidad predictiva

El cálculo de las curvas ROC es de gran utilidad a la hora de valorar la capacidad predictiva. En este tipo de gráficas (**Figura 11**), se representan los pares de valores, (1-especificidad) y sensibilidad, obtenidos al considerar todos los posibles “puntos de corte”, permitiendo determinar:

i) El valor del punto de corte óptimo que, de forma general, corresponde al valor para el que la suma de la sensibilidad y especificidad alcanza un valor máximo.

ii) La capacidad de predicción del modelo, proporcional al valor del área bajo la curva (AUC). El AUC corresponde a la probabilidad de clasificar correctamente un par de muestras seleccionadas al azar y puede tomar valores entre 1 y 0.5, siendo 0.5 un valor de predicción nulo. Se considera un valor de AUC aceptable a partir de 0.75 (Fan et al., 2006).



**Figura 11:** Curva ROC.

Para evaluar la capacidad predictiva de un modelo se utilizan, además del valor de AUC, distintos parámetros entre los que se encuentran:

- **Sensibilidad:** Capacidad de predecir como positivos los casos realmente positivos. Corresponde a la proporción de enfermos correctamente

clasificados.

- **Especificidad:** Capacidad de predecir como negativos los casos realmente negativos. Corresponde a la proporción de individuos sanos correctamente clasificados.

En la **Figura 12** se incluyen las fórmulas utilizadas para calcular estos valores.

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \qquad \text{Especificidad} = \frac{VN}{VN + FP}$$

**Figura 12:** Cálculo de los parámetros habitualmente utilizados para la valoración de la capacidad predictiva de un método diagnóstico (VP: verdaderos positivos, VN: verdaderos negativos, FP: falsos positivos, FN: falsos negativos).

### l) Interpretación biológica de los resultados

Una vez identificados los metabolitos que juegan un papel importante en la separación entre grupos, y confirmada su relevancia estadística, la última etapa en un estudio de metabolómica consiste en la interpretación biológica de las variaciones encontradas entre los distintos perfiles metabólicos. Existen diversas bases de datos con información sobre rutas metabólicas relevantes, en las que se detallan los metabolitos que intervienen en ellas. Una de las bases de datos más útil en metabolómica es la HMDB, también utilizada para la asignación de los metabolitos en los espectros de RMN. Otras bases de datos específicas que aportan información complementaria para la interpretación de los resultados incluyen la *ConsensusPathDB-human* (Kamburov et al., 2009, 2011), y KEGG pathway analysis database (Kanehisa, 2013).

De forma complementaria, también es necesaria la búsqueda en la literatura de información relacionada que permita situar los metabolitos encontrados en el contexto adecuado.

## 5- APLICACIONES DE LA METABOLÓMICA POR RMN EN ONCOLOGÍA

La oncología es una de las áreas terapéuticas que está atrayendo más interés desde el punto de vista de la aplicación de la metabolómica al análisis de biofluidos y tejidos. Entre otros estudios, esta técnica se ha empleado en la identificación de nuevos biomarcadores de diagnóstico temprano en cáncer renal (Kim et al., 2009; Kind et al., 2007), cáncer colorectal (Ritchie et al., 2010), cáncer de páncreas (Urayama et al., 2010), leucemia (MacIntyre et al., 2010), cáncer de ovario (Guan et al., 2009; Odunsi et al., 2005) y cáncer oral (Tiziani et al., 2009). Más recientemente, han empezado a publicarse estudios que describen la utilidad de la metabolómica en la identificación/desarrollo de biomarcadores asociados al pronóstico, respuesta a tratamiento o reacciones adversas en distintos procesos patológicos (Bertini et al., 2012; Tomita & Kami, 2012).

Los estudios incluidos en esta Tesis Doctoral están orientados a la identificación de biomarcadores oncológicos de utilidad clínica en Mieloma Múltiple (MM) y Cáncer de Pulmón (CP). En los siguientes apartados se explica brevemente la problemática asociada a cada una de estas enfermedades y el enfoque de los proyectos planteados.

### 5.1- Mieloma Múltiple

MM es una neoplasia hematológica caracterizada por una proliferación anormal de células plasmáticas malignas en la médula ósea y por la presencia de una proteína monoclonal (proteína M) en el suero y/u orina. En función del tipo de inmunoglobulina monoclonal, se puede clasificar en: IgG (50-60%), IgA (30%), IgM (2%), IgD (0.5%), y excepcionalmente IgE. En el 15% de los casos estas inmunoglobulinas están formadas por cadenas ligeras que, al ser filtradas a través del riñón, se detectan sólo en orina (Mieloma de Bence-Jones). Se trata de la segunda enfermedad hematológica más frecuente, por detrás del linfoma no-Hodgking, con una incidencia anual de 4-5 nuevos pacientes por cada 100.000 personas.

A menudo, los pacientes con MM no presentan síntomas específicos de la enfermedad hasta no haber alcanzado un estado avanzado de la enfermedad. En la mayoría de los casos, la progresión de la enfermedad va desde un estadio premaligno asintomático, conocido como Gammapatía Monoclonal de Significación Indeterminada (MGUS), que progresa a MM latente y, finalmente, en un porcentaje

del 1% anual, a MM sintomático (Landgren et al., 2011). Existe un conocimiento muy limitado sobre los posibles biomarcadores moleculares y/o bioquímicos característicos de los pacientes con MGUS que presentan un elevado riesgo de progresar a MM. La ausencia o poca especificidad de los síntomas, unida a la elevada incidencia de MGUS entre la población mayor de 50 años (Wadhera & Rajkumar, 2010), hace que sea muy complicado el diagnóstico de la enfermedad en estadios iniciales.

En los últimos años se han logrado grandes avances relacionados con esta enfermedad, que abarcan desde un conocimiento básico de su etiología hasta nuevos procedimientos de diagnóstico, clasificaciones de los pacientes en función del pronóstico, y nuevas aproximaciones al tratamiento. La administración de terapias en altas dosis, seguida del trasplante de células madre (HDT-ASCT), unido a la introducción de nuevos agentes, tales como los inmunomoduladores (talidomida y lenalidomida) y los inhibidores del proteosoma (bortezomib), han contribuido a mejorar significativamente la esperanza de vida de los pacientes con MM.

A pesar de todos estos avances, MM sigue siendo considerada una enfermedad incurable en la mayoría de los casos, ya que únicamente una minoría de pacientes responde de manera completa a los tratamientos administrados (Fonseca et al., 2009). Esto hace que la evolución de los pacientes esté habitualmente caracterizada por etapas de remisión de la enfermedad (remisión completa, RC), seguida por recaídas, hasta alcanzar una situación en la que dejan de responder a los tratamientos (enfermedad refractaria). Así pues, el desarrollo de nuevas herramientas que permitan un diagnóstico temprano de la enfermedad, una evaluación del pronóstico y una monitorización de la evolución de cada paciente, así como una evaluación de la eficacia de los tratamientos administrados sería de gran valor en el tratamiento de los pacientes con MM.

## **5.2- Cáncer de pulmón**

El cáncer de pulmón representa aproximadamente el 12% de todos los casos de cáncer, con una incidencia de más de un millón y medio de casos nuevos al año en todo el mundo (Jemal et al., 2011). A pesar de todos los avances, tanto en el diagnóstico precoz como en el tratamiento, continúa siendo una enfermedad con mal pronóstico en la mayoría de los casos.

En función de las características anatomopatológicas, el cáncer de pulmón se clasifica en cuatro tipos histológicos distintos: el carcinoma microcítico de pulmón o de célula pequeña (CMP) (20% de los casos), el carcinoma epidermoide (30%), el adenocarcinoma (35-40%) y el carcinoma de células grandes (alrededor de un 10%). Desde un punto de vista clínico, los distintos tipos histológicos pueden clasificarse como: carcinoma microcítico de pulmón, sensible a los agentes citotóxicos y no tratable quirúrgicamente; y cáncer de pulmón no microcítico o de célula no pequeña (CPNM), que engloba a los otros tres subgrupos histológicos, y cuyo principal tratamiento es la cirugía si no existe diseminación (Travis, 2002). Para proceder al diagnóstico completo es necesario además la determinación del estadio en función de la clasificación TNM (Greene, F. L., Page D.L, Fleming I.D, Fritz A.G, Balch C.M, Haller D.G, 2002). La determinación de estas variables en cada paciente proporciona además una importante información pronóstica y terapéutica.

La detección del cáncer de pulmón, actualmente basada en la aparición de síntomas como tos, dolor torácico, hemoptisis o disnea, se produce en la mayoría de los casos en fases avanzadas, cuando ya no es posible la intervención curativa (cirugía) y las tasas de supervivencia a los cinco años se ven reducidas hasta 2-5% (Jantus-lewintre et al., 2012).

En los últimos años, se han llevado a cabo diversos ensayos controlados para evaluar la eficacia de la citología de esputo, radiografía de tórax y tomografía computarizada (TC) a dosis bajas en el cribado de cáncer de pulmón. Dados los resultados no concluyentes obtenidos en estos estudios, existe un interés creciente en el desarrollo de modelos de predicción de riesgo que incorporen biomarcadores genéticos y/o moleculares mediante el uso de técnicas no-invasivas o mínimamente invasivas tales como sangre, aire exhalado o esputo (Jantus-lewintre et al., 2012). La identificación de nuevos biomarcadores característicos de la enfermedad y el desarrollo de nuevas técnicas diagnósticas que permitieran la detección temprana de la enfermedad, así como una mejor evaluación del riesgo de cada paciente, podrían traducirse en un aumento significativo de la esperanza de vida de estos pacientes.

## **6- OBJETIVOS**

Dadas las ventajas que ofrece la metabolómica por RMN en la obtención de un conocimiento más detallado de los mecanismos moleculares que subyacen a los procesos patológicos, y considerando la problemática particular de cada una de las patologías expuestas en los anteriores apartados, la finalidad de este trabajo de investigación es evaluar las posibilidades de la metabolómica en la identificación de biomarcadores oncológicos de utilidad clínica en Mieloma Múltiple y Cáncer de Pulmón.

Objetivos concretos:

### **MIELOMA MÚLTIPLE**

a.- Evaluar la existencia de perfiles metabólicos específicos asociados a pacientes con MM en el momento del diagnóstico y tras la remisión completa de la enfermedad, y comparados con el correspondiente a un conjunto de individuos sanos.

b.- Profundizar en los mecanismos moleculares ligados a la aparición de MM y a la respuesta al tratamiento en pacientes de MM.

c.- Caracterizar los cambios metabólicos producidos por el tratamiento y discriminar los distintos tipos de efectos producidos, como consecuencia de éste, en los perfiles metabólicos de los pacientes con MM.

### **CÁNCER DE PULMÓN**

d.- Evaluar las diferencias en los perfiles metabólicos de pacientes con CPNM en relación al de individuos sanos, y las posibles variaciones asociadas con características clínicas de los pacientes.

e.- Comparar las diferencias encontradas entre los perfiles metabólicos de individuos sanos en relación al de pacientes con CPNM en distintos estadios de la enfermedad.

f.- Caracterizar, a través de la comparación de los perfiles metabólicos observados en los distintos estadios de la enfermedad, los mecanismos moleculares asociados a la progresión de la enfermedad.



## **II. METODOLOGÍA**



## 1.- DISEÑO DE LOS ESTUDIOS

### 1.1- Mieloma Múltiple

Las muestras de MM se obtuvieron de pacientes incluidos en dos ensayos clínicos llevados a cabo por el Grupo Español de Mieloma (Mateos et al., 2010):

- El ensayo clínico en fase III GEM2005 < 65y para pacientes jóvenes que recibieron de manera aleatoria tratamiento de inducción con VBMCP (vincristina, carmustina, melfalán, ciclofosfamida, y prednisona) / VBAD (vincristina, carmustina, doxorubicina, y dexametasona) con bortezomib en los dos últimos ciclos, o talidomida / dexametasona o bortezomib / talidomida / dexametasona, seguido de tratamiento con dosis elevadas y trasplante autólogo de células madre (ClinicalTrials.gov, NCT00461747).
- El ensayo clínico en fase III GEM 2005 > 65y para pacientes de edad avanzada, en el que el tratamiento consistía en bortezomib con melfalán / prednisona o talidomida / prednisona (ClinicalTrials.gov, NCT00443235).

Ambos ensayos clínicos incluyeron pacientes con un primer diagnóstico de MM (definido por niveles en suero de proteína monoclonal superiores a 10 g/L, o superiores a 0.2 g diarios en orina), sintomático, y no tratado con anterioridad. Para este estudio se seleccionaron muestras de suero de pacientes con MM, recogidas en el momento del diagnóstico y una vez alcanzada la RC, definida como ausencia de proteína monoclonal determinada por inmunofijación y menos del 5% de células plasmáticas en la médula ósea.

En el estudio se incluyó también un grupo de muestras de suero de individuos sanos, seleccionados para conseguir una distribución de edades y sexos similar a la del grupo de pacientes de MM estudiado. Los pacientes incluidos en este grupo procedieron de voluntarios sanos de la Clínica Universidad de Navarra.

### 1.2- Cáncer de Pulmón

En el estudio del perfil metabólico de pacientes con CPNM se incluyeron un total de 143 muestras de suero al diagnóstico procedentes de pacientes con CPNM, 70 pertenecientes a pacientes en estadios avanzados y 72 de pacientes en estadios tempranos de la enfermedad. En función del estadio de la enfermedad (Greene, F. L., Page D.L, Fleming I.D, Fritz A.G, Balch C.M, Haller D.G, 2002), las muestras se

clasificaron en dos grupos con las siguientes características:

- CPNM estadios avanzados: Pacientes procedentes de un estudio multicéntrico coordinado por el Grupo español de Cáncer de Pulmón (GECP) (Iranzo et al., 2011; Jantus-Lewintre et al., 2011). Los criterios de selección fueron el diagnóstico de CPNM en estadio avanzado (IIIB con derrame pleural o IV) y no haber recibido quimioterapia previa. A lo largo del ensayo, las respuestas de los pacientes a los distintos tratamientos fueron categorizadas siguiendo los criterios RECIST (Therasse et al., 2000).
- CPNM estadios tempranos: Pacientes con CPNM resecable (estadios IA-IIIB) procedentes del Consorcio Hospital General Universitario de Valencia (CHGUV). Los criterios de selección fueron el diagnóstico de CPNM en estadio temprano y no haber recibido quimioterapia previa a la cirugía.

De forma paralela, se incluyó en el estudio un grupo control de 74 individuos sanos anónimos procedente de la colección de la Red de Biobancos de la Comunidad Valenciana. La selección de muestras se realizó con la finalidad de conseguir un grupo control con una distribución similar a la de los grupos de pacientes en cuanto a criterios de sexo, edad y hábito tabáquico.

## **2.- RECOGIDA DE MUESTRAS**

La recogida de muestras para todos los estudios se realizó tras la obtención del consentimiento informado de los pacientes, de conformidad con los principios fundamentales establecidos en la Declaración de Helsinki y con la aprobación de los Comités Éticos de Investigación de los centros que participaron en los estudios.

Para el procesamiento de las muestras de suero se siguieron los protocolos estandarizados que se emplean habitualmente en todos los estudios de metabolómica y que incluyen: i) la recogida de 1-2mL sangre en un tubo libre de anticoagulante o de cualquier otro aditivo (BD Vacutainer, sin aditivos), hasta completar el vacío, ii) el mantenimiento del tubo en hielo o, en su defecto, en nevera 30 minutos hasta su coagulación, iii) la centrifugación a 1.600 g durante 15 minutos, 4°C, iv) la recogida cuidadosa la parte superior del sobrenadante, evitando que la fracción celular se resuspenda y v) la congelación rápida en nitrógeno líquido y su conservación a -80°C hasta su análisis por RMN.

### 3.- PREPARACIÓN DE MUESTRAS PARA RMN

En el momento de realizar las medidas de RMN, las muestras fueron descongeladas en hielo. Posteriormente, se añadió a cada muestra de 500  $\mu\text{L}$ , un volumen equivalente de una solución tampón (140 mM  $\text{Na}_2\text{HPO}_4$ , 5 mM TSP, 0.04%  $\text{NaN}_3$ , pH 7.4) preparada con un 10% de  $\text{D}_2\text{O}$ . De la mezcla resultante, 550  $\mu\text{L}$  de cada muestra fueron transferidos a un tubo de RMN de 5mm de diámetro para su análisis.

### 4.- ADQUISICIÓN DE ESPECTROS DE RMN

Los experimentos de RMN se adquirieron utilizando un espectrómetro Bruker Avance II 600 MHz equipado con una criosonda TCI de 5mm. Tanto la adquisición como el procesado de estos experimentos se realizó con el programa TOPSPIN 3.0 (Bruker Biospin, Rheinstetten, Alemania).

Siguiendo los protocolos recomendados por Bruker para estudios de metabolómica con muestras biológicas, y antes de proceder a la adquisición de cada conjunto de experimentos, se procedió a la optimización y calibración de todos los parámetros críticos en este tipo de estudios, entre los que se encuentran la temperatura, los shims, la frecuencia de la señal del agua (o1), etc.

De manera general, los protocolos de optimización comprenden los siguientes pasos:

- Calibración de la temperatura con una muestra de referencia de metanol deuterado (99.8%). Secuencia de pulso: zg30.
- Optimización de los shims y de la frecuencia de la señal del agua (o1) con una muestra patrón de sacarosa (2 mM sacarosa en 90%  $\text{H}_2\text{O}$  + 10%  $\text{D}_2\text{O}$ ). Secuencia de pulso: zgpr.
- Optimización de los shims y de la frecuencia de la señal del agua (o1) para una muestra representativa del estudio (50% suero + 50% solución tampón). Secuencia de pulso: zgpr.
- Comprobación del ajuste de los parámetros en un espectro 1D-NOESY de la muestra representativa del estudio y evaluación de las siguientes características:
  - Línea base plana
  - Buena corrección de la fase

- Ajuste de los shims: Señal simétrica del TSP, con señales satélite pronunciadas, y una anchura de señal inferior a 1Hz.
- Buena supresión de la señal del agua: La señal residual del agua debe ser significativamente menor que la obtenida en el correspondiente espectro zgpr, con una anchura de línea inferior a 100 Hz.

En todos los casos, la adquisición de los espectros de RMN se realizó de manera conjunta, y de forma aleatoria, para los distintos grupos de muestras incluidos en cada estudio. Los experimentos de RMN se realizaron a 310K y el pulso de pre-saturación que se aplicó para mejorar la supresión de la señal del agua fue de 25 Hz. Estos dos parámetros se mantuvieron constantes para todas las muestras de cada uno de los estudios. Únicamente se ajustó de manera individual para cada muestra la impedancia de la sonda, utilizando la rutina de oscilación *wobble*, y la duración del pulso de 90° (p1).

El número de repeticiones necesario para conseguir una adecuada relación señal/ruido (S/N) en los experimentos de RMN fue optimizado con una muestra representativa de cada grupo de estudio.

Para todas las muestras incluidas en los diferentes estudios se adquirió un experimento **1D-CPMG**:

Secuencia de pulso: cpmgpr1d.comp

AUNM: au\_prof

AUNMP: proc\_profcpmg

TD (<sup>1</sup>H): 73728

NS: 64

SW: 20 ppm

d1: 4 s

d20: 0.3 ms

L4: 128

RG: 80.6

Adicionalmente, para algunas de las muestras se adquirieron otros experimentos complementarios:

**1D-NOESY:**

Secuencia de pulso: noesygppr1d.comp

AUNM: au\_prof1d

AUNMP: proc\_plasma1d

TD (<sup>1</sup>H): 98304

NS: 64

SW: 30 ppm

d1: 4 s

d8: 10 ms

d16: 0.2 ms

RG: 80.6

**Difusión:**

Secuencia de pulso: ledbgppr2s1d.comp

AUNM: au\_prof

AUNMP: proc\_profdif

TD (<sup>1</sup>H): 98304

NS: 32

SW: 30 ppm

d1: 4 s

d16: 0.2 ms

d20: 0.12 ms

d21: 5ms

RG: 80.6

**J-resuelto:**

Secuencia de pulso: jresgpprqf

AUNM: au\_prof

AUNMP: proc\_jres

TD F2 ( $^1\text{H}$ ): 10240

TD F1 ( $^1\text{H}$ ): 40

NS: 2

SW: 14/0.13 ppm

d1: 2 s

d16: 0.2 ms

RG: 80.6

**$^1\text{H}$ - $^1\text{H}$  TOCSY:**

Secuencia de pulso: dipsi2gpqh19

TD F2 ( $^1\text{H}$ ): 2048

TD F1 ( $^1\text{H}$ ): 256

NS: 56

SW: 12/12 ppm

d1: 1.5 s

d9: 60 ms

**$^1\text{H}$ - $^{13}\text{C}$  HSQC:**

Secuencia de pulso: hsqcetgpprsisp2.2

TD F2 ( $^1\text{H}$ ): 2048

TD F1 ( $^{13}\text{C}$ ): 256

NS: 128

SW: 16/195 ppm

d1: 1.5 s

d16: 0.2 ms



## 5.- PRE-PROCESADO DE LOS ESPECTROS DE RMN

Para el procesado de los datos se utilizó un filtro de apodización (ensanchamiento de señal, lb) de 1 Hz. La corrección de la fase y la línea base de todos los experimentos monodimensionales se realizó de manera automática empleando el comando “apk0.noe”.

Por otro lado, cada espectro se referenció internamente tomando como referencia para los valores de desplazamiento químico la señal correspondiente al grupo metilo de la alanina (1.47 ppm).

## 6.- ASIGNACIÓN DE LOS ESPECTROS

La asignación de las señales de los espectros se basó en la información disponible en bases de datos especializadas (HMDB, BMRB, PRIME, etc.) combinada con la de la base de datos comercial BBIORFCODE (Bruker BioSpin, Rheinstetten, Alemania). Esta información fue complementada con la disponible en la literatura sobre la asignación de metabolitos en muestras de suero (Koppelaar D; Lindon JC; Tranter, 2010; MacIntyre et al., 2010; J K Nicholson et al., 1995). De forma complementaria, se adquirieron espectros de RMN de compuestos patrón que facilitaron la confirmación de algunas asignaciones.

## 7.- PROCESADO DE DATOS

Los experimentos de RMN que se emplearon para la obtención de los modelos estadísticos multivariantes fueron, en todos los casos, los correspondientes a los <sup>1</sup>H-RMN CPMG. La integración de los espectros se realizó utilizando el programa AMIX (Analysis of MIXtures, Bruker BioSpin, Rheinstetten, Alemania).

Los parámetros utilizados para la integración de los espectros en regiones equidistantes (*buckets*) fueron:

- **Mieloma Múltiple:**

Método integración:	Suma de Intensidades
Región del espectro:	0.26 – 10.5 ppm
Regiones excluidas:	4.35 – 5.12 ppm (agua) 5.65 – 5.85 ppm (urea).

Tamaño buckets: 0.04 ppm  
Normalización: Sin normalizar

• **Cáncer de Pulmón:**

Método integración: Suma de Intensidades Absolutas

Región del espectro: 0.14 – 9.02 ppm

Regiones excluidas: 4.30 – 5.06 ppm (agua)

5.60 – 5.85 ppm (urea).

Tamaño buckets: 0.01 ppm

Normalización: Área total

## 8.- ANÁLISIS ESTADÍSTICO MULTIVARIANTE

El análisis estadístico multivariante se realizó empleando SIMCA-P 12.0 (Umetrics AB, Suecia). El tipo de escalado empleado en cada estudio se ajustó a los datos obtenidos en los espectros de RMN con el fin de obtener los modelos estadísticos que mejor se ajustaran a los grupos de muestras analizados y a las existentes entre los distintos grupos. Se optó por utilizar Varianza Unitaria en el caso de MM y Pareto en el estudio de CPNM.

## 9.- CUANTIFICACIÓN DE LOS METABOLITOS RELEVANTES

La integración de las señales correspondientes a los metabolitos de interés en cada modelo se realizó, al igual que la integración de los *buckets*, utilizando el programa AMIX (Analysis of MIXtures, Bruker BioSpin, Rheinstetten, Alemania). La intensidad de las señales integradas se normalizó, o no, respecto al área total del espectro, en función del procedimiento seguido en cuanto a la normalización utilizada en la segmentación del espectro (*bucketing*) con regiones equidistantes.

## 10.- ANÁLISIS ESTADÍSTICO UNIVARIANTE

Los análisis estadísticos univariantes se realizaron con el paquete estadístico SPSS 15.0, considerando un valor de  $p < 0.05$  como estadísticamente significativo.

En función del objetivo del análisis, se realizaron los siguientes test estadísticos:

- Prueba de *Kolmogorov-Smirnov*: Evaluación de la normalidad de los

datos obtenidos para cada metabolito tras la integración de los espectros de  $^1\text{H}$ -RMN.

- Prueba de *Student* (paramétrica): Comparación de medias en grupos independientes de muestras, en aquellos casos en que los datos se ajustaron a una distribución normal.
- Prueba de *Mann Whitney* (no paramétrica): Comparación de medias en grupos independientes de muestras, en aquellos casos en que los datos no se ajustaron a una distribución normal.
- Correlación de *Pearson*: Medida del grado de relación lineal entre dos variables cuantitativas.
- Regresión logística univariante (método *Introducir*): Obtención de los coeficientes de ponderación de los metabolitos de interés en la ecuación de probabilidad univariante.
- Regresión logística multivariante (método *Adelante:RV*): Obtención de los coeficientes de ponderación ajustados de los metabolitos de interés en la ecuación de probabilidad multivariante.
- Curvas ROC: Evaluación de la capacidad de predicción, punto de corte óptimo y los valores de sensibilidad y especificidad asociados.



### **III. RESULTADOS Y DESARROLLO ARGUMENTAL**



## 1.- MIELOMA MÚLTIPLE

A pesar de los esfuerzos destinados a intentar conocer mejor tanto la patogénesis de la enfermedad como los mecanismos de respuesta a los nuevos fármacos empleados para el tratamiento del MM, la comprensión detallada de estos procesos sigue siendo un reto en este área (Bhattacharyya et al., 2006; Rajpal et al., 2011; Terpos et al., 2005). En este sentido, este trabajo representa el primer estudio, empleando la metabolómica mediante  $^1\text{H}$ -RMN de alta resolución, dirigido a la caracterización del perfil metabólico en suero de pacientes con MM. Esta aproximación experimental permite el análisis integrado de múltiples variables (metabolitos) en un grupo de muestras, facilitando así una caracterización global del perfil metabólico en suero de pacientes con MM. Los resultados obtenidos demuestran que el perfil de los pacientes con MM representa una importante fuente de información sobre los mecanismos moleculares característicos de la enfermedad, que podría ser utilizada para la identificación de biomarcadores específicos, relacionados con la respuesta a los tratamientos.

### 1.1.- Análisis y asignación de los espectros

El estudio incluyó el análisis por RMN de muestras de suero de 27 pacientes con MM recogidas en el momento del diagnóstico, 23 muestras de pacientes con MM recogidas tras la remisión completa de la enfermedad, y 31 voluntarios sanos (Tabla 7).

**Tabla 7:** Muestras incluidas en el estudio de metabolómica en MM

Grupo	n	Sexo	Edad $\pm$ s.e.m. <sup>a</sup>	Proteína M
Control	31	H = 16 M = 15	67.1 $\pm$ 1.6	-
Diagnosis	27	H = 14 M = 13	63.6 $\pm$ 2.2	IgG = 13 IgA = 9 Bence-Jones = 5
Remisión	23	H = 12 M = 11	63.7 $\pm$ 2.3	-

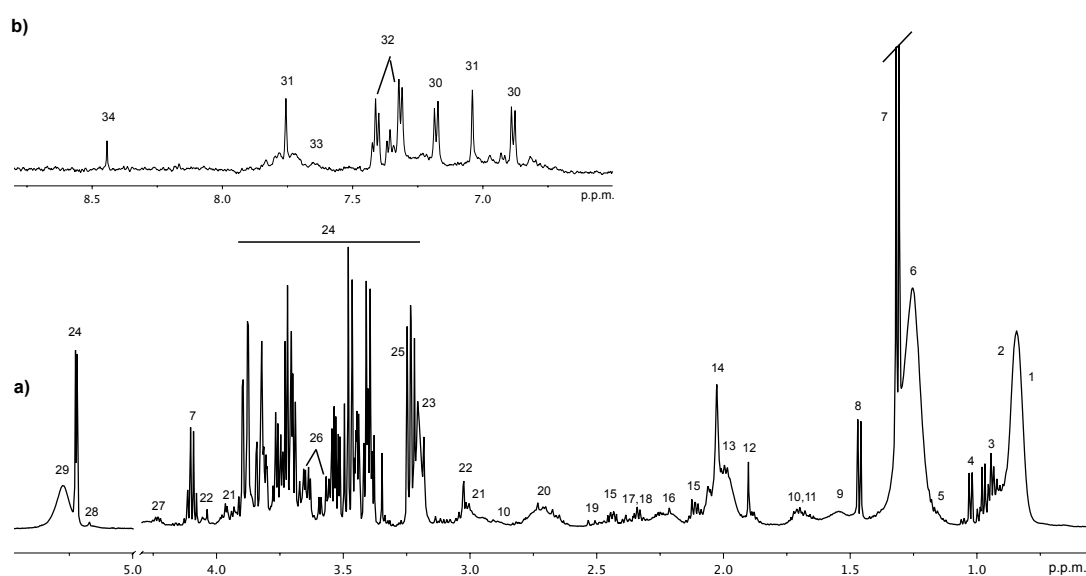
n, número de muestras

s.e.m., error estándar de la media

<sup>a</sup> Edad media en el momento de la recogida de la muestra (para el grupo control) y en el momento del diagnóstico (para el grupo a diagnóstico y en remisión completa)  $\pm$  s.e.m.

Se obtuvieron espectros de buena calidad tanto para el grupo control, como para las muestras de pacientes con MM en el momento del diagnóstico y tras la remisión

completa. En la **Figura 13** se muestra un espectro  $^1\text{H}$ -RMN CPMG típico para un paciente con MM en el momento del diagnóstico. Este tipo de espectros están dominados por señales correspondientes a metabolitos de bajo peso molecular. Las señales identificadas corresponden a compuestos presentes típicamente en muestras de suero, incluyendo varios aminoácidos (p.ej., lisina, arginina, glutamina, fenilalanina, histidina, tirosina, alanina), glucosa y otros azúcares, creatina, creatinina, ácidos orgánicos (acetato, lactato, citrato, formato, etc.). También se pueden observar otras señales correspondientes a metabolitos de alto peso molecular: protones de los grupos acetilo de la cadena lipídica de lipoproteínas, ácidos grasos insaturados y glicoproteínas.



**Figura 13:** Asignación de las señales más representativas en un espectro  $^1\text{H}$ -RMN CPMG, adquirido a 600MHz, para una muestra de un paciente con MM en el momento del diagnóstico, medida a 310K. a) Región alifática del espectro ( $\delta=5.50\text{-}0.5$  ppm) y b) región aromática ( $\delta=9.00\text{-}6.50$  ppm). Asignación de los picos: 1, colesterol ( $\text{C}_{18}\text{-CH}_3$ ); 2, lípidos ( $\text{-CH}_3$ ) (principalmente LDL/VLDL); 3, leucina/isoleucina; 4, valina; 5, 3-hidroxibutirato; 6, lípidos  $(\text{CH}_2)_n$  (principalmente LDL/VLDL); 7, lactato; 8, alanina; 9, ácido adípico; 10, lisina; 11, arginina; 12, acetato; 13, lípidos ( $\text{CH}_2\text{-C=C}$ ); 14, N-acetil cisteína (grupos acetilo de las glicoproteínas); 15, glutamina; 16, lípidos ( $\text{CH}_2\text{-CO}$ ); 17, glutamato; 18, piruvato; 19, citrato; 20, lípidos ( $\text{CH=CH-CH}_2\text{-CH=CH-}$ ); 21, creatina; 22, creatinina; 23, colina; 24, glucosa; 25, N-óxido de trimetilamina; 26, glicerol; 27, treonina; 28, glicerol de los lípidos; 29, lípidos ( $\text{-CH=CH-}$ ); 30, tirosina; 31, histidina; 32, fenilalanina; 33, 3-metilhistidina; 34, formato.

## 1.2.- Análisis no supervisado

En el análisis no supervisado de los datos se busca, en primer lugar, poder identificar y estudiar cualquier muestra que presente un comportamiento anómalo en relación al de su grupo (posibles *outliers*). Una vez identificadas estas muestras y tras decidir, de manera justificada, su inclusión/exclusión en el estudio, se estudia

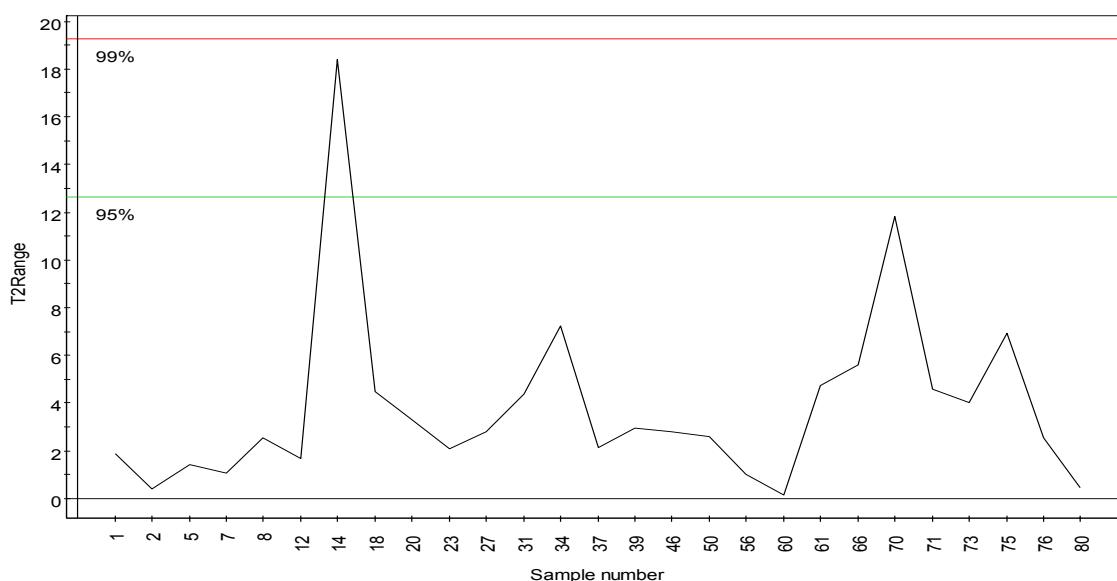


la existencia de posibles tendencias o agrupaciones que puedan existir entre las muestras, y que sean debidas a otras variables distintas de la del objeto del estudio (control/diagnóstico/remisión).

La identificación y análisis de este tipo de muestras y/o tendencias es muy importante en los estudios de metabolómica ya que podrían representar fuentes de heterogeneidad no deseadas, pudiendo así interferir en los resultados del mismo. Por esta razón, este análisis se lleva a cabo al inicio del estudio, previo al análisis supervisado de los resultados, en el que se tendrá en cuenta toda la información disponible.

### a) Identificación de *outliers*

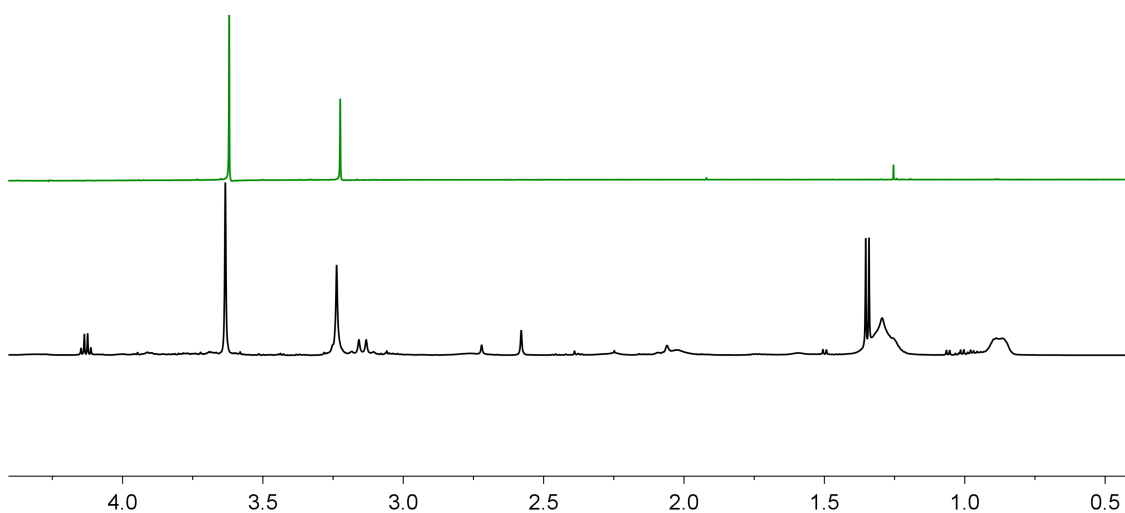
Para la identificación de *outliers*, a partir del análisis de componentes principales (PCA) de cada grupo de muestras, se analizaron los *score plot* y el gráfico  $T^2$  de *Hotelling* en cada caso. En el grupo de muestras de pacientes con MM en el momento del diagnóstico, se identificó la presencia de una muestra (#14) fuera del intervalo de confianza del 95% en el gráfico  $T^2$  de *Hotelling*, que mostraba un comportamiento estadísticamente distinto al del resto de su grupo (**Figura 14**).



**Figura 14:** Gráfico  $T^2$  de *Hotelling* derivado del análisis PCA de los espectros  $^1\text{H-RMN}$  CPMG correspondientes a las muestras de suero de pacientes con MM recogidas en el momento del diagnóstico. Los límites marcados en rojo y verde denotan los intervalos de confianza del 99% y 95%, respectivamente. La muestra #14 fue identificada como un *outlier* extremo dentro de su grupo.

Con objeto de entender el comportamiento anómalo de esta muestra, se realizó una revisión exhaustiva de los datos incluidos en la historia clínica de este paciente, no encontrándose ninguna variación biológica, en relación con la del resto de su

grupo, que pudiera explicar una desviación en su comportamiento estadístico. La inspección visual de la muestra tampoco mostró ninguna coloración rosa, que pudiera indicar la presencia de hemólisis. La revisión del espectro de  $^1\text{H}$ -RMN CPMG de esa muestra, reveló dos señales intensas, a 3.20 y 3.60 ppm, que no se encontraban presentes en el resto de espectros del grupo. Con el fin de determinar si el origen de esas señales era un error en la recogida de la muestra, se adquirieron espectros de  $^1\text{H}$ -RMN de muestras de tampón fosfato salino (PBS), preparadas empleando distintos tubos comerciales utilizados habitualmente en el ámbito hospitalario para la recogida de muestras de sangre. Tal y como se muestra en la **Figura 15**, a través de este análisis se pudo determinar que la presencia de las señales intensas en el espectro de la muestra #14 se debía a que ésta había sido recogida en un tubo "BD Vacutainer® K3E 7.2 mg" (ref. 368860). En concreto, ambas señales son compatibles con la presencia de EDTA en la muestra, procedente del tubo de recogida.

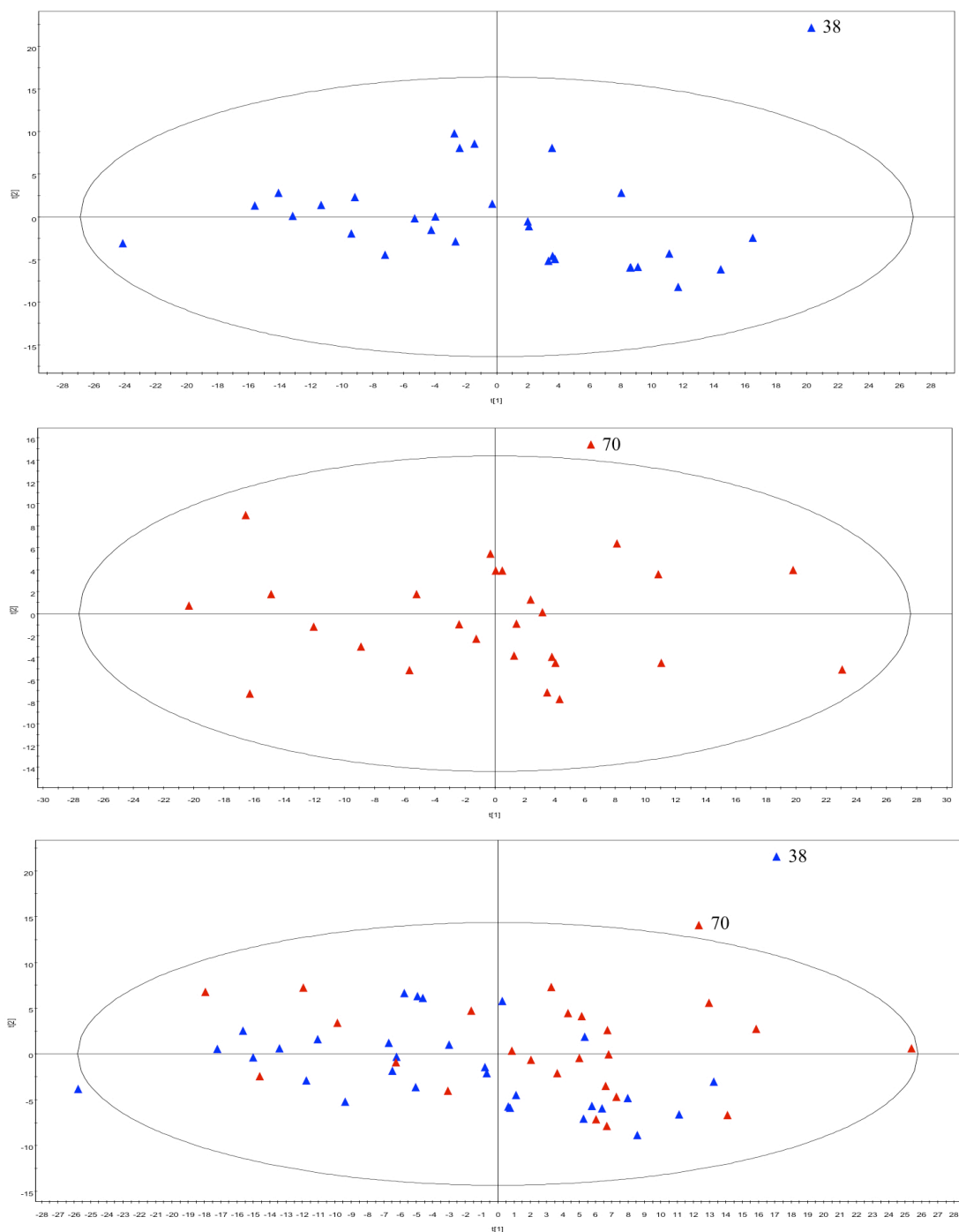


**Figura 15:** Comparación del espectro de la muestra identificada como *outlier* en el grupo de pacientes con MM en el momento del diagnóstico (#14, negro) con el correspondiente a la muestra de PBS recogida en un tubo "BD Vacutainer® K3E 7.2 mg" tubes (ref 368860, verde). Como se puede apreciar, las señales intensas situadas a 3.20 y 3.60 ppm, responsables del comportamiento anómalo de la muestra, están presentes en ambos espectros.

De acuerdo a estos datos, la muestra #14, a diferencia de las demás muestras incluidas en el estudio (muestras de suero), habría sido recogida en un tubo que contenía EDTA, habitualmente empleado para la recogida de muestras de plasma. Al margen de las perturbaciones espectrales causadas por la presencia de las señales intensas correspondientes al EDTA, existen varios estudios que señalan diferencias significativas entre los niveles de distintos metabolitos cuando se comparan muestras equivalentes de suero y plasma de un mismo paciente (Dettmer

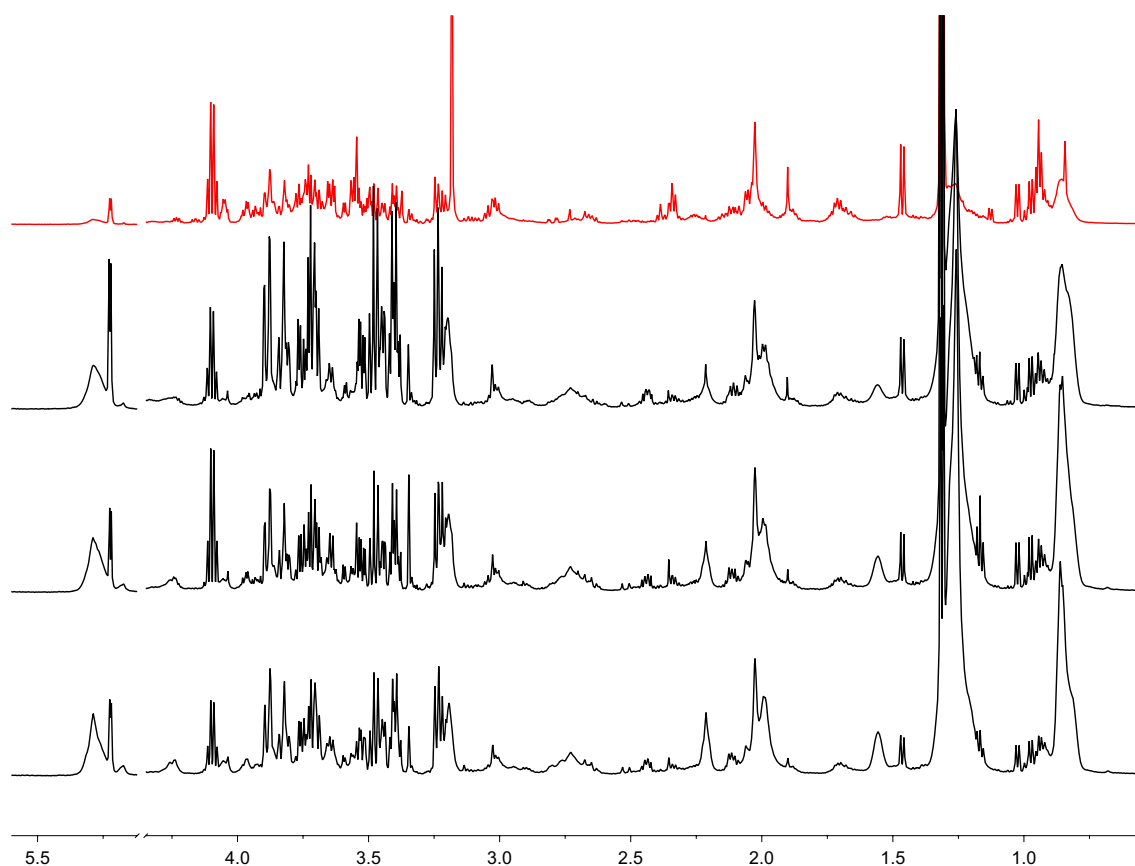
et al., 2010; Teahan et al., 2011; Wedge et al., 2011; Yu et al., 2011). Por esta razón se decidió excluir la muestra #14 del estudio. En general, los procedimientos que se siguen para la recogida de muestras en los estudios de metabolómica son los mismos que los utilizados en la práctica clínica diaria. Éste es un ejemplo de cómo, a través del análisis no supervisado previo de los datos, es posible detectar la aparición de determinados contaminantes y/o errores que pueden haber surgido durante la recogida/procesamiento de las muestras.

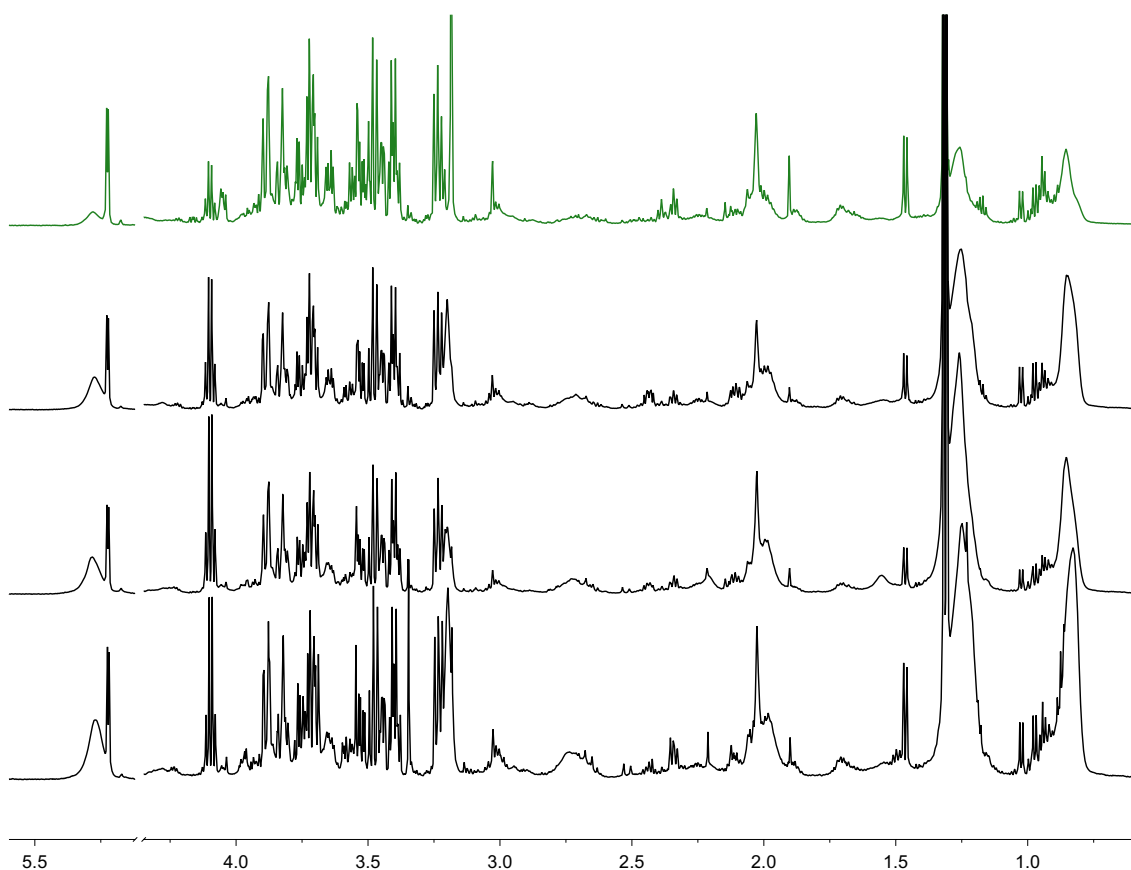
En el curso de este mismo análisis, una muestra en el grupo control (#38) y otra del grupo de pacientes con MM en el diagnóstico (#70) fueron identificadas como *outliers* dentro de sus grupos, estando fuera del intervalo de confianza del 95% en el *score plot* derivado del análisis PCA de cada grupo. Además, ambas muestras continuaban comportándose como *outliers* cuando el análisis se llevaba a cabo un análisis de forma conjunta de las muestras del grupo control y del grupo de pacientes con MM en el momento del diagnóstico (**Figura 16**).



**Figura 16:** *Score plots* derivados del análisis PCA de los espectros  $^1\text{H}$ -RMN CPMG de muestras del grupo control (azul), muestras del grupo de pacientes con MM en el momento del diagnóstico (rojo) y muestras del grupo control y pacientes con MM al diagnóstico de forma conjunta (azul y rojo, respectivamente). En todos ellos se puede apreciar cómo las muestras #38 (control) y #70 (diagnóstico) se sitúan fuera de la elipse que marca el 95% del intervalo de confianza en los modelos estadísticos generados, siendo identificadas ambas muestras como *outliers*. En el último plot, hay otra muestra perteneciente al grupo control (muestra #25, cuadrante inferior izquierdo) que también se sitúa fuera de la elipse. Esta muestra no fue identificada como *outlier* dentro de su propio grupo (en el primer gráfico de la figura).

Estos resultados indican que ambas muestras (#38 y #70), siendo *outliers* en sus correspondientes grupos, comparten ciertas características que hacen que tengan un comportamiento estadístico similar en el análisis PCA, cuando se analizan los dos grupos de forma conjunta. El análisis visual de los espectros correspondientes a estas muestras reveló que ambas muestras, comparados con los de su grupo, presentaban un perfil metabólico caracterizado por cambios en señales correspondientes a lípidos y lipoproteínas, así como señales en la región del grupo  $-N(CH_3)_3^+$  de la colina, acetato, lactato, alanina, y otros (**Figura 17**).





**Figura 17:** Comparación de los espectros  $^1\text{H}$ -RMN CPMG de la muestra #38 (rojo) con la de otras muestras representativas de su grupo (grupo control, negro), y de la muestra #70 (verde) con la de otras muestras representativas de su grupo (grupo de pacientes con MM en el momento del diagnóstico, negro).

El análisis de estos perfiles metabólicos y su comparación con los resultados obtenidos en otros estudios previos (Brindle et al., 2003; Griffin et al., 2011) reveló su relación con un cuadro de hipertensión. Tras la revisión detallada de la historia clínica del paciente con MM, se encontró que éste era diabético e hipertenso, coincidiendo con los datos apuntados por el espectro de  $^1\text{H}$ -RMN, razón por la que se decidió excluir justificadamente estas dos muestras del estudio.

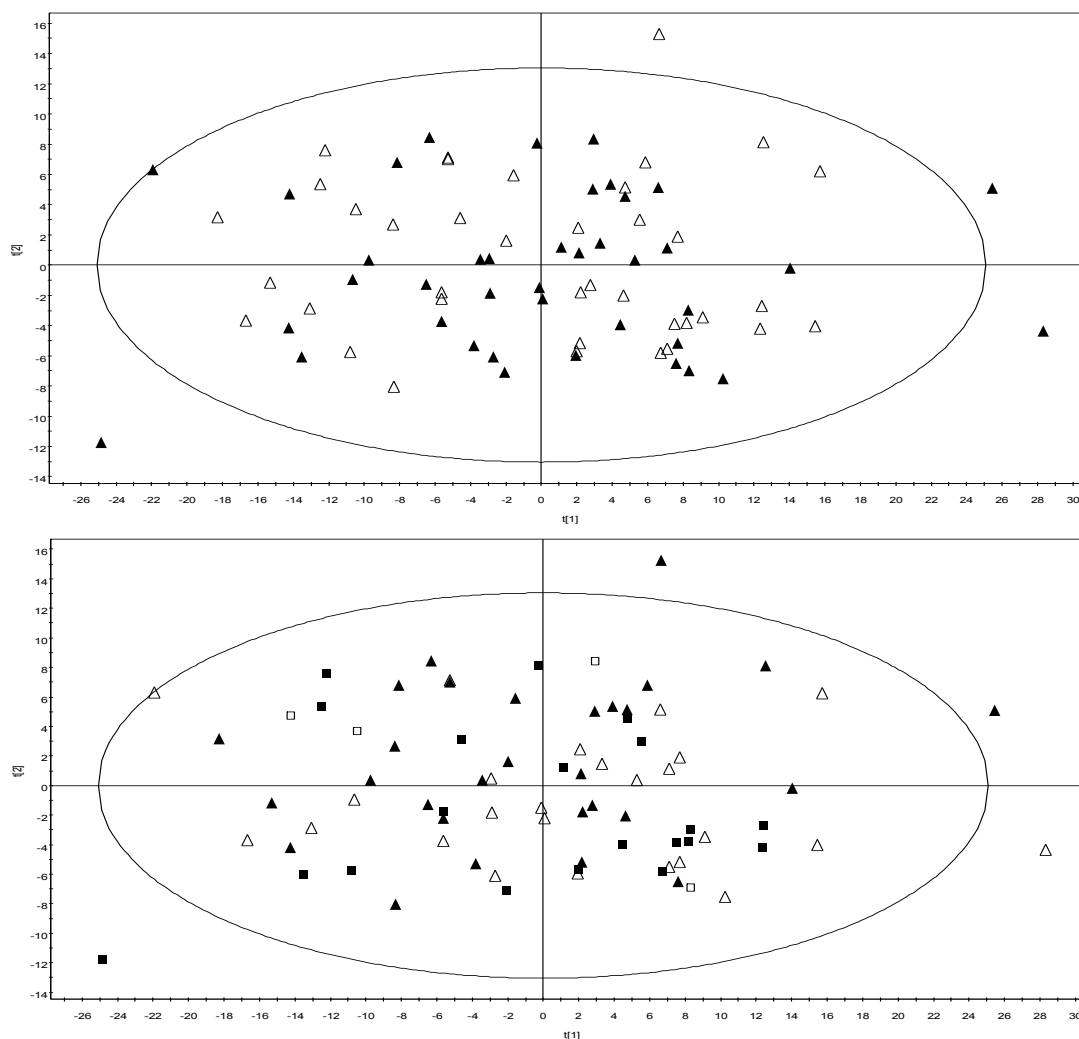
La estrategia que se siguió para la identificación y la decisión de incluir/excluir estas muestras del análisis posterior de este estudio es una estrategia comúnmente utilizada en los estudios de metabolómica (Duarte et al., 2010; Patel et al., 2012; Tea et al., 2012; Tredwell et al., 2011; Triba et al., 2010).

### **b) Identificación de otras agrupaciones**

Tras la exclusión justificada del estudio de las tres muestras identificadas como *outliers* dentro de sus grupos (muestras #14 y #70 en el grupo de pacientes con MM en el momento del diagnóstico, y muestra #38 en el grupo control), se prosiguió con

el análisis no supervisado de todas las muestras y de forma conjunta, con el fin de identificar posibles agrupaciones dentro de los grupos debidas a otras variables.

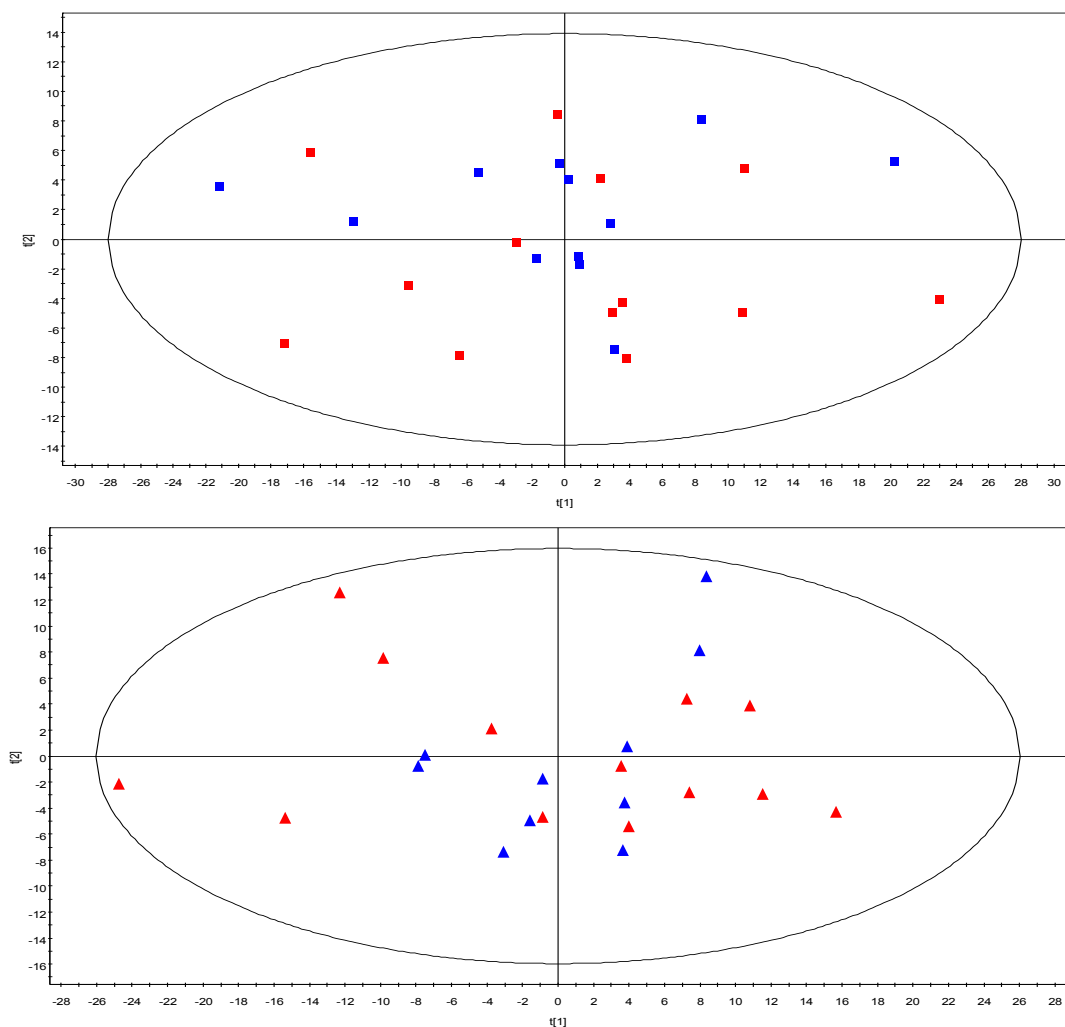
El análisis PCA de todas las muestras incluidas en el estudio se empleó para evaluar si existía alguna tendencia debida a la distribución de edades y/o sexos (**Figura 18**). Como se puede apreciar, no se observa que la distribución de edades y sexo pueda influir en la homogeneidad de las muestras incluidas en el estudio.



**Figura 18:** Análisis PCA de los espectros  $^1\text{H}$ -RMN CPMG de todas las muestras incluidas en el estudio según el sexo (arriba; triángulos llenos: hombres, triángulos vacíos: mujeres), y según la edad (abajo; triángulos llenos: > 70 años, triángulos vacíos: 60 - 70 años, cuadrados llenos: 50 - 60 años, cuadrados vacíos: < 50 años).

Tal y como se ha indicado en la sección de Materiales y Métodos, las muestras de pacientes incluidos en los grupos de MM al diagnóstico y tras la remisión completa, procedían de dos ensayos clínicos distintos: GEM2005 < 65y y GEM 2005 > 65y. Se evaluó también en el análisis PCA de ambos grupos la posible influencia de esta variable en la distribución de las muestras dentro del espacio

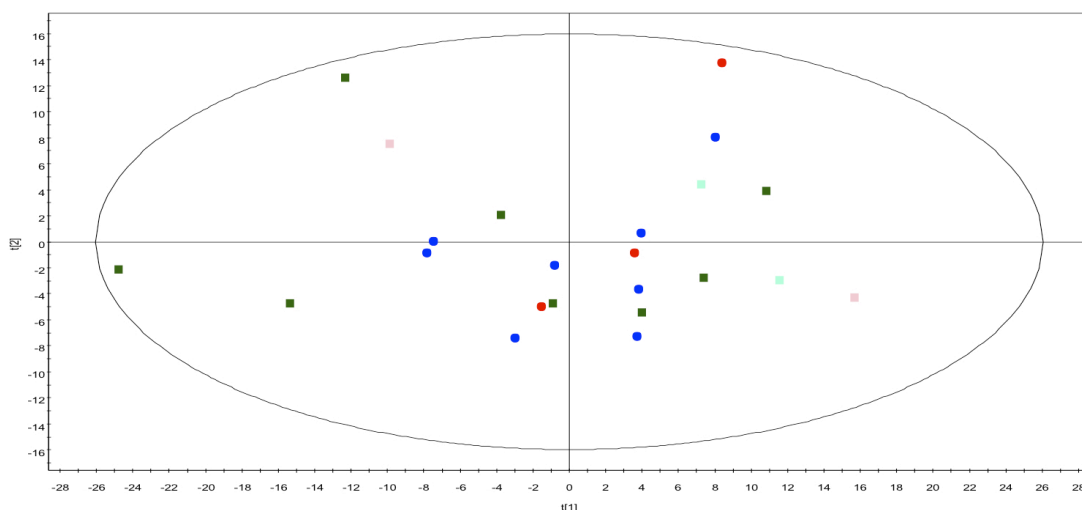
multidimensional. Como se puede apreciar en los *score plots* de la **Figura 19**, en ninguno de los dos grupos se pudieron apreciar agrupaciones debidas a la procedencia de las muestras en relación con un ensayo clínico u otro.



**Figura 19:** Análisis PCA de los espectros  $^1\text{H-RMN CPMG}$  de las muestras del grupo de pacientes con MM al diagnóstico (arriba, cuadrados) y tras la remisión completa (abajo, triángulos) coloreados según el ensayo clínico del que procedían (rojo: GEM2005 < 65y, azul: GEM2005 > 65y).

En último lugar, se realizó este mismo análisis para el grupo de pacientes con MM tras la remisión completa. Este grupo de muestras, procedentes de pacientes que habían alcanzado la remisión completa siguiendo distintos regímenes terapéuticos, fueron evaluados en función del ensayo clínico del que procedían y de la rama de tratamiento que les había sido asignada en el mismo. Tal y como se muestra en la **Figura 20** no se observó ninguna agrupación asociada al tratamiento recibido por cada paciente.





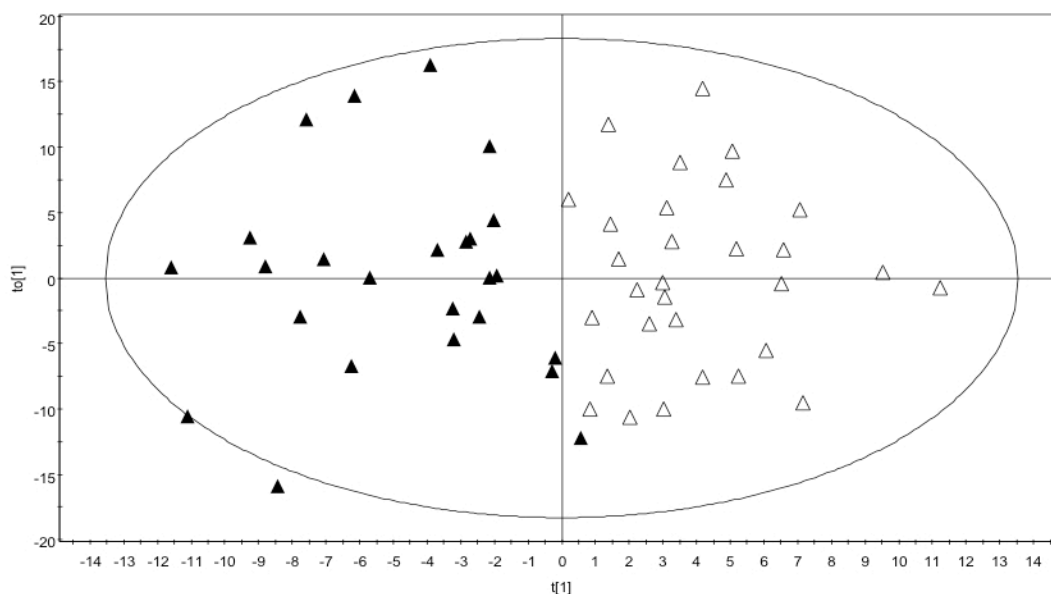
**Figura 20:** Análisis PCA de los espectros  $^1\text{H}$ -RMN CPMG de las muestras del grupo de pacientes con MM tras la remisión completa (cuadrados: GEM2005 < 65y, círculos: GEM2005 > 65y). Las muestras se han coloreado según la rama de tratamiento recibida por cada paciente durante el ensayo clínico (rojo: Melfalán/Prednisona/Velcade, 6 ciclos; azul: Melfalán/Talidomida/Velcade, 6 ciclos; rosa: quimioterapia VBAD/VMCP, 4 ciclos + Velcade/Dexametasona, 2 ciclos; verde claro: Talidomida/Dexametasona, 6 ciclos; verde oscuro: Talidomida/Dexametasona/Velcade, 6 ciclos)

### 1.3.- Análisis supervisado

Una vez estudiadas las posibles variables que podían interferir en el correcto análisis de los resultados, se procedió al análisis supervisado de los resultados de RMN. Como se ha explicado en apartados anteriores, este análisis emplea el conocimiento previo que se tiene de las muestras (la pertenencia a cada grupo del estudio) para construir modelos estadísticos (OPLS-DA) que permitan, posteriormente, identificar cuáles son las variables que mejor predicen la clase a la que pertenece cada muestra, identificando así las variables características de cada uno de los grupos de muestras.

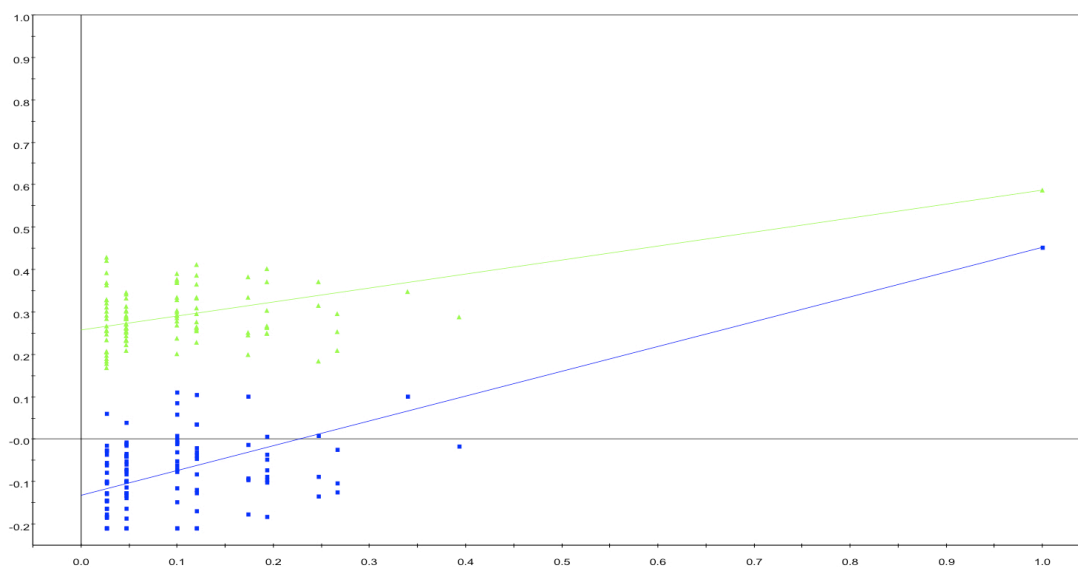
#### a) Pacientes con MM en el momento del diagnóstico vs voluntarios sanos

En el modelo OPLS-DA obtenido para las muestras de voluntarios sanos y las pertenecientes a pacientes con MM recogidas en el momento del diagnóstico, representado en la **Figura 21**, se obtuvo una buena separación entre los dos grupos de muestras ( $R^2 = 0.613$ ,  $Q^2 = 0.462$ ).



**Figura 21:** *Score plot* correspondiente al modelo OPLS-DA resultante de la comparación entre los perfiles metabólicos de los espectros  $^1\text{H}$ -RMN CPMG de los pacientes con MM en el momento del diagnóstico (triángulos llenos) y el grupo de muestras de voluntarios sanos (triángulos vacíos).

Con el fin de validar los resultados obtenidos para el modelo estadístico multivariante (OPLS-DA), se calculó el test de permutación ( $n=100$ , siendo  $n$  el número de repeticiones) para el modelo PLS-DA equivalente (**Figura 22**). Tanto los valores obtenidos para la intersección de  $R^2(Y)$  y  $Q^2(Y)$  (0.258 y -0.133, respectivamente), como el  $p$ -valor obtenido en el test ANOVA ( $p < 0.05$ ), reflejan la validez de los resultados obtenidos.



**Figura 22:** Resultado del test de permutación obtenido para el modelo PLS-DA de pacientes con MM en el momento del diagnóstico y de muestras correspondientes a voluntarios sanos. En el eje horizontal se representa el valor del coeficiente de correlación entre el valor del modelo original y los valores de cada permutación. El eje vertical representa los valores obtenidos para  $R^2(Y)$  (verde) y  $Q^2(Y)$  (azul).

Una vez evaluada la significación estadística del modelo obtenido, se procedió al análisis de la lista VIP y del *loading plot* correspondientes a este modelo estadístico, para lo cual se tuvieron en cuenta todos los *buckets* con un valor de VIP superior a uno. Este análisis permitió identificar a través del análisis de los espectros 1D-RMN y los espectros 2D de las muestras representativas de cada grupo las regiones del espectro que contribuían mayoritariamente a la separación entre los grupos (**Tabla 8**).

**Tabla 8:** Lista de los metabolitos más representativos en la discriminación en el modelo OPLS-DA obtenido para pacientes con MM en el momento del diagnóstico y voluntarios sanos.

Metabolito	$\delta$ <sup>1</sup> H (ppm) <sup>a</sup>
3-Hidroxiбутирато	1.15-1.19
Colesterol/LDL/VLDL (CH <sub>3</sub> )	0.70-0.90
Lisina	2.87-2.93
Lípidos (CH=CH)	5.24-5.37
Arginina	1.63-1.77
Fenilalanina	7.29-7.33
Lípidos (CH <sub>2</sub> -C=C)	1.93-2.01
Acetato	1.88-1.91
Lípidos (CH=CH-CH <sub>2</sub> -CH=CH)	2.60-2.84
Isoleucina	0.90-1.04
Glutamina	2.41-2.46
Tirosina	7.16-7.20
Colina	3.19-3.21
Lípidos (CH <sub>2</sub> -CO)	2.17- 2.25

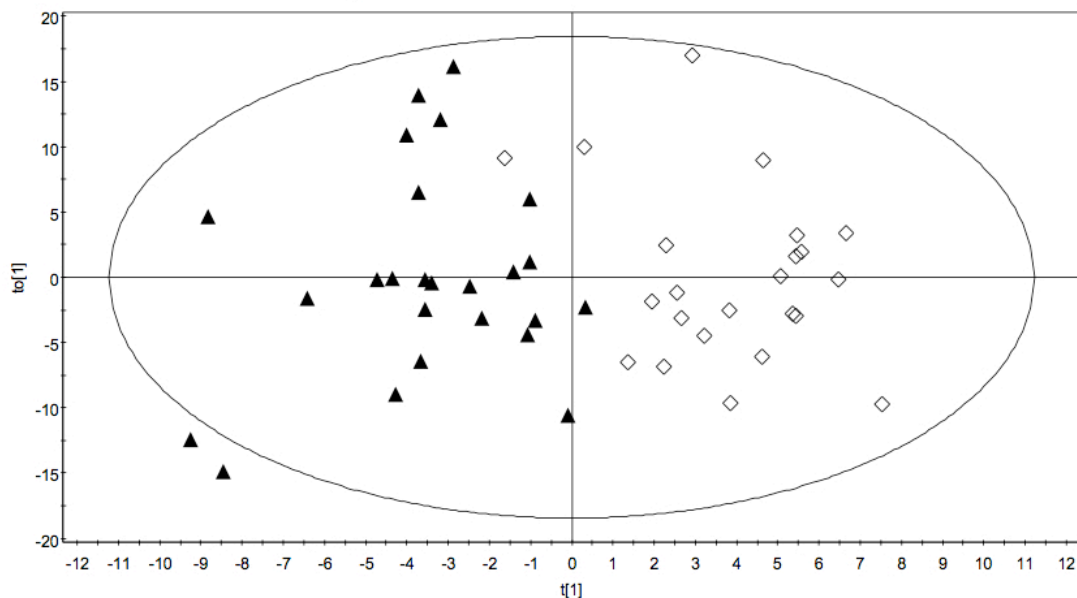
<sup>a</sup> Desplazamientos químicos ( $\delta$ ) expresados en partes por millón (ppm)

Las regiones espectrales indicadas en la **Tabla 8** se emplearon posteriormente para cuantificar de manera específica las señales correspondientes a esos metabolitos en los espectros 1D-RMN.

#### **b) Pacientes con MM en el momento del diagnóstico vs pacientes con MM tras la remisión completa**

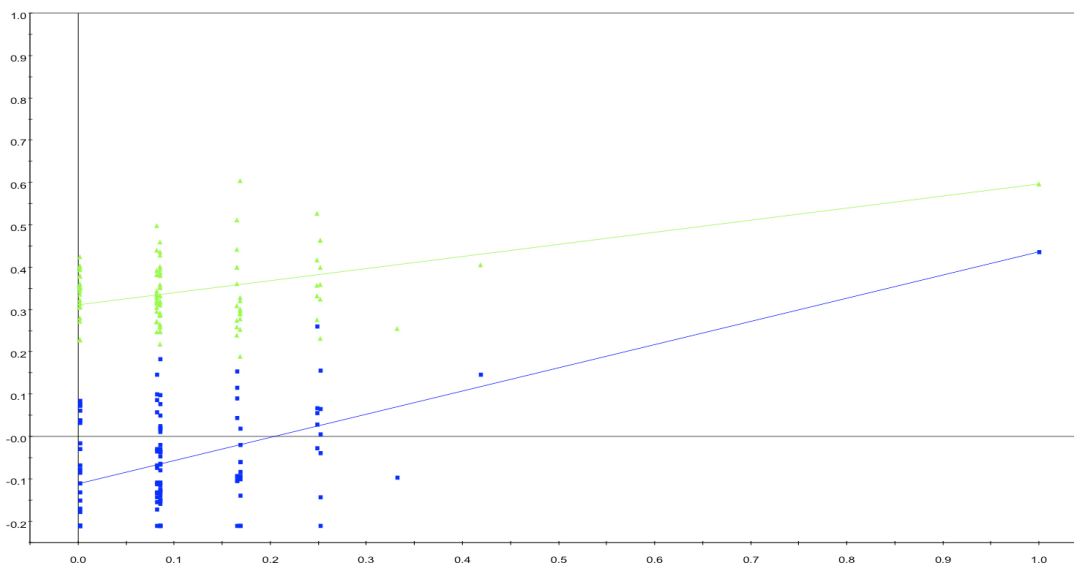
Una vez determinado que los pacientes con MM en el momento del diagnóstico presentan un perfil metabólico específico y claramente distinto del obtenido para voluntarios sanos, se llevó a cabo un análisis equivalente para los perfiles metabólicos de las muestras de pacientes con MM recogidas en el momento del diagnóstico y las obtenidas tras la remisión completa de la enfermedad (**Figura 23**).

En este caso, los valores obtenidos para el ajuste y la capacidad de predicción del modelo ( $R^2=0.553$ ,  $Q^2=0.358$ ), aún siendo inferiores a los obtenidos en la comparación entre pacientes con MM en el diagnóstico e individuos sanos, reflejan una buena separación entre los grupos.



**Figura 23:** Score plot correspondiente al modelo OPLS-DA resultante de la comparación entre los perfiles metabólicos de los espectros  $^1\text{H-RMN}$  CPMG de muestras de suero de pacientes con MM en el momento del diagnóstico (triángulos llenos) y tras la remisión completa de la enfermedad (rombos vacíos).

En este caso, los resultados obtenidos en el test de permutación ( $R^2(Y) = 0.31$  y  $Q^2(Y) = -0.112$ ) (**Figura 24**) y el test ANOVA ( $p < 0.05$ ) para el modelo PLS-DA equivalente confirmaron también la validez del modelo estadístico.



**Figura 24:** Resultados del test de permutación obtenido para el modelo PLS-DA de pacientes con MM en el momento del diagnóstico y tras la remisión completa. En el eje horizontal se representa el valor del coeficiente de correlación entre el valor del modelo

original y los valores de cada permutación. El eje vertical representa los valores obtenidos para  $R^2(Y)$  (verde) y  $Q^2(Y)$  (azul).

De igual modo a lo que se indicó en el apartado anterior, se seleccionaron las regiones del espectro que contribuían de manera más significativa a la separación de los perfiles metabólicos de estos dos grupos de muestras empleando la información obtenida del *loading plot* y la lista VIP del modelo OPLS-DA correspondiente. El listado de los metabolitos identificados y sus correspondientes regiones en el espectro 1D-RMN se muestra en la **Tabla 9**.

**Tabla 9:** Lista de los metabolitos más representativos en la discriminación en el modelo OPLS-DA obtenido para pacientes con MM en el momento del diagnóstico y tras la remisión completa.

Metabolito	$\delta$ $^1\text{H}$ (ppm) <sup>a</sup>
Lisina	2.87-2.93
Lípidos (CH=CH)	5.24-5.37
Creatinina	2.96-3.06
Citrato	2.50-2.55
Colesterol/LDL/VLDL (CH <sub>3</sub> )	0.70-0.90
Glucosa	5.21-5.24
N-óxido de trimetilamina (TMAO)	3.17-3.19
Lactato	4.13-4.18
Colina	3.19-3.21

<sup>a</sup> Desplazamientos químicos ( $\delta$ ) expresados en partes por millón (ppm)

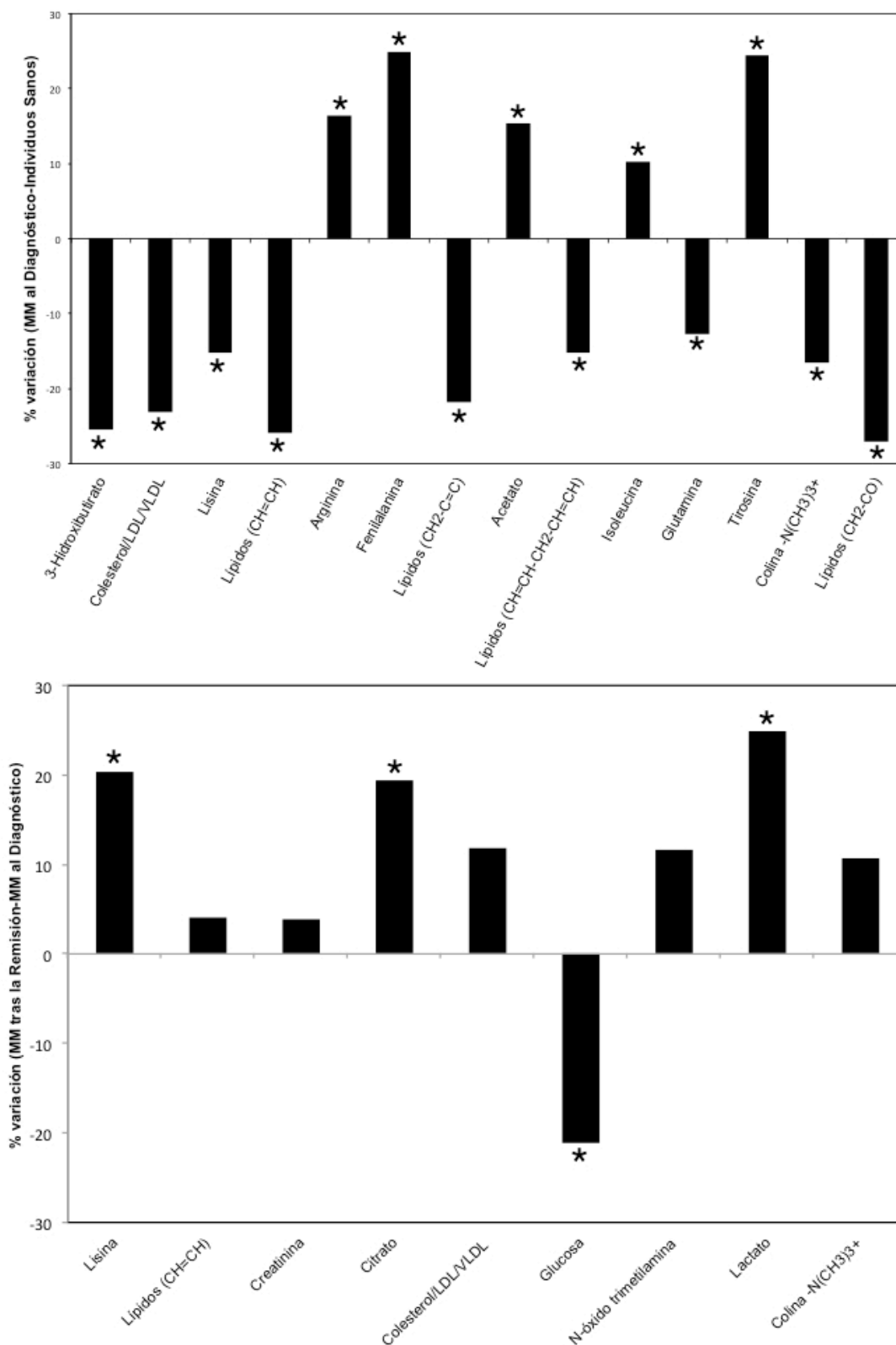
### c) Pacientes con MM tras la remisión completa vs voluntarios sanos

Finalmente, el modelo OPLS-DA correspondiente a la comparación de los perfiles metabólicos del grupo de muestras de pacientes con MM obtenidas tras la remisión completa de la enfermedad con el del grupo de muestras de voluntarios sanos mostró que la discriminación estadística entre esos dos grupos era significativamente menor ( $Q^2(Y) < 0.25$ ) que la obtenida en los modelos anteriores (pacientes con MM al diagnóstico vs voluntarios sanos, y pacientes con MM al diagnóstico vs tras la remisión completa). Es decir, a diferencia de los perfiles metabólicos de los pacientes con MM, que difieren significativamente de los correspondiente a los voluntarios sanos, los perfiles metabólicos de los pacientes con MM, una vez alcanzada la remisión completa, de la enfermedad son más parecidos a los perfiles obtenidos para voluntarios sanos.

#### 1.4.- Integración de los metabolitos más relevantes

Una vez identificadas, a partir de cada modelo estadístico OPLS-DA, las regiones del espectro  $^1\text{H-RMN CPMG}$  que más contribuían a la discriminación entre los distintos grupos de muestras incluidos en el estudio y los metabolitos a los que correspondían esas señales se procedió a la integración de esas señales.

Para cada modelo estadístico, se procedió a la integración de las señales que más contribuían a la discriminación de los grupos de muestras que se comparaban. Se calcularon las medias y valores de s.e.m para cada grupo, así como el coeficiente de variación entre ambos grupos. Con el fin de valorar la significación estadística de las variaciones encontradas entre grupos se calculó, para cada metabolito, el p-valor mediante la prueba de *Student*. En la **Figura 25** se representan gráficamente las variaciones obtenidas para los distintos metabolitos cuantificados en los dos modelos OPLS-DA.



**Figura 25:** Diferencias (%) obtenidas entre las medias calculadas para los distintos grupos de muestras incluidos en los modelos OPLS-DA. Arriba: Comparación de los perfiles metabólicos de individuos sanos frente a pacientes con MM al diagnóstico. Abajo: Comparación de los perfiles metabólicos de pacientes con MM al diagnóstico y tras la remisión completa. (\*) Diferencias estadísticamente significativas ( $p < 0.05$ ).

El análisis de estos resultados revela que el perfil metabólico en suero de pacientes con MM en el momento del diagnóstico presenta, en relación al de los individuos sanos, niveles más elevados de metabolitos como isoleucina, arginina, acetato, fenilalanina y tirosina, e inferiores de 3-hidroxiacetato, lisina, glutamina, así como algunas señales correspondientes a lípidos (**Figura 25**, panel superior).

Por otro lado, los perfiles metabólicos de las muestras obtenidas de pacientes con MM en el momento del diagnóstico y tras la remisión completa, difieren principalmente en los niveles de lactato, lisina, TMAO, colina y colesterol, los cuales se encuentran más elevados en las muestras obtenidas tras la remisión completa (**Figura 25**, panel inferior).

Los datos obtenidos reproducen las variaciones observadas en los *loading plots* de los correspondientes modelos OPLS-DA. En las páginas siguientes, se incluyen los valores obtenidos para la integración de estos metabolitos en los distintos modelos (**Tablas 10 y 11**). Por otro lado, en la **Tabla 12** se indican los valores de las intensidades medias y su error estándar (s.e.m.) obtenidos para los metabolitos que, globalmente, mostraron variaciones más significativas entre los tres grupos de muestras incluidos en el estudio. Se incluye también una representación gráfica de las variaciones de las intensidades medias de estos metabolitos entre los tres grupos de muestras (**Figura 26**).



**Tabla 10:** Cuantificación relativa de los metabolitos más significativos en el modelo estadístico OPLS-DA obtenido para la comparación de los perfiles metabólicos de individuos sanos y pacientes con MM en el momento del diagnóstico.

Metabolito	$\delta^1\text{H}$ (ppm) <sup>a</sup>	Intensidad media $\pm$ s.e.m. (unidades arbitrarias)		% variación	$t^b$
		Grupo Control	Pacientes con MM al Diagnóstico		
3-Hidroxiacetato	1.15-1.19	1.95E07 $\pm$ 6.59E05	1.45E07 $\pm$ 6.97E05	-25.45	1.97 E-06*
Colesterol/LDL/VLDL (CH <sub>3</sub> )	0.70-0.90	1.62E08 $\pm$ 7.42E06	1.24E08 $\pm$ 7.53E06	-23.18	4.00 E-04*
Lisina	2.87-2.93	3.70E06 $\pm$ 5.93E04	3.14E06 $\pm$ 1.11E05	-15.19	9.99 E-06*
Lípidos (CH=CH)	5.24-5.37	4.09E07 $\pm$ 2.74E06	3.03E07 $\pm$ 2.59E06	-25.93	4.00 E-03*
Arginina	1.63-1.77	1.18E07 $\pm$ 3.38E05	1.37E07 $\pm$ 5.64E05	16.45	2.00 E-03*
Fenilalanina	7.29-7.33	1.06E06 $\pm$ 4.65E04	1.33E06 $\pm$ 7.62E04	24.90	2.00 E-03*
Lípidos (CH <sub>2</sub> -C=C)	1.93-2.01	4.69E07 $\pm$ 2.51E06	3.67E07 $\pm$ 2.20E06	-21.84	2.00 E-03*
Acetato	1.88-1.91	4.03E06 $\pm$ 1.62E05	4.65E06 $\pm$ 2.78E05	15.39	2.50 E-02*
Lípidos (CH=CH-CH <sub>2</sub> -CH=CH)	2.60-2.84	3.22E07 $\pm$ 1.21E06	2.73E07 $\pm$ 1.52E06	-15.17	7.00 E-03*
Isoleucina	0.90-1.04	3.58E07 $\pm$ 9.47E05	3.95E07 $\pm$ 1.47E06	10.27	2.00 E-02*
Glutamina	2.41-2.46	6.76E06 $\pm$ 1.31E05	5.90E06 $\pm$ 3.57E05	-12.66	1.00 E-02*
Tirosina	7.16-7.20	5.64E05 $\pm$ 3.46E04	7.02E05 $\pm$ 7.02E05	24.40	2.20 E-02*
Colina -N(CH <sub>3</sub> ) <sub>3</sub> <sup>+</sup>	3.19-3.21	3.31E07 $\pm$ 1.60E06	2.77E07 $\pm$ 1.70E06	-16.55	1.20 E-02*
Lípidos (CH <sub>2</sub> -CO)	2.17- 2.25	1.81E07 $\pm$ 2.11E06	1.32E07 $\pm$ 1.38E06	-27.04	3.50 E-02*

\* Diferencias estadísticamente significativas ( $p < 0.05$ )

<sup>a</sup> Región del espectro de RMN utilizada para la cuantificación

<sup>b</sup>  $p$  valor calculado mediante la prueba de Student

**Tabla 11:** Cuantificación relativa de los metabolitos más significativos en el modelo estadístico OPLS-DA obtenido para la comparación de los perfiles metabólicos de pacientes con MM recogidas en el momento del diagnóstico y tras la remisión completa de la enfermedad.

Metabolito	$\delta^1\text{H}$ (ppm) <sup>a</sup>	Intensidad media $\pm$ s.e.m. (unidades arbitrarias)		% variación	$t^b$
		Pacientes con MM al Diagnóstico	Pacientes con MM tras Remisión		
Lisina	2.87-2.93	3.14E06 $\pm$ 1.11E05	3.77E06 $\pm$ 8.45E04	20.32	2.15 E-05*
Lípidos (CH=CH)	5.24-5.37	3.03E07 $\pm$ 2.59E06	3.15E07 $\pm$ 2.03E06	4.15	3.50 E-01
Creatinina	2.96-3.06	1.67E07 $\pm$ 4.41E05	1.73E07 $\pm$ 4.52E05	3.86	1.60 E-01
Citrato	2.50-2.55	1.51E06 $\pm$ 6.65E04	1.80E06 $\pm$ 1.00E05	19.33	9.00 E-003*
Colesterol/LDL/VLDL (CH <sub>3</sub> )	0.70-0.90	1.24E08 $\pm$ 7.53E06	1.39E08 $\pm$ 4.91E06	11.77	6.00 E-02
Glucosa	5.21-5.24	1.41E07 $\pm$ 1.25E06	1.11E07 $\pm$ 9.67E05	-21.17	3.40 E-02*
N-óxido trimetilamina	3.17-3.19	1.17E07 $\pm$ 1.37E06	1.31E07 $\pm$ 8.70E05	11.58	2.10 E-01
Lactato	4.13-4.18	2.01E06 $\pm$ 2.00E05	2.51E06 $\pm$ 2.04E05	24.98	4.30 E-02*
Colina -N(CH <sub>3</sub> ) <sub>3</sub> <sup>+</sup>	3.19-3.21	2.77E07 $\pm$ 1.70E06	3.06E07 $\pm$ 1.60E06	10.78	1.00 E-01

\* Diferencias estadísticamente significativas ( $p < 0.05$ )

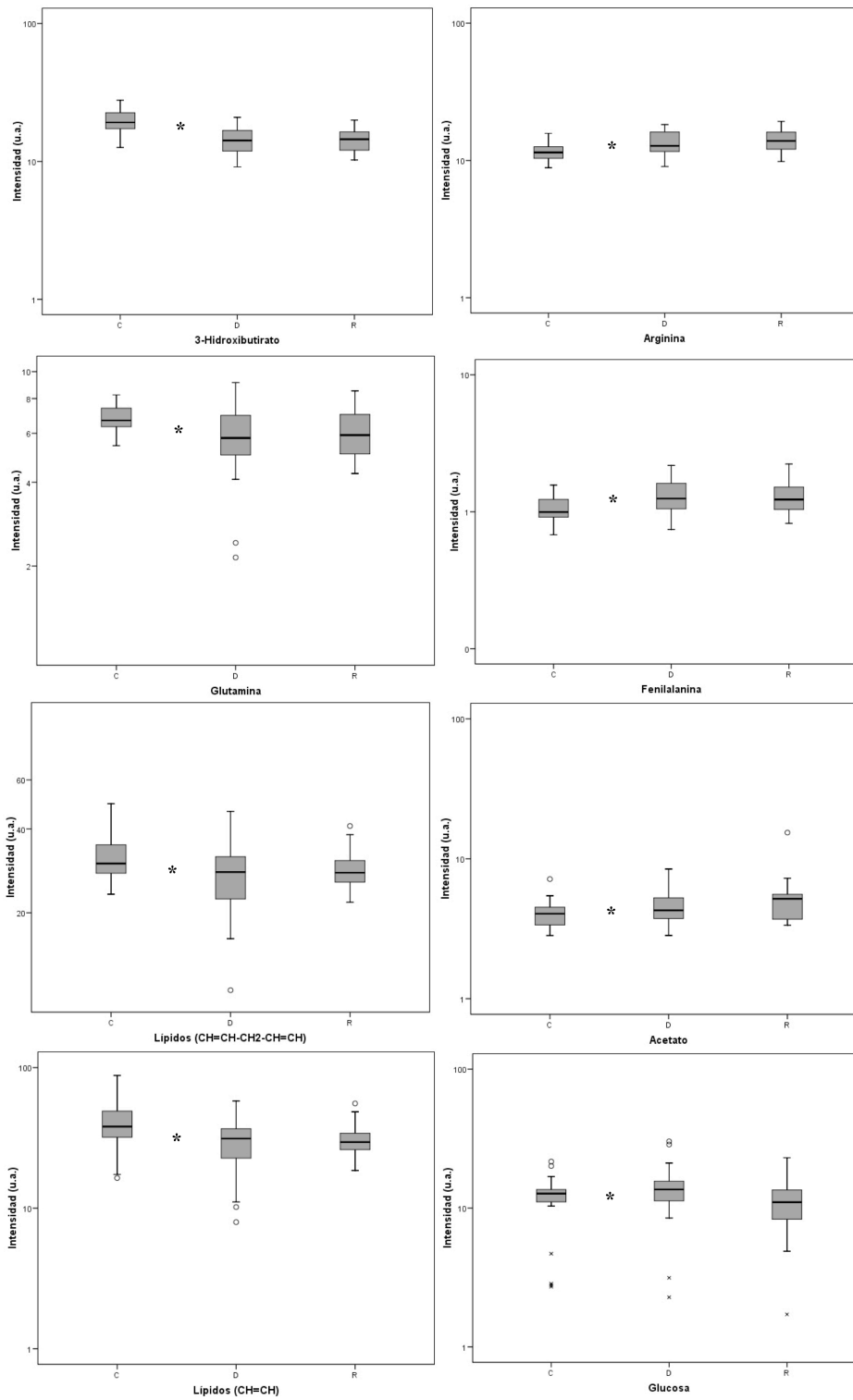
<sup>a</sup> Región del espectro de RMN utilizada para la cuantificación

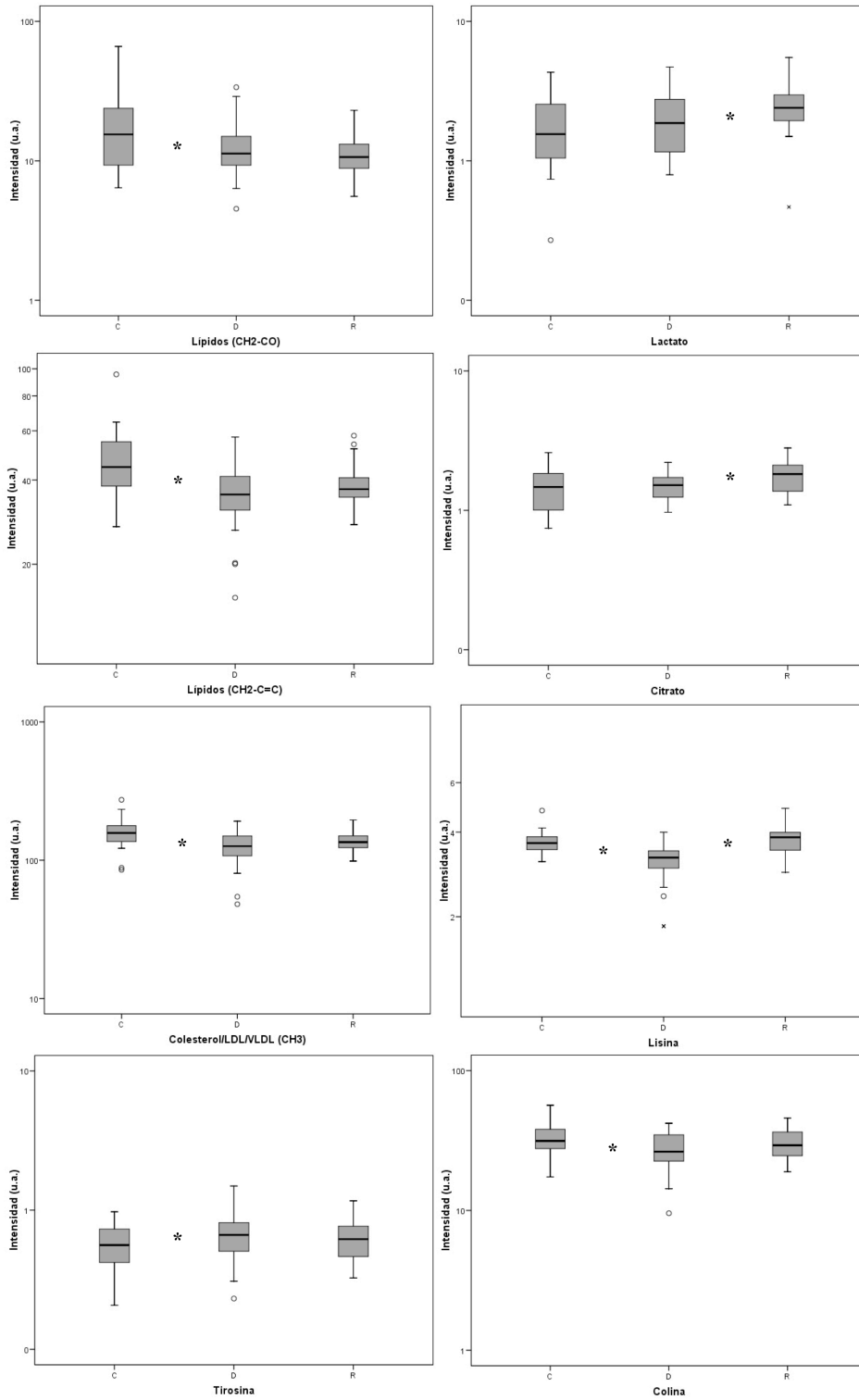
<sup>b</sup> p valor calculado mediante la prueba de Student

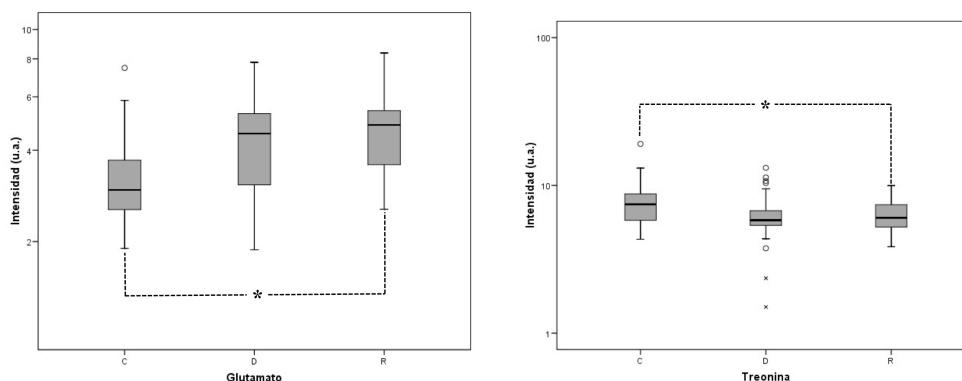
**Tabla 12:** Intensidades medias de los metabolitos que presentan variaciones más significativas en las comparaciones correspondientes entre los tres grupos incluidos en el estudio.

Metabolito	$\delta^1\text{H}$ (ppm) <sup>a</sup>	Intensidad media $\pm$ s.e.m. (unidades arbitrarias)		
		Grupo control	Pacientes con MM al Diagnóstico	Pacientes con MM tras Remisión
3-Hidroxiбутирато	1.15-1.19	1.95E07 $\pm$ 6.59E05	1.45E07 $\pm$ 6.97E05	1.44E07 $\pm$ 6.45E05
Arginina	1.63-1.77	1.18E07 $\pm$ 3.38E05	1.37E07 $\pm$ 5.64E05	1.42E07 $\pm$ 5.10E05
Glutamina	2.41-2.46	6.76E06 $\pm$ 1.31E05	5.90E06 $\pm$ 3.57E05	6.13E06 $\pm$ 2.52E05
Fenilalanina	7.29-7.33	1.06E06 $\pm$ 4.65E04	1.33E06 $\pm$ 7.62E04	1.33E06 $\pm$ 8.38E04
Lípidos (CH=CH-CH <sub>2</sub> -CH=CH)	2.60-2.84	3.22E07 $\pm$ 1.21E06	2.73E07 $\pm$ 1.52E06	2.90E07 $\pm$ 1.04E06
Acetato	1.88-1.91	4.03E06 $\pm$ 1.62E05	4.65E06 $\pm$ 2.78E05	5.28E06 $\pm$ 5.14E05
Lípidos (CH=CH)	5.24-5.37	4.09E07 $\pm$ 2.74E06	3.03E07 $\pm$ 2.59E06	3.15E07 $\pm$ 2.03E06
Glucosa	5.21-5.24	1.22E07 $\pm$ 8.00E05	1.41E07 $\pm$ 1.25E06	1.11E07 $\pm$ 9.67E05
Lípidos (CH <sub>2</sub> -CO)	2.17- 2.25	1.81E07 $\pm$ 2.11E06	1.32E07 $\pm$ 1.38E06	1.18E07 $\pm$ 9.77E05
Lactato	4.13-4.18	1.84E06 $\pm$ 1.87E05	2.01E06 $\pm$ 2.00E05	2.51E06 $\pm$ 2.04E05
Lípidos (CH <sub>2</sub> -C=C)	1.93-2.01	4.69E07 $\pm$ 2.51E06	3.67E07 $\pm$ 2.20E06	3.92E07 $\pm$ 1.67E06
Citrato	2.50-2.55	1.46E06 $\pm$ 9.17E04	1.51E06 $\pm$ 6.65E04	1.80E06 $\pm$ 1.00E05
Colesteroi/LDL/VLDL (CH <sub>3</sub> )	0.70-0.90	1.62E08 $\pm$ 7.42E06	1.24E08 $\pm$ 7.53E06	1.39E08 $\pm$ 4.91E06
Lisina	2.87-2.93	3.70E06 $\pm$ 5.93E04	3.14E06 $\pm$ 1.11E05	3.77E06 $\pm$ 8.45E04
Tirosina	7.16-7.20	5.64E05 $\pm$ 3.46E04	7.02E05 $\pm$ 7.02E05	6.30E05 $\pm$ 4.80E04
Colina -N(CH <sub>3</sub> ) <sub>3</sub> <sup>+</sup>	3.19-3.21	3.31E07 $\pm$ 1.60E06	2.77E07 $\pm$ 1.70E06	3.06E07 $\pm$ 1.60E06
Glutamato	2.32-2.35	3.40E06 $\pm$ 2.35E05	4.35E06 $\pm$ 3.09E05	4.71E06 $\pm$ 2.98E05
Treonina	4.21-4.29	7.94E06 $\pm$ 5.55E05	6.41E06 $\pm$ 5.41E05	6.44E06 $\pm$ 3.66E05

<sup>a</sup> Región del espectro de RMN utilizada para la cuantificación.







**Figura 26:** Diagramas de cajas de las intensidades relativas obtenidas para los metabolitos que mostraron las variaciones más significativas entre los tres grupos de muestras incluidas en el estudio: voluntarios sanos (C), pacientes con MM en el momento del diagnóstico (D), y pacientes con MM tras la remisión completa (R). Las cajas incluyen el 50% de los casos de cada grupo, la línea central indica la mediana del grupo, y los extremos de las líneas indican los límites del intervalo de confianza del 95%. “o” representa los *outliers* moderados y “x” los *outliers* extremos. Los niveles de los metabolitos corresponden a la integral del área de la señal identificada en el espectro expresada en unidades arbitrarias (u.a.) y dividida por  $10^6$ . (\*) Diferencias estadísticamente significativas ( $p < 0.05$ ).

El análisis global de estas variaciones reveló tres comportamientos distintos:

- A. Algunos metabolitos, entre los que se encuentran el 3-hidroxiacetato, la arginina y el acetato, varían entre los perfiles metabólicos de voluntarios sanos y pacientes con MM en el momento del diagnóstico, y lo hacen en el mismo sentido al comparar los pacientes con MM al diagnóstico y tras la remisión completa. Estas variaciones podrían ser indicativas de cambios metabólicos provocados por la enfermedad, que no revierten cuando los pacientes han alcanzado la remisión completa de la enfermedad.
- B. Otros metabolitos, como el citrato, el lactato, etc., muestran diferencias significativas entre el grupo de pacientes con MM al diagnóstico y tras la remisión completa, pero no entre los valores en voluntarios sanos y pacientes con MM en el momento del diagnóstico, y pueden ser indicativos de cambios provocados por el tratamiento terapéutico no necesariamente ligados a la recuperación del paciente.
- C. Un tercer grupo de metabolitos, como la glutamina, el colesterol, algunos lípidos, la lisina, etc., presentan variaciones en ambos casos, pero en direcciones opuestas una vez iniciado el tratamiento. Estos metabolitos pueden estar relacionados con cambios en el perfil metabólico de los pacientes debidos a la respuesta al tratamiento (recuperación de la enfermedad).

## 1.5.- Interpretación biológica

### a) Pacientes con MM en el momento del diagnóstico vs voluntarios sanos

Como se ha indicado anteriormente, el modelo estadístico supervisado OPLS-DA obtenido para la comparación de pacientes con MM en el momento del diagnóstico, frente a individuos sanos reveló que las diferencias metabólicas más significativas se deben a la presencia de niveles más elevados en los pacientes con MM de ciertos aminoácidos (isoleucina, fenilalanina, arginina y tirosina) y acetato, así como de niveles inferiores de 3-hidroxibutirato, colesterol, lisina, lípidos, colina y glutamina.

Las alteraciones en el perfil sérico de ácidos grasos ya han sido descritas con anterioridad en relación a diversos procesos neoplásicos (Body, 1999). En este tipo de procesos, tanto los ácidos grasos esenciales como los saturados, monoinsaturados y sus derivados poli-insaturados se ven disminuidos, aparentemente como consecuencia de la enfermedad (Baró et al., 1998). El descenso observado en los niveles de colesterol en pacientes con MM, que está de acuerdo con lo descrito en la bibliografía, podría estar relacionado con un aumento en la unión a LDL y un mayor consumo de lípidos y colesterol por las células de mieloma (Yavasoglu et al., 2008). De hecho, se ha demostrado experimentalmente que las células de mieloma NS0 dependen, de forma constitutiva, de colesterol, probablemente debido a una deficiencia bioquímica en la desmetilación del lanosterol para dar lugar a esteroides C-29 (Luz-Hdez, 2012; Sato et al., 1988). Otros estudios previos (Roberts et al., 1976; Weiss et al., 2003) han puesto de manifiesto la necesidad de lisina y glutamina para el crecimiento de las células de mieloma. Por otro lado, la glutamina, además de ser rápidamente utilizada por las células de mieloma, es eliminada de manera espontánea mediante su conversión al ácido carboxílico de la pirrolidona y a través de su modificación enzimática en ácido glutámico (Wein & Goetz, 1973).

Las diferencias observadas en los niveles de algunos aminoácidos en pacientes con MM podrían estar además asociadas a las características específicas de las proteínas producidas en el transcurso de este proceso patológico. A este respecto, existen estudios que demuestran que las proteínas séricas aisladas a partir del suero de pacientes con MM presentan una mayor proporción de ácido aspártico, ácido glutámico, lisina, prolina y serina, mostrando además una menor utilización de arginina, isoleucina, fenilalanina, triptófano y tirosina (Grisolia & Cohen, 1953).

El modelo OPLS-DA obtenido para muestras de pacientes con MM en el momento del diagnóstico e individuos sanos no reveló diferencias significativas en los niveles de glucosa. Este resultado concuerda con los resultados obtenidos en la evaluación de pacientes con MM mediante PET/CT. En este sentido, se ha observado que, a pesar de la sensibilidad de esta técnica y de su utilidad en el diagnóstico de MM, la captación de fluoro-deoxiglucosa ( $^{18}\text{F}$ -FDG) en pacientes con MM, tomando como medida los valores estandarizados de captación máxima ( $\text{SUV}_{\text{max}}$ ), puede ser relativamente baja e incluso en ocasiones comparable al valor de  $\text{SUV}_{\text{max}}$  de una lesión benigna, siendo habitualmente difícil la distinción entre una lesión benigna y una lesión de MM con una tasa metabólica baja (Nanni et al., 2007). Sin embargo, el PET/CT basado en la captación de  $^{11}\text{C}$ -colina parece ser más sensible que el basado en la captación de  $^{18}\text{F}$ -FDG para la detección de lesiones óseas mielomatosas. La colina es una molécula pequeña, precursora de fosfolípidos, cuya utilización se ve incrementada en células en estado de proliferación debido a su implicación en el metabolismo de la membrana y el crecimiento celular. Los niveles séricos de colina que se encontraron en el estudio para pacientes con MM, significativamente inferiores a los niveles en voluntarios sanos, se correlacionan con los datos publicados en los estudios mediante PET/CT. A su vez, los niveles de colina una vez alcanzada la remisión completa son superiores cuando se comparan con los niveles en el momento del diagnóstico, probablemente reflejando la desaparición de las lesiones mielomatosas.

#### **b) Pacientes con MM en el momento del diagnóstico vs pacientes con MM tras la remisión completa**

En este caso, los cambios más significativos que se encontraron entre los perfiles metabólicos de los pacientes con MM en el momento del diagnóstico y tras la remisión completa de la enfermedad estaban relacionados con los niveles de glucosa, citrato, lactato y lisina.

Por un lado, los niveles inferiores de glucosa unidos a niveles superiores de lactato en los pacientes con MM tras la remisión completa podrían estar relacionados con el efecto de tratamiento de quimioterapia administrado a los pacientes tras el diagnóstico de la enfermedad. Numerosos estudios indican que los niveles séricos de lactato deshidrogenasa (LDH), la enzima que cataliza la interconversión de piruvato en lactato, podrían encontrarse elevados debido al tratamiento con bortezomib, un fármaco utilizado en la mayoría de los regímenes



terapéuticos que recibieron los pacientes incluidos en este estudio, y no con la progresión de la enfermedad (Chang et al., 2011).

Este modelo estadístico permitió detectar además que los niveles de citrato en los pacientes con MM se veían incrementados una vez alcanzada la remisión completa. Este hecho podría explicarse por el efecto inhibitorio de los tratamientos contra el MM, en particular del bortezomib (Shin et al., 2008), sobre los factores inducibles por hipoxia (HIFs), una familia de factores de transcripción que promueven la respiración anaerobia (Kirito, 2012) a través de distintos mecanismos. Uno de estos mecanismos incluye la activación de la piruvato deshidrogenasa quinasa (PDK), una enzima que inactiva la piruvato deshidrogenasa (E1) a través de una fosforilación mediada por ATP. E1 es el primer componente del complejo piruvato deshidrogenasa (PDC), el cual cataliza la transformación del piruvato en acetil-CoA, una molécula clave en el ciclo del ácido cítrico. En este contexto, la inhibición de HIF provocada por fármacos como el bortezomib podría jugar un papel importante en el cambio de una respiración anaeróbica a una respiración aeróbica, explicando así los niveles elevados de citrato observados tras el tratamiento en el grupo de muestras de pacientes con MM recogidas tras la remisión completa.

### **c) Pacientes con MM tras la remisión completa vs voluntarios sanos**

En contraste con los cambios observados en la comparación de los perfiles metabólicos de los pacientes con MM en el momento del diagnóstico con respecto al de los individuos sanos o de los pacientes con MM tras la remisión completa, el modelo estadístico resultante del estudio de los perfiles metabólicos de los pacientes con MM tras la remisión completa en comparación con los individuos sanos mostró, como era de esperar, menor poder de discriminación entre los dos grupos de muestras. Este resultado sugiere que el perfil metabólico de los pacientes con MM, una vez alcanzada la remisión completa de la enfermedad, tiene un mayor parecido con el correspondiente a individuos sanos.

## **1.6.- Relevancia de los resultados**

El objetivo principal de este estudio es proporcionar una visión complementaria de los cambios metabólicos que caracterizan la aparición y la respuesta al tratamiento de los pacientes con MM. Los estudios metabólicos realizados con biofluidos tienen la ventaja cualitativa, cuando se comparan con otras plataformas

analíticas, de evaluar de una forma sistemática, empleando una aproximación holística (un único experimento de RMN proporciona información sobre múltiple metabolitos y diferentes rutas de señalización), la complejidad de los cambios que caracterizan la aparición de una enfermedad, y la posible relación entre los metabolitos circulantes y los procesos patológicos. Por otro lado, y aunque el objetivo de este estudio no es reemplazar el procedimiento actual de diagnóstico de MM, se han evaluado las posibles diferencias en las sensibilidades analíticas entre la aproximación clásica de diagnóstico y la establecida en este estudio. Esta evaluación se ha realizado teniendo en cuenta el peso molecular promedio de las proteínas monoclonales (aproximadamente 150 kDa), el umbral de proteína M por encima del cual se diagnostica MM actualmente (10 g/L) y el límite de sensibilidad del espectrómetro de RMN (600 MHz equipado con una criosonda TCI) empleado en este estudio (bajo micromolar). El análisis de estos datos indica que, desde un punto de vista analítico, no existen diferencias significativas entre los dos métodos. Esto no significa que los dos métodos sean equivalentes desde el punto de vista de la sensibilidad clínica, aspecto éste que podrá determinarse en estudios prospectivos. De manera independiente, se ha realizado también un análisis estadístico complementario (correlación de *Pearson*) de los datos para evaluar si existe alguna correlación entre los niveles de la proteína M, medida en el momento del diagnóstico, y las correspondientes fluctuaciones metabólicas identificadas en este estudio. En este sentido, se ha encontrado que existe una correlación negativa, estadísticamente significativa ( $p < 0.05$ ), entre los niveles de la proteína M y los correspondientes a las señales de colesterol/LDL/VLDL, isoleucina, lisina, lípidos ( $\text{CH}_2\text{-CH=CH}$ ), y lípidos ( $\text{CH}_2\text{-CO}$ ), lo que viene a respaldar los resultados obtenidos en este estudio. El signo negativo de esta correlación indica que niveles relativamente altos de proteína M están ligados a niveles relativamente bajos de estos metabolitos, que es de hecho el comportamiento observado para los pacientes de MM en el momento del diagnóstico.

Globalmente, y aún siendo necesarios estudios prospectivos de validación que incluyan un mayor número de pacientes, los resultados obtenidos en este estudio, ponen de manifiesto que el análisis estadístico multivariante de los perfiles metabólicos obtenidos por RMN de pacientes con MM permite distinguir de forma objetiva individuos que presentan o no la enfermedad, así como identificar aquellos que han alcanzado la remisión completa. La caracterización del perfil metabólico específico asociado a MM presenta un gran potencial en la identificación y desarrollo

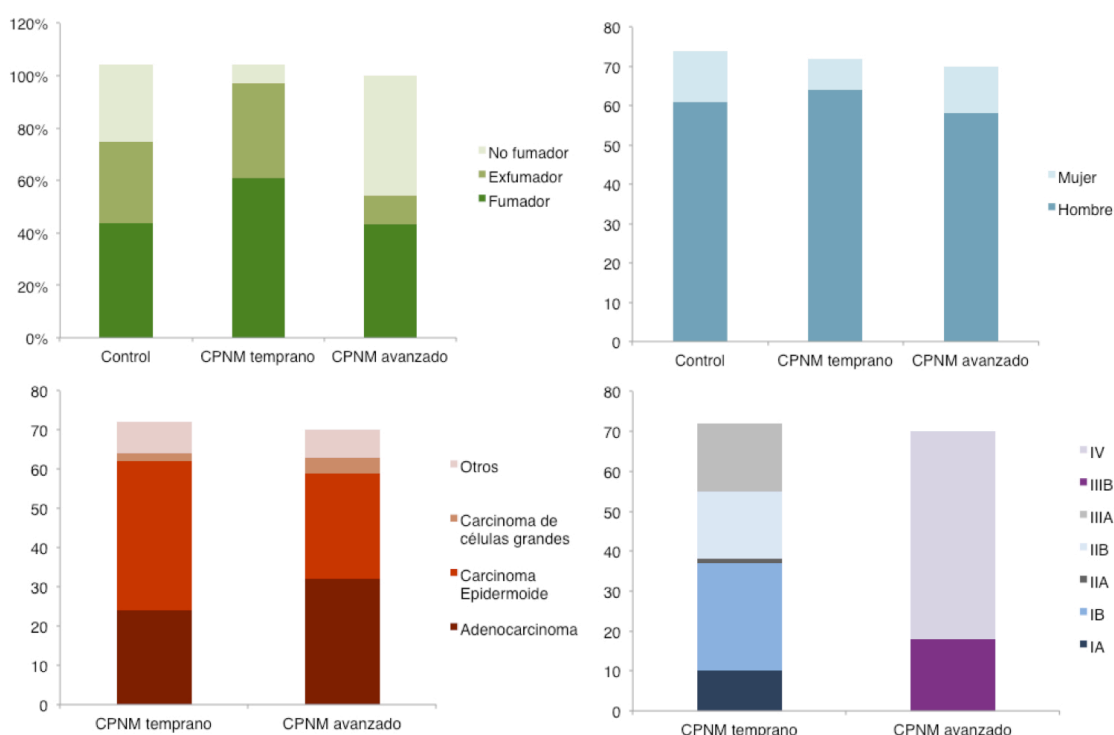
de nuevos biomarcadores útiles en el diagnóstico y tratamiento de la enfermedad. A través del estudio de los perfiles obtenidos mediante  $^1\text{H}$ -RMN podrían identificarse, en el suero de los pacientes, cambios relacionados con la enfermedad de una manera rápida y no invasiva.

Desde un punto de vista clínico, los resultados asociados con la respuesta al tratamiento presentan un interés especial porque podrían representar, en combinación con otros parámetros clínicos y moleculares, un punto de partida en la individualización de los tratamientos terapéuticos para distintos grupos de pacientes. Actualmente, uno de los mayores problemas en el tratamiento de los pacientes con MM es el desarrollo de resistencias a los fármacos, así como la aparición de efectos adversos. La utilización de estos biomarcadores podría ayudar a evitar el tratamiento por exceso o por defecto de los pacientes con MM, así como la duración de los tratamientos, evitando toxicidades innecesarias y permitiendo explorar distintas aproximaciones terapéuticas específicas para los pacientes (San-Miguel & Mateos, 2011).

## 2.- CÁNCER DE PULMÓN

### 2.1.- Análisis y asignación de los espectros

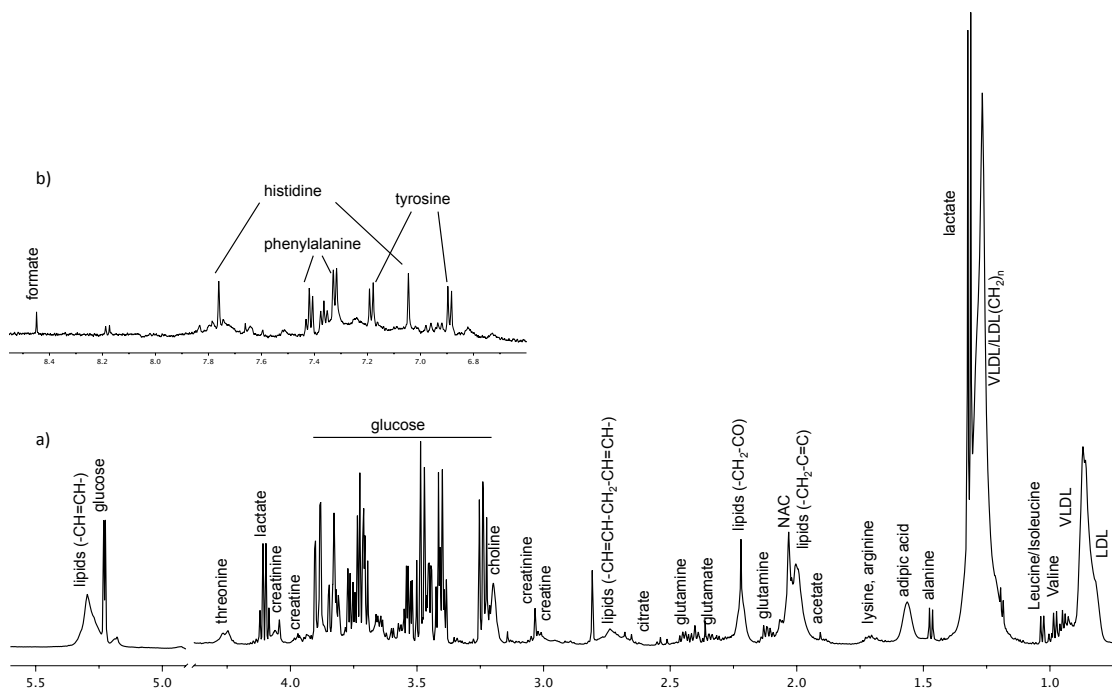
Se analizaron mediante  $^1\text{H-RMN}$  los perfiles metabólicos en suero de un grupo de pacientes diagnosticados con CPNM, 72 de ellos en estadios tempranos, resecables (I-III A) y 70 en estadios avanzados (IIIB-IV), y un grupo de 74 muestras de suero de voluntarios sanos (emparejados en edad, sexo y hábito tabáquico con el grupo de pacientes), cedidas por la Red de Biobancos de la Comunidad Valenciana. En la **Figura 27** se muestra la distribución de algunas de las características clínico-patológicas de los grupos de pacientes incluidos en el estudio.



**Figura 27:** Características clínico-patológicas de los grupos de pacientes incluidos en el estudio.

Los espectros obtenidos del suero de pacientes con CPNM en diferente estadio, así como los correspondientes a los individuos sanos incluidos en el estudio, contienen una serie de señales derivadas de una mezcla compleja de compuestos, entre los que destacan metabolitos, lípidos de bajo peso molecular y proteínas. Las señales identificadas en los mismos (**Figura 28**) corresponden mayoritariamente a aminoácidos, intermediarios metabólicos, ácidos grasos y triglicéridos saturados e insaturados, N-acetil proteínas y ácidos orgánicos. La excelente calidad de los mismos, evaluada tanto por la existencia de un gran número de señales como por la

dispersión y resolución de las mismas, permitió el análisis estadístico multivariante de los datos, dirigido en primer lugar a la identificación de posibles agrupaciones o tendencias dentro de cada grupo (análisis no supervisado), y finalmente a la identificación de aquellos metabolitos que contribuyen significativamente a la discriminación entre los perfiles metabólicos de los distintos grupos de pacientes incluidos en el estudio (análisis supervisado).



**Figura 28:** Espectro  $^1\text{H}$ -RMN CPMG correspondiente a una muestra de suero de un paciente diagnosticado con CPNM en estadio temprano. (a) Región alifática del espectro ( $\delta = 0.15\text{-}5.5$  ppm) y (b) amplificación ( $\times 25$ ) de la región aromática ( $\delta = 6.60\text{-}8.60$  ppm), con la asignación de las señales más representativas del espectro.

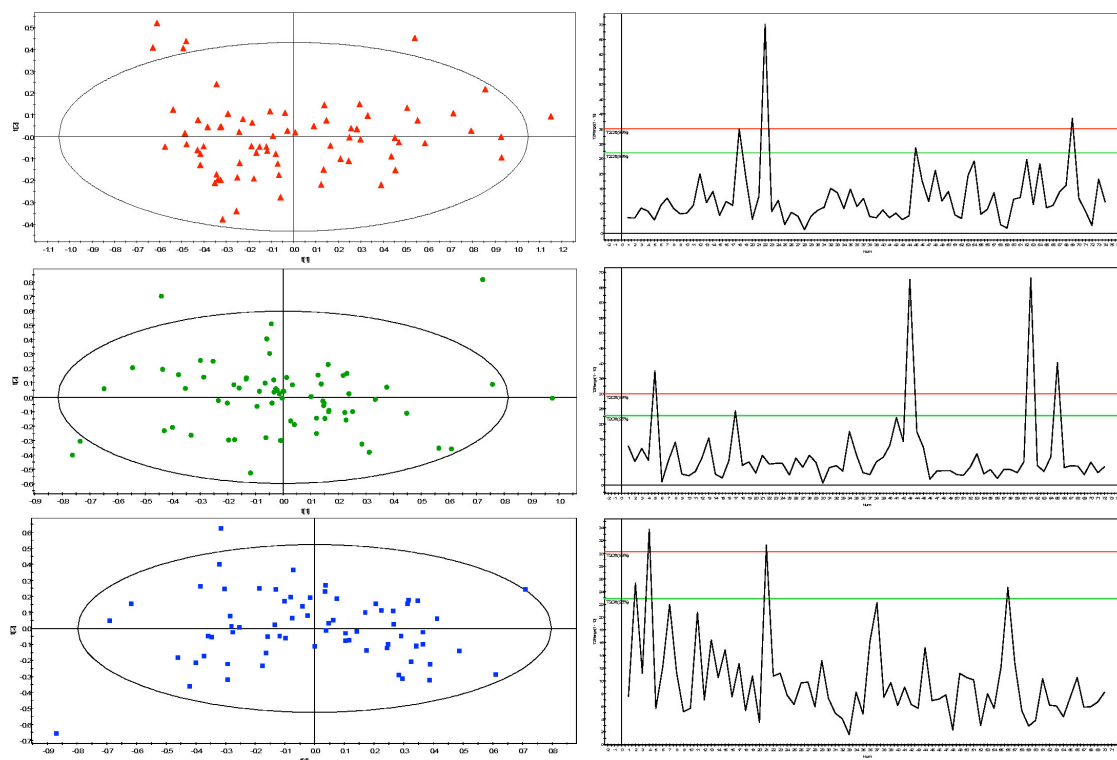
El análisis estadístico multivariante de los espectros obtenidos se basó en la evaluación de los espectros  $^1\text{H}$ -RMN CPMG ya que éstos presentan una serie de ventajas relacionadas con una línea base óptima para este tipo de análisis, así como unas características espectrales que facilitan la identificación de metabolitos. Los espectros fueron posteriormente divididos en regiones de  $0.01$  ppm desde  $10.5$  ppm a  $0.26$  ppm, y se excluyeron las regiones del agua ( $5.60\text{-}4.30$  ppm) y la urea ( $5.85\text{-}5.60$  ppm) para evitar interferencias relacionadas con diferencias en la supresión del agua y/o variabilidad en la señal de la urea. Finalmente, antes de llevar a cabo el análisis estadístico se escalaron los datos usando Pareto mediante la división de cada variable por la raíz cuadrada de la desviación estándar de la misma.

## 2.2.- Análisis no supervisado

Tras la integración de los espectros, utilizando un tamaño de *bucket* constante de 0.01 ppm, se procedió al estudio de la homogeneidad de cada grupo del estudio, a través de los modelos PCA correspondientes, y al estudio de los *outliers* identificados en el gráfico  $T^2$  de *Hotelling*.

### a) Identificación de *outliers*

Para la identificación de *outliers*, se analizaron los *score plot* y el gráfico  $T^2$  de *Hotelling* obtenidos para cada grupo de pacientes. En la **Figura 29** se muestran los gráficos obtenidos para cada grupo, generados a partir de su correspondiente modelo PCA.

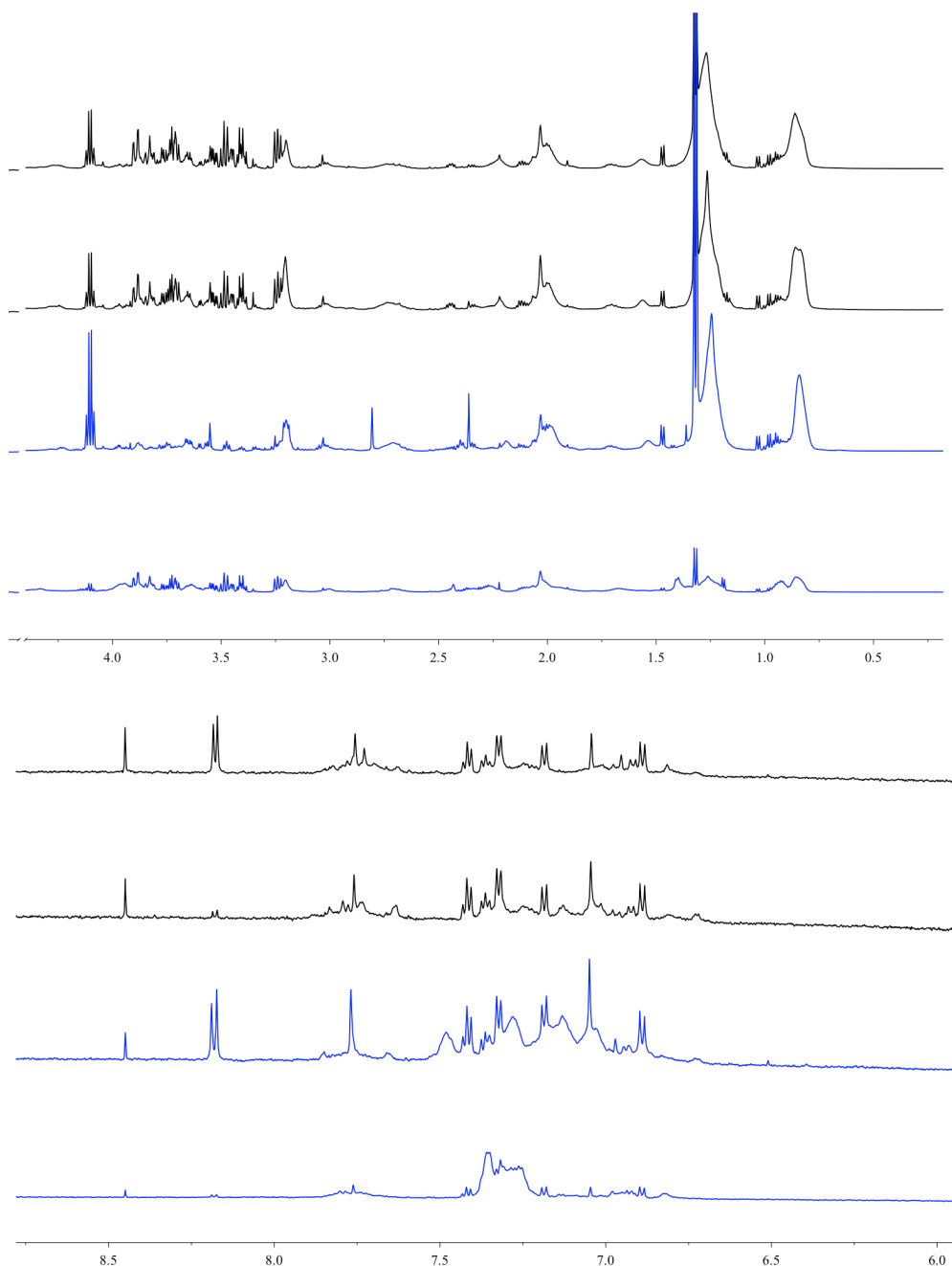


**Figura 29:** Análisis no supervisado de los grupos incluidos en el estudio. Izquierda, *score plots* derivados del análisis PCA de los espectros  $^1\text{H-RMN}$  CPMG de muestras del grupo control (rojo), muestras del grupo de pacientes con CPNM en estadios tempranos (verde) y pacientes con CPNM en estadios avanzados (azul). La elipse marca el 95% del intervalo de confianza en los modelos estadísticos generados. Derecha, gráfico  $T^2$  de *Hotelling* derivado del análisis PCA correspondiente. Los límites marcados en rojo y verde denotan, respectivamente, los intervalos de confianza del 99% y 95%.

El análisis detallado de los espectros correspondientes a las muestras identificadas como *outliers* dentro de cada grupo permitió identificar las características que distinguían estas muestras en relación a las de su grupo. Atendiendo a las principales

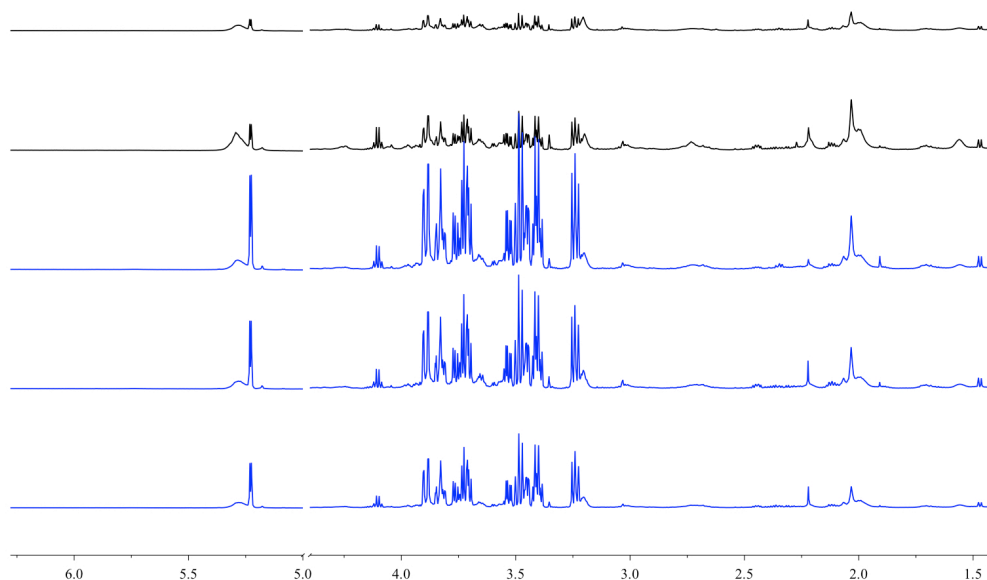
diferencias que presentaban, los *outliers* de todos los grupos se clasificaron en cuatro categorías:

- a) Problemas en la conservación de las muestras (**Figura 30**): Dos muestras del grupo de pacientes con CPNM en estadios tempranos no dieron lugar a espectros de suficiente calidad, probablemente debido a la degradación de las muestras de suero.



**Figura 30:** Comparación de los espectros  $^1\text{H}$ -RMN CPMG de dos muestras con signos de degradación (azul) y otras dos muestras representativas de este mismo grupo (negro).

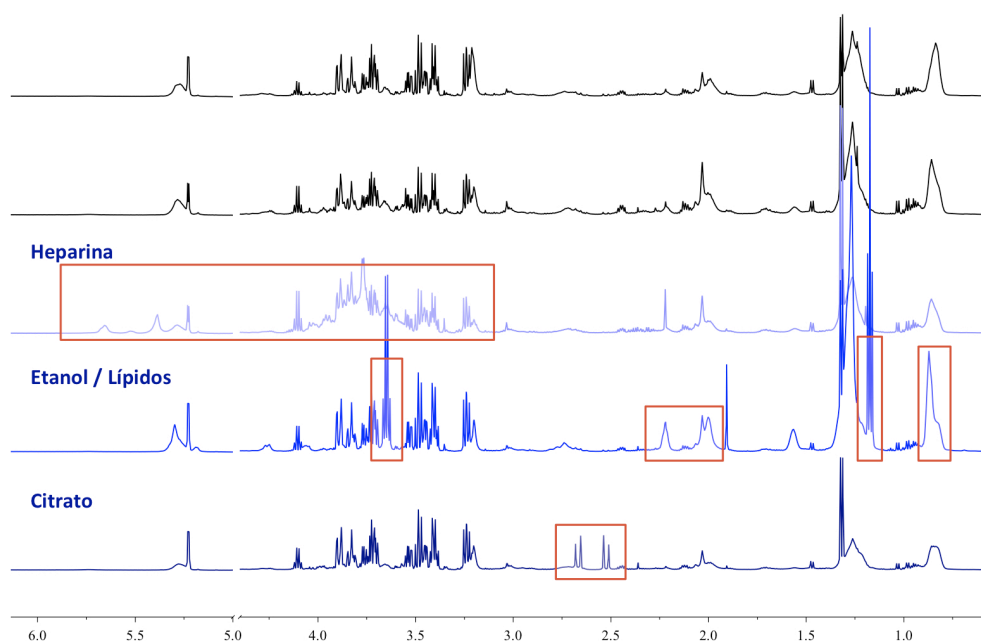
- b) Pacientes que no estaban en ayunas (**Figura 31**): Entre las muestras de suero de los pacientes diagnosticados con CPNM, tres de ellas (dos en estadios tempranos y una en estadios avanzados) presentaron señales mucho más intensas correspondientes a glucosa. Los niveles elevados de glucosa en estos pacientes probablemente sean debidos a que la extracción de la sangre no se realizó en condiciones de ayuno.



**Figura 31:** Comparación de los espectros  $^1\text{H}$ -RMN CPMG de las tres muestras pertenecientes al grupo de pacientes con CPNM que presentaban niveles de glucosa elevados en suero (azul) y otras dos muestras representativas de ese mismo grupo (negro).

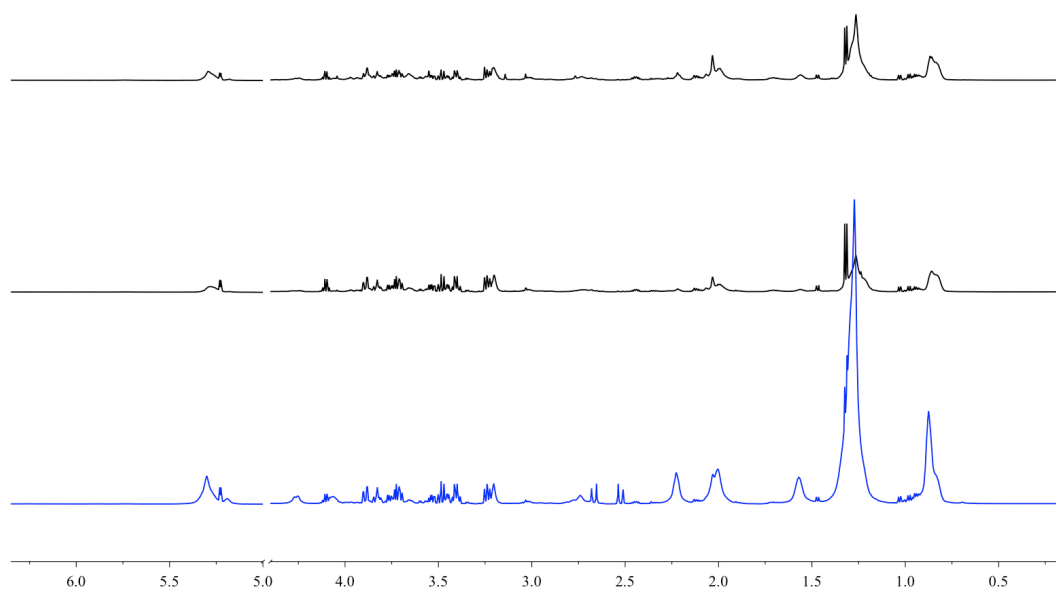
- c) Muestras contaminadas (**Figura 32**): En dos muestras del grupo control y dos del grupo de pacientes con CPNM en estadios tempranos, se identificó la presencia de contaminantes: isopropanol (procedente de desinfectantes empleados durante la punción), citrato (procedente de la utilización de un tubo de recogida de sangre con anticoagulantes, debido a un error durante la recogida de muestras), niveles elevados de lípidos y etanol (probablemente debido a la ingesta), y señales correspondientes a heparina (procedente de la heparinización de la vía venosa en pacientes hospitalizados).





**Figura 32:** Comparación de los espectros  $^1\text{H}$ -RMN CPMG de tres de las muestras contaminadas (azul) y otras dos muestras representativas del estudio (negro). Se han resaltado en rojo las señales correspondientes a los distintos contaminantes que, como se puede apreciar en la comparación, están ausentes en las demás muestras del estudio.

- d) Pacientes con triglicéridos elevados (**Figura 33**): Se descartó del análisis el espectro correspondiente a una de las muestras del grupo control debido a la presencia de señales correspondientes a lípidos muy intensas. Los elevados datos de triglicéridos en este voluntario ya habían sido registrados con anterioridad en la historia clínica del paciente.

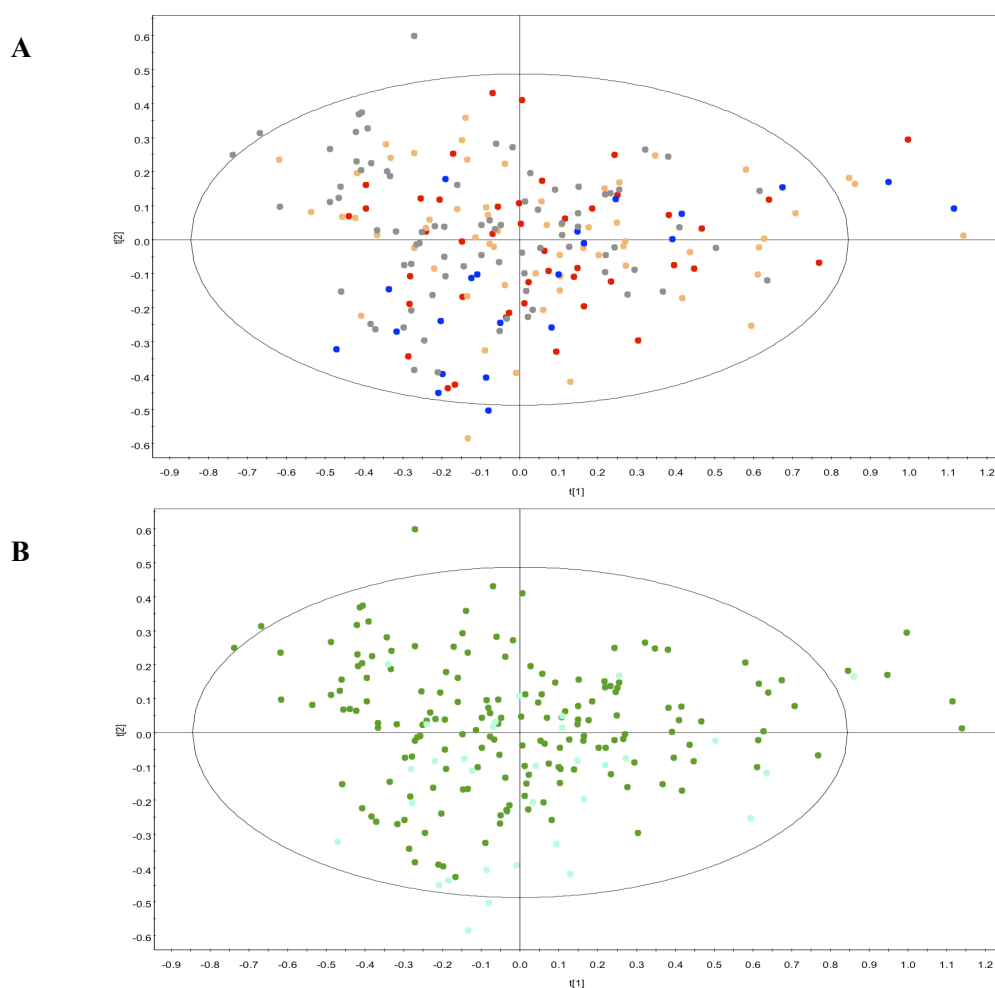


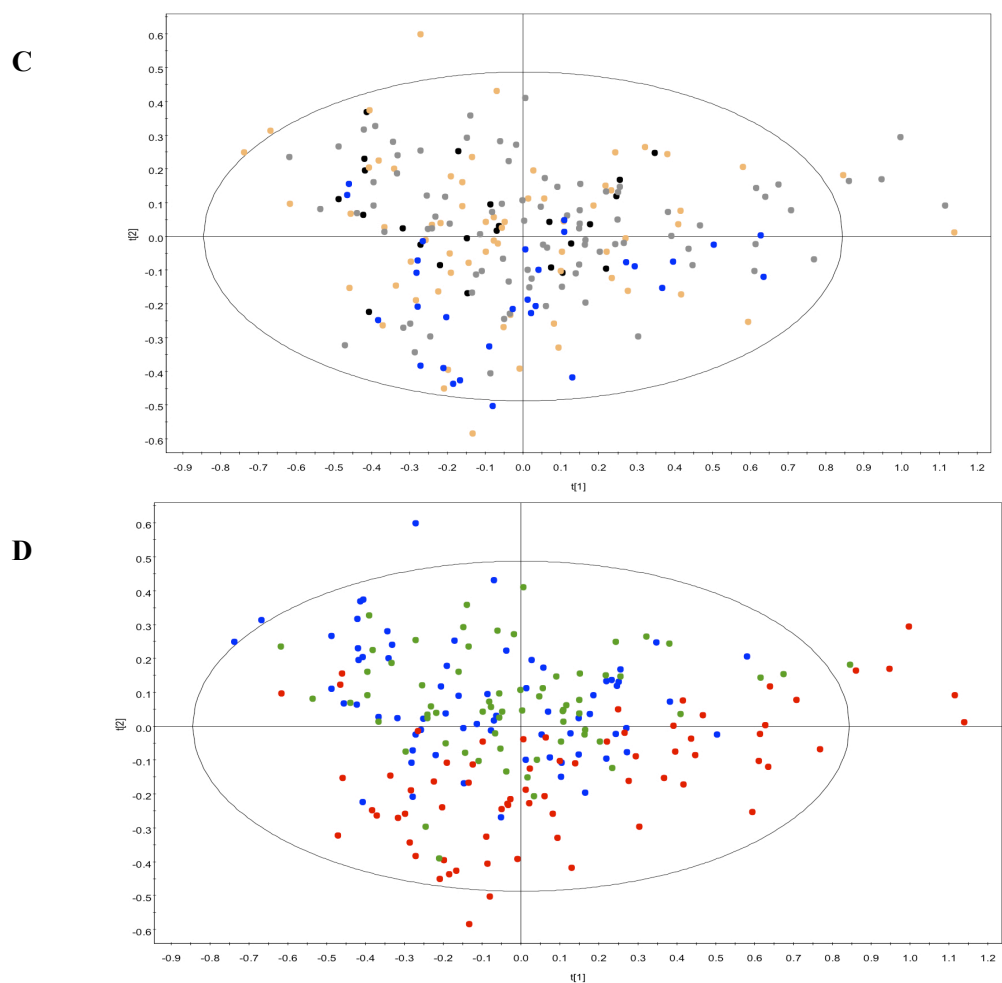
**Figura 33:** Comparación del espectro  $^1\text{H}$ -RMN CPMG de la muestra correspondiente al paciente con hiperlipidemia (azul) y otras dos muestras representativas del grupo control (negro).

## b) Análisis de homogeneidad

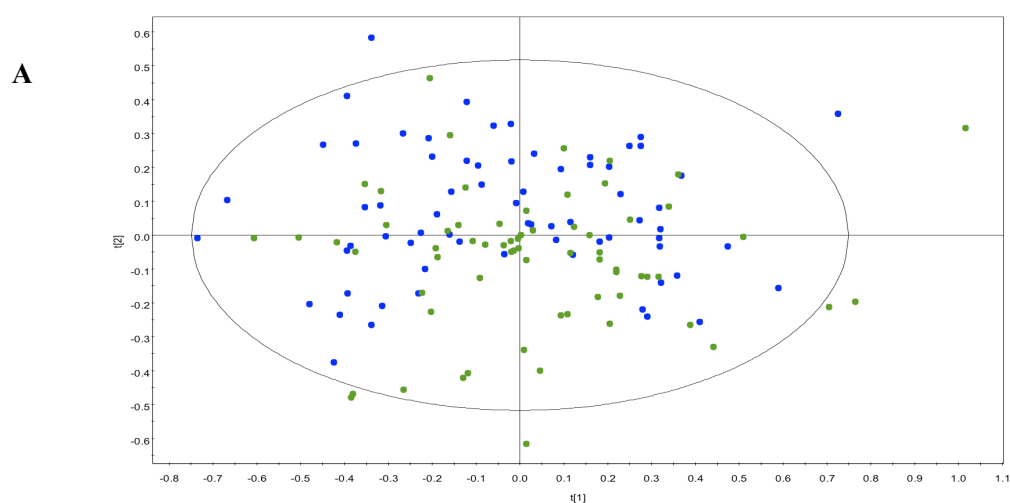
Una vez eliminadas del análisis las muestras que fueron identificadas como *outliers* dentro de su grupo y que, de una manera justificada, se consideró que no debían formar parte del análisis, se procedió a examinar la posible influencia de otras variables distintas a la variable principal del estudio (sano/CPNM). Este análisis se llevó a cabo utilizando un modelo PCA generado a partir de todas las muestras incluidas en los tres grupos del estudio (n=206) para el análisis de las variables demográficas, y el modelo PCA correspondiente al grupo de muestras de pacientes diagnosticados con CPNM (n=135) para el análisis de la influencia de otras variables clínicas.

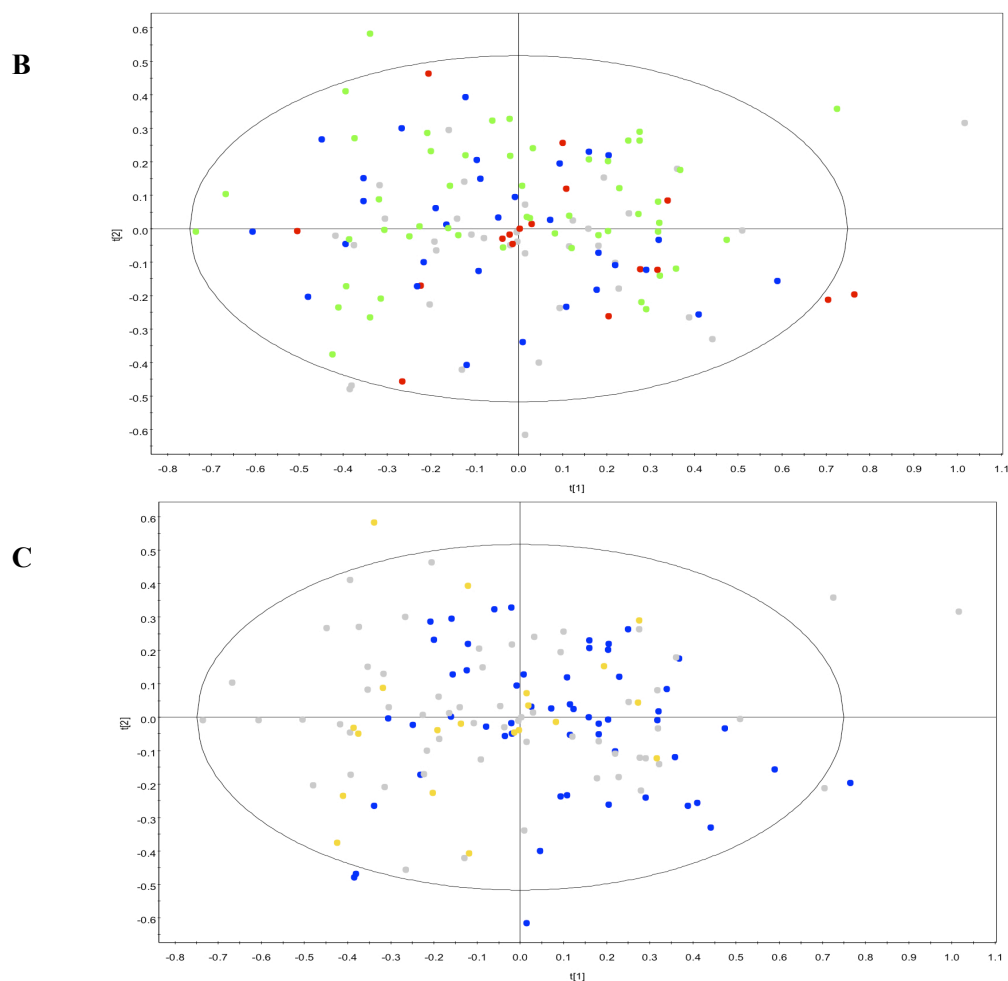
En las siguientes figuras se incluyen los *score plots* de los PCA correspondientes a los tres grupos de muestras incluidos en el estudio (**Figura 34**) y al conjunto de muestras de pacientes diagnosticados con CPNM (**Figura 35**) utilizados para analizar las distintas variables demográficas y clínicas.





**Figura 34:** Análisis PCA de los espectros  $^1\text{H}$ -RMN CPMG correspondientes a todas las muestras incluidas en el estudio. Clasificación de las muestras, atendiendo a códigos de colores, según: **A)** edad (gris: > 65 años, naranja:  $\leq 65$  - > 55 años, rojo:  $\leq 55$  - > 45 años, azul:  $\leq 45$  años), **B)** sexo (verde: hombres, azul: mujeres), **C)** hábito tabáquico (gris: fumadores, naranja: exfumadores, azul: no fumadores, negro: sin información), y **D)** grupo del estudio (rojo: voluntarios sanos, verde: pacientes con CPNM en estadios tempranos; azul: pacientes con CPNM en estadios avanzados).





**Figura 35:** Análisis PCA de los espectros  $^1\text{H-RMN CPMG}$  correspondientes a todas las muestras incluidas en el estudio. Clasificación de las muestras, atendiendo a códigos de colores, según: **A)** grupo del estudio (verde: pacientes con CPNM en estadios tempranos, azul: pacientes con CPNM en estadios avanzados), **B)** estadiaje según TNM (gris: IA/IB; rojo: IIA/IIB; azul: IIIA/IIIB; verde: IV), **C)** histología (azul: adenocarcinoma, gris: carcinoma epidermoide, amarillo: otros).

Entre todas las variables estudiadas, únicamente las variables voluntario sano/CPNM (**Figura 34D**) y CPNM estadio temprano/CPNM estadio avanzado (**Figura 35A**), mostraron la presencia de agrupaciones entre las muestras incluidas en el estudio. No se observó una relación estadísticamente significativa entre ninguna de las demás variables y la posición de las distintas muestras en los modelos PCA generados.

### 2.3.- Análisis supervisado

Los objetivos principales del análisis supervisado de los datos fueron dos:

1. La identificación de las diferencias fundamentales entre los perfiles metabólicos de individuos sanos y pacientes diagnosticados con CPNM, incluyendo los distintos estadios de la enfermedad. Mediante este análisis se pretendía llegar a

la identificación de los principales marcadores característicos de la presencia de enfermedad, y que potencialmente pudiesen ser empleados como biomarcadores de detección temprana de esta patología.

2. La caracterización del perfil metabólico específico de los distintos estadios de la enfermedad. Este estudio podría proporcionar información sobre los cambios metabólicos que se producen durante la progresión de la enfermedad, obteniendo información sobre las posibles rutas metabólicas implicadas.

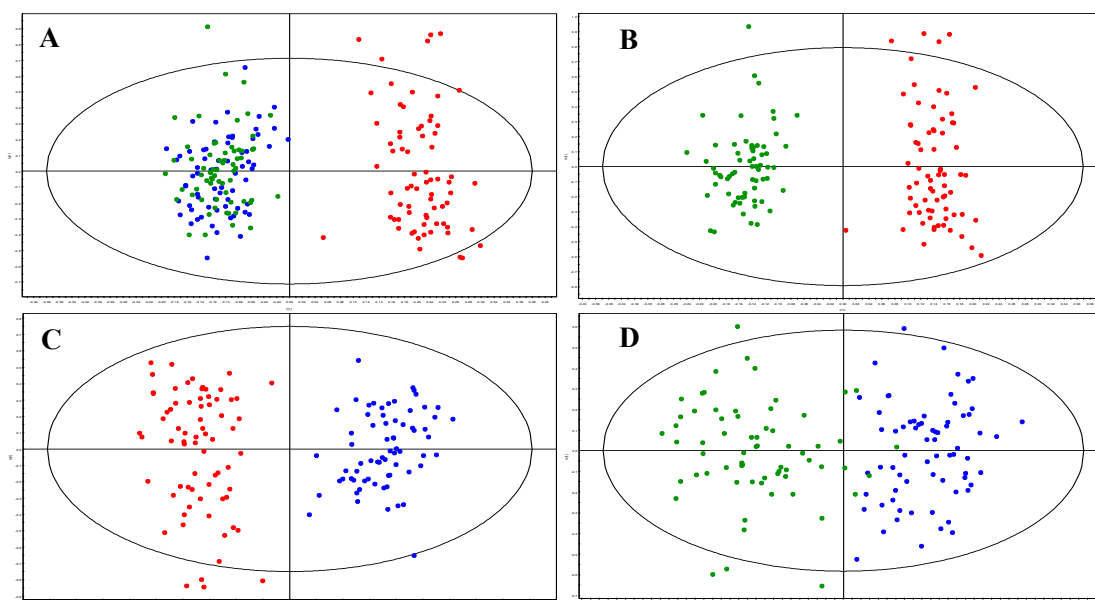
Con esta finalidad, y con carácter exploratorio, se construyeron cuatro modelos OPLS-DA (**Figura 36**) basados en la clasificación de las muestras incluidas en el estudio atendiendo a distintos criterios, con el fin de estudiar las diferencias existentes entre los perfiles metabólicos de:

**A)** Individuos sanos frente a pacientes con CPNM:  $R^2(Y) = 0.930$ ,  $Q^2(Y) = 0.871$ .

**B)** Individuos sanos frente a pacientes con CPNM en estadios tempranos:  $R^2(Y) = 0.950$ ,  $Q^2(Y) = 0.877$ .

**C)** Individuos sanos frente a pacientes con CPNM en estadios avanzados:  $R^2(Y) = 0.911$ ,  $Q^2(Y) = 0.839$ .

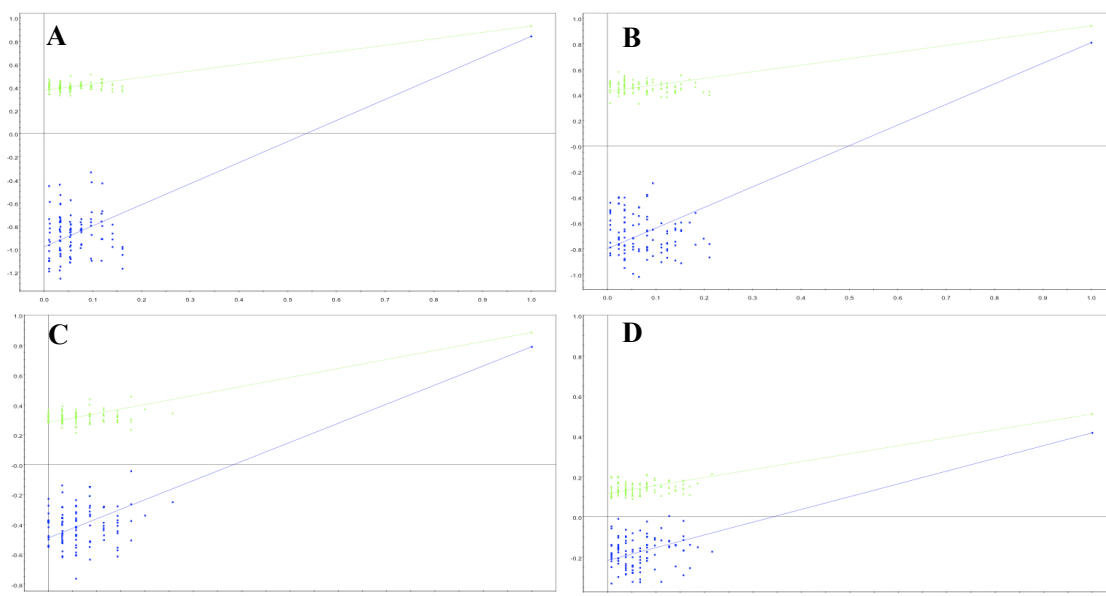
**D)** Pacientes con CPNM en estadios tempranos frente a pacientes con CPNM en estadios avanzados:  $R^2(Y) = 0.780$ ,  $Q^2(Y) = 0.595$ .



**Figura 36:** Modelos estadísticos resultantes del análisis supervisado (OPLS-DA) de los espectros  $^1\text{H-RMN}$  CPMG obtenidos para las muestras de suero de pacientes diagnosticados con CPNM: **A)** individuos sanos (rojo) frente a pacientes con CPNM (azul y verde); **B)** individuos sanos (rojo) frente a pacientes con CPNM en estadios tempranos (verde); **C)** individuos sanos (rojo) frente a pacientes con CPNM en estadios avanzados (azul); **D)** pacientes con CPNM en estadios tempranos (verde) frente a pacientes con CPNM en estadios avanzados (azul).

En todos los casos, los valores obtenidos para el ajuste y la capacidad de predicción del modelo ( $R^2(Y)$  y  $Q^2(Y)$ , respectivamente) indican la existencia de diferencias significativas entre los grupos de muestras analizados. De manera complementaria, y antes de continuar con el análisis de los modelos multivariantes, se calcularon los test de permutación ( $n=100$ , siendo  $n$  el número de repeticiones) para los modelos PLS-DA equivalentes (**Figura 37**), obteniendo los siguientes resultados:

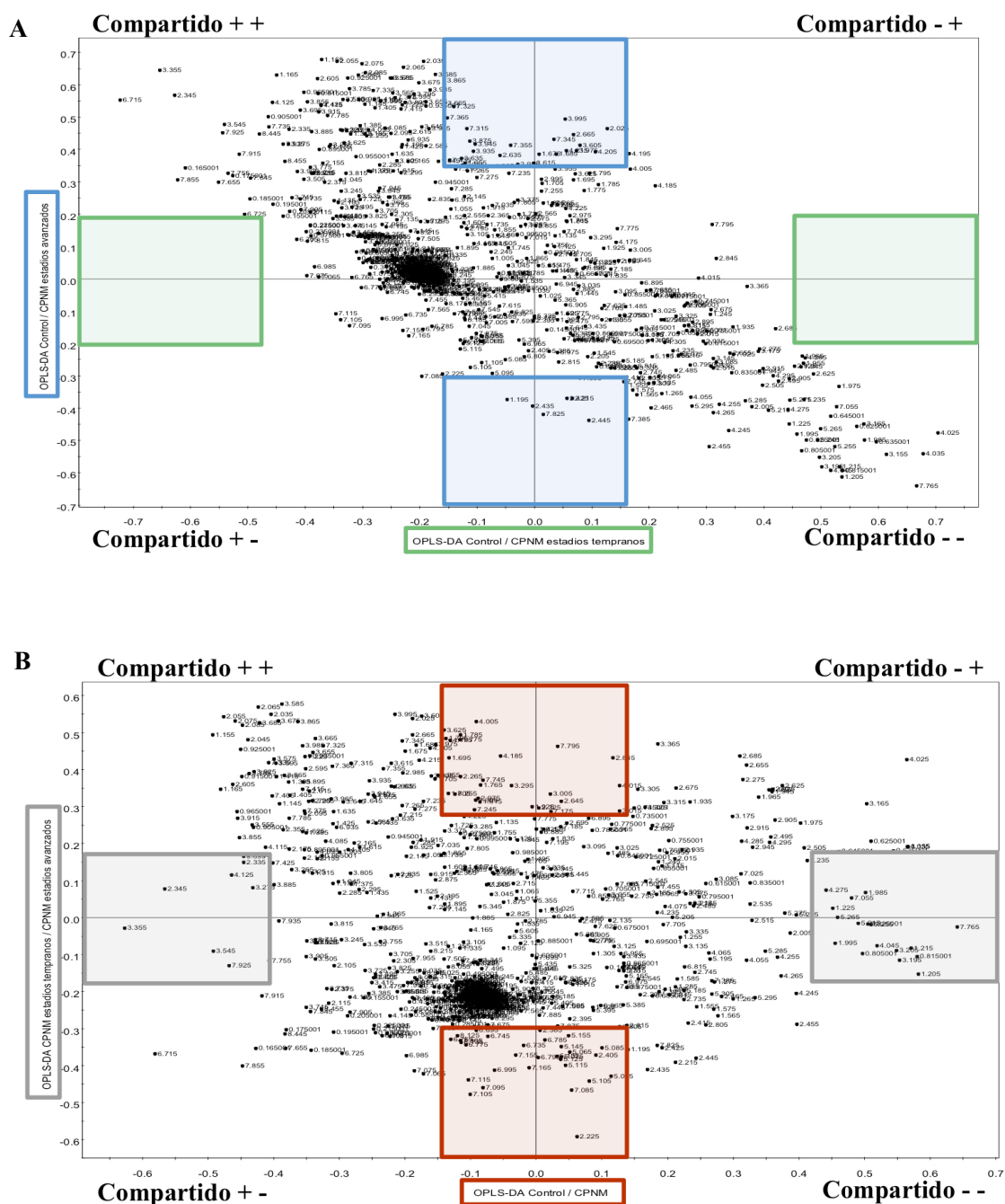
- A)** Individuos sanos frente a pacientes con CPNM: intersección  $R^2(Y) = 0.366$ , intersección  $Q^2(Y) = -0.932$ .
- B)** Individuos sanos frente a pacientes con CPNM en estadios tempranos: intersección  $R^2(Y) = 0.426$ , intersección  $Q^2(Y) = -0.801$ .
- C)** Individuos sanos frente a pacientes con CPNM en estadios avanzados: intersección  $R^2(Y) = 0.279$ , intersección  $Q^2(Y) = -0.491$ .
- D)** Pacientes con CPNM en estadios tempranos frente a pacientes con CPNM en estadios avanzados: intersección  $R^2(Y) = 0.116$ , intersección  $Q^2(Y) = -0.216$ .



**Figura 37:** Resultados del test de permutación para el modelo PLS-DA entre: **A)** individuos sanos frente a pacientes con CPNM; **B)** individuos sanos frente a pacientes con CPNM en estadios tempranos; **C)** individuos sanos frente a pacientes con CPNM en estadio avanzados; **D)** pacientes con CPNM en estadios tempranos frente a pacientes con CPNM en estadios avanzados. En el eje horizontal se representa el valor del coeficiente de correlación entre el valor del modelo original y los valores de cada permutación. El eje vertical representa los valores obtenidos para  $R^2(Y)$  (verde) y  $Q^2(Y)$  (azul).

En todos los casos, tanto los valores obtenidos para la intersección de  $R^2(Y)$  y  $Q^2(Y)$  ( $< 0.3-0.4$  y  $< 0.05$ , respectivamente) como el p-valor obtenido en el test ANOVA ( $p < 0.05$ ) reflejan la validez de los modelos obtenidos.





**Figura 39:** *SUS*-plot correspondientes a la comparación de los modelos OPLS-DA entre: **A)** individuos sanos frente a pacientes con CPNM estadios tempranos y entre individuos sanos frente a pacientes con CPNM estadios avanzados; **B)** individuos sanos frente a pacientes con CPNM y entre pacientes con CPNM en estadios tempranos y pacientes con CPNM en estadios avanzados.

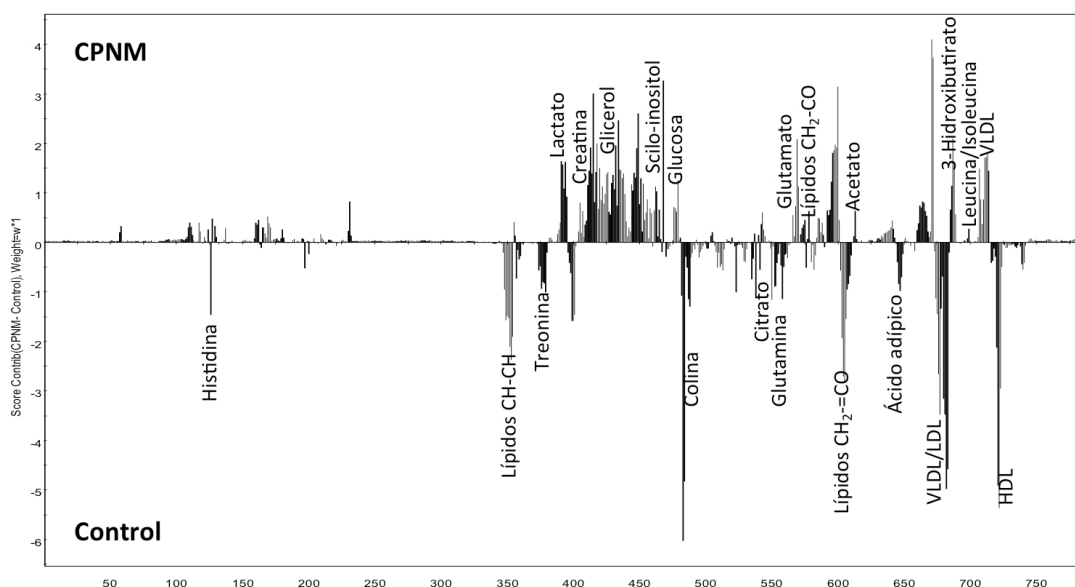
A través del análisis global de los *S*-plot de cada modelo junto con los *SUS*-plot de las comparaciones entre los distintos modelos, se consiguieron identificar las regiones del espectro que contribuían de forma significativa a la discriminación de los distintos grupos, así como aquellas que eran comunes a distintos modelos. Estos resultados se irán detallando en los siguientes apartados. De manera general, los modelos en los



que se comparan los perfiles metabólicos de individuos sanos frente a pacientes con CPNM, independientemente de cual sea su estadio, comparten la mayoría de regiones relevantes, mientras que el modelo en el que se comparan los distintos estadios dentro de los pacientes diagnosticados con CPNM presentan un mayor número de regiones que únicamente son relevantes en esa comparación (**Figura 39B**).

En base a estos resultados, se procedió, por un lado, a la identificación de los metabolitos más significativos que contribuyen a la discriminación entre individuos sanos y pacientes con CPNM (sin diferenciar entre estadios), lo que tendría como objetivo identificar un perfil característico del cáncer de pulmón y, por otro lado, a la de aquellos que contribuyen de forma más significativa a la separación observada en el modelo OPLS-DA para los distintos estadios de CPNM. Con esta finalidad, se analizaron los *contribution plot* de los dos modelos citados. Utilizando este tipo de gráficos se puede representar, de manera cuantitativa, las diferencias existentes en las intensidades de cada variable entre los grupos que se comparan en el modelo estadístico.

El siguiente paso consistió en la asignación de las señales correspondientes en el espectro de RMN de las regiones más representativas en cada modelo (**Figura 40**). Para este análisis, se tuvieron en cuenta todas las señales que presentaban un valor de VIP superior a 1.



**Figura 40:** *Contribution plot* correspondiente a la comparación del modelo OPLS-DA entre individuos sanos frente a pacientes con CPNM. Se indican los metabolitos que presentan las diferencias más importantes entre los dos grupos comparados.

#### 2.4.- Integración de los metabolitos más relevantes

Una vez identificadas las regiones más relevantes en los espectros, y los metabolitos a los que correspondían, se procedió a la integración de las señales en los espectros de RMN originales, correspondientes a cada uno de los grupos analizados en los dos modelos (control/CPNM y CPNM temprano/CPNM avanzado). Los datos obtenidos fueron de nuevo normalizados respecto al área total de cada espectro, al igual que se hizo antes de construir los modelos estadísticos multivariantes.

El análisis estadístico univariante de los distintos metabolitos identificados se realizó empleando la prueba de *Mann Whitney*, ya que muchas de las variables no cumplían los criterios de normalidad (prueba de *Kolmogorov-Smirnov*,  $p < 0.05$ ). En las **Tablas 13** y **14** se incluyen los resultados del análisis univariante de los metabolitos identificados como relevantes. La mayoría de los metabolitos identificados presentaron diferencias estadísticamente significativas entre los grupos de muestras analizados. En la **Figura 41** se incluyen los diagramas de cajas, mostrando la variación en las intensidades de estos metabolitos en los distintos grupos de muestras.

**Tabla 13:** Cuantificación de los metabolitos más significativos derivados del modelo OPLS-DA obtenido para el grupo control y el de pacientes con CPNM.

Metabolito	$\delta^1\text{H}$ (ppm) <sup>a</sup>	Intensidad media $\pm$ s.e.m. (unidades arbitrarias)		% variación	t <sup>b</sup>
		Grupo Control	Pacientes con CPNM		
HDL	0.79-0.85	55.14 $\pm$ 1.47	45.55 $\pm$ 0.71	-17.39	0.0000*
VLDL	0.85-0.91	66.10 $\pm$ 1.46	65.17 $\pm$ 0.76	-1.40	0.9070
Leucina/Isoleucina	0.91-0.97	14.40 $\pm$ 0.24	16.67 $\pm$ 0.25	15.75	0.0000*
3-Hidroxitirato	1.15-1.18	4.95 $\pm$ 0.14	6.54 $\pm$ 0.15	32.09	0.0000*
LDL/VLDL (CH <sub>2</sub> ) <sub>n</sub>	1.20-1.31	229.06 $\pm$ 5.98	197.77 $\pm$ 3.18	-13.66	0.0000*
Compuesto desconocido 1	1.36-1.40	4.01 $\pm$ 0.11	5.04 $\pm$ 0.14	25.65	0.0000*
Ácido Adípico	1.52-1.60	15.97 $\pm$ 0.68	13.79 $\pm$ 0.33	-13.64	0.0410*
Acetato	1.90-1.91	1.29 $\pm$ 0.04	1.56 $\pm$ 0.04	21.22	0.0000*
Lípidos CH <sub>2</sub> -C=C	1.93-2.02	40.15 $\pm$ 0.45	35.56 $\pm$ 0.31	-11.44	0.0000*
N-Acetilcisteína	2.02-2.05	23.50 $\pm$ 0.31	27.01 $\pm$ 0.43	14.94	0.0000*
Lípidos CH <sub>2</sub> CO	2.19-2.26	14.19 $\pm$ 0.64	13.17 $\pm$ 0.30	-7.15	0.5780
Glutamato	2.33-2.36	2.14 $\pm$ 0.07	2.95 $\pm$ 0.07	37.65	0.0000*
Glutamina	2.42-2.47	5.82 $\pm$ 0.19	4.98 $\pm$ 0.11	-14.37	0.0000*
Colina -N(CH <sub>3</sub> ) <sub>3</sub> <sup>+</sup>	3.18-3.21	23.56 $\pm$ 0.65	17.60 $\pm$ 0.33	-25.30	0.0000*
Glucosa	3.22-3.26	26.77 $\pm$ 0.75	28.87 $\pm$ 0.46	7.85	0.0620
Scilo-inositol	3.35-3.36	1.01 $\pm$ 0.05	1.81 $\pm$ 0.05	78.81	0.0000*
Glicerol	3.78-3.80	2.90 $\pm$ 0.09	3.62 $\pm$ 0.08	25.16	0.0000*
Creatina	3.91-3.92	3.22 $\pm$ 0.10	3.93 $\pm$ 0.07	21.99	0.0000*
Lactato	4.08-4.13	10.45 $\pm$ 0.37	13.74 $\pm$ 0.41	31.53	0.0000*
Treonina	4.21-4.30	6.92 $\pm$ 0.12	5.86 $\pm$ 0.08	-15.34	0.0000*
Lípidos CH=CH	5.24-5.36	32.92 $\pm$ 0.74	28.08 $\pm$ 0.44	-14.70	0.0000*
Histidina	7.74-7.78	0.37 $\pm$ 0.01	0.29 $\pm$ 0.01	-20.76	0.0000*

\* Diferencias estadísticamente significativas (p &lt; 0.05)

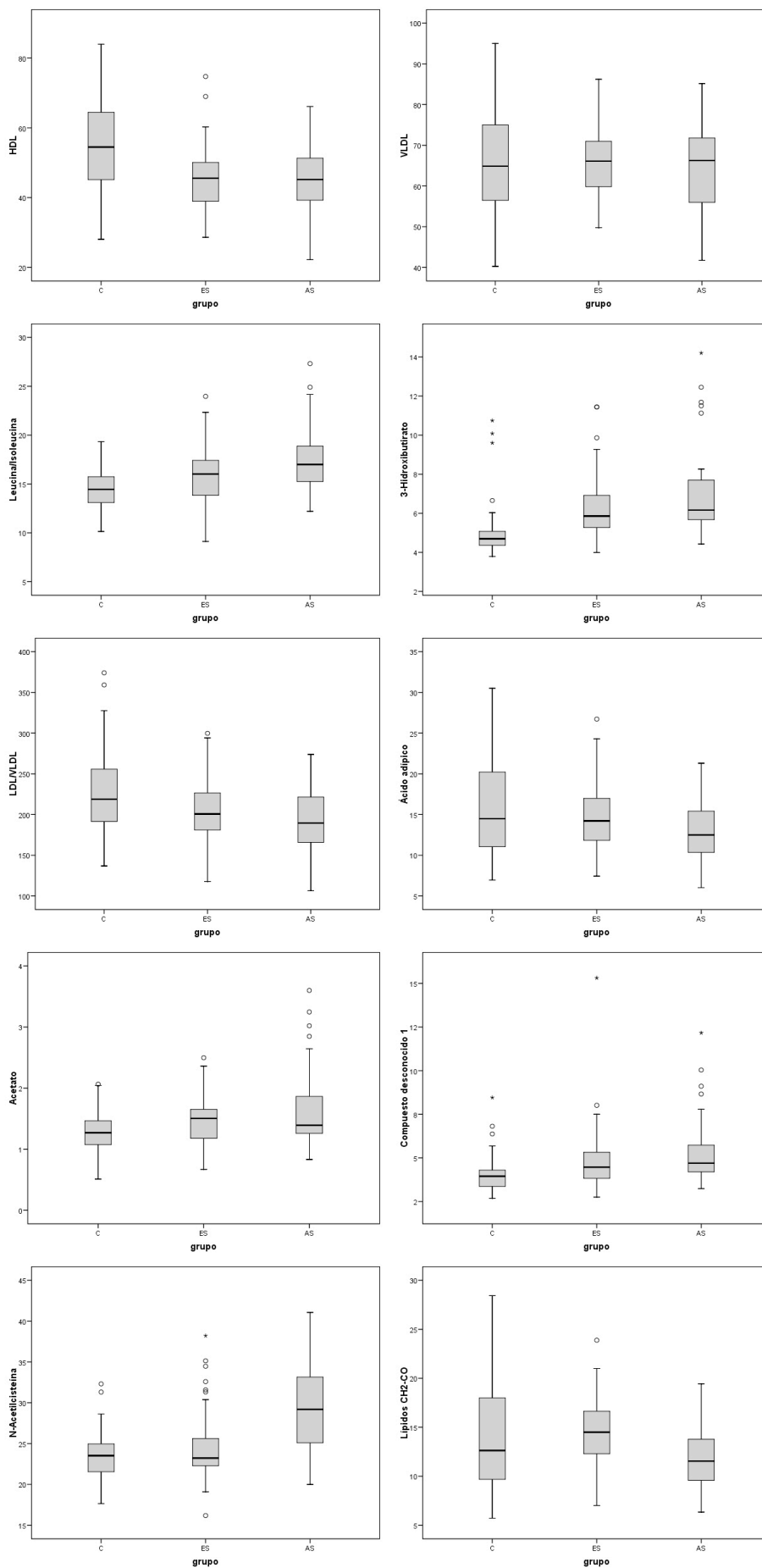
<sup>a</sup> Región del espectro de RMN utilizada para la cuantificación<sup>b</sup> p valor calculado mediante la prueba de Mann Whitney

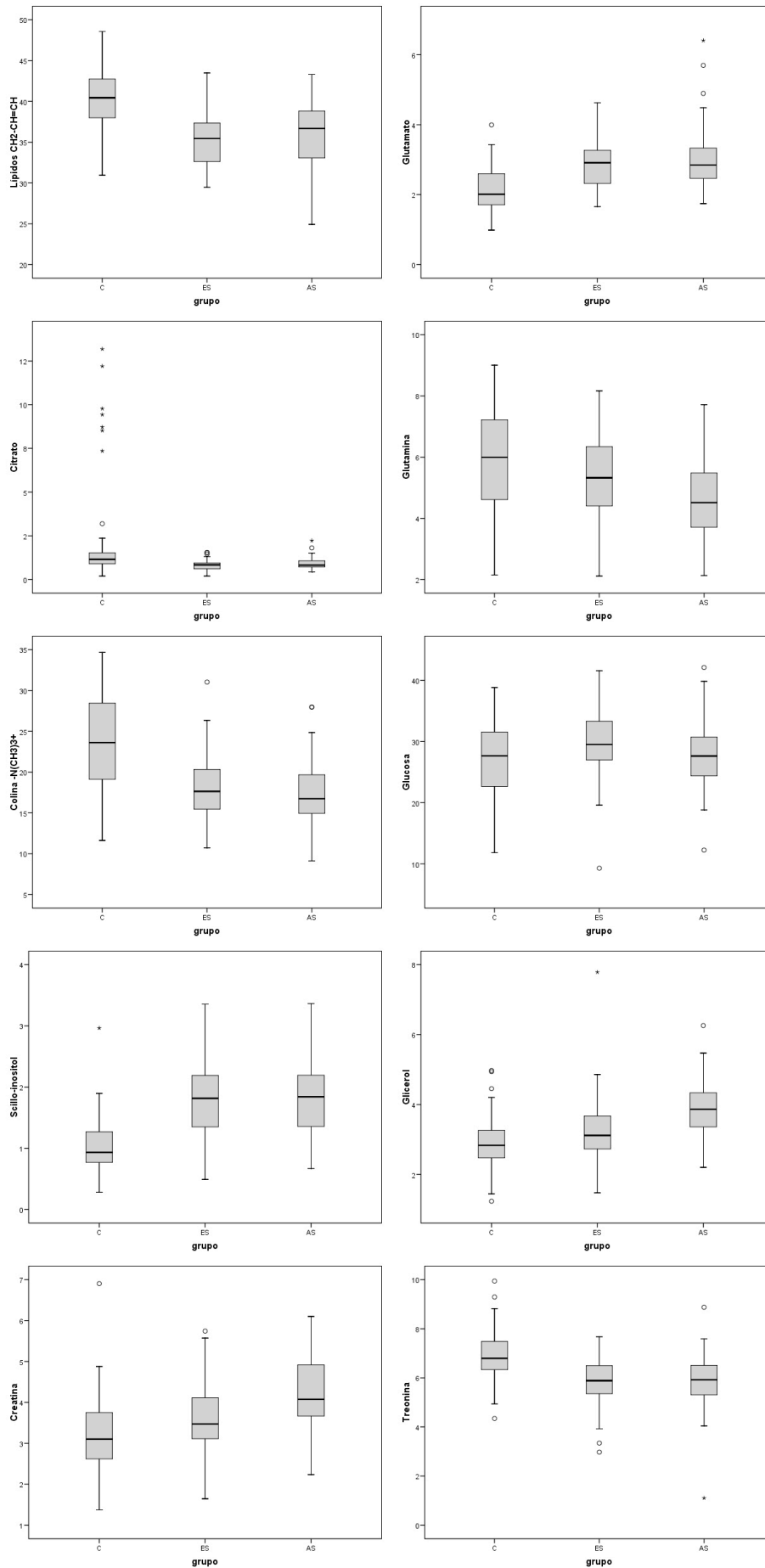
**Tabla 14:** Cuantificación de los metabolitos más significativos derivados del modelo OPLS-DA obtenido para los grupos de pacientes con CPNM en estadios tempranos y avanzados de CPNM.

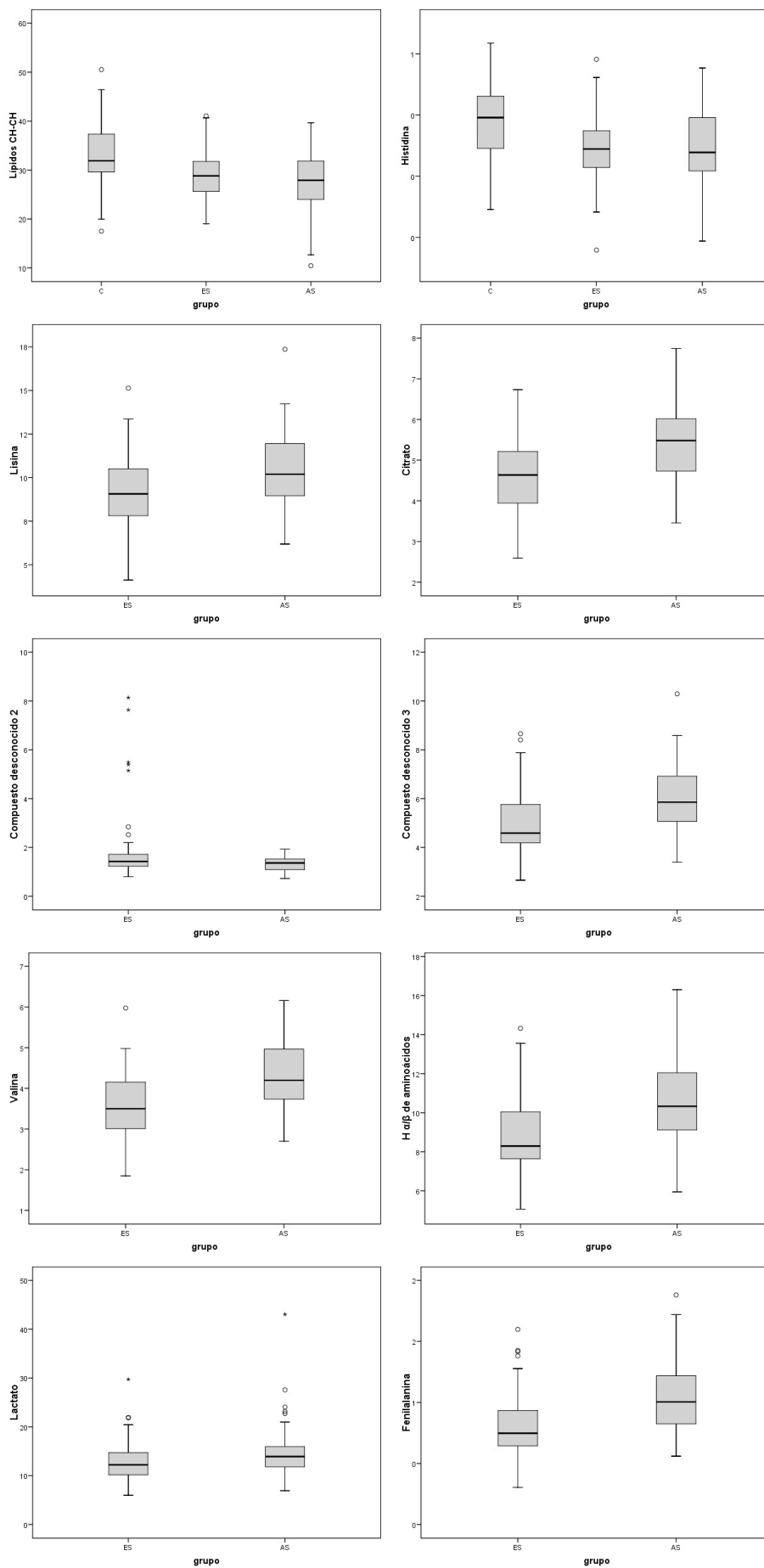
Metabolito	$\delta$ <sup>1</sup> H (ppm) <sup>a</sup>	Intensidad media $\pm$ s.e.m. (unidades arbitrarias)		% variación	t <sup>b</sup>
		Pacientes con CPNM temprano	Pacientes con CPNM avanzado		
HDL	0.79-0.85	45.72 $\pm$ 1.03	45.70 $\pm$ 1.00	-0.06	0.9054
VLDL	0.85-0.91	65.66 $\pm$ 0.90	64.71 $\pm$ 1.22	-1.43	0.8602
Leucina/Isoleucina	0.91-0.97	15.89 $\pm$ 0.32	17.41 $\pm$ 0.36	9.59	0.0039*
3-Hidroxi butirato	1.15-1.18	12.60 $\pm$ 0.34	12.09 $\pm$ 0.32	-4.02	0.2161
LDL/VLDL (CH <sub>2</sub> ) <sub>n</sub>	1.20-1.31	203.92 $\pm$ 4.29	191.89 $\pm$ 4.61	-5.90	0.1151
Ácido adipico	1.52-1.60	14.61 $\pm$ 0.48	13.00 $\pm$ 0.44	-11.01	0.0248*
Lisina	1.65-1.75	9.15 $\pm$ 0.25	10.36 $\pm$ 0.26	13.16	0.0025*
Lípidos CH <sub>2</sub> -C=C	1.93-2.02	35.27 $\pm$ 0.38	35.84 $\pm$ 0.48	1.60	0.1403
N-acetilcisteína	2.02-2.05	24.51 $\pm$ 0.48	29.39 $\pm$ 0.59	19.90	0.0000*
Glutamato	2.05-2.09	10.09 $\pm$ 0.23	12.30 $\pm$ 0.27	21.86	0.0000*
Lípidos CH <sub>2</sub> CO	2.19-2.26	14.55 $\pm$ 0.40	11.85 $\pm$ 0.37	-18.54	0.0000*
Glutamina	2.42-2.47	5.91 $\pm$ 0.16	5.03 $\pm$ 0.15	-14.85	0.0005*
Citrato	2.64-2.69	4.66 $\pm$ 0.10	5.41 $\pm$ 0.11	16.26	0.0000*
Compuesto Desconocido 2	2.79-2.82	1.81 $\pm$ 0.17	1.34 $\pm$ 0.03	-26.30	0.0620
Colina N(CH <sub>3</sub> ) <sub>3</sub> <sup>+</sup>	3.18-3.21	18.05 $\pm$ 0.48	17.17 $\pm$ 0.46	-4.86	0.2081
Compuesto Desconocido 3	3.56-3.58	4.98 $\pm$ 0.16	5.96 $\pm$ 0.16	19.71	0.0000*
Valina	3.59-3.61	3.55 $\pm$ 0.09	4.30 $\pm$ 0.10	21.05	0.0000*
Glicerol	3.63-3.67	13.49 $\pm$ 0.36	15.76 $\pm$ 0.38	16.82	0.0001*
Creatina	3.91-3.92	1.25 $\pm$ 0.04	1.43 $\pm$ 0.04	14.43	0.0031*
H $\alpha/\beta$ de aminoácidos	3.92-4.02	8.86 $\pm$ 0.25	10.54 $\pm$ 0.26	18.92	0.0000*
Lactato	4.08-4.13	13.01 $\pm$ 0.50	14.59 $\pm$ 0.64	12.15	0.0274*
Glucosa	5.21-5.24	13.06 $\pm$ 0.36	11.86 $\pm$ 0.36	-9.19	0.0058*
Lípidos CH=CH	5.24-5.37	28.89 $\pm$ 0.55	27.51 $\pm$ 0.69	-4.79	0.2948
Fenilalanina	7.40-7.43	0.31 $\pm$ 0.02	0.46 $\pm$ 0.03	52.10	0.0000*

\* Diferencias estadísticamente significativas (p &lt; 0.05)

<sup>a</sup> Región del espectro de RMN utilizada para la cuantificación<sup>b</sup> p valor calculado mediante la prueba de Mann Whitney







**Figura 41:** Diagramas de cajas correspondientes a las intensidades relativas (expresado en unidades arbitrarias; u.a.) obtenidas para los metabolitos que presentaron variaciones significativas en los tres grupos de muestras incluidos en el estudio: voluntarios sanos (C), pacientes con CPNM en estadios tempranos (ES) y pacientes con CPNM en estadios avanzados (AS). Las cajas incluyen el 50% de los casos de cada grupo, la línea central indica la mediana del grupo, y los extremos el límite superior e inferior del intervalo de confianza del 95%. “o” representa los *outliers* moderados y “\*” *outliers* extremos.

## 2.5.- Modelo de regresión logística

El análisis de regresión logística es un método estadístico de gran utilidad en la generación de modelos predictivos en los que una o varias variables independientes puedan explicar matemáticamente la probabilidad de que ocurra un evento (p.ej., sano/enfermo). Con esta finalidad, se procedió al análisis por regresión logística de las variables identificadas como estadísticamente significativas en los modelos de discriminación entre individuos sanos y pacientes diagnosticados con CPNM (19 metabolitos) y entre pacientes con CPNM en estadios tempranos y avanzados (20 metabolitos).

El primer paso consistió en el análisis por regresión logística de cada una de las variables de manera individual. Por tanto, para cada metabolito incluido en este análisis univariante se obtuvo tanto el coeficiente de la regresión logística correspondiente ( $\beta$ ) como su significación estadística (p valor). En las **Tablas 15** y **16** se incluyen los resultados obtenidos.

**Tabla 15:** Metabolitos significativos en la discriminación de individuos sanos y pacientes con CPNM. Análisis univariante de regresión logística.

METABOLITO	$\beta$	OR <sup>a</sup>	1/OR	p-valor
HDL	-0.08	0.92	<b>1.02</b>	7.92E-08 <sup>*</sup>
Leucina/Isoleucina	0.36	<b>1.43</b>	0.70	6.28E-07 <sup>*</sup>
3-Hidroxibutirato	1.22	<b>3.38</b>	0.30	7.88E-09 <sup>*</sup>
LDL/VLDL	-0.02	0.98	<b>1.02</b>	1.12E-05 <sup>*</sup>
Compuesto desconocido 1	0.92	<b>2.50</b>	0.40	4.22E-06 <sup>*</sup>
Ácido Adípico	-0.09	0.91	<b>1.10</b>	2.53E-03 <sup>*</sup>
Acetato	1.74	<b>5.71</b>	0.18	1.18E-04 <sup>*</sup>
Lípidos CH <sub>2</sub> C=C	-0.33	0.72	<b>1.40</b>	3.21E-10 <sup>*</sup>
N-Acetilcisteína	0.22	<b>1.24</b>	0.81	3.43E-06 <sup>*</sup>
Glutamato	1.86	<b>6.45</b>	0.16	9.81E-10 <sup>*</sup>
Glutamina	-0.40	0.67	<b>1.50</b>	1.80E-04 <sup>*</sup>
Colina -N(CH <sub>3</sub> ) <sup>3+</sup>	0.25	0.78	<b>1.28</b>	6.44E-11 <sup>*</sup>
Scillo-inositol	2.89	<b>18.00</b>	0.06	1.62E-11 <sup>*</sup>
Glicerol	1.12	<b>3.07</b>	0.33	5.90E-07 <sup>*</sup>
Creatina	0.97	<b>2.65</b>	0.38	1.40E-06 <sup>*</sup>
Lactato	0.25	<b>1.28</b>	0.78	1.58E-06 <sup>*</sup>



METABOLITO	$\beta$	OR <sup>a</sup>	1/OR	p-valor
Treonina	-1.14	0.32	<b>3.08</b>	1.07E-08*
Lípidos CH=CH	-0.15	0.86	<b>1.17</b>	6.34E-07*
Histidina	-4.61	0.01	<b>201.70</b>	5.29E-05*

<sup>a</sup> Exponente del coeficiente de regresión logística: *Odds Ratio* (OR)

\* Coeficiente de regresión logística estadísticamente significativo ( $p < 0.05$ )

De manera general, cuando se aplica el análisis por regresión logística al estudio de la probabilidad de padecer una enfermedad, o de que ésta progrese a un estadio más avanzado, se consideran variables que aumentan el riesgo todas aquellas con valores de OR superiores a 1, y variables protectoras aquellas con un OR inferior a 1.

**Tabla 16:** Metabolitos significativos en la discriminación entre pacientes con CPNM en estadios tempranos/ avanzados. Análisis univariante de regresión logística.

METABOLITO	$\beta$	OR <sup>a</sup>	1/OR	p-valor
Leucina/Isoleucina	0.19	<b>1.21</b>	0.83	4.22E-03*
Ácido adípico	-0.10	0.90	<b>1.12</b>	1.94E-02*
Lisina	0.27	<b>1.31</b>	0.77	2.45E-03*
N-acetilcisteína	0.24	<b>1.27</b>	0.79	6.12E-07*
Glutamato	0.50	<b>1.65</b>	0.60	1.06E-06*
Lípidos CH <sub>2</sub> CO	-0.25	0.78	<b>1.29</b>	2.86E-05*
Glutamina	-0.49	0.61	<b>1.65</b>	4.43E-04*
Citrato	0.90	<b>2.46</b>	0.41	3.26E-05*
Compuesto desconocido 3	0.54	<b>1.72</b>	0.58	1.55E-04*
Valina	1.14	<b>3.12</b>	0.32	7.31E-06*
Glicerol	0.24	<b>1.27</b>	0.79	1.56E-04*
Creatina	1.55	<b>4.72</b>	0.21	4.27E-03*
H de aminoácidos	0.36	<b>1.43</b>	0.70	7.12E-05*
Lactato	0.08	1.08	0.93	6.77E-02
Glucosa	-0.14	0.87	<b>1.14</b>	2.56E-02*
Fenilalanina	4.34	<b>76.72</b>	0.01	2.62E-04*

<sup>a</sup> Exponente del coeficiente de regresión logística: *Odds Ratio* (OR)

\* Coeficiente de regresión logística estadísticamente significativo ( $p < 0.05$ )

El siguiente paso del análisis consistió en la regresión logística multivariante, en la que se incluyó para cada modelo todos los metabolitos que resultaron estadísticamente significativos en la regresión logística univariante. Este análisis permitió identificar los metabolitos que es necesario considerar para la correcta discriminación entre los perfiles metabólicos de individuos sanos y pacientes diagnosticados con CPNM (**Tabla 17**) y entre pacientes con CPNM en estadios tempranos y avanzados (**Tabla 18**).

**Tabla 17:** Regresión logística multivariante de los metabolitos identificados como significativos en la discriminación entre individuos sanos y pacientes con CPNM.

METABOLITO	$\beta$	OR <sup>a</sup>	1/OR	p valor
3-Hidroxibutirato	1.33	<b>3.77</b>	0.27	3.36E-03*
Glutamato	2.29	<b>9.91</b>	0.10	1.07E-02*
Glutamina	-1.48	0.23	<b>4.40</b>	3.66E-03*
Colina -N(CH <sub>3</sub> ) <sup>3+</sup>	-0.44	0.65	<b>1.55</b>	3.93E-04*
Scillo-inositol	2.97	<b>19.39</b>	0.05	1.86E-03*
Treonina	-1.61	0.20	<b>5.02</b>	7.29E-03*
Constante del modelo	11.33	83026.88	1.20E-05	4.39E-02*

<sup>a</sup> Exponente del coeficiente de regresión logística multivariante: *Odds Ratio* ajustado (OR)

\* Coeficiente de regresión logística estadísticamente significativo ( $p < 0.05$ )

**Tabla 18:** Regresión logística multivariante de los metabolitos identificados como significativos en la discriminación entre pacientes con CPNM en estadios tempranos y avanzados.

METABOLITO	$\beta$	OR <sup>a</sup>	1/OR	p valor
Ácido adípico	1.26	<b>3.51</b>	0.29	2.59E-05*
Lípidos CH <sub>2</sub> CO	-1.27	0.28	<b>3.58</b>	1.38E-05*
Glutamina	-1.38	0.25	<b>3.98</b>	9.65E-06*
Citrato	2.34	<b>10.37</b>	9.64E-02	1.33E-03*
Valina	3.41	<b>30.23</b>	3.31E-02	1.46E-03*
H $\alpha/\beta$ de aminoácidos	-0.99	0.37	<b>2.68</b>	9.06E-03*
Constante	-8.52	2.00E-04	5024.92	2.23E-02*

<sup>a</sup> Exponente del coeficiente de regresión logística multivariante: *Odds Ratio* ajustado (OR)

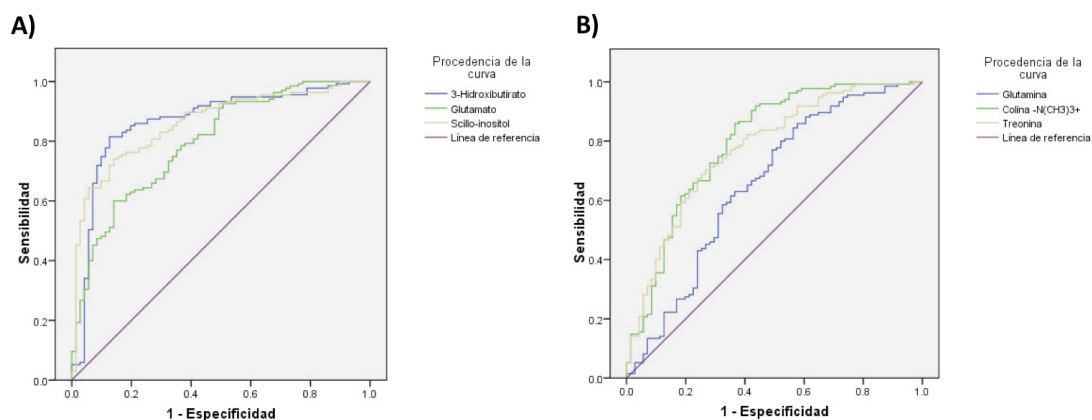
\* Coeficiente de regresión logística estadísticamente significativo ( $p < 0.05$ )

En este análisis, los valores de OR calculados para cada metabolito se conocen como OR ajustados, ya que incorporan el efecto de la presencia del resto de las variables. Por tanto, quedaron excluidas de la ecuación de la regresión logística multivariante todas aquellas variables que no presentaron niveles de significación estadística en el modelo de regresión multivariante.

## 2.6.- Valoración de la capacidad predictiva de los metabolitos identificados

A partir de los resultados obtenidos en la regresión logística multivariante, se calcularon los valores de AUC en las curvas ROC de los seis metabolitos identificados como esenciales en cada uno de los modelos evaluados. La **Figura 42** muestra las curvas ROC obtenidas para los metabolitos incluidos en la ecuación de regresión logística entre pacientes con CPNM e individuos sanos. Por otra parte, en la **Tabla 19** se incluyen los valores de AUC calculados para cada una de estas curvas ROC así como los puntos de corte óptimos, obtenidos para el valor máximo

de sensibilidad+especificidad (índice de Youden), y los valores de sensibilidad y especificidad que se obtendrían al utilizar ese valor como punto de corte para la predicción.



**Figura 42:** Curvas ROC calculadas para los metabolitos identificados como esenciales en la discriminación entre individuos sanos y pacientes con CPNM: **A)** metabolitos que presentan niveles más elevados en el grupo de pacientes con CPNM y **B)** metabolitos que presentan niveles inferiores en el grupo de pacientes con CPNM.

Como se ha comentado anteriormente, en general, se consideran valores de AUC aceptables para un marcador de diagnóstico a partir de 0.75 (Fan et al., 2006). Estos son los valores que sería deseable obtener en el caso de la comparación entre individuos sanos y pacientes diagnosticados con CPNM, donde el objetivo es poder identificar biomarcadores que puedan hacer posible la detección de la enfermedad de manera temprana y no invasiva. En este caso, todos los metabolitos identificados como esenciales en la predicción, a excepción de la glutamina, presentaron valores de AUC > 0.75.

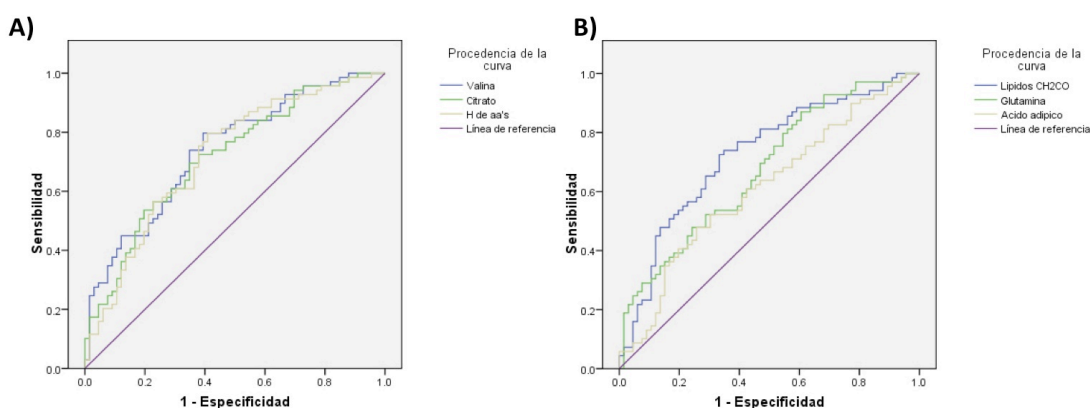
**Tabla 19:** Valores de AUC, puntos de corte, especificidad y sensibilidad obtenidos a partir las curvas ROC de los metabolitos identificados como esenciales en la discriminación entre individuos sanos y pacientes con CPNM.

METABOLITO	AUC	I.C. 95% <sup>a</sup>		Punto de corte	Sensibilidad	Especificidad
		Inf.	Sup.			
3-Hidroxi-butirato	0.87	0.81	0.92	5.35	81%	87%
Glutamato	0.80	0.73	0.86	2.75	60%	86%
Glutamina	0.66	0.57	0.74	6.44	86%	42%
Colina -N(CH <sub>3</sub> ) <sup>3+</sup>	0.80	0.73	0.86	21.17	86%	63%
Scillo-inositol	0.86	0.81	0.91	1.36	74%	86%
Treonina	0.77	0.71	0.84	6.37	70%	73%

<sup>a</sup> Intervalo de confianza asintótico al 95% del valor de AUC, límites inferior y superior.

Por otro lado, se determinaron también los valores de AUC para los metabolitos identificados a partir del modelo OPLS-DA obtenido para la comparación entre los perfiles metabólicos de pacientes con CPNM en distintos estadios. La **Figura 43** y

**Tabla 20** incluyen los resultados obtenidos para el cálculo de las curvas ROC de los metabolitos incluidos en el modelo de regresión logística multivariante.



**Figura 43:** Curvas ROC calculadas para los metabolitos identificados como esenciales en la discriminación entre pacientes con CPNM en estadios tempranos y avanzados: **A)** metabolitos que presentan niveles más elevados en el grupo de pacientes con CPNM en estadios avanzados y **B)** metabolitos que presentan niveles inferiores en el grupo de pacientes con CPNM en estadios avanzados.

A diferencia del modelo entre individuos sanos y pacientes con CPNM, y tal y como se expuso al inicio de este apartado, la identificación de los metabolitos implicados en la discriminación entre estadios estaba dirigido a conocer mejor los procesos moleculares implicados en la progresión de la enfermedad.

**Tabla 20:** Valores de AUC, puntos de corte, especificidad y sensibilidad obtenidos a partir las curvas ROC de los metabolitos identificados como esenciales en la discriminación entre entre pacientes con CPNM en estadios tempranos y avanzados.

METABOLITO	AUC	I.C. 95% <sup>a</sup>		Punto de corte	Sensibilidad	Especificidad
		Inf.	Sup.			
Ácido adípico	0.61	0.52	0.71	12.30	48%	74%
Lípidos CH <sub>2</sub> CO	0.72	0.64	0.81	13.51	73%	67%
Glutamina	0.68	0.59	0.77	6.34	87%	39%
Citrato	0.72	0.63	0.80	4.93	70%	65%
Valina	0.74	0.66	0.82	3.68	80%	61%
H $\alpha/\beta$ de aminoácidos	0.72	0.63	0.80	8.70	80%	59%

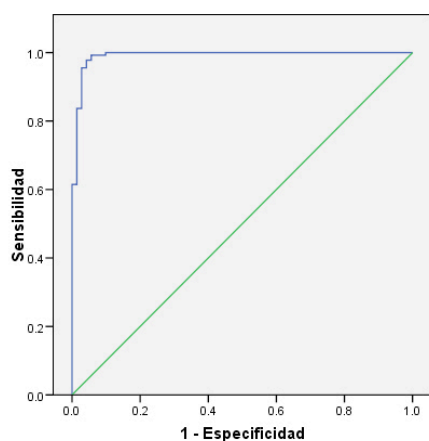
<sup>a</sup> Intervalo de confianza asintótico al 95% del valor de AUC, límites inferior y superior.

## 2.7.- Valoración de la capacidad predictiva del modelo de regresión

Con la finalidad de poder comparar la capacidad predictiva obtenida a través de los metabolitos identificados en cada modelo de manera individual y la obtenida utilizando la combinación obtenida a través de la regresión logística, en la siguiente fase del análisis se calcularon las curvas ROC correspondientes a la regresión logística multivariante de los datos obtenidos a partir de los modelos OPLS-DA

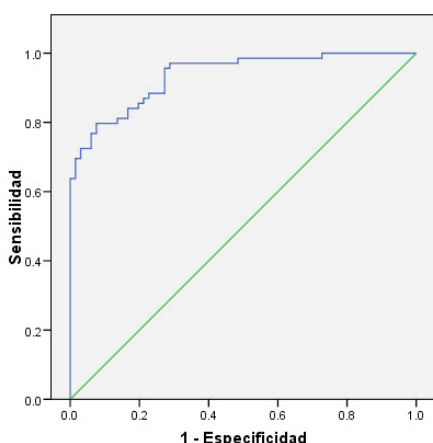
entre individuos sanos y pacientes con CPNM, y entre pacientes con CPNM en distintos estadios. Para ello se calcularon partiendo de los valores de  $\beta$  y las constantes de regresión multivariante incluidas en las **Tablas 17 y 18**, el valor correspondiente a la ecuación de regresión para cada una de las muestras incluidas en los dos modelos.

En las **Figuras 44 y 45** se incluyen los gráficos obtenidos para las curvas ROC de ambos modelos de regresión logística, así como los valores calculados para el AUC. Como se puede apreciar, los valores de AUC para estas curvas ROC combinadas supera a los obtenidos de manera individual para cualquiera de los metabolitos identificados como revelantes en los modelos OPLS-DA correspondientes.



AUC	Intervalo de confianza asintótico al 95%	
	Límite superior	Límite inferior
0.99	0.98	1.00

**Figura 44:** Curva ROC calculada usando los datos obtenidos para la ecuación de la regresión logística en el análisis de la discriminación entre individuos sanos y pacientes con CPNM.



AUC	Intervalo de confianza asintótico al 95%	
	Límite superior	Límite inferior
0.94	0.90	0.97

**Figura 45:** Curva ROC calculada usando los datos obtenidos para la ecuación de la regresión logística en el análisis de la discriminación entre pacientes con CPNM en estadios tempranos y avanzados.

A partir de estos datos, se determinaron los valores óptimos para el punto de corte en el que el valor de sensibilidad+especificidad era máximo (índice de Youden). La **Tabla 21** incluye los valores de especificidad y sensibilidad asociados a estos puntos de corte en los dos modelos analizados.

**Tabla 21:** Capacidad predictiva calculada a partir de las curvas ROC de los modelos de regresión logística multivariante.

MODELO REGRESIÓN MULTIVARIANTE	Punto de corte	Sensibilidad	Especificidad
Individuos sanos / pacientes CPNM	0.33	99.26 %	94.37 %
CPNM temprano / CPNM avanzado	63.62	79.70 %	92.4 %

En general, los resultados obtenidos ponen de manifiesto la existencia de importantes diferencias en el perfil metabólico de pacientes con CPNM comparado con el de individuos sanos, y de cambios significativos en el perfil en función del estadio de la enfermedad. Se ha estudiado el valor predictivo de los cambios observados, tanto en el diagnóstico como en la estadificación de la enfermedad, a través de las ecuaciones de regresión logística multivariantes obtenidas, en las que cada uno de los metabolitos identificados como esenciales es ponderado en función de su correlación con la probabilidad de padecer la enfermedad o pertenecer al grupo de pacientes con CPNM en estadio avanzado, respectivamente. En cualquier caso, será necesario realizar futuros estudios de validación con un mayor número de muestras independientes para poder evaluar el valor predictivo de estos metabolitos en una nueva cohorte de pacientes. Disponer de un mayor número de muestras, recogidas de manera prospectiva, permitiría además poder estudiar las diferencias encontradas entre distintos subgrupos de pacientes que pudieran ser de interés (p.ej., histología, estado mutacional, etc.).

## 2.8.- Interpretación biológica

Recientemente han aparecido varios trabajos dirigidos a la caracterización del perfil metabólico de muestras biológicas obtenidas de pacientes diagnosticados con CP. En general, la mayoría de esos trabajos incluyen, dentro del mismo grupo de estudio, pacientes con histologías y/o estadios de la enfermedad muy distintos y, muchos de ellos, basan los análisis en un número muy reducido de muestras. En este sentido, algunos de los trabajos más relevantes son los realizados por Gil y colaboradores (Barros et al., 2010; Carrola et al., 2011; Rocha et al., 2011), en los

que se incluyeron un número de muestras relativamente alto (70-80 muestras por grupo en los estudios con plasma y orina). En la selección de pacientes en estos estudios no se tuvo en cuenta ni la histología del tumor, ni el estadio de la enfermedad, lo que puede explicar que el análisis de los perfiles metabólicos en plasma y orina no revelase diferencias significativas entre las distintas histologías y/o estadios incluidos. Sin embargo, en un estudio reciente realizado por EM, empleando muestras de plasma, se observaron diferencias entre los perfiles metabólicos de individuos sanos (n=28) y pacientes con adenocarcinoma en estadio I (n=31) (Wen et al., 2013).

Hasta el momento, el único trabajo realizado con muestras de suero es el publicado por Hori y colaboradores (Hori et al., 2011). En este estudio, realizado por EM, se incluyeron 33 muestras de suero de pacientes diagnosticados con CP (adenocarcinomas, carcinoma epidermoide y CMP) y fue posible observar diferencias entre los perfiles metabólicos característicos de distintas histologías y estadio.

El trabajo aquí presentado representa el primer estudio dirigido a caracterizar el perfil metabólico específico de pacientes con CPNM en el que se han observado diferencias significativas entre individuos sanos y pacientes con CPNM, así como entre diferentes estadios de la enfermedad. Este análisis ha revelado que el perfil metabólico de pacientes diagnosticados con CPNM, comparado con el de individuos sanos, se caracteriza por la presencia de niveles más elevados de 3-hidroxiacetil-CoA, glutamato y scillo-inositol, y por niveles inferiores de glutamina, colina y treonina. Por otro lado, el análisis de las diferencias entre los perfiles metabólicos en suero de pacientes con CPNM en distintos estadios (tempranos vs avanzados) ha permitido identificar un conjunto de metabolitos (leucina, isoleucina, ácido adípico, lípidos, N-acetilcisteína, valina, creatina) que presentan niveles significativamente diferentes en esos dos grupos, y que pueden estar implicados en los cambios metabólicos asociados a la enfermedad.

Los resultados de este estudio revelan que el perfil metabólico en suero de pacientes con CPNM está caracterizado por un conjunto de metabolitos que varían de forma muy específica. Algunos de los metabolitos identificados como relevantes en este estudio coinciden con los identificados en los estudios de Rocha y colaboradores (Rocha et al., 2011) (HDL, glutamina, lactato, histidina, citrato, lisina), Wen y colaboradores (Wen et al., 2013) (glutamina y treonina) y Hori y

colaboradores (Hori et al., 2011) (3-hidroxiubutirato). Sin embargo, otros de los metabolitos identificados aportan nueva información sobre los cambios bioquímicos asociados a la aparición de la enfermedad.

La existencia de niveles inferiores de HDL/LDL/VLDL y, en general, de lípidos en los pacientes con CPNM coincide con lo descrito anteriormente en el caso de MM. En ese sentido, y aunque no se conocen en detalle los mecanismos moleculares responsables de este fenómeno, la relación entre niveles inferiores de lipoproteínas en suero y los procesos oncológicos ha sido también descrita en otros tumores (Muntoni et al., 2009). En general, se ha encontrado que el factor con mayor capacidad de discriminación entre pacientes oncológicos e individuos sanos está relacionado con los niveles de HDL, y se especula que las variaciones en las concentraciones de lípidos podrían estar relacionadas con una mayor utilización de colesterol, componente esencial de las membranas celulares, por parte de las células tumorales. Probablemente, estos cambios en el metabolismo lipídico podrían también explicar las variaciones observadas en las concentraciones de ácido adípico entre pacientes en diferentes estadios de CPNM, un metabolito que está relacionado con anomalías en el metabolismo de los ácidos grasos. De acuerdo con estos resultados, Barros y colaboradores (Barros et al., 2010) observaron una mayor presencia, en relación a los niveles encontrados en individuos sanos, de lípidos, así como de colina, glicerofosfocolina, y fosfocolina en tejido tumoral de pacientes con CPNM. Así, los niveles inferiores de colina observados en suero, en el grupo de pacientes con CPNM, podrían estar relacionados con la mayor demanda, dada la elevada tasa de proliferación celular, que este metabolito juega como precursor de fosfolípidos de membrana en las células tumorales (Bañez-Coronel et al., 2008; Gallego-Ortega et al., 2009). De forma coincidente también con los resultados obtenidos en nuestro estudio en suero de pacientes con CPNM, Barros y colaboradores (Barros et al., 2010) observaron niveles reducidos de scillo-inositol en tejido tumoral, al igual que ha sido descrito en otros estudios en los que se analizaron muestras de tejidos correspondientes a neoplasmas renales (Righi et al., 2007). Otros estudios realizados con tumores de mama (Beckonert et al., 2003) han revelado niveles inferiores de mio-inositol en tejido tumoral en comparación con el tejido sano. Estos osmolitos juegan un papel importante en la regulación del volumen celular, por lo que el aumento observado en los niveles de scillo-inositol en suero de los pacientes con CPNM sería compatible con los niveles inferiores encontrados en el análisis de distintos tejidos



tumorales.

Por otro lado, de acuerdo con los resultados obtenidos por otros autores empleando muestras de plasma y tejido (Barros et al., 2010; Rocha et al., 2011), el perfil metabólico de pacientes con CPNM en suero se caracteriza por la presencia de niveles significativamente elevados de lactato y disminuidos de glucosa. Estos resultados concuerdan con lo esperado, dado que la elevada captación de glucosa y su conversión a lactato se ha observado en distintos tejidos tumorales (Chan et al., 2009; Yang et al., 2007), un fenómeno relacionado con el conocido efecto Warburg (Warburg, 1956).

Existen en la literatura numerosos estudios que reportan variaciones en los perfiles aminoacídicos en suero de pacientes con distintos tipos de cáncer (Lai et al., 2005). A pesar de la heterogeneidad de los resultados obtenidos en función del tipo de cáncer y del estadio de la enfermedad en los pacientes estudiados, todos ellos muestran alteraciones en el perfil de aminoácidos en comparación con el de individuos sanos, relacionándolo con el estado hipermetabólico y la mayor demanda de aminoácidos como consecuencia de la presencia y crecimiento del tumor (Pisters & Pearlstone, 1993). En nuestro estudio, los pacientes con CPNM presentan niveles superiores, comparados con los individuos sanos, de leucina, isoleucina y glutamato, e inferiores de glutamina, treonina e histidina, lo que coincide con observaciones realizadas en estudios anteriores (Rocha et al., 2011) llevados a cabo en plasma. Por otro lado, se observa que los estadios avanzados de CPNM, en relación a los pacientes en estadios tempranos de la enfermedad, presentan perfiles metabólicos caracterizados por niveles elevados de citrato, lisina, valina y fenilalanina. Estos resultados están de acuerdo con los obtenidos por Maeda y colaboradores (Maeda et al. 2010), pero no con los de Rocha y colaboradores (Rocha et al. 2011). Esta discrepancia puede deberse a la distinta distribución de estadios de la enfermedad en los grupos de pacientes estudiados, ya que mientras en el estudio de Maeda se incluyeron pacientes con CPNM en estadios III y IV, el de Rocha se centró en pacientes en estadio I-III.

En relación con los cambios observados en la composición aminoacídica, cabe destacar el descenso en los niveles de glutamina en los pacientes con CPNM, lo que coincide con lo observado en otros pacientes oncológicos (Gao et al., 2008; Urayama et al., 2010; Zira et al., 2010), y que se ha asociado al aumento de la actividad metabólica (glucólisis y glutaminólisis) derivada de las condiciones de

hipoxia e hipermetabolismo observadas en el entorno tumoral (Eigenbrodt et al., 1998). Estos resultados pueden también juzgarse en el contexto de estudios recientes de proteómica realizados con células derivadas de adenocarcinoma pancreático y que ponen de manifiesto el papel que juega, en estas condiciones, la glutamina como fuente de nitrógeno en la síntesis de nucleótidos y aminoácidos (Zhou et al., 2012). En estos estudios también se observó una disminución en los niveles de la alanina aminotransferasa, la enzima que cataliza la producción de alanina a partir de glutamato. Este hecho, unido a la posible hidrólisis de la glutamina para la generación de glutamato y amonio, que contribuiría a equilibrar el pH en las células tumorales contrarrestando el exceso de lactato producido como consecuencia del efecto Warburg, podría explicar los niveles de glutamato elevados que se observan en el grupo de pacientes diagnosticados con CPNM incluidos en nuestro estudio.

Otro metabolito que juega un papel fundamental en la discriminación de perfiles metabólicos en suero es el 3-hidroxiacetato, cuyos niveles se encuentran significativamente elevados en los pacientes con CPNM. Este resultado coincide con los obtenidos por Hori y colaboradores (Hori et al., 2011) en pacientes con CPNM en estadios tempranos de la enfermedad, y con los de Carrola y colaboradores (Carrola et al., 2011) en muestras de orina de pacientes con CP. Los niveles de 3-hidroxiacetato, junto a los de acetato en el suero de pacientes con CPNM, podrían estar relacionados con la activación de cuerpos cetónicos y del ciclo de Krebs como consecuencia del elevado metabolismo energético en las células tumorales.

En líneas generales, puede decirse que los resultados obtenidos en el análisis de los perfiles metabólicos en suero de pacientes con CPNM muestra un enorme grado de concordancia con el descrito en estudios previos, y ha permitido la identificación de una “huella metabólica” que permite discriminar pacientes con CPNM de individuos sanos con una sensibilidad+especificidad del 97%. Por otro lado, el análisis de los perfiles de pacientes en distintos estadios de CPNM aporta nueva información sobre los mecanismos moleculares implicados en la progresión de la enfermedad.

## **2.8.- Relevancia de los resultados**

Los resultados ponen de manifiesto la existencia de importantes diferencias en el perfil metabólico en suero de pacientes con CPNM cuando se compara con el de individuos sanos, pudiendo además diferenciar los perfiles característicos de los distintos estadios de la enfermedad.

La comparación con estudios previos, realizados con distintos conjuntos de muestras y pacientes, ha revelado la existencia de similitudes que permiten aventurar la utilidad diagnóstica de los metabolitos identificados como relevantes en la discriminación (Wen et al., 2013). Por otro lado, la información extraída de este análisis ha permitido identificar por primera vez una combinación concreta de seis metabolitos que, en caso de ser validada en estudios prospectivos, podría constituir un método sensible, específico, mínimamente invasivo y fácilmente implementable en la práctica clínica para el diagnóstico temprano del CPNM. Finalmente, se ha caracterizado un conjunto de metabolitos que podrían jugar un papel importante en el seguimiento de la progresión de la enfermedad.



## **IV. CONCLUSIONES FINALES**



I.- La metabolómica por RMN permite caracterizar el perfil metabólico de muestras biológicas y la identificación de biomarcadores asociados a situaciones patológicas.

II.- La estrecha relación entre el perfil metabólico y el fenotipo permite la monitorización de cambios dinámicos relacionados con: a) la progresión de una enfermedad, b) la remisión en respuesta al tratamiento farmacológico y c) los efectos del tratamiento no ligados a la respuesta terapéutica.

III.- Los pacientes diagnosticados con MM presentan, en el momento del diagnóstico y en comparación con individuos sanos, un perfil metabólico en suero caracterizado por niveles superiores de isoleucina, fenilalanina, arginina, tirosina y acetato, e inferiores de 3-hidroxiacetato, colesterol, lisina, lípidos, colina y glutamina.

IV.- Los cambios observados entre los perfiles metabólicos de los pacientes con MM en el momento del diagnóstico y tras la remisión completa están fundamentalmente asociados a la respuesta al tratamiento.

V.- El análisis global de los perfiles metabólicos correspondientes a los distintos grupos incluidos en el estudio de MM ha permitido la identificación de cambios metabólicos asociados a la enfermedad que no revierten tras la remisión y que podrían ser indicativos de enfermedad residual.

VI.- Un algoritmo de riesgo compuesto por 6 metabolitos (3-hidroxiacetato, treonina, glutamina, glutamato, colina y scillo-inositol) permite discriminar con un 97% de sensibilidad y especificidad individuos sanos de pacientes diagnosticados con CPNM, por lo que podría utilizarse como un método mínimamente invasivo con valor diagnóstico en esta patología.

VII.- La comparación de los perfiles metabólicos en suero de pacientes diagnosticados con CPNM en estadios tempranos y avanzados revela la existencia de diferencias significativas. Estos cambios están caracterizados por variaciones en los niveles de ácido adípico, lípidos, glutamina, citrato, valina y otros aminoácidos, y aportan información sobre rutas metabólicas que podrían estar implicadas en la progresión de la enfermedad.





## **V. BIBLIOGRAFÍA**



- Arn, P. H. Newborn screening: current status. *Health Affairs (Project Hope)*. 2007; **26**(2): 559–66.
- Bañez-Coronel, M., Ramírez de Molina, A., Rodríguez-González, A., Sarmentero, J., Ramos, M. A., García-Cabezas, M. A., García-Oroz, L., & Lacal, J. C. Choline kinase alpha depletion selectively kills tumoral cells. *Current Cancer Drug Targets*. 2008; **8**(8): 709–19.
- Baró, L., Hermoso, J. C., Núñez, M. C., Jiménez-Rios, J. A., & Gil, A. Abnormalities in plasma and red blood cell fatty acid profiles of patients with colorectal cancer. *British Journal of Cancer*. 1998; **77**(11): 1978–83.
- Baron, A., Migita, T., Tang, D., & Loda, M. Fatty acid synthase: a metabolic oncogene in prostate cancer? *Journal of Cellular Biochemistry*. 2004; **91**(1): 47–53.
- Barros, S., Gil, A. M., Goodfellow, B. J., Humpfer, E., Spraul, M., Carreira, I. M., Melo, J. B., & Gomes, A. Metabolic Profiling of Human Lung Cancer Tissue by 1 H High Resolution Magic Angle Spinning ( HRMAS ) NMR Spectroscopy research articles 2010; 319–332.
- Bathen, T. F., Sitter, B., Sjøbakk, T. E., Tessem, M.-B., & Gribbestad, I. S. Magnetic resonance metabolomics of intact tissue: a biotechnological tool in cancer diagnostics and treatment evaluation. *Cancer Research*. 2010; **70**(17): 6692–6.
- Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C., & Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*. 2007; **2**(11): 2692–703.
- Beckonert, O., Monnerjahn, J., Bonk, U., & Leibfritz, D. Visualizing metabolic changes in breast-cancer tissue using 1H-NMR spectroscopy and self-organizing maps. *NMR in Biomedicine*. 2003; **16**(1): 1–11.
- Bertini, I., Cacciatore, S., Jensen, B. V., Schou, J. V., Johansen, J. S., Kruhøffer, M., Luchinat, C., Nielsen, D. L., & Turano, P. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Research*. 2012; **72**(1): 356–64.
- Bhatt, A. N., Mathur, R., Farooque, A., Verma, A., & Dwarakanath, B. S. Cancer biomarkers - current perspectives. *The Indian Journal of Medical Research*. 2010; **132**(August): 129–49.
- Bhattacharyya, S., Epstein, J., & Suva, L. J. Biomarkers that discriminate multiple myeloma patients with or without skeletal involvement detected using SELDI-TOF mass spectrometry and statistical and machine learning tools. *Disease Markers*. 2006; **22**(4): 245–55.
- BMRB - Biological Magnetic Resonance Bank. [www.bmrb.wisc.edu](http://www.bmrb.wisc.edu)
- Body, J. J. Metabolic sequelae of cancers (excluding bone marrow transplantation). *Current Opinion in Clinical Nutrition and Metabolic Care*. 1999; **2**(4): 339–44.
- Brindle, J. T., Antti, H., Holmes, E., Tranter, G., Nicholson, J. K., Bethell, H. W. L., Clarke, S., Schofield, P. M., McKilligin, E., Mosedale, D. E., & Grainger, D. J. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. *Nature Medicine*. 2002; **8**(12): 1439–44.
- Brindle, J. T., Nicholson, J. K., Schofield, P. M., Grainger, D. J., & Holmes, E. Application of chemometrics to 1H NMR spectroscopic data to investigate a

relationship between human serum metabolic profiles and hypertension. *The Analyst*. 2003; **128**(1): 32–6.

Carrola, J., Rocha, M., Barros, S., Gil, A. M., Goodfellow, B. J., Carreira, I. M., Gomes, A., Sousa, V., Carvalho, L., & Duarte, I. F. Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Urine research articles 2011; 221–230.

Chan, E. C. Y., Koh, P. K., Mal, M., Cheah, P. Y., Eu, K. W., Backshall, A., Cavill, R., Nicholson, J. K., & Keun, H. C. Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *Journal of Proteome Research*. 2009; **8**(1): 352–61.

Chang, H., Lee, S.-Y., & Tang, T.-C. Tumor lysis syndrome in patients with light chain multiple myeloma: report of two cases. *Chang Gung Medical Journal*. 2011; **34**(6 Suppl): 70–5.

Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R., & Nicholson, J. K. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; **106**(34): 14728–33.

D'Alessandro, A., & Zolla, L. Metabolomics and cancer drug discovery: let the cells do the talking. *Drug Discovery Today*. 2012; **17**(1-2): 3–9.

DeBerardinis, R. J., & Cheng, T. Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*. 2010; **29**(3): 313–24.

Dettmer, K., Almstetter, M. F., Appel, I. J., Nürnbergger, N., Schlamberger, G., Gronwald, W., Meyer, H. H. D., & Oefner, P. J. Comparison of serum versus plasma collection in gas chromatography--mass spectrometry-based metabolomics. *Electrophoresis*. 2010; **31**(14): 2365–73.

Duarte, I. F., & Gil, A. M. Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2012; **62**: 51–74.

Duarte, I. F., Rocha, C. M., Barros, A. S., Gil, A. M., Goodfellow, B. J., Carreira, I. M., Bernardo, J., Gomes, A., Sousa, V., & Carvalho, L. Can nuclear magnetic resonance (NMR) spectroscopy reveal different metabolic signatures for lung tumours? *Virchows Archiv: an International Journal of Pathology*. 2010; **457**(6): 715–25.

Edwards, B. J. C., & Ph, D. Principles of NMR. In *Process NMR Associates* 1946; (1–51).

Eigenbrodt, E., Kallinowski, F., Ott, M., Mazurek, S., & Vaupel, P. Pyruvate kinase and the interaction of amino acid and carbohydrate metabolism in solid tumors. *Anticancer Research*. 1998; **18**(5A): 3267–74.

Elmer, P., Karley, D., Gupta, D., & Tiwari, A. Review Biomarker for Cancer: A great Promise for Future 2011; **2**(4): 151–157.

Fan, J., Upadhye, S., & Worster, A. Understanding receiver operating characteristic (ROC) curves. *CJEM*. 2006; **8**(1): 19–20.

Fiehn, O., Garvey, W. T., Newman, J. W., Lok, K. H., Hoppel, C. L., & Adams, S. H. Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women. *PloS One*. 2010; **5**(12): e15234.

Fonseca, R., Bergsagel, P. L., Drach, J., Shaughnessy, J., Gutierrez, N., Stewart, A. K., Morgan, G., Van Ness, B., Chesi, M., Minvielle, S., Neri, A., Barlogie, B., Kuehl, W. M., Liebisch, P., Davies, F., Chen-Kiang, S., Durie, B. G. M., Carrasco, R., Sezer, O., Reiman, T., Pilarski, L., & Avet-Loiseau, H. International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. *Leukemia*. 2009; **23**(12): 2210–21.

Gallego-Ortega, D., Ramirez de Molina, A., Ramos, M. A., Valdes-Mora, F., Barderas, M. G., Sarmentero-Estrada, J., & Lacal, J. C. Differential role of human choline kinase alpha and beta enzymes in lipid metabolism: implications in cancer onset and treatment. *PLoS One*. 2009; **4**(11): e7819.

Gao, H., Dong, B., Liu, X., Xuan, H., Huang, Y., & Lin, D. Metabonomic profiling of renal cell carcinoma: high-resolution proton nuclear magnetic resonance spectroscopy of human serum with multivariate data analysis. *Analytica Chimica Acta*. 2008; **624**(2): 269–77.

Genoma España. *Informe de Vigilancia Tecnológica: Biomarcadores para Uso Clínico* 2010; .

GLOBOCAN: Country Fast Stat.  
<http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900>.

Greene, F. L., Page D.L, Fleming I.D, Fritz A.G, Balch C.M, Haller D.G, (2002). AJCC Cancer Staging Manual 6th ed. In Springer (Ed.), 2002; New York.

Griffin, J. L., Atherton, H., Shockcor, J., & Atzori, L. Metabolomics as a tool for cardiac research. *Nature Reviews. Cardiology*. 2011; **8**(11): 630–43.

Grisolia, F. T., & Cohen, P. P. Amino acid analysis of serum proteins in multiple myeloma. *Cancer Research*. 1953; **13**(12): 851–54.

Gu, H.-W., Qi, Y.-P., Xu, N., Ding, J.-H., an, Y.-B., & Chen, H.-W. Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry-Based Metabolomics for Cancer Diagnosis. *Chinese Journal of Analytical Chemistry*. 2012; **40**(12): 1933–1937.

Guan, W., Zhou, M., Hampton, C. Y., Benigno, B. B., Walker, L. D., Gray, A., McDonald, J. F., & Fernández, F. M. Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics*. 2009; **10**: 259.

Halouska, S., & Powers, R. Negative impact of noise on the principal component analysis of NMR data. *Journal of Magnetic Resonance (San Diego, Calif. : 1997)*. 2006; **178**(1): 88–95.

Hanahan, D., & Weinberg, R. a. (a). Hallmarks of cancer: the next generation. *Cell*. 2011; **144**(5): 646–74.

Hanahan, D., & Weinberg, R. A. (b). Hallmarks of cancer: the next generation. *Cell*. 2011; **144**(5): 646–74.

Hanahan, D., Weinberg, R. A., & Francisco, S. The Hallmarks of Cancer Review University of California at San Francisco 2000; **100**: 57–70.

Hori, S., Nishiumi, S., Kobayashi, K., Shinohara, M., Hatakeyama, Y., Kotani, Y., Hatano, N., Maniwa, Y., Nishio, W., Bamba, T., Fukusaki, E., Azuma, T., Takenawa, T., Nishimura, Y., & Yoshida, M. A metabolomic approach to lung cancer. *Lung Cancer (Amsterdam, Netherlands)*. 2011; **74**(2): 284–92.

Iranzo, V., Sirera, R., Bremnes, R. M., Blasco, A., Jantus-Lewintre, E., Tarón, M., Berrocal, A., Blasco, S., Caballero, C., Del Pozo, N., Rosell, R., & Camps, C.

Chemotherapy-induced neutropenia does not correlate with DNA repair gene polymorphisms and treatment efficacy in advanced non-small-cell lung cancer patients. *Clinical Lung Cancer*. 2011; **12**(4): 224–30.

Jantus-Lewintre, E., Sanmartín, E., Sirera, R., Blasco, A., Sanchez, J. J., Tarón, M., Rosell, R., & Camps, C. Combined VEGF-A and VEGFR-2 concentrations in plasma: diagnostic and prognostic implications in patients with advanced NSCLC. *Lung Cancer (Amsterdam, Netherlands)*. 2011; **74**(2): 326–31.

Jantus-lewintre, E., Usó, M., Sanmartín, E., & Camps, C. Update on biomarkers for the detection of lung cancer 2012; .

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: a Cancer Journal for Clinicians*. **61**(2): 69–90.

Kamburov, A., Pentchev, K., Galicka, H., Wierling, C., Lehrach, H., & Herwig, R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Research*. 2011; **39**(Database issue): D712–7.

Kamburov, A., Wierling, C., Lehrach, H., & Herwig, R. ConsensusPathDB--a database for integrating human functional interaction networks. *Nucleic Acids Research*. 2009; **37**(Database issue): D623–8.

Kanehisa, M. Molecular network analysis of diseases and drugs in KEGG. *Methods in Molecular Biology (Clifton, N.J.)*. 2013; **939**: 263–75.

Keun, H. C., Beckonert, O., Griffin, J. L., Richter, C., Moskau, D., Lindon, J. C., & Nicholson, J. K. Cryogenic probe <sup>13</sup>C NMR spectroscopy of urine for metabolomic studies. *Analytical Chemistry*. 2002; **74**(17): 4588–93.

Kim, K., Aronov, P., Zakharkin, S. O., Anderson, D., Perroud, B., Thompson, I. M., & Weiss, R. H. Urine metabolomics analysis for kidney cancer detection and biomarker discovery. *Molecular & Cellular Proteomics : MCP*. 2009; **8**(3): 558–70.

Kind, T., Tolstikov, V., Fiehn, O., & Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry*. 2007; **363**(2): 185–95.

Kirito, K. Therapeutic Approaches for Targeting Hypoxia- Inducible Factor in Multiple Myeloma 2012; .

Kirschenlohr, H. L., Griffin, J. L., Clarke, S. C., Rhydwen, R., Grace, A. A., Schofield, P. M., Brindle, K. M., & Metcalfe, J. C. Proton NMR analysis of plasma is a weak predictor of coronary artery disease. *Nature Medicine*. 2006; **12**(6): 705–10.

Koppenaal D; Lindon JC; Tranter. *Encyclopedia of Spectroscopy and Spectrometry*,. (A. Press, Ed.) 2010; London.

Kroemer, G., & Pouyssegur, J. Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell*. 2008; **13**(6): 472–82.

Kumar, A., Ernst, R. R., & Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochemical and Biophysical Research Communications*. 1980; **95**(1): 1–6.

Lai, H.-S., Lee, J.-C., Lee, P.-H., Wang, S.-T., & Chen, W.-J. Plasma free amino acid profile in cancer patients. *Seminars in Cancer Biology*. 2005; **15**(4): 267–76.

Landgren, O., Kyle, R. A., & Rajkumar, S. V. From myeloma precursor disease to multiple myeloma: new diagnostic concepts and opportunities for early intervention.

- Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*. 2011; **17**(6): 1243–52.
- Lenz, E. M., & Wilson, I. D. Analytical strategies in metabolomics. *Journal of Proteome Research*. 2007; **6**(2): 443–58.
- Lindon, J.C., Holmes, E., & Nicholson, J. K. Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2001; **39**(1): 1–40.
- Lindon, John C., Nicholson, J. K., Holmes, E., & Everett, J. R. Metabolomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance*. 2000; **12**(5): 289–320.
- Luz-Hdez, K. D. la. *Metabolomics*. (U. Roessner, Ed.) 2012; InTech.
- MacIntyre, D. a, Jiménez, B., Lewintre, E. J., Martín, C. R., Schäfer, H., Ballesteros, C. G., Mayans, J. R., Spraul, M., García-Conde, J., & Pineda-Lucena, a. Serum metabolome analysis by 1H-NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. *Leukemia*. 2010; **24**(4): 788–97.
- Mamas, M., Dunn, W. B., Neyses, L., & Goodacre, R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Archives of Toxicology*. 2011; **85**(1): 5–17.
- Manna, S. K., Patterson, A. D., Yang, Q., Krausz, K. W., Idle, J. R., Fornace, A. J., & Gonzalez, F. J. UPLC-MS-based urine metabolomics reveals indole-3-lactic acid and phenyllactic acid as conserved biomarkers for alcohol-induced liver disease in the Ppara-null mouse model. *Journal of Proteome Research*. 2011; **10**(9): 4120–33.
- Mashima, T., Seimiya, H., & Tsuruo, T. De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy. *British Journal of Cancer*. 2009; **100**(9): 1369–72.
- Mateos, M.-V., Oriol, A., Martínez-López, J., Gutiérrez, N., Teruel, A.-I., de Paz, R., García-Laraña, J., Bengoechea, E., Martín, A., Mediavilla, J. D., Palomera, L., de Arriba, F., González, Y., Hernández, J.-M., Sureda, A., Bello, J.-L., Bargay, J., Peñalver, F.-J., Ribera, J.-M., Martín-Mateos, M.-L., García-Sanz, R., Cibeira, M.-T., Ramos, M.-L. M., Vidriales, M.-B., Paiva, B., Montalbán, M.-A., Lahuerta, J.-J., Bladé, J., & Miguel, J.-F. S. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma. *The Lancet Oncology*. 2010; **11**(10): 934–41.
- Meiboom, S., & Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments*. 1958; **29**(8): 688.
- Muñoz-Pinedo, C., El Mjiyad, N., & Ricci, J.-E. Cancer metabolism: current perspectives and future directions. *Cell Death & Disease*. 2012; **3**: e248.
- Muntoni, S., Atzori, L., Mereu, R., Satta, G., Macis, M. D., Congia, M., Tedde, A., & Desogus, A. Serum lipoproteins and cancer. *Nutrition, Metabolism, and Cardiovascular Diseases: NMCD*. 2009; **19**(3): 218–25.
- Nanni, C., Zamagni, E., Cavo, M., Rubello, D., Tacchetti, P., Pettinato, C., Farsad, M., Castellucci, P., Ambrosini, V., Montini, G. C., Al-Nahhas, A., Franchi, R., & Fanti, S. 11C-choline vs. 18F-FDG PET/CT in assessing bone involvement in patients with multiple myeloma. *World Journal of Surgical Oncology*. 2007; **5**: 68.

- Nicholson, J K, Foxall, P. J., Spraul, M., Farrant, R. D., & Lindon, J. C. 750 MHz <sup>1</sup>H and <sup>1</sup>H-<sup>13</sup>C NMR spectroscopy of human blood plasma. *Analytical Chemistry*. 1995; **67**(5): 793–811.
- Nicholson, Jeremy K, Wilson, I. D., & Lindon, J. C. Pharmacometabonomics as an effector for personalized medicine. *Pharmacogenomics*. 2011; **12**(1): 103–11.
- Odunsi, K., Wollman, R. M., Ambrosone, C. B., Hutson, A., McCann, S. E., Tammela, J., Geisler, J. P., Miller, G., Sellers, T., Cliby, W., Qian, F., Keitz, B., Intengan, M., Lele, S., & Alderfer, J. L. Detection of epithelial ovarian cancer using <sup>1</sup>H-NMR-based metabonomics. *International Journal of Cancer. Journal International Du Cancer*. 2005; **113**(5): 782–8.
- Patel, N. R., McPhail, M. J. W., Shariff, M. I. F., Keun, H. C., & Taylor-Robinson, S. D. Biofluid metabonomics using (<sup>1</sup>H) NMR spectroscopy: the road to biomarker discovery in gastroenterology and hepatology. *Expert Review of Gastroenterology & Hepatology*. 2012; **6**(2): 239–51.
- Pisters, P. W., & Pearlstone, D. B. Protein and amino acid metabolism in cancer cachexia: investigative techniques and therapeutic interventions. *Critical Reviews in Clinical Laboratory Sciences*. 1993; **30**(3): 223–72.
- Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T., & Wishart, D. S. The human serum metabolome. *PloS One*. 2011; **6**(2): e16957.
- Rajendran, J. G., Mankoff, D. A., O’Sullivan, F., Peterson, L. M., Schwartz, D. L., Conrad, E. U., Spence, A. M., Muzi, M., Farwell, D. G., & Krohn, K. A. Hypoxia and glucose metabolism in malignant tumors: evaluation by [<sup>18</sup>F]fluoromisonidazole and [<sup>18</sup>F]fluorodeoxyglucose positron emission tomography imaging. *Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*. 2004; **10**(7): 2245–52.
- Rajpal, R., Dowling, P., Meiller, J., Clarke, C., Murphy, W. G., O’Connor, R., Kell, M., Mitsiades, C., Richardson, P., Anderson, K. C., Clynes, M., & O’Gorman, P. A novel panel of protein biomarkers for predicting response to thalidomide-based therapy in newly diagnosed multiple myeloma patients. *Proteomics*. 2011; **11**(8): 1391–402.
- Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification y 2007; (February): 341–351.
- Research Advocacy Network. Biomarkers in Cancer An Introductory Guide for Advocates 2010; .
- Righi, V., Mucci, A., Schenetti, L., Tosi, M. R., Grigioni, W. F., Corti, B., Bertaccini, A., Franceschelli, A., Sanguedolce, F., Schiavina, R., Martorana, G., & Tugnoli, V. Ex vivo HR-MAS magnetic resonance spectroscopy of normal and malignant human renal tissues. *Anticancer Research*. 2007; **27**(5A): 3195–204.
- Ritchie, S. A., Heath, D., Yamazaki, Y., Grimmalt, B., Kavianpour, A., Krenitsky, K., Elshoni, H., Takemasa, I., Miyake, M., Sekimoto, M., Monden, M., Tomonaga, T., Matsubara, H., Sogawa, K., Matsushita, K., Nomura, F., & Goodenowe, D. B. Reduction of novel circulating long-chain fatty acids in colorectal cancer patients is independent of tumor burden and correlates with age. *BMC Gastroenterology*. 2010; **10**: 140.



- Roberts, R. S., Hsu, H. W., Lin, K. D., & Yang, T. J. Amino acid metabolism of myeloma cells in culture. *Journal of Cell Science*. 1976; **21**(3): 609–15.
- Rocha, C. M., Carrola, J., Barros, A. S., Gil, A. M., Goodfellow, B. J., Carreira, I. M., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L., & Duarte, I. F. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. *Journal of Proteome Research*. 2011; **10**(9): 4314–24.
- San-Miguel, J. F., & Mateos, M.-V. Can multiple myeloma become a curable disease? *Haematologica*. 2011; **96**(9): 1246–8.
- Sato, J. D., Cao, H. T., Kayada, Y., Cabot, M. C., Sato, G. H., Okamoto, T., & Welsh, C. J. Effects of proximate cholesterol precursors and steroid hormones on mouse myeloma growth in serum-free medium. *In Vitro Cellular & Developmental Biology: Journal of the Tissue Culture Association*. 1988; **24**(12): 1223–8.
- Shin, D. H., Chun, Y.-S., Lee, D. S., Huang, L. E., & Park, J.-W. Bortezomib inhibits tumor adaptation to hypoxia by stimulating the FIH-mediated repression of hypoxia-inducible factor-1. *Blood*. 2008; **111**(6): 3131–6.
- Spratlin, J. L., Serkova, N. J., & Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*. 2009; **15**(2): 431–40.
- Tea, I., Le Gall, G., Küster, A., Guignard, N., Alexandre-Gouabau, M.-C., Darmaun, D., & Robins, R. J. 1H-NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants. *PLoS One*. 2012; **7**(1): e29947.
- Teahan, O., Bevan, C. L., Waxman, J., & Keun, H. C. Metabolic signatures of malignant progression in prostate epithelial cells. *The International Journal of Biochemistry & Cell Biology*. 2011; **43**(7): 1002–9.
- Terpos, E., Politou, M., & Rahemtulla, A. The role of markers of bone remodeling in multiple myeloma. *Blood Reviews*. 2005; **19**(3): 125–42.
- The Human Metabolome Project website. [www.metabolomics.ca](http://www.metabolomics.ca)
- Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Verweij, J., Van Glabbeke, M., van Oosterom, A. T., Christian, M. C., & Gwyther, S. G. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *JNCI Journal of the National Cancer Institute*. 2000; **92**(3): 205–216.
- Tiziani, S., Lopes, V., & Günther, U. L. Early stage diagnosis of oral cancer using 1H NMR-based metabolomics. *Neoplasia (New York, N.Y.)*. 2009; **11**(3): 269–76, 4p following 269.
- Tomita, M., & Kami, K. Cancer. Systems biology, metabolomics, and cancer metabolism. *Science (New York, N.Y.)*. 2012; **336**(6084): 990–1.
- Travis, W. D. Pathology of lung cancer. *Clinics in Chest Medicine*. 2002; **23**(1): 65–81, viii.
- Tredwell, G. D., Behrends, V., Geier, F. M., Liebeke, M., & Bundy, J. G. Between-person comparison of metabolite fitting for NMR-based quantitative metabolomics. *Analytical Chemistry*. 2011; **83**(22): 8683–7.
- Triba, M. N., Starzec, a, Bouchemal, N., Guenin, E., Perret, G. Y., & Le Moyec, L. Metabolomic profiling with NMR discriminates between biphosphonate and doxorubicin effects on B16 melanoma cells. *NMR in Biomedicine*. 2010; **23**(9): 1009–16.

- Trygg, J., & Wold, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*. 2002; **16**(3): 119–128.
- Urayama, S., Zou, W., Brooks, K., & Tolstikov, V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Communications in Mass Spectrometry: RCM*. 2010; **24**(5): 613–20.
- 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*. 2006; **7**: 142.
- Viant, M. R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and Biophysical Research Communications*. 2003; **310**(3): 943–8.
- Vizán, P., Mazurek, S., & Cascante, M. Robust metabolic adaptation underlying tumor progression. *Metabolomics*. 2008; **4**(1): 1–12.
- Wadhera, R. K., & Rajkumar, S. V. Prevalence of monoclonal gammopathy of undetermined significance: a systematic review. *Mayo Clinic Proceedings. Mayo Clinic*. 2010; **85**(10): 933–42.
- Warburg, O. On the origin of cancer cells. *Science (New York, N.Y.)*. 1956; **123**(3191): 309–14.
- Wedge, D. C., Allwood, J. W., Dunn, W., Vaughan, A. A., Simpson, K., Brown, M., Priest, L., Blackhall, F. H., Whetton, A. D., Dive, C., & Goodacre, R. Is serum or plasma more appropriate for intersubject comparisons in metabolomic studies? An assessment in patients with small-cell lung cancer. *Analytical Chemistry*. 2011; **83**(17): 6689–97.
- Wein, J., & Goetz, I. E. (1973). Asparaginase and glutaminase activities in culture media containing dialyzed fetal calf serum. *In Vitro*. **9**(3): 186–93.
- Weiss, S. A., Godwin, G. P., Gorfien, S. F., & Whitford, W. G. NS0 Serum-free culture and applications. *BioProcess International*. 2003; 36–47.
- Wen, T., Gao, L., Wen, Z., Wu, C., Tan, C. S., Toh, W. Z., & Ong, C. N. Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. *Molecular bioSystems*. 2013; **9**(9): 2370–8.
- Westerhuis, J. a., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., Duijnhoven, J. P. M., & Dorsten, F. a. Assessment of PLS-DA cross validation. *Metabolomics*. 2008; **4**(1): 81–89.
- Wikoff, W. R., Gangoiti, J. A., Barshop, B. A., & Siuzdak, G. Metabolomics identifies perturbations in human disorders of propionate metabolism. *Clinical Chemistry*. 2007; **53**(12): 2169–76.
- Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorn Dahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., & Scalbert, A. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*. 2013; **41**(Database issue): D801–7.
- Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. a, Lim, E., Sobsey, C. a, Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga,

- A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., & Forsythe, I. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research*. 2009; **37**(Database issue): D603–10.
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., Macinnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J., & Querengesser, L. HMDB: the Human Metabolome Database. *Nucleic Acids Research*. 2007; **35**(Database issue): D521–6.
- Yang, Y., Li, C., Nie, X., Feng, X., Chen, W., Yue, Y., Tang, H., & Deng, F. Metabonomic studies of human hepatocellular carcinoma using high-resolution magic-angle spinning <sup>1</sup>H NMR spectroscopy in conjunction with multivariate data analysis. *Journal of Proteome Research*. 2007; **6**(7): 2605–14.
- Yavasoglu, I., Tombuloglu, M., Kadikoylu, G., Donmez, A., Cagirgan, S., & Bolaman, Z. Cholesterol levels in patients with multiple myeloma. *Annals of Hematology*. 2008; **87**(3): 223–8.
- Yu, Z., Kastenmüller, G., He, Y., Belcredi, P., Möller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., Polonikov, A., Dahmen, N., Prokisch, H., Xie, L., Li, Y., Wichmann, H.-E., Peters, A., Kronenberg, F., Suhre, K., Adamski, J., Illig, T., & Wang-Sattler, R. Differences between human plasma and serum metabolite profiles. *PloS One*. 2011; **6**(7): e21230.
- Zhou, W., Capello, M., Fredolini, C., Racanicchi, L., Piemonti, L., Liotta, L. a, Novelli, F., & Petricoin, E. F. Proteomic analysis reveals Warburg effect and anomalous metabolism of glutamine in pancreatic cancer cells. *Journal of Proteome Research*. 2012; **11**(2): 554–63.
- Zira, A. N., Theocharis, S. E., Mitropoulos, D., Migdalis, V., & Mikros, E. (<sup>1</sup>H NMR metabonomic analysis in renal cell carcinoma: a possible diagnostic tool. *Journal of Proteome Research*. 2010; **9**(8): 4038–44.



## **VI. ANEXOS**



## 1. Comunicaciones a congresos nacionales e internacionales

Autores: **Puchades-Carrasco L**

Título: **Metabolomics: Basic concepts and some applications**

Tipo de participación: Comunicación oral

Congreso: *IV Curso Superior de Biotecnología en la Empresa*

Lugar celebración: BIOVAL, Valencia, Spain

Fecha: 2011

Autores: **Puchades-Carrasco, L**

Título: **Metabolomics by NMR**

Tipo de participación: Comunicación oral

Congreso: *Seminarios de Investigación*

Lugar celebración: BiOncoTech Therapeutics, Valencia, España

Fecha: 2011

Autores: **Puchades-Carrasco L**

Título: **Metabolomics: Basic concepts and some applications**

Tipo de participación: Comunicación oral

Congreso: *Seminarios de Investigación*

Lugar celebración: Centre for Cellular and Molecular Biology, Hyderabad, India

Fecha: 2012

Autores: **Puchades-Carrasco L**, Lecumberri R, Martínez-López J, Lahuerta JJ, Mateos MV, Prósper F, San Miguel JF, Pineda-Lucena A.

Título: **Metabolomic profile of multiple myeloma patients.**

Tipo de participación: Póster

Congreso: *IV Ibero-American NMR Meeting - VI GERMN Bienal Meeting*

Lugar celebración: University of Aveiro, Portugal.

Fecha: 2012

Autores: **Puchades-Carrasco L**, Lecumberri R, Martínez-López J, Lahuerta JJ, Mateos MV, Prósper F, San Miguel JF, Pineda-Lucena A.

Título: **El perfil metabolómico de los pacientes con Múltiple Mieloma varía al alcanzar la remisión completa de la enfermedad.**

Tipo de participación: Presentación oral

Congreso: *LIV REUNIÓN NACIONAL DE LA SEHH Y XXVIII CONGRESO NACIONAL DE LA SETH*

Lugar celebración: Palacio de Congresos y Exposiciones de Castilla y León, Salamanca.

Fecha: 2012

Autores: **Puchades-Carrasco L**, Lecumberri R, Martínez-López J, Lahuerta JJ, Mateos MV, Prósper F, San Miguel JF, Pineda-Lucena A.

Título: **Metabolomic profile of multiple myeloma patients.**

Tipo de participación: Póster

Congreso: *Symposium CIPF'12*

Lugar celebración: Centro de Investigación Príncipe Felipe, Valencia, Spain.

Fecha: 2012

Autores: **Puchades-Carrasco L**, Jantus-Lewintre E, Lucas R, Blasco A, Pineda-Lucena A, Camps C.

Título: **Metabolomic profile of lung cancer patients.**

Tipo de participación: Póster discutido

Congreso: *European Multidisciplinary Conference in Thoracic Oncology (EMCTO 2013)*

Lugar celebración: Lugano, Suiza.

Fecha: 2013

Autores: **Puchades-Carrasco L.**

Título: **Metabolomics by NMR: Applications in disease diagnosis and treatment monitoring.**

Tipo de participación: Comunicación oral

Congreso: TRANSBIO Emergence Forum.

Lugar celebración: Montpellier, Francia.

Fecha: 2013

Autores: **Puchades-Carrasco L.**, Jantus-Lewintre E, Lucas R, Blasco A, Camps C, Pineda-Lucena A.

Título: **Caracterización por RMN del perfil metabólico de pacientes con cáncer de pulmón.**

Tipo de participación: Póster expuesto

Congreso: XIV Congreso Nacional SEOM.

Lugar celebración: Salamanca, España.

Fecha: 2013

Autores: **Puchades-Carrasco L.**, Lecumberri R, Martínez-López J, Lahuerta JJ, Mateos MV, Prósper F, San Miguel JF, Pineda-Lucena A.

Título: **Metabolomic profile of Multiple Myeloma patients.**

Tipo de participación: Comunicación oral

Congreso: ***Markers in Cancer Meeting***

Lugar celebración: Bruselas, Bélgica.

Fecha: 2013



## **2. Publicaciones derivadas de la Tesis Doctoral**



## Multiple Myeloma Patients Have a Specific Serum Metabolomic Profile That Changes after Achieving Complete Remission

Leonor Puchades-Carrasco<sup>1</sup>, Ramón Lecumberri<sup>2</sup>, Joaquín Martínez-López<sup>4</sup>, Juan-José Lahuerta<sup>4</sup>, María-Victoria Mateos<sup>5</sup>, Felipe Prósper<sup>3</sup>, Jesús F. San-Miguel<sup>5</sup>, and Antonio Pineda-Lucena<sup>1</sup>

### Abstract

**Purpose:** Multiple myeloma remains an incurable disease. New approaches to develop better tools for improving patient prognostication and monitoring treatment efficacy are very much needed. In this study, we aimed to evaluate the potential of metabolomics by <sup>1</sup>H-NMR to provide information on metabolic profiles that could be useful in the management of multiple myeloma.

**Experimental Design:** Serum samples were collected from multiple myeloma patients at the time of diagnosis and after achieving complete remission. A matched control set of samples was also included in the study. The <sup>1</sup>H-NMR measurements used to obtain the metabolic profile for each patient were followed by the application of univariate and multivariate statistical analyses to determine significant differences.

**Results:** Metabolic profiles of multiple myeloma patients at diagnosis exhibited higher levels of isoleucine, arginine, acetate, phenylalanine, and tyrosine, and decreased levels of 3-hydroxybutyrate, lysine, glutamine, and some lipids compared with the control set. A similar analysis conducted in multiple myeloma patients after achieving complete remission indicated that some of the metabolic changes (i.e., glutamine, cholesterol, lysine) observed at diagnosis displayed a variation in the opposite direction upon responding to treatment, thus contributing to multiple myeloma patients having a closer metabolic profile to those of healthy individuals after the disappearance of major disease manifestations.

**Conclusions:** The results highlight the potential of metabolic profiles obtained by <sup>1</sup>H-NMR in identifying multiple myeloma biomarkers that may be useful to objectively discriminate individuals with and without multiple myeloma, and monitor response to treatment. *Clin Cancer Res*; 19(17); 4770–9. ©2013 AACR.

### Introduction

Multiple myeloma is a malignant disorder of postgerminal center B cells, characterized by the clonal proliferation of malignant plasma cells within the bone marrow. It is the second most frequent hematologic malignancy, after non-Hodgkin lymphoma with an incidence of 4 to 5 new patients per 100,000 people and year. In the past few years, many significant advances have occurred in the field of

multiple myeloma, about all aspects of research, from basic knowledge to supportive care, novel diagnostic procedures, and prognostic classifications, as well as novel therapeutic approaches. The administration of high-dose therapy followed by stem cell support (HDT-ASCT) and the introduction of novel agents, such as immunomodulatory drugs (thalidomide and lenalidomide) and proteasome inhibitors (bortezomib) have contributed to improve the overall survival of multiple myeloma patients. Nevertheless, multiple myeloma remains an incurable disease and new efforts are necessary not only to improve survival but also to develop better tools to evaluate the prognosis and to monitor treatment efficacy.

Metabolomics, a new and powerful tool based on the combination of metabolic profiling techniques with multivariate statistical approaches, has been proved to be extremely useful in the identification of early biomarkers related with inborn errors of metabolism, cardiovascular disease, metabolic disorders, or cancer (1). This approach involves the systematic study of small molecules that characterize the metabolic pathways of biological systems. The most commonly employed analytical

**Authors' Affiliations:** <sup>1</sup>Structural Biochemistry Laboratory, Centro de Investigación Príncipe Felipe, Valencia; <sup>2</sup>Hematology Service, <sup>3</sup>Hematology and Cell Therapy Area, Clínica Universidad de Navarra, Pamplona; <sup>4</sup>Hematology Service, Hospital 12 de Octubre, Madrid; and <sup>5</sup>Hematology Service, Hospital Universitario de Salamanca, IBSAL, IBMCC (USAL-CSIC), Salamanca, Spain

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

**Corresponding Author:** Antonio Pineda-Lucena, Structural Biochemistry Laboratory, Centro de Investigación Príncipe Felipe, Eduardo Primo Yúfera, 3, 46012, Valencia, Spain. Phone: +34 963289681; Fax: +34 963289701; E-mail: apineda@cifpf.es

doi: 10.1158/1078-0432.CCR-12-2917

©2013 American Association for Cancer Research.

### Translational Relevance

Multiple myeloma is a hematological neoplasm difficult to diagnose in its early stages. Despite the introduction of novel agents that have contributed to improve the overall survival of multiple myeloma patients, the disease is still incurable. This scenario explains why new efforts are necessary not only to improve survival but also to develop better tools to evaluate the prognosis and to monitor treatment efficacy. In this context, we assessed the potential of metabolomics to provide information that could be useful in the clinical management of multiple myeloma patients. Although validation in a prospective fashion using a larger sample set is required, our results provide evidence that metabolomics could be a promising tool for a better understanding of the molecular mechanisms of multiple myeloma with implications as a diagnostic tool, as well as for elucidating the biochemical pathways involved in the response to chemotherapy.

techniques used for metabolic profiling include nuclear magnetic resonance ( $^1\text{H-NMR}$ ) spectroscopy and mass spectrometry (MS; ref. 2). High-resolution  $^1\text{H-NMR}$  spectroscopy provides quantitative analysis of metabolite concentrations and reproducible information with minimal sample handling. Particularly in the cancer setting, metabolomics has been applied to develop novel early diagnostic biomarkers in renal cancer (3, 4), colorectal cancer (5), pancreatic cancer (6), leukemia (7), ovarian cancer (8, 9) and oral cancer (10). More recently, this approach has also been used to provide predictive biomarkers associated with prognosis, response to treatment and toxicity (11, 12).

In this study, we assessed the potential role of metabolomics by  $^1\text{H-NMR}$  spectroscopy to characterize the specific metabolic profile of multiple myeloma patients by conducting a comparative analysis of healthy subjects and patients with multiple myeloma. Furthermore, to extend the clinical applications of metabolomics, a com-

parison between the metabolic profiles of multiple myeloma patients at the time of diagnosis and after achievement of complete remission was conducted. Overall, we found that this approach provides a promising tool for a better understanding of the molecular mechanisms of multiple myeloma with implications as a diagnostic tool, as well as for elucidating the biochemical pathways involved in the response to chemotherapy, an information that could have a relevant impact in treatment design and monitoring.

### Materials and Methods

#### Selection and description of participants

Serum samples were obtained from multiple myeloma patients (Table 1) included in 2 Spanish Myeloma Group clinical trials: the phase 3 GEM2005<65y trial, in which young patients were randomized to receive induction with VBMCP (vincristine, carmustine, melphalan, cyclophosphamide, and prednisone)/VBAD (vincristine, carmustine, doxorubicin, and dexamethasone) plus bortezomib in the last 2 cycles or thalidomide/dexamethasone or bortezomib/thalidomide/dexamethasone followed by HDT-ASCT, and the phase 3 GEM 2005>65y for elderly patients, consisting on bortezomib plus melphalan/prednisone or thalidomide/prednisone (13). Both trials included patients with newly diagnosed, untreated, symptomatic, measurable multiple myeloma (defined as serum monoclonal protein of more than 10 g/L or urine monoclonal protein of 0.2 g or more per day). For most patients included in this study, 2 serum samples were analyzed; the first one was collected at the time of diagnosis and the second after achieving complete remission, defined as absence of the original M component by immunofixation and <5% plasma cells in bone marrow. A matched control set of 31 serum samples (age, sex) from healthy subjects was also included. Samples were collected after informed consent was obtained in accordance with the Declaration of Helsinki and with approval from the Ethics Committees of all participating institutions. Both clinical trials were registered with ClinicalTrials.gov, number NCT 00461747 for GEM2005<65y and NCT00443235 for GEM2005>65y.

**Table 1.** Samples included in the metabolomic study

Group	n	Gender	Age $\pm$ SEM <sup>a</sup>	M protein
Control	31	M = 16 F = 15	67.1 $\pm$ 1.6	–
Diagnosis	27	M = 14 F = 13	63.6 $\pm$ 2.2	IgG = 13 IgA = 9 Bence–Jones = 5
Remission	23	M = 12 F = 11	63.7 $\pm$ 2.3	–

n = sample size.

<sup>a</sup>Mean years at time of sample collection (for the control group) and diagnosis (for diagnosis and remission groups)  $\pm$  SEM.

### Sample preparation and $^1\text{H-NMR}$ acquisition

Serum was collected and immediately stored at  $-80^\circ\text{C}$ . At the time of analysis, the serum samples were thawed on ice. Five hundred microliters of 10%  $\text{D}_2\text{O}$  buffer (5 mmol/L TSP, 140 mmol/L  $\text{Na}_2\text{HPO}_4$ , 0.04%  $\text{NaN}_3$ , pH 7.4) were added to 500  $\mu\text{L}$  of serum sample. After this, 550  $\mu\text{L}$  of the mixture were transferred to a 5-mm NMR tube for analysis.  $^1\text{H-NMR}$  spectra were acquired using a Bruker Avance II 600 MHz spectrometer equipped with triple resonance cryo-probe with a cooled  $^{13}\text{C}$  preamplifier (TCI).  $^1\text{H-NMR}$  experiments were acquired at 310 K ( $37^\circ\text{C}$ ) for every sample. Carr–Purcell–Meiboom–Gill (CPMG) spin-echo pulse sequence (14), which generates spectra edited by T2 relaxation times with reduced signals from high molecular weight species and giving improved resolution of low molecular weight metabolite resonances, was collected for each sample with a total of 16 accumulations and 72 K data points over a spectral width of 16 ppm. A 4-second relaxation delay was included between free induction decays (FIDs). The total spin-spin relaxation delay was 40 milliseconds. A one-dimensional (1D) NOESY pulse sequence that generates an unedited spectrum with improved solvent peak suppression (15) was collected using the same parameters as the CPMG experiment, with a 4-second relaxation delay and 10 milliseconds of mixing time. For both experiments, a water presaturation pulse of 25 Hz was applied throughout the relaxation delays to improve solvent suppression. In addition, for assignment purposes, 2D  $J$ -resolved spectra, homonuclear 2D  $^1\text{H-}^1\text{H}$  total correlation spectroscopy and 2D  $^1\text{H}$ ,  $^{13}\text{C}$  heteronuclear single quantum correlation were acquired for selected samples. All spectra were multiplied by a line-broadening factor of 1 Hz and Fourier transformed. Spectra were automatically phased and baseline corrected, chemical shift referenced internally to the methyl group signal of alanine at 1.47 ppm using TOPSPIN 3.0 (Bruker Biospin) and analyzed with Analysis of MIXtures (AMIX; Bruker) for metabolites assignment. Metabolites of interest were identified using Amix v 3.9.7 in combination with the Bruker NMR Metabolic Profiling Database BBIORFCODE 2.0.0 database (Bruker Biospin), as well as other existing public databases and literature reports (7, 15, 16).

### Multivariate statistical analysis

All samples, but one having high-intensity signals derived from a contamination during the sample collection/preparation process, were considered valid samples for the statistical analysis. CPMG spectra were used for the multivariate statistical analysis because of improved baseline and metabolite detection. Furthermore, any reduction in signal intensity induced by relaxation effects would be consistent across samples.  $^1\text{H-NMR}$  spectra were binned using Amix 3.9.7 (Bruker Biospin) into 0.04-ppm-wide rectangular buckets over the region  $\delta$  10.5 to 0.26 ppm. The water ( $\delta$  5.12–4.35 ppm) and urea signal ( $\delta$  5.85–5.65 ppm) regions were excluded from the analysis to avoid interference arising from dif-

ferences in water suppression and variability from urea signal, respectively. Bucket tables generated were imported into SIMCA-P 12.0 software (Umetrics AB). Because statistical analysis data were scaled to unit variance by dividing each variable by  $1/(\text{SD})$ , where SD represents the standard deviation value of each variable, so that all the 234 variables were given equal weight regardless of their absolute value.

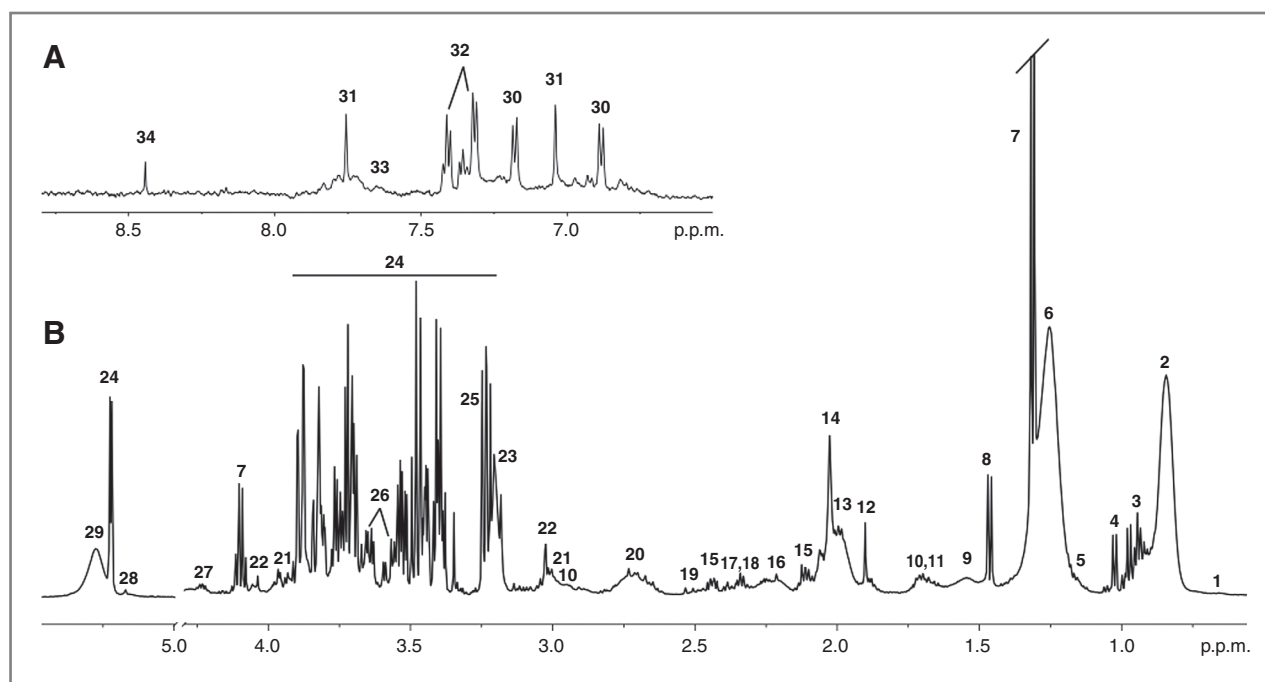
Then, principal components analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) were conducted. PCA, an unsupervised pattern recognition method that does not use prior knowledge of a system, was used to examine the intrinsic variability within the dataset, to observe clustering or separation trends and for the identification of outliers. Next, OPLS, a supervised approach used to produce models of "best fit" was applied to minimize the possible contribution of intergroup variability and to further improve separation between the groups of samples. The default method of seven-fold internal cross validation of the software was applied, from which  $Q^2Y$  (predictive ability parameter, estimated by cross-validation) and  $R^2Y$  (goodness of fit parameter) values were extracted. Those parameters were used for the evaluation of the quality of the OPLS-DA models obtained. To identify the variations responsible for the separation between groups of samples, the corresponding loading plots and the variable importance in projection (VIP) list of each OPLS-DA model were carefully inspected, and used to identify which variables were important for discrimination between groups.

### Quantitative analysis of variations of selected metabolites

The main metabolites contributing to group discrimination in each model (healthy individuals vs. multiple myeloma patients at the time of diagnosis, multiple myeloma patients at the time of diagnosis vs. multiple myeloma patients after complete remission, and healthy individuals vs. multiple myeloma patients after complete remission) were integrated to enable comparison between samples. To this end, the variable size bucketing module in Amix 3.9.7 was used to obtain the exact integral value corresponding to the NMR signal of each selected metabolite. The statistical significance of the differences between the means of the 2 groups was assessed using the Student  $t$  test. A  $P$ -value  $<0.05$  (confidence level 95%) was considered statistically significant.

### Results

Thirty-one serum samples from healthy donors, 27 of multiple myeloma patients at the time of diagnosis and 23 paired samples from these same multiple myeloma patients after achieving complete remission were included in the analysis. The characteristics of the subjects included in this study are summarized in Table 1. A typical  $^1\text{H-NMR}$  CPMG spectrum of a sample corresponding to a multiple myeloma patient collected at the time of diagnosis is illustrated in Fig. 1. Spectra of similar quality were obtained



**Figure 1.** Assignment of the  $^1\text{H-NMR}$  signals of a representative 600 MHz one-dimensional Carr–Purcell–Meiboom–Gill (1D-CPMG)  $^1\text{H-NMR}$  spectra of a multiple myeloma patient serum sample measured at 310K. A, full spectrum ( $\delta = 5.50\text{--}0.5$  ppm) and B, magnification of aromatic region ( $\delta = 9.00\text{--}6.50$  ppm). Peak assignments: 1, cholesterol (C18-CH<sub>3</sub>); 2, lipids (–CH<sub>3</sub>) (mainly LDL/VLDL); 3, leucine/isoleucine; 4, valine; 5, 3-hydroxybutyrate; 6, lipids (CH<sub>2</sub>)<sub>n</sub> (mainly LDL/VLDL); 7, lactate; 8, alanine; 9, adipic acid; 10, lysine; 11, arginine; 12, acetate; 13, lipids (CH<sub>2</sub>–C = C); 14, NAC (acetyl signals from glycoproteins); 15, glutamine; 16, lipids (CH<sub>2</sub>–CO); 17, glutamate; 18, pyruvate; 19, citrate; 20, lipids (CH = CH–CH<sub>2</sub>–CH = CH–); 21, creatine; 22, creatinine; 23, choline; 24, glucose; 25, trimethylamine N-oxide; 26, glycerol; 27, threonine; 28, glycerol of lipids; 29, lipids (–CH = CH–); 30, tyrosine; 31, histidine; 32, phenylalanine; 33, 3-methylhistidine; 34, formate.

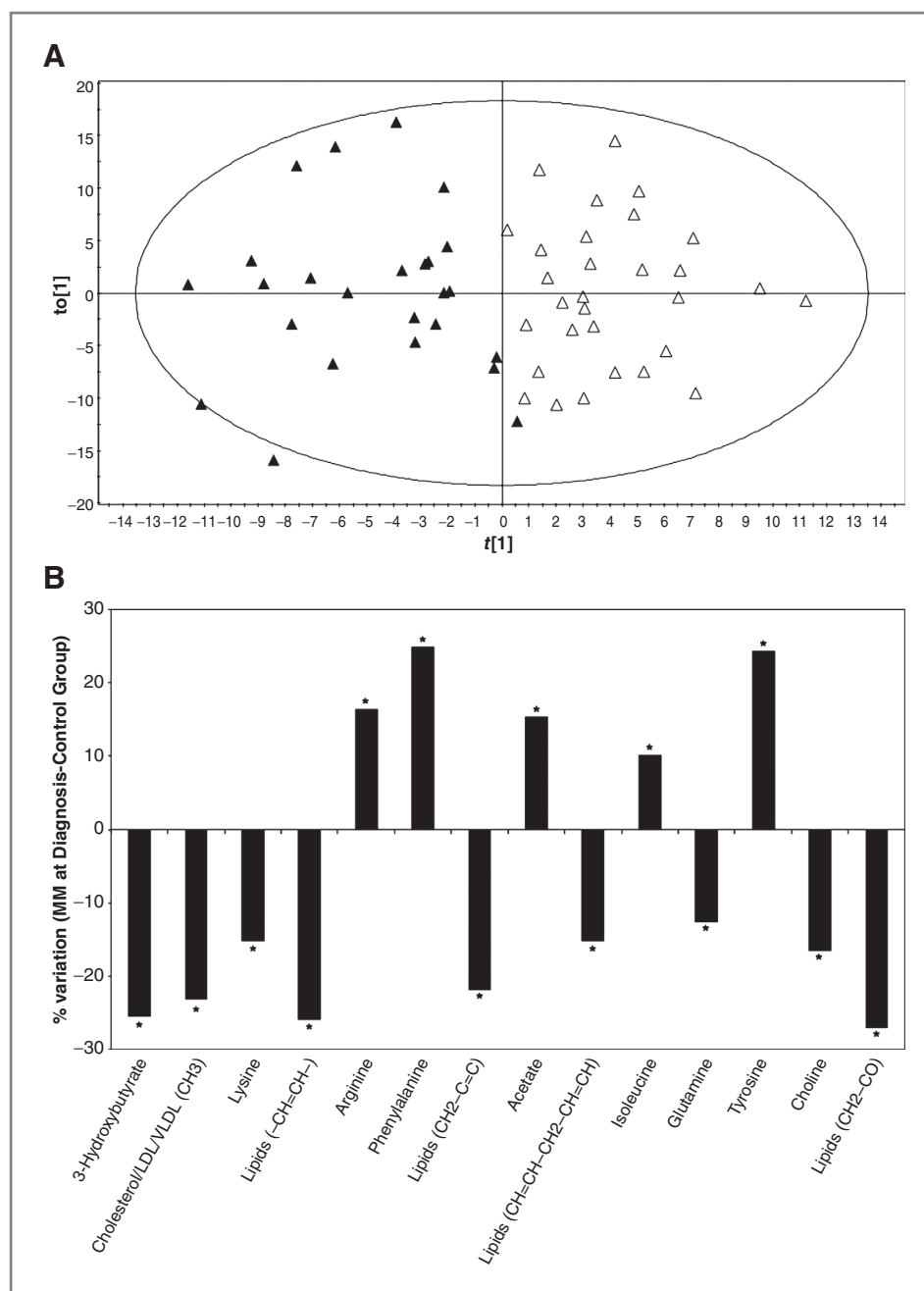
for multiple myeloma patients after achieving complete remission and control individuals. Spectra were dominated by signals of low-molecular weight metabolites. Assigned signals represent a typical range of metabolites normally found in serum, including several amino acids (e.g., lysine, arginine, glutamine, phenylalanine, histidine, tyrosine, alanine), glucose and sugars, creatine, creatinine, organic acids (e.g., acetate, lactate, citrate, formate). Higher molecular weight metabolites were also observable in the CPMG spectrum: fatty acyl chain protons from lipoproteins, unsaturated acids, and glycoproteins.

PCA analyses were applied independently to the full set of control samples, patients samples collected at the time of diagnosis and after complete remission to detect potential intrinsic clustering and possible outliers. No age or gender-related differences were observed to influence the homogeneity of the groups of samples included in the study (Supplementary Fig. S1). Only one sample from the control set and one from the group of multiple myeloma patients at the time of diagnosis were identified as strong outliers on their group, outside the Hotellings 95% confidence limit. They displayed a metabolic profile that has been previously reported to be related to hypertensive patients (17, 18), characterized by changes in signals attributed to lipids and lipoproteins, as well as the choline-containing region. Furthermore, the detailed revision of the clinical history of the multiple myeloma patient at the time of diagnosis revealed that the patient was diabetic and hypertensive.

These findings justified the exclusion of these 2 samples from the study.

#### Multiple myeloma patients at diagnosis versus healthy donors

To better examine differences between the groups of samples an OPLS-DA model was used. The score plot (Fig. 2A) shows a clear separation between healthy individuals and multiple myeloma patients at the time of diagnosis ( $R^2 = 0.613$ ;  $Q^2 = 0.462$ ). Examination of the corresponding loading plot and VIP list indicated the buckets that were mainly contributing to the discrimination between healthy and multiple myeloma individuals. The metabolites corresponding to those regions of the spectra were identified through a combination of their  $^1\text{H}$  chemical shift in the 1D-CPMG experiment and the spin system patterns obtained from the 2D spectra acquired for representative samples of each group. Metabolic profiles of multiple myeloma patients exhibited higher levels of isoleucine, arginine, acetate, phenylalanine, and tyrosine and decreased levels of 3-hydroxybutyrate, lysine, glutamine, as well as some lipids. Supplementary Table S1 summarizes the mean and SEM values obtained from the integration of the defined regions in each group of samples and the percentage of variation for the selected metabolite between these 2 groups. The percentage of variation found for the above-mentioned metabolites between the 2 groups is



**Figure 2.** Multivariate modeling resulting from the analysis of serum  $^1\text{H-NMR}$  spectra of control individuals (open symbols) vs. multiple myeloma patients at the time of diagnosis (full symbols). A, OPLS-DA score plot showing the separation between the 2 groups based on their metabolic profile. B, mean signal intensity variation (%) of the metabolites that contributed most to the separation between groups. Positive bars in the graphic correspond to signals that are increased in multiple myeloma patients at the time of diagnosis. \*, Statistically significant differences compared to control ( $P < 0.05$ ).

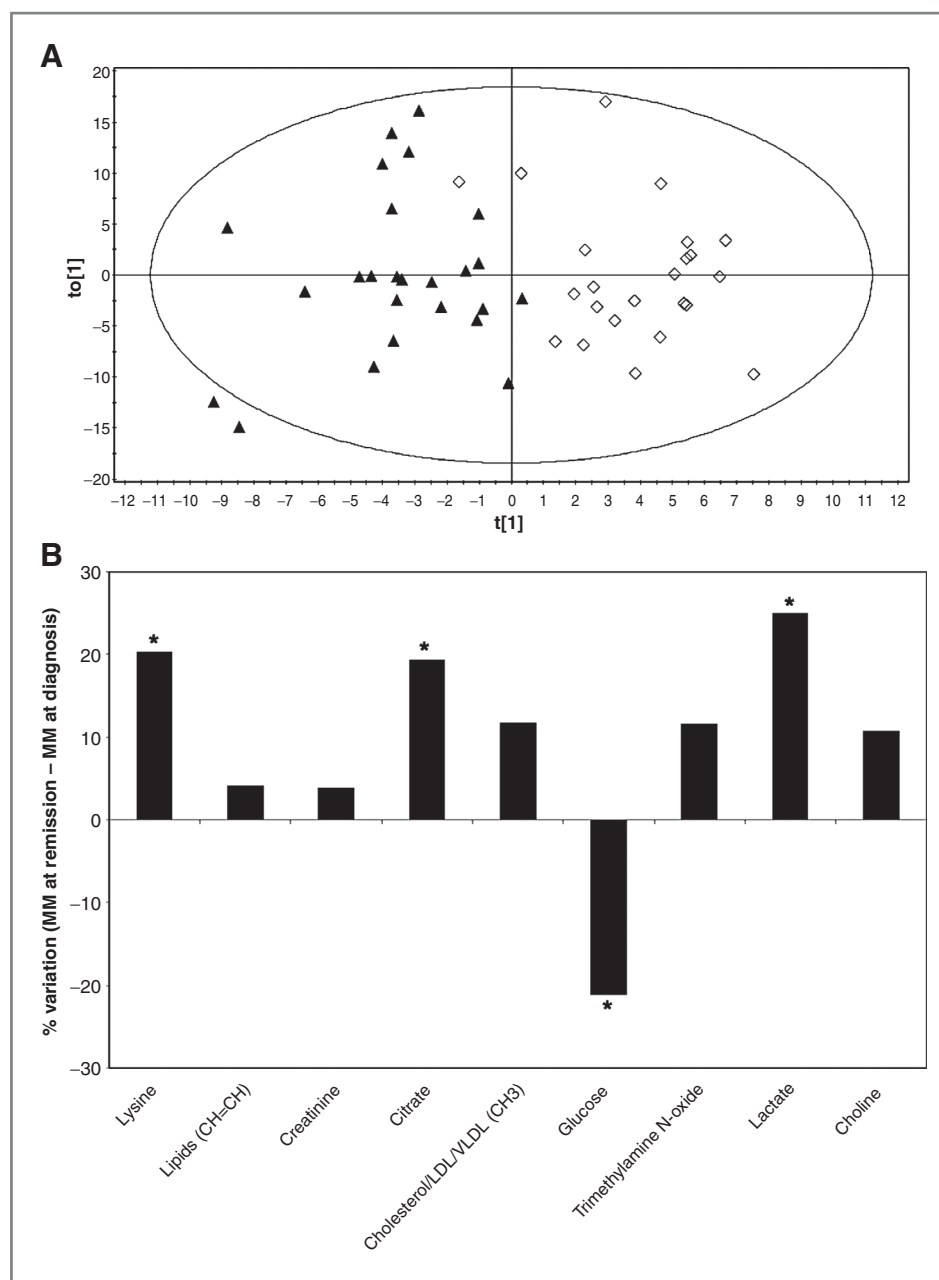
graphically represented in Fig. 2B. The results obtained are consistent with the OPLS-DA model built for the discrimination of both groups of samples.

#### Multiple myeloma patients at diagnosis versus multiple myeloma patients after complete remission

A different OPLS-DA model was built to determine if it was possible to distinguish when a multiple myeloma patient has achieved complete remission of the disease based on its  $^1\text{H-NMR}$  metabolic profile. Figure 3A and B shows the score plot for this model, which illustrates the clear separation obtained between the 2 groups ( $R^2 =$

0.553;  $Q^2 = 0.358$ ), and a graphic representation of the main metabolites contributing to this separation, based on the results obtained from the loading plot and the most representative buckets defined in the VIP list. Based on these results, the  $^1\text{H-NMR}$  serum metabolic profile of multiple myeloma patients after achieving complete remission differed from the profile at the time of diagnosis mainly in the levels of lactate, lysine, TMAO, choline, and cholesterol, which were significantly higher in patients after complete remission. To validate the statistical modeling, the metabolites identified were quantified using variable size bucketing for the integration of the corresponding  $^1\text{H-NMR}$

**Figure 3.** Multivariate modeling resulting from the analysis of serum  $^1\text{H-NMR}$  spectra of multiple myeloma (MM) patients at the time of diagnosis (full symbols) vs. MM patients after complete remission (open symbols). A, OPLS-DA score plot showing the separation between the 2 groups based on their metabolic profile. B, mean signal intensity variation (%) of the metabolites that contributed most to the separation between groups. Positive bars in the graphic correspond to signals that are increased in MM patients after the complete remission of the disease. \*, Statistically significant differences compared to diagnosis ( $P < 0.05$ ).



signals in the spectra. The summary of the metabolites that were quantified for this model is shown in Supplementary Table S2, including the percentage of variation found in the group of multiple myeloma patients after complete remission compared to that of multiple myeloma patients at the time of diagnosis. A graphical representation of the percentage of variation found in the level of these metabolites between the 2 groups is shown in Fig. 3B.

The reliability of the statistical OPLS-DA models was evaluated based on the parameters obtained from the internal cross validation, summarized in Table 2, in terms of  $Q^2$  and  $R^2$  values (representing, respectively, the predictive capability and the explained variance),

sensitivity (percentage of true positive results), specificity (percentage of true negative results), and accuracy (total percentage of samples correctly classified) values. This information was extracted from the misclassification table obtained when the class labels of all samples were predicted using the corresponding model.

#### Multiple myeloma patients after complete remission versus healthy donors

Finally, an OPLS-DA analysis conducted to compare the metabolic profiles of multiple myeloma patients after complete remission and healthy individuals showed that the statistical discrimination between these 2 groups was



**Table 2.** Prediction results obtained for OPLS-DA models built for control vs. diagnosis samples and diagnosis vs. remission samples

	<i>n</i>	<i>R</i> <sup>2</sup>	<i>Q</i> <sup>2</sup>	Sensitivity <sup>a</sup> (%)	Specificity <sup>a</sup> (%)	Accuracy <sup>a</sup> (%)	Fishers prob. <sup>b</sup>
Control vs. diagnosis	55	0.613	0.462	88	100	94.5	1.8E-012
Diagnosis vs. remission	48	0.553	0.358	95.7	96	95.8	1.9E-011

*n* = sample size.

<sup>a</sup>Classification results were used to define true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) and therefore calculate as percent sensitivity [TP/(TP + FN) × 100], specificity [TN/(TN + FP) × 100], and accuracy [TP + TN/(TP + TN + FP + FN) × 100].

<sup>b</sup>Probability of a particular model occurring by chance and is satisfied when *P* < 0.05 for 95% confidence.

smaller ( $Q^2 < 0.25$ ) to that obtained in the previous comparisons (i.e., healthy controls vs. multiple myeloma patients at the time of diagnosis, and multiple myeloma patients at the time of diagnosis vs. multiple myeloma patients after complete remission). Table 3 summarizes the mean and SEM values obtained from the integration of the metabolites that showed the most significant variations between the 3 groups of samples.

## Discussion

Efforts to identify multiple myeloma early biomarkers that could help to better understand disease pathogenesis

and to evaluate response to treatment remain a fundamental goal in this area (19–21). In that sense, to the best of our knowledge, this study represents the first application of high field <sup>1</sup>H-NMR metabolomics to the characterization of serum samples from multiple myeloma patients. The multivariate statistical approach used in this study takes into account numerous variables (metabolites) across multiple samples, facilitating a global description of the serum metabolic profile of multiple myeloma patients. Overall, the results suggest that the serum metabolic profile of multiple myeloma patients could provide useful information about the molecular

**Table 3.** Mean intensities of most significant variations found in the metabolites for the 3 groups of samples included in the study

Metabolite	$\delta^1\text{H}$ (ppm) <sup>a</sup>	Mean spectral intensity ± SEM (arbitrary units)		
		Control group	Multiple myeloma patients at diagnosis	Multiple myeloma patients at remission
3-Hydroxybutyrate	1.15–1.19	1.95E07 ± 6.59E05	1.45E07 ± 6.97E05	1.44E07 ± 6.45E05
Arginine	1.63–1.77	1.18E07 ± 3.38E05	1.37E07 ± 5.64E05	1.42E07 ± 5.10E05
Glutamine	2.41–2.46	6.76E06 ± 1.31E05	5.90E06 ± 3.57E05	6.13E06 ± 2.52E05
Phenylalanine	7.29–7.33	1.06E06 ± 4.65E04	1.33E06 ± 7.62E04	1.33E06 ± 8.38E04
Lipids (CH = CH–CH <sub>2</sub> –CH = CH)	2.60–2.84	3.22E07 ± 1.21E06	2.73E07 ± 1.52E06	2.90E07 ± 1.04E06
Acetate	1.88–1.91	4.03E06 ± 1.62E05	4.65E06 ± 2.78E05	5.28E06 ± 5.14E05
Lipids (CH = CH)	5.24–5.37	4.09E07 ± 2.74E06	3.03E07 ± 2.59E06	3.15E07 ± 2.03E06
Glucose	5.21–5.24	1.22E07 ± 8.00E05	1.41E07 ± 1.25E06	1.11E07 ± 9.67E05
Lipids (CH <sub>2</sub> –CO)	2.17–2.25	1.81E07 ± 2.11E06	1.32E07 ± 1.38E06	1.18E07 ± 9.77E05
Lactate	4.13–4.18	1.84E06 ± 1.87E05	2.01E06 ± 2.00E05	2.51E06 ± 2.04E05
Lipids (CH <sub>2</sub> –C = C)	1.93–2.01	4.69E07 ± 2.51E06	3.67E07 ± 2.20E06	3.92E07 ± 1.67E06
Citrate	2.50–2.55	1.46E06 ± 9.17E04	1.51E06 ± 6.65E04	1.80E06 ± 1.00E05
Cholesterol/LDL/VLDL (CH <sub>3</sub> )	0.70–0.90	1.62E08 ± 7.42E06	1.24E08 ± 7.53E06	1.39E08 ± 4.91E06
Lysine	2.87–2.93	3.70E06 ± 5.93E04	3.14E06 ± 1.11E05	3.77E06 ± 8.45E04
Tyrosine	7.16–7.20	5.64E05 ± 3.46E04	7.02E05 ± 7.02E05	6.30E05 ± 4.80E04
Choline	3.19–3.21	3.31E07 ± 1.60E06	2.77E07 ± 1.70E06	3.06E07 ± 1.60E06
Glutamate	2.32–2.35	3.40E06 ± 2.35E05	4.35E06 ± 3.09E05	4.71E06 ± 2.98E05
Threonine	4.21–4.29	7.94E06 ± 5.55E05	6.41E06 ± 5.41E05	6.44E06 ± 3.66E05

<sup>a</sup>Chemical shift of region used for quantification.

mechanisms behind this disease and could have a potential role in the identification of new biomarkers indicative of response to treatment.

As a proof-of-principle, serum metabolic profiles of healthy individuals and multiple myeloma patients at the time of diagnosis were examined. Initial analysis using unsupervised PCA did not indicate any obvious disease-related variation in the metabolic profile. However, a supervised multivariate OPLS-DA modeling enabled good separation of healthy individuals and multiple myeloma patients revealing clear differences in their metabolic profiles. The major metabolic differences detected in multiple myeloma patients at the time of diagnosis included increased levels of amino acids such as isoleucine, phenylalanine, arginine and tyrosine, and acetate, as well as decreased levels of 3-hydroxybutyrate, cholesterol, lysine, lipids, choline, and glutamine. Alterations in the fatty acid profiles of plasma lipids are quite common in cancer and have been shown in a variety of neoplastic processes (22). All saturated, monounsaturated, and essential fatty acids and their polyunsaturated derivatives are decreased, apparently as a consequence of the disease itself (23). Decreased cholesterol levels in multiple myeloma patients have been previously described, probably representing increased LDL clearance and utilization of lipids and cholesterol by myeloma cells (24). In fact, it has been shown that NS0 myeloma cells are naturally cholesterol dependent; probably due to a biochemical deficiency in the demethylation of lanosterol to C-29 sterols (25, 26). Previous studies (27, 28) have also shown that lysine and glutamine are absolutely required for myeloma cells growth. Glutamine is not only rapidly used by myeloma cells but also depleted by spontaneous conversion to pyrrolidone carboxylic acid and by enzymatic modification to glutamic acid (29). Differences in the observed amino acid serum levels of multiple myeloma patients could be associated with the specific features of the proteins that are produced in the course of this disease. Thus, it has been reported (30) that serum proteins isolated from multiple myeloma patients display a preferential utilization of aspartic acid, glutamic acid, lysine, proline, and serine, and comparatively show a low utilization of arginine, isoleucine, phenylalanine, tryptophan, and tyrosine. The comparison of the metabolic profiles of multiple myeloma patients at the time of diagnosis and healthy controls did not reveal statistically significant differences in the glucose levels of the 2 groups, a result that is in agreement with PET/CT observations obtained during evaluation of multiple myeloma patients. Thus, it has been reported that, despite its sensitivity and usefulness in multiple myeloma diagnosis, the uptake of fluorodeoxyglucose ( $^{18}\text{F}$ -FDG) in multiple myeloma assessed with the maximum standardized uptake value ( $\text{SUV}_{\text{max}}$ ) can be very low, sometimes even comparable to the  $\text{SUV}_{\text{max}}$  of a benign lesion, making sometimes very difficult to distinguish between a benign lesion and a low-metabolic multiple myeloma lesion (31). However,  $^{11}\text{C}$ -Choline PET/CT seems to be more sensitive than  $^{18}\text{F}$ -FDG PET/CT for the detection of bony myelomatous lesions. Cho-

line, a small molecule precursor of phospholipids whose uptake is increased in proliferating cells as it is involved in membrane metabolism and growth, shows a statistically significant decrease when the metabolic profile of multiple myeloma patients and healthy donors were compared, thus correlating with the PET/CT data. Interestingly, choline serum levels show an opposite tendency after complete remission has been achieved, perhaps reflecting the disappearance of myeloma lesions. Taken together, these results provide evidence that serum  $^1\text{H}$ -NMR profiling harbors the capacity to identify, using a rapid noninvasive assay, disease-related changes in the serum of multiple myeloma patients.

Having determined that multiple myeloma patients present a serum metabolic signature distinguishable from healthy controls, the metabolic profiles of multiple myeloma patients at the time of diagnosis and after complete remission were examined using the same methodology. Here again, an OPLS-DA multivariate model provided a clear separation between these 2 groups of samples. In this case, the levels of glucose, citrate, lactate, and lysine were found to be significantly different. The decreased levels of glucose and concomitant increased lactate levels observed in the multiple myeloma group after complete remission, when compared to the multiple myeloma patients at diagnosis, could be related to the chemotherapy treatment. A number of reports have suggested that increased levels of serum lactate dehydrogenase (LDH), the enzyme that catalyzes the interconversion of pyruvate into lactate, could be associated to treatment with bortezomib, a drug common to most of the therapeutic regimes used in this study, and not by progressive disease (32). Moreover, the increased levels of citrate observed for the multiple myeloma group after complete remission could be explained by the inhibitory effects of antimyeloma drugs, bortezomib in particular (33), on hypoxia-inducible factors (HIFs), a family of transcription factors that promote anaerobic respiration (34) through different mechanisms. One of them involves the upregulation of pyruvate dehydrogenase kinase (PDK), an enzyme that inactivates the pyruvate dehydrogenase (E1) by ATP-mediated phosphorylation. E1 is the first component of pyruvate dehydrogenase complex that catalyzes the transformation of pyruvate into acetyl-CoA, a key molecule in the citric acid cycle. In this context, HIF inhibition promoted by bortezomib could play an important role in shifting from anaerobic to aerobic respiration, thus explaining the increased levels of citrate observed after chemotherapy treatment.

In contrast to findings obtained when comparing the metabolic profiles of multiple myeloma patients at the time of diagnosis with those of controls or multiple myeloma patients after complete remission of the disease, the statistical model obtained from the comparison of the metabolic profiles of multiple myeloma patients after complete remission and healthy individuals showed, as it would be expected, a lower discriminatory capability. This result tend to suggest that the metabolic profile of a multiple myeloma patient gets closer to the metabolic

profile of a healthy subject after having achieved complete remission.

The analysis of the variations in the mean intensities of the metabolites that were identified to play an important role in the discrimination between groups revealed 3 different behaviors (Table 3 and Supplementary Fig. S2). Some of the metabolites that differed between the healthy individuals and multiple myeloma patients at the time of diagnosis showed the same tendency when comparing the variation between multiple myeloma patients at diagnosis and after complete remission (e.g., 3-hydroxybutyrate, arginine, acetate, etc.). This behavior could reflect metabolic changes caused by the disease that are not reversed after achieving complete remission. However, there are other metabolites (e.g., citrate, lactate, etc.) that showed a statistically significant variation when comparing multiple myeloma patients at the time of diagnosis and after achieving complete remission, but not between healthy donors and multiple myeloma patients at the time of diagnosis, perhaps indicating metabolic changes, as suggested earlier, caused by the treatment but not necessarily linked to the response to treatment. Finally, metabolites, such as glutamine, cholesterol, lipids, lysine, etc., displayed a variation in the opposite direction after treatment has been initiated. This suggests that these metabolites could be reflecting the changes in the metabolic profiles of the multiple myeloma patients in response to treatment. Interestingly, these results highlight the potential value of serum metabolic profiling in the identification of specific metabolites that could be used to objectively monitoring treatment efficacies.

Although further validation of the results, using additional patient samples, will be necessary to increase the robustness of this analysis, our data support the idea that multivariate statistical analysis of  $^1\text{H-NMR}$  serum metabolic profiles obtained from multiple myeloma patients comprises both subtle and more obvious metabolic differences that could be used for objectively discriminating individuals with and without multiple myeloma, and identifying patients who achieved complete remission. The characterization of a specific metabolic profile associated to multiple myeloma holds great promise in the identification and development of new serum biomarkers for the diagnosis and treatment of this disease. Furthermore, the results associated with the response to treatment are particularly relevant because they could represent a starting point in the individualization of the therapeutic regimes for different patient groups. The fact that we found only few false negatives (88% and 95.7% sensitivity for the 2 sets of comparisons) and almost no false positives (100% and 96%

specificity for both comparisons) suggest that the measurement of serum metabolic profiles could become, if validated in prospective studies based on a much larger patient cohort, an effective screening tool in conjunction with other clinical and molecular parameters. The use of these biomarkers would help to avoid both under- and overtreatment, particularly when evaluating consolidation or maintenance therapies. Finally, a better understanding of the multiple myeloma metabolic profiles could be useful in identifying those cases that will be resistant to specific agents, would avoid repetitive errors and cumulative toxicity, and allow to explore new experimental strategies in these cohorts (35).

From a pharmaceutical point of view, one of the major problems in the therapeutic management of multiple myeloma is the development of resistance to drugs and the occurrence of adverse events. Future studies would likely focus on exploring the identification of biomarkers that could be informative about the prediction of drug toxicity, which is often measured by the presence of aberrant drug-metabolizing enzymes. Such biomarkers could be used to select drug doses and dose regimens, identify the best duration of treatments, as well as potential toxicities and efficacies of several drug candidates.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Authors' Contributions

**Conception and design:** R. Lecumberri, J.J. Lahuerta, F. Prósper, J.S. Miguel, A. Pineda-Lucena

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** L. Puchades-Carrasco, R. Lecumberri, J. Martínez-López, J.J. Lahuerta, M.V. Mateos, A. Pineda-Lucena

**Analysis and interpretation of data (e.g., statistical analysis, bio-statistics, computational analysis):** L. Puchades-Carrasco, A. Pineda-Lucena

**Writing, review, and/or revision of the manuscript:** L. Puchades-Carrasco, R. Lecumberri, J. Martínez-López, J.J. Lahuerta, M.V. Mateos, F. Prósper, J.S. Miguel, A. Pineda-Lucena

**Study supervision:** J.J. Lahuerta, A. Pineda-Lucena

#### Acknowledgments

The authors thank Centro de Investigación Príncipe Felipe for their economic support, and Bruker BioSpin for technical contributions.

#### Grant Support

This study was partially supported by the Spanish Ministerio de Ciencia e Innovación (MICINN, now MINECO, SAF2011-28350) and the Fondo de Investigaciones Sanitarias (FIS) Intrasalud Project (PS2009/01897).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 11, 2012; revised May 2, 2013; accepted July 8, 2013; published OnlineFirst July 19, 2013.

#### References

- Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch Toxicol* 2011;85:5-17.
- Gebregiorgis T, Powers R. Application of NMR metabolomics to search for human disease biomarkers. *Comb Chem High Throughput Screen* 2012 Apr 5. [E-pub ahead of print]

3. Kind T, Tolstikov V, Fiehn O, Weiss RH. A comprehensive urinary metabolomic approach for identifying kidney cancer. *Anal Biochem* 2007;363:185–95.
4. Kim K, Aronov P, Zakharkin SO, Anderson D, Perroud B, Thompson IM, et al. Urine metabolomics analysis for kidney cancer detection and biomarker discovery. *Mol Cell Proteom* 2009;8:558–70.
5. Ritchie SA, Heath D, Yamazaki Y, Grimmlt B, Kavianpour A, Krenitsky K, et al. Reduction of novel circulating long-chain fatty acids in colorectal cancer patients is independent of tumor burden and correlates with age. *BMC Gastroenterol* 2010;10:140.
6. Urayama S, Zou W, Brooks K, Tolstikov V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Commun Mass Spectrom* 2010;24:613–20.
7. MacIntyre DA, Jiménez B, Lewintre EJ, Martín CR, Schäfer H, Ballesteros CG, et al. Serum metabolome analysis by <sup>1</sup>H-NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. *Leukemia* 2010;24:788–97.
8. Guan W, Zhou M, Hampton CY, Benigno BB, Walker LD, Gray A, et al. Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinform* 2009;10:259.
9. Odunsi K, Wollman RM, Ambrosone CB, Hutson A, McCann SE, Tammela J, et al. Detection of epithelial ovarian cancer using <sup>1</sup>H-NMR-based metabolomics. *Int J Cancer* 2005;113:782–8.
10. Tiziani S, Lopes V, Günther UL. Early stage diagnosis of oral cancer using <sup>1</sup>H NMR-based metabolomics. *Neoplasia* 2009;11:269–76.
11. Tomita M, Kami K. Cancer. Systems biology, metabolomics, and cancer metabolism. *Science* 2012;336:990–1.
12. Bertini I, Cacciatore S, Jensen BV, Schou JV, Johansen JS, Kruhøffer M, et al. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res* 2012;72:356–64.
13. Mateos MV, Oriol A, Martínez-López J, Gutiérrez N, Teruel AI, de Paz R, et al. Bortezomib, melphalan, and prednisone versus bortezomib, thalidomide, and prednisone as induction therapy followed by maintenance treatment with bortezomib and thalidomide versus bortezomib and prednisone in elderly patients with untreated multiple myeloma: a randomised trial. *Lancet Oncol* 2010;11:934–41.
14. Meiboom S, Gill D. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 1958;29:688–701.
15. Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz <sup>1</sup>H and <sup>1</sup>H-<sup>13</sup>C NMR spectroscopy of human blood plasma. *Anal Chem* 1995;67:793–811.
16. Lindon JC, Tranter GE, Koppenaal D. *Encyclopedia of spectroscopy and spectrometry*. London: Academic Press; 2010.
17. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nat Rev Cardiol* 2011;8:630–43.
18. Brindle JT, Nicholson JK, Schofield PM, Grainger DJ, Holmes E. Application of chemometrics to <sup>1</sup>H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *Analyst* 2003;128:32–6.
19. Rajpal R, Dowling P, Meiller J, Clarke C, Murphy WG, O'Connor R, et al. A novel panel of protein biomarkers for predicting response to thalidomide-based therapy in newly diagnosed multiple myeloma patients. *Proteomics* 2011;11:1391–402.
20. Terpos E, Politou M, Rahemtulla A. The role of markers of bone remodeling in multiple myeloma. *Blood Rev* 2005;19:125–42.
21. Bhattacharyya S, Epstein J, Suva LJ. Biomarkers that discriminate multiple myeloma patients with or without skeletal involvement detected using SELDI-TOF mass spectrometry and statistical and machine learning tools. *Dis Markers* 2006;22:245–55.
22. Body JJ. Metabolic sequelae of cancers (excluding bone marrow transplantation). *Curr Opin Clin Nutr Metab Care* 1999;2:339–44.
23. Baro L, Hermoso JC, Nunez MC, Jimenez-Rios JA, Gil A. Abnormalities in plasma and red blood cell fatty acid profiles of patients with colorectal cancer. *Br J Cancer* 1998;77:1978–83.
24. Yavasoglu I, Tombuloglu M, Kadikoylu G, Donmez A, Cagircan S, Bolaman Z. Cholesterol levels in patients with multiple myeloma. *Ann Hematol* 2008;87:223–8.
25. De la Luz-Hdez K. *Metabolomics and mammalian cell culture*. Croatia: Ute Roessner; 2012.
26. Sato JD, Cao HT, Kayada Y, Cabot MC, Sato GH, Okamoto T, et al. Effects of proximate cholesterol precursors and steroid hormones on mouse myeloma growth in serum-free medium. *In Vitro Cell Dev Biol* 1988;12:1223–8.
27. Whitford W. NS0 serum-free culture and applications. *BioProcess Intl* 2003;2003:36–47.
28. Robert RS, Hsu H. Amino acid metabolism of myeloma cells in culture. *J Cell Sci* 1976;21:609–15.
29. Wein J, Goetz IE. Asparaginase and glutaminase activities in culture media containing dialyzed fetal calf serum. *In vitro* 1973;9:186–93.
30. Grisolia FT, Cohen PP. Amino acid analysis of serum proteins in multiple myeloma. *Cancer Res* 1953;13:851–4.
31. Nanni C, Zamagni E, Cavo M, Rubello D, Tacchetti P, Pettinato C, et al. <sup>11</sup>C-choline vs. <sup>18</sup>F-FDG PET/CT in assessing bone involvement in patients with multiple myeloma. *World J Surg Oncol* 2007;5:68.
32. Chang H, Lee SY, Tang TC. Tumor lysis syndrome in patients with light chain multiple myeloma: report of two cases. *Chang Gung Med J* 2011;34:70–5.
33. Shin DH, Chun YS, Lee DS, Huang LE, Park JW. Bortezomib inhibits tumor adaptation to hypoxia by stimulating the FIH-mediated repression of hypoxia-inducible factor-1. *Blood* 2008;111:3131–6.
34. Kiritto K. Therapeutic approaches for targeting hypoxia-inducible factor in multiple myeloma. Multiple myeloma—an overview. Croatia: InTech; 2012.
35. San-Miguel JF, Mateos MV. Can multiple myeloma become a curable disease? *Haematologica* 2011;96:1246–8.