UNIVERSITAT DE BARCELONA

Estudio de perfiles metabolómicos asociados al consumo de Legumbres y Café. Aplicación de una Aproximación Metabolómica por Resonancia Magnética Nuclear en Estudios de Intervención Nutricional y Observacionales

Francisco Javier Madrid Gambín

UNIVERSIDAD DE BARCELONA

FACULTAD DE FARMACIA Y CIENCIAS DE LA ALIMENTACIÓN

DEPARTAMENTO DE NUTRICIÓN, CIENCIAS DE LA ALIMENTACIÓN Y GASTRONOMÍA

ESTUDIO DE PERFILES METABOLÓMICOS ASOCIADOS AL CONSUMO DE LEGUMBRES Y CAFÉ. APLICACIÓN DE UNA APROXIMACIÓN METABOLÓMICA POR RESONANCIA MAGNÉTICA NUCLEAR EN ESTUDIOS DE INTERVENCIÓN NUTRICIONAL Y OBSERVACIONALES.

FRANCISCO JAVIER MADRID GAMBÍN
2017

UNIVERSITAT DE BARCELONA

FACULTAD DE FARMACIA Y CIENCIAS DE LA ALIMENTACIÓN

PROGRAMA DE DOCTORADO

ALIMENTACIÓN Y NUTRICIÓN

2014-2017

ESTUDIO DE PERFILES METABOLÓMICOS ASOCIADOS AL CONSUMO DE LEGUMBRES Y CAFÉ. APLICACIÓN DE UNA APROXIMACIÓN METABOLÓMICA POR RESONANCIA MAGNÉTICA NUCLEAR EN ESTUDIOS DE INTERVENCIÓN NUTRICIONAL Y OBSERVACIONALES.

Memoria presentada por FRANCISCO JAVIER MADRID GAMBÍN para optar al título de doctor por la Universidad de Barcelona

Dra. Mar Garcia Aloy
(Directora)

Dr. Rafael Llorach Asunción
(Director)

Dra. Cristina Andrés Lacueva
(Tutora)

FRANCISCO JAVIER MADRID GAMBÍN
2017

"*En algún lugar, algo increíble está esperando ser descubierto".*

Carl Sagan (1934-1996). Astrofísico estadounidense.

Este trabajo ha sido financiado por:

Ministerio de Economía y Competitividad y Fondo Europeo de Desarrollo Regional (FEDER).

Food Biomarker Alliance (FOODBALL) apoyado por la iniciativa "EU Joint Programming Initiative: A Healthy Diet for a Healthy Life".

European cooperation in science and technology. POSITIVe COST Action (FA1403): Interindividual variation in response to consumption of plant food bioactives and determinants involved.

Universidad de Barcelona – Ayudas de personal investigador predoctoral en formación (APIF).

Agradecimientos

Haber llegado hasta aquí ha sido un camino interesante en mi vida. Me resulta fascinante como las circunstancias te llevan a un sitio u otro en función de las oportunidades y elecciones que hacemos. Al principio de esta etapa predoctoral me costaba tomar decisiones por miedo al coste de oportunidad, sabiendo que por el mero hecho de estar eligiendo algo, estaba renunciando a otra cosa, y nunca sabré que te deparaba ese camino. Hoy día he aprendido a dejarme llevar en el micro-universo y tal como aprendí, de lo malo se aprende, de lo bueno se avanza. Así es como un día, las circunstancias y elecciones me llevaron al doctorado.

Primero de todo, me gustaría agradecer a mis directores y tutora de Tesis, los cuales me han enseñado y formado casi como persona a día de hoy. Cristina, gracias por darme la oportunidad de doctorarme en este grupo, el cuál siempre me ha parecido un grupo bastante luchador e inquieto en la ciencia. Mar, no sabría explicar lo mucho que he aprendido ti. Ha sido un honor ser tu primera Tesis, y espero que no la última. Rafa, gracias por ser un apoyo en la fase predoctoral, especialmente cuando necesitaba un cable extra, también por las enseñanzas multivariates o por la visión de ciencia que me has inculcado.

A mis compañeros de laboratorio. A todos. A Nicole, Magalí y Enrique mis compañeros de viaje predoctoral a día de hoy, que siempre me han apoyado y escuchado y con los que he compartido alegrías y tristezas. A Marisa, Arantxa, Ascen, Anna, Montse y Ximena por esos cafés y charlas vespertinas. A Rosa, Patricia, Mireia, Raúl, Sara, Olha, Anika Sheila y María por vuestros conocimientos, charlas y meriendas.

Agradezco también las enseñanzas que me han llevado también hasta aquí. A Maria Antonia, Teresa y Anders Pedersen, los técnicos de RMN que han sido como padres en esta aventura a nivel experimental. A Calle y Rikard, los suecos que durante más de medio año me han ofrecido su visión (no solo científica y estadística).

Gracias también a mis amigos de toda la vida que siempre me han apoyado: la prima Lydia, Carlos, Irene, la prima Carmen y Elena. Por supuesto a mi manada del Altillo: Aaron, Xipi, Iván, Mario e Irene. También a mis amigos más recientes, especialmente a Marta, mi compa murciana; a Sergio, el valenciano-catalán; a Tilo; a Gioelle y Sergio; Anne; Renzo y Mariluz. A mis compañeros de carrera, especialmente Anarro, Alex y Bea.

A mi madre y a mi padre, a los que tengo mucho que agradecer.

Dedico esta última línea de agradecimientos para decir que, sin duda, esta Tesis va dedicada a Juanjo, Isaac y Xema, mi familia.

Lista de abreviaturas

2 D	2 dimensiones
AAR/BCAA	Aminoácido ramificado (del inglés, *branched-chain amino acid*)
ACM	Concentración absoluta de los metabolitos (del inglés, *absolute concentration of metabolites*)
ANOVA	Análisis de la Variancia
AUC	Área bajo la curva (del inglés, *Area Under the Curve*)
CIM	Capacidad de identificar metabolitos (del inglés, *Capacity to Identify Metabolites*)
CFCA/FFQ	Cuestionario de frecuencia de consumo de alimentos (del inglés, *Food Frequency Questionnaire*)
COSY	Espectroscopia de correlación (del inglés, *COrrelation SpectroscopY*)
CVD	Enfermedad cardiovascular (del inglés, *Cardiovascular Disease*)
D	Doblete
FDR	Tasa de falsos positivos (del inglés, *False Discovery Rate*)
FID	Decaimiento libre de inducción (del inglés, *Free Induction Decay*)
HMDB	Base de datos del metaboloma humano (del inglés, *Human Metabolome DataBase*)
CEB	Bebida de café (del inglés, *Coffee Extract Beverage*)
CGA/ACG	Ácidos clorogénicos (del inglés, *chlorogenic acids*)
CI	Intervalo de confianza (del inglés, *Confidence Interval*)
IMC/BMI	Índice de masa corporal (del inglés, *Body Mass Index*)
IQR	Rango intercuartil (del inglés, *InterQuartile Range*)
m	Multiplete
MD	Dieta mediterránea (del inglés, *Mediterranean Diet*)
NOESY	Espectroscopia de efecto nuclear "*Overhauser"* (del inglés, *Nuclear Overhauser Effect SpectroscopY*)
OSC	Corrección ortogonal de las señales (dei inglés, *Orthogonal Signal Correction*)

PC	Componente principal (del inglés, *Principal Component)*
PCA	Análisis de componentes principales (del inglés, *Principal Component Analysis)*
PLS-DA	Análisis discriminante por mínimos cuadrados parciales (del inglés, *Partial Least Squares Discriminant Analysis*)
PR	Precipitación de proteínas con metanol (del inglés, *Protein Precipitation with metanol)*
PREDIMED	Prevención con dieta Mediterránea
RMN	Resonancia magnética nuclear (en inglés, *Nuclear Magnetic Resonance*)
ROC	Característica Operativa del Receptor (del inglés, *Receiver Operating Characteristic*)
s	Singlete
SPE	Fase de extracción sólida (del inglés, *Solid-Phase Extraction*)
STOCSY	Espectroscopia de correlación estadística total (del inglés, *Statistical TOtal Correlation SpectroscopY*)
t	Triplete
T2D	Diabetes Mellitus tipo 2
TCA	Ácido tricarboxílico (del inglés, *tricarboxylic acid*)
TMA	Trimetilamina
TMAO	Trimethilamina-*N*-óxido
U	No identificado/desconocido (del inglés, *Unknown*)
UF	Ultrafiltración (del inglés, ultrafiltration)
VIP	Importancia de las variables en la proyección (del inglés, *Variable Importance in the Projection*)

Índice

RESUMEN Y OBJETIVOS

ANTECEDENTES BIBLIOGRÁFICOS

Biomarkers of intake for non-alcoholic beverages. [*En proceso de publicación*].

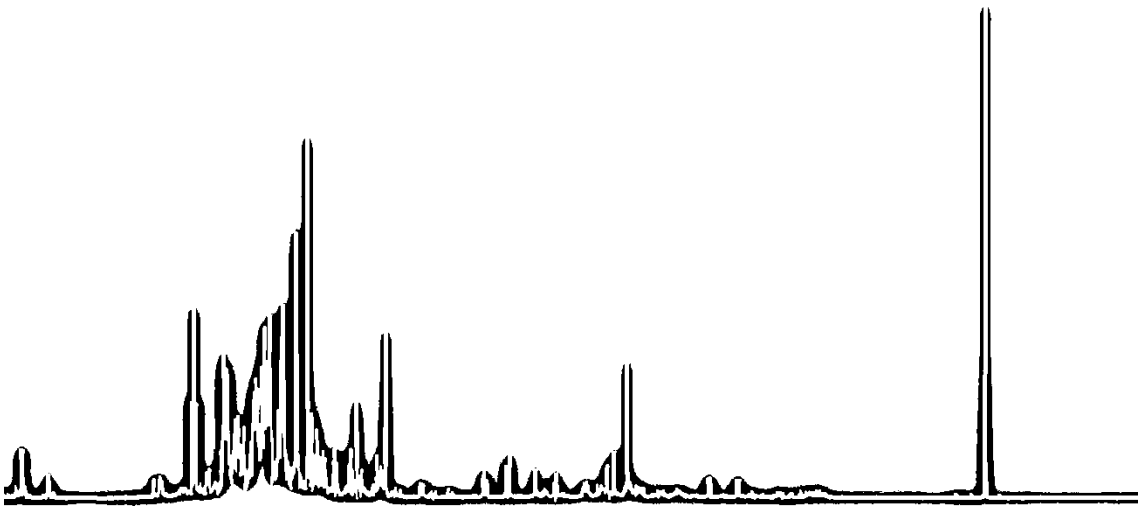
RESULTADOS

DISCUSIÓN Y CONCLUSIONES

REFERENCIAS

ANNEXOS

Resumen y objetivos



Resumen

Los biomarcadores nutricionales proporcionan una medida precisa y objetiva de la exposición dietética (biomarcadores de ingesta) y de su impacto en el individuo (biomarcadores de efecto), que unidos a las medidas tradicionales de consumo de alimentos pueden mejorar la medida de la exposición dietética y permitir la formulación de nuevas hipótesis sobre el impacto que tiene la dieta en la salud. Sin embargo, actualmente sólo unos pocos alimentos cuentan con biomarcadores de ingesta validados. Esta Tesis Doctoral se enmarca dentro del consorcio "*Food Biomarker Alliance*" (FOODBALL), el cual propone llevar a cabo una exploración sistemática de biomarcadores nutricionales y su validación para obtener una evaluación precisa de la ingesta de alimentos en diferentes grupos poblacionales dentro de Europa, mediante el descubrimiento de nuevos biomarcadores nutricionales en el marco de estudios de intervención controlados y el análisis de las respectivas muestras biológicas a través de aproximaciones metabolómicas; la revisión bibliográfica de los biomarcadores nutricionales propuestos hasta la actualidad y el desarrollo de un sistema de validación de éstos; la aplicación de este sistema de validación en los nuevos biomarcadores descubiertos en los estudios de intervención; y la exploración de los efectos biológicos por parte de los alimentos mediante el descubrimiento de biomarcadores de efecto.

Esta Tesis Doctoral tiene como objetivo principal contribuir en el descubrimiento de biomarcadores nutricionales, tanto de ingesta como de efecto, a través del análisis del perfil metabolómico asociado al consumo de legumbres y de café en muestras de orina y/o de sangre mediante Resonancia Magnética Nuclear (RMN), en estudios nutricionales con diferentes diseños.

Para la evaluación del metaboloma asociado a la ingesta de legumbres (concretamente lentejas, garbanzos y alubias) se realizó un estudio de intervención aguda en individuos sanos. El perfil metabolómico urinario se caracterizó por estar asociado a distintos tipos de metabolitos, tanto con aquellos provenientes del "*food metabolome"* (como son trigonelina, dimetilglicina, 3-metilhistidina, trimetilamina, y lisina); como con metabolitos endógenos (mayores niveles de aminoácidos ramificados, glutamina, y ácido xanturénico; y menores niveles de glucosa). El análisis

del metaboloma en suero reveló una menor cantidad de metabolitos discriminantes, mayormente endógenos (asparagina, histidina, dimetilglicina, lisina, 3-hidroxiisovalerato, 2-hidroxibutirato y glucosa). Los principales candidatos a biomarcadores urinarios fueron adicionalmente explorados a través de muestras biológicas recogidas hasta 48 h tras la ingesta de los alimentos en estudio, revelando un retraso en la excreción de la mayoría de los metabolitos después del consumo de alubias y garbanzos, seguido de lentejas.

Para el análisis de la huella metabolómica urinaria asociada con el consumo de legumbres en individuos con condiciones de vida libre, se estratificaron los sujetos según el consumo reportado mediante un cuestionario de frecuencia de consumo de alimentos validado. La exposición a legumbres se asoció con metabolitos provenientes del metabolismo de la colina, el metabolismo de los aminoácidos, y con el metabolismo energético. Con el objetivo de mejorar la capacidad predictiva del consumo de legumbres, se diseñó un modelo multimetabolito que dio como resultado la agrupación de glutamina, dimetilamina y 3-metilhistidina. El modelo fue evaluado por una curva característica de funcionamiento del receptor (ROC) exhibiendo mayor área bajo la curva (>90%) en comparación con los metabolitos evaluados de forma individual (<90% en todos los casos). Los metabolitos replicados en los estudios de intervención aguda y observacional fueron dimetilglicina, 3-metilhistidina y trimetilamina como precursor de trimetilamina-*N*-óxido (provenientes del "*food metabolome*"); y glutamina y aminoácidos ramificados, así como unos menores niveles de glucosa como metabolitos endógenos. Por lo tanto, debido al cumplimiento de varios aspectos relacionados con la validez de un biomarcador, estos metabolitos podrían ser los candidatos más robustos a biomarcadores nutricionales de legumbres.

En el caso del estudio de café, el análisis del metaboloma urinario tras el consumo agudo de una bebida a base de café con alto contenido en compuestos bioactivos indicó un incremento en la excreción urinaria de 2-furoilglicina y compuestos endógenos, tales como ácidos succínico, cítrico, 3-metil-2-oxovalérico e isobutírico. El consumo sostenido de café exhibió un aumento en la excreción urinaria de compuestos derivados de la acción de la microbiota, como los ácidos hipúrico, 3-(3-hidroxifenil)-3-hidroxipropiónico y 3-hidroxihipúrico. Por otra parte, la trigonelina se encontró en la orina tras el consumo tanto agudo como sostenido del café.

En ambos estudios de intervención (con legumbres y café) se estudió el perfil metabolómico de los propios alimentos, lo que facilitó la interpretación biológica como

biomarcadores derivados del "*food metabolome"*. En el caso de las legumbres, trigonelina, histidina, lisina, y aminas cuaternarias se observaron previamente en el alimento. Las diferencias de estos metabolitos en orina tras la ingesta de las diferentes legumbres concordaban con las distintas concentraciones de éstos en los alimentos de origen. Asimismo, en el caso del café, se identificaron compuestos furanos como precursores de la 2-furoilglicina, mientras que la trigonelina fue excretada aparentemente sin ninguna biotransformación, lo que sugiere una menor variación interindividual de este compuesto.

Esta Tesis Doctoral presenta diferentes aproximaciones a través de estudios de intervención nutricional y observacionales, y mediante diferentes biofluídos (orina y sangre) para el descubrimiento de biomarcadores nutricionales asociados al consumo de legumbres y café mediante la aplicación de una estrategia metabolómica no dirigida usando la RMN.

Abstract

Nutritional biomarkers provide an accurate and objective measure of dietary exposure (biomarkers of intake) and its impact on the individual (biomarkers of effect), that coupled with traditional measures of food consumption can improve the measure of the dietary exposure and enable the formulation of new hypotheses of the impact of diet on health. However, there are only few foods that hold validated biomarkers of intake. This Doctoral Thesis has been developed within the consortium "*Food Biomarker Alliance*" (FOODBALL), which proposes to identify and validate food intake biomarkers for a range of foods in different populations across Europe, through the discovery of new nutritional biomarkers in acute intervention studies and the analysis of the corresponding biological samples by metabolomic approaches; extensive reviews of the bibliography about existing nutritional biomarkers and the development of a new validation system for food intake biomarkers; the application of this validation system in the new discovered biomarkers in the acute intervention studies; and the exploration of the biological effects of foods through the discovery of biomarkers of effect.

This Doctoral Thesis aims to contribute to the discovery of nutritional biomarkers, both biomarkers of intake and effect, through the analysis of the metabolomic profile associated with the consumption of pulses and coffee in urine and/or blood samples by nuclear magnetic resonance (NMR) in nutritional studies with different designs.

To investigate the metabolomic fingerprint associated with dietary pulses intake (specifically lentils, chickpeas and beans) an acute intervention study was carried out in healthy individuals. The urinary metabolic fingerprint was associated with different types of metabolites, those coming from the food metabolome (such as trigonelline, dimethylglycine, 3-methylhistidine, trimethylamine, and lysine); and with endogenous metabolites (higher levels of branched-chain amino acids, glutamine, and xanthurenic acid; and lower levels of glucose). The analysis of the serum metabolome revealed a lower amount of significant compounds, mostly endogenous metabolites (asparagine, histidine, dimethylglycine, lysine, 3-hydroxyisovalerate, 2-hydroxybutyrate and glucose). The main candidates for urinary biomarkers were further explored through biological samples collected up to 48 h after ingestion of the studied foods revealing a

delay in the excretion of most metabolites after consumption of beans and chickpeas followed by lentils.

For the analysis of the urinary fingerprint associated with the consumption of dietary pulses in a free-living population, subjects were stratified according to the reported consumption by a validated food frequency questionnaire. The exposure to pulses was associated with metabolites derived from the metabolism of choline, metabolism of amino acids, and the energy metabolism. In order to improve the predictive capacity of the pulses consumption, a multi-metabolite biomarker model was built that resulted in the clustering of glutamine, dimethylamine and 3-methylhistidine. The model was assessed by a receiver operating characteristic (ROC) curve exhibiting a larger area under the curve (> 90%) compared to individual metabolites (<90% in all cases). The metabolites from the intervention study replicated in the observational study were dimethylglycine, 3-methylhistidine and trimethylamine as the precursor of trimethylamine-*N*-oxide (biomarkers from food metabolome); and glutamine, branched-chain amino acids, and lower glucose as endogenous metabolites. Therefore, due to the fulfillment of several aspects related to the validity of a biomarker, these metabolites could be the most robust candidates for nutritional biomarkers of pulses.

Concerning the metabolomic study of the urinary fingerprint after a coffee with a high content of bioactive compounds, the acute consumption indicated a higher urinary excretion of 2-furoylglycine; and endogenous compounds such as succinic, citric, 3-methyl-2-oxovaleric and isobutyric acids. Sustained consumption of coffee exhibited an increase in compounds derived from the microbiota activity, such as hippuric, 3-(3-hydroxyphenyl)-3-hydroxypropionic and 3-hydroxyhippuric acids. On the other hand, trigonelline was found in the urine after both acute and sustained consumption of coffee.

In both intervention studies (with legumes and coffee) the metabolic profile of the food was studied indicating certain precursors of compounds from the food metabolome. Pulses showed differential concentrations of trigonelline, histidine, lysine, and quaternary amines. The differences of these metabolites in urine after the ingestion of the different pulses were in agreement with the differential concentrations in the foods. Likewise, in the case of coffee, furan compounds were identified which resulted in 2-furoylglycine, whereas the trigonelline was apparently excreted without any biotransformation, suggesting a lower inter-individual variation of this compound.

This Doctoral Thesis presents different approaches through intervention and observational studies, through different biofluids (urine and blood) for the discovery of nutritional biomarkers associated with the consumption of pulses and coffee through the application of an untargeted metabolomic strategy using NMR.

Objetivos

El **objetivo principal** de esta Tesis Doctoral es identificar biomarcadores asociados con la ingesta de legumbres y café, así como biomarcadores relacionados con el efecto de esta ingesta mediante la aplicación de una estrategia metabolómica no dirigida usando Resonancia Magnética Nuclear (RMN) en estudios nutricionales con diferentes diseños.

Para la consecución de este objetivo principal se han planteado los siguientes **objetivos específicos:**

1. Desarrollar un método analítico para la identificación y cuantificación de metabolitos en muestras de suero y plasma mediante RMN (publicación 1).
2. Identificar el perfil metabolómico en orina y/o sangre asociado a la ingesta de legumbres y café en estudios de intervención dietética (publicaciones 2 y 3).
3. Identificar el perfil metabolómico de los alimentos objeto de los estudios de intervención dietética, tres clases de legumbres y una bebida a base de café, como herramienta para corroborar los biomarcadores asociados a su ingesta (publicaciones 2 y 3).
4. Estudiar la idoneidad del uso de diferentes especímenes biológicos: determinar la utilidad de las muestras de suero y/o plasma frente a la utilización de muestras de orina en el descubrimiento de biomarcadores dietéticos (publicación 2).
5. Identificar el perfil metabolómico asociado al consumo de legumbres a partir del estudio de biomarcadores de ingesta y efecto en muestras de orina y mediante el uso de cuestionarios de frecuencia de consumo de alimentos en un estudio observacional (publicación 4).
6. Determinar el poder predictivo de biomarcadores dietéticos combinados en modelos multi-metabolito frente al de los metabolitos individuales (publicación 4).
7. Confrontar los biomarcadores asociados al consumo de legumbres obtenidos a partir de estudios de intervención dietética con los provenientes de estudios observacionales (publicaciones 2 y 4).

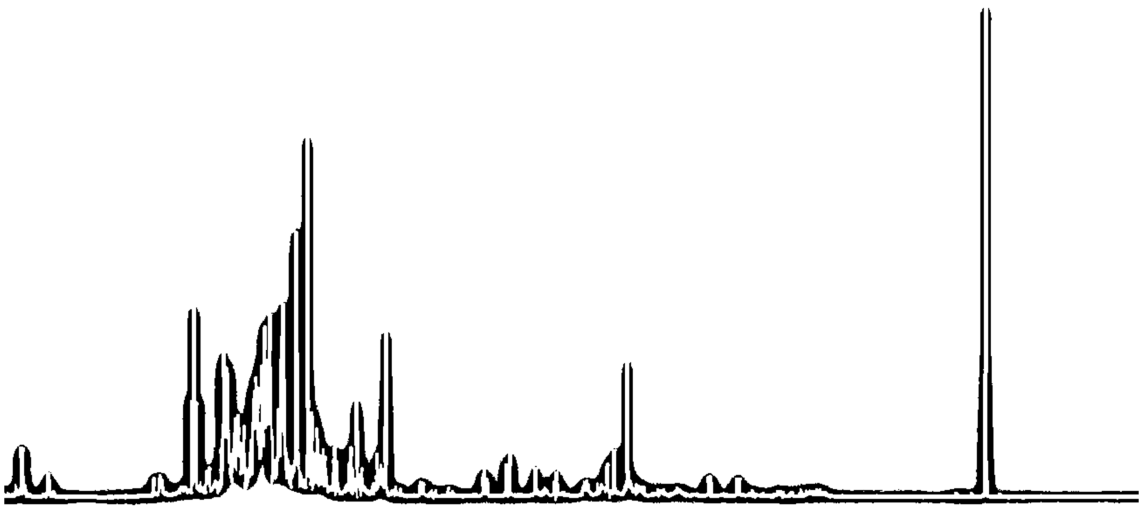
Aims

The **main objective** of this Doctoral Thesis is to identify biomarkers related to the intake of certain foods, as well as biomarkers related to the effect of this intake through the application of an untargeted metabolomics approach using the Nuclear Magnetic Resonance (NMR) in nutritional studies with different study designs.

In order to achieve this main objective, the following **specific objectives** have been proposed**:**

1. To develop an analytical method for the identification and quantification of metabolites in serum and plasma samples by NMR (manuscript 1).
2. To identify the metabolomic fingerprint in urine and/or blood associated with the intake of dietary pulses (legumes) and coffee through intervention studies (manuscripts 2 and 3).
3. To identify the metabolic profile of the foods used in the intervention studies, a coffee-based beverage and three types of pulses, as a tool to corroborate the biomarkers associated with their intake (manuscripts 2 and 3).
4. To study the suitability of using different biological specimens: to determine the utility of serum and/or plasma samples against the use of urine samples in the discovery of dietary biomarkers (manuscript 2).
5. To identify the metabolomic fingerprint associated with the consumption of pulses through the study of biomarkers of intake and effect in urine samples using food frequency questionnaires in an observational study (manuscript 4).
6. To determine the predictive power of combined dietary biomarkers in multi-metabolite models versus individual metabolites (manuscript 4).
7. To compare the nutritional biomarkers associated with the consumption of pulses from the observational study with those coming from the intervention study with pulses (manuscripts 2 and 4).

Antecedentes bibliográficos



Introducción

La presente Tesis Doctoral se enmarca dentro del proyecto europeo "*Food Biomarker Alliance*" (FOODBALL; http://foodmetabolome.org/), el cual está formado por un consorcio de 22 grupos de investigación provenientes de 11 países, y es apoyado por la iniciativa "*European Joint Programming Initiative: A Healthy Diet for a Healthy Life*" (JPI-HDHL; https://www.healthydietforhealthylife.eu/). El consorcio FOODBALL tiene como objetivos: i) contribuir a la validación de las herramientas de evaluación dietética existentes; ii) proporcionar nuevos biomarcadores nutricionales; y iii) mejorar la fiabilidad de los estudios observacionales para el estudio del papel de la dieta en la salud humana. Para ello, propone llevar a cabo, entre otras actividades, una exploración sistemática de biomarcadores nutricionales presentes en la bibliografía científica; y el descubrimiento y validación de nuevos biomarcadores a través de la nutrimetabolómica. En este marco, la presente Tesis Doctoral se ha centrado en el estudio de los biomarcadores nutricionales asociados con el consumo de legumbres y de café, cuyos antecedentes serán presentados a continuación como parte de la contextualización.

En primer lugar se presentan los biomarcadores nutricionales, así como los diferentes criterios que deben ser considerados para su validación, tanto biológica como analítica, para dar pie a los diferentes diseños de estudio que suportan esta validación.

A continuación, se describe la metabolómica en su conjunto como herramienta para el estudio de los biomarcadores nutricionales, para después describir de forma general el diagrama de flujo que se sigue en todo estudio metabolómico, así como una reseña de las principales plataformas analíticas utilizadas en la actualidad para abordar un análisis metabolómico.

Posteriormente, se describe el uso de la metabolómica aplicada a la nutrición bajo el término de nutrimetabolómica, dentro del cual se habla de los conceptos de metaboloma y componentes de este, así como de su aplicabilidad a través de los conceptos de "biomarcador de ingesta" y "biomarcador de efecto", los cuales se encuentran en esta especialidad.

Para acabar este apartado introductorio, se exponen las evidencias científicas existentes sobre la aplicación de la nutrimetabolómica para el descubrimiento y validación de biomarcadores de ingesta de los alimentos estudiados en la presente

Tesis Doctoral. Esta información se presenta mediante dos revisiones bibliográficas dentro del marco del consorcio FOODBALL (JPI-HDHL) tituladas "*Biomarkers of legume intake in human intervention and observational studies: A systematic review*" y "*Biomarkers of intake for non-alcoholic beverages*" (ambas en proceso de publicación). En ellas se presenta una recopilación de los estudios existentes en la actualidad dónde se exponen los potenciales biomarcadores de ingesta de los alimentos analizados mediante la aproximación nutrimetabolómica.

Capítulo 01.

BIOMARCADORES NUTRICIONALES

A lo largo de los años, el estudio de la inherente relación entre alimentación y salud ha tomado importancia en la prevención de numerosas enfermedades crónicas, pero también en el aumento de la longevidad y la mejora de la calidad de vida (Fontana y Partridge 2015). Los mecanismos específicos implicados en esta relación no están claramente establecidos, por lo que para su correcta investigación resulta de gran importancia disponer de una evaluación de la exposición dietética altamente precisa. Por lo tanto, en los estudios de nutrición y salud es esencial que las exposiciones dietéticas sean determinadas con una precisión suficiente que permita la correcta evaluación del impacto de la alimentación sobre la salud (Dragsted 2017). Aunque en los últimos años los instrumentos de evaluación dietética basados en encuestas alimentarias han sido altamente mejorados, a menudo concurren distorsiones en la ingesta derivadas de la subjetividad en la estimación de la dieta (Illner et al. 2012).

Ante la necesidad de una valoración de la ingesta dietética más precisa se recurre al uso de biomarcadores que puedan monitorizar la dieta de una forma más precisa. La Organización Mundial de la Salud clasifica los biomarcadores como: i) biomarcadores de exposición, definidos como aquellos elementos exógenos, o sus metabolitos, o el producto de la interacción entre un agente xenobiótico y una molécula o célula diana, que pueden ser medidos en alguna muestra biológica del organismo; ii) biomarcadores de efecto, definidos como aquellas medidas relacionadas con una alteración bioquímica, fisiológica, de comportamiento o cualquier otra alteración en el organismo que, en función de su magnitud, puede ser asociada con una enfermedad; y iii) biomarcadores de susceptibilidad, definidos como sustancias que indican la capacidad del organismo para responder a una determinada exposición (OMS 1993). En el ámbito de la nutrición es importante tener presente que mediante la alimentación se producen cambios en el metabolismo, los cuales pueden ser medidos mediante el análisis cuantitativo y cualitativo de los metabolitos en una muestra biológica, y que pueden utilizarse como biomarcadores de exposición, de efecto o de susceptibilidad. Al aplicar el concepto de biomarcador a las ciencias de la alimentación obtenemos un nuevo concepto: el biomarcador nutricional (Potischman y Freudenheim 2003).

En el trabajo de Potischman y Freudenheim (2003) se definió un biomarcador nutricional como cualquier indicador bioquímico, funcional o clínico observado en una muestra biológica que refleja el estado nutricional con respecto a la ingesta o metabolismo de los constituyentes de la dieta, así como las consecuencias fisiológicas de dicha dieta. Respecto a sus aplicaciones, un biomarcador nutricional puede ser utilizado para la validación de encuestas alimentarias, siempre y cuando este biomarcador tenga una fuerte relación directa con la correspondiente ingesta que ése está evaluando (Potischman y Freudenheim 2003). Una segunda aplicación de los biomarcadores nutricionales sería la de servir de indicadores de la ingesta o exposición dietética a un determinado nutriente o constituyente de la dieta, ya que el uso complementario de un biomarcador y la evaluación dietética mediante los métodos tradicionales basados en cuestionarios y registros podría proporcionar una estimación de la exposición nutricional más completa (Garcia-Aloy et al. 2017). Finalmente, los biomarcadores nutricionales también pueden ser utilizados para la monitorización del estado nutricional tras la ingesta dietética y que reflejan su impacto en el organismo (Potischman y Freudenheim 2003).

A la luz de lo anteriormente expuesto, los biomarcadores nutricionales conforman una herramienta precisa a la hora de evaluar la ingesta y exposición dietética.

1.1. Validación de los biomarcadores nutricionales

Antes de que un biomarcador nutricional pueda ser utilizado de forma sistemática es fundamental que éste haya sido validado. Para la validación de dichos biomarcadores nutricionales es esencial considerar aspectos tanto analíticos como biológicos (Andersen 2014; Vázquez Fresno 2015a). Durante el proceso tanto de descubrimiento como de validación de los biomarcadores nutricionales es importante que estos sean valorados a través de diferentes estudios con distintos diseños y en diferentes poblaciones, ya que la información que se deriva de estos datos indicará el grado de cumplimiento de las diferentes condiciones que se requieren para valorar la utilidad de un biomarcador (Scalbert et al. 2014). Por lo tanto, el uso de distintas aproximaciones con diferentes diseños de estudio ayudará a validar cada biomarcador nutricional en función de los criterios resumidos en la Tabla 1, y expuestos en el siguiente apartado.

Tabla 1. Parámetros a considerar para la validación de un biomarcador nutricional. Adaptación de Lay 2006 (Lay et al. 2006), Kuhnle 2012 (Kuhnle 2012), Andersen 2014 (Andersen 2014) y Scalbert 2014 (Scalbert et al. 2014)

Validación biológica	Relación con el alimento	Debe haber una relación de causalidad del biomarcador con respecto a la exposición dietética.
	Sensibilidad y especificidad	El biomarcador debe ser altamente particular del alimento (especificidad) y detectable en caso de haber ingerido el alimento (sensibilidad).
	Dosis-respuesta	Debe haber una correlación positiva entre el nivel de ingesta y el nivel del biomarcador.
	Tiempo-respuesta	Debe determinarse el comportamiento del biomarcador desde su aparición tras la ingesta dietética hasta su eliminación, lo que determinará si se trata de un biomarcador que refleja una ingesta reciente, o bien, una ingesta prolongada.
	Acumulación	Es necesario conocer si existe una correlación positiva entre la ingesta repetida y el acúmulo del biomarcador.
	Variación inter-individual	La variación inter-individual (debida, por ejemplo, al genotipo, edad, sexo, microbiota, etc.) debe ser mínima y conocida.
	Robustez	El biomarcador debe mantener su capacidad incluso tras ingestas complejas incluyendo varios alimentos.
Validación analítica	Muestras	Se debe definir qué tipo de muestra es necesaria para la determinación del biomarcador, así como en qué momento se debe recoger, y las condiciones de almacenamiento y preparación.
	Estabilidad	El biomarcador debe ser química y biológicamente estable.
	Medidas definidas	El biomarcador debe ser cuantificable con un error analítico definido.
	Reproducibilidad	El método analítico empleado para la determinación del biomarcador debe ser reproducible entre distintos laboratorios.

I) Validación biológica

Relación con la exposición: El biomarcador nutricional debe estar asociado con la correspondiente exposición dietética a través de una relación de causalidad, es decir, el biomarcador debe ser un compuesto o un metabolito de un componente característico del alimento. Para abordar este parámetro es necesaria la caracterización de la composición química del alimento, así como el uso de estudios de intervención que permitan proporcionar el asentamiento de esta relación causal (Heinzmann et al. 2010; Garcia-Perez et al. 2016).

Sensibilidad y especificidad: El biomarcador nutricional debe ser propio de una exposición dietética determinada, en la medida de lo posible. Para ello, la proporción de resultados verdaderos positivos (sensibilidad) y de resultados verdaderos negativos (especificidad) tras la exposición debe ser lo más alta posible (Vázquez Fresno 2015).

Dosis-respuesta: Tras la ingesta de un determinado alimento, el biomarcador aparecerá en la muestra estudiada. Sin embargo, la bondad de un biomarcador también viene determinada por una correlación positiva entre la concentración de éste y el nivel de la exposición al alimento estudiado. Por tanto, es necesario el estudio del comportamiento del biomarcador tras la ingesta de diferentes cantidades (dosis) del alimento (Brennan 2017). Esta relación se puede determinar tanto mediante estudios de intervención con un alimento en diferentes dosis (Garcia-Perez et al. 2016), como en estudios observacionales donde se dispone de información dietética cuantitativa de los participantes (Andersen et al. 2014).

Tiempo-respuesta: Para determinar la aplicabilidad de un biomarcador es importante conocer con precisión su comportamiento a lo largo del tiempo tras la exposición dietética. Consecuentemente, los biomarcadores pueden reflejar la exposición (i) a corto plazo, expresando la ingesta de horas o días atrás; (ii) medio plazo, reflejando la ingesta a lo largo de semanas; y por último, (iii) a largo plazo, indicando la exposición de meses o años atrás (Potischman 2003). Este aspecto está relacionado con el estudio del comportamiento del biomarcador en un biofluido a lo largo del tiempo desde su aparición tras la ingesta del alimento estudiado hasta su completa eliminación del organismo. Para conocer esta relación son necesarios estudios de intervención aguda o puntual de un alimento con la correspondiente investigación de la cinética del biomarcador. Para ello, es necesaria la toma de muestras biológicas en diferentes puntos temporales tras la intervención (Heinzmann et al. 2010).

Acumulación: Es necesario conocer si la exposición continuada y repetida al alimento en estudio conlleva a una acumulación del biomarcador en la muestra biológica estudiada, especialmente a medio o largo plazo. Para ello, los estudios de intervención prolongada y repetida con varias medidas de los biomarcadores a lo largo del tiempo son de mucha utilidad para determinar existencia de acumulación (Clarke et al. 2016).

Variación inter-individual: Las características de la población donde se aplique el biomarcador deben ser consideradas para el estudio de posibles fuentes de variación inter-individual (Manach et al. 2017). Por lo tanto, a la hora de estudiar un biomarcador es importante resaltar que un compuesto que no sea metabolizado puede presentar una menor variabilidad inter-individual en comparación con un compuesto susceptible de ser metabolizado (Primrose et al. 2011).

Robustez: Un buen biomarcador nutricional debería seguir estando relacionado con la exposición dietética incluso tras ingestas complejas que incluyan diferentes alimentos o patrones alimentarios. En los estudios de intervención, la dieta habitualmente es controlada tanto en los alimentos a ingerir como en los alimentos a evitar, restringiéndose alimentos que también podrían estar asociados con el biomarcador estudiado (Scalbert et al. 2014). Por el contrario, en los estudios observacionales la dieta no está controlada, lo que implica que si un biomarcador sigue teniendo la capacidad de caracterizar a los consumidores del alimento estudiado, éste estará indicando alta robustez. La presencia de esta cualidad es fácilmente visible cuando se replica un biomarcador en diferentes estudios de intervención y observacionales (Garcia-Aloy et al. 2015). Mientras que la relación con la exposición se basa en la composición química de los alimentos, la robustez se evalúa en base a cuanto particular es el biomarcador del alimento evaluado bajo condiciones en las que se consumen otros alimentos a la vez.

II) Validación analítica

Muestras: El tipo de muestra en la cual se va a medir el biomarcador nutricional está relacionado con la duración de la exposición y las características del biomarcador, reflejando diferentes grados de la antigüedad de la exposición en función del tipo de muestra seleccionada. Por ejemplo, los biomarcadores medidos en la orina, el plasma o el suero reflejan una exposición a corto plazo, mientras que las mediciones de biomarcadores en los eritrocitos dan como resultado una exposición dietética a medio

plazo (Kuhnle 2012). Asimismo, las mediciones de biomarcadores en pelo, uñas o dientes se emplean con mayor frecuencia como biomarcadores a largo plazo (Figura 1).

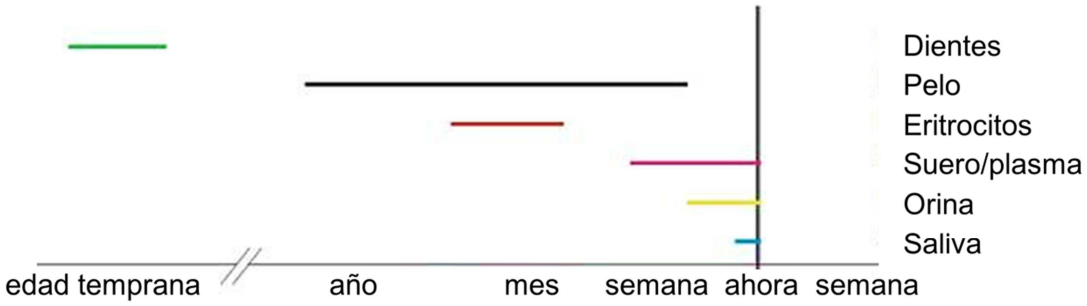


Figura 1. Representación temporal de los diferentes métodos de evaluación dietética a partir de diferentes tipos de muestra. Adaptación de Kuhnle 2012 (Kuhnle 2012).

Estabilidad: Habitualmente suele transcurrir un tiempo desde que las muestras biológicas son recogidas hasta que éstas son analizadas para la determinación de los biomarcadores. A la vez, las muestras pueden sufrir algunos ciclos de congelación-descongelación antes de su utilización. En este sentido, el biomarcador no debería degradarse ni verse afectado por las condiciones de conservación y/o tratamiento de la muestra biológica hasta el momento de su análisis. Tampoco deben afectarle las condiciones en las que se mantiene durante el análisis experimental para la determinación del biomarcador. En este sentido, es necesario conocer el grado de estabilidad del biomarcador a lo largo del tiempo en determinadas condiciones de almacenamiento y tratamiento (Gika et al. 2008).

Método analítico: La medida de un biomarcador va condicionada por un método analítico válido que la determine, el cual debe presentar una reducida variabilidad analítica (por ejemplo, expresado como coeficiente de variación), y una alta exactitud, linealidad y precisión; así como se deben conocer los límites de detección y cuantificación (Shrivastava y Gupta 2011).

Reproducibilidad: El método analítico diseñado para la determinación del biomarcador ha de ser reproducible en diferentes laboratorios. Este aspecto se puede abordar mediante análisis de comparación entre distintos laboratorios ("*inter-laboratory*

comparison tests") del método analítico utilizado para la cuantificación del biomarcador (Albermann et al. 2012).

1.2. El diseño de los estudios nutricionales

En el apartado anterior se ha resaltado la importancia de la combinación de aproximaciones con diferentes diseños de estudios para la validación de un biomarcador nutricional (Garcia-Aloy et al. 2015). La cantidad de información que se puede extraer de un estudio nutricional también está determinada por el tipo de aproximación que se realice (Manach 2013; Pujos-Guillot et al. 2013). Los datos y muestras a utilizar pueden provenir de estudios de intervención (a corto, medio y/o largo plazo) o de estudios observacionales (Figura 2).

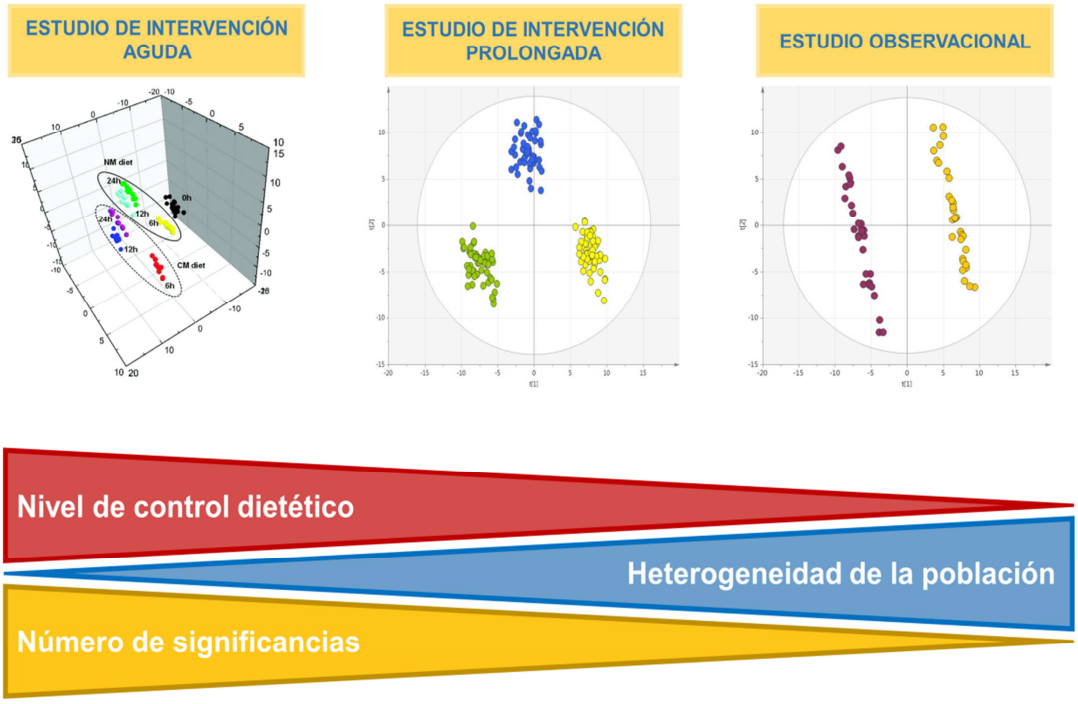


Figura 2. Comparación de los diferentes diseños de estudios. Adaptación de Pujos-Guillot *et al.* 2013 (Pujos-Guillot et al. 2013) y Manach 2013 (Manach 2013).

I) Estudios de intervención dietética

Los estudios de intervención son los estudios en los que los participantes reciben un tratamiento dado, en este caso, dietético, consistiendo en la ingesta de un determinado alimento o grupo de alimentos en una única ocasión (estudio de intervención aguda o puntual) o tras ingestas repetidas durante un periodo de tiempo acotado que puede ir desde días o semanas (estudio de intervención media) hasta meses (estudio de intervención prolongada). Las muestras biológicas se recogen al principio del estudio y tras la intervención, ya sea de una forma puntual o durante un determinado periodo de hasta 24 h, normalmente (Manach et al. 2009). Frecuentemente, en este tipo de estudios existe un gran número de biomarcadores nutricionales significativos identificados que deben de ser validados para asegurar la efectividad de los mismos. Por ello, a menudo las intervenciones dietéticas son controladas con un alimento control. La principal limitación de este tipo de estudios reside en la incerteza de la especificidad del biomarcador con respecto a la exposición, ya que, como se ha comentado con anterioridad, también suelen restringirse determinados alimentos durante el período de intervención dietética, los cuales también podrían estar asociados con el biomarcador estudiado (Scalbert et al. 2014). Por este motivo, la información extraída en estudios observacionales acerca de la dieta en poblaciones en condiciones de vida libre, al no estar restringidas, pueden incluir otros alimentos que contienen precursores de los mismos biomarcadores propuestos, lo que indicaría una baja especificad de los mismos (Edmands et al. 2015; Garcia-Aloy et al. 2015). Consiguientemente, la combinación de los estudios de intervención con estudios de observacionales, en los que la población estudiada se dispone en condiciones de vida libre sin controlar su alimentación, puede solventar esta limitación (Scalbert et al. 2014; Garcia-Aloy et al. 2015).

II) Estudios de cohorte u observacionales

Los estudios observacionales o de cohorte son también cruciales en la determinación de biomarcadores nutricionales. Habitualmente, en estos estudios los participantes se estratifican en función del consumo de un alimento, es decir, individuos con un bajo consumo o no consumidores, y con alto consumo o consumidores habituales (Llorach et al. 2012). Para tal tarea se suele recurrir a los datos de ingesta disponibles (como cuestionarios de frecuencia de consumo de alimentos, registros de 24h o diarios dietéticos) u otras herramientas de evaluación dietética para un estudio dado (Garcia-

Aloy et al. 2014a). La comparación de los subgrupos previamente definidos permite la determinación de biomarcadores dietéticos que son reflejo de la ingesta habitual en condiciones de vida libre, siempre que estos biomarcadores tengan una vida media suficiente en el organismo o que los alimentos se consuman regularmente (Scalbert et al. 2014). Sin embargo, la gran limitación que presenta este tipo de estudios anida en la colinealidad en el consumo de varios alimentos o patrones dietéticos, dadas las condiciones de vida libre y, por lo tanto, es complicado discernir exactamente si el alimento en estudio es precursor del biomarcador propuesto (Garcia Aloy 2014b). Por lo tanto, la evaluación dietética en estudios observacionales con poblaciones en condiciones de vida libre sigue siendo un reto importante (Satija et al. 2015). Cabe destacar que al utilizar este tipo de estudios (principalmente de manera transversal) no puede establecerse una relación de causalidad entre ingesta del alimento y el biomarcador, al contrario de lo que sucede con los estudios de intervención. Nuevamente, el uso de estudios con diferentes diseños permitiría la evaluación de la funcionalidad de un determinado biomarcador independientemente del diseño del estudio (Scalbert et al. 2014; Garcia-Aloy et al. 2015).

Capítulo 02.

LA METABOLÓMICA PARA EL ESTUDIO DE BIOMARCADORES

La *metabolómica* se focaliza en el análisis sistemático del metaboloma, definido como el conjunto de metabolitos de bajo peso molecular (<1500 daltons) presente en una muestra biológica bajo unas condiciones fisiológicas y/o perturbaciones determinadas (variaciones genéticas, estados patológicos o respuestas a estímulos externos) con el objetivo de proporcionar una visión general del estado metabólico y de los eventos bioquímicos asociados con un sistema celular o biológico (Kell et al. 2005). En este contexto, la metabolómica ha demostrado permitir elucidar los mecanismos bioquímicos y biológicos e identificar el impacto que tienen las enfermedades o los factores ambientales en el organismo (Dettmer et al. 2007).

2.1. Diagrama de flujo en metabolómica

Todo análisis metabolómico sigue un determinado diagrama de flujo de trabajo, el cual se desglosa en los siguientes 5 pasos generales (Figura 3) (Brennan 2014; Vázquez Fresno 2015a):

1. Recolección y preparación de las muestras biológicas
2. Adquisición de los datos
3. Procesado y análisis estadístico de los datos
4. Identificación de los metabolitos discriminantes
5. Interpretación biológica

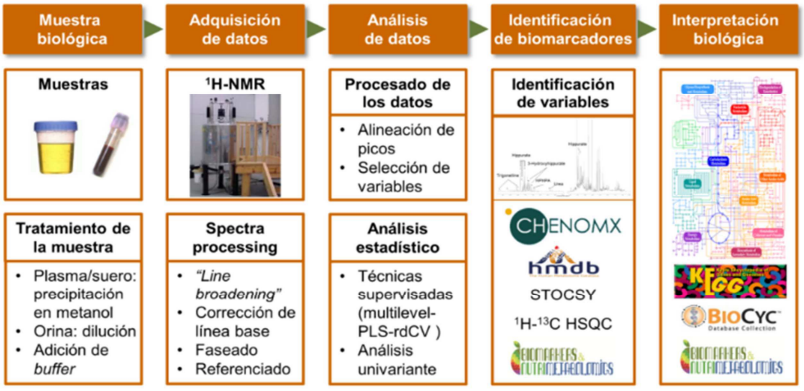


Figura 3. Diagrama de flujo del proceso de trabajo en metabolómica. Adapatación de Vázquez-Fresno 2015 (Vázquez Fresno 2015a).

- **Recolección y preparación de la muestra biológica**

Existe una amplia variedad de muestras biológicas que pueden utilizarse para realizar el estudio metabolómico de las mismas. Estas muestras pueden abarcar orina, sangre (plasma/suero), saliva, fluidos seminales, líquido amniótico, liquido cerebroespinal, liquido sinovial, fluidos digestivos y aspirados pulmonares, entre otros (Claus and Swann 2013). Los más utilizados en estudios con humanos y menos invasivos son plasma/suero, orina y saliva (Gibney et al. 2005; Llorach et al. 2012).

La preparación de la muestra consiste en la extracción de los analitos provenientes de las matrices biológicas, o dependiendo del tipo de muestra, en la simple dilución de ésta, y la consecuente eliminación de los componentes que puedan obstaculizar el análisis en función de la técnica analítica que vaya a ser usada (Beckonert et al. 2007; Want et al. 2010).

- **Adquisición de los datos**

La adquisición de los datos va a estar condicionada por la naturaleza de la aproximación metabolómica, la cual puede ser dirigida o no dirigida en función del conocimiento previo de los metabolitos susceptibles de formar parte de los resultados del análisis (Oresic et al. 2009). Las aproximaciones dirigidas o perfilado metabólico ("***metabolic profiling***") hacen referencia al análisis, generalmente cuantitativo, de un grupo de metabolitos que ya ha sido predefinido previamente y que pertenecen a una familia de compuestos o a una ruta metabólica concreta (Cevallos-Cevallos et al. 2009).

Las aproximaciones no-dirigidas o análisis de la huella metabólica ("***metabolic fingerprinting***") refieren al análisis del metaboloma en un sentido amplio y que comprenda el mayor número de metabolitos posible sin un conocimiento previo de los mismos. Por tanto, también incluye aquellos metabolitos poco caracterizados o desconocidos ("*unknowns*") (Oresic et al. 2009). Este enfoque también nos da una visión más completa a nivel cualitativo o semi-cuantitativo del metaboloma, con el objetivo de detectar el máximo número de metabolitos que expliquen un fenómeno (Cevallos-Cevallos et al. 2009). Esta técnica es una herramienta útil y necesaria en la metabolómica puesto que permite el descubrimiento de nuevos biomarcadores y es entorno a la que gira la presente Tesis Doctoral.

El metaboloma humano incluye una importante variabilidad en cuanto al tamaño, concentración, composición, etc., de los distintos tipos de compuestos presentes en el mismo. A consecuencia de esto, hasta el momento no existe una técnica analítica que abarque el análisis completo del metaboloma y, por lo tanto, se recomienda el uso de varias técnicas para obtener una visión complementa de éste. La Resonancia Magnética Nuclear (RMN) y la espectrometría de masas (EM) son las técnicas analíticas más ampliamente empleadas en metabolómica (Zhang et al. 2012). Ambas técnicas poseen diferencias fundamentales en el tipo de resultados obtenidos (Lenz and Wilson 2007). Sin embargo, aunque ambas tienen ventajas e inconvenientes (Álvarez Sánchez 2012), la RMN y la EM no deben considerarse necesariamente como una disyuntiva sino que pueden considerarse técnicas complementarias, ya que ambas técnicas combinadas proporcionan una mayor cobertura del metaboloma con una mayor recopilación de información (Deng et al. 2016), extrayendo grandes cantidades de información bioquímica con volúmenes de muestra relativamente pequeños (Hertkorn et al. 2007). La espectroscopia de ^{1}H-RMN es utilizada en metabolómica para el análisis de metabolitos de mayor abundancia en los biofluídos y que además posean protones no intercambiables (Vázquez Fresno 2015a). Esta técnica tiene como ventaja una preparación de la muestra rápida y sencilla, así como una alta reproducibilidad y robustez (Wishart 2008a). Además, este método es también no destructivo y no requiere separación previa de los metabolitos, permitiendo la recuperación de la muestra para experimentos adicionales. Sin embargo, un inconveniente de la espectroscopia de RMN es la baja sensibilidad de esta técnica, además de un gran grado de superposición espectral en las resonancias de muestras biológicas complejas (Wishart 2008b). Surgen así, por un lado, los espectrómetros modernos equipados con sondas criogénicas que aumentan la sensibilidad en gran medida, y por otro, técnicas de deconvolución espectral para la superposición de picos, además de técnicas espectroscópicas bidimensionales (Claus and Swann 2013). La EM se basa en la ionización de los compuestos presentes en la muestra que se está analizando y proporciona información relativa a la masa exacta del ion molecular de dichos compuestos, así como de los iones correspondientes a sus principales fragmentos y/o aductos producidos en la fuente de ionización (Dunn et al. 2013). Esta técnica es más sensible que la RMN, pero una gran cantidad de compuestos detectados a menudo son desconocidos (Wishart 2008b). La EM es también una técnica que requiere una mayor preparación de la muestra biológica

antes del análisis (Wishart 2008b). Esta serie de ventajas e inconvenientes se sintetiza en la Tabla 2.

Tabla 2. Principales características de las plataformas analíticas usadas en metabolómica. Adaptación de Lenz 2007 (Lenz and Wilson 2007), Wishart 2008 (Wishart 2008b) y Álvarez Sánchez 2012 (Álvarez Sánchez 2012)

Características	**RMN**	**EM**
Preparación rápida y sencilla de la muestra	+	-
Ausencia de separación cromatográfica previa	+	-
Sensibilidad de la técnica analítica	-	+
Volumen mínimo de muestra	-	+
Recuperación de la muestra	+	-
Detección de determinados compuestos	-	+
Ausencia de derivatización	+	-
Ausencia de control de pH	-	+
Alta reproducibilidad y repetitividad	+	-
Flexibilidad de la técnica	-	+

+: mayor; -: menor.

- **Análisis de los datos**

Esta etapa presenta dos pasos importantes: el primero es el procesamiento de los datos y el segundo es el análisis estadístico de los mismos (Yi et al. 2016). En este sentido, una vez obtenidos los datos, es necesario su procesado para focalizar los análisis estadísticos posteriores con la información biológicamente relevante y minimizando la variación que no está relacionada con el estudio. Estas variaciones pueden consistir en (i) diferencias en varios órdenes de magnitud entre las intensidades de las señales; (ii) diferencias en la magnitud de los cambios biológicos; (iii) variaciones biológicas que no se deben al fenómeno estudiado; (iv) variaciones técnicas debidas a errores analíticos; y/o (v) variaciones que no son constantes en todas las muestras analizadas (heterocedasticidad) (van den Berg et al. 2006). Los

procedimientos de tratamiento de datos más extendidos en metabolómica son la transformación, el centrado y el escalado. La transformación es utilizada para la corrección de la heterocedasticidad de los datos, incrementando la simetría de su distribución. El centrado modifica las oscilaciones dispersas alrededor de la media para que se dispongan alrededor de cero. Por último, el escalado adapta cada variable a una constante para que todas participen en el análisis estadístico con la misma importancia independientemente de la magnitud original de su intensidad (van den Berg et al. 2006).

A continuación, los datos previamente procesados son analizados a nivel estadístico. Para ello, pueden emplearse análisis estadísticos univariantes o multivariantes en función de si se desean testar las variables de forma individual e independiente o de forma conjunta e interrelacionada(reduciendo la dimensionalidad de los datos), respectivamente (Trygg et al. 2007; Xia et al. 2015). Los análisis estadísticos univariantes se agrupan en pruebas paramétricas (por ejemplo, análisis de t-Student y ANOVA) y no paramétricas (por ejemplo, prueba de Mann Whitney y Wilcoxon) dependiendo de la distribución de los datos. Los análisis estadísticos multivariantes se subclasifican en métodos supervisados y métodos no supervisados en función de si se considera o no la clase a la que pertenecen cada una de las muestras (Goodacre et al. 2004; Yi et al. 2016). Entre los métodos no supervisados se hallan el análisis de componentes principales (en inglés *Principal Component Analysis*, PCA) o el análisis de agrupación jerárquica (en inglés, *Hierarchical Clustering Analysis*, HCA), los cuales determinan patrones de comportamiento de los datos de acuerdo con sus características. Se utilizan para la exploración de agrupamientos de individuos o de variables en forma de "*clusters*" o bien para determinar la presencia de valores atípicos (en inglés *outliers*). Uno de los métodos supervisados más utilizados para determinar las variables relacionadas con un fenómeno dado (por ejemplo, una intervención dietética) es el análisis discriminante por mínimos cuadrados parciales (en inglés, *Partial Least Square Discriminant Analysis*, PLS-DA). Sin embargo, existe el riesgo de sobreajuste del modelo elaborado sobre los datos, con lo que a menudo son necesarios análisis adicionales, como el test de permutaciones, para corroborar que el modelo es válido (Yi et al. 2016). Una vez realizado el análisis de los datos, el siguiente paso es la identificación de los metabolitos discriminantes o significativos.

- **Identificación de metabolitos**

La identificación de metabolitos es una parte decisiva en metabolómica. La RMN permite extraer información estructural sobre un metabolito desconocido (Wishart 2008b), siempre que la superposición espectral sea mínima.

La identificación de los metabolitos se lleva a cabo a través de una serie de procesos analíticos y computacionales. Asimismo, las bases de datos de metabolómica, tales como "*the Human Metabolome DataBase"* (http://www.hmdb.ca/), "*the Food DataBase"* (http://foodb.ca/) y "*Phenol-Explorer"* (http://phenol-explorer.eu/), pueden facilitar el trabajo de identificación por comparación de resultados con los alojados en éstas (Wishart et al. 2013). La comparación de las señales con las de los estándares disponibles en los diferentes laboratorios ayuda en determinados casos en los que no se dispone el compuesto en las respectivas bases de datos. En este contexto, ha surgido la nueva librería "*the Food Compound Exchange"* (http://foodcomex.org/) que facilita el acceso a diferentes estándares mediante el intercambio de éstos entre laboratorios, como parte del proyecto FOODBALL.

Los experimentos adicionales de cada técnica ayudan al proceso de identificación y/o confirmación de un compuesto. Algunos experimentos adicionales son la fragmentación en tándem para la EM (Dunn et al. 2013) y experimentos bidimensionales para la RMN (Beckonert et al. 2007; Wu et al. 2008), los cuales permiten el estudio estructural y/o comparación con compuestos estándares.

- **Interpretación biológica**

Una vez finalizado el proceso de identificación, continúa la interpretación biológica en la que se recurre tanto a la búsqueda bibliográfica como la búsqueda en las bases de datos relacionadas con vías metabólicas visualizando estos metabolitos en las rutas correspondientes, y apoyando a la interpretación biológica de los resultados (Frolkis et al. 2010; Kanehisa et al. 2017).

Cabe destacar que la información que brinda el uso de la metabolómica en muestras de alimentos puede complementar la información que proporciona el análisis del metaboloma humano en los estudios de intervención nutricional para el descubrimiento de biomarcadores de exposición dietética (Heinzmann et al. 2010; Garcia-Perez et al. 2016). Además de los nutrientes característicos de los alimentos, éstos habitualmente

también suelen presentar una importante cantidad de otros compuestos no esenciales, los cuales, debido a que también se ven reflejados en el metaboloma humano tras su ingesta, pueden permitir una mejor caracterización de la huella metabólica asociada a su consumo y, por tanto, a la caracterización de los biomarcadores nutricionales asociados a su ingesta (Gibney et al. 2005).

Capítulo 03.

APLICACIÓN DE LA NUTRIMETABOLÓMICA PARA EL DESCUBRIMIENTO DE BIOMARCADORES NUTRICIONALES

La metabolómica presenta una amplia gama de aplicaciones en la investigación biomédica. Una de ellas es la capacidad de evaluar los mecanismos moleculares asociados a la aparición y progresión de distintas enfermedades o de controlar el cumplimiento y la respuesta a una determinada intervención terapéutica mediante el uso de biomarcadores (Patti et al. 2012). En el campo de la investigación nutricional, la metabolómica presenta una interesante aplicación para analizar el estado nutricional de un individuo, evaluar la ingesta dietética, estudiar los cambios fisiopatológicos que se producen como respuesta a una intervención nutricional, o determinar los mecanismos metabólicos asociados a las diferencias en las respuestas a las mismas intervenciones nutricionales entre distintos individuos (Llorach et al. 2012).

Para el desarrollo de la metabolómica es necesario el estudio del metaboloma. El conjunto del metaboloma está formado por diferentes subconjuntos: (i) el metaboloma endógeno, el cual incluye aquellos metabolitos provenientes del metabolismo de las células y tejidos del individuo; (ii) el "*food metabolome*", compuesto por aquellos metabolitos y compuestos procedentes de la absorción, digestión y metabolismo (tanto del propio individuo como de su microbiota intestinal) de los compuestos nutricionales y fitoquímicos presentes en los alimentos; (iii) el "*drug metabolome*" procedente de compuestos xenobióticos derivados de fármacos, entre otros; y (iv) el "*pollutant metabolome*", que engloba aquellos metabolitos provenientes del metabolismo derivado los compuestos presentes en el ambiente, producidos por la polución, etc. El metaboloma endógeno y el "*food metabolome"* se ven afectados por la dieta, provocando cambios en estos subconjuntos del metaboloma, los cuales son susceptibles de conformar biomarcadores nutricionales (Scalbert et al. 2014).

De la unión de la metabolómica y las ciencias de la nutrición ha surgido una herramienta que permite investigar la relación entre alimentación y salud (Rezzi et al. 2007). Nace así la metabolómica nutricional o la "***nutrimetabolómica*"** (en inglés *nutrimetabolomics*), la cual se define como la disciplina ómica que investiga el impacto de la dieta sobre el metaboloma (Zhang et al. 2008). Por tanto, implícitamente se

centra en el estudio de la función que ejercen los componentes de la dieta en el mantenimiento de la salud y el desarrollo de procesos patológicos, así como revelar nuevos biomarcadores que permiten monitorizar la ingesta de estos compuestos (Zhang et al. 2008).

3.1. Perfiles metabolómicos asociados al consumo de alimentos para la caracterización de biomarcadores nutricionales de ingesta y de efecto

El análisis de los perfiles metabolómicos asociados al consumo de determinados alimentos permite profundizar en el estudio de la relación entre la dieta y el metaboloma. Habitualmente estos perfiles metabolómicos acostumbran a clasificar los metabolitos discriminantes en dos grupos principales, a saber, biomarcadores de ingesta y biomarcadores de efecto (Trespalacios et al. 2012).

I) Biomarcadores de ingesta

Los biomarcadores de ingesta están relacionados con el "*food metabolome"* y representan una fuente de potenciales nuevos biomarcadores dietéticos precisos (Llorach et al. 2012). En este campo es importante tener presente que el "*food metabolome"* presenta una elevada complejidad y variabilidad, ya que se estima que existen más de 25.000 compuestos distintos en los diversos alimentos, la gran mayoría de los cuales sufren diversos procesos metabólicos en el organismo. Todos estos compuestos y metabolitos son susceptibles de desempeñar la función de biomarcador de ingesta (Scalbert et al. 2014).

Esta realidad hace que el análisis del "*food metabolome"* resulte una tarea compleja. En primer lugar, como se acaba de resaltar, comprende una amplia variedad química. Una segunda peculiaridad del "*food metabolome"* es la gran escala de concentraciones que abarca, que pueden ir desde concentraciones picomolares hasta concentraciones milimolares en función del compuesto. Por último, se desconoce una enorme cantidad de compuestos que pertenecen al "*food metabolome"*, especialmente compuestos que no son nutrientes. Algunos ejemplos de biomarcadores de ingesta son el ácido tartárico y el etilglucurónido para el vino (Vázquez-Fresno et al. 2015b); urolitina A para nueces (Garcia-Aloy et al. 2014a), 7-metilxantina y 5-(3',4'-

dihidroxiphenil)-valerolactona para el cacao (Garcia-Aloy et al. 2015), y benzoxazinoides y alquilresorcinoles para el pan (Garcia-Aloy et al. 2014c).

Hasta hace poco, el "*food metabolome"* se analizaba mediante métodos dirigidos optimizados para compuestos determinados de los que se disponía de un conocimiento previo (Rotches-Ribalta et al. 2012). Sin embargo, el uso de la metabolómica no dirigida facilita la identificación de nuevos biomarcadores nutricionales abarcando el metaboloma ampliamente y sin focalizarse en compuestos específicos previamente seleccionados (Tulipani et al. 2011; Scalbert et al. 2014).

II) Biomarcadores de efecto

Los metabolitos presentes de forma intrínseca en el metaboloma corresponden a compuestos endógenos que pueden ser modulados a través de una exposición dietética como consecuencia de diferentes procesos fisiológicos y/o patológicos derivados de esta exposición (van Ravenzwaay et al. 2007). Por tanto, son considerados biomarcadores de efecto de dicha exposición (Rezzi et al. 2007). En este tipo de biomarcadores, la RMN resulta una técnica especialmente competente debido a su capacidad de detección de metabolitos endógenos en orina (Bouatra et al. 2013) y sangre (Psychogios et al. 2011), generalmente de características diferentes a los provenientes del "*food metabolome".* El descubrimiento de varios metabolitos endógenos provenientes de una ruta metabólica determinada ofrece una valiosa herramienta de información en la interpretación biológica y el testado de un papel beneficioso o perjudicial de una determinada exposición dietética (van Ravenzwaay et al. 2007; Vázquez Fresno 2015a). Algunos ejemplos de biomarcadores de efecto son el ácido 3-metil-2-oxovalérico y el ácido 4-hidroxifenilacético para el consumo de vino (Vázquez-Fresno et al. 2015b), el ácido pirrol-2-carboxílico para la exposición a pectinas (Kristensen et al. 2012), y componentes del ciclo de Krebs y diferentes aminoácidos para el té (Assadi-Porter et al. 2015).

3.2. Uso de matrices biológicas en nutrimetabolómica para el estudio de los perfiles metabolómicos

Para el desarrollo de la nutrimetabolómica se pueden emplear diversos biofluídos tales como orina, sangre, saliva, fluido cerebroespinal, etc. (Brennan 2014). En la presente

Tesis Doctoral nos centramos en la orina y la sangre cuyas particularidades se describen a continuación.

La principal función de la orina como biofluido es la eliminación de compuestos externos en el cuerpo. Estos compuestos se vuelven más polares para facilitar la excreción renal. Por lo tanto, constituye una matriz rica de compuestos externos que pueden provenir de la dieta (por ejemplo, fitoquímicos) y, por lo tanto, servir como biomarcadores de ingesta (Claus y Swann 2013).

La sangre es un biofluido rico en nutrientes y compuestos que son transportados de un órgano o tejido a otros. La recolección de la sangre como biofluido para la identificación de biomarcadores resulta ser una técnica más invasiva pero proporciona una fotografía metabólica del estado fisiológico del organismo en el momento del muestreo (Claus y Swann 2013).

Capítulo 04.

BIOMARCADORES NUTRICIONALES DE LEGUMBRES

La dieta mediterránea (DM) se caracteriza por una alta ingesta de verduras, cereales, legumbres, nueces, pescado y aceite de oliva, una baja ingesta de carne roja y productos cárnicos procesados, y un consumo moderado de aves de corral, vino y productos lácteos (Bach-Faig et al. 2011). Se ha demostrado que este patrón dietético tiene un rol beneficioso para la salud, incluyendo prevención en las enfermedades crónicas no transmisibles como son diabetes tipo 2, obesidad, enfermedades inflamatorias, enfermedades cardiovasculares y cáncer (Salas-Salvadó et al. 2011; Estruch et al. 2013; Marlow et al. 2013; Lopez-Legarrea et al. 2014).

Las legumbres juegan un papel importante en la DM, constituyendo una buena fuente de proteína vegetal (Tabla 3), fibra alimentaria, vitaminas y minerales, así como una gran variedad de fitoquímicos (Campos-Vega et al. 2010; Roy et al. 2010; Bouchenak y Lamri-Senhadji 2013). A la vez, el consumo habitual de legumbres también se incluye en las principales directrices dietéticas en todo el mundo, como son las Guías Alimentarias para los Estadounidenses (Britten et al. 2006; U.S. Department of Health and Human Services and U.S. Department of Agriculture 2015) y en la Nueva Dieta Nórdica (Mithril et al. 2013), entre otras. El consumo de legumbres ha demostrado ejercer efectos beneficiosos sobre la salud (Dilis y Trichopoulou 2009; Bouchenak y Lamri-Senhadji 2013; Faris et al. 2013; Ramalingam et al. 2015; Becerra-Tomás et al. 2017). Sin embargo, aunque las legumbres gozan de excepcionales cualidades nutricionales, en la actualidad se está observando una disminución en la ingesta de este grupo de alimentos, pasando de los 41,0 g consumidos por persona y día en los años 60 a 10,6 g por persona y día en el año 2015 (Agencia Española de Consumo 2015; Fundación Española de la Nutrición 2017; Ministerio de Agricultura Medioambiente y Alimentación 2016).

Además, las legumbres también han sido recomendadas en la rotación de cultivos de cereales para mejorar la economía y el impacto ambiental en la producción de cereal gracias a que potencian la fijación del nitrógeno al suelo donde se cultivan y la diversificación de los cultivos (MacWilliam et al. 2014). A esto se le añade el hecho de que no requieren de un elevado consumo de recursos naturales, ya que una vez

recolectadas no necesitan ser procesadas ni refrigeradas para su conservación (Fundación Española de la Nutrición 2017).

A la luz de la importancia que tienen las legumbres en la salud, surge la necesidad de disponer de medidas precisas de su ingesta así como de medidas sobre el impacto fisiológico en el individuo derivado de tal ingesta. Debido a esto, la evaluación del conocimiento existente acerca de biomarcadores nutricionales en este grupo de alimentos resulta de gran utilidad.

En el marco del proyecto FOODBALL se presenta la siguiente revisión bibliográfica sobre los biomarcadores nutricionales de legumbres estudiados hasta la actualidad. Esta revisión se ha desarrollado en colaboración con el *Institute of Food and Health* de la *University College Dublin* (Irlanda), y el *Department of Physiology and Biochemistry of Nutrition* del *Max Rubner-Institut* (Alemania).

Este trabajo está en proceso de publicación en una edición especial de la revista *Genes and Nutrition* indexada en el *Journal Citation Reports* con un factor de impacto de 2.398 (2015) y situada en el tercer cuartil de las categorías *Nutrition & Dietetics* (41/80) (2015) y *Genetics & Heredity* (88/166).

Tabla 3. Composición nutricional por 100 g de porción comestible de las legumbres principales. Extraído de Moreiras 2013 (Moreiras et al. 2013).

Alimento	Energía (Kcal)	Proteínas (g)	Hidratos de carbono (g)	Fibra (g)	Grasas (g)
Lentejas	351	23,8	54,0	11,7	1,8
Garbanzos	373	19,4	55,0	15,0	5
Alubias	349	19,0	52,5	25,4	1,4
Guisantes	91,0	6,0	13,1	5,2	0,5
Soja	406	35,9	15,8	15,7	18,6

Biomarkers of legume intake in human intervention and observational studies: A systematic review

Siva Charan Sri Harsha Pedapati[1], Mar Garcia Aloy[2,3], Roshaida Abdul Wahab[1], Francisco Madrid-Gambin[2], Sheila Estruel-Amades[2], Bernhard watzl[4], Cristina Andrés-Lacueva[2,3] and Lorraine Brennan[1]

[1]UCD School of Agriculture and Food Science, UCD Institute of Food and Health, UCD, Belfield, Dublin 4, Ireland.

[2]Biomarkers and Nutrimetabolomic Laboratory, Department of Nutrition, Food Sciences and Gastronomy, Campus Torribera, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain

[3]CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain.

[4]Department of Physiology and Biochemistry of Nutrition, Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Karlsruhe, Germany.

Keywords: Legumes, Metabolomics, Biomarkers

Abstract

Legumes are important sources of nutrients and bioactive compounds that may influence human health and increase or decrease the risk of chronic diseases. Due to the health benefits of legumes, there is a growing interest in assessing their dietary intake across numerous populations. Monitoring specific food intakes or compliance to dietary components is complicated by dietary assessment tools based on self-reporting methods that are innately inaccurate.

Here, biomarkers of legumes are reviewed. Although some compounds have been suggested as biomarkers for soy in general, the validation of these compounds against other markers for the same food/food group needs to be performed. In addition, only few compounds have been described after pulses intake, concretely dry beans. Proposed biomarkers did not fulfill a developed scoring scheme for legume intake biomarker validity. Hence, further discoveries as well as validation studies are needed in this area and to identify reliable biomarkers of pulses intake as there is a current dearth of information

Background

Legumes are fruits or seeds of a plant belonging to the family Fabaceae and are a popular food source in the traditional diets of many regions in the world. Well-known legumes include peas, beans, lentils, lupins, chickpeas, carob, soybeans, peanuts and tamarind. They provide proteins, complex carbohydrates, soluble and insoluble fibers, a variety of phytochemicals and micronutrients such as iron, copper and manganese. Furthermore, legumes have a low glycaemic index, ranging from 10 to 40. In general, a serving of legumes (~100g fresh weight) provides 115 calories, 20g of carbohydrates, 7-9g of fiber, 8g of protein and 1g of fat [1].

Legumes, and in particular soy, constitute an important part of the diet for the majority of Asian population and have been investigated for potential health promoting effects. Consumption of an isoflavone rich soy diet has been linked to improved health outcomes in a number of studies [2-5]. These health benefits may arise in part due to the presence of various soy isoflavonoid components such as daidzein, genistein and glycitein [4]. Both daidzein and genistein are further metabolised by intestinal bacteria to equol, *O*-desmethylangolensin, dihydrogenistein and dihydrodaidzein [6, 7]. The beneficial effects of soy seem to be related to the combination of these compounds and not any one in particular. For example, daidzein and genistein were reported to have a synergistic effect on inhibiting cell proliferation and inducing apoptosis of prostate cancer cells [8]. In addition, these compounds and equol were also proposed to be antiestrogenic, antioxidative, anticarcinogenic and may protect against chronic diseases such as hormone-dependent cancer, cardiovascular diseases and osteoporosis [9-13]. On the other hand, intake of dietary pulses, a sub-group of legumes, has also been associated with reduced chronic disease risks for cardiovascular disease and cancer; as well as with improvements in intermediate cardiovascular disease risk factors, such as blood pressure, dyslipidemia, glycaemic control, and weight management [14-17]. Pulses also provide a valuable means of lowering the glycaemic-index (GI) of the diet [18].

Due to the health benefits of legumes, there is a growing interest in assessing their dietary intake across numerous populations. In order to achieve this with a high standard, it is pertinent that we have high accurate and reliable assessment tools to monitor their intake. The classical approaches of data collection tools such as food frequency questionnaires (FFQ), food dairies and 24h dietary recalls are associated with a number of errors [19-23]. Hence, there is a growing need for more objective measures of intake and biomarkers have emerged as having great potential in this

field. Such biomarkers should be able to reflect the differences in dietary intakes, should be non-invasive and responsive to the intervention study carried out [24].

The objective of this paper was to perform a systematic review of the literature and summarize the information from observational and human intervention studies on the biomarkers of legumes intake and also evaluate the validity, reproducibility and sensitivity of the proposed markers that could potentially be useful indicators of legumes consumption.

Search methodology

The literature review was conducted using the PRISMA guidelines. The reviewing process made use of all elements of PRISMA statement [25], that were relevant for a search for literature on biomarkers. In brief, original research papers and reviews were searched in three databases (PubMed, Scopus and ISI Web of Knowledge) using combinations of the grouped search terms (legume OR bean OR pea) AND (biomarker* OR marker* OR metabolite* OR biokinetics OR biotransformation) AND (trial OR experiment OR study OR intervention) AND (human* OR men OR women OR patient* OR volunteer*) AND (urine OR plasma OR serum OR blood OR excretion) AND (intake OR meal OR diet OR ingestion OR consumption OR eating OR drink*). The research was limited to papers in English language, while no restriction was applied for the publication dates. The research papers identifying or using potential biomarkers of intake for the foods were selected by one or more skilled researcher from the list of retrieved references taking some inclusion and exclusion criteria into consideration in a process outlined in Figure 1. Additional papers if any were identified from reference lists in these papers.

A secondary search was performed to identify if any other foods contain the same biomarkers or its precursor and determine the apparent specificity of the compound of interest. In this second step, Pubmed, Scopus and Web of Science were used as search platforms and the compounds checked for their specificity were genistein, daidzein, dihydrogenistein, dihydrodaidzein, enterodiol, enterolactone, matairesinol, *O*-desmethylglycitein, glycitein, kaempferol, trigonelline, pipecolic acid, indolepropionate, *S*-Methylcysteine and *N*-acetyl-ornithine and their synonyms (Table 2). For each of these potential biomarkers this additional search was performed following the strategy: (“the name and synonyms of the compound” OR “the name and synonyms of any parent compound”) AND (biomarker* OR marker* OR metabolite* OR biokinetics OR biotransformation) AND (trial OR experiment OR study OR intervention) AND (human*

OR men OR women OR patient* OR volunteer*) AND (urine OR plasma OR serum OR blood OR excretion) AND (intake OR meal OR diet OR ingestion OR consumption OR eating OR drink*).

Results and discussion

The literature search performed identified a total of 1967 articles from the three databases and the flow diagram of the study selection is represented in Fig 1. A total of 1780 articles were obtained after removal of duplicates using Endnote X7.4. Of this a total of 68 articles were selected after screening on the basis of title and abstract. Exclusion criteria for the remaining 1712 articles included the following: effect on physiology, effect on drug metabolism, *in vitro* studies, food analysis and other articles related to antioxidant markers, disease/health markers, oxidative stress markers, articles not relevant to intake biomarkers and any animal studies. Full texts of the 68 papers were downloaded and assessed further for exclusion/inclusion criteria. Exclusion criteria at this stage included animal studies, inappropriate study design and articles not specific to legumes intake. A total of 41 articles were retained and one additional paper was identified from the reference list of these papers. In total, 42 articles were used for the development of the tables. Table 1 provides a summary of these studies, including the candidate biomarkers for legumes/beans/peas intake identified through this search process.

Cross-sectional studies reporting on isoflavones and their metabolites

Examination of cross-sectional studies revealed that a number of studies investigated the isoflavones and their metabolites such as genistein, daidzein, glycitein and *O*-desmethylangolensin (O-DMA) in biological samples such as urine and blood (serum and plasma) following the consumption of soy or soy-based foods (Table 1).

A study conducted in 147 Singaporean Chinese with spot urine samples demonstrated a statistically significant, dose-dependent association between frequency of overall soy intake and levels of urinary daidzein (P= 0.03) and sum of urinary daidzein, genistein and glycitein (P= 0.04) [26]. A dose-response relationship was also observed between dietary soy consumption and urinary excretion rate in fasting first void urine samples of all individual major isoflavonoids and total isoflavonoids ($p \leq 0.05$) in Chinese women (n=60) [27]. There were also positive correlations established between urinary isoflavonoids excretion and the amount of soy food, soy protein and

soy isoflavones intake (r=0.50, p<0.001, r=0.53, p<0.001 and r=0.54, p<0.001, respectively). Similarly, another study examining a western population (n=100), showed significant correlations between soy protein intake from 24 hour recalls with daidzein (r=0.72, (CI):0.43, 0.96), genistein (r=0.67, (CI):0.43, 0.91) and total isoflavonoids (r=0.72, (CI): 0.47, 0.98) [28]. Additionally, between FFQs and urinary excretion, the correlations were (r=0.50, (CI):0.32, 0.65), (r=0.48 (CI): 0.29, 0.61) and (r=0.50 (CI): 0.32, 0.64) for daidzein, genistein and total isoflavonoids, respectively [28]. Significant correlations were reported between intake of soy foods through FFQ and 5-day diet records, with urinary genistein (r=0.40, p=0.0001), O-DMA (r=0.37, P=0.0002), daidzein (r=0.34, p=0.0007) and the sum of isoflavonoids (r=0.39, p=0.0001) in US men and women (n=98) [29]. In another US population study, positive correlations were demonstrated between self-reported soy intake and excretion of urinary isoflavones (r=0.52, p<0.001 for dietary recall and r=0.29, p<0.01 for FFQ) [30]. In a US study (n=451 women) [31], 24h urine sample measures were shown to be strongly correlated with overnight urine excretion for daidzein (r=0.84) and genistein (r=0.93). The 24h urine sample measures were also correlated with soy food questionnaire (SFQ) estimates of daidzein (r=0.48) and genistein (r=0.54) intake [31]. A strong correlation between isoflavones measured in an overnight urine and soy protein intake estimated by self-reported intake (dietary questionnaire) was also established in multiethnic population (n=102), both in the previous 24 hour (r=0.61, p<0.0001) and in the past year (r=0.32, p<0.0012) [32]. Overall, significant correlations between soy intake and urinary isoflavonoids excretion either in spot, overnight or 24h urine samples was demonstrated suggesting that these compounds have the potential to serve as dietary biomarkers.

While the above studies have focused on isoflavone levels in urine, there was also evidence to support relationships in plasma and serum. A study of four groups of 20 premenopausal British women (n=80) demonstrated significant correlations between dietary total soy intake estimated by FFQ and food diaries with plasma daidzein (r=0.74-0.78, p<0.001) and genistein (r=0.73-0.78, p<0.001) [33]. Similarly, significant correlations were reported between genistein and daidzein intakes as determined by soy FFQ with plasma concentrations (r=0.53 and 0.45) respectively in a Western population group (n=77) [34]. Furthermore, similar results were found in US postmenopausal women (n=96) with correlations varying from 0.35 to 0.43 depending on the dietary intake instrument [35]. A statistically significant (p=0.002) 3-fold difference in mean plasma levels of total isoflavones was observed between women

with high and low soy isoflavone intake levels as determined from FFQ [36]. For the correlation between serum isoflavones levels and soy intake, a significant linear trend ($p<0.01$) was observed in serum isoflavones (daidzein and genistein) concentrations across increasing categories of soy food consumption estimated by FFQ in Asian women (n=1823) [37].

In summary, the cross-sectional studies demonstrated that there were positive correlations between soy intake with urinary, plasma or serum isoflavonoids levels, mainly daidzein and genistein or total isoflavonoids, in different population groups.

Acute & intervention studies reporting relationships between soy intake and isoflavones and their metabolites

Soy isoflavones

The literature search identified a number of intervention studies which focused on soy-based diets and isoflavonoid excretion in different population groups (Table 1). Various acute studies have reported increased isoflavonoid concentrations in blood and urine following consumption of soy based foods (Table 1). To understand the metabolic fate of dietary isoflavones in humans, a study examined 24-hour urines from 12 healthy Caucasian male and female participants following three days of soy challenge: the urinary isoflavone levels (genistein, daidzein, glycitein) peaked more than 3.8 fold and returned to basal levels by day 4, while the major urinary metabolites (O-DMA, equol, 6-hydroxy-O-DMA, dihydrodaidzein) demonstrated a more significant increase of over 5 to 40 fold and progressively fell over days 4 and 5 [38]. This marked variation among the major urinary isoflavonoid metabolites may reflect variability in an individual's ability to ferment isoflavones and the fat content of the diet [39]. A randomized controlled crossover feeding study demonstrated that the urinary excretion of total isoflavonoids significantly increased with soy diet consumption (26.01 ± 2.30 µmol/day) as compared to the vegetable free (0.75 µmol/day), carotenoid (0.51 µmol/day) and cruciferous vegetable diet (1.03 µmol/day) [40]. The soy diet consisted of a normal basal diet plus 100 g tofu and 45 g of soy protein isolate served per day. The urinary excretion of lignans was higher with the consumption of carotenoid (1.99 µmol/day) and cruciferous vegetable-rich diets (3.86 µmol/day) compared to soy diet (0.84 µmol/day). Overall the results from this study provides information on the utility of urinary isoflavonoids as biomarkers of soy intake.

In another study, the urinary recovery of daidzein levels were significantly higher than genistein ($p<0.001$), while the plasma concentrations of both isoflavones did not differ significantly ($p>0.1$) after single doses of 0.7, 1.3 and 2.0 mg isoflavones/kg body weight in soybean milk [13]. Subsequently, a randomized , double-blind, crossover study involving four 9-day soy protein beverage supplementation periods established a positive dose-response between urinary isoflavones excretion and soy intake ($p=0.0001$) with no significant difference between equol excreters and non-excreters [41].

A study examining the repeated intake of consuming soy protein powder of about 60 g/day in a controlled intervention trial for over a period of 28 days demonstrated that the plasma isoflavone levels markedly increased on day 28 compared to day 0 with no change in the control group on a casein supplement diet [42]. This result was in agreement with other related studies, which demonstrated an increase in isoflavonoid excretion following a soy challenge [13, 38, 43, 44].

Glycitein is another soy isoflavonoid which constitutes 5 to 10% of the total isoflavones in the soy beans [45]. Few cross-sectional studies have reported urinary excretion of glycitein [26, 27, 32] but the excretion levels are low compared to genistein and daidzein. Maximum serum concentration for glycitein attained after ingestion of aglycone and glucosidic forms of soy beverage did not differ significantly (0.07-0.09 µmol/L) and maximum urinary excretion was reported to be ~3 µmol [46]. Similarly, maximum glycitein plasma concentration of ~200 ng/ml and maximum glycitein urinary concentration of 11,000 ng/ml was reported in a bioavailability study [47]. A large soy intervention trial demonstrated a 3-4 fold increase in glycitein levels in plasma, overnight urine and spot urine was observed in the soy group compared to the placebo. However, the magnitude of increase was smaller compared to major soy isoflavones genistein and daidzein [48].

Several studies have also reported differences in isoflavone excretion with respect to the type of soy foods with most interest in the difference between fermented and non-fermented sources. A study comparing the effects of fermented and non-fermented soy product consumption demonstrated that the urinary isoflavone recovery of genistein and daidzein was higher ($p<0.002$) when the subjects consumed tempeh (fermented) compared to the soybean pieces diet [44]. This suggests that the fermented products, due to the hydrolysis of isoflavone glucosides to their corresponding aglycones could have increased availability of the isoflavones. Contrastingly, urinary isoflavonoid

excretion showed no significant difference upon consumption of soymilk (non-fermented) compared to miso soup (fermented) ($p=0.87$) [49].

Similarly, a study comparing the effects of fermented and non-fermented soy product consumption demonstrated that the plasma concentrations of genistein and daidzein was more than twice and five times higher ($p<0.05$) when subjects consumed fermented soybean extract compared to non-fermented soybean extract [50]. Similar results were demonstrated with higher serum isoflavone concentrations (~ 2 µmol/L) attained with fermented soymilk ingestion compared to non-fermented soymilk (~0.94 µmol/L) [46]. In contrast, reports after ingestion of commercial soy supplements have shown that plasma isoflavone concentrations were higher following ingestion of glucosidic forms compared to aglycone forms [51, 52]. However, hydrolysis of isoflavone glycosides to their corresponding aglycones did not seem to alter plasma concentrations in some studies [53, 54].

Overall, the urinary excretion levels and plasma concentrations of soy isoflavones seems to alter considerably upon consumption of isoflavone aglycone and isoflavone glucoside rich foods also demonstrating the effects of fermentation on the soy products. This further reveals the importance of gut microbiota where human gut bacterial *β*-glucosidases play a role in the hydrolysis of glycosides releasing the aglycone forms of isoflavones.

Pharmacokinetics of isoflavones

A number of studies investigated the pharmacokinetic behaviour of isoflavones following soy intake. In all studies, the shapes of the plasma appearance and disappearance curves with respect to time exhibited biphasic pattern as a result of enterohepatic circulation of the compounds. Peak plasma concentration of isoflavones following consumption of a soy-based meal was reported for genistein at 8.42 ± 0.69h ($t_{1/2}$ = 5.7±1.3h) and daidzein at 7.42±0.74h ($t_{1/2}$ = 4.7±1.1h) [55]. Similarly, genistein was reported to having a longer half-life (8.36h) as compared to daidzein (5.79h) following consumption of 60g of kinako (baked soybean powder) [46]. Peak serum concentrations of daidzein and genistein on average were attained at 6.9 ± 0.7 h and 6.5 ± 1.0 h, respectively, and their corresponding elimination half-lives were reported as 8 and 10.1 hours, respectively, following consumption of 10, 20 or 40 g of soy nuts [56]. A curvilinear relationship was established between bioavailability for daidzein and genistein and the increased amount of soy nuts intake. This suggests a decrease in serum concentrations measured at increasing dosage levels. The same study revealed

that most of the excreted urinary isoflavones were eliminated within the first two days following consumption of soy nuts at different doses. However, the urinary daidzein excretion decreased from approximately 63 to 44%, while the urinary genistein excretion decreased from 25 to 15 % upon increased amount of soy nuts intake from 10 to 40 g. This non-linear pharmacokinetic behaviour over a dose range reveals that optimum steady state isoflavone concentrations can be achieved by multiple intake of soy foods at regular intervals of time than a single high dosage soy product [56]. Similarly, a randomized two-phase crossover study reported peak plasma concentrations of daidzein and genistein on average attained at 6.08 h and 6.37 h, respectively, and their corresponding half-lives were 7.17 h and 7.7 h, respectively [57]. Typically in all of the studies, urinary recovery of genistein and daidzein is complete within 24-36 h [58].

Chronic ingestion of soy isoflavones

A study in post-menopausal women who chronically ingested the commercial soy based preparation named “Prevastein” for 30 and 60 days demonstrated that the urinary and plasma concentrations of genistein, daidzein and equol remained constant from day 15 until the end of experimental period [59]. The data suggest that chronic ingestion could lead to a saturation point and optimum steady state biofluid concentrations can be achieved consistently with adequate intake of soy foods or supplements. However, from a biomarker view point, this may indicate a limitation for estimation of high intakes [56, 59]. An intervention study in young girls who consumed either one daily serving of soymilk (8.5 oz) or soy nuts (1 oz) for 8 weeks period demonstrated that urinary excretion of soy isoflavones increased by almost 6-fold from baseline (23.3 to 142 nmol/mg creatinine). This finding was also reported to be consistent with the 3-day food record which showed a significant increase in isoflavone intake (5.4 to 32.6 mg/day) during the intervention period [60].

While there are many studies focused on concentration levels of isoflavones present in the biofluids following consumption of a range of soy foods, these studies have limited number of subjects and some of the studies lack repeated collection of biofluids. A large randomized, double-blind soy intervention trial with 350 postmenopausal women for 3 years established high correlations between isoflavone measurements of overnight urine, spot urine and plasma with Pearson correlations ranging between 0.60 and 0.94 [48]. All three matrices showed significantly high isoflavone quantitative differences of upto 3-19 fold between placebo and soy group and also highly significant

correlations between mean isoflavone values and soy doses, but not in the placebo group. In another two randomized soy trials conducted among 256 premenopausal women consuming high (~50 mg isoflavones/day) and low (~10 mg isoflavones/day) soy diets, urinary isoflavonoid excretion significantly correlated to dietary isoflavone intake ($r=0.51$, AUC=0.85; $p<0.0001$) [61]. Overall, these studies provide further support for the use of isoflavones as biomarkers of dietary soy intake.

Soybeans are consumed mainly as processed soy products such as tofu, milk, nuts, protein isolate powder etc. The influence of soy food matrix and the effect of industrial processing has resulted in varied isoflavone contents in soy based products [62, 63]. The varied degree of processing conditions has also influenced the metabolism, pharmacokinetics and bioavailability of soy isoflavones [43, 52, 64, 65]. However, our review concentrated on covering information on potential biomarkers obtained from soy and soy based food products rather than trying to understand the influence of above mentioned conditions on the bioavailability of biomarkers.

Soy isoflavone metabolites

A number of soy isoflavone metabolites are found in the circulation. The following section will highlight the key isoflavone metabolites found in the literature review.

Equol is the major isoflavonoid estrogen metabolite produced from daidzein by gut microflora and is produced by ~30-40% of individuals after a soy challenge (named “equol producers”) [58]. A chronic soy exposure study demonstrated that the urinary recovery of equol increased by 3-100 fold ($p<0.05$) over 4 weeks of daily soy ingestion [66]. Similarly, a study examining the prevalence of equol excretion in both male and female individuals revealed that 35% of the participants among the 60 were found to excrete equol following soy protein beverage consumption after 3 days [67]. However, the common isoflavonoids excreted after ingestion of soy based foods such as daidzein, genistein and O-DMA was similar between equol excretors and non-excretors in both men and women [67]. A similar study demonstrated an increase in equol production in older women, while the total excretion of isoflavonoids remained the same after a standardized dose of soy milk among three generations of American-Japanese women [68]. The differences in excretion could in part be attributed to differential gut floral composition with age and differential habitual dietary compositions [39, 69, 70]. The metabolism of equol from daidzein was investigated and demonstrated that daidzein excretion is significantly lower in the equol producers (~ 65 µmol) compared with equol non-producers (~85 µmol) over the entire elimination period

of the soy isoflavones and the difference disappeared when equol excretion was added to daidzein excretion in equol producers suggesting the role of daidzein in equol production [71]. With the same experimental group used [48], equol production changes were studied over a period of three years and demonstrated a high intraindividual variability of upto 35 % with respect to changes in the equol production status [72]. In contrast, equol production was found to be relatively stable over time [73, 74]. Such differences in equol production could be due to dietary factors such as minor differences in intake of micronutrients [75], but further research needs to be done to consider other factors responsible for variation in equol production.

O-DMA is another isoflavonoid estrogen metabolite formed when daidzein is metabolized to dihydrodaidzein by intestinal bacteria in the large intestine and further undergoes ring cleavage [76]. Urinary excretion of *O*-DMA was reported to be generally higher when subjects consumed soy based foods [38, 40, 44, 77, 78]. A randomized crossover study also demonstrated that *O*-DMA appears in plasma after ~6-8 h post-consumption of daidzein rich soy isoflavone preparation and also observed almost a two fold increase after ingestion of pure daidzein glucoside compared to the aglycone form. Urinary excretion levels of *O*-DMA were also two times higher following ingestion of the glucoside form compared to aglycone form [51]. In a large soy intervention trial, ~10-fold increase in *O*-DMA levels in plasma, overnight urine and spot urine was observed in the soy group fed with soy beverage powder and soy bars compared to the placebo fed with protein isolates and bars with no isoflavone content [48]. A lower urinary isoflavone excretion value for *O*-DMA was reported especially in the equol excretors probably due to prefered metabolic pathway converting daidzein to equol [79].

Additionally, few studies have reported the urinary dihydrogenistein and dihydrodaidzein (intermediate products of soy isoflavone metabolism) levels post-consumption of soy based foods [38, 80, 81]. Dihydrogenistein and dihydrodaidzein levels were reported ~4-7 fold higher in the soy intervention group compared to the placebo [48]. While these metabolites reveal interesting metabolic information it remains to be determined if they are useful as biomarkers of soy intake.

Lignan phytoestrogens

A number of studies have reported the presence of lignans in biofluids following consumption of soy or soy-based foods. The urinary excretion of lignans enterodiol and enterolactone was reported to be low after consumption of soy rich diet [82] and

fermented and unfermented soy products [44]. The enterolactone levels were found to remain unaffected following a soy challenge [38]. No differences in excretion levels of lignans were found following soy based diet and basal diet consumption [40]. Subsequently, the urinary excretion levels of lignans is high following consumption of other sources of lignans such as cruciferous vegetable diets [40] and fruit based diet [29] and hence lignans cannot be considered as markers of soy consumption.

Studies relating pulses intake to metabolites in biofluids

While most of the studies are focused on soybeans, there are very few studies which proposed metabolites related to pulses intake. A randomized controlled crossover human feeding study involving 46 middle aged men following consumption of a high dry bean enriched diet (250 g/day) for 4 weeks led to elevated serum levels of pipecolic acid, *S*-methyl cysteine, *N*-acetylornithine, trigonelline and indole propionate [83]. Based on a further study in which participants self-reported their dry bean intake, only pipecolic acid and *S*-methyl cysteine reflected dry bean consumption. Therefore, these two metabolites were proposed as useful markers of dry bean consumption [83]. In another study, maximum peak urinary excretion of kaempferol was observed after 2-8 h following consumption of cooked beans (*Phaseolus vulgaris* L.) [84]. The average excretion was 6.1% and 5.4% of kaempferol dose for males and females respectively. However, although the excretion profiles were similar between subjects, a 6.72-fold inter-individual variation in excretion concentrations was reported, which was ascertained to variations in intestinal physiology [84]. A study examining urinary exposure markers of a wide range of individual foods and food groups revealed that the most probable food exposure marker for green beans was an unsaturated aliphatic hydroxyl-dicarboxylic acid [85]. Notwithstanding the substantial research performed on soy isoflavones, research still needs to be performed in order to identify potential biomarkers of pulses and pulse-based foods in general.

Apparent specificity of the biomarkers

Biomarker validity was further assessed through a previously presented scoring scheme addressing key questions related to chemical/biological plausibility, quantitative aspect, single-dose kinetics, cumulative aspects, robustness, reliability and also analytical validation parameters such as feasibility, analytical performance and reproducibility (Table 2).

The assessment of data presented in table 2 and the secondary search performed revealed that the compounds genistein and daidzein are present in very high concentrations in soybeans and in moderate concentrations in legume based vegetables such as beansprouts, chickpeas, lentils, fava bean, roots of kudzu wine etc. [86, 87]. They are also present in lower concentrations in other fruits and vegetables such as potato, tomato, cabbage, turnip, pumpkin, asparagus etc. [87]. Although there are some reports of excretion of daidzein and genistein and their precursors following consumption of red clover [88], the higher concentrations following soy consumption makes them highly specific markers of soy consumption as evidenced in the published studies. Moreover, genistein and daidzein were observed as having a dose-response relationship with soy intake in various observational studies [26, 27, 32, 34, 89]. Dose-response effects after a single bolus ingestion of three different doses of soy isoflavones was also reported [56]. The time-response relationship explaining the elimination half-life of genistein and daidzein [52, 78] as well as kinetics of repeated intake [59] were explained in the previous section. Both genistein and daidzein are also proven to be stable in urine and plasma at -20^{o} C for almost three months [90] and various quantification methods through LCMS and GCMS platforms for genistein and daidzein have been developed (Table.1). The recovery %, limit of detection, sensitivity and specificity of genistein and daidzein were reported in different analytical methods [91]. A multi-laboratory validation study across seven different laboratories proposed to determine and quantify the total isoflavones content in three soybean varieties showed a satisfactory interlaboratory precision [92]. However, there are no reports of individual isoflavones reported for interlaboratory reproducibility.

Regarding isoflavone metabolites *O*-DMA, dihydrogenistein and dihydrodaidzein, the urinary excretion of these compounds was weakly associated with soy food intake [93] and moreover, these metabolized isoflavonoids are also reported to be present in human urine following red clover supplementation [94]. This suggests the non-specificity of these metabolites after soy intake.

The lignan phytoestrogens enterodiol, enterolactone, matairesinol are widely distributed in many plant classes and their presence is more prevalent in whole grains and fibre containing plant foods (wheat, oats, rye), oilseeds (flax seeds and sesame seeds) and various other fruits and vegetables [58, 82, 95]. Moreover, the urinary excretion levels of lignans is high after consumption of sources other than soy based foods suggesting the limited role of soy as a specific source of lignan phytoestrogens.

In parallel, an additional search was conducted for compounds identified following pulses consumption to examine the candidate biomarkers for specificity for pulses. The assessment of this search revealed that kaempferol is present in a wide range of edible plants such as tea, broccoli, cabbage, kale, leek, tomato, strawberries and grapes [96]. Consequently, kaempferol has been detected in biological samples after intake of other foods such as fruits and vegetables [97, 98]; onions [99, 100]; tea [99, 101-104]; other phenol-rich foods [105], and other food sources [106-108]. On the other hand, trigonelline has also been proposed as a biomarker of coffee with high concentrations reported after coffee intake [109-111]. Regarding pipecolic acid, it can be found in the urine or serum after consumption of black soybean peptide [112] and whole grain-enriched diet [113]; and indole propionate was reported as marker of red meat and eggs intake [114]. On the other hand, no relevant papers were found to evaluate the specificity of *S*-methyl cysteine and *N*-acetyl-ornithine. Based on the above information, the compounds detected in the biofluids after consumption of beans are also present in other food sources and hence none of these compounds could be considered as specific biomarkers of beans intake when evaluated alone.

Conclusion

Although many compounds have been suggested as biomarkers for soy and pulses and for legumes in general, the validation of these compounds against other markers for the same food/food group needs to be performed. Genistein and daidzein could potentially be considered as specific markers of soy and there are reports of dose-response relationship with the amount of soy ingested. Further discovery as well as validation studies are needed in this area and to identify reliable biomarkers of pulses intake as there is a current dearth of information.

Acknowledgments

FoodBAll is a project funded by the BioNH call (grant number 529051002) under the Joint Programming Initiative, “A Healthy Diet for a Healthy Life”. The project is funded nationally by the respective Research Councils: the work has been funded in part by a grant from Science Foundation Ireland (SFI 14/JPI-HDHL/B3076 & SFI 11/PI/1119) to laboratory of LB; and to laboratory of CAL by a grant from the Spanish National Grants from the Ministry of Economy and Competitiveness (MINECO) (PCIN-2014-133-MINECO, Spain), an award from the Generalitat de Catalunya’s Agency AGAUR

(2014SGR1566), and by CIBERFES (co-funded by the FEDER Program from EU). MGA received financial support from the COST Action FA1403 “POSITIVe” to conduct a short-term scientific mission at UCD (COST-STSM-ECOSTSTSM-FA1403-010715-061235).

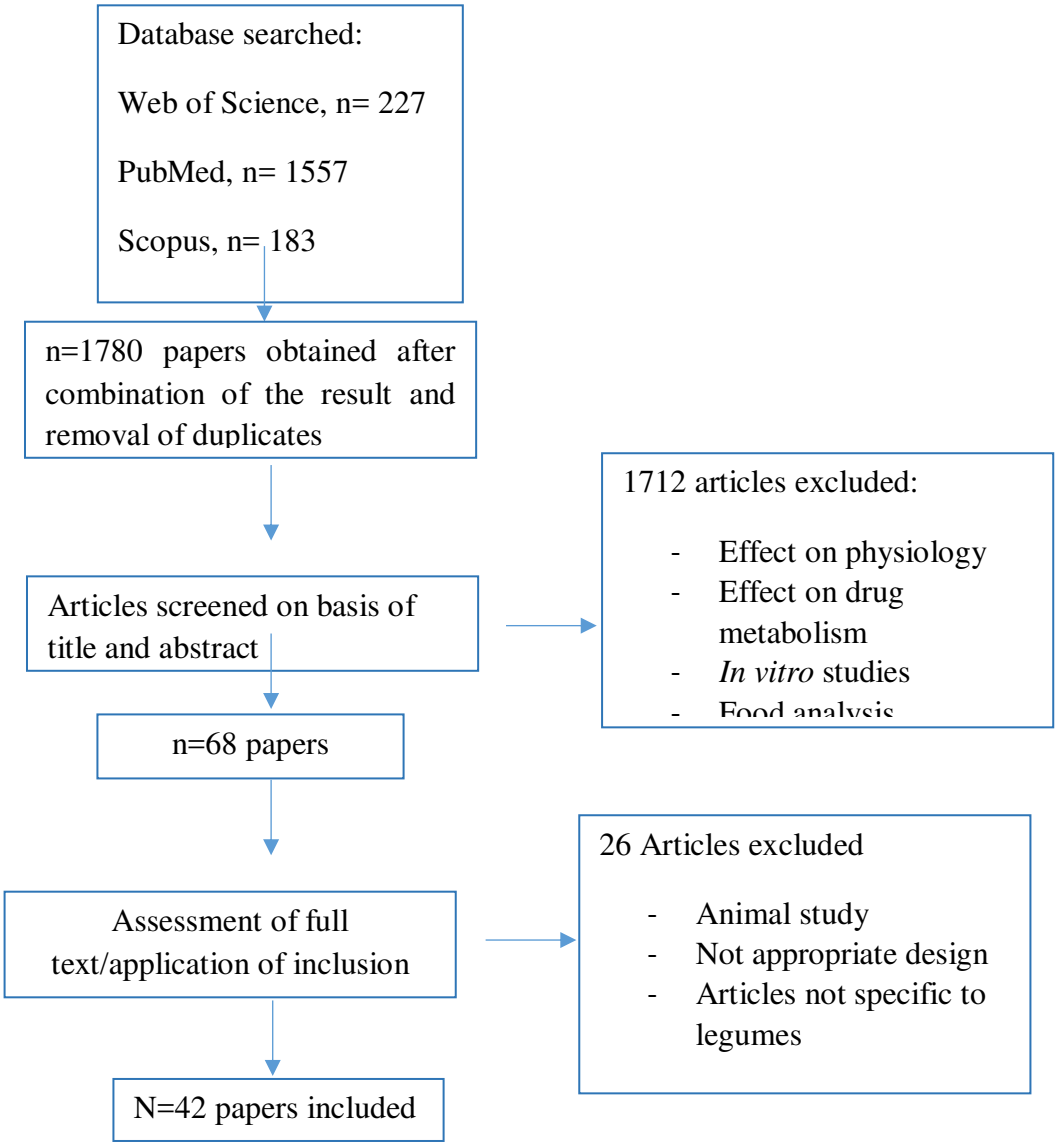


Fig. 1 Flow diagram of study selection

Table 1 List of reported putative legume biomarkers of intake

Dietary factor	Subject	Study design	Number of subjects	Analytical method	Sample type	Discriminating metabolites / Candidate biomarkers	Primary Reference(s)
Soy beans / tofu	Human (M/F)	Cross-sectional study	98	GC-MS	Urine (24 h)	Genistein Daidzein O-Desmethylangolensin	Lampe et al., 1999
Soy protein	Human (M/F)	Intervention study	14	GC-MS	Urine (24 h)	Genistein Daidzein O-Desmethylangolensin Equol	Karr et al., 1997
Soy based foods	Human (F)	Cross-sectional study	102	HPLC	Urine (24 h)	Genistein Daidzein O-Desmethylangolensin Glycitein	Maskarinec et al., 1998
Soy based foods	Human (F)	Cross-sectional study	60	HPLC	Urine	Genistein Daidzein O-Desmethylangolensin Glycitein	Chen et al., 1999
Soy based foods	Human (M/F)	Cross-sectional study	147	HPLC	Urine (spot)	Genistein Daidzein O-Desmethylangolensin Glycitein Equol	Seow et al., 1998
Soy protein	Human (M/F)	Cross-sectional study	100	HPLC-MS	Urine (overnight)	Genistein Daidzein Equol	Jaceldo-Siegl et al., 2008

Soy based foods	Human (F)	Intervention study	18	HPLC-MS	Urine	Genistein Daidzein O-Desmethylangolensin Glycitein Equol	Franke et al., 2006
Soy based foods	Human (M/F)	Cross-sectional study	77	LC-MS	Plasma (fasting)	Genistein Daidzein	Frankenfeld et al., 2002
Soy based foods	Human (F)	Cross-sectional study	80	Flouroimmuno assay	Plasma	Genistein Daidzein	Verkasalo et al., 2001
Soy based foods	Human (F)	Cross sectional study	27/451	GC-MS	Urine (24 h, overnight)	Genistein Daidzein	Tseng et al., 2008
Soy based foods	Human (F)	Cross-sectional study	1823	LC-coularray, LC-MS	Serum (non-fasting)	Genistein Daidzein	Frankenfeld et al., 2004
Soy based foods	Human (F)	Intervention study	17	HPLC	Urine (spot)	Genistein, Daidzein, O-Desmethylangolensin, Equol, Glycitein, Dihydrogenistein, and Dihydrodaidzein	Maskarenic et al., 2005
Soy flour-based meal	Human (M)	Acute study	6	HPLC	Plasma (0-35 h after intake) Urine (0-35h after intake)	Genistein Daidzein	King et al., 1998
Soy extract capsule / Soy beverage	Human (F)	Acute study	12	HPLC	Plasma (0-32 h after intake)	Genistein Daidzein	Anupongsanugool et al., 2005

Soy extract capsule	Human (F)	Acute study	6	HPLC	Plasma (0-24 h after intake)	Genistein	Joshi et al., 2007
Soymilk powder	Human (F)	Acute study	12	HPLC	Plasma (0-24 h after intake) Urine (0-24 h after intake)	Genistein Daidzein	Xu et al., 1994
Soy protein powder	Human (M)	Intervention study (28d)	20	GC-MS	Plasma	Genistein Daidzein	Gooderham et al., 1996
Soy bean powder (Kinako)	Human (M)	Acute study	7	GC-MS	Plasma (0-72 h after intake) Urine (0-24 h) Feces (24-72h)	Genistein Daidzein O-desmethylanglolensin Equol	Watanabe et al., 1998
Soy nuts	Human (F)	Acute study	10	HPLC-MS GC-MS	Serum (0-48 h) Urine (0-12 h for 5d)	Genistein Daidzein Equol	Setchell et al., 2003
Soymilk-based beverages	Human (M/F)	Acute study	12	LC-MS	Serum (0-24 h) Urine (0-48 h)	Genistein Daidzein Glycitein	Kano et al., 2006
Soy bean products	Human (M)	Intervention study (9d)	17	GC-MS	Urine (24 h)	Genistein Daidzein O-desmethylanglolensin Equol Enterodiol Enterolactone Matairesinol	Hutchins et al., 1995

Soy isoflavones	Human (F) Human (M)	Intervention study (30d) Intervention study (60d)	27 12	ELISA	Urine (24 h) Plasma (fasting)	Genistein Daidzein Equol	Mathey et al., 2006
Soy capsules (Phytosoya)	Human (M)	Acute study	12	ELISA	Plasma (0-48 h) Urine (0-24h)	Genistein Daidzein Equol	Vergne et al., 2007
Soy based foods	Human (M/F)	Intervention study (9d)	20	GC-MS	Urine (24 h)	Genistein Daidzein O-desmethylanglolensin Equol Enterodiol Enterolactone	Kirkman et al., 1995
Soy based foods and supplements	Human (F)	Cross-sectional study	96	LC-MS	Plasma (fasting)	Genistein Daidzein	Frankenfelfd et al., 2003
Soy bean extracts	Human (M/F) Human (M)	Acute study Intervention study (4wk)	16 8	HPLC HPLC	Plasma (0-24h after intake) Plasma (non-fasting)	Genistein Daidzein Genistein Daidzein	Izumi et al., 2000
Soy based foods	Human (F)	Intervention study	43	LC-MS	Urine (overnight)	Genistein Daidzein Equol	Maskarenic et al. 2007

Soy based foods	Human (F)	Intervention study	350	LC-MS	Urine (Overnight, spot) Plasma	Genistein Daidzein O-Desmethylangolensin Equol Glycitein Dihydrogenistein Dihydrodaidzein	Franke et al., 2010
Soy flour	Human (M/F)	Acute study	12	GC-MS	Urine (0-24 h on 0d, 3d, 4d 5d)	Genistein Daidzein O-Desmethylangolensin Equol Glycitein Dihydrodaidzein Enterolactone	Kelly et al., 1993
Soy protein beverage	Human (M/F)	Acute study	60	GC-MS	Urine (0-24 h on 4d)	Equol Daidzen Genistein O-Desmethylangolensin	Lampe et al., 1998
Soy milk, Miso soup	Human (F)	Intervention study	21	LC-MS	Urine (overnight, spot) for 6 days	Daidzein Genistein Equol	Maskarinec 2008
Soy milk	Human (F)	Acute study	6	GC-FID	Urine (1-4d,16-18d and 30-32d)	Daidzein Genistein Equol	Lu et al., 1996
Soy based foods	Human (F)	Intervention study	350	LC-MS	Urine (Overnight, spot) Plasma	Equol Daidzein	Franke et al., 2012
Soy based	Human	Intervention	256	HPLC	Urine	Total isoflavone	Morimoto et al., 2014

diet	(F)	study		LC-MS/MS	(Overnight)	excretion	
Soy milk	Human (F)		159		Urine (24h)	Equol	Setchell et al., 2013
Commercial Soy isoflavone supplements	Human (F)	Acute study	19	GC-MS	Plasma (0-48 h)	Genistein Daidzein Genistin Daidzin	Setchell et al., 2001
Commercial soy preparation	Human (M)	Intervention study	7	GC-MS	Urine (0-24 h) Plasma (0-48 h)	Daidzein Daidzein glucoside Dihydrodaidzein O-desmethylangolensin (O-DMA) Equol	Rufer et al., 2008
Soy based foods	Human (F)	Cross section	363	GC-MS	Urine (2 Overnight, 48h apart)	Genistein Daidzein O-Desmethylangolensin Equol	Atkinson et al., 2002
Soy based foods	Human (F)	Case Control Study	97 (Cases), 97 (Control)	Isotope Dilution Electrospray Tandem Mass Spectrometry	Plasma (randomly timed)	Genistein Daidzein Equol Dihydrogenistein Dihydrodaidzein	Wu et al., 2004
Dry beans	Human (M) Human (M/F)	Intervention study Observational study	46 106	LC-MS; GC-MS LC-MS; GC-MS	Serum (fasting) Serum (fasting)	Pipecolic acid *S*-Methyl cysteine *N*-Acetylornithine Trigonelline Indole propionate Pipecolic acid *S*-Methyl cysteine *N*-Acetylornithine	T. Perera et al. 2015

	Mouse (M)	Intervention study	12	LC-MS; GC-MS	Serum, Fecal	Trigonelline Indole propionate Pipecolic acid *S*-Methyl cysteine *N*-Acetylornithine (only in serum)	
Green beans	Human (M/F)	Intervention study	181	UPLC-qTOF-MS	Urine (24h)	Unsaturated aliphatic hydroxyl-dicarboxylic acid	Andersen et al., 2014
Beans	Human (M/F)	Acute study	7	HPLC	Urine (0-24 h after intake)	Kaempferol	Bonetti et al., 2007

Table 2: Possible scoring scheme for legume intake biomarker validity

Food item	Metabolites	Biofluid	Questions								
			1	2	3	4	5	6	7	8	9
Soy	Genistein	Urine	Y	Y	Y	Y	N	N	Y	Y	U
		Plasma/serum	Y	Y	Y	Y	N	N	Y	Y	U
	Dihydrogenistein	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	U	U	N	N	U	Y	U
	Dihydrodaidzein	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	Y	U	N	N	U	Y	U
	Enterodiol	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	U	U	N	N	U	Y	U
	Enterolactone	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	U	U	N	N	U	Y	U
	Matairesinol	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	U	U	N	N	U	Y	U
	Daidzein	Urine	Y	Y	Y	Y	N	N	Y	Y	U
		Plasma/serum	Y	Y	Y	Y	N	N	Y	Y	U
	Equol	Urine	N	U	Y	U	N	N	U	Y	U
		Plasma/serum	N	U	Y	U	N	N	U	Y	U
	O-desmethylangolensin	Urine	N	U	U	U	N	N	U	Y	U
		Plasma/serum	N	U	Y	U	N	N	U	Y	U
	Glycitein	Urine	Y	U	Y	U	N	N	U	Y	U
		Plasma/serum	Y	U	Y	U	N	N	U	Y	U
Dry beans	Kaempferol	Urine	N	U	Y	U	U	U	U	Y	U
	Trigonelline	Plasma/serum	N	U	U	U	N	U	U	N	U
	Pipecolic acid	Plasma/serum	N	U	U	U	Y	U	U	N	U
	Indolepropionate	Plasma/serum	U	U	U	U	N	U	U	N	U
	S-Methylcysteine	Plasma/serum	N	U	U	U	Y	U	U	N	U
	N-Acetyl-ornithine	Plasma/serum	U	U	U	U	N	U	U	N	U

*Possible answers are Y (yes), N (No), or U (unknown or uncertain)

Biological/nutritional validation and applicability

1. Is the marker compound known as unique for the food or food group?
2. Is there a dose-response relationship at relevant intake levels of the targeted food?
3. Is the single-meal time-response relationship described adequately?
4. Is the biomarker kinetics for repeated intakes of the food/food group adequate?
5. Has the marker been shown to be robust after intake of complex meals?
6. Has the marker been shown to compare well with other markers for the same food/food group?

Analytical validation

7. Is the marker chemically and biologically stable, making measurement reliable?
8. Are analytical variability (CV%), accuracy, sensitivity and specificity known as adequate?
9. Has the analysis been successfully reproduced in another laboratory?

References

1. Polak R, Phillips EM, Campbell A. Legumes: Health Benefits and Culinary Approaches to Increase Intake. Clin Diabetes. 2015; 33(4):198-205.
2. Bazzano LA, Thompson AM, Tees MT, Nguyen CH, Winham DM. Non-soy legume consumption lowers cholesterol levels: a meta-analysis of randomized controlled trials. Nutr Metab Cardiovasc Dis. 2011; 21(2):94-103.
3. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet. 2014; 383(9933):1999-2007.
4. Messina MJ. Legumes and soybeans: overview of their nutritional profiles and health effects. Am J Clin Nutr. 1999; 70(3):439s-450s.
5. Heber D. Plant Foods and Phytochemicals in human health: CRC Press; 2008.
6. Adlercreutz CH, Goldin BR, Gorbach SL, Hockerstedt KA, Watanabe S, Hamalainen EK, Markkanen MH, Makela TH, Wahala KT, Adlercreutz T. Soybean phytoestrogen intake and cancer risk. J Nutr. 1995; 125(3 Suppl):757S-770S.
7. Setchell KD, Borriello SP, Hulme P, Kirk DN, Axelson M. Nonsteroidal estrogens of dietary origin: possible roles in hormone-dependent disease. Am J Clin Nutr. 1984; 40(3):569-578.
8. Dong X. Anticancer effects of daidzein, genistein and soy extracts on human prostate cancer cells. University of Delaware; 2011.
9. Birru RL, Ahuja V, Vishnu A, Evans RW, Miyamoto Y, Miura K, Usui T, Sekikawa A. The impact of equol-producing status in modifying the effect of soya isoflavones on risk factors for CHD: a systematic review of randomised controlled trials. J Nutr Sci. 2016; 5:e30.
10. Goldwyn S, Lazinsky A, Wei H. Promotion of health by soy isoflavones: efficacy, benefit and safety concerns. Drug Metabol Drug Interact. 2000; 17(1-4):261-289.
11. Adlercreutz, H. Epidemiology of phytoestrogens. Baillieres Clin Endocrinol Metab. 1998; 12(4):605-623.
12. Wiseman H. The therapeutic potential of phytoestrogens. Expert Opin Investig Drugs. 2000; 9(8):1829-1840.
13. Xu X, Wang HJ, Murphy PA, Cook L, Hendrich S. Daidzein Is a More Bioavailable Soymilk Isoflavone Than Is Genistein in Adult Women. J Nutr. 1994; 124(6):825-832.
14. Jayalath VH, de Souza RJ, Sievenpiper JL, Ha V, Chiavaroli L, Mirrahimi A, Di Buono M, Bernstein AM, Leiter LA, Kris-Etherton PM et al. Effect of dietary pulses on blood pressure: a systematic review and meta-analysis of controlled feeding trials. Am J Hypertens. 2014; 27(1):56-64.
15. Kim SJ, de Souza RJ, Choo VL, Ha V, Cozma AI, Chiavaroli L, Mirrahimi A, Blanco Mejia S, Di Buono M, Bernstein AM et al. Effects of dietary pulse consumption on body weight: a systematic review and meta-analysis of randomized controlled trials. Am J Clin Nutr. 2016; 103(5):1213-1223.
16. Li SS, Kendall CW, de Souza RJ, Jayalath VH, Cozma AI, Ha V, Mirrahimi A, Chiavaroli L, Augustin LS, Blanco Mejia S et al. Dietary pulses, satiety and food

intake: a systematic review and meta-analysis of acute feeding trials. Obesity. 2014; 22(8):1773-1780.

17. Nothlings U, Schulze MB, Weikert C, Boeing H, van der Schouw YT, Bamia C, Benetou V, Lagiou P, Krogh V, Beulens JW et al. Intake of vegetables, legumes, and fruit, and risk for all-cause, cardiovascular, and cancer mortality in a European diabetic population. J Nutr. 2008; 138(4):775-781.

18. Sievenpiper JL, Kendall CW, Esfahani A, Wong JM, Carleton AJ, Jiang HY, Bazinet RP, Vidgen E, Jenkins DJ. Effect of non-oil-seed pulses on glycaemic control: a systematic review and meta-analysis of randomised controlled experimental trials in people with and without diabetes. Diabetologia. 2009; 52(8):1479-1495.

19. Bingham SA. Biomarkers in nutritional epidemiology. Public Health Nutr. 2002; 5(6a):821-827.

20. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, Ferrari P, Carroll RJ. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr. 2002; 5(6a):915-923.

21. Kuhnle GG. Nutritional biomarkers for objective dietary assessment. J Sci Food Agric. 2012; 92(6):1145-1149.

22. O'Gorman A, Gibbons H, Brennan L. Metabolomics in the identification of biomarkers of dietary intake. Comput Struct Biotechnol J. 2013; 4:e201301004.

23. Tucker KL, Smith CE, Lai CQ, Ordovas JM. Quantifying diet for nutrigenomic studies. Annu Rev Nutr. 2013; 33:349-371.

24. Crews H, Alink G, Andersen R, Braesco V, Holst B, Maiani G, Ovesen L, Scotter M, Solfrizzo M, van den Berg R et al. A critical assessment of some biomarker approaches linked with dietary intake. Br J Nutr. 2001; 86 Suppl 1:S5-35.

25. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009, 6(7):e1000097.

26. Seow A, Shi CY, Franke AA, Hankin JH, Lee HP, Yu MC. Isoflavonoid levels in spot urine are associated with frequency of dietary soy intake in a population-based sample of middle-aged and older Chinese in Singapore. Cancer Epidemiol Biomarkers Prev. 1998; 7(2):135-140.

27. Chen Z, Zheng W, Custer LJ, Dai Q, Shu XO, Jin F, Franke AA. Usual dietary consumption of soy foods and its correlation with the excretion rate of isoflavonoids in overnight urine samples among Chinese women in Shanghai. Nutr Cancer. 1999; 33(1):82-87.

28. Jaceldo-Siegl K, Fraser GE, Chan J, Franke A, Sabate J. Validation of soy protein estimates from a food-frequency questionnaire with repeated 24-h recalls and isoflavonoid excretion in overnight urine in a Western population with a wide range of soy intakes. Am J Clin Nutr. 2008; 87(5):1422-1427.

29. Lampe JW, Gustafson DR, Hutchins AM, Martini MC, Li S, Wahala K, Grandits GA, Potter JD, Slavin JL. Urinary isoflavonoid and lignan excretion on a Western diet: Relation to soy, vegetable, and fruit intake. Cancer Epidem Biomar. 1999; 8(8):699-707.

30. Atkinson C, Skor HE, Fitzgibbons ED, Scholes D, Chen C, Wahala K, Schwartz SM, Lampe JW. Overnight urinary isoflavone excretion in a population of women living in the United States, and its relationship to isoflavone intake. Cancer Epidem Biomar. 2002; 11(3):253-260.

31. Tseng M, Olufade T, Kurzer MS, Wahala K, Fang CY, van der Schouw YT, Daly MB. Food frequency questionnaires and overnight urines are valid indicators of daidzein and genistein intake in U.S. women relative to multiple 24-h urine samples. Nutr Cancer. 2008; 60(5):619-626.

32. Maskarinec G, Singh S, Meng LX, Franke AA: Dietary soy intake and urinary isoflavone excretion among women from a multiethnic population. Cancer Epidem Biomar. 1998; 7(7):613-619.

33. Verkasalo PK, Appleby PN, Allen NE, Davey G, Adlercreutz H, Key TJ. Soya intake and plasma concentrations of daidzein and genistein: validity of dietary assessment among eighty British women (Oxford arm of the European Prospective Investigation into Cancer and Nutrition). Brit J Nutr. 2001; 86(3):415-421.

34. Frankenfeld CL, Patterson RE, Kalhorn TF, Skor HE, Howald WN, Lampe JW. Validation of a soy food frequency questionnaire with plasma concentrations of isoflavones in US adults. J Am Diet Assoc. 2002; 102(10):1407-1413.

35. Frankenfeld CL, Patterson RE, Horner NK, Neuhouser ML, Skor HE, Kalhorn TF, Howald WN, Lampe JW. Validation of a soy food-frequency questionnaire and evaluation of correlates of plasma isoflavone concentrations in postmenopausal women. Am J Clin Nutr. 2003; 77(3):674-680.

36. Wu AH, Yu MC, Tseng CC, Twaddle NC, Doerge DR. Plasma isoflavone levels versus self-reported soy isoflavone levels in Asian-American women in Los Angeles County. Carcinogenesis. 2004; 25(1):77-81.

37. Frankenfeld CL, Lampe JW, Shannon J, Gao DL, Ray RM, Prunty J, Kalhorn TF, Wahala K, Patterson RE, Thomas DB. Frequency of soy food consumption and serum isoflavone concentrations among Chinese women in Shanghai. Public Health Nutr. 2004; 7(6):765-772.

38. Kelly GE, Nelson C, Waring MA, Joannou GE, Reeder AY. Metabolites of Dietary (Soya) Isoflavones in Human Urine. Clin Chim Acta. 1993; 223(1-2):9-22.

39. Rowland IR, Wiseman H, Sanders TAB, Adlercreutz H, Bowey EA. Interindividual variation in metabolism of soy isoflavones and lignans: Influence of habitual diet on equol production by the gut microflora. Nutr Cancer. 2000; 36(1):27-32.

40. Kirkman LM, Lampe JW, Campbell DR, Martini MC, Slavin JL. Urinary Lignan and Isoflavonoid Excretion in Men and Women Consuming Vegetable and Soy Diets. Nutr Cancer. 1995; 24(1):1-12.

41. Karr SC, Lampe JW, Hutchins AM, Slavin JL. Urinary isoflavonoid excretion in humans is dose dependent at low to moderate levels of soy-protein consumption. Am J Clin Nutr. 1997; 66(1):46-51.

42. Gooderham MJ, Adlercreutz H, Ojala ST, Wahala K, Holub BJ. A soy protein isolate rich in genistein and daidzein and its effects on plasma isoflavone concentrations, platelet aggregation, blood lipids and fatty acid composition of plasma phospholipid in normal men. J Nutr. 1996; 126(8):2000-2006.

43. Cassidy A, Bingham S, Setchell KDR. Biological Effects of a Diet of Soy Protein-Rich in Isoflavones on the Menstrual-Cycle of Premenopausal Women. Am J Clin Nutr. 1994; 60(3):333-340.

44. Hutchins AM, Slavin JL, Lampe JW. Urinary Isoflavonoid Phytoestrogen and Lignan Excretion after Consumption of Fermented and Unfermented Soy Products. J Am Diet Assoc. 1995; 95(5):545-551.

45. Klejdus B, Vacek J, Benesova L, Kopecky J, Lapcik O, Kuban V. Rapid-resolution HPLC with spectrometric detection for the determination and identification of isoflavones in soy preparations and plant extracts. Anal Bioanal Chem. 2007; 389(7-8):2277-2285.

46. Kano M, Takayanagi T, Harada K, Sawada S, Ishikawa F. Bioavailability of isoflavones after ingestion of soy beverages in healthy adults. J Nutr. 2006; 136(9):2291-2296.

47. Shinkaruk S, Durand M, Lamothe V, Carpaye A, Martinet A, Chantre P, Vergne S, Nogues X, Moore N, Bennetau-Pelissero C. Bioavailability of glycitein relatively to other soy isoflavones in healthy young Caucasian men. Food Chem. 2012; 135(3):1104-1111.

48. Franke AA, Hebshi SM, Pagano I, Kono N, Mack WJ, Hodis HN. Urine Accurately Reflects Circulating Isoflavonoids and Ascertains Compliance During Soy Intervention. Cancer Epidem Biomar. 2010; 19(7):1775-1783.

49. Maskarinec G, Watts K, Kagihara J, Hebshi SM, Franke AA. Urinary isoflavonoid excretion is similar after consuming soya milk and miso soup in Japanese-American women. Brit J Nutr. 2008; 100(2):424-429.

50. Izumi T, Piskula MK, Osawa S, Obata A, Tobe K, Saito M, Kataoka S, Kubota Y, Kikuchi M. Soy isoflavone aglycones are absorbed faster and in higher amounts than their glucosides in humans. J Nutr. 2000; 130(7):1695-1699.

51. Rufer CE, Bub A, Moseneder J, Winterhalter P, Sturtz M, Kulling SE. Pharmacokinetics of the soybean isoflavone daidzein in its aglycone and glucoside form: a randomized, double-blind, crossover study. Am J Clin Nutr. 2008; 87(5):1314-1323.

52. Setchell KD, Brown NM, Desai P, Zimmer-Nechemias L, Wolfe BE, Brashear WT, Kirschner AS, Cassidy A, Heubi JE. Bioavailability of pure isoflavones in healthy humans and analysis of commercial soy isoflavone supplements. J Nutr. 2001; 131(4 Suppl):1362S-1375S.

53. Richelle M, Pridmore-Merten S, Bodenstab S, Enslen M, Offord EA. Hydrolysis of isoflavone glycosides to aglycones by beta-glycosidase does not alter plasma and urine isoflavone pharmacokinetics in postmenopausal women. J Nutr. 2002; 132(9):2587-2592.

54. Zubik L, Meydani M. Bioavailability of soybean isoflavones from aglycone and glucoside forms in American women. Am J Clin Nutr. 2003; 77(6):1459-1465.

55. King RA, Bursill DB. Plasma and urinary kinetics of the isoflavones daidzein and genistein after a single soy meal in humans. Am J Clin Nutr. 1998; 67(5):867-872.

56. Setchell KDR, Brown NM, Desai PB, Zimmer-Nechimias L, Wolfe B, Jakate AS, Creutzinger V, Heubi JE. Bioavailability, disposition, and dose-response effects

of soy isoflavones when consumed by healthy women at physiologically typical dietary intakes. J Nutr. 2003; 133(4):1027-1035.

57. Anupongsanugool E, Teekachunhatean S, Rojanasthien N, Pongsatha S, Sangdee C.Pharmacokinetics of isoflavones, daidzein and genistein, after ingestion of soy beverage compared with soy extract capsules in postmenopausal Thai women. BMC Clin Pharmacol. 2005; 5:2.

58. Lampe JW. Isoflavonoid and lignan phytoestrogens as dietary biomarkers. J Nutr. 2003; 133(3):956s-964s.

59. Mathey J, Lamothe V, Coxam V, Potier M, Sauvant P, Bennetau-Pelissero C. Concentrations of isoflavones in plasma and urine of post-menopausal women chronically ingesting high quantities of soy isoflavones. J Pharm Biomed Anal. 2006; 41(3):957-965.

60. Maskarinec G, Oshiro C, Morimoto Y, Hebshi S, Novotny R, Franke AA. Urinary isoflavone excretion as a compliance measure in a soy intervention among young girls: a pilot study. Eur J Clin Nutr. 2005; 59(3):369-375.

61. Morimoto Y, Beckford F, Franke AA, Maskarinec G. Urinary isoflavonoid excretion as a biomarker of dietary soy intake during two randomized soy trials. Asia Pac J Clin Nutr. 2014; 23(2):205-209.

62. Genovese MI, Lopes Barbosa AC, Pinto MD, Lajolo FM. Commercial soy protein ingredients as isoflavone sources for functional foods. Plant Food Hum Nutr. 2007; 62(2):53-58.

63. Wang C, Ma Q, Pagadala S, Sherrard MS, Krishnan PG. Changes of isoflavones during processing of soy protein isolates. J Am Oil Chem Soc. 1998; 75(3):337-341.

64. Allred CD, Twaddle NC, Allred KF, Goeppinger TS, Churchwell MI, Ju YH, Helferich WG, Doerge DR. Soy processing affects metabolism and disposition of dietary isoflavones in ovariectomized balb/c mice. J Agr Food Chem. 2005; 53(22):8542-8550.

65. Andrade JE, Twaddle NC, Helferich WG, Doerge DR. Absolute Bioavailability of Isoflavones from Soy Protein Isolate-Containing Food in Female Balb/c Mice. J Agr Food Chem. 2010; 58(7):4529-4536.

66. Lu LJW, Lin SN, Grady JJ, Nagamani M, Anderson KE. Altered kinetics and extent of urinary daidzein and genistein excretion in women during chronic soya exposure. Nutr Cancer. 1996; 26(3):289-302.

67. Lampe JW, Karr SC, Hutchins AM, Slavin JL. Urinary equol excretion with a soy challenge: Influence of habitual diet. P Soc Exp Biol Med. 199;, 217(3):335-339.

68. Maskarinec G, Yamakawa R, Hebshi S, Franke AA. Urinary isoflavonoid excretion and soy consumption in three generations of Japanese women in Hawaii. Eur J Clin Nutr. 2007; 61(2):255-261.

69. Saunier K, Dore J. Gastrointestinal tract and the elderly: functional foods, gut microflora and healthy ageing. Digest Liver Dis. 2002; 34:S19-S24.

70. Van Tongeren SP, Slaets JPJ, Harmsen HJM, Welling GW. Fecal microbiota composition and frailty. Appl Environ Microb. 2005; 71(10):6438-6442.

71. Vergne S, Titier K, Bernard V, Asselineau J, Durand M, Lamothe V, Potier M, Perez P, Demotes-Mainard J, Chantre P et al. Bioavailability and urinary

excretion of isoflavones in humans: Effects of soy-based supplements formulation and equol production. J Pharmaceut Biomed. 2007; 43(4):1488-1494.

72. Franke AA, Lai JF, Halm BM, Pagano I, Kono N, Mack WJ, Hodis HN. Equol production changes over time in postmenopausal women. J Nutr Biochem. 2012; 23(6):573-579.

73. Frankenfeld CL, Atkinson C, Thomas WK, Gonzalez A, Jokela T, Wahala K, Schwartz SM, Li SS, Lampe JW. High concordance of daidzein-metabolizing phenotypes in individuals measured 1 to 3 years apart. Brit J Nutr. 2005; 94(6):873-876.

74. Nettleton JA, Greany KA, Thomas W, Wangen KE, Adlercreutz H, Kurzer MS. Plasma phytoestrogens are not altered by probiotic consumption in postmenopausal women with and without a history of breast cancer. J Nutr. 2004; 134(8):1998-2003.

75. Setchell KDR, Brown NM, Summer S, King EC, Heubi JE, Cole S, Guy T, Hokin B. Dietary Factors Influence Production of the Soy Isoflavone Metabolite S-(-)Equol in Healthy Adults. J Nutr. 2013; 143(12):1950-1958.

76. Frankenfeld CL. O-Desmethylangolensin: The Importance of Equol's Lesser Known Cousin to Human Health. Adv Nutr. 2011; 2(4):317-324.

77. Karr SC, Lampe JW, Hutchins AM, Slavin JL. Urinary isoflavonoid excretion in humans is dose dependent at low to moderate levels of soy-protein consumption. Am J Clin Nutr. 1997; 66(1):46-51.

78. Watanabe S, Yamaguchi M, Sobue T, Takahashi T, Miura T, Arai Y, Mazur W, Wahala K, Adlercreutz H. Pharmacokinetics of soybean isoflavones in plasma, urine and feces of men after ingestion of 60 g baked soybean powder (kinako). J Nutr. 1998; 128(10):1710-1715.

79. Franke AA, Morimoto Y, Yeh LM, Maskarinec G. Urinary isoflavonoids as a dietary compliance measure among premenopausal women. Asia Pac J Clin Nutr. 2006; 15(1):88-94.

80. Heinonen S, Wahala K, Adlercreutz H. Identification of isoflavone metabolites dihydrodaidzein, dihydrogenistein, 6 '-OH-O-dma, and cis-4-OH-equol in human urine by gas chromatography-mass spectroscopy using authentic reference compounds. Anal Biochem. 1999; 274(2):211-219.

81. Zhang Y, Hendrich S, Murphy PA. Glucuronides are the main isoflavone metabolites in women. J Nutr. 2003; 133(2):399-404.

82. Adlercreutz H, Honjo H, Higashi A, Fotsis T, Hamalainen E, Hasegawa T, Okada H. Urinary-Excretion of Lignans and Isoflavonoid Phytoestrogens in Japanese Men and Women Consuming a Traditional Japanese Diet. Am J Clin Nutr. 1991; 54(6):1093-1100.

83. Perera T, Young MR, Zhang ZY, Murphy G, Colburn NH, Lanza E, Hartman TJ, Cross AJ, Bobe G. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. Mol Nutr Food Res. 2015; 59(4):795-806.

84. Bonetti A, Marotti I, Dinelli G. Urinary excretion of kaempferol from common beans (Phaseolus vulgaris L.) in humans. Int J Food Sci Nutr. 2007; 58(4):261-269.

85. Andersen MBS, Kristensen M, Manach C, Pujos-Guillot E, Poulsen SK, Larsen TM, Astrup A, Dragsted L. Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. Anal Bioanal Chem. 2014; 406(7):1829-1844.

86. Kaufman PB, Duke JA, Brielmann H, Boik J, Hoyt JE. A comparative survey of leguminous plants as sources of the isoflavones, genistein and daidzein: implications for human nutrition and health. J. Altern. Complement. Med. 1997; 3(1):7-12.

87. Liggins J, Bluck LJ, Runswick S, Atkinson C, Coward WA, Bingham SA. Daidzein and genistein contents of vegetables. Br J Nutr. 2000; 84(5):717-725.

88. Tsunoda N, Pomeroy S, Nestel P. Absorption in humans of isoflavones from soy and red clover is similar. J Nutr. 2002; 132(8):2199-2201.

89. Frankenfeld CL, Patterson RE, Horner NK, Neuhouser ML, Skor HE, Kalhorn TF, Howald WN, Lampe JW. Validation of a soy food-frequency questionnaire and evaluation of correlates of plasma isoflavone concentrations in postmenopausal women. Am J Clin Nutr. 2003; 77(3):674-680.

90. Busby MG, Jeffcoat AR, Bloedon LT, Koch MA, Black T, Dix KJ, Heizer WD, Thomas BF, Hill JM, Crowell JA et al. Clinical characteristics and pharmacokinetics of purified soy isoflavones: single-dose administration to healthy men. Am J Clin Nutr. 2002; 75(1):126-136.

91. Wilkinson AP, Wahala K, Williamson G. Identification and quantification of polyphenol phytoestrogens in foods and human biological fluids. J Chromatogr B Analyt Technol Biomed Life Sci. 2002; 777(1-2):93-109.

92. Ogita T, Watanabe J, Wakagi M, Nakamichi K, Komiyama S, Takebayashi J, Mano J, Kitta K, Koyano S, Takano-Ishikawa Y. Evaluation of a Method to Quantify Isoflavones in Soybean by Single and Multi-laboratory Validation Studies. Food Sci Technol Res. 2015; 21(3):473-477.

93. Wu X, Cai H, Gao YT, Dai Q, Li H, Cai Q, Yang G, Franke AA, Zheng W, Shu XO. Correlations of urinary phytoestrogen excretion with lifestyle factors and dietary intakes among middle-aged and elderly Chinese women. Int J Mol Epidemiol Genetics. 2012; 3(1):18-29.

94. Lipovac M, Pfitscher A, Hobiger S, Laschitz T, Imhof M, Chedraui P, Jungbauer A. Red clover isoflavone metabolite bioavailability is decreased after fructooligosaccharide supplementation. Fitoterapia. 2015; 105:93-101.

95. Milder IEJ, Arts ICW, van de Putte B, Venema DP, Hollman PCH. Lignan contents of Dutch plant foods: a database including lariciresinol, pinoresinol, secoisolariciresinol and matairesinol. Brit J Nutr. 2005; 93(3):393-402.

96. Calderon-Montano JM, Burgos-Moron E, Perez-Guerrero C, Lopez-Lazaro M. A review on the dietary flavonoid kaempferol. Mini Rev Med Chem. 2011; 11(4):298-344.

97. Brevik A, Rasmussen SE, Drevon CA, Andersen LF. Urinary excretion of flavonoids reflects even small changes in the dietary intake of fruits and vegetables. Cancer Epidemiol Biomarkers Prev. 2004; 13(5):843-849.

98. Mennen LI, Sapinho D, Ito H, Bertrais S, Galan P, Hercberg S, Scalbert A. Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. Brit J Nutr. 2006; 96(1):191-198.

99. de Vries JHM, Hollman PCH, Meyboom S, Buysman MNCP, Zock PL, van Staveren WA, Katan MB. Plasma concentrations and urinary excretion of the antioxidant flavonols quercetin and kaempferol as biomarkers for dietary intake. Am J Clin Nutr. 1998; 68(1):60-65.

100. Hong Y, Mitchell AE. Metabolic profiling of flavonol metabolites in human urine by liquid chromatography and tandem mass spectrometry. J Agr Food Chem. 2004; 52(22):6794-6801.

101. Brantsaeter AL, Haugen M, Rasmussen SE, Alexander J, Samuelsen SO, Meltzer HM. Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). Public Health Nutr. 2007; 10(8):838-847.

102. Hollman PC, Van Het Hof KH, Tijburg LB, Katan MB. Addition of milk does not affect the absorption of flavonols from tea in man. Free radical research. 2001; 34(3):297-300.

103. Krogholm KS, Bysted A, Brantsaeter AL, Jakobsen J, Rasmussen SE, Kristoffersen L, Toft U. Evaluation of flavonoids and enterolactone in overnight urine as intake biomarkers of fruits, vegetables and beverages in the Inter99 cohort study using the method of triads. Brit J Nutr. 2012; 108(10):1904-1912.

104. Nielsen SE, Freese R, Kleemola P, Mutanen M. Flavonoids in human urine as biomarkers for intake of fruits and vegetables. Cancer Epidem Biomar. 2002; 11(5):459-466.

105. Kim HY, Kim OH, Sung MK. Effects of phenol-depleted and phenol-rich diets on blood markers of oxidative stress, and urinary excretion of quercetin and kaempferol in healthy volunteers. J Am Coll Nutr. 2003; 22(3):217-223.

106. DuPont MS, Day AJ, Bennett RN, Mellon FA, Kroon PA. Absorption of kaempferol from endive, a source of kaempferol-3-glucuronide, in humans. Eur J Clin Nutr. 2004; 58(6):947-954.

107. Grinder-Pedersen L, Rasmussen SE, Bugel S, Jorgensen LV, Dragsted LO, Gundersen V, Sandstrom B. Effect of diets based on foods from conventional versus organic production on intake and excretion of flavonoids and markers of antioxidative defense in humans. J Agr Food Chem. 2003; 51(19):5671-5676.

108. Wang FM, Yao TW, Zeng S. Disposition of quercetin and kaempferol in human following an oral administration of Ginkgo biloba extract tablets. Eur J Drug Metab Pharmacokinet. 2003; 28(3):173-177.

109. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q, Huang WY, Xiong XQ, Freedman ND, Cross AJ et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. Am J Clin Nutr. 2015; 101(5):1000-1011.

110. Lang R, Wahl A, Stark T, Hofmann T. Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. Mol Nutr Food Res. 2011; 55(11):1613-1623.

111. Rothwell JA, Fillatre Y, Martin JF, Lyan B, Pujos-Guillot E, Fezeu L, Hercberg S, Comte B, Galan P, Touvier M et al. New Biomarkers of Coffee Consumption Identified by the Non-Targeted Metabolomic Profiling of Cohort Study Subjects. Plos One. 2014; 9(4).

112. Kim MJ, Yang HJ, Kim JH, Ahn CW, Lee JH, Kim KS, Kwon DY. Obesity-related metabolomic analysis of human subjects in black soybean peptide intervention study by ultraperformance liquid chromatography and quadrupole-time-of-flight mass spectrometry. J Obes. 2013; 2013:874981.

113. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, de Mello V, Sola R, Lehtonen M, Poutanen K et al. Nontargeted metabolite profiling discriminates diet-specific biomarkers for consumption of whole grains, fatty fish, and bilberries in a randomized controlled trial. J Nutr. 2015; 145(1):7-17.

114. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr. 2014; 100(1):208-217.

Capítulo 05.

BIOMARCADORES NUTRICIONALES DE CAFÉ

El café es una de las bebidas más consumidas en el mundo (Ozen et al. 2012). En España, el consumo medio por persona y año de cafés e infusiones asciende a 1,73 kg/persona/año suponiendo un gasto anual per cápita de 24,26 € (2015) (Ministerio de Agricultura Medioambiente y Alimentación 2015). Consecuentemente, existe un gran interés sobre los efectos del consumo de esta bebida en la salud, por lo que en la bibliografía científica existe un importante número de estudios que han evaluado los beneficios del consumo de café (Guertin et al. 2015; Pan et al. 2016; Loader et al. 2017).

Los granos de café contienen una gran variedad de compuestos biológicamente activos, siendo los ácidos clorogénicos (CGA) y la cafeína los predominantes (Johnston et al. 2003). En relación con esto, diversas investigaciones epidemiológicas sugieren que el consumo moderado de café puede ayudar a prevenir enfermedades crónicas no transmisibles, como la enfermedad cardiovascular, la hipertensión, la obesidad, la diabetes mellitus tipo 2, el síndrome metabólico, la enfermedad de Parkinson, y la enfermedad hepática (La Vecchia 2005; Sääksjärvi et al. 2008; Nordestgaard et al. 2015; Chrysant 2017). Debido a la importancia que tiene este alimento sobre su popular consumo y sus efectos sobre la salud, surge la conveniencia de determinar su consumo mediante el uso de biomarcadores de ingesta precisos, así como apoyar sus efectos beneficiosos mediante biomarcadores de efecto, ya provengan de la cafeína contenida en el café, o bien de sus compuestos fitoquímicos.

En el marco del proyecto FOODBALL, cuyos objetivos principales incluyen la revisión de la bibliografía científica de alimentos destacados por su consumo y/o efectos sobre la salud, se presenta la siguiente publicación redactada en forma de revisión bibliográfica sobre los biomarcadores nutricionales de bebidas no alcohólicas, incluyendo el correspondiente apartado de café, en colaboración con la *International Agency for Research on Cancer* (Francia), el *Northern Ireland Centre for Food and Health* de la *Ulster University* (Irlanda), el *Federal Research Institute of Nutrition and*

Food del *Max Rubner-Institut* (Alemania) y el *Department of Nutrition, Exercise and Sports* de la *University of Copenhagen* (Dinamarca).

Este trabajo está en proceso de publicación en una edición especial de la revista *Genes and Nutrition* indexada en el Journal Citation Reports con un factor de impacto de 2.398 (2015) y situada en el tercer cuartil de las categorías *Nutrition & Dietetics* (41/80) (2015) y *Genetics & Heredity* (88/166).

Biomarkers of intake for some non-alcoholic beverages

Joseph A. Rothwell[1], Francisco Madrid-Gambin[2], Mar Garcia-Aloy[2,3], Cristina Andres-Lacueva[2,3] C. Logue[4], A.M. Gallagher[4], Carina Mack[5], Sabine E. Kulling[5], Qian Gao[6], Giulia Praticò[6], Lars O. Dragsted[6], Augustin Scalbert[1]*

[1]International Agency for Research on Cancer (IARC). Nutrition and Metabolism Section, Biomarkers Group ,150 cours Albert Thomas, F-69372 Lyon Cedex 08, France.

[2]Biomarkers and Nutrimetabolomics Laboratory, Department of Nutrition, Food Sciences and Gastronomy, Campus Torribera, Faculty of Pharmacy and Food Sciences, University of Barcelona, Spain.

[3]CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain.

[4]Northern Ireland Centre for Food and Health (NICHE), Biomedical Sciences Research Institute, Ulster University, Cromore Road, Coleraine, Northern Ireland.

[5]Department of Safety and Quality of Fruit and Vegetables. Federal Research Institute of Nutrition and Food. Max Rubner-Institut, Karlsruhe, Germany.

[6]Department of Nutrition, Exercise and Sports. Faculty of Science. University of Copenhagen, Denmark.

*Corresponding author.

Email : scalberta@iarc.fr

Abstract

Non-alcoholic beverages are important sources of nutrients and bioactive compounds that may influence human health and increase or decrease the risk of chronic diseases. A wide variety of beverage constituents are absorbed in the gut, found in the systemic circulation and excreted in urine. They may be used as compliance markers in intervention studies or as biomarkers of intake to improve measurements of beverage consumption in cohort studies and reveal new associations with disease outcomes that may have been overlooked when using dietary questionnaires. Here, biomarkers of intake of some non-alcoholic beverages - coffee, tea, sugar-sweetened beverages and low-calorie sweetened beverages - are reviewed. Results from dietary intervention studies and observational studies are reviewed and analysed and respective strengths and weaknesses of the various identified biomarkers discussed. A variety of compounds derived from phenolic acids, alkaloids and terpenes were shown to be associated with coffee intake and trigonelline and cyclo(isoleucylprolyl) showed a particularly high specificity for coffee intake. Epigallocatechin and 4'-*O*-methylepigallocatechin appear to be the most sensitive and specific biomarkers for green or black tea, while 4-*O*-methylgallic acid may be used to assess black tea consumption. Intake of sugar-sweetened beverages has been assessed through the measurement of carbon-13 enrichment of whole blood or of blood alanine in North America where sugar from sugar cane or corn is used as a main ingredient. The most useful biomarkers for low-calorie sweetened beverages are the low-calorie sweeteners themselves. Further studies are needed to validate these biomarkers in larger and independent populations and to further evaluate their specificity, reproducibility over time, and fields of application.

Keywords: Non-alcoholic beverages, coffee, tea, sugar-sweetened beverages, low-calorie sweetened beverages, biomarkers, intake.

Introduction

Human adults consume about 3 to 4 litres of water per day, which originate from plain water, solid foods, and other beverages like coffee, tea, soda, fruit juices, sport and energy drinks, milk, and alcoholic beverages [1]. Such beverages are not only a source of water but also of other constituents which may influence health in a positive or negative way. Some of these constituents like polyphenols in tea, coffee and wine are thought to contribute to the prevention of chronic diseases such as cardiovascular diseases or diabetes. Other constituents of beverages may have detrimental effects on health. Compounds like sugars in sodas, may provide an excess of energy and increase the risk of obesity and associated diseases.

In order to understand the effects of beverages on disease risk in large cohort studies, accurate measurement of intake is needed. Beverage consumption is most often measured with self-administered questionnaires. However these measurements are known to be prone to random or systematic errors which may prevent the detection of associations of intake with disease outcomes [2]. These errors are linked to limitations in the capacity of subjects to accurately describe their beverage intake, and to difficulties in accounting for variations in the nature of the beverages consumed within or between populations. For example, concentrations of polyphenols in coffee and in tea vary with the type of coffee beans or tea leaves and with the method of coffee brewing or tea leaf infusion, information most often not recorded in dietary questionnaires. In contrast to questionnaires, biomarkers are objective measures that provide estimates of beverage intake [3-5]. They are directly derived from beverage constituents absorbed in the gut during digestion. Thus they provide more direct estimates of exposure to beverage constituents and are notably independent of the dilution of the beverages, or on any aspect of the recipes used for their preparation.

Biomarkers of beverage intake can be identified using two separate approaches. Initially, biomarkers were hypothesized based on knowledge of beverage constituents and their metabolism. Beverages are fed to volunteers in controlled intervention studies and known

metabolites of interest measured in biofluids to test for increases with intake. Until a few years ago, this targeted approach was the only viable method of biomarker discovery. More recently, metabolome profiling has taken precedence as a more comprehensive and agnostic approach. Rather than measuring known compounds only, biofluids are profiled to measure the relative intensities of as many signals as can be detected, either by nuclear magnetic resonance (NMR) or mass spectrometry (MS) coupled to liquid or gas chromatography. Signals associated with intake of the target food across the most possible subjects are retained as candidate biomarkers. Initial studies used an intervention design with a standardized dose and controlled diet, but more sensitive analytical techniques have enabled the analysis of biofluids from free-living subjects who have reported their beverage intakes via dietary questionnaires. This approach encourages the discovery of biomarkers that are valid in the presence of interfering dietary factors and that also account for differences in the time period between beverage intake and biofluid collection. Biomarkers can only be considered valid for the populations and biofluids in which they are discovered as different populations drink different brands or brews of a beverage with different compositions.

The purpose of this review is to describe the various biomarkers proposed to evaluate intake of non-alcoholic beverages, including coffee, tea, sugar-sweetened beverages (SSB) and low-calorie sweetened beverages (LCSB), all beverages that have raised much interest for their effects on the risk of various chronic diseases. Other non-alcoholic beverages such as fruit juices and milk as well as alcoholic beverages will be discussed in other chapters in this special volume. For each beverage, the main metabolites formed from their constituents and identified in blood or urine in human dietary intervention studies as putative biomarkers are first described. Their eventual detection in observational studies and respective value as intake biomarkers is then discussed.

Materials and methods

The reviewing process made was described in details recently [6] and use all elements of the PRISMA statement [7] that were relevant for a literature search on dietary biomarkers. Original research papers and reviews were searched for in PubMed, Scopus and ISI Web of Knowledge in December 2015 using the following search terms: (biomarker* OR marker* OR metabolite* OR biokinetics OR biotransformation) AND (human* OR men OR women OR male OR female OR patient* OR volunteer* OR participant*) AND (urine OR plasma OR serum OR blood OR excretion) AND (intake OR meal OR diet OR ingestion OR consumption OR eating OR drink* OR administration) AND ('name of beverage'). Name of beverages were (coffee), (tea), or (diet soda OR diet beverage* OR soft drink* OR artificial* sweet* beverage* OR low sugar beverage* OR low calorie sweet* beverage*).

The research was limited to papers in English language, and no restrictions were applied to publication date. The research papers identifying or using potential biomarkers of intake for the foods were selected by one or more skilled researchers from the list of retrieved references in the process outlined in Supplemental Figures 1 and 2. Additional papers were identified from reference lists in these papers and from reviews or book chapters identified through the search. For each potential biomarker identified an additional search was conducted with ("the name and synonyms of the compound" OR "the name and synonyms of any parent compound") AND (biomarker* OR marker* OR metabolite* OR biokinetics OR biotransformation) to identify potential other foods containing the biomarker or its precursor. In this second step, Scifinder and Google Scholar were also used as search platforms, as well as the databases listed above. This second search was used to evaluate the apparent specificity of the marker.

Results and discussion

Biomarkers have been systematically searched in the scientific literature independently for the four types of non-alcoholic beverages (coffee, tea, SSBs and LCSBs). The literature was particularly abundant for biomarkers of coffee and tea intake (Supplemental Figures 1 and 2). Results are successively presented below.

- **Coffee biomarkers**

Coffee is one of the most widely consumed beverages in the world. It is consumed widely in Europe and North America, and although coffee is not traditionally drunk in the Far East, the beverage is increasingly consumed in countries such as China and Japan. Per capita consumption is highest in Northern Europe, particularly Scandinavia [8]. Due to the high content of bioactive phytochemicals in coffee, much attention has been paid to the effects of regular consumption on human health and disease. Although coffee drinking was initially thought to induce negative effects on health, for example increasing blood pressure and the risk of cardiovascular disease risk in some cohorts [9, 10], the main interest in recent years is the potential for positive health effects. These include plausible reductions in risk of type 2 diabetes, Parkinson disease, Alzheimer's disease and liver and colorectal cancer [11-16].

The health effects of coffee are thought to be mediated by its distinct families of phytochemicals [8]. The alkaloid caffeine is the most well-known and abundant phytochemical constituent of coffee, but the beverage is also one of the greatest contributors to phenolic acid intake due to its high concentrations of chlorogenic acid (5-caffeoylquinic acid) and other hydroxycinnamate derivatives [17]. Coffee is also a major source of diterpenes, particularly cafestol, kahweol and atractyloside, which contribute towards the bitter flavour. Proline-, leucine- and valine-containing diketopiperazines (cyclic amino acid dimers) are also found in coffee, as well as short peptides, free amino acids, quaternary ammonium compounds, and a wide variety of small aromatic Maillard and Strecker products. Owing to this wide variety of specific phytochemicals,

many at high concentrations, a number of coffee compounds have been proposed as biomarkers of coffee intake.

Coffee metabolites in controlled intervention studies

Many controlled intervention studies have been performed on coffee to study the metabolism of known coffee constituents (Table 1). Most aimed to study the metabolism of caffeoylquinic acid derivatives. Typically, in these studies, a small number of subjects consumed coffee after a washout period, and blood or urine samples were taken at intervals for analysis. In early studies, metabolites were quantified by HPLC after enzymatic hydrolysis of glucuronide and sulfate conjugates. Isoferulic acid concentrations most markedly increased in urine samples taken periodically of volunteers repeatedly dosed with coffee and was thus proposed as a potential intake biomarker [18]. Another such intervention indicated that caffeic acid, as well as the microbial metabolite *m*-coumaric acid, appear in 24-hour urine after coffee intake [19]. With mass spectrometers, many more metabolites derived from coffee chlorogenic acids were later characterized without enzyme treatment. For example, individual caffeic and ferulic acid conjugates were measured by liquid chromatography coupled to MS (LC-MS) in the plasma and urine of subjects fed instant coffee [20]. Dihydroferulic acid 4-*O*-sulfate and dihydrocaffeic acid 3-*O*-sulfate attained the highest plasma concentrations after coffee intake. Dihydrocaffeic acid 3-*O*-sulfate and feruolylglycine, were reported as the most sensitive urinary biomarkers of intake. Further, in a double-blind randomized controlled trial investigating bioavailability of chlorogenic acids from coffee, all chlorogenic acid metabolites increased in a dose-dependent manner in plasma and urine after the administration of coffee containing three different levels of chlorogenic acids [21]. Concentrations of caffeic and ferulic acid sulfates were most markedly increased, but coffee intake also caused increases of intact caffeoylquinic acids and sulfated caffeoylquinic acid lactones. 3-Feruloylquinic acid, in both 24-hour urine and plasma, was highly correlated with the caffeoylquinic acids consumed from coffee (Spearman r=0.81 and r=0.73 respectively).

Despite providing a wealth of information on potential markers of coffee intake, the main purpose of these controlled intervention studies was usually to investigate the metabolism of hydroxycinnamic acid derivatives rather than search for novel biomarkers of intake. Other authors have followed the appearance of a variety of coffee compounds in blood or urine in subjects administered a standardized dose of coffee [22]. Dihydroferulic acid conjugates, trigonelline, caffeine and its primary metabolites were found to persist in plasma for long enough that they should never be fully cleared in individuals drinking three cups of coffee over a day. *N*-2-Furoylglycine was identified as a promising biomarker of coffee intake after the untargeted NMR profiling of spot urine samples from five volunteers administered a dose of espresso coffee [23]. The highest concentration was observed two hours after intake. N-2-furoylglycine is derived from furans formed during coffee roasting.

Biomarkers of coffee intake in observational studies

Markers uncovered in controlled intervention studies may be sensitive but not sufficiently specific to the food of interest since other possible food sources of these metabolites are excluded throughout the intervention. Caffeoylquinic and feruloylquinic acids and their derivatives, for example, are also present in fruits, vegetables and grains [24]. Although coffee is the principal dietary source, high intake of other confounding foods could lead to inaccurate estimates of intake. Biomarkers lacking specificity are better excluded in observational studies, in which subjects consume their usual diets. Food intake is estimated with dietary questionnaires before blood or urine collection. The first such study on coffee hypothesized that urinary isoferulic acid, a metabolite of caffeoylquinic and caffeic acids, would reflect habitual coffee intake (Table 1) [25]. Study participants, consuming their usual diets, recorded coffee intake via questionnaires and 24-hour pooled urine was collected. Urinary isoferulic acid excretion varied substantially between coffee consumers and was not strongly related to coffee intake as reported by food frequency questionnaire (FFQ) (r=0.26) or 24-hour dietary recall (r=0.18). Another study on 53 free-living French subjects showed a stronger correlation (r=0.63,

p<0.001) between 5-caffeoylquinic acid concentrations in spot urine and coffee intake, but this association was not statistically significant when tested in 24-hour urine [26]. High correlations were observed between coffee intake and caffeic acid (r=0.65), protocatechuic acid (r=0.60) and ferulic acid (r=0.58) concentrations measured in 24-hour urine collected in 475 adult participants from the European Prospective Investigation into Cancer and Nutrition (EPIC) cross-sectional study [27].

Later metabolomic studies found novel markers of coffee intake in observational studies using untargeted approaches. In a first such study, N-methylpyridinium and trigonelline, products of the coffee roasting process, were found to best distinguish coffee-drinkers from non-coffee drinkers after analysis of urine by LC-MS [28]. Both compounds remained elevated in urine for at least two days after coffee consumption and were thus proposed as stable biomarkers of intake. *N*-Methylnicotinamide also distinguished the two groups, although it was not considered specific to coffee intake, being a metabolite of niacin found in a wide range of foods. In another study, dihydrocaffeic acid and its 3-glucuronide, measured in 24-hour urine by LC-MS, was found to discriminate groups of high and low coffee consumers with high sensitivity and specificity [29]. A later cross-sectional study in 481 subjects of the EPIC cohort, also using 24-hour urine samples, showed significant correlations between concentrations of 16 phenolic acids, mostly glucuronide or sulfate esters, with acute coffee intake [30]. Dihydroferulic acid sulfate concentrations correlated most strongly with coffee intake whether assessed by FFQ (r=0.62) or 24-hour dietary recall (r=0.52). Dihydroferulic acid sulfate, feruloylquinic acid glucuronide, ferulic acid sulfate and guaiacol glucuronide were the metabolites whose measured intensities best classified subjects into the highest or lowest quintiles of coffee intake (ROC AUC > 94%). Non-phenolic metabolites were not investigated in this study.

The use of 24-hour urine samples yields the widest range of potential biomarkers, as all coffee metabolites accumulate in urine after coffee intake regardless of metabolite pharmacokinetics. Biomarkers identified in 24-hour urine samples cannot be assumed to be effective markers when

measured in spot urine or in blood samples taken at a single time point, since many food-derived metabolites are excreted rapidly after absorption. A few authors have searched for markers of coffee intake in spot urine or blood collections (Table 1). For example, biomarkers of coffee intake were searched for in morning spot urines of French subjects from the SUVIMAX cross-sectional study [31]. The intensities of several coffee-derived metabolites accurately classified consumers into high and low intake groups. The most effective of these were the diterpene, atractyligenin glucuronide (r=0.534, ROC AUC=0.98), the cyclic amino acid cyclo(isoleucylprolyl) (r=0.543, ROC AUC=0.969) and the caffeine metabolite 1-methylxanthine (r=0.561, ROC AUC=0.965). Also, urinary concentrations of 1,7-dimethyluric acid, 1-methyluric acid and trigonelline each classified subjects with excellent sensitivity and specificity (ROC AUC>0.9). Combining cyclo(isoleucylprolyl), 1-methylxanthine and trigonelline concentrations as a single biomarker increased classification performance relative to any one single compound. Hippuric acid was elevated in the urine of coffee consumers, while caffeoylquinic acid-derived metabolites were not reported as discriminants in this study.

Two further studies have identified biomarkers of coffee intake in blood. Trigonelline, 1-methylxanthine and paraxanthine were identified as serum biomarkers of coffee intake in an American nested case-control study, along with *N*-2-furoylglycine and catechol sulfate [32]. A more detailed study on coffee in the same cohort additionally reported that plasma trigonelline (partial r=0.608) and quinic acid (partial r=0.59) concentrations best correlated with coffee intake as reported by FFQ [16]. In contrast, concentrations of unmetabolized caffeine correlated moderately with coffee intake (partial r=0.327).

The studies described above proposed biomarkers of intake exclusively in urine, plasma or serum. A small number of studies have considered coffee-derived metabolites in other biospecimens. For example, after a dose of coffee, a Japanese group was able to measure caffeine and three isomers of dimethylxanthine in fingerprints [33]. Also, pyridine was found to increase in breath after consumption of a large cup of espresso [34]. Such techniques may not be

applicable to epidemiological studies at present but represent possible future alternatives to measuring biomarkers of coffee intake in blood and urine.

In summary, many exogenous blood and urinary metabolites have been proposed as coffee intake markers but their validity depends on study design, study population, biofluid and analytical method. In intervention studies where urine or blood samples are taken shortly after the administration of coffee, hydroxycinnamic acid derivatives such as caffeic and dihydroferulic acid (and their phase II conjugates) have been most commonly proposed as biomarkers of coffee intake. Most recent studies in free-living subjects suggest that several phenolic acids (ferulic, isoferulic, dihydroferulic, caffeic and dihydrocaffeic acids and their glucuronides and sulfate esters), alkaloids (caffeine, trigonelline and their metabolites), cyclo(isoleucylprolyl) and atractyligenin glucuronide, measured in urine, are the most sensitive and specific biomarkers of coffee intake. Fewer studies in free-living subjects were conducted on blood and trigonelline and quinic acid were found to best correlate with coffee intake.

Still it will be important to check the specificity of these biomarkers in the populations where implemented. Indeed some of these biomarkers like caffeine or ferulic acid may also arise from other dietary sources such as tea or soft drinks for caffeine or wholegrain cereals for ferulic acid. Trigonelline and cyclo(isoleucylprolyl) or their precursors may be preferred considering their high specificity for coffee.

The evaluation of the specificity of dietary biomarker requires a good knowledge of the composition of the beverages and of other foods that may act as confounders. This evaluation is often difficult as food composition data for secondary metabolites is often missing or scattered across a large number of publications and not easily analysed. More databases on food composition such as those developed for polyphenols or for caffeine [35, 36] will be needed to identify all possible confounders [30].

- **Tea biomarkers**

Tea (*Camellia sinensis*) is another beverages widely consumed worldwide. Tea is classified into three main types according to the level of leaf fermentation, green (unfermented), oolong (semi-fermented) and black (fully fermented) tea. The main constituents of tea are polyphenols, caffeine and other methylxanthines, proteins, nucleic acids and amino acids such as theanine [37]. Catechins are the most abundant polyphenols in tea, and epigallocatechin gallate (EGCG) is the predominant compound in this class, followed by epigallocatechin (EGC), epicatechin gallate (ECG), epicatechin (EC), and their epimeric forms catechin (C), gallocatechin (GC), catechin gallate (CG), and gallocatechin gallate (GCG). During the fermentation process they are largely oxidized into theaflavins and thearubigins, which are then the major constituents of black and oolong tea. Flavonols, mainly glycosides of quercetin, kaempferol and myricetin, are also present in tea. Gallic acid and theogallin are the two predominant phenolic acids and their concentrations in tea respectively increase and decrease during the fermentation process. Tea also contains alkaloids, particularly caffeine but also small quantities of theophylline and theobromine [37].

Tea consumption has been associated with a reduction of the risk of chronic diseases and more particularly cardiovascular diseases, type 2 diabetes, cancers and neurodegenerative diseases [38-41]. The tea constituents described above may mediate these effects. However, some of these results are not fully consistent among the different studies and it has been suggested that this lack of agreement could be associated with the traditional methods used to assess tea consumption, usually measured in number of cups of tea consumed per day [42]. Therefore, the use of accurate biomarkers reflecting tea exposure is a promising tool to better assess the role of tea consumption in human health.

Tea metabolites in controlled intervention studies

Metabolism and pharmacokinetics of catechins have been studied in a large number of intervention studies with green or black tea (Table 2). The main compounds detected in biofluids after green tea ingestion are catechins and their metabolites formed in phase II biotransformations (methylation, glucuronidation and sulfation) and ring-fission reactions. EGCG, EGC, ECG and EC were the main compounds detected in plasma. They are quickly absorbed and peak concentrations are observed about 2 hours after ingestion [43]. Tea catechins are also quickly excreted and their elimination half-lives usually do not exceed 3 hours. EGC is the most abundant catechin in plasma after tea intake, mainly found in its glucuronidated form [43]. EGC is also methylated in the liver and 4′-*O*-methyl-EGC is found in both sulfated and glucuronidated forms. Galloylated catechins (EGCG and ECG) are present in plasma in their non-conjugated forms.

Catechins not absorbed in the small intestine reach the colon where they can be degraded by the microbiota into low-molecular weight metabolites such as hydroxyphenylvalerolactones, hydroxyphenylvaleric acids, phenolic acids and hippuric acid. These metabolites show longer elimination half-lives in urine where they persist for 48 hrs after tea intake [44].

Most studies showed a linear relationship between plasma concentrations or urine excretion of tea catechins and the ingested dose [45-49] although concentrations of some catechins were also shown to reach a plateau at a high level of intake in some studies [50, 51].

Many untargeted metabolomics studies have been conducted to elucidate additional candidate biomarkers of tea intake [52-54]. In a placebo-controlled cross-over intervention study based on high-resolution mass spectrometry, twelve male subjects consumed a single capsule of tea extract (2.65 g) or a placebo after one day of a polyphenol-poor diet [55]. Fifty-nine polyphenol metabolites were identified and increased in concentration after black tea ingestion. These metabolites were catechins, phenolic acids, valerolactones and simple phenols, most in the form

of glucuronides and sulfate esters. The highest concentrations in plasma were observed within 1-4 hrs (catechins, kaempferol, gallic acid) or 5-10 hrs (microbial metabolites such as phenylvalerolactones, pyrogallol and hippuric acid), suggesting that they would be better indicators of tea intake if measured in 24-h urine samples rather than spot urine samples.

Some markers may be used to differentiate intake of green and black tea. In particular 4-*O*-methylgallic acid, formed by *O*-methylation of gallic acid, has frequently been reported in both urine and plasma after black tea intake [56-58]. This metabolite is also detected in urine after green tea intake but in much lower concentrations [59]. This difference in concentrations is explained by the low content of gallic acid in green tea when compared to black tea. Measurement of 4-*O*-methylgallic acid in urine or plasma could thus be useful to differentiate consumption of black and green tea, particularly if used in combination with catechins more abundant in green tea than in black tea [35] and present at higher concentrations in plasma or urine after consumption of green tea when compared to black tea [50, 60-62]. Theaflavins and thearubigins are abundant in black tea and absent from green tea. However their high molecular weight considerably limits or prevents their absorption in the gut [63]. Therefore they cannot be used as biomarkers of black tea intake.

Catechin metabolites formed by the gut microbiota are, in comparison with catechins or gallic acid, less useful as biomarkers of tea intake. Some of these metabolites (some phenolic acids and hippuric acid) were present in plasma in the absence of tea consumption [55]. This is explained by the existence of precursors of these metabolites in other dietary sources such as coffee, cocoa, fruits or vegetables beside tea [64-66]. Other metabolites derived from tea catechins such as 5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone, 5-(3',4'-dihydroxyphenyl)-γ-valerolactone and/or 5-(3',5'-dihydroxyphenyl)-γ-valerolactone may be more specific for tea intake [49, 67]. However, formation of these metabolites from catechins shows high interindividual variations, compromising their potential use as biomarkers of intake [68].

Similarly, other tea constituents such as quercetin or caffeine may also be of limited utility as biomarkers due to their abundance in other foods and beverages [69, 70].

Biomarkers of tea intake in observational studies

Cross-sectional studies allow the investigation of biomarker sensitivity and specificity. In the Shanghai Cohort Study, EGC, 4′-*O*-methyl-EGC, EC and two phenylvalerolactones measured in urine were significantly associated with self-reported tea intake [71, 72]. In another study carried out in 481 subjects from the European Prospective Investigation into Cancer and Nutrition cohort, metabolic profiles were compared in 24-hour pooled urine samples by high resolution MS in tea consumers and non-consumers [30]. Several phenolic compounds were found to be associated with recent tea intake as assessed with a 24-hour dietary recall. The compound that best distinguished non- and high consumers was found to be 4-*O*-methylgallic acid (ROC AUC: 84%). 4-*O*-Methylgallic acid was also associated with habitual tea intake as assessed with a FFQ despite its short elimination half-life, and this was explained by frequent tea intake in this population. However some confounding with red wine intake was also observed. Other polyphenol metabolites were also found to be associated with tea intake in the same study (methyl(epi)catechin sulfate, dihydroxyphenyl-γ-valerolactone sulfate, hydroxyphenylvaleric acid glucuronide and pyrogallol sulfate) but they may not be more specific for tea than 4-*O*-methylgallic acid, as other dietary sources of the same polyphenol metabolites are also known. Excretion of catechin metabolites in urine were also found to be correlated with intake of chocolate products, apples and pears. Pyrogallol sulfate has also been identified in plasma after intake of nuts [73] or mixed berry fruit [74]. In another cross-sectional study, 24-hour urinary kaempferol was correlated with tea intake ($r = 0.41$; $p < 0.01$) [75], but kaempferol was also found to be correlated with onion in another study [76]. Therefore, most of these markers associated with tea intake may not be specific enough for tea, depending on the possible consumption of confounding foods in the population considered.

Some catechins might be the most specific biomarkers for tea intake. Three catechins, EGCG, EGC and ECG, are mainly or exclusively found in tea according to the Phenol-Explorer database [35]. Galloylated catechins EGCG and ECG and their methylated metabolites have been detected in both plasma and urine in tea intervention studies but they may be more difficult to measure in population studies due their limited absorption in the gut when compared with other catechins [77, 78]. EGC and its phase II derivative, 4'-*O*-methyl-EGC have been measured in several cohort studies and might be a useful biomarker for green and black tea intake [71, 72]. 4-*O*-Methylgallic acid, a metabolite of gallic acid (particularly abundant in black tea), has been associated with black tea consumption in both intervention and observational studies (Table 2). Ratio of 4-*O*-methylgallic acid over EGC is expected to be higher in black tea consumers when compared to green tea consumers and could help to discern which of the two types of tea has been consumed. A similar approach using ratios of specific alkylresorcinols was used for discriminating between wholegrain wheat and rye intakes [79, 80]. Other constituents characteristic of tea like theanine have not received much attention and should also be tested as possible biomarkers of tea intake.

- **Sugar-sweetened beverage biomarkers**

Identification and validation of markers reflecting the consumption of SSB is an important task to better assess the association between the consumption of SSB and related health effects. Major challenges to identify such biomarkers are linked to the broad definition of what a SSB is, and to the variability of their composition. One major point is the nature of the caloric sweeteners used. Beverages containing added caloric sweeteners such as sucrose or high-fructose corn syrup (HFCS) are the main types of SSB. These beverages form a very heterogeneous group comprising soft drinks, fruit drinks, sports drinks, energy drinks, flavored water drinks and iced teas [81-83]. Given this diversity of products and their different composition, having specific markers that reflect the intake of individual products or the total daily consumption of SSB is challenging.

The dominating ingredient in all SSB is, by definition, the added sugar. Exposure markers for the intake of sugar have been described using two different approaches, either through the measurement of the carbon isotope ratio $^{13}C/^{12}C$ (expressed as $\delta^{13}C$ value) or through the determination of sugars in urine. The first approach is based on the different discrimination against carbon dioxide formed from the carbon-13 and carbon-12 isotopes in plants. Crop species have been classified as C3 and C4 plants depending on their photosynthetic pathway. The photosynthetic pathway of C3 plants like sugar beet discriminates against $^{13}CO_2$ compared with $^{12}CO_2$ and thus the resulting plant mass carbon has a lower $^{13}C/^{12}C$ ratio than atmospheric CO_2. In contrast, the C4 pathway is almost non-discriminating against ^{13}C, resulting in a plant mass higher in ^{13}C compared to C3 plants. Sugar cane and corn, the main sources for sugar production in the U.S., are C4 crops. As a consequence sugar from these crops is enriched in carbon-13, compared to sugar produced by C3 plants. This enrichment can be seen in whole human biospecimens or specific metabolites in these biospecimens after consumption of sugars from C4 plants.

The $^{13}C/^{12}C$ isotope ratio of blood plasma, finger-prick blood or of the amino acid alanine either from hair protein or red blood cells have been proposed in various studies to predict intake of SSB and added sugar in the U.S. (Table 3) [84-88]. However, this approach also has limitations depending on the nature of the sugar sources consumed in various populations. While in the U.S. added sugar is mostly derived (78%) from sugar cane or corn [89], the situation in Europe is the opposite with around 80% added sugar derived from the C3 plant sugar beet [90]. In consequence, the use of carbon-13 as a potential exposure marker for added sugar or SSB is limited to the U.S. Additionally, carbon-13 in whole blood is also influenced by the intake of corn products and meat from livestock fed on corn and thus the intake assessment of sugar might be confounded. Two methods have been described to correct for these confounding food items. The first method uses nitrogen-15 which is elevated in marine foods as well as in meat products. Therefore, it was suggested to use this second isotopic marker as control for the intake

of animal protein [86, 87, 91, 92]. Nash et al. [87, 92] found favorable results in a study population of Yup'ik eskimos consuming low amounts of sugar from sources not ^{13}C-enriched such as sugar beet, honey, fruits or dairy products, but high amounts of fish and marine mammals compared to other U.S. populations. They were able to explain three times as much variation in the sweetener intake by using a model including both, the carbon and nitrogen isotope ratios than by using $\delta^{13}C$ alone. In contrast, Fakhouri et al. [86] and Hedrick et al. [91] found no significant improvement in their results after correcting for the animal protein intake using nitrogen-15 in a population where more corn-fed meat is consumed. The second method to correct for confounders is based on the use of a specific metabolite to measure carbon-13, which favors the incorporation of glucose carbon like alanine as described by Choy et al. [84]. They found no association between the carbon-13 of alanine and dietary confounders like commercial meat, fish and corn products. At the same time they found that a dual-isotope model using carbon-13 and nitrogen-15 in red blood cells was associated with meat intake as well as sweeteners. The use of alanine as a specific metabolite shows favorable results in this respect, but further research is needed, especially concerning different populations.

The second approach for calculating sugar intake uses urinary sucrose and fructose as exposure markers [93, 94]. Details on the different interventional studies and surveys were recently reviewed [93]. For this exposure marker no information on its applicability for SSB is currently available.

Both approaches, the carbon isotope ratio of different biospecimens and urinary fructose and sucrose, show promise as exposure markers for sugar and sucrose intake. However, sugar alone does not appear suitable as an exposure marker for SSB consumption because it is also contained in hundreds of other food items and is thus not specific. The use of an exposure marker for sugars in combination with additional substances used as ingredients might prove to be more specific. However ingredients used in SSB vary depending upon the type of SSB. Citric acid is added to many types of widely distributed SSB as acidity regulator. However, citric acid

seems not suitable as a marker for SSB consumption since it is produced in large quantities in the human metabolism. Moreover, it also naturally occurs in large concentrations in fruits and fruit juices. Other typical ingredients in SSB are natural or artificial flavors such as ginger extract for ginger ale or caffeine for cola type beverages. Using these compounds or their metabolites as markers would consequently only cover specific subgroups of SSB. Their specificity should also be carefully assessed as some of these substances may also be ingested with other foods, like coffee in case of caffeine.

A more reliable approach to reach a high specificity when assessing intake of a whole food group such as SSB might be to use a combination of exposure markers as described by Gibbons et al. [95]. In this study a subcohort of 565 participants from the National Adult Nutrition Survey was divided into four quartiles. Four metabolites detected by NMR spectroscopy in urine - formate, citrulline, taurine and isocitrate - showed concentration levels that differed significantly between SSB consumers (n=146) and non-consumers (n=391) as identified via a 4-day semi-weighed food record. Levels of these same four metabolites also showed transient and modest increase in an acute intervention study (n=10) after the consumption of one can of cola. The authors detected all four markers in the cola drink used for the acute intervention study. Although found in the cola drink, none of these compounds are normally added in a pure form so that they would most likely originate from added flavor extracts. It also remains questionable whether these markers are indicative of intake of the whole group of SSBs or cola consumption only. So-called energy drinks often contain taurine, a sulphur-containing amino acid. Thus taurine might serve as a marker for this specific type of beverage. However, it is unlikely that taurine is present in all types of SSB. Furthermore, other sources of taurine cannot be excluded. Taurine is formed in human metabolism from methionine and cysteine with an endogenous average daily synthesis of 0.4-1.0 mmol (50-125 mg) [96]. Meat and seafood are significant dietary sources of this amino acid [93]. Similarly, isocitrate is an intermediate of the citric cycle - like citrate - and therefore occurs in all humans, animals and plants. These examples

demonstrate that the urinary excretion of these compounds can have different origins that interfere with their use as markers of exposure for SSB. Further validation of the proposed panel of biomarkers in other populations is still necessary. Furthermore, more research is needed to identify new markers or marker panels with higher specificity and selectivity.

In summary, biomarkers of sugar intake would not be ideal biomarkers of SSB intake due the diversity of dietary sources of sugar. One exception may be the $^{13}C/^{12}C$ isotope ratio in countries where sugar cane and corn sugar is more specifically used as SSB sugar. Other SSB ingredients or combinations of ingredients might be used as biomarkers of intake for SSB or for specific types of SSB. However none of these biomarkers have yet been validated in population studies.

- **Low-calorie sweetened beverage biomarkers**

The use of LCSB has increased over the years as an approach for reducing the free sugar content of the diet while maintaining palatability. Given that the overconsumption of free sugars, particularly from sugar-sweetened beverages (SSB), has been implicated in the development of adverse health outcomes such as obesity and type 2 diabetes [97], LCSB are often considered as potentially useful in the prevention of such outcomes. Research interest to date has focused on assessing the effect of LCSB consumption on a variety of health outcomes such as metabolic health, weight management and renal impairment [98, 99]. To carry out such research, reliable assessment of LCSB intake is highly desirable and the successful implementation of a biomarker approach could facilitate this.

No studies were identified which specifically aimed to identify biomarkers of LCSB intake. Rather, many of the studies investigated the impact of LCSB consumption on health via the measurement of effect biomarkers in relation to metabolic syndrome [100], blood pressure [101] and glycaemic control [102, 103]. LCSB consumption was found to influence the concentration of various compounds including urinary hippuric acid excretion [104], serum uric acid concentrations [105] and plasma free fatty acid concentrations [106]. However, these biomarkers are not specific to LCSB; hippuric acid is a potential marker of toluene exposure as

well as of fruit and vegetable intake, while serum uric acid and plasma free fatty acids are endogenous compounds influenced by factors other than LCSB intake. Therefore, utilisation of these biomarkers is unlikely to provide useful information on LCSB intake owing to a lack of specificity. Several AGEs were found to be present in regular and, to a lesser extent, in diet varieties of cola drinks. They are normally excreted via the urine, but they are also confounded by other more significant dietary sources of these compounds [107]. Brominated vegetable oil (BVO) is used in North America as a clouding agent in some soft drinks but is not permitted for use in many other areas, including the European Union, owing to concerns about bromine toxicity. Bendig et al. assessed the BVO content of commonly consumed soft drinks and BVO was detected in only three out of the ten drinks investigated indicating that tissue bromine concentrations would not serve as a reliable biomarker of LCSB intake [108]. Eisner et al. assessed the citrate and malate content of diet beverages in the context of the treatment of nephrolithiasis (kidney stones) [109]. Secondary literature searches revealed that concentrations of these compounds may be directly influenced by other dietary and non-dietary factors or may not be relevant for all LCSB.

Given that the term LCSB encompasses an array of products including carbonated beverages, fruit cordials, sweetened dairy products as well as flavoured teas, all of which are likely to differ significantly in their composition, the identification of a specific biomarker of intake which is specific to all LCSB as a single food group is likely to be challenging. The low-calorie sweeteners (LCSs) themselves, common to a large diversity of LCSB, may prove to be the most useful biomarkers of LCSB intake. There are eleven LCSs currently approved for use on the European market (Table 4) and their use extends to a wide range of foods apart from LCSB, as well as non-dietary products such as oral hygiene products and e-cigarette fluids [110]. Although this may complicate the use of LCSs as valid markers of LCSB intake, LCSB are widely considered to be the primary source of LCSs in the diet and therefore measurement of these as biomarkers may provide a viable option for assessing LCSB intake.

The potential application of a biomarker approach for investigating intake of LCSs has recently been reviewed [110]. The metabolic fate of each LCS is assessed prior to approval for use as a food additive (Table 5) and therefore a targeted approach may be implemented for those that lend themselves well to analyses. A number of LCSs are excreted in urine in the same form as found in the LCSB. Acesulfame-K [111] and saccharin [112-114] are almost completely absorbed and excreted unchanged via the urine. The usefulness of urinary excretions of these two LCSs as biomarkers of intake was investigated by Wilson et al. who measured levels of excretion in 24-hour urine samples and found high correlation with acute intakes of saccharin and acesulfame-K in an intervention study with five different doses of the LCS (R^2 >0.99 for both compounds), demonstrating a clear dose-response relationship for both compounds [115]. Two other LCSs, cyclamate [116] and sucralose [117, 118], are partially absorbed (respectively 30-50 and 10-15%) and subsequently excreted unchanged in urine. A proportion of cyclamate is converted to cyclohexylamine via bacterial hydrolysis in the gut in around 20% of the population. This is then absorbed and excreted via the urine [119]. The process of conversion to cyclohexylamine is highly variable between and within individuals, particularly during chronic exposure [116] and therefore some doubt must be cast on its usefulness as a reliable marker of cyclamate intake in the free-living population. A relatively new and increasingly used LCS, a mixture of steviol glycosides isolated from the leaf of *Stevia rebaudiana*, undergoes bacterial hydrolysis in the gut prior to absorption and subsequent excretion in urine as steviol glucuronide [120-122]. Advantame is deesterified to advantame-acid and a small proportion is absorbed (~6%) and excreted in urine [123]. The excretion products of acesulfame-K, saccharin, sucralose, cyclamates, steviol glycosides and advantame are highly specific to exposures to the parent compound *per se* as they are not produced endogenously. Therefore any detection of these in urine would indicate intake of the related sweeteners. However further work would be required to assess their specificity as biomarkers of LCSB intake.

Other LCS, namely aspartame, neohesperidine dihydrochalcone (NHDC) and thaumatin, undergo extensive metabolism into metabolites commonly found in the diet or in the body, which suggests that no obvious specific biomarkers of intake exist for these LCS. Aspartame is broken down to its three constituents, aspartic acid, phenylalanine and methanol, all of which are often present in greater quantities in other dietary sources [124, 125]. Thaumatin, a naturally occurring protein complex, undergoes normal protein digestion and therefore identification of a specific biomarker of intake is not likely to be possible [126]. Finally, NHDC has a similar structure to naturally occurring flavonoids with similar metabolites and for this reason, specificity of any putative biomarker is likely to be problematic [127].

In summary, a biomarker approach for assessing intake of specific LCS used in LCSB may prove useful as several of them are excreted unchanged in urine following ingestion. Expected biomarkers should be based on LCSs most commonly used in LCSB. However, considering the variety of LCSs used, any biomarker approach should consider assessing a range of commonly consumed LCSs. More work will be needed to validate their use as biomarkers of LCSB intake and to identify possible confounding by other foods containing the same LCSs.

Conclusions

A wide diversity of biomarkers has been proposed to estimate the consumption of non-alcoholic beverages. Metabolism of major constituents of non-alcoholic beverages has been studied in a large number of controlled intervention studies, and many metabolites were identified mainly in blood and urine (Tables 1 and 2). Participants in these studies have usually consumed a high amount of a particular beverage with a fully controlled diet after a proper wash-out period. These studies are useful to identify putative biomarkers of intake, but they provide limited evidence on their potential value as biomarkers of intake in a population and more particularly on their sensitivity and specificity.

Several of these candidate biomarkers have been measured in observational studies with individuals following their own diet and their sensitivity and specificity as indicators of coffee, tea and SSB intake evaluated (Tables 1-3). Various compounds including phenolic acids, alkaloids and terpenes measured in urine or plasma samples were shown to accurately predict coffee intake in various populations and EGC and 4-*O*-methylgallic acid were also shown to be good indicators of tea intake. These two last biomarkers could also possibly be used to differentiate intake of green and black tea. Several biomarkers have been proposed to estimate SSB intake but none for LCSB intake. Difficulties met in finding biomarkers for these two last classes of beverages are explained by the lack of constituents that would be at the same time characteristic of each of these two groups of beverages and absent in all other foods.

Combinations of biomarkers may be needed to assess intake of SSBs and LCSBs. Combinations of biomarkers may also allow estimating intake of different types of beverages within a particular group, as suggested for green and black tea. Combination of a generic biomarker for coffee with caffeine should help assessing relative intake of caffeinated and decaffeinated coffee. Panels of biomarkers of intake have been proposed but none has yet been validated.

The new biomarkers, identified through metabolomic or other approaches, will need to be validated in populations sharing similar lifestyle and diet to the ones where these biomarkers will be used. All possible confounders (e.g. foods containing the same biomarker or a precursor transformed into the biomarker in the body) will need to be carefully considered. More extensive food composition data for these compounds, will have to be collected and made easily accessible in new databases. Biomarkers with long elimination half-lives should be preferred particularly to assess intakes of foods more episodically consumed [128] although this may be less important for beverages like coffee or tea most often consumed on a daily basis [30]. Finally, the practicality of biomarkers will need to be assessed, to identify their performance according to the type of biospecimens (e.g. urine vs. blood, fasting vs. non-fasting blood

samples or 24-hr urine samples vs. spot urine samples), performance of analytical methods and cost of analyses.

Overall biomarkers should help estimating intake of non-alcoholic beverages and this may be particularly useful to overcome some of the limitations met with dietary questionnaires. It is expected that these biomarkers will be increasingly used in cohort studies to evaluate the effects of non-alcoholic beverages on disease risk. However it will also be important to carefully evaluate the respective advantages of biomarkers and questionnaires, an evaluation that has not been done yet.

Competing interests

The authors declare that they have no competing interests.

Funding

FoodBAll is a project funded by the BioNH call (grant number 529051002) under the Joint Programming Initiative, "A Healthy Diet for a Healthy Life". The project is funded nationally by the respective Research Councils; the work was funded in part by a grant from the Danish Innovation Foundation (#4203-00002B) and a Semper Ardens grant from the Carlsberg Foundation to LOD; a grant from the China Scholarship Council (201506350127) to QG; a postdoc grant from the University of Rome La Sapienza ("Borsa di studio per la frequenza di corsi o attività di perfezionamento all'estero" erogata ai sensi della legge 398/89) to GP; a grant from the Spanish National Grants from the Ministry of Economy and Competitiveness (MINECO) (PCIN-2014-133-MINECO Spain) and an award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR to CAL.

Table 1. Metabolites identified in coffee human intervention studies and biomarkers of intake measured in observational studies.

Beverage	Study design	No. subjects	Sample type	Analytical method	Enzymatic hydrolysis	Discriminating metabolites / candidate biomarkers	Association with	Reference
Instant coffee (repeated intake)	Intervention	5	Urine	HPLC	Yes	Ferulic, isoferulic, dihydroferulic, vanillic acids, 3-hydroxyhippuric acid		[18]
Instant coffee	Intervention	9	Urine	LC-MS	Yes	Chlorogenic acid, caffeic acid, m-coumaric acid		[19]
Coffee	Intervention	13	Urine	CE-MS	No	Coumaric acid, caffeic acid		[129]
Instant coffee	Intervention	11	Plasma, urine	LC-MS	No	Ferulic and dihydroferulic acid sulfates (blood and urine), feruloylglycine (urine only)		[20]
Filter coffee	Intervention	9	Urine	LC-MS	No	Trigonelline, N-methylpyridium		[28]
Filter coffee	Intervention	13	Plasma	LC-MS	No	Trigonelline, dimethylxanthines, methyl xanthines, dihydroferulic acid, dihydrocaffeic acid sulfate, ferulic acid glucuronide, ferulic acid sulfate, ferulic acid, dihydroferulic acid sulfate, dihydroferulic acid glucuronide, N- feruloylglycine		[22]
Instant coffee	Intervention	11	Plasma, urine	LC-MS	No	3-, 4- and 5-Feruloylquinic acid, 3 and 4-Caffeoylquinic acid lactone sulfate (urine and plasma), isoferulic acid-3-glucuronide (urine only)		[21]
Coffee (repeated intake)	Intervention	8	Urine	1H-NMR	No	2-Furoylglycine		[23]
Coffee	Observational	111	Urine (24h)	GC-MS	Yes	Isoferulic acid	FFQ	[25]
Coffee	Observational	344	Urine (24h)	GC-MS	Yes	Isoferulic acid	24-HDR	[25]
Coffee	Observational	53	Urine (24h and spot)	LC-MS	Yes	Chlorogenic acid, caffeic acid	2-Day dietary record	[26]
Coffee	Observational	68	Urine (24h)	FIA-MS	Yes	Dihydrocaffeic acid, dihydrocaffeic acid 3-glucuronide	FFQ	[29]

Coffee	Observational	39	Urine (morning spot)	LC-MS	No	Atractyligenin glucuronide, cyclo(isoleucylprolyl), trigonelline, paraxanthine, theobromine, theophylline, 1-methylxanthine, hippuric acid, trimethyluric acid, 3-hydroxyhippuric acid, AFMU, 1,3 or 3,7 dimethyluric acid, caffeine	FFQ	[31]
Coffee	Observational	502	Serum	LC-MS, GC-MS	No	Trigonelline, Quinic acid, 1-methylxanthine, paraxanthine, N-2-furoylglycine, catechol sulfate	FFQ	[32]
Coffee	Observational	498	Serum	LC-MS	No	Trigonelline, quinic acid, paraxanthine, N-2-furoylglycine, catechol sulfate, caffeine, 1-methylxanthine, theophylline, trimethyluric acid, hydroxyhippuric acid, 1-7-dimethyluric acid, 1-methyluric acid, cyclo(leu-pro), 4-vinylphenol sulfate, hydroxyphenylpropionate, theobromine, cinnamoylglycine	FFQ	[16]
Coffee	Observational	475	Urine (24h)	LC-MS	No	Dihydroferulic acid sulfate, guaiacol glucuronide, feruoylquinic acid, ferulic acid sulfate, feruoylquinici acid glucuronide, 3-caffeoylquinic acid, p-coumaric acid sulfate, caffeic acid sulfate, ferulic acid glucuronide, hydroxyhippuric acid, dihydrocaffeic acid sulfate, m-coumaric acid sulfate, dihydroferulic acid glucuronide, p-hydroxyphenyllactic acid, guaiacol sulfate, ethylcatechol glucuronide	24-HDR	[30]

Table 2. Metabolites identified in tea human intervention studies and biomarkers of intake measured in observational studies.

Type of beverage	Study design	No. subjects	Sample type	Analytical method	Enzymatic hydrolysis	Discriminating metabolites / candidate biomarkers	Association with	Reference
Green tea, black tea	Intervention	18	Urine	LC-ECD	Yes	Epicatechin, epigallocatechin		[50]
Green tea, black tea	Intervention	20	Urine	GC-MS	Yes	4-O-Methylgallic acid		[130]
Black tea	Intervention	10	Plasma, urine	HPLC	?	4-*O*-Methylgallic acid, gallic acid		[56]
Green tea, black tea (extracts)	Intervention	17	Urine	LC-MS	No	Hippuric acid		[131]
Green tea, black tea (decaffeinated)	Intervention	133	Urine	LC-ECD	Yes	(-)-Epigallocatechin		[60]
Green tea, black tea	Intervention	30	Plasma	LC-ECD	Yes	(-)-Epicatechin, (-)-epicatechin-3-gallate, (-)-epigallocatechin, (-)-epigallocatechin-3-gallate		[61]
Green tea	Intervention	10	Plasma, urine	LC-MS	No	(-)-Epicatechin-3'-O-glucuronide, (epi)catechin-O-sulfates, 3'-O-methyl-(epi)catechin-O-sulfates, 4'-O-methyl-(epi)catechin-O-sulfate, (epi)gallocatechin-O-glucuronide, 4'-O-methyl-(epi)gallocatechin-O-sulfates (urine and plasma); (-)-epicatechin-3-O-gallate, 4'-O-methyl-(epi)gallocatechin-O-glucuronide, (-)-epigallocatechin-3-O-gallate (plasma only); (epi)gallocatechin-O-sulfates (urine only)		[43]

Green tea	Intervention	20	Urine	LC-MS	No	(Epi)catechin glucuronide, (epi)catechin sulfate, (epi)catechin sulfoglucuronide, methyl(epi)catechin sulfate, (epi)gallocatechin glucuronide,(epi)gallocatechin sulfate, methyl(epi)gallocatechin glucuronide, methyl(epi)gallocatechin sulfate, methyl(epi)gallocatechin sulfoglucuronide, 5-(hydroxyphenyl)-γ-valerolactone glucuronide, 5-(hydroxyphenyl)-γ-valerolactone glucuronide sulfate, 5-(3′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide,5-(3′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide disulfate, 5-(4′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide, 5-(4′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide disulfate, 5-(4′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide sulfoglucuronide, methyl-5-(4′,5′-dihydroxyphenyl)-γ-valerolactone glucuronide, 5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone glucuronide, 5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone glucuronide sulfate, methyl-5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone glucuronide, methyl-5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone sulfate		[44]
Black tea	Intervention	4	Urine	LC-MS	No	(Epi)catechin sulfate, O-methylcatechin sulfate, O-methyl(epi)catechin sulfates, O-methyl(epi)gallocatechin sulfates, di-O-methyl(epi)gallocatechin sulfates, dihydronaringenin sulfates, 3′-O-methyl-5-(3′,4′-dihydroxyphenyl)-γ-valerolactone 4′-O-glucuronide, 4′-O-methyl-5-(3′,4′-dihydroxyphenyl)-γ-valerolactone 3′-O-glucuronide, 5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone 3′-O-glucuronide, 4′-O-glucuronide, 3′-O-sulfate, 4′-O-sulfate & sulfoglucuronide, O-methyl-5-(3′,4′,5′-trihydroxyphenyl)-γ-valerolactone 3′- or 4′-O-glucuronides, 3′ or 5′-O-glucuronide & O-sulfate, 5-(3′,4′-dihydroxyphenyl)-γ-valerolactone 3′-O-glucuronide, 4′-O-glucuronide & sulfates, 5-(3′,5′-dihydroxyphenyl)-γ-valerolactone 3′-O-glucuronide & 3′-		[132]

						O-sulfate, O-methyl-5-(3’,5’-dihydroxyphenyl)-γ-valerolactone glucuronides & sulfates, 5-(3’-hydroxyphenyl)-γ-valerolactone 3’-O-glucuronide, 4’-O-glucuronide, 5-(hydroxyphenyl)-γ-valerolactone sulfate, O-methyl-4-hydroxy-5-(3’,4’,5’-trihydroxyphenyl)valeric acid glucuronides, 4-hydroxy-5-(3’,4’-dihydroxyphenyl)valeric acid glucuronides, O-methyl-4-hydroxy-5-(3’,4’-dihydroxyphenyl)valeric acid glucuronides, 4-hydroxy-5-(3’,5’-dihydroxyphenyl)valeric acid glucuronide, O-methyl-4-hydroxy-5-(3’,5’-dihydroxyphenyl)valeric acid glucuronides, 4-hydroxy-5-(3’-hydroxyphenyl)valeric acid 3’-O-sulfate & 4’-O-sulfate, O-methyl-4-hydroxy-5-(dihydroxyphenyl)valeric acid sulfates, 4-hydroxy-5-(dihydroxyphenyl)valeric acid sulfates, O-methyl-4-hydroxy-5-(hydroxyphenyl)valeric acid sulfate, 4-hydroxy-5-(phenyl)valeric acid glucuronides & sulfates, hippuric acid, indole-3-acetic-acid glucuronide, indole-3-carboxylic acid glucuronide, p-cresol sulphate & glucuronide, pyrogallol 2-O-glucuronide, 1-O-sulfate & 2-O-sulfate, urolithin A -3-O-glucuronide, 8-O-glucuronide & sulfoglucuronide, urolithin B-O-glucuronide, vanilloylglycine, vanillic acid-4-O-glucuronide & 4-O-sulfate, phenylacetylglycine	
Green tea (extract)	Intervention	14	Urine	LC-MS	Yes	Catechin, , epicatechin, 3’-O-methylepicatechin, 4’-O-methylepicatechin, epicatechin-3-O-gallate, gallocatechin, gallocatechin gallate, epigallocatechin, 3’-O-methylepigallocatechin, 4’-O-methylepigallocatechin, epigallocatechin-3-O-gallate, 5-(3’,4’,5’-trihydroxyphenyl)- γ-valerolactone, 5-(3’,4’-dihydroxyphenyl)valerolactone, 5-(3’,5’-dihydroxyphenyl)valerolactone, gallic acid, 3-O-methyl gallic acid, 3-hydroxybenzoic acid, syringic acid, benzoic acid, hippuric acid	[133]

Black tea (extract)	Intervention	12	Plasma	LC-MS	No	(Epi)catechin sulfate, O-methyl(epi)catechin sulfates, di-O-methyl(epi)catechin sulfate, O-methyl(epi)gallocatechin sulfate, (epi)catechin gallate sulfate & sulfoglucuronide, O-methyl(epi)catechin gallate sulfate & sulfoglucuronide, (epi)gallocatechin gallate, (epi)gallocatechin gallate sulfate, O-methyl(epi)gallocatechin gallate sulfate, δ-(3',4'-dihydroxyphenyl)-γ-valerolactone) 3'-O-glucuronide, 4'-O-glucuronide, 3'-O-sulfate & sulfoglucuronide, 5-(3'-hydroxyphenyl)-γ-valerolactone 3'-O-glucuronide & 3'-O-sulfate, 5-(4'-hydroxyphenyl)-γ-valerolactone 4'-O-glucuronide, 5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone 3'-O-glucuronide, 4'-O-glucuronide, 3'-O-sulfate & sulfoglucuronide, O-methyl-5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone 3'/4'-O-glucuronide, 3'/5'-O-glucuronide & O-sulfates, 5-(3',5'-dihydroxyphenyl)-γ-valerolactone 3'-O-glucuronide & 3'-O-sulfate, O-methyl-5-(3',5'-dihydroxyphenyl)-γ-valerolactone 3'-O-glucuronide & 3'-O-sulfate, 4-hydroxy-5-(dihydroxyphenyl)valeric acid sulfates, O-methyl-4-hydroxy-5-(hydroxyphenyl)valeric acid sulfates, 4-hydroxy-5-(3',4'-dihydroxyphenyl)valeric acid glucuronides, 4-hydroxy-5-(3',5'-dihydroxyphenyl)valeric acid glucuronide, O-methyl-4-hydroxy-5-(3',5'-dihydroxyphenyl)valeric acid glucuronide, O-methyl-4-hydroxy-5-(3',4',5'-trihydroxyphenyl)valeric acid glucuronide, kaempferol glucuronide, O-methylgallic acid sulfates, pyrogallol 2-O-glucuronide & 2-O-sulfate, O-methylcatechol sulfates, resorcinol glucuronide & sulfate, hippuric acid	[55]
Black tea (extract)	Intervention	19	Plasma	LC-MS	Yes	(-)-Catechin, (-)-epicatechin, (-)-(epi)catechin gallate, (-)-epigallocatechin, (-)-epigallocatechin gallate, isorhamnetin, 3/4-O-methylgallic acid, 5-(3',4'-dihydroxyphenyl)-γ-valerolactone, 5-(3'-methoxy-4'-	[57]

						hydroxyphenyl)-γ-valerolactone		
Black tea (extract)	Intervention	12	Plasma	GC-MS	Yes	4-O-Methylgallic acid, gallic acid, hippuric acid, pyrogallol		[58]
Green tea, black tea	Intervention	93	Urine	LC-ECD	Yes	(-)-Epicatechin, (-)-epigallocatechin, 4'-O-methylepigallocatechin		[62]
Black tea	Observational	232	Urine (24h)	GC-MS	Yes	4-*O*-Methylgallic acid	24-HDR	[134]
Black tea	Observational	53	Urine (24h and spot)	LC-MS	Yes	Gallic acid, 4-O-methylgallic acid	2-Day dietary record	[26]
Green tea	Observational	968	Urine	LC-MS	Yes	Epicatechin, (-)-epigallocatechin, 4'-O-methyl-epigallocatechin, 5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone, 5-(3',4'-dihydroxyphenyl)-γ-valerolactone	FFQ	[71]
Green tea, black tea	Observational	119	Urine (24h)	LC-MS	Yes	Kaempferol	FFQ, 4-day food diary	[75]
Green tea	Observational	660	Urine (spot, non-fasting)	HPLC	Yes	Epicatechin, (-)-epigallocatechin, 4'-O-methylepigallocatechin, 5-(3',4',5'-trihydroxyphenyl)-γ-valerolactone, 5-(3',4'-dihydroxyphenyl)-γ-valerolactone	FFQ	[72]
Tea	Observational	191	Urine (24h, overnight)	LC-MS	Yes	Kaempferol	FFQ	[135]
Tea	Observational	476	Urine (24h)	LC-MS	No	Methyl(epi)catechin sulfate, hydroxyphenylvaleric acid glucuronide, dihydroxyphenyl-γ-valerolactone glucuronide & sulfate, 4-O-methylgallic acid, methylgallic acid sulfate, pyrogallol sulfate	24-HDR	[30]

Table 3. Biomarkers of intake for sugar sweetened beverages measured in observational studies.

Study design	No. subjects	Analytical method	Sample type	Discriminating metabolites / candidate biomarkers	Association with	Reference
Observational	144	IRMS	Serum	δ13C	24-HDR (x2)	[86]
Observational	60	IRMS	Blood	δ13C	4-day dietary record	[85]
Observational	68	GC-IRMS	Red blood cells	δ13C of alanine	24-HDR (x2)	[84]
Observational	68	IRMS	Red blood cells, plasma, hair	δ13C	24-HDR	[87]
Observational	68	IRMS	Fasting plasma	δ13C of glucose	24-HDR	[87]
Observational	257	IRMS	Blood	δ13C	24-HDR (x3), 4-day dietary record	[91]
Observational	565	1H-NMR	Urine	Formate, citrulline, taurine, isocitrate	4-day dietary record	[95]

Table 4. Low-calorie sweeteners approved for use in the European Union.

Intense sweeteners	E-number	Sweetness*	ADI (mg/kg BW)	Year of approval
Saccharin & its salts	E954	300-500	0-5	1977
Aspartame	E951	180-200	0-40	1984
Acesulfame-K	E950	200	0-9	1984
Cyclamates	E952	30	0-7	1984
Thaumatin	E957	2000-3000	No ADI	1984
NHDC	E959	1900	0-5	1988
Aspartame-acesulfame salt	E962	350	See aspartame & acesulfame-K	2000
Sucralose	E955	600	0-15	2000
Steviol glycosides	E960	300	0-4	2011
Advantame	E969	37000	0-5	2014

NHDC, neohesperidine dihydrochalcone; ADI, acceptable daily intake; BW, body weight.
* Relative to sucrose.

Table 5. Metabolic fates of low-calorie sweeteners approved in the European Union (adapted from [110]).

Sweetener (CAS Registry No.)	Metabolic fate	Route(s) of excretion*	References
Saccharin (81-07-2)	Not metabolised, excreted unchanged.	Urine	[112, 113]
Acesulfame-K (55589-62-3)	Not metabolised, excreted unchanged.	Urine	[111]
Aspartame (22839-47-0)	Hydrolysed to aspartic acid, phenylalanine and methanol.	N/A	[124, 125]
Cyclamate (139-05-9)	80% of the population do not metabolise cyclamate. In 20%, it undergoes partial hydrolysis in the gut to cyclohexylamine. Extent of hydrolysis vary between and within individuals.	Faeces, urine	[116]
Thaumatin (53850-34-3)	Undergoes normal protein digestion.	N/A	[126]
NHDC (20702-77-6)	Metabolised by gut microflora to similar metabolites to naturally occurring flavonoids.	Urine	[127]
Salt of aspartame-acesulfame (106372-55-8)	Dissociates to individual sweeteners in digestive fluids and undergoes same metabolic fates.	See information for acesulfame-K and aspartame	[136]
Sucralose (56038-13-2)	Not metabolised, excreted mainly unchanged but 2% of absorbed portion excreted as conjugates.	Faeces, urine	[117, 118]
Steviol glycosides†	Bacterial hydrolysis in the gut to steviol which is then absorbed and excreted as steviol glucuronide.	Urine	[120-122]
Advantam (714229-20-6)	Converted to advantame acid and mainly excreted as such with 2 minor metabolites.	Faeces, urine	[123]

CAS, Chemical Abstract Service; NHDC, neohesperidine dihydrochalcone; N/A, not applicable as broken down to normal dietary components; JECFA, Joint FAO/WHO Expert Committee on Food Additives.

* Principal route of excretion listed.

† No CAS Registry No available.

References

1. Drewnowski A, Rehm CD, Constant F (2013) Water and beverage consumption among adults in the United States: cross-sectional study using data from NHANES 2005-2010. BMC Public Health 13.

2. Rosner B, Willett WC, Spiegelman D (1989) Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 8: 1051-1069; discussion 1071-1053.

3. Potischman N, Freudenheim JL (2003) Biomarkers of Nutritional Exposure and Nutritional Status: An Overview. J Nutr 133: 873S-874S.

4. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA (2009) Biomarkers in nutritional epidemiology: applications, needs and new horizons. Hum Genet 125: 507-525.

5. Gao Q, Scalbert A, Vergeres G, Kolehmainen M, Manach C, Brennan L, Afman LA, Wishart D, Andres-Lacueva C, Garcia-Aloy M, et al. (2017) A scheme for a flexible classification of food and health biomarkers. Genes Nutr: to be submitted.

6. Practico G, Gao Q, Scalbert A, Vergeres G, Kolehmainen M, Manach C, Brennan L, Pedapati SH, Afman LA, Wishart D, et al. (2017) Food Intake biomarker – Guidelines for systematic REview (FIGuRE): Allium vegetable intake biomarkers as an example. submitted.

7. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6.

8. Ludwig IA, Clifford MN, Lean MEJ, Ashihara H, Crozier A (2014) Coffee: biochemistry and potential impact on health. Food & Function 5: 1695-1717.

9. Noordzij M, Uiterwaal CSPM, Arends L, Kok F, Grobbee D, Geleijnse JM (2005) Blood pressure response to chronic intake of coffee and caffeine: a meta-analysis of randomized controlled trialsJ Hypertens, pp. 921-928.

10. Cornelis MC, El-Sohemy A (2007) Coffee, caffeine, and coronary heart disease. Curr Opin Clin Nutr Metab Care 10: 745-751.

11. Van Dam RM, Scalbert A (2008) Dietary Polyphenols, Urinary Biomarkers and Risk of Type 2 Diabetes, pp. 82 pp.

12. van Dam RM, Willett WC, Manson JE, Hu FB (2006) Coffee, caffeine, and risk of type 2 diabetes - A prospective cohort study in younger and middle-aged US women. Diabetes Care 29: 398-403.

13. Zhang R, Wang Y, Song B, Jørgensen HS, Xu Y (2012) Coffee consumption and risk of stroke: a meta-analysis of cohort studies. Central European Journal of Medicine 7: 310-316.

14. Butt MS, Sultan MT (2011) Coffee and its Consumption: Benefits and Risks. Crit Rev Food Sci Nutr 51: 363 - 373.

15. Aleksandrova K, Bamia C, Drogan D, Lagiou P, Trichopoulou A, Jenab M, Fedirko V, Romieu I, Bueno-de-Mesquita HB, Pischon T, et al. (2015) The association of coffee intake with liver cancer risk is mediated by biomarkers of inflammation and hepatocellular injury: data from the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr 102: 1498-1508.

16. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q, Huang W-Y, Xiong X, Freedman ND, Cross AJ, et al. (2015) Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. The American Journal of Clinical Nutrition 101: 1000-1011.

17. Perez-Jimenez J, Neveu V, Vos F, Scalbert A (2010) Identification of the 100 richest dietary sources of polyphenols: an application of the Phenol-Explorer database. Eur J Clin Nutr 64 Suppl 3: S112-120.

18. Rechner AR, Spencer JP, Kuhnle G, Hahn U, Rice-Evans CA (2001) Novel biomarkers of the metabolism of caffeic acid derivatives in vivo. Free Radic Biol Med 30: 1213-1222.

19. Ito H, Gonthier M-P, Manach C, Morand C, Mennen L, Rémésy C, Scalbert A (2005) Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. Br J Nutr 94: 500-509.

20. Stalmach A, Mullen W, Barron D, Uchida K, Yokota T, Cavin C, Steiling H, Williamson G, Crozier A (2009) Metabolite Profiling of Hydroxycinnamate Derivatives in Plasma and Urine after the Ingestion of Coffee by Humans: Identification of Biomarkers of Coffee Consumption. Drug Metab Dispos 37: 1749-1758.

21. Stalmach A, Williamson G, Crozier A (2014) Impact of dose on the bioavailability of coffee chlorogenic acids in humans. Food Funct 5: 1727-1737.

22. Lang R, Dieminger N, Beusch A, Lee YM, Dunkel A, Suess B, Skurk T, Wahl A, Hauner H, Hofmann T (2013) Bioappearance and pharmacokinetics of bioactives upon coffee consumption. Anal Bioanal Chem 405: 8487-8503.

23. Heinzmann SS, Holmes E, Kochhar S, Nicholson JK, Schmitt-Kopplin P (2015) 2-Furoylglycine as a candidate biomarker of coffee consumption. J Agric Food Chem 63: 8615-8621.

24. Clifford MN (1999) Chlorogenic acids and other cinnamates - Nature, occurence and dietary burden. J Sci Food Agric 79: 362-372.

25. Hodgson JM, Chan SY, Puddey IB, Devine A, Wattanapenpaiboon N, Wahlqvist ML, Lukito W, Burke V, Ward NC, Prince RL, et al. (2004) Phenolic acid metabolites as biomarkers for tea- and coffee-derived polyphenol exposure in human subjects. Br J Nutr 91: 301-306.

26. Mennen L, Sapinho D, Ito H, Galan P, Hercberg S, Scalbert A (2006) Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. Br J Nutr 96: 191-198.

27. Zamora-Ros R, Achaintre D, Rothwell JA, Rinaldi S, Assi N, Ferrari P, Leitzmann M, Boutron-Ruault M-C, Fagherazzi G, Auffret A, et al. (2016) Urinary excretions of 34 dietary polyphenols and their associations with lifestyle factors in the EPIC cohort study. Scientific Reports 6: 26905.

28. Lang R, Wahl A, Stark T, Hofmann T (2011) Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. Mol Nutr Food Res 55: 1613-1623.

29. Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J (2013) Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. Am J Clin Nutr 97: 377-389.

30. Edmands WM, Ferrari P, Rothwell JA, Rinaldi S, Slimani N, Barupal DK, Biessy C, Jenab M, Clavel-Chapelon F, Fagherazzi G, et al. (2015) Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. Am J Clin Nutr 102: 905-913.

31. Rothwell JA, Fillâtre Y, Martin J-F, Lyan B, Pujos-Guillot E, Fezeu L, Hercberg S, Comte B, Galan P, Touvier M, et al. (2014) New Biomarkers of Coffee Consumption Identified by the Non-Targeted Metabolomic Profiling of Cohort Study Subjects. PLoS ONE 9: e93474.

32. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ (2014) Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. The American Journal of Clinical Nutrition 100: 208-217.

33. Kuwayama K, Tsujikawa K, Miyaguchi H, Kanamori T, Iwata YT, Inoue H (2013) Time-course measurements of caffeine and its metabolites extracted from fingertips after coffee intake: a preliminary study for the detection of drugs from fingerprints. Anal Bioanal Chem 405: 3945-3952.

34. Berchtold C, Meier L, Steinhoff R, Zenobi R (2013) A new strategy based on real-time secondary electrospray ionization and high-resolution mass

spectrometry to discriminate endogenous and exogenous compounds in exhaled breath. Metabolomics 10: 291-301.

35. Neveu V, Perez-Jiménez J, Vos F, Crespy V, du Chaffaut L, Mennen L, Knox C, Eisner R, Cruz J, Wishart D, et al. (2010) Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. Database 2010: bap024.

36. U.S. Department of Agriculture ARS (2009) USDA National Nutrient Database for Standard Reference, Release 28. Nutrient Data Laboratory Home Page, http://www.ars.usda.gov/nutrientdata.

37. Sang SM, Lambert JD, Ho CT, Yang CS (2011) The chemistry and biotransformation of tea constituents. Pharmacol Res 64: 87-99.

38. Yang CS, Hong J (2013) Prevention of Chronic Diseases by Tea: Possible Mechanisms and Human Relevance. In: Cousins RJ (ed) Annual Review of Nutrition, Vol 33. pp. 161-181.

39. Yang W-S, Wang W-Y, Fan W-Y, Deng Q, Wang X (2014) Tea consumption and risk of type 2 diabetes: a dose-response meta-analysis of cohort studies. Br J Nutr 111: 1329-1339.

40. Onakpoya I, Spencer E, Heneghan C, Thompson M (2014) The effect of green tea on blood pressure and lipid profile: A systematic review and meta-analysis of randomized clinical trials. Nutrition Metabolism and Cardiovascular Diseases 24: 823-836.

41. Yuan J-M, Sun C, Butler LM (2011) Tea and cancer prevention: Epidemiological studies. Pharmacol Res 64: 123-135.

42. Sang S, Lambert JD, Ho CT, Yang CS (2011) The chemistry and biotransformation of tea constituents. Pharmacol Res 64: 87-99.

43. Stalmach A, Troufflard S, Serafini M, Crozier A (2009) Absorption, metabolism and excretion of Choladi green tea flavan-3-ols by humans. Mol Nutr Food Res 53 Suppl 1: S44-53.

44. Calani L, Del Rio D, Luisa Callegari M, Morelli L, Brighenti F (2012) Updated bioavailability and 48 h excretion profile of flavan-3-ols from green tea in humans. Int J Food Sci Nutr 63: 513-521.

45. Nakagawa K, Okuda S, Miyazawa T (1997) Dose-dependent incorporation of tea catechins, (-)-epigallocatechin-3-gallate and (-)-epigallocatechin, into human plasma. Biosci Biotechnol Biochem 61: 1981-1985.

46. August DA, Landau J, Caputo D, Hong J, Lee MJ, Yang CS (1999) Ingestion of green tea rapidly decreases prostaglandin E2 levels in rectal mucosa in humans. Cancer Epidemiol Biomarkers Prev 8: 709-713.

47. Kimura M, Umegaki K, Kasuya Y, Sugisawa A, Higuchi M (2002) The relation between single/double or repeated tea catechin ingestions and plasma antioxidant activity in humans. Eur J Clin Nutr 56: 1186-1193.

48. Luo H, Cox SB, Gao W, Yu J, Tang L, Wang J-S (2006) Metabolic profiling in validation of plasma biomarkers for green tea polyphenols. Metabolomics 2: 235-241.

49. Wu AH, Spicer D, Stanczyk FZ, Tseng C-C, Yang CS, Pike MC (2012) Effect of 2-Month Controlled Green Tea Intervention on Lipoprotein Cholesterol, Glucose, and Hormone Levels in Healthy Postmenopausal Women. Cancer Prev Res (Phila Pa) 5: 393-402.

50. van het Hof KH, Wiseman SA, Yang CS, Tijburg LB (1999) Plasma and lipoprotein levels of tea catechins following repeated tea consumption. Proc Soc Exp Biol Med 220: 203-209.

51. Renouf M, Marmet C, Guy PA, Beaumont M, Lepage M, Williamson G, Dionisi F (2013) Dose-response plasma appearance of green tea catechins in adults. Mol Nutr Food Res 57: 833-839.

52. Daykin CA, Van Duynhoven JP, Groenewegen A, Dachtler M, Van Amelsvoort JM, Mulder TP (2005) Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. J Agric Food Chem 53: 1428-1434.

53. Van Dorsten FA, Daykin CA, Mulder TPJ, Van Duynhoven JPM (2006) Metabonomics approach to determine metabolic differences between green tea and black tea consumption. J Agric Food Chem 54: 6929-6938.

54. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Grun CH, Jacobs DM, Duchateau GS, Vis DJ, Smilde AK (2009) Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. J Proteome Res 8: 3317-3330.

55. van Duynhoven J, van der Hooft JJJ, van Dorsten FA, Peters S, Foltz M, Gomez-Roldan V, Vervoort J, de Vos RCH, Jacobs DM (2014) Rapid and Sustained Systemic Circulation of Conjugated Gut Microbial Catabolites after Single-Dose Black Tea Extract Consumption. J Proteome Res 13: 2668-2678.

56. Shahrzad S, Aoyagi K, Winter A, Koyama A, Bitsch I (2001) Pharmacokinetics of gallic acid and its relative bioavailability from tea in healthy humans. J Nutr 131: 1207-1210.

57. van Velzen EJJ, Westerhuis JA, Grun CH, Jacobs DM, Eilers PHC, Mulder TP, Foltz M, Garczarek U, Kemperman R, Vaughan EE, et al. (2014) Population-based nutrikinetic modeling of polyphenol exposure. Metabolomics 10: 1059-1073.

58. van der Pijl PC, Foltz M, Glube ND, Peters S, Duchateau GSMJE (2015) Pharmacokinetics of black tea-derived phenolic acids in plasma. J Funct Foods 17: 667-675.

59. Hodgson JM, Morton LW, Puddey IB, Beilin LJ, Croft KD (2000) Gallic acid metabolites are markers of black tea intake in humans. J Agric Food Chem 48: 2276-2280.

60. Hakim IA, Harris RB, Brown S, Chow HH, Wiseman S, Agarwal S, Talbot W (2003) Effect of increased tea consumption on oxidative DNA damage among smokers: a randomized controlled study. J Nutr 133: 3303s-3309s.

61. Henning SM, Niu Y, Lee NH, Thames GD, Minutti RR, Wang H, Go VL, Heber D (2004) Bioavailability and antioxidant activity of tea flavanols after

consumption of green tea, black tea, or a green tea extract supplement. Am J Clin Nutr 80: 1558-1564.

62. Henning SM, Wang P, Said JW, Huang M, Grogan T, Elashoff D, Carpenter CL, Heber D, Aronson WJ (2015) Randomized clinical trial of brewed green and black tea in men with prostate cancer prior to prostatectomy. Prostate 75: 550-559.

63. Mulder TP, van Platerink CJ, Wijnand Schuyl PJ, van Amelsvoort JM (2001) Analysis of theaflavins in biological fluids using liquid chromatography-electrospray mass spectrometry. J Chromatogr B Biomed Sci Appl 760: 271-279.

64. Urpi-Sarda M, Monagas M, Khan N, Lamuela-Raventos RM, Santos-Buelga C, Sacanella E, Castell M, Permanyer J, Andres-Lacueva C (2009) Epicatechin, procyanidins, and phenolic microbial metabolites after cocoa intake in humans and rats. Anal Bioanal Chem 394: 1545-1556.

65. Krupp D, Doberstein N, Shi L, Remer T (2012) Hippuric Acid in 24-Hour Urine Is a Potential Biomarker for Fruit and Vegetable Consumption in Healthy Children and Adolescents. The Journal of Nutrition 142: 1314-1320.

66. Madrid-Gambin F, Garcia-Aloy M, Vázquez-Fresno R, Vegas-Lozano E, de Villa Jubany MCR, Misawa K, Hase T, Shimotoyodome A, Andres-Lacueva C (2016) Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects. Food Research International: doi: 10.1016/j.foodres.2016.1003.1038.

67. Del Rio D, Calani L, Cordero C, Salvatore S, Pellegrini N, Brighenti F (2010) Bioavailability and catabolism of green tea flavan-3-ols in humans. Nutrition 26: 1110-1116.

68. Perez-Jimenez J, Hubert J, Ashton K, Hooper L, Cassidy A, Manach C, Williamson G, Scalbert A (2010) Urinary metabolites as biomarkers of polyphenol intake in humans - A systematic review. Am J Clin Nutr 92: 801-809.

69. Hertog MGL, Hollman PCH, Katan MB, Kromhout D (1993) Intake of potentially anticarcinogenic flavonoids and their determinants in adults in The Netherland. Nutr Cancer 20: 21-29.

70. Bispo MS, Veloso MC, Pinheiro HL, De Oliveira RF, Reis JO, De Andrade JB (2002) Simultaneous determination of caffeine, theobromine, and theophylline by high-performance liquid chromatography. J Chromatogr Sci 40: 45-48.

71. Yuan JM, Gao YT, Yang CS, Yu MC (2007) Urinary biomarkers of tea polyphenols and risk of colorectal cancer in the Shanghai Cohort Study. Int J Cancer 120: 1344-1350.

72. Inoue-Choi M, Yuan JM, Yang CS, Van Den Berg DJ, Lee MJ, Gao YT, Yu MC (2010) Genetic Association Between the COMT Genotype and Urinary Levels of Tea Polyphenols and Their Metabolites among Daily Green Tea Drinkers. Int J Mol Epidemiol Genet 1: 114-123.

73. Tulipani S, Llorach R, Jáuregui O, López-Uriarte P, Garcia-Aloy M, Bullo M, Salas-Salvadó J, Andrés-Lacueva C (2011) Metabolomics Unveils Urinary Changes in Subjects with Metabolic Syndrome following 12-Week Nut Consumption. J Proteome Res 10: 5047-5058.

74. Pimpao RC, Ventura MR, Ferreira RB, Williamson G, Santos CN (2015) Phenolic sulfates as new and highly abundant metabolites in human plasma after ingestion of a mixed berry fruit puree. Br J Nutr 113: 454-463.

75. Brantsaeter AL, Haugen M, Rasmussen SE, Alexander J, Samuelsen SO, Meltzer HM (2007) Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). Public Health Nutr 10: 838-847.

76. de Vries JH, Hollman PC, Meyboom S, Buysman MN, Zock PL, van Staveren WA, Katan MB (1998) Plasma concentrations and urinary excretion of the antioxidant flavonols quercetin and kaempferol as biomarkers for dietary intake. Am J Clin Nutr 68: 60-65.

77. Meng X, Lee MJ, Li C, Sheng S, Zhu N, Sang S, Ho CT, Yang CS (2001) Formation and identification of 4'-o-methyl-(-)-epigallocatechin in humans. Drug Metab Dispos 29: 789-793.

78. Manach C, Williamson G, Morand C, Scalbert A, Remesy C (2005) Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. Am J Clin Nutr 81: 230S-242S.

79. Landberg R, Aman P, Friberg LE, Vessby B, Adlercreutz H, Kamal-Eldin A (2009) Dose response of whole-grain biomarkers: alkylresorcinols in human plasma and their metabolites in urine in relation to intake. Am J Clin Nutr 89: 290-296.

80. Linko-Parvinen AM, Landberg R, Tikkanen MJ, Adlercreutz H, Penalvo JL (2007) Alkylresorcinols from whole-grain wheat and rye are transported in human plasma lipoproteins. J Nutr 137: 1137-1142.

81. Bleich SN, Wang YC, Wang Y, Gortmaker SL (2009) Increasing consumption of sugar-sweetened beverages among US adults: 1988-1994 to 1999-2004. Am J Clin Nutr 89: 372-381.

82. Bremer AA, Auinger P, Byrd RS (2009) Relationship between insulin resistance-associated metabolic parameters and anthropometric measurements with sugar-sweetened beverage intake and physical activity levels in US adolescents: findings from the 1999-2004 National Health and Nutrition Examination Survey. Arch Pediatr Adolesc Med 163: 328-335.

83. Wang YC, Ludwig DS, Sonneville K, Gortmaker SL (2009) Impact of change in sweetened caloric beverage consumption on energy intake among children and adolescents. Arch Pediatr Adolesc Med 163: 336-343.

84. Choy K, Nash SH, Kristal AR, Hopkins S, Boyer BB, O'Brien DM (2013) The Carbon Isotope Ratio of Alanine in Red Blood Cells Is a New Candidate Biomarker of Sugar-Sweetened Beverage Intake. J Nutr 143: 878-884.

85. Davy BM, Jahren AH, Hedrick VE, Comber DL (2011) Association of delta C-13 in Fingerstick Blood with Added-Sugar and Sugar-Sweetened Beverage Intake. J Am Diet Assoc 111: 874-878.

86. Fakhouri TH, Jahren AH, Appel LJ, Chen L, Alavi R, Anderson CA (2014) Serum carbon isotope values change in adults in response to changes in sugar-sweetened beverage intake. J Nutr 144: 902-905.

87. Nash SH, Kristal AR, Hopkins SE, Boyer BB, O'Brien DM (2014) Stable isotope models of sugar intake using hair, red blood cells, and plasma, but not fasting plasma glucose, predict sugar intake in a Yup'ik study population. J Nutr 144: 75-80.

88. Schoeller DA (2013) A Novel Carbon Isotope Biomarker for Dietary Sugar. J Nutr 143: 763-765.

89. Jahren AH, Bostic JN, Davy BM (2014) The potential for a carbon stable isotope biomarker of dietary sugar intake. Journal of Analytical Atomic Spectrometry 29: 795-816.

90. European Starch Industry Association (2015) Facts about sugar, http://www.starch.eu/wp-content/uploads/2015/05/Facts-about-sugars-Brochure.pdf, accessed 13.06.2016.

91. Hedrick VE, Zoellner JM, Jahren AH, Woodford NA, Bostic JN, Davy BM (2015) A Dual-Carbon-and-Nitrogen Stable Isotope Ratio Model Is Not Superior to a Single-Carbon Stable Isotope Ratio Model for Predicting Added Sugar Intake in Southwest Virginian Adults. J Nutr 145: 1362-1369.

92. Nash SH, Kristal AR, Bersamin A, Hopkins SE, Boyer BB, O'Brien DM (2013) Carbon and nitrogen stable isotope ratios predict intake of sweeteners in a Yup'ik study population. J Nutr 143: 161-165.

93. Tasevska N (2015) Urinary Sugars-A Biomarker of Total Sugars Intake. Nutrients 7: 5816-5833.

94. Tasevska N, Runswick SA, McTaggart A, Bingham SA (2005) Urinary sucrose and fructose as biomarkers for sugar consumption. Cancer Epidemiology Biomarkers & Prevention 14: 1287-1294.

95. Gibbons H, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, Brennan L (2015) A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. Am J Clin Nutr 101: 471-477.

96. Lourenco R, Camilo ME (2002) Taurine: a conditionally essential amino acid in humans? An overview in health and disease. Nutr Hosp 17: 262-270.

97. Scientific_Advisory_Committee_on_Nutrition_(SACN) (2015) Carbohydrates and Health, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445503/SACN_Carbohydrates_and_Health.pdf, accessed Last accessed July 2016.

98. Nettleton JA, Lutsey PL, Wang Y, Lima JA, Michos ED, Jacobs DR, Jr. (2009) Diet soda intake and risk of incident metabolic syndrome and type 2 diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). Diabetes Care 32: 688-694.

99. Lin J, Curhan GC (2011) Associations of sugar and artificially sweetened soda with albuminuria and kidney function decline in women. Clin J Am Soc Nephrol 6: 160-166.

100. Crichton G, Alkerwi Aa, Elias M (2015) Diet Soft Drink Consumption is Associated with the Metabolic Syndrome: A Two Sample Comparison. Nutrients 7: 3569-3586.

101. Souza Bda S, Cunha DB, Pereira RA, Sichieri R (2016) Soft drink consumption, mainly diet ones, is associated with increased blood pressure in adolescents. J Hypertens 34: 221-225.

102. de Koning L, Malik VS, Rimm EB, Willett WC, Hu FB (2011) Sugar-sweetened and artificially sweetened beverage consumption and risk of type 2 diabetes in men. Am J Clin Nutr 93: 1321-1327.

103. Imamura F, O'Connor L, Ye Z, Mursu J, Hayashino Y, Bhupathiraju SN, Forouhi NG (2015) Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ : British Medical Journal 351: h3576.

104. Christiani DC, Chang SH, Chun BC, Lee WJ (2000) Urinary excretion of hippuric acid after consumption of nonalcoholic beverages. Int J Occup Environ Health 6: 238-242.

105. Choi JW, Ford ES, Gao X, Choi HK (2008) Sugar-sweetened soft drinks, diet soft drinks, and serum uric acid level: the Third National Health and Nutrition Examination Survey. Arthritis Rheum 59: 109-116.

106. Bellet S, Kershbaum A, Roman L (1968) Effect of cola drinks on serum free fatty acids. Arch Environ Health 17: 803-806.

107. Ahmed N, Mirshekar-Syahkal B, Kennish L, Karachalias N, Babaei-Jadidi R, Thornalley PJ (2005) Assay of advanced glycation endproducts in selected beverages and food by liquid chromatography with tandem mass spectrometric detection. Mol Nutr Food Res 49: 691-699.

108. Bendig P, Maier L, Vetter W (2012) Brominated vegetable oil in soft drinks – an underrated source of human organobromine intake. Food Chem 133: 678-682.

109. Eisner BH, Asplin JR, Goldfarb DS, Ahmad A, Stoller ML (2010) Citrate, malate and alkali content in commonly consumed diet sodas: implications for nephrolithiasis treatment. J Urol 183: 2419-2423.

110. Logue C, Dowey LC, Strain JJ, Verhagen H, Gallagher AM (2016) The potential application of a biomarker approach for the investigation of low-calorie sweetener exposure. Proc Nutr Soc 75: 216-225.

111. Christ O, Rupp W (1976) Human experiments with Acetosulfam-14C. Pharmacokinetics after oral administration of 30mg to three healthy male probands. Unpublished report,

http://www.inchem.org/documents/jecfa/jecmono/v28je13.htm accessed Last accessed June 2015.

112. Byard JL, McChesney EW, Golberg L, Coulston F (1974) Excretion and metabolism of saccharin in man. II. Studies with 14-C-labelled and unlabelled saccharin. Food Cosmet Toxicol 12: 175-184.

113. Ball LM, Renwick AG, Williams RT (1977) The fate of [14C]saccharin in man, rat and rabbit and of 2-sulphamoyl[14C]benzoic acid in the rat. Xenobiotica 7: 189-203.

114. Sweatman TW, Renwick AG, Burgess CD (1981) The pharmacokinetics of saccharin in man. Xenobiotica 11: 531-540.

115. Wilson LA, Wilkinson K, Crews HM, Davies AM, Dick CS, Dumsday VL (1999) Urinary monitoring of saccharin and acesulfame-K as biomarkers of exposure to these additives. Food Addit Contam 16: 227-238.

116. Renwick AG, Thompson JP, O'Shaughnessy M, Walter EJ (2004) The metabolism of cyclamate to cyclohexylamine in humans during long-term administration. Toxicol Appl Pharmacol 196: 367-380.

117. Grice HC, Goldsmith LA (2000) Sucralose--an overview of the toxicity data. Food Chem Toxicol 38 Suppl 2: S1-6.

118. Roberts A, Renwick AG, Sims J, Snodin DJ (2000) Sucralose metabolism and pharmacokinetics in man. Food Chem Toxicol 38 Suppl 2: S31-41.

119. Bopp BA, Sonders RC, Kesterson JW (1986) Toxicological aspects of cyclamate and cyclohexylamine. CRC Crit Rev Toxicol 16: 213-306.

120. Geuns JM, Buyse J, Vankeirsbilck A, Temme EH (2007) Metabolism of stevioside by healthy subjects. Exp Biol Med (Maywood) 232: 164-173.

121. Geuns JM, Buyse J, Vankeirsbilck A, Temme EH, Compernolle F, Toppet S (2006) Identification of steviol glucuronide in human urine. J Agric Food Chem 54: 2794-2798.

122. Wheeler A, Boileau AC, Winkler PC, Compton JC, Prakash I, Jiang X, Mandarino DA (2008) Pharmacokinetics of rebaudioside A and stevioside after single oral doses in healthy men. Food Chem Toxicol 46 Suppl 7: S54-60.

123. Ubukata K, Nakayama A, Mihara R (2011) Pharmacokinetics and metabolism of N-[N-[3-(3-hydroxy-4-methoxyphenyl) propyl]-alpha-aspartyl]-L-phenylalanine 1-methyl ester, monohydrate (advantame) in the rat, dog, and man. Food Chem Toxicol 49 Suppl 1: S8-29.

124. Butchko HH, Stargel WW, Comer CP, Mayhew DA, Benninger C, Blackburn GL, de Sonneville LM, Geha RS, Hertelendy Z, Koestner A, et al. (2002) Aspartame: review of safety. Regul Toxicol Pharmacol 35: S1-93.

125. Magnuson BA, Burdock GA, Doull J, Kroes RM, Marsh GM, Pariza MW, Spencer PS, Waddell WJ, Walker R, Williams GM (2007) Aspartame: a safety evaluation based on current use levels, regulations, and toxicological and epidemiological studies. Crit Rev Toxicol 37: 629-727.

126. Joint_FAO/WHO_Expert_Committee_on_Food_Additives_(JECFA) (1985) Thaumatin. WHO Food Additives Series 20, http://www.inchem.org/documents/jecfa/jecmono/v20je15.htm, accessed Last accessed June 2015.

127. Borrego F, Montijano H (2001) Neohesperidine dihydrochalcone. In: O'Brien Nabors L (ed) Alternative Sweeteners, 3rd ed. Marcel Dekker, New York, pp. 87-105.

128. Scalbert A, Rothwell JA, Keski-Rahkonen P, Neveu V (2017) The food metabolome and dietary biomarkers. In: Schoeller DA, Westerterp-Plantenga M (eds) Advances in the Assessment of Dietary Intake. CRC Press, pp. In press.

129. Allard E, Backstrom D, Danielsson R, Sjobberg JR, Bergquist J (2008) Comparing Capillary Electrophoresis - Mass Spectrometry Fingerprints of Urine Samples Obtained after Intake of Coffee, Tea, or Water. Anal Chem 80: 8946-8955.

130. Hodgson JM, Puddey IB, Croft KD, Burke V, Mori TA, Caccetta RA, Beilin LJ (2000) Acute effects of ingestion of black and green tea on lipoprotein oxidation. Am J Clin Nutr 71: 1103-1107.

131. Mulder TP, Rietveld AG, van Amelsvoort JM (2005) Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. Am J Clin Nutr 81: 256S-260S.

132. van der Hooft JJ, de Vos RC, Mihaleva V, Bino RJ, Ridder L, de Roo N, Jacobs DM, van Duynhoven JP, Vervoort J (2012) Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake. Anal Chem 84: 7263-7271.

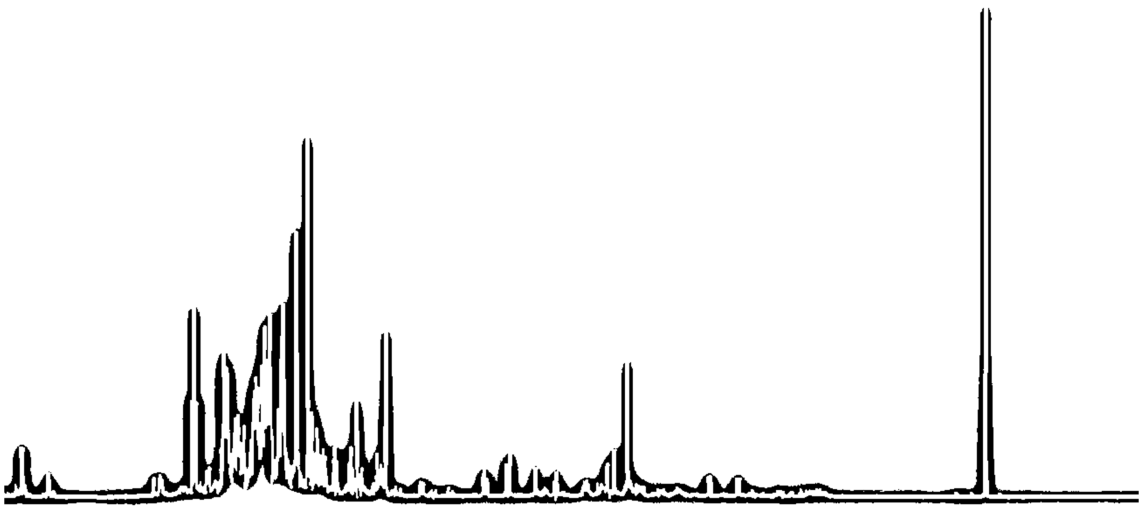
133. Rhodes LE, Darby G, Massey KA, Clarke KA, Dew TP, Farrar MD, Bennett S, Watson RE, Williamson G, Nicolaou A (2013) Oral green tea catechin metabolites are incorporated into human skin and protect against UV radiation-induced cutaneous inflammation in association with reduced production of pro-inflammatory eicosanoid 12-hydroxyeicosatetraenoic acid. Br J Nutr 110: 891-900.

134. Hodgson JM, Devine A, Puddey IB, Beilby J, Prince RL (2006) Drinking tea is associated with lower plasma total homocysteine in older women. Asia Pac J Clin Nutr 15: 253-258.

135. Krogholm KS, Bysted A, Brantsaeter AL, Jakobsen J, Rasmussen SE, Kristoffersen L, Toft U (2012) Evaluation of flavonoids and enterolactone in overnight urine as intake biomarkers of fruits, vegetables and beverages in the Inter99 cohort study using the method of triads. Br J Nutr 108: 1904-1912.

136. European_Commission (2000) Minutes of the 120th Meeting of the Scientific Committee on Food held on 8–9 March 2000 in Brussels, https://ec.europa.eu/food/sites/food/files/safety/docs/sci-com_scf_out57_en.pdf, accessed January 2017.

Resultados



Capítulo 06.

ESTUDIO COMPARATIVO DE LA METODOLOGÍA UTILIZADA EN PLASMA/SUERO

En el marco del proyecto FOODBALL se ha realizado un estudio comparativo de la metodología más utilizada en RMN para el desarrollo de la metabolómica en muestras de sangre (concretamente plasma y suero). Los resultados de este estudio se incluyeron en el Protocolo Normalizado de Trabajo (PNT) general del consorcio del FOODBALL apoyado por la iniciativa "*EU Joint Programming Initiative: A Healthy Diet for a Healthy Life*". El presente trabajo se encuentra en proceso de publicación:

Francisco Madrid-Gambin, Rafael Llorach, Cristina Andres-Lacueva. Metabolomics comparison of plasma treatments by quantitative ^{1}H-NMR profiling: precipitation, ultrafiltration and SPE. [*En proceso de publicación*].

Resumen:

El tratamiento de las muestras de plasma y suero para el posterior análisis metabolómico cuantitativo mediante RMN conlleva una de las cuestiones más desafiantes del procedimiento analítico. El elevado número de proteínas (y fosfolípidos) en muestras de plasma/suero implica un reto importante a la hora de explorar metabolitos de bajo peso molecular. Dado que existen varios procedimientos para el tratamiento de las muestras de plasma/suero para el análisis metabolómico y dado que no existe ningún método internacionalmente aceptado, esta cuestión sigue siendo crítica para futuras investigaciones.

Para el análisis metabolómico se utilizaron muestras de plasma que fueron mezcladas en un *pool* para atenuar la variabilidad biológica. Las muestras se sometieron a tres procedimientos diferentes, que consistieron en la extracción de los metabolitos mediante (i) la precipitación de la fracción proteica con metanol (PR), (ii) la ultrafiltración (UF), y (iii) la extracción en fase sólida (SPE) para la eliminación de proteínas y fosfolípidos; para posteriormente proceder con el análisis metabolómico mediante ^{1}H-RMN. El análisis comparativo entre la preparación de las muestras se

basó en la capacidad de identificar metabolitos (CIM), en la comparación de la concentración absoluta de los metabolitos identificados (ACM), y en el cálculo de un puntaje final considerando también la capacidad de automatización, el tiempo empleado y los costes en las diferentes metodologías.

En relación a la CIM, 39 metabolitos fueron identificados en todos los procedimientos, mientras que el número máximo de 41 metabolitos se identificó después de la PR. La comparación de la ACM mostró que la SPE fue el procedimiento menos efectivo considerando la menor concentración de varios metabolitos después de la extracción; sin embargo, en esta técnica se observó la mayor concentración de triptófano. La PR mostró una concentración intermedia de la mayoría de los compuestos, con un tratamiento mínimo y rápido de las muestras. Sin embargo, en esta técnica hubo una menor concentración en compuestos volátiles tales como acetona y acetoacetato. La ultrafiltración exhibió mayores concentraciones de algunos de los metabolitos identificados, especialmente en los metabolitos volátiles antes mencionados, histidina y citrato. Sin embargo, se encontró un residuo prominente de glicerol proveniente de la membrana de los filtros que cubrió parte de los espectros. Por último, el puntaje final calculado indicó que PR fue la técnica más adecuada teniendo en cuenta los aspectos prácticos estudiados para la metabolómica cuantitativa basada en la RMN.

En conclusión, de acuerdo con los resultados presentados, el uso de la precipitación con metanol para eliminar las proteínas en plasma parece ser un método reproducible y eficaz para realizar análisis cuantitativos metabolómicos mediante RMN.

Metabolomics comparison of plasma sample procedures by quantitative ^{1}H-NMR profiling: precipitation, ultrafiltration and solid phase extraction.

Francisco Madrid-Gambin[1], Rafael Llorach[1,2], Cristina Andres-Lacueva[1,2]*

[1]Biomarkers & Nutrimetabolomics Laboratory. Department of Nutrition, Food Sciences and Gastronomy, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona 08028, Spain.

[2]CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona 08028, Spain.

*E-mail: candres@ub.edu. Tel: +34.934034840. Fax: +34.93403593.

Abbreviations: ACM, absolute concentration of metabolites; CIM, capacity to identify metabolites; FID, free induction decay; KOD, potassium deuteroxide; NMR, nuclear magnetic resonance; PC, principal component, PCA, principal component analysis; SPE, solid-phase extraction; TSP, (3-(trimethylsilyl)-proprionate-2,2,3,3-d_4.

ABSTRACT

Plasma and serum preparation is one of the most critical issues of the metabolomics analytical procedure. Since there are several treatments for protein removal of the plasma/serum samples for quantitative analysis without a universal accepted method, this issue keeps being interesting for further investigations. In this work, plasma samples were subjected to three different procedures consisting in extraction of the metabolites through: (i) protein precipitation with methanol (PR), (ii) ultrafiltration (UF), and (iii) solid phase extraction (SPE); prior to NMR-driven metabolomics analysis. The comparative analysis among procedures was based on the capacity to identify metabolites (CIM), on the comparison of the absolute concentration of those metabolites (ACM), and on the calculation of a final score also considering the capability of automation, time consuming and costs. Concerning CIM, 39 metabolites were identified in all procedures while 41 metabolites were identified after PR. The comparison of the ACM showed that SPE was the less effective procedure considering the lower concentration of several metabolites after the extraction. UF displayed the highest retaining for volatile compounds, while PR exhibited an intermediate concentration for most of the compounds. The final score indicated that PR was the most suitable technique considering all studied aspects. According to the presented results, the use of PR seemed to be an effective method for quantitative NMR-based metabolomics.

1. INTRODUCTION

Human plasma and serum are biological samples highly relevant in metabolomics studies (Psychogios et al. 2011). Blood transports dissolved gases, nutrients, hormones and metabolic wastes, and also plays an important role in the regulation of the body temperature, stabilization of the pH, and the defence against pathogens or toxins, among other functions (Martini and Ober 2006).

The use of technology-driven approaches such as metabolomics provides a comprehensive fingerprinting of low-molecular-weight metabolites present in these biological samples (Scalbert et al. 2014). Metabolomics is still in its rapid development stage; therefore development of its methodology in both nuclear magnetic resonance (NMR) and mass spectrometry (MS) is important for wide range of applications including serum/plasma analysis (Li et al. 2016). Plasma and serum are suitable samples for biomarker discovery and hypothesis generation. A previous study using a metabolomics kit (AbsoluteIDQ™ kit p150, Biocrates Life Sciences AG, Innsbruck, Austria) based on FIA-MS by Yu et al suggests that serum could be more sensitive for biomarker detection compared to plasma, whereas measuring metabolites in plasma may be more reproducible (Yu et al. 2011). However, non-substantial differences could be observed between serum and plasma, regarding the concentration of detected metabolites, in a quantitative ^{1}H-NMR metabolomic study with patients with inflammatory bowel disease (Schicho et al. 2012).

On the other hand, the quantity of proteins and phospholipids that are present in serum/plasma samples implicates a major challenge to explore low-molecular-weight metabolites. The quantification of metabolites remains a difficult subject in metabolomics, and the development of new methods is being required to overcome this challenge (Nagana Gowda et al. 2015a). Proteins and phospholipids display a high number of mostly broad NMR peaks, which overlap with the resonances from the low-molecular-weight metabolites (de Graaf et al. 2015). In addition, they introduce an unstable baseline of spectra, which hinders the identification and quantification of metabolites, and also certain proteins that may interact with some of the metabolites (Nagana Gowda and Raftery 2014; Tiziani et al. 2008). Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (Wang et al. 2003) could partially overcome this issue by a spin–spin relaxation edited NMR approach. However, several metabolites such as lactate, histidine or tyrosine may be bound to plasma/serum proteins (Nicholson and

Gartland 1989). Consequently, they are undervalued and unreliable compounds using this approach.

To solve this issue several approaches for removing serum/plasma proteins have emerged, improving the quantification capability and the number of detected metabolites (Wishart 2008). Physico-chemical methods such as ultrafiltration or solvent extraction can be more robust, repeatable and quantitative than the application of the CPMG pulse sequence (Daykin et al. 2002; Wishart 2008). Ultrafiltration (UF) and protein precipitation using methanol (PR) have been applied on plasma/serum samples for quantification of metabolites in NMR, before the analysis (Nagana Gowda and Raftery 2017). Also solid phase extraction (SPE) for the removal of phospholipids improved the signal response of most non-lipid metabolites of blood fluids, to remove phospholipids that are not extracted, and to be a fast and simple method compatible with automation (Tulipani et al. 2013). Hence, this method may be an interesting approach, which would be worth trying by NMR.

With the background that there are several sample treatments in quantitative plasma analysis and given the importance for reliable analysis, this topic continues to be of significant interest for further investigations and improvements. In the present work we focused in the number of detected metabolites and quantification aspects among the different treatments of the samples. To our knowledge, no study has evaluated these parameters together with other important practical aspects among methodologies. We selected plasma samples considering them as more reproducible than serum (Li et al. 2016). Hence, observed differences should be related to the studied procedures. The aim of the present study was to evaluate the performance of UF, PR and SPE, which are previously published as techniques for protein and/or phospholipid removal, comparing the capacity to identify metabolites (CIM), the absolute concentration of those metabolites (ACM), and a final score considering the capability of automation on-line, time consuming and costs using an ^{1}H-NMR untargeted metabolomic analysis of plasma.

2. MATERIAL AND METHODS

2.1. Biological Samples

Blood samples from 5 fasting healthy donors were collected into heparin (anticoagulant) vials. Plasma was obtained after removal of cells by centrifugation at 1600 g during 15 min at room temperature. Pooled biological samples were used to avoid biological variability in the comparative analysis among sample preparation procedures. Aliquots were stored at −80 °C until analysis and processed as described in the next section.

2.2. Sample preparation

Preparation of plasma samples was carried out separately.

UF was applied based on a published methodology (Nagana Gowda and Raftery 2014). Centrifugal filters (3 kDa cutoff; Amicon Microcon, YM-3; Sigma-Aldrich) were washed with water and centrifuged thrice with 500 µL of MiliQ water at 13 400 g for 20 min, each time. Three 200 µL of plasma samples were then transferred to filter tubes and centrifuged for 20 min at 13 400 g. The filtrates were mixed with a 100 µL solution of phosphate buffer in D_2O (pH 7.0) containing 2.32 mM of TSP. The solutions were made up to 600 µL with D_2O and transferred to 5 mm NMR tubes.

For PR, a procedure based on previously published methodology was applied (Nagana Gowda and Raftery 2014). Three 200 µL of plasma samples were mixed with methanol in 2:1 solvent-to-plasma ratio (v/v), vortexed, and incubated at −20 °C for 20 min. The mixtures were centrifuged at 13 400 g for 30 min to pellet proteins. Supernatants were decanted into fresh vials and dried. The dried samples were mixed with 100 µL of phosphate buffer in D_2O (pH 7.0) containing 2.32 mM of 3-(trimethylsilyl)-proprionate-2,2,3,3-d_4 (TSP). The solutions were made up to 600 µL with D_2O and transferred to 5 mm NMR tubes.

SPE approach was carried out using a method based on a previously published procedure (Tulipani et al. 2013). Three 200 µL of plasma samples were thawed, vortexed, and mixed with acidic solvent extraction, followed by glycerophospholipid SPE, using Ostro 96-well plates with pressure valves (Ostro plates, Waters). Samples were pipetted into the wells, followed by the forceful addition of 1% formic acid in

acetonitrile (3:1 solvent/sample) for in-well protein precipitation. Then, after drying the solvents, samples were mixed with 100 µL of phosphate buffer in D_2O (pH 7.0) containing 2.32 mM of TSP. The solutions were made up to 600 µL with D_2O and then transferred to 5 mm NMR tubes.

2.3. Data acquisition and processing

Spectra were acquired on a Bruker 400 MHz Avance III spectrometer equipped with a cryoprobe using a NOESYPRESAT pulse sequence. An exponential window function was applied to the free induction decay (FID) with a line-broadening factor of 0.3 Hz prior to the Fourier transformation. For each sample, FIDs were collected into 32 K data points (128 scans) with a spectral width of 14 ppm, an acquisition time of 3 s, a relaxation delay of 5 s and a mixing time of 10 ms. Parameters were adjusted for enabling quantification by Chenomx NMR Suite 8.1 profiler (Chenomx Edmonton, Canada) (Chenomx NMR Suite 2017). All the NMR spectra were phased, baseline corrected, and calibrated (TSP, 0.0 ppm) using TopSpin software (version 3.0, Bruker, BioSpin, Germany).

2.4. Profiling and quantitation

For calculating the CIM, the total number of identified metabolites after each procedure was considered (profiling), while the differences related to the ACM were assessed for methodology comparison (quantitation).

Metabolite identification and quantitation of the spectra were performed using Chenomx NMR Suite 8.1 profiler. This software allows the determination of concentrations in overlapped spectral regions and enables the fit of spectral peaks using a standard metabolite library. Further contributions to the proton peak assignment were provided by comparing the chemical shifts with those available in the Human Metabolome Database (http://www.hmdb.ca). In addition, a Pearson's correlation test was performed to assess correlations between signals of the same metabolite.

2.5. Statistical analysis

Data from absolute concentration were submitted to MetaboAnalyst 3.0 (Xia et al. 2015). Data were row-wise normalized (rows were samples) by the sum of the intensities of the spectra and column-wise normalized (columns were metabolites) using Pareto scaling and cube root transformation. Principal component analysis (PCA) score plot and loading biplot were performed on transformed data to detect the presence of outliers and data dispersion among the different sample treatments.

Univariate statistical analyses were performed with R version 3.1.2. ANOVA test was performed for the comparison between methodologies to assess the statistical significances in the ACM identified using the Geepack R package v 1.2-1 (Højsgaard et al. 2006) for. A FDR correction for multiple comparisons was also applied. Statistical significance was considered at a p-value <0.05. For calculating the differences in ACM, UF was considered as control procedure due to the higher retaining of metabolite concentrations (Tiziani et al. 2008).

2.6. Practical aspects

Finally, practical aspects of each procedure was discussed regarding: (i) CIM, sorted by the number of identified metabolites (ii) ACM, sorted by the absolute concentration of significant metabolites; (iii) automation, in which procedures were sorted by the reliable capability of automation on-line; (iv) quickness, in which the score was assigned according to the shorter time for preparing 100 samples; and (v) low-cost, in which procedure was sorted ascending by global costs derived of preparing 100 samples. A final score was obtained derived of the sum of individual scores at each item. Individual scores were based on the suitability of each item (3 points, the most appropriate technique; 2 points, intermediate technique; 1 point, the worst technique).

3. RESULTS AND DISCUSSION

All methods displayed highly resolved ^{1}H NMR spectra for metabolite quantitation. As we expected, PCA showed greater variability between methods than between plasma samples, showing a clear separation inter-treatment (Figure 1). In addition, PR was the method with less dispersion in the first two principal components (81.9 % of total

variability) denoting a high reproducibility. Loading biplot exhibited separation over the PCA mainly as consequence of methanol, glycerol, formate and tryptophan.

3.1. Capacity to identify metabolites

A total number of 41 metabolites were identified in the plasma samples. Figure 2 shows the typical spectra for UF, PR and SPE methods along with the annotations for all 41 metabolites. Thirty-nine metabolites were identified in all procedures while the 41 metabolites were identified after PR, showing the highest CIM in this approach. UF and SPE displayed a total of 40 metabolites, of which tryptophan was not detected after ultrafiltration, and histidine after SPE (Figure 2). This was also observed through the PCA loading biplot, in which tryptophan and histidine moved in opposite directions but still close to PR procedure (Figure 1). The present results suggest that the PR enables a slightly higher number of metabolites than the other ones. Among the measured metabolites, we present 41 identified metabolites against from 37 metabolites in plasma/serum by Wishart and colleagues (Wishart 2008) until the 67 detected compounds presented by Gowda and colleagues (Nagana Gowda et al. 2015b). We attribute these discrepancies to the different volume used in the present work (200 µL) and the differences in the sensitivity related to the equipment.

3.2. Concentration of metabolites

Based on the comprehensive analysis of pooled samples, 17 of the metabolites were statistically different in at least one of the method used (Table 1). In fact, here UF showed a high concentration in the majority of metabolites (Table 1). Due to its efficiency for protein removal, UF has been largely used as in NMR-plasma metabolomics studies (Bathe et al. 2011; Farshidfar et al. 2012; Nagana Gowda et al. 2015a; Psychogios et al. 2011). In this context, Table 2 presents the behavior of the significant metabolites, as well as the p-values. For visualization purposes, UF was taken as reference for comparing the three methods (Table 2).

Using UF, there was a lower concentration of 2-hydroxybutyrate and a total absence of tryptophan, also shown on the spectra in **Figure 2** (aliphatic and aromatic regions, respectively). In another study comparing UF and PR, the authors found equal concentrations of 2-hydroxybutyrate between methods (Gowda and Raftery 2014).

However, in line with our results, they also found much lower concentration of tryptophan after UF. Hence, we hypothesize that tryptophan and other compounds bounded to proteins, require protein degradation (e.g. precipitation with methanol) to be correctly detected in NMR. In addition, the glycerol normally present in filters of UF needs several washes prior the use implicating a high time consuming. Consequently, the concentration of glycerol was higher than the other treatments suggesting an unreliable metabolite coming from plasma. In addition, this metabolite was observed through the PCA loading biplot (Figure 1) indicating a strong influence by ultrafiltration.

PR approach showed a lower ACM for citrate, histidine, acetoacetate and acetone, than UF. The decrease of the volatile compounds (acetoacetate and acetone) could be related to the methanol evaporation step. In line with these findings, Tiziani and co-workers (2008) found a lower concentration of citrate when a mix of methanol and chloroform was used for precipitating the plasmatic proteins. In that study, the signals of alanine (δ1.49), citrate (δ2.52, δ2.69) and myo-inositol (δ3.29, δ3.52, δ3.63) showed a reduced intensity compared to the spectra corresponding to ultrafiltration (Tiziani et al. 2008). Although the mix with chloroform should improve the extraction of lipoproteins and lipids, the final ratio of methanol-plasma was set to 1:1 (v/v). However, for an improved ACM of the metabolites, the minimum ratio methanol-plasma should be set to 2:1 (methanol-to-sample ratio, v/v) (Nagana Gowda et al. 2015a), increasing the spectral quality with higher ratios. Protein precipitation by using a 1:1 methanol-to-sample ratio retains a high level of residual proteins that complicates the identification/quantification.

A high concentration of methanol, shown on the spectra through a singlet at δ3.34 (**Figure 2**), may be related with a residue of the methanol used for the protein precipitation with methanol. Hence, the quantification of this metabolite using this procedure may be unreliable (**Table 2**). In other studies, PR using different ratios of methanol/sample displayed a higher performance over other organic solvents enabling a huge increase in the number of metabolites identified (Gowda and Raftery 2014; Nagana Gowda et al. 2015a; Nagana Gowda and Raftery 2017). Then, the use of a higher ratio of methanol-to-sample could improve the accuracy in the measure of these metabolites. Nevertheless, it should be considered a longer evaporation step, and consequently, the sample stability. In another study, the comparison between ultrafiltration and protein precipitation showed that precipitation with methanol exhibited

superior performance over ultrafiltration in the quantitative recovering of blood metabolites from the protein matrix (Nagana Gowda and Raftery 2014, 2015).

SPE showed the lowest concentration of the most of the significant metabolites, with the exception of tryptophan. These results suggest that the breaking of the protein binding with acetonitrile and formic acid (also observed in PR) enabled the measure of the highest concentration in this type of compounds. However, a prominent singlet at δ8.49 corresponding to formate was present in the aromatic region of the spectra (**Figure 2**). The high concentration of formate is related with a residue of the formic acid used in SPE, instead of meaning a high concentration of the compound as an endogenous metabolite. Hence, it is critical the consideration of this metabolite when this procedure was used. The concentration of creatinine, glycine, lactate, serine, isobutyrate, ethanol, lysine and ornithine was lower than ultrafiltration. Besides, the concentrations of histidine, citrate and tryptophan concentrations were lower than those quantified in all other techniques. Interestingly, the concentration of tryptophan was the highest in the SPE, probably because of a better release from proteins as we previously highlighted. In contrast, other authors found a better retaining of metabolites when SPE was compared to precipitation with methanol or ultrafiltration (Tulipani et al. 2013). The authors showed that the combination of solvent extraction and SPE-mediated removal of phospholipids, prior analysis by MS, was the most suitable procedure for detecting quantitative changes. Particularly, the recovery of highly polar compounds such as acetylcholine, acetyl-l-carnitine, (iso)leucine and phenylalanine was highly accurate. Nevertheless, none of these metabolites were statistically different across the methodologies used in the present study. Further optimization of the procedure may enhance the global accuracy of absolute concentrations in this approach.

3.3. Score of practical aspects

In addition to CIM and ACM, Table 3 summarizes practical aspects considered for a final score. UF required more laborious work due to the need of washing the filters coupled to Eppendorf vials. This also implicated a lower capability of automation. On the other hand, PR is an automatable procedure (Almanza-Aguilera et al. 2017) (to be published). In addition, the washing of filters also required longer time for preparing 100 samples. After 3 washes with MiliQ water a residue of glycerol was still visible on the

spectra (Figure 2). On the other hand, the use of Ostro plates of SPE facilitated the handling of samples. Concerning the CIM and ACM, the score was assigned in relation to the findings observed above. The highest score of CIM was for PR and “medium” score for both UF and SPE because they have the same number of identified metabolites. It is important to note that only 1 additionally metabolite was found in PR compared with the other procedures. ACM displayed higher concentration of several metabolites (especially volatile compounds) after UF compared to PR and SPE. Hence, the highest score was for UF and the lowest for SPE. Concerning the low-cost of procedures, the final costs for analysing 100 samples indicated that PR was the cheapest procedure whereas SPE was the most expensive one. Finally, the sum of each score showed the PR as the most suitable option considering all these aspects.

4. CONCLUSION

In the present study, the three main methodologies already published for plasma/serum treatment were compared to each other, and therefore evaluated to determine the most appropriate methodology. The results suggest that fewer steps in the procedure may be critical for the final concentration of the metabolites. The CIM values showed PR as the technique with the highest number of detected metabolites. However, the need of evaporation in the PR and SPE approaches could modify the final concentration of volatile metabolites such as acetone and acetoacetate (ACM). In addition, PR displayed lower histidine and citrate than UF. Other metabolites such as tryptophan may be undetected without protein degradation with organic solvents such as methanol or acetonitrile. Hence, these types of metabolites should be carefully treated when UF is used. Regarding the SPE, the concentration of several metabolites was generally lower, denoting a less suitable sample preparation technique in NMR. The calculated score showed lower score for UF and SPE due to a longer time for sample preparation and higher costs than PR. In conclusion, the PR for removing the protein fraction denoted an appropriated technique, fast and reproducible, with the compromise of sacrificing part of the concentration of some volatile metabolites.

REFERENCES

Almanza-Aguilera, E., Brunius, C., Bernal-Lopez, M. R., Garcia-Aloy, M., Madrid-Gambin, F., Gomez-Huelgas, et al. (2017). A lifestyle intervention for weight-loss among metabolically healthy obese women alters the plasma metabolome. *To be published.*

Bathe, O. F., Shaykhutdinov, R., Kopciuk, K., Weljie, A. M., McKay, A., Sutherland, F. R., et al. (2011). Feasibility of identifying pancreatic cancer based on serum metabolomics. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *20*(1), 140–7. doi:10.1158/1055-9965.EPI-10-0712

Chenomx NMR Suite. (2017). User guide (2017). http://www.chenomx.com/support/nmrsuite/8.3/#?sampleHandling.dataAcquisition. Accessed 26 May 2017

Daykin, C. A., Foxall, P. J. D., Connor, S. C., Lindon, J. C., & Nicholson, J. K. (2002). The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by (1)H nuclear magnetic resonance spectroscopy. *Analytical biochemistry*, *304*(2), 220–30. doi:10.1006/abio.2002.5637

de Graaf, R. A., Prinsen, H., Giannini, C., Caprio, S., & Herzog, R. I. (2015). Quantification of 1H NMR spectra from human plasma. *Metabolomics*, *11*(6), 1702–1707. doi:10.1007/s11306-015-0828-1

Farshidfar, F., Weljie, A. M., Kopciuk, K., Buie, W. D., Maclean, A., Dixon, E., et al. (2012). Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome medicine*, *4*(5), 42. doi:10.1186/gm341

Gowda, G. A. N., & Raftery, D. (2014). Quantitating Metabolites in Protein Precipitated Serum Using NMR Spectroscopy. *Analytical Chemistry*, *86*(11), 5433–5440. doi:10.1021/ac5005103

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, *15*(2), 1–11. doi:10.18637/jss.v015.i02

Li, N., Song, Y. peng, Tang, H., & Wang, Y. (2016). Recent developments in sample preparation and data pre-treatment in metabonomics research. *Archives of*

Biochemistry and Biophysics, *589*, 4–9. doi:10.1016/j.abb.2015.08.024

Martini, F., & Ober, W. C. (2006). *Fundamentals of anatomy & physiology* (6th ed.). San Francisco: Benjamin Cummings.

Nagana Gowda, G. A., Gowda, Y. N., & Raftery, D. (2015a). Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. *Analytical chemistry*, *87*(1), 706–15. doi:10.1021/ac503651e

Nagana Gowda, G. A., Gowda, Y. N., & Raftery, D. (2015b). Expanding the Limits of Human Blood Metabolite Quantitation Using NMR Spectroscopy. *Analytical Chemistry*, *87*(1), 706–715. doi:10.1021/ac503651e

Nagana Gowda, G. A., & Raftery, D. (2014). Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Analytical chemistry*, *86*(11), 5433–40. doi:10.1021/ac5005103

Nagana Gowda, G. A., & Raftery, D. (2015). Can NMR solve some significant challenges in metabolomics? *Journal of magnetic resonance (San Diego, Calif. : 1997)*, *260*, 144–60. doi:10.1016/j.jmr.2015.07.014

Nagana Gowda, G. A., & Raftery, D. (2017). Recent Advances in NMR-Based Metabolomics. *Analytical Chemistry*, *89*(1), 490–510. doi:10.1021/acs.analchem.6b04420

Nicholson, J. K., & Gartland, K. P. R. (1989). 1H NMR studies on protein binding of histidine, tyrosine and phenylalanine in blood plasma. *NMR in Biomedicine*, *2*(2), 77–82. doi:10.1002/nbm.1940020207

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., et al. (2011). The human serum metabolome. *PloS one*, *6*(2), e16957. doi:10.1371/journal.pone.0016957

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., et al. (2014). The food metabolome: a window over dietary exposure. *American Journal of Clinical Nutrition*, *99*(6), 1286–1308. doi:10.3945/ajcn.113.076133

Schicho, R., Shaykhutdinov, R., Ngo, J., Nazyrova, A., Schneider, C., Panaccione, R., et al. (2012). Quantitative Metabolomic Profiling of Serum, Plasma, and Urine by ^{1}H NMR Spectroscopy Discriminates between Patients with Inflammatory Bowel Disease and Healthy Individuals. *Journal of Proteome Research*, *11*(6), 3344–3357. doi:10.1021/pr300139q

Tiziani, S., Emwas, A.-H., Lodi, A., Ludwig, C., Bunce, C. M., Viant, M. R., & Günther, U. L. (2008). Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, *377*(1), 16–23. doi:10.1016/j.ab.2008.01.037

Tulipani, S., Llorach, R., Urpi-Sarda, M., & Andres-Lacueva, C. (2013). Comparative Analysis of Sample Preparation Methods To Handle the Complexity of the Blood Fluid Metabolome: When Less Is More. *Analytical Chemistry*, *85*(1), 341–348. doi:10.1021/ac302919t

Wang, Y., Bollard, M. E., Keun, H., Antti, H., Beckonert, O., Ebbels, T. M., et al. (2003). Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning 1H nuclear magnetic resonance spectroscopy of liver tissues. *Analytical biochemistry*, *323*(1), 26–32. http://www.ncbi.nlm.nih.gov/pubmed/14622955. Accessed 25 April 2017

Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry*, *27*(3), 228–237. doi:10.1016/j.trac.2007.12.001

Xia, J., Sinelnikov, I. V, Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Research*. doi:10.1093/nar/gkv380

Yu, Z., Kastenmüller, G., He, Y., Belcredi, P., Möller, G., Prehn, C., et al. (2011). Differences between Human Plasma and Serum Metabolite Profiles. *PLoS ONE*, *6*(7), e21230. doi:10.1371/journal.pone.0021230

Table 1. Compounds and absolute concentrations (mg/dL) in plasma pools (mean±SD) using ^{1}H-NMR spectroscopy. Metabolites are sorted alphabetically.

	UF	PR	SPE
2-Hydroxybutyrate *	0.022±0.002^{b}	0.058±0.021^{a}	0.075±0.018^{a}
3-Hydroxybutyrate	0.385±0.053	0.370±0.161	0.290±0.031
3-M-2-oxovalerate	0.070±0.007	0.074±0.035	0.062±0.006
Acetate	0.584±0.357	0.388±0.276	0.114±0.035
Acetoacetate*	0.139±0.034^{a}	0.074±0.033^{b}	0.047±0.012^{b}
Acetone*	0.028±0.007^{a}	0.014±0.002^{b}	0.021±0.004^{ab}
Alanine	0.533±0.065	0.421±0.108	0.466±0.086
Asparagine	0.166±0.030	0.202±0.008	0.172±0.037
Betaine	0.068±0.062	0.026±0.005	0.035±0.010
Carnitine	0.099±0.025	0.105±0.009	0.136±0.019
Choline	0.022±0.004	0.023±0.003	0.022±0.003
Citrate*	0.379±0.050^{a}	0.139±0.024^{b}	0.011±0.018^{c}
Creatine	0.074±0.005	0.083±0.016	0.063±0.007
Creatinine*	0.195±0.023^{a}	0.166±0.014^{a}	0.030±0.004^{b}
Ethanol*	0.061±0.051^{a}	0.110±0.039^{a}	0.012±0.004^{b}
Formate*	0.471±0.409^{b}	0.451±0.121^{b}	8.056±1.282^{a}
Glucose	19.419±3.344	17.871±3.261	15.950±0.574
Glutamine	1.881±0.127	1.836±0.032	1.646±0.126
Glycerol*	10.404±4.173^{a}	0.169±0.020^{b}	0.218±0.045^{b}
Glycine*	0.357±0.007^{a}	0.396±0.074^{a}	0.060±0.018^{b}
Histidine*	0.291±0.036^{a}	0.246±0.008^{b}	0.000±0.000^{c}
Isobutyrate*	0.013±0.005^{a}	0.018±0.004^{a}	0.007±0.001^{b}
Isoleucine	0.164±0.021	0.156±0.019	0.167±0.009
Lactate*	2.455±0.254^{a}	2.145±0.309^{a}	1.097±0.026^{b}
Leucine	0.334±0.019	0.299±0.023	0.331±0.019
Lysine*	0.567±0.122^{a}	0.604±0.019^{a}	0.414±0.034^{b}
Mannose	0.188±0.062	0.104±0.009	0.087±0.036
Methanol*	0.150±0.049^{b}	2.346±0.292^{a}	0.124±0.012^{b}
Methionine	0.089±0.025	0.090±0.014	0.053±0.024
Myo-inositol	0.154±0.108	0.196±0.024	0.114±0.030
Dimethylglycine	0.003±0.002	0.004±0.000	0.004±0.000
O-Acetylcarnitine	0.043±0.007	0.042±0.001	0.048±0.006
Ornithine*	0.126±0.021^{a}	0.165±0.036^{a}	0.057±0.039^{ab}
Phenylalanine	0.175±0.035	0.152±0.010	0.241±0.057
Proline	0.318±0.056	0.369±0.041	0.334±0.079
Serine*	0.323±0.039^{a}	0.317±0.051^{a}	0.159±0.021^{b}
Succinate	0.013±0.006	0.017±0.003	0.012±0.006
Threonine	0.267±0.114	0.360±0.023	0.316±0.072
Tryptophan*	0.000±0.000^{c}	0.184±0.030^{b}	0.287±0.063^{a}
Tyrosine	0.209±0.039	0.191±0.003	0.213±0.005
Valine	0.533±0.100	0.525±0.074	0.462±0.047

UF: ultrafiltration, PR: protein precipitation, SPE: solid phase extraction, 3-M-2-oxovalerate: 3-Methyl-2-oxovalerate. *Statistical significant. Values with different letters (a, b, c) are significantly different between methods (one-way ANOVA followed by FDR correction, $P < 0.05$).

Table 2. Behaviour of the significant metabolites among methods compared with ultrafiltration.

Metabolite	PR	SPE	p-Value[1]	FDR[2]
Histidine	↓	↓↓	1.75E-08	7.17E-07
Methanol*	↑	-	1.00E-06	2.01E-05
Tryptophan	↑	↑↑	1.47E-06	2.01E-05
Creatinine	-	↓	3.29E-06	3.38E-05
Glycine	-	↓	3.28E-05	2.69E-04
Glycerol*	↓	↓	7.30E-05	4.99E-04
Lactate	-	↓	3.15E-04	1.65E-03
Formate*	-	↑	3.22E-04	1.65E-03
Citrate	↓	↓↓	4.17E-04	1.90E-03
Serine	-	↓	1.60E-03	6.55E-03
2-Hydroxybutyrate	↑	↑	6.07E-03	2.26E-02
Isobutyrate	-	↓	1.01E-02	3.46E-02
Ethanol	-	↓	1.31E-02	4.15E-02
Acetone	↓	-	1.74E-02	4.94E-02
Acetoacetate	↓	↓	1.81E-02	4.94E-02
Ornithine	-	↓	2.45E-02	6.27E-02
Lysine	-	↓	3.71E-02	8.95E-02

PR: protein precipitation, SPE: solid phase extraction, -: no change respect to ultrafiltration, ↑: statistically higher than ultrafiltration, ↓: statistically lower than ultrafiltration, ↑↑: statistically higher than ultrafiltration and precipitation with methanol, ↓↓: statistically lower than ultrafiltration and precipitation with methanol.* Changes due to the methodology. [1]p-Value of one-way ANOVA. [2]False discovery rate.

Table 3. Score derived of considering practical aspects in the sample treatments.

Procedure	CIM	ACM	Automation	Quickness	Low-cost	Final score
UF	medium	high	low	low	medium	**9/15**
PR	high	medium	medium	high	high	**13/15**
SPE	medium	low	high	medium	low	**9/15**

CIM: capacity to identify metabolites, ACM: absolute concentration of identified metabolites, UF: ultrafiltration, PR: protein precipitation, SPE: solid phase extraction. Final score was calculated as the sum of individual parameters as follows: High= 3 points, medium= 2 points, low= 1 point.

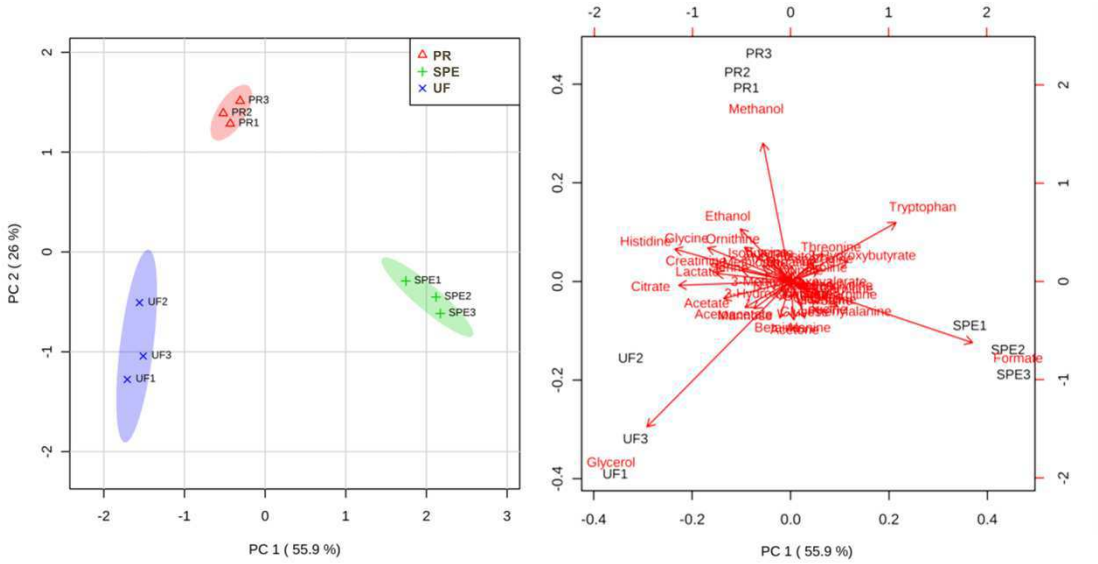


Figure 1. Principal component analysis (PCA) score plot and loading biplot of plasmatic samples across methodologies. First and second principal components explained 81.9% of the variability. UF: ultrafiltration, PR: protein precipitation, SPE: solid phase extraction.

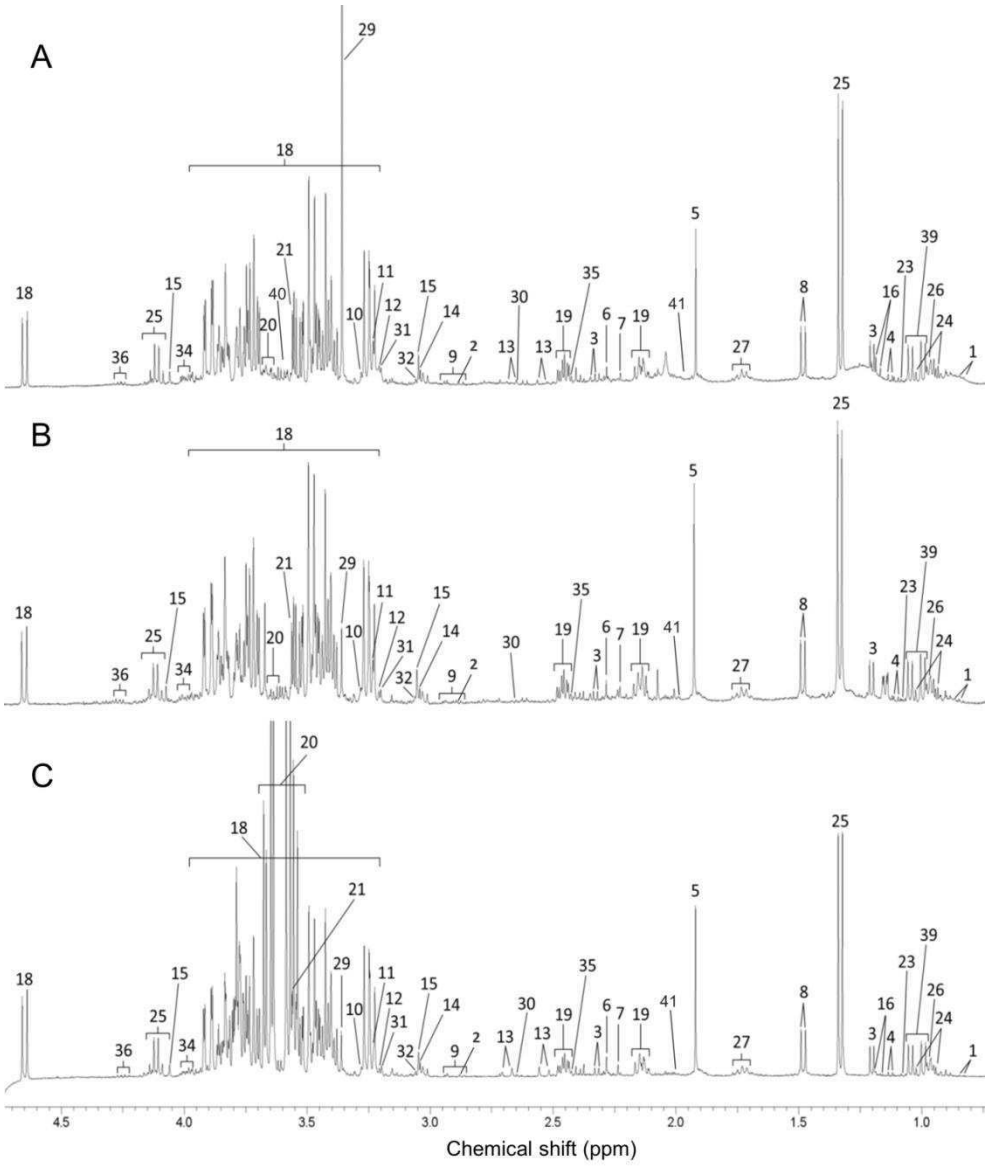


Figure 2. 400 MHz ^{1}H-NMR spectra of plasma treated with (A) precipitation with methanol, (B) solid phase extraction and (C) ultrafiltration (aliphatic region). 1, 2-Hydroxybutyrate; 2, Dimethylglycine; 3, 3-Hydroxybutyrate; 4, 3-Methyl-2-oxovalerate; 5, Acetate; 6, Acetoacetate; 7, Acetone; 8, Alanine; 9, Asparagine; 10, Betaine; 11, Carnitine; 12, Choline; 13, Citrate; 14, Creatine; 15, Creatinine; 16, Ethanol; 17, Formate; 18, Glucose; 19, Glutamine; 20, Glycerol; 21, Glycine; 22, Histidine; 23, Isobutyrate; 24, Isoleucine; 25, Lactate; 26, Leucine; 27, Lysine; 28, Mannose; 29, Methanol; 30, Methionine; 31, O-Acetylcarnitine; 32, Ornithine; 33,Phenylalanine; 34, Serine; 35, Succinate; 36, Threonine; 37, Tryptophan; 38, Tyrosine; 39, Valine; 40, Myo-Inositol; 41, Proline.

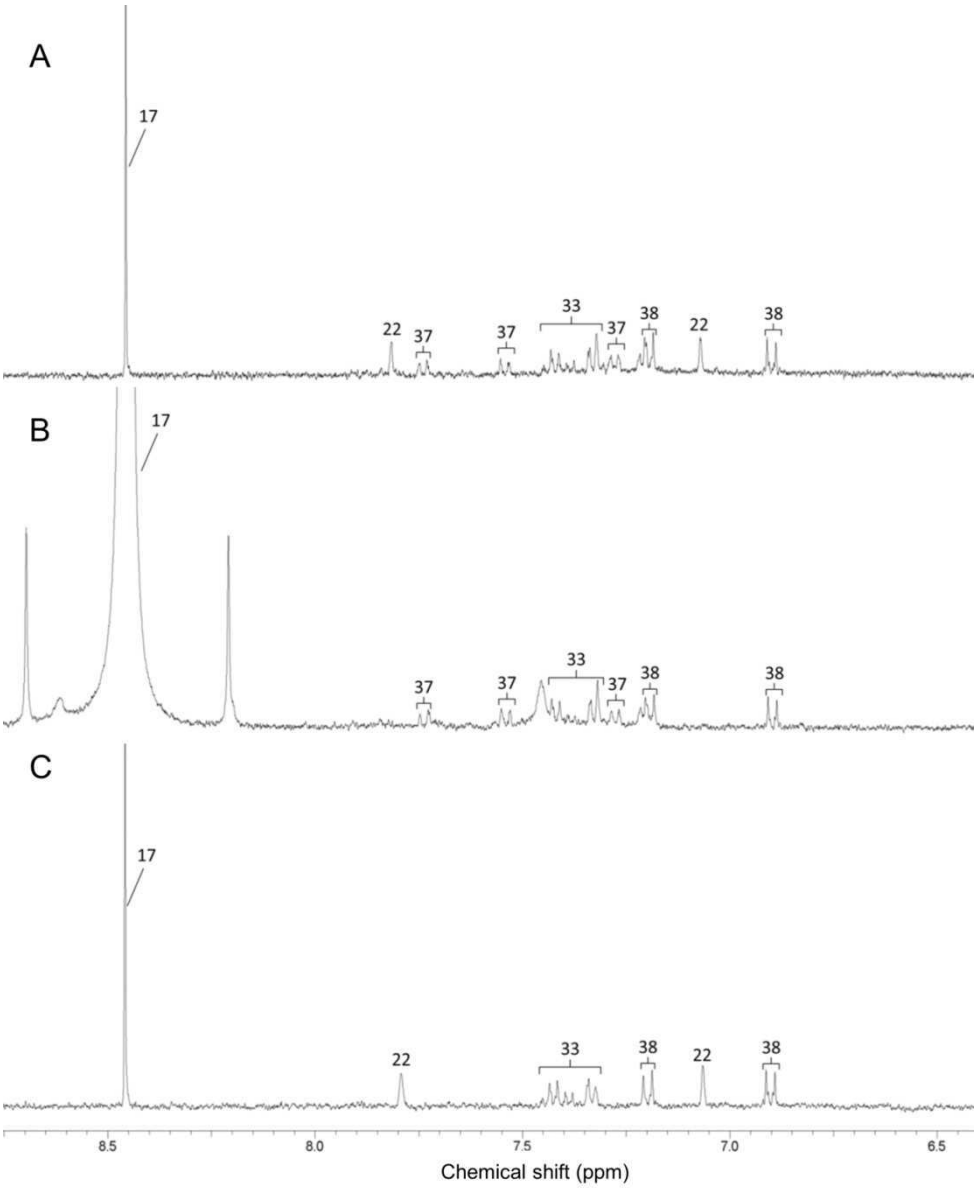


Figure 2. Continued (aromatic region).

Capítulo 07.

ESTUDIOS DE INTERVENCIÓN NUTRICIONAL

7.1. Estudio de intervención nutricional en humanos tras el consumo agudo de legumbres.

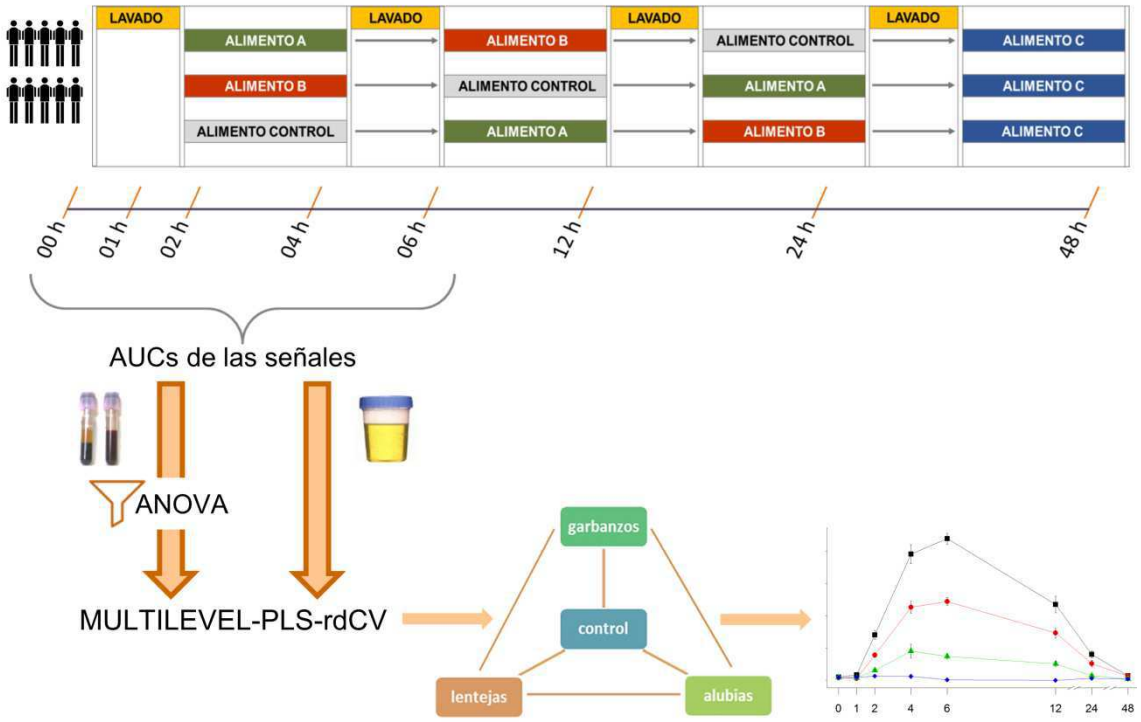
En el marco del proyecto europeo "*Food Biomarker Alliance*" (FOODBALL; http://foodmetabolome.org/), y concretamente del subobjetivo "*discovery of novel biomarkers of food intake*", se ha analizado el perfil metabolómico de la orina y el suero tras el consumo de legumbres (específicamente lentejas, garbanzos y alubias). Dicho estudio de intervención se realizó en la Universidad de Barcelona, y el procesado experimental de las muestras se llevó a cabo en colaboración con la *Swedish University of Agricultural Sciences* (Uppsala, Suecia) y el *Swedish NMR Centre* (Gotemburgo, Suecia) mediante una estancia internacional con el grupo del Dr. Rikard Landberg y con el seguimiento científico del Dr. Carl Brunius. Esta estancia pudo contar con la financiación de una beca de movilidad para investigadores predoctorales que otorga la Acción Europea POSITIVe COST Action (FA1403) "*Interindividual variation in response to consumption of plant food bioactives and determinants involved*", de una beca de movilidad para estancias en el extranjero para los investigadores predoctorales APIF de la Universidad de Barcelona, y con la financiación del proyecto PCIN-2014-133 FOODBALL-MINECO/UB de la JPI-HDHL para el análisis experimental de las muestras. Este trabajo se ha enviado a una revista indexada en el *Journal Citation Reports* y actualmente se encuentra en proceso de revisión.

Francisco Madrid-Gambin, Carl Brunius, Mar Garcia-Aloy, Sheila Estruel-Amades, Rikard Landberg, Cristina Andres-Lacueva. Untargeted ^{1}H-NMR based metabolomics analysis of urine and serum profiles after consumption of dietary pulses: an extended meal study. [*En proceso de publicación*].

Resumen:

La evaluación precisa y objetiva de la exposición dietética es uno de los actuales retos en la investigación nutricional. El presente estudio tuvo como objetivo descubrir biomarcadores nutricionales asociados con el consumo de legumbres (concretamente lentejas, garbanzos y alubias) a través un enfoque metabolómico no dirigido mediante RMN en un estudio cruzado y aleatorizado. Dos raciones de lentejas, garbanzos, alubias o pasta (alimento control) se sirvieron como desayuno (300 g) y se tomaron muestras de orina y suero. Dichas muestras fueron analizadas mediante una estrategia metabolómica no dirigida por RMN. Se calculó el área bajo la curva (AUC, del inglés *Area Under the Curve*) de las señales provenientes de las muestras de orina y suero desde antes de la ingesta hasta las 6 h después de la ingesta ($AUC_{0\text{-}6\ h}$) para cada uno de los cuatro alimentos. El análisis estadístico multivariante ("*multilevel PLS-rdCV*") mostró como principales biomarcadores nutricionales asociados al consumo de legumbres la trigonelina, la 3-metilhistidina, dimetilglicina, trimetilamina y lisina. También se observaron menores niveles de glucosa y urea tras el consumo de las legumbres en comparación al consumo de pasta (alimento control). Los metabolitos urinarios que fueron discriminantes fueron adicionalmente investigados a lo largo de las 48 h tras la ingesta. La intensidad máxima de la mayoría de metabolitos se produjo de manera más tardía después del consumo de alubias y garbanzos comparado con lentejas. El consumo de legumbres afecta al perfil metabolómico urinario y sérico, el análisis del cual permite la elucidación de potenciales biomarcadores nutricionales asociados a su ingesta. Las diferencias observadas en la excreción de los metabolitos endógenos reflejan alteraciones en el metabolismo que pueden estar relacionadas con los efectos beneficiosos asociados a la ingesta de legumbres.

Figura 4. Gráfico-resumen del estudio de intervención nutricional en humanos tras el consumo de legumbres.



Research Article

Untargeted ^{1}H-NMR based metabolomics analysis of urine and serum profiles after consumption of dietary pulses: an extended meal study

Francisco Madrid-Gambin[1], Carl Brunius[2,3], Mar Garcia-Aloy[1,4], Sheila Estruel-Amades[1], Rikard Landberg[2,3], Cristina Andres-Lacueva[1,4*]

[1]Biomarkers & Nutrimetabolomics Laboratory, Department of Nutrition, Food Sciences and Gastronomy, Food Technology Reference Net (XaRTA), Nutrition and Food Safety Research Institute (INSA), Campus Torribera, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain

[2]Department of Molecular Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

[3]Department of Biology and Biological Engineering, Food and Nutrition Science, Chalmers University of Technology, Gothenburg, Sweden

[4]CIBER de Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona, Spain

Correspondence: Cristina Andres-Lacueva (candres@ub.edu)

Keywords: biomarkers; dietary pulses; legumes; metabolomics; NMR

Abbreviations: AUC, area under the curve; BCAA, branched-chain amino acid; BMI, body mass index; FDR, false discovery rate; PCA, principal component analysis; PLS, partial least squares analysis; rdCV, repeated double cross-validation; TMA, trimethylamine; TMAO, trimethylamine-N-oxide.

Abstract

Scope: Legumes have been associated with positive effects on health. The use of metabolomics helps to evaluate these effects and their mechanisms. The current study aimed to investigate potential nutritional biomarkers of dietary pulses (including lentils, chickpeas and beans) consumption through an untargeted metabolomics approach.

Methods and results: Eleven healthy subjects consumed three different pulse-rich meals in a randomized, 3-way crossover study. Lentils, chickpeas or pasta (control food) were served as breakfast (300 g cooked product). Eight of the 11 individuals also performed a similar intervention with beans. An untargeted NMR-based metabolomics approach was applied to urine and serum samples. Multivariate statistical analysis of samples from baseline to 6 h after intake showed that trigonelline, 3-methylhistidine, dimethylglycine, trimethylamine and lysine could be potential biomarkers of these intakes. Decreased levels of glucose in urine and serum were observed after the consumption of legumes compared to control food. The discriminant urinary metabolites were further explorer until 48 h post intervention.

Conclusions: This study shows potential candidates for biomarkers of pulses intake. Differences in endogenous metabolites can be linked to the observed health benefits associated with intake of pulses. The use of a combined metabolite panel may improve the specificity in observational studies.

1 Introduction

Dietary pulses refer to the dried seeds of legumes such as lentils, chickpeas and beans, among others [1]. High intake of pulses has been associated with lower body weight [2], lower blood pressure [3], lower LDL cholesterol levels [4] and improvement of glycaemic control [5] in meta-analyses of controlled feeding studies. Due to these effects, pulses are increasingly being recognized for their role in promoting health [6, 7]. Indeed, habitual pulse consumption is included in the main dietary guidelines worldwide, including the Mediterranean Diet [8], the Dietary Guidelines for Americans [9, 10] and the Nordic Diet [11], among others, and they are also advocated in view of their low environmental impact compared with other protein sources [12]. However, monitoring specific food intakes or compliance to dietary components is complicated by dietary assessment tools based on self-reporting methods that are innately inaccurate [13]. Therefore, there is a growing need for reliable biomarkers which correctly reflect the food intakes to allow accurate risk assessment but also for the elucidation of mechanisms of observed effects [14]. Discovery and validation of specific biomarkers related to pulses could improve the evaluation of pulses intake in relation to hard endpoints in epidemiological studies.

Metabolomics aims to identify specific molecular activities occurring in biological systems by the identification and quantitation of a large number of metabolites with high-throughput techniques [15]. Hence, metabolomics may be a powerful tool for identification of food exposure biomarkers in humans [16], either single biomarkers or multivariate biomarker panels [17]. Additionally, reflecting diet-induced changes in the metabolome, new biomarkers of the effect of the intake may reveal potential mechanisms in diet-related physiology [18]. Therefore, metabolomic approaches have been proposed for evaluating the relationship between nutrition and health status [19].

Previously, a study investigating exposure markers in serum following dietary intervention with dry beans proposed pipecolic acid and S-methyl cysteine as biomarker candidates [20]. However, that study did not investigate potential differences between pulses. Moreover, dietary exposure is often better reflected in the urinary metabolome [21]. On the other hand, urinary exposure markers of pulses were previously investigated in free-living subjects by our group [22], but to our knowledge no study has so far evaluated the effects of exposure on both urinary and serum metabolome after controlled dietary intervention with specific pulses. Untargeted metabolomics approaches applied to single meal studies with a time series of samples taken after intake is a useful strategy for discovery of new dietary biomarkers [23, 24]. Moreover, this technique allows for the study of perturbations in endogenous metabolites as a result of the acute pulses intake, which may relate to health effects.

In the present study, an untargeted ^{1}H-NMR metabolomics approach was applied with the aim to perform an exploratory discovery of dietary and endogenous biomarkers related to a single meal intake of specific pulses namely lentils, chickpeas and white beans and to investigate the time dependent profiles of these biomarker candidates in urine and serum.

2 Materials and Methods

2.1. Subjects and study design

The study was conducted as a randomized crossover study (Figure 1). In total, 14 healthy participants were enrolled and 11 of them (7 women and 4 men) completed the intervention with control, chickpea and lentil diets. The individuals were between 19 and 37 years old with a body mass index (BMI) of 24.3±3.8 kg/m^2. In addition, 8 of the 11 individuals were also recruited to consume beans. Subjects of the subgroup were between 19 and 37 years old and

with a BMI of 24.2±3.8 kg/m^2. Exclusion criteria included intake of supplements, serious illness (such as heart disease, kidney disease or diabetes) and allergy to legumes or pasta.

After a wash-out period of one week, the participants were instructed to avoid all legumes and legume-related products, alcohol beverages and foods containing phenolic compounds starting 2 d before the study and throughout the study for 5 d in connection with consumption of each diet. Participants received standardized meals from the dinner preceding intervention until 48 h after of the intervention (Supplemental Table 1). After the initial 2-d run-in period, all participants ingested a single dose of the test meal as breakfast. The test meals consisted of 300 g cooked lentils, chickpeas, beans or pasta (control), with 7 g of butter, allowing 20 min for the ingestion. The nutritional content of the advised foods is reported in Table 1. Individuals consumed all foods in a crossover design, except for the individuals who underwent the bean intervention, in which all subjects consumed this food at the final place of the study (Figure 1). The experiment was conducted in the Hospital Clinic of Barcelona (Spain), where urine samples were collected for the following time intervals: baseline, 0–1 h, 1–2 h, 2–4 h, 4–6 h, 6–12 h, 12–24 h and 24–48 h. Blood samples were drawn at baseline, and after 1 h, 2 h, 4 h, 6 h, 24 h and 48 h. All urine samples were collected in plastic specimen containers, and then aliquoted and stored at −80°C until analysis. Blood samples were collected in vacutainer tubes, and stored on ice during 30 min. Then, samples were centrifuged at 1600 g during 15 min and supernatants were aliquoted and stored at −80°C until analysis.

The study protocol was approved by the Bioethical Committee of the University of Barcelona and all participants provided the written informed consent. The study was registered in the International Standard Randomized Controlled Trial Number as ISRCTN17200423.

2.2. Sample preparation

Cooked pulses were homogenized, after which 40 mg was mixed with methanol/water (1.5 ml 80:20 v/v). Mixtures were centrifuged at 1320 rpm and 4º C for 15 min and supernatants were

then decanted to fresh vials and dried under nitrogen. Food extracts were reconstituted in 200 µl of buffer (37.5 mM sodium phosphate, 0.02% sodium azide, 0.25 mM DSS-d_6 and 1 mM imidazole in D_2O (D 99%)) and shaked for 30 min. Samples (180 µl) were transferred into 3 mm SampleJet NMR tubes for analysis.

Serum (150 µl) was transferred into 96-deepwell plates containing 150 µl of ultrapure water in each well. Then, 600 µl of pure cold (-20º C) methanol was added to each well followed by incubation on thermomixer at 12º C, 800 rpm during 10 min. The samples were kept at −20 °C for 30 min prior to centrifugation at 2250 *g*, 4º C during 60 min. Supernatants (600 µl) were lyophilized, washed with 50 µl of MeOD and re-lyophilized to remove solvent remains. Dried samples were reconstituted in 200 µl of buffer (37.5 mM sodium phosphate, 0.02% sodium azide, 0.25 mM DSS-d_6 and 1 mM imidazole in D_2O (D 99%)). Samples (180 µl) were transferred into 3 mm SampleJet NMR tubes for analysis.

Urine samples were thawed at 4°C, spun down, and 180 µl were added in a 96-deep well plate containing 20 µL of internal standard solution (1.5 M potassium phosphate, 0.5% sodium azide and 0.5 mM DSS-d_6 in D_2O (D 99%)). After shaking in an Eppendorf Thermomixer (22°C, 800 rpm, during 30 s), the mixture was transferred into 3 mm SampleJet NMR tubes using a SamplePro L liquid handling robot (Bruker BioSpin, Rheinstetten, Germany).

2.3. NMR spectroscopy analysis

Spectra were recorded using an Oxford 800 MHz magnet equipped with a Bruker Avance III HD console and a 3 mm TCI cryoprobe. Acquisition temperature was 298K. ^{1}H spectra were acquired using a water suppression pulse program collecting 128 scans into 65 K data points with a spectral width of 20 ppm, an acquisition time of 4 s and a relaxation delay of 1 s. The FIDs were multiplied by an exponential weighting function equivalent to line broadening of 0.3 Hz prior to Fourier transformation.

For assignment purposes, ^{1}H-^{13}C HSQC spectra were acquired on the same spectrometer, collected into 16 scans, spectral widths of 20 (^{1}H) and 100 (^{13}C) ppm and relaxation delay of 3 s. Spectra were processed using TopSpin 3.5pI6 (Bruker GmbH, Rheinstetten, Germany).

2.4. Processing of NMR data

Processed spectra were imported into R v 3.3.1 [25] and aligned with the 'speaq' R-package v 1.2.1 [26]. Metabolic features were obtained using continuous wavelet transformation peak picking and extracting peak intensities at corresponding chemical shifts. Data sets were probabilistic quotient normalized [27] for dilution factors adjustment. Area under the curve (AUC) of each feature was calculated from t_{0h} until t_{6h}.

2.5. Statistical analysis

Principal component analysis (PCA) on baseline samples coloured by the sequence that participants followed and by the immediately previous intervention were performed to investigate the presence of potential carryover effects. No carryover effect could be observed since any type of cluster by sequence nor by the previous intervention could be observed (Supplemental Figure 1). Additionally, PCAs of AUCs of each intervention was performed to evaluate the presence of outliers. No outliers were observed through the PCA (Supplemental Figure 2). To exclude noisy variables from the serum data, a univariate pre-selection of variables was performed: repeated measures ANOVA was applied on variable-wise to serum data, using treatment, time and treatment x time as fixed factors with individual within treatment as repeating unit. Only those variables with significant treatment effect after FDR adjustment were retained in multivariate analysis. Predictive modelling of differences between treatments was performed by Partial Least Squares (PLS) regression of $AUC_{0\text{-}6\ h}$ data in a pairwise multilevel design to take into consideration the crossover design, having each individual as its own control [28, 29]. The modelling was performed in a repeated double cross-validation (rdCV) framework [30, 31] optimised for unbiased variable selection [32]. The

double cross-validation procedure consists of an inner "tuning" (or validation) loop nested within an outer "testing" loop to reduce bias from overfitting models to experimental data. Variables were ranked and selected in the inner loop as described elsewhere [32]. Model performance was further assessed by permutation analysis (serum, n=400; urine, n=200) [33]. All statistical analyses were performed in the open source statistical programming environment R v 3.3.1 [25] and all scripts are available from the authors upon request.

2.6. Identification

Metabolite identification and deconvolution were performed using Chenomx NMR Suite 8.2 profiler (Chenomx Inc., Edmonton, Canada). Further contributions to the proton peak assignment were provided by comparing the chemical shifts with those available in the Human Metabolome Database (http://www.hmdb.ca). Finally, an in-house R script for statistical total correlation spectroscopy (STOCSY) [34] and a heteronuclear ^{1}H-^{13}C HSQC experiment were used to assign tentative metabolite signals to specific metabolites.

2.7. Exploration of metabolite time profiles

The identified variables unbiasedly extracted from the multivariate models were combined into single metabolites which were examined individually by repeated measures ANOVA across the time points, using the Geepack R package v 1.2-1 [35]. Treatment (control, lentil, chickpea and bean), time (t_{0h}–t_{48h}) and treatment x time interaction were used as fixed factors and individual x treatment as repeating unit with an autoregresive covariance structure. The Benjamini-Hochberg false discovery rate (FDR) procedure was used to adjust for multiple testing. Statistical significance was considered at an FDR-adjusted *p*-value <0.05.

3 Results and Discussion

In addition to the nutritional content (Table 1), metabolomic analysis of extracts from pulses (Supplemental Table 2) revealed precursor compounds of the metabolites that could be

considered biomarkers of intake [36, 37]. A total of 41 compounds were identified showing different concentrations among pulses. Some examples are raffinose, stachyose and pipecolate, which presented diverse levels in the pulses extracts. Regarding the biological samples, multilevel classification resulted in zero misclassified subjects using urine data, while for serum only one misclassified subject was observed (Supplemental Figure 3) and the high correct classification rate was confirmed by permutation tests ($p<0.05$ for all models, Table 2). Assignment of significant features from the pairwise models enabled the identification of several metabolites that distinguished dietary treatments that were either exogenous compounds or altered endogenous metabolites (Table 2).

3.1. Serum biomarkers of pulse consumption

The low amount of significant features in serum data did not enable a robust modelling without a pre-selection of variables to remove noise in the data. Modelling after univariate variable pre-selection, multivariate classification models resulted in a total of 8 metabolites chosen as discriminant compounds of the consumption of dietary pulses (Table 2). Most of these metabolites may not derive from the food items *per se*, but rather reflect a response in endogenous metabolism. After lentils consumption, there was an increase of histidine and dimethylglycine compared with control. The levels of histidine in serum were also altered between pulses intakes, showing the highest intensity after beans, followed by chickpeas and lentils. Interestingly, these findings may mirror the food composition (Supplemental Table 2). On the other hand, variations of the levels of metabolites coming from the microbiota, such as 3-hydroxyisovalerate, may reflect differences in the metabolism of aromatic amino acids and polyphenols by colon microorganisms [39, 40].

Dietary pulses have a low glycemic index [2]. This is further supported by the lower glucose in both urine and serum samples after the ingestion of pulses compared with control. This compound was also different among pulses. The lower levels of glucose after beans compared

to chickpeas, and chickpeas compared to lentils suggest a low glycaemic index, sort as beans < chickpeas < lentils (and finally pasta). However, García-Alonso et al. (1998), calculated a higher glycaemic index after beans compared to lentils [41].

Chickpea intake resulted in higher lysine levels, which could reflect the food composition (Supplemental Table 2). Asparagine, 2-hydroxybutyrate and were higher after bean consumption. Although Esko et al. (2017) reported 2-hydroxybutyrate and asparagine as discriminants of a low-fat diet [38], this is not likely in the present study, since the beans contained more fat than the control food. Moreover, asparagine content was also highest in the beans and observed differences among pulses could then reflect dietary intake (Supplemental Table 2).

3.2. Urinary biomarkers of pulse consumption

Trigonelline was the dominant excreted compound after the three different dietary pulses (highest loading score, data not shown), and was thus the most discriminant biomarker of all pulses when they were compared against control. However, the levels of urinary trigonelline were different between the three pulses, reflecting the differences in intake from the test foods (Supplemental Table 2). In fact, beans had almost twice trigonelline as much as the chickpeas, and about six times more than the lentils (Supplemental Table 2). However, these results differ with those presented in a study by Zwart et al. (2003), in which they observed that the concentration of trigonelline was higher in chickpeas, followed by lentils and beans [42]. We attribute these discrepancies to biological variability from choice of cultivars and other agronomical factors. In another study, trigonelline was observed in serum after bean consumption using MS-based analysis [20]. However, we could not successfully identify this compound in serum in the present study, presumably due to the lower sensitivity of NMR compared to MS. Previous short- and long-term studies have shown that consumption of e.g. coffee leads to the presence of urinary trigonelline [37, 43], which has in fact been proposed as

biomarker of coffee consumption. The lack of specificity may limit its use as a selective biomarker for pulses consumption.

3-Methylhistidine increased after the consumption of all pulses, but especially after the consumption of beans and closely mirrored the histidine content in the food (Supplemental Table 2) and also shown in serum (Table 2). Despite that this metabolite was associated with the ingestion of pulses in our previous observational study [22], it has also been found in urine after consumption of meat, especially chicken, and fish [44] and thus it could not be specific to pulses.

The level of lysine was increased after the consumption of all pulses compared to control. There were also differences in the excretion among pulses, according to the previous composition of lysine in the foods (Supplemental Table 2). Therefore, urinary lysine may come from the lysine present in pulses. Higher plasmatic amino acids such as lysine were observed after intake of whole-grain [45] indicating poor specificity to legumes. Although pipecolate was identified in the beans of the present study, and it has also been described in human serum after its consumption in MS [20], it was not detected in either the urine or the serum samples of the participants by NMR. This compound was also not identified in other legume types.

Higher levels of dimethylglycine, trimethylamine (TMA) and methylamine were observed after the consumption of chickpeas (Table 2). Indeed, in our previous research, these metabolites were associated with total pulses consumption composed by lentils, chickpeas and beans in an observational setting (in that study we were not able to study the differences between the different types of pulses due to intrinsic characteristics of the data used) [22]. These metabolites may appear in the urine because of the exposure to quaternary amines such as choline in legumes [22, 46]. However, choline was not significantly higher in the chickpeas compared with the other legume sources (Supplemental Table 2) and these metabolites may instead reflect other dietary quaternary amines such as trimethylglycine (i.e. betaine), carnitine and/or lecithin

[47] or metabolism perturbations related to dietary exposure. TMA is generated by gut microbial activity from precursors such as choline, carnitine and trimethylglycine [48, 49] and in turn causes the formation of both trimethylamine-*N*-oxide (TMAO) and dimethylamine [50], which we previously showed that they were associated with global pulses consumption in a population with cardiovascular risk [22]. Thus, although dimethylglycine, TMA and TMAO may seem to be good candidates of pulses consumption in general [22] or more likely to chickpea consumption as indicated in the present study, these metabolites should be considered with caution because they may reflect the exposure to quaternary amines rather than dietary pulses. In fact, TMAO has been considered a biomarker of red meat consumption [44], fish consumption [51] and cardiovascular risk [52] and thus most likely has very limited use as a biomarker of pulses consumption.

Xanthurenate (3-hydroxykynurenate), which is a compound involved in the kynurenine pathway of the metabolism of tryptophan [53], was increased after chickpeas and lentils. This metabolite was also observed in urine after a sustained consumption of cheese [54] and increased accordingly to the amount of tryptophan ingested as a supplement in a crossover intervention study [55]. Also ascorbic acid appeared as a discriminant compound for chickpea consumption (Table 2). However, this was attributed to its use as preservative and thus does not reflect the intake of chickpeas *per se*.

The increase in glutamine and *N*-acetylglutamine after the consumption of lentils could reflect the food composition [56]. However, these results may also reflect a modulation of the production of this glucogenic amino acid as a result of perturbations in the BCAAs metabolism and the Krebs cycle [57, 58]. The low concentration of this metabolite in the foods suggests an endogenous response rather than a direct excretion, as has also been suggested in other studies after meat consumption [59], soy consumption [60], and after the exposure to the Mediterranean diet [61].

Regarding biomarkers of effects associated with the intake of legumes, several amino acids were found in urine after pulses consumption. The presence of branched amino acids (BCAAs) valine and leucine were higher after consumption of beans compared to the other food meals. Leucine has a ketogenic role in the body, while valine is glycogenic, ingoing to the Krebs cycle as succinyl-CoA. Valine is metabolized to succinate via methylmalonate and methylmalonyl-CoA [57]. However, although the excretion of valine was higher in legumes, the excretion of methylmalonate was statistically much lower than in the control. On the other hand, 3-aminoisobutanoate and 3-hydroxyisobutyrate are formed during the metabolism of valine and can be excreted. Nevertheless, the low levels of these metabolites compared to the excretion after the control meal suggest a slowdown in BCAAs metabolism, with the consequent increase in the urinary BCAAs. In agreement with our previous observational study with pulses, we observed a lower urinary glucose in pulses replicating this finding in healthy subjects [22].

Overall, among identified metabolites there was a lack of specificity to dietary pulses. In summary, trigonelline, 3-methylhistidine, trimethylamine, dimethylglycine, and lysine were the most adequated metabolites to consider as biomarkers of pulse intake. However, the metabolites observed in the present study to be discriminative of total or specific pulses consumption were not specific to the consumption of pulses. Their potential use as individual biomarkers could therefore be limited. Using multiple biomarkers in combination has the potential to enable more specific and reliable estimation of dietary exposure [17]. However, the present study setting did not allow for such investigations due to the study design.

3.3. Monitoring ^{1}H NMR significant urinary metabolites over 48 h

Excretion patterns of the most strongly discriminating urinary metabolites are shown in Figure 2 and time behaviour of all discriminant metabolites is reported in Supplemental Table 3 and Supplemental Figure 4. Overall, there was a delay in the excretion peak of most metabolites in the bean and chickpea intervention arms (Supplemental Figure 4). This is likely explained by be

a delayed digestion, absorption and subsequent excretion of several metabolites due to the high fiber content in beans and chickpeas (Table 1) compared to the other meals [62, 63].

The different concentrations of trigonelline in the foods were reflected in the urine of the volunteers, suggesting a dose-response relationship (Figure 2; Supplemental Table 4). Differences in excretion were apparent from two hours post ingestion and peaking after four to six hours. Increased levels were sustained at least 24 h and Lang and co-workers in fact observed this elevated levels of trigonelline up to three days after coffee intake [43].

TMA increased after chickpeas and beans, showing differences 4 h post-ingestion (Figure 2). The non-significant increase after lentils suggests a higher specificity towards chickpeas and beans. In parallel to this results, in our previous observational study TMAO, the downstream product of TMA, was associated with total pulses consumption [22].

The excretion pattern of 3-methylhistidine may be affected by the different meals, in which again was earlier after the lentil consumption (maximum excretion to T_{2h}) than after chickpea and bean intake (maximum excretion at T_{6h}). Interestingly, the similar behavior was observed for histidine (Figure 2), in which the highest level was at $T_{4\text{-}6h}$ after beans intake also according to the histidine present in the foods. According to the results presented, this metabolite may reflect the intake of pulses, but also the intakes of meat and fish from 2 h after consumption [44].

Likewise, the excretion of lysine after beans, whose maximum intensity occurred at T_{4h}, was later than in the case of lentils, whose maximum excretion occurred at T_{2h} after ingestion (Supplemental Figure 4, Supplemental Table 4).

Dimethylglicine increased from T_{4h} until T_{12h}. Therefore, this metabolite only would reflect either the recent consumption or habitual repeated consumption. In addition, despite the multivariate analysis of $AUC_{0\text{-}6h}$ indicated dimethylglicine as discriminant of chickpeas

consumption, the univariate analysis revealed the increase of this metabolite after all pulses (Figure 2, Supplemental Table 4). We attribute these discrepancies to the variability derived of the AUC performing included in the models.

4 Concluding remarks

Untargeted metabolomics revealed several metabolites that were discriminative of total or specific pulses consumption, being the urine a more suitable biofluid than serum for discovering of nutritional biomarkers of pulses. Among the identified metabolites, trigonelline was the most discriminating metabolite. On the other hand, 3-methylhistidine, downstream products of quaternary amine metabolism (dimethylglycine and TMA as precursor of TMAO), and glutamine matched previous findings from our observational study with free-living subjects. Other amino acid metabolism intermediates, e.g. lysine, also managed to discriminate pulses consumption. Although these metabolites emerged as candidates of biomarkers of intake of pulses in the present study, they are not specific to dietary pulses consumption which will severely hamper their use. The use of a combined metabolite panel may be used to improve the predictive power of the exposure in observational studies.

Acknowledgements

This study is supported by PCIN-2014-133 (FOODBALL project) from the Spanish Ministry of Economy and Competitiveness (MINECO) together with the Joint Programming Initiative "Healthy Diet for Healthy Life" (JPI-HDHL, http://www.healthydietforhealthylife.eu) and the award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR. F.M-G. acknowledges the APIF Ph.D. fellowship (University of Barcelona) and the POSITIVe COST Action (FA1403).

F.M-G. conducted sample measurements, data analysis and interpretation, and drafted the manuscript. C.B. supervised the data analysis and manuscript writing. M.G-A. designed and conducted the human intervention study, and supervised the metabolomics analyses, data interpretation and manuscript writing. S.S-A. designed and conducted the human intervention study. R.L. supervised data interpretation and manuscript writing. C.A-L., the principal investigator of the study, conceived and designed the human intervention study, and supervised data interpretation and manuscript writing. All authors reviewed the final version of the manuscript.

The authors have declared no conflict of interest.

References

[1] Food and Agriculture Organization, Definition and classification of commodities -Roots and tubers and derived products. 1994.

[2] Kim, S.J., de Souza, R.J., Choo, V.L., Ha, V., et al., Effects of dietary pulse consumption on body weight: a systematic review and meta-analysis of randomized controlled trials. *Am. J. Clin. Nutr.* 2016, *103*, 1213–1223.

[3] Jayalath, V.H., de Souza, R.J., Sievenpiper, J.L., Ha, V., et al., Effect of Dietary Pulses on Blood Pressure: A Systematic Review and Meta-analysis of Controlled Feeding Trials. *Am. J. Hypertens.* 2014, *27*, 56–64.

[4] Ha, V., Sievenpiper, J.L., de Souza, R.J., Jayalath, V.H., et al., Effect of dietary pulse intake on established therapeutic lipid targets for cardiovascular risk reduction: a systematic review and meta-analysis of randomized controlled trials. *Can. Med. Assoc. J.* 2014, *186*, E252–E262.

[5] Sievenpiper, J.L., Kendall, C.W.C., Esfahani, A., Wong, J.M.W., et al., Effect of non-oil-seed pulses on glycaemic control: a systematic review and meta-analysis of randomised controlled experimental trials in people with and without diabetes. *Diabetologia* 2009, *52*, 1479–1495.

[6] Bouchenak, M., Lamri-Senhadji, M., Nutritional quality of legumes, and their role in cardiometabolic risk prevention: a review. *J. Med. Food* 2013, *16*, 1–14.

[7] Faris, M.A.I.E., Takruri, H.R., Issa, A.Y., Role of lentils (Lens culinaris L.) in human health and nutrition: A review. *Med. J. Nutrition Metab.* 2013, *6*, 3–16.

[8] Bach-Faig, A., Berry, E.M., Lairon, D., Reguant, J., et al., Mediterranean diet pyramid today. Science and cultural updates. *Public Health Nutr.* 2011, *14*, 2274–84.

[9] U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2015–2020 Dietary Guidelines for Americans. *8th Ed.* 2015.

[10] Britten, P., Marcoe, K., Yamini, S., Davis, C., Development of food intake patterns for the MyPyramid Food Guidance System. *J. Nutr. Educ. Behav.* 2006, *38*, S78-92.

[11] Mithril, C., Dragsted, L.O., Meyer, C., Tetens, I., et al., Dietary composition and nutrient

content of the New Nordic Diet. *Public Health Nutr.* 2013, *16*, 777–85.

[12] MacWilliam, S., Wismer, M., Kulshreshtha, S., Life cycle and economic assessment of Western Canadian pulse systems: The inclusion of pulses in crop rotations. *Agric. Syst.* 2014, *123*, 43–53.

[13] Livingstone, M.B.E., Black, A.E., Markers of the validity of reported energy intake. *J. Nutr.* 2003, 895S–920S.

[14] Ismail, N.A., Posma, J.M., Frost, G., Holmes, E., et al., The role of metabonomics as a tool for augmenting nutritional information in epidemiological studies. *Electrophoresis* 2013, *34*, n/a-n/a.

[15] German, J.B., Hammock, B.D., Watkins, S.M., Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 2005, *1*, 3–9.

[16] O'Gorman, A., Gibbons, H., Brennan, L., Metabolomics in the Identification of Biomarkers of Dietary Intake. *Comput. Struct. Biotechnol. J.* 2013, *4*, 1–7.

[17] Garcia-Aloy, M., Rabassa, M., Casas-Agustench, P., Hidalgo-Liberona, N., et al., Novel strategies for improving dietary exposure assessment: Multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. *Trends Food Sci. Technol.* 2017.

[18] Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., et al., The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* 2014, *99*, 1286–1308.

[19] McNiven, E.M.S., German, J.B., Slupsky, C.M., Analytical metabolomics: nutritional opportunities for personalized health. *J. Nutr. Biochem.* 2011, *22*, 995–1002.

[20] Perera, T., Young, M.R., Zhang, Z., Murphy, G., et al., Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol. Nutr. Food Res.* 2015, *59*, 795–806.

[21] Claus, S.P., Swann, J.R., Nutrimetabonomics:Applications for Nutritional Sciences, with Specific Reference to Gut Microbial Interactions. *Annu. Rev. Food Sci. Technol.* 2013, *4*, 381–399.

[22] Madrid-Gambin, F., Llorach, R., Vázquez-Fresno, R., Urpi-Sarda, M., et al., Urinary 1H

Nuclear Magnetic Resonance Metabolomic Fingerprinting Reveals Biomarkers of Pulse Consumption Related to Energy-Metabolism Modulation in a Subcohort from the PREDIMED study. *J. Proteome Res.* 2017, *16*, 1483–1491.

[23] Zhu, Y., Shurlknight, K.L., Chen, X., Sang, S., Identification and Pharmacokinetics of Novel Alkylresorcinol Metabolites in Human Urine, New Candidate Biomarkers for Whole-Grain Wheat and Rye Intake. *J. Nutr.* 2014, *144*, 114–122.

[24] Van Velzen, E.J.J., Westerhuis, J.A., Van Duynhoven, J.P.M., Van Dorsten, F.A., et al., Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *J. Proteome Res.* 2009, *8*, 3317–3330.

[25] R Core Team, R: A language and environment for statistical computing. *R Found. Stat. Comput.* 2015.

[26] Vu, T.N., Valkenborg, D., Smets, K., Verwaest, K.A., et al., An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics* 2011, *12*, 405.

[27] Dieterle, F., Ross, A., Schlotterbeck, G., Senn, H., Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in ^{1}H NMR Metabonomics. *Anal. Chem.* 2006, *78*, 4281–4290.

[28] Westerhuis, J.A., van Velzen, E.J.J., Hoefsloot, H.C.J., Smilde, A.K., Multivariate paired data analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* 2010, *6*, 119–128.

[29] Van Velzen, E.J.J., Westerhuis, J. a., Van Duynhoven, J.P.M., Van Dorsten, F. a., et al., Multilevel data analysis of a crossover designed human nutritional intervention study. *J. Proteome Res.* 2008, *7*, 4483–4491.

[30] Westerhuis, J.A., Hoefsloot, H.C.J., Smit, S., Vis, D.J., et al., Assessment of PLSDA cross validation. *Metabolomics* 2008, *4*, 81–89.

[31] Filzmoser, P., Liebmann, B., Varmuza, K., Repeated double cross validation. *J. Chemom.* 2009, *23*, 160–171.

[32] Hanhineva, K., Brunius, C., Andersson, A., Marklund, M., et al., Discovery of urinary biomarkers of whole grain rye intake in free-living subjects using nontargeted LC-MS metabolite profiling. *Mol. Nutr. Food Res.* 2015, *59*, 2315–2325.

[33] Szymańska, E., Saccenti, E., Smilde, A.K., Westerhuis, J.A., Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 2012, *8*, 3–16.

[34] Olivier Cloarec, †, Marc-Emmanuel Dumas, †, Andrew Craig, †, Richard H. Barton, †, et al., Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic 1H NMR Data Sets. 2005.

[35] Højsgaard, S., Halekoh, U., Yan, J., The R Package geepack for Generalized Estimating Equations. *J. Stat. Softw.* 2006, *15*, 1–11.

[36] Heinzmann, S.S., Brown, I.J., Chan, Q., Bictash, M., et al., Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *Am. J. Clin. Nutr.* 2010, *92*, 436–443.

[37] Madrid-Gambin, F., Garcia-Aloy, M., Vázquez-Fresno, R., Vegas-Lozano, E., et al., Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects. *Food Res. Int.* 2016, *89*, 1064–1070.

[38] Esko, T., Hirschhorn, J.N., Feldman, H.A., Hsu, Y.-H.H., et al., Metabolomic profiles as reliable biomarkers of dietary composition. *Am. J. Clin. Nutr.* 2017, *105*, 547–554.

[39] Moco, S., Martin, F.-P.J., Rezzi, S., Metabolomics View on Gut Microbiome Modulation by Polyphenol-rich Foods. *J. Proteome Res.* 2012, *11*, 4781–4790.

[40] Rechner, A.R., Kuhnle, G., Bremner, P., Hubbard, G.P., et al., The metabolic fate of dietary polyphenols in humans. *Free Radic. Biol. Med.* 2002, *33*, 220–235.

[41] García-Alonso, A., Goñi, I., Saura-Calixto, F., Resistant starch and potential glycaemic index of raw and cooked legumes (lentils, chickpeas and beans). *Eur. Food Res. Technol.* 1998, *206*, 284–287.

[42] De Zwart, F.J., Slow, S., Payne, R.J., Lever, M., et al., Glycine betaine and glycine betaine analogues in common foods. *Food Chem.* 2003, *83*, 197–204.

[43] Lang, R., Wahl, A., Stark, T., Hofmann, T., Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Mol. Nutr. Food Res.* 2011, *55*, 1613–1623.

[44] Cheung, W., Keski-Rahkonen, P., Assi, N., Ferrari, P., et al., A metabolomic study of biomarkers of meat and fish intake. *Am. J. Clin. Nutr.* 2017, *105*, 600–608.

[45] Hanhineva, K., Lankinen, M.A., Pedret, A., Schwab, U., et al., Nontargeted Metabolite Profiling Discriminates Diet-Specific Biomarkers for Consumption of Whole Grains, Fatty Fish, and Bilberries in a Randomized Controlled Trial. *J. Nutr.* 2015, *145*, 7–17.

[46] Zhang, A.., Mitchell, S.., Smith, R.., Dietary Precursors of Trimethylamine in Man: A Pilot Study. *Food Chem. Toxicol.* 1999, *37*, 515–520.

[47] Zhao, X., Zeisel, S.H., Zhang, S., Rapid LC-MRM-MS assay for simultaneous quantification of choline, betaine, trimethylamine, trimethylamine ***N*** -oxide, and creatinine in human plasma and urine. *Electrophoresis* 2015, *36*, 2207–2214.

[48] Tang, W.H.W., Wang, Z., Levison, B.S., Koeth, R.A., et al., Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk. *N. Engl. J. Med.* 2013, *368*, 1575–1584.

[49] Wang, Z., Tang, W.H.W., Buffa, J.A., Fu, X., et al., Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-N-oxide. *Eur. Heart J.* 2014, *35*, 904–910.

[50] Dumas, M.-E., Barton, R.H., Toye, A., Cloarec, O., et al., Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* 2006, *103*, 12511–6.

[51] Lloyd, A.J., Fave, G., Beckmann, M., Lin, W., et al., Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am. J. Clin. Nutr.* 2011, *94*, 981–991.

[52] Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., et al., Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 2011, *472*, 57–63.

[53] Schwarcz, R., Bruno, J.P., Muchowski, P.J., Wu, H.-Q., Kynurenines in the mammalian brain: when physiology meets pathology. *Nat. Rev. Neurosci.* 2012, *13*, 465–477.

[54] Hjerpsted, J.B., Ritz, C., Schou, S.S., Tholstrup, T., et al., Effect of cheese and butter intake on metabolites in urine using an untargeted metabolomics approach. *Metabolomics* 2014, *10*, 1176–1185.

[55] Hiratsuka, C., Fukuwatari, T., Sano, M., Saito, K., et al., Supplementing Healthy Women with up to 5.0 g/d of L-Tryptophan Has No Adverse Effects. *J. Nutr.* 2013, *143*, 859–866.

[56] Marinangeli, C.P.F., Jones, P.J.H., Pulse grain consumption and obesity: effects on energy expenditure, substrate oxidation, body composition, fat deposition and satiety. *Br. J. Nutr.* 2012, *108*, S46–S51.

[57] Hutson, S.M., Sweatt, A.J., Lanoue, K.F., Branched-chain amino acid metabolism: implications for establishing safe intakes. *J Nutr* 2005, *135*, 1557S–64S.

[58] Newsholme, P., Procopio, J., Ramos Lima, M.M., Pithon-Curi, T.C., et al., Glutamine and glutamate - Their central role in cell metabolism and function. *Cell Biochem. Funct.* 2003, *21*, 1–9.

[59] Stella, C., Beckwith-Hall, B., Cloarec, O., Holmes, E., et al., Susceptibility of human metabolic phenotypes to dietary modulation. *J. Proteome Res.* 2006, *5*, 2780–2788.

[60] Solanky, K.S., Bailey, N.J., Beckwith-Hall, B.M., Bingham, S., et al., Biofluid 1H NMR-based metabonomic techniques in nutrition research — metabolic effects of dietary isoflavones in humans. *J. Nutr. Biochem.* 2005, *16*, 236–244.

[61] Vázquez-Fresno, R., Llorach, R., Urpi-Sarda, M., Lupianez-Barbero, A., et al., Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *J. Proteome Res.* 2015, *14*, 531–540.

[62] Tharanathan, R.., Mahadevamma, S., Grain legumes—a boon to human nutrition. *Trends Food Sci. Technol.* 2003, *14*, 507–518.

[63] Holt, S., Carter, D., Tothill, P., Heading, R., et al., EFFECT OF GEL FIBRE ON GASTRIC EMPTYING AND ABSORPTION OF GLUCOSE AND PARACETAMOL. *Lancet* 1979, *313*, 636–639.

Table 1. Nutritional composition of the cooked test foods (300 g) used in the study.

Food (label)	Energy (kcal)	Protein (g)	Carbohidrate (g)	Fibre (g)	Sugars (g)	Fat (g)	SFA (g)
Lentils	267	20	40	8	1	1	0
Chickpeas	196	27	38	26	4	5	1
Beans	612	46	75	51	1	4	1
Pasta	367	12	74	4	1	1	0

SFA: saturated fatty acids

Table 2. Discriminant metabolites in serum and urine of subjects exposed to different pulses.

SERUM	Misclassification (%)	P-value (permutation)	Metabolites	
			Increased	**Decreased**
Lentils vs Control	9.1% (n=1)	0.02	Histidine Dimethylglycine 3-Hydroxyisovalerate	Glucose
Chickpeas vs Control	9.1% (n=1)	<0.01	Lysine	Glucose
Beans vs Control	0.0% (n=0)	0.01	2-Hydroxybutyrate Asparagine	Glucose Methylguanidine
Lentils vs Chickpeas	9.1% (n=1)	0.02	Glucose 3-Hydroxyisovalerate 2-Hydroxybutyrate	Histidine
Beans vs Lentils	0.0% (n=0)	0.02	2-Hydroxybutyrate Asparagine	-
Beans vs Chickpeas	0.0% (n=0)	0.02	2-Hydroxybutyrate Asparagine Histidine	Glucose

Table 2. Continued.

URINE	Misclassification (%)	P-value (permutation)	Metabolites	
			Increased	**Decreased**
Lentils vs Control	0.0% (n=0)	<0.01	Trigonelline Xanthurenate 3-Methylhistidine[a] Unknown A	Urea glucose[a]
Chickpeas vs Control	0.0% (n=0)	<0.01	Ascorbate Dimethylglycine[a] Lactate Lysine Trigonelline Trimethylamine Xanthurenate Unknown A 3-Methylhistidine[a]	Methylmalonate 3-Hydroxyisovalerate N-Acetylglutamine[a] glucose[a]
Beans vs Control	0.0% (n=0)	0.02	Lactate Leucine[a] Lysine Trigonelline Trimethylamine 3-Methylhistidine Unknown A	glucose[a]
Lentils vs Chickpeas	0.0% (n=0)	<0.01	Glutamine[a] Lysine Methanol N-Acetylglutamine[a]	Ascorbate Methylamine Trigonelline 3-Methylhistidine[a]
Beans vs Lentils	0.0% (n=0)	0.01	Leucine[a] Valine Lysine Trigonelline 3-Methylhistidine[a]	Unknown A
Beans vs Chickpeas	9.1% (n=1)	<0.01	3-Aminoisobutirate 3-Hydroxyisovalerate Histidine Leucine[a] Lysine Trigonelline 3-Methylhistidine[a]	Trimethylamine Unknown A

[a]Result in agreement with Madrid-Gambin et al. [22]

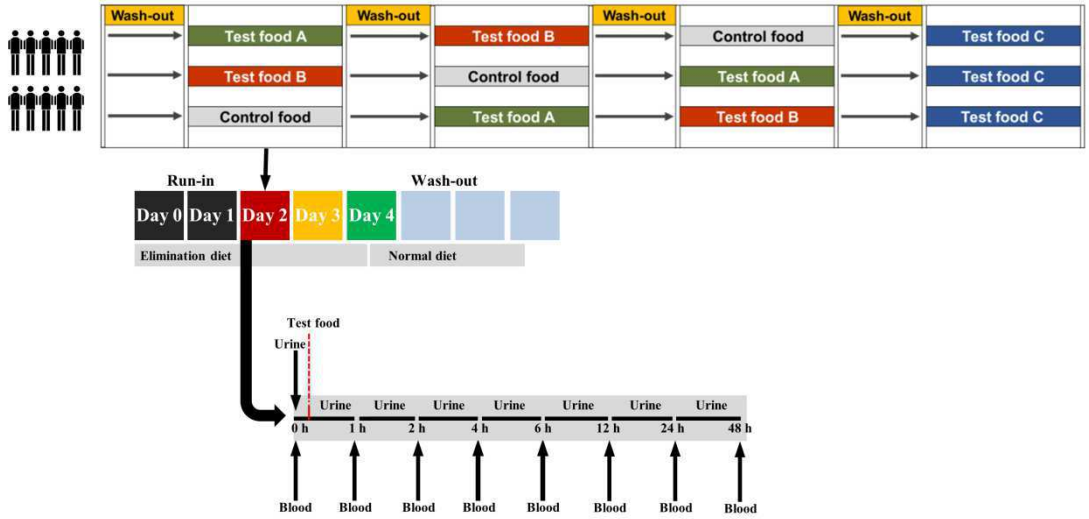


Figure 1. Study design. Test food A, B or C corresponded with the pulses studied: lentils, chickpeas or beans, respectively. Control food corresponded with pasta.

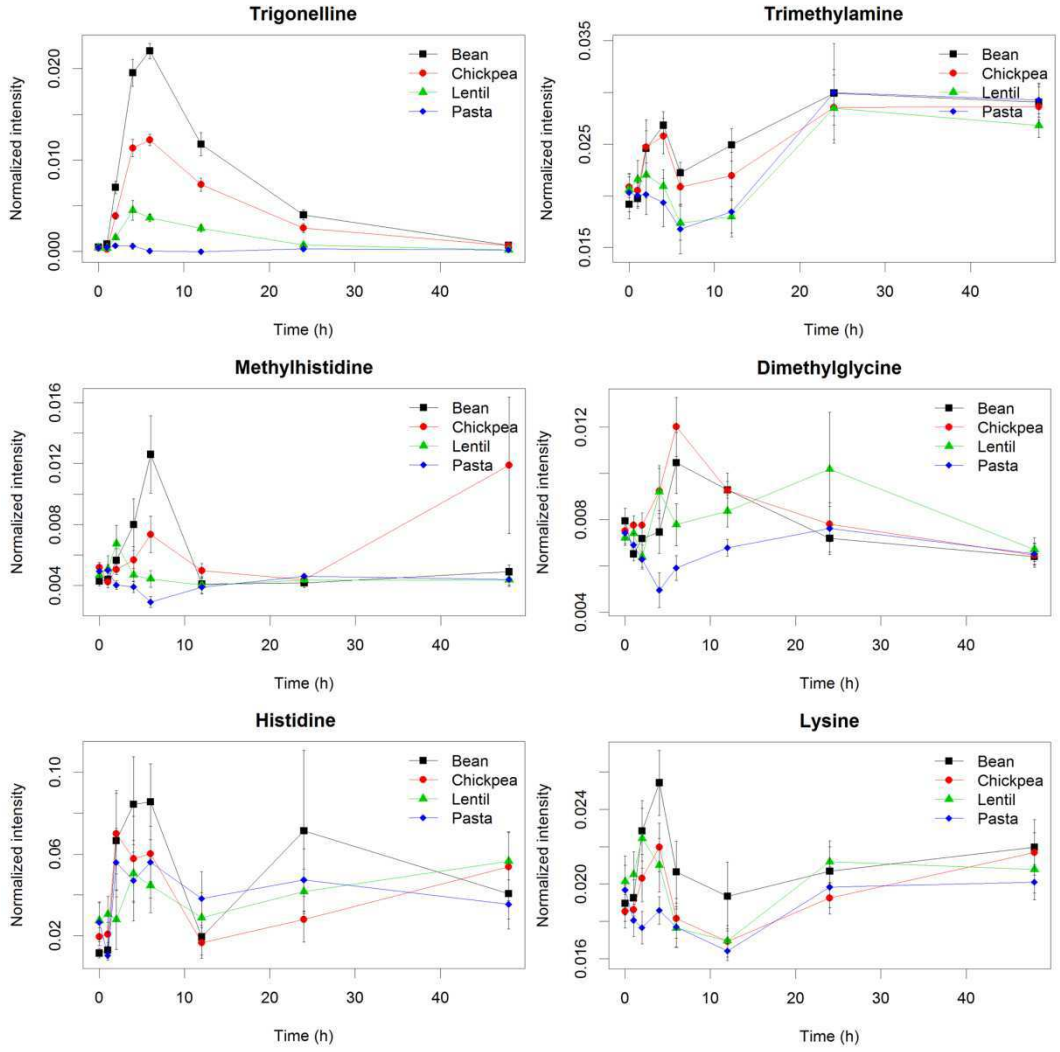


Figure 2. Urinary excretion profiles of selected metabolites.

Supplemental Material

Untargeted ^{1}H-NMR based metabolomics analysis of urine and serum profiles after consumption of dietary pulses: an extended meal study.

Supplemental Table 1. Full meal scheme followed in the study.

	Day before intervention	Intervention day	Day after intervention
Breakfast	-	Lentils / chickpeas / beans / pasta[a]	Sandwich with cheese and cooked ham Yogurt
Lunch	-	Rice with cooked ham and eggs Yogurt	Rice with chicken Yogurt
Snack	-	Yogurt Biscuits (4)	Yogurt Biscuits (4)
Dinner	Rice with chicken Yogurt	Couscous with turkey Yogurt	Pasta with turkey and bacon Yogurt

[a]Control meal.

Supplemental Table 2. Concentrations (mM) of identified compounds from 1H-NMR profiling of extracts of the pulses used in the study.

Compound	Beans	Chickpeas	Lentils
2-Oxoglutarate	0.009±0.000[a]	0.019±0.003[b]	0.003±0.000[c]
4-Aminobutyrate	0.205±0.007[a]	0.017±0.002[b]	0.064±0.005[c]
Acetate	0.191±0.010[a]	0.109±0.016[b]	0.123±0.004[b]
Adenosine	0.020±0.001[a]	0.024±0.002[b]	0.011±0.000[c]
Alanine	0.101±0.004[a]	0.006±0.000[b]	0.036±0.002[c]
Arabinose	0.005±0.001[a]	0.007±0.001[b]	0.008±0.001[b]
Arginine	0.313±0.018[a]	0.152±0.021[b]	0.125±0.013[b]
Ascorbate	0.000±0.000[a]	0.191±0.025[b]	0.000±0.000[a]
Asparagine	0.077±0.003[a]	0.018±0.000[b]	0.033±0.001[c]
Choline	0.280±0.016[a]	0.072±0.013[b]	0.070±0.007[b]
Citrate	0.066±0.013[a]	0.021±0.001[b]	0.129±0.018[c]
Formate	0.015±0.001[a]	0.025±0.004[b]	0.025±0.003[b]
Fructose	0.140±0.001[a]	0.170±0.021[b]	0.109±0.007[c]
Galactose	0.023±0.002[a]	0.067±0.008[b]	0.080±0.005[c]
Glucuronate	0.034±0.002[a]	0.037±0.002[a]	0.014±0.001[b]
Glutamate	0.048±0.002[a]	0.097±0.013[b]	0.051±0.008[a]
Glutamine	0.021±0.001[a]	0.008±0.003[b]	0.010±0.001[b]
Histidine	0.058±0.004[a]	0.005±0.000[b]	0.008±0.001[b]
Isoleucine	0.008±0.000[a]	0.004±0.001[b]	0.004±0.001[b]
Lactate	0.261±0.014[a]	0.026±0.004[b]	0.029±0.002[b]
Leucine	0.021±0.001[a]	0.007±0.001[b]	0.006±0.001[b]
Lysine	0.013±0.000[a]	0.011±0.002[b]	0.008±0.002[b]
Malate	0.056±0.005[a]	0.037±0.005[b]	0.039±0.001[b]
Maltitol	0.012±0.000[a]	0.517±0.042[b]	0.224±0.024[c]
Mannose	0.005±0.000[a]	0.005±0.001[a]	0.005±0.000[a]
Methanol	0.037±0.004[a]	0.064±0.007[b]	0.085±0.005[c]
Phenylalanine	0.018±0.001[a]	0.018±0.002[a]	0.013±0.000[b]
Pipecolate	0.642±0.038[a]	0.000±0.000[b]	0.000±0.000[b]
Proline	0.041±0.001[a]	0.036±0.001[a]	0.056±0.007[b]
Pyroglutamate	0.171±0.011[a]	0.045±0.003[b]	0.035±0.004[b]
Raffinose	1.518±0.002[a]	1.418±0.159[a]	0.599±0.081[b]
Stachyose	4.773±0.087[a]	3.027±0.481[b]	2.483±0.362[b]
Succinate	0.044±0.002[a]	0.015±0.002[b]	0.017±0.001[b]
Sucrose	1.733±0.081[a]	1.254±0.051[b]	0.344±0.024[c]
Threonine	0.017±0.001[a]	0.009±0.000[b]	0.009±0.001[b]
Trigonelline	0.156±0.005[a]	0.081±0.010[b]	0.024±0.001[c]
Tryptophan	0.030±0.001[a]	0.031±0.003[a]	0.006±0.001[b]
Tyrosine	0.010±0.001[a]	0.014±0.001[b]	0.006±0.001[c]
Valine	0.023±0.001[a]	0.006±0.001[b]	0.009±0.001[c]
myo-Inositol	0.046±0.003[a]	0.075±0.010[b]	0.040±0.008[a]
γ-Glutamylphenylalanine	0.003±0.000[a]	0.040±0.006[b]	0.005±0.000[a]

Values are expressed as mean ± SD. Values with different letters (a, b, c) are significantly different between foods (one-way ANOVA followed by Fisher's LSD correction, $P < 0.05$).

Supplemental Table 3. Comparison of metabolites across the treatment (control, lentils, chickpeas or white beans), time (t_{0h}-t_{48h}) and interaction treatment-time.

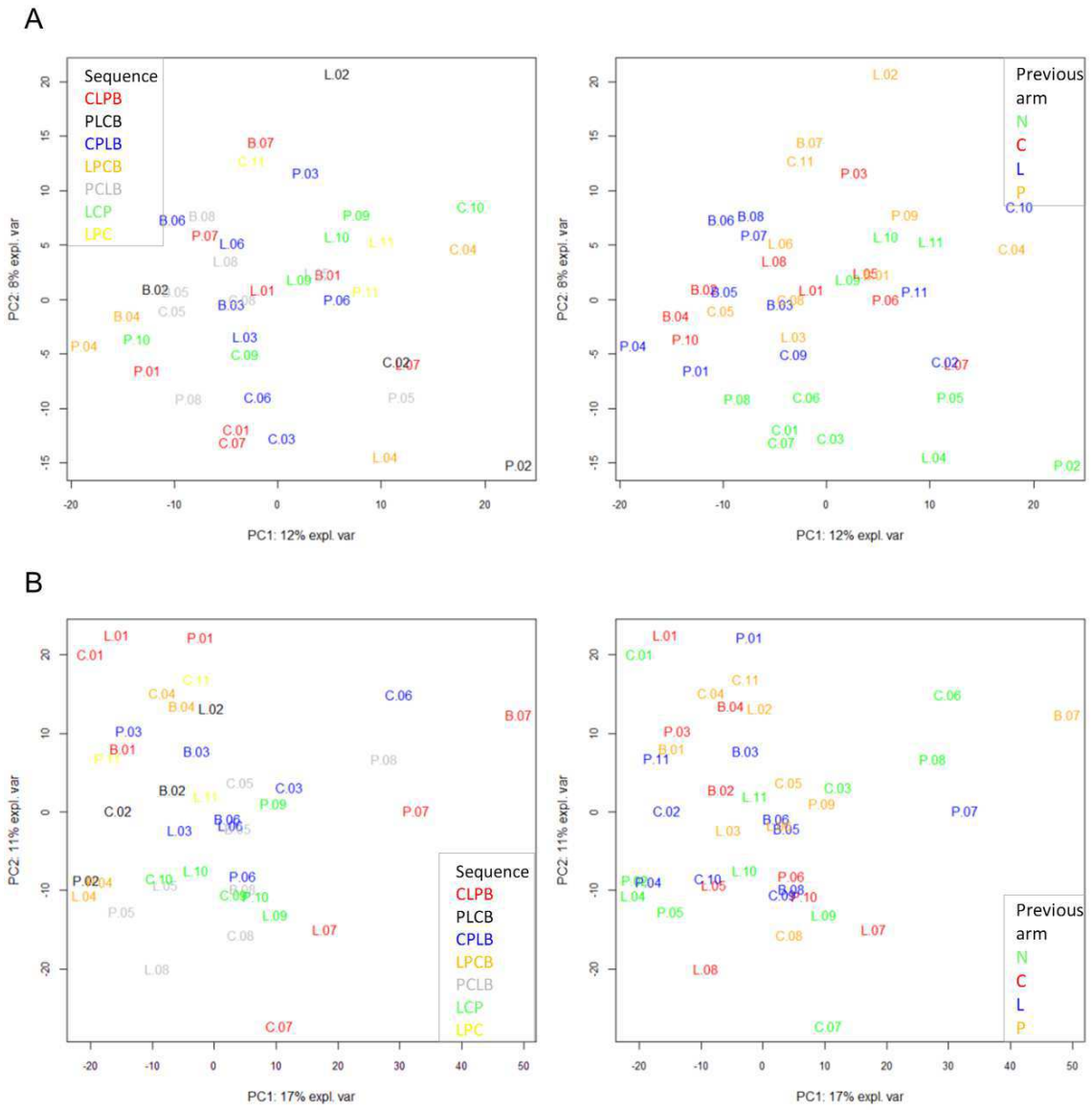
	Treatment	**Time**	**Treatment * time**
3-Aminoisobutyrate	0.94	<0.01	<0.01
3-Hydroxyisovalerate	0.94	<0.01	<0.01
3-Methylhistidine	<0.01	<0.01	<0.01
Ascorbate	<0.01	0.02	<0.01
Dimethylglycine	<0.01	<0.01	<0.01
Glucose	0.06	<0.01	<0.01
Glutamine	0.13	<0.01	<0.01
Histidine	0.94	<0.01	<0.01
Lactate	0.87	<0.01	<0.01
Leucine	0.94	<0.01	<0.01
Lysine	0.87	<0.01	<0.01
Methanol	0.02	<0.01	<0.01
Methylamine	0.94	<0.01	<0.01
Mathylmalonate	0.85	<0.01	<0.01
n-Acetylglutamine	0.13	<0.01	<0.01
Trigonelline	<0.01	<0.01	<0.01
Trimethylamine	0.68	<0.01	<0.01
Urea	0.94	<0.01	<0.01
Valine	0.94	<0.01	<0.01
Xanthurenate	0.04	<0.01	<0.01
Unknown A	<0.01	<0.01	<0.01

Examined by repeated measures ANOVA test. The Benjamini-Hochberg false discovery rate (FDR) procedure was used to adjust for multiple testing. Statistical significance was considered at an FDR-adjusted p-value <0.05.

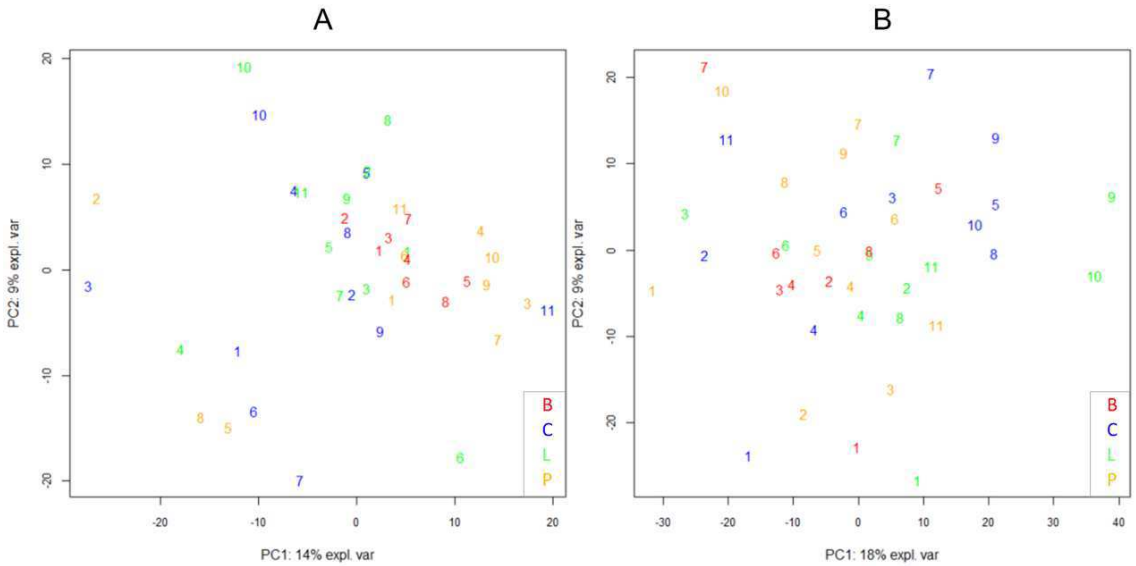
Supplemental Table 4. P-values of comparisons of selected biomarkers across the treatment (control, lentils, chickpeas or white beans) and individual time points (t_{0h}-t_{48h}).

Contrast	Time	Trigonelline	Trimethylamine	3-MH	Dimethylglycine	Histidine	Lysine
Bean - Chickpea	0	0.63	0.34	0.05	0.49	0.11	0.75
Bean - Lentil	0	0.80	0.50	0.27	0.25	0.09	0.49
Bean - Pasta	0	0.84	0.63	0.15	0.46	0.12	0.67
Chickpea - Lentil	0	0.69	0.90	0.19	0.50	0.44	0.33
Chickpea - Pasta	0	0.72	0.80	0.63	0.87	0.50	0.48
Lentil - Pasta	0	0.97	0.90	0.50	0.70	0.95	0.81
Bean - Chickpea	1	0.01	0.48	0.73	0.01	0.27	0.61
Bean - Lentil	1	0.04	0.33	0.04	0.11	0.07	0.42
Bean - Pasta	1	0.12	0.83	0.52	0.50	0.57	0.36
Chickpea - Lentil	1	0.37	0.57	0.09	0.57	0.35	0.19
Chickpea - Pasta	1	0.21	0.75	0.45	0.19	0.10	0.62
Lentil - Pasta	1	0.60	0.47	0.94	0.47	0.03	0.10
Bean - Chickpea	2	<0.01	0.97	0.47	0.45	0.91	0.22
Bean - Lentil	2	<0.01	0.47	0.45	0.29	0.18	0.86
Bean - Pasta	2	<0.01	0.18	0.05	0.16	0.72	<0.01
Chickpea - Lentil	2	<0.01	0.32	0.18	0.07	0.09	0.30
Chickpea - Pasta	2	<0.01	0.06	0.02	0.02	0.59	0.09
Lentil - Pasta	2	<0.01	0.52	0.03	0.87	0.21	0.01
Bean - Chickpea	4	<0.01	0.63	0.22	0.19	0.39	0.11
Bean - Lentil	4	<0.01	<0.01	0.06	0.23	0.21	0.05
Bean - Pasta	4	<0.01	0.01	0.02	0.04	0.22	<0.01
Chickpea - Lentil	4	<0.01	0.04	0.31	0.98	0.78	0.62
Chickpea - Pasta	4	<0.01	0.03	0.06	<0.01	0.71	0.02
Lentil - Pasta	4	<0.01	0.58	0.16	<0.01	0.88	0.13
Bean - Chickpea	6	<0.01	0.50	0.06	0.40	0.27	0.21
Bean - Lentil	6	<0.01	0.01	<0.01	0.10	0.07	0.12
Bean - Pasta	6	<0.01	0.04	<0.01	<0.01	0.25	0.14
Chickpea - Lentil	6	<0.01	0.15	0.03	0.01	0.42	0.73
Chickpea - Pasta	6	<0.01	0.17	<0.01	<0.01	0.86	0.76
Lentil - Pasta	6	<0.01	0.84	0.02	0.07	0.60	0.97
Bean - Chickpea	12	<0.01	0.28	0.14	0.98	0.80	0.23
Bean - Lentil	12	<0.01	<0.01	0.91	0.23	0.53	0.22
Bean - Pasta	12	<0.01	0.03	0.72	<0.01	0.24	0.13
Chickpea - Lentil	12	<0.01	0.15	0.19	0.38	0.39	0.98
Chickpea - Pasta	12	<0.01	0.29	0.08	<0.01	0.15	0.60
Lentil - Pasta	12	<0.01	0.86	0.86	0.04	0.60	0.52
Bean - Chickpea	24	0.05	0.81	0.68	0.52	0.29	0.36
Bean - Lentil	24	<0.01	0.77	0.58	0.24	0.47	0.77
Bean - Pasta	24	<0.01	0.99	0.19	0.73	0.57	0.62
Chickpea - Lentil	24	<0.01	0.98	0.97	0.36	0.38	0.17
Chickpea - Pasta	24	<0.01	0.71	0.59	0.89	0.31	0.67
Lentil - Pasta	24	0.02	0.59	0.29	0.34	0.76	0.39
Bean - Chickpea	48	0.88	0.86	0.12	0.93	0.53	0.88
Bean - Lentil	48	0.02	0.27	0.33	0.61	0.41	0.54
Bean - Pasta	48	0.02	0.93	0.45	0.84	0.76	0.29
Chickpea - Lentil	48	<0.01	0.42	0.09	0.72	0.90	0.58
Chickpea - Pasta	48	<0.01	0.79	0.10	0.93	0.37	0.26
Lentil - Pasta	48	0.95	0.22	0.94	0.78	0.26	0.67

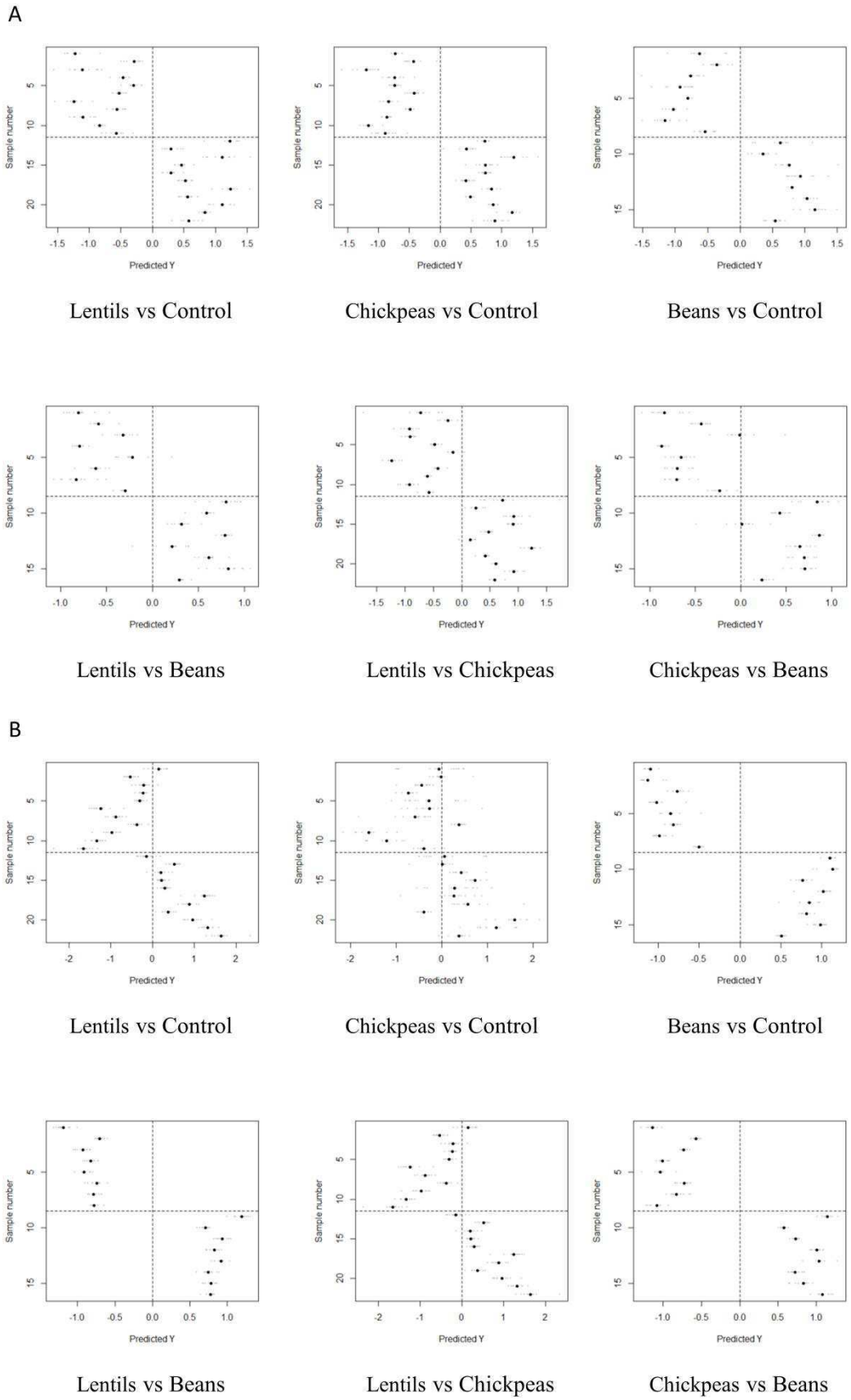
Examined by repeated measures ANOVA test. The Benjamini-Hochberg false discovery rate (FDR) procedure was used to adjust for multiple testing. Statistical significance was considered at an FDR-adjusted p-value <0.05.



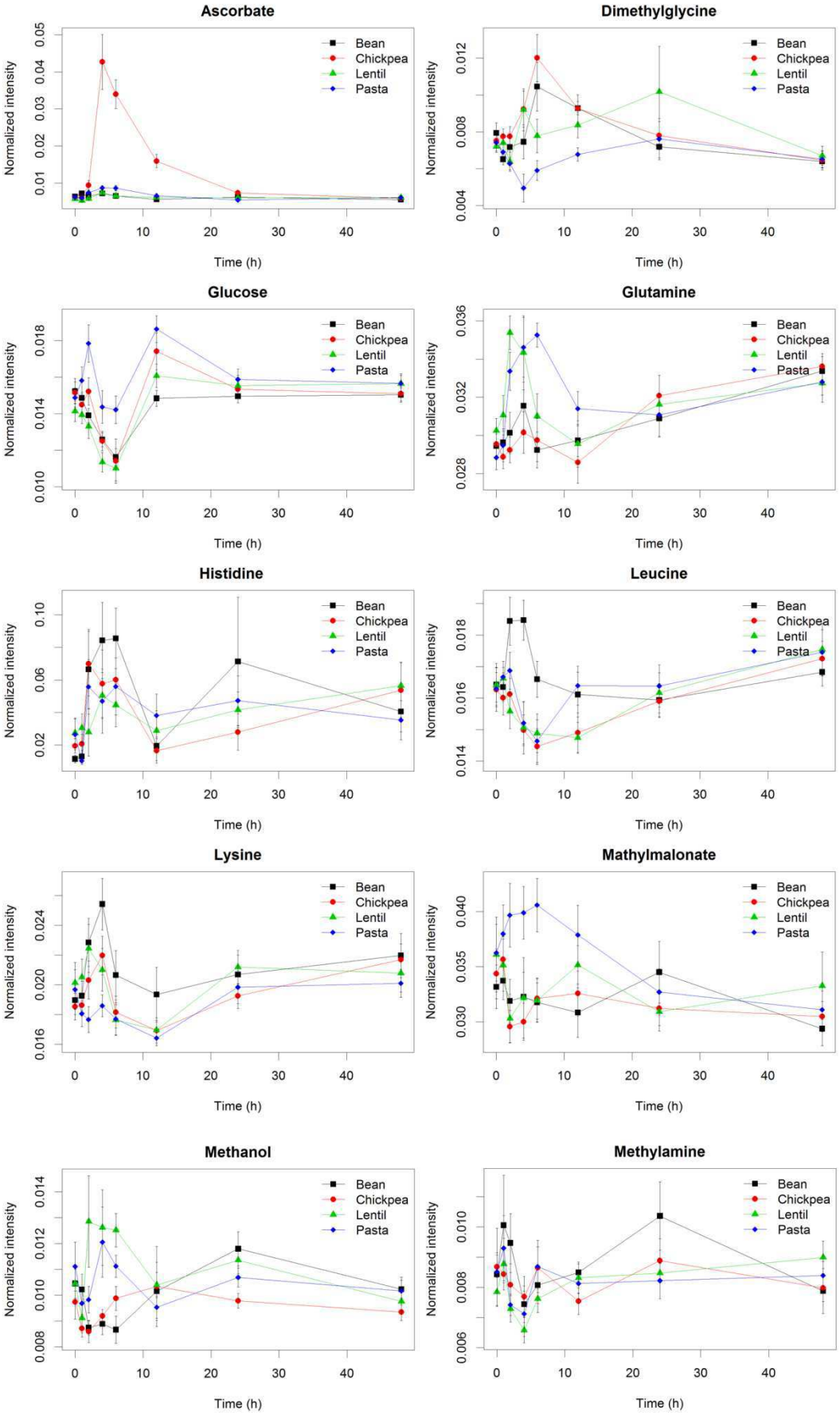
Supplemental Figure 1. Principal component analysis (PCA) of baseline points of each subject colored by the sequence followed in the study (left) and the immediately previous intervention (right) in serum (A) and urine (B) samples. B: beans, C: control, L: lentils, N: no intervention, P: chickpeas. Sequence is described by initial of each intervention. Subjects were labelled with numbers (1-11).

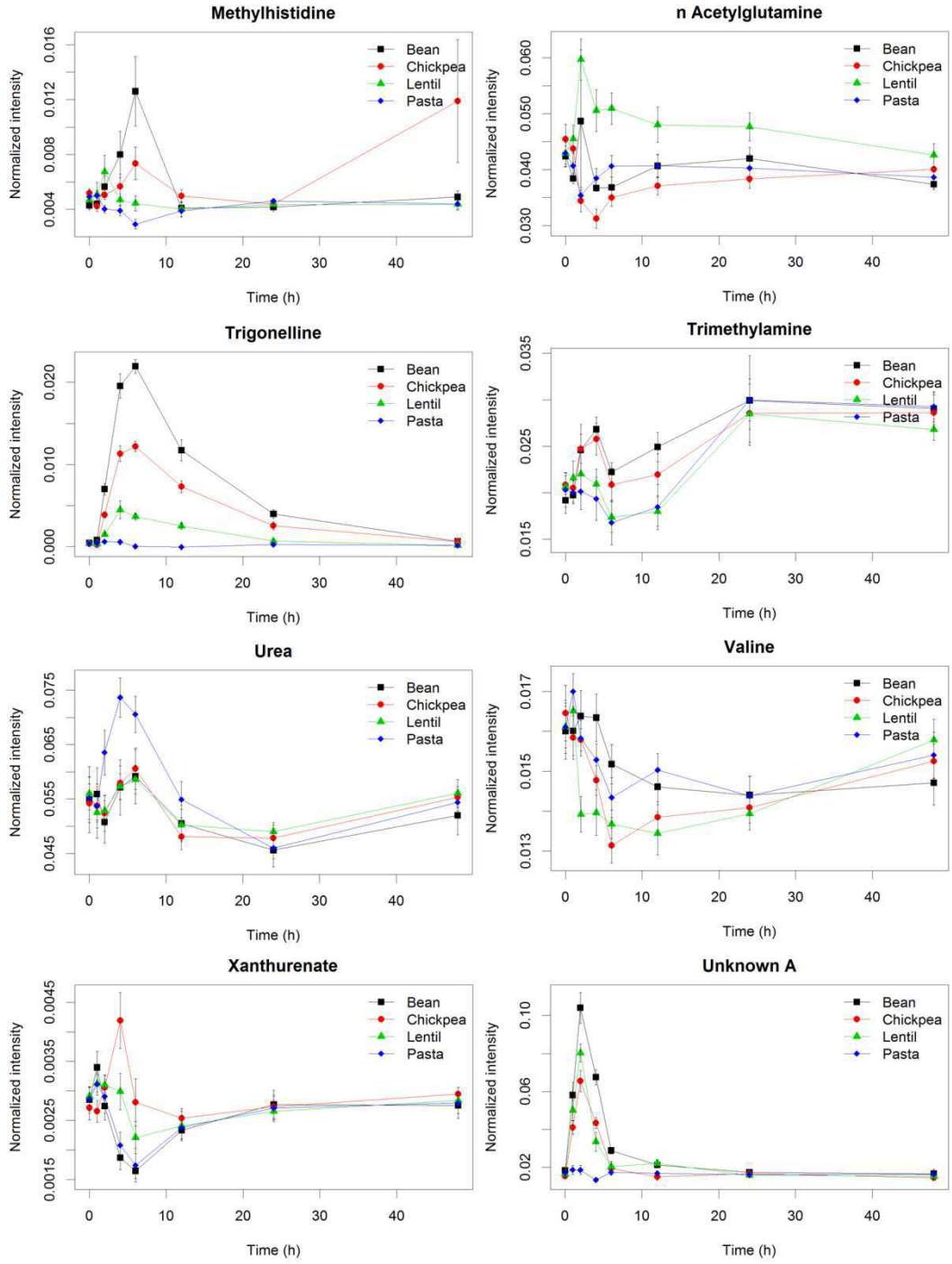


Supplemental Figure 2. Principal component analysis (PCA) of AUCs used in the multivariate analysis in serum (A) and urine (B) samples. Samples were coloured by intervention. B: beans, C: control, L: lentils, P: chickpeas. Subjects were labelled with numbers (1-11).



Supplemental Figure 3. Misclasification plots from the PLS modelling in (A) urine and (B) serum data.


Ascorbate
Dimethylglycine
Glucose
Glutamine
Histidine
Leucine
Lysine
Mathylmalonate
Methanol
Methylamine
Bean
Chickpea
Lentil
Pasta
Normalized intensity
Time (h)



Supplemental Figure 4. Response of significant urinary metabolites until 48 h.

7.2. Estudio de intervención nutricional en humanos tras el consumo agudo y sostenido de una bebida de café con alto contenido en compuestos bioactivos.

Los resultados de este estudio se publicaron en la revista *Food Research International* indexada en el *Journal Citation Reports* con un factor de impacto de 3,182 (2015) y situada en el primer cuartil de la categoría *Food Science & Technology* (18/125) (2015).

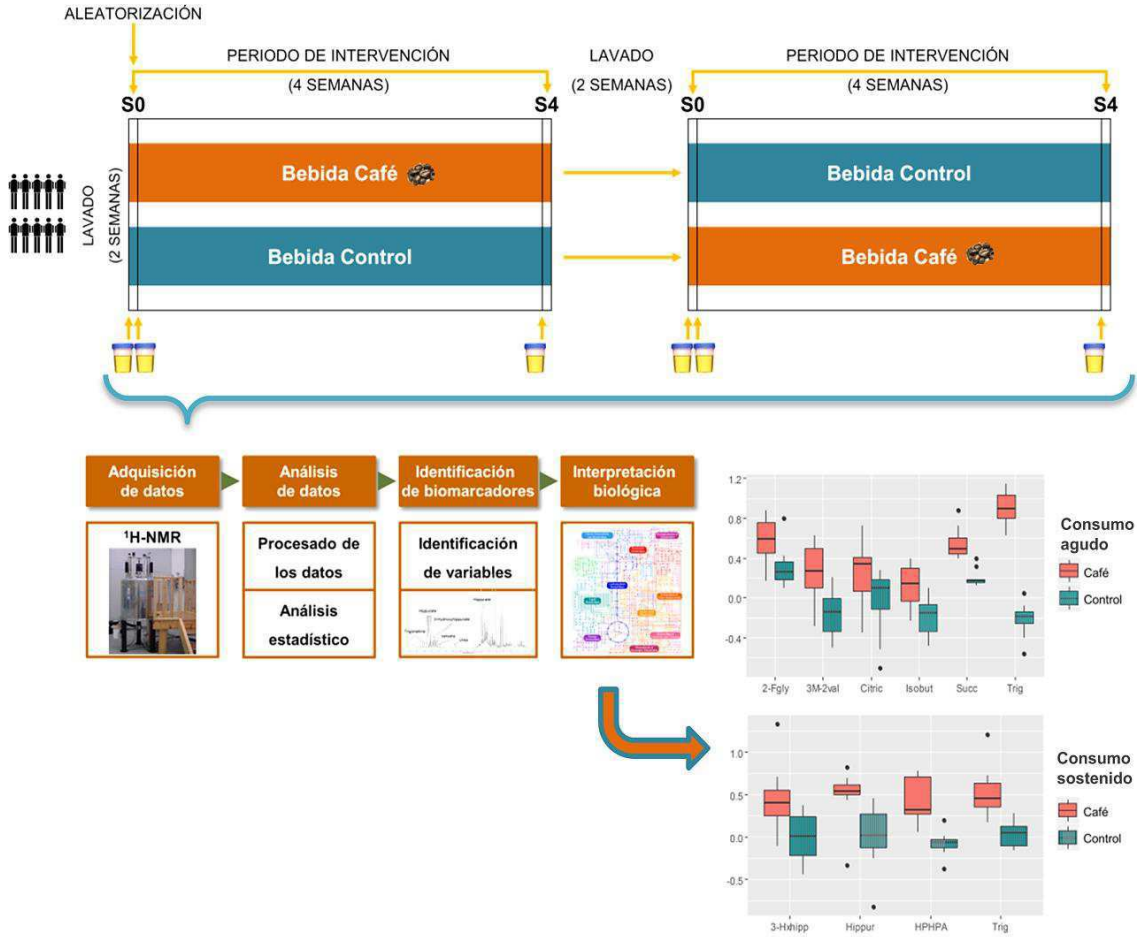
Francisco Madrid-Gambin, Mar Garcia-Aloy, Rosa Vázquez-Fresno, Esteban Vegas-Lozano, M. Carmen Ruiz de Villa Jubany, Koichi Misawa, Tadashi Hase, Akira Shimotoyodome, Cristina Andres-Lacueva. Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects. *Food Research International*. 2016; 89, 1064–1070. doi:10.1016/j.foodres.2016.03.038.

Resumen:

Numerosos estudios sugieren que el café ejerce efectos beneficiosos sobre la salud. Sin embargo, el papel específico que desempeñan únicamente los principales polifenoles del café, concretamente los ácidos clorogénicos (ACGs) y sin interacción con la cafeína, son todavía muy poco conocidos. En el presente estudio se realizó un análisis metabolómico no dirigido por resonancia magnética nuclear de protón (^{1}H-RMN) para evaluar el efecto de una bebida de café rica en ACG sobre el metaboloma urinario. Diez hombres sanos fueron incluidos en un estudio de intervención aguda y sostenida, aleatorizado, cruzado, y a doble ciego. El estudio consistió en una ingesta puntual de una bebida de café rica en ACGs (CEB, del inglés *coffee extract beverage*: 223 mg/100 ml de ACG) o una bebida de control con igual dosis de cafeína. Luego, los participantes consumieron a diario la bebida a la que fueron asignados para el estudio agudo durante los siguientes 28 días. Posteriormente, y tras un período de "*washout*" de 14 días, se repitió este mismo protocolo alternando los individuos entre los grupos. Se recogieron las muestras de orina en ayunas en el primer y último día de cada periodo del estudio y se analizaron usando un enfoque metabolómico no dirigido mediante ^{1}H-RMN (estudio tras el consumo sostenido). Además, también se analizaron las muestras de orina postprandiales de hasta 4 horas después de la

primera ingesta de cada bebida (estudio agudo). Para el análisis estadístico se utilizaron enfoques uni- y multi-variantes con el objetivo de fortalecer los resultados. En el enfoque multivariante, se utilizó el análisis discriminante de mínimos cuadrados parciales en varios niveles (*multilevel PLS-DA*) para comparaciones pareadas a través del diseño cruzado del estudio. Para el análisis univariante, se aplicó un modelo estadístico cruzado para evaluar las diferencias significativas entre los grupos en los cambios observados en el metaboloma. En este modelo estadístico, se consideraron como covariables el individuo, el período de intervención y la secuencia en la que se asignaron las intervenciones. El consumo agudo de la CEB dio como resultado una excreción elevada de 2-furoilglicina, así como compuestos endógenos tales como ácido succínico, ácido cítrico, ácido 3-metil-2-oxovalérico y ácido isobutírico. El consumo sostenido de la CEB mostró un aumento en los niveles urinarios de los compuestos derivados de la microbiota, tales como ácido hipúrico, ácido 3-(3-hidroxifenil)-3-hidroxipropiónico y ácido 3-hidroxihipúrico. Además, se observó un aumento estadísticamente significativo en la excreción urinaria de trigonelina tras la ingesta aguda y sostenida de la CEB. Este compuesto también fue observado en la composición propia de la bebida de café, lo que podría indicar una excreción directa de este compuesto, sugiriendo la ausencia de metabolismo y variación interindividual en su excreción. Este estudio se encuentra registrado en el *International Standard Randomized Controlled Trial Number* con el código ISRCTN15516017.

Figura 5. Gráfico-resumen del estudio de intervención nutricional en humanos tras el consumo agudo sostenida de una bebida de café con alto contenido en compuestos bioactivos.



Food Research International 89 (2016) 1064–1070

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects

Francisco Madrid-Gambin [a], Mar Garcia-Aloy [a,1], Rosa Vázquez-Fresno [a,*,1], Esteban Vegas-Lozano [b], M. Carmen Ruiz de Villa Jubany [b], Koichi Misawa [c], Tadashi Hase [c], Akira Shimotoyodome [c], Cristina Andres-Lacueva [a,**]

[a] Biomarkers & Nutrimetabolomic Lab., Nutrition and Food Science Department, XaRTA, INSA, Campus Torribera, Pharmacy Faculty, University of Barcelona, Barcelona, Spain
[b] Department of Statistics, University of Barcelona, Barcelona, Spain
[c] Biological Science Laboratories, Kao Corporation, 2606, Akabane, Ichikai, Haga, Tochigi 321-3497, Japan

ARTICLE INFO

Article history:
Received 29 January 2016
Accepted 28 March 2016
Available online 9 April 2016

Keywords:
Coffee
Chlorogenic acids
Metabolomics
NMR
Biomarkers
Multilevel PLS-DA

ABSTRACT

Several studies suggest that coffee has some benefits for health; however, little is known about the specific role of the main polyphenol compounds of coffee, chlorogenic acids (CGAs), without caffeine interaction. A ^{1}H-Nuclear Magnetic Resonance (^{1}H-NMR)-based metabolomics approach was used to assess the effect of CGAs from coffee on the human urine metabolome. Ten male volunteers participated in a dietary crossover randomized intervention study with a rich CGAs coffee extract beverage (CEB: 223 mg/100 ml of CGAs). The study consisted of a daily intake of CEB or a control beverage with equal caffeine dose during 28 days. Fasting urines collected at the first and last days of each period of the study were analyzed using an CGAs untargeted ^{1}H-NMR approach. Additionally, 4-hour postpandrial urines after the first intake of each beverage were also analyzed. Uni- and multi-variate statistic approaches were used to strengthen the results. Multilevel partial least squares discriminant analysis (ML-PLS-DA) was used to paired comparisons across the crossover design. A further univariate analysis model for crossover studies was performed to assess the significant changes. Acute consumption of CEB resulted in high excretion of 2-furoylglycine, likewise endogenous compounds such as succinic, citric, 3-methyl-2-oxovaleric and isobutyric acids. Sustained consumption of CEB showed an increase of microbiota-derived compounds such as hippuric, 3-(3-Hydroxyphenyl)-3-hydroxypropionic and 3-hydroxyhippuric acids in urine. Moreover, trigonelline was found in urine after both acute and sustained intakes, as well as in the composition of the beverage exhibiting a direct excretion of this biomarker without any biotransformation, suggesting a non-interindividual variation.

1. Introduction

Coffee is one of the most widely consumed beverages worldwide, and therefore there are a huge number of studies concerning its benefits for health (Cowan et al., 2014; Guertin et al., 2015; R M van Dam & Hu, 2005). Coffee beans contain a large variety of biologically active compounds, the predominant ones being chlorogenic acids (CGAs) and caffeine (Johnston, Clifford, & Morgan, 2003). Several epidemiological researches suggest that coffee consumption may help prevent chronic diseases, including type 2 diabetes mellitus (Van Dam & Feskens, 2002), obesity and metabolic syndrome (Nordestgaard, Thomsen, & Nordestgaard, 2015), Parkinson's disease (Sääksjärvi et al., 2008) and liver disease (La Vecchia, 2005). Moreover, coffee is rich in many polyphenols and diterpenes (Urgert & Katan, 1997), which have antioxidant properties (Yanagimoto, Ochi, Lee, & Shibamoto, 2004) and may also mitigate harmful gut microbiota species with its regular consumption (Cowan et al., 2014). Although these beneficial effects are attributed to caffeinated coffee, high intakes of decaffeinated coffee have also been associated with a reduced risk of type 2 diabetes mellitus (Ding, Bhupathiraju, Chen, van Dam, & Hu, 2014; Huxley et al., 2009) and positive effects on cognition and psychomotor behavior (Shukitt-Hale, Miller, Chu, Lyle, & Joseph, 2013), indicating that some compounds in coffee other than caffeine may have a protective effect. Few

Abbreviations: CEB, coffee extract beverage; CGAs, chlorogenic acids; FID, free induction decay; HPHPA, 3-(3-Hydroxyphenyl)-3-hydroxypropionic acid; LR, loading rank; ML-PLS-DA, multilevel partial least squares discriminant analysis; NMP, *N*-methylpyridinium; NMR, Nuclear Magnetic Resonance; PCA, principal component analysis; TCA, tricarboxylic acid; TSP, 3-(trimethylsilyl)-proprionate-2,2,3,3-d_4.

* Correspondence to: R. Vázquez-Fresno, Biomarkers & Nutrimetabolomic Lab., Nutrition and Food Science Department, Campus Torribera, Pharmacy Faculty, University of Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.
** Correspondence to: C. Andres-Lacueva, Biomarkers & Nutrimetabolomic Lab., Nutrition and Food Science Department, Campus Torribera, Pharmacy Faculty, University of Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.
E-mail addresses: rosavfr@gmail.com (R. Vázquez-Fresno), candres@ub.edu (C. Andres-Lacueva).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.foodres.2016.03.038
0963-9969/

interventional studies focused on the effect of bioactive compounds in coffee in both the short- and long-term have been able to provide an explanation for physiological mechanisms. However, in most epidemiological studies there is no distinction between caffeinated coffee consumers and newer or occasional decaffeinated coffee consumers (Higdon & Frei, 2006). Therefore the role of CGAs alone in this beverage is still unclear.

The application of metabolomics in the nutrition field can provide new information on dietary components, with the discovery of new biomarkers of food exposure, thereby revealing potential modifications in diet-related pathways in early disease stages likewise in healthy individuals (Puiggròs, Canela, & Arola, 2015; Scalbert et al., 2014). Nuclear Magnetic Resonance (NMR) in metabolomics is a fast, robust, reproducible, non-destructive technique (Wishart, 2008). Metabolomics allows a global description of metabolites that gives detailed information on metabolic pathways and in turn on biological processes, thereby clarifying associations with health benefits and elucidating underlying mechanisms (Brennan, 2014). Untargeted metabolomics is generally used for the discovery of metabolites and modifications in global metabolic pathways, whereas targeted metabolomics is focused on specific metabolic pathways (Xu, Wang, Ho, & Ong, 2014).

Several biomarkers of coffee intake have previously been proposed, including trigonelline, 2-furoylglycine, caffeine and several polyphenols (Heinzmann, Holmes, Kochhar, Nicholson, & Schmitt-Kopplin, 2015; Ito et al., 2005; Lang, Wahl, Stark, & Hofmann, 2011; Lloyd et al., 2013; Rothwell, Fillâ Tre, et al., 2014; Stalmach et al., 2009). Coffee is the main dietary source of CGAs (Clifford, 2000), but although CGAs and caffeic acid are the main phenolic compounds identified in urine after coffee intake in short-term (Ito et al., 2005), derived compounds of polyphenol metabolism are dependent on high interindividual variation in humans in long-term because of gut microbiota (Hervert-Hernández & Goñi, 2011).

The present study was conducted to determine whether CGAs from coffee may impact on the human urine metabolome through an untargeted metabolomic approach, and to identify the changes on the metabolome after both acute and sustained consumptions on healthy human volunteers.

2. Material and methods

2.1. Subjects and study design

Ten healthy male between 25 and 44 years old with a body mass index of 23.4 ± 2.1 kg/m^2 (mean ± SD) participated in a randomized, double-blind, placebo-controlled, crossover clinical trial. Exclusion criteria included caffeine intoxication, intake of CGAs supplements, serious illness (such as heart disease, kidney disease or diabetes) and food allergies. The study protocol was approved by the Human Research Ethics Committee of Biological Science Laboratories of the KAO Corporation (ref: 507-20131218). This clinical trial was registered as International Standard Randomized Controlled Trial Number 15516017.

During the washout and the study periods, the subjects were forbidden to consume coffee and tea beverages other than test drink. Subjects were not allowed to consume alcoholic beverages and to practice exercise from 2 days before the beginning of the intervention. After a 14-day washout period, subjects were asked to consume a coffee extract beverage (CEB) containing 223 mg/100 ml of CGAs (KAO Corporation, Japan) or a control beverage (caffeine-containing beverage). The daily dose of caffeine was similar between the intervention and control groups (see Table S1, Supporting information). Then, the participants consumed the corresponding beverage every day for the next 28 days (period I). The same procedure was repeated switching the individuals between the groups (period II) after a second 14-day washout period, in accordance with the crossover design. To analyze acute consumption, urine samples were collected during the first 4 postprandial hours after beverage intake on the first day of the intervention. For the analysis of sustained consumption, fasting urines on the first and last days of each period of the study were collected. All urine samples were stored in aliquots at −80 °C prior to analysis.

2.2. Sample preparation

Both urine and beverage samples were thawed, vortexed and centrifuged at 13,200 rpm for 5 min. The supernatant (600 μl) from each sample was mixed with an internal standard solution [120 μl, consisting of 0.1% 3-(trimethylsilyl)-proprionate-2,2,3,3-d_4 (TSP), chemical shift reference), 2 mM of sodium azide (NaN_3, bacteriostatic agent) and 1.5 M KH_2PO_4 in 99% deuterium water (D_2O)]. The optimized pH of the buffer was set at 7.0, with a potassium deuteroxide solution, to minimize variations in the chemical shifts of the NMR resonances. This mixture was transferred to a 5 mm NMR tube.

2.3. ^{1}H-NMR data acquisition and processing

The ^{1}H-NMR urinary spectra were acquired on a Varian-Inova-500 MHz NMR spectrometer with presaturation of the water resonance using a NOESYPRESAT pulse sequence. During the acquisition, the internal temperature was kept constant at 298 K. An exponential window function was applied to the free induction decay (FID) with a line-broadening factor of 0.3 Hz prior to the Fourier transformation. For each sample, FIDs were collected into 32 K data points, 128 scans, with a spectral width of 14 ppm, an acquisition time of 2 s, a relaxation delay of 5 s and a mixing time of 100 ms (Vázquez-Fresno et al., 2012).

A CEB sample was acquired on a Bruker Avance III 400 MHz NMR spectrometer equipped with a cryoprobe with presaturation of the water resonance using a NOESYPRESAT pulse sequence. The internal temperature was kept constant at 298 K and each sample was processed with a line-broadening factor of 0.3 Hz, 64 scans with a spectral width of 15 ppm, an acquisition time of 3 s, a relaxation delay of 5 s and a mixing time of 10 ms.

NMR spectra were phased, baseline corrected and referenced (TSP, 0.0 ppm) using TopSpin software (version 3.0, Bruker, BioSpin, Germany). After baseline correction, original spectral data were bucketed in intelligent bucketing domains of 0.005 ppm with ACD/NMR Processor 12.0 software (Advanced Chemistry Development, Toronto, Canada). The water signal and the regions above 9.5 ppm and below 0.5 ppm were excluded from the analysis.

2.4. Data pre-processing and statistical analysis

Data from acute and sustained interventions were submitted individually to MetaboAnalyst 3.0 for filtering and normalization purposes (Xia, Sinelnikov, Han, & Wishart, 2015). To exclude data points showing little variance across experimental conditions, the matrix was interquartile range filtered. Data were row-wise normalized (rows were samples) by the sum of the intensities of the spectra and column-wise normalized (columns were metabolites) using Pareto scaling and cube root transformation. Both data sets were used for further statistical analyses.

All the statistical analyses were performed with R version 3.1.2. Principal component analysis (PCA) was performed to detect the presence of outliers and to evaluate a potential carryover effect. Multilevel partial least squares discriminant analysis (ML-PLS-DA) was used to paired comparisons of the effects of CEB versus control beverage exploiting the crossover design with the R package mixOmics (Lê Cao, González, & Déjean, 2009). ML-PLS-DA is an extension of ordinary PLS-DA described by Van Velzen et al. (2008) which allows separation of the within-subject variation from the inter-subject variation that could obscure nutrition-related metabolic effects. A "leave-one-subject-out" cross-validation was performed to assess the ML-PLS-DA model and the classification error rate was determined by comparing the predicted class with the original one. Discriminant variables were determined

Table 1
Compounds identified by NMR spectroscopy in the coffee beverage (CEB) used in the present study.

Compound	mg/100 ml
2-Furoic acid	0.56
3-*O*-Caffeoylquinic acid	48.02
4-*O*-Caffeoylquinic acid	47.03
5-*O*-Caffeoylquinic acid	44.23
Acetic acid	20.36
Alanine	0.61
Caffeic acid	0.43
Caffeine	22.65
Choline	2.08
Citrate	1.75
Formic acid	9.45
Fumaric acid	0.64
Furfuryl alcohol	2.63
Lactic acid	3.79
Malic acid	22.78
Methanol	0.20
N-Methylpyridinium	–[a]
Nicotinic acid	0.36
Myo-inositol	9.52
Propionic acid	1.03
Sucrose	10.22
Quinic acid	15.83
Trigonelline	18.83

[a] –: identified but not quantified.

based on the minimum value of the classification error rate and also on the absolute value of loading scores from the first latent variable. Loading scores were ranked and the top 5% of total variables were selected as discriminant ones as long as this value was within the range of the minimum classification error rate. These variables were used to build a sparse ML-PLS-DA, and the corresponding score plot was drawn. On the sparse ML-PLS-DA, the signs were ranked based on the absolute values of their loading scores (named loading ranks, LRs), giving the lowest LR-values to signs that had the highest loading scores.

The statistical univariate crossover model (Jones & Kenward, 2014; Senn, 2002) of differences between baseline and intervention data was performed between groups to assess the statistical significances. In addition to individual effect, period and sequence variables were included in the crossover model for further correction on potential carry-over effects. Statistical significance was considered at a p-value <0.05.

2.5. Metabolite identification

Metabolite identification was performed using Chenomx NMR Suite Professional Software package (version 8.1; Chenomx Inc., Edmonton, Canada) and by comparing NMR spectral data to those available in databases such as the Human Metabolome Database (http://www.hmdb.ca), the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu) and the Madison Metabolomics Consortium Database (www.mmcd.nmrfam.wisc.edu), along with the existing NMR-based metabolomics literature. In addition, quantification of the beverage metabolites was performed with Chenomx using the methyl resonance of TSP as an internal standard. Peak fitting with reference to the TSP signal allowed quantification of absolute concentrations for the identified compounds. Further, a Pearson's correlation test was performed to test the correlation between signals of the same metabolite.

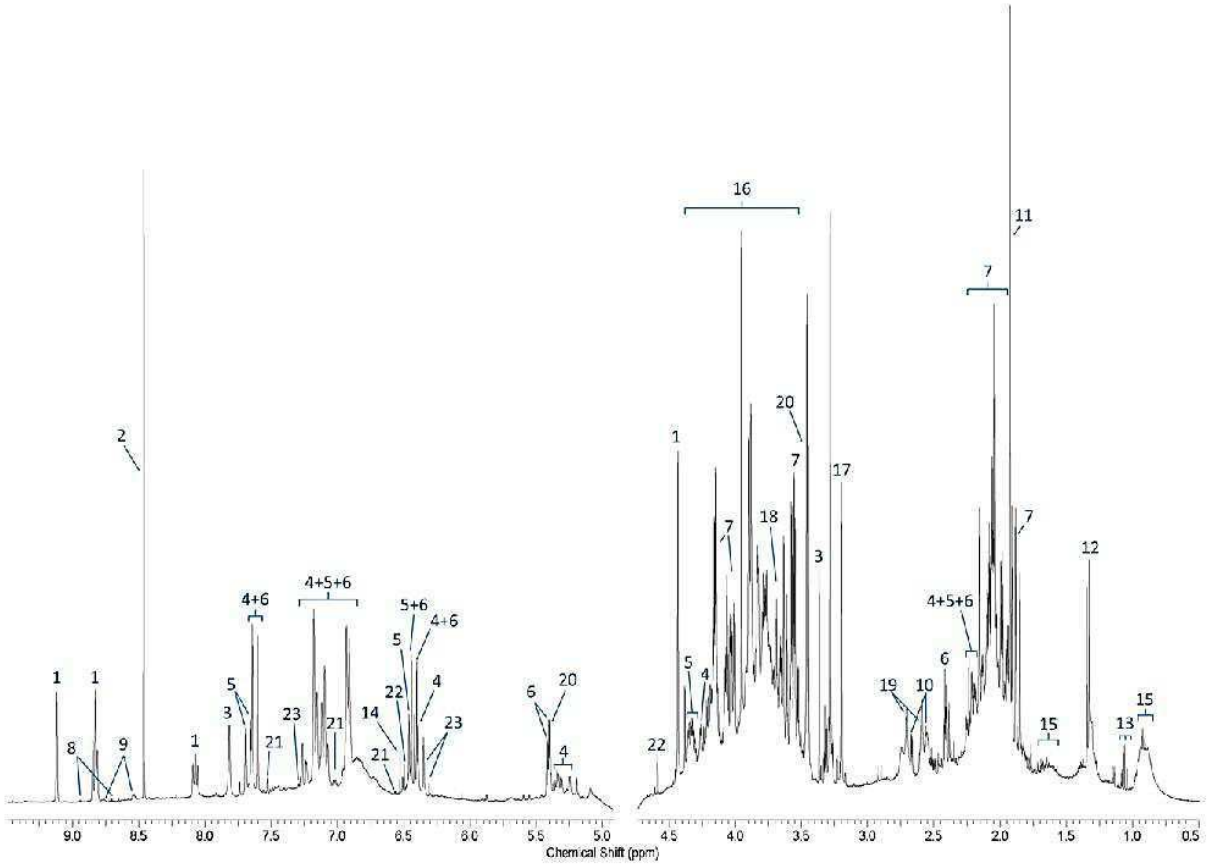


Fig. 1. 400 MHz ^{1}H-NMR spectra of coffee beverage; 1, trigonelline; 2, formic acid; 3, caffeine; 4, 3-*O*-caffeoylquinic acid; 5, 4-*O*-caffeoylquinic acid; 6, 5-*O*-caffeoylquinic acid; 7, quinic acid; 8, nicotinic acid; 9, *n*-methylpyridinium; 10, citric acid; 11, acetic acid; 12, lactic acid; 13, propionic acid; 14, fumaric acid; 15, lipids; 16, polysaccharides; 17, choline; 18, myo-inositol; 19, malic acid; 20, sucrose; 21, 2-furoic acid; 22, furfuryl alcohol; 23, caffeic acid.

3. Results and discussion

3.1. ^{1}H-NMR profiling of the coffee extract beverage

In order to analyze the urine metabolites coming from the coffee beverage, metabolic profiling of the CEB was carried out. The identified and quantified metabolites are summarized in Table 1 and are shown on spectra in Fig. 1.

Two furan compounds naturally present in coffee (2-furoic acid and furfuryl alcohol) were found in the CEB as precursors of 2-furoylglycine (Heinzmann et al., 2015), a compound found in urine after the acute intervention (see Section 3.4).

N-methylpyridinium (NMP) and nicotinic acid, which were previously described in coffee by NMR (Wei et al., 2012), were observed in the CEB but not detected in urine in this study. NMP and nicotinic acid are the main products of trigonelline degradation during pyrolysis and roasting (Stadler, Varga, Hau, Arce Vera, & Welti, 2002). We assumed that the low concentration of NMP seen in the CEB (Fig. 1) would not be enough to detect this metabolite in urine samples in the present study, but it was observed in urine after coffee intake by other authors (Lang et al., 2011).

Caffeine, 3-*O*-caffeoylquinic, 4-*O*-caffeoylquinic and 5-*O*-caffeoylquinic, quinic and caffeic acids, which are common compounds in coffee (Wei et al., 2012), were also found in the CEB.

3.2. Evaluation of carryover effects

PCA was performed to explore the distribution of baseline samples according to the period (I and II) and the sequence (control ➔ CEB; or CEB ➔ control; Fig. S1) of the crossover study. A point cloud would indicate homogeneity at baseline. Fig. S1 shows that the sample of the period I from subject #5 was outside of the point cloud and this may be due to a carryover effect. Therefore, all the statistical analyses were performed including and excluding subject #5. This subject was included after ruling out biological differences in all analyses without any output variation. Accordingly, they did not indicate any carryover effect. However, additional correction for period and sequence was included in the univariate statistical crossover models.

Table 2
Metabolites detected in urine after acute and sustained coffee extract beverage (CEB) in comparison with control beverage.

Intervention type	Metabolite	δ (multiplicity)	CBE vs. control	*P* value[a]	LR[b]
Acute	Trigonelline	9.13 (s)	↑	7.88×10^{-8}	4, 5
		8.84 (d)		8.44×10^{-7}	10, 11, 14
		8.08 (t)		1.26×10^{-2}	–
		4.44 (s)		2.51×10^{-6}	15, 25
	2-Furoylglycine	7.70 (s)	↑	5.90×10^{-4}	39
		7.18 (d)		9.68×10^{-2}	–
		6.64 (m)		4.14×10^{-2}	–
		3.93 (d)		3.00×10^{-4}	31, 47
	Citric acid	2.68 (d)	↑	8.72×10^{-3}	–
		2.54 (d)		1.31×10^{-3}	55
	Succinic acid	2.41 (s)	↑	7.78×10^{-4}	43
	3-Methyl-2-oxovaleric acid	1.10 (d)	↑	1.35×10^{-3}	50
		0.90 (t)		1.30×10^{-2}	–
	Isobutyric acid	1.07 (d)	↑	8.91×10^{-3}	–
		2.39 (m)		2.68×10^{-2}	–
Sustained	Trigonelline	9.13 (s)	↑	1.64×10^{-3}	8
		8.84 (d)		6.65×10^{-4}	7
		8.08 (t)		9.54×10^{-3}	–
		4.44 (s)		9.98×10^{-3}	18
	Hippuric acid	8.53 (s)	↑	2.74×10^{-3}	11, 17, 24
		7.84 (d)		7.30×10^{-3}	–
		7.64 (t)		6.66×10^{-5}	2, 5, 10
		7.56 (t)		8.38×10^{-3}	28, 34, 56
		3.97 (d)		1.17×10^{-2}	22, 35
	HPHPA	7.31 (t)	↑	2.29×10^{-4}	1, 3, 27
		6.98 (d)		1.69×10^{-2}	44, 47
		6.91 (s)		7.40×10^{-3}	16, 21, 49
		6.84 (d)		2.89×10^{-2}	–
		5.03 (m)[c]		–	–
	3-Hydroxyhippuric acid	7.40 (t)	↑	1.77×10^{-2}	55
		7.36 (s)		9.32×10^{-4}	6
		7.28 (s)		3.18×10^{-1}	–
		7.13 (dd)		2.42×10^{-2}	42
		4.15 (d)		3.37×10^{-2}	60
		2.22 (s)		1.94×10^{-2}	–

HPHPA: 3-(3-Hydroxyphenyl)-3-hydroxypropionic acid, s: singlet, d: doublet, t: triplet, dd: double doublet, m: multiplet.

[a] *P*-value of univariate statistical crossover model of differences.

[b] Loading rank of ML-PLS-DA.

[c] Signal not considered because of proximity to water region.

3.3. Uni- and multi-variate analyses

Discriminant metabolites and chemical shifts identified with *p*-value and LR corresponding to uni- and multi-variate, respectively, are presented in Table 2. Fig. 3 presents box plots comparing the intensities of significant metabolites between CEB and control beverage.

The comparison of the differences between interventions resulted in ML-PLS-DA models with a minimum classification error rate of 0.1 and 0.0 for sustained and acute interventions, respectively (Fig. S2). The lowest rate was 0.1 for a range between the 6th and the 94th variables after the sustained intervention, whereas the first discriminant variables after the acute intervention were strong enough for keeping the rate to 0 for all variables. The top 5% of loading scores were within the range of variables for these classification error rates. Therefore, two

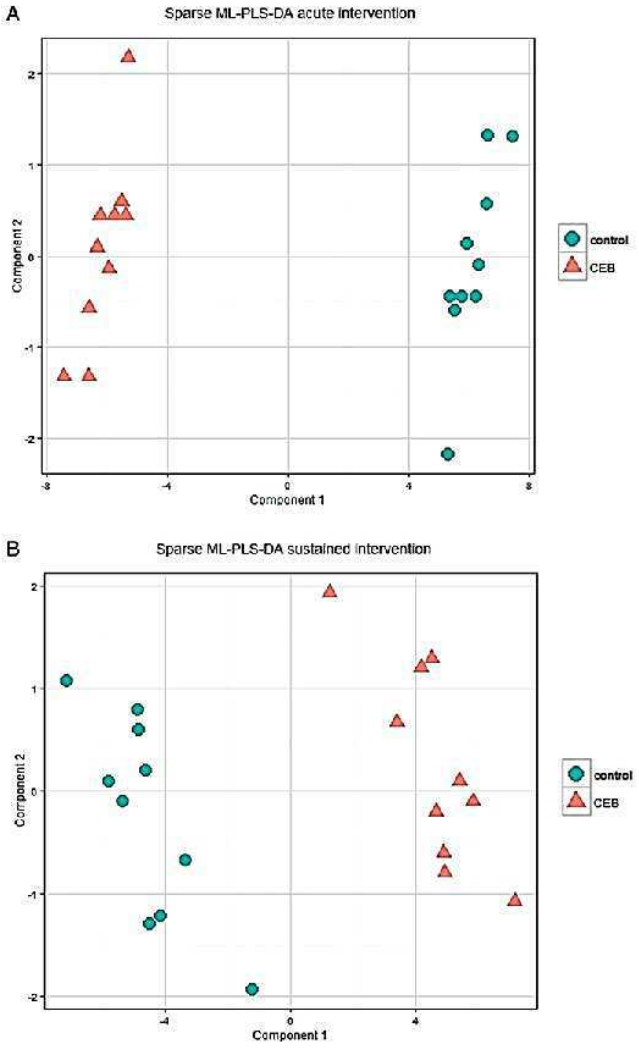


Fig. 2. Score plots of the subjects projected onto the first principal components, PC1 and PC2, after coffee extract beverage (CEB) in acute (A) and sustained (B) interventions.

sparse ML-PLS-DA models were built with them. Fig. 2 presents the sparse ML-PLS-DAs which show a clear separation of treatments. The main discriminatory metabolites were projected in the first component and the contribution to treatment differences of each sign was set in LRs presented in Table 2.

3.4. Acute intervention

2-Furoylglycine observed after CBE intake is a product of furan derivatives contained in roasting products like coffee (Moon & Shibamoto, 2009) or cocoa (Bonvehí, 2005) conjugated with glycine (Ginz & Engelhardt, 2000; Pettersen & Jellum, 1972). The furan product 2-furoylglycine has been identified after cocoa intake (Garcia-Aloy et al., 2015) and also associated with coffee intake in epidemiological studies (Guertin et al., 2014). Heinzmann and co-workers investigated its excretion kinetics, observing that maximal urinary excretion occurred 2 h after the coffee intake, returning to baseline levels after 24 h (Heinzmann et al., 2015). It explains its absence in the sustained intervention in the present study.

The increase of tricarboxylic acid (TCA) cycle intermediates suggests that CGAs coming from coffee may affect the central energy metabolism in humans. High levels of isobutyric and 3-methyl-2-oxovaleric acids were observed. They are products of branched-chain amino acid (BCAA) catabolism: valine and isoleucine, respectively. Both valine (glucogenic BCAA) and isoleucine (ketogenic BCAA) are metabolized to succinic acid via the methylmalonyl-CoA (Hutson, Sweatt, & LaNoue, 2005) TCA cycle, supported also by a significant higher excretion of succinic acid after CEB acute intake. In addition, an increase in the urinary level of citric acid was detected, which is also a TCA cycle intermediate, leading to the conclusion that CGAs coming from coffee may affect the central energy metabolism in humans, as was seen in a study with dietary supplementation with CGAs, which affected the circulating levels of intermediates of the TCA cycle in rats (Ruan, Yang, Zhou, Wen, et al., 2014). In accordance with energy metabolism, habitual coffee consumption was previously associated with weight control

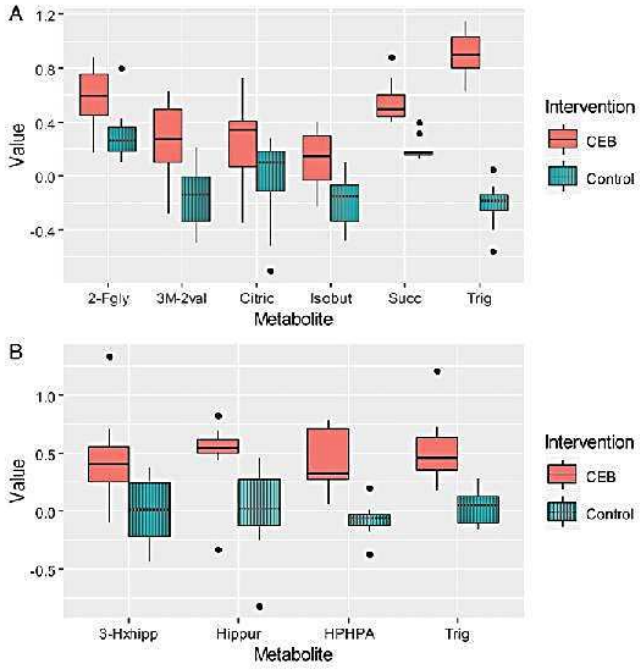


Fig. 3. Box plots of significant metabolites in urine after coffee extract beverage (CEB) in acute (A) and sustained (B) interventions. 2-Fgly: 2-furoylglycine, 3M-2val: 3-methyl-2-oxovaleric acid, Citric: citric acid, Isobut: isobutyric acid, Succ: succinic acid, Trig: trigonelline, 3-Hxhipp: 3-hydroxyhippuric acid, Hippur: hippuric acid, HPHPA: 3-(3-hydroxyphenyl)-3-hydroxypropionic acid.

(Bakuradze et al., 2011) and a lower risk of type 2 diabetes (R M van Dam & Hu, 2005; Wedick et al., 2011).

3.5. Sustained intervention

Hippuric, 3-hydroxyhippuric and 3-(3-Hydroxyphenyl)-3-hydroxypropionic (HPHPA) acids were detected after CEB in comparison with control beverage. These three compounds have been observed previously in epidemiological studies after coffee intake (Guertin et al., 2015). It has been suggested that CGAs are metabolized by microbiota in the lower gastrointestinal tract (Gonthier et al., 2006; M.-P. Gonthier, Verny, Besson, Rémésy, & Scalbert, 2003; Ludwig, Paz de Peña, Concepción, & Alan, 2013), obtaining as a product caffeic acid due to cleavage of the ester bond between caffeic and quinic acids by an esterase enzyme provided by the colonic microbiota (Plumb et al., 1999). The double bond of caffeic acid is reduced to form HPHPA prior to dihydroxylation of 3-(3,4-dihydroxyphenyl)-propionic acid (Rechner et al., 2004). HPHPA may be absorbed from the colon to form hippuric and 3-hydroxyhippuric acids through β-oxidation and glycination in the liver (Rechner et al., 2002). Therefore, microbiota metabolites were downstream metabolites of 3-*O*-caffeoylquinic, 4-*O*-caffeoylquinic and 5-*O*-caffeoylquinic acids which were detected and identified in the CEB (see Section 3.1). Hippuric, 3-hydroxyhippuric and HPHPA acids are metabolites of microbiota fermentation of CGAs which are subsequently absorbed in the colon (Higdon & Frei, 2006), playing a role in health effects (Williamson & Clifford, 2010) and leading to the conclusion that CGAs may have an important function as precursors of microbiota-derived compounds observed after sustained consumption.

3.6. Common metabolites in urine and the coffee extract beverage

Trigonelline was identified in urine after intake of the CEB in both acute and sustained interventions. According to the results, this metabolite comes from trigonelline naturally present in coffee beans (Beans, Allred, Yackley, Vanamala, & Allred, 2009; Guertin et al., 2015), as was observed in ^{1}H-NMR analysis of the CEB (Fig. 1). Lang and co-workers observed this candidate biomarker up to three days after coffee consumption (Lang et al., 2011); therefore, trigonelline in the sustained intervention could correspond to trigonelline from the previous punctual intake of the day before. Thus, according to the results, trigonelline reflects coffee intake of both a punctual and a sustained consumption. This metabolite has been proposed as a candidate for dietary biomarker coffee intake (Lang et al., 2011; Rothwell, Fillâtre, et al., 2014). Trigonelline was also a common metabolite present in urine after both acute and sustained intakes as well as in the CEB, indicating a possible direct excretion (without any metabolism occurring) from coffee intake, as suggested by K. A. Guertin et al. (2015). A direct excretion of this metabolite suggests a non-interindividual variation in the metabolism of this compound, in contrast to microbiota-derived compounds of polyphenol metabolism, which are dependent on a high interindividual variation in humans (Hervert-Hernández & Goñi, 2011). Hence, trigonelline seems to be a suitable biomarker of coffee intake after both acute and sustained interventions.

4. Conclusions

The results of the current study show the capability of an NMR-based metabolomic approach to detect significant changes in urinary metabolite profiling and help to clarify the role of specific coffee compounds other than caffeine in human metabolism after both acute and sustained interventions. Firstly, the increase of endogenous metabolites related to the TCA cycle such as succinic, citric, isobutyric and 3-methyl-2-oxovaleric acids after the acute study indicates that coffee CGAs impact on the central energy metabolism. Secondly, the present results exhibit two potential CEB biomarkers: 2-furoylglycine, after acute

intake, and trigonelline, after both acute and sustained CEB intakes. Finally, there are other long-term metabolites after sustained intervention coming from the degradation of CGAs by gut microbiota observed through the identification of hippuric acid, 3-hydroxyhippuric acid and HPHPA. In conclusion, this study shows biomarkers of both short- and long-term intakes of coffee products along with other metabolites involved in energy metabolism (after acute intake) and microbiota metabolism (after sustained intake).

5. Acknowledgments

This research was funded by: the Fondo Europeo de Desarrollo Regional (FEDER) (AGL2009-13906-C02-01, AGL2010-10084-E) and the JPI HDHL FOODBALL (PCIN-2014-133-MINECO-Spain). The clinical study was supported in part by the KAO Corporation and Fundació Bosch i Gimpera (FBG 307683). We also thank the award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR. FM acknowledges the APIF fellowship (University of Barcelona).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.foodres.2016.03.038.

References

Bakuradze, T., Boehm, N., Janzowski, C., Lang, R., Hofmann, T., Stockis, J. -P., ... Eisenbrand, G. (2011). Antioxidant-rich coffee reduces DNA damage, elevates glutathione status and contributes to weight control: Results from an intervention study. *Molecular Nutrition & Food Research*, *55*(5), 793–797. http://dx.doi.org/10.1002/mnfr.201100093.

Beans, C., Allred, K. F., Yackley, K. M., Vanamala, J., & Allred, C. D. (2009). *Trigonelline is a novel phytoestrogen in, 1833–1838. http://dx.doi.org/10.3945/jn.109.108001.1833.*

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder. *European Food Research and Technology*, *221*(1–2), 19–29. http://dx.doi.org/10.1007/s00217-005-1147-y.

Brennan, L. (2014). NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *83*, 42–49. http://dx.doi.org/10.1016/j.pnmrs.2014.09.001.

Clifford, M. N. (2000). Chlorogenic acids and other cinnamates — Nature, occurrence, dietary burden, absorption and metabolism. *Journal of the Science of Food and Agriculture*, *1043*(October 1999), 1033–1043 (http://doi.org/http://dx.doi.org/10.1002/(SICI)1097-0010(20000515)80:7<1033::AID-JSFA595>3.0.CO;2-T).

Cowan, T. E., Palmnäs, M. S. a, Yang, J., Bomhof, M. R., Ardell, K. L., Reimer, R. a., ... Shearer, J. (2014). Chronic coffee consumption in the diet-induced obese rat: Impact on gut microbiota and serum metabolomics. *Journal of Nutritional Biochemistry*, *25*(4), 489–495. http://dx.doi.org/10.1016/j.jnutbio.2013.12.009.

Ding, M., Bhupathiraju, S. N., Chen, M., van Dam, R. M., & Hu, F. B. (2014). Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: A systematic review and a dose–response meta-analysis. *Diabetes Care*, *37*(2), 569–586. http://dx.doi.org/10.2337/dc13-1203.

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Auregui, O., Corella, D., Ruiz-Canela, M., ... Andres-Lacueva, C. (2015). A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED study. *Mol. Nutr. Food Research*, *59*, 212–220. http://dx.doi.org/10.1002/mnfr.201400434.

Ginz, M., & Engelhardt, U. H. (2000). Identification of proline-based diketopiperazines in roasted coffee. *Journal of Agricultural and Food Chemistry*, *48*(8), 3528–3532. http://dx.doi.org/10.1021/jf991256v.

Gonthier, M. P., Remesy, C., Scalbert, a, Cheynier, V., Souquet, J. M., Poutanen, K., & Aura, a M. (2006). Microbial metabolism of caffeic acid and its esters chlorogenic and caftaric acids by human faecal microbiota in vitro. *Biomedicine and Pharmacotherapy*, *60*(9), 536–540. http://dx.doi.org/10.1016/j.biopha.2006.07.084.

Gonthier, M. -P., Verny, M. -A., Besson, C., Rémésy, C., & Scalbert, A. (2003). Chlorogenic acid bioavailability largely depends on its metabolism by the gut microflora in rats. *The Journal of Nutrition*, *133*(6), 1853–1859.

Guertin, K. a, Loftfield, E., Boca, S. M., Sampson, J. N., Moore, S. C., Xiao, Q., ... Sinha, R. (2015). *Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancerVols. 1 – 5, (C). (pp. 1–12), 1–12. http://dx.doi.org/10.3945/ajcn.114.096099.*

Guertin, K. A., Moore, S. C., Sampson, J. N., Huang, W. -Y., Xiao, Q., Stolzenberg-Solomon, R. Z., ... Cross, A. J. (2014). Metabolomics in nutritional epidemiology: Identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *The American Journal of Clinical Nutrition*, *100*, 208–217. http://dx.doi.org/10.3945/ajcn.113.078758.

Heinzmann, S. S., Holmes, E., Kochhar, S., Nicholson, J. K., & Schmitt-Kopplin, P. (2015). 2-Furoylglycine as a candidate biomarker of coffee consumption. *Journal of Agricultural and Food Chemistry*, *63*(38), 8615–8621. http://dx.doi.org/10.1021/acs.jafc.5b03040.

Hervert-Hernández, D., & Goñi, I. (2011). Dietary polyphenols and human gut microbiota: A review. *Food Reviews International*, *27*(2), 154–169. http://dx.doi.org/10.1080/87559129.2010.535233.

Higdon, J., & Frei, B. (2006). Coffee and health: A review of recent human research. *Critical Reviews in Food Science and Nutrition*, *46*(2), 101–123. http://dx.doi.org/10.1080/10408390500400009.

Hutson, S. M., Sweatt, A. J., & LaNoue, K. F. (2005). Branched-chain amino acid metabolism: Implications for establishing safe intakes. *The Journal of Nutrition*, *135*(6), 1557S–11564 Retrieved from http://jn.nutrition.org/cgi/content/long/135/6/1557S

Huxley, R., Lee, C. M. Y., Barzi, F., Timmermeister, L., Czernichow, S., Perkovic, V., ... Woodward, M. (2009). Coffee, decaffeinated coffee, and tea consumption in relation to incident type 2 diabetes mellitus: A systematic review with meta-analysis. *Archives of Internal Medicine*, *169*(22), 2053–2063. http://dx.doi.org/10.1001/archinternmed.2009.439.

Ito, H., Gonthier, M. -P., Manach, C., Morand, C., Mennen, L., Rémésy, C., & Scalbert, A. (2005). Polyphenol levels in human urine after intake of six different polyphenol-rich beverages. *The British Journal of Nutrition*, *94*, 500–509. http://dx.doi.org/10.1079/BJN20051522.

Johnston, K. L., Clifford, M. N., & Morgan, L. M. (2003). Coffee acutely modifies gastrointestinal hormone secretion and glucose tolerance in humans: Glycemic effects of chlorogenic acid and caffeine. *American Journal of Clinical Nutrition*, *78*(4), 728–733. http://dx.doi.org/10.1021/jf980737w.

Jones, B., & Kenward, M. G. (2014). *Design and analysis of cross-over trials* (3rd ed.). CRC Press Retrieved from https://books.google.com/books?id=aErNBQAAQBAJ&pgis=1

La Vecchia, C. (2005). Coffee, liver enzymes, cirrhosis and liver cancer. *Journal of Hepatology*, *42*(4), 444–446. http://dx.doi.org/10.1016/j.jhep.2005.01.004.

Lang, R., Wahl, A., Stark, T., & Hofmann, T. (2011). Urinary *N*-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Molecular Nutrition and Food Research*, *55*(11), 1613–1623. http://dx.doi.org/10.1002/mnfr.201000656.

Lê Cao, K. -A., González, I., & Déjean, S. (2009). integrOmics: An R package to unravel relationships between two omics datasets. *Bioinformatics (Oxford, England)*, *25*(21), 2855–2856. http://dx.doi.org/10.1093/bioinformatics/btp515.

Lloyd, A. J., Beckmann, M., Haldar, S., Seal, C., Brandt, K., & Draper, J. (2013). Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *American Journal of Clinical Nutrition*, *97*(C), 377–389. http://dx.doi.org/10.3945/ajcn.112.048033.

Ludwig, I. a., Paz de Peña, M., Concepción, C., & Alan, C. (2013). Catabolism of coffee chlorogenic acids by human colonic microbiota. *BioFactors*, *39*(6), 623–632. http://dx.doi.org/10.1002/biof.1124.

Moon, J. -K., & Shibamoto, T. (2009). Role of roasting conditions in the profile of volatile flavor chemicals formed from coffee beans. *Journal of Agricultural and Food Chemistry*, *57*(13), 5823–5831 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19579294

Nordestgaard, a. T., Thomsen, M., & Nordestgaard, B. G. (2015). Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: A Mendelian randomization study. *International Journal of Epidemiology*, *01*, 1–15. http://dx.doi.org/10.1093/ije/dyv083.

Pettersen, J. E., & Jellum, E. (1972). The identification and metabolic origin of 2-furoylglycine and 2.5-furandicarboxylic acid in human urine. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, *41*, 199–207 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4630229

Plumb, G. W., Garcia-Cones, M. T., Kroon, P. A., Rhodes, M., Ridley, S., & Williams, G. (1999). Metabolism of chlorogenic acid by human plasma, liver, intestine and gut microflora [1]. *Journal of the Science of Food and Agriculture*, *79*, 390–392.

Puiggròs, F., Canela, N., & Arola, L. (2015). Metabolome responses to physiological and nutritional challenges. *Current opinion in food science*. http://dx.doi.org/10.1016/j.cofs.2015.06.001.

Rechner, A. R., Kuhnle, G., Bremner, P., Hubbard, G. P., Moore, K. P., & Rice-Evans, C. A. (2002). The metabolic fate of dietary polyphenols in humans. *Free Radical Biology and Medicine*, *33*(2), 220–235. http://dx.doi.org/10.1016/S0891-5849(02)00877-8.

Rechner, A. R., Smith, M. A., Kuhnle, G., Gibson, G. R., Debnam, E. S., Srai, S. K. S., ... Rice-Evans, C. A. (2004). Colonic metabolism of dietary polyphenols: Influence of structure on microbial fermentation products. *Free Radical Biology and Medicine*, *36*(2), 212–225. http://dx.doi.org/10.1016/j.freeradbiomed.2003.09.022.

Rothwell, J. A., Fillâtre, Y., Martin, J. -F., Lyan, B., Pujos-Guillot, E., Fezeu, L., ... Manach, C. (2014). New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PloS One*, *9*(4), e93474. http://dx.doi.org/10.1371/journal.pone.0093474.

Ruan, Z., Yang, Y., Zhou, Y., Wen, Y., Sheng, D., ... Yin, Y. (2014). *Metabolomic analysis of amino acid and energy metabolism in rats supplemented with chlorogenic acid. http://dx.doi.org/10.1007/s00726-014-1762-7.*

Sääksjärvi, K., Knekt, P., Rissanen, H., Laaksonen, M. a, Reunanen, a, & Männistö, S. (2008). Prospective study of coffee consumption and risk of Parkinson's disease. *European Journal of Clinical Nutrition*, *62*(7), 908–915. http://dx.doi.org/10.1038/sj.ejcn.1602788.

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., ... Wishart, D. S. (2014). The food metabolome: A window over dietary exposure. *The American Journal of Clinical Nutrition*, *99*(6), 1286–1308. http://dx.doi.org/10.3945/ajcn.113.076133.

Senn, S. (2002). Cross-over trials in clinical research. Retrieved from https://books.google.com/books?hl=es&lr=&id=MWbHzwpPTgEC&pgis=1

Shukitt-Hale, B., Miller, M. G., Chu, Y. F., Lyle, B. J., & Joseph, J. a. (2013). Coffee, but not caffeine, has positive effects on cognition and psychomotor behavior in aging. *Age*, *35*(6), 2183–2192. http://dx.doi.org/10.1007/s11357-012-9509-4.

Stadler, R. H., Varga, N., Hau, J., Arce Vera, F., & Welti, D. H. (2002). Alkylpyridiniums. 1. Formation in model systems via thermal degradation of trigonelline. *Journal of Agricultural and Food Chemistry*, *50*, 1192–1199. http://dx.doi.org/10.1021/jf011234k.

Stalmach, A., Mullen, W., Barron, D., Uchida, K., Yokota, T., Cavin, C., ... Crozier, A. (2009). Metabolite profiling of hydroxycinnamate derivatives in plasma and urine after the ingestion of coffee by humans: Identification of biomarkers of coffee consumption. *Drug Metabolism and Disposition*, *37*(8), 1749–1758. http://dx.doi.org/10.1124/dmd.109.028019.

Urgert, R., & Katan, M. B. (1997). The cholesterol-raising factor from coffee beans. *Annual Review of Nutrition*, *17*, 305–324. http://dx.doi.org/10.1146/annurev.nutr.17.1.305.

van Dam, R. M., & Feskens, E. J. M. (2002). *Coffee consumption and risk of type 2 diabetes mellitus treatment of active Crohn's disease with recombinant humanVol. 360. (pp. 1477–1478), 1477–1478.*

Van Dam, R. M., & Hu, F. B. (2005). Coffee consumption and risk of type 2 diabetes: A systematic review. *JAMA*, *294*(1), 97–104. http://dx.doi.org/10.1001/jama.294.1.97.

Van Velzen, E. J. J., Westerhuis, J. a., Van Duynhoven, J. P. M., Van Dorsten, F. a., Hoefsloot, H. C. J., Jacobs, D. M., ... Smilde, A. K. (2008). Multilevel data analysis of a crossover designed human nutritional intervention study. *Journal of Proteome Research*, *7*(10), 4483–4491. http://dx.doi.org/10.1021/pr800145j.

Vázquez-Fresno, R., Llorach, R., Alcaro, F., Rodríguez, M.Á, Vinaixa, M., Chiva-Blanch, G., ... Andrés-Lacueva, C. (2012). (1)H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors. *Electrophoresis*, *33*(15), 2345–2354. http://dx.doi.org/10.1002/elps.201100646.

Wedick, N. M., Brennan, A. M., Sun, Q., Hu, F. B., Mantzoros, C. S., & Van Dam, R. M. (2011). Effects of caffeinated and decaffeinated coffee on biological risk factors for type 2 diabetes: A randomized controlled trial. *Nutrition Journal*, *10*, 93. http://dx.doi.org/10.1186/1475-2891-10-93.

Wei, F., Furihata, K., Koda, M., Hu, F., Miyakawa, T., & Tanokura, M. (2012). Roasting process of coffee beans as studied by nuclear magnetic resonance: Time course of changes in composition. *Journal of Agricultural and Food Chemistry*, *60*(4), 1005–1012. http://dx.doi.org/10.1021/jf205315r.

Williamson, G., & Clifford, M. N. (2010). Colonic metabolites of berry polyphenols: The missing link to biological activity? *The British Journal of Nutrition*, *104 Suppl*, S48–S66. http://dx.doi.org/10.1017/S0007114510003946.

Wishart, D. S. (2008). Metabolomics: Applications to food science and nutrition research. *Trends in Food Science & Technology*, *19*(9), 482–493. http://dx.doi.org/10.1016/j.tifs.2008.03.003.

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). Metaboanalyst 3.0—Making metabolomics more meaningful. *Nucleic Acids Research.*. http://dx.doi.org/10.1093/nar/gkv380.

Xu, Y. -J., Wang, C., Ho, W. E., & Ong, C. N. (2014). Recent developments and applications of metabolomics in microbiological investigations. *TrAC Trends in Analytical Chemistry*, *56*, 37–48. http://dx.doi.org/10.1016/j.trac.2013.12.009.

Yanagimoto, K., Ochi, H., Lee, K. -G., & Shibamoto, T. (2004). Antioxidative activities of fractions obtained from brewed coffee. *Journal of Agricultural and Food Chemistry*, *52*(3), 592–596. http://dx.doi.org/10.1021/jf030317t.

Supplementary material

Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects

Table S1. Composition of beverages used in the study (determined by the manufacturer)

Composition	CEB	Control
CGA, mg/100 ml	223	0
Caffeine, mg/100 ml	37.3	35.1
Carbohydrates, g/100 ml	1.0	0.3
Proteins, g/100 ml	0.3	<0.1
Fats, g/100 ml	<0.1	<0.1
Sodium, mg/100 ml	18.5	6.1

CEB: coffee extract beverage, CGA: chlorogenic acid

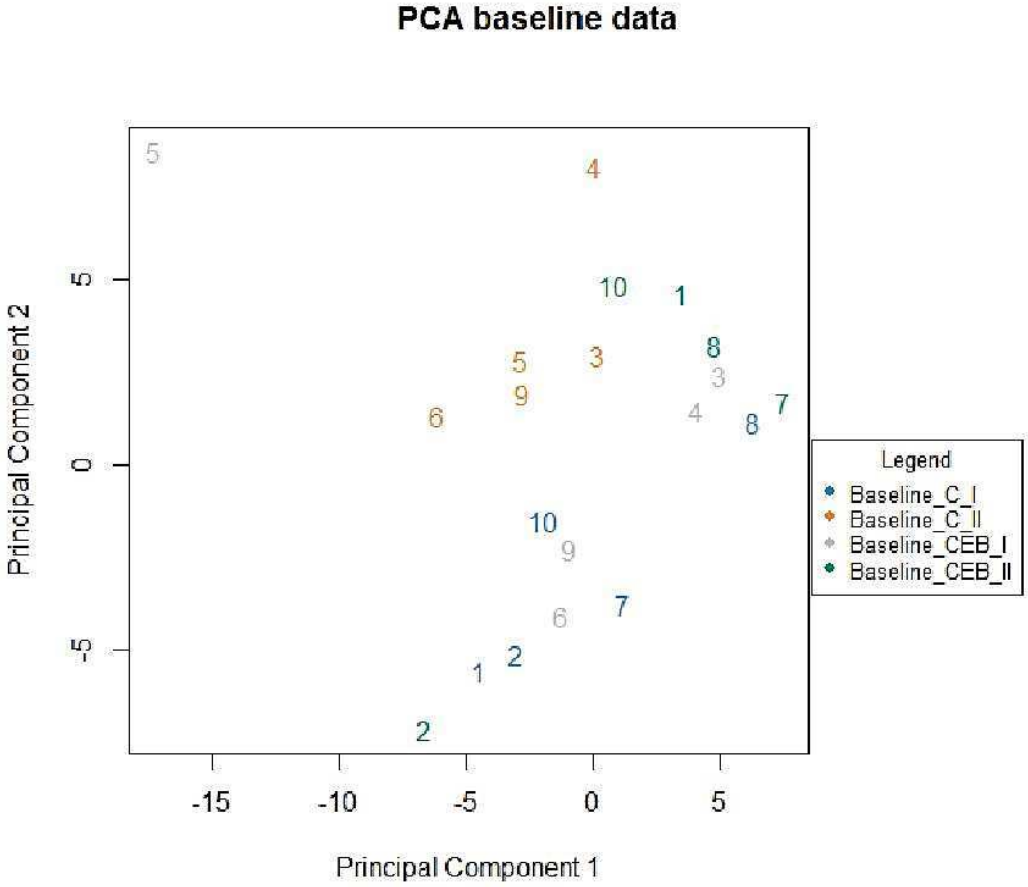
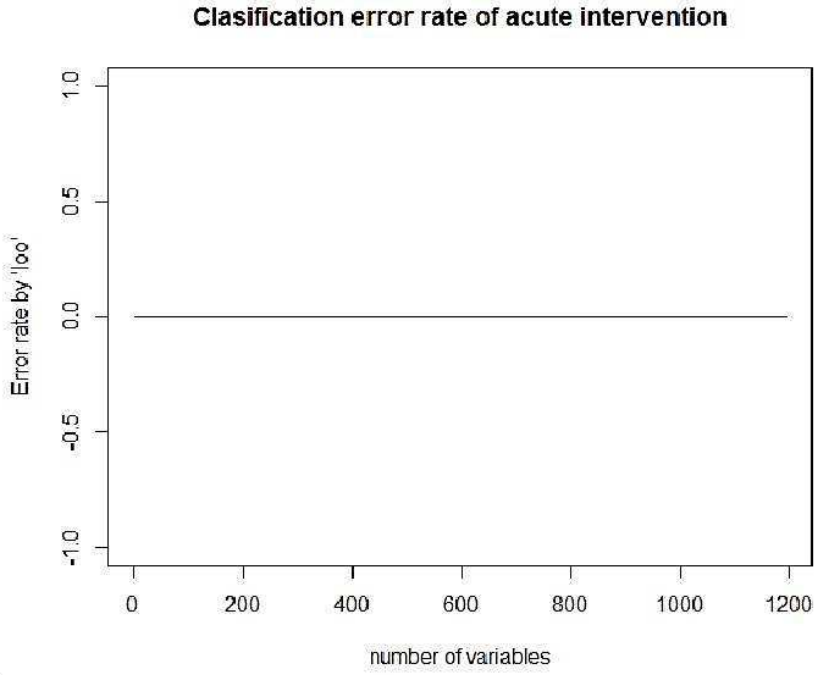
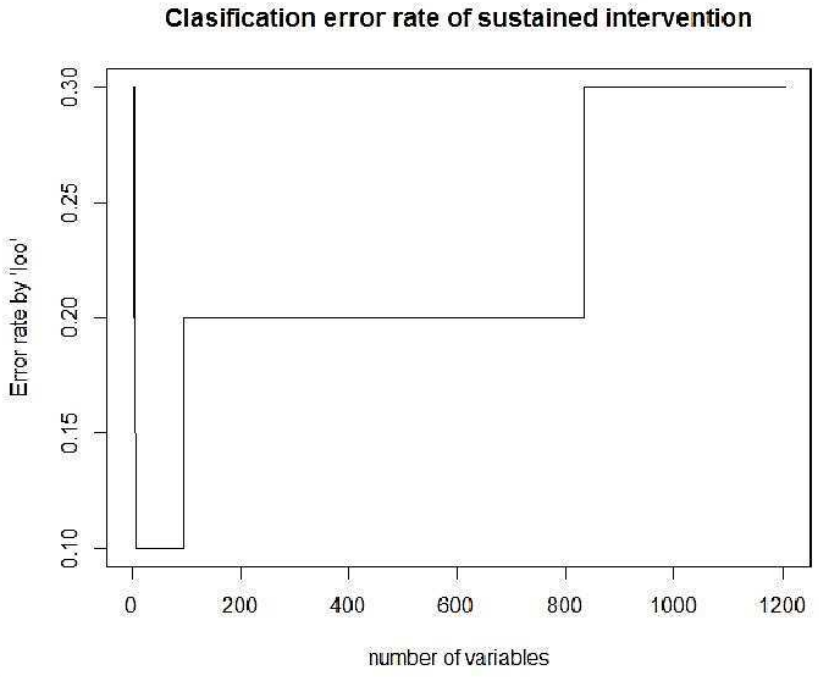


Figure S1. Principal component analysis (PCA) of baseline points in CEB (Baseline_CEB) and control (Baseline_C) groups in both I and II periods of the study.



A



B

Figure S2. Clasification error rate by "leave-one-subject-out" method through the number of variables in the statistic models after acute (A) and sustained (B) interventions.

Capítulo 08.

ESTUDIOS OBSERVACIONALES

8.1. Análisis de la huella metabolómica tras el consumo habitual de legumbres en individuos en condiciones de vida libre

En el marco del proyecto FOODBALL se ha realizado un estudio observacional para el análisis de la huella metabolómica asociada con el consumo declarado de legumbres en individuos en condiciones de vida libre. Los resultados de este estudio se publicaron en la revista *Journal of Proteome Research* indexada en el *Journal Citation Reports* con un factor de impacto de 4,172 (2015) y situada en el primer cuartil de la categoría *Biochemical Research Methods* (12/77) (2015).

Francisco Madrid-Gambin, Rafael Llorach, Rosa Vázquez-Fresno, Mireia Urpi-Sarda, Enrique Almanza-Aguilera, Mar Garcia-Aloy, Ramon Estruch, Dolores Corella, Cristina Andres-Lacueva. Urinary ^{1}H Nuclear Magnetic Resonance Metabolomic Fingerprinting Reveals Biomarkers of Pulse Consumption Related to Energy-Metabolism Modulation in a Subcohort from the PREDIMED study. *Journal of Proteome Research*. 2017; 16(4), 1483–1491. doi:10.1021/acs.jproteome.6b00860

Resumen:

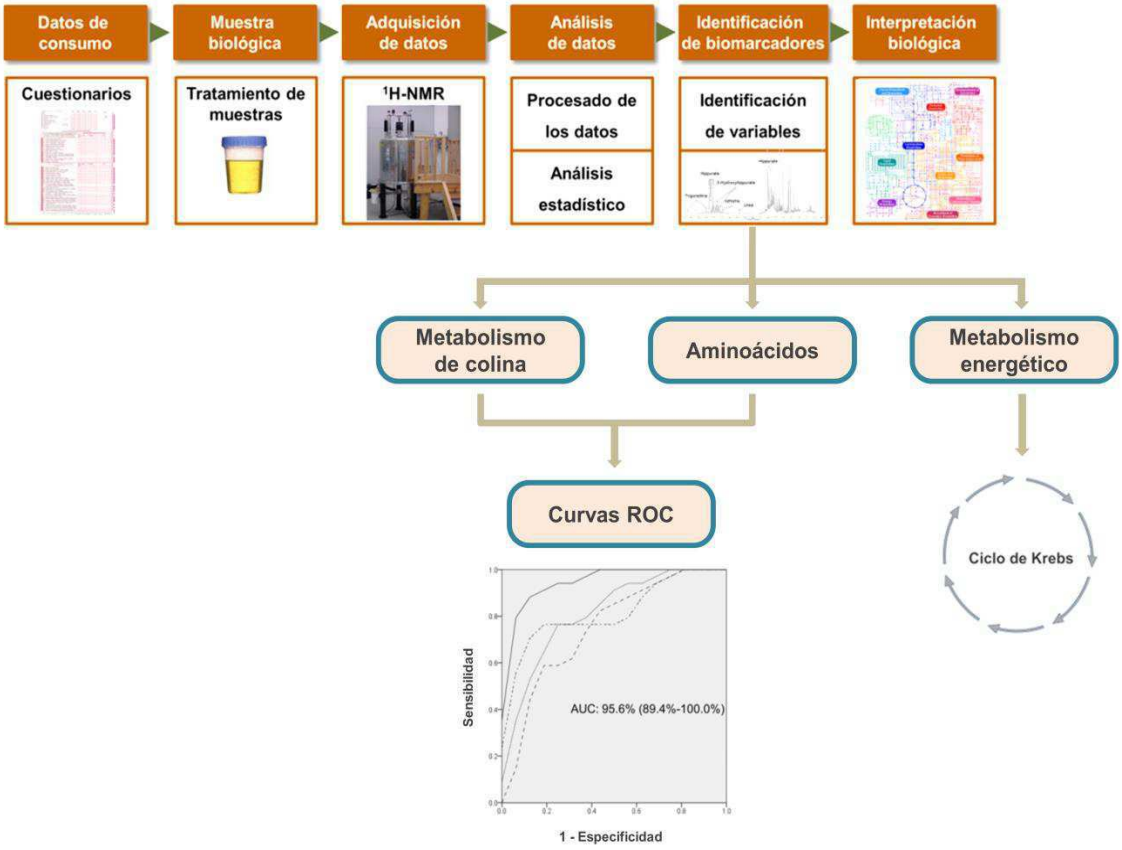
Las legumbres han sido reconocidas por ejercer efectos beneficiosos sobre la salud. Éstas podrían estar jugando un papel importante en la Dieta Mediterránea ya que constituyen un alimento de excelente calidad nutricional, proporcionando proteínas, fibra dietética, muchas vitaminas y minerales, así como una gran variedad de fitoquímicos. En este contexto, es de gran importancia disponer de medidas de ingesta objetivas, como son los biomarcadores nutricionales, para evaluar dichos efectos, así como sus mecanismos de acción. Sin embargo, el perfil metabolómico asociado al consumo de legumbres es aún poco conocido.

Para caracterizar la huella metabolómica del consumo de legumbres (concretamente garbanzos, lentejas y alubias), se estratificaron los sujetos según el consumo

reportado mediante un cuestionario de frecuencia de consumo de alimentos. Posteriormente se analizaron las muestras de orina de los consumidores habituales de legumbres (≥25 g/día de ingesta de legumbres) y los no consumidores de legumbres (≤ 4 g/día de consumo de legumbres) utilizando un enfoque metabolómico no dirigido mediante resonancia magnética nuclear (RMN), combinado con análisis estadísticos multi- y uni-variantes. El consumo de legumbres se asoció con diferencias estadísticamente significativas en los niveles urinarios de 16 metabolitos provenientes del metabolismo de la colina, de compuestos relacionados con proteínas, y con el metabolismo energético (incluyendo una menor concentración de glucosa excretada). Para diseñar un modelo combinado de exposición global a legumbres se aplicó un análisis de regresión logística por pasos, que dio como resultado un modelo que incluyó la glutamina, la dimetilamina y la 3-metilhistidina. Este modelo fue evaluado por una curva característica de funcionamiento del receptor (ROC) exhibiendo una mayor área bajo la curva para el modelo combinado (>90%) en comparación con los metabolitos individualmente (<90%, todos los casos).

El perfil metabolómico urinario, basado en la RMN, asociado el consumo de legumbres desveló nuevos candidatos a biomarcadores de consumo de legumbres y el impacto de estas sobre el metabolismo energético, generando nuevas hipótesis sobre los mecanismos moleculares por los cuales ejercen efectos en la salud. Otros estudios de intervención serían necesarios para confirmar estos hallazgos.

Figura 6. Gráfico-resumen del estudio observacional de biomarcadores de exposición dietética tras el consumo declarado de legumbres.



Journal of proteome research

Article

pubs.acs.org/jpr

Urinary ^{1}H Nuclear Magnetic Resonance Metabolomic Fingerprinting Reveals Biomarkers of Pulse Consumption Related to Energy-Metabolism Modulation in a Subcohort from the PREDIMED study

Francisco Madrid-Gambin,[†] Rafael Llorach,*[,†,‡] Rosa Vázquez-Fresno,[†,§] Mireia Urpi-Sarda,[†,‡] Enrique Almanza-Aguilera,[†] Mar Garcia-Aloy,[†,‡] Ramon Estruch,[∥,⊥] Dolores Corella,[⊥,#] and Cristina Andres-Lacueva*[,†,‡]

[†]Biomarkers & Nutrimetabolomics Laboratory, Nutrition, Food Science and Gastronomy Department, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona 08028, Spain

[‡]CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Barcelona 08028, Spain

[§]Department of Biological Sciences and Department of Computing Science, University of Alberta, Edmonton AB T6G 2E9, Canada

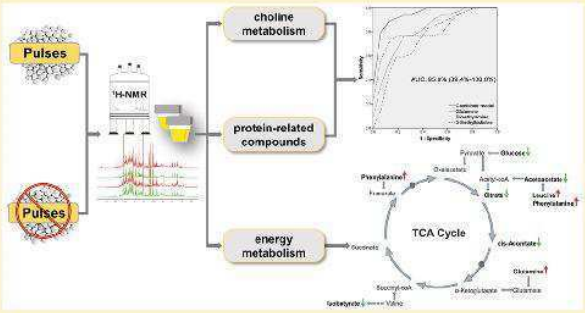
[∥]Department of Internal Medicine, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona 08036, Spain

[⊥]CIBER OBN, The Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid 28029, Spain

[#]Department of Preventive Medicine and Public Health, University of Valencia, Valencia 46010, Spain

Supporting Information

ABSTRACT: Little is known about the metabolome fingerprint of pulse consumption. The study of robust and accurate biomarkers for pulse dietary assessment has great value for nutritional epidemiology regarding health benefits and their mechanisms. To characterize the fingerprinting of dietary pulses (chickpeas, lentils, and beans), spot urine samples from a subcohort from the PREDIMED study were stratified using a validated food frequency questionnaire. Urine samples of nonpulse consumers (≤4 g/day of pulse intake) and habitual pulse consumers (≥25 g/day of pulse intake) were analyzed using a ^{1}H nuclear magnetic resonance (NMR) metabolomics approach combined with multi- and univariate data analysis. Pulse consumption showed differences through 16 metabolites coming from (i) choline metabolism, (ii) protein-related compounds, and (iii) energy metabolism (including lower urinary glucose). Stepwise logistic regression analysis was applied to design a combined model of pulse exposure, which resulted in glutamine, dimethylamine, and 3-methylhistidine. This model was evaluated by a receiver operating characteristic curve (AUC > 90% in both training and validation sets). The application of NMR-based metabolomics to reported pulse exposure highlighted new candidates for biomarkers of pulse consumption and the impact on energy metabolism, generating new hypotheses on energy modulation. Further intervention studies will confirm these findings.



KEYWORDS: *pulses, legumes, metabolomics, NMR, choline metabolism, energy, biomarkers, ROC curve*

1. INTRODUCTION

The Mediterranean diet (MD) is a dietary pattern characterized by a high intake of vegetables, cereals, pulses, nuts, fish and olive oil; a low intake of red meat and processed meat products; and low to moderate consumption of poultry, wine, and dairy products.[1] Moreover, the MD has been demonstrated to be useful in the prevention of type 2 diabetes, obesity, inflammatory diseases, cardiovascular diseases (CVD), and even cancer.[2−5]

One of the components of the MD is pulses, which constitute an excellent food, providing protein, dietary fiber, and many vitamins and minerals as well as a great variety of phytochemicals.[6−8] Thus, they could contribute to the beneficial effects reported for this dietary pattern.[9] In addition, pulses are increasingly being recognized for their role in promoting good health.[6,10−12] Indeed, habitual pulse consumption is included in the main dietary guidelines worldwide, including the MD,[13] the Dietary Guidelines for Americans[14,15] and the Nordic Diet,[16] among others, and they are also

Received: September 28, 2016
Published: January 9, 2017

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

advocated in view of their low environmental impact compared with other protein sources.[17]

Metabolomics is a powerful tool for identifying food-exposure biomarkers in humans[18] and provides new information on dietary components and dietary patterns.[19] In this regard, the evaluation of dietary exposure through a combination of biomarkers enables a better understanding of compliance to a dietary exposure.[20] Moreover, little is known about the metabolome fingerprint from pulse consumption either individually or as a complex food group, with only a few tentative biomarkers being described.[21,22]

Determining the changes in the urinary metabolome, new biomarkers of intake, and the effect may reveal potential modifications in diet-related physiology both in healthy and diseased individuals.[23] Furthermore, metabolomic approaches have been proposed for evaluating the relationship between nutrition and health status.[24] In light of this connection, recent scientific publications have pointed out the potential health benefits of legumes in chronic diet-related diseases, such as CVD and type 2 diabetes mellitus.[6,8,25,26] Thus, the application of metabolomics to a high-cardiovascular-risk population could provide new insights into this potential relationship.

In the present work, we compared the metabolome profiles among reported high versus low legume consumers in a free-living population to find putative biomarkers reflecting intake, the effect of intake on health status, or both. Analysis of individuals under free-living conditions enables more-representative data to be obtained on the metabolome fingerprints of pulse consumers. In light of this, a better understanding of the specific role of pulse consumption in terms of health benefits, beyond their excellent nutritional profile, is expected. Therefore, the aim of the present study was to investigate dietary pulse fingerprinting in spot urine using an untargeted ^{1}H nuclear magnetic resonance (NMR) metabolomic approach on a free-living subcohort from the PREDIMED study. For this purpose, we mainly focused on urinary biomarkers of a complex pulse exposure composed of chickpeas, lentils, and beans in a combined urinary biomarker model.

2. MATERIALS AND METHODS

2.1. PREDIMED Subcohort Study

For the present study, a subsample of 50 participants from the PREDIMED study (ISRCTN 35739639; http://www.predimed.org) was selected. The PREDIMED study is a large, parallel-group, multicenter, randomized and controlled clinical trial assessing the effects of a MD on the primary prevention of CVD. The trial protocol was conducted according to the Declaration of Helsinki and was approved by the Institutional Review Boards of all the centers involved. Briefly, free-living participants (55−80 years old) without CVD that fulfilled at least one of the two criteria (type 2 diabetes mellitus or three or more major cardiovascular risk factors) were included for a MD supplemented either with extra virgin olive oil or mixed nuts.[27] The exclusion criteria were CVD, any severe chronic illness, drug or alcohol addiction, a history of allergies, or intolerance to olive oil or nuts. The subcohort consisted of a sample of participants at high cardiovascular risk, recruited from the Barcelona and Valencia PREDIMED centers. The PREDIMED study design and 137 item validated food-frequency questionnaires (FFQs) used have been reported elsewhere.[28,29] Data reported from the FFQs included information on total legume consumption and disaggregated type of legume consumed.

2.2. Stratification of the Study Population

2.2.1. Defining Potential Consumers. Both the use of FFQs and the population stratification of a cohort of individuals by consumption have demonstrated an effective approach for the study of biomarkers of food consumption.[30−32] Participants were classified into two levels (consumers and nonconsumers) of habitual intake of dietary pulses (chickpeas, lentils, or beans) on the basis of the analysis of the validated FFQs (Table S1). Intake of pulses was calculated as the sum of consumed chickpeas, lentils, and beans. Nonpulse (NP) consumers were defined as subjects with nonconsumption or sporadic consumption (≤4.00 g/day) of pulses. Habitual pulse (HP) consumers were set at a consumption of ≥25.71 g/day, regularly. To explore global pulse consumption, individuals that did not consume the three kinds of pulses simultaneously were excluded. Additionally, the condition of sporadic or nonintake of peas (≤4 g/day) was taken into consideration because the features of this type of legume are not similar to the others.[33] No other legume types were considered.

2.2.2. Selecting Individuals by Consumption. Spot urine samples were matched to corresponding individual FFQ data. From a cohort of 828 individuals, 25 subjects were defined as NP consumers and 37 as HP consumers (none of the other participants from both pulse consumer groups fulfilled any criteria). To reduce the potential sources of variability not related to pulse exposure, the number of HP consumers was balanced against NP consumers (HP = 25, NP = 25). Finally, dietary data, anthropometry, biochemical parameters, health status, and medication were explored with a view to discarding any variability unrelated to pulse consumption.

2.3. Metabolomics Analysis

2.3.1. Urine Sample Analysis and Data Processing. Morning fasting spot urine samples were collected, aliquoted, encoded, and frozen at −80 °C until they were used. Sample preparation was based on the methodology previously published.[19] The ^{1}H NMR urinary spectra were acquired using a Varian Inova 500 MHz NMR spectrometer with presaturation of the water resonance using a NOESYPRESAT pulse sequence. During the acquisition, the internal temperature was kept constant at 298 K. An exponential window function was applied to the free induction decay (FID) with a line-broadening factor of 0.3 Hz prior to Fourier transformation. For each sample, a total of 128 scans were collected into 32 K data points with a spectral width of 14 ppm at 300 K, an acquisition time of 3.2 s, and a relaxation delay of 3 s.

^{1}H NMR spectra were phased, baseline-corrected, and calibrated (TSP, 0.0 ppm) using TopSpin software (version 3.0, Bruker, BioSpin, Germany). After baseline correction, original spectral data were bucketed in intelligent bucketing domains of 0.005 ppm with ACD NMR Processor 12.0 software (Advanced Chemistry Development, Toronto, Canada). The water signal and noise regions above 9.5 ppm and below 0.5 ppm were excluded from the analysis. Data were submitted to MetaboAnalyst 3.0 for interquartile range filtering and normalization by the sum of the intensities of the spectra.[34]

2.3.2. Statistical Analysis. The NMR data set was log-transformed, Pareto scaled, and posteriorly analyzed in a multivariate approach using SIMCA-P+13.0 software (Umetrics, Umeå, Sweden). Interindividual variation may confuse the effects of intervention, particularly in multivariate data of high

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

Table 1. Tentative Discriminant Metabolites Derived from the Multi- and Univariate Analysis of ^{1}H NMR Signal Intensities in Urine From HP Consumers[a]

source	metabolite	HP vs NP	δ (multiplicity)	FDR p-value[b]	Cliff's delta[c]
choline metabolism	choline	↑	3.19 (s)	3.27×10^{-2}	0.475
	dimethylglycine	↑	2.93 (s)	3.81×10^{-2}	0.386
	TMAO	↑	3.27 (s)	7.29×10^{-3}	0.485
	dimethylamine	↑	2.72 (s)	1.05×10^{-2}	0.488
protein-related compounds	*N*-acetylglutamine	↑	2.04 (s)	2.55×10^{-2}	0.706
			2.08 (m)		
			2.26 (m)		
			4.18 (m)		
	glutamine	↑	2.12 (m)	1.17×10^{-6}	0.814
			2.46 (m)		
			3.76 (t)		
	phenylalanine	↑	3.19 (m)	3.21×10^{-2}	0.354
			3.98 (dd)		
			7.32 (d)		
			7.36 (m)		
			7.42 (m)		
	methylguanidine	↑	2.83 (s)	3.72×10^{-4}	0.635
	3-methylhistidine	↑	7.18 (s)	1.73×10^{-4}	0.658
			7.92 (s)		
energy metabolism	citric acid	↓	2.55 (dd)	8.43×10^{-5}	−0.690
			2.69 (dd)		
	cis-aconitic acid	↓	5.74 (s)	1.11×10^{-3}	−0.629
			3.12 (s)		
	glucose	↓	3.50 (m)	7.89×10^{-5}	−0.718
			4.66 (d)		
			5.25 (d)		
	acetoacetic acid	↓	2.27 (s)	1.95×10^{-2}	−0.408
	isovalerylglycine	↑	0.92 (d)	2.83×10^{-4}	0.635
			2.16 (d)		
			3.74 (d)		
	leucine	↑	0.94 (t)	1.18×10^{-3}	0.626
			1.70 (m)		
			3.72 (m)		
	isobutyric acid	↓	1.06 (d)	1.29×10^{-2}	−0.446

[a]All features have VIP values of ≥1.0 in the corresponding OSC-PLS-DA model. [b]The p-value of Student's t test with false discovery rate correction. [c]Estimation of the effect size by Cliff's delta with thresholds: $|n| < 0.330$ "small", $0.330 > |n| < 0.474$ "medium" and $|n| > 0.474$ "large". TMAO, trimethylamine-*N*-oxide. s: singlet, d: doublet, t: triplet, dd: double doublet, m: multiplet.

dimensionality. Therefore, partial least-squares discriminant analysis with orthogonal signal correction (OSC-PLS-DA) was used to explore the differences in metabolomes among the pulse consumption.[35] OSC filtration was used to reduce the variability not associated with dietary classification, as has been done in other published nutrimetabolomic studies.[19,31,36] The quality of the models was evaluated by the proportion of the variance of the response variable that is explained by the model (R^2Y) and the predictive ability (Q^2) parameters.[35] Validation of the models and the evaluation of the degree of overfitting were carried out using a permutation test (n = 200), and the correlation coefficient between the original Y and the permuted Y plotted against the cumulative R^2 and Q^2 was calculated. Those NMR signals with variable importance for projection (VIP) values ≥1 in the component of the OSC-PLS-DA model were selected as being relevant for explaining the differences in metabolic profiles. These variables were further studied through the univariate Student's t test among HP and NP consumers to assess the statistical significances. Multiple tests were controlled by the false discovery rate (FDR). Statistical significance was considered at a FDR-adjusted p-value of <0.05. Next, Cliff's delta was chosen for estimation of the effect size[37] and calculated for each feature.

2.3.3. Metabolite Identification. Metabolite identification was performed using the Chenomx NMR Suite Professional Software package (version 8.1; Chenomx Inc., Edmonton, Canada) and by comparing NMR spectral data to those available in databases such as the Human Metabolome Database (http://www.hmdb.ca), the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu), and the Madison Metabolomics Consortium Database (www.mmcd.nmrfam.wisc.edu), along with the existing NMR-based metabolomics literature. Further, a Pearson's correlation test and clustering analysis with Pearson distance and Ward's minimum variance using PermutMatrix 1.9.3.0 software[38] were applied to identify the signals corresponding to the same metabolite.

2.4. Study of Combined Urinary Biomarker Model

The interaction between gender and the resulting metabolites was evaluated by a logistic regression for discarding any effect on the biomarkers. Then, the discriminant metabolites were submitted to a stepwise logistic regression analysis (IBM SPSS

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

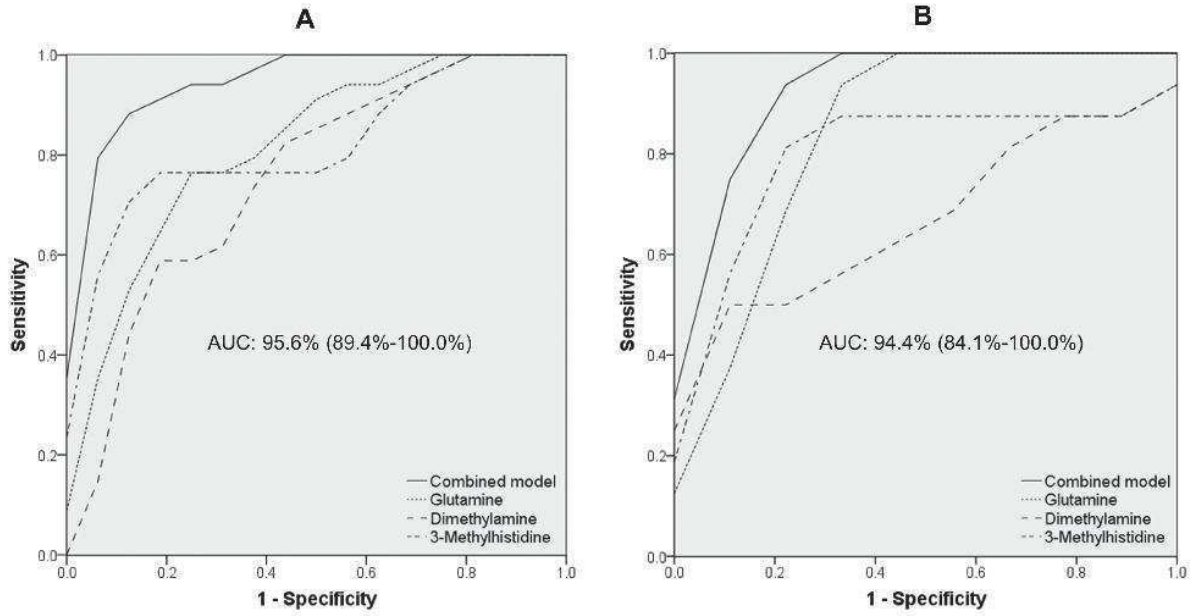


Figure 1. Receiver operating characteristic (ROC) curves of combined model (continuous line) with the area under the ROC curve and of included individual metabolites (discontinuous lines) in the training (A) and validation (B) sets.

Statistics 20 software, SPSS, Inc., Chicago, IL) to evaluate whether the combination of more than one biomarker improves the discrimination[20] of pulse consumption. The models were constructed through a dichotomous variable of pulse consumption as the dependent variable and identified metabolites as independent variables, with a p-value of <0.05 as a condition required for entering and remaining in the model. For validation of models, the analysis with a training set of 2/3 of the samples (removing 1/3 of the individuals as the validation set) was permuted 20 times. Spearman's rank correlation coefficient was used to assess correlations between the combined models and pulse consumption.

The global performance of the models was evaluated by receiver operating characteristic (ROC) curve and estimation of the area under the curve (AUC) values. The optimum cutoff for sensitivity and specificity of the biomarkers was determined as the minimum distance to the top-left corner.[39]

3. RESULTS

A flowchart of the participants allocated in the present study is presented in the Figure S1. Anthropometric measurements and biochemical analyses were performed using standardized methods.[28] HP consumers showed a pulse consumption of 38.45 ± 14.68 g/day, while NP consumers reported a consumption of 3.75 ± 3.95 g/day (mean ± SD). The characteristics of participants classified by pulse consumption are presented in Table S2. The stratified populations were not different in terms of disease (type 2 diabetes mellitus or cardiovascular risk factors), medications or biochemical parameters, among other data. Subjects who were HP consumers showed higher amounts of both dietary fiber (p < 0.01) and polyunsaturated fatty acid (p < 0.05) intakes, probably as a consequence of legume macronutrient composition.[6,40] No significances other than pulses were found with regard to food intake.

3.1. Selection of Significant Biomarkers Related to Pulse Consumption

For the analysis of the features belonging to pulse consumption in the urinary metabolome of the HP and NP consumers, an orthogonal signal correction was applied before PLS-DA analysis. The OSC-PLS-DA analysis of the two groups resulted in a latent variable model with R^2Y and Q^2 values of 0.954 and 0.809, respectively, indicating that the model was able to classify most of the individuals in the correct consumption group. The corresponding permutation tests showed negative Q^2 intercepts with a value of −0.164, implicating validation of the model.[35] With the purpose of selecting the most discriminative urinary markers of consumption, only the statistically significant variables coming from both multi- and univariate analyses simultaneously were considered.

3.2. Identified Biomarkers of Habitual Pulse Consumption

A total of 16 compounds were identified as discriminant metabolites of pulse consumption. Metabolites and chemical shifts identified corresponding to statistical analyses are presented in Table 1. The total number of metabolites related to pulse consumption was divided into categories as follows: (i) choline metabolism: choline, dimethylglycine, trimethylamine-*N*-oxide (TMAO) and dimethylamine; (ii) protein-related compounds: 3-methylhistidine, methylguanidine, phenylalanine, glutamine and *n*-acetylglutamine; and (iii) energy metabolism: glucose, leucine, isovalerylglycine, and isobutyric, acetoacetic, citric, and *cis*-aconitic acids.

3.3. Combined Urinary Biomarker Approach

Logistic regression analysis revealed that there was no significant interaction between gender and the metabolites (p > 0.05; all) as shown in Table S3. For the improvement of the discrimination between groups (HP and NP consumers), a conditional stepwise variable selection method, through a binary logistic regression analysis, was used on a combination of more than one discriminant metabolite. Table S4 shows the

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

Table 2. Receiver Operating Characteristic Curve Parameters of Combined Models and of Individual Metabolites in Both Training and Validation Sets

	set[b]	sensitivity (%)	specificity (%)	AUC (95% CI)
combined model	training	88.2	93.7	95.6 (89.4−100.0)
	validation	87.5	88.9	94.4 (84.1−100.0)
3-methylhistidine	training	76.5	87.5	82.4 (67.7−97.0)
	validation	87.5	77.8	80.6 (56.2−100.0)
glutamine	training	76.5	81.2	81.6 (67.0−96.3)
	validation	87.5	77.8	84.7 (65.3−100.0)
dimethylamine	training	82.4	62.5	75.0 (57.8−92.2)
	validation	50.0	66.7	68.1 (40.4−100.0)

[a]AUC: area under the ROC curve. CI: confidence interval. [b]Corresponding to 2/3 of the population for the training and 1/3 for the validation set.

resulting metabolites included in all 20 permuted models and the contribution to the model. A total of three metabolites were included in the fitted model according to the maximum AUC, which contained two protein-related metabolites (glutamine and 3-methylhistidine) and one choline-related metabolite (dimethylamine). These three metabolites correlated individually with the pulse consumption. However, the combined model exhibited the strongest correlation ($r = 0.73$, $p < 0.01$) with the pulse exposure, as shown in Table S5.

The ROC curve analysis was used to evaluate the combined metabolite model and their metabolites using both training and validation sets separately. The highest AUC was for the combined metabolite model for both training (AUC = 95.6%) and validation (AUC = 94.4%) sets, including glutamine, 3-methylhistidine, and dimethylamine followed by the individual metabolites 3-methylhistidine (AUC = 82.4%), glutamine (AUC = 81.6%), and dimethylamine (AUC = 75.0%), as shown in Figure 1. The equations generated from the logistic regression and the AUCs from the models with their sensitivity and specificity are shown in Table 2.

4. DISCUSSION

In this study, we present a panel of different urinary metabolites related to habitual pulse exposure using a ^{1}H NMR based untargeted nutrimetabolomic approach in a free-living population. In addition, high correlations were found when the exposure was assessed as a continuous variable (defined by the combined biomarker panel).

4.1. Characterization of Pulse Fingerprinting in Urine

4.1.1. Pulse Metabolomic Fingerprinting and Choline Metabolism. Several compounds found in the spot urine of pulse consumers are related to choline. Thus, pulses, as a rich source of choline,[41] may be the precursor of additional metabolites that are susceptible to microbial degradation generating new compounds.[42] Therefore, the increase of several intermediates of choline metabolism, such as TMAO and dimethylamine, appears to be a consequence of the microbial activity in HP consumers. The increase of dimethylamine, which is also a downstream product of choline, supports the microbial degradation of TMAO from choline. Furthermore, TMAO was identified as a major source of urinary dimethylamine in humans,[43] directly related to gut microbiota metabolism.[44] However, the increase of urinary dimethylglycine may also come from the choline contained in pulses. The enzymes choline dehydrogenase, betaine aldehyde dehydrogenase, and betaine homocysteine methyltransferase lead to dimethylglycine from choline.[45] Therefore, the results of the present study suggest the impact on urinary metabolome by choline from pulses that is degraded via both (i) mammalian pathways in which choline is converted to dimethylglycine through betaine, and (ii) microbial metabolism in which choline is degraded to trimethylamine, TMAO, and dimethylamine. For this reason, we propose dimethylamine and dimethylglycine in spot urine as potential candidates of biomarkers of pulse consumption. Nevertheless, these choline-related metabolites need to be further explored in controlled studies confirming that they are food intake biomarkers instead of reflecting metabolic differences due to the pulse consumption. Figure 2 shows both proposed pathways for downstream products of choline.

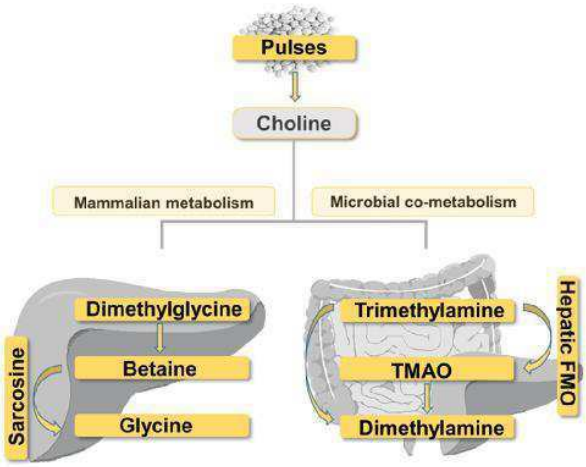


Figure 2. Proposed pathways for choline degradation from pulses including significant metabolites in HP consumers in the present study.

4.1.2. Pulse Metabolomic Fingerprinting and Protein-Related Compounds. With regard to the increases in glutamine and the acetylated form *n*-acetylglutamine, several explanations may be proposed. Glutamine and *n*-acetylglutamine could come from dietary sources because glutamine is found in high-protein foods, such as pulses.[46] Another explanation could be the alteration of urinary levels previously shown in this type of population,[47] affected by pulse consumption. There was a higher excretion of 3-methylhistidine in HP consumers. This metabolite is a biomarker of meat and fish consumption,[48] denoting a potential role as a biomarker of consumption of high-protein foods. Interestingly, all food sources of this metabolite are also protein sources, including pulses as a vegetable source, as highlighted in the

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

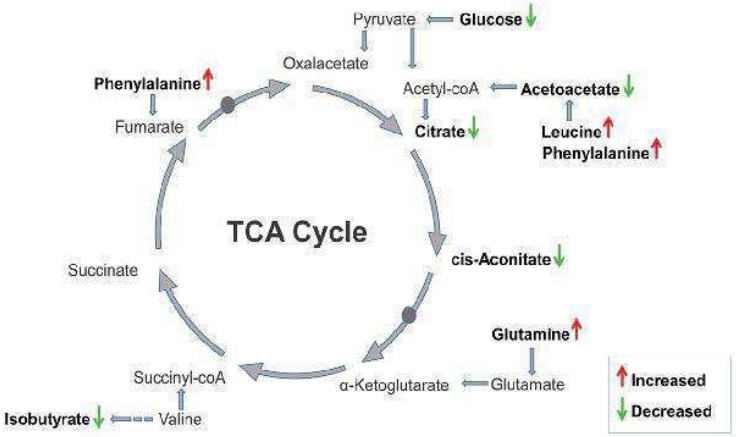


Figure 3. Modified metabolites found in HP consumers connected to energy metabolism.

present study. However, 3-methylhistidine is also a muscle protein breakdown that is sensitive to gender and age.[49] Therefore, its use as biomarker should be considered carefully. Methylguanidine is derived from protein catabolism and from the breakdown of creatinine;[50,51] therefore, it may be related to protein from pulses.

4.1.3. Pulse Metabolomic Fingerprinting and Energy Metabolism. The signals of several usual metabolites were altered between the two groups. However, their definition as food intake biomarkers is controversial. Instead, they probably reflect metabolic differences associated with being a low and high consumer, determined on the basis of the study design. Most of the biomarkers found in the present study are metabolites related to energy metabolism. The lower excretion of acetoacetic acid, glucose, and tricarboxylic acid (TCA) cycle intermediates (citric and *cis*-aconitic acids) appears to involve a different energy modulation according to the pulse consumption. This fact is in part reinforced by changes in BCAAs and subproducts, which are implicated in energy metabolism. For example, isobutyric acid is a short-chain fatty acid that is a product of BCAA catabolism of valine, which is a glucogenic BCAA metabolized via the methylmalonyl−CoA in the TCA cycle.[52] However, acetoacetic acid is a ketone body produced in the human liver for fatty acid breakdown,[53] which serves as a source of energy when normal glycolysis is altered. Interestingly, acetoacetic acid was shown to be increased in diabetes mellitus.[54] Hence, we hypothesize that gluconeogenesis may be diminished in pulse consumers, supported by the urinary reduction of acetoacetic and isobutyric acids, and the reduction of TCA cycle intermediates and urinary glucose. Furthermore, it was observed that pulse consumption has a glucose-lowering role in diabetes mellitus,[55,56] thereby explaining the lower plasma glucose concentration and lower urinary excretion. Figure 3 shows the resulting endogenous metabolites connected to the TCA cycle. Nevertheless, the small sample size that resulted after the stratification of the population leads to only exploratory results that should be confirmed.

The role of other findings such as increases of leucine and phenylalanine in pulse consumers is unclear. On the one hand, these habitual urinary compounds could be increased as a consequence of pulses being the source. On the other hand, another explanation of these findings could support the hypothesis above. Leucine, which is an acetoacetic acid precursor, may modulate glucose metabolism through oxidation, as well as insulin signaling and release. In addition, stimulation of glucose recycling via the glucose−alanine cycle by leucine may inhibit protein breakdown.[57,58] Alterations in urinary leucine have also been proposed for the prediction of diabetes mellitus, probably related to the perturbed energy metabolism.[54] The origin of increased phenylalanine is also uncertain. This ketogenic amino acid is also altered in an insulin-resistant state and obesity.[59] Overall, the consumption of pulses seems to affect the energy metabolism in the studied population.

4.2. New Biomarker Panel to Characterize Habitual Pulse Consumption

To delimit the prediction of habitual pulse intake, composed of lentils, chickpeas, and beans, a combination of more than one discriminatory metabolite had to be studied. The combination of three metabolites enhanced considerably the AUC and the confidence interval of the model in comparison with individual metabolites, as shown in Table 2. The developed model indicated that glutamine, 3-methylhistidine, and dimethylamine were strong candidates for exposure biomarkers. It is important to note that the role of the component coming from choline metabolism suggests the importance of this metabolite as a biomarker of intake. Interestingly, metabolites displaying changes in energy metabolism were scarcely considered by the stepwise logistic regression. None of the other metabolites entered the model, probably as a result of collinearity in the evidence provided by these compounds, which may originate from the same metabolic pathways, giving similar biological or dietary information.[36] Instead, two metabolites related to protein coming from pulses and one connected to microbiota choline degradation were established in the combined metabolite model, giving complementary information, showing a better discrimination (AUC > 90% in both training and validation sets) than each metabolite individually (AUC < 90% in all cases) and reinforcing the improved capacity of biomarker patterns to distinguish between different dietary exposures.

5. CONCLUSIONS

We applied an untargeted ^{1}H NMR-based metabolomic strategy to distinguish the urinary metabolome of habitual pulse consumption in a free-living population. Stepwise logistic regression analysis exhibited a useful approach to designing a combined urinary biomarker model, taking into consideration

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

the different characteristics of pulses. With regard to food metabolome, this study points to a central role of choline contained in pulses and breakdown products such as dimethylglycine, TMAO, and dimethylamine. Protein-related compounds such as glutamine, 3-methylhistidine, and methylguanidine were also increased in the urine of HP consumers. The combined metabolite model indicated that dimethylamine, 3-methylhistidine, and glutamine were the strongest candidates for exposure prediction. In relation to energy metabolism, numerous compounds connected to the TCA cycle, including BCAAs and acetoacetic acid, were modified, denoting a substantial impact on energy metabolism modulation and on urinary glucose in this population. However, because the status of type 2 diabetes mellitus or three or more major cardiovascular risk factors in the studied population could have a distinctive energy modulation, properly controlled interventions could confirm the findings observed in this cross-sectional study.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.6b00860.

Tables showing criteria for stratifying participants by frequency of consumption, characteristics of the study population according to pulse consumption, interaction between gender and the metabolites found in the present study, permuted models used in training and validation sets with the resulting metabolites, and correlations between legume consumption and the combined model for prediction of legume exposure and considered individual metabolites; and a flowchart of subjects from the PREDIMED subcohort included in the study. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: rafallorach@ub.edu. Tel: +34.934033798.
*E-mail: candres@ub.edu. Tel: +34.934034840. Fax: +34.93403593.

ORCID

Cristina Andres-Lacueva: 0000-0002-8494-4978

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study is supported by PCIN-2014-133 (FOODBALL project) from the Spanish Ministry of Economy and Competitiveness (MINECO) together with the Joint Programming Initiative "Healthy Diet for Healthy Life" (JPI-HDHL, http://www.healthydietforhealthylife.eu) and the award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR. F.M-G. acknowledges the APIF Ph.D. fellowship (University of Barcelona). M.U-S. thank thes "Ramón y Cajal" programme (RYC-2011-09677) from MINECO and the Fondo Social Europeo. E.A.-A. thanks CONACYT (Mexico) for the Ph.D. fellowship.

■ ABBREVIATIONS

AUC, area under the curve; FFQ, food frequency questionnaire; FID, free induction decay; HP, habitual pulses; ISRCTN, International Standard Randomized Controlled Trial Number; KOD, potassium deuteroxide; MD, Mediterranean diet; NMR, nuclear magnetic resonance; NP, nonpulses; OSC-PLS-DA, partial least-squares discriminant analysis with orthogonal signal correction; ROC, receiver operating characteristic; TCA, tricarboxylic acid; TMAO, trimethylamine-*N*-oxide; TSP, 3-(trimethylsilyl)-proprionate-2,2,3,3-d_4; VIP, variable importance projection

■ REFERENCES

(1) Bergouignan, A.; Momken, I.; Schoeller, D. A.; Simon, C.; Blanc, S. Metabolic fate of saturated and monounsaturated dietary fats: the Mediterranean diet revisited from epidemiological evidence to cellular mechanisms. *Prog. Lipid Res.* **2009**, *48* (3−4), 128−147.

(2) Salas-Salvadó, J.; Bulló, M.; Babio, N.; Martínez-González, M. Á.; Ibarrola-Jurado, N.; Basora, J.; Estruch, R.; Covas, M. I.; Corella, D.; Arós, F.; et al. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care* **2011**, *34* (1), 14−19.

(3) Marlow, G.; Ellett, S.; Ferguson, I. R.; Zhu, S.; Karunasinghe, N.; Jesuthasan, A. C.; Han, D. Y.; Fraser, A. G.; Ferguson, L. R. Transcriptomics to study the effect of a Mediterranean-inspired diet on inflammation in Crohn's disease patients. *Hum. Genomics* **2013**, *7* (1), 24.

(4) Lopez-Legarrea, P.; Fuller, N. R.; Zulet, M. A.; Martinez, J. A.; Caterson, I. D. The influence of Mediterranean, carbohydrate and high protein diets on gut microbiota composition in the treatment of obesity and associated inflammatory state. *Asia Pac. J. Clin. Nutr.* **2014**, *23* (3), 360−368.

(5) Estruch, R.; Ros, E.; Salas-Salvadó, J.; Covas, M.-I.; Corella, D.; Arós, F.; Gómez-Gracia, E.; Ruiz-Gutiérrez, V.; Fiol, M.; Lapetra, J.; et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N. Engl. J. Med.* **2013**, *368* (14), 1279−1290.

(6) Bouchenak, M.; Lamri-Senhadji, M. Nutritional quality of legumes, and their role in cardiometabolic risk prevention: a review. *J. Med. Food* **2013**, *16* (3), 185−198.

(7) Roy, F.; Boye, J. I.; Simpson, B. K. Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. *Food Res. Int.* **2010**, *43* (2), 432−442.

(8) Campos-Vega, R.; Loarca-Piña, G.; Oomah, B. D. Minor components of pulses and their potential impact on human health. *Food Res. Int.* **2010**, *43* (2), 461−482.

(9) Sofi, F.; Abbate, R.; Gensini, G. F.; Casini, A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. *Am. J. Clin. Nutr.* **2010**, *92* (5), 1189−1196.

(10) Dilis, V.; Trichopoulou, A. Nutritional and health properties of pulses. *Mediterranean Journal of Nutrition and Metabolism*; IOS Press January 1, 2009; pp 149−157.

(11) Ramalingam, A.; Kudapa, H.; Pazhamala, L. T.; Weckwerth, W.; Varshney, R. K. Proteomics and Metabolomics: Two Emerging Areas for Legume Improvement. *Front. Plant Sci.* **2015**, *6*, 1116.

(12) Faris, M. A. I. E.; Takruri, H. R.; Issa, A. Y. Role of lentils (Lens culinaris L.) in human health and nutrition: A review. *Mediterranean Journal of Nutrition and Metabolism*. IOS Press: Amsterdam, The Netherlands, 2013; pp 3−16.

(13) Bach-Faig, A.; Berry, E. M.; Lairon, D.; Reguant, J.; Trichopoulou, A.; Dernini, S.; Medina, F. X.; Battino, M.; Belahsen, R.; Miranda, G.; et al. Mediterranean diet pyramid today. Science and cultural updates. *Public Health Nutr.* **2011**, *14* (12A), 2274−2284.

(14) U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015−2020 Dietary Guidelines for Americans. http://health.gov/dietaryguidelines/2015/ (accessed May 9, 2016).

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

(15) Britten, P.; Marcoe, K.; Yamini, S.; Davis, C. Development of food intake patterns for the MyPyramid Food Guidance System. *J. Nutr. Educ. Behav.* **2006**, *38* (6), S78−92.
(16) Mithril, C.; Dragsted, L. O.; Meyer, C.; Tetens, I.; Biltoft-Jensen, A.; Astrup, A. Dietary composition and nutrient content of the New Nordic Diet. *Public Health Nutr.* **2013**, *16* (5), 777−785.
(17) MacWilliam, S.; Wismer, M.; Kulshreshtha, S. Life cycle and economic assessment of Western Canadian pulse systems: The inclusion of pulses in crop rotations. *Agric. Syst.* **2014**, *123*, 43−53.
(18) O'Gorman, A.; Gibbons, H.; Brennan, L. Metabolomics in the Identification of Biomarkers of Dietary Intake. *Comput. Struct. Biotechnol. J.* **2013**, *4* (5), 1−7.
(19) Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Lupianez-Barbero, A.; Estruch, R.; Corella, D.; Fitó, M.; Arós, F.; Ruiz-Canela, M.; Salas-Salvadó, J.; et al. Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *J. Proteome Res.* **2015**, *14* (1), 531−540.
(20) Garcia-Aloy, M.; Llorach, R.; Urpi-Sarda, M.; Tulipani, S.; Estruch, R.; Martínez-González, M. a.; Corella, D.; Fitó, M.; Ros, E.; Salas-Salvadó, J.; et al. Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: The predimed study. *J. Proteome Res.* **2014**, *13* (7), 3476−3483.
(21) Perera, T.; Young, M. R.; Zhang, Z.; Murphy, G.; Colburn, N. H.; Lanza, E.; Hartman, T. J.; Cross, A. J.; Bobe, G. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol. Nutr. Food Res.* **2015**, *59* (4), 795−806.
(22) Bonetti, A.; Marotti, I.; Dinelli, G. Urinary excretion of kaempferol from common beans (Phaseolus vulgaris L.) in humans. *Int. J. Food Sci. Nutr.* **2007**, *58* (4), 261−269.
(23) Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; van der Hooft, J. J. J.; Wishart, D. S. The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* **2014**, *99* (6), 1286−1308.
(24) McNiven, E. M. S.; German, J. B.; Slupsky, C. M. Analytical metabolomics: nutritional opportunities for personalized health. *J. Nutr. Biochem.* **2011**, *22* (11), 995−1002.
(25) Souza, R. G. M.; Gomes, A. C.; Naves, M. M. V; Mota, J. F. Nuts and legume seeds for cardiovascular risk reduction: scientific evidence and mechanisms of action. *Nutr. Rev.* **2015**, *73* (6), 335−347.
(26) Wilson, C. Nutrition: Consumption of legumes might be beneficial in type 2 diabetes mellitus. *Nat. Rev. Endocrinol.* **2013**, *9* (1), 3.
(27) Martinez-Gonzalez, M. A.; Corella, D.; Salas-Salvado, J.; Ros, E.; Covas, M. I.; Fiol, M.; Warnberg, J.; Aros, F.; Ruiz-Gutierrez, V.; Lamuela-Raventos, R. M.; et al. Cohort Profile: Design and methods of the PREDIMED study. *Int. J. Epidemiol.* **2012**, *41* (2), 377−385.
(28) Estruch, R.; Martinez-Gonzalez, M. A.; Corella, D.; Salas-Salvado, J.; Ruiz-Gutierrez, V.; Covas, M. I. Effects of a Mediterranean-Style Diet on Cardiovascular Risk Factors. *Ann. Int. Med.* **2006**, *145* (1), 1−11.
(29) Fernández-Ballart, J. D.; Piñol, J. L.; Zazpe, I.; Corella, D.; Carrasco, P.; Toledo, E.; Perez-Bauer, M.; Martínez-González, M. A.; Salas-Salvadó, J.; Martín-Moreno, J. M. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. *Br. J. Nutr.* **2010**, *103*, 1808−1816.
(30) Khan, N.; Khymenets, O.; Urpí-Sardà, M.; Tulipani, S.; Garcia-Aloy, M.; Monagas, M.; Mora-Cubillos, X.; Llorach, R.; Andres-Lacueva, C. Cocoa polyphenols and inflammatory markers of cardiovascular disease. *Nutrients* **2014**, *6* (2), 844−880.
(31) Pujos-Guillot, E.; Hubert, J.; Martin, J. F.; Lyan, B.; Quintana, M.; Claude, S.; Chabanas, B.; Rothwell, J. a.; Bennetau-Pelissero, C.; Scalbert, A.; et al. Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: Citrus fruit as a case study. *J. Proteome Res.* **2013**, *12* (4), 1645−1659.
(32) Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; Khymenets, O.; Bullo, M.; Corella, D.; Fito, M.; Martinez-Gonzalez, M. A.; Estruch, R.; Andres-Lacueva, C. An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics* **2015**, *11* (4), 797−806.
(33) Food and Agriculture Organization. Definition and classification of commodities, 1994. http://www.fao.org/ES/faodef/fdef04e.htm (accessed May 1, 2016).
(34) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res.* **2015**, *43* (W1), W251−W257.
(35) Llorach-Asunción, R.; Jauregui, O.; Urpi-Sarda, M.; Andres-Lacueva, C. Methodological aspects for metabolome visualization and characterization: a metabolomic evaluation of the 24 h evolution of human urine after cocoa powder consumption. *J. Pharm. Biomed. Anal.* **2010**, *51* (2), 373−381.
(36) Garcia-Aloy, M.; Llorach, R.; Urpi-Sarda, M.; Jáuregui, O.; Corella, D.; Ruiz-Canela, M.; Salas-Salvadó, J.; Fitó, M.; Ros, E.; Estruch, R.; et al. A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED study. *Mol. Nutr. Food Res.* **2015**, *59* (2), 212−220.
(37) Cliff, N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin* **1993**, *114* (3), 494−509.
(38) Caraux, G.; Pinloche, S. PermutMatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* **2005**, *21* (7), 1280−1281.
(39) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **2013**, *9* (2), 280−299.
(40) De Almeida Costa, G. E.; Da Silva Queiroz-Monici, K.; Pissini Machado Reis, S. M.; De Oliveira, A. C. Chemical composition, dietary fibre and resistant starch contents of raw and cooked pea, common bean, chickpea and lentil legumes. *Food Chem.* **2006**, *94* (3), 327−330.
(41) Lewis, E. D.; Kosik, S. J.; Zhao, Y.-Y.; Jacobs, R. L.; Curtis, J. M.; Field, C. J. Total choline and choline-containing moieties of commercially available pulses. *Plant Foods Hum. Nutr.* **2014**, *69* (2), 115−121.
(42) Tremaroli, V.; Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **2012**, *489* (7415), 242−249.
(43) Zhang, A. Q.; Mitchell, S. C.; Ayesh, R.; Smith, R. L. Dimethylamine formation in man. *Biochem. Pharmacol.* **1993**, *45* (11), 2185−2188.
(44) Dumas, M.-E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C.; et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (33), 12511−12516.
(45) Friesen, R. W.; Novak, E. M.; Hasman, D.; Innis, S. M. Relationship of Dimethylglycine, Choline, and Betaine with Oxoproline in Plasma of Pregnant Women and Their Newborn Infants. *J. Nutr.* **2007**, *137* (12), 2641−2646.
(46) Marinangeli, C. P. F.; Jones, P. J. H. Pulse grain consumption and obesity: effects on energy expenditure, substrate oxidation, body composition, fat deposition and satiety. *Br. J. Nutr.* **2012**, *108* (S1), S46−S51.
(47) Roberts, L. D.; Koulman, A.; Griffin, J. L. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. *Lancet Diabetes Endocrinol.* **2014**, *2* (1), 65−75.
(48) Brennan, L.; Gibbons, H.; O'Gorman, A. An Overview of the Role of Metabolomics in the Identification of Dietary Biomarkers. *Curr. Nutr. Rep.* **2015**, *4* (4), 304−312.
(49) Aranibar, N.; Vassallo, J. D.; Rathmacher, J.; Stryker, S.; Zhang, Y.; Dai, J.; Janovitz, E. B.; Robertson, D.; Reily, M.; Lowe-Krentz, L.; et al. Identification of 1- and 3-methylhistidine as biomarkers of skeletal muscle toxicity by nuclear magnetic resonance-based metabolic profiling. *Anal. Biochem.* **2011**, *410* (1), 84−91.
(50) Yokozawa, T.; Fujitsuka, N.; Oura, H. Studies on the precursor of methylguanidine in rats with renal failure. *Nephron* **1991**, *58* (1), 90−94.

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

(51) Ando, A.; Orita, Y.; Nakata, K.; Tsubakihara, Y.; Takamitsu, Y.; Ueda, N.; Yanase, M.; Abe, H. Effect of low protein diet and surplus of essential amino acids on the serum concentration and the urinary excretion of methylguanidine and guanidinosuccinic acid in chronic renal failure. *Nephron* **1979**, *24* (4), 161−169.
(52) Hutson, S. M.; Sweatt, A. J.; Lanoue, K. F. Branched-chain amino acid metabolism: implications for establishing safe intakes. *J. Nutr.* **2005**, *135* (6), 1557S−64S.
(53) Stern, J. R. Enzymes of acetoacetate formation and breakdown. *Methods Enzymol.* **1955**, *1*, 573−585.
(54) Urpi-Sarda, M.; Almanza-Aguilera, E.; Tulipani, S.; Tinahones, F. J.; Salas-Salvadó, J.; Andres-Lacueva, C. Metabolomics for Biomarkers of Type 2 Diabetes Mellitus: Advances and Nutritional Intervention Trends. *Curr. Cardiovasc. Risk Rep.* **2015**, *9* (3), 12.
(55) Mudryj, A. N.; Yu, N.; Aukema, H. M. Nutritional and health benefits of pulses. *Appl. Physiol., Nutr., Metab.* **2014**, *39* (11), 1197−1204.
(56) Singhal, P.; Kaushik, G.; Mathur, P. Antidiabetic potential of commonly consumed legumes: a review. *Crit. Rev. Food Sci. Nutr.* **2014**, *54* (5), 655−672.
(57) Layman, D. K.; Walker, D. A. Potential importance of leucine in treatment of obesity and the metabolic syndrome. *J. Nutr.* **2006**, *136* (1), 319S−23S.
(58) Shearer, J.; Duggan, G.; Weljie, A.; Hittel, D. S.; Wasserman, D. H.; Vogel, H. J. Metabolomic profiling of dietary-induced insulin resistance in the high fat-fed C57BL/6J mouse. *Diabetes, Obes. Metab.* **2008**, *10* (10), 950−958.
(59) Adams, S. H. Emerging Perspectives on Essential Amino Acid Metabolism in Obesity and the Insulin-Resistant State. *Adv. Nutr.* **2011**, *2* (6), 445−456.

DOI: 10.1021/acs.jproteome.6b00860
J. Proteome Res. 2017, 16, 1483−1491

Urinary ^{1}H-NMR metabolomic fingerprinting reveals biomarkers of pulse consumption related to energy metabolism modulation in a subcohort from the PREDIMED study

Francisco Madrid-Gambin[†], Rafael Llorach*[†], Rosa Vázquez-Fresno[†], Mireia Urpi-Sarda[†], Enrique Almanza-Aguilera[†], Mar Garcia-Aloy[†], Ramon Estruch[§#], Dolores Corella[⊥#] and Cristina Andres-Lacueva*[†]

[†]Biomarkers & Nutrimetabolomics Laboratory. Nutrition, Food Science and Gastronomy Department, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona, Spain.

[§]Department of Internal Medicine, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Barcelona, Spain.

[⊥]Department of Preventive Medicine and Public Health, University of Valencia, Valencia, Spain.

[#]CIBER OBN, The Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid, Spain.

*E-mail: rafallorach@ub.edu. Tel: +34.934033798. *E-mail: candres@ub.edu. Tel: +34.934034840. Fax: +34.93403593.

Table of contents

Table S5 – Correlations between legume consumption and the combined model for prediction of legume exposure and considered individual metabolites.

Figure S1 – Flow chart of subjects from the PREDIMED subcohort included in the study.

Table S1. Criteria for stratification of participants by frequency of consumption

Pulses	Minimum individual intake (g/day)	Total pulse intake (g/day)[§]	Classification
Chickpeas	≥ 8.57	≥ 25.71[†]	Consumers
Lentils	≥ 8.57		
Beans	≥ 8.57		
Chickpeas	≤ 4.00	≤ 4.00	Non-consumers
Lentils	≤ 4.00		
Beans	≤ 4.00		

[§]Values corresponding to the total sum of the consumption of the three pulse types.
[†]Threshold used for classification.

Table S2. Characteristics of the study population according to pulse consumption (37 females, 13 males)[a]

Parameters	NP consumers	HP consumers	p–value[†]
Gender (F)*	23 (92%)‡	14 (56%)	<0.01
Age (Y)	65.76 ± 4.93§	68.80 ± 6.23	0.06
Current smokers	8 (32%)	9 (36%)	0.77
Dietary data			
Energy (Kcal/day)	2209.25 ± 575.47	2474.65 ± 489.10	0.09
Protein (g/day)	97.09 (77.07–111.79)[ι]	106.90 (95.29–120.56)	0.11
Carbohydrates (g/day)	233.69 ± 60.77	259.03 ± 54.64	0.13
Fat (g/day)	94.23 (70.07–108.85)	102.06 (92.41–122.57)	0.13
MUFA (g/day)	47.01 (36.48–55.93)	52.02 (46.56–59.96)	0.08
PUFA (g/day)*	12.06 (9.86–18.93)	17.84 (12.42–22.99)	0.03
SFA (g/day)	24.38 (18.09–28.88)	24.62 (22.13–34.35)	0.20
Fibre (g/day)*	22.21 ± 5.05	30.26 ± 6.23	<0.001
Alcohol (g)	0.68 (0–6.64)	3.04 (0.69–8.85)	0.20
Cholesterol	372.61 (286.16–480.33)	432.71 (384.26–495.18)	0.18
Food			
Legumes (g/day)*	4 (0–4)	29.71 (25.71–46.86)	<0.001
Chickpeas (g/day)*	0 (0–0)	8.57 (8.57–8.57)	<0.001
Lentils (g/day)*	0 (0–4)	8.57 (8.57–25.71)	<0.001
Beans (g/day)*	0 (0–0)	8.57 (8.57–8.57)	<0.001
Vegetables (g/day)	348.29 ± 111.47	402.64 ± 83.24	0.06
Fruits (g/day)	482.90 ± 258.12	419.74 ± 175.60	0.32
Cereals (g/day)	245.31 ± 86.82	241.02 ± 89.24	0.86
Dairy products (g/day)	303.57 (210.24–581.67)	296.43 (221.43–623.21)	1.00
Meat and products (g/day)	132.38 (93.33–199.05)	121.43 (106.29–168.57)	0.90
Fish and products (g/day)	108.50 ± 42.86	118.88 ± 53.36	0.45
Biscuits and products (g/day)	7.14 (0–27.52)	10.48 (3.33–28.09)	0.49
Olive oil (g/day)	30 (25–50)	50 (25–50)	0.41
Nuts (g/day)	2 (0–12.86)	6.28 (4.28–14.86)	0.16
Coffee (g/day)	0 (0–3.33)	0 (0–50)	0.67
Anthropometrical parameters			
Height (m)	1.57 ± 0.07	1.59 ± 0.09	0.35
Weight (kg)	76.91 ± 14.82	76.98 ± 13.77	0.99
BMI (kg/m2)	30.38 (27.81–34.91)	28.87 (26.98–32.59)	0.69
Waist (cm)	101.12 ± 12.82	102.72 ± 11.1	0.64
Waist to height	0.64 ± 0.069	0.65 ± 0.065	0.48
Biochemical parameters			
Glycaemia (mg/dl)	117 (99–170)	105 (96–125)	0.11
Total cholesterol (mg/dl)	208.88 ± 41.09	208.88 ± 41.19	1.00
LDL cholesterol (mg/dl)	131.22 ± 36.58	131.62 ± 30.90	0.97
HDL cholesterol (mg/dl)	52.64 ± 12.00	52.48 ± 11.56	0.96
Triglycerides (mg/dl)	95 (84–177)	109 (87–147)	0.73

Table S2. Continued

Diseases and Medication			
% Type 2 diabetes mellitus	12 (48%)	15 (60%)	0.39
Antidiabetic agents	13 (52%)	8 (32%)	0.15
% Hypertension	21 (84%)	19 (76%)	0.48
Antihypertensives	18 (72%)	18 (72%)	1.00
% Dyslipidaemia	17 (68%)	16 (64%)	0.77
Lipid-lowering drugs	12 (48%)	10 (40%)	0.39
Anti-inflammatory drugs	11 (44%)	8 (32%)	0.38

[a]Descriptive analyses were compared between NP and HP consumers. [†]Statistical comparisons are from Student's t-test, Mann-Whitney U or chi-square, as appropriate. *Statistical significance between groups. [‡]n (%). [§]Mean ± SD (all such values). [ɻ]Median; IQR in parentheses (all such values).

Table S3. Interaction between gender and the urinary metabolites related to pulse consumption found in the present study[a]

Pathway	Metabolite[†]	p-value
Choline metabolism	Choline	0.998
	Dimethylglycine	0.314
	TMAO	0.472
	Dimethylamine	0.633
Protein-related compounds	N-acetylglutamine	0.281
	Glutamine	0.998
	Phenylalanine	0.858
	Methylguanidine	0.481
	3-Methylhistidine	0.557
Energy metabolism	Citric acid	0.208
	Cis-aconitic acid	0.854
	Glucose	0.999
	Acetoacetic acid	0.918
	Isovalerylglycine	0.995
	Leucine	0.529
	Isobutyric acid	0.533

[a]Interaction was studied using a logistic regression. [†]Significant metabolites related to pulse consumption.

Table S4. Permuted models used in training/validation sets with the resulting metabolites[a]

Model	Metabolites[†]	Beta	p-Value
1	Glutamine	24.942	0.004
	Methylguanidine	23.576	0.012
2	Glutamine	21.998	0.010
	Methylguanidine	24.513	0.005
3	Leucine	29.209	0.022
	Glutamine	28.483	0.011
4	Dimethylamine	6.743	0.011
	Methylhistidine	9.837	0.004
5	Glutamine	44.728	0.031
	TMAO	4.068	0.047
6	Dimethylamine	15.956	0.014
	Methylhistidine	12.041	0.013
7	Methylguanidine	25.297	0.012
	Dimethylglycine	5.672	0.041
8	Dimethylamine	15.373	0.034
	Choline	22.325	0.030
9	Glutamine	18.504	0.047
	Methylguanidine	16.400	0.021
	Methylhistidine	8.948	0.030
10	Leucine	14.686	0.018
	Glutamine	13.836	0.008
11	Glutamine	18.855	0.018
	Methylguanidine	19.026	0.016
	Choline	24.987	0.029
12	Glutamine	13.488	0.043
	Dimethylamine	6.329	0.019
	Methylhistidine	11.513	0.022
13	Leucine	15.574	0.004
	Dimethylglycine	12.516	0.010
14	Glutamine	18.122	0.007
	Methylguanidine	18.358	0.006
15	Glutamine	24.992	0.003
	Methylguanidine	23.523	0.014
16	Glutamine	21.605	0.008
	Methylguanidine	22.382	0.008
17	Leucine	18.529	0.041
	Glutamine	26.631	0.014
	Methylguanidine	21.028	0.017
18	Glutamine	18.339	0.009
	Methylguanidine	18.149	0.006
19	Glutamine	17.022	0.008
	TMAO	3.575	0.025
20	Leucine	17.906	0.045
	Glutamine	26.238	0.011
	Methylguanidine	14.422	0.036

[a]Models through dichotomous variable of pulse consumption vs metabolites. [†]Resulting metabolites that remained in the model.

Table S5. Correlations between legume consumption and the combined model for prediction of pulse exposure and considered individual metabolites

	$r^{\dagger}$	*P value*
Combined model	0.73	1.64×10^{-9}
Glutamine	0.53	8.03×10^{-5}
3-Methylhistidine	0.45	9.65×10^{-4}
Dimethylamine	0.44	1.57×10^{-3}

[†]Spearman's rank correlation coefficient.

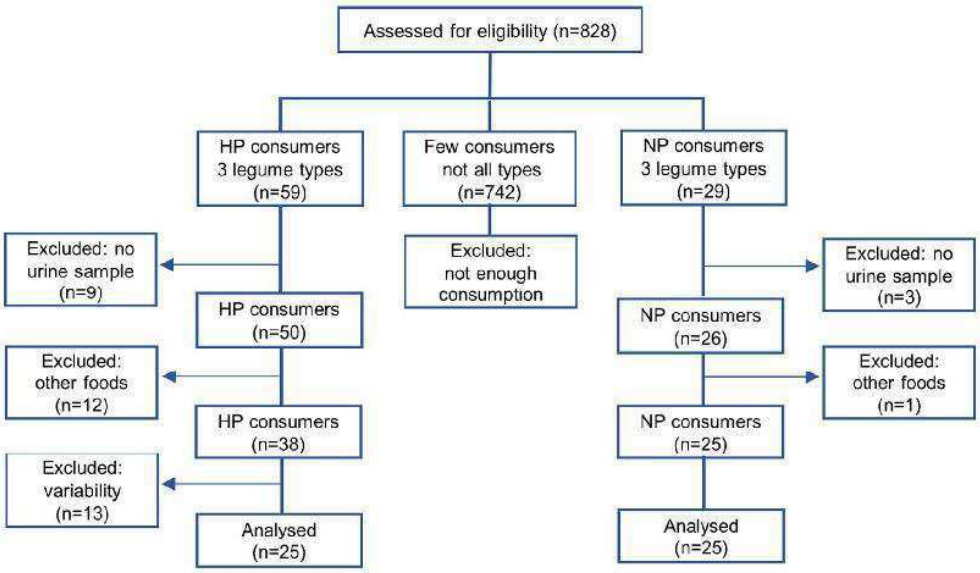
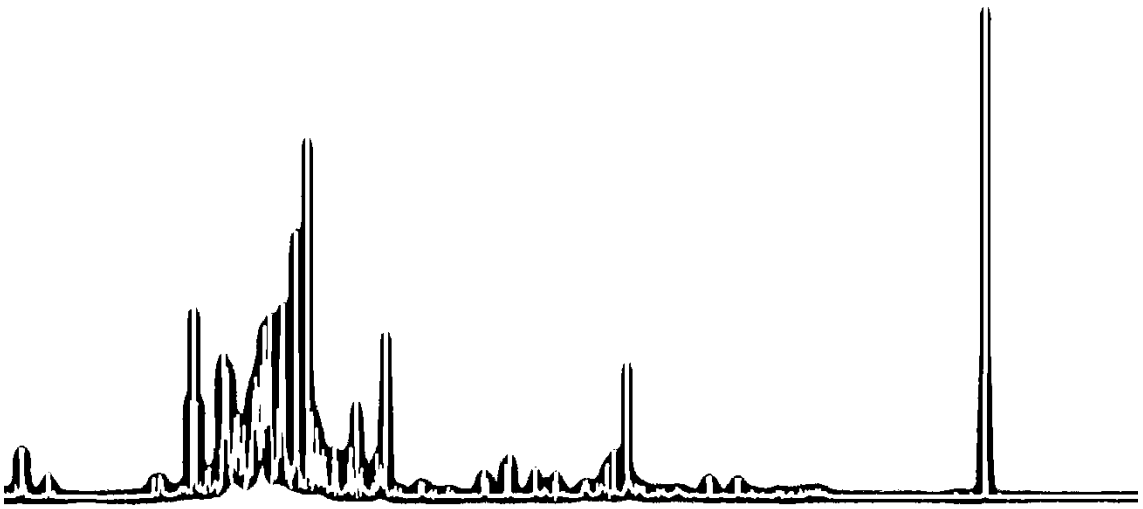


Figure S1. Flow chart of subjects from the PREDIMED subcohort included in the study.

Discusión y conclusiones



Capítulo 09.

DISCUSIÓN GENERAL

Este capítulo pretende ofrecer una visión global a modo de discusión general de todos los resultados obtenidos. La presente Tesis Doctoral se centró en el estudio del perfil metabolómico en muestras de orina y sangre asociado al consumo de legumbres (concretamente lentejas, garbanzos y alubias) y café.

En primer lugar, se compararon las principales metodologías utilizadas en el procesado de muestras sanguíneas (plasma/suero) para determinar la metodología más apropiada (publicación 1) para ser aplicada a estudios controlados. Posteriormente, se llevaron a cabo dos estudios de intervención: (i) una intervención aguda con legumbres del que se analizaron muestras de sangre y orina tomadas en diferentes puntos temporales tras una única ingesta (publicación 2), y (ii) una intervención con una bebida de café con alto contenido en compuestos bioactivos del que se analizó la orina tras una ingesta puntual y tras una ingesta prolongada durante 4 semanas (publicación 3). También se realizó un estudio observacional de la exposición a legumbres en una población con condiciones de vida libre (publicación 4).

9.1. Estudio comparativo de metodologías de RMN para la preparación de muestras sanguíneas (plasma/suero)

La identificación y cuantificación de metabolitos es uno de los principales desafíos en el desarrollo de la metabolómica. De hecho, existen diferentes iniciativas para el estudio y perfeccionamiento de métodos para superar este reto (Nagana Gowda et al. 2015a). En el caso de los estudios con RMN, la gran cantidad de proteínas y fosfolípidos que están presentes en muestras sanguíneas aumenta la dificultad para explorar metabolitos de bajo peso molecular (Nagana Gowda y Raftery 2015b). En este sentido, en este trabajo se utilizaron las metodologías más empleadas en el campo de la metabolómica mediante RMN: la precipitación de las proteínas con metanol y la ultrafiltración (Nagana Gowda y Raftery 2015b), y el procedimiento de SPE, el cual constituye una metodología validada y competente para el tratamiento de muestras de plasma/suero en EM (Tulipani et al. 2013). Para nuestro conocimiento,

ningún estudio ha evaluado estas metodologías teniendo en cuenta la capacidad de identificar el mayor número de metabolitos, la concentración absoluta de los metabolitos identificados, y aspectos prácticos importantes para el desarrollo de la técnica como la capacidad de automatización, el tiempo empleado y los costes globales del análisis metabolómico a través de estos de estos procedimientos.

Para el estudio comparativo, se utilizaron muestras de plasma de cinco individuos, las cuales fueron mezcladas en un único "*pool"* para ser analizado por triplicado a través de estos tres procedimientos (publicación 1). El hecho de utilizarse el mismo "*pool"* para realizar la comparación entre métodos permite, entre otras ventajas, evitar el problema de la variación inter-individual que podría influir en la comparación de la concentración de los metabolitos. Este fenómeno podría indicar diferencias provenientes de las variaciones biológicas en la concentración de los metabolitos en vez de indicar diferencias relacionadas con los diferentes tratamientos de muestra analizados. El uso de un "*pool"* evita la posible influencia de covariables confusoras, como género y edad, que pueden modificar el metaboloma (Slupsky et al. 2007).

El análisis comparativo entre los procedimientos estudiados se basó en la capacidad de identificar el mayor número de metabolitos y en la comparación de la concentración absoluta de los metabolitos previamente identificados con respecto a la ultrafiltración por constituir la metodología más frecuentemente usada en estudios metabolómicos con muestras de plasma/suero (Bathe et al. 2011; Psychogios et al. 2011; Farshidfar et al. 2012; Nagana Gowda et al. 2015a). Además, se elaboró un puntaje final considerando también la capacidad de automatización del procedimiento, el tiempo total empleado y los costes derivados del procedimiento en las diferentes metodologías. La metodología más adecuada entre los procedimientos analizados fue la precipitación de macromoléculas con metanol. La precipitación con metanol exhibió el mayor número de metabolitos identificados. Sin embargo, en relación a la comparación de la concentración absoluta de los metabolitos, en esta técnica hubo una menor concentración en compuestos volátiles (acetona y acetoacetato). En cambio, la ultrafiltración presentó mayores concentraciones de estos metabolitos, además de histidina y citrato. El cálculo del puntaje final otorgó la máxima puntuación a la precipitación con metanol debido a su compatibilidad de automatización y el bajo coste de la técnica, entre otros aspectos. Por consiguiente, esta metodología fue posteriormente aplicada al estudio de intervención con legumbres (publicación 2).

La metodología fue ligeramente optimizada para poder ser aplicada en un sistema automatizado que permitió el análisis de un gran número de muestras en un tiempo reducido gracias a una colaboración establecida con el *Swedish NMR Centre* (Gotemburgo, Suecia). En este centro se tuvo acceso a un equipamiento de alta sensibilidad (crio-sonda NMR 800 mHz) con preparación on-line de las muestras bajo la tutela del Dr. Anders Pedersen. Posteriormente, se realizó el pretratamiento de los datos conjuntamente con el desarrollo estadístico gracias a la colaboración con el grupo del Dr. Rikard Landberg y con el seguimiento científico del Dr. Carl Brunius en la *Swedish University of Agricultural Sciences* (Uppsala, Suecia). Como fruto de esta estancia de siete meses, se produjeron las siguientes optimizaciones en el flujo de trabajo: (i) se redujo el volumen necesario para el análisis como consecuencia del acceso a un equipo altamente sensible; (ii) se automatizó la preparación de la muestra, lo que redujo considerablemente el tiempo final del análisis; (iii) se desarrolló e implementó un algoritmo de alineación de las señales espectrales de la RMN que mejoró la potencia estadística de los resultados; y (iv) se implementaron técnicas de identificación y confirmación de metabolitos a través de la técnica *STOCSY* (del inglés, *Statistical TOtal Correlation SpectroscopY*), la cual se basa en correlaciones estadísticas que permiten la construcción de un espectro 2D (Cloarec et al. 2005).

9.2. Biomarcadores de ingesta: "*food metabolome*"

El estudio de la huella que dejan los alimentos en el perfil metabolómico se ha definido como una estrategia para identificar nuevos biomarcadores y mejorar la precisión de la medida de exposiciones dietéticas (Llorach et al. 2012; Scalbert et al. 2014). El "*food metabolome"* hace referencia a aquellos metabolitos derivados directamente de la ingesta de alimentos, así como aquellos derivados de su absorción y biotransformación tisular y/o por parte de la microbiota (Fardet et al. 2008). La identificación de estos compuestos puede informar de la exposición a los alimentos y de qué forma han sido metabolizados antes de ser excretados (Llorach et al. 2012). De esta manera, el uso de la metabolómica ha permitido identificar nuevos biomarcadores asociados a la ingesta de determinados alimentos, los cuales podrían proporcionar una medida objetiva de la exposición dietética que carece de los sesgos y errores asociados a los métodos basados en el auto-reporte de la ingesta (Scalbert et al. 2014). En la presente Tesis Doctoral se han identificado los biomarcadores nutricionales asociados a la ingesta de legumbres [lentejas, garbanzos y alubias (publicación 2)] y café (publicación 3).

9.2.1. Intervención con legumbres

En la publicación 2 se presentan los resultados del análisis del metaboloma urinario y sérico asociado a una ingesta única de legumbres. El alcaloide trigonelina fue el principal biomarcador urinario proveniente del "*food metabolome"*, el cual se halló originariamente en los tres tipos de legumbres.

Respecto a la cantidad presente de trigonelina en los alimentos ensayados, las alubias presentaron casi el doble de concentración de este compuesto que los garbanzos, y unas seis veces más que las lentejas. Sin embargo, estos resultados difieren de los presentados en un estudio realizado por Zwart y colaboradores, en el que estos autores mostraron que la concentración de trigonelina fue superior en los garbanzos, seguida de lentejas y finalmente alubias (De Zwart et al. 2003). En relación a la concentración urinaria, se observaron diferencias estadísticamente significativas en la excreción de este compuesto según el tipo de legumbre ingerido a partir de la segunda hora (T_{2h}) tras la ingesta acordes con la concentración de trigonelina en los alimentos. Estos hallazgos sugieren una relación "dosis-respuesta" de este compuesto en el que una mayor ingesta del compuesto proveniente de diferentes legumbres supone una mayor excreción del biomarcador. A la vez, se observó que en el caso de las alubias y

los garbanzos la concentración máxima de trigonelina fue en T_{6h}, mientras que para las lentejas ésta fue en T_{4h}. De esta forma, estos datos permiten establecer la relación "tiempo-respuesta" respecto a este metabolito, el cual fue detectable a partir de 2 h después de la ingesta de legumbres. De manera general, y acorde con lo observado con la trigonelina, se observó una excreción retardada de la mayoría de los metabolitos provenientes del "*food metabolome*" para el caso de alubias y garbanzos, en comparación con las lentejas. Una posible explicación en los patrones de excreción de estos metabolitos podría ser la digestión ralentizada, absorción y la subsiguiente excreción derivada de las diferentes cantidades de fibra presentes en los alimentos (3 g/100g en lentejas, frente a los 9 g/100g en garbanzos y 17 g/100g en alubias) utilizados en el estudio, acorde con lo que ya se ha observado previamente en la bibliografía científica (Holt et al. 1979; Tharanathan y Mahadevamma 2003). En otro estudio, la trigonelina también fue identificada tras el consumo de alubias en muestras de suero analizado por EM (Perera et al. 2015). Sin embargo, ésta no pudo ser identificada en el suero analizado por RMN en los participantes de este estudio, probablemente debido a una menor sensibilidad de la RMN en comparación con la EM. La excreción de trigonelina sin ninguna biotransformación sugiere una menor "variación inter-individual", cumpliendo así con varios de los parámetros propuestos para la validación biológica del biomarcador (detallado en el capítulo 01 de la presente Tesis Doctoral). Sin embargo, debido al diseño del estudio, existe una carencia en otros parámetros como "sensibilidad y especificidad", y "robustez", como se ha comentado previamente en el apartado de antecedentes bibliográficos. El cumplimiento o no de estos parámetros es intuido en la publicación 4, y son comentados posteriormente en esta sección de discusión.

Las señales espectrales correspondientes a 3-metilhistidina incrementaron después del consumo de los tres tipos de legumbres, pero especialmente tras la ingesta de alubias. Este metabolito ha sido descrito en orina tras el consumo de carne (especialmente pollo) y pescado (Cheung et al. 2017), y tras la ingesta de legumbres en el estudio observacional incluido en la presente Tesis Doctoral (publicación 4). En el presente estudio de intervención, los incrementos de 3-metilhistidina, además de resultar estadísticamente significativos en comparación con el alimento control también fueron estadísticamente significativos entre los distintos tipos de legumbres. Una explicación podría basarse en que los niveles de este metabolito en orina reflejen la concentración de histidina en el alimento de partida, la cual se halla en los alimentos en la misma distribución que la 3-metilhistidina urinaria (alubias > garbanzos /

lentejas). La posterior metilación y excreción podría repercutir en las diferencias observadas en el patrón de excreción, el cual nuevamente es más precoz tras la ingesta de lentejas (excreción máxima a T_{2h}) que en la de garbanzos y alubias (excreción máxima a T_{4h}). Acorde con los resultados presentados, este metabolito puede reflejar la ingesta de legumbres (publicación 2), carne y pescado (Cheung et al. 2017), a partir de 2 h después de su consumo.

Los compuestos dimetilglicina, trimetilamina y metilamina fueron identificados tras el consumo de legumbres. Estos metabolitos pueden aparecer en la orina debido a la exposición a aminas cuaternarias como la colina presente en leguminosas (Zhang et al. 1999; Madrid-Gambin et al. 2017) o a la degradación de lecitina y/o carnitina (Zhao et al. 2015). Dado que la concentración de colina no fue significativamente mayor en esta legumbre resulta más plausible que provengan de la degradación de lecitina y/o carnitina. Sin embargo, estos compuestos no pudieron ser identificados en ninguna legumbre, probablemente debido a una fuerte superposición de picos en el área correspondiente a éstos en los espectros. En el estudio observacional incluido en la presente Tesis Doctoral (publicación 4) estos metabolitos también se asociaron con la ingesta de legumbres.

El nivel urinario de lisina aumentó tras el consumo de todas las legumbres en comparación con el alimento control. También se observaron diferencias en la excreción entre los diferentes tipos de legumbres, acorde con la composición previa de lisina en el alimento (alubias > lentejas > garbanzos). Una vez más, las concentraciones máximas se produjeron a diferentes tiempos (T_{2h} tras lentejas; T_{4h} tras alubias y garbanzos).

El ácido pipecólico se identificó en las alubias (publicación 2) y pese a que ha sido descrito en el suero de humanos tras su consumo (Perera et al. 2015), éste no fue detectado ni en la orina ni en el suero de los participantes en los estudios de esta Tesis Doctoral.

En lo que se refiere a suero, se identificó aproximadamente la mitad de metabolitos en comparación con la orina. Además, estos metabolitos fueron en su mayoría compuestos endógenos (publicación 2), sugiriendo una menor adecuación en comparación con la orina para estudios de biomarcadores nutricionales, especialmente de ingesta (Gibney et al. 2005; Claus y Swann 2013). Histidina y dimetilglicina incrementaron tras el consumo de lentejas. Sin embargo, estos metabolitos no

presentaron diferencias estadísticamente significativas en suero tras la ingesta de garbanzos y alubias en comparación con el alimento control a pesar de haber sido identificados en orina tras la ingesta de estas legumbres. Los niveles séricos de lisina también incrementaron después de la ingesta de garbanzos. El 2-hidroxibutirato, 3-hidroxiisovalerato y asparagina incrementaron en suero tras el consumo de alubias. En un estudio recientemente publicado sobre diferentes patrones dietéticos, los autores observaron que los metabolitos plasmáticos asparagina y 2-hidroxibutirato, entre otros, eran discriminantes de una dieta baja en grasa (Esko et al. 2017). Sin embargo, en este estudio, las legumbres contenían más grasa que el alimento control.

Respecto a los metabolitos que mostraron una reducción en su concentración, la prolina disminuyó tras la ingesta de garbanzos mientras que los niveles de glucosa fueron menores tras la ingesta de todas las legumbres en comparación con la ingesta del alimento control (pasta). Este hecho estaría en línea con los resultados obtenidos, en los que la glucosa, tanto en orina como en suero, se encontró en menores niveles tras la ingesta de legumbres, las cuales tienen un bajo índice glicémico (Kim et al. 2016), en comparación con la ingesta del alimento control.

Las variaciones en los niveles de metabolitos provenientes de la microbiota, como el 3-hidroxiisovalerato, reflejan diferencias en el metabolismo de aminoácidos aromáticos y polifenoles por parte de los microorganismos del colon (Rechner et al. 2002; Moco et al. 2012). Estos resultados apuntan a establecer relaciones entre el consumo de legumbres y su impacto sobre la microbiota intestinal (Faris et al. 2013).

9.2.2. Intervención con café

El estudio de intervención con una bebida de café con alto contenido en compuestos bioactivos (publicación 3) permitió detectar e identificar cinco metabolitos provenientes del "*food metabolome*" en orina tras el consumo de dicha bebida**.** Respecto al análisis del alimento, la trigonelina y los compuestos furanos en el café se identificaron como posibles precursores de metabolitos del "*food metabolome"*. En orina, la trigonelina y la 2-furoilglicina fueron identificadas tras la ingesta aguda del café. Además, la excreción urinaria de trigonelina también resultó estadísticamente significativa tras la ingesta prolongada de la bebida utilizada en este estudio. Los metabolitos excretados sin sufrir grandes procesos de metabolización, como ocurre con la trigonelina, resultarían ser biomarcadores potenciales, debido a que probablemente presentarán una menor variabilidad inter-individual. La concentración inicial que podemos encontrar de este

compuesto en el café es alrededor de unas seis veces superior con respecto a las legumbres (De Zwart et al. 2003). Pese a esta diferencia, aun existiría una cierta limitación acerca del uso de la trigonelina como biomarcador tanto de café como de legumbres ya que ambos alimentos podrían ser precursores dietéticos de este metabolito. Se han descrito situaciones similares en la bibliografía donde el mismo compuesto puede presentar diferentes fuentes dietéticas. Un ejemplo serían los metabolitos microbianos de las procianidinas (como las fenil-valerolactonas o los ácidos fenil-valéricos) que han sido identificados en orina de voluntarios tras el consumo de alimentos ricos en estas sustancias como por ejemplo té o cacao (Chen y Sang 2014; Garcia-Aloy et al. 2015). El uso de un panel multi-metabolito puede solventar esta limitación mediante la combinación de varios biomarcadores (Garcia-Aloy et al. 2015, 2017).

La 2-furoilglicina es un producto derivado de los compuestos furanos presentes en el café (Moon y Shibamoto 2009), como son el ácido 2-furóico y el alcohol furfurílico encontrados en la bebida de café utilizada en este estudio. Este metabolito también ha sido asociado con la ingesta de café en estudios previos (Guertin et al. 2014; Heinzmann et al. 2015).

En relación a los metabolitos procedentes de la microbiota intestinal, se identificaron el ácido hipúrico, el ácido 3-hidroxihipúrico y el ácido 3-(3-hidroxifenil)-3-hidroxipropiónico tras el consumo prolongado de la bebida. De manera general es asumido que estos compuestos derivan del metabolismo microbiano de los ácidos clorogénicos presentes en el café (Gonthier et al. 2006).

El análisis metabolómico de los alimentos reveló los propios compuestos o bien los precursores de los metabolitos susceptibles de considerarse biomarcadores, particularmente de ingesta (“*food metabolome”*), lo que facilitó la identificación e interpretación del compuesto desde el alimento hasta la muestra biológica (Scalbert et al. 2014) como ya se ha visto en la bibliografía científica (Heinzmann et al. 2010; Garcia-Perez et al. 2016). Algunos ejemplos son la trigonelina encontrada tanto en la bebida de café como en las legumbres analizadas, la cual fue posteriormente encontrada en la orina en los estudios de intervención (publicaciones 2 y 3); aminas cuaternarias como la colina, lecitina y/o carnitina que probablemente resultaron en el incremento de dimetilglicina, trimetilamina y metilamina (publicaciones 2 y 4); y compuestos furanos en el café, que dieron como resultado la 2-furoilglicina (publicación 3).

9.3. Biomarcadores de efecto: impacto sobre el organismo

Los biomarcadores de efecto reflejan el impacto de una exposición dietética sobre el metaboloma endógeno. Las diferencias en los metabolitos endógenos podrían estar relacionadas con el efecto en el organismo tras su ingesta, permitiendo generar nuevas hipótesis sobre los mecanismos de acción (Llorach et al. 2012). Asimismo, resulta de interés la capacidad para explicar o apoyar los potenciales efectos sobre la salud derivados de la exposición a un determinado alimento mediante el estudio de los biomarcadores del efecto y su situación en las rutas metabólicas afectadas (Scalbert et al. 2014).

9.3.1. Intervención con legumbres

En la intervención con legumbres (publicación 2) se encontraron diferencias estadísticamente significativas en la excreción urinaria de varios aminoácidos. La presencia de los aminoácidos ramificados (AARs) valina y leucina fue mayor tras el consumo de alubias en comparación con la ingesta de lentejas y garbanzos, y el alimento control (pasta). La leucina tiene un papel cetogénico en el organismo (catabolismo de ácidos grasos), formando acetil-CoA y acetoacetato, mientras que la valina es glucogénica (síntesis de glucógeno), entrando en el ciclo de Krebs como succinil-CoA. La valina se metaboliza a succinato a través del metilmalonato y la metilmalonil-CoA (Hutson et al. 2005) con lo que se sugieren consecuencias clínicas relacionadas con el metabolismo energético que podrían estar relacionadas con la menor glucosa urinaria observada en las publicaciones 2 y 4. Sin embargo, aunque la excreción de valina fue superior en las legumbres, la excreción de metilmalonato fue menor que en el alimento control. Estos resultados junto con los bajos niveles de 3-aminoisobutanoato y 3-hidroxiisobutirato en comparación con el alimento control sugieren una disminución en el metabolismo de ARRs ya que éstos se forman durante el metabolismo de la valina. Este hecho estaría en línea con el aumento de AARs en la orina.

El incremento de los aminoácidos glutamina y *N*-acetil-glutamina tras el consumo de lentejas podría reflejar un aumento de estos metabolitos debido a la presencia de éstos en las lentejas, o bien, una modulación de la producción de este aminoácido glucogénico como consecuencia de los cambios en los AARs y posteriormente en el ciclo de Krebs (Newsholme et al. 2003; Hutson et al. 2005). La baja concentración de este metabolito en el alimento de partida sugiere una respuesta endógena en lugar de

una excreción directa, como también se ha sugerido en otros estudios tras el consumo de carne (Stella et al. 2006), consumo de soja (Solanky et al. 2005) y la exposición a la dieta Mediterránea (Vázquez-Fresno et al. 2015c), lo que también podría estar relacionado con los bajos niveles de glucosa en orina (publicaciones 2 y 4).

9.3.2. Intervención con café

Respecto a los compuestos endógenos se observó un aumento de la excreción en los metabolitos implicados en el ciclo de Krebs (ácidos succínico, cítrico, isobutírico y 3-metil-2-oxovalérico). Este resultado apunta a una modulación en el metabolismo energético, en concreto, el incremento de los productos de valina (ácido isobutírico) e isoleucina (ácido 3-metil-2-oxovalérico) podría estar conectado con el metabolismo de éstos a succinato vía metilmalonil-CoA (Hutson et al. 2005). Esta hipótesis también reforzaría la explicación del porqué del aumento de ácido succínico. Junto a esto, también cabe destacar un aumento de la excreción de ácido cítrico, el cual también es un intermediario del ciclo de Krebs. Resultados similares se observaron en un estudio en ratas, en el que aumentaron los intermediarios del ciclo de Krebs tras el consumo de un suplemento con compuestos bioactivos del café, concretamente los CGAs (Ruan et al. 2013).

9.4. Biomarcadores nutricionales en estudios con diseños diferenciados

La replicación de los biomarcadores obtenidos en los diferentes estudios de intervención en otras poblaciones permite incrementar el grado de evidencia de los hallazgos encontrados, como ya se sugirió previamente en estudios de nutrimetabolómica (Garcia-Aloy et al. 2015; Vázquez-Fresno et al. 2015b), proporcionando una valiosa información a la hora de evaluar la capacidad de un biomarcador nutricional (Garcia Aloy 2014b; Scalbert et al. 2014).

Las publicaciones desarrolladas durante la presente Tesis Doctoral han incluido la investigación del metaboloma tras el consumo de legumbres y café a través de estudios con diseños diferenciados. Inicialmente, los biomarcadores nutricionales tras el consumo agudo de legumbres fueron determinados en un estudio de intervención controlado (publicación 2). Con la investigación del impacto sobre el perfil metabolómico asociado al consumo de legumbres en un estudio observacional (publicación 4) se determinaron los biomarcadores nutricionales en una población con condiciones de vida libre. Además, también se estudiaron los biomarcadores nutricionales derivados tanto de una exposición aguda como de una exposición prolongada durante 4 semanas a café con alto contenido en compuestos bioactivos (publicación 3).

Respecto al consumo de legumbres, los biomarcadores de ingesta como 3-metilhistidina, dimetilglicina y varios compuestos relacionados con la acción microbiana sobre aminas cuaternarias provenientes de los propios alimentos (TMA/TMAO) fueron observados tanto en el estudio de intervención puntual como en el estudio observacional. Por lo tanto, a pesar de que estos metabolitos no son altamente específicos del consumo de legumbres, la combinación de estos biomarcadores en un panel multi-metabolito podría aumentar la especificidad y sensibilidad en la predicción de la exposición, como ya se demostró en estudios previos del grupo de investigación donde se ha desarrollado la presente Tesis Doctoral (Vázquez-Fresno et al. 2015b; Garcia-Aloy et al. 2015, 2017). Además, a través del estudio de intervención, se observaron indicios de los parámetros de “dosis-respuesta”, “causalidad” y “tiempo-respuesta”. Cabe mencionar el caso de la trigonelina, que presentó una gran relevancia en el contexto de estudios controlados y que posteriormente este comportamiento no fue replicado en el estudio observacional, exhibiendo el no cumplimiento del parámetro “sensibilidad y especificidad”, así como el de “robustez”. En el estudio observacional (publicación 4), la trigonelina urinaria no fue

significativa, por lo que posteriormente se analizó la ingesta reportada de café entre los consumidores y no consumidores habituales de legumbres. En este análisis, se observó que la ingesta de café fue similar entre los consumidores y los no consumidores habituales de legumbres. Por lo tanto, este metabolito probablemente no apareció como discriminativo del consumo de legumbres en el estudio observacional, debido al enmascaramiento derivado de la fuerte presencia de trigonelina en el café, muy superior a la de legumbres (De Zwart et al. 2003), dado que este metabolito también aparece en orina tras el consumo agudo y prolongado de café (publicación 3). Con ello, se establece una baja especificidad y robustez de la trigonelina como biomarcador de legumbres, especialmente si no se considera el consumo de café. En lo que refiere a la lisina como biomarcador, este metabolito no fue discriminante de la exposición a legumbres en el estudio con sujetos en condiciones de vida libre, con lo que denota también una baja sensibilidad y especificidad como biomarcador. Además, se observaron concentraciones diferenciales de este metabolito en función del tipo de legumbre sugiriendo una excreción directa. Sin embargo, la lisina es un compuesto poco específico de un alimento, al igual que en el caso de la trigonelina.

En relación a los biomarcadores de efecto replicados, cabe destacar que tanto los AARs como los menores niveles de glucosa urinaria fueron replicados en el estudio de intervención y observacional de legumbres. Estos resultados denotan la validez de estos biomarcadores de efecto y aumentan la evidencia del impacto beneficioso sobre el metabolismo energético y/o del metabolismo de la glucosa por parte de las legumbres, como ha sido asociado recientemente en un estudio prospectivo sobre diabetes tipo 2 (Becerra-Tomás et al. 2017).

En relación al estudio de café, la trigonelina fue estadísticamente significativa tanto en la intervención aguda como la intervención sostenida de la bebida, implicando una relación de “causalidad” y “acumulación”. Sin embargo, son necesarios estudios observacionales adicionales para investigar otros parámetros relacionados con la especificidad y robustez de este biomarcador, así como el estudio de otros compuestos que han de ser replicados en otras poblaciones. En este contexto, la trigonelina fue asociada con el consumo de café en un estudio observacional correlacionando este metabolito con el consumo reportado en una población en condiciones de vida libre proveniente de la cohorte *French SU.VI.MAX2* (Rothwell et al. 2014). La 2-furoilglicina también fue asociada al consumo de café en otro estudio

que consistió en una intervención con proteína animal, ingesta de frutas y consumo de uva (Heinzmann et al. 2015), por lo que nuevamente se establecen los parámetros de “especificidad” y “robustez”.

Estos resultados ponen de manifiesto la importancia de la aproximación metabolómica no dirigida, ya que puede ayudar a nuevos descubrimientos útiles tanto en la determinación de la exposición a los alimentos, como en la interpretación biológica de los posibles mecanismos implicados en los efectos relacionados con la salud por parte de los componentes de la dieta.

Capítulo 10.

CONCLUSIONES

De los diferentes trabajos realizados durante la presente Tesis Doctoral derivan las siguientes conclusiones:

1. Se ha implementado un método altamente reproducible y aplicable a un gran número de muestras que ha permitido identificar biomarcadores nutricionales representativos del perfil metabólico sérico y/o plasmático asociados a una determinada exposición dietética. Este método se basó en la precipitación de las proteínas con metanol.

2. La huella metabolómica procedente de la ingesta de legumbres y una bebida a base de café en estudios de intervención ha permitido la identificación de biomarcadores nutricionales asociados a su consumo (biomarcadores de ingesta) y biomarcadores de efecto relacionados con el metabolismo endógeno tras el consumo de los alimentos estudiados.

3. La huella metabolómica asociada al consumo agudo de legumbres, concretamente lentejas, garbanzos y alubias, ha permitido identificar compuestos del metaboloma de los alimentos ("*food metabolome*") y potenciales biomarcadores de ingesta, tales como trigonelina, 3-metilhistidina, dimetilglicina, lisina y trimetilamina. Así mismo, también se han identificado compuestos asociados al consumo, pero procedentes del metabolismo endógeno, potenciales candidatos a biomarcadores de efecto, tales como como glucosa, aminoácidos ramificados y compuestos relacionados con el ciclo de Krebs.

4. La huella metabolómica derivada del consumo agudo y sostenido de café con alto contenido en compuestos bioactivos ha permitido identificar compuestos del "*food metabolome"*, tales como trigonelina y 2-furoilglicina, así como compuestos relacionados con la acción microbiana sobre los fitoquímicos del café [ácidos hipúrico, 3-hidroxihipúrico y 3-(3-hidroxifenil)-3-hidroxipropiónico]. En paralelo, se han identificado potenciales biomarcadores de efecto, concretamente compuestos endógenos relacionados con el ciclo de Krebs.

5. El análisis metabolómico por RMN de los alimentos utilizados en los estudios de intervención ha permitido confirmar la procedencia de biomarcadores nutricionales específicos como parte del "*food metabolome*".

6. El análisis metabolómico de legumbres (lentejas, garbanzos y alubias) permitió la identificación y cuantificación diferencial de trigonelina, lisina e histidina en cada legumbre. Esta cuantificación diferencial se mantuvo posteriormente también discriminante en la orina de los voluntarios tras el consumo de cada una de las legumbres consumidas en el estudio de intervención dietética.

7. El análisis metabolómico de la bebida a base de café permitió la identificación de trigonelina y de determinados compuestos furanos, los cuales fueron posteriormente encontrados en orina como trigonelina y 2-furoilglicina, potenciales biomarcadores de ingesta.

8. El análisis metabolómico de la orina por RMN se propone de elección respecto al plasma/suero en estudios encaminados al descubrimiento de biomarcadores nutricionales, especialmente biomarcadores de ingesta. El análisis de la orina permitió la identificación de hasta el doble de compuestos discriminantes de la ingesta de legumbres (endógenos y exógenos) respecto a los identificados en el suero (mayoritariamente metabolitos endógenos).

9. Los biomarcadores nutricionales asociados a la ingesta de legumbres en estudios de intervención fueron comparados con los resultados obtenidos en estudios observacionales al considerar la ingesta de legumbres a partir del cuestionario de frecuencia de consumo de alimentos (CFCA).

10. La huella metabolómica derivada del consumo habitual reportado de legumbres mediante CFCAs, en concreto lentejas, garbanzos y alubias, ha replicado la presencia de compuestos relacionados con el "*food metabolome*", como son dimetilglicina, 3-metilhistidina, y compuestos derivados de la acción microbiana sobre aminas cuaternarias en los alimentos (trimetilamina como precursor de trimetilamina-*N*-óxido). Los compuestos endógenos asociados a la ingesta de legumbres que fueron replicados en el estudio observacional fueron una menor glucosa urinaria, así como compuestos relacionados con el ciclo de Krebs.

11. En estudios observacionales, la combinación de varios biomarcadores nutricionales en modelos multi-metabolito, aunque *a priori* pudieran

considerarse poco específicos del consumo de un alimento, mejora la capacidad predictiva de la exposición dietética respecto a la utilización de estos biomarcadores individualmente. La combinación de 3-metilhistidina, dimetilglicina y glutamina proporcionó una mejor capacidad predictiva como biomarcador de exposición a legumbres que éstos mismos evaluados de forma individual.

CONCLUSIONS

From the results presented in this Doctoral Thesis, the derived conclusions are as follows:

1. A highly reproducible and widely applicable methodology to identify nutritional biomarkers from the metabolomic fingerprint of serum and/or plasma associated with a specific dietetic exposure has been implemented. This methodology was based on the protein precipitation using methanol.

2. The metabolomic fingerprint derived of the intake of legumes and a coffee-based beverage in intervention studies enabled the identification of nutritional biomarkers associated with their consumption (biomarkers of intake) and biomarkers of effect related to endogenous metabolism after the consumption of the studied foods.

3. The metabolomic fingerprint derived from the acute consumption of legumes (namely lentils, chickpeas and beans) enabled the identification of compounds from the food metabolome, potential biomarkers of intake, such as trigonelline, 3-methylhistidine, dimethylglycine, lysine and trimethylamine. Similarly, compounds associated with the consumption of legumes from the endogenous metabolism, potential biomarkers of effect, such as glucose, branched-chain amino acids and compounds related to the Krebs cycle were also identified.

4. The metabolomic fingerprint derived from the acute and sustained consumption of a coffee with a high content of bioactive compounds enabled the identification of compounds from the food metabolome, such as trigonelline and 2-furoylglycine, as well as compounds related to microbial activity on phytochemicals from coffee [hippuric, 3-hydroxyhippuric and 3-(3-hydroxyphenyl)-3-hydroxypropionic acids]. Similarly, potential biomarkers of effect have been identified. Specifically, endogenous compounds related to the Krebs cycle.

5. The metabolomic analysis by NMR of the foods used in the intervention studies confirmed the origin of nutritional biomarkers as part of the food metabolome.

6. The metabolomic analysis of pulses (lentils, chickpeas and beans) enabled the identification and differential quantification of trigonelline, lysine and histidine in each legume. This differential quantification was subsequently observed in the urine of the volunteers after the consumption of each pulse consumed in the intervention study.

7. The metabolomic analysis of the coffee-based beverage allowed the identification of trigonelline and certain furan compounds, which were subsequently found in urine as trigonelline *per se* and 2-furoylglycine, potential biomarkers of intake.

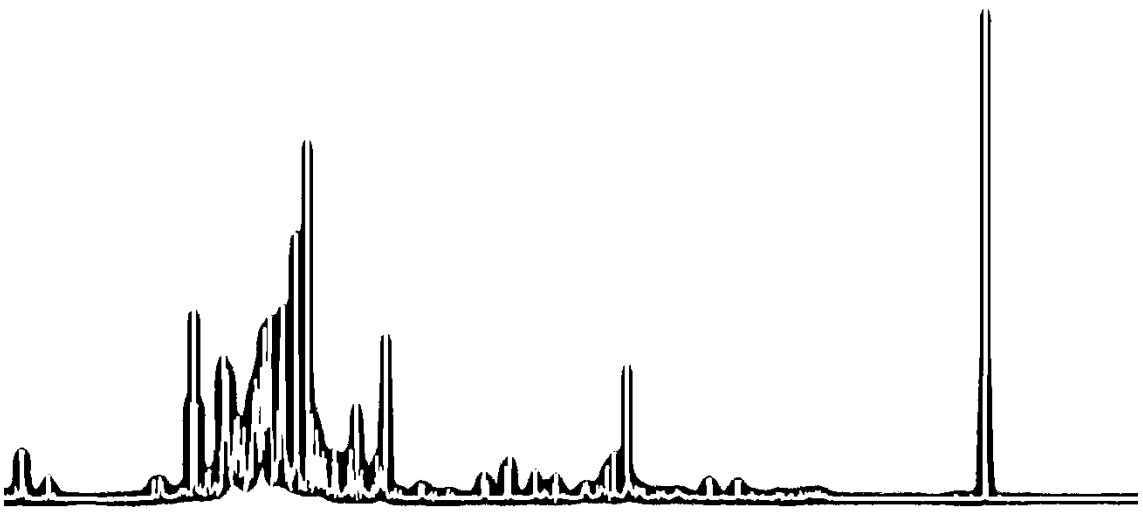
8. The NMR-based metabolomic analysis of urine has shown it as a more suitable biofluid than plasma/serum in studies aimed to discover nutritional biomarkers, especially biomarkers of intake. The analysis of the urine after the intake of pulses enabled the identification of up to twice the number of discriminant compounds (endogenous and exogenous) than serum (mainly endogenous metabolites).

9. The nutritional biomarkers associated with legume intake in intervention studies were compared with the results displayed in observational studies considering the food frequency questionnaire (FFQ) to assess the intake of pulses.

10. The metabolomic fingerprint derived from the reported consumption of pulses in FFQs, namely lentils, chickpeas and beans, replicated the presence of compounds related to the food metabolome, such as dimethylglycine, 3-methylhistidine, and compounds derived from the microbial action on quaternary amines in foods (trimethylamine as precursor of trimethylamine-N-oxide). The endogenous compounds associated with the consumption replicated in the observational study were lower urinary glucose, as well as compounds related to the Krebs cycle.

11. In observational studies, the combination of different nutritional biomarkers that may be considered to be unspecific of the consumption of certain foods, improves the capacity of prediction of the dietary exposure with respect to the use of these biomarkers individually. The combination of 3-methylhistidine, dimethylglycine and glutamine provided higher predictive power for the pulses exposure than these metabolites individually.

Referencias



Referencias bibliográficas

Agencia Española de Consumo. (2015). Encuesta ENALIA 2. Encuesta Nacional de Alimentación en población adulta, mayores y embarazadas. http://www.aecosan.msssi.gob.es/AECOSAN/web/seguridad_alimentaria/subdetalle/enalia_2.htm. Accessed 6 April 2017

Albermann, M. E., Musshoff, F., & Madea, B. (2012). A High-Performance Liquid Chromatographic-Tandem Mass Spectrometric Method for the Determination of Ethyl Glucuronide and Ethyl Sulfate in Urine Validated According to Forensic Guidelines. *Journal of Chromatographic Science*, *50*(1), 51–56. doi:10.1093/chromsci/bmr012

Álvarez Sánchez, B. (2012). *New analytical platforms in metabolomics.* [Tesis doctoral]. Universidad de Córdoba, España.

Andersen, M.-B. S. (2014). *Discovery of food exposure* markers in urine and evaluation of dietary compliance by untargeted LC-MS metabolomics. [Tesis doctoral]. Faculty of Science, University of Copenhagen.

Andersen, M.-B. S., Kristensen, M., Manach, C., Pujos-Guillot, E., Poulsen, S. K., Larsen, T. M., et al. (2014). Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Analytical and Bioanalytical Chemistry*, *406*(7), 1829–1844. doi:10.1007/s00216-013-7498-5

Assadi-Porter, F., Selen, E., & Shen, C. (2015). NMR-based metabolomics analysis in muscle and serum of middle-aged ovariectomized rats supplemented with 6-month green tea polyphenols. *The FASEB Journal*, *29*(1 Supplement), 745.2.

Bach-Faig, A., Berry, E. M., Lairon, D., Reguant, J., Trichopoulou, A., Dernini, S., et al. (2011). Mediterranean diet pyramid today. Science and cultural updates. *Public health nutrition*, *14*(12A), 2274–84. doi:10.1017/S1368980011002515

Bathe, O. F., Shaykhutdinov, R., Kopciuk, K., Weljie, A. M., McKay, A., Sutherland, F. R., et al. (2011). Feasibility of identifying pancreatic cancer based on serum metabolomics. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *20*(1), 140–7. doi:10.1158/1055-9965.EPI-10-0712

Becerra-Tomás, N., Díaz-López, A., Rosique-Esteban, N., Ros, E., Buil-Cosiales, P., Corella, D., et al. (2017). Legume consumption is inversely associated with type 2 diabetes incidence in adults: A prospective assessment from the PREDIMED study. *Clinical Nutrition*. doi:10.1016/j.clnu.2017.03.015

Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C., & Nicholson, J. K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols*, *2*(11), 2692–703. doi:10.1038/nprot.2007.376

Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., et al. (2013). The Human Urine Metabolome. *PLoS ONE*, *8*(9), e73076. doi:10.1371/journal.pone.0073076

Bouchenak, M., & Lamri-Senhadji, M. (2013). Nutritional quality of legumes, and their role in cardiometabolic risk prevention: a review. *Journal of medicinal food*, *16*(3), 1–14. doi:10.1089/jmf.2011.0238

Brennan, L. (2014). NMR-based metabolomics: From sample preparation to applications in nutrition research. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *83*, 42–49. doi:10.1016/j.pnmrs.2014.09.001

Brennan, L. (2017). The nutritional metabolomics crossroads: how to ensure success for dietary biomarkers. *The American Journal of Clinical Nutrition*, *105*(2), 293–294. doi:10.3945/ajcn.116.150847

Britten, P., Marcoe, K., Yamini, S., & Davis, C. (2006). Development of food intake patterns for the MyPyramid Food Guidance System. *Journal of nutrition education and behavior*, *38*(6 Suppl), S78-92. doi:10.1016/j.jneb.2006.08.007

Campos-Vega, R., Loarca-Piña, G., & Oomah, B. D. (2010). Minor components of pulses and their potential impact on human health. *Food Research International*, *43*(2), 461–482. doi:10.1016/j.foodres.2009.09.004

Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., & Rodrick, G. E. (2009). Metabolomic analysis in food science: a review. *Trends in Food Science & Technology*, *20*(11), 557–566. doi:10.1016/j.tifs.2009.07.002

Chen, H., & Sang, S. (2014). Biotransformation of tea polyphenols by gut microbiota. *Journal of Functional Foods*, *7*, 26–42. doi:10.1016/j.jff.2014.01.013

Cheung, W., Keski-Rahkonen, P., Assi, N., Ferrari, P., Freisling, H., Rinaldi, S., et al. (2017). A metabolomic study of biomarkers of meat and fish intake. *The American Journal of Clinical Nutrition*, *105*(3), 600–608. doi:10.3945/ajcn.116.146639

Chrysant, S. G. (2017). The impact of coffee consumption on blood pressure, cardiovascular disease and diabetes mellitus. *Expert Review of Cardiovascular Therapy*, *15*(3), 151–156. doi:10.1080/14779072.2017.1287563

Clarke, K. A., Dew, T. P., Watson, R. E. B., Farrar, M. D., Osman, J. E., Nicolaou, A., et al. (2016). Green tea catechins and their metabolites in human skin before and after exposure to ultraviolet radiation. *The Journal of nutritional biochemistry*, *27*, 203–10. doi:10.1016/j.jnutbio.2015.09.001

Claus, S. P., & Swann, J. R. (2013). Nutrimetabonomics:Applications for Nutritional Sciences, with Specific Reference to Gut Microbial Interactions. *Annual Review of Food Science and Technology*, *4*(1), 381–399. doi:10.1146/annurev-food-030212-182612

De Zwart, F. J., Slow, S., Payne, R. J., Lever, M., George, P. M., Gerrard, J. A., & Chambers, S. T. (2003). Glycine betaine and glycine betaine analogues in common foods. *Food Chemistry*, *83*(2), 197–204. doi:10.1016/S0308-8146(03)00063-3

Deng, L., Gu, H., Zhu, J., Nagana Gowda, G. A., Djukovic, D., Chiorean, E. G., & Raftery, D. (2016). Combining NMR and LC/MS Using Backward Variable Elimination: Metabolomics Analysis of Colorectal Cancer, Polyps, and Healthy Controls. *Analytical Chemistry*, *88*(16), 7975–7983. doi:10.1021/acs.analchem.6b00885

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, *26*(1), 51–78. doi:10.1002/mas.20108

Dilis, V., & Trichopoulou, A. (2009, January 1). Nutritional and health properties of pulses. *Mediterranean Journal of Nutrition and Metabolism*. IOS Press. doi:10.3233/s12349-008-0023-2

Dragsted, L. O. (2017). Relying on biomarkers for intake assessment in nutrition. *The American Journal of Clinical Nutrition*, *105*(1), 8–9. doi:10.3945/ajcn.116.148320

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, *9*(S1), 44–66. doi:10.1007/s11306-012-0434-4

Edmands, W. M. B., Ferrari, P., Rothwell, J. a., Rinaldi, S., Slimani, N., Barupal, D. K., et al. (2015). Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *American Journal of Clinical Nutrition*, *102*(4), 905–913. doi:10.3945/ajcn.114.101881

Esko, T., Hirschhorn, J. N., Feldman, H. A., Hsu, Y.-H., Deik, A. A., Clish, C. B., et al. (2017). Metabolomic profiles as reliable biomarkers of dietary composition. *The American Journal of Clinical Nutrition*, *105*(3), 547–554. doi:10.3945/ajcn.116.144428

Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., Arós, F., et al. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. *The New England journal of medicine*, *368*(14), 1279–90. doi:10.1056/NEJMoa1200303

Fardet, A., Llorach, R., Orsoni, A., Martin, J.-F., Pujos-Guillot, E., Lapierre, C., & Scalbert, A. (2008). Metabolomics provide new insight on the metabolism of dietary phytochemicals in rats. *The Journal of nutrition*, *138*(7), 1282–7.

Faris, M. A. I. E., Takruri, H. R., & Issa, A. Y. (2013). Role of lentils (Lens culinaris L.) in human health and nutrition: A review. *Mediterranean Journal of Nutrition and Metabolism*. IOS Press. doi:10.1007/s12349-012-0109-8

Farshidfar, F., Weljie, A. M., Kopciuk, K., Buie, W. D., Maclean, A., Dixon, E., et al. (2012). Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome medicine*, *4*(5), 42. doi:10.1186/gm341

Fontana, L., & Partridge, L. (2015). Promoting Health and Longevity through Diet: From Model Organisms to Humans. *Cell*, *161*(1), 106–118. doi:10.1016/j.cell.2015.02.020

Frolkis, A., Knox, C., Lim, E., Jewison, T., Law, V., Hau, D. D., et al. (2010). SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Research*, *38*(Database), D480–D487. doi:10.1093/nar/gkp1002

Fundación Española de la Nutrición. (2017). Informe sobre Legumbres, Nutrición y Salud. http://fen.org.es/index.php/actividades/publicacion/informe-sobre-legumbres-nutricion-y-salud. Accessed 6 April 2017

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Jáuregui, O., Corella, D., Ruiz-Canela, M., et al. (2015). A metabolomics-driven approach to predict cocoa product consumption by designing a multimetabolite biomarker model in free-living subjects from the PREDIMED

study. *Molecular Nutrition and Food Research*, *59*(2), 212–220. doi:10.1002/mnfr.201400434

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Tulipani, S., Estruch, R., Martínez-González, M. a., et al. (2014a). Novel multimetabolite prediction of walnut consumption by a urinary biomarker model in a free-living population: The predimed study. *Journal of Proteome Research*, *13*(7), 3476–3483. doi:10.1021/pr500425r

Garcia Aloy, M. (2014b). *Biomarcadors d'exposició dietètica en estudis nutricionals d'intervenció i observacionals. Aplicació d'una aproximació metabolòmica HPLC-q-ToF-MS per a la millora de la capacitat predictiva a través de models combinats multi-metabòlit.* [Tesis doctoral]. Universitat de Barcelona.

Garcia-Aloy, M., Llorach, R., Urpi-Sarda, M., Tulipani, S., Salas-Salvadó, J., Martínez-González, M. A., et al. (2014c). Nutrimetabolomics fingerprinting to identify biomarkers of bread exposure in a free-living population from the PREDIMED study cohort. *Metabolomics*, *11*(1), 155–165. doi:10.1007/s11306-014-0682-6

Garcia-Aloy, M., Rabassa, M., Casas-Agustench, P., Hidalgo-Liberona, N., Llorach, R., & Andres-Lacueva, C. (2017). Novel strategies for improving dietary exposure assessment: Multiple-data fusion is a more accurate measure than the traditional single-biomarker approach. *Trends in Food Science & Technology*. doi:10.1016/j.tifs.2017.04.013

Garcia-Perez, I., Posma, J. M., Chambers, E. S., Nicholson, J. K., C. Mathers, J., Beckmann, M., et al. (2016). An Analytical Pipeline for Quantitative Characterization of Dietary Intake: Application To Assess Grape Intake. *Journal of Agricultural and Food Chemistry*, *64*(11), 2423–2431. doi:10.1021/acs.jafc.5b05878

Gibney, M. J., Walsh, M., Brennan, L., Roche, H. M., German, B., & van Ommen, B. (2005). Metabolomics in human nutrition: opportunities and challenges. *The American journal of clinical nutrition*, *82*(3), 497–503. http://www.ncbi.nlm.nih.gov/pubmed/16155259. Accessed 9 March 2017

Gika, H. G., Theodoridis, G. A., & Wilson, I. D. (2008). Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry fingerprinting of human urine. *Journal of Chromatography A*, *1189*(1–2), 314–322. doi:10.1016/j.chroma.2007.10.066

Gonthier, M. P., Remesy, C., Scalbert, A., Cheynier, V., Souquet, J. M., Poutanen, K., & Aura, A. M. (2006). Microbial metabolism of caffeic acid and its esters chlorogenic and caftaric acids by human faecal microbiota in vitro. *Biomedicine and Pharmacotherapy*, *60*(9), 536–540. doi:10.1016/j.biopha.2006.07.084

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology*, *22*(5), 245–252. doi:10.1016/j.tibtech.2004.03.007

Guertin, K. A., Loftfield, E., Boca, S. M., Sampson, J. N., Moore, S. C., Xiao, Q., et al. (2015). Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *The American journal of clinical nutrition*, *101*(5), 1000–11. doi:10.3945/ajcn.114.096099

Guertin, K. A., Moore, S. C., Sampson, J. N., Huang, W.-Y., Xiao, Q., Stolzenberg-Solomon, R. Z., et al. (2014). Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr*, *100*, 208–17. doi:10.3945/ajcn.113.078758

Heinzmann, S. S., Brown, I. J., Chan, Q., Bictash, M., Dumas, M.-E., Kochhar, S., et al. (2010). Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *American Journal of Clinical Nutrition*, *92*(2), 436–443. doi:10.3945/ajcn.2010.29672

Heinzmann, S. S., Holmes, E., Kochhar, S., Nicholson, J. K., & Schmitt-Kopplin, P. (2015). 2-Furoylglycine as a Candidate Biomarker of Coffee Consumption. *Journal of Agricultural and Food Chemistry*, *63*(38), 8615–8621. doi:10.1021/acs.jafc.5b03040

Hertkorn, N., Ruecker, C., Meringer, M., Gugisch, R., Frommberger, M., Perdue, E. M., et al. (2007). High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Analytical and Bioanalytical Chemistry*, *389*(5), 1311–1327. doi:10.1007/s00216-007-1577-4

Holt, S., Carter, D., Tothill, P., Heading, R., & Prescott, L. (1979). Effect of gel fibre on gastric emptying and absorption of glucose and paracetamol. *The Lancet*, *313*(8117), 636–639. doi:10.1016/S0140-6736(79)91079-1

Hutson, S. M., Sweatt, A. J., & Lanoue, K. F. (2005). Branched-chain amino acid metabolism: implications for establishing safe intakes. *J Nutr*, *135*(6 Suppl), 1557S–64S. doi:135/6/1557S [pii]

Illner, A.-K., Freisling, H., Boeing, H., Huybrechts, I., Crispim, S., & Slimani, N. (2012). Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *International Journal of Epidemiology*, *41*(4), 1187–1203. doi:10.1093/ije/dys105

Johnston, K. L., Clifford, M. N., & Morgan, L. M. (2003). Coffee acutely modifies gastrointestinal hormone secretion and glucose tolerance in humans: Glycemic effects of chlorogenic acid and caffeine. *American Journal of Clinical Nutrition*, *78*(4), 728–733. doi:10.1021/jf980737w

JPI. A healthy diet for a healthy life. (2017). http://www.healthydietforhealthylife.eu/. Accessed 10 March 2017

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. doi:10.1093/nar/gkw1092

Kell, D. B., Brown, M., Davey, H. M., Dunn, W. B., Spasic, I., & Oliver, S. G. (2005). Metabolic footprinting and systems biology: the medium is the message. *Nature Reviews Microbiology*, *3*(7), 557–565. doi:10.1038/nrmicro1177

Kim, S. J., de Souza, R. J., Choo, V. L., Ha, V., Cozma, A. I., Chiavaroli, L., et al. (2016). Effects of dietary pulse consumption on body weight: a systematic review and meta-analysis of randomized controlled trials. *American Journal of Clinical Nutrition*, *103*(5), 1213–1223. doi:10.3945/ajcn.115.124677

Kristensen, M., Engelsen, S. B., & Dragsted, L. O. (2012). LC–MS metabolomics top-down

approach reveals new exposure and effect biomarkers of apple and apple-pectin intake. *Metabolomics*, *8*(1), 64–73. doi:10.1007/s11306-011-0282-7

Kuhnle, G. G. (2012). Nutritional biomarkers for objective dietary assessment. *Journal of the Science of Food and Agriculture*, *92*(6), 1145–1149. doi:10.1002/jsfa.5631

La Vecchia, C. (2005). Coffee, liver enzymes, cirrhosis and liver cancer. *Journal of Hepatology*, *42*(4), 444–446. doi:10.1016/j.jhep.2005.01.004

Lay, J. O., Liyanage, R., Borgmann, S., & Wilkins, C. L. (2006). Problems with the "omics." *TrAC Trends in Analytical Chemistry*, *25*(11), 1046–1056. doi:10.1016/j.trac.2006.10.007

Lenz, E. M., & Wilson, I. D. (2007). Analytical Strategies in Metabonomics. *Journal of Proteome Research*, *6*(2), 443–458. doi:10.1021/pr0605217

Llorach, R., Garcia-Aloy, M., Tulipani, S., Vazquez-Fresno, R., & Andres-Lacueva, C. (2012). Nutrimetabolomic strategies to develop new biomarkers of intake and health effects. In *Journal of Agricultural and Food Chemistry* (Vol. 60, pp. 8797–8808). doi:10.1021/jf301142b

Loader, T. B., Taylor, C. G., Zahradka, P., & Jones, P. J. H. (2017). Chlorogenic acid from coffee beans: evaluating the evidence for a blood pressure–regulating health claim. *Nutrition Reviews*, nuw057. doi:10.1093/nutrit/nuw057

Lopez-Legarrea, P., Fuller, N. R., Zulet, M. A., Martinez, J. A., & Caterson, I. D. (2014). The influence of Mediterranean, carbohydrate and high protein diets on gut microbiota composition in the treatment of obesity and associated inflammatory state. *Asia Pacific journal of clinical nutrition*, *23*(3), 360–8.

MacWilliam, S., Wismer, M., & Kulshreshtha, S. (2014). Life cycle and economic assessment of Western Canadian pulse systems: The inclusion of pulses in crop rotations. *Agricultural Systems*, *123*, 43–53. doi:10.1016/j.agsy.2013.08.009

Madrid-Gambin, F., Llorach, R., Vázquez-Fresno, R., Urpi-Sarda, M., Almanza-Aguilera, E., Garcia-Aloy, M., et al. (2017). Urinary ^{1}H Nuclear Magnetic Resonance Metabolomic Fingerprinting Reveals Biomarkers of Pulse Consumption Related to Energy-Metabolism Modulation in a Subcohort from the PREDIMED study. *Journal of Proteome Research*, *16*(4), 1483–1491. doi:10.1021/acs.jproteome.6b00860

Manach, C. (2013). The Food metabolome. https://www.slideshare.net/cmanach/diapo-glasgowpdf-24420507. Accessed 26 May 2017

Manach, C., Hubert, J., Llorach, R., & Scalbert, A. (2009). The complex links between dietary phytochemicals and human health deciphered by metabolomics. *Molecular Nutrition & Food Research*, *53*(10), 1303–1315. doi:10.1002/mnfr.200800516

Manach, C., Milenkovic, D., Van de Wiele, T., Rodriguez-Mateos, A., de Roos, B., Garcia-Conesa, M. T., et al. (2017). Addressing the inter-individual variation in response to consumption of plant food bioactives: Towards a better understanding of their role in healthy aging and cardiometabolic risk reduction. *Molecular Nutrition & Food Research*, *61*(6), 1600557. doi:10.1002/mnfr.201600557

Marlow, G., Ellett, S., Ferguson, I. R., Zhu, S., Karunasinghe, N., Jesuthasan, A. C., et al. (2013).

Transcriptomics to study the effect of a Mediterranean-inspired diet on inflammation in Crohn's disease patients. *Human genomics*, *7*(1), 24. doi:10.1186/1479-7364-7-24

Ministerio de Agricultura Medioambiente y Alimentación. Panel de consumo alimentario. (2016). Informe del consumo de alimentación en España 2015. http://www.mapama.gob.es/es/alimentacion/temas/consumo-y-comercializacion-y-distribucion-alimentaria/informeconsumoalimentacion2015_tcm7-422694.pdf. Accessed 7 April 2017

Mithril, C., Dragsted, L. O., Meyer, C., Tetens, I., Biltoft-Jensen, A., & Astrup, A. (2013). Dietary composition and nutrient content of the New Nordic Diet. *Public health nutrition*, *16*(5), 777–85. doi:10.1017/S1368980012004521

Moco, S., Martin, F.-P. J., & Rezzi, S. (2012). Metabolomics View on Gut Microbiome Modulation by Polyphenol-rich Foods. *Journal of Proteome Research*, *11*(10), 4781–4790. doi:10.1021/pr300581s

Moon, J.-K., & Shibamoto, T. (2009). Role of roasting conditions in the profile of volatile flavor chemicals formed from coffee beans. *Journal of agricultural and food chemistry*, *57*(13), 5823–31.

Moreiras, O., Carbajal, A., Cabrera, L., & Cuadrado, C. (2013). *Tablas de composición de alimentos : guía de prácticas* (16º.). Madrid: Pirámide.

Nagana Gowda, G. A., Gowda, Y. N., & Raftery, D. (2015a). Expanding the limits of human blood metabolite quantitation using NMR spectroscopy. *Analytical chemistry*, *87*(1), 706–15. doi:10.1021/ac503651e

Nagana Gowda, G. A., & Raftery, D. (2015b). Can NMR solve some significant challenges in metabolomics? *Journal of magnetic resonance (San Diego, Calif. : 1997)*, *260*, 144–60. doi:10.1016/j.jmr.2015.07.014

Newsholme, P., Procopio, J., Ramos Lima, M. M., Pithon-Curi, T. C., & Curi, R. (2003, March). Glutamine and glutamate - Their central role in cell metabolism and function. *Cell Biochemistry and Function*. John Wiley & Sons, Ltd. doi:10.1002/cbf.1003

Nordestgaard, a. T., Thomsen, M., & Nordestgaard, B. G. (2015). Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: a Mendelian randomization study. *International Journal of Epidemiology*, *1*, 1–15. doi:10.1093/ije/dyv083

Cloarec, O., Dumas, M. E., Craig, A., Barton, R. H., Trygg, J., Hudson, J., et al. (2005). Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Analytical Chemistry*, *77*(5), 1282–1289. doi:10.1021/ac048630x

OMS. (1993). Biomarkers and risk assessment: concepts and principles. *Organización Mundial de la Salud. ISBN: 9241571551*, *154*(106). http://apps.who.int/iris/bitstream/10665/39037/1/9241571551-eng.pdf. Accessed 10 March 2017

Oresic, M., Wiest, M. M., Watkins, S. M., Teegarden, D., Zemel, M. B., German, J. B., & al., et. (2009). Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, metabolism, and cardiovascular diseases : NMCD*, *19*(11), 816–24.

doi:10.1016/j.numecd.2009.04.018

Ozen, A. E., Pons, A., & Tur, J. A. (2012). Worldwide consumption of functional foods: a systematic review. *Nutrition Reviews*, *70*(8), 472–481. doi:10.1111/j.1753-4887.2012.00492.x

Pan, M.-H., Tung, Y.-C., Yang, G., Li, S., & Ho, C.-T. (2016). Molecular mechanisms of the anti-obesity effect of bioactive compounds in tea and coffee. *Food & function*, *7*(11), 4481–4491. doi:10.1039/c6fo01168c

Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology, Published online: 22 March 2012; | doi:10.1038/nrm3314*, *13*(4), 263. doi:10.1038/nrm3314

Perera, T., Young, M. R., Zhang, Z., Murphy, G., Colburn, N. H., Lanza, E., et al. (2015). Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Molecular nutrition & food research*, *59*(4), 795–806. doi:10.1002/mnfr.201400847

Potischman, N. (2003). Biologic and Methodologic Issues for Nutritional Biomarkers. *J. Nutr.*, *133*(3), 875S–880. doi:10.1091/mbc.E05

Potischman, N., & Freudenheim, J. L. (2003). Biomarkers of nutritional exposure and nutritional status: an overview. *The Journal of nutrition*, (3), 873S–874S.

Primrose, S., Draper, J., Elsom, R., Kirkpatrick, V., Mathers, J. C., Seal, C., et al. (2011). Metabolomics and human nutrition. *British Journal of Nutrition*, *105*(8), 1277–1283. doi:10.1017/S0007114510004812

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., et al. (2011). The human serum metabolome. *PloS one*, *6*(2), e16957. doi:10.1371/journal.pone.0016957

Pujos-Guillot, E., Hubert, J., Martin, J. F., Lyan, B., Quintana, M., Claude, S., et al. (2013). Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: Citrus fruit as a case study. *Journal of Proteome Research*, *12*(4), 1645–1659. doi:10.1021/pr300997c

Ramalingam, A., Kudapa, H., Pazhamala, L. T., Weckwerth, W., & Varshney, R. K. (2015). Proteomics and Metabolomics: Two Emerging Areas for Legume Improvement. *Frontiers in plant science*, *6*(December), 1116. doi:10.3389/fpls.2015.01116

Rechner, A. R., Kuhnle, G., Bremner, P., Hubbard, G. P., Moore, K. P., & Rice-Evans, C. A. (2002). The metabolic fate of dietary polyphenols in humans. *Free Radical Biology and Medicine*, *33*(2), 220–235. doi:10.1016/S0891-5849(02)00877-8

Rezzi, S., Ramadan, Z., Fay, L. B., & Kochhar, S. (2007). Nutritional metabonomics: Applications and perspectives. *Journal of Proteome Research*. American Chemical Society. doi:10.1021/pr060522z

Rezzi, S., Ramadan, Z., Martin, F. P. J., Fay, L. B., Van Bladeren, P., Lindon, J. C., et al. (2007). Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals. *Journal of Proteome Research*, *6*(11), 4469–4477. doi:10.1021/pr070431h

Rotches-Ribalta, M., Andres-Lacueva, C., Estruch, R., Escribano, E., & Urpi-Sarda, M. (2012). Pharmacokinetics of resveratrol metabolic profile in healthy humans after moderate consumption of red wine and grape extract tablets. *Pharmacological Research*, *66*(5), 375–382. doi:10.1016/j.phrs.2012.08.001

Rothwell, J. A., Fillâtre, Y., Martin, J.-F., Lyan, B., Pujos-Guillot, E., Fezeu, L., et al. (2014). New biomarkers of coffee consumption identified by the non-targeted metabolomic profiling of cohort study subjects. *PloS one*, *9*(4), e93474. doi:10.1371/journal.pone.0093474

Roy, F., Boye, J. I., & Simpson, B. K. (2010). Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. *Food Research International*, *43*(2), 432–442. doi:10.1016/j.foodres.2009.09.002

Ruan, Z., Yang, Y., Zhou, Y., Wen, Y., Ding, S., Liu, G., Wu, X., et al. (2014). Metabolomic analysis of amino acid and energy metabolism in rats supplemented with chlorogenic acid. doi:10.1007/s00726-014-1762-7

Sääksjärvi, K., Knekt, P., Rissanen, H., Laaksonen, M. a, Reunanen, a, & Männistö, S. (2008). Prospective study of coffee consumption and risk of Parkinson's disease. *European journal of clinical nutrition*, *62*(7), 908–915. doi:10.1038/sj.ejcn.1602788

Salas-Salvadó, J., Bulló, M., Babio, N., Martínez-González, M. Á., Ibarrola-Jurado, N., Basora, J., et al. (2011). Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes care*, *34*(1), 14–9. doi:10.2337/dc10-1288

Satija, A., Yu, E., Willett, W. C., & Hu, F. B. (2015). Understanding Nutritional Epidemiology and Its Role in Policy. *Advances in Nutrition: An International Review Journal*, *6*(1), 5–18. doi:10.3945/an.114.007492

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L. O., Draper, J., et al. (2014). The food metabolome: a window over dietary exposure. *American Journal of Clinical Nutrition*, *99*(6), 1286–1308. doi:10.3945/ajcn.113.076133

Shrivastava, A., & Gupta, V. (2011). Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles of Young Scientists*, *2*(1), 21–21.

Slupsky, C. M., Rankin, K. N., Wagner, J., Fu, H., Chang, D., Weljie, A. M., et al. (2007). Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. *Anal Chem*, *79*(18), 6995–7004. doi:10.1021/ac0708588

Solanky, K. S., Bailey, N. J., Beckwith-Hall, B. M., Bingham, S., Davis, A., Holmes, E., et al. (2005). Biofluid 1H NMR-based metabonomic techniques in nutrition research — metabolic effects of dietary isoflavones in humans. *The Journal of Nutritional Biochemistry*, *16*(4), 236–244. doi:10.1016/j.jnutbio.2004.12.005

Stella, C., Beckwith-Hall, B., Cloarec, O., Holmes, E., Lindon, J. C., Powell, J., et al. (2006). Susceptibility of human metabolic phenotypes to dietary modulation. *Journal of Proteome Research*, *5*(10), 2780–2788. doi:10.1021/pr060265y

Tharanathan, R. ., & Mahadevamma, S. (2003). Grain legumes—a boon to human nutrition. *Trends in Food Science & Technology*, *14*(12), 507–518. doi:10.1016/j.tifs.2003.07.002

Trespalacios, M. del P., Tulipani, S., Garcia-Aloy, M., Zamora-Ros, R., Rotchés-Ribalta, M., Rabassa, M., et al. (2012). Nutritional Biomarkers: Applications and Challenges. In J. L. Morrison (Ed.), *Food intake : regulation, assessing and controlling.* Nova Science Publishers.

Trygg, J., Holmes, E., & Lundstedt, T. (2007). Chemometrics in Metabonomics. *Journal of Proteome Research*, *6*(2), 469–479. doi:10.1021/pr060594q

Tulipani, S., Llorach, R., Jáuregui, O., López-Uriarte, P., Garcia-Aloy, M., Bullo, M., et al. (2011). Metabolomics Unveils Urinary Changes in Subjects with Metabolic Syndrome following 12-Week Nut Consumption. *Journal of Proteome Research*, *10*(11), 5047–5058. doi:10.1021/pr200514h

Tulipani, S., Llorach, R., Urpi-Sarda, M., & Andres-Lacueva, C. (2013). Comparative Analysis of Sample Preparation Methods To Handle the Complexity of the Blood Fluid Metabolome: When Less Is More. *Analytical Chemistry*, *85*(1), 341–348. doi:10.1021/ac302919t

U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2015). 2015–2020 Dietary Guidelines for Americans. *8th Edition*. http://health.gov/dietaryguidelines/2015/. Accessed 9 May 2016

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, *7*(1), 142. doi:10.1186/1471-2164-7-142

van Ravenzwaay, B., Cunha, G. C.-P., Leibold, E., Looser, R., Mellert, W., Prokoudine, A., et al. (2007). The use of metabolomics for the discovery of new biomarkers of effect. *Toxicology Letters*, *172*(1), 21–28. doi:10.1016/j.toxlet.2007.05.021

Vázquez Fresno, R. (2015a). *Estudio metabolómico de biomarcadores nutricionales en estudios de intervención mediante resonancia magnética nuclear. Intervención con alimentos, complementos y patrones alimentarios.* [Tesis doctoral]. Universitat de Barcelona.

Vázquez-Fresno, R., Llorach, R., Urpi-Sarda, M., Khymenets, O., Bullo, M., Corella, D., et al. (2015b). An NMR metabolomics approach reveals a combined-biomarkers model in a wine interventional trial with validation in free-living individuals of the PREDIMED study. *Metabolomics*, *11*(4), 797–806. doi:10.1007/s11306-014-0735-x

Vázquez-Fresno, R., Llorach, R., Urpi-Sarda, M., Lupianez-Barbero, A., Estruch, R., Corella, D., et al. (2015c). Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: A 1- and 3-year follow-up in the PREDIMED study. *Journal of Proteome Research*, *14*(1), 531–540. doi:10.1021/pr5007894

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., et al. (2010). Global metabolic profiling procedures for urine using UPLC–MS. *Nature Protocols*, *5*(6), 1005–1018. doi:10.1038/nprot.2010.50

Wishart, D. S. (2008a). Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry*, *27*(3), 228–237. doi:10.1016/j.trac.2007.12.001

Wishart, D. S. (2008b). Metabolomics: applications to food science and nutrition research. *Trends in Food Science & Technology*, *19*(9), 482–493. doi:10.1016/j.tifs.2008.03.003

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., et al. (2013). HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*, *41*(D1), D801–D807. doi:10.1093/nar/gks1065

Wu, H., Southam, A. D., Hines, A., & Viant, M. R. (2008). High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Analytical Biochemistry*, *372*(2), 204–212. doi:10.1016/j.ab.2007.10.002

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research*, *43*(W1), W251–W257. doi:10.1093/nar/gkv380

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*, *914*, 17–34. doi:10.1016/j.aca.2016.02.001

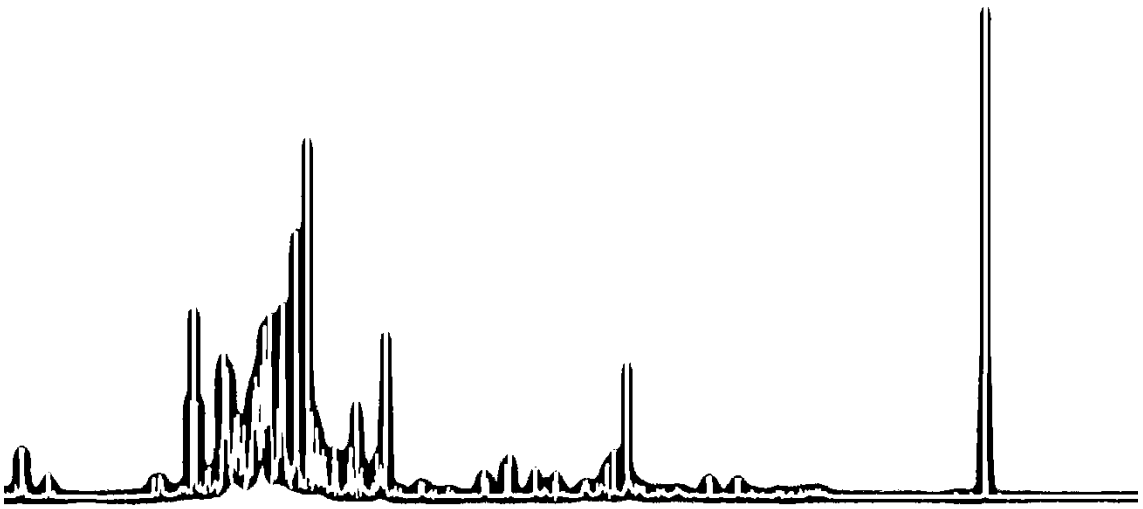
Zhang, A. ., Mitchell, S. ., & Smith, R. . (1999). Dietary Precursors of Trimethylamine in Man: A Pilot Study. *Food and Chemical Toxicology*, *37*(5), 515–520. doi:10.1016/S0278-6915(99)00028-9

Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *The Analyst*, *137*(2), 293–300. doi:10.1039/C1AN15605E

Zhang, X., Yap, Y., Wei, D., Chen, G., & Chen, F. (2008). Novel omics technologies in nutrition research. *Biotechnology Advances*, *26*(2), 169–176. doi:10.1016/j.biotechadv.2007.11.002

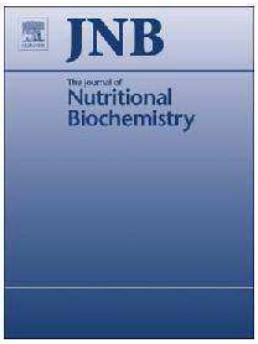
Zhao, X., Zeisel, S. H., & Zhang, S. (2015). Rapid LC-MRM-MS assay for simultaneous quantification of choline, betaine, trimethylamine, trimethylamine ***N*** -oxide, and creatinine in human plasma and urine. *Electrophoresis*, *36*(18), 2207–2214. doi:10.1002/elps.201500055

Anexos



Otras publicaciones en revistas

Publicación I: Almanza-Aguilera, E.; Urpi-Sarda, M.; Llorach, R.; Vázquez-Fresno, R.; Garcia-Aloy, M.; Carmona, F.; Sanchez, A.; Madrid-Gambin, F.; Estruch, R.; Corella, D.; Andres-Lacueva, C. Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a ^{1}H-NMR-based untargeted metabolomics approach. *The Journal of Nutritional Biochemistry.* [*En prensa*].
doi: doi.org/10.1016/j.jnutbio.2017.06.001



Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a ^{1}H–NMR-based untargeted metabolomics approach

Enrique Almanza-Aguilera, Mireia Urpi-Sarda, Rafael Llorach, Rosa Vázquez-Fresno, Mar Garcia-Aloy, Francesc Carmona, Alex Sanchez, Francisco Madrid-Gambin, Ramon Estruch, Dolores Corella, Cristina Andres-Lacueva

PII: S0955-2863(16)30793-8
DOI: doi: 10.1016/j.jnutbio.2017.06.001
Reference: JNB 7789

To appear in: *The Journal of Nutritional Biochemistry*

Received date: 30 December 2016
Revised date: 9 May 2017
Accepted date: 1 June 2017

Please cite this article as: Almanza-Aguilera Enrique, Urpi-Sarda Mireia, Llorach Rafael, Vázquez-Fresno Rosa, Garcia-Aloy Mar, Carmona Francesc, Sanchez Alex, Madrid-Gambin Francisco, Estruch Ramon, Corella Dolores, Andres-Lacueva Cristina, Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a ^{1}H–NMR-based untargeted metabolomics approach, *The Journal of Nutritional Biochemistry* (2017), doi: 10.1016/j.jnutbio.2017.06.001

Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a ^{1}H-NMR-based untargeted metabolomics approach

Enrique Almanza-Aguilera[a,], Mireia Urpi-Sarda[a,b]*, Rafael Llorach[a,b], Rosa Vázquez-Fresno[a], Mar Garcia-Aloy[a,b], Francesc Carmona[c], Alex Sanchez[c], Francisco Madrid-Gambin[a,b,], Ramon Estruch[d,e], Dolores Corella[d,f], Cristina Andres-Lacueva[a,b]*

[a]Biomarkers and Nutrimetabolomics Laboratory, Nutrition, Food Science and Gastronomy Department, XaRTA, INSA, Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Barcelona 08028, Spain; [b]CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Instituto de Salud Carlos III, Madrid 28028, Spain; [c]Department of Statistics, University of Barcelona, Barcelona, Spain, [d] CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid 28028, Spain; [e]Department of Internal Medicine, Institut d'Investigacions Biomediques August Pi Sunyer (IDIBAPS), Hospital Clinic, University of Barcelona, Barcelona, 08036, Spain; [f]Department of Preventive Medicine, University of Valencia, Valencia 46010, Spain.

Corresponding authors: Dr. Mireia Urpi-Sarda and Dr. Cristina Andres-Lacueva, Biomarkers and Nutrimetabolomics Laboratory, Nutrition, Food Science and Gastronomy Department, Food Campus Torribera, Faculty of Pharmacy and Food Science, University of Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain. E-mail addresses: murpi@ub.edu (M. Urpi) and candres@ub.edu (C. Andres). Phone: +34.934034844, +34.934034840.

Short running title: Metabolomic profile of high adherence to a Mediterranean dietary pattern

This work was supported by Spanish National Grants from the Ministry of Economy and Competitiveness (MINECO), and co-funded by FEDER (Fondo Europeo de Desarrollo Regional): the EU Joint Programming Initiative 'A Healthy Diet for a Healthy Life' on Biomarkers BioNHFOODBALL (PCIN-2014-133-MINECO-Spain), AGL2009-13906-C02-01, CIBERFES, Associació Catalana de Diabetis (ACD2015) and the award of 2014SGR1566 from the Generalitat de Catalunya's Agency AGAUR.

Keywords: microbiota, biomarkers, metabolomics, dietary patterns, Mediterranean diet, high adherence

Abbreviations: AUC, area under the curve; CVD, cardiovascular disease; FFQ, food frequency questionnaire; ^{1}H-NMR, proton nuclear magnetic resonance; H-MDA, high Mediterranean diet adherence; HPHPA, 3-(3-hydroxyphenyl)-3-hydroxypropanoate; L-MDA, low Mediterranean diet adherence; MedDiet, Mediterranean diet; MetS, metabolic syndrome; PREDIMED, Prevención con Dieta Mediterránea; OSC-PLS-DA, orthogonal signal correction – partial least-squares discriminant analysis; PAGN, phenylacetylglutamine; ROC, receiver operating characteristic; VIP, variable importance in projection.

Abstract

The study of biomarkers of dietary patterns including the Mediterranean diet (MedDiet) is scarce and could improve the assessment of these patterns. Moreover, it could provide a better understanding of health benefits of dietary patterns in nutritional epidemiology.

We aimed to determine a robust and accurate biomarker associated with a high adherence to a MedDiet pattern that included dietary assessment and its biological effect.

In this cross-sectional study, we included 56 and 63 individuals with high (H-MDA) and low (L-MDA) MedDiet adherence categories, respectively, all from the PREDIMED (Prevención con Dieta Mediterránea) trial. An ^{1}H-NMR-based untargeted metabolomics approach was applied to urine samples. Multivariate statistical analyses were conducted to determine the metabolite differences between groups. A stepwise logistic regression and receiver operating characteristic curves were used to build and evaluate the prediction model for H-MDA.

Thirty-four metabolites were identified as discriminant between H-MDA and L-MDA.

The fingerprint associated with H-MDA included higher excretion of proline betaine and phenylacetylglutamine, among others, and decreased amounts of metabolites related to glucose metabolism. Three microbial metabolites – phenylacetylglutamine, *p*-cresol and 4-hydroxyphenylacetate – were included in the prediction model of H-MDA (95% specificity, 95% sensitivity and 97% area under the curve).

The model composed of microbial metabolites was the biomarker that defined high adherence to a Mediterranean dietary pattern. The overall metabolite profiling identified reflects the metabolic modulation produced by H-MDA. The proposed biomarker may be a better tool for assessing and aiding nutritional epidemiology in future associations between H-MDA and the prevention or amelioration of chronic diseases.

Keywords: microbiota, biomarkers, metabolomics, dietary pattern, Mediterranean diet, high adherence

1. Introduction

High adherence to healthy dietary patterns is associated with lower risk of chronic diseases [1]. The measurement of dietary intake is an essential component in studies attempting to establish links between dietary exposure and health outcomes [1]. Currently, the quality and adherence of dietary patterns in nutritional epidemiology are measured by the use of self-reported questionnaires such as the dietary indexes/scores [2]. A healthy dietary score is based on dietary recommendations as a result of scientific consensus or as proposed by investigators using an evidence-based approach [3]. Several dietary scores have been developed and applied to populations to evaluate the role of diet in a more holistic perspective in the risk of mortality, cardiovascular disease (CVD), and cancer [4]. However, because dietary patterns are complex not only in terms of composition but also in terms of amounts and frequency of food intake, it is well recognized that in addition to the conventional methods, the emergence of novel biomarkers of food exposure may help to improve the accuracy of the assessment of compliance and adherence [5]. In this regard, metabolomics has emerged as a valuable tool in nutrition research for the discovery of novel dietary biomarkers for both single foods [6,7] and food patterns [5,8] and in addition, is able to evaluate their effect in the organism [9]. The use of these biomarkers is a more specific tool and complementary to traditional indexes/scores. Moreover, although progress has been made in the metabolite characterization of dietary patterns, most of the metabolomic studies have applied approaches based on multivariate analyses of food and nutritional data intake [10,11], leaving the complementary use of dietary scores and metabolomics as a new field to explore in dietary patterns characterization.

Up to now, several Mediterranean diet (MedDiet) adherence scores have been described in the literature, such as the 9-item MedDiet score [12] or the PREDIMED 14-item MedDiet score [13]. These scores were developed with the aim of appraising the adherence to a traditional

MedDiet of several populations as well as to evaluate the effect of adherence to a MedDiet on microbiota composition [14], CVD risk factors [15], aging diseases [16] or total mortality [12]. The application of targeted and untargeted metabolomics approaches in the study of the effects of the protective mechanisms of a MedDiet on CVD has been poorly studied but is now beginning to attract more interest [9,17]. Furthermore, currently there are some reports about metabolic profiling in biological samples (feces or urine) that enable the characterization of high adherence to a MedDiet pattern [14,18], but to our knowledge there are no reports calculating a prediction model of MedDiet adherence. In addition, the study of biomarkers to explain the assessment of the pattern and gain a better understanding of its health benefits in nutritional epidemiology is limited. The characterization of dietary patterns by using metabolomic approaches is important because it would allow insights into the relationship between diet, taking into account the bioavailability of MedDiet bioactives, and the risk of chronic diseases [14,19].

In the current study, we aimed to determine a robust and accurate urinary biomarker associated with a high adherence to MedDiet pattern that included dietary assessment and its biological effect on the organism by using a ^{1}H-NMR-based untargeted metabolomics approach which can be usefully applied in nutritional epidemiology.

2. Subjects and Methods

2.1. Study population and study design

The PREDIMED study is a multicentre, randomized, parallel and controlled clinical trial conducted in Spain and aimed at assessing the effects of a MedDiet on primary prevention of CVD. Full details of the design and methods have been published elsewhere [20,21]. Briefly, the study population included men (55–80 y) and women (60–80 y) without a previous history of CVD at enrolment but with either type 2 diabetes mellitus or at least three or more of the following CVD risk factors: current smoking, hypertension, high LDL cholesterol, low HDL cholesterol, overweight/obesity or family history of premature CVD. Exclusion criteria were the presence of any severe chronic illness, alcohol or drug abuse, BMI (in kg/m^2) $\geq$40, and allergy or intolerance to olive oil or nuts. The trial was registered at http://www.controlled-trials.com (ISRCTN35739639).

For the current work, we conducted a cross-sectional study with baseline dietary data and urine samples of 119 individuals recruited in two PREDIMED trial centres (Hospital Clinic of Barcelona, and University of Valencia). At baseline, one morning urine sample was collected from all participants and immediately aliquoted and stored at -80 ºC until the day of analysis.

2.2. Assessment of Mediterranean Diet adherence and other parameters

In order to appraise the adherence to a MedDiet among participants, the validated 14-item PREDIMED MedDiet score questionnaire was administered. In detail, the MedDiet score questionnaire consists of 12 questions on food consumption frequency and two questions on food intake habits considered characteristic of the Spanish MedDiet [13]. Each item/question is scored as 1 or 0, according to whether it is met or not, respectively (Supplementary Table 1). Thus, the total MedDiet score ranges from 0 to 14 points, meaning that the higher the score, the higher the adherence to a MedDiet. The MedDiet adherence score was calculated

for all participants and used for their subsequent stratification, which was done by using the proposed cut-off values previously reported in the PREDIMED study. [15]: MedDiet score $\leq$ 7 indicated low MedDiet adherence (L-MDA) (n= 63) and MedDiet score $\geq$ 10 (n=59) indicated high MDA (H-MDA).

All participants were also asked to complete a validated semi-quantitative 137-item food frequency questionnaire (FFQ) [22], and the Spanish version of the Minnesota Leisure-Time Physical Activity Questionnaire [23]. Trained dieticians in the PREDIMED study administered all questionnaires, including the PREDIMED MedDiet score questionnaire. The nutrient composition and energy intakes were calculated from the FFQ data by using Spanish food composition tables [24]. Also, anthropometrical measurements were taken directly by qualified nurses.

2.3. Metabolite profiling

^{1}H-NMR analysis and spectra processing were performed by following previous methodology [8]. Briefly, the urine samples were thawed, vortexed and centrifuged at 13,200 rpm for 5 min. From each supernatant, a volume of 300 µl was taken and diluted with 200 µl of H_2O/D_2O and mixed with a buffer solution [8]. The optimized pH of the buffer was set at 7.0, with a potassium deuteroxide solution, to minimize variations in the chemical shifts of the ^{1}H-NMR resonances. This mixture was transferred to a 5 mm NMR tube. ^{1}H-NMR spectra acquisition was performed using a Varian-Inova-500 MHz NMR spectrometer with pre-saturation of the water resonance using a NOESYPRESAT pulse sequence. The spectra data processed were binned to 1165 variables with bin widths of 0.005 ppm and integrated with ACD/NMR Processor 12.0 software (Advanced Chemistry Development, Toronto, Canada). The spectral region containing water (δ 4.75–5.00 ppm) was excluded before normalization to avoid spectral interference. Integrated spectra were row-wise normalized by sum using

MetaboAnalyst 3.0 (www.metaboanalyst.ca), a web server designed to permit comprehensive metabolomics data analysis.

Metabolites were identified following a multi-step procedure as previously reported [8]. This multi-step includes: 1) Comparison of the experimental NMR spectra with those in the library of Chenomx Suite 8.1 profiler software (Chenomx, Inc., Edmonton, Canada), which includes access to the Human Metabolome Database (HMDB) library [25]. Other databases such as the Biological Magnetic Resonance Data Bank and the Madison Metabolomics Consortium Database were also consulted along with existing NMR-based metabolomics literature. 2) At the same time, the NMR peak assignments were correlated using Pearson's correlation coefficient ($r \geq 0.7$, $p \leq 0.05$) to confirm the multiplicity and identify clusters of metabolites which were then compared to databases with NMR data [8,25]. Biological interpretation was made by consulting the HMDB and Kyoto Encyclopedia of Genes and Genomes databases.

2.4. Statistical analyses

Demographic characteristics, medication usage and dietary intake were compared between groups by conducting Student's *t*-tests and chi-square analyses for continuous and categorical variables, respectively. Variables with a non-normal distribution were log transformed before analyses. Multivariate data analysis was performed using SIMCA-P+13.0 (Umetrics, Umeå, Sweden) software. Data sets containing the integrated NMR spectral bins were log transformed and Pareto scaled before performing a principal component analysis to explore the quality of data acquisition. An orthogonal signal correction (OSC) filter was applied to the data sets in order to reduce the variability not associated with the dietary classification [8], in this case the category of adherence to a MedDiet. Afterwards, a partial least squares discriminant analysis (PLS-DA) was performed to examine the difference in metabolite profile between subjects at H-MDA and L-MDA. The quality and validation of resultant

model was also appraised; the first through the R^2Y (cum) and Q^2 (cum) parameters, and the second with a permutation test (n=200). Discriminant variables between groups were selected based on their variable importance in projection (VIP) values >1.0, which is a generally accepted threshold [26]. The normality of discriminant variables was assessed using a Kolmogorov–Smirnov test with Lilliefors significance correction, and additionally Mann–Whitney or independent Student's *t*-tests were performed, according to the normality of the data. To account for multiple comparisons in the metabolomic analysis we used a corrected P value with the Benjamini–Hochberg procedure.

Aiming to identify the metabolites with the best discriminant capability between H-MDA and L-MDA, we designed and assessed the performance of a prediction model with the H-MDA as dependent variable and the discriminant metabolites identified from the OSC-PLS-DA model as independent variables [6]. To this end, first the data set of individuals (n=119) was randomly split into two-thirds to build one training set (n=79), and one-third for a validation set (n=40). A stepwise binary logistic regression analysis was performed in the training set in order to identify the metabolites with the most significant predictive capacity, and from these a combined model was built. Subsequently, a receiver operating characteristic (ROC) curve analysis was performed, first in the training set and then in the validation set, to evaluate the obtained model, as well as that of the individual metabolites included. Then the performance of both the combined model and the individual metabolites was determined by the area under the curve (AUC) of the ROC curves, as well as by the sensitivity and specificity at the optimal cut-off point defined as the minimum distance to the top-left corner in the ROC curve. Finally, the association between the combined multi-metabolite model and individual metabolites with food groups was tested by a Spearman's rank correlation analysis with correction of *p* value using Benjamini–Hochberg procedure.

All univariate analyses, including normality, Student's *t*-tests, Spearman's rank correlation, logistic regression and ROC curve analyses, were performed on IBM SPSS 21® statistics software (IBM Corp., Armonk, NY, USA).

3. Results

3.1. Demographic and dietary intake measurements

For the current study, we included individuals who, according to their individual MedDiet score were assigned to L-MDA (≤7 points, n=63) or H-MDA (≥10 points, n=56). Our population had a mean (± SD) age of 67 ± 6 y, a mean (± SD) BMI of 30.3 ± 4.5 kg/m^2 and 68.9% of the participants were women (Table 1). Cardiovascular risk factors as well as medication use were similar between both groups (Table 1). With regard to food and nutrient intake (Table 2), individuals in H-MDA consumed higher amounts of olive oil, nuts, vegetables, fruits, legumes and fish, and total dietary fiber, than L-MDA participants (p<0.05). However, no statistically significant differences were found for total energy, total fat, carbohydrates and protein intakes.

3.2. Discriminant metabolite profile of high adherence to a Mediterranean diet pattern

The OSC-PLS-DA analysis resulted in one latent component model with R^2Y (cum) and Q^2Y (cum) values of 0.913 and 0.764, respectively, indicating a good ability to classify individuals according to their MedDiet adherence. In addition, a permutation test (n=200), with intercept R^2 and Q^2 values of 0.346 and -0.154, respectively, showed the validity of the model. Table 3 shows a list of 34 metabolites that were identified after the selection of discriminant variables from the OSC-PLS-DA (based on VIP>1.0 values). Individuals in H-MDA group had a marked excretion of metabolites involved in protein/amino acid metabolism characterized by a higher excretion of anserine, carnosine, creatine, creatinine, guanidoacetate, histidine and N-acetylglutamine, as well as a lower excretion of 3-methylhistidine, alanine, glycine and lysine, than those in L-MDA. Similarly, another major group of metabolites derived from gut microbiota was identified. Concretely, H-MDA was characterized by a higher urinary amount of 4-hydroxyhippurate, 4-hydroxyphenylacetate, dimethylsulfone, 3-(3-hydroxyphenyl)-3-hydroxypropanoate (HPHPA), *p*-cresol and phenylacetylglutamine (PAGN) and a lower

urinary amounts of 3-indoxyl sulfate, hippurate and isobutyrate, compared to the L-MDA. Furthermore, participants in the H-MDA group excreted lower levels of metabolites involved in the energy pathway, and in the propanoate and purine and caffeine pathways, than those in L-MDA. In addition, participants in the H-MDA group had higher levels of metabolites involved in the choline pathway (except betaine), as well as the inositol, niconitate, nicotidamide and pyrimidine pathways (Table 3).

3.3. Prediction model for high adherence to Mediterranean dietary pattern

We performed a model based on a stepwise binary logistic regression analysis including the previous 34 metabolites identified from the OSC-PLS-DA model and further ROC curve analyses to evaluate the resulting model and the individual metabolites included in this. To this end, the data set of individuals from the H-MDA and L-MDA groups was divided into training and validation sets, as indicated above. The resulting model included three metabolites derived from gut microbiota, namely PAGN, *p*-cresol and 4-hydroxyphenylacetate (Supplementary Table 2). The specificity and sensitivity of the model were higher than 90% in both the training and validation sets. The PAGN had 85.7% and 68.4% of specificity and sensitivity, respectively in validation set, while *p*-cresol and 4-hydroxyphenylacetate showed values between 66% and 84% in these parameters (Table 4) for the validation set. Furthermore, the global performance of the combined model in terms of AUC was 97.7% for the training and 97.0% for the validation set, while individually each metabolite had values of AUC between 59% and 86% (Table 4). Figure 1 illustrates that the model improves the classification of MedDiet adherence (H-MDA and L-MDA) in comparison with the use of each metabolite individually. In the analysis of correlation, the combined metabolite model showed a strong correlation with the MedDiet score ($r=0.7$; $p<0.001$), as well as with the intake of vegetables, fruits, legumes, fish ($r=0.2 – 0.3$; $p<0.01$) (Supplementary Table 3) and dietary fiber ($r=0.3$; $p<0.01$). Otherwise, the three individual

metabolites had good correlation with the MedDiet score (r=0.3 – 0.6; p<0.001) but weaker or not significant correlations with the intake of individual foods.

4. Discussion

In the present metabolomic study, we identified the urinary metabolite profile of 34 metabolites that enable discrimination between two groups of individuals with high or low adherence to the score of Mediterranean diet adherence validated in the PREDIMED study [13]. Participants in the L-MDA group had a cut-off of ≤ 7 points and those in the H-MDA group ≥ 10 points, as previously proposed [15].

The set of metabolites that discriminated between H-MDA and L-MDA suggested the metabolic modulation of the MedDiet. These metabolites are involved in multiple molecular mechanisms and metabolic pathways, which together provide a holistic view of variations in the urine metabolome due to the effect of dietary pattern. Some of these metabolites have previously been proposed as putative biomarkers of single food intake and also related to foods included in MedDiet, as well as, up- and down-regulated endogenous metabolites. We found that the higher excretion of proline betaine and scyllo-inositol in H-MDA correlated significantly with the intake of citrus fruits (r=0.36 and r=0.35; $p<0.001$, respectively), which is consistent with previous studies that have proposed them as markers of citrus fruit intake [27]. Recently, in our previous work, we also found positive correlations of proline betaine signals with orange consumption in long-term MedDiet intervention and low-fat diet groups in a subsample of non-diabetic participants of the Predimed study [8]. In this regard, in the H-MDA group, we also identified other metabolites derived from the microbial metabolism: 4-hydroxyhippurate, 3-(3-hydroxyphenyl)-3-hydroxypropanoate and 4-hydroxyphenylacetate. These metabolites have been described after interventions with mixed red wine/grape juice extracts [28], as well as linked to the intake of other polyphenol-rich foods such as cocoa and almond, among others (http://phenol-explorer.eu/). It is interesting to note that these microbial metabolites were not found in our previous work evaluating the intervention with MedDiet [8].

Other metabolites have been related to the intake of meat or fish. Both dimethylamine and trimethylamine-N-oxide (TMAO), which were higher in the H-MDA group than in the L-MDA, have been related to the intake of fish and seafood [29,30]. Nevertheless, only TMAO had a significant correlation with the intake of total fish (r=0.24, $p<0.01$) in our study sample and this correlation was not previously observed in the subset of samples of non-diabetics in the Predimed study [8]. In addition, we identified some compounds related to the histidine pathway: carnosine, anserine and methylhistidine. While 3-methylhistidine and carnosine are proposed as good biomarkers of red meat intake [31–33], anserine has alternatively been proposed as a marker of white meat (i.e. poultry) [34] or salmon intake [35]. In our study, subjects in the H-MDA group excreted lower amounts of 3-methylhistidine and higher amounts of carnosine and anserine than those in the L-MDA group. We found significant correlations between anserine, with antioxidant properties, and the intake of white meats (r=0.231, $p<0.01$) but not with fish ($p>0.05$). In addition, we did not find significant correlations between methylhistidine or carnosine and foods. Nevertheless, it should be noted that carnosine is a normal constituent in human urine, which occurs naturally in the skeletal muscle of mammals and has antioxidant properties and therapeutic potential against numerous diseases [36]. To the best of our knowledge, there is limited information about which mechanisms could increase these antioxidants in the organism, but a recent study has shown that dietary vitamin B6 could determine the carnosine concentration in the skeletal muscle of rats [37]. In this regard, we observed significantly higher ingested concentrations of vitamin B6 in subjects in the H-MDA group (2.9 ± 0.7 mg/d) than in those in the L-MDA group (2.4 ± 0.5 mg/d, $p<0.001$). Moreover, methylhistidine is a metabolite of anserine and carnosine in the histidine pathway. Methylhistidine is also found in urine as anserine and carnosine and its concentration was inversely correlated with them in our study (r=-0.32, $p<0.05$; r=-0.27;

$p<0.05$, respectively). Thus, the contribution of a MedDiet in the modulation of the histidine pathway could be an interesting field to explore further.

Besides the metabolites related to the intake of foods, we also identified other endogenous metabolites involved in energy metabolism. In particular, individuals in the H-MDA group showed lower excretion of glucose, lactate and succinate than those in the L-MDA group, whose metabolites are related to pathways affected in diabetic patients and other diseases [38]. This finding is interesting because although our population included diabetic individuals, their distribution and the use of medication between the H-MDA and L-MDA groups were well balanced (Table 1). This fact suggested, therefore, that in comparison to the L-MDA group, individuals in the H-MDA group could show relatively better glycaemic control. Supporting this notion, several studies have previously found an inverse association between adherence to a MedDiet and indices of glucose homeostasis in the general population, including elderly people, and high-risk patients [39]. Previous works comparing intervention with MedDiet and low-fat diet in non-diabetic Predimed participants did not found changes for these metabolites [8].

After identifying 34 metabolites in the H-MDA pattern, we studied the prediction of high adherence to MedDiet pattern. To improve the prediction of H-MDA, a model with a combination of more than one discriminatory metabolite was developed. For this purpose, population was split into training and validation sets. The model included PAGN, 4-hydroxyphenylacetate and *p*-cresol. Interestingly and in accordance with this finding, in our previous study, we also reported associations of PAGN and *p*-cresol with long-term interventions with MedDiet [8]. Currently, the use of multi-metabolite biomarkers is still limited. In a recent study, Marklund *et al.* (2014) combined six serum metabolites to create a dietary biomarker score [19]. This score was aimed at assessing the compliance with a healthy Nordic diet in a population with metabolic syndrome (MetS) and subsequently used to

analyze the effects of diet on cardiometabolic risk factors among those individuals with highest compliance to this diet [19]. In our study, the inclusion of PAGN, 4-hydroxyphenylacetate and *p*-cresol in the model, as well as the identification of other metabolites derived from gut microbiota in the previous set of discriminant metabolites, highlight the role of a MedDiet in the modulation of gut microbiota. Supporting this notion, De Filippis *et al.* (2015) demonstrated the relationship between the level of MedDiet adherence, gut microbiota and microbial metabolites [18]. Later, Haro *et al.* (2016) demonstrated that a long-term intervention with a MedDiet partially restores the alteration in the gut microbiota composition in individuals with MetS [40]. Interestingly, by using the same PREDIMED MedDiet score, authors found a weak but significant correlation between this score and the abundance of *F. prausnitzii* and *B. adolescentis*, which were observed to be decreased in MetS patients compared with non-MetS patients [40]. More recently, Gutierrez-Díaz *et al.* (2017) found increased fecal concentrations of benzoic and 3-hydroxyphenylacetic acids in individuals with higher adherence to MedDiet [14]. Interestingly and in accordance with our results, a hydroxyphenylacetic acid metabolite entered in our multi-metabolite model, thus supporting the notion that H-MDA modulates gut microbiota metabolism. Nowadays, there exists an increasing interest in the study of the relationship between MedDiet adherence and food-derived alterations of the gut microbiota in order to use these data in the prevention of food-related diseases [41].

The three metabolites included in our multi-metabolite model may arise from several food sources. PAGN is formed by the conjugation of glutamine with phenylacetate, which can arise from endogenous *β*-oxidation of phenyl containing fatty acids or phenylalanine metabolism [42], or be obtained through the exogenous intake contained in plant-food sources [43]. In line with our data, O'Sullivan *et al.* (2011) found a positive association between the urinary concentration of PAGN and vegetable intake, suggesting that PAGN in urine may be

a useful biomarker of vegetable intake [11]. *p*-Cresol is a product of microbial tyrosine breakdown via 4-hydroxyphenylacetate [44]. Moreover, 4-hydroxyphenylacetate has been found to be related to the intake of vegetarian diets [31] and polyphenol-rich foods, including red wine [45] and dark chocolate [46]. A further correlation analysis showed that although we found weak or non-significant correlations between individual metabolites and some food groups, the combined multi-metabolite model was significantly associated with the intake of vegetables, fruits, legumes, fish and total fiber (Supplementary Table 3).

The main limitation of this cross-sectional study is that our results may not be generalized or extrapolated to other populations, mainly because of the age of our population as well as their high risk of CVD. However, the validity of the PREDIMED MedDiet score in distinguishing between individuals at high or low MedDiet adherence has been ascertained within the prominent PREDIMED study [15]. Moreover, the current complementary use of this score with an NMR-based untargeted metabolomics approach deserves mention, first, because dietary scores are still the main tool to assess dietary patterns adherence, and second, because this robust analytical platform allows us to identify differences in urinary metabolome at micro- to millimolar levels [47].

In conclusion, the model composed of microbial metabolites was the biomarker that defined high adherence to a Mediterranean dietary pattern. This fact highlights the role of microbiota in the study of the biomarkers associated to the MedDiet pattern. The effect of the MedDiet involves several interconnected molecular mechanisms through complex regulatory networks, which are reflected in the microbiota metabolism as the metabolic modulation of H-MDA. Future studies in nutritional research should have to include the measurement of these dietary biomarkers in order not only to improve the assessment of dietary intake, but also to understand in depth the molecular mechanisms involved in the effects associated with food intake. The proposed biomarker may assess and aid nutritional epidemiology in future

associations between adherence to the MedDiet and the prevention or amelioration of chronic diseases.

Conflict of interest

The authors declare no financial or personal conflicts of interest.

Role of the funding sources

The funding sources had no role in the study design, data collection and analysis, decision to publish or the preparation of the manuscript.

Authorship

The authors' contributions to the manuscript were as follows: MUS, RL and CAL designed the research; RE and DC provided the clinical samples from the PREDIMED study; RVF, EAA, and FMG conducted the NMR experimental data and the NMR data analysis; EAA, MGA, MUS, RL, FC and AS conducted the statistical analyses; EAA and MUS wrote the paper; All authors provided critical revision. MUS and CAL have the primary responsibility for the final content. All authors read and approved the final manuscript.

Acknowledgements

MUS would like to thank the 'Ramón y Cajal' programme from MINECO and the Fondo Social Europeo (RYC-2011-09677). EAA would like to thank CONACYT (México) for the Ph.D. fellowship.

REFERENCES

[1] Boeing H. Nutritional epidemiology: New perspectives for understanding the diet-disease relationship? Eur J Clin Nutr 2013;67:424–9. doi:10.1038/ejcn.2013.47.

[2] Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002;13. doi:10.1097/00041433-200202000-00002.

[3] Scientific Report of the 2015 Dietary Guidelines Advisory Committee (DGAC). Part D. Chapter 2: Dietary Patterns, Foods and Nutrients, and Health Outcomes 2015. doi:10.1017/CBO9781107415324.004.

[4] Harmon BE, Boushey CJ, Shvetsov YB, Ettienne R, Reedy J, Wilkens LR, et al. Associations of key diet-quality indexes with mortality in the Multiethnic Cohort: the Dietary Patterns Methods Project. Am J Clin Nutr 2015;101:587–97. doi:10.3945/ajcn.114.090688.

[5] Andersen M-BS, Rinnan A, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J Proteome Res 2014;13:1405–18. doi:10.1021/pr400964s.

[6] Garcia-Aloy M, Llorach R, Urpi-Sarda M, Tulipani S, Estruch R, Martínez-González MA, et al. Novel Multimetabolite Prediction of Walnut Consumption by a Urinary Biomarker Model in a Free-Living Population: the PREDIMED Study. J Proteome Res 2014;13:3476–83. doi:10.1021/pr500425r.

[7] Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas M-E, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. Am J Clin Nutr 2010;92:436–43. doi:10.3945/ajcn.2010.29672.

[8] Vazquez-Fresno R, Llorach R, Urpi-Sarda M, Lupianez-Barbero A, Estruch R, Corella D, et al. Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: a 1- and 3-year follow-up in the PREDIMED study. J Proteome Res 2015;14:531–40. doi:10.1021/pr5007894.

[9] Bondia-Pons I, Martinez JA, de la Iglesia R, Lopez-Legarrea P, Poutanen K, Hanhineva K, et al. Effects of short- and long-term Mediterranean-based dietary treatment on plasma LC-QTOF/MS metabolic profiling of subjects with metabolic syndrome features: The Metabolic Syndrome Reduction in Navarra (RESMENA) randomized controlled trial. Mol Nutr Food Res 2015;59:711–28. doi:10.1002/mnfr.201400309.

[10] Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, Illig T, et al. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. Eur J Epidemiol 2011;26:145–56. doi:10.1007/s10654-010-9524-7.

[11] O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: Potential role in dietary assessment studies. Am J Clin Nutr 2011;93:314–21. doi:10.3945/ajcn.110.000950.

[12] Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med 2003;348:2599–608. doi:10.1056/NEJMoa025039.

[13] Schroder H, Fito M, Estruch R, Martinez-Gonzalez MA, Corella D, Salas-Salvado J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011;141:1140–5. doi:10.3945/jn.110.135566.

[14] Gutierrez-Diaz I, Fernandez-Navarro T, Salazar N, Bartolome B, Moreno-Arribas MV, de Andres-Galiana EJ, et al. Adherence to a Mediterranean Diet Influences the Fecal Metabolic Profile of Microbial-Derived Phenolics in a Spanish Cohort of Middle-Age and Older People. J Agric Food Chem 2017;65:586–95. doi:10.1021/acs.jafc.6b04408.

[15] Martinez-González MA, Garcia-Arellano A, Toledo E, Salas-Salvado J, Buil-Cosiales P, Corella D, et al. A 14-item Mediterranean diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. PLoS One 2012;7:e43134. doi:10.1371/journal.pone.0043134.

[16] Talegawkar SA, Bandinelli S, Bandeen-Roche K, Chen P, Milaneschi Y, Tanaka T, et al. A higher adherence to a Mediterranean-style diet is inversely associated with the development of frailty in community-dwelling elderly men and women. J Nutr 2012;142:2161–6. doi:10.3945/jn.112.165498.

[17] Guasch-Ferre M, Zheng Y, Ruiz-Canela M, Hruby A, Martinez-Gonzalez MA, Clish CB, et al. Plasma acylcarnitines and risk of cardiovascular disease: effect of Mediterranean diet interventions. Am J Clin Nutr 2016. doi:10.3945/ajcn.116.130492.

[18] De Filippis F, Pellegrini N, Vannini L, Jeffery IB, La Storia A, Laghi L, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut 2015. doi:10.1136/gutjnl-2015-309957.

[19] Marklund M, Magnusdottir OK, Rosqvist F, Cloetens L, Landberg R, Kolehmainen M, et al. A dietary biomarker approach captures compliance and cardiometabolic effects of a healthy Nordic diet in individuals with metabolic syndrome. J Nutr 2014;144:1642–9. doi:10.3945/jn.114.193771.

[20] Martínez-González MÁ, Corella D, Salas-Salvadó J, Ros E, Covas MI, Fiol M, et al.

Cohort profile: design and methods of the PREDIMED study. Int J Epidemiol 2012;41:377–85. doi:10.1093/ije/dyq250.

[21] Estruch R, Ros E, Salas-Salvado J, Covas M-I, Corella D, Aros F, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med 2013;368:1279–90. doi:10.1056/NEJMoa1200303.

[22] Fernandez-Ballart JD, Pinol JL, Zazpe I, Corella D, Carrasco P, Toledo E, et al. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. Br J Nutr 2010;103:1808–16. doi:10.1017/S0007114509993837.

[23] Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the Minnesota Leisure Time Physical Activity Questionnaire in Spanish men. The MARATHOM Investigators. Am J Epidemiol 1994;139:1197–209.

[24] Mataix Verdú J. Tabla de Composición de Alimentos. Univ Granada 2003.

[25] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res 2013;41:D801-7. doi:10.1093/nar/gks1065.

[26] Feng Q, Liu Z, Zhong S, Li R, Xia H, Jie Z, et al. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. Sci Rep 2016;6:22525. doi:10.1038/srep22525.

[27] Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, et al. Metabolomics in nutritional epidemiology: Identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr 2014;100:208–17. doi:10.3945/ajcn.113.078758.

[28] van Dorsten FA, Grun CH, van Velzen EJJ, Jacobs DM, Draijer R, van Duynhoven JPM. The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. Mol Nutr Food Res 2010;54:897–908. doi:10.1002/mnfr.200900212.

[29] Mitchell SC, Zhang AQ, Smith RL. Dimethylamine and diet. Food Chem Toxicol 2008;46:1734–8. doi:10.1016/j.fct.2008.01.010.

[30] Schmedes M, Aadland EK, Sundekilde UK, Jacques H, Lavigne C, Graff IE, et al. Lean-seafood intake decreases urinary markers of mitochondrial lipid and energy metabolism in healthy subjects: metabolomics results from a randomized crossover intervention study. Mol Nutr Food Res 2016. doi:10.1002/mnfr.201500785.

[31] Stella C, Beckwith-hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of Human Metabolic Phenotypes to Dietary Modulation. J Proteome Res 2006;5:2780–8. doi:10.1021/pr060265y.

[32] Altorf-van der Kuil W, Brink EJ, Boetje M, Siebelink E, Bijlsma S, Engberink MF, et al. Identification of biomarkers for intake of protein from meat, dairy products and grains: a controlled dietary intervention study. Br J Nutr 2013;110:810–22. doi:10.1017/S0007114512005788.

[33] Cross AJ, Major JM, Sinha R. Urinary biomarkers of meat consumption. Cancer Epidemiol Biomarkers Prev 2011;20:1107–11. doi:10.1158/1055-9965.EPI-11-0048.

[34] Yeum K-J, Orioli M, Regazzoni L, Carini M, Rasmussen H, Russell RM, et al. Profiling histidine dipeptides in plasma and urine after ingesting beef, chicken or chicken broth in humans. Amino Acids 2010;38:847–58. doi:10.1007/s00726-009-0291-2.

[35] Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J. Data-driven strategy for

the discovery of potential urinary biomarkers of habitual dietary exposure. Am J Clin Nutr 2013;97:377–89. doi:10.3945/ajcn.112.048033.

[36] Boldyrev AA, Aldini G, Derave W. Physiology and pathophysiology of carnosine. Physiol Rev 2013;93:1803–45. doi:10.1152/physrev.00039.2012.

[37] Suidasari S, Stautemas J, Uragami S, Yanaka N, Derave W, Kato N. Carnosine Content in Skeletal Muscle Is Dependent on Vitamin B6 Status in Rats. Front Nutr 2015;2:39. doi:10.3389/fnut.2015.00039.

[38] Urpi-Sarda M, Almanza-Aguilera E, Tulipani S, Tinahones FJ, Salas-Salvadó J, Andres-Lacueva C. Metabolomics for Biomarkers of Type 2 Diabetes Mellitus: Advances and Nutritional Intervention Trends. Curr Cardiovasc Risk Rep 2015;9. doi:10.1007/s12170-015-0440-y.

[39] Salas-Salvadó J, Bulló M, Estruch R, Ros E, Covas M-I, Ibarrola-Jurado N, et al. Prevention of diabetes with Mediterranean diets: a subgroup analysis of a randomized trial. Ann Intern Med 2014;160:1–10. doi:10.7326/M13-1725.

[40] Haro C, Garcia-Carpintero S, Alcala-Diaz JF, Gomez-Delgado F, Delgado-Lista J, Perez-Martinez P, et al. The gut microbial community in metabolic syndrome patients is modified by diet. J Nutr Biochem 2016;27:27–31. doi:10.1016/j.jnutbio.2015.08.011.

[41] Del Chierico F, Vernocchi P, Dallapiccola B, Putignani L. Mediterranean diet and health: food effects on gut microbiota and disease control. Int J Mol Sci 2014;15:11678–99. doi:10.3390/ijms150711678.

[42] Andrew W. Nicholls John N. Haselden, Catherine J. Waterfield JKN. A metabonomic approach to the investigation of drug-induced phospholipidosis: an NMR spectroscopy and pattern recognition study. Biomarkers 2000;5:410–23.

doi:10.1080/135475000750052420.

[43] Wightman F, Lighty DL. Identification of phenylacetic acid as a natural auxin in the shoots of higher plants. Physiol Plant 1982;55:17–24. doi:10.1111/j.1399-3054.1982.tb00278.x.

[44] Ramakrishna BS, Roberts-Thomson IC, Pannall PR, Roediger WE. Impaired sulphation of phenol by the colonic mucosa in quiescent and active ulcerative colitis. Gut 1991;32:46–9.

[45] Vazquez-Fresno R, Llorach R, Alcaro F, Rodriguez MA, Vinaixa M, Chiva-Blanch G, et al. (1)H-NMR-based metabolomic analysis of the effect of moderate wine consumption on subjects with cardiovascular risk factors. Electrophoresis 2012;33:2345–54. doi:10.1002/elps.201100646.

[46] Martin F-PJ, Rezzi S, Pere-Trepat E, Kamlage B, Collino S, Leibold E, et al. Metabolic effects of dark chocolate consumption on energy, gut microbiota, and stress-related metabolism in free-living subjects. J Proteome Res 2009;8:5568–79. doi:10.1021/pr900607v.

[47] Liu H, Tayyari F, Edison AS, Su Z, Gu L. NMR-based metabolomics reveals urinary metabolome modifications in female Sprague-Dawley rats by cranberry procyanidins. J Nutr Biochem 2016;34:136–45. doi:10.1016/j.jnutbio.2016.05.007.

Table 1.

Demographic characteristics and medication usage of 119participants classified in low and high Mediterranean diet adherence at baseline

	Total	L-MDA (≤ 7 MedDiet score)	H-MDA (≥10 MedDiet score)	*P* value
Participants, *n*	119	63	56	
Age, y	66.7 ± 5.9	66.8 ± 6.1	66.6 ± 5.7	0.83
Women, *n* (%)	82 (68.9)	42 (66.7)	40 (71.4)	0.58
BMI, kg/m^2	30.3 ± 4.5	31.0 ± 4.1	29.5 ± 4.7	0.08
Current smoker, *n* (%)	12 (10.1)	7 (11.1)	5 (8.9)	0.69
Family history of CHD, *n* (%)	39 (32.8)	23 (36.5)	16 (28.6)	0.34
Diabetes, *n* (%)	63 (52.9)	36 (57.1)	26 (46.4)	0.24
Hypertension, *n* (%)	102 (85.7)	51 (81.0)	51 (91.1)	0.12
Hypercholesterolemia, *n* (%)	85 (71.4)	44 (69.8)	41 (73.2)	0.68
Medication use				
Oral antidiabetic agents, *n* (%)	46 (38.7)	26 (41.3)	20 (35.7)	0.53
Insulin, *n* (%)	18 (15.1)	10 (16.0)	8 (14.3)	0.81
Antihypertensive agents, *n* (%)	56 (47.1)	27 (42.9)	29 (51.8)	0.33
Lipid-lowering medication, *n* (%)	58 (48.7)	29 (46.0)	29 (51.8)	0.53

Values are mean ± SD or n (%) as appropriate. Differences between low and high Mediterranean diet adherence groups were tested by Student's t-test and chi-square test for continuous and categorical variables, respectively ($p<0.05$). Abbreviations: CHD, coronary heart disease; H-MDA, high Mediterranean diet adherence; L-MDA, low Mediterranean diet adherence; MedDiet, Mediterranean Diet; METs, metabolic equivalent of task.

Table 2.

Dietary intake of 119 participants with low and high adherence to Mediterranean Diet

Food and nutrient intake	Total	L-MDA (≤ 7 MedDiet score)	H-MDA (≥10 MedDiet score)	*P* value
Participants, *n*	119	63	56	
Foods				
Olive oil, g/d	38 ± 15	35 ± 15	41 ± 13	0.024
Nuts, g/d	12 ± 12	9 ± 12	15 ± 12	0.018
Vegetables, g/d	402 ± 192	377 ± 215	430 ± 160	0.032
Fruits, g/d	480 ± 242	419 ± 235	550 ± 233	0.003
Legumes, g/d	20 ± 10	17 ± 9	22 ± 10	0.004
Cereals, g/d	269 ± 111	275 ± 112	263 ± 111	0.56
Fish, g/d	103 ± 50	84 ± 47	124 ± 45	<0.001
Meat, g/d	147 ± 69	152 ± 81	142 ± 52	0.41
Dairy, g/d	358 ± 213	366 ± 207	350 ± 220	0.70
Pastries, g/d	18 ± 21	19 ± 22	16 ± 20	0.44
Wine, mL/d	56 ± 116	54 ± 4	47 ± 3	0.71
Nutrients				
Total energy intake, kcal/d	2397 ± 569	2399 ± 635	2394 ± 490	0.96
Carbohydrates, % of total energy	43.3 ± 6.8	43.4 ± 6.6	43.2 ± 7.0	0.87
Proteins, % of total energy	16.7 ± 2.6	16.4 ± 2.8	17.1 ± 2.4	0.13
Total fat, % of total energy	38.0 ± 6.2	37.8 ± 6.5	38.2 ± 5.8	0.71
Dietary fiber, g/d	28.4 ± 8.0	26.2 ± 7.0	31.0 ± 8.3	0.001
MedDiet score	8.3 ± 2.2	6.4 ± 0.9	10.5 ± 0.6	<0.001

Values are mean ± SD. Differences between low and high Mediterranean diet adherence groups were tested by Student's t-test ($p<0.05$).
Abbreviations: H-MDA, high Mediterranean diet adherence; L-MDA, low Mediterranean diet adherence; MedDiet, Mediterranean diet.

Table 3.

List of discriminant urinary metabolites between individuals with low and high Mediterranean diet adherence[1]

Metabolite	^{1}H δ, ppm (multiplicity)	Excretion in H-MDA	*P* value[2]
Protein/amino acid metabolism			
3-Methylhistidine	7.95 (s)	↓	2.9×10^{-2}
Alanine	1.49 (d)	↓	9.4×10^{-3}
Anserine	3.79 (s), 8.29 (s)	↑	7.8×10^{-4}
Carnosine	7.18 (s), 8.15 (s)	↑	3.8×10^{-2}
Creatine	3.94 (s)	↑	7.6×10^{-3}
Creatinine	3.05 (s), 4.07 (s)	↑	2.3×10^{-3}
Glycine	3.57 (s)	↓	9.1×10^{-3}
Guanidoacetate	3.78 (s)	↑	6.7×10^{-3}
Histidine	7.09 (s), 7.93 (s)	↑	4.8×10^{-4}
Lysine	1.72 (m)	↓	2.1×10^{-2}
N-Acetylglutamine	2.08 (m), 2.30 (m)	↑	6.0×10^{-3}
Proline betaine	3.11 (s), 3.30 (s), 4.08 (m)	↑	6.2×10^{-7}
Gut microbiota metabolites			
3-Indoxyl sulfate	7.27 (t), 7.51 (d), 7.71 (d)	↓	1.6×10^{-4}
4-Hydroxyhippurate	6.98 (d), 7.75 (d)	↑	5.3×10^{-4}
4-Hydroxyphenylacetate	6.88 (d), 7.16 (s)	↑	5.0×10^{-4}
Dimethylsulfone	3.16 (s)	↑	6.7×10^{-3}
Hippurate	7.55 (tt), 7.82 (dd)	↓	2.0×10^{-3}
HPHPA	5.02 (t), 6.86 (d), 6.93 (br s), 7.30 (t)	↑	3.4×10^{-2}
Isobutyrate	1.07 (d)	↓	2.2×10^{-4}
p-Cresol	2.34 (s), 7.22 (d), 7.28 (d)	↑	9.6×10^{-5}
Phenylacetylglutamine	1.93 (m), 2.27 (m), 3.66 (m), 4.18 (m), 7.36 (t), 7.43 (t), 8.0 (d)	↑	1.5×10^{-11}

Energy metabolism (glycolysis/gluconeogenesis, TCA and ketone bodies)			
β-glucose	3.39 (m), 3.44 (m), 3.49 (m), 3.90 (m), 4.66 (d)	↓	1.9×10^{-4}
Lactate	1.35 (d)	↓	1.1×10^{-2}
Succinate	2.42 (s)	↓	4.7×10^{-3}
Choline metabolism			
Dimethylamine	2.72 (s)	↑	5.2×10^{-4}
Betaine	3.27 (s), 3.91 (s)	↓	1.2×10^{-2}
TMAO	3.28 (s)	↑	2.6×10^{-3}
Inositol phosphate metabolism			
Scyllo-inositol	3.36 (s)	↑	7.8×10^{-4} 9.3×10^{-3}
Nicotinate and nicotinamide metabolism			
N-Methylnicotinamide	4.49 (s), 8.85 (m), 9.13 (s)	↑	
Propanoate metabolism			
Isopropanol	1.15 (d)	↓	1.9×10^{-2}
Purine and caffeine metabolism			
Xanthosine	7.86 (s)	↓	1.1×10^{-2}
Pyrimidine metabolism			
Urea	5.82 (br.s)	↑	3.8×10^{-3}
Methylguanidine	2.82 (s)	↑	2.0×10^{-2}
Malonate	3.12 (s)	↑	2.7×10^{-2}

[1]Obtained from the OSC-PLS-DA model (VIP > 1.0) and using the high Mediterranean diet adherence group as reference. (↑) and (↓) indicate a relatively higher and lower excretion, respectively, in H-MDA with respect to L-MDA.

[2]Differences between NMR-signals were tested by Student's t-test or Mann-Whitney U tests, according to their normal distribution. Significance was corrected by the Benjamini–Hochberg procedure. Abbreviations: HPHPA, 3-(3-Hydroxyphenyl)-3-hydroxypropanoate; multiplicity (s, singlet; br. s, broad singlet; d, doublet; dd, doublet of doublets; t, triplet; tt, triplet of triplets; m, multiplet); TMAO, Trimethylamine N-oxide.

Table 4.

Receiver operating characteristic (ROC) curve parameters of combined multi-metabolite model and individual metabolites for prediction of high Mediterranean diet adherence

	ROC curve parameters							
	Training set				Validation set			
	Specificity (%)	Sensitivity (%)	AUC (95% CI)	*P value*	Specificity (%)	Sensitivity (%)	AUC (95% CI)	*P value*
Multi-metabolite model	93.0	94.6	97.7 (95.0-100)	<0.001	95.2	94.7	97.0 (92.3-100)	<0.001
Phenylacetylglutamine	85.7	83.8	89.7 (82.7-96.7)	<0.001	85.7	68.4	81.7 (68.5-94.9)	0.001
p-cresol	59.5	81.1	73.2 (61.7-84.8)	<0.001	76.2	73.7	77.2 (62.4-92.0)	0.003
4-Hydroxyphenylacetate	62.0	73.0	68.1 (56.4-79.9)	0.006	66.7	84.2	77.7 (63.1-92.3)	0.003

Abbreviations: AUC, area under the curve

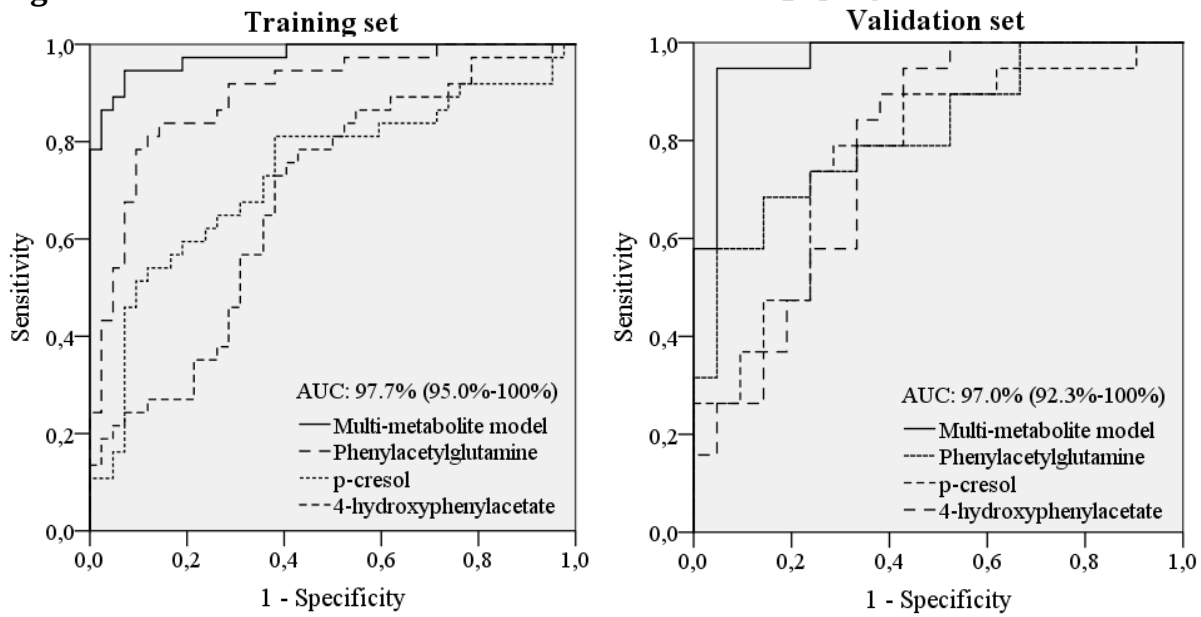
Figure legends:

Figure 1.

Receiver operating characteristic (ROC) curves of the combined multi-metabolite model and individual metabolites included for discrimination between high and low Mediterranean diet adherence in both training and validation sets.

Abbreviations: PAGN, phenylacetylglutamine.

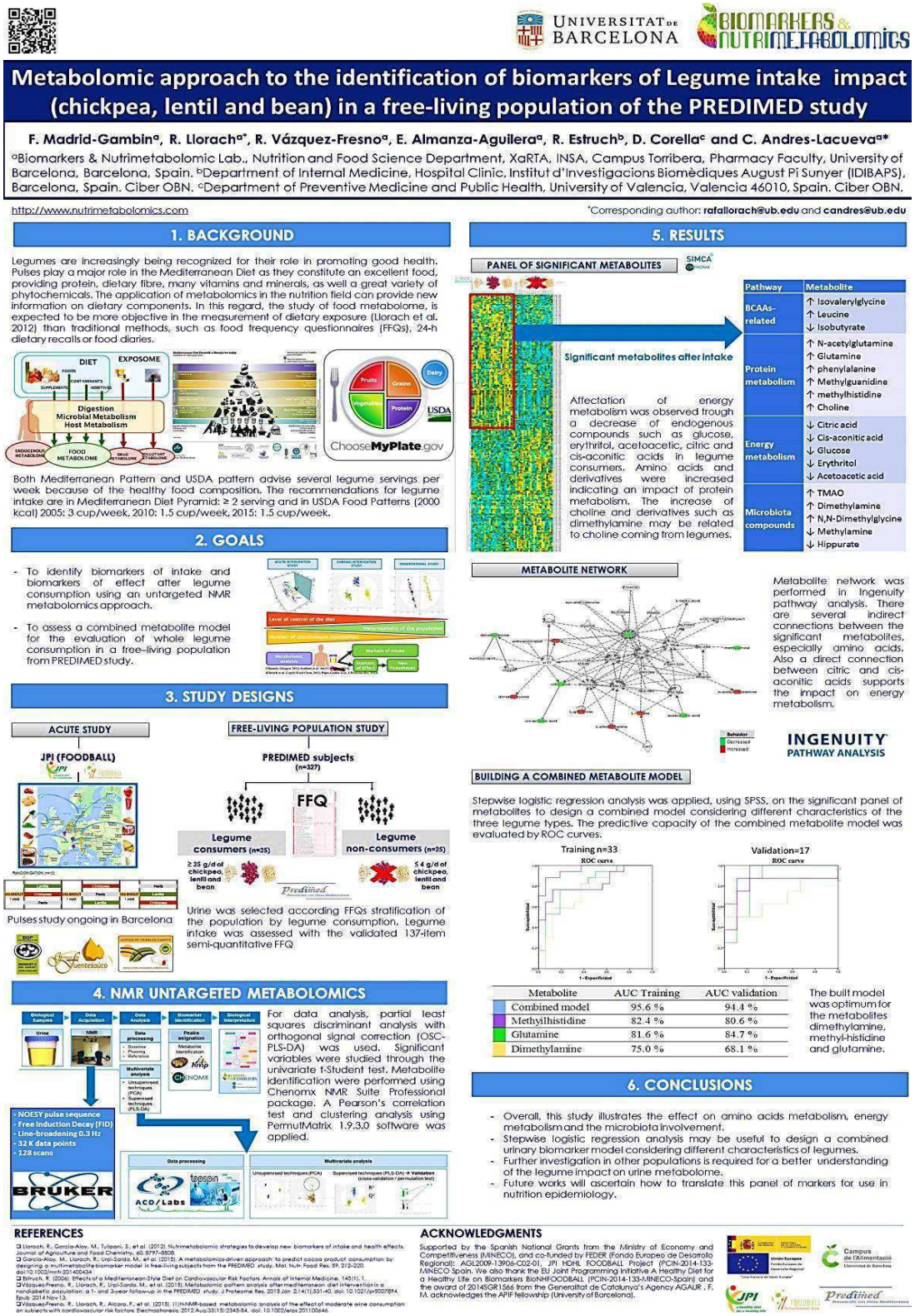
Figure 1



Comunicaciones en congresos

Durante el periodo predoctoral se han presentado 4 comunicaciones en congresos, las cuales presentan algunos de los resultados recogidos en la presente Tesis Doctoral. Se han presentado 3 pósters en 3 congresos internacionales: NuGOweek: Mechanisms of a long-life health (2015), XI Congreso Internacional sobre Dieta Mediterránea (2016), y NuGOweek: Phenotypes and prevention (2016). También se ha presentado una comunicación oral en el 4th JPI-HDHL-FOODBALL Meeting (2017).

Póster: Madrid-Gambin, F.; Llorach, R.; Vázquez-Fresno, R.; Urpi-Sarda, M.; Almanza-Aguilera, E.; Garcia-Aloy, M.; Estruch, R.; Corella, D.; Andres-Lacueva, C. Metabolomic approach to the identification of biomarkers of legume intake impact (chickpea, lentil and bean) in a free-living population of the PREDIMED study. *XI Congreso Internacional sobre Dieta Mediterránea*. Barcelona (España), 2016.



Póster: Garcia-Aloy, M.; Vázquez-Fresno, R.; Madrid-Gambin, F.; Vegas-Lozano, E.; de Villa Jubany, M. C. R.; Misawa, K.; Hase, T.; Shimotoyodome, A.; Andres-Lacueva, C. NMR-based metabolomic analysis of the impact of chlorogenic acid enriched coffee on urine metabolome. *NuGOweek 2016*: Copenhagen (Dinamarca), 2016.

NMR-BASED METABOLOMIC ANALYSIS OF THE IMPACT OF CHLOROGENIC ACID ENRICHED COFFEE ON URINE METABOLOME

BIOMARKERS & NUTRIMETABOLOMICS

UNIVERSITAT DE BARCELONA

Mar Garcia-Aloy[†,1], Rosa Vázquez-Fresno[†,1,*], Francisco Madrid-Gambin[†], Esteban Vegas-Lozano[‡], M. Carmen Ruiz de Villa Jubany[‡], Alex Sanchez-Pla[‡], Koichi Misawa[§], Tadashi Hase[§], Akira Shimotoyodome[§], Cristina Andres-Lacueva[†*]

[†]Biomarkers & Nutrimetabolomic Lab, Nutrition & Food Science Department XaRTA, INSA-UB, Campus Torribera, Pharmacy Faculty, University of Barcelona, Barcelona, Spain. [‡]Department of Statistics, University of Barcelona, Barcelona, Spain. [§]Biological Science Laboratories, Kao Corporation, 2606, Akabane, Ichikai, Haga, Tochigi 321-3497, Japan.
[1]Equal contribution; *Corresponding authors: rosavfr@gmail.com; candres@ub.edu.

BACKGROUND

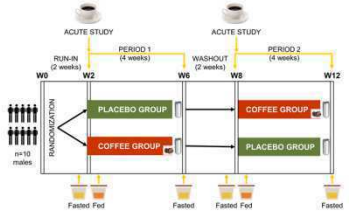
Many studies have highlighted that coffee could provide some benefits for human health. However, little is known about the role of chlorogenic acids (CGAs), the major polyphenol compounds of coffee.

GOAL

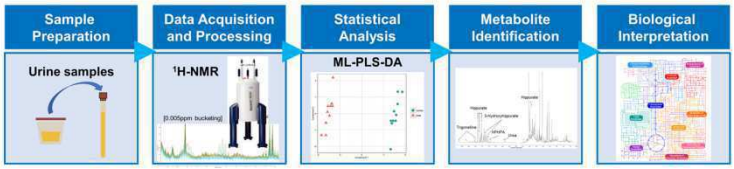
To assess the effect of CGAs from coffee in the urinary metabolome of 10 male volunteers that participated in a crossover randomized and placebo-controlled intervention study with a rich CGAs coffee extract beverage (CEB: 223 mg/100 ml of CGAs).

METHODOLOGY

STUDY DESIGN



^{1}H-NMR UNTARGETED METABOLOMICS



RESULTS

Figure 1. Score plots of the subjects projected onto the first principal components, PC1 and PC2, after CEB in acute (A) and sustained (B) interventions.

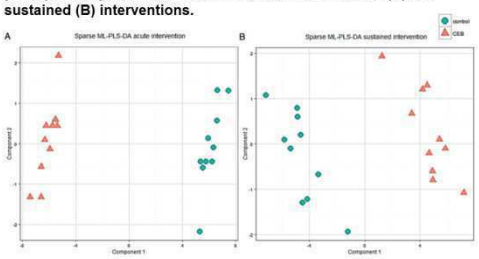


Figure 2. Representative 500 mHz ^{1}H-NMR spectrum of urine after CEB in sustained intervention.

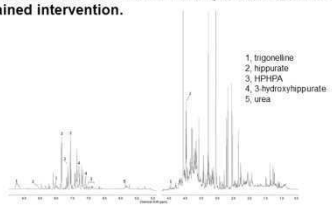


Figure 3. Representative 500 mHz ^{1}H-NMR spectrum of urine after CEB in acute intervention.

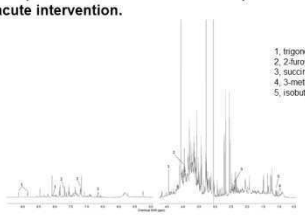


Table 1. Metabolites detected in urine after acute and sustained coffee extract beverage (CEB) in comparison with control beverage (CNT).

ACUTE INTERVENTION					SUSTAINED INTERVENTION				
Metabolite	δ	CBE vs. CNT	P value[a]	LR[b]	Metabolite	δ	CBE vs. CNT	P value[a]	LR[b]
Trigonelline	9.13 (s) 8.84 (d) 8.08 (t) 4.44 (s)	↑	7.88×10^{-8} 8.44×10^{-7} 1.26×10^{-2} 2.51×10^{-6}	04, 05 10, 11, 14 – 15, 25	Trigonelline	9.13 (s) 8.84 (d) 8.08 (t) 4.44 (s)	↑	1.64×10^{-3} 6.65×10^{-4} 9.54×10^{-3} 9.98×10^{-3}	08 07 – 18
2-Furoylglycine	7.70 (s) 7.18 (d) 6.64 (m) 3.93 (d)	↑	5.90×10^{-4} 9.68×10^{-2} 4.14×10^{-2} 3.00×10^{-4}	39 – – 31, 47	Hippuric acid	8.53 (s) 7.84 (d) 7.64 (t) 7.56 (t) 3.97 (d)	↑	2.74×10^{-3} 7.30×10^{-3} 6.66×10^{-5} 8.38×10^{-3} 1.17×10^{-2}	11, 17, 24 – 02, 05, 10 28, 34, 56 22, 35
Citric acid	2.68 (d) 2.54 (d)	↑	8.72×10^{-3} 1.31×10^{-3}	– 55	HPHPA	7.31 (t) 6.98 (d) 6.91 (s) 6.84 (d) 5.03 (m)[c]	↑	2.29×10^{-4} 1.69×10^{-2} 7.40×10^{-3} 2.89×10^{-2} –	01, 03, 27 44, 47 16, 21, 49 – –
Succinic acid	2.41 (s)	↑	7.78×10^{-4}	43		7.40 (t)		1.77×10^{-2}	55
3-Methyl-2-oxovaleric acid	1.10 (d) 0.90 (t)	↑	1.35×10^{-3} 1.30×10^{-2}	50 –		7.36 (s)		9.32×10^{-4}	06
					3-Hydroxyhippuric acid	7.28 (s) 7.13 (dd)	↑	3.18×10^{-1} 2.42×10^{-2}	– 42
Isobutyric acid	1.07 (d) 2.39 (m)	↑	8.91×10^{-3} 2.68×10^{-2}	– –		4.15 (d) 2.22 (s)		3.37×10^{-2} 1.94×10^{-2}	60 –

HPHPA: 3-(3-Hydroxyphenyl)-3-hydroxypropionic acid, s: singlet, d: doublet, t: triplet, dd: double doublet, m: multiplet. [a]P-value of univariate statistical crossover model of differences. [b]Loading rank of ML-PLS-DA. [c]Signal not considered because of proximity to water region.

CONCLUSIONS

- The increase of endogenous metabolites related to the TCA cycle (succinic, citric, isobutyric and 3-methyl-2-oxovaleric acids) after the acute study indicates that coffee CGAs impact on the central energy metabolism.
- Sustained consumption of CEB showed an increase of microbiota-derived compounds (hippuric, HPHA and 3-hydroxyhippuric acids) coming from the degradation of CGAs.
- 2-Furoylglycine is a potential biomarker of an acute intake of CEB.
- Trigonelline was found in urine after both acute and sustained intakes, as well as in the composition of the beverage exhibiting a direct excretion of this biomarker without any biotransformation, suggesting a non-interindividual variation.

The results of the current study show the capability of the NMR-based metabolomics approach to identify nutritional biomarkers for compliance monitoring in dietary interventions, and to clarify the role of specific food compounds in health.
Further work will allow to translate these urinary metabolites for use in epidemiological research to enhance their utility as dietary biomarkers, and to study their relationships with human health.

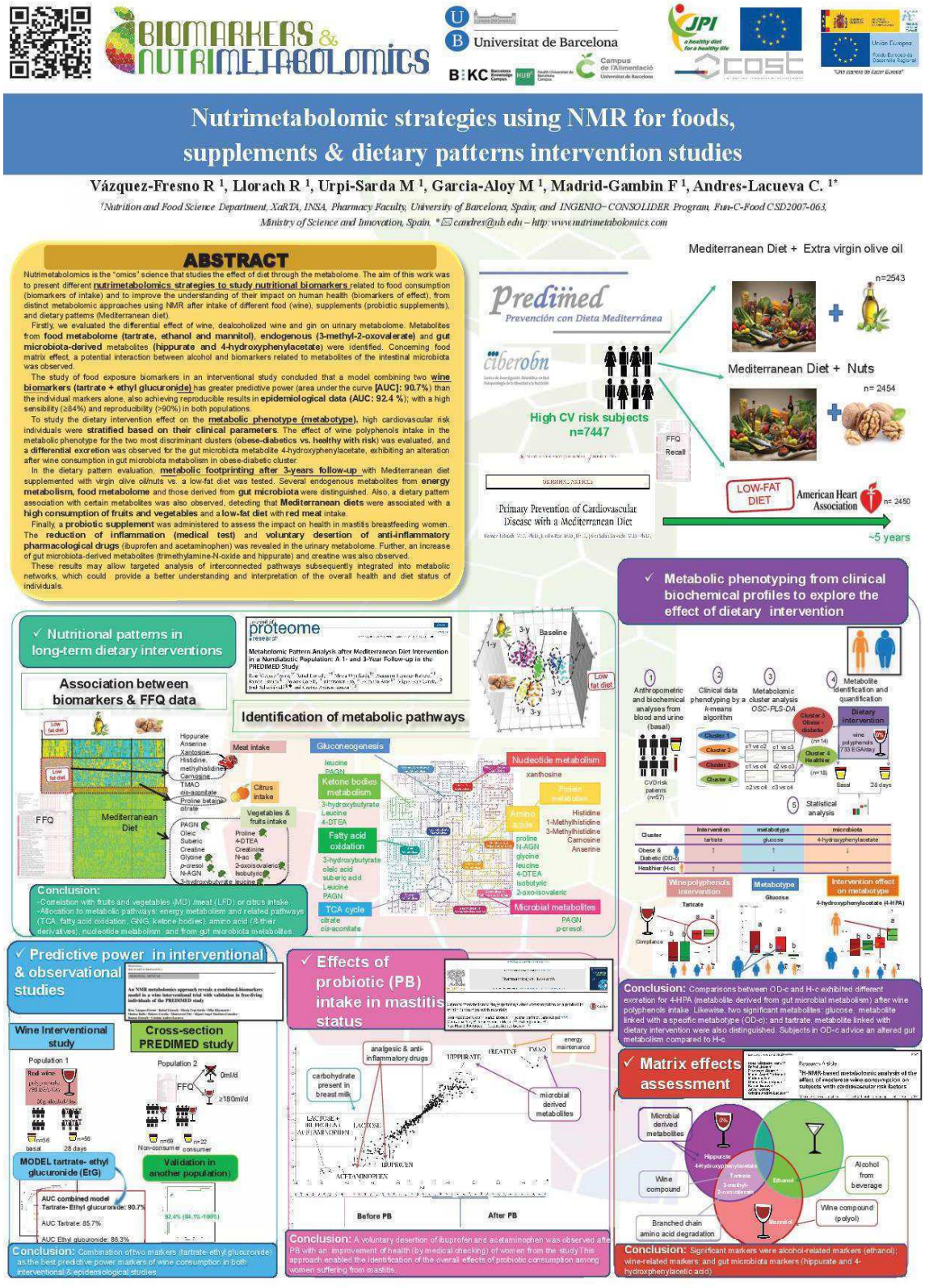
ACKNOWLEDGMENTS

This research was funded by: the Fondo Europeo de Desarrollo Regional (FEDER) (AGL2009-13906-C02-01, AGL2010-10084-E) and the JPI HDHL-FOODBALL (PCIN-2014-133-MINECO-Spain). The clinical study was supported in part by the KAO Corporation and Fundació Bosch i Gimpera (F6G-30768). We also thank the award of 2014-SGR-1566 from the Generalitat de Catalunya's Agency AGAUR. PM acknowledges the APIF fellowship (University of Barcelona).
The assistance to this conference has been supported by a travel grant from INSA-UB (University of Barcelona's Nutrition and Food Safety Research Institute).

Unión Europea
Fondo Europeo de Desarrollo Regional
Una manera de hacer Europa

Campus de l'Alimentació
Universitat de Barcelona
INSA

Póster: Vázquez-Fresno, R.; Llorach, R.; Urpi-Sarda, M.; García-Aloy, M.; Madrid-Gambin, F.; Andres-Lacueva, C. Nutrimetabolomic strategies using NMR for foods, supplements & dietary patterns intervention studies. *NuGOweek 2015*: Barcelona (España), 2015.



Comunicación oral: Madrid-Gambin, F.; Garcia-Aloy, M.; Brunius, C.; Pedersen, A.; Landberg, R.; Andres-Lacueva C. Metabolomic analysis of pulse consumption by 1H-NMR: statistical approach. *4th JPI-HDHL-FOODBALL Meeting*. Karlsruhe (Alemania), 2017.