Universitat de València

Instituto de Investigación Sanitaria La Fe

Metabolomic and lipidomic approaches for fact-based nutrition assessment in infants and lactating mothers

Victoria Ramos García

Programa de doctorado en Técnicas Experimentales en Química - 3158

Supervised by: Dr. Julia Kuligowski & Dr. Isabel Ten Doménech

Tutored by: Dr. Salvador Garrigues Mateo

Valencia, June 2023

Dra. Julia Kuligowski y Dra. Isabel Ten Doménech, ambas investigadoras postdoctorales pertenecientes al Grupo de Investigación en Perinatología del Instituto de Investigación Sanitaria La Fe, en calidad de directoras y Dr. Salvador Garrigues Mateo, catedrático del departamento de Química Analítica de la Universidad de Valencia, en calidad de tutor de la Tesis Doctoral presentada por Dña. Victoria Ramos Garcia, con el título *"Metabolomic and lipidomic approaches for fact-based nutrition assessment in infants and lactating mothers",*

CERTIFICAN:

Que han dirigido y supervisado tanto los aspectos científicos como los aspectos formales de dicha Tesis Doctoral, así como la elaboración de la memoria adjunta que, en su opinión, cumple los requisitos necesarios para que pueda procederse a su defensa.

Por ello, **INFORMAN FAVORABLEMENTE** la solicitud de depósito de la Tesis Doctoral.

Valencia, a 16 de junio de 2023

Fdo.: Julia Kuligowski
(Directora)

Fdo.: Isabel Ten Doménech
(Directora)

Fdo.: Salvador Garrigues Mateo

(Tutor)

A mis padres, a mi hermana y a Abel,
por confiar en mí y apoyarme al
máximo en todo momento.

Agradecimientos

Me siento tan afortunada de tener a tantas personas a las que agradecer tantas cosas que no sé ni por dónde empezar, así que empezaré por el principio.

A Boro, por haber sido tan buen tutor de TFG y TFM que tenía que elegirte también como tutor de tesis doctoral. Gracias por haberme descubierto el mundo de la investigación y, sobre todo, por haber visto potencial en mí desde los inicios.

A mis directoras de tesis, por haberme guiado y aconsejado tan bien a lo largo de estos cuatro años. Julia, gracias por haber depositado en mi tanta confianza, por haberme dado la libertad de descubrir mi forma de trabajar y gestionar mi tiempo. Espero que siempre recuerdes a tu "chica de oro" como la más risueña del laboratorio y no como la más quejica (que también). Isabel, mi fiel compañera de escritorio, te agradezco muchísimo todo lo que he aprendido trabajando a tu lado y no solamente sobre química. Nuestros "tesis chapter", dramas y cotilleos es de lo que más voy a echar de menos (aunque no te creas que te vas a librar de mi tan rápido).

Al Grupo de Investigación en Perinatología, especialmente a Máximo Vento, por ser el mejor líder de equipo y por habernos sabido guiar tan bien a todos. A todos los clínicos, por haberme enseñado tanto sobre la realidad de los pacientes, los que realmente dan sentido a nuestra investigación y, sobre todo, al Laboratorio 6.26. Alba, Merce, Laura, David, Ángel, Guillermo y José Luís, gracias por acompañarme durante este proyecto y por todos los buenos ratos que me habéis regalado en el laboratorio. Ha sido un verdadero honor coincidir con vosotros.

Al servicio de espectrometría de masas del SCSIE, en concreto a Sales y a Isabel. No os hacéis una idea de lo que me habéis alegrado los días de análisis de muestras, que no han sido pocos. Gracias por haber estado en pie de guerra conmigo para desarrollar todos los métodos desde cero y medir las tropecientas mil muestras. Os agradezco muchísimo todo lo que habéis hecho por mí, desde los tours guiados por los equipos hasta esperarme a deshoras para recibir mis preciadas muestras.

A mis amigos, por haber sido mi vía de desconexión total y por animarme sin ni siquiera saberlo. Gracias por esperarme cada finde con un "esmorzaret", un "soparet" o un "atenció bando".

A mis queridos químicos, a mis pipironisas Alba, Raquel, Dunia y Roberto. A vosotros os debo los años más felices de mi vida y sumando. Porque como muy bien dijo la poeta del grupo, "nosotros no ganaremos todas las batallas, pero la guerra es nuestra". Y tanto que ha sido nuestra. Te digo yo a ti, que sí.

A Miriam, por haber estado ahí siempre a pesar de la distancia y el tiempo. Gracias por ser la persona que mejor ha sabido entenderme en cuestión de segundos.

A mis abuelos, por ser mi motivación en la vida. A mi abuelo Vicente, gracias por confiar ciegamente en mí. A mi abuelo Enrique, no me cabe duda de lo orgulloso que estás de mí y lo mucho que me proteges cada día, desde allá donde estés. A mi abuela Virgi, por transmitirme esa calma tan positiva que te caracteriza y por enseñarme que lo más importante en la vida, por suerte, ya lo tengo. Y, como no, a mi abuela Victoria, por ser ese rayo de sol cuándo todo lo veo oscuro y por enseñarme que todo lo que pasa en la vida es para bien.

A mis padres, no sé cómo resumir todo lo que os agradezco. A vosotros os debo la vida entera. Vosotros sois los testigos de que llevo la ciencia conmigo desde pequeñita. Gracias por criarme con los valores que me habéis inculcado, por apostar por mi educación, por confiar en mi capacidad sin dudar ni lo más mínimo y por apoyarme en cada decisión que he tomado, incluso cuando escogí química, esa carrera "tan difícil".

A Virginia (Virxin para los amigos, Shinit para mí), eres el mejor regalo que nuestros padres me han hecho nunca. Gracias por estar siempre, incondicionalmente, por celebrar conmigo todos mis éxitos, llorar conmigo todos mis fracasos y hacerme ser mejor persona. Eres mi definición de amor verdadero.

A Abel, mi persona favorita del universo. No tengo palabras suficientes para describir lo feliz que me haces todos y cada uno de los días de mi vida. Gracias por escucharme siempre, por preocuparte, por ser tan atento, por dar el 180% de ti en los momentos en los que yo estoy al 20%, por querer construir nuestra vida juntos, siempre de la mano, y por quererme tanto y tan bien. Sin duda eres mi persona vitamina. Te amo muchísimo, always.

"And last but not least, I wanna thank me". Por último, quería darme las gracias a mí misma. Por haber luchado y trabajado tanto y tan duro para cumplir mis objetivos y mis metas, por haber confiado tanto en mí y por haber permanecido siempre fiel a mí misma.

Gràcies. Gracias. Thanks. Danke.

Prefacio

La presente tesis doctoral se enmarca en la modalidad de compendio de publicaciones, de acuerdo con el reglamento sobre depósito, evaluación y defensa de la tesis doctoral de la Escuela de Doctorado y el Comité Académico del Programa de Doctorado en Técnicas Experimentales en Química de la Universitat de València. Esta tesis doctoral está compuesta por seis artículos científicos y un capítulo de libro originales de primer autor, publicados en revistas indexadas en "*Journal Citation Reports*". Además de otros dos artículos (Capítulos 4 y 8) en proceso de publicación.

Artículo 1

V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, A. Ramón-Beltrán, M.J. Vaya, F. Mohareb, C. Molitor, P. Refinetti, A. Silva, L.A. Rodrigues, S. Rezzi, A.C.C. Hodgson, S. Canarelli, E. Bathrellou, E. Mamalaki, M. Karipidou, D. Poulimeneas, M. Yannakoulia, C.K. Akhgar, A. Schwaighofer, B. Lendl, J. Karrer, D. Migliorelli, S. Generelli, M. Gormaz, M. Vasileiadis, J. Kuligowski, M. Vento. Fact-based nutrition for infants and lactating mothers—The NUTRISHIELD study. Frontiers in Pediatrics 2023, 11, 1130179. https://doi.org/10.3389/fped.2023.1130179

Índice de impacto (JCR 2021): 3.569 (Q2-Pediatrics)

Artículo 2

V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, M. Gormaz, M. Vento, M. Karipidou, D. Poulimeneas, E. Mamalaki, E. Bathrellou, J. Kuligowski. Joint Microbiota Activity and Dietary Assessment through Urinary Biomarkers by LC-MS/MS. Nutrients 2023, 15, 1894. https://doi.org/10.3390/nu15081894

Índice de impacto (JCR 2021): 6.706 (Q1-Nutrition & dietetics).

Artículo 3

V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, Á. Solaz-García, I. Lara-Cantón, A. Pinilla-Gonzalez, M. Gormaz, M. Vento, J. Kuligowski, G. Quintás. GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers. Clinica Chimica Acta 2022, 532, 172–180. https://doi.org/10.1016/j.cca.2022.05.005

Índice de impacto (JCR 2021): 6.315 (Q1- Medical laboratory technology).

Artículo 4

V. Ramos-Garcia, I. Ten-Doménech, M. Vento, C. Bullich-Vilarrubias, M. Romaní-Pérez, Y. Sanz, A. Nobili, M. Falcone, M. Di Stefano, G. Quintás, J. Kuligowski. Fast profiling of primary, secondary, conjugated and sulphated bile acids in human urine and murine feces samples.

Analytical and Bioanalytical Chemistry 2023. https://doi.org/10.1007/s00216-023-04802-8

Índice de impacto (JCR 2021): 4.478 (Q2- Biochemical research methods; Q2- Chemistry, analytical).

Artículo 5

Ten-Doménech, **V. Ramos-Garcia**, J.D. Piñeiro-Ramos, M. Gormaz, A. Parra-Llorca, M. Vento, J. Kuligowski, G. Quintás. Current Practice in Untargeted Human Milk Metabolomics. Metabolites 2020, 10, E43. https://doi.org/10.3390/metabo10020043

Índice de impacto (JCR 2020): 4.932 (Q2- Biochemistry & molecular biology).

Artículo 6

Ten-Doménech, **V. Ramos-Garcia**, M. Moreno-Torres, A. Parra-Llorca, M. Gormaz, M. Vento, J. Kuligowski, G. Quintás. The effect of Holder pasteurization on the lipid and metabolite composition of human milk. Food Chemistry 2022, 384, 132581. https://doi.org/10.1016/j.foodchem.2022.132581

Índice de impacto (JCR 2021): 9.231 (Q1- Chemistry, applied; Q1- Food science & technology; Q1 - Nutrition & dietetics).

Artículo 7

V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, M. Gormaz, A. Parra-Llorca, A.P. Shephard, P. Sepúlveda, D. Pérez-Guaita, M. Vento, B. Lendl, G. Quintás, J. Kuligowski. ATR-FTIR spectroscopy for the routine quality control of exosome isolations. Chemometrics and Intelligent Laboratory Systems 2021, 217, 104401. https://doi.org/10.1016/j.chemolab.2021.104401

Índice de impacto (JCR 2021): 4.175 (Q2- Automation & control systems; Q2- Chemistry, analytical; Q2 – Computer science, artificial intelligence; Q2 – Instruments & instrumentation; Q1 – Mathematics, interdisciplinar applications; Q1 – Statistics & probability).

Content

Resumen global

I. Introducción

La medición de la exposición a la dieta es crucial para determinar las asociaciones entre la ingesta de alimentos y el estado de salud. Además, optimizar el asesoramiento nutricional para grupos de población específicos, como son los pacientes con enfermedades crónicas, niños en edad de crecimiento, madres embarazadas y lactantes, deportistas, etc., se ha convertido recientemente en un gran desafío [1]. La identificación de biomarcadores en muestras biológicas que sean capaces de caracterizar distintos patrones dietéticos ha impulsado el desarrollo de una serie de técnicas analíticas destinadas a este fin, como son las técnicas cromatográficas, espectrométricas y espectrofotométricas, entre otras. Sin embargo, el análisis de muestras biológicas no es sencillo debido a la complejidad de las matrices, las cuales requieren una cuidadosa preparación de las muestras, una amplia gama de técnicas analíticas especializadas y una interpretación compleja de los resultados. Es por ello que los enfoques "ómicos" han arrojado luz en el estudio integral del metaboloma y lipidoma de diferentes fluidos biológicos, que incluyen los productos intermedios y finales del metabolismo.

La metabolómica y la lipidómica son dos de las ciencias "ómicas" más recientes que se definen como el estudio sistemático del conjunto completo de metabolitos de bajo peso molecular (típicamente < 1500 Da) y especies lipídicas, respectivamente, presentes en un sistema biológico. Cualquier perfil metabólico detectable en un fluido biológico es causado por la interacción

entre la expresión génica, el medio ambiente y el propio metabolismo. El enfoque de la metabolómica y la lipidómica ofrece la posibilidad de identificar variaciones en el perfil de metabolitos y lípidos que pueden usarse para la búsqueda de nuevos biomarcadores, la realización de estudios longitudinales y la discriminación de enfermedades, entre otros. Sin embargo, la generación, el análisis y la integración de los grandes y complejos conjuntos de datos obtenidos en estos estudios van de la mano de diferentes desafíos, tales como la variedad de compuestos estructuralmente heterogéneos presentes conjuntamente en concentraciones que varían cubriendo varios órdenes de magnitud, los pasos preanalíticos relacionados con el muestreo, el almacenamiento y el preprocesamiento de las muestras y la diversidad de plataformas actualmente disponibles. La resonancia magnética nuclear (RMN), la cromatografía de gases (GC, del inglés *gas chromatography*), la cromatografía líquida (LC, del inglés *liquid chromatography*) y la electroforesis capilar (CE, del inglés *capillary electrophoresis*) acopladas a espectrometría de masas (MS, del inglés *mass spectrometry*) y MS en tándem (MS/MS) son las técnicas primordiales para determinar el estado metabólico de un organismo en un momento específico [1]. Además, la espectroscopía infrarroja por transformada de Fourier con reflectancia total atenuada (ATR-FTIR, del inglés *attenuated total reflectance - fourier transform infrared*) es una herramienta emergente en el campo biomédico, ya que se basa en la adquisición rápida del espectro infrarrojo de la muestra sin necesidad de un tratamiento previo y sin utilizar reactivos o consumibles tóxicos y costosos.

Los análisis metabolómicos y lipidómicos se pueden clasificar en dirigidos, semidirigidos y no dirigidos. El **análisis dirigido** se centra en la determinación y cuantificación de un número limitado de metabolitos (típicamente decenas), los cuales son previamente seleccionados por el analista. En un análisis dirigido, la concentración de cada metabolito seleccionado se obtiene a partir

de la preparación de la correspondiente curva de calibrado con el reactivo patrón. En este escenario, y con el fin de garantizar la fiabilidad de los resultados, es importante validar la metodología en base a diferentes parámetros como son la precisión, la exactitud, la sensibilidad y la selectividad, entre otros. Para ello, existen diferentes guías de validación de métodos, como p. ej. la guía de la FDA estadounidense (FDA, del inglés *food and drug administration*) [2] o la guía de la EMA europea (EMA, del inglés *european medicines agency*) [3]. Esta aproximación permite, por tanto, evaluar el comportamiento de un grupo específico de compuestos en la muestra problema bajo unas condiciones determinadas. El **análisis semidirigido** parte de un marco de referencia preestablecido, como lo es un listado de compuestos de interés, normalmente cientos de ellos, de los cuales sus estándares no están comercialmente disponibles o su total adquisición supondría un coste demasiado elevado, pero se dispone de información sobre su comportamiento químico (p.ej. transición, tiempo de retención). Los resultados de un análisis semidirigido se pueden obtener tanto en unidades de área o área relativa, como mediante la extrapolación en un calibrado realizado con un compuesto estructuralmente similar. Además, este tipo de análisis se puede llevar a cabo tanto en equipos de alta como de baja resolución.

La metabolómica y lipidómica **no dirigida** se centran en el análisis de la huella metabólica y lipídica, intentando captar el mayor número de metabolitos posibles presentes en la muestra, sin cuantificarlos. La principal dificultad asociada a los estudios no dirigidos es el procesado del gran volumen de datos generado, el cual implica varias etapas, además de ser necesaria una correcta estrategia de análisis de acuerdo con la plataforma analítica seleccionada. Como resultado de este tipo de análisis se obtiene una tabla de dos entradas, conocida como tabla de picos, la cual recoge las áreas del conjunto de variables extraídas en cada muestra analizada [4,5]. Dada su

complejidad, el análisis e interpretación de los datos se basa en técnicas de reconocimiento de patrones, análisis discriminante, métodos univariantes con tasa de descubrimientos falsos (FDR, de inglés *false discovery rate*) o como entrada en análisis de rutas, entre otros [6]. Este enfoque es más útil para la generación de nuevas hipótesis y en el descubrimiento de biomarcadores diagnósticos o patrones metabólicos y lipídicos específicos de estados fisiológicos, entre otros. En la **Tabla 1** se reúnen los parámetros habitualmente empleados según el tipo de análisis lipidómico y metabolómico.

Tabla 1. Parámetros habitualmente empleados según el tipo de análisis lipidómico y metabolómico.

Parámetro	**Tipo de análisis**		
	Dirigido	**Semidirigido**	**No dirigido**
Normativas aplicables	ICH FDA EMA EP	-	Sugerencias de anotación [7–9]
Hipótesis	Basado en hipótesis	Basado en hipótesis	Sin hipótesis / Generación de hipótesis
Procesado de muestra	SPE Derivatización Extracción	SPE Derivatización Extracción	Dilución Extracción
Unidades	Concentración	Concentración Área/área relativa	Área/área relativa
Número de analitos	~10^1	~10^2	~10^3
Sensibilidad	nM-µM	nM-µM	µM
Rango lineal	10^3-10^5	$\leq 10^5$	$\leq 10^3$
Exactitud de masa	> 20 ppm	Depende del equipo	< 20 ppm
Identificación / Anotación	Patrón analítico	RT y transición de masa	Librería (RT, MS^n, CCS)

Nota: SPE, extracción en fase sólida (del inglés, *solid phase extraction*); ICH (del inglés, i*nternational council for harmonisation*); FDA (del inglés, f*ood and drug administration*); EMA (del inglés, e*uropean medicines agency*); EP, farmacopea europea (del inglés, e*uropean pharmacopoeia*); RT (del inglés, *retention time*); MSn (del inglés, *sequential mass spectrometry*); y CCS (del inglés, *collision cross section*).

El análisis metabolómico y lipidómico en la diada madre-recién nacido prematuro (PI, del inglés *preterm infant*) presenta particularidades que

lo hacen extremadamente delicado y susceptible. Las dificultades a las que se enfrentan la mayoría de madres de PI, hacen que la participación en estudios clínicos sea una sobrecarga y, por tanto, la tasa de reclutamiento se vea comprometida. Además, la obtención de muestras de PI, a diferencia de los adultos, no es trivial; pudiéndose llegar a ser una ardua tarea, especialmente si se requieren procedimientos invasivos o si derivado de la prematuridad tiene lugar un ingreso en la unidad de cuidados intensivos neonatal.

Durante las últimas décadas, la incidencia de partos prematuros (<37 semanas de gestación) ha ido en constante aumento [10], así como la supervivencia de los PI gracias a los avances en los cuidados y las intervenciones médicas. Según la Organización Mundial de la Salud (OMS), más de 1 de cada 10 partos son prematuros, siendo la prematuridad la principal causa de muerte en niños menores de cinco años, mientras que el 80% de los supervivientes se enfrentan a secuelas duraderas, que incluyen discapacidades de aprendizaje, visuales y auditivas, entre otras [11,12].

La nutrición infantil temprana es uno de los factores más importantes en la mejora de los resultados clínicos de los PI [13]. En este sentido, la lactancia materna es el estándar de oro de la nutrición infantil, sobre todo para PI, ya que se ha asociado a una reducción del riesgo a sufrir diferentes enfermedades como pueden ser la enterocolitis necrotizante o la sepsis [14,15] y a una mejora en el crecimiento y el desarrollo cognitivo [16,17]. Es por todo ello que la OMS recomienda encarecidamente la leche materna (HM, del inglés *human milk*) para todos los recién nacidos de hasta seis meses, nacidos tanto a término (TI, del inglés *term infants*) [18,19], como pretérmino [20]. La HMno solo contiene todos los micro y macronutrientes necesarios para cubrir las necesidades del recién nacido, sino que también contiene varios componentes bioactivos, como inmunoglobulinas secretoras, células

inmunológicas (como los leucocitos) y factores antimicrobianos (como lisozimas, oligosacáridos y péptidos) (**Figura 1**), los cuales contribuyen al óptimo crecimiento y desarrollo del lactante [17]. Además de todos estos componentes, la HM también es una fuente rica en vesículas extracelulares (EVs, del inglés *extracellular vesicles*) [21,22], pequeños cuerpos vesiculares endocíticos (30–200 nm) que tienen importantes funciones biológicas mediando la comunicación célula-célula entre células distantes [23], ya que estas HM-EVs pueden encapsular diferentes tipos de cargas, como proteínas, ARNm y microARNs [24].

Figura 1. Composición de la HM y los factores que la modifican.

Existen muchos factores que afectan a la composición de la HM (**Figura 1**), como son la edad gestacional, la edad postnatal, la nutrición, el ritmo circadiano, el estilo de vida (ejercicio, tabaco, alcohol, etc.), algunas enfermedades, el uso de medicamentos y suplementos alimenticios, y la exposición a contaminantes [17,25]. Además, el método de extracción, manejo y almacenamiento de la HM también puede influir en su contenido [25].

En ocasiones, las madres de PI no producen volúmenes de HM suficiente para alimentar a su bebé. En estos casos, la leche de madres donantes (DHM, del inglés *donor human milk*) es la alternativa preferida frente a la leche de fórmula para PI [26,27]. Para garantizar la seguridad biológica de la DHM, los bancos de leche emplean rutinariamente la pasteurización de Holder (HoP, del inglés *Holder pasteurization*), ya que ésta destruye o inactiva las posibles bacterias patógenas y virus presentes [28]. Durante la HoP, la HM se calienta a 62.5 ºC durante 30 minutos, seguido de un enfriamiento rápido a 4 ºC. Debido a este proceso, la composición de la HM puede verse afectada, ya que algunos compuestos pueden degradarse [17,29]. Estudios previos demostraron que el contenido total de carbohidratos y proteínas se mantiene relativamente estable, mientras que el contenido total de grasas parece verse notablemente afectado [28,30,31]. Además, varios componentes bioactivos, como enzimas, inmunoglobulinas o células inmunitarias, también se ven comprometidos o destruidos durante este proceso [30,32].

Además de las adaptaciones fisiológicas, la investigación se ha centrado en el potencial efecto de la dieta materna sobre la composición de la HM. En una reciente revisión bibliográfica de más de cien estudios, se llegó a concluir que la composición de la HM se relaciona con el consumo materno de ácidos grasos, vitaminas liposolubles y la ingesta de vitamina B1 y C [33,34]. En este contexto, los datos de ingesta dietética se recopilan mayoritariamente utilizando herramientas basadas en autoinformes, como cuestionarios de frecuencia alimenticia (FFQ, del inglés *food frequency questionnaire*) para la evaluación del consumo habitual y diarios de alimentos o recordatorios de 24 horas (R24h, del inglés *24 h recall*) para la evaluación del consumo a corto plazo. Sin embargo, dichos métodos son propensos a errores e inexactitudes debido a su naturaleza subjetiva [35]. Por otra parte, la

influencia que la dieta de la madre tiene sobre los constituyentes no nutritivos de la HM, como por ejemplo, el contenido de prebióticos [36] o la microbiota [34,37] ha sido menos estudiada, y por regla general, se ha centrado en TI.

Asimismo, la dieta es uno de los factores clave que intervienen en la conformación del microbioma intestinal, afectando la diversidad microbiana, así como la abundancia de microbios específicos [38–40]. Los productos finales metabólicos de los micro y macronutrientes dietéticos, como los ácidos biliares (BA, del inglés *bile acids*) (por ejemplo, ácidos cólico y litocólico), ácidos grasos de cadena corta (SCFA, del inglés *short chain fatty acids*) (por ejemplo, ácido acético y butírico), indoles y derivados polifenólicos (por ejemplo, ácido 3-indolpropiónico y ácido hipúrico), y ácidos fenólicos (por ejemplo, ácido gálico y ácido ferúlico) se utilizan para evaluar la actividad del microbioma intestinal [41–45].

Es por ello que se hace necesario tanto el estudio la influencia de la HoP y la dieta materna sobre la composición de la HM en beneficio último del PI, como el desarrollo de métodos analíticos para llevar a cabo una evaluación conjunta de la nutrición y la actividad microbiana.

II. Hipótesis y objetivos

En la presente Tesis doctoral, la hipótesis que se propone es que el desarrollo de metodologías metabolómicas y lipidómicas puede contribuir en la evaluación nutricional tanto de los PI como de sus madres. Asimismo, se han propuesto tres objetivos principales:

- Evaluar las asociaciones entre la dieta materna y los biomarcadores de nutrición y de la actividad microbiana en el contexto de la prematuridad
- Evaluar el efecto de la HoP y el almacenamiento en la composición de la DHM.
- Optimizar el aislamiento y control de calidad de HM-EVs, así como evaluar su composición lipídica.

Los artículos que conforman la presente Tesis doctoral por compendio de publicaciones se han dividido en tres secciones. La Sección I se titula "*Biomarkers relevant in nutrition and microbiota studies in the field of neonatology*" e incluye cuatro capítulos: i) la descripción del estudio NUTRISHIELD, enfocado en la evaluación de la nutrición de los PI y sus madres lactantes; ii) la evaluación conjunta de la actividad microbiana y la dieta a través de biomarcadores en orina mediante LC-MS/MS; iii) la evaluación de la actividad microbiana en muestras de orina y heces de PI, TI y sus madres, mediante GC-MS; y iv) el análisis de biomarcadores de la actividad microbiana en muestras de orina humana y heces de ratón mediante LC-MS/MS.

La Sección II se titula "*Effect of pasteurization on the metabolome and lipidome of HM*" e incluye dos capítulos: i) la revisión bibliográfica de las metodologías actuales para el análisis metabolómico no dirigido en muestras de HM; y ii) la evaluación del efecto de la HoP sobre la composición metabolómica y lipidómica de la DHM mediante LC-MS/MS y GC-MS.

La Sección III se titula "*Novel analytical tools for the characterization of the molecular composition of HM-EVs*" e incluye tres capítulos: i) la descripción del protocolo a seguir para el aislamiento de HM-EVs y su posterior análisis lipidómico no dirigido mediante LC-MS/MS; ii) la

evaluación del perfil lipidómico de HM-EVs mediante LC-MS/MS y evaluación de las diferentes estrategias de normalización; y iii) el desarrollo de un método analítico mediante ATR-FTIR para el control de calidad rutinario del procedimiento de extracción de HM-EVs.

III. Metodología

i. Diseño de estudio

Todas las muestras empleadas en la presente Tesis doctoral se han obtenido de tres estudios clínicos que fueron aprobados por los correspondientes comités éticos y todos los métodos se realizaron de acuerdo con las directrices y normativas pertinentes. Se obtuvieron consentimientos informados por escrito de los participantes o de sus representantes legales antes de la recogida de la muestra y del análisis de la información clínica y demográfica. En la **Figura 2** se describe el diseño de estudio llevado a cabo en cada sección.

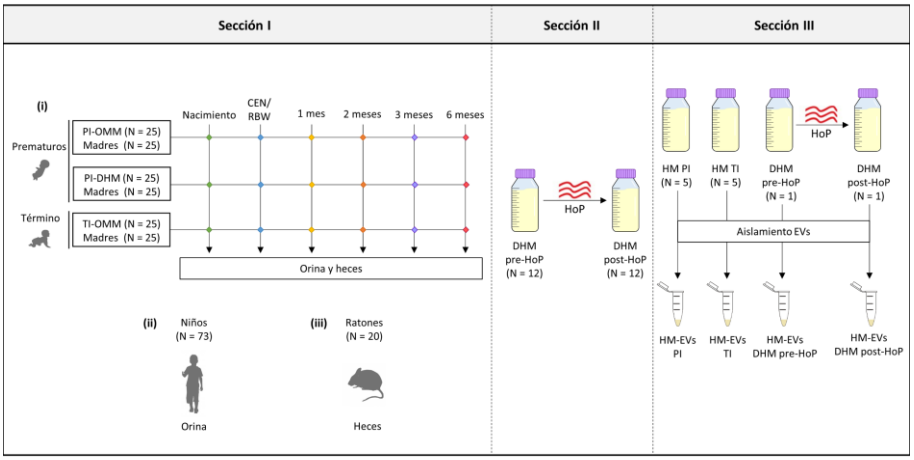


Figura 2. Diseño de los estudios llevados a cabo en cada sección. Nota: CEN, nutrición enteral completa (150 mL/kg/día) (del inglés, *complete enteral nutrition*); RBW, recuperación del peso de nacimiento (del inglés, *recovery of birth weight*); OMM, leche de la propia madre (del inglés, *own mother's milk*); DHM, leche de donante (del inglés, *donor human milk*); HoP, pasteurización de Holder (del inglés, *Holder pasteurization*); HM, leche materna (del inglés, *human milk*); PI, recién nacido prematuro (del inglés, *preterm infant*) y TI, recién nacido a término (del inglés, *term infant*); EVs, vesículas extracelulares (del inglés, *extracelular vesicles*).

En la Sección I, se emplearon tres conjuntos de muestras: (i) muestras del proyecto NUTRISHIELD aprobado por el Comité de Ética de la Investigación con medicamentos (CEIm #2019-289-1) (https://nutrishield-project.eu/), un estudio prospectivo, observacional, de cohortes (NCT05646940) realizado en el Servicio de Neonatología del Hospital Universitario y Politécnico La Fe (Valencia, España), que incluye tres grupos de estudio: PI <32 semanas de gestación que reciben exclusivamente (es decir,

>80% de la ingesta total) leche de su propia madre (OMM, del inglés *own mother's milk*), PI que reciben exclusivamente DHM, y TI que reciben exclusivamente OMM, así como sus madres. Se recogieron muestras de orina y heces a seis tiempos, cubriendo el período desde el nacimiento hasta los seis meses de edad del bebé; (ii) muestras de orina de niños reclutados en el Departamento de Pediatría del Hospital San Raffaele de Milán (Italia) de un estudio aprobado por el Comité Ético Institucional del Instituto Científico IRCCS San Raffaele (NUTRI-T1D, 2019); y (iii) muestras de heces de ratones macho CC57BL/6J de un estudio conforme a las directrices de la Unión Europea 2010/63/UE y España RD53/2013 y aprobado por el comité de ética de la Universidad de Valencia (Sección de Producción Animal, SCSIE, Universidad de Valencia, España) y autorizado por la Dirección General de Agricultura, Ganadería y Pesca (Generalitat Valenciana) (2021/VSC/PEA/0273).

En la Sección II, se utilizaron muestras de DHM (N = 12) antes y después de pasteurizar pertenecientes al estudio PREMILK (CEIm #2014/0247).

En la Sección III, se emplearon muestras de HM-EVs extraídas de HM de madres de PI (N = 5) y TI (N = 5) y una muestra de DHM antes y después de pasteurizar preparada a partir de alícuotas de 20 muestras de DHM individuales pertenecientes al estudio CALMNESS (CEIm #2020-052-1).

ii. Recogida de muestras biológicas

La recogida de las muestras ha seguido un procedimiento operativo estándar (SOP, del inglés *standard operating procedure*) para cada matriz,

como se indica en la **Figura 3**. La primera orina de la mañana y las heces de las madres se recogieron en un recipiente de polipropileno, mientras que las muestras de orina de los bebés se recogieron colocando almohadillas de algodón estériles en el pañal que, después de orinar, se escurrieron con una jeringa estéril en tubos estériles. Las muestras de heces del bebé se recogieron directamente del pañal en recipientes estériles empleando pinzas estériles.

Respecto a las muestras de OMM y DHM, las madres se extrajeron la leche utilizando un sacaleches y siguiendo un procedimiento estandarizado empleado de forma rutinaria en el Banco de Leche Materna de la Comunidad Valenciana. Las partes extraíbles del sacaleches, así como los biberones de recolección, se esterilizaron antes de su uso y se dieron las siguientes pautas para la extracción: i) la extracción de HM se lleva a cabo al menos tres horas después de la última toma del bebé; ii) preferiblemente entre las 7 y las 10 de la mañana; iii) extracción completa de un pecho; y iv) se registran los detalles sobre el método de extracción (la marca del extractor de leche, etc.), la fecha y el volumen extraído. En el caso de las muestras de DHM, se recogieron alícuotas de 25-50 mL antes y después de la HoP.

Todas las muestras se dividieron en alícuotas para evitar los ciclos de congelación y descongelación y se almacenaron a -80 ºC hasta su uso.

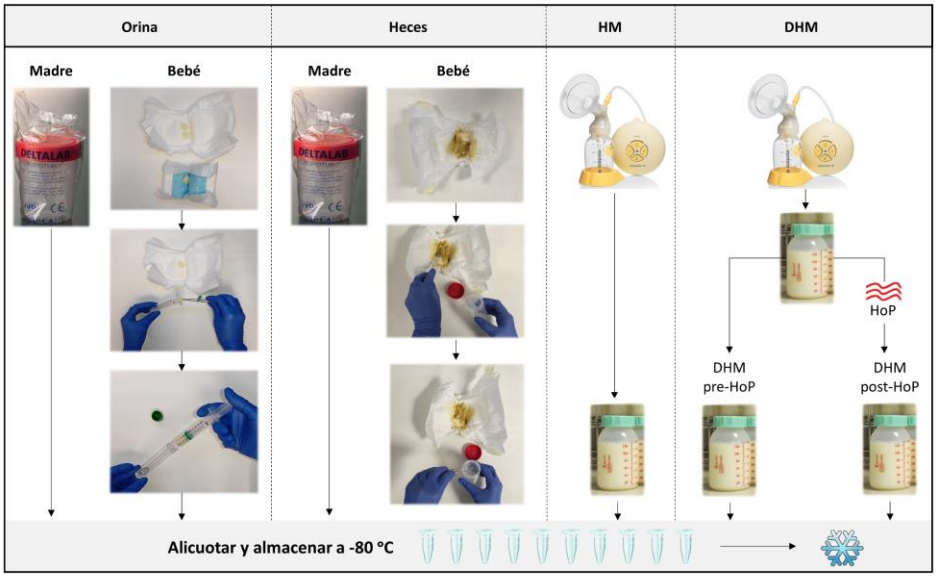


Figura 3. Procedimientos de recogida de muestras. Nota: DHM, leche de donante (del inglés, *donor human milk*); HoP, pasteurización de Holder (del inglés, *Holder pasteurization*) y HM, leche materna (del inglés, *human milk*).

iii. Aislamiento de HM-EVs

En la **Figura 4** se indica el procedimiento para la extracción de HM-EVs, el cual consiste en centrifugar consecutivamente la HM para eliminar los glóbulos de grasa y precipitar las proteínas. Tras filtrar el sobrenadante resultante, éste se somete a dos ultracentrifugaciones seriadas (10000 rpm) para sedimentar las proteínas restantes y, tras otra filtración, a tres ultracentrifugaciones seriadas (30000 rpm) para precipitar y lavar las EVs. Las HM-EVs aisladas se almacenaron a -80 ºC hasta su uso.

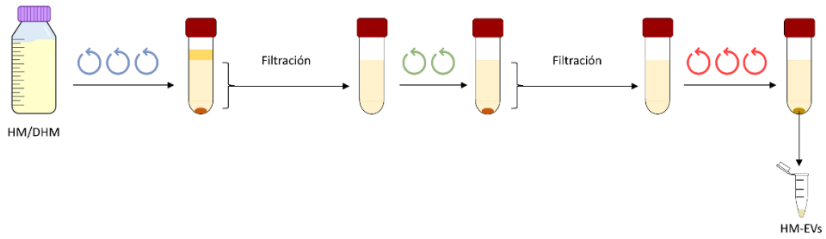


Figura 4. Procedimiento de extracción de HM-EVs. Nota: centrifugación; ultracentrifugación (10000 rpm), ultracentrifugación (30000 rpm); DHM, leche de donante (del inglés, *donor human milk*); HM, leche materna (del inglés, *human milk*); EVs, vesículas extracelulares (del inglés, *extracelular vesicles*).

iv. Preparación y análisis de las muestras

Con el objetivo de compatibilizar las diferentes técnicas analíticas empleadas en esta Tesis doctoral con la complejidad de las muestras biológicas, se han llevado a cabo diferentes preparaciones de muestra, tal y como se muestra en la **Figura 5**. Para los análisis mediante LC-MS, la dilución y/o desproteinización de la muestra fue suficiente para reducir la matriz, sin embargo, los análisis mediante GC-MS requirieron además una derivatización de los analitos para aumentar la volatilidad de los mismos.

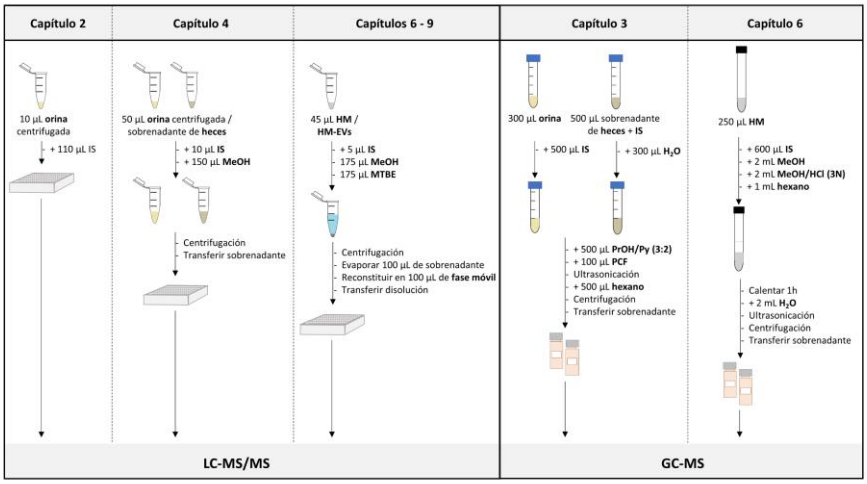


Figura 5. Procesados de muestra llevados a cabo en cada capítulo. Nota: IS, patrón interno (del inglés, *internal standard*); MeOH, metanol; MTBE, metil tert-butil éter; PrOH, propanol; Py, piridina y PCF, propil cloroformiato.

En la presente Tesis doctoral, se han abordado los tres tipos de metodologías (dirigida, semidirigida y no dirigida), todas ellas resumidas en la **Tabla 2**.

Tabla 2. Metodologías utilizadas en los diferentes capítulos.

Sección	Capítulo	Metodología	Análisis	Técnica	Matriz
I. Biomarkers relevant in nutrition and microbiota studies in the field of neonatology	2. Joint microbiota activity and dietary assessment through urinary biomarkers by LC-MS/MS	Dirigido y Semidirigido	Biomarcadores de nutrición y de la actividad microbiana	LC-MS/MS	Orina
	3. GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers	Dirigido	Biomarcadores de la actividad microbiana	GC-MS	Orina Heces
	4. Fast profiling of primary, secondary, conjugated and sulphated bile acids in urine and faeces samples	Dirigido	Biomarcadores de la actividad microbiana	LC-MS/MS	Orina Heces
II. Effect of pasteurization on HM	6. The effect of Holder pasteurization on the lipid and metabolite composition of human milk	No dirigido y dirigido	Análisis metabolómico, lipidómico y de ácidos grasos	LC-MS/MS GC-MS	HM
III. Novel analytical tools for the characterization of the molecular composition of HM-EVs	8. The lipidomic profile of human milk extracellular vesicles from mothers of term and preterm infants	No dirigido	Análisis metabolómico y lipidómico	LC-MS/MS	HM-EVs
	9. ATR-FTIR spectroscopy for the routine quality control of exosome isolations	No dirigido y semidirigido	Análisis metabolómico y lipidómico	LC-MS/MS ATR-FTIR	HM-EVs

a. Análisis dirigidos y semidirigidos

En el análisis de biomarcadores de nutrición y actividad microbiana por LC-MS/MS (Capítulo 2), la muestra de orina se ha diluido con patrón interno (IS, del inglés *internal standard*) y se ha inyectado en el sistema de LC-MS/MS QTRAP 6500+ de la casa Sciex (Framingham, EE. UU.). La separación cromatográfica se realizó mediante una columna en fase reversa Luna Omega Polar C18 equipada con un cartucho de protección de seguridad C18 completamente poroso de la casa Phenomenex (Torrance, EE. UU.) y la posterior detección de MS se realizó usando el modo de monitorización de reacciones múltiples (MRM, del inglés *multiple reaction monitoring*).

En el caso del análisis de biomarcadores de la actividad microbiana mediante GC-MS (Capítulo 3), se realizó una derivatización con propil cloroformiato para obtener los respectivos ésteres propílicos y, tras una extracción con hexano, las muestras se inyectaron en un sistema GC 7890B acoplado a un MS de cuadrupolo simple 5977A. La columna capilar utilizada para la separación cromatográfica fue una HP-5 MS, todo ello de la casa Agilent (Santa Clara, California, US). La adquisición de los datos del espectrómetro de masas se realizó mediante monitorización selectiva de iones (SIM, del inglés *selected ion monitoring*).

Para el análisis de biomarcadores de la actividad microbiana mediante LC-MS/MS (Capítulo 4), las muestras se diluyeron con IS y se precipitaron las proteínas mediante la adición de metanol. Tras una incubación en frío y una centrifugación, los sobrenadantes se inyectaron en un sistema UPLC-MS/MS ACQUITY UPLC - Xevo TQS de la casa Waters (Milford, Massachusetts, US). La separación cromatográfica tuvo lugar en una columna

en fase reversa Acquity BEH C18 y la detección por MS se realizó mediante MRM.

En el método desarrollado para el estudio de los ácidos grasos en muestras de OMM y DHM mediante GC-MS (Capítulo 6), estos se derivatizaron a sus respectivos ésteres metílicos con ácido clorhídrico metanólico. Una vez extraídos con hexano, se inyectaron en el sistema GC 7890B acoplado a un MS de cuadrupolo simple 5977A de la casa Agilent utilizando una columna Zebron™ ZB-WAXplus™ de la casa Phenomenex. La adquisición de los datos del espectrómetro de masas se realizó mediante SIM.

En todos los análisis dirigidos la cuantificación se llevó a cabo mediante calibrados externos con IS y se validaron siguiendo la guía de la FDA [2], que consistió en comprobar los parámetros de precisión, exactitud, rango de linealidad, contaminación por arrastre, selectividad, especificidad y estabilidad de los patrones y muestras. Previo al análisis estadístico de los resultados, se realizó un control de calidad mediante el cual se sustituyeron los valores inferiores al límite de cuantificación por la mitad del valor del límite de detección. En los análisis semidirigidos, conociendo las transiciones de masa de los iones y el tiempo de retención, se utilizaron las áreas relativas, es decir, el área del analito normalizado por el área del IS.

b. Análisis no dirigidos

Para el análisis de metabolómica y lipidómica no dirigida de las muestras de HM y HM-EVs (Capítulos 6- 9), se realizó una dilución con IS, seguido de una extracción líquido-líquido de una única fase (*single phase*) con

metanol y metil tert-butil éter para favorecer la extracción de lípidos y otros metabolitos polares. El sobrenadante resultante de la centrifugación se llevó a sequedad y se reconstituyó en fase móvil (FM) para ser inyectado en el sistema UPLC1290 Infinity acoplado a un sistema de MS de cuadrupolo tiempo de vuelo (Q-TOF, del inglés *quadrupole time-of-flight*) iFunnel 6550 de la casa Agilent (Santa Clara, EE. UU.). La separación cromatográfica para la metabolómica se llevó a cabo empleando una columna en fase reversa Synergi^{TM} Hydro-RP 80 Å LC C18 de la casa Phenomenex (Torrance, CA, USA), mientras que para la lipidómica se utilizó una columna en fase reversa Acquity BEH C18 de la casa Waters (Milford, EE. UU.) y una FM específica para la detección de lípidos que contiene un 98% de FM A (5:1:4 isopropanol:metanol:agua 5 mM acetato de amonio, 0.1% v/v ácido fórmico) y un 2% de FM B (99:1 isopropanol:agua 5 mM acetato de amonio, 0.1% v/v ácido fórmico)). Los datos de MS de escaneo completo (*full scan*) se adquirieron entre 70 y 1500 *m/z* con ionización por electropulverización (*electrospray*) tanto positiva como negativa (ESI, del inglés *electrospray ionization*). Se preparó una muestra de control de calidad (QC, del inglés *quality control*) mezclando una alícuota de cada extracto de muestra con el fin de representar el conjunto de muestras. Se inyectaron varios QC al comienzo de la secuencia para el acondicionamiento del sistema y la adquisición de datos de MS^2, y a lo largo de la secuencia cada seis muestras para la evaluación y corrección de la intensidad de la señal.

Tras la adquisición de los datos, se utilizó el software XCMS (https://xcmsonline.scripps.edu/) para extraer y alinear los picos detectados, basándose en la precisión y exactitud del tiempo de retención y la *m/z* de los IS. Tras la detección, deconvolución e integración de los datos, se obtuvieron las tablas de picos. Posteriormente se utilizó el software matemático MATLAB de MathWorks (Natrick, EE.UU), para llevar a cabo la anotación

de los picos, la corrección del efecto lote utilizando los QCs con la regresión de vectores de soporte (SVR, del inglés *support vector regression*), y el filtrado de los datos con los blancos y QCs.

c. Otras técnicas

Finalmente, para llevar a cabo el análisis del perfil espectral de HM-EVs mediante ATR-FTIR, se depositaron únicamente 2 µL sobre el cristal ATR y, tras secarlos con una corriente de aire a temperatura ambiente, los espectros de IR se adquirieron en el rango de 4000 a 400 cm^{-1} utilizando un espectrómetro Alpha II equipado con un ATR de platino con un diamante monolítico como elemento de interfaz de medición, una fuente de infrarrojos CenterGlow™ y un detector DTGS (del inglés, *deuterated alanine doped tri-glycine sulphate*) de la casa Bruker (Ettlingen, Alemania).

v. Análisis estadístico

Tal y como se indica en la **Tabla 3**, los métodos estadísticos univariantes utilizados a lo largo de la presente Tesis doctoral son las pruebas de *t* de Student y de los rangos con signo de Wilcoxon, y los coeficientes de correlación de Pearson y Spearman, en caso necesario, con corrección de FDR. Los métodos estadísticos multivariantes de análisis de agrupamiento jerárquico (HCA, del inglés *hierarchical clustering analysis*), componentes principales (PCA, del inglés *principal component analysis*) y heterospectroscopia estadística (SHY, del inglés *statistical*

heterospectroscopy) son los que se han utilizado. Respecto a las herramientas bioinformáticas para en análisis de redes, se ha utilizado el algoritmo de *mummichog* de MetaboAnalyst 5.0 [46], la herramienta *web-based ontology enrichment tool for lipidomic analysis: Lipid Ontology analysis* (LION) (www.lipidontology.com) [47] y el algoritmo *prize-collecting Steiner forest algorithm for integrative analysis of untargeted metabolomics* (PIUMet) (http://fraenkel-nsf.csbi. mit.edu/piumet2/) [48].

Tabla 3. Métodos estadísticos y herramientas bioinformáticas utilizadas en los diferentes capítulos.

Método estadístico	Tipo	Capítulo	Finalidad
Prueba *t* de Student ($\alpha = 0.05$)	Univariante	2, 3 y 4	Comparar el efecto de los ciclos de congelación en las muestras biológicas
		6	Comparar las composición metabólica y lipídica en las muestras de DHM antes y después de HoP
		8	Comparar el efecto de la HoP sobre los lípidos en HM-EVs
Prueba de los rangos con signo de Wilcoxon ($\alpha = 0.05$)	Univariante	2	Comparar la ingesta de grupos de alimentos entre madres lactantes
		3	Comparar los niveles de biomarcadores de la actividad microbiana en muestras de orina y heces de PI, TI y sus madres
		6	Comparar los niveles de ácidos grasos en muestras de DHM antes y después de HoP
		8	Comparar la concentración, el tamaño y el contenido total de proteínas en HM-EVs de madres de PI y TI

Tabla 3 (continuación). Métodos estadísticos y herramientas bioinformáticas utilizadas en los diferentes capítulos.

Método estadístico	Tipo	Capítulo	Finalidad
Coeficiente de correlación de Pearson ($\alpha = 0.05$)	Univariante	2	Evaluar las asociaciones pareadas entre concentraciones de biomarcadores de nutrición y actividad microbiana en muestras de orina
		3	Evaluar las asociaciones pareadas entre concentraciones de biomarcadores de la actividad microbiana en muestras de orina y heces
		3	Evaluar las asociaciones entre los niveles de biomarcadores de la actividad microbiana en muestras de orina y heces y la edad gestacional o postnatal
		8	Evaluar las asociaciones entre la edad gestacional y el contenido total de lípidos en HM-EVs
Coeficiente de correlación de Spearman	Univariante	2	Evaluar las asociaciones pareadas entre concentraciones de biomarcadores de nutrición y actividad microbiana en orina y la ingesta de grupos de alimentos
		8	Evaluar las asociaciones ente la edad gestacional y corregida por la edad post natal y las subclases de lípidos en HM-EVs

Tabla 3 (continuación). Métodos estadísticos y herramientas bioinformáticas utilizadas en los diferentes capítulos.

Método estadístico	Tipo	Capítulo	Finalidad
HCA	Multivariante	2	Detectar patrones de dieta en madres lactantes
		4	Detectar agrupaciones según los niveles de biomarcadores de la actividad microbiana en orina humana y heces murinas
		9	Detectar agrupaciones según los perfiles lipidómicos en muestras de HM-EVs
PCA	Multivariante	2	Detectar patrones de dieta y de actividad microbiana en madres lactantes
		3	Detectar agrupaciones según los niveles de biomarcadores de la actividad microbiana en orina y heces de PI, TI y sus madres
		4	Detectar agrupaciones según los niveles de biomarcadores de la actividad microbiana en orina humana y heces murinas
		9	Detectar agrupaciones según los perfiles lipidómicos en muestras de HM-EVs
SHY	Multivariante	9	Evaluar las asociaciones entre los espectros ATR-FTIR y los picos de UPLC-MS de cada lípido anotado en muestras de HM-EVs
Mummichog de Metaboanalyst	Análisis de redes	6	Evaluar las rutas metabólicas alteradas en muestras de DHM antes y después de HoP
LION	Análisis de redes	6	Evaluar el efecto de HoP sobre el lipidoma en muestras de DHM
PIUMet	Análisis de redes	6	Evaluar las alteraciones moleculares causadas por HoP en muestras de DHM

vi. Data availability

Los datos de los análisis dirigidos y semidirigidos de los diferentes estudios se pueden encontrar en el material suplementario de sus respectivos capítulos (Capítulo 2: https: //www.mdpi.com/article/10.3390/nu15081894/s1; Capítulo 3: https://doi. org/10.1016/j.cca.2022.05.005; Capítulo 4: https://link.springer.com/article/10.1007/s00216-023-04802-8; Capítulo 5: http://www.mdpi.com/2218-1989/10/2/43/s1; y Capítulo 6: https://doi.org/10.1016/j.foodchem.2022.132581).

Las tablas de picos extraídas de los análisis no dirigidos del Capítulo 6 están disponibles en el repositorio de datos Mendeley en https://data.mendeley.com/datasets/fnzbxmkv83/1 (lipidómica) y https://data.mendeley.com/datasets/jymtst88jm/1 (metabolómica).

Los datos sobre el experimento del Capítulo 8 están disponibles en la base de conocimiento de EV-TRACK (EV230598). El conjunto de datos estará disponible tras la publicación del manuscrito a través del repositorio de Zenodo (doi: 10.5281/zenodo.8042647).

El conjunto de datos de los análisis no dirigidos del Capítulo 9 se encuentran en el repositorio de datos Zenodo en zenodo.org/deposit/5148582.

IV. Resultados principales

En el Capítulo 2 se validó, según el criterio de la guía de la FDA, un método analítico que permite el análisis de un panel de biomarcadores de

nutrición y de la actividad microbiana en orina de madres lactantes. Se reportaron los niveles de estos metabolitos y se encontraron correlaciones entre biomarcadores de un mismo grupo de alimentos (coeficiente de correlación de Pearson, p-valor < 0.05). Además, se diferenciaron tres grupos de madres con respecto a las concentraciones de estos metabolitos y se encontraron correlaciones entre los niveles de biomarcadores de frutas, carne y pescado y los cuestionarios tradicionales de recordatorios dietéticos de 24 horas (coeficiente de correlación de Pearson, p-valor < 0.05).

El método de análisis de biomarcadores de la actividad microbiana desarrollado en el Capítulo 3, se validó con éxito según la guía de la FDA, desde la recogida de muestra hasta su análisis, revelando una contaminación muy significativa de ácido acético en las gasas y algodones que se emplean típicamente para la recogida de muestras de orina de recién nacidos. Se observó una correlación significativa entre SCFA y BCAA en muestras de heces y de orina y un total de cinco metabolitos mostraron una correlación significativa entre matrices (coeficiente de correlación de Pearson, p-valor < 0.05). Además, se obtuvieron rangos de referencia de estos metabolitos en madres, PI y TI y se encontraron diferentes patrones entre los grupos. Los bebés mostraron concentraciones significativamente más bajas de SCFA y BCAA que las madres en las heces y concentraciones más altas en la orina. Se identificaron valores de concentración ligeramente más altos en PI en comparación con TI, excepto para leucina e isoleucina en muestras fecales que fueron más altas en TI (Prueba de suma de rangos de Wilcoxon, $\alpha = 0.05$).

El método de análisis de biomarcadores de la actividad microbiana desarrollado en el Capítulo 4, se validó según el criterio de la guía de la FDA para la determinación de BA primarios, secundarios, conjugados y sulfatados en muestras de orina humana y heces murinas, quedando demostrada la

aplicabilidad del mismo en el campo (pre)clínico. Se observaron diferencias en los perfiles de concentración de BA entre ambas matrices, observándose que la mayoría de los BA presentes en muestras de orina humana corresponden a BA conjugados secundarios, seguidos de BA sulfatados, siendo estas clases las menos concentradas en muestras de heces murinas. Además, los BA encontrados en muestras fecales murinas eran en su mayoría conjugados primarios y BA primarios, siendo estas clases menos abundantes en las muestras de orina humana. Además, se identificaron dos grupos con respecto a las concentraciones de BA en muestras de ambas matrices y se llevó a cabo una comparación entre las vías metabólicas y los mecanismos de excreción de los BA urinarios humanos y fecales murinos, de acuerdo con las diferencias entre especies.

En el análisis bibliográfico llevado a cabo en el Capítulo 5, se han podido compilar los diferentes análisis metabolómicos realizados hasta la fecha en muestras de HM, con especial énfasis a la etapa de recogida de muestra, las técnicas utilizadas y los metabolitos detectados. Además, se han señalado los puntos débiles encontrados en la literatura.

En el Capítulo 6, centrado en el estudio del efecto de la HoP en la composición lipídica de la HM y DHM, se observó que los metabolitos que se ven afectados están relacionados con los esteroides (la biosíntesis de esteroides y la biosíntesis de hormonas esteroideas), así como con las rutas metabólicas de los ácidos grasos (la biosíntesis de ácidos grasos insaturados y el metabolismo del ácido linoleico). Además, los niveles de metabolitos estructuralmente relacionados se vieron afectados por el tratamiento térmico, así como los ácidos grasos, ya que las concentraciones del 76% de éstos se vieron alteradas después de la pasteurización, con una disminución significativa de la concentración relativa del 10% (Prueba de suma de rangos

de Wilcoxon, $\alpha = 0.05$). Finalmente, los resultados de este estudio también demuestran que la HoP tiene un impacto no solo en la composición, sino también en las propiedades fisicoquímicas, los componentes celulares y la funcionalidad de los lípidos.

En el Capítulo 7, se ha podido elaborar una guía para el aislamiento de HM-EVs y la posterior caracterización de su huella lipídica remarcando los aspectos prácticos relevantes.

En el Capítulo 8, en el cual los HM-EVs se caracterizaron y sometieron a análisis lipidómico mediante LC-MS, el conjunto de datos destaca la importancia de emplear una amplia gama de métodos de caracterización complementarios para todas las muestras de estudio. El aislamiento de HM-EVs puros en este estudio se vio obstaculizado por el co-aislamiento de glóbulos de grasa de leche y/o caseínas. Además, la elección de los enfoques de normalización de datos influyó en la comparabilidad de las huellas lipidómicas entre muestras. Los resultados demuestran que, cuando se normalizan las huellas lipidómicas de las HM-EVs, es fundamental seleccionar cuidadosamente un procedimiento adecuado de normalización teniendo en cuenta la cuestión científica específica a que se quiere dar respuesta, así como la pureza y la distribución del tamaño de las HM-EVs.

En el estudio del análisis de control de calidad del proceso de aislamiento de HM-EVs del Capítulo 9, se anotaron con éxito un total de 377 variables detectadas mediante LC-MS/MS, pertenecientes mayoritariamente a las clases de glicerofosfolípidos (50%), esfingolípidos (31%) glicerolípidos (17%). Se observaron asociaciones significativas entre las abundancias de las diferentes clases de lípidos en LC-MS/MS y regiones específicas de los espectros ATR-FTIR de las mismas muestras. También se evaluó la similitud entre las tendencias observadas mediante ambas técnicas, resultando ésta

significativa (prueba de Mantel, *p*-valor <0.005). Adicionalmente, se llevó a cabo el análisis cuantitativo simultáneo de proteínas y lípidos totales mediante ATR-FTIR en HM-EVs.

V. Conclusiones y trabajo futuro

En la presente Tesis doctoral, se ha llevado a cabo el desarrollo de diferentes métodos analíticos y se ha demostrado su aplicación en el ámbito clínico mediante el análisis de muestras biológicas de bebés y sus madres. Las conclusiones derivadas se resumen a continuación:

- La determinación de biomarcadores de nutrición y de la actividad microbiana en muestras de orina de madres lactantes se puede llevar a cabo mediante LC-MS/MS. Asimismo, es posible su uso como herramienta complementaria a los cuestionarios tradicionales de recordatorios dietéticos de 24 horas, presentando un gran potencial para la evaluación integral de los patrones de nutrición y su efecto en el microbioma

- El análisis simultáneo de los SCFAs y BCAAs en muestras de orina y heces se puede llevar a cabo mediante GC-MS. La aplicación de este método permite la cuantificación precisa de estos biomarcadores mediante un único procedimiento de derivatización, proporcionando información que permite distinguir entre los diferentes grupos de estudio, en este caso

PI, TI y sus madres. Además, se ha revelado una contaminación muy significativa de ácido acético en las gasas y algodones que se emplean típicamente para la recolección de muestras de orina de recién nacidos que se debe considerar en futuros estudios clínicos

- Se ha validado un método de LC-MS/MS para el análisis cuantitativo de BA en muestras de orina humana y heces murinas, demostrando la aplicabilidad del mismo en el campo (pre)clínico. Además, se ha obtenido una comparación entre las vías metabólicas y los mecanismos de excreción de los BA urinarios humanos y fecales murinos, de acuerdo con las diferencias entre especies

- Los análisis lipidómico y metabolómico no dirigido y el análisis dirigido del perfil de ácidos grasos de la DHM mediante LC-MS/MS y GC-MS, respectivamente, han demostrado que la HoP afecta significativamente tanto a la composición lipídica, como a las propiedades fisicoquímicas y la funcionalidad de la HM

- La evaluación de diferentes técnicas de normalización en el análisis lipidómico de HM-EVs ha permitido ahondar en el correcto aislamiento y caracterización de las mismas, así como en las potenciales implicaciones derivadas de la selección de la técnica de normalización en cuestión

- El control de calidad rápido y directo del procedimiento de aislamiento de HM-EVs es posible mediante los lípidos y proteínas totales derivados del perfil espectral determinado por ATR-FTIR

Además del desarrollo y la aplicación de los diferentes métodos analíticos presentados, se ha realizado un artículo descriptivo del estudio clínico del proyecto NUTRISHIELD, una revisión bibliográfica sobre las metodologías para el análisis de metabolómica no dirigida en HM y un capítulo de libro sobre el procedimiento para el correcto aislamiento y análisis lipidómico no dirigido de HM-EVs.

Finalmente, tras el desarrollo de la presente Tesis, han surgido nuevos aspectos a estudiar que se están llevando a cabo actualmente por nuestro grupo de investigación, los cuales se describen a continuación:

- El desarrollo de un algoritmo capaz de proporcionar un consejo de nutrición personalizado en base a los resultados de los biomarcadores de nutrición y de la actividad microbiana obtenidos, junto con otras variables clínicas y la composición de la microbiota intestinal y de la leche. Este algoritmo se aplicaría en consultas médicas para dar un consejo nutricional a madres lactantes con el fin de mejorar la composición de la HM, y de esta manera mejorar los resultados clínicos de los PI a corto y largo plazo

- El análisis de los BA como biomarcadores de la actividad microbiana en muestras de orina y heces en el contexto

neonatal, para estudiar las correlaciones entre ambas matrices y determinar los niveles en esta población. De esta manera, se podrá determinar si es posible utilizar los niveles en muestras de orina como biomarcadores de la actividad microbiana, facilitando así el proceso de recogida no invasiva y el tratamiento de la muestra

- El estudio de la relevancia biológica y el impacto de los cambios observados en la composición y funcionalidad de los componentes de HM tras su pasteurización, así como impulsar el desarrollo de alternativas tecnológicas a la HoP para la estabilización microbiológica de la HM

- Desarrollar y optimizar un protocolo alternativo a la ultracentrifugación para el aislamiento de los HM-EVs mediante cromatografía de exclusión por tamaños, para reducir la cantidad de muestra necesaria, las impurezas y el tiempo requerido para el aislamiento. Asimismo, evaluar las propiedades de las HM-EVs con el fin de implementar su uso como fortificante en la nutrición de PI

- Establecer el análisis ATR-FTIR para la caracterización rutinaria de proteína y lípidos totales en EVs aislados de diferentes matrices

List of abbreviations

ATR-FTIR	Attenuated Total Reflectance – Fourier Transform Infrared
BCA	Bicinchoninic acid assay
BCAA	Branched-chain amino acid
BMI	Body mass index
CE	Capillary electrophoresis
CEIm	Scientific and Ethics Committee for Biomedical Research
CE-MS	Capillary electrophoresis coupled to mass spectrometry
CI	Confident interval
COSY	Correlation Spectroscopy
DG	Diradylglycerol
DHM	Donor human milk
DOSY	Diffusion-Ordered Spectroscopy
EI	Electron impact
EML	Evidence-based metabolome library
ESI	Electrospray ionization

EV	Extracellular vesicle
FA	Formic acid
FAME	Fatty acid methyl ester
FA	Fatty acid
FDA	Food and Drug Administration
GC	Gas chromatography
GC-MS	Gas chromatography coupled to mass spectrometry
GPI	Glycosylphosphatidylinositol
GP	Glycerophospholipid
HCA	Hierarchical Cluster Analysis
HILIC	Hydrophilic interaction liquid chromatography
HM	Human milk
HMBC	Heteronuclear Multiple Bond Correlation
HMDB	Human Metabolome Database
HM-EV	Human milk extracellular vesicle
HMO	Human milk oligosaccharide
HoP	Holder pasteurization

HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
HUiP La Fe	University and Polytechnic Hospital La Fe
IPA	Isopropanol
IPO	Isotopologue Parameter Optimization
IQR	Interquartile range
IRE	Internal reflection element
IS	Internal standard
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LCFA	Long-chain fatty acid
LC-HRMS	liquid chromatography coupled to high resolution mass spectrometry
LC-MS	Liquid chromatography coupled to mass spectrometry
LC-QTOF-MS	Liquid chromatography coupled to quadrupole time of flight mass spectrometry
LLE	Liquid-liquid extraction
LMSD	LipidMAPS Structure Database

LOD	Limit of detection
LOQ	Limit of quantification
M	Mothers
MCDB	Milk Composition Database
MeOH	Methanol
MetDNA	Metabolic reaction network-based recursive algorithm
miRNAs	microRNAs
MS	Mass spectrometry
MSI	Metabolomics Standards Initiative
MTBE	Methyl tert-butyl ether
mTOR	Mammalian target of rapamycin
MUFA	Monounsaturated fatty acid
NIST	National Institute of Science and Technology
NMR	Nuclear magnetic resonance
NTA	Nanoparticle tracking analysis
OMM	Own mother's milk
PBS	Phosphate buffered saline

PC	Phosphatidylcholine
PCA	Principal component analysis
PCF	Propylchloroformate
PI	Preterm infant
PIUMet	Network-based Prize-collecting Steiner forest algorithm for integrative analysis of untargeted metabolomics
PNP	Purine nucleoside phosphorylase
PP	Polypropylene
PrOH	Propanol
PUFA	Polyunsaturated fatty acid
Py	Pyridine
QA	Quality assurance
QC	Quality control
QC-SVRC	Quality control-Supported vector regression correction
QqTOF/QTOF	Quadrupole time-of-flight
RI	Retention index
RMSECV	Root mean square error of cross validation

RSD	Relative standard deviation
RT	Retention time
SCFA	Short chain fatty acid
SD	Standard deviation
SFA	Saturated fatty acid
SHY	Statistical heterospectroscopy
SIM	Selected ion monitoring
SOP	Standard operational procedure
SPE	Solid-phase extraction
TG	Triradylglycerol
TI	Term infant
TOCSY	Homonuclear Correlation Spectroscopy
TQD	Triple quadrupole
UPLC	Ultra-performance liquid chromatography

List of tables

List of figures

Abstract

This PhD thesis addressed the different metabolomic and lipidomic approaches for nutrition assessment in infants and their lactating mothers in the frame of prematurity.

Preterm infants' (PI) early nutrition is one of the most important factors in improving their clinical outcomes. In this sense, breastfeeding is the gold standard of child nutrition due to human milk (HM) nutrients and bioactive components, which contribute to the optimal growth and development of the infant. When not enough own mother's milk is available, the administration of pasteurized donor human milk (DHM) is considered a valuable alternative for PIs. To ensure the safety of DHM, milk banks routinely employ Holder pasteurization (HoP) in order to destroy bacteria and viruses present. However, its composition can be affected by this process. In addition, this work has focused on the potential effect of maternal diet on HM composition. Likewise, diet is one of the key factors involved in the conformation of the intestinal microbiome, affecting microbial diversity, as well as the abundance of specific microbes.

Different sample preprocessing methods have been applied to urine, faeces and HM samples before analysis using targeted, semi-targeted and untargeted metabolomic and lipidomic analytical methods. Liquid chromatography-tandem mass spectrometry (LC-MS/MS), gas chromatography coupled to MS (GC-MS), and attenuated total reflectance - Fourier transform infrared spectroscopy (ATR-FTIR) methods have been developed for the analysis of nutrition and microbiome activity biomarkers, as well as to study HM composition and the effect of HoP on it.

The different developed analytical methodologies have been successfully applied to different sets of biological samples, demonstrating their applicability in the clinical field.

Hypothesis and objectives

In this PhD thesis, the hypothesis that is proposed is that the development of metabolomic and lipidomic methodologies can contribute to the nutritional evaluation of both PIs and their mothers. Likewise, three main objectives have been proposed:

- To evaluate associations between maternal diet and biomarkers of nutrition and microbial activity in the context of prematurity
- To evaluate the effect of HoP and storage on the composition of DHM
- To optimize the isolation and quality control of human milk extracellular vesicles (HM-EVs), as well as evaluate their lipid composition

The articles that make up this doctoral thesis by compendium of publications have been divided into three sections. Section I is entitled "*Biomarkers relevant in nutrition and microbiota studies in the field of neonatology*" and includes four chapters: i) the description of the NUTRISHIELD study, focused on the evaluation of nutrition of PIs and their lactating mothers; ii) joint evaluation of microbial activity and diet through urine biomarkers using LC-MS/MS; iii) the evaluation of microbial activity in urine and feces samples of PIs, TIs and their mothers by GC-MS; and iv) the analysis of biomarkers of microbial activity in human urine and murine faeces samples by LC-MS/MS.

Section II is entitled "*Effect of pasteurization on the metabolome and lipidome of HM*" and includes two chapters: i) a bibliographic review of current methodologies for untargeted metabolomic analysis in HM samples;

and ii) the evaluation of the effect of HoP on the metabolomic and lipidomic composition of DHM by LC-MS/MS and GC-MS.

Section III is entitled "*Novel analytical tools for the characterization of the molecular composition of HM-EVs*" and includes three chapters: i) the description of the protocol to be followed for the HM-EVs isolation and untargeted lipidomic analysis using LC -MS/MS; ii) the evaluation of the HM-EVs lipidomic profile by LC-MS/MS and the evaluation of the different normalization strategies; and iii) the development of an analytical method using ATR-FTIR for routine quality control of the HM-EVs extraction procedure.

Section I. Biomarkers relevant in nutrition and microbiota studies in the field of neonatology

Chapter 1. Fact-based nutrition for infants and lactating mothers – The NUTRISHIELD study

1.1. Abstract

Background: Human milk (HM) is the ideal source of nutrients for infants. Its composition is highly variable according to the infant's needs. When not enough own mother's milk (OMM) is available, the administration of pasteurized donor human milk (DHM) is considered a suitable alternative for preterm infants. This study protocol describes the NUTRISHIELD clinical study. The main objective of this study is to compare the % weight gain/month in preterm and term infants exclusively receiving either OMM or DHM. Other secondary aims comprise the evaluation of the influence of diet, lifestyle habits, psychological stress, and pasteurization on the milk composition, and how it modulates infant's growth, health, and development.

Methods and design: NUTRISHIELD is a prospective mother-infant birth cohort in the Spanish-Mediterranean area including three groups: preterm infants <32 weeks of gestation (i) exclusively receiving (i.e., >80% of total intake) OMM, and (ii) exclusively receiving DHM, and (iii) term infants exclusively receiving OMM, as well as their mothers. Biological samples and nutritional, clinical, and anthropometric characteristics are collected at six time points covering the period from birth and until six months of infant's age. The genotype, metabolome, and microbiota as well as the HM composition (i.e., macronutrients, fatty acids, vitamins, human milk oligosaccharides, and steroids) are characterized. Portable sensor prototypes for the analysis of HM and urine are benchmarked. Additionally, maternal psychosocial status is measured at the beginning of the study and at month six,

including social support, family functioning, perceived stress, anxiety, and depression symptoms, and traumatic life events. Mother-infant postpartum bonding and parental stress are also examined. At six months, infant neurodevelopment scales are applied. Mother's concerns and attitudes to breastfeeding are registered through a specific questionnaire.

Discussion: NUTRISHIELD provides an in-depth longitudinal study of the mother-infant-microbiota triad combining multiple biological matrices, newly developed analytical methods, and ad-hoc designed sensor prototypes with a wide range of clinical outcome measures. Data obtained from this study will be used to train a machine-learning algorithm for providing dietary advice to lactating mothers and will be implemented in a user-friendly platform based on a combination of user-provided information and biomarker analysis. A better understanding of the factors affecting milk's composition, together with the health implications for infants plays an important role in developing improved strategies of nutraceutical management in infant care.

Clinical trial registration: https://register.clinicaltrials.gov, identifier: NCT05646940

1.2. Introduction

Breastfeeding (BF) is the optimal feeding practice for all infants, associated with numerous health benefits for the mother–infant dyad, as well as, being an ecologic practice [49]. The World Health Organization highly recommends exclusive BF for all infants up to six months, born either full–term [18,19] or prematurely [20]. Human Milk (HM) is a dynamic fluid that meets the offspring's needs [50–53], and is ample in nutrients and bio-active

compounds intended to enhance immunity and growth [54]. HM also hosts a complex ecosystem of microbiota [55], which is of paramount importance for the offspring's immunity [50].

During the last decades, the incidence of preterm deliveries (<37 weeks of gestation) and survival rate of preterm infants (PI) have been steadily increasing [10], hand in hand with the interest in preterm and early infant nutrition. Associated birth complications are the leading cause of death among children under five years of age, responsible for ~1 million global deaths per year [56]. Progress in medical interventions has allowed to enhance survival of an increasing proportion of extremely low gestational age newborns and low birth-weight infants. Early infant nutrition has become a major player in improving clinical outcomes of survivors [13]. HM is recommended for PI based on an impressive array of benefits provided to this highly vulnerable population, including ameliorated immunological and gastrointestinal outcomes [14,57–62]. Optimal growth in the postnatal period is challenging dure to increased nutritional demands and metabolic and digestive immaturity of PI. Greater weight gain is associated with better neurodevelopment outcomes [63].

In situations where mothers are unable to produce sufficient milk quantities to exclusively or partially breastfeed, pasteurized donor human milk (DHM) is a viable option to avoid formula feeding, especially for low birth-weight children [64]. DHM is a valuable but limited resource and Human Milk Banks prioritize its distribution to the patients with highest risk of necrotizing enterocolitis, generally PI <32 weeks of gestation or with a birthweight <1500 g [19,65]. To date, most studies focus on the benefits of using DHM over formula in PI, when own mother's milk (OMM) is limited or unavailable. There is clear evidence of the superiority of OMM against formula [66];

however, although DHM apparently protects against necrotizing enterocolitis as compared to formula, results are not conclusive [67] and there are only few reports directly comparing DHM and OMM. While providing some bioactive agents, DHM consumption is associated with slower growth rates in comparison to the administration of OMM or formula [66]. This finding may be attributable to several factors. Most DHM is provided by women who have delivered at term and donate their milk in later stages of lactation up to several months after delivery. The effect of gestational age on HM composition has been reported to affect a wide array of compounds including lipids [68], lactoferrin [69], amino acids [70], and HM oligosaccharides [71]. In comparison to preterm milk during the first weeks after delivery, studies on the mean composition of DHM show a lower content in total protein, fat, and other bioactive molecules [72]. Its composition is also affected by the processing of expressed milk, including stringent protocols applied in HM banks, i.e. pasteurization, freezing, and storage [17,29], necessary to reduce the potential risk to transmit infectious agents.

Beyond the physiological adaptations, research has focused on the potential effects of the maternal diet on the HM composition. In an earlier report, HM composition was found rather different among mothers of diverse ethnic backgrounds, with different dietary habits [73]. The fatty acid profile of HM varies in relation to maternal diet, particularly, in the long chain polyunsaturated fatty acids (LCPUFAs) [17,29,74,75]. Additionally, in a recent comprehensive systematic review of 104 observational and interventional studies, HM composition was found to relate to the maternal consumption of fatty acids, fat-soluble vitamins, and vitamin B1 and C intake [33,34]. No similar relationships were found for dietary intake of iron, folate, calcium, selenium, or proteins [33].On the contrary, another recent systematic

review determined that information regarding the relationship between dietary patterns during lactation and HM content is insufficient for total fat, vitamins B and C, and choline content and inexistent for total proteins, macronutrient distribution, human milk oligosaccharides (HMOs), vitamins A, D, E and K, iodine, and selenium [76]. Hence, more studies are needed to conclude for several nutrients and little is known on how the maternal diet may also affect non-nutritive constituents of HM, e.g., prebiotics content [36] or microbiota [37,77]. Finally, existing evidence comes mainly from studies on full-term infants (TI); the impact of maternal nutrition on the composition of HM for the preterm offspring has not been adequately documented.

Therefore, the main objective of this study is to compare the % weight gain/month in PI and TI exclusively receiving either OMM or DHM. In addition, secondary objectives are (i) to evaluate associations between the mother's diet, physical, and psychosocial status (e.g., lifestyle, perceived stress, anxiety, and depression) and HM composition in PI and TI, (ii) to evaluate the effect of pasteurization/storage on DHM composition, (iii) to assess the interplay of microbiota and microbiota activity and HM composition, (iv) to assess the interplay of microbiota and microbiota activity detected in HM and PIs and TIs, (v) to evaluate the impact of microbiota and microbiota activity on the vitamin status of PIs and TIs, (vi) to test the performance of novel sensor devices developed within this project, and (vii) the development of a personalized nutrition algorithm for lactating mothers.

1.3. Methods and analysis

1.3.1. Study design

The NUTRISHIELD study is a parallel group, non-randomized, observational study performed at the Division of Neonatology of the University & Polytechnic Hospital La Fe (HULAFE), including TI and PI and their mothers, covering the period from birth to 24 months of infant's age. In addition, mothers providing DHM to study participants are also included.

The study protocol has been approved by the Scientific and Ethics Committee for Biomedical Research (CEIm) of the HULAFE (#2019-289-1). All methods have been performed in accordance with the relevant guidelines and regulations and written permission has been obtained from mothers or legal representatives by signing an informed consent form.

Confidentiality of subjects is maintained during the study. All participants have been assigned a code and data allowing personal identification will not be shared at any time. Participants may withdraw consent for participating in the study at any time.

1.3.2. Study population and recruitment

Participants were recruited at the end of pregnancy (when admitted to the obstetric ward) or within one week after birth in case of hospitalized infants at HULAFE. Recruitment started in October 2020 and was completed in July 2022. Three mother-infant groups were enrolled, i.e., PI fed with OMM

(PI-OMM), PI fed with pasteurized DHM (PI-DHM), and TI receiving OMM (TI-OMM), and their mothers.

The inclusion criteria were (i) acceptance of the mother to participate and sign an informed consent form, (ii) a gestational age < 32 weeks for the group of PI and > 37 weeks for the group of TI, and (iii) exclusive consumption (i.e., >80% of total intake) of either OMM or DHM at time point CEN (complete enteral nutrition) for PI or RBW (recovery of birth weight) for TI. Although the type of feeding of PIs included in either group (PI-OMM or PI-DHM) may have changed at later time points, they remained in the same study group, so the effect of feeding of each type of milk at CEN and its long-term effects can be studied. Likewise, the use of formula milk and the initiation of weaning at later time points was recorded only without affecting follow-up assessments. The exclusion criteria were (i) non-compliance with any of the inclusion criteria, (ii) the requirement of a special diet for the mother (e.g., celiac disease, diabetes) or maternal consumption of probiotics, (iii) the need of intestinal surgery, severe congenital malformations, or chromosomopathies of the infant, (iv) mother's residence outside the Valencian Community and (v) severe language barriers hampering the collection of necessary data from mothers not speaking Spanish and/or English.

Sample size estimation was performed for the primary outcome (i.e., % weight gain/month from birth until hospital discharge of PI fed with DHM and OMM) and was based on result of a previous study (42). Considering the median % weight gain/month (interquartile range, IQR) of 52% (IQR 30) in the PI-DHM group and 63% (IQR 21) in the PI-OMM, 18 infants per group were needed to achieve a power of 90% with an alpha error of 5% between

the % weight gain/month of PI exclusively receiving either OMM or DHM at CEN.

On the other hand, HM donors were recruited during their regular visits at the hospital's HM Bank of the Valencian Community. All participants met ordinary criteria for HM donation (e.g., negative screening results of a series of transmissible diseases, toxic habits, or some chronic medication) and accepted to sign an informed consent form and participate in the study.

1.3.3. Assessment points and biological samples

Samples from lactating mothers and their infants are collected at different timepoints, as shown in **Figure 1.1**. At birth, cord blood, and urine, and faeces from both infants and mothers are collected when possible. When PIs achieve CEN and TIs RBW, as well as one, two, three, and six months after delivery, urine and faeces from mothers and infants, as well as, HM from mothers who breastfeed, are collected. Infants' and mothers' buccal swabs are collected at month six. During hospitalization of participants, samples are collected at the hospital by healthcare providers. After discharge, samples are collected at home and either transported to the hospital by study participants or staff of the NUTRISHIELD study. Mothers receive Standard Operation Procedures (SOPs) with detailed instructions and all the required material for sample collection at home. Additionally, hospital's staff instructs mothers on how to collect and store samples correctly.

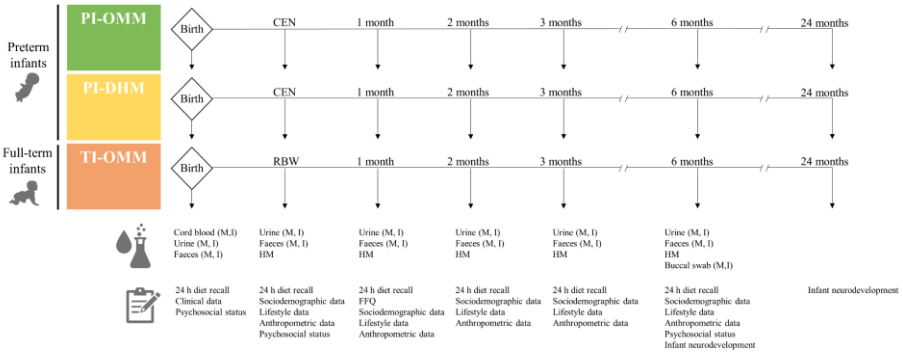


Figure 1.1. Design of the NUTRISHIELD study. Note: PI: preterm infant; TI: term infant; M: mother; I: infant; OMM: own mother's milk; DHM: donor human milk; CEN: complete enteral nutrition (150 mL/kg/day); RBW: recovery of birth weight; HM: human milk; FFQ: food frequency questionnaire.

Samples collected at home are stored at 4 ºC during a maximum of 24 h before their transport to the hospital on ice, where they are registered, aliquoted, labelled, and stored at -80 ºC until analysis. A HM and faeces aliquot is directly stored in a tube containing a microbiome preservative solution. For DHM, collection times respective to delivery are heterogeneous, since from each donor several aliquots are collected over time, pooled, and then subjected to Holder pasteurization in batches (i.e., 62.5 ºC during 30 min followed by rapid cooling to 4 ºC). The lactation time was estimated as the mean (standard deviation, SD) of time elapsed during the collection of different aliquots from one DHM batch, being 16 (SD 12) days.

Mothers' and infants' buccal swab samples are collected by thoroughly rubbing a flocked sterile swab up and down the inner side of both

cheeks for 30 seconds each. Swabs are cut and stored in a sterile tube at -80 ℃ until further analysis. Discarded arterial and venous blood from the umbilical cord is collected in EDTA blood sample tubes after the placenta is delivered and separated from the baby. Blood is centrifuged (1300 x g for 10 min at 20 ℃) and the upper plasma layer is stored in opaque vials at -80 ℃ until analysis. Mother's first morning urine is collected in a polypropylene container, and infant's urine is collected by placing sterile cotton pads in the diaper and, after urinating, squeezing them with a sterile plastic syringe into sterile tubes (see **Figure 1.2**). Mother's faeces are collected in a polypropylene container, and infant's faeces are collected directly from the diaper into sterile tubes using sterile tweezers.

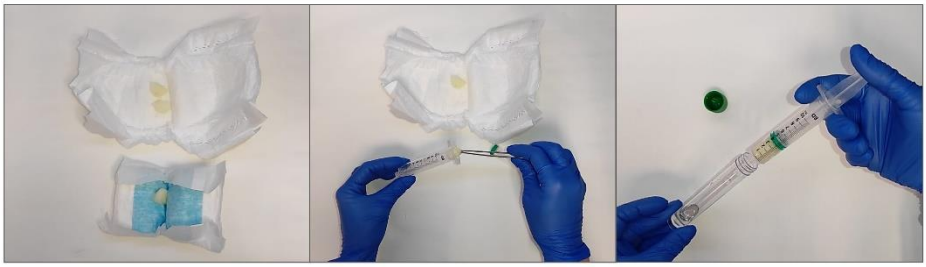


Figure 1.2. Infants' urine sample collection procedure. Sterile cotton pads are placed in the diaper (left); cotton pads soaked with urine are collected using sterile tweezers (middle); cotton pads are squeezed with a sterile syringe to collect the urine sample (right).

Mothers express milk using breast milk pumps following an SOP employed routinely in the hospital and the HM bank. The removable parts of

the breast milk pump as well as the collection bottles are sterilized before their use. HM is collected at least three hours after breastfeeding. Mothers are required to wash their hands with soap and water and clean nipples with water. HM should be preferably expressed between 7 and 10 a.m., full expression of one breast is required (a minimum volume of 50 mL is recommended), and details regarding the extraction method (e.g., breast pump brand), date, and extracted volume are registered. For DHM samples, aliquots of each pooled DHM sample are collected before and after Holder pasteurization.

1.3.4. Dietary and psychological status assessment

For all three study groups, information obtained from clinical records is collected. Anthropometric data (e.g., weight, height, and head circumference), sociodemographic, dietary, other relevant information related to lifestyle (e.g., smoking habits, sleep hours, and physical activity), and psychosocial status are provided through questionnaires by participating mothers as summarized in **Table 1.1**. As shown in **Figure 1.1**, 24 h dietary recalls are recorded at all timepoints except at birth, food frequency questionnaire (FFQ) at month 1, maternal psychosocial status at CEN/RBW and month six (corrected age for PI groups), and infant neurodevelopment at months six and twenty-four. The questionnaires are administered online, on paper, or through direct interviews with the mothers.

Table 1.1. Recorded parameters from lactating mothers and infants.

Assessment	Parameter
Nutrition	FFQ
	24-hour dietary recall
Clinical data	Maternal conditions
	Pregnancy complications
	Delivery type
	Medical information
	Maternal medication
	IIFAS
	Infant's gestational age
	Infant sex
	Infant diagnosis
	Perinatal complications
Sociodemographic data	Origin
	Education
	Household income
	Employment status
Lifestyle	Smoking habits
	Physical activity
	Sleeping habits
Anthropometric data	Maternal weight
	Infant height
	Infant head circumference
Psychosocial status	MSPSS
	FACES
	PSS-10
	STAI
	EPDS
	BDI
	TEQ
	PBQ
	PSI

Table 1.1 (continuation). Recorded parameters from lactating mothers and infants.

Assessment	Parameter
Infant neurodevelopment	ASQ-3 IBQ Merril-Palmer revised scales

Notes: FFQ: Food Frequency Questionnaire, IIFAS: Iowa Infant Feeding Attitude Scale, BMI: Body Mass Index, MSPSS: Multidimensional Scale of Perceived Social Support, FACES: Family Adaptability and Cohesion Evaluation Scale, PSS-10: Perceived Stress Scale-10, STAI: State-Trait Anxiety Inventory, EPDS: Edinburgh Postnatal Depression Scale, BDI: Beck Depression Inventory, TEQ: Traumatic Experience Questionnaire, PBQ: Postpartum Bonding Questionnaire, PSI: Parenting Stress Index, ASQ-3: Ages & Stages Questionnaires, IBQ: Infant Rothbart's Temperament Questionnaire.

Regarding the 24 h recall, trained researchers ask for all foods and beverages participants consumed during the previous day, using the multiple-pass method [78]. Recall data are analyzed in terms of nutrients using the dietary analysis software Nutritionist Pro™ (2007, Axxya Systems, Texas, USA). Additionally, dietary intake is grouped into food groups, (i.e., fruits, vegetables, bread/starch, meat/high fat, meat/medium fat, meat/low fat, meat/very low fat, milk/non-fat, milk/low fat, milk/full fat, eggs, fish, soy, dairy products, soft drinks, and other carbohydrate-rich foods).

The FFQ is administered by trained personnel, and it comprises 142 questions on the consumption of foods that are commonly eaten by the Spanish population throughout a year, including dairy products, cereals, fruits,

vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants, and sweets. Using a 9-grade scale ("never or less than 1 time/month", "1-3 times/month", "1 time/week", "3-4 times/week", "5-6 times/week", "1 time/day", "2-3 times/day", "4-5 times/day", "≥6 times/day") participants are required to indicate the absolute frequency of consuming a certain amount of food, expressed in grams, milliliters or in other common measures, such as slice, tablespoon, or cup, depending on the food. The previous month is set as the timeframe. The FFQ is an easy-to-use questionnaire and is not expected to increase the burden of lactating mothers.

Based on the FFQ-responses, adherence to the Mediterranean diet is evaluated by using the MedDietScore, a composite score calculated for each participant [79]. For food groups presumed to be part of the Mediterranean pattern (i.e., those with a recommended intake of 4 servings per week or more, such as non-refined cereals, fruits, vegetables, legumes, olive oil, fish, and potatoes) higher scores are assigned when the consumption is according to the rationale of the Mediterranean pattern, while lower scores are assigned when participants report no, rare, or moderate consumption. For the consumption of foods presumed to be eaten less frequently within the Mediterranean diet (i.e., consumption of meat and meat products, poultry, and full fat dairy products), scores are assigned on a reverse scale. As this study is focused on lactating mothers, the original score was modified by removing the alcohol consumption component. Thus, the range of this modified MedDietScore is between 0 and 50, with higher values of the score indicating greater adherence to the Mediterranean diet.

Maternal psychosocial status is measured at recruitment and at six months, including social support (Multidimensional Scale of Perceived Social Support, MSPSS [80]), family functioning (Family Adaptability and Cohesion

Evaluation Scale, FACES [81]), perceived stress (Perceived Stress Scale -10, PSS-10 [82]), anxiety (State-Trait Anxiety Inventory, STAI [83]), depression (Edinburgh Postnatal Depression Scale, EPDS [84], Beck Depression Inventory (BDI) [85]) symptoms, and traumatic life events (Traumatic Experience Questionnaire, TEQ [86,87]). Mother-infant postpartum bonding (Postpartum Bonding Questionnaire, PBQ [88]) and parental stress (Parenting Stress Index, PSI [89]) are also examined.

As for the infants, at six and twenty-four months (corrected age for PIs) assessment of neurodevelopment is conducted during a hospital visit by a trained psychologist and psychiatrist from the team. Parents complete the Ages & Stages Questionnaires, Third Edition (ASQ-3) [90] for psychomotor development evaluation and the Infant Rothbart's Temperament Questionnaire (IBQ) [91] for infant temperament. The ASQ is a well standardized instrument used in clinical and research practice to examine children's psychomotor development gathering items in five domains: communication, gross motor, fine motor, problem solving, and personal-social. Clinical observation and Merrill-Palmer revised scales [92] for development are applied during the session.

Mother's concerns and attitudes to breastfeeding are also registered through the Iowa Infant Feeding Attitude Scale (IIFAS) [93] at CEN/RBW and six months.

Milk donors are asked to provide sociodemographic, medical, dietary, and other lifestyle data. PSS-10 is evaluated when entering the study in order to correlate with hormones and other compounds present in the milk. In particular, each time a batch of milk is prepared, a FFQ is administered (representing the period of time that the batch of milk covered) and the donors are asked to provide medical, anthropometric, and lifestyle data.

1.3.5. Analysis of biological samples employing laboratory methods

A range of state-of-the-art as well as novel laboratory methods are employed for the analysis of the collected biological samples, as summarized in **Table 1.2**.

Table 1.2. Analysis of biological samples employed.

Analysis	Parameters	Technique	Matrix
Genome	Whole genome sequencing	Illumina NovaSeq	Buccal swab
Microbiota	Bb species	ATGC	Faeces HM
Microbiota activity	SCFAs BCAAs	GC-MS	Faeces Urine HM
	BAs	LC-MS/MS	Urine
Nutrition	Flavonoids Isoflavones Arylglycines Amino acids	LC-MS/MS	Urine
Metabolome	Metabolomic fingerprinting	LC-HRMS	Urine
Macronutrients	Fat Carbohydrates Crude & true protein Total solids Energy	Miris HM analyzer	HM

Table 1.2 (continuation). Analysis of biological samples employed.

Analysis	Parameters	Technique	Matrix
Fatty acids	34 fatty acids (C6-C24) SAT MONOs PUFAs UNSAT SCFAs MCFAs LCFAs	GC-MS HM fatty acid sensor	HM
Vitamins	Vitamin B, D and K groups Retinol forms Carotenoids and vitamin E groups	LC-MS/MS LC-UV LC-UV/Fluorescence	HM Umbilical cord blood
HMOs	HMO screening	LC-HRMS	HM
Steroids	19 steroid hormones	LC-MS/MS	HM Urine
Proteins	Total protein Alpha-lactalbumin Lactoferrin Casein	HM protein sensor	HM
pH	pH	Urine pH sensor	Urine
Phosphate and creatinine	Phosphate and creatinine	Phosphate and creatinine sensor	Urine

Note: ATGC: Advanced Testing for Genetic Composition, Bb: Bifidobacterium, GC: Gas Chromatography, (HR)MS: (High Resolution) Mass Spectrometry, SCFAs: Short Chain Fatty Acids, BCAAs: Branched Chain Amino Acids, BAs: Bile Acids, SAT: saturated fatty acids, MONOs: monounsaturated fatty acids, PUFAs: polyunsaturated fatty acids, UNSAT: unsaturated fatty acids, MCFAs: medium-chain fatty acids, LCFAs: long-chain fatty acids, LC: Liquid Chromatography, UV: Ultraviolet, HMOs: Human Milk Oligosaccharides.

1.3.5.1. Genome sequencing

For genome sequencing analysis, DNA extraction from buccal swab samples takes place according to the protocol developed by the sequencing provider facility. Polymerase chain reaction (PCR) quality control (QC) is performed before samples are shipped to the sequencing service provider, and again upon receival by the sequencing facility. QC and library preparation follow the Illumina NovaSeq protocol (Paired End, 150bp). Three polygenic risk scores (PRS) models are developed for body mass index (BMI), diabetes type 2, and lactose intolerance, calculated by Plink software [94], and the resulting genotype files, one per sample, are then used as separate input to the PRS models.

This score can then be compared to the scores obtained from the UK Biobank set [95], to estimate the relative risk for this individual compared to the rest of the UK Biobank individuals (~370.000 individuals), which are split in 10 quantiles according to their genetic risk.

1.3.5.2. Microbiota analysis

Advanced Testing for Genetic Composition (ATGC) is used for microbiota analysis of faeces and HM samples [96]. It is a targeted measurement technique that provides greater precision, specificity, and versatility than rtPCR, while being just as cost-effective and fast. It has been developed to translate discovery data from untargeted analysis (e.g., 16S or shotgun meta-genomics) into tests for routine use.

The Bifidobacterium (Bb) assay used is specifically designed to analyze the main species and subspecies varieties in HM and newborn gut microbiomes. For sample processing, the HM lipid layer is solubilized using a detergent, while for faeces, a lysis step using a lysis buffer with bead beating for 20 seconds is required. Then, an in-house DNA extraction procedure and a PCR are carried out. Cycling temperature capillary electrophoresis (CTCE) is made for each primer on an independent capillary. It is performed on a MegaBace 1000 instrument (General Electrics, Boston MA, USA) as described earlier [97].

1.3.5.3. Microbiota activity biomarkers analysis

The microbiota activity is evaluated with the following methods:

1. Short chain fatty acids (SCFAs) and branched chain amino acids (BCAAs) in faeces, urine and HM samples

SCFAs (i.e., acetic, propionic, isobutyric, butyric, 2-methylbutyric, isovaleric, valeric, caproic and heptanoic acids) and BCAAs (i.e., valine, leucine and isoleucine) are determined by targeted gas chromatography coupled to mass spectrometry (GC-MS), as described elsewhere [98,99], using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector (Agilent Technologies, Santa Clara, CA, USA).

2. Bile acids (BAs) in urine samples

In total, 34 BAs (6 primary, 7 primary conjugated, 5 secondary, 10 secondary conjugated and 6 sulphated) are determined in urine samples by liquid chromatography coupled to tandem MS (LC-MS/MS), as described

elsewhere [100] with minor modifications, using an ACQUITY LC chromatograph (Waters Ltd, Elstree, UK) coupled to a Xevo TQ-S MS detector (Waters, Manchester, UK).

1.3.5.4. Nutrition biomarkers analysis

Quantification and semi-quantification of 20 and 205 urinary nutrition biomarkers (e.g., flavonoids, isoflavones, and arylglycines), respectively, related to nine food groups such as fruits, vegetables, meat, fish, dairy products, milk, seeds, coffee, and soft drinks is carried out [101,102]. Additionally, seven microbiota activity biomarkers are determined using this method (i.e., phenylpropionylglycine, L-kynurenine, L-tyrosine, hippuric acid, 3-indolepropionic acid, ferulic acid sulphate and 3-indoleacetic acid). LC-MS/MS analysis is conducted using a Sciex QTRAP 6500+ system (Sciex, Framingham, Massachusetts, USA).

1.3.5.5. Untargeted metabolomic fingerprinting

Untargeted metabolomic analysis is carried out in urine samples by LC-MS/MS, employing an Agilent 1290 Infinity HPLC system coupled to an Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS detector (Agilent Technologies, Santa Clara, CA, USA), as previously described [103].

1.3.5.6. Macronutrients analysis

Direct measurement of HM macronutrients including fat, carbohydrates, crude and true proteins, total solids, and energy are determined using a Miris HM analyzer (Miris AB, Uppsala, Sweden). Determinations, QC, and instrument calibration are performed following the SOP provided by the manufacturer [104].

1.3.5.7. Fatty acid profile analysis

The targeted analysis of 36 fatty acids in HM samples is carried out by GC-MS as described elsewhere [105,106], using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector (Agilent Technologies, Santa Clara, CA, USA).

1.3.5.8. Vitamin analysis

Quantification of water-soluble (group of vitamin B: thiamine, thiamine monophosphate, riboflavin, flavin adenine dinucleotide, nicotinamide, pyridoxal and pyridoxal phosphate) and lipid-soluble vitamins (group of vitamin A: retinol forms, β-carotene, β -cryptoxanthin, lutein, lycopene and zeaxanthin; group of vitamin E: α-tocopherol and γ-tocopherol; group of vitamin K: phylloquinone and menaquinone-4; group of vitamin D: cholecalciferol and calcifediol) in HM and umbilical cord blood samples is carried out.

The analysis of B, D and K vitamins are determined as described somewhere [107–111], using an Acquity LC chromatographic system (Waters AG, Switzerland) hyphenated to a Xevo TQ-S mass spectrometer (Waters AG, Switzerland).

The retinol forms are analyzed as previously described [110], using a LC using a quaternary Flexar chromatographic system (Perkin Elmer, Switzerland) with UV detection (325 nm).

Carotenoids and vitamin E analysis is carried out as previously described [112], using a binary Acquity LC chromatographic system (Waters AG, Switzerland) with multiple UV detection (295 nm, 450 nm and 472 nm) and fluorescence detection (exc. 296 nm / em. 330 nm).

1.3.5.9. Oligosaccharides analysis

HMOs analysis is carried out by LC-MS in accordance with a previously described protocol [113,114] with some modifications, employing a Vanquish LC Binary Pump coupled to an Orbitrap QExactive Plus MS detector (ThermoFisher, Waltham, MA, USA).

1.3.5.10. Steroid analysis

Steroid analysis of HM and urine samples is carried out on an Acquity UPLC system (Waters Ltd, Elstree, UK) coupled to a Waters Xevo TQ-S MS detector (Waters, Manchester, UK) as previously described [115,116]. A

panel of 19 steroids is targeted (i.e., cortisol, 5β-tetrahydrocortisol, 6β-hydrocortisol, 20α-dihydrocortisol, 20β-dihydrocortisol, corticosterone, aldosterone, estrone, androstenedione, dehydroepiandrosterone, progesterone, 17-hydorxy-progesterone, pregnenolone, cortisone, 20α-dihydrocortisone, 20β-dihydrocortisone, 6α-hydroxycortisone, testosterone, and 5α-dihydrotestosterone).

1.3.6. Sensor prototypes

Due to the special requirements of the study, three sensor prototypes enabling the determination of complementary parameters in HM and urine samples are being developed and benchmarked within the NUTRISHIELD study.

1.3.6.1. HM protein sensor

Direct quantification of the most abundant proteins (i.e., casein, α-lactalbumin, and lactoferrin) in HM milk is performed using quantum cascade laser-based mid-infrared (QCL-IR) spectroscopy [117,118]. Laser-based IR spectroscopy enables more robust and sensitive analysis in the spectral region of IR signatures of proteins compared to conventional Fourier-transform infrared (FTIR) spectrometers [117]. Reference protein analysis is carried out using specific Kjeldahl and HPLC analysis. Within the study, a dedicated QCL-based protein analyzer for HM is developed and benchmarked.

1.3.6.2. HM fatty acids sensor

HM fatty acid profiling is performed based on mid-IR spectroscopy of an extracted lipid HM fraction. For analysis, 15 µL of the pure fat are transferred onto a diamond single bounce attenuated total reflection (ATR) accessory (Platinum ATR, Bruker, Ettlingen) connected to a Tensor 37 (Bruker, Ettlingen) FTIR spectrometer [119,120].

1.3.6.3. Urine pH sensor

A custom-made portable system based on potentiometric measurements and screen-printed electrode technology is developed to measure pH in urine. The system works with small sample volumes (i.e., 50 □L) and no dilution is needed. An additive is spiked into samples prior to measurement in order to preserve and regenerate the sensor's surface, which extends sensor lifetime to multiple uses with no carry-over between urine samples.

1.3.6.4. Urine phosphate and creatinine sensor

Quantification of phosphate and creatinine in urine samples is performed using a method based on FTIR transmission spectroscopy. Within the study, a dedicated QCL-based urine analyzer is developed. Reference values for phosphate are obtained by a colorimetric determination after

reaction of inorganic phosphorus with ammonium molybdate, while creatinine reference values are obteined following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA).

1.3.7. Statistics and data integration

The Student's t-test, Wilcoxon ranksum test or X2 test with $\alpha = 0.05$ will be used for between-group comparisons and fold changes (FC) will be calculated as the ratio of means or medians between groups in accordance to the underlying distribution of the data. Pearson's linear or Spearman rank correlation coefficients will be determined between continuous variables. If necessary, partial correlation coefficients adjusting for known confounding factors will be computed. The false discovery rate (FDR) from the *p*-values of multiple-hypothesis testing will be estimated using the Benjamini and Hochberg procedure [121] and adjusted *p*-values < 0.05 will be considered statistically significant.

One of NUTRISHIELD secondary objectives is the development of a personalized nutrition system based on measured biomarkers. The system is to be made available to medical personnel, as well as mothers, to improve infant-mother dyads' health and wellbeing. The results and the associations revealed from the study will be used to train a personalized nutritional algorithm with data obtained from clinical settings. This will be used to build and validate the NUTRISHIELD platform, describing how a theoretical framework designed and fed by the patients' data translates to clinical practice. Toward that goal, a comprehensive data integration system has been developed as the Clinical Trial App (CTA). The CTA is a system intended for

personalized medicine/nutrition, implementing the following main features: (i) data acquisition from questionnaires, (ii) data acquisition from laboratory biomarker analysis, (iii) processing data to train machine learning algorithms, and (iv) deploying trained algorithms to produce reports for final users.

Figure 1.3 shows the data flow and functionalities of the CTA. Questionnaires data are collected by medical staff directly using the phone app. Alternatively, questionnaires that have been collected on paper or on spreadsheets, can be uploaded as well. All samples can be identified with a unique QR code and, at the moment of collection, the phone app can be used to scan the QR code and input sample data. This way the samples can travel between labs with the QR code as only identification. The QR code can be scanned again to associate the results with the original sample when uploading it through the web portal. The collected data is stored on a database and used to feed the personalized recommendation algorithm. This algorithm produces a report and feeds it back to the phone app.

Using the CTA, all study data is consolidated in a structured database, of which an anonymized version can be later stored on public repositories for re-use in further research.

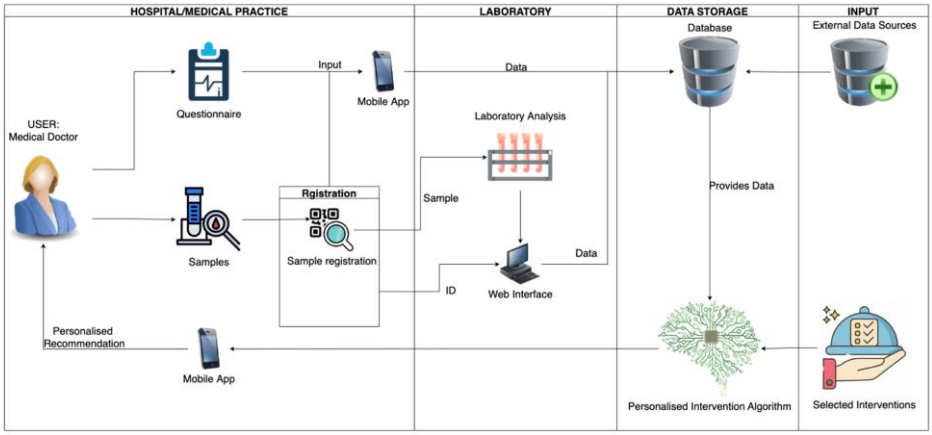


Figure 1.3. Data flow and functionalities of the Clinical Trial App (CTA).

1.4. Results and discussion

Figure 1.4 summarizes the recruitment process, including the eligible PI (n = 104), excluded (n = 49), lost to follow up (n = 10) and the number of PI that were finally included in the study (n = 45). In addition, a control group comprising 31 full term OMM infant-mother dyads was recruited, with five participants being lost during follow-up. Altogether, a total of 71 infant-mother dyads were included and completed the study (28 PI-OMM, 17 PI-DHM and 26 TI-OMM), after losing 15 infant participants during follow-up. In total 662 urine, 594 faeces, 134 buccal swab, 33 venous and arterial cord blood, and 234 HM samples, as well as aliquots from 147 DHM batches before and after pasteurization have been collected. An ample array of laboratory methods for biomarker analysis and screening of biological samples and three ad-hoc designed sensors were developed. Currently, the analysis of all collected study samples is in progress at the different participating institutions.

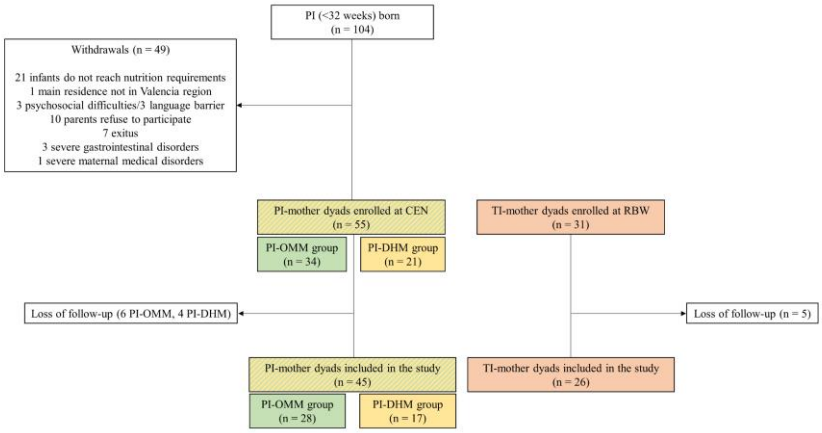


Figure 1.4. Prospective flow chart of the NUTRISHIELD study. Note: PI: preterm infant; CEN: complete enteral nutrition (150 mL/kg/day); RBW: recovery of birth weight; OMM: own mother's milk; DHM: donor human milk.

The results obtained from HM analysis will be used to study how macronutrients, proteins, vitamins, HMOs, steroids, fatty acids, and microbiota are affected by maternal nutrition, psychosocial status (e.g., social support, family functioning, and perceived stress), other clinical variables (e.g., gestational age, BMI, and type of delivery), and pasteurization, in the case of DHM. Additionally, it is assessed how HM composition affects growth and other health parameters in TI and PI. The results obtained from urine samples analysis will also be used to study how urinary nutrition biomarkers and the metabolic profile are affected by nutrition, psychosocial status, the PRS obtained from genome sequencing and other clinical variables. In addition, HM and faeces microbiome analysis will be used to study its correlation with microbiota biomarker analysis (i.e., SCFAs, BCAAs, and

BAs). HM and urine samples will also be used to test the pH, protein, fatty acids, and phosphate and creatinine sensors developed within the study.

This study presents several strengths and limitations. During the study design, special emphasis was put on the comfort of participants. Hence, the collected biological samples, including faeces, urine, HM, buccal swap, and umbilical cord blood samples, were obtained by non-invasive procedures that had been tested and employed in earlier studies carried out at HULAFE. SOPs for sample collection and handling were designed and reviewed by the study team ahead of time. Sample collection kits with detailed instructions and pictures were prepared for sample collection at home. Finally, all procedures were tested in a small pilot study conducted before the initiation of the main study, allowing for minor amendments in the protocols and leading to a homogeneous process conducted all through the main study. During follow-up, some visits were carried out at home avoiding the additional effort for participants to travel to the hospital. These efforts helped to encourage mothers to participate in the study and enhanced adherence during the follow-up period.

One of the most outstanding strengths of this study is the rich array of information obtained from participants during the first six months of life, a pivotal period for health programming. Direct access to participants through trained staff of the research team at the hospital greatly aided to reduce the rate of missing data. The promotion of good relationships with participants could facilitate to further extending the life of the study cohort. Nevertheless, the follow-up period presented challenges. We would like to highlight the loss of some samples due to technical issues when collected at home as well as the temporary loss of some sampling time points, e.g., due to illness of participants or overburdening of lactating mothers. Our preliminary results

need to be confirmed in a greater cohort of infants and in different European settings.

This study will provide a better understanding of the impact of maternal nutrition in HM composition, and the interplay of HM composition, microbiota, and newborn physiology, especially in PI's development. More knowledge and a better understanding could allow for a personalized nutrition or even individualized care at an early stage of the neonatal period to improve the outcome of these vulnerable newborns. Additionally, data obtained from this study will be used to train a machine-learning algorithm for providing dietary advice to lactating mothers and will be implemented in a user-friendly platform based on a combination of user-provided information and biomarker analysis.

1.5. Conclusion

A better understanding of the factors affecting the milk's composition, together with the health implications for infants plays an important role in developing improved strategies of nutraceutical management in infant care. The development of a user-friendly platform based on a combination of user-provided information and biomarker analysis for providing dietary advice would make latest scientific advances directly accessible to a wide range of lactating mothers. This development is anticipated to be of special importance for health, growth, and development of prematurely born infants.

Chapter 2. Joint microbiota activity and dietary assessment through urinary biomarkers by LC-MS/MS

2.1. Abstract

Accurate dietary assessment in nutritional research is a huge challenge, but essential. Due to the subjective nature of self-reporting methods, the development of analytical methods for food intake and microbiota biomarkers determination is needed. This work presents an ultra-high performance liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) method for the quantification and semi quantification of 20 and 201 food intake biomarkers (BFIs), respectively, as well as seven microbiota biomarkers applied to 208 urine samples from lactating mothers (M) (N=59). Dietary intake was assessed through a 24-hour dietary recall (R24h). BFI analysis identified three distinct clusters among samples: samples from clusters 1 and 3 presented higher concentrations of most biomarkers than cluster 2, being dairy products and milk biomarkers more concentrated in cluster 1, and seeds and garlic and onion in cluster 3. Significant correlations were observed between three BFIs (fruits, meat, and fish) and R24h data ($r > 0.2$, p-values < 0.01, Spearman correlation). Microbiota activity biomarkers were simultaneously evaluated and subgroup patterns detected were compared to clusters from dietary assessment. These results evidence feasibility, usefulness, and complementary nature of the determination of BFIs and R24h as well as microbiota activity biomarkers in observational nutrition cohort studies.

2.2. Introduction

The measurement of diet exposure is crucial for determining associations between food intake and health status. Additionally, optimizing nutritional advice to specific population groups (e.g. chronic diseases patients, lactating mothers, etc.) has recently become a major challenge [122]. The incidence of preterm deliveries (<37 weeks of gestation) and survival rate of preterm infants have been steadily increasing in the last decades [10] and early infant nutrition is key for improving clinical outcomes. As human milk (HM) is recommended as the gold standard for infant nutrition [13] and the impact of maternal nutrition on the composition of HM has been evidenced [17,29,74], it is important to develop tools that can help to evaluate maternal dietary patterns with the aim of providing dietary advice that could enhance preterm infant's growth rate.

Dietary intake data is most commonly collected using tools based on self-reporting, such as food frequency questionnaires (FFQ) for the assessment of regular consumption and food diaries or 24-hour recalls (R24h) for the assessment of short-term consumption. This food intake data is then translated into quantitative information regarding the specific nutrients or food groups using food composition databases. However, such methods are prone to errors due to their subjective nature [35]. In particular, foods perceived as being unhealthy (e.g., processed foods and high fats) are typically under-reported [123], while foods perceived as being healthy (e.g., fruits and vegetables) are often being over-reported. Therefore, the generation of robust data of regular dietary intake is essential to improve the accuracy of dietary assessment. The measurement of metabolomic food intake biomarkers (BFIs) in biological samples [124,125] has emerged in recent years for supporting a

more accurate assessment of nutritional intake [126,127]. Metabolomic fingerprinting has opened new opportunities for the discovery of specific BFIs in body fluids [128]. For example, flavonoids, methyl histidine, isoflavones, trimethylamine N-oxide (TMAO), and arylglycines have been described recently as biomarkers of fruits and vegetables, meat, seeds, fish, and dairy products, respectively [125,129].

After ingestion, these compounds are metabolized by phase I and II enzymes, yielding glucuronidated, sulfated, and methylated metabolites [126]. Although recent studies illustrated that an individual's dietary habits can predict the levels of specific metabolites present in plasma [130] and research has identified hundreds of mediation linkages that provide insight into diet–microbiome interactions [131], urine is the most practical and feasible biofluid used for BFIs identification, since many metabolic byproducts are excreted through it [132]. Compared to the collection of other biofluids, such as blood and plasma, the collection of urine is easier, cheaper, allows quicker collection of large volumes, and is less burdensome and invasive for the participants [133].

Due to the different nature of BFIs, high sensitivity and wide coverage analytical methods are needed to assess the metabolic fingerprint of food intake. Liquid (LC) or gas (GC) chromatography coupled to mass spectrometry (MS) and H1 nuclear magnetic resonance (NMR) spectroscopy are the most commonly used techniques for the analysis of BFIs [134], being targeted assays that enable the quantification of metabolites the preferred strategy [135]. The vast majority of previously published studies about BFI fingerprinting consist in complex and time-consuming sample preparation, such as enzymatic hydrolysis [136,137] or solid phase extraction [101,138], resulting in incomplete hydrolysis and low recoveries, respectively.

Moreover, the lack of commercial standards of metabolic byproducts entangle the targeted approach [139] and therefore an attractive strategy to overcome this drawback could be the semi-quantification of these urinary metabolic byproducts.

In addition, diet is one of the key factors involved in shaping the gut microbiota, affecting the microbial diversity, as well as the abundance of specific microbes [38–40]. Metabolic end products of dietary micro and macronutrients, such as bile acids (e.g., cholic and lithocholic acids), short chain fatty acids (e.g., acetic and butyric acid), indole and polyphenyl derivatives (e.g., 3-indolepropionic acid (3-IPA) and hippuric acid), and phenolic acids (e.g., gallic acid and ferulic acid) are used to assess the activity of gut microbiota [41–45]. A joint assessment of food intake and microbiota activity biomarkers would be of interest.

The aim of this work was to develop an ultra-high performance LC coupled to tandem MS (UHPLC-MS/MS) method for the quantification and semi-quantification of 20 and 201 BFIs, respectively, of different food groups (e.g., fruits and vegetables, meat, etc.) and the simultaneous determination of seven microbiota biomarkers. The applicability of the method was evaluated by the analysis of 208 urine samples from lactating mothers (N = 59). The performance of the determined BFIs as a complementary tool to self-reported R24h dietary assessment was evaluated. Furthermore, nutrition patterns were compared to microbiota activity biomarker diversity clusters.

2.3. Materials and methods

2.3.1. Study design, population, and sample collection

Samples were collected in the frame of the NUTRISHIELD project (https://nutrishield-project.eu/), in a prospective, observational, cohort study (NCT05646940) performed at the Division of Neonatology of the University and Polytechnic Hospital La Fe (Valencia, Spain), including infant-mother dyads. Urine samples from lactating mothers were collected at six time points, covering the period from birth to six months of infant's age. R24h were recorded at all timepoints except at birth, while FFQs were recorded at month one. During data analysis, urine samples and R24h from each time point were treated individually, as maternal food intake varied between time points. Demographic and clinical characteristics of participants are shown in **Table 2.1**. Mother's first morning urine was collected in sterile polypropylene PP containers. In total, 208 urine samples from 59 lactating mothers were collected. For additional information on the NUTRISHIELD study the reader is referred to Ramos-Garcia et al. [140].

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry #2019-289-1 and all methods were performed in accordance with relevant guidelines and regulations. Written informed consents were obtained from lactating mothers prior to sample collection and analysis of clinical and demographic information.

Table 2.1. Demographic and clinical characteristics of participants at time point 'Recovery of Birth Weight' or 'Complete Enteral Nutrition' for mothers of term and preterm infants, respectively.

Parameters	**Mothers (N = 59)**
Age (years), mean (SD)	36 (5)
Weight (kg), mean (SD)	62 (14)
BMI (kg/m^2), mean (SD)	24 (5)
C-Section delivery, N (%)	28 (48)
Antibiotic therapy, N (%)	3 (5)
Dietary supplements, N (%)	52 (88)
MedDietScore*, mean (SD)	30 (5)

Note: *Recorded at time point month one; SD = standard deviation; BMI = body mass index; C-Section = caesarean section; MedDietScore = adherence to the Mediterranean diet.

2.3.2. Dietary assessment

Regarding the dietary assessment methods, a R24h and a validated FFQ [141] were performed. For the R24h collection, trained researchers asked for all foods and beverages participants consumed the previous day, using the multiple-pass method [78]. Recall data were analyzed in terms of nutrients using the dietary analysis software Nutritionist Pro™ (2007, Axxya Systems, Texas, USA). Additionally, dietary intake was grouped into food groups, namely fruits, vegetables, meat, fish, egg, bread/starch, seeds, milk, dairy products, fat, other carbohydrates, and soft drinks.

Regarding the FFQ, it comprises 142 questions on the consumption of foods that are commonly eaten by the Spanish population throughout the year, including dairy products, cereals, fruits, vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants and sweets. Using a 9-grade scale ("never or less than 1 time/month", "1-3 times/month", "1 time/week", "3-4 times/week", "5-6 times/week", "1 time/day", "2-3 times/day", "4-5 times/day", "≥6 times/day") participants were required to indicate the absolute frequency of consuming a certain amount of food, expressed in g, milliliters or in other common measures, such as slice, tablespoon or cup, depending on the food. The previous month was set as the timeframe.

Based on the FFQ-responses, adherence to the Mediterranean diet was evaluated by using the MedDietScore, a composite score calculated for each participant [79]. For food groups presumed to be part of the Mediterranean pattern (i.e., those with a recommended intake of 4 servings per week or more, such as non-refined cereals, fruits, vegetables, legumes, olive oil, fish, and potatoes) higher scores are assigned when the consumption is according to the rationale of the Mediterranean pattern, while lower scores are assigned when participants report no, rare, or moderate consumption. For the consumption of foods presumed to be eaten less frequently within the Mediterranean diet (i.e., consumption of meat and meat products, poultry, and full fat dairy products), scores are assigned on a reverse scale. As the sample of the study is lactating mothers, the original score was modified by removing the component regarding alcohol consumption. Thus, the range of this modified MedDietScore is between 0 and 50, with higher values of the score indicating greater adherence to the Mediterranean diet.

2.3.3. Standards and reagents

HPLC grade acetonitrile (ACN) (≥99.9%), ammonium formate (≥99.0%), formic acid (≥95%), phenylpropionylglycine (PPG) (≥99%), 3-indolepropionic acid (3-IPA) (≥99%), L-kynurenine (≥98%), 3-indoleacetic acid (3-IAA) (≥98%), hippuric acid (≥98%), ferulic acid sulphate (≥99%), proline betaine (≥99%), hesperetin (≥95%), phloretin (≥99%), quercetin (≥95%), kaempferol (≥97%), O-desmethylangolensin (O-DMA) (≥97%), daidzein (≥98%), equol (≥99%), glycitein (≥97%), genistein (≥98%), trimethylamine N-oxide (TMAO) (≥95%), isovalerylglycine (≥97%), isobutyrylglycine (≥95%), galactitol (≥99%), gallic acid (≥98%), and the "Amino Acid Standards, physiological" solution containing the amino acids L-tyrosine, 1-methylhistidine, 3-methylhistidine, anserine, citrulline, and taurine, as well as the isotopically labelled internal standards (IS) caffeine-D9 (≥99%), phenylalanine-D5 (≥99%), and taxifolin (≥95%) were purchased from Sigma-Aldrich Química SL (Madrid, Spain). Betaine-D11 (≥99%) was purchased from Cambridge Isotope Laboratories, Inc (Massachusetts, USA). Standard solutions were prepared in ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany).

2.3.4. UHPLC-MS/MS determination of nutrition and microbiota biomarkers

The UHPLC-MS/MS method for the quantification of 20 nutrition biomarkers (i.e., proline betaine, hesperetin, phloretin, quercetin, kaempferol, O-DMA, daidzein, equol, glycitein, genistein, 1-methylhistidine, 3-methylhistidine, anserine, TMAO, isovalerylglycine, isobutyrylglycine,

galactitol, gallic acid, citrulline, and taurine) and seven microbiota biomarkers (i.e., PPG, 3-IPA, L-kynurenine, 3-IAA, L-tyrosine, hippuric acid, and ferulic acid sulphate) and semi-quantification of 201 nutrition biomarkers of fruits and vegetables (N = 105), meat (N = 8), fish (N = 3), seeds (N = 17), olive oil (N = 4), coffee (N = 10), curcuma (N = 2), garlic and onion (N = 1), grains (N = 9), soft drinks (N = 3), alcoholic beverages (N = 13), and other groups (i.e., potato, cocoa, mushrooms, legumes, and nuts; N = 26) (see Supplementary Table S2.1) was developed based on previous results [101]. 110 µL of 1:20 (v/v) diluted urine samples in deionized water were added to a 96-well plate and mixed with 10 µL of IS (i.e., caffeine-D9, phenylalanine-D5, betaine-D11, and taxifolin at 15 µM each).

UHPLC-MS/MS analysis was conducted using a Sciex QTRAP 6500+ system (Sciex, Framingham, Massachusetts, USA) operating in positive and negative ionization modes (ESI+/-). Separations were performed using a Luna Omega Polar C18 column (100 mm x 2.1 mm, 1.6 µm) equipped with a fully porous polar C18 security guard cartridge (Phenomenex, Torrance, California, USA). Conditions were as follows: column temperature, 40 °C; autosampler temperature, 10 °C; injection volume, 10 µL. 0.1% formic acid and 10 mM ammonium formate in water and pure ACN were used as aqueous (A) and organic (B) mobile phases (MP), respectively. The gradient program was as follows: 0−8 min, 5−20% B; 8−10 min, 20−100% B; 10−12 min, 100% B; 12−12.1 min, 100−5% B; and 12.1−14 min, 5% B.MS detection was performed using the multiple reaction monitoring (MRM) mode (see Supplementary Table S2.2)). The mass spectrometer operated using the following parameters: ion spray voltage, (±)4500 V; source temperature, 600 °C; curtain gas, 35 psi; ion source gases 1 and 2, 60 psi each; and entrance potential, (±)10 V. When available, the transitions were optimized by infusing 5 µM individual solutions of commercial standards dissolved in MP into the

mass spectrometer. For the semi-quantification of nutrition biomarkers identified in urine samples for which authentic standards were not available, their corresponding transitions, as reported in the literature [101], were included (see Supplementary Table S2.1) and it was carried out using the relative peak areas. Quantification was carried out using the relative peak areas with an external calibration line using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. For monitoring the instrument's performance, a quality control (QC) sample (5 µL of each sample pooled) was analysed every 20 samples in the randomized analytical batch. The batch acceptance criterion was QC relative standard deviation (RSD) < 25%. In addition, calibration blanks (i.e., H_2O) and a process blank (i.e., processing of H_2O as described for urine samples) were injected at the beginning of the sequence for system suitability testing.

Urinary biomarker concentrations were normalized to creatinine concentrations quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA). Samples were diluted with deionized water prior to measurements employing a 1:20 (v/v) dilution.

2.3.5. Method validation

Method validation was based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [2], including the following bioanalytical parameters: accuracy, precision, linearity range, carryover, selectivity, specificity, and stability.

Replicates (N = 3) of standards at three concentration levels and replicates (N = 3) of spiked samples at three concentration levels (low, medium, and high) on three validation days were analysed to assess accuracy. RSD of replicate standards/samples within one validation batch (intra-day) and between validation batches (inter-day) estimated precision. The calibration curves included a zero calibrator (i.e., blank with IS) and, at least, six standards covering the selected concentration ranges ensuring linearity. Carryover between samples was assessed by the analysis of a calibration blank after the injection of the standards. The analysis of calibration and process blank samples from multiple (n=6) sources were used to demonstrate selectivity and specificity. Analytes' freeze-thaw and long-term stabilities were tested by comparing concentrations observed in a freshly prepared sample to sample extracts after three freeze-thaw cycles and in a sample stored for six months (-80 ℃).

2.3.6. Data availability and statistical analysis

UHPLC-MS/MS data were acquired and processed using SCIEX OS Software (Sciex, Framingham, Massachusetts, USA). Data analysis was carried out in MATLAB R2019b (MathWorks, Natick, MA, USA) and using the PLS Toolbox 8.9 (Eigenvector Research Inc., Manson, WA, USA). Biomarker levels normalized to creatinine in urine samples determined in this work are available in Supplementary Table S2.3. Continuous variables were expressed as mean ± standard deviation or medians with interquartile range, depending on underlying data distribution. Student's t-test ($\alpha = 0.05$) or Wilcoxon rank-sum test ($\alpha = 0.05$) were used for inter-group comparison.

Pearson's and Spearman's correlation coefficients were used for assessing paired associations among metabolite concentrations, and among metabolite concentrations and R24h food groups, respectively. Hierarchical clustering analysis was carried out using autoscaled data.

Diversity of microbiota biomarkers within a sample was assessed with Simpson's Index of Diversity (1-D), where D was defined as:

$$D = \frac{\sum n(n-1)}{N(N-1)} \quad (1)$$

being n the concentration of a particular microbiota activity biomarker in the sample; and N the total concentration of microbiota activity biomarkers. Thus, the higher the value for this index (1-D), the higher the microbiota activity biomarker diversity.

2.4. Results and discussion

2.4.1. UHPLC-MS/MS method validation

Recommended guidelines were used to perform the validation of the analytical method [2]. Linearity of response was assessed covering up to four orders of magnitude with limits of detection (LODs) and limits of quantification (LOQs) in the 0.03-8 μM range (see Supplementary Table S2.2) and carryover did not exceed 20% of LOQ. Regarding accuracy and precision, intra- and inter-day recoveries in standard solutions spiked at the different

concentration levels were between 82 and 119%, with an average precision of 7% (min-max precision of 1-20%) (see Supplementary Table S2.4). Similarly, in spiked urine samples, intra-day and inter-day recoveries and precisions ranged between 80 and 120% and between 1 and 20% (average precision of 10%), respectively, with exception of galactitol and genistein spiked at the low concentration level, that presented an inter-day recovery of 123% and an inter-day accuracy %RSD of 21, respectively. This evidences an adequate method performance for all quantified metabolites. Additionally, all compounds were stable after freezing stock solutions for one year at -80 ℃, and after three freeze-thaw cycles (-80 ℃) (t-test, *p*-values > 0.05). Process blanks using sterile PP containers were analyzed and compared to calibration blanks to assess the selectivity and specificity, as well as test the compatibility of the sample collection procedure with the analysis of these metabolites, and no contaminations were detected.

2.4.2. R24h results

As shown in **Figure 2.1**, results from R24h evidence that mothers' intake of fruits, vegetables, meat, milk, egg, fish, dairy, and other carbohydrates ranged between 1 and 5 portions/day, while bread/starch and fat were the food groups of higher intakes, with 5 to 10 portions/day. However, seeds, and soft drinks were rarely present in their diets. The majority of the mothers participating in the study reported lower intake of fruits and vegetables, meat and fish, bread/starch, and milk, and dairy products than the recommended intake according to the World Health Organization diet guidelines for lactating mothers (5, 2, 8.5, and 3 portions/day, respectively) [142]. Additionally, results from FFQ indicate that 51 of the 59 mothers (86%) included in the study presented adherence to the Mediterranean diet (i.e.,

MedDietScore between 25 and 35), while 8 women showed lower adherence (7 and 1 participants < 35 and < 25, respectively).

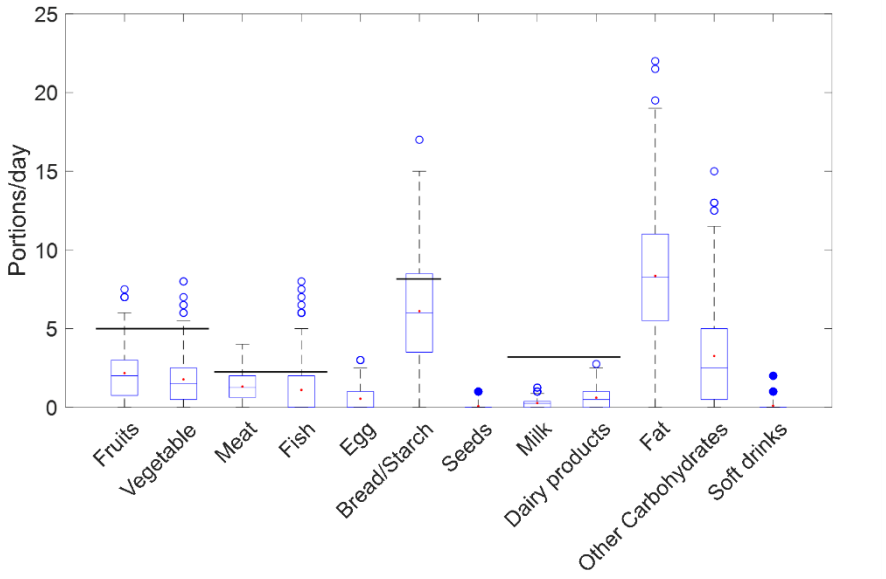


Figure 2.1. R24h food groups intake in lactating mothers. Note: Horizontal black line represents recommended intake according to the World Health Organization diet guidelines for lactating mothers.

2.4.3. Biomarker profiles of lactating mothers

The concentrations of the nutrition and microbiota biomarkers determined in urine samples, after normalization to creatinine using the validated UHPLC-MS/MS method are shown in **Table 2.2**.

Table 2.2. Microbiota and nutrition biomarker concentrations in urine samples from lactating mothers.

Category	Metabolite	Range (µmol/g creat)	Median (µmol/g creat)	IQR (25-75)	Detection frecuency (%)
Microbiota	Phenylpropionylglycine	0.007 - 58	0.03	0.4	72
	3-IPA	0.012 - 8	2	3	3
	L-Kynurenine	0.009 - 21	0.5	2	80
	3-IAA	0.012 - 176	1.3	7	99
	L-Tyrosine	0.014 - 2829	16	43	98
	Hippuric Acid	11 - 8496	255	598	100
	Gallic Acid	0.013 - 10	0.4	0.4	62
	Ferullic Acid Sulphate	0.02 - 126	0.4	3	74
Fruits	Proline betaine	1.4 - 2568	88	315	100
	Hesperetin	0.02 - 11	0.05	0.13	64
	Phloretin	0.004 - 2	0.02	0.10	29
Vegetables	Quercetin	0.03 - 76	0.4	0.2	48
	Kaempferol	0.02 - 2	0.2	0.6	93

Table 2.2 (continuation). Microbiota and nutrition biomarker concentrations in urine samples from lactating mothers.

Category	Metabolite	Range (µmol/g creat)	Median (µmol/g creat)	IQR (25-75)	Detection frecuency (%)
Seeds	O-DMA	0.007 - 5	0.02	0.3	61
	Daidzein	0.005 - 3	0.4	0.9	96
	Equol	0.012 - 10	0.3	0.05	34
	Glycitein	0.006 - 1	0.05	0.2	25
	Genistein	0.4 - 2	0.03	0.02	2
Meat	1-Methylhistidine	4 - 1784	48	95	100
	3-Methylhistidine	0.2 - 4555	31	87	100
	Anserine	0.6 - 304	2	10	88
Fish	TMAO	3 - 3368	142	473	100
Dairy products	Isovalerylglycine	0.005 - 91	3	13	87
	Isobutyrylglycine	0.7 - 125	3	9	93
Milk	Galactitol	1.2 - 2424	22	67	92
Soft drinks	Citrulline	0.09 - 60	7	11	95
	Taurine	0.2 - 2341	55	207	100

Note: 3-IPA = 3-indolepropionic acid; 3-IAA = indole-3-acetic acid; O-DMA = O-desmethylangolensin; TMAO = trimethylamine N-oxide; IQR = interquartile range.

Results show that, overall, median concentrations of microbiota biomarkers were relatively low, with exception of hippuric acid and L-tyrosine, that stood out over the rest. As observed, a high variability (two to four orders of magnitude) was encountered for these two microbiota biomarkers. It is noteworthy that 3-IPA, an end product of tryptophan

metabolism, was only detected in 3% of the samples, whereas the other microbiota biomarkers were detected in over 62% of the samples. For targeted analysis of nutrition biomarkers, the highest concentrations were found for TMAO, proline betaine, taurine, and 1- and 3-methylhistidine, being phloretin, O-DMA, genistein, hesperetin, and glycitein the less abundant BFIs. As for the microbiota biomarkers, broad concentration ranges for the more prevalent metabolites were observed while lower detection frequencies for certain low abundant biomarkers (e.g., genistein, phloretin, and glycitein) were found.

Figure 2.2 depicts the significant paired correlations among BFIs (*p*-value < 0.05). Some metabolites that are biomarkers of the same food group are significantly correlated among each other, such as 1- and 3-methylhistidine (meat); isovalerylglycine and isobutyrylglycine (dairy products); quercetin, phloretin and kaempferol (fruits and vegetables); and genistein, equol and glycitein (seeds). However, additional correlations between biomarkers from different food groups were also significant, such as meat BFI anserine with fruits and vegetables BFIs (i.e., quercetin, phloretin and kaempferol), soft drink BFI taurine with phloretin, and some meat BFIs (i.e., 1- and 3-methylhistidine) with quercetin. This could be due to the paired intake of both food groups or chance correlations.

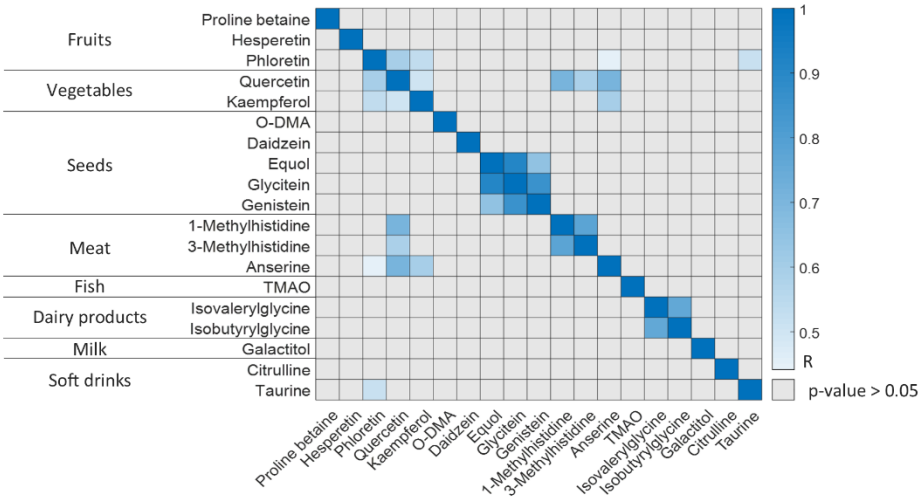


Figure 2.2. Significant paired correlations among BFIs (Pearson's correlation, *p*-value < 0.05).

Regarding the semi-quantitative analysis, a higher number of detected metabolites belonged to fruits and vegetables, followed by other groups (i.e., potato, cocoa, mushrooms, legumes, and nuts), coffee, meat, and seeds, as shown in **Figure 2.3**.

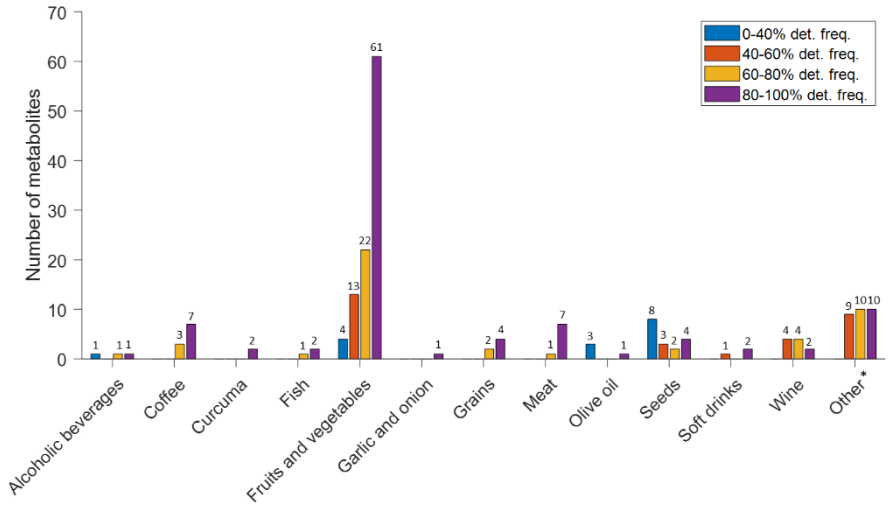


Figure 2.3. Number of metabolites and detection frequencies of the semi-quantitative analysis. Note: (*) Potato, cocoa, mushrooms, legumes and nuts.

2.4.4. R24h and BFIs in lactating mothers

In **Figure 2.4** (top, left), BFI patterns in mothers' urine samples detected by hierarchical clustering analysis (HCA) are shown. Three clusters were identified as clusters 1, 2, and 3, that were depicted in the Principal Component Analysis (PCA) scores plot in **Figure 2.4** (top, right). Samples from clusters 1 and 3 presented higher concentrations of all biomarkers than samples from cluster 2, while dairy product and milk biomarkers were more concentrated in cluster 1, and seeds and garlic and onion BFIs in cluster 3 as shown in the PCA loadings plot in **Figure 2.4** (bottom, left).

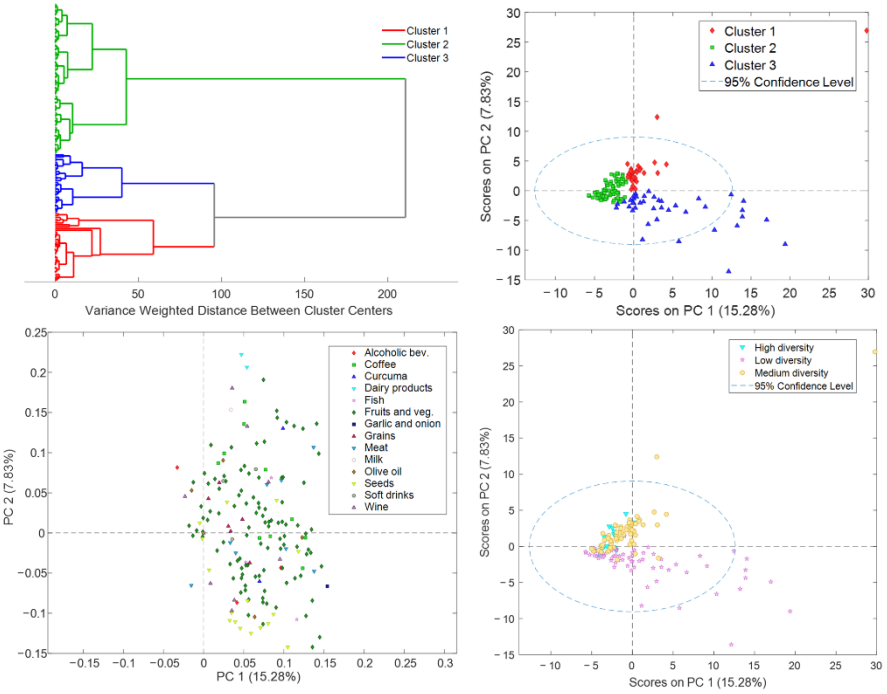


Figure 2.4. BFIs patterns in mothers' urine samples. Top: Hierarchical clustering analysis revealing three sub-groups within the study samples (left) and PCA scores plot (right). Bottom: loadings plot (left) and PCA scores plot with Simpson's Index of Diversity classes (right).

Figure 2.5 depicts the mean R24h food group values for each of the three clusters detected by analysing BFI patterns in urine samples. It shows that mothers from cluster 1 had a higher intake of some food groups (i.e., vegetables, meat, egg, and other carbohydrates) than mothers from cluster 3, who presented a lower intake, evidencing that they followed a different dietary pattern. Conversely, mothers in cluster 3 had highest intakes of fruits, fish, bread/starch, seeds, milk, dairy products, fat, and soft drinks. Mothers from

cluster 2 showed a mean intake in between mothers from clusters 1 and 3 for all food classes, except for other carbohydrates, egg, bread/starch, and seeds.

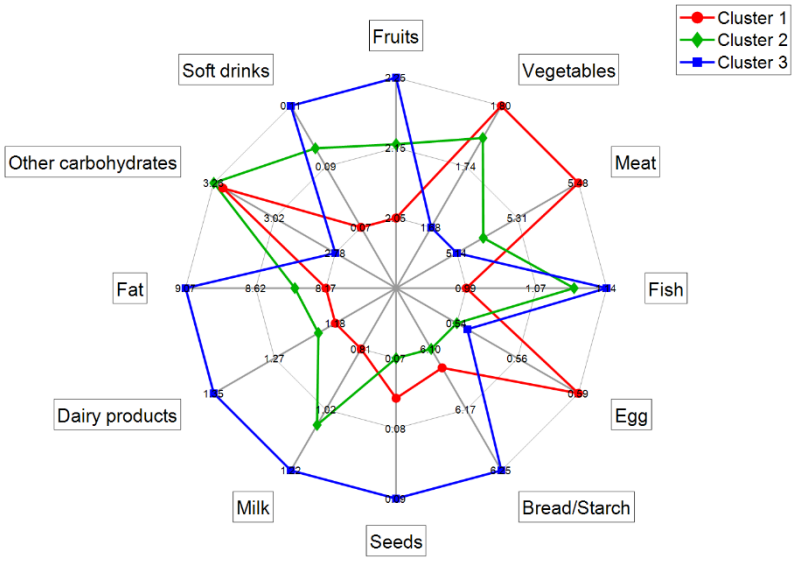


Figure 2.5. Mean R24h food group values for clusters 1 to 3.

In concordance with the results from the HCA, clusters 1 and 3 present higher intakes of the majority of food groups and cluster 3 is characterized by a high intake of seeds. However, R24h results show that the class with lowest consumption of milk and dairy products is cluster 1, contradictory to HCA's results. These divergences could be due to the small absolute difference in the mean R24h food group values between clusters, as the majority of the mothers follow a similar Mediterranean diet (p-value > 0.05, Wilcoxon rank-sum test), and/or errors due to the subjective nature of the R24h recording. Also, we would like to highlight the complementary information provided by R24h food groups and BFIs, as the information provided by BFIs is more detailed reporting food groups not specified in R24h (e.g., garlic and onion, alcoholic

beverages, coffee, curcuma, olive oil, wine). Also, these differences hamper the comparison of the information accessible through both R24h and BFIs.

Correlations between urinary BFIs and the food groups obtained in the R24h are depicted in **Figure 2.6**, and show significant correlations between three biomarkers (i.e., proline betaine, anserine, and TMAO) and their corresponding food groups (i.e., fruits, meat, and fish) from the R24h, with a p-value < 0.01 for fruits, and p-values < 0.001 for meat and fish ($r > 0.2$, Spearman's correlation). These results strengthen findings of Lloyd et al. reporting urinary excretion of proline betaine as a potentially useful biomarker of habitual citrus consumption [143], and conclusions from Cheung et al. in which anserine and TMAO are presented as potential biomarkers of chicken and fish intake, respectively [144]. Additionally, there are three other metabolites that are significantly correlated to meat (i.e., phloretin, kaempferol, and taurine), which are described as fruit, vegetable and soft drink BFIs [145,146]. This could be due to a paired intake of vegetables and meat (see **Figure 2.5**) and/or due to the correlation of taurine levels with meat intake reported in other studies, as ingestion of foods with animal origin can be another source of taurine [147–149]. It should be remarked that not all food groups recorded in the R24h have their corresponding urinary BFI (i.e., egg, bread/starch, fat and other carbohydrates), due to the difficulty in finding specific BFIs [150].

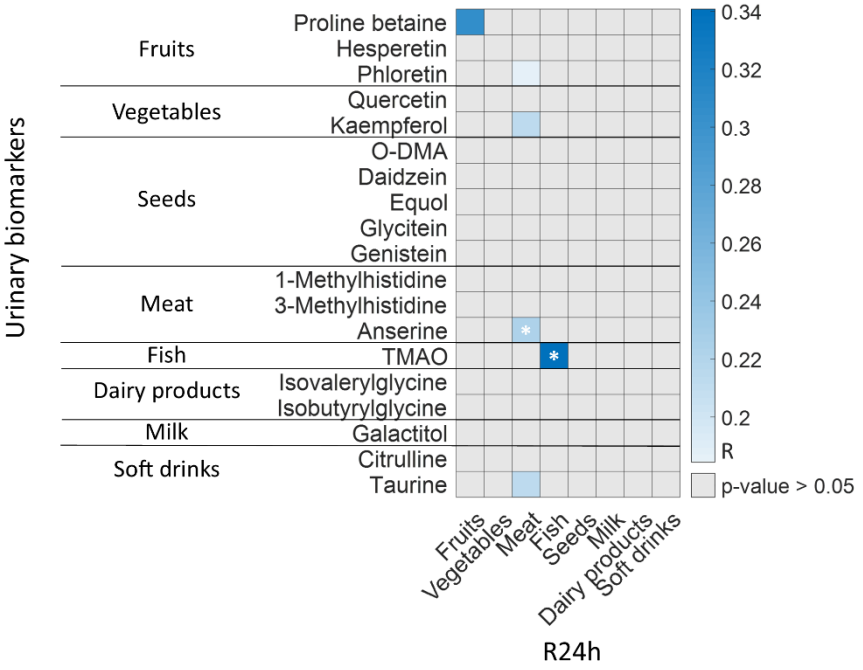


Figure 2.6. Significant correlations among BFIs and R24h food groups (Spearman's correlation, *p*-value < 0.05). Note: (*) *p*-value < 0.001.

On the other hand, the specificity of most of the already proposed BFIs is limited, as many dietary compounds are present in different types of foods and similar compounds from different food sources can result in the same metabolite [150]. There are additional concerns, such as the lack of knowledge about the dose-response relationships, the limited quantitative data available, the interindividual variation that can lead to different BFIs levels with the same intake levels [150,151], and the fact that some metabolites are derived not only from diet, but also from other parallel endogenous pathways affected by intake [152]. However, self-reported surveys are known to be

prone to misreporting issues and, therefore, can lead to uncertain research findings. Consequently, due to the high complexity of accurate dietary assessments, the implementation of a combined strategy bringing together complementary tools might allow to achieve an enhanced performance.

Despite the fact that mothers included in the study are similar ages, the same diet could lead to different metabolite by-products due to different metabolism between individuals, as recent studies have found significant associations between food intake metabolites and body mass index (BMI) [153,154]. However, in this study we did not find any significant correlation between mother's BMI and their food intake (p-value > 0.05, Spearman's correlation). In this study, repeated sampling from the same individual was performed, corresponding to the collection of urine samples and R24h at different study time points, but it has to be taken into consideration that lactating mothers are a particular population, as the body readjusts after giving birth, usually resulting in a weight reduction.

2.4.5. Microbiota activity biomarkers in lactating mothers

The diversity of microbiota activity biomarkers was estimated through the Simpson's Index of Diversity (Equation 1) and ranged between 0 and 0.6 with a median value of 0.08. Samples were classified as low (N = 111), medium (N = 60), and high (N = 37) diversity for Simpson's Index of Diversity under 0.05, between 0.05 and 0.25, and over 0.25, respectively. In **Figure 2.4** (bottom, right), a PCA scores plot of the urinary BFIs concentrations with the microbiota activity biomarker diversity classes assigned is shown. Samples classified as high and medium diversity presented

different BFI levels from the samples classified as low diversity, reflecting a nexus between dietary patterns and microbiota activity. This corroborates findings reported by Turnbaugh et al. indicating that diet has a significant impact on sculpting the microbial communities in the gut, and changes in dietary patterns can directly influence the composition and functionality of the gut microbiota, through the availability of macro- and micronutrients in the gut [155]. More specifically, the high number of samples from dietary cluster 3 that show a low microbiota activity diversity (43%) is noticeable. However, BFIs and microbiome analysis should be assessed with more detail in further studies on the impact of diet on the gut microbiome.

2.5. Conclusions

A UHPLC-MS/MS method for the analysis of BFIs and microbiota activity biomarkers was developed and validated. Levels of microbiota activity biomarkers were reported and a panel of both quantitative and semi-quantitative BFIs was determined in an observational study involving lactating mothers. Correlations between some BFIs of the same food group were found (i.e., meat, dairy products, fruits and vegetables, and seeds). Additionally, the BFI profile allowed to discern three clearly distinct dietary patterns present in the study population evidencing the capability of urinary food intake assessment as a complementary tool to traditionally employed questionnaires. The presence of some highly specific BFIs in urine samples (e.g., curcuma and garlic and onion) highlights its complementary nature. Correlations between fruit, meat, and fish biomarker concentrations and R24h results were found. This evidences that the use of some biomarker measurements for the assessment of some food group intakes is possible, even

though more studies are needed in order to expand the number of food groups that can be assessed by BFIs. Furthermore, the diversity of microbiome activity biomarkers reflected dietary patterns detected in lactating mothers. In future studies, biomarkers and microbiome analysis will be integrated for a joint assessment of diet, microbiome, and ultimately health.

Supplementary materials

The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/nu15081894/s1.

Chapter 3. GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers

3.1. Abstract

Short chain fatty acids (SCFAs) and branched chain amino acids (BCAAs) are frequently determined in faeces, and widely used as biomarkers of gut-microbiota activity. However, collection of faeces samples from neonates is not straightforward, and to date levels of these metabolites in newborn's faeces and urine samples have not been described. A targeted gas chromatography – mass spectrometry (GC-MS) method for the determination of SCFAs and BCAAs in both faeces and urine samples has been validated. The analysis of 210 urine and 137 faeces samples collected from preterm (PI), term infants (TI) and their mothers was used to report faecal and urinary SCFAs and BCAAs levels in adult and neonatal populations.A significant correlation among five SCFAs and BCAAs in faeces and urine samples were observed, reference ranges of SCFAs and BCAAs in mothers, PI and TI were reported showing infant's lower concentrations in faeces and higher concentrations in urine. In conclusion, this method presents a non-invasive approach for the simultaneous assessment of SCFAs and BCAAs in faecal and urine samples and the results will serve as a knowledge base for future experiments that will focus on the study of the impact of nutrition on the microbiome of lactating mothers and their infants.

3.2. Introduction

The host-gut microbiota crosstalk is vital for the development and maturation of the immune system [156–158] and it has been widely acknowledged that this crosstalk is regulated by nutrition [159]. Gut microbiota provides essential nutrients, such as vitamins, and helps catabolizing dietary fibres, human milk oligosaccharides, and other carbohydrates to produce short chain fatty acids (SCFAs), which are an important energy source for the intestinal epithelium, and modulate epithelial integrity [160]. These compounds provide general protection against pathogen gut colonization [161]; moreover, taken up by the host, they stimulate the development of the immune system and are used as energy sources or regulators [162], being particularly relevant in the perinatal period.

The characterization of SCFAs profiles in faeces is widely used to study the activity of gut microbiota in clinical studies. The most abundant SCFAs are acetate, propionate, and butyrate produced by anaerobic fermentation in the intestine. Valerate, caproate, heptanoate, and branched chain SCFAs (i.e., 2-methyl-butyrate, isobutyrate, and isovalerate) are also produced by intestinal microbiota, but in considerably lower amounts. Branched chain SCFAs are derived from the bacterial metabolism of dietary proteins, human milk oligosaccharides and branched chain amino acids (BCAAs) [156,162–164]. BCAAs (i.e., valine, leucine, and isoleucine) are considered essential amino acids thus they that cannot be synthesized endogenously and must be provided by the diet. They serve as substrates for protein synthesis or energy production and are of key importance in several metabolic and signalling functions, e.g. the activation of the mammalian target of rapamycin (mTOR) signalling pathway [165].

Mass spectrometry (MS) hyphenated with separation techniques, such as gas chromatography (GC) or liquid chromatography (LC), is frequently used for the metabolic analysis of biological samples due to its sensitivity and selectivity. LC-MS requires minimal sample preparation for SCFA and BCAA analysis. However, harsh experimental conditions, such as a very acidic aqueous mobile phases, need to be employed in LC-MS for the quantitation of SCFAs without chemical derivatization [166]. In addition, poor chromatographic separation, and insufficient ionization due to the hydrophilicity of SCFAs are an additional burden [167]. To overcome these problems, several chemical derivatization methods have been developed for SCFAs and BCAAs quantification with LC-MS [168–170], requiring long reaction times or specific reaction conditions. Alternatively, SCFAs and BCAAs have been derivatized by methyl-, ethyl-, and propyl-chloroformate, as well as trimethylsilylation, and determined by GC-MS [99,171–173].

Preterm birth is a serious event affecting 5-13 % of births with an increasing tendency in recent years [174]. Prematurely born infants are characterized by an immature immune system and gut barrier, and hence, they are more vulnerable to suffer infections and inflammatory processes such as necrotizing enterocolitis, associated with high mortality rates [175]. In studies targeting the newborn population, the collection and handling of stool samples can be troublesome, and the determination of absolute concentrations typically requires the normalization of the detected concentrations to wet or dry weight of faecal samples. Hence, alternative procedures supporting straightforward sample collection and enabling the quantitative analysis of these biomarkers are warranted. The assessment of newborns' gut microbiota activity using metabolic profiles of urine samples as surrogates could enable an effective simplification of sample collection protocols in clinical studies, which would be especially useful in neonatology.

In the present study we show the applicability of a novel GC-MS method for the determination of SFCAs and BCAAs in both faecal and urine samples, specifically tailored to the very low sample volumes typically available in studies involving newborns. The method encompasses a 1-step derivatization of SCFAs and BCAAs to propyl-esters and provides appropriate sensitivity, linearity, and accuracy for their quantification. The applicability of the method for the determination of SCFA and BCAA concentration levels is assessed through the analysis of a set of 137 faecal and 210 urine samples from a clinical cohort comprising lactating mothers and term (TI) and preterm infants (PI), providing for the first time reference concentration ranges of these metabolites, and an initial study of the association between the levels found in faeces and urine in paired samples. Results obtained during method development include the pre-screening of sterile polypropylene (PP) bags, cotton pads, and gauzes widely used for urine collection from the newborn for assessing their suitability with respect to the determination of SCFAs and BCAAs in future clinical studies.

3.3. Material and methods

3.3.1. Standards and reagents

Certified ACS grade sodium hydroxide, HPLC grade propanol (PrOH), pyridine (Py), propyl-chloroformate (PCF), and n-hexane, and SCFAs (acetic acid (≥ 99.8%), propionic acid (≥ 99.5%), isobutyric acid (≥ 99%), butyric acid (≥ 99%), 2-methyl butyric acid (≥ 98%), isovaleric acid (≥ 99%), valeric acid (≥ 99%), caproic acid (≥ 99%), and heptanoic acid (≥

99%)), and BCAAs (valine (≥ 99.5%), leucine (≥ 99.5%), and isoleucine (≥ 99%)) standards as well as the isotopically labelled internal standard (IS) caproic acid-D3 (≥ 99%) were purchased from Sigma-Aldrich Química SL (Madrid, Spain). Standard solutions were prepared in ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany).

3.3.2. Study design, population, and sample collection

Samples were collected in the frame of the Nutrishield project (https://nutrishield-project.eu/), in a prospective, observational, cohort study performed at the Division of Neonatology of the University and Polytechnic Hospital La Fe (HUiP La Fe), including three groups of infants: i) PI (<32 weeks of gestation) exclusively (i.e., >80% v/v) receiving own mother's milk (N=30), ii) PI exclusively receiving donor human milk (N=7) and iii) TI (>37 weeks of gestation) exclusively receiving own mother's milk (N=7), as well as mothers of infants of all three groups (N=45). Urine and faeces samples from infants and mothers were collected at six time points, covering the period from achieving complete enteral nutrition, i.e., a stable daily intake of 150 mL/kg, in PI and recovery of birth weight in TI, followed by sampling at one, two, three and six months of infant's age. Additional faeces and matched urine samples from neonates (N=15) during their stay at the hospital were collected. A total of 45 mothers, 50 PI and 9 TI were included in the study. Demographic, clinical, and perinatal characteristics from participants were recorded as shown in **Table 3.1**. Mother's first morning urine was collected in sterile PP containers and infants' urine was collected using sterile cotton pads placed in the diaper. Cotton pads were collected after 1 h and squeezed with a sterile PP syringe. The process was repeated until collecting a minimum of 1 mL. Urine

samples were aliquoted to avoid freeze-thawing cycles and stored at -80 ºC until analysis. Faeces samples were collected in sterile tubes using sterile tweezers. In total, 210 urine and 137 faeces samples collected were employed for method development and for the comparison between SCFA and BCAA concentrations in faeces and urine.

The study protocol was approved by the Scientific and Ethics Committee for Biomedical Research (CEIm) of the HUiP La Fe (study approvals #2019-289-1 and #2019/0312). All methods were performed in accordance with the relevant guidelines and regulations and written permission was obtained from mothers or legal representatives by signing an informed consent form.

Table 3.1. Characteristics of the population and general confounders.

Parameter	**M (N=45)**	**TI (N = 9)**	**PI (N = 50)**	***p*-value**
Maternal age (years)/Gestational age (weeks), median (SD)/(5-95% CI)	35 (5)	37 (1.2)	30 (0.9)	< 0.001
Maternal weight (kg)/ infant birth weight (g), mean (SD)	67 (12)	2486 (776)	1420 (445)	< 0.001
BMI (kg/m^2), median (SD)	26 (7)	-	-	-
Sex male, n (%)	0 (0)	8 (89)	29 (58)	0.09
C-Section delivery, n (%)	22 (48)	4 (44)	38 (76)	0.06
Apgar[1] 1 min (median; 5-95% CI)		9 (1.4)	7 (0.8)	0.6
Apgar[1] 5 min (median; 5-95% CI)		10 (1.2)	9 (0.4)	0.8
Antibiotic therapy, n (%)	4 (9)	3 (38)	2 (5)	0.006

Note: Body mass index (BMI) = weight/height^2; SD = standard deviation; CI = confident interval; M = mothers; TI = term infants; PI = preterm infants; [1][176].

3.3.3. GC-MS determination of SCFAs and BCAAs

The GC-MS method for the determination of SCFAs (i.e., acetic acid, propionic acid, butyric acid, valeric acid, caproic acid, heptanoic acid, isobutyric acid, 2-methylbutyric acid, and isovaleric acid) and BCAAs (i.e. valine, leucine, and isoleucine) was developed based on previous results [99]. 1000 µL of 5 mM aqueous NaOH containing IS (5 µL mL-1 caproic acid-D3) was added to 100 mg of wet faeces, then the sample was homogenized for 10 min and centrifuged (13200 x g for 20 min at 4 °C). 500 µL of faecal supernatant was transferred into a 15 mL falcon tube, and 300 µL of water were added. For urine, 300 µL of sample and 500 µL of 5 mM aqueous NaOH containing IS (5 µL mL-1 caproic acid-D3) were mixed in a 15 mL falcon tube. Both faecal and urine samples were derivatized as follows. An aliquot of 500 µL PrOH/Py solvent mixture (v/v = 3:2) and 100 µL of PCF were added and vortexed briefly. Derivatization was carried out during 1 min in an ultrasonic water bath prior to a two-step extraction by adding 300 and 200 µL of n-hexane, respectively, followed by centrifugation (2000 x g for 5 min at 25 °C). The upper n-hexane layers containing the extracted derivatives were collected and pooled followed by thorough mixing during 3 s prior to analysis.

GC-MS analysis was conducted using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole mass spectrometric detector (Agilent Technologies, Santa Clara, CA, USA) operating in selected ion monitoring (SIM) mode. Separations were performed using an HP-5 MS capillary column coated with 5% phenyl-95% methylpolysiloxane (30 m x 250 µm i.d., 0.25 µm film thickness, Agilent J & W Scientific, Folsom, CA, USA). One microliter of derivatized sample extracts was injected in split mode with a ratio of 10:1, and the solvent delay time was set to 2.36 min. The initial oven temperature was held at 50 °C for 2 min, ramped to 70 °C at a rate of 10 °C

min^{-1}, to 85 ℃ at a rate of 3 ℃ min^{-1}, to 110 ℃ at a rate of 5 ℃ min^{-1}, to 290 ℃ at a rate of 30 ℃ min^{-1}, and finally held at 290 ℃ for 8 min. Helium was used as a carrier gas at a constant flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source were set at 260, 290, and 230 ℃, respectively, and the electron energy was -70 eV.

The measurement parameters used for the studied analytes are summarized in **Table 3.2**. For quantification, an external calibration was carried out using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. Aliquots of a quality control (QC) sample of each matrix were analysed every 10 samples in the randomized analytical batch for monitoring the instrument's performance. QC RSD < 25% was the batch acceptance criterion. In addition, calibration blanks (i.e., addition of H_2O instead of sample into the tubes) and a process blank (addition of H_2O instead of sample to a PP bag/container, cotton or gauze pad and squeezed before adding it to the tubes) were injected.

SCFA and BCAA concentrations in faecal and urine samples were normalized by dividing by sample weight and creatinine concentration, respectively. Creatinine was quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA). Samples were diluted with deionized water prior to measurements employing a 1:20 and a 1:4 dilution for adult and infant urine, respectively.

Table 3.2. Measurement parameters and main figures of merit of the GC-MS method.

Compound class	Metabolite	*m/z*	RT ± s (min)	Faeces				Urine			
				Calibration range (µM)	**R^2**	**LOD (µM)**	**LOQ (µM)**	**Calibration range (µM)**	**R^2**	**LOD (µM)**	**LOQ (µM)**
SCFA	Acetic acid	61	2.75 ± 0.02	2 - 8593	0.9996	0.3	1	0.7 - 655	0.997	0.5	2
	Propionic acid	75	3.98 ± 0.02	0.5 - 8927	0.998	0.1	0.2	0.6 - 547	0.994	0.6	2
	Isobutyric acid	89	4.763 ± 0.012	0.5 - 8149	0.998	0.12	0.4	0.6 - 609	0.996	0.12	0.4
	Butyric acid	89	5.59 ± 0.02	0.10 - 1619	0.999	0.05	0.2	0.2 - 171	0.997	0.11	0.4
	2-Methylbutyric acid	103	6.733 ± 0.014	0.08 - 1332	0.998	0.07	0.2	0.13 - 121	0.996	0.1	0.3
	Isovaleric acid	85	6.83 ± 0.03	0.05 - 783	0.9997	0.02	0.07	0.07 - 65	0.998	0.05	0.2
	Valeric acid	103	8.24 ± 0.03	0.07 - 1077	0.9996	0.08	0.3	0.10 - 98	0.997	0.05	0.2
	Caproic acid	117	11.22 ± 0.02	0.08 - 1309	0.9998	0.01	0.04	0.10 - 98	0.998	0.04	0.13
	Caproic acid-D_3 (IS)	120	11.281 ± 0.014	-	-	-		-	-	-	-
	Heptanoic acid	131	14.31 ± 0.03	0.05 - 896	0.998	0.012	0.04	0.05 - 52	0.998	0.02	0.05
BCAA	Valine	158	17.843 ± 0.012	0.14 - 2328	0.996	0.2	0.8	1.0 - 908	0.998	0.5	2
	Leucine	172	18.164 ± 0.012	0.13 - 2165	0.997	0.112	0.4	0.9 - 844	0.997	0.5	2
	Isoleucine	172	18.240 ± 0.015	0.12 - 1926	0.9995	0.11	0.2	0.8 - 751	0.998	0.4	1.3

Note: RT = retention time; R = coefficient of determination; Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20%, taking into account the preconcentration factor achieved during sample processing; Limit of detection (LOD) = 3/10*LOQ.

3.3.4. Method validation

The method validation was based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [2], including the following bioanalytical parameters: linearity range, accuracy, precision, recovery, selectivity, specificity, carry over, and stability. The ranges of the analytical method were selected to enable the quantification of the metabolites in faeces and urine samples with adequate precision, accuracy, and linearity. **Table 3.2** summarizes the employed working concentration ranges, which were chosen considering the expected values and a wide inter- and intra-individual variability. The calibration curves included a zero calibrator (i.e., blank with IS) and, at least, 6 standards covering the selected concentration ranges. Accuracy, precision, and recovery were assessed by replicate (n=3) analysis of standards at three concentration levels and replicate (n=3) analysis of spiked samples at three concentration levels (low, medium, and high) on three validation days. Precision was estimated as the percentage of relative standard deviation (RSD) of replicate standards within one validation batch (intra-day) and between validation batches (inter-day). Selectivity and specificity were demonstrated by analyzing calibration and process blank samples from multiple (n=6) sources. Carry-over between samples was assessed by the analysis of zero-injections after the analysis of high concentrated standards and spiked samples (n=3), and by the analysis of a blank without analytes or IS after the injection of the standards. Autosampler sample stability was assessed by comparing concentrations observed in a freshly prepared sample and in the same processed sample after 20 h stored in the autosampler (sealed vial, 25 ℃). Analytes' freeze-thaw stability and long-term stability were established by comparing concentrations observed in

sample extracts after three freeze-thaw cycles and in a sample stored for one year (-80 ℃), respectively, to a freshly prepared sample.

3.3.5. Data availability and statistical analysis

GC-MS data were acquired and processed using MassHunter B.07.01 (Agilent Technologies, Santa Clara, CA). Data analysis was carried out in MATLAB R2019b (MathWorks, Natick, MA, USA) and using the PLS Toolbox 8.9 (Eigenvector Research Inc., Manson, WA, USA). SCFAs and BCAAs normalized levels in faeces and urine samples determined in this work are available as Supplementary Tables 3.1 and 3.2. Categorical variables were compared using Pearson's $\chi 2$ test ($\alpha = 0.05$). Continuous variables were expressed as mean ± standard deviation or medians with interquartile range, depending on underlying data distribution and Student's t-test ($\alpha = 0.05$) or Wilcoxon rank-sum test ($\alpha = 0.05$) were used for inter-group comparisons, respectively. The Pearson's correlation coefficient was used for assessing paired associations among metabolite concentrations. Principal Component Analysis (PCA) was carried out using autoscaled, normalized data using creatinine and wet weight as normalization factor for urine and faeces, respectively.

3.4. Results and discussion

3.4.1. Method validation

Analytical method validation was performed following recommended guidelines [2]. Linearity of response was assessed covering up to four orders

of magnitude with LODs and LOQs in the 0.02-0.6 and 0.05-2 μM ranges in urine, and in the 0.01-0.3 and 0.04-1.0 μM ranges in faeces, respectively (see **Table 1.2**). Appropriate accuracies with recoveries between 95 and 122% and 85 and 119% at the LOQ and precisions ranging between 1 and 11 %RSD and 1 and 13 %RSD at the LOQ were observed in standard solutions in urine and faeces, respectively, as shown in **Table 3.3**. Stock solutions from all compounds showed good stability after 1 year at -80 ℃, and after three freeze-thaw cycles (-80 ℃), and derivatized compounds were stable for 20 h in autosampler storage conditions (t-test, *p*-values > 0.05). Extraction efficiency and matrix effects were assessed in spiked samples. Data summarized in **Table 3.3** displayed excellent intra- and inter-day accuracy and precision for all compounds in spiked faeces and urine samples, evidencing an adequate method performance for all nine SCFAs and three BCAAs.

The impact of different sample collection procedures on the retrieved SCFAs and BCAAs profiles was evaluated. Several options are available for the collection of urine samples and their selection largely depends on the study population. Typically, sterile PP containers are used for the collection of urine samples from adults. Likewise, for faeces samples PP containers are also typically employed. However, collecting urine samples from newborns can be a very difficult task as they are obviously non-toilet trained. The recommended procedure for clean catch urine collection involves waiting for a nappy-free child to void spontaneously [177]. To overcome the limitations of this approach, different methods to stimulate voiding were recently proposed, involving gently rubbing the lower abdomen in circular motions with a piece of gauze soaked in cold liquid [178] and bladder stimulation followed by paravertebral lumbar massage, to trigger urination [179,180]. However, the efficiency of this method has not been demonstrated in preterm

infants and so, either sterile plastic bags or cotton pads and gauzes that are placed in the diaper are typically used in this population. To test the compatibility of the sample collection procedure with the determination of SCFAs and BCAAs, process blanks using sterile PP bags, cotton pads, and gauzes were analysed and compared to calibration blanks. Using PP materials (i.e., either containers or bags), no SCFAs or BCAAs were detected in blanks. Remarkably, an intense background concentration of acetic acid was observed when cotton pads or gauzes were used (520 ± 3 and 850 ± 20 µM, respectively). Hence, in case of urine samples collected with cotton pads or gauzes, as it is typically the case for urine samples from newborns, acetic acid concentrations must be excluded from further data analysis. The observed peak area values of the other considered metabolites and the IS were within the acceptability criteria (i.e., <1/10 of the signal at LOQ).

Table 3.3. Calculated intra- and inter-day accuracy (i.e., recovery) and precision (i.e., RSD) of the method in standard solutions and spiked urine and faeces samples.

Matrix	Compound class	Metabolite	Concentration levels (µM)			Accuracy ± RSD											
						Standard solutions						Spiked samples					
						Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
			Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Urine	SCFA	Acetic acid	15	101	150	99 ± 3	100 ± 7	120 ± 2	116 ± 3	113 ± 8	105 ± 6	105 ± 11	106 ± 2	100 ± 5	102 ± 17	100 ± 7	106 ± 8
		Propionic acid	13	84	126	112 ± 5	99 ± 11	115 ± 9	115 ± 6	115 ± 2	115 ± 4	114 ± 3	100 ± 7	102 ± 6	106 ± 4	102 ± 3	109 ± 7
		Isobutyric acid	14	94	140	108 ± 2	96 ± 3	116 ± 5	102 ± 3	108 ± 5	103 ± 9	93 ± 4	88 ± 3	94 ± 6	92 ± 11	85 ± 3	93 ± 5
		Butyric acid	4	26	39	116 ± 7	118 ± 6	112 ± 5	118 ± 2	96 ± 1	117 ± 8	107 ± 4	106 ± 3	106 ± 4	94 ± 6	98 ± 7	100 ± 5
		2-Methylbutyric acid	3	19	28	115 ± 6	119 ± 4	104 ± 8	99 ± 5	109 ± 2	110 ± 5	110 ± 4	97 ± 3	102 ± 5	99 ± 4	100 ± 7	97 ± 4
		Isovaleric acid	2	10	15	112 ± 9	103 ± 5	118 ± 2	95 ± 4	99 ± 7	116 ± 4	114 ± 10	110 ± 6	112 ± 4	94 ± 12	93 ± 17	95 ± 16
		Valeric acid	2	15	23	100 ± 4	116 ± 11	112 ± 4	120 ± 10	107 ± 8	100 ± 6	102 ± 3	97 ± 4	99 ± 1	96 ± 10	96 ± 5	98 ± 3
		Caproic acid	2	15	23	100 ± 11	117 ± 2	115 ± 9	122 ± 7	110 ± 3	110 ± 7	102 ± 8	102 ± 2	100 ± 1	90 ± 10	97 ± 4	98 ± 2
		Heptanoic acid	1.2	8	12	95 ± 4	98 ± 5	110 ± 2	100 ± 11	97 ± 5	97 ± 9	93 ± 5	99 ± 2	99 ± 3	92 ± 11	97 ± 4	100 ± 3
	BCAA	Valine	21	140	209	107 ± 3	107 ± 9	113 ± 11	112 ± 4	104 ± 3	112 ± 10	110 ± 10	103 ± 8	107 ± 4	101 ± 6	103 ± 5	102 ± 8
		Leucine	19	130	195	103 ± 2	102 ± 1	97 ± 10	103 ± 7	95 ± 3	102 ± 7	110 ± 9	105 ± 8	106 ± 5	86 ± 12	96 ± 11	88 ± 12
		Isoleucine	17	116	173	97 ± 7	97 ± 3	107 ± 3	95 ± 7	114 ± 3	108 ± 7	103 ± 11	94 ± 10	102 ± 9	99 ± 8	96 ± 12	92 ± 9

Table 3.3 (continuation). Calculated intra- and inter-day accuracy (i.e., recovery) and precision (i.e., RSD) of the method in standard solutions and spiked urine and faeces samples.

Matrix	Compound class	Metabolite	Concentration levels (µM)			Accuracy ± RSD											
						Standard solutions						Spiked samples					
						Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
			Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Faeces	SCFA	Acetic acid	26	103	155	93 ± 7	90 ± 5	85 ± 4	99 ± 8	90 ± 5	105 ± 5	93 ± 5	109 ± 4	99 ± 4	99 ± 4	105 ± 4	112 ± 9
		Propionic acid	27	107	162	116 ± 10	104 ± 4	116 ± 1	103 ± 9	110 ± 3	93 ± 4	92 ± 3	112 ± 2	95 ± 6	106 ± 7	108 ± 2	102 ± 6
		Isobutyric acid	24	98	147	89 ± 4	108 ± 9	105 ± 1	92 ± 1	116 ± 7	104 ± 2	110 ± 4	102 ± 1	92 ± 3	99 ± 3	104 ± 8	96 ± 10
		Butyric acid	5	19	29	117 ± 3	95 ± 5	95 ± 5	99 ± 7	88 ± 2	98 ± 7	100 ± 1	113 ± 10	118 ± 7	108 ± 5	110 ± 5	110 ± 7
		2-Methylbutyric acid	4	16	24	105 ± 8	103 ± 10	116 ± 1	97 ± 5	89 ± 1	98 ± 1	110 ± 5	114 ± 5	116 ± 6	108 ± 1	108 ± 6	114 ± 4
		Isovaleric acid	2	9	14	106 ± 7	86 ± 1	112 ± 5	110 ± 8	109 ± 2	117 ± 9	95 ± 4	107 ± 4	95 ± 4	100 ± 5	108 ± 4	107 ± 6
		Valeric acid	3	13	19	104 ± 11	88 ± 4	106 ± 1	97 ± 2	104 ± 6	98 ± 8	106 ± 10	103 ± 2	94 ± 6	106 ± 10	103 ± 1	94 ± 9
		Caproic acid	4	16	24	110 ± 13	92 ± 9	102 ± 3	110 ± 5	90 ± 8	119 ± 4	108 ± 6	99 ± 3	97 ± 6	108 ± 5	99 ± 9	97 ± 9
		Heptanoic acid	3	11.2	16	119 ± 9	100 ± 10	107 ± 1	115 ± 9	102 ± 3	108 ± 5	112 ± 5	110 ± 5	92 ± 4	112 ± 7	110 ± 9	92 ± 1
	BCAA	Valine	35	140	209	100 ± 5	86 ± 4	95 ± 5	118 ± 3	116 ± 10	112 ± 1	105 ± 6	115 ± 4	113 ± 9	106 ± 8	103 ± 10	94 ± 2
		Leucine	32	130	195	115 ± 9	104 ± 2	95 ± 4	117 ± 3	116 ± 1	93 ± 9	103 ± 5	108 ± 7	112 ± 5	108 ± 3	99 ± 3	97 ± 8
		Isoleucine	29	116	173	103 ± 11	103 ± 2	103 ± 11	90 ± 8	116 ± 3	90 ± 5	106 ± 2	103 ± 3	94 ± 6	112 ± 3	110 ± 10	92 ± 10

3.4.2. Analysis of SCFAs and BCAAs in faeces and urine

Table 3.4 summarizes the concentrations of SCFAs determined in faeces and urine samples, after normalization to sample weight and to creatinine, respectively, using the validated GC-MS method. Results showed that acetic acid is the most concentrated SCFA present in faeces and together with isobutyric acid represents 86% of SCFAs molarity in this matrix. With respect to urine, isobutyric acid is the most abundant SCFA followed by acetic and butyric acids, and they account for 99% of total SCFAs measured in this matrix. Unfortunately, although acetic acid contributes noticeably to total SCFAs in faeces, the levels of this metabolite in newborn's urine were excluded due to the abovementioned contamination during sample collection, and therefore its contribution could not be studied. The lowest relative concentrations among the considered metabolites were found for caproic acid in faeces, isovaleric acid in urine and heptanoic acid in both matrices. While all metabolites presented a detection frequency higher than 93% in faeces samples, propionic, 2-methylbutyric, isovaleric, valeric, and caproic acids were detected in 42 to 88 % of urine samples.

Table 3.4. SCFA concentrations in faecal and urine samples.

Compound class	Metabolite	Faeces (N = 137), µmol/g faeces				Urine (N = 210), µmol/g creatinine			
		Range	Median	IQR (25-75)	Detection frequency (%)	Range	Median	IQR (25-75)	Detection frequency (%)
SCFA	Acetic acid	0.09 - 72	21	18	100	23 - 4935	76	60	100
	Propionic acid	0.08 - 18	0.6	2	98	0.08 - 3761	5	47	72
	Isobutyric acid	0.012 - 84	13	25	99	2 - 47841	1534	4004	100
	Butyric acid	0.013 - 17	3	5	100	7 - 844	42	72	96
	2-Methylbutyric acid	0.012 - 5	0.8	2	96	0.12 - 67	3	15	88
	Isovaleric acid	0.014 - 3	0.5	0.6	96	0.02 - 15	0.2	3	42
	Valeric acid	0.03 - 3	0.6	1	93	0.02 - 252	2	7	80
	Caproic acid	0.012 - 2	0.02	0.2	95	0.03 - 193	2	13	84
	Heptanoic acid	0.02 - 0.3	0.0012	0.03	90	0.02 - 19	0.8	2	93
BCAA	Valine	0.03 - 8	0.3	0.3	100	4 - 1132	88	137	100
	Leucine	0.013 - 1.3	0.4	0.4	99	may-55	154	266	100
	Isoleucine	0.02 - 2	0.3	0.2	100	2 - 678	53	108	100

Note: IQR = interquartile range.

Figure 3.1 depicts the paired correlations among SCFAs and BCAAs found in faecal and urine samples separately, and **Table 3.5** summarizes results from a regression analysis between faecal and urinary SCFA and BCAA concentrations. As a general trend, it can be observed that SCFAs in faecal supernatants are significantly correlated among each other. Interestingly, similar patterns of significant correlations among SCFAs were observed in urine samples, except in the case of propionic, butyric and valeric acids. BCAAs in both faeces and urine samples followed a similar pattern, being the correlation between leucine and isoleucine higher than with valine. Between both matrices, i.e., faeces and urine, significant positive correlations were found for propionic acid and valine and significant negative correlations were found for butyric, 2-methylbutyric, and isovaleric acid as well as leucine.

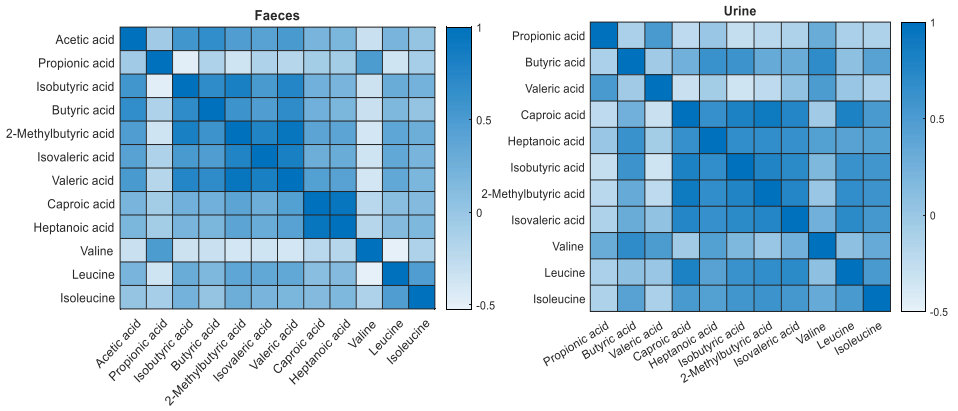


Figure 3.1. Pearson correlation coefficients of SCFAs and BCAAs in faecal supernatants and urine samples.

Table 3.5. Linear regression parameters among faecal and urinary SCFAs and BCAAs concentrations in paired samples (N=80).

Compound class	Metabolite	Slope	*p*-value	R^2
SCFA	Acetic acid	-0.6 ± 2.0	0.8	0.003
	Propionic acid	45 ± 12	< 0.001	0.2
	Isobutyric acid	-8 ± 10	0.4	0.009
	Butyric acid	-5 ± 3	0.04	0.05
	2-Methylbutyric acid	-4.3 ± 1.0	< 0.001	0.2
	Isovaleric acid	-2.3 ± 0.6	< 0.001	0.2
	Valeric acid	-4 ± 3	0.2	0.03
	Caproic acid	-8 ± 5	0.12	0.03
	Heptanoic acid	-3 ± 2	0.12	0.03
BCAA	Valine	44 ± 12	< 0.001	0.2
	Leucine	-360 ± 83	< 0.001	0.2
	Isoleucine	-42 ± 30	0.2	0.02

3.4.3. SCFAs and BCAAs in samples from newborn infants and their mothers

As shown in **Table 3.1**, demographic, clinical, and perinatal characteristics from participants were recorded and no differences in prenatal demographic characteristics or confounders during the hospitalization and antibiotic therapy were found between both infant groups (*p*-value > 0.05), except from gestational age and birth weight, according to the definition of PI and TI. Metabolite levels in both matrices were compared between lactating mothers, TI, and PI. **Figure 3.2** (top, left) shows PCA scores plots of faeces samples from mothers and infants. Three clusters for mothers, PI and TI can be observed when considering faecal SCFA and BCAA concentrations. For

urinary SCFA and BCAA concentrations (**Figure 3.2** bottom, left), a clear separation of the tightly clustered group of urine samples collected from lactating mothers from those of infants is evidenced, although SCFA and BCAA profiles of infants were overlapping in the PC1-PC2 scores space summarizing 53% of the variance. The joint analysis of the scores and loadings plot (**Figure 3.2**, top) shows that faecal SCFAs and BCAAs are present at higher concentrations in maternal samples, and a negative correlation of valine and propionic acid with isoleucine and leucine, linked to the clustering of PI samples in the scores plot. On the other hand, the analysis of the scores and loadings plots from the analysis of urinary SCFAs and BCAAs (**Figure 3.2**, bottom) shows a high positive correlation between isoleucine and leucine, and higher levels of the metabolites in infant (PI and TI) samples. Hence, the determination of both SCFAs and BCAAs in urine and faeces samples using the developed GC-MS method provide information that allows to distinguish between groups in the present data set.

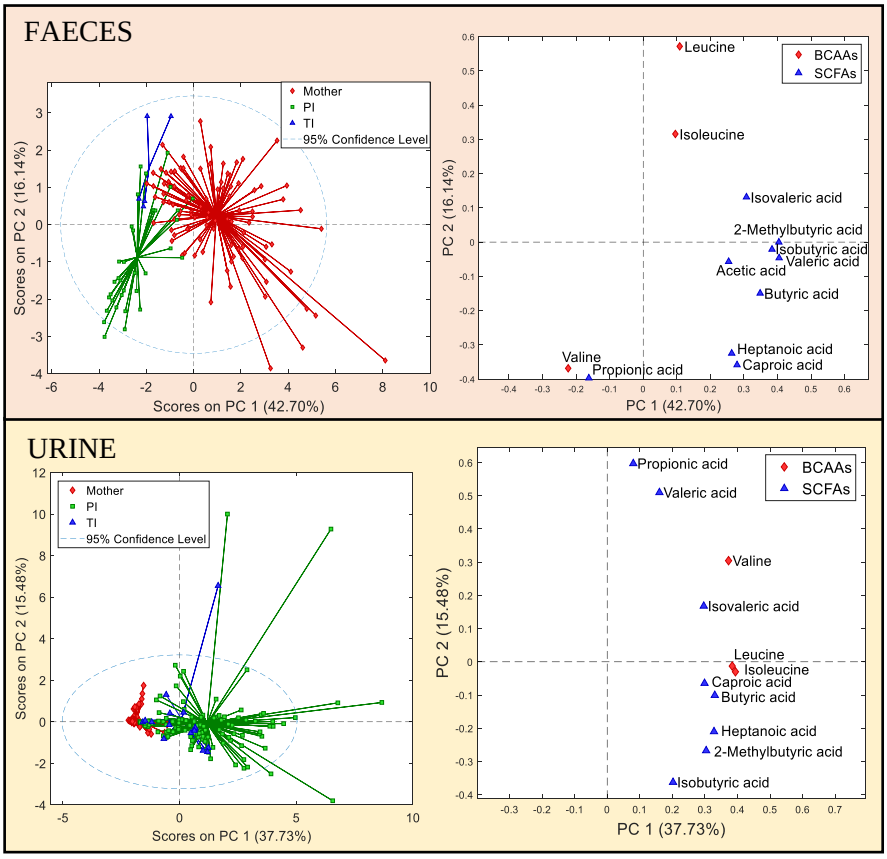


Figure 3.2. PCA scores (left) and loadings plots (right) in faeces (top) and urine (bottom) samples.

Figure 3.3 shows the biomarker concentration ranges in the three study groups. In faeces, higher concentration values were detected in mothers in comparison to PI and TI, except from propionic acid and BCAAs (Wilcoxon rank-sum test, $\alpha = 0.05$). In urine, higher concentration values of some SCFAs (i.e., butyric, valeric, heptanoic and isobutyric acids) and all BCAAs were detected when comparing M to PI and TI, respectively. Regarding PI vs. TI, slightly higher faecal concentration values in PI were

identified in two SCFAs (i.e., propionic and caproic acids) and lower faecal concentration values in PI were identified in two BCAAs (i.e., leucine and isoleucine). In urine, higher concentration values of caproic and 2-methylbutyric acids and leucine were found in PI. No statistically significant correlations between biomarker concentrations of mothers and their infants or between biomarker concentrations and gestational or postnatal age were found in this study (Pearson's $\chi 2$ test, $\alpha = 0.05$). Likewise, no statistically significant differences were observed between male and female infants in any of the assessed metabolites (Wilcoxon rank-sum test, $\alpha = 0.05$).

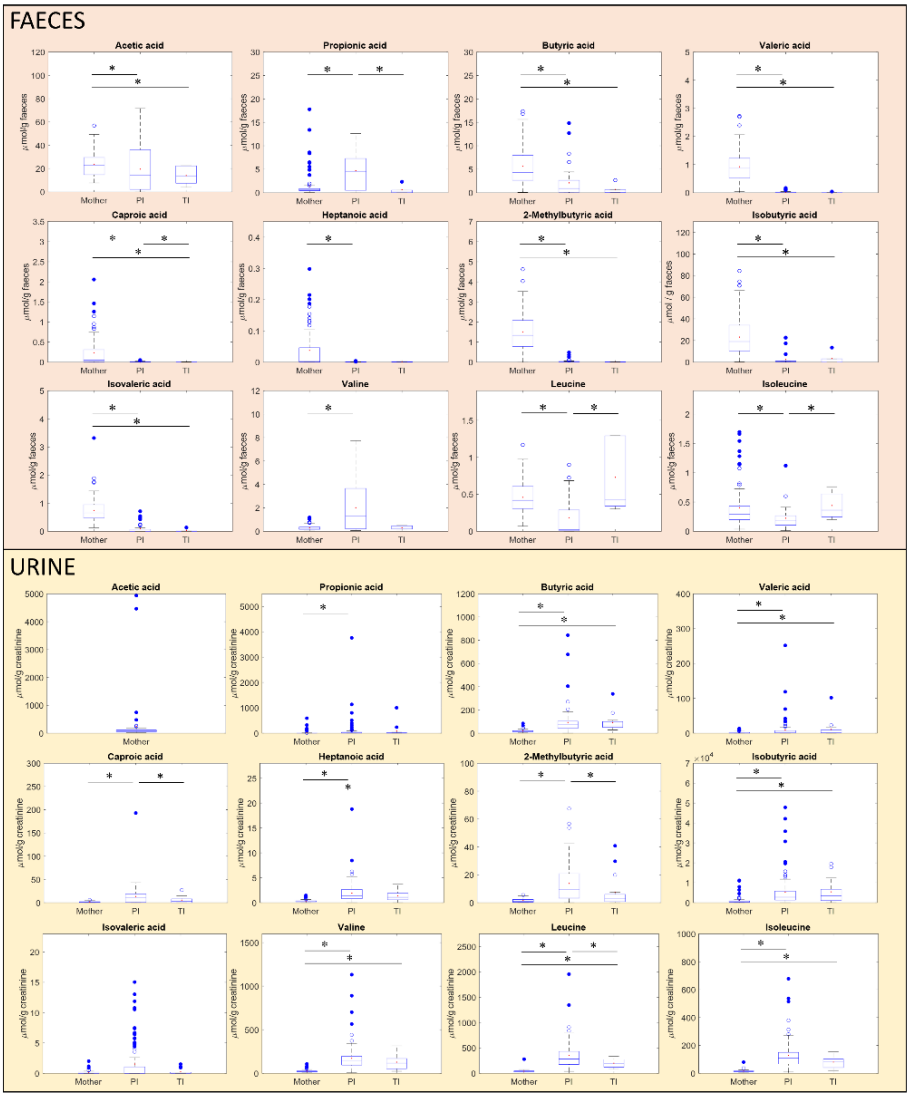


Figure 3.3. Faecal and urinary SCFA and BCAA levels in mothers and their TI and PI. Note: * indicates significant differences (Wilcoxon rank-sum, *p*-value < 0.05).

Further studies are needed to gain a deeper understanding of the biological meaning of SCFAs and BCAAs levels in faeces and urine samples and the effect of possible confounders such as, e.g., gestational, and postnatal age, type of nutrition. The biochemical interpretation of the results is hampered by the difficulty to assess the origin of SCFAs and BCAAs in urine due to the possibility of their generation by microbial activity in other parts of the body. Nonetheless, previous work showed an association between the SCFA urinary profiles and the disease status of patients with ulcerative colitis, where higher butyrate levels in patients at remission were observed, and it was suggested that the urinary output could reflect the levels of butyrate produced in the gut [181].

3.5. Conclusions

A method for the analysis of nine SCFAs and three BCAAs in faeces and urine samples has been reported. The method enabled an accurate quantification of these biomarkers in urine and faeces using a single derivatization procedure. It was successfully validated including the entire test system, from sample collection to analysis, revealing a very significant contamination with acetic acid from cotton and gauze pads typically employed for newborn urine samples' collection that should be considered in future clinical studies. The suitability of the developed method to the analysis of clinical samples has been demonstrated through the analysis of a set of 137 faecal and 210 urine samples from a broad clinical cohort comprising adults, term and preterm infants. A significant correlation among SCFAs and BCAAs in faecal supernatants and urine samples were observed and a total of five metabolites showed a significant correlation across matrices. Furthermore,

reference ranges of SCFAs and BCAAs in mothers, PI and TI were reported and different patterns were found among groups. Infants showed lower concentrations of SCFAs and BCAAs than mothers in faeces and higher concentrations in urine. Slightly higher concentration values in PI compared to TI were identified, except for leucine and isoleucine in faecal samples that were higher in TI. The presented results will serve as a knowledge base for future studies that will examine the interplay of human milk composition, microbiota, newborn physiology, and the study of serious gastrointestinal diseases that mostly affect preterm babies such as necrotizing enterocolitis.

Supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2022.05.005.

Chapter 4. Fast profiling of primary, secondary, conjugated and sulphated bile acids in urine and feces samples

4.1. Abstract

Bile acids (BAs) are a complex class of metabolites that have been described as specific biomarkers of gut microbiota activity. The development of analytical methods allowing the quantification of an ample spectrum of BAs in different biological matrices is needed to enable a wider implementation of BAs as complementary measures in studies investigating the functional role of the gut microbiota. This work presents results from the validation of a targeted ultra-high performance liquid chromatography – tandem mass spectrometry (UHPLC-MS/MS) method for the determination of 28 BAs and six sulphated BAs, covering primary, secondary, and conjugated BAs. The analysis of 73 urine and 20 feces samples was used to test the applicability of the method. Concentrations of BAs in human urine and murine feces were reported, ranging from 0.5 to 50 nmol/g creatinine and from 0.012 to 332 nmol/g, respectively. 79% of BAs present in human urine samples corresponded to secondary conjugated BAs, while 69% of BAs present in murine feces corresponded to primary conjugated BAs. Glycocholic acid sulphate (GCA-S) was the most abundant BA in human urine samples, while taurolithocholic acid was the lowest concentrated compound detected. In murine feces, the most abundant BAs were α-murocholic, deoxycholic, dehydrocholic and β-murocholic acids, while GCA-S was the lowest concentrated BA. The presented method is a non-invasive approach for the simultaneous assessment of BAs and sulphated BAs in urine and feces

samples and the results will serve as a knowledge base for future translational studies focusing on the role of the microbiota in health.

4.2. Introduction

Bile acids (BAs) are a group of molecules from the family of steroids with a hydrophobic core structure of four fused carbon rings with hydroxyl and methyl groups, and a carbon side chain ending in a carboxylic acid group. Primary BAs (e.g., cholic (CA) and chenodeoxycholic acids (CDCA)) are produced from cholesterol in the liver through a series of enzymatic reactions and are modified thereafter through the addition of glycine or taurine to form conjugated BAs (e.g., glycocholic (GCA) and taurocholic (TCA) acids) before being released to the bile. BAs are then secreted into the intestine through the gall bladder. During their transit through the intestine, BAs undergo modifications due to the action of gut microbiota (e.g., deconjugation, dehydroxylation, hydroxyl group oxidation, and epimerization), producing secondary BAs (e.g., litocholic (LCA) and deoxycholic acids (DCA)), making BAs more hydrophobic and, consequently, less dangerous to bacteria [182]. Both primary and secondary BAs are reabsorbed in the terminal ileum and large intestine before being transported back to the liver, where they are recycled [45]. This process, known as enterohepatic circulation , is important for the efficient use of BAs as it helps to maintain a balance of these substances in the body and plays an important role in various physiological processes, such as emulsifying ingested lipids [183]. Additionally, BAs can be found in small amounts distributed across other parts of the body, as they can function as signaling molecules in other physiological processes apart from the digestive system, such as cardiac system or renal water regulation [184,185].

The BAs metabolism is known to differ between species (e.g., humans and murine animals) [186]. In humans, the classical pathway for BA metabolism (80%) is by the downstream metabolism of CA and DCA, while in murine animals, BAs metabolism is dominated by the production of α-muricholic (α-MCA) and β-muricholic acid (β-MCA) from CDCA, LCA and ursodeoxycholic acid (UDCA) [187].

Only a small portion of BAs, approximately 5% of the total pool, is not reabsorbed but excreted [188]. BAs excreted though feces have been reported to be mostly present in the unconjugated form. However, urine is the primary route for BA elimination, and BAs not recycled from the liver or renal cells are excreted through it in the main detoxification pathways (i.e., glucuronidation and sulfation) [189–191], mostly in the conjugated and/or sulphated form [192,193].

BAs and gut microbiota are considered a dynamic equilibrium, as both have a clear impact on one another. Changes in BA levels or types in biofluids may be indicative of certain health conditions or response to treatment. In addition, the measurement of BAs can help to identify potential biomarkers for diseases such as liver disease, gastrointestinal disorders, and metabolic disorders [194]. Overall, the analysis of BAs provides valuable information for clinical monitoring.

There are various methods that are currently used to identify and quantify BAs in biological samples [195]. Liquid (LC) or gas (GC) chromatography coupled to mass spectrometry (MS) have been the most commonly used techniques for the analysis of BAs in human and animal biofluids in recent years [100,196–203]. Despite GC-MS providing a very high peak resolution, its main disadvantage is the time-consuming sample preparation that requires group fractionation, enzymatic or chemical

hydrolysis of conjugates and preparation of volatile methyl ester trimethylsilyl ether derivatives [195]. In LC-MS, BAs are typically resolved using reversed-phase chromatography columns and gradient elution programs. LC-MS offers excellent sensitivity and specificity, which may reduce the need for extensive sample preparation and ultimately improve analytical throughput. However, the majority of the studies regarding BA analysis by LC-MS involve solid phase extraction and/or preconcentration steps. Furthermore, previously reported LC-MS methods for BAs quantification do not include sulphated BAs [197–199].

The aim of this work was to develop a straightforward ultra-high performance liquid chromatography – tandem mass spectrometry (UHPLC-MS/MS) method for the determination of 28 BAs and six sulphated BAs, covering a wide range of primary, secondary, conjugated, and sulphated BAs. The applicability of the method in the clinical field and in animal models was evaluated by the analysis of 73 human urine samples and 20 murine feces samples.

4.3. Material and methods

4.3.1. Standards and reagents

LC-MS grade methanol, propanol, acetic acid, ammonium acetate, and acetonitrile were obtained from Merck Life Science S.L.U. (Madrid, Spain). Bile acid standards (CA, CDCA, LCA, DCA, GCA, TCA, α-MCA, β-MCA, UDCA, hyodeoxycholic (HDCA), murocholic (MCA), 3,7,12-dehydrocholic (DHCA), hyocholic (HCA), glycolithocholic (GLCA),

glycoursodeoxycholic (GUDCA), glycohyodeoxycholic (GHDCA), glycochenodeoxycholic (GCDCA), glycodeoxycholic (GDCA), 3,7,12-glycodehydrocholic (GDHCA), glycohyocholic (GHCA), taurolithocholic (TLCA), tauroursodeoxycholic (TUDCA), taurohyodeoxycholic (THDCA), taurochenodeoxycholic (TCDCA), taurodeoxycholic (TDCA), 3,7,12-taurodehydrocholic (TDHCA), taurohyocholic (THCA), tauro-α-muricholic (T-α-MCA) acids, as well as deuterated internal standards (IS) (lithocholic-D4 (LCA-D4), cholic-D4 (CA-D4), glycochenodeoxycholic-D4 (GCDCA-D4), glycocholic-D4 (GCA-D4) acids) were purchased from Steraloids (Newport, RI, US). Sulphated BAs (lithocholic (LCA-S), chenodeoxycholic (CDCA-S), ursodeoxycholic (UDCA-S), deoxycholic (DCA-S), cholic (CA-S), glycocholic (GCA-S) acid sulphates), as well as lithocholic-D4 acid sulphate (LCA-S-D4) as IS were obtained from Qmx Laboratories Ltd. (Essex, UK). Individual 5 mM standard solutions were prepared in methanol.

4.3.2. Study design, population, and sample collection

Urine samples were collected from 73 children, i.e., 44 male and 29 female individuals, aged 13 ± 3 years (mean ± standard deviation) with a range between 8 and 18 years recruited at the Pediatric Department of the San Raffaele Hospital in Milan (Italy). During the enrolment phase, pediatricians informed the parents about the purpose of this study, the lack of reported risks related to the collection of samples, the effort required to take part in this study and their right to deny or withdraw their consent at any time. Parents accepted to participate in the study and signed an informed consent form. The study was approved by the Institutional Ethical Committee of the IRCCS San Raffaele Scientific Institute (Protocol: NUTRI-T1D, 2019). Individual

identifiable private information was protected according to the EU General Data Protection Regulation (EU-GDPR) with the supervision of the Institutional Data Protection Officer. The Clinical Research Investigator assigned a code to each patient and identifiers that link to protected information. Each subject collected urine samples at home and kept them at 4 °C until transported to the hospital on the same day. Sterile plastic containers were used for sample collection and a minimum of 10 mL were required. Samples were aliquoted, labelled and stored at -80 °C until analysis.

Feces samples were collected from CC57BL/6J male mice (n = 20, 7 weeks old) provided by Charles River Laboratories (Wilmington, USA). Mice were collectively housed (n = 5/cage) and maintained under constant conditions of humidity and temperature (i.e., 23 ± 2 °C), and regular 12-h light-dark cycles. Mice fed with control diet (CD, D12450K, Research Diets, New Brunswick, NJ) were fasted overnight to collect feces 2 h after the administration of a mixed solution of lipids (intralipid, 0.06% per mouse, Merck KGaA) by oral gavage. Collected samples were immediately frozen in liquid nitrogen until further analysis. This experimental procedure using animals was in accordance with European Union 2010/63/UE and Spanish RD53/2013 guidelines and approved by the ethics committee of the University of Valencia (Animal Production Section, SCSIE, University of Valencia, Spain) and authorized by Dirección General de Agricultura, Ganadería y Pesca (Generalitat Valenciana) (approval ID 2021/VSC/PEA/0273).

4.3.3. UHPLC-MS/MS determination of BAs

The UHPLC-MS/MS method for the quantification of 34 BAs was developed based on previously reported method [100] with modifications. For feces, 1 mL of ultrapure water was added to 100 mg of wet samples, then the mixture was homogenized for 10 min and centrifuged at 13200 × g for 20 min at room temperature (RT). Urine samples were centrifuged at 3500 × g for 15 min at RT. 50 µL of fecal or urinary supernatants were mixed with 10 µL of IS solution containing LCA-D4, CA-D4, GCDCA-D4, GCA-D4 and LCA-S-D4 at 2.1 µM each and 150 µL of ice-cold methanol were added for protein precipitation. The mixture was then agitated for 15 min at 4 °C, incubated for 20 min at -20 °C, and centrifugated at 3500 × g for 15 min at 4 °C. Urine samples were diluted 1:20 (v/v) with ultrapure H_2O prior to analysis. Supernatants were transferred to a 96-well plate for analysis.

UHPLC-MS/MS analysis was conducted using an ACQUITY UPLC instrument (Waters Ltd, Elstree, UK) coupled to a Xevo TQ-S MS (Waters, Manchester, UK) operating in negative electrospray ionization mode (ESI-). Chromatographic separation was performed using an ACQUITY BEH C8 column (100 mm x 2.1 mm, 1.7 µm) (Waters) at an operating temperature of 60 °C and a flow rate of 0.6 mL min^{-1}. Mobile phase component A consisted of 100 mL of acetonitrile added to 1 mM ammonium acetate in 1 L ultrapure water, adjusting the pH to 4.15 with acetic acid. Mobile phase component B consisted of an acetonitrile:propanol (1:1, v/v) solution. The gradient separation was performed as follows: 0-0.1 min, 10% B, 0.1-9.25 min, 10-35% B; 9.25-11.5 min, 35-85% B; 11.5-11.8 min, 85-100% B; 11.8-14.3 min, 100-10% B; and 14.3-16 min, 10% B. The injection volume was 5 µL. MS/MS detection was performed using the multiple reaction monitoring (MRM) mode and the following ionization conditions: ion spray voltage, 1.5 kV; cone

voltage, 60V; source temperature, 150 °C; desolvation temperature, 600 °C; desolvation gas flow, 1000 L/h; and cone gas flow, 150 L/h. The MRM transitions for each BA and IS summarized in **Table 4.1** were optimized by infusing individual solutions of commercial standards.

Table 4.1. Measurement parameters and main figures of merit of the LC-MS method for BA determination.

Class	Metabolite	MRM transition	CE (V)	RT ± s (min)	R^2	Urine 115imples (nM)		Feces samples (nM)	
						LOD	LOQ	LOD	LOQ
Primary	MCA	391.3 → 391.3	10	9.489 ± 0.002	0.997	0.5	1.7	0.8	3
	CDCA	391.3 → 391.3	10	10.760 ± 0.014	0.997	0.5	1.7	0.6	2
	CA	407.3 → 407.3	10	10.155 ± 0.002	0.998	0.2	0.7	0.3	1.12
	α-MCA	407.3 → 407.3	10	8.782 ± 0.015	0.997	0.2	0.7	0.5	1.6
	β-MCA	407.3 → 407.3	10	8.883 ± 0.012	0.997	0.3	1.0	0.2	0.7
	HCA	407.3 → 407.3	10	9.857 ± 0.014	0.998	0.3	1.0	0.3	0.8
Primary conjugated	GCDCA	448.4 → 74.0	40	9.347 ± 0.014	0.996	1.0	3	0.002	0.01
	GHCA	464.4 → 74.0	40	6.820 ± 0.002	0.996	0.2	0.7	0.13	0.4
	GCA	464.4 → 74.0	40	7.573 ± 0.012	0.996	0.3	1.0	0.013	0.04
	TCDCA	498.4 → 80.0	60	8.721 ± 0.004	0.9993	0.5	1.7	0.02	0.06
	TCA	514.3 → 80.0	60	7.047 ± 0.013	0.998	0.2	0.7	0.02	0.07
	T-α-MCA	514.3 → 80.0	60	5.238 ± 0.002	0.998	0.2	0.7	0.03	0.08
	THCA	514.3 → 80.0	60	6.308 ± 0.014	0.9992	0.2	0.7	0.03	0.10

Table 4.1 (continuation). Measurement parameters and main figures of merit of the LC-MS method for BA determination.

.Class	Metabolite	MRM transition	CE (V)	RT ± s (min)	R^2	Urine samples (nM)		Feces samples (nM)	
						LOD	LOQ	LOD	LOQ
Secondary	LCA	375.3 → 375.3	10	11.140 ± 0.002	0.998	1.0	3	0.6	1.8
	DCA	391.3 → 391.3	10	10.812 ± 0.003	0.998	1.0	3	0.4	1.3
	UDCA	391.3 → 391.3	10	10.060 ± 0.012	0.996	0.2	0.7	0.3	1.12
	HDCA	391.3 → 391.3	10	10.283 ± 0.012	0.998	0.2	0.7	0.2	0.8
	DHCA	468.4 → 74.0	10	5.440 ± 0.003	0.996	0.2	0.7	0.3	0.8
Secondary conjugated	GLCA	432.4 → 74.0	40	10.444 ± 0.003	0.9992	1.0	3	0.010	0.03
	GDCA	448.4 → 74.0	40	9.772 ± 0.002	0.998	0.3	1.0	0.02	0.03
	GUDCA	448.4 → 74.0	40	7.122 ± 0.016	0.9992	0.2	0.7	0.003	0.01
	GHDCA	448.4 → 74.0	40	7.490 ± 0.004	0.998	1.0	3	0.013	0.04
	GDHCA	458.3 → 74.0	40	2.831 ± 0.002	0.997	0.2	0.7	0.02	0.06
	TLCA	466.3 → 80.0	40	10.220 ± 0.003	0.998	0.2	0.7	0.02	0.07
	TDCA	498.4 → 80.0	60	9.123 ± 0.016	0.9995	0.3	1.0	0.03	0.08
	TUDCA	498.4 → 80.0	60	6.640 ± 0.015	0.998	0.5	1.7	0.13	0.4
	THDCA	498.4 → 80.0	60	6.945 ± 0.002	0.998	0.3	1.0	0.04	0.13
	TDHCA	508.3 → 80.0	60	2.464 ± 0.013	0.997	0.2	0.7	0.010	0.03

Table 4.1 (continuation). Measurement parameters and main figures of merit of the LC-MS method for BA determination.

Class	Metabolite	MRM transition	CE (V)	RT ± s (min)	R^2	Urine samples (nM)		Feces samples (nM)	
						LOD	LOQ	LOD	LOQ
Sulphated	GCA-S	566.3 → 97	40	6.931 ± 0.014	0.996	0.3	1.0	0.005	0.02
	UDCA-S	471.2 → 97	40	7.320 ± 0.012	0.996	0.2	0.7	0.008	0.03
	LCA-S	455.2 → 97	40	10.320 ± 0.003	0.997	0.2	0.7	0.006	0.02
	DCA-S	471.3→97	40	9.288 ± 0.002	0.998	0.2	0.7	0.08	0.3
	CDCA-S	471.2 → 97	40	9.488 ± 0.014	0.998	0.2	0.7	0.13	0.4
	CA-S	487.2 → 97	40	7.571 ± 0.012	0.996	0.3	1.0	0.4	1.4
IS	LCA-d4	379.6 → 97	10	11.143 ± 0.002	-	-	-	-	-
	CA-d4	411.6 → 97	10	10.151 ± 0.002	-	-	-	-	-
	GCDCA-d4	451.5 → 74	40	9.330 ± 0.002	-	-	-	-	-
	GCA-d4	468.6 → 74	40	7.572 ± 0.016	-	-	-	-	-
	LCA-S-d4	459.2 → 97	10	10.312 ± 0.013	-	-	-	-	-

Note: IS = internal standard; CE = collision energy; RT = retention time; R = coefficient of determination; Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20% and taking into account the preconcentration factor achieved during sample processing; Limit of detection (LOD) = 3/10*LOQ.

Standard solutions were prepared by serial dilution of a working solution containing mixtures of pure analytical standards in methanol. For monitoring instrument performance, aliquots of quality control (QC) samples obtained by pooling 5 µL of each urine or fecal supernatant, respectively, were analyzed after every 20 samples in each analytical batch. The batch acceptance criterion for each analyte was a QC RSD <25%. In addition, calibration (i.e., ultrapure H_2O) and process blanks (generated replacing the sample by H_2O) were included in each batch to assess method specificity and the lack of carry over and cross-contamination.

Fecal BA concentrations were normalized to sample weight while urinary BA concentrations were normalized to creatinine concentrations quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA).

4.3.4. Method validation

The method validation included the determination of linearity range, accuracy, precision, selectivity, specificity, carry over, and analyte stability, based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [2] and the International Council for Harmonisation (ICH) M10 on bioanalytical method validation – scientific guideline [204]. The range of concentrations employed ranged from 1.2 nM to 1.25 µM, selected to accommodate for the anticipated values in real samples and a significant inter- and intra-individual variability, thereby enabling precise, accurate, and linear quantification of the metabolites. At least six

standards covering the concentrations ranges, as well as a process blank were included in the calibration curves. Accuracy and precision in standards were evaluated, respectively, by comparing absolute concentrations values and calculating the relative standard deviation (%RSD) of replicates of standards within a single validation batch (intra-day) and between different validation batches (inter-day). To assess the accuracy and precision in the sample matrix, three replicates of both standard and spiked samples at three different concentrations (i.e., low, medium, and high) were analyzed on three separate validation days and recoveries as well as %RSD were calculated. The analysis of calibration and process blank samples from six different sources was used to assess selectivity and specificity of the analytical method. Additionally, the analysis of a process blank after the injection of the high concentrated standard mixture was used for evaluating carry-over and cross contamination between consecutive samples. The freeze-thaw and long-term stability of the analytes were assessed by comparing the concentrations in a sample that had undergone three freeze-thaw cycles and a sample that had been stored for six months at -80 ^{o}C, respectively, to a freshly prepared sample.

4.3.5. Data availability and statistical analysis

UHPLC-MS/MS data were acquired and processed using MassLynx software version 4.0 (Waters Ltd, Elstree, UK). Data analysis was carried out in MATLAB R2018b (MathWorks, Natick, MA, USA). Biomarker normalized levels in urine and feces samples determined in this work are available in Tables S4.1 and S4.2. Continuous variables were expressed as mean ± standard deviation or medians with interquartile (IQR) ranges.

4.4. Results and discussion

4.4.1. Method validation

The chromatographic peaks of the different BAs are depicted in **Figure 4.1**. Primary, secondary, and sulphated BAs presented different chromatographic retention clusters. Primary BAs RTs ranged between 8.7 and 10.8 min, and secondary BAs eluted between 10.1 and 11.2 min, except DHCA at 5.4 min. Sulphated BAs eluted in the 6.9-10.3 min range. Due to their structural and physicochemical heterogeneity, primary and secondary conjugated BAs eluted in a wider time ranging between 2.4 and 10.4 min. Results depicted in **Figure 4.1** indicated an adequate chromatographic and/or spectral resolution among BAs that enabled their simultaneous quantification.

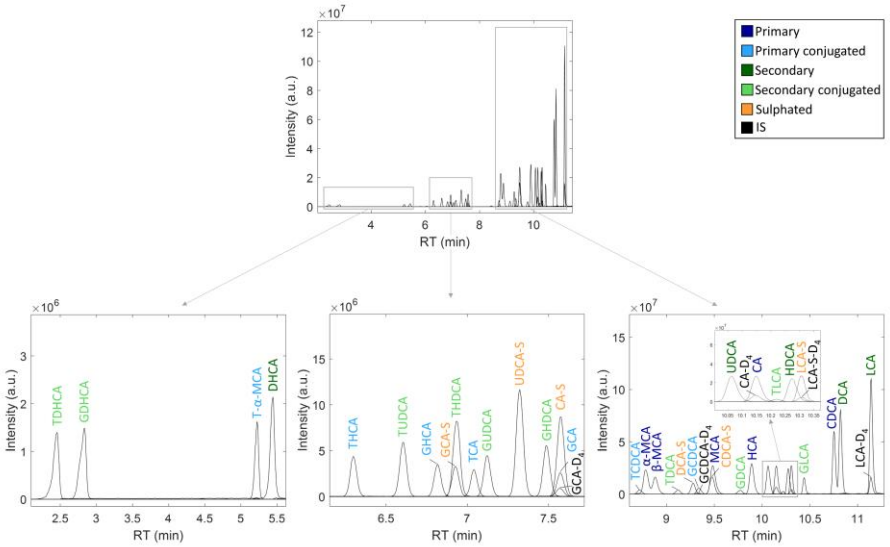


Figure 4.1. Separation of BAs in the developed UPLC-MS/MS method. Note: TDHCA = 3,7,12-taurodehydrocholic acid, GDHCA = 3,7,12-glycodehydrocholic acid, T-α-MCA = tauro-α-muricholic acid, DHCA = 3,7,12-dehydrocholic acid, THCA = taurohyocholic acid, TUDCA = tauroursodeoxycholic acid, GHCA = glycohyocholic acid, GCA-S = glycocholic acid sulphate, THDCA = taurohyodeoxycholic acid, TCA = taurocholic acid, GUDCA = glycoursodeoxycholic acid, UDCA-S = ursodeoxycholic acid sulphate, GHDCA = glycohyodeoxycholic acid, CA-S = cholic acid sulphate, GCA = glycocholic acid, GCA-D4 = deuterated glycocholic acid, TCDCA = taurochenodeoxycholic acid, α-MCA = α-Muricholic acid, β-MCA = β-muricholic acid, TDCA = taurodeoxycholic acid, DCA-S = deoxycholic acid sulphate, GCDCA = glycochenodeoxycholic acid, GCDCA-D4 = deuterated glycochenodeoxycholic acid, MCA = murocholic acid, CDCA-S = chenodeoxycholic acid sulphate, GDCA = glycodeoxycholic acid, HCA = hyocholic acid, UDCA = ursodeoxycholic acid, CA = cholic acid, CA-D4 = deuterated cholic acid, TLCA = taurolithocholic acid, HDCA = hyodeoxycholic acid, LCA-S = lithocholic

acid sulphate, LCA-S-D4 = deuterated lithocholic acid sulphate, GLCA = glycolithocholic acid, CDCA = chenodeoxycholic acid, DCA = deoxycholic acid, LCA = litocholic acid, LCA-D4 = deuterated litocholic acid.

Analytical method validation was performed following recommended FDA guidelines for bioanalytical method validation [2]. The evaluation of the linearity of response was carried out by covering up to four orders of magnitude with LODs and LOQs in the 0.2-3 and 0.002-3 nM range in urine and feces, respectively (see Table 4.1). As shown in Table 4.2, appropriate accuracies with recoveries between 84 and 120% and precisions ranging between 1 and 19 %RSD at the three levels were observed in standard solutions, except for the UDCA-S and CDCA-S inter-day accuracy and precision at the medium concentration level with a recovery of 117% and an RSD of 16%, respectively. In spiked urine and feces samples, recoveries between 80 and 119%, 85 and 115%, and 85 and 115% and precisions between 2 and 19%, 1 and 17%, and 1 and 16% at low, medium, and high levels were found, respectively, except for the inter-day accuracy of the medium concentration level of HDCA (118%) in urine and the precision of HCA (17%, medium level, urine), GCA (24%, low level, feces), and TCDCA (23%, low level, feces). After six months at -80 ^{o}C and after undergoing three freeze-thaw cycles (-80 °C), all compounds remained stable in samples (t-test, p-values >0.05). The compatibility of the sample collection procedure, as well as the selectivity, specificity, and carry over were assessed, and no interfering contamination was detected (<1/10 of the signal at LOQ). The analysis of blanks and process blanks confirmed the lack of cross-contamination or carry over in the assessed chromatographic conditions. The presented data show adequate analytical method performance for all compounds in spiked urine

samples and fecal supernatants, supporting the suitability of the method for the analysis of BAs in the two considered matrices.

Table 4.2. Calculated intra- and inter- day accuracies (i.e. recovery) and precisions (i.e. %RSD) of the LC-MS method in standard solutions and spiked urine and feces samples.

		Accuracy ± RSD																	
		Standard solutions						Spiked urine samples						Spiked feces samples					
		Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
Class	**Metabolite**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Primary	MCA	87 ± 18	109 ± 3	114 ± 2	102 ± 13	100 ± 12	112 ± 2	109 ± 16	90 ± 2	103 ± 12	111 ± 5	97 ± 7	106 ± 5	109 ± 5	95 ± 15	113 ± 3	104 ± 15	101 ± 11	107 ± 10
	CDCA	89 ± 11	97 ± 5	108 ± 6	100 ± 11	93 ± 6	105 ± 8	101 ± 6	92 ± 12	109 ± 9	110 ± 7	106 ± 12	103 ± 11	98 ± 11	92 ± 15	106 ± 3	96 ± 8	93 ± 11	103 ± 4
	CA	105 ± 9	101 ± 11	115 ± 7	103 ± 16	102 ± 10	108 ± 6	112 ± 8	92 ± 7	102 ± 5	114 ± 2	98 ± 5	102 ± 7	118 ± 18	85 ± 3	96 ± 12	100 ± 16	93 ± 8	99 ± 8
	α-MCA	110 ± 11	101 ± 13	104 ± 11	111 ± 6	91 ± 10	103 ± 7	110 ± 10	94 ± 14	107 ± 2	107 ± 3	105 ± 9	106 ± 1	96 ± 15	104 ± 3	86 ± 6	99 ± 4	103 ± 5	98 ± 15
	β-MCA	84 ± 3	111 ± 6	87 ± 2	92 ± 15	107 ± 3	98 ± 14	82 ± 18	86 ± 14	98 ± 9	98 ± 18	101 ± 13	102 ± 6	81 ± 12	105 ± 7	94 ± 5	96 ± 13	106 ± 2	97 ± 3
	HCA	112 ± 17	87 ± 5	109 ± 8	111 ± 7	99 ± 12	106 ± 11	85 ± 8	85 ± 13	104 ± 6	94 ± 16	105 ± 17	100 ± 12	117 ± 4	107 ± 7	92 ± 6	99 ± 18	94 ± 13	92 ± 1
Primary conjugated	GCDCA	112 ± 8	96 ± 12	92 ± 12	98 ± 13	91 ± 4	94 ± 5	96 ± 13	114 ± 7	114 ± 12	91 ± 5	104 ± 15	106 ± 10	102 ± 18	92 ± 3	105 ± 12	95 ± 7	97 ± 7	100 ± 5
	GHCA	117 ± 8	110 ± 2	103 ± 8	102 ± 16	108 ± 5	108 ± 4	112 ± 4	90 ± 12	85 ± 3	98 ± 13	91 ± 2	95 ± 11	104 ± 15	110 ± 9	96 ± 11	101 ± 7	100 ± 12	98 ± 9
	GCA	111 ± 6	99 ± 13	115 ± 1	104 ± 6	98 ± 1	106 ± 9	119 ± 14	98 ± 14	106 ± 3	110 ± 7	100 ± 3	110 ± 5	115 ± 6	101 ± 12	113 ± 8	101 ± 24	102 ± 2	98 ± 13
	TCDCA	114 ± 19	90 ± 7	108 ± 11	102 ± 12	102 ± 11	99 ± 10	113 ± 7	88 ± 2	99 ± 6	98 ± 15	100 ± 11	102 ± 4	117 ± 7	96 ± 2	107 ± 10	101 ± 23	96 ± 1	100 ± 5
	TCA	92 ± 8	93 ± 8	108 ± 5	102 ± 11	92 ± 5	99 ± 8	95 ± 11	85 ± 15	96 ± 5	96 ± 10	103 ± 15	97 ± 11	108 ± 14	95 ± 5	111 ± 7	104 ± 8	99 ± 5	105 ± 8
	T-α-MCA	103 ± 8	94 ± 6	111 ± 4	107 ± 4	100 ± 10	107 ± 8	89 ± 9	88 ± 12	89 ± 10	102 ± 15	95 ± 8	98 ± 13	80 ± 14	90 ± 15	88 ± 3	100 ± 19	100 ± 11	94 ± 5
	THCA	104 ± 12	115 ± 7	113 ± 7	101 ± 12	100 ± 14	107 ± 6	93 ± 14	95 ± 9	96 ± 9	102 ± 10	101 ± 11	97 ± 10	100 ± 4	98 ± 2	103 ± 11	98 ± 7	99 ± 14	104 ± 12

Table 4.2 (continuation). Calculated intra- and inter- day accuracies (i.e. recovery) and precisions (i.e. %RSD) of the LC-MS method in standard solutions and spiked urine and feces samples.

		Accuracy ± RSD																	
		Standard solutions						Spiked urine samples						Spiked feces samples					
		Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
Class	**Metabolite**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Secondary	LCA	101 ± 5	89 ± 12	105 ± 12	90 ± 9	99 ± 10	96 ± 9	101 ± 19	113 ± 6	103 ± 8	103 ± 7	112 ± 2	106 ± 3	87 ± 10	95 ± 10	94 ± 2	103 ± 14	105 ± 9	102 ± 11
	DCA	106 ± 14	96 ± 8	102 ± 7	104 ± 2	92 ± 4	103 ± 1	92 ± 6	91 ± 10	113 ± 11	99 ± 13	102 ± 12	106 ± 11	96 ± 12	111 ± 11	103 ± 4	108 ± 11	105 ± 10	109 ± 5
	UDCA	103 ± 4	114 ± 15	107 ± 2	107 ± 5	95 ± 15	102 ± 5	100 ± 4	105 ± 12	94 ± 1	106 ± 5	107 ± 3	103 ± 8	100 ± 10	95 ± 4	101 ± 6	109 ± 9	103 ± 9	105 ± 7
	HDCA	112 ± 4	103 ± 6	104 ± 10	102 ± 11	95 ± 8	98 ± 6	112 ± 14	112 ± 12	100 ± 10	101 ± 11	118 ± 5	104 ± 6	115 ± 17	108 ± 15	111 ± 8	107 ± 9	100 ± 15	107 ± 7
	DHCA	91 ± 8	90 ± 12	101 ± 6	99 ± 14	98 ± 10	109 ± 7	110 ± 8	105 ± 9	109 ± 8	111 ± 2	101 ± 14	105 ± 6	106 ± 12	105 ± 4	114 ± 9	107 ± 13	104 ± 8	102 ± 12
Secondary conjugated	GLCA	112 ± 9	100 ± 9	109 ± 10	101 ± 10	99 ± 12	106 ± 6	98 ± 10	93 ± 2	104 ± 10	105 ± 9	101 ± 9	102 ± 2	101 ± 12	93 ± 15	108 ± 12	104 ± 13	106 ± 11	106 ± 7
	GDCA	109 ± 9	106 ± 12	87 ± 3	96 ± 12	101 ± 10	100 ± 14	106 ± 3	102 ± 7	97 ± 6	105 ± 14	106 ± 5	100 ± 12	100 ± 19	93 ± 4	115 ± 6	102 ± 15	100 ± 11	106 ± 10
	GUDCA	104 ± 7	94 ± 3	107 ± 2	101 ± 6	104 ± 10	102 ± 6	97 ± 13	101 ± 3	105 ± 12	111 ± 12	102 ± 11	108 ± 6	95 ± 2	85 ± 2	104 ± 6	105 ± 10	115 ± 15	109 ± 5
	GHDCA	115 ± 5	96 ± 9	114 ± 6	105 ± 10	96 ± 3	107 ± 9	95 ± 2	104 ± 12	101 ± 10	98 ± 10	105 ± 1	103 ± 7	103 ± 18	114 ± 12	104 ± 4	110 ± 8	110 ± 6	108 ± 4
	GDHCA	84 ± 4	85 ± 9	104 ± 11	98 ± 19	86 ± 2	105 ± 9	109 ± 5	100 ± 10	113 ± 12	108 ± 14	106 ± 5	101 ± 11	115 ± 7	95 ± 6	111 ± 9	103 ± 12	99 ± 10	103 ± 15
	TLCA	87 ± 8	100 ± 4	99 ± 9	96 ± 11	96 ± 6	101 ± 13	99 ± 13	90 ± 4	107 ± 3	101 ± 17	96 ± 7	100 ± 8	106 ± 19	103 ± 2	113 ± 7	106 ± 12	106 ± 5	102 ± 14
	TDCA	119 ± 9	88 ± 5	106 ± 11	102 ± 15	90 ± 5	105 ± 6	105 ± 11	108 ± 2	95 ± 11	104 ± 15	100 ± 10	100 ± 7	96 ± 5	96 ± 3	112 ± 7	109 ± 12	103 ± 7	106 ± 10
	TUDCA	87 ± 12	114 ± 3	110 ± 7	94 ± 7	108 ± 7	103 ± 7	94 ± 6	93 ± 2	110 ± 8	105 ± 11	95 ± 7	109 ± 3	102 ± 13	104 ± 11	111 ± 8	107 ± 5	102 ± 15	105 ± 5
	THDCA	90 ± 4	105 ± 7	97 ± 5	96 ± 10	100 ± 13	102 ± 5	109 ± 18	104 ± 15	113 ± 1	110 ± 3	101 ± 9	112 ± 1	112 ± 11	109 ± 8	112 ± 10	104 ± 13	106 ± 4	107 ± 8
	TDHCA	92 ± 16	94 ± 11	87 ± 2	99 ± 8	99 ± 4	88 ± 3	93 ± 3	97 ± 14	115 ± 4	104 ± 10	105 ± 8	106 ± 15	117 ± 17	105 ± 8	113 ± 2	113 ± 9	100 ± 11	103 ± 12

Table 4.2 (continuation). Calculated intra- and inter- day accuracies (i.e. recovery) and precisions (i.e. %RSD) of the LC-MS method in standard solutions and spiked urine and feces samples.

		Accuracy ± RSD																	
		Standard solutions						Spiked urine samples						Spiked feces samples					
		Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
Class	**Metabolite**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**	**Low**	**Medium**	**High**
Sulphated	GCA-S	111 ± 12	105 ± 7	106 ± 8	112 ± 7	102 ± 5	96 ± 10	102 ± 16	113 ± 1	103 ± 4	101 ± 5	103 ± 9	109 ± 5	98 ± 7	99 ± 2	89 ± 16	105 ± 7	101 ± 4	102 ± 13
	UDCA-S	91 ± 12	95 ± 10	101 ± 1	95 ± 9	117 ± 8	97 ± 8	96 ± 3	107 ± 9	101 ± 4	101 ± 12	107 ± 9	105 ± 6	91 ± 6	111 ± 12	111 ± 12	100 ± 11	106 ± 5	107 ± 4
	LCA-S	104 ± 19	108 ± 13	89 ± 3	108 ± 4	100 ± 7	91 ± 2	96 ± 2	111 ± 13	104 ± 8	102 ± 13	104 ± 14	105 ± 4	109 ± 7	91 ± 7	103 ± 9	104 ± 6	100 ± 9	104 ± 8
	DCA-S	107 ± 15	112 ± 10	114 ± 10	94 ± 12	104 ± 8	110 ± 4	91 ± 16	88 ± 6	105 ± 4	100 ± 8	101 ± 11	106 ± 3	90 ± 16	105 ± 5	101 ± 5	96 ± 11	99 ± 6	104 ± 10
	CDCA-S	120 ± 4	87 ± 15	102 ± 4	111 ± 9	95 ± 16	106 ± 8	81 ± 11	103 ± 6	107 ± 2	91 ± 10	98 ± 11	105 ± 9	109 ± 19	107 ± 9	104 ± 4	107 ± 6	108 ± 1	102 ± 11
	CA-S	113 ± 10	111 ± 4	95 ± 1	104 ± 15	101 ± 12	94 ± 2	115 ± 4	94 ± 11	92 ± 2	109 ± 12	102 ± 7	106 ± 13	102 ± 6	93 ± 11	95 ± 1	98 ± 6	100 ± 11	99 ± 4

4.4.2. BAs in urine and fecal samples

In order to test the applicability of the validated UHPLC-MS/MS method, a set of 73 human urine samples and 20 murine fecal samples were analyzed. **Table 4.3** summarizes the results obtained from the quantification of BAs in human urine and murine fecal samples, after normalization to creatinine concentration and sample weight, respectively. Results show that GCA-S and CA are the most abundant BAs in human urine samples, while TLCA, THCA, UDCA, HDCA, GDHCA and LCA-S were the lowest concentrated compounds detected. In murine feces samples, the most abundant metabolites were the murine specific BAs α-MCA and β-MCA and the secondary BAs DCA and DHCA, while the lowest concentrated BAs were TLCA and GCA-S. 20 and 31 BAs were detected in more than 50% of the urine and feces samples, respectively, being TLCA the less frequently detected in both matrices and GLCA in urine.

Table 4.3. BAs concentrations in human urine and murine feces samples, respectively.

Class	Metabolite	Urine (N = 73), nmol/g creatinine				Feces (N = 20), nmol/g faeces			
		Range	Median	IQR (25-75)	Detection frecuency (%)	Range	Median	IQR (25-75)	Detection frecuency (%)
Primary	MCA	0.8 - 40	1.8	1.3	57	2 - 26	5	8	100
	CDCA	0.8 - 30	2	2	41	4 - 72	17	19	100
	CA	1.2 - 500	14	49	99	5 - 376	38	78	100
	α-MCA	0.3 - 13	1.2	2	81	134 - 791	332	290	100
	β-MCA	0.4 - 20	3	5	77	35 - 610	103	207	100
	HCA	0.6 - 20	1.7	1.2	18	0.6 - 162	3	4	95
Primary conjugated	GCDCA	2 - 30	4	6	70	0.0005 - 1.3	0.06	0.19	100
	GHCA	0.7 - 20	3	3	82	0.3 - 2	0.4	0.6	60
	GCA	0.8 - 16	5	5	82	0.003 - 12	0.2	0.4	100
	TCDCA	0.6 - 80	2	2	43	0.004 - 1.6	0.05	0.2	80
	TCA	0.3 - 13	0.8	0.7	77	0.005 - 6	0.17	0.5	90
	T-α-MCA	0.4 - 70	1.8	3	88	0.006 - 26	0.19	4	90
	THCA	0.3 - 20	0.7	0.7	28	0.007 - 1.6	0.02	0.14	55

Table 4.3 (continuation). BAs concentrations in human urine and murine feces samples, respectively.

Class	Metabolite	Urine (N = 73), nmol/g creatinine				Feces (N = 20), nmol/g faeces			
		Range	Median	IQR (25-75)	Detection frecuency (%)	Range	Median	IQR (25-75)	Detection frecuency (%)
Secondary	LCA	1.9 - 30	4	1.8	100	1.3 - 14	4	3	100
	DCA	2 - 40	5	6	49	35 - 461	182	120	100
	UDCA	0.3 - 14	0.7	0.6	50	35 - 244	66	52	100
	HDCA	0.3 - 12	0.7	0.7	53	21 - 126	61	46	100
	DHCA	0.4 - 30	1.3	1.2	72	19 - 644	117	301	95
Secondary conjugated	GLCA	3 - 3	3	0	1.4	0.002 - 1.12	0.02	0.16	80
	GDCA	0.6 - 20	2	3	68	0.02 - 3	0.12	0.17	100
	GUDCA	0.7 - 45000	4	11	73	0.0006 - 12	0.07	0.18	100
	GHDCA	1.2 - 340	7	15	11	0.003 - 1.1	0.02	0.16	75
	GDHCA	0.3 - 9	0.7	0.9	9	0.004 - 1	0.02	0.10	80
	TLCA	0.5 - 0.5	0.7	0	1.4	< LOD	< LOD	< LOD	0
	TDCA	1.2 - 20	1.5	12	4	0.006 - 1.09	0.05	0.2	80
	TUDCA	1.2 - 20	2	1.2	8	0.3 - 2	1.0	2	60
	THDCA	0.8 - 30	2	9	7	0.09 - 1.5	0.2	0.2	100
	TDHCA	0.3 - 12	0.8	3	7	0.002 - 1.14	0.13	0.06	45
Sulphated	GCA-S	0.5 - 200	50	50	100	0.013 - 1.05	0.012	0.07	45
	UDCA-S	0.3 - 20	4	5	92	0.02 - 1.4	0.8	1	90
	LCA-S	0.3 - 6	0.7	0.5	54	0.014 - 1.09	0.03	0.2	70
	DCA-S	0.3 - 600	1.4	4	81	0.2 - 23	5	6	100
	CDCA-S	0.3 - 30	1.6	3	92	0.3 - 22	5	6	100
	CA-S	0.7 - 5	3	5	3	1 - 46	11	13	100

Note: IQR = interquartile range; LOD = limit of detection

Differences in concentration profiles of BAs in human urine and murine feces samples were observed.

Figure 4.2 depicts the BA metabolism pathways in conjunction with BA concentration ranges found in the studied samples according to species and matrix. It shows that the most concentrated metabolites in human urine belong to the CA metabolism (the classical pathway for BA metabolism in humans), and that the most concentrated metabolites in murine feces belong to the CDCA and UDCA metabolism (the classical pathway in murine species). These findings are in agreement with Lin et al. results, where CA metabolism prevails over CDCA metabolism in humans, whereas CDCA metabolism is significantly enhanced in murine species [187]. Additionally, most of BAs present in human urine samples correspond to secondary conjugated BAs (e.g., GLCA and TUDCA), followed by sulphated BAs (e.g., GCA-S and CA-S), being those classes the less concentrated in murine feces samples. These results are in concordance with the literature [195], where most concentrated urinary BAs were conjugated and sulphated, confirming the urinary BA excretion mechanisms in humans. Thus, differences in BA profiles observed between human urine and murine feces samples demonstrate the different BA excretion mechanisms [186,187,205], what must be taken into consideration in studies focused on the interrelation of gut microbiota and BA metabolism.

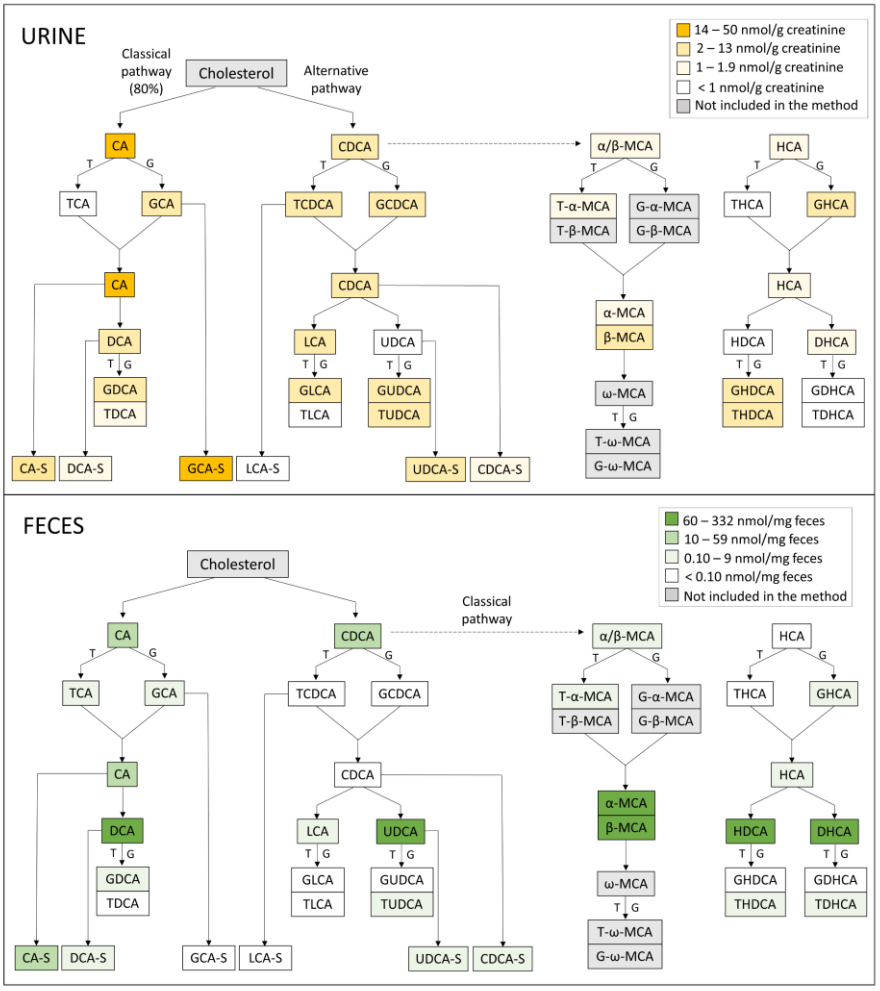


Figure 4.2. BA pathways and concentration ranges in human urine and murine feces samples. Note: T = taurine, G = glycine.

Additionally, two clusters were identified after a hierarchical clustering analysis (HCA) in both urine and feces samples and projected on a Principal Component Analysis (PCA) scores plot, as shown in **Figure 4.3**. As

observed in the loadings plot, urine samples corresponding to cluster 1 presented higher concentrations of primary, sulphated and secondary conjugated BAs, while samples from cluster 2 were more concentrated in secondary and primary conjugated BAs. This can also be evidenced in the BAs classes abundance distribution, as exemplified in the pie charts from four of the samples in each cluster. Regarding feces, samples included in cluster 2 present higher concentrations in primary and secondary BAs, while samples from cluster 1 were less concentrated in those BA classes, as evidenced in the sample pie charts.

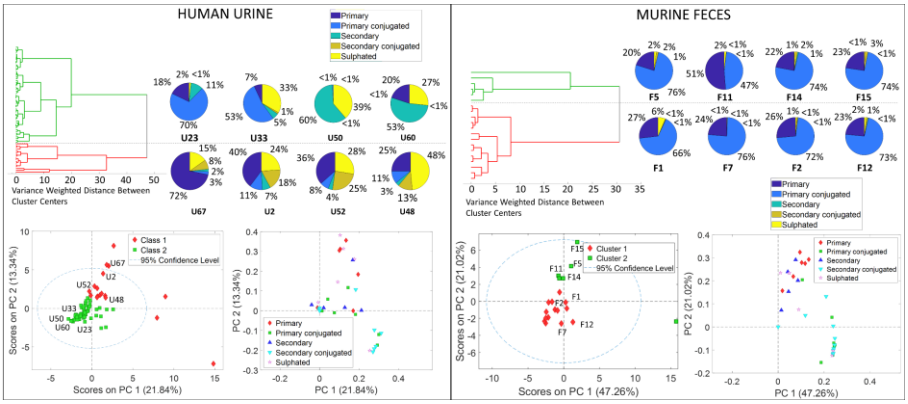


Figure 4.3. BAs patterns in human urine and murine feces samples. Top: Hierarchical clustering analysis revealing two sub-groups within the study samples (left) and pie charts of four samples from each cluster as examples (right). Bottom: PCA scores (left) and loadings (right) plots.

4.5. Conclusions

An UHPLC-MS/MS method suitable for the quantitative analysis of BAs in urine and feces samples has been validated. Levels of human urinary BAs and murine fecal samples were reported proving the applicability of the method in the (pre-)clinical and clinical fields. Furthermore, a comparison between human urinary and murine fecal BAs metabolic pathways and excretion mechanisms was provided. Additionally, two clusters regarding BA concentrations in both urine and feces samples were observed.

The novel approach of BA analysis in both human urine and murine feces samples provides valuable insights into the metabolic processes, allowing for a comprehensive understanding of BA metabolism across different species. In future studies, integrating BAs and microbiota analysis could provide a more comprehensive understanding of the relationship between diet, microbiome, health, and disease.

Supplementary materials

Supplementary data to this article can be found online at https://link.springer.com/article/10.1007/s00216-023-04802-8.

Section II. Effect of pasteurization on the metabolome and lipidome of HM

Chapter 5. Current practice in untargeted human milk metabolomics

5.1. Abstract

Human milk (HM) is considered the gold standard for infant nutrition. HM contains macro- and micronutrients, as well as a range of bioactive compounds (hormones, growth factors, cell debris, etc.). The analysis of the complex and dynamic composition of HM has been a permanent challenge for researchers. The use of novel, cutting-edge techniques involving different metabolomics platforms has permitted to expand knowledge on the variable composition of HM. This review aims to present the state-of-the-art in untargeted metabolomic studies of HM, with emphasis on sampling, extraction and analysis steps. Workflows available from the literature have been critically revised and compared, including a comprehensive assessment of the achievable metabolome coverage. Based on the scientific evidence available, recommendations for future untargeted HM metabolomics studies are included.

5.2. Introduction

Human milk (HM) has been markedly established as the optimal way of providing infants with the necessary nutrients and bioactive factors for their early development. Many health associations and organisms, including World

Health Organization, recommend exclusive breastfeeding for the first six months of life [206]. Health benefits of HM for infants include reduced mortality and morbidity, including sepsis, respiratory diseases, otitis media, gastroenteritis, and urinary tract infections, among others [207]. In addition, studies reporting on long-term benefits of HM consumption such as lower risk of suffering from type 1 diabetes and inflammatory bowel disease or overweight in adulthood emerged [208]. HM may also be associated with a slightly improved neurological outcome as cohort studies report [209], especially in preterm infants [210], although potential confounders must be accounted for [211].

HM composition is dynamic and influenced by several factors including genetics, gestational and infant's age, circadian rhythm, maternal nutrition, or ethnicity. It provides a series of nutrients such as lipids, proteins, carbohydrates, and vitamins, jointly with a number of bioactive factors that contribute to several physiological activities in the newborn infant as well as to short- and long-term outcomes [212,213]. Living cells including stem cells, hormones, growth factors, enzymes, microbiota and even genetic material are part of this vast array of HM components with impact in early development, particularly the immune system [17]. In addition, HM appears to be one of the richest sources of microRNAs [214]. On the other hand, because of the maternal environmental exposure and lifestyle, the presence of some contaminants such as persistent organic pollutants or pharmacologically active substances in HM has been described [215,216].

Due to its complex composition, the analysis of HM is not straightforward. While the advent of "omics" approaches has offered valuable insights into the composition of this unique biofluid, untargeted metabolomic and lipidomic studies have only recently been applied to HM [217]. The

comprehensive study of the HM metabolome, which includes the intermediate and end products of metabolism, can shed light on maternal status or phenotype [218,219]. The generation, analysis and integration of large and complex data sets obtained in metabolomic studies go hand in hand with the following challenges: i) the intrinsic complexity of the sample: a rich variety of jointly present, structurally heterogeneous compounds at concentrations that strongly vary covering several orders of magnitude; ii) pre-analytical steps related to sampling, storage, and pre-processing (e.g. extraction, clean-up); and iii) the diversity of platforms currently available including nuclear magnetic resonance (NMR), as well as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) coupled to mass spectrometry (MS). The analysis of the HM metabolome has been approached employing a variety of extraction and analytical techniques to respond to a spectrum of clinically relevant questions. Several studies have compared HM metabolome with formula milk [217,220–224] or with milk from other mammalian species including monkey [225], donkey [221], and cow [222], whereas others have made efforts in defining the metabolome of preterm milk [217,220,226–230] and the evaluation of the HM metabolome during the course of lactation [219,227,231–234]. Furthermore, the influence of maternal diet [218,219,235], phenotype [218,236], obesity [234] or atopy status [237], as well as geographical location [237,238], time of the day [233,239], chemotherapy [240] or preeclampsia during pregnancy [235] on the HM metabolome have been reported.

Recent review articles that address the HM metabolome or lipidome [216,241–245] are available. For information on the compounds and compound families typically found in HM and their function the reader is referred to [241–244]. Readers with a particular interest in HM lipidomics are referred to a recent compilation study [245]. Technical aspects of HM analysis

when performing metabolomics studies in HM have been recently described [216]. This review article gathers recent literature available on metabolomic analysis of HM, particularly focusing on untargeted approaches as indicated in **Figure 5.1**, to provide an up-to-date overview of the key factors that may influence HM metabolome coverage. Based on the information provided within the available literature, recommendations to guide study design and analytical method development of untargeted HM metabolomics assays were developed.

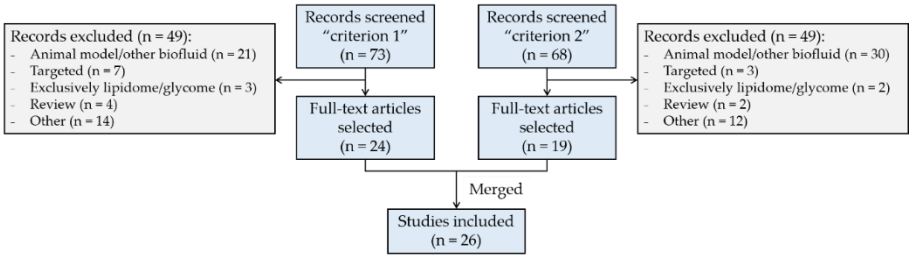


Figure 5.1. Flow diagram of literature selection and review process. Search criterion (i): term ("human milk" OR "breast milk"), AND "metabolom*", AND "infant"; only articles. Search criterion (ii): term ("human milk" OR "breast milk"), AND "metabolom*", AND ("GC" OR "LC" OR "NMR" OR "CE"); only articles. Web of Science database was employed for literature search.

5.3. Considerations regarding the study design

HM is a biofluid characterized by a dynamically varying composition according to several factors including lactation time, time of the day, throughout each feed, maternal status, and the environmental exposure.

Although compositional variations have been mainly studied regarding the protein content of HM [53], changes of other compound classes such as fat or vitamins have been also reported [246,247]. Considering the intrinsic variability of HM, the complexity of obtaining representative HM samples is not negligible. Sources of variation related to sample manipulation and compositional variation can be minimized using standard operational procedures (SOPs). SOPs are fundamental to maintain quality control (QC) and quality assurance (QA) process and facilitate repeatable and reproducible research within and across laboratories. However, biologically meaningful results across studies will only be obtained if several key factors during the sample collection process are successfully controlled. This is of special importance in untargeted approaches, where the interpretation of results is especially challenging, and confounding factors introduced by a non-exhaustive sampling protocol can be wrongly attributed to differences between subjects of a studied population. Conversely, biologically meaningful information can be missed or remain unnoticed due to unwanted bias introduced during sample collection.

The information regarding study design provided in HM metabolomics studies varies considerably [217–228,230–236,238–240,248–250] as shown in **Figure 5.2**. Repeatedly reported factors have been grouped into three categories and are discussed in detail in the following sections: 1) maternal-infant-related factors (blue bars), 2) time-related factors (green bars), and 3) HM collection-related factors (orange bars). It should be noted that, although the importance of each factor might vary with the scientific question of each study, the authors encourage i) the use of SOPs employed during sample collection to assure homogenous and representative sampling and ii) the reporting of all documented factors in order to enhance comparability between results of metabolomic studies on HM. In case of HM,

samples are typically collected, handled and sometimes temporary stored and transported by the mothers and not, such as it is the case for other biofluids (e.g. plasma, serum), by health professionals. During study design it is therefore very important to assure that mothers receive detailed instructions and/or training for the correct handling of collected samples. In addition, one should keep in mind that sampling protocols should neither interfere with infant feeding nor negatively impact on the mother-baby bonding. Hence, the collection of transitional and mature milk is usually preferred over colostrum, especially in studies involving mothers of preterm infants, where colostrum is usually kept exclusively for the infant's supply.

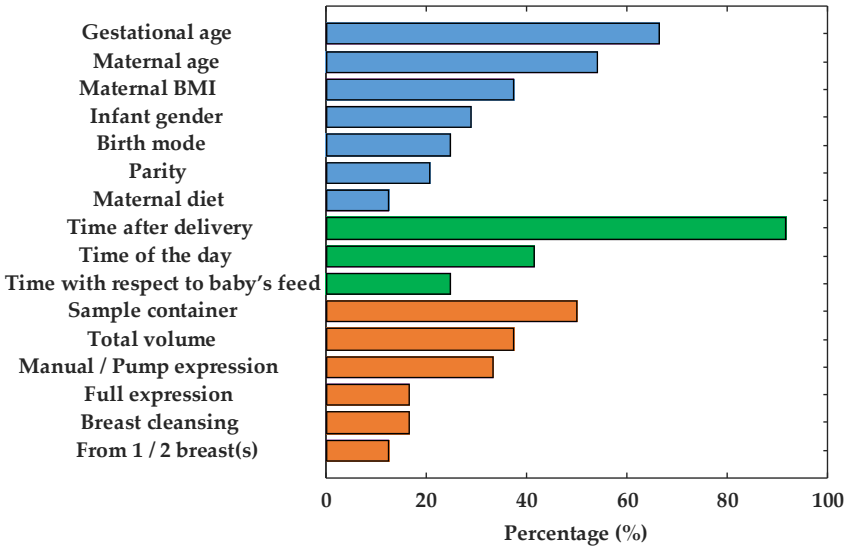


Figure 5.2. Reporting frequency of factors relevant to the HM sampling process: Maternal-infant-related factors (blue bars), time-related factors (green bars), and HM collection-related factors (orange bars). Note: BMI = body mass index.

5.3.1. Maternal-infant-related factors

In HM metabolomic studies, gestational age is frequently reported (see **Figure 5.2**), although the impact of this factor on the HM metabolome has not been fully characterized. Studies focused on preterm milk showed that, analogously to full-term milk, its composition is dynamic throughout the first month of lactation [217,220,226]. However, after 5-7 weeks, metabolite composition of HM from mothers of preterm infants resembled that collected from mothers of full-term infants [227]. On the other hand, Marincola et al. [217] observed that HM from mothers of early preterm infants (26 weeks of gestation) differentiated from milk samples from term infants. However, the low number of samples involved in the study (N = 20 and N = 3 mothers of preterm and term infants, respectively) hindered the assessment of the statistical significance of the impact of gestational age on the milk metabolite composition. Sundekilde et al. [227] carried out a longitudinal study on milk from mothers of preterm and full-term infants covering similar lactation periods (3-14 weeks and 3-26 weeks after birth, respectively) and showed that some metabolites were present at significantly different levels in full-term milk compared to preterm milk. On the contrary, Longini et al. [220] did not observed significant differences between preterm and full-term milk within the first week after delivery, only being able to discriminate milk samples from early preterm infants (<29 weeks of gestation). It is worth noting that the effect of gestational age on the HM metabolome has been mainly studied employing NMR platforms [217,220,226,227], in which metabolite coverage is limited (see **Figure 5.5**) and some metabolite classes (e.g. lipids) are barely accessible. For this reason, and in order to further evaluate the impact of gestational age on the milk metabolome, we warrant more comprehensive metabolomic studies employing complementary analytical platforms.

Other potentially relevant, miscellaneous information about the studied population of mother-infant pairs such as infant gender, parity, and birth mode, have been frequently reported in metabolomic studies (see **Figure 5.2**), although these characteristics often remain in the background since the studies focus on other aspects. The influence of these factors on the metabolite composition of HM has not been elucidated yet, and this might be addressed in forthcoming studies. An additional factor that is not typically reported in HM metabolomics studies is maternal secretor status. Significant differences in the oligosaccharides profile of milk between so-called secretors (Se+), which are those mothers that provide a functional FUT2 gene, and non-secretors (Se-) have been reported [251]. Secretor status is mainly established based on the presence (Se+) or absence (Se-) of 2'-fucosyllactose, with a prevalence rate of approximately 80% of secretors over non-secretors [218,226,227,230,236,252]. Maternal secretor status is therefore usually determined a posteriori during data processing and analysis. Oligosaccharides are polar compounds that are present at concentrations in the mM range that will likely be preserved during sample extraction procedures employed for metabolomics studies. As their presence/absence might potentially affect clustering of milk based on maternal secretor status [228], to provide this information, when available, might be of interest.

5.3.2. Time-related factors

HM undergoes significant changes over time, having established three differentiated lactation stages: colostrum, transitional milk and mature milk. As can be seen in **Figure 5.2**, lactation time is reported in the vast majority of

metabolomic studies. In particular, several studies have focused on the HM metabolome throughout lactation [219,227,231–234], all of them concluding that significant differences in the metabolic profile over time exist. Therefore, it seems reasonable to report this factor. On the other hand, although it has been demonstrated that diurnal variation affects HM fat content [253], its effect on the overall metabolite composition is a controversial issue which has not yet been adequately addressed in the available literature. Whereas no significant changes in some lipids and small polar metabolites have been observed [233,239], differences in some micronutrients (e.g. vitamins) could be evidenced [247]. The use of a pool of a 24-h expression of HM should compensate for changes due to diurnal variation, thus obtaining more representative samples [217,228,229] in longitudinal studies. However, the drawback is that this practice is incompatible with breastfeeding of the infant, which, in turn, might raise severe ethical concerns. A feasible compromise for ameliorating diurnal variations is the use of pooled morning and evening samples [233,239].

Regarding time of collection with respect to baby's feed, the influence of this variable has not been studied to date, but given the differences found between fore- and hindmilk [254], it seems reasonable to assume that this factor might be potentially relevant.

5.3.3. HM collection-related factors

Any uncontrolled variable within an experiment can result in a potential source of bias. In this sense, although less attention has been payed to other factors related to the expression and storage of HM (see **Figure 5.2**, orange bars), they may be relevant to the outcomes of metabolomic studies.

As it can be seen, the type of sample container is indicated in 50% of the studies, whereas other specifications regarding HM expression are included scarcely. The latter factor deserves some special attention, since differences in the milk fat content between foremilk (initial milk of a feed) and hindmilk (last milk of a feed) have been reported [254]. Therefore, full expression of breast(s) is desirable in order to obtain a representative HM aliquot [255]. The influence of all other factors, to date, remains unstudied.

5.3.4. Pasteurization and storage

HM banks rely on stringent protocols in which pasteurization, indispensable for minimizing the potential to transmit infectious agents, as well as freezing and long-term storage procedures are established. The pasteurization process affects some of the nutritional and biological properties of HM [256–258]. In this review, three studies that use milk from HM banks are included [220,224,227], but only one specifies whether or not HM has undergone pasteurization [227]. Variability of the metabolite profile of HM caused by pasteurization has not been comprehensively explored to date. Future studies focused on the systematic exploration of the effect of thermal treatment on HM are warranted.

HM is usually stored frozen employing -20 ℃ and -80 ℃ for short- and long-term storage, respectively. However, duration of storage and the effect of repeated freeze-thaw cycles are identified as additional factors with potential impact on HM composition that are missing in most published studies. In lipidomic studies, the integrity of HM samples is preserved by subjecting HM to inactivation of endogenous enzymes such as lipases in order to minimize lipolysis and lipogenesis. In this sense, immediate storage at -80

℃ is advisable [245]. Particularly for metabolite composition analysis, storage at -80 ℃ is widespread [217,219,222,223,228–230,232,235,248,249], sometimes with a prior short-term storage at -20 ℃ [218,225,226,231,233,237,238]. Wu et al. [233] investigated the effect of storage conditions by keeping samples for different times at -20 ℃ and then transferring them to -80 ℃ vs storing samples directly at -80 ℃. Variations in duration of storage at -20 ℃ vs -80 ℃ showed no detectable effect on the metabolites considered (e.g. lactose and other carbohydrates, choline and its derivatives, and a variety of amino acids) by visual inspection of sample clusters in principal component analysis scores plots. However, ANOVA analysis evidenced differences in butyrate, caprate, and acetate contents. However, time of storage considered in this study was limited to two weeks, which is not representative for conditions employed in clinical studies or standard home routines. It is therefore clear that further studies are required in this regard.

On the other hand, HM employed for research studies is typically stored in small aliquots. When working with raw milk, this procedure might introduce bias due to phase separation prior to the preparation of the aliquots. Hence, the homogenization of HM with a disruptor, resulting in a stable emulsion with reduced size of milk fat globules [56], as employed prior to the quantitation of macronutrients with HM analyzers, might be advisable.

5.4. Metabolite extraction from HM

For metabolite extraction from HM, an array of methods has been reported. An overview of the employed approaches is shown in **Figure 5.3**. The selection of the extraction method is conditioned by the study objective

and the subsequent analysis method. As in other untargeted metabolomics workflows, for HM metabolomics the selected sample preparation approach should enable a high degree of metabolome coverage while making the sample matrix compatible with the analytical platform. Other considerations might include the available amount of sample volume and the use of one sample extraction procedure for subsequent analysis by multiple, complementary analytical platforms [217,231,232].

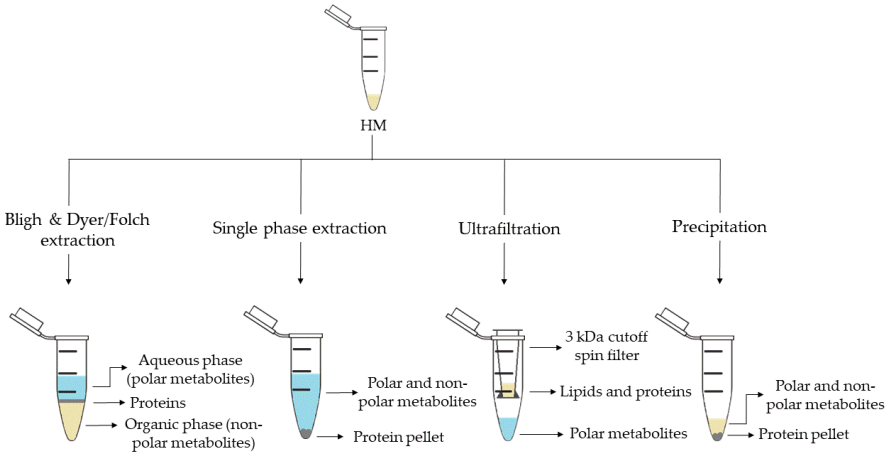


Figure 5.3. Sample preparation approaches employed in HM metabolomics.

Liquid-liquid extraction (LLE) is the classical extraction method employed in metabolomics and lipidomics. This method, developed by Folch et al. [259] in 1957, uses a chloroform-methanol mixture (2:1, v/v), which results in two differentiate phases: an upper phase containing polar metabolites and a lower phase containing nonpolar metabolites. Subsequently, in 1959 Bligh and Dyer [260] developed a modified method using a miscible

chloroform-methanol-water mixture and later separated into two phases by adding chloroform or water. Both approaches enable the separation of polar and nonpolar metabolites, thus allowing the analysis of a wide range of metabolites and making them compatible with several analytical platforms. While the use of Bligh and Dyer LLE is widely extended for HM metabolomics studies (see **Table 5.1**) [217,220–223,228,229,233,236], only Andreas et al. [232] used a modified Folch extraction protocol for processing HM samples.

Methyl tert-butyl ether (MTBE) in combination with methanol has recently been proposed for single-phase extraction [231]. MTBE is a nontoxic and noncarcinogenic solvent and it is therefore considered a safe and environmentally friendly alternative to harmful solvents employed in traditional LLE methods, such as chloroform, which is a suspected human carcinogen. In this extraction method, a unique phase containing both polar and nonpolar metabolites is obtained with a protein pellet at the bottom (see **Figure 5.3**). Thus, the simultaneous analysis of lipidome and metabolome in a very small amount of biological sample is achievable. This method has been successfully employed to determine polar metabolites and fatty acids (FAs) in HM by GC-MS [231,232], as well as lipids and polar metabolites by LC-MS [219,231,232], thus increasing the metabolome coverage by the combined use of complementary analytical platforms.

Ultrafiltration makes use of centrifugal molecular weight cutoff filters. Different molecular weight cut-off filters are commercially available for this purpose and repeated centrifugation steps might be employed to remove proteins and lipids (see **Table 5.1**). Unlike single-phase extraction, ultrafiltration allows to separate polar metabolites from the HM without dilution [218,225,226,233], however this method does not have the capacity

to study the global metabolome of HM. At present, this extraction method has only been used in combination with NMR analyses [218,225,226,233].

Precipitation with organic solvents separates the polar and nonpolar metabolites of the proteins that settle at the bottom of the tube which can then be easily removed by centrifugation. This simple method has been employed for the analysis of polar metabolites by GC-MS after derivatization [240] as well as for the analysis of polar and nonpolar metabolites by LC-MS without further pre-processing [222]. Furthermore, this approach has been implemented in more sophisticated workflows as recently shown by Hewelt-Belka et al. [239]. Here the authors combined LLE and a protein precipitation and solid-phase extraction (SPE) procedure to prepare HM samples, thereby enabling the detection of high- and low-abundant lipid species (e.g. glycerolipids and phospholipids) in one LC-MS run.

5.5. Analytical platforms employed in HM metabolomics

As reflected in **Table 5.1**, the use of all analytical platforms that are commonly employed in untargeted metabolomics studies such as LC-MS, GC-MS, NMR and, to a lesser extent, CE-MS, has been reported for performing HM metabolomics. 1H-NMR is the most frequently used technique [217,218,220,224–226,230,232,233,235–237] for both, the analysis of polar and hydrophobic metabolites in HM. 13C-NMR has been reported for the detection of triacylglycerols [224]. NMR is a highly reproducible technique that allows a straightforward library matching after spectral alignment while at the same time supporting structural elucidation of detected metabolites. However, it presents lower sensitivity and hence, the achievable metabolome coverage is low in comparison to other analytical platforms.

All other analytical platforms rely on the use of MS detection. LC-MS provides high sensitivity and is characterized by a huge versatility due to the availability of i) a large selection of chromatographic columns with a variety of stationary phases, that in combination with appropriate mobile phases achieve compound separation based on different retention mechanisms and ii) an array of different instrumental configurations (i.e. different ion sources and mass analyzers). For example, reversed phase (C8, C18) LC-quadrupole time of flight MS (LC-QTOF-MS) [219,222,231,232,239,249] and LC-Orbitrap-MS [228,229] have been reported for the detection of both, polar and lipidic metabolites in HM; and hydrophilic interaction LC (HILIC)-QTOF-MS for polar metabolite detection [239]. On the other hand, for the successful screening of HM oligosaccharides, a porous graphitic carbon column installed on a LC-triple quadrupole (TQD)-MS instrument was used [228].

GC-MS is the most suitable platform for measuring volatile compounds, while other non-volatile compounds must be derivatized prior to analysis. For HM analysis, methoximation followed by silylation or methylation are commonly employed (see **Table 5.1**). The most frequently used column is the DB-5ms column for both, polar and FA detection [217,221–223], in some cases with an integrated 10 m pre-column (deactivated fused silica) [231,232,240].

Regarding CE-MS, only one study has been reported for polar metabolite detection in HM [232]. CE provides a series of advantages over other techniques, mainly due to the small sample volumes employed and the efficient separation of polar compounds that is difficult to be achieved by LC columns. However, issues with poor reproducibility, matrix effects and sensitivity may be hindering a widely extended use of this technique for the analysis of complex biological samples such as HM.

Due to the diversified composition of HM, no single analytic technique can resolve the entire HM metabolome. Only multiplatform approaches enable a comprehensive characterization providing a high metabolome coverage including polar and nonpolar metabolites present in HM. In this sense, two studies performing a multiplatform approach were found in the literature combining LC-MS and GC-MS [222,231], and only one study that performed analysis using four different techniques (LC-MS, GC-MS, NMR and CE-MS) was reported [232].

Table 5.1. Sample preparation steps and platforms employed in untargeted analysis of HM metabolome.

Sample preparation (1st. step)	Sample preparation (2nd. step)	Compound class	Platform	Column/Capillary	References
Bligh & Dyer extraction	Deuterated solvent addition to aqueous phase	Polar metabolites	^{1}H-NMR	-	[41,44,57,60]
	Derivatization of aqueous phase: methoximation and silylation	Polar metabolites and FAs	GC-MS	DB-5ms	[45–47]
	Derivatization of organic phase: methylation	FAs	GC-MS	DB-5ms	[41]
	Direct injection of aqueous phase	Polar metabolites	LC-QTOF-MS (+)	HILIC	[63]
	Redissolution of aqueous phase in H_2O:ACN (95:5)	Polar metabolites	LC-Orbitrap-MS (+, -)	C18	[52]
	Redissolution of organic phase in (ACN:IPA:H_2O (65:30:5)	Lipidic metabolites	LC-Orbitrap-MS (+, -)	C18	[53]
Folch extraction	Deuterated solvent addition to aqueous and organic phases	Hydrophobic and polar metabolites	^{1}H-NMR	-	[56]
	Redissolution of aqueous phase in formic acid and centrifugation	Polar metabolites (amino acids)	CE-TOF-MS (+)	60 m x 50 µm I.D.	
	Redissolution of organic phase in IPA:H_2O:ACN (2:1:1) and centrifugation	Lipidic metabolites	UPLC-QTOF-MS (+, -)	C18	
Single phase extraction	Derivatization: methoximation and silylation	Polar metabolites and FAs	GC-MS	DB-5ms	[55,56]
	Direct injection	Lipidic (and polar) metabolites	LC-QTOF-MS (+, -)	C8	[55,56]
			UPLC-QTOF-MS (+)	C18	[43]
Fat extraction with n-hexane / IPA	Deuterated solvent addition	TGs	^{13}C-NMR; ^{1}H-NMR	-	[48]
Filtration 3 kDa cutoff spin filter	Deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[42,49,50,57,61]

Table 5.1 (continuation). Sample preparation steps and platforms employed in untargeted analysis of HM metabolome.

Protein precipitation	Derivatization: methoximation and silylation	Polar metabolites	GC-MS	DB-5ms	[64]
	Hybrid SPE-Phospholipid extraction and redissolution in diluted organic phase of Bligh & Dyer extraction	Lipidic metabolites	LC-QTOF-MS (+)	C8	[63]
	Fat removal with CH_2Cl_2 and dansylation of aqueous phase	Polar metabolites (amine/phenol submetabolome)	Chemical isotope labelling LC-QTOF-MS (+)	C18	[73,74]
	Direct injection	Polar metabolites and FAs	UPLC-QTOF-MS (+, -)	C18	[46]
Fat removal by centrifugation	Two additional centrifugations and deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[62]
	Filtration 10 kDa cutoff spin filter and deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[51,54]
Homogenization	Deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[59]
H_2O-dilution	$NaBH_4$-reduction and PGC cartridge	Oligosaccharides	UPLC-TQD-MS (+)	Hypercarb®	[52]

Note: CE, capillary electrophoresis; FAs, fatty acids; GC, gas chromatography; HILIC, hydrophilic interaction liquid chromatography; IPA, 2-propanol; I.D., inner diameter; LC, liquid chromatography; MS, mass spectrometry; NMR, nuclear magnetic resonance; PGC, porous graphitic carbon; QTOF, quadrupole time of flight; TGs, triacylglycerols; TQD, triple quadrupole; UPLC, ultraperformance liquid chromatography.

The use of high-end analytical platforms requires the implementation of QA and QC processes to improve data quality, repeatability and reproducibility, especially in untargeted metabolomics. For practical guidelines on the use of QC measures in untargeted, MS-based assays the reader is referred to [261]. Pooled QC samples are prepared by mixing small aliquots of the study samples and therefore, they are considered representative in terms of matrix composition and concentration ranges of the metabolites present in the study samples. QC samples are analyzed repeatedly throughout the analytical sequence alongside the study samples. The signal of each feature detected in QC samples can be used to model and correct systematic changes in the instrument response during the analytical sequence. Also, the obtained data can be used to perform intra-study reproducibility assessments and to correct for systematic variation across batches. In HM metabolomics, Smilowitz et al. [218], Andreas et al. [232], and Gay et al. [237] used QC samples for NMR studies, while Villaseñor et al. [231], Mung et al. [249], Hewelt-Belka et al. [239], and Alexandre-Gouabau et al. [228,229] used pooled HM samples for QC purposes in LC-MS-based assays. Considering the highly complex sample matrix of HM, the authors strongly recommend the implementation of QC measures, including the analysis of QC samples, to increase reproducibility and facilitate the joint analysis of data from different studies.

5.6. The HM metabolome: compound annotation & coverage

As in other areas of metabolomic research, compound identification is still a major bottleneck in data analysis and interpretation. The

Metabolomics Standards Initiative's (MSI) defines four levels of metabolite identification, which include: identified metabolites (level 1); putatively annotated compounds (level 2); putatively annotated compound classes (level 3); and unknown compounds (level 4) [9]. Due to the limited availability of pure analytical standards required to reach level 1, biological databanks and spectral databases are the most important resources for metabolite annotation (levels 2 and 3). A large number of databases are available today, providing different levels of information and complementary data on chemical structures, physicochemical properties, biological functions, and pathway mapping of metabolites [262]. The metabolomics community classifies these resources in several categories: i) chemical databases; ii) spectral libraries; iii) pathway databases; iv) knowledge databases; and v) references repositories [263].

Regarding HM metabolomics, the most frequently used databases and libraries are: Human Metabolome Database (HMDB) [264], METLIN [265], National Institute of Science and Technology (NIST) library, Fiehn RTL Library [266], LipidMAPS Structure Database (LMSD) [267], Milk Metabolome Database (MCDB) [268,269], Kyoto Encyclopedia of Genes and Genomes (KEGG) [270], MycompoundID with the evidence-based metabolome library (EML) [271], Chenomx NMR Suite Profiles and other online university databases, such as CEU-mass mediator [272,273].

Metabolite assignment in NMR spectra has been performed based on literature data and commercial resonance databases, such as Chenomx NMR Suite Profiles. Metabolite annotation was contrasted with in-house libraries containing pure compound spectra. Some of the proposed assignments were confirmed by two-dimensional NMR spectra, such as Correlation Spectroscopy (COSY) [217,233,235,236], Homonuclear Correlation

Spectroscopy (TOCSY) [217,235,236,238], Diffusion-Ordered Spectroscopy (DOSY) [236], Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) [236,238] and Heteronuclear Multiple Bond Correlation (HMBC) [236].

In LC-MS and CE-MS-based studies of the HM metabolome, tentative metabolite annotation has been carried out by matching of accurate masses, isotopic profiles, and/or fragmentation patterns to candidate metabolites in online databases such as KEGG, METLIN, LipidMAPS, and HMDB [222,228,229,231,232,239]. In-house built databases generated by the analysis of commercial standards are also commonly employed [228,229]. In GC-MS, retention index (RI) corrections are made by analyzing a fatty acid methyl ester (FAME) mixture standard solution and assigning a match score between the experimental FAME mixture and theoretical RI values based on the values contained in the Fiehn RTL library. Furthermore, metabolites were complementarily annotated by comparing their mass fragmentation patterns with those available in Fiehn RTL and NIST libraries [217,221–223,231,232,240].

A comprehensive list of annotated and/or identified metabolites in HM from untargeted metabolomics studies [218,219,221–223,225–233,235–240] is reported in Table S5.1. This table contains information about the metabolites reported in each reference, such as their molecular formula, IDs (LipidMAPS and/or HMDB IDs), the extraction procedure performed, the analytical platform used, and the detected metabolite class. Readers can select metabolites dynamically by filtering data according to the latter information. A total of 1187, 111 and 128 metabolites were reported using LC-MS, GC-MS and NMR, respectively (see **Figure 5.4**). As shown in the Venn diagram, LC-MS and GC-MS allowed the detection of 36 common metabolites (mainly

carbohydrates and FAs); a total of 29 metabolites overlapped between LC-MS and NMR (principally oligosaccharides); and 21 metabolites (predominantly amino acids and organic acids) were commonly reported in GC-MS and NMR based studies. Only 13 metabolites were reported by all three platforms, i.e. creatine, tyrosine, arabinose, galactose, glucose, lactose, maltose, capric acid/caprate, caprylic acid/ caprylate, citric acid/citrate, pyruvic acid/pyruvate, hippuric acid/hippurate, and myo-inositol. These metabolites were assigned to different classes including amino acids, carbohydrates, FAs, and organic acids.

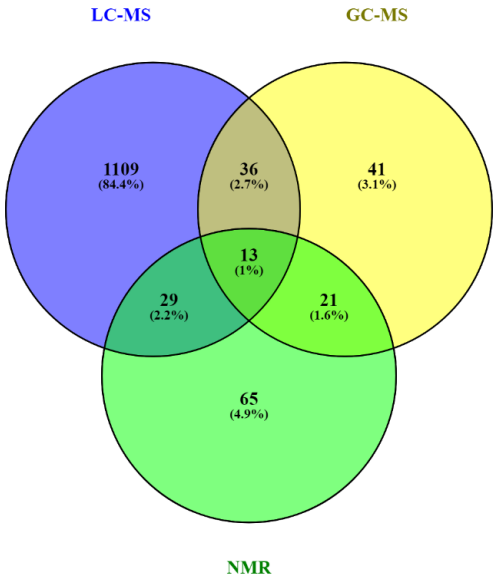


Figure 5.4. Venn diagram of metabolites reported in HM according to technique [274]. Note: GC-MS, gas chromatography—mass spectrometry; LC-MS, liquid chromatography—mass spectrometry; NMR, nuclear magnetic resonance.

Based on the available data from the literature, the distribution of metabolite classes present in HM according to each technique was assessed. As it can be seen in **Figure 5.5**, the difference in detected metabolite classes as observed by LC-MS in comparison to GC-MS and NMR is evident. Using GC-MS and NMR, carbohydrates are the most reported metabolites in HM, followed by amino acids, organic acids, organooxygen compounds, and organoheterocyclic compounds, with all these metabolite classes being certainly less abundant in LC-MS studies. In the case of NMR, organonitrogen compounds have also been reported, as well as nucleosides and nucleotides on a smaller scale. In the case of lipid classes, fatty acyls have been identified by LC-MS and GC-MS with similar incidence and in lesser extent by NMR. It is indubitable that lipid classes are more comprehensively studied by LC-MS assays, where glycerophospholipids, glycerolipids, and fatty acyls are detected at relatively high abundances, followed by sphingolipids, sterol lipids and, to a lesser extent, prenol lipids.

Table 5.2 shows a list of metabolites reported in >80% of studies employing either LC-MS, GC-MS, or NMR-based assays. This table is intended to aid method development of future untargeted metabolomics workflows tailored to the study of the HM metabolome, as it shows a shortlist of metabolites that should be detected by each platform regardless of the instrumental settings employed. It should be noted that due to the high versatility of LC-MS, there is a greater variation in metabolites recorded and in return, the list of consistently reported metabolites in HM across studies is shorter than for NMR and GC-MS, where differences in experimental conditions and variations between the employed detection parameters and instruments are smaller. Again, this table represents the high orthogonality

between the detected metabolites using NMR and LC-MS. While the use of LC-MS clearly is of advantage for the measurement of different lipids, NMR provides information on amino acids and small organic acids. Metabolome coverage provided by GC-MS falls in-between the other two platforms, consistently providing information on lipids, sugars, amino acids and organic acids.

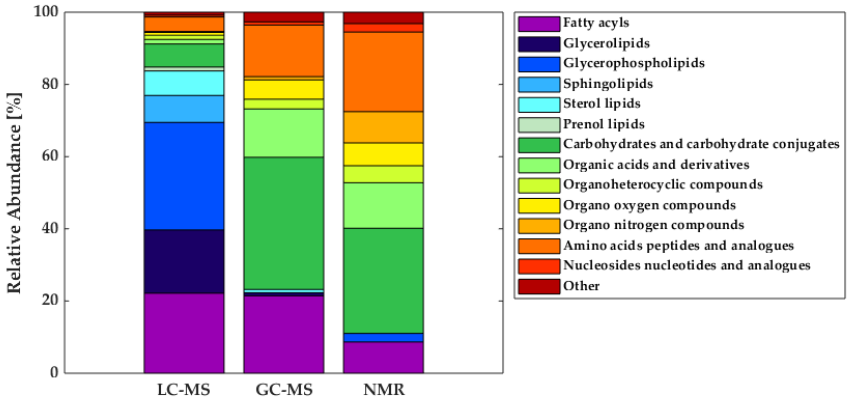


Figure 5.5. Distribution of metabolite classes annotated and/or identified in HM according to technique. Note: GC-MS, gas chromatography-mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; NMR, nuclear magnetic resonance.

Table 5.2. Most frequently reported metabolites (>80% of studies) according to technique.

Metabolite class	LC-MS	GC-MS	NMR
Fatty acyls	Linoleic acid (C18:2) Oleic acid (C18:1) Palmitoleic acid (C16:1)	Oleic acid (C18:1) Palmitic acid (C16:0) Stearic acid (C18:0)	-
Glycerolipids	DG (36:1)	-	-
Glycerophospholipids	LysoPC (16:0)	-	-
Carbohydrates and carbohydrate conjugates	-	Fructose Fucose Ribose	Lactose
Organic acids and derivatives	-	Malic acid Urea	Acetate Citrate Lactate
Organo nitrogen compounds	-	-	Choline
Amino acids, peptides, and analogues	-	Alanine Glutamate Glycine Pyroglutamic acid Serine Valine	Alanine Creatine Glutamate Glutamine Isoleucine Leucine Tyrosine Valine

Note: GC-MS, gas chromatography—mass spectrometry; LC-MS, liquid chromatography—mass spectrometry; NMR, nuclear magnetic resonance; DG, diacylglycerol; PC, phosphatidylcholine.

5.7. Conclusions and future perspectives

In less than a decade, 26 research papers have been published trying to shed light on the complex and dynamic composition of HM and the

feasibility of different options for sample extraction and metabolite detection has been demonstrated. Due to the many factors that influence HM composition, a thorough study design including SOPs for milk extraction, collection, and storage is indispensable for obtaining biologically meaningful results. Multi-platform approaches are encouraged for providing adequate metabolome coverage, as the diversity of compounds contained in HM will not be properly reflected using one single assay. In line with metabolomics workflows tailored to other sample types, the reproducibility of HM metabolomics studies will benefit from the implementation of QA/QC procedures. Automated metabolite annotation and identification with pure chemical standards is warranted and the authors encourage the use of publicly accessible platforms for enabling the exchange of raw data for comparison between studies.

Chapter 6. The effect of holder pasteurization on the lipid and metabolite composition of human milk

6.1. Abstract

Human milk (HM) is the gold standard for newborn nutrition. When own mother's milk is not sufficiently available, pasteurized donor human milk becomes a valuable alternative. In this study we analyzed the impact of Holder pasteurization (HoP) on the metabolic and lipidomic composition of HM. Metabolomic and lipidomic profiles of twelve paired HM samples were analysed before and after HoP by liquid chromatography – mass spectrometry (MS) and gas chromatography-MS. Lipidomic analysis enabled the annotation of 786 features in HM out of which 289 were significantly altered upon pasteurization. Fatty acid analysis showed a significant decrease of 22 out of 29 detectable fatty acids. The observed changes were associated to five metabolic pathways. Lipid ontology enrichment analysis provided insight into the effect of pasteurization on physical and chemical properties, cellular components, and functions. Future research should focus on nutritional and/or developmental consequences of these changes.

6.2. Introduction

Human milk (HM) is the gold standard for infant nutrition and provides the optimum composition of nutritional elements needed for growth and development. Consequently, HM offers numerous short and long-term benefits to the infant-mother dyad [275]. Of note, the survival of extremely low gestational age newborns has consistently improved in the last decades

[276]. In this scenario, early infant nutrition has become a major player in improving clinical outcomes of survivors and based on an impressive array of benefits, HM is the first choice for feeding preterm infants [277].

When own mother's milk (OMM) renders insufficient, pasteurized donor human milk (DHM) rather than preterm infant formula is the preferred alternative for preterm infants [27]. The use of DHM in comparison to the use of formula milk might reduce the incidence of necrotizing enterocolitis [67,277], and although it has been linked to lower rates of weight gain, linear growth, and head growth during hospital admission, no differences in long term growth have been described [67]. The exclusive feeding of pasteurized DHM undoubtedly has an impact on different aspects of preterm biology. Hence, DHM modifies the gut-microbiota composition of preterm infants as compared to the use of OMM [278], although it does not compromise the protection against oxidative stress [279]. Furthermore, preterm infants exclusively receiving DHM exhibited different urinary steroid hormone levels when compared to infants receiving OMM. These findings might be potentially linked to steroid hormone concentrations present in fresh and pasteurized HM [280].

The composition of HM is affected by processing and handling of expressed HM following stringent protocols applied in HM banks involving pasteurization, necessary for minimizing the potential to transmit infectious agents, as well as freezing, long-term storage, and multiple container passages [28]. Low-temperature (62.5 ℃) long-time (30 min) pasteurization known as "Holder" pasteurization (HoP) is routinely employed at HM banks to avoid newborn infection through transmission of pathogens that might be present in DHM as it destroys vegetative forms of bacteria and most viruses including human immunodeficiency virus, herpes, and cytomegalovirus [28]. The use

of softer alternatives (e.g., high-temperature short-time, high pressure processing, ultraviolet and microwave irradiation, and thermosonic processing [32] has been in the spotlight of on-going investigations. However, to our knowledge, none of these techniques has been yet implemented for the routine processing of small volume milk batches (approximately 3 to 6 L) at HM banks.

While total carbohydrates and proteins contents remain relatively stable upon HoP, there is controversy regarding the degree of loss of total fat, as percentages between -6.2 and -25% have been reported in studies replicating current milk bank procedures [28]. Furthermore, several bioactive components, such as enzymes, immunoglobulins, cytokines, microRNAs, and immune cells are inactivated or destroyed [32]. Scientific evidence documenting the effects of HoP on changes of specific lipid and metabolite classes is scant. The present study aims at the characterization of the impact of HoP on the global HM metabolome and lipidome. Metabolomic and lipidomic profiles of twelve paired HM samples collected before and after HoP were analyzed by liquid chromatography – high resolution mass spectrometry (LC-HRMS) and gas chromatography-MS (GC-MS). Results obtained enabled the identification of potentially relevant alterations at the metabolic pathway level. Finally, the potential implications of the observed compositional changes in the properties and functionality of HM are discussed.

6.3. Materials and methods

6.3.1. Human milk samples

The study was conducted in accordance with relevant guidelines and regulations including the Declaration of Helsinki. The Ethics Committee for Biomedical Research of the Health Research Institute La Fe (Valencia, Spain) approved the study protocol (approval number 2014/0247), and mothers gave their written consent to participate. DHM (N=12) samples were provided by healthy HM donors admitted after routine screening and interview to the HM bank at the University and Polytechnic Hospital La Fe (Valencia, Spain). The median gestational age at which milk donors had given birth was 40 (interquartile range, IQR=4) weeks. Milk expression was accomplished with breast milk pumps following the standard operating procedure routinely employed at the hospital and the HM bank. Prior to extraction, removable parts of the breast milk pump and collection bottles were sterilized. In addition, mothers washed their hands with soap and water, and their nipples with water. After extraction, bottles were stored at -20 ºC. Milk bottles from the same mother were defrosted and pooled together to form a batch of approximately 2 L and after gentle shaking, a 1 mL aliquot of it was collected in a dry, 1.5 mL microcentrifuge tube before and after HoP (30 min at 62.5 ºC followed by fast cooling to 4 ºC). The median elapsed time between the first and last expression of a pooled DHM sample was 25 (IQR=36) days. Time of collection in relation to the infants' age was established from this median value and ranged between 21 and 164 days after delivery with a median value of 87 days (IQR=83).

6.3.2. Lipidomic and metabolomic analyses of HM samples

6.3.2.1. Standards and reagents

LC-MS grade acetonitrile (CH_3CN), methanol (CH_3OH), n-hexane (>98%), and isopropanol (IPA), reagent grade methyl tert-butyl ether (MTBE), ammonium acetate (≥98%), formic acid (HCOOH, >98%), methanolic HCl (3N), Supelco 37-component fatty acid methyl ester (FAME) mix, lauric acid-D23 (≥ 98%), and nonadecanoic acid (≥ 98%), were obtained from Merck Life Science S.L.U. (Madrid, Spain). Ultra-pure water (>18.2 MΩ) was generated using a Milli-Q Water Purification System (Merck Millipore, Darmstadt, Germany). (15,15,16,16,17,17,18,18,18-D9) oleic acid-D9 (>99%) was purchased from Avanti Polar Lipids (Alabaster, AL, USA) and prostaglandin $F_{2\alpha}$-D4 (≥98 %, deuterated incorporation ≥99%) was purchased from Cayman Chemical Company (Ann Arbor, MI, USA).

6.3.2.2. HM untargeted metabolomic and lipidomic analyses

HM samples were thawed at room temperature and then heated in a water bath for 10 min at 33 °C. 5 µL of an internal standard (IS) solution containing oleic acid-D9 (80 µM) and prostaglandin F2α-D4 (39 µM) in H_2O were added to 45 µL of HM. A single-phase extraction procedure adding 175 µL of CH_3OH and 175 µL of MTBE [231] to each HM sample followed by a 4 fold dilution of the supernatant with a CH_3OH:MTBE (1:1, v/v) solution was applied. The resulting extract was used for both lipidomic and metabolomic analyses. In addition, a blank extract was prepared following the

steps described for HM samples but replacing HM with water. A pooled quality control (QC) sample was prepared by mixing 20 µL of each HM sample extract.

For untargeted lipidomic and metabolomic analysis, a 1290 Infinity ultra-performance LC (UPLC) system coupled to a 6550 Spectrometer iFunnel quadrupole time-of-flight (QqTOF) MS system from Agilent Technologies (Santa Clara, CA, USA) was used. A detailed description of the parameters used for lipidomic fingerprinting of HM samples can be found elsewhere [103,281]. In short, an Acquity BEH C18 column (50 x 2.1 mm, 1.7 µm) from Waters (Milford, MA, USA) running a binary mobile phase gradient with (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) and (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) as mobile phase components was used. Column and autosampler were kept at 55 and 4 °C, respectively, the flow rate was 0.4 mL min^{-1}, and the injection volume was 2 µL. For untargeted metabolomics screening of HM extracts, a SynergiTM Hydro-RP 80Å LC C18 column (150 x 2 mm, 4 µm) from Phenomenex (Torrance, CA, USA) employing a stepwise gradient with solvent A (H_2O, 0.1% v/v HCOOH) and solvent B (CH_3CN, 0.1% v/v/ HCOOH) as mobile phase components was used as follows: 1% B was held for 2 min followed by a linear gradient from 1 to 80% B in 8 min and from 80 to 98% B in 0.1 min; 98% B were held for 1.9 min before returning to initial conditions in 0.1 min and column equilibration with 1% B during 2.9 min. The flow rate was set to 0.4 mL min^{-1}, column and autosampler to 40 and 4 °C, respectively, and the injection volume was 3 µL.

MS detection was carried out in the ESI+ mode (lipidomics) and the ESI+ and ESI- modes (metabolomics). Full scan MS data was acquired between 70 and 1500 m/z using the following ionization parameters: gas T,

200 °C; drying gas, 14 L min^{-1}; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11 L min^{-1}. Mass reference standards were introduced into the source for automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (background contaminant), 121.050873 (purine), and 922.009798 (HP-0921) m/z as references in ESI+ and 119.036 (purine) and 980.0163 ($[HP^{-}0921^{+}CH_3COOH^{-}H]^{-}$) m/z in ESI-. MS2 data were acquired using data dependent acquisition methods as explained elsewhere [103] using centroid mode at a rate of 5 Hz in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, medium isolation window (~4 amu), MS2 fragmentation with automated selection of five precursor ions per cycle, and an exclusion window of 0.15 min after two consecutive selections of the same precursor. UPLC-QqTOF-MS data acquisition was carried out employing MassHunter Workstation (version B.07.00) from Agilent.

Before UPLC-QqTOF-MS experiments, a system suitability check was carried out by analyzing a 2 μM IS solution. Then, 2 blanks and several QCs (9 and 20 replicates for metabolomics and lipidomic analysis, respectively) were injected at the beginning of the sequence for system conditioning and MS2 data acquisition. HM sample extracts were injected in random order. The QC was injected every six samples and twice at the beginning and end of the batch for assessment and correction of instrumental performance [261]. The blank extract was injected twice at the end of the measurement sequence to identify signals from other than biological origin, and possible carry-over [282].

6.3.2.3. Analysis of fatty acid methyl esters (FAMEs)

The determination of 36 FAMEs was performed employing a previously described GC-MS method [105] with slight modifications. Briefly, a HM aliquot was defrosted on ice and gently shaken to avoid phase separation. Then, 250 µL of HM and 600 µL of n-hexane containing ISs (12 µM lauric acid-D23 and 26 µM nonadecanoic acid) were mixed in a 15 mL test tube equipped with Teflon-lined screw caps. An aliquot of 2 mL of CH_3OH, 2 mL of CH_3OH /HCl (3N), and 1 mL of n-hexane were added and vortexed vigorously. Derivatization was carried out in a water bath at 90 ℃ for 60 min, with occasional additional shaking. After cooling down to room temperature, 2 mL of water were added and shaken vigorously prior to centrifugation (1200 x g for 5 min at 4 ℃). The upper hexane layer containing the extracted derivatives was transferred into GC-MS vials. GC-MS analysis was conducted using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector operating in selected ion monitoring (SIM) mode. Separations were performed using a Zebron™ ZB-WAXplus™ column (30 m x 250 µm i.d., 0.25 µm film thickness, Phenomenex, Torrance, CA, USA). Two microliters of derivatives were injected in split mode with a ratio of 40:1, and the solvent delay time was set to 2.6 min. The initial oven temperature was held at 60 ℃ for 2 min, ramped to 150 ℃ at a rate of 13 ℃ min^{-1} and held for 15 min, and to 240 ℃ at a rate of 2 ℃ min^{-1} and held for 2 min. Helium was used as a carrier gas at a constant flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source were set at 250, 290, and 230 ℃, respectively, and the electron energy was -70 eV. Further measurement parameters used for the studied analytes are summarized in Supplementary Table S5.1. For

quantification, an external calibration line was employed using standard solutions obtained from different volumes of the Supelco 37-component FAME mix after evaporation and reconstitution in n-hexane containing derivatized IS compounds. This procedure was used to remove the 37-component FAME mix solvent (i.e., dichloromethane) and consequently, the most volatile FAME (i.e., FAME of butyric acid) was lost and could no longer be quantified.

The method was validated based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [2] including the following bioanalytical parameters: linearity range, selectivity and specificity, sensitivity, carry-over, accuracy, precision, recovery, and stability. The linear range was selected according to the expected concentrations ranges in HM samples. Calibration curves included a blank without analytes nor IS, a zero calibrator (i.e., blank with IS) and, at least, six standards covering the selected concentration ranges. Accuracy, precision, specificity, and recovery of the method were established from the replicate analysis (N=3) of standards and spiked samples at three concentration levels (low, medium, and high) conducted on three different days to ensure that the extraction of the metabolites was efficient and reproducible. Selectivity and specificity were demonstrated by analyzing multiple blanks. Carry-over was assessed by the analysis of zero-injections after the analysis of high concentrated standards and spiked samples (N=3). Autosampler and processed sample stability were assessed by comparing concentrations observed in a freshly prepared sample and in the same processed sample after 20 h stored in the autosampler (sealed vial, 25 ℃). Analytes' freeze-thaw stability was established by comparing concentrations observed in a sample after three freeze-thaw cycles (-80 ℃) to a freshly prepared sample.

6.3.3. Data processing and statistics

6.3.3.1. Lipidomics and metabolomics data pre-processing

Centroid UPLC-QqTOF-MS raw data were converted to mzXML format employing ProteoWizard [283] (http://proteowizard.sourceforge.net). XCMS software (http://metlin.scripps.edu/xcms/) [284] and CAMERA [285] in R 3.6.1 were employed for the generation of peak tables. Peak table extraction for lipidomics is described elsewhere [103]. The selection of parameters for peak table extraction and alignment for metabolomics was based on the observed variation of RT and m/z values of ISs. The centWave method with the following settings was used for peak detection: m/z range = 70-1500, ppm = 15, peakwidth = (5 and 20), snthr = 6, prefilter = (3, 100). A minimum difference in m/z of 0.01 Da was selected for overlapping peaks. Intensity-weighted m/z values of each feature were calculated using the wMean function. Peak limits used for integration were found through descent on the Mexican hat filtered data. Peak grouping was carried out using the "density" method using mzwid = 0.015 and bw = 5. RT correction was carried out using the "obiwarp" method. After peak grouping, the fillPeaks method with the default parameters was applied to fill missing peak data. Automatic integration was assessed by comparison to manual integration using IS signals. A total of 18401 (lipidomic analysis), 1826 (metabolomic analysis, ESI+) and 893 (metabolomic analysis, ESI-) features were initially detected after peak detection, integration, chromatographic deconvolution, and alignment in HM samples.

Further data processing and statistical analysis were carried out in MATLAB 2019b (Mathworks Inc., Natick, MA, USA) using in-house written

scripts available from the authors, and the PLS Toolbox 8.7 (Eigenvector Research Inc., Wenatchee, USA). Data were annotated by automatic matching of experimental MS2 spectra to publicly available databases (i.e., HMDB, MS-DIAL, and LipidBlast) as described elsewhere [103]. In addition, Lipidex [286] was used for the annotation of lipidomic data (mass error 20 ppm). Enlargement of metabolite annotation was achieved with the Metabolic reaction network-based recursive algorithm (MetDNA) [287].

Intra-batch effect correction was performed using the Quality Control-Support Vector Regression algorithm employing a Radial Basis Function kernel [288] and the LIBSVM library [289] with the following parameters: ε–range = 2 to 5%; γ-range = 1 to 105; C = 90%. Features with RSD(QC) > 20% after QC-SVRC, and those for which the ratio between the median peak area values in QCs and blanks was lower than a threshold value (i.e., 9 and 3 in metabolomic and lipidomic experiments, respectively) were classified as unreliable and removed from further analysis.

6.3.3.2. Network-based analysis

Differences between HM before and after pasteurization on the pathway level were studied employing the "Functional Analysis" tool (version 2.0) available in MetaboAnalyst 5.0 [46] using a mass accuracy of 10 ppm, the mummichog algorithm with a *p*-value cut-off of 0.005, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) Homo Sapiens pathway library [290]. Mummichog analysis was carried out using the m/z and RT values of each feature and the FDR-corrected *p*-values from a t-test to test whether the unknown population means of the two groups were equal, accounting for unequal variances. Using the results of the mummichog algorithm, matched

compounds (i.e., hits) were assigned to their corresponding m/z-RT features of the dataset. Then, KEGG compounds and KEGG glycans of each significantly altered pathway were searched in the matched mummichog compound list and assigned to unique features in the dataset. Fold changes (FC) were calculated as the ratio of medians between groups. The pasteurization effect on the HM lipidome was studied with the web-based ontology enrichment tool for lipidomic analysis: Lipid Ontology analysis - LION (www.lipidontology.com) [47]. As an input, peak areas of annotated features in HM samples were used. The "ranking mode" with log2FC values as local statistic and a two-tailed alternative hypothesis testing was employed. To uncover molecular alterations caused by HoP, an integrative analysis of untargeted lipidomic and metabolomic data was carried out with the network-based Prize-collecting Steiner forest algorithm for integrative analysis of untargeted metabolomics (PIUMet) (http://fraenkel-nsf.csbi.mit.edu/piumet2/) [48]. As an input, m/z features with p-values < 0.01 from a Wilcoxon signed-rank test of peak areas between pre- and post-pasteurized samples were introduced either as metabolomic or lipidomic peaks, also using the -log(*p*-value) as prize to each input data point.

6.3.3.3. Data availability

Peak tables extracted from HM UPLC-QqTOF-MS analysis are accessible via the Mendeley Data repository (https://data.mendeley.com/) under https://doi.org/10.17632/fnzbxmkv83.1 (lipidomics) and https://doi.org/10.17632/jymtst88jm.1 (metabolomics). Supplementary Table S5.2 summarizes FAME concentrations in all samples before and after pasteurization.

6.4. Results and discussion

6.4.1. Compositional and functional alteration of pasteurized vs raw DHM samples

Lipidomic profiles of DHM samples were acquired by UPLC-QqTOF-MS retrieving a total of 7109 metabolic features after peak detection, deconvolution, integration, alignment, within-batch effect correction, and clean-up. In total, 786 (11%) of all features were annotated. The classes with the largest numbers of annotated lipids were triradylglycerols (TGs) and diradylglycerols (DGs) followed by alkenyl-DGs and ceramides (see **Figure 6.1**, left). 3259 features (45.8% of the total) had significantly different mean values (t-test, FDR-adjusted p-value<0.05) and |log2(FC)|>1, including 289 (9%) annotated features using spectral libraries and MetDNA. The sub-class distribution depicted in **Figure 6.1** (right) shows that lipid levels across multiple lipid classes were affected by pasteurization, including DGs and TGs, alkenyl-DGs, and linoleic acids and derivatives, among others. Interestingly, 83% of DGs were affected by pasteurization while only 13% of TGs changed. In addition, it must be highlighted that, although the UPLC-QqTOF-MS method was specifically tailored to the detection of lipids, the use of a single-phase extraction procedure in combination with an untargeted detection allowed the detection of compounds from non-lipid classes such as amino acids, peptides, and analogues, and carbohydrates and carbohydrate conjugates.

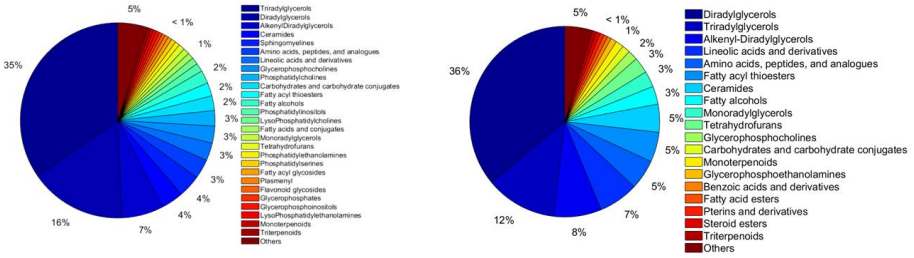


Figure 6.1. Distribution by sub-classes of the metabolites annotated using HMDB/METLIN or LipidBlast spectral libraries and MetDNA detected in HM (left) and with significantly different mean values (t-test, FDR-adjusted *p*-value<0.05) and |log2(FC)| >1 before and after pasteurization (right).

Lipidomic network analysis was employed for a functional interpretation of the observed changes in the lipidomic profiles within relevant networks using the mummichog algorithm [291]. Pathway analysis detected 14 pathways with at least two significantly altered metabolites in DHM samples collected before and after pasteurization. Results summarized in **Table 6.1** indicated a significant impact (*p*-values < 0.05) on metabolites included in the steroid hormone biosynthesis, glycosylphosphatidylinositol (GPI)-anchor biosynthesis, linoleic acid metabolism, biosynthesis of unsaturated fatty acids, and mucin type O-glycan biosynthesis pathways.

Table 6.1. Altered pathways in DHM linked to the pasteurization process.

Pathway name	Hits (total)	Hits (sig.)	*p*-value
Steroid hormone biosynthesis	6	4	0.02
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	4	3	0.03
Linoleic acid metabolism	7	4	0.03
Biosynthesis of unsaturated fatty acids	17	7	0.04
Mucin type O-glycan biosynthesis	2	2	0.04

Note: *p*-values from Fisher's exact t-test; only detected pathways with at least 2 significantly altered features are reported.

Between group comparison of the features' intensities found in the dataset, that corresponded to KEGG compounds and KEGG glycans of each altered pathway (i.e., pre-pasteurization vs post-pasteurization) revealed significantly different mean values (FDR-adjusted *p*-value<0.05) and |log2(FC)| >1. Remarkably, for all features in all altered pathways, higher intensities were detected in DHM samples before pasteurization compared with DHM samples after pasteurization, except for (4Z,7Z,10Z,13Z,16Z,19Z)-Docosahexaenoyl-CoA (KEGG ID C16169, m/z-RT: 1146.3436-7.36) in the biosynthesis of unsaturated fatty acids pathway. The results are summarized in Supplementary Table S5.3 and Supplementary Figures S5.1-S5.5.

Metabolic analysis of DHM samples retrieved a total of 466 and 379 metabolic features in the ESI+ and ESI- mode, respectively, after peak detection, deconvolution, integration, alignment, within-batch effect correction, and clean-up.

Figure 6.2 shows the resulting network for the joint analysis of lipidomic (1033 features with *p*-value < 0.01; Wilcoxon signed-rank test) and

metabolomic (35 features with p-value < 0.01; Wilcoxon signed-rank test) data [48], in which the relation between detected features and hidden metabolites and proteins is displayed. Several features associated to the steroid hormone biosynthesis (i.e., tetrahydrocortisol, tetrahydrocorticosterone, 7a-hydroxydehydroepiandrosterone, tetrahydrodeoxycorticosterone) and the biosynthesis of unsaturated fatty acids (i.e., palmitic acid, eicosapentaenoic acid) were identified, in agreement with results shown in **Table 6.1** and Supplementary Table S5.3. In addition, several metabolites associated to the steroid biosynthesis (i.e., 4,4-dimethylcholesta-8,14,24-trienol, 5-dehydroavenasterol, 7-dehydrodesmosterol, 7-dehydrocholesterol, lathosterol) and purine metabolism (i.e., guanosine, adenine, hypoxanthine, deoxyguanosine) pathways were significantly altered.

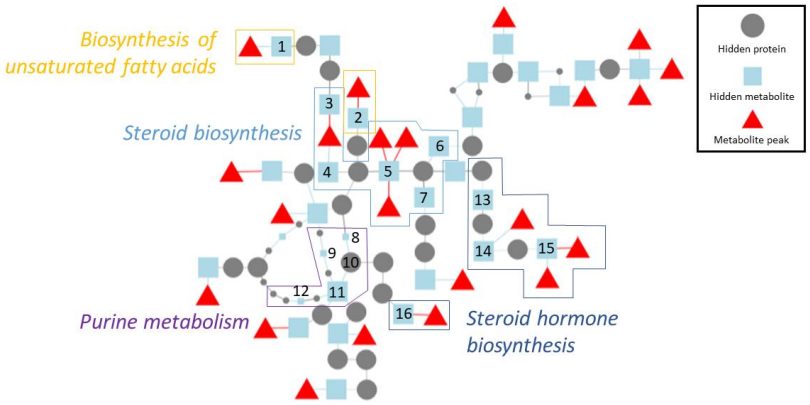


Figure 6.2. PIUMet network showing metabolic pathways altered by pasteurization. Note: Pathways were highlighted if three or more hidden metabolites from the same pathway were interconnected as well as pathways those that were identified earlier by pathway analysis shown in **Table 6.1** (biosynthesis of unsaturated fatty acids). Hidden metabolite / hidden protein #1: Palmitic acid; #2: Eicosapentaenoic acid; #3: 4,4-Dimethylcholesta-8,14,24-trienal; #4: 5-Dehydroavenasterol; #5: 7-Dehydrodesmosterol; #6: 7-Dehydrocholesterol; #7: Lathosterol; #8: Deoxyguanosine; #9: Guanosine; #10: purine nucleoside phosphorylase (PNP) protein; #11: Adenine; #11: Hypoxanthine; #12: 7a-Hydroxydehydroepiandrosterone; #13: Tetrahydrocorticosterone; #14: Tetrahydrocortisol; #15: Tetrahydrodeoxycorticosterone.

A recent study of the impact of the type of nutrition on the urinary metabolome of preterm infants has shown a significant alteration of the steroid hormone biosynthesis pathway associated to the intake of fresh OMM or pasteurized DHM [280]. This result suggested either that those steroid hormones present in HM may have a significant influence in the activity of

the steroid hormone biosynthesis pathway in preterm infants, either directly or via the modification of gut-microbiota crosstalk, or that ingested levels of those compounds differ between pasteurized and fresh HM. In the present study, we focused on the analysis of DHM samples before and after HoP. Previous results found no significant differences of the levels of cortisol and cortisone in HM before and after HoP [292]. However, in the present study, the alteration of structurally closely related compounds included in the steroid and steroid hormone biosynthesis pathways upon pasteurization was identified (see **Table 6.1** and **Figure 6.2**).

The variation of total fat and fatty acid composition upon pasteurization has been studied but there is considerable variation in the reported effects [31]. A recent review on this topic concluded that the total fat content and total fatty acid composition of expressed HM was not generally influenced by storage and handling process as most changes were found below 10% and within the expected random methodological variation [30]. On the contrary, in a recent study with more than four hundred DHM samples, Piemontese et al. [293] confirmed that pasteurization reduced macronutrient composition, especially in terms of lipids and protein. Vincent et al. [294] attributed these differences to the adherence of disrupted milk fat globules to container surfaces and to whether thermal treatment took place in a laboratory environment or in industrial pasteurizers routinely used in HM banks. Thus, Fidler et al. [295] did not find significant changes of fatty acids in HM after HoP, but a trend toward slightly lower percentage values of some fatty acids, including eicosapentaenoic acid was found. This observation was in agreement with the results from PIUMet network analysis presented in this study (see **Figure 6.2**).

Furthermore, alterations of the purine metabolism were detected. A significant increase in nucleotide monophosphates associated to HoP of HM, which could be produced from the break-down of polymeric nucleotides and/or nucleotide adducts was reported [296]. Here, we observed changes in guanosine and deoxyguanosine, which participate as building blocks in the synthesis of oligonucleotides. In addition, in the network-based approach, the purine nucleoside phosphorylase (PNP) protein, which catalyses the phosphorolytic breakdown of the N-glycosidic bond in the beta-(deoxy)ribonucleoside molecules, with the formation of the corresponding free purine bases and pentose-1-phosphate, appeared altered (see **Figure 6.2**).

Finally, no literature reports regarding changes of glycans in pasteurized HM have been found. Although HM oligosaccharides (HMOs) are reported to be unaffected by HoP [297], in this work, mucin type O-glycans appeared altered upon pasteurization. As for HMOs, it has been hypothesized that mucin type O-glycans may have a certain role in mucus barrier function by promoting mutualism with host microbiota [298]. Hence, additional studies on specific glycans are needed.

In the lipid ontology enrichment analysis performed in pursuit of functional alterations of the lipidomic profile of HM after pasteurization, 634 out of 786 (80.66%) annotated features were matched to LION entries. The LION enrichment graph with the four major branches (lipid classification, chemical and physical properties, function, and sub-cellular component) is shown in **Figure 6.3**.

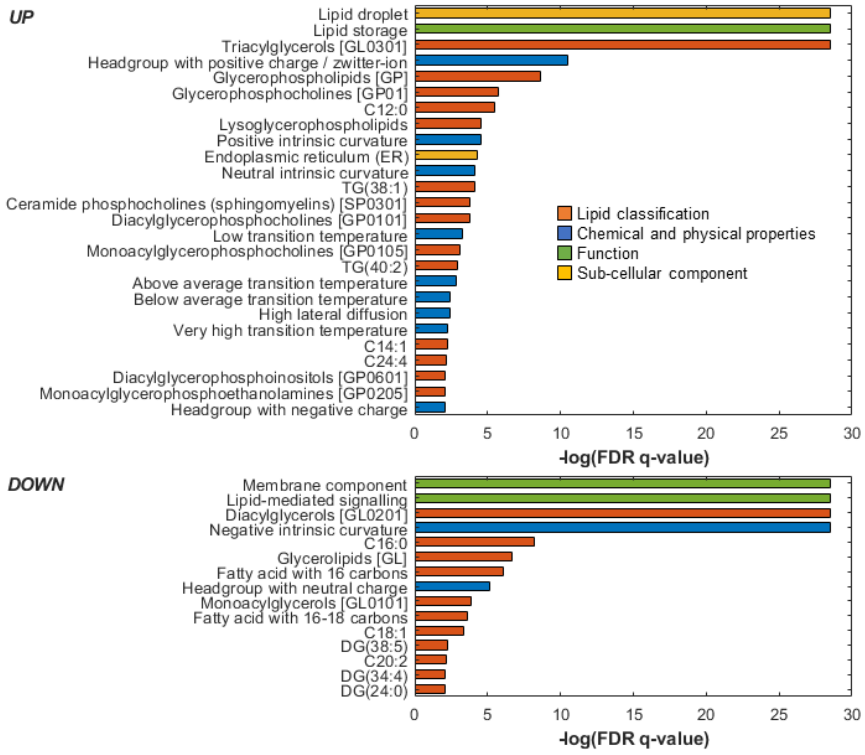


Figure 6.3. Lipid ontology enrichment analysis of the pasteurization process performed in the "ranking mode" (DHM-Post vs DHM-Pre). Only the 40 most enriched LION-term have been represented.

The terms "lipid droplet" and "lipid storage", belonging to the category "cellular component" and "function", respectively, were enriched in pasteurized DHM. Inspection of the lipid species responsible of these LION-terms revealed the presence of TGs. It has been described that the heating process very likely causes the disruption of the fat globule membranes. In fact, a significant decreased in fat globule size and, hence, an increase in the overall

surface area upon pasteurization have been previously reported [299]. This surface increase of the TG/water interface jointly with potential changes in the interface composition might play an important role with respect to chemical and enzymatic reactions that take place at the interface. In this work, the terms "membrane component" and "lipid-mediated signaling", belonging to the ontology root "function", were down-regulated by pasteurization. The first term is associated with lipids primarily regarded as structural components of lipid bilayers (i.e., DGs, glycerophosphocholines, phosphosphingolipids, and lyso-glycerophospholipids (GPs)) indicating a decay of membrane lipids during pasteurization. In this study, a total of 258 identifiers were ascribed to this LION-term, distributed as DGs (59%), GPs (24%), sphingomyelins (12%), and lyso-GPs (5%). The term "lipid-mediated signaling", associated with lipids implicated in signaling, such as DGs, monoradylglycerols, and lyso-GPs, was also down-regulated because of processing. Changes in the oxylipin composition during HoP have been previously observed [300]. This potential functional alteration should be addressed in future studies, as it could potentially be important for cellular processes, and especially relevant for the brain development of preterm infants. Lipids related to the term "negative intrinsic curvature" were downregulated in pasteurized DHM. LION assumes curvature to be predominantly head-group-dependent and neglects fatty acid composition. In this work, 184 identifiers, distributed in DGs (80%), ceramides (15%) and GPs (4%), were ascribed to this term indicating a change in physico-chemical properties of HM upon pasteurization.

Lipid ontology enrichment analysis underpinned the alteration of different lipid classes upon pasteurization as discussed earlier (see also **Figure 6.1**). In addition, it allowed to draw conclusions regarding the consequences of those changes that were associated to the physical and chemical properties (i.e., negative intrinsic curvature), cellular components (i.e., lipid droplet) and

function (i.e., lipid storage, membrane component, and lipid-mediated signaling) of altered lipids.

6.4.2. Fatty acid analysis

A method for the quantification of derivatized free fatty acids after acidic hydrolysis of lipids in HM was developed and validated following the FDA guidelines for bioanalytical method validation. Supplementary Table S5.4 summarizes the employed concentration intervals, which were selected considering the wide expected inter- and intra-individual variability. Intra and inter-day accuracy and precision of the method varied between 80% and 121%, and 1.0% and 20%, respectively, in standards and between 80% and 122%, and 1.0% and 14%, respectively, in spiked samples. Selectivity was confirmed and no carry-over was detected.

A total of 29 fatty acids from the 36 fatty acids included in the analytical method were found at higher levels than the lower limit of quantification and were successfully quantified in HM samples. Mean concentrations of five fatty acids were unaffected by pasteurization, two fatty acids showed an increase, and 22 fatty acids showed a statistically significant decrease (median decrease 10% (14% IQR)) (see **Table 6.2**).

Table 6.2. Individual and total content of fatty acids in DHM samples before (DHM-Pre) and after (DHM-Post) pasteurization.

Fatty acid	Median (IQR) (mM)		*p*-value
	DHM-Pre	DHM-Post	
Hexanoate (C6:0)	56 (20 - 95)	35 (18 - 75)	0.0002
Octanoate (C8:0)	41 (19 - 55)	29 (11 - 47)	0.0005
Decanoate (C10:0)	42 (22 - 53)	35 (15 - 45)	0.0002
Undecanoate (C11)	< LOQ	< LOQ	-
Laurate (C12:0)	14 (10 - 18)	10 (9 - 16)	0.02
Tridecanoate	< LOQ	< LOQ	-
Myristate (C14:0)	4 (3 - 4)	3 (2 - 4)	0.013
Myristoleate (C14:1)	0.10 (0.06 - 0.15)	0.08 (0.05 - 0.14)	0.0005
Pentadecanoate (C15:0)	0.10 (0.08 - 0.14)	0.09 (0.06 - 0.13)	0.0007
cis-10-Pentadecenoic (C15:1)	< LOQ	< LOQ	-
Palmitate (C16:0)	9 (7 – 10)	8 (5 - 10)	0.0005
Palmitoleate (C16:1)	0.9 (0.6 - 1.2)	0.9 (0.5 - 1.0)	0.0012
Heptadecanoate (C17:0)	0.3 (0.2 - 0.3)	0.3 (0.2 - 0.3)	0.003
cis-10-Heptadecenoic (C17:1)	0.2 (0.15 - 0.3)	0.2 (0.14 - 0.2)	0.0002
Stearate (C18:0)	3 (2 - 5)	3 (3 - 4)	0.0002
Oleic (C18:1n9c)	30 (26 - 35)	26 (18 - 30)	0.002
Elaidic (C18:1n9t)	0.8 (0.7 – 1.0)	0.8 (0.5 - 0.9)	0.013
Linoleic (C18:2n6c)	6 (5 - 8)	5 (4 - 7)	0.0002
Linolelaidic (C18:2n6t)	5 (1.1 - 8)	4 (0.8 - 7)	0.0007
gamma-Linolenic (C18:3n6)	0.2 (0.15 - 0.3)	0.2 (0.15 - 0.2)	0.0012
Linolenic (C18:3n3)	0.4 (0.4 - 0.5)	0.4 (0.3 - 0.4)	0.0002
Eicosanoic (C20:0)	0.3 (0.3 - 0.3)	0.3 (0.3 - 0.3)	0.02
cis-11-Eicosenoic (C20:1)	0.2 (0.13 - 0.2)	0.15 (0.13 - 0.2)	0.02
cis-11,14-Eicosadienoic (C20:2)	0.09 (0.05 - 0.2)	0.07 (0.04 - 0.2)	0.02
cis-8,11,14-Eicosatrienoic (C20:3n6)	0.4 (0.3 - 0.5)	0.4 (0.3 - 0.4)	0.05
Arachidonic (C20:4n6)	0.3 (0.3 - 0.4)	0.3 (0.3 - 0.4)	0.002
Heneicosanoate (C21:0)	< LOQ	< LOQ	-
cis-11,14,17-Eicosatrienoic (C20:3n3)	0.15 (0.11 - 0.2)	0.15 (0.11 - 0.2)	0.13

Table 6.2 (continuation). Individual and total content of fatty acids in DHM samples before (DHM-Pre) and after (DHM-Post) pasteurization.

Fatty acid	Median (IQR) (mM)		*p*-value
	DHM-Pre	DHM-Post	
cis-5,8,11,14,17-Eicosapentaenoic (C20:5n3)	< LOQ	< LOQ	-
Docosanoate (C22:0)	0.5 (0.4 - 0.5)	0.5 (0.4 - 0.5)	0.02
Erucic acid (C22:1)	0.006 (0.005 - 0.13)	0.007 (0.004 - 0.13)	0.13
cis-13,16-Docosadienoic (C22:2)	0.3 (0.2 – 0.3)	0.3 (0.2 - 0.3)	0.5
Tricosanoate (C23:0)	< LOQ	< LOQ	-
cis-4,7,10,13,16,19-Docosahexaenoic (C22:6n3)	0.5 (0.4 - 0.7)	0.5 (0.4 - 0.6)	0.008
Lignocerate (C24:0)	0.2 (0.2 - 0.3)	0.2 (0.2 - 0.3)	0.05
Nervonic acid (C24:1)	< LOQ	< LOQ	-
TOTAL	230 (120 – 290)	169 (104 – 252)	0.0002

Besides, the relative decrease in the concentration was higher for saturated fatty acids (SFAs) (25%), followed by polyunsaturated fatty acids (PUFAs) (18%), long-chain fatty acids (LCFAs) (15%), and monounsaturated fatty acids (MUFAs) (12%) (see **Figure 6.4**).

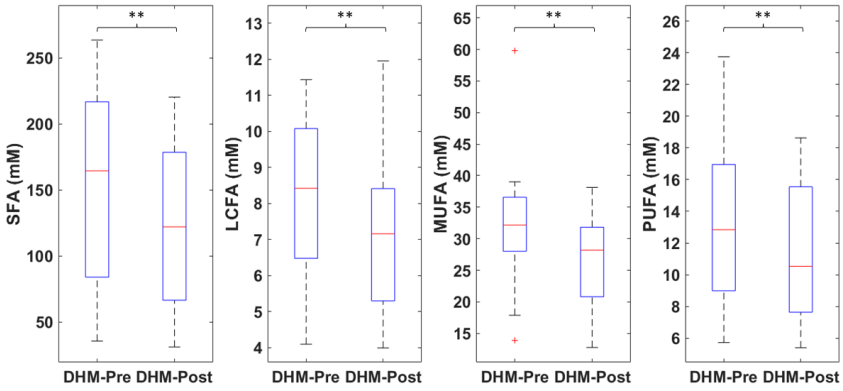


Figure 6.4. Total saturated fatty acids (SFA), long chain fatty acids (LCFA), monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA) of DHM-samples before (DHM-Pre) and after (DHM-Post) pasteurization. ***p*-value <0.01, one-tailed Wilcoxon signed-rank test.

The observed reduction in the fatty acids concentrations was in agreement with previous results [30]. The small decrease in concentration found in this work might be, at least partially, responsible for the controversy in literature reports [31] on the effect of pasteurization on fatty acid concentration in HM and underlines the importance of method validation for enabling the detection of subtle changes between groups.

6.5. Conclusion

This study provides the first comprehensive assessment of the impact of pasteurization on the HM lipidome and metabolome. Results identified affected metabolites that were related to steroids (i.e., steroid biosynthesis and

steroid hormone biosynthesis) as well as fatty acids (i.e., biosynthesis of unsaturated fatty acids and linoleic acid metabolism) metabolic pathways. Earlier studies solely focusing on cortisone and cortisol did not find alterations due to HoP. In this work, however, the levels of structurally related metabolites were affected by the thermal treatment. The present work furthermore demonstrates that HoP has an impact not only on the composition, but also on the physico-chemical properties, cellular components, and the functionality of lipids. Finally, the concentrations of the 76% of the analyzed fatty acids were altered after pasteurization, with a median decrease in the relative concentration of 10%. Results obtained encourage further research into the analysis of the biological relevance and impact of the observed changes in composition and functionality of HM components.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.132581.

Section III. Novel analytical tools for the characterization of the molecular composition of HM-EVs

Chapter 7. Isolation and lipidomic screening of human milk extracellular vesicles

7.1. Abstract

Extracellular vesicles (EVs) are secreted by cells and can be found in biological fluids (e.g., blood, saliva, urine, cerebrospinal fluid, and milk). EVs isolation needs to be optimized carefully depending on the type of biofluid and tissue. Human milk (HM) is known to be a rich source of EVs, and they are thought to be partially responsible for benefits associated to breastfeeding. Here, a workflow for the isolation and lipidomic analysis of HM-EVs is described. The procedure encompasses initial steps such as sample collection and storage, a detailed description for HM-EV isolation by multi-stage ultracentrifugation, metabolite extraction, and analysis by liquid chromatography coupled to mass spectrometry, as well as data analysis and curation.

7.2. Introduction

Eukaryotic and prokaryotic cells release a variety of nano- and micron-sized membrane-containing vesicles into their extracellular environment, which are collectively referred to as extracellular vesicles (EVs). EVs can be harvested from cell culture supernatants and from all body fluids including plasma, saliva, urine, cerebrospinal fluid, and human milk (HM). HM is known to be a rich source of EVs, with early milk containing a greater EV concentration compared to mature milk.

The physiological purpose of generating EVs remains largely unknown and questions surrounding the function of EVs are mostly focused on understanding the fate of their constituents and the phenotypic and molecular alterations that they induce in recipient cells. The effects of EVs on recipient cells can vary due to the expression of different cell surface receptors, resulting in an array of possible biological functions including the induction of cell survival, apoptosis, and immunomodulation. In addition, recent studies indicate a functional, targeted, mechanism-driven accumulation of specific cellular components such as RNAs, small RNAs, nucleic acids, lipids, proteins, and metabolites in EVs, suggesting that they have a role in regulating intercellular communication [301]. Furthermore, encapsulation in EVs confers protection against enzymatic and nonenzymatic degradation of cargos and provides a pathway for cellular uptake of cargos by endocytosis of EVs [24,302].

The composition of HM-EVs is still an open question. The richness of microRNAs (miRNAs) within HM-EVs was first discovered and the dynamics during lactation stages was surveyed [303,304]. Proteins have been studied in HM-EVs [305,306] showing an enrichment of pathways involved in early-life immunity, as well as intestinal cell proliferation and migration. Although metabolites are known to be part of the EV cargo [307–309], literature reports providing insight into the metabolite cargo or lipid composition of HM-EVs are scarce. Hence, we developed a pipeline for the isolation and Liquid Chromatography – Mass Spectrometry (LC-MS) based lipidomic screening of HM-EVs [310].

7.3. Materials

Prepare all solutions using ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany) using analytical grade reagents. Prepare all reagents at room temperature.

1. Phosphate buffered saline (PBS) solution: dissolve one commercially available phosphate buffered saline tablet (e.g., Fisher BioReagentsTM, Ref: BP2944-100) in 200 mL of ultrapure water for a 10 mM phosphate buffer, 2.7 mM potassium chloride and 137 mM sodium chloride solution, pH 7.4, at 25 ℃ (see Note 1). Filter this solution with a 0.40 μm syringe filter. Store at 4 ℃.

2. Internal standard (IS) mixture: prepare individual 5 mM stock solutions of internal standards by weighing e.g., 1.46 mg of oleic acid-D9 and 1.49 mg of prostaglandin $F_{2\alpha}$-D_4 and dissolving them in 1 mL of ultrapure water. Put 976 μL of ultrapure water into a microcentrifuge tube. Add 16 and 7.8 μL of oleic acid-D9 and prostaglandin $F_{2\alpha}$-D4 5 mM stock solutions, respectively (resulting concentrations: 80 and 39 μM, respectively) (see Note 2). Aliquot to avoid freeze-thaw cycles and store at -20 ℃.

3. Methanol (MeOH). Store at room temperature.

4. Methyl tert-butyl ether (MTBE). Store at room temperature.

5. Solution A (1 L of (5:1:4) Isopropanol (IPA):MeOH:H_2O, 5mM CH_3COONH_4, 0.1% formic acid (FA)): add 500 mL IPA and 100 mL MeOH into a 1 L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 20 mL of ultrapure

water and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with ultrapure water. Store at 4 ℃.

6. Solution B (1 L of (99:1) IPA:H_2O, 5mM CH_3COONH_4, 0.1% FA): add 900 mL IPA into a 1 L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 9 mL of ultrapure water and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with IPA. Store at 4 ℃.

7. Solution (1:1) A:B: mix 10 mL of solution A with 10 mL of solution B. Store at 4 ℃.

8. For the performance of untargeted screening experiments, an LC (preferably: ultra performance LC)-high resolution MS instrument and a reversed phase chromatographic column should be employed. The following protocol was developed using a 1290 Infinity HPLC system from Agilent Technologies (CA, USA) equipped with a UPLC BEH C18 column (50 x 2.1 mm, 1.7 µm, Waters, Wexford, Ireland) coupled to an Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS.

7.4. Methods

Carry out all procedures at room temperature unless otherwise specified.

7.4.1. HM sample collection and storage (see Note 3)

1. Clean the hands with water and soap during at least 15 seconds followed by drying with a clean towel.

2. Sterilize hands with hand sanitizer.

3. Clean and sterilize the milk pump (see Note 4).

4. Clean the skin area that gets into contact with the milk pump with water and dry it with sterile gauzes (see Note 5).

5. Extract the milk with the milk pump into clean sterile bottles following the manufacturer's instructions (see Note 6).

6. Manual shake gently during 30 s for sample homogenization.

7. Put the sample into 50 mL plastic Falcon tubes using a 25 mL plastic pipette (see **Figure 7.1A**).

8. Centrifuge at 3000 x g for 10 min at 22 ºC for cream, cell, and platelet removal [311,312].

9. Remove and discard the cream layer with a spatula (see Note 7 and **Figure 7.1B**).

10. Transfer the liquid into a new 50 mL plastic Falcon tube.

11. Repeat steps 8 and 9 (see **Figure 7.1C**) and store milk at -80 ºC until further processing (see **Figure 7.1D**).

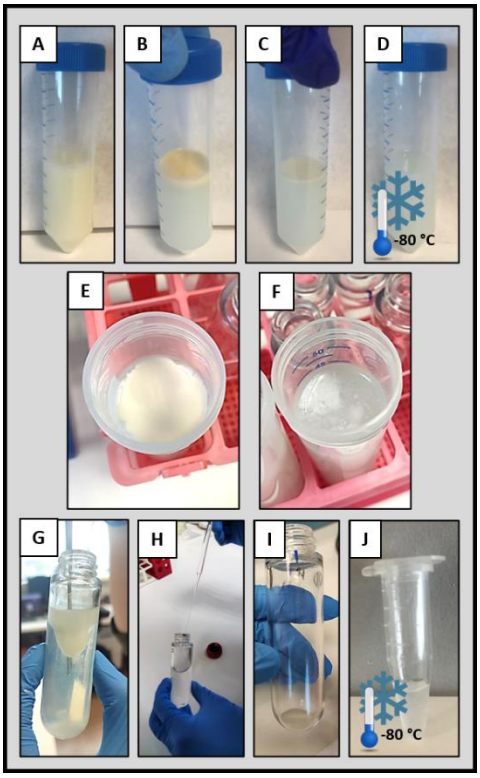


Figure 7.1. Isolation of extracellular vesicles (EVs) from human milk (HM). (a) Raw HM sample. (b and c) HM sample after consecutive centrifugation steps with fat layer on the top. (d) Partially defatted HM sample for storage t -80℃ prior to EV isolation. (e) HM sample after first centrifugation step with remaining fat layer on the top. (f) Skimmed HM sample. (g) Supernatant aspiration after the first ultracentrifugation (10,000 rpm, 1 h, 4℃) (protein pellet is discarded). (h) Supernatant removal after the second ultracentrifugation (30,000 rpm, 2 h, 4℃) with the pellet containing EVs at the bottom. (i) Pellet containing EVs. (j) Reconstituted EVs in phosphate-buffered saline for storage at -80℃ until use.

7.4.2. Isolation of EVs from HM

1. Thaw HM samples at 4 ℃ overnight by putting them in the refrigerator.

2. Centrifuge at 3000 x g for 10 min at 4 ℃ to remove the remaining milk fat and milk fat globules.

3. Remove the upper fat layer with a spatula (see Note 7) and discard (see **Figure 7.1E** and **7.1F**).

4. Transfer the liquid into a 25 mL polycarbonate bottle appropriate for the ultracentrifugation.

5. Continue filling the ultracentrifuge tube with PBS until an estimated volume of 25 mL is reached and make sure all tubes have the same weight (including the caps) (see Note 8). This step is crucial as any ultracentrifuge can easily become unbalanced if equal masses are not located opposite to each other in the rotor.

6. Ultracentrifuge at 10000 rpm for 1 h at 4 ℃ to pellet cellular debris, large size vesicles and protein aggregates using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti rotor (Indianapolis, United States) or similar.

7. Collect the supernatant with a 25 mL plastic pipette or a syringe with a needle (see Note 9) and transfer it to a new 25 mL ultracentrifuge tube (see **Figure 7.1G**).

8. Repeat steps 5 and 6.

9. Collect the supernatant into a different 25 mL ultracentrifuge tube filtering it with a 0.45 μm syringe filter (see Note 10).

10. Repeat step 5.

11. Ultracentrifuge at 30000 rpm for 2 h at 4 ℃ to pellet HM EVs using the same centrifuge rotor as in step 6.

12. Slowly aspirate and discard the supernatant using e.g., a pipette (see **Figure 7.1H**).

13. Wash and reconstitute the remaining pellet with 25 mL of PBS (see Note 11).

14. Repeat steps 5, 11-13 for a second ultracentrifugation.

15. Repeat steps 5 and 11 for a third ultracentrifugation.

16. Slowly aspirate the PBS supernatant and save it to be later used as a blank. Store at -80 ℃.

17. Reconstitute the pellet containing the isolated EVs (see **Figure 7.1I**) with 200 μL PBS (see Note 12).

18. Aliquot the isolated EVs in microtubes for quality control (QC) assays such as, bicinchoninic acid assay (BCA) for protein quantification, nanoparticle tracking analysis (NTA) for vesicle concentration and size determination, etc. and for LC-MS analysis. Store at -80 ℃ (see **Figure 7.1J**).

7.4.3. Extraction of the HM-EVs lipid fraction

Lipids and other polar metabolites were extracted from HM-EVs using a single-phase extraction procedure [103,280,313].

1. Thaw the isolated HM-EV suspension in PBS obtained in section 3.2 on ice and vortex for sample homogenization.

2. Mix 45 µL of isolated EVs with 5 µL of IS mixture.

3. Sonicate for 2 min.

4. Add 175 µL of MeOH followed by 175 µL of MTBE (see Note 13).

5. Vortex for 30 s for protein precipitation and compound extraction.

6. Sonicate for 2 min to assist the release of metabolites from EVs during extraction.

7. Centrifugate at 4000 x g for 15 min at 4 ℃.

8. Transfer 100 µL of supernatant containing the extracted lipids and metabolites to a different microtube.

9. Dry at 35 ℃ using a centrifugal vacuum concentrator (see Note 14).

10. Reconstitute in 100 µL of solution (1:1) A:B.

11. Prepare a pooled QC sample by mixing 5 μL of each reconstituted sample extract.

12. Prepare a calibration blank by repeating steps 1-10, replacing the initial 45 μL of sample with 45 μL of ultrapure water.

13. Prepare a procedural blank by repeating steps 1-10, replacing the initial 45 μL of sample with 45 μL of aspirated PBS supernatant from step 18 in section 3.2.

14. Analyse by LC-MS or, alternatively, store at -80 ℃ for further analysis.

7.4.4. LC-MS lipidomics method (see Note 15)

1. Perform a system suitability check of the MS (e.g., resolution, accuracy, and sensitivity) and UPLC performances (e.g., retention time (RT) of standards, resolution, lack of contamination of the analytical system) using blanks, standard mixtures, and QC samples (see Note 15).

2. Thaw the samples and place them in the autosampler.

3. Equilibrate the system by injecting the QC sample repeatedly. MS/MS data for peak annotation is typically acquired at the beginning or end of the batch [103] (see Note 16).

4. Run sample sequence including QC samples, study samples, and blanks, acquiring data using a suitable ionization interface (e.g., positive and/or negative electrospray ionization) (see Notes 17 and 18).

7.4.5. Data processing and analysis

1. Perform an initial data quality check through the manual integration of IS and several (e.g., 5 to 10) endogenous compounds with different intensities and RTs, that are expected to be detected in the sample under study. Assess the stability of peak areas, m/z accuracy, RT and chromatographic parameters (e.g. peak width, resolution) throughout the analytical sequence.

2. Convert raw data into mzXML format using e.g., ProteoWizard (http://proteowizard.sourceforge.net/) (optional) and generate a peak table (see Notes 19 and 20).

3. Annotate lipids using MS/MS information and available spectral and/or in-house LC-MS libraries (see Note 21 and **Figures 2A-C**).

4. Identify and correct within-batch effect (see Note 22 and **Figure 7.2D**).

5. Perform data-clean up: (i) filtering of features found in blanks and (ii) filtering of features with high relative standard deviation (RSD) detected in QC samples (see Note 23).

6. Check data quality (see Note 24 and **Figure 7.2E**).

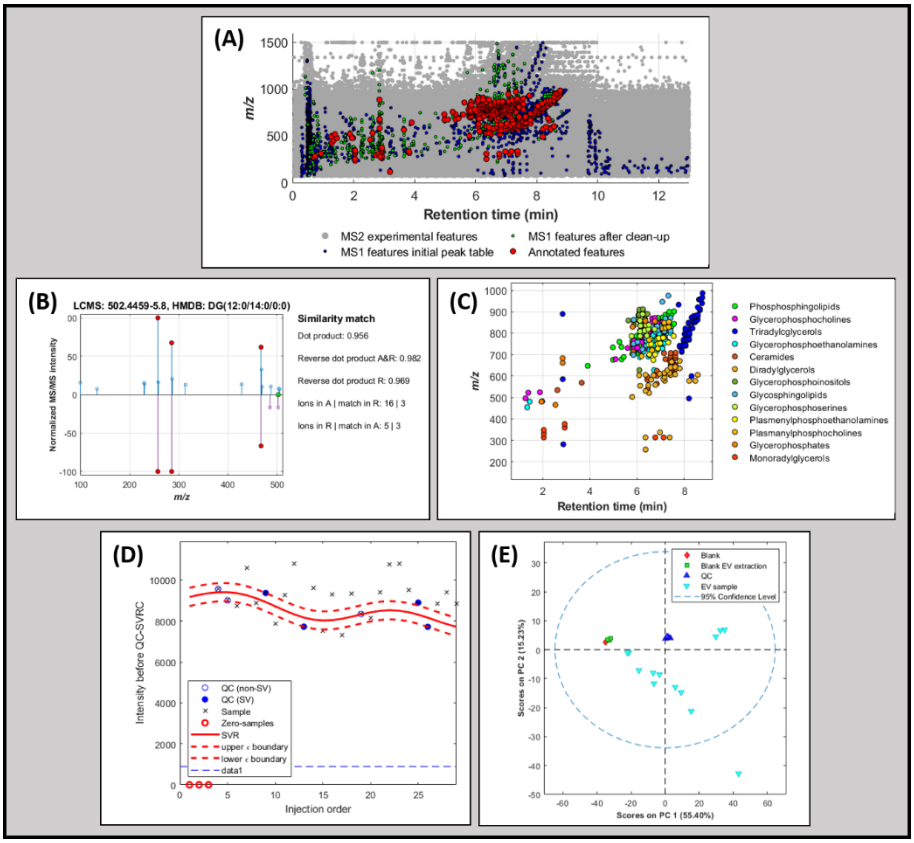


Figure 7.2. LC-MS lipidomic data processing. (a) Distribution of MS1 (blue dots) and MS2 (gray dots) experimental features measured in extracellular vesicles (EVs) from human milk (HM) samples in the m/z-retention time space and MS1 experimental features and annotated features after applying cleanup steps (green and red dots, respectively). (b) Similarity match of a feature (m/z-retention time: 502.4489-5.8) annotated as DG (12: 0/14:0/0:0) from the HMDB database. MS2 experimental spectrum (up). MS2 database spectrum (down). (c) Distribution in the m/z-retention time space of annotated features by sub-classes in EVs from HM samples. Only sub-classes containing at least five features are represented. (d) Intensity of a selected feature as a

function of the injection order before intra-batch effect correction with the quality control-supported vector regression correction (QC-SVRC) approach. (e) Score plot of the principal component analysis model using the preprocessed data set. Note: DG diradylglycerol, QC quality control.

7.5. Notes

1. Prepare this solution fresh each time.

2. It is convenient to immerse the pipette tip in the water when adding small volumes.

3. HM samples are usually collected by lactating mothers and not by the hospital staff. Hence, the staff must provide detailed instructions to the mothers to avoid sampling bias and procedural differences between the different HM samples collected within a study. We recommend establishing a Standard Operation Procedure for the collection of milk samples and their processing to avoid experimental bias between samples. Details regarding the sample collection procedure need to be indicated including e.g., extraction of foremilk, hindmilk, or full expression of one or both breasts; time of the day; time since last feeding; extraction method (i.e., manual expression, milk pump).

4. We recommend the use of electric milk pumps as indicated in this protocol, however if participants prefer manual expression, this could also be accommodated within the protocol. Milk pumps can be sterilized in the microwave by using dedicated plastic bags or in a

water bath (15 min at 100 ℃). All parts need to be dried with sterile gauzes. The milk pump should be cleaned with hot water and soap immediately after every use.

5. No ointments should be applied to the skin before HM extraction. If they had been applied, clean the skin carefully with water and soap.

6. Take notes of date, time, and extracted HM volume. When HM samples are manipulated, gloves must be used in order to avoid sample contamination.

7. Remove the upper fat layer with a spatula and aspirate the supernatant with a pipette. Be careful not to aspirate the remaining fat.

8. If the tubes do not have the same weight, compensate by adding PBS.

9. Pellet should not be aspirated.

10. More than one filter may be needed due to obturation.

11. Reconstitute the pellet in 1 mL PBS and add the remaining 24 mL when the pellet is dissolved.

12. Do it gently to avoid foam formation.

13. The order of the solvents' addition is important.

14. This step usually takes 3-4 hours.

15. Basic considerations for untargeted, LC-MS-based lipidomics and metabolomics studies should be implemented throughout the pipeline. For further information, the reader is referred to [261].

16. A set of approximately 5 to 10 QCs should be injected at the beginning of each batch for system conditioning and MS/MS data acquisition. MS/MS spectra need to be acquired for subsequent annotation purposes as described elsewhere [103]. Depending on the employed MS system, parameters need to be adjusted. For this example, a rate of 5 spectra/s in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, an automated selection of five precursor ions per cycle and an exclusion window of 0.15 min after two consecutive selections of the same precursor was used.

17. Samples should be injected in randomized order and a QC sample should be intercalated every 5 to 10 samples for subsequent correction of the batch effect [261,288,314]. In our pipeline we use a binary mobile phase gradient starting at 98% of solution A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) during 0.5 min followed by a linear gradient from 2 to 20% of solution B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) during 3.5 min and from 20 to 95% v/v of solution B in 4 min; 95% v/v of solution B are maintained during 1 min; return to initial conditions in 0.25 min and maintain for a total run time of 13 min. Column and autosampler are kept at 55 and 4 °C, respectively, an injection volume of 2 µL is used, and the flow rate is set to 400 µL/min. Full scan MS data in the range between 70 and 1500 m/z are acquired at a scan frequency of 5 Hz using the following parameters: gas T, 200 °C; drying gas, 14 L/min;

nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11 / min. Mass reference standards are introduced into the source for on-the-fly automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (phtalic anhydride), 121.050873 (purine), and 922.009798 (HP-0921) m/z in ESI+, and 119.036 (purine) and 980.0163 (HP-0921, $[M^{-}H^{+}CH_3COOH]^{-}$) m/z in ESI-, as references. ESI+ and ESI- analysis were carried out in independent batches and between them, the instrument was cleaned and calibrated according to manufacturer guidelines.

18. Blank injections can affect the performance of the LC-MS system and hence, the position of the blanks within the analytical sequence should be chosen carefully [282].

19. We use MassHunter Workstation (version B.07.00) from Agilent for UPLC-TOFMS data acquisition and manual integration. In our workflow peak detection, integration, deconvolution, alignment and pseudospectra identification are carried out using XCMS [315] and CAMERA [285] in R 3.6.1. See reference [310] for an example of XCMS and CAMERA settings. Employed parameters depend on the LC-MS system and the method's performance and might need to be optimized in advance using e.g. Isotopologue Parameter Optimization (IPO) [316]. Alternatively, peak tables can be generated using vendor and other available software such as MZmine2 or MS-DIAL. We recommend comparing manual vs. automatic integration for IS and endogenous metabolites considered in step 1 to assess the accuracy of the automated integration.

20. If ESI+ and ESI- data were acquired in separate injections, perform data clean-up and filtering independently.

21. Metabolite annotation (Level ID: 2, putatively annotated compounds without matching to data for chemical standards acquired under the same experimental conditions) can be performed by matching experimentally acquired MS/MS spectra with the experimental MS/MS databases (e.g., HMDB, METLIN) in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [317]. Metabolite annotation can also exploit in-silico databases (e.g., LipidBlast [318]). Annotation and subsequent correction and clean-up steps can be programmed in different programming languages for data science such as MATLAB (Mathworks Inc., Natick, MA, USA), R or Python. Besides, metabolite annotation can also be performed using vendor software and other widely used tools such as MZMine2, MS-DIAL, GNPS (Global Natural Products Social Molecular Networking) or LipidDex [103] [286].

22. Features detected in QCs samples can efficiently be used to monitor the instrument performance and correct within-batch effects [261]. We developed a non-parametric approach based on QC-Supported vector regression correction (QC-SVRC) approach employing a Radial Basis Function kernel for within batch effect removal [288,314]. The selection of the tolerance threshold (ε), the penalty term applied to margin slack values (C) and the kernel width (γ) can be carried out using a grid search. QC-SVRC is robust to hyperparameter selection and a pre-selection of C and optimization of ε and γ using a grid search by leave-one-out cross validation, using

the root mean square error of cross validation (RMSECV) as target function allows a significant reduction in computation time. C can be selected for each LC-MS feature as the median value of the intensities observed in QC replicates. The ε search range is chosen based on the expected instrumental precision (e.g., 5-10% of the signal intensity). The γ search interval can be set to [1, 105].

23. It is important to remove variables that are likely originating from background contaminants or cross-contamination, as well as those from unstable features to enhance performance of subsequent data analysis steps. We typically remove features with more than 20% missing values in QCs, those with RSD(QC)>20% after within-batch effect correction, and those for which the ratio between the median peak area values in QCs and blanks is lower than 6.

24. Different qualitative and quantitative tools are available for checking data quality [319], including Principal Component Analysis, which is the most widely employed tool for this purpose. We recommend comparing data quality of raw data and pre-processed data.

Chapter 8. Normalization approaches for extracellular vesicle-derived lipidomic fingerprints – A human milk case study

8.1. Abstract

Human milk (HM) extracellular vesicles (EVs) are nano-sized, cell-derived particles sheathed in a lipid bilayer that encase specific cargo for delivery from mother to infant. The aim of this study was to expand our understanding of the lipidomic profile of HM-EVs, with a specific focus on the data normalization process. To this end, a set of EVs with varying characteristics was simulated and the effect of different normalization strategies on these populations was evaluated. Then, HM samples from mothers of preterm (N = 5) and term infants (N = 5) and a pool of donor human milk from 20 mothers (before and after pasteurization) were collected. EVs were isolated by multi-stage ultracentrifugation and characterized in terms of total protein content, total particle count and size, surface tetraspanin profile and protein markers, and morphology. Lipidomic analysis after single-phase extraction was performed by liquid chromatography-mass spectrometry (LC-MS). Lipid annotation was based on MS/MS information using HMDB and LipidBlast databases. Different data normalization strategies were adopted, compared, and discussed. Results obtained show that the selection of the normalization approach should consider the specific study aims, as well as the purity and homogeneity of size distribution of EV isolates. Our findings exemplify the need for guidance with respect to data processing in untargeted, LC-MS-based lipidomics studies of EVs to fully exploit the potential of this valuable information in this area of research.

8.2. Introduction

Extracellular vesicles (EVs) are nano-sized, cell-derived particles sheathed in a lipid bilayer that are secreted from all types of mammalian cells into the extracellular space. In 2007, Admyre, C. et al. [320] isolated for the first time EVs from colostrum and mature human milk (HM) and hypothesized their potential role in influencing the immune system of the infant. HM-EVs are now considered a bioactive component of HM and the body of evidence involving their handling and isolation [321,322], composition [323,324] as well as functionality [325–327] is growing.

Regarding the molecular composition of HM-EVs, the prevalence of proteins, nucleic acids, and lipids derived from the plasma membrane and cytoplasm of the cells of origin is expected. Previous studies of the protein and miRNA content of HM-EVs revealed the presence of 920 proteins [305] and 1523 miRNAs [328], respectively. However, little is known about the lipid composition of HM-EVs. The lipid composition of EVs has been reported to be dynamic, depending on several factors (e.g., producer cell type, physiological stage of the producer cell, fate and function of the EV) [329], and may provide useful information about the purity of isolated EVs [330]. The potential use of Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) spectroscopy for a fast and direct determination of the total lipid and protein content of isolated EVs from HM was demonstrated [310]. Significant correlations between ATR-FTIR spectra and lipidomic fingerprints obtained by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), reporting the presence of glycerolipids, sphingolipids, and glycerophospholipids, were observed. Chen et al. [323] assessed the lipidomic fingerprints of colostrum and transitional HM-EVs

from mothers of preterm and term infants by LC-MS/MS. Despite the limited sample size (N = 3), their findings showed that some lipids were distinct between preterm and term HM-EVs, with high abundance of specific lipids (i.e., phosphatidylethanolamine (PE) (18:1/18:1) phosphatidylcholine (PC) (18:0/18:2), PC (18:1/16:0), phosphatidylserine (PS) (18:0/18:1), and PS (18:0/22:6)) in both groups.

Data normalization plays a crucial role in 'omics' studies, including lipidomics. The main purpose of data normalization is to minimize systematic bias and technical variations that can arise during data generation and processing in order to enhance the reliability and biological interpretability of the results [331]. Thus, normalization helps to remove unwanted variation caused by biological factors such as differences in EV abundance across samples, as well as technical or instrumental factors such as batch effects, sample size differences, or experimental conditions. Additionally, normalization facilitates the integration of multiple datasets from different studies or platforms, enabling meta-analysis and the joint analysis of data to enhance statistical power, thereby supporting the identification of consistent patterns and phenotypes across diverse datasets.

Although several statistical strategies have been proposed in the field of metabolomics and lipidomics [332], normalization is still a big challenge and there is no guidance on the implementation of normalization strategies for the lipidomic analysis of EVs. Although the International Society for Extracellular Vesicles (ISEV) survey comments on the importance of normalization of EVs in the context of comparative functional studies, it evidences broad disagreement in this regard [333]. The aim of this work was to showcase the impact of the use of different normalization strategies by means of simulated data as well as experimental lipidomic fingerprints of EVs

acquired by LC-MS and their implications for outcome interpretation. As a case study, we report potential changes of the lipidome of HM-EVs associated with biological (preterm delivery, postnatal age) and technical (heat treatment during Holder pasteurization) effects.

8.3. Material and methods

8.3.1. Simulated data

A hypothetical scenario (i.e., simulation) with four samples with different EV sizes and/or particle counts assuming a homogeneous population of EVs (i.e., narrow size distribution) was constructed (see Table AI.1). Besides, morphology of EVs was assumed to be spherical.

In all four samples, the protein and lipid (i.e., lipid 1 and 2) contents of EVs were considered directly proportional to the EV surface (i.e., integral and peripheral EV proteins) and volume (i.e., encased EV protein cargo). Thus, total protein and lipid contents in the simulated EV samples were calculated according to Eq. 1:

$$\text{Total content of a particular compound} = N \times \left(\alpha 4\pi r^2 + \beta \frac{4}{3}\pi r^3\right) \quad \text{(Eq. 1)}$$

where N refers to the total number of EVs (i.e., particles), r to the EV radius and α and β are the number of proteins or lipids per surface and volume in the EVs, respectively (see Table AI.1). In this simulation, α and β values for proteins and lipids were equal across all four simulated samples, i.e., the number of proteins ad lipids per surface and volume is constant, while the EV count and size vary. Lipids were simulated to be mainly located in integral

and peripheral locations as compared to the cargo (α>β), whereas for proteins a slightly higher proportion in the encased location as compared to the surface was selected (α<β).

8.3.2. HM samples

HM samples were collected from mothers of preterm infants (gestational age < 32 weeks; N = 5) and term infants (gestational age > 37 weeks; N = 5) in addition to a pooled donor human milk (DHM) sample. For milk expression, electric breast milk pumps with sterilized removable parts and bottles in accordance with the standard operating procedure of the hospital and the HM bank were employed. After extraction, bottles were stored at -20 °C. A total of N=20 DHM samples were defrosted and pooled together to form a batch of approximately 2 L, from which a 25 mL aliquot was collected before (i.e., DHM-Pre) and after (i.e., DHM-Post) Holder pasteurization (i.e., 30 min at 62.5 °C followed by fast cooling to 4 °C). The study was conducted in accordance with relevant guidelines and regulations including the Declaration of Helsinki. The Ethics Committee for Biomedical Research of the Health Research Institute La Fe (Valencia, Spain) approved the study protocol (approval number 2020-052-1), and mothers gave their written consent to participate.

8.3.3. Isolation of HM-EVs

EVs from HM samples were isolated through multi-stage ultracentrifugation as described elsewhere [334]. Briefly, after gentle manual shaking of a 25 mL HM aliquot during 30 s, the upper fat layer resulting from

two consecutive centrifugations at 3000 x g for 10 min at 4 °C employing an Eppendorf 5804 benchtop centrifuge equipped with an A-4-62 rotor (Hamburg, Germany), was removed. Then the supernatant was syringe-filtered (0.40 μm) and centrifuged (3000 x g, 10 min, 4 °C) to pellet proteins. The supernatant was collected and ultracentrifuged twice at 10000 rpm for 1 h, at 4 °C using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti rotor (Indianapolis, United States) to pellet remaining proteins. The supernatant was syringe-filtered (0.45 μm). Then, three ultracentrifugation steps at 30000 rpm for 2 h at 4 °C were performed to pellet HM-EVs. Between ultracentrifugation steps, supernatants were discarded, and pellets were washed with 25 mL of phosphate buffered saline solution (PBS). After the last ultracentrifugation step, supernatants were aspirated, and the isolated HM-EVs pellets were suspended in 200 μL PBS and stored at -80 °C.

8.3.4. Reference characterization of HM-EVs

Morphology of HM-EVs was assessed through transmission electron microscopy (TEM), size distribution and quantification of vesicles were analyzed by nanoparticle tracking analysis (NTA) and ExoView technology, while common canonical exosome markers were determined by western blot (WB) and by the ExoView device. TEM and WB were performed as described [335]. Briefly, EVs pellets were suspended in PBS, loaded onto Formvar carbon-coated grids, and contrasted with 2% uranyl acetate. Preparations were examined with a FEI Tecnai G2 Spirit transmission electron microscope (FEI Europe, Eindhoven, The Netherlands) and images were recorded using a Morada CCD camera (Olympus Soft Image Solutions GmbH, Münster, Germany). For WB, HM-EVs were lysed in RIPA buffer containing protease

(Complete Mini, Sigma-Aldrich) and phosphatase (PhosSTOP, Sigma-Aldrich) inhibitors. Protein concentration was determined following the Bicinchoninic Acid (BCA) Protein Assay Kit (Merck Life Science S.L.U.). 30 µg of denatured protein were separated on 10% SDS-polyacrylamide gels and transferred to polyvinylidene difluoride membranes (Immobilon-P; Millipore). After blocking, membranes were incubated overnight with following human primary antibodies: anti-HSP70 (dilution 1/500; Cell Signaling Technology; D69), anti CD63 (dilution 1/200; Immunostep; TEA3/18), anti-TSG101 (dilution 1/200; Santa Cruz; C-2), anti-CD81 (dilution 1/500; Santa Cruz; B-11) and anti-CD9 (dilution 1/500; Santa Cruz; C-4). Peroxidase-conjugated secondary antibodies used were anti-IgG rabbit (dilution 1/4000; Dako; P0448) and anti-IgG mouse (dilution 1/10,000; Sigma-Aldrich; A9044). Detection was carried out using the SuperSignalTM West Femto (Thermo Fisher Scientific). Reactions were visualized using an Amersham Imager 600 (GE Healthcare). Prior to NTA analysis using the NanoSight-NTA NS300 (Malvern Panalytical), isolated EVs were diluted 1:1000 with PBS, except for the DHM sample, that was diluted 1:2000 (DHM-Pre) and 1:4000 (DHM-Post). Besides, the ExoView platform (NanoView[1] Biosciences, MA, USA), in which through a multiplexed microarray chip EVs are immune-captured (in this study C63, CD81, and CD9) and their size, quantity and surface-protein characteristics are determined at a single EV level by interferometric reflectance imaging sensing, was employed. For the analysis with the ExoView platform, isolated HM-EVs were diluted 1:1 x 106 in 0.22 µm pre-filtered PBS and then incubated on 'ExoView Human Tetraspanin Plasma' chips prior to counter-staining with CD9, CD63, and CD81 fluorescent antibodies, following manufacturer's protocol.

[1] NanoView Biosciences has been acquired by Unchained Labs (CA, USA)

8.3.5. Lipidomic analysis of HM-EVs

For details regarding the extraction of lipids and other polar metabolites from HM-EVs, the analysis by LC-MS, and the pre-processing of MS data for peak table generation, readers are referred to [310]. In brief, lipids and metabolites were extracted from 45 µL of isolated HM-EVs using a single-phase extraction with methanol and tert-butyl methyl ether. Extraction of the raw and pasteurized pool of DHM was performed in triplicate, while the remaining samples were extracted only once. Untargeted lipidomic analysis was carried out using reverse phase LC followed by detection with an Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS system working in the positive and negative electrospray (ESI) modes. MS/MS spectra were acquired in the Data Dependent Acquisition (DDA) mode to support metabolite annotation as described elsewhere [103]. Compounds were putatively annotated (Level ID: 2) by matching experimentally acquired MS/MS spectra with experimental and in silico MS/MS databases in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [9,317] using the Human Metabolome Data Base (www.hmdb.ca) spectral library with the following parameters: m/z accuracy error between experimental and database parent ion, 20 ppm; difference in RT between parent ion and MS/MS spectra, 0.15 min; intensity threshold to remove low intensity MS/MS signals, absolute intensity <250 AU or relative intensity <0.01% of base peak; MS/MS spectra normalization, base peak; minimum number of ions in each experimental and database spectrum, 5 and 4, respectively; minimum number of matching ions between experimental and database spectrum, 3. In addition, LipiDex [286] was used employing LipidBlast [318] with 0.01 Da tolerances in both MS (precursor) and MS/MS

(fragments) data for the matching against spectra gathered in 'LipidBlast Acetate' and 'LipidBlast Formic' libraries.

Results from the lipidomic analysis were normalized using five different strategies:

Strategy A: sample volume normalization. This strategy employs the sample volume as normalization factor. In this study, since the milk volume for the isolation of EVs was the same for all studied samples, peak areas of lipids were handled without any further processing.

Strategy B: NTA particle count normalization. This strategy uses the total number of particles per sample determined by NTA for normalization.

Strategy C: ExoView particle count normalization. This strategy employs the total number of particles per sample estimated using the ExoView platform for normalization. In this study, the mean total particle count per capture probe (i.e., CD63, CD81 and CD9) in each sample was adopted as an estimate for total particle count in each sample.

Strategy D: Total protein content normalization. This strategy is based on the use of the concentration of total protein in each sample determined by the BCA assay as normalization factor.

Strategy E: Total lipid signal normalization. This strategy uses the sum of peak areas of all LC-MS features in a particular sample. Thus, peak areas of individual lipids in each sample were divided by the sum of all signals in this sample determined by LC-MS.

8.3.6. Data availability

We have submitted all relevant data of our experiments to the EV-TRACK knowledgebase (EV-TRACK ID: EV230598) [336]. MS1 and MS2 data, metadata and the dataset generated and analyzed during the current study are accessible via the Zenodo repository (https://zenodo.org/) under doi: 10.5281/zenodo.8042647 (will be made available upon publication of the manuscript).

8.4. Results

8.4.1. Normalization of simulated EV lipidomic fingerprints

To showcase the effect of employing different normalization strategies, simulated data were used. **Figure 8.1** shows the retrieved lipid content under the different normalization strategies.

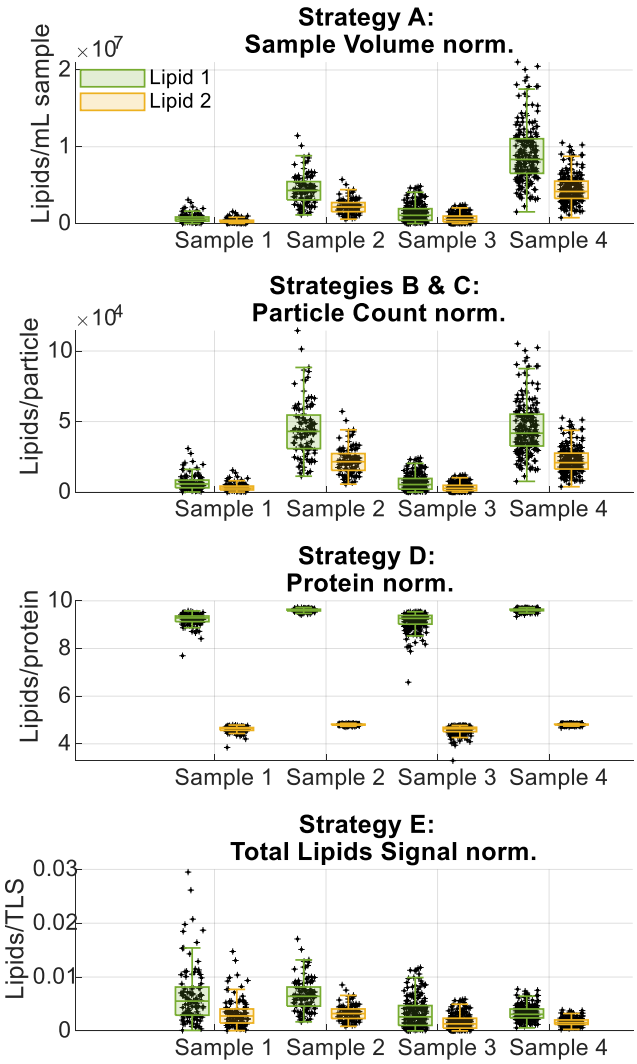


Figure 8.1. Effect of different normalization strategies on the assessment of the lipidomic fingerprint of four simulated EV samples.

As it can be seen, when normalization to sample volume is used, the lipid contents largely depend on the EV number and size. When strategy B/C is employed, i.e., data is normalized to the particle count, differences due to

the number of particles between samples are compensated, however, as the particle sizes is not accounted for, still different lipid concentrations across the four samples are observed. Of note, the impact on between sample differences due to differences in particle size (samples 1&3 (r = 50 nm) vs samples 2&4 (r = 100 nm)) is bigger than due to particle count. Hence, the use of normalization strategy B/C requires that all study samples show homogeneous and similar particle size distributions. The use of total protein or lipid content accounts for differences in particle count and size, however, the distribution of proteins and lipids between EV membrane and EV cargo will have an impact on the result. While in the simulated data sets proteins are assumed to be present in both membrane and cargo, lipids were predominately assumed to be located in the membrane. **Figure 8.1** shows the impact of these differences in the location of proteins and lipids on the outcome of normalization. However, when employing normalization strategy D (total protein normalization) differences between lipids 1 and 2 within a particular sample are augmented, while this does not occur when employing strategy E (total lipid normalization), although in both cases differences across samples 1 to 4 are compensated.

8.4.2. Characteristics of isolated HM-EVs

EVs were isolated from five HM samples of mothers of term and preterm infants as well as from a pool of DHM samples collected from 20 mothers before and after Holder pasteurization. To characterize isolated EVs, the expression of typical exosome markers such as Hsp70, TSG101, CD63, CD9, and CD81 (**Figure 8.2A**) were determined in a pooled sample of EVs.

Besides, a round-shaped morphology in the 100-250 nm range could be evidenced (**Figure 8.2B**).

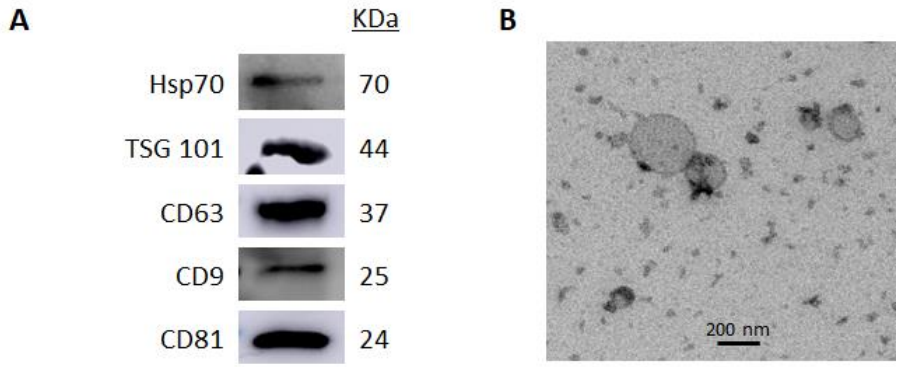


Figure 8.2. Representative Western blot of Hsp70, TSG101, CD63, CD9, and CD81 proteins in 30 µg of loaded HM-EVs (A); representative TEM image of isolated HM-EVs (scale bar = 200 nm) (B).

Characteristics in terms of particle count and particle size obtained by NTA, particle count per capture probe (i.e., CD63, CD81, and CD9) and mean particle sizes obtained with the ExoView platform, total protein content determined by the BCA assay, and total lipids, referred to as the sum of total signals (arbitrary units, a.u.) determined by LC-MS are depicted in **Figure 8.3**.

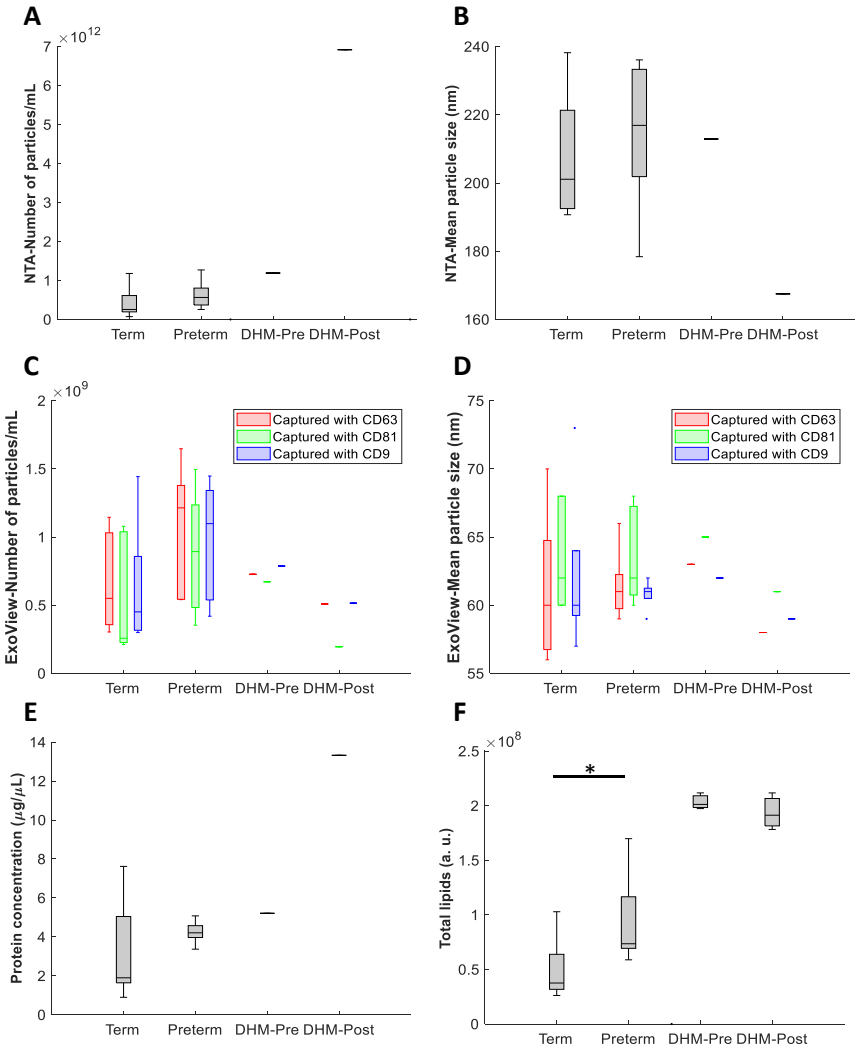


Figure 8.3. Characteristics of isolated HM-EVs. Mean particle count (A) and mean particle size (B) obtained by NTA; mean particle count per capture probe (C) and mean particle size (D) obtained with the ExoView platform; total protein content (E) obtained by BCA assay; and total signal of lipids (arbitrary units, a.u.) (F) obtained by LC-MS.

No statistically significant differences between HM-EVs derived from mothers of preterm and term infants as per particle count, particle size, and total protein content were found (Wilcoxon ranksum test p-value > 0.05) and no statistically significant correlations were found between ExoView/NTA-particle counts or BCA protein content and gestational age and postnatal age (data not shown). However, the total lipid signal was significantly higher in the preterm group (**Figure 8.3F**) and a significant, negative correlation with gestational age was found (r = -0.6; p-value < 0.05; Pearson's correlation; Figure AI.2). Figure AI.3 shows captured HM-EVs based on the expression of the ubiquitous tetraspanin markers CD9, CD63, and CD81 from the analysis of HM-EV isolates derived from selected term and preterm samples. As can be observed, tetraspanin profiles clearly varied among samples.

The pasteurized DHM sample exhibited a significantly higher count of NTA particles and total protein content, as depicted in **Figure 8.3A** and **8.3E**. However, it is worth noting that while NTA indicated a noticeable rise in particle count, the findings from the ExoView platform showed the opposite, as shown in **Figure 8.3C**. Both data from NTA and the ExoView platform showed a decreasing trend in particle size when comparing DHM-Post to DHM-Pre and no changes in the total lipid content were observed. **Figure 8.4** shows captured HM-EVs based on the expression of the ubiquitous tetraspanin markers CD9, CD63, and CD81 from the analysis of HM-EVs isolates derived from the raw (i.e., DHM-Pre) (top) and from the pasteurized (i.e., DHM-Post) (bottom) DHM samples. The marker colocalization analysis (see pie charts on **Figure 8.4**) shows the distribution of the different tetraspanins positive particles across the membranes of the extracted HM-EVs. The thermal treatment slightly influenced the tetraspanin profile of particles captured with CD63 and CD81. However, a reduction in CD81+ particles was observed for particles captured with CD9.

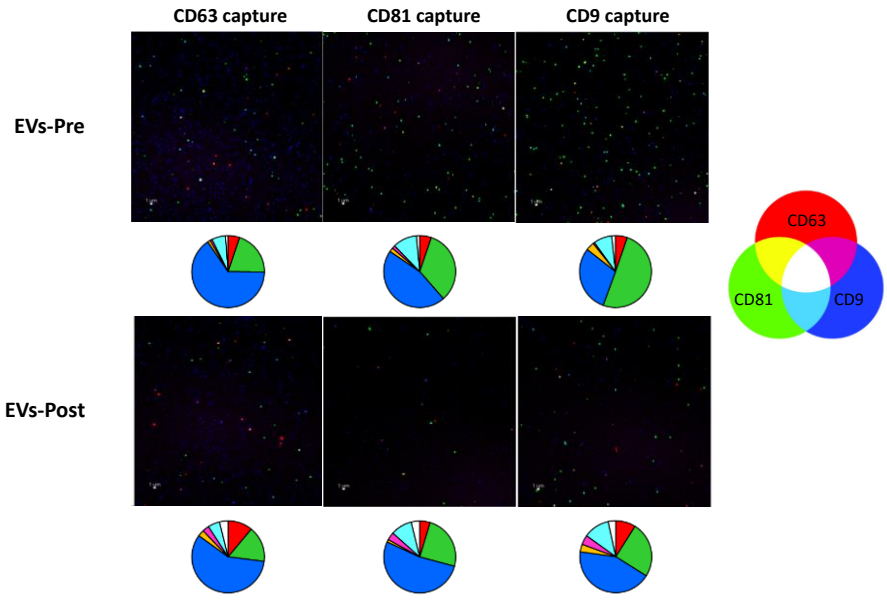


Figure 8.4. Tetraspanin fluorescent staining and marker colocalization analysis (pie charts) of EVs isolated from DHM-Pre sample (top) and DHM-Post sample (bottom). Note: scale bar, 1 µm; on the right, color code for fluorescent single positive particles (i.e., CD81+ (green), CD63+ (red) and CD9+ (blue)) and for colocalized fluorescent positive particles (i.e., CD81+/CD63+ (yellow), CD81+/CD9+ (cyan), CD9+/CD63+ (magenta), and CD81+/CD63+/CD9+ (white)).

Figure AI.4 shows Pearson's correlations between the characteristics of EVs (i.e., NTA- and ExoView-particle count, protein content and total signal of lipids) that will be applied later on as normalization factors. As it can be seen, only the results from the ExoView platform did not correlate with any other characteristic of the isolated EVs.

8.4.3. Lipid composition of HM-EVs

After peak table generation, pre-processing, and clean-up, 1085 and 696 features were detected using ESI+ and ESI- ionization, respectively. Using MS/MS data, 277 (ESI+) and 160 (ESI-) features, corresponding to 379 unique entries, were successfully annotated. **Figure 8.5** (top) summarizes the main sub-classes of the annotated features in the LC-MS dataset and **Figure 8.5** (bottom) their corresponding abundance (i.e., sum of peak intensities for all features assigned to a particular sub-class).

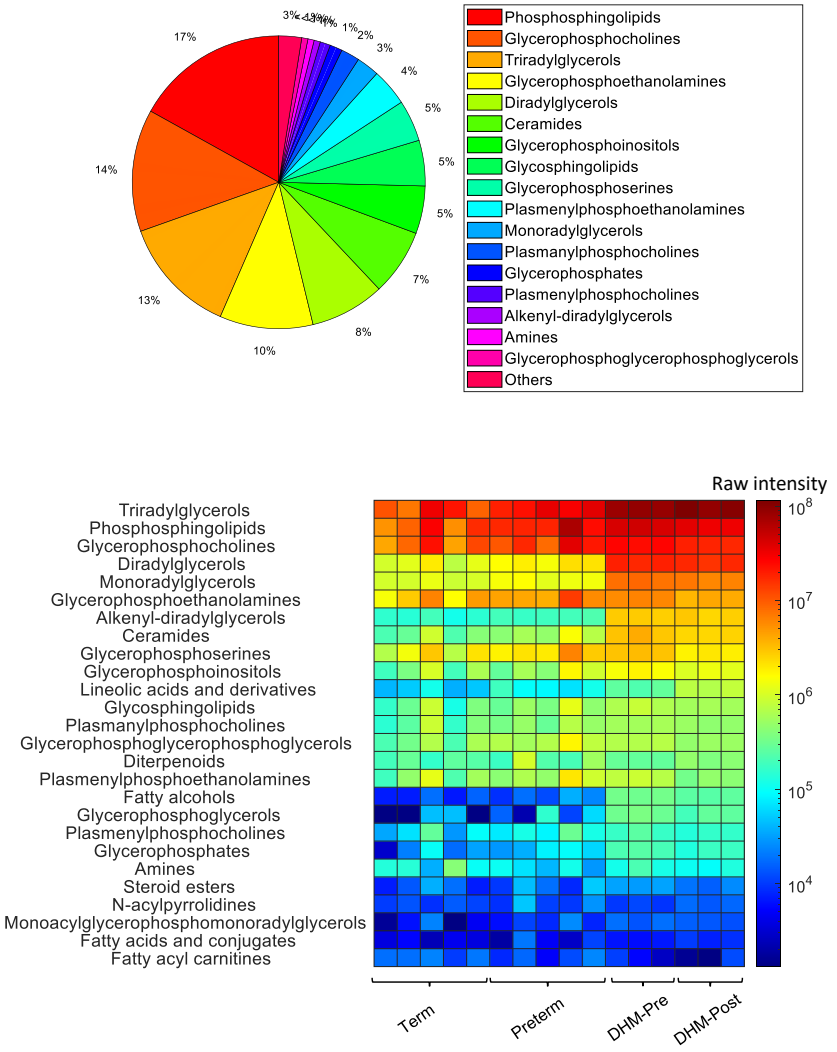


Figure 8.5. Distribution by sub-classes of the lipids annotated using HMDB or LipidBlast spectral libraries in HM-EVs (top) and abundance by sub-classes observed in HM-EVs derived from mothers of term and preterm infants, and from a DHM sample before (DHM-Pre) and after (DHM-Post) Holder pasteurization (bottom).

The sub-classes with the largest number of annotated features were phosphosphingolipids (17%), glycerophosphocholines (14%), triradylglycerols (TGs) (13%), glycerophosphoethanolamines (10%), diradylglycerols (8%), and ceramides (7%). On the other hand, the sub-classes with the highest prevalence in HM-EVs were TGs, followed by phosphosphingolipids and glycerophosphocholines (see **Figure 8.5**, bottom).

The correlation between the different lipid sub-classes determined in isolated HM-EVs of term and preterm mothers and the clinical parameters collected (i.e., gestational age and postnatal age) employing the different normalization strategies was also evaluated. Figure AI.5 shows Spearman's correlation of lipid sub-classes with gestational age corrected by postnatal age and vice versa. On the other hand, the effect of Holder pasteurization on the individual lipids encountered in HM-EVs isolates was evaluated. **Figure 8.6** shows a Venn diagram with the number of lipids that significantly changed upon pasteurization (pFDR-value < 0.05, paired Student's t-test) according to the normalization strategy adopted. As it can be seen, a higher number of altered lipids is identified in case of normalization with NTA-particle count (i.e., 348 lipids or 92%) and the total protein content (i.e., 316 lipids or 83%), while changes observed with all other normalization strategies are rather small, affecting only between 7 (2%) and 13 (3%) of the 379 unique lipids detected.

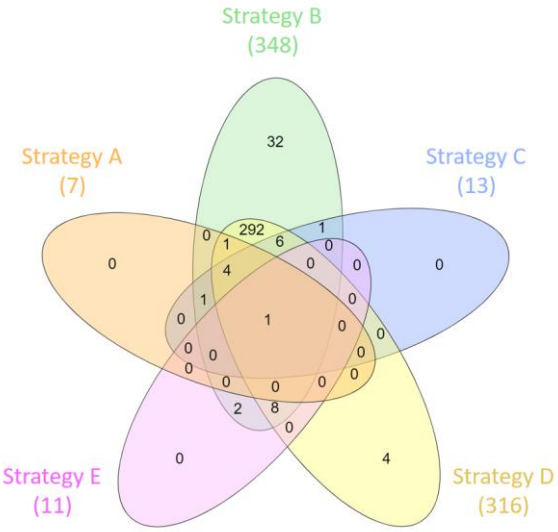


Figure 8.1. Venn diagram [337] of lipids found in HM-EVs that significantly change upon Holder pasteurization (pFDR-value < 0.05, paired Student's t-test) with the different normalization strategies. Note: fragments and duplicate entries were not considered. For the different normalization strategies (i.e., A to E), see legend of Figure 8.5.

8.5. Discussion

This work reports on the lipidomic fingerprint of HM-EVs from mothers of preterm and term infants as well as DHM samples before and after pasteurization. The experimental procedure allowed to detect a wide range of lipid classes in HM-EVs including typical membrane lipids, as well as storage lipids. Whereas phosphosphingolipids and glycerophosphocholines are lipids commonly ascribed to membrane lipids, here, a high abundance of TGs, which

are mainly found in lipoproteins and lipid droplets was reported. This could be associated to the presence of a subtype of vesicles/particles with a very high content of these lipids, whose presence may be related to co-isolation of lipid droplets or autophagosomes [330]. Although this contamination arises as a possibility, one should keep in mind that reports regarding lipidomic analysis of HM-EVs are scarce and that the high content of TGs could be an inherent characteristic of these EVs as for some EVs released from other cell types [307]. Chen W. et al. (Chen et al., 2021) found that out of the 395 lipids identified, TGs accounted for 22.5% of the total. However, their abundance was comparatively lower with respect to other lipid sub-classes. On the contrary, Blans et al. observed that EV isolates from size-exclusion chromatography were devoid of TGs and presented a phospholipid profile different from milk fat globules [338].

The normalization of lipidomics data is an essential step for facilitating meaningful comparisons between samples and enabling biological interpretation of the results. Despite its crucial importance, the normalization of EVs used for comparative functional studies lacks a defined strategy and remains an open field of research. The ISEV survey [333] highlights a widespread disagreement concerning the normalization of EVs. Furthermore, the literature specifically addressing normalization strategies in lipidomic analysis of EVs is particularly scarce. The set of normalization strategies evaluated in this study included normalization to sample volume as normalization factor, as well as widely used characteristics of isolated EVs (i.e., particle count, total protein content), and the total signal of lipids observed by LC-MS. While NTA, BCA and total lipids detected across different samples showed correlation, ExoView particle counts were uncorrelated (see Figure AI.4). This might indicate the presence of impurities

in EV isolates, as the ExoView platform, contrary to all other employed methods, specifically detects EVs.

Strategy A (i.e., sample volume normalization) reflects the huge variability encountered in the different samples involved in a particular study. Therefore, to capture the true variability of a particular factor and to carry out comparisons between samples is very challenging. For our field of study, this strategy best reflects the lipid content encountered in the EVs that is taken up by the infant. Many significant negative correlations with gestational age (Figure AI.5, A) are observed with this approach. Other authors have described the effect of gestational age on lipid composition of both HM and the milk fat [339–341]. However, it should be noted that: (i) in this study, HM samples of the preterm and term group were not collected at the same postnatal age (median (interquartile range), 7 (4) for the term group vs 60 (45) for the preterm group; p-value < 0.05, Wilcoxon ranksum test) and (ii) the number of EVs in the preterm group was slightly higher than in the term group (**Figure 8.3A** and **8.3C**). Indeed, a negative trend in the particle count (see **Figure 8.3A** and **8.3C**) and particle size (**Figure 8.3B** and **8.3D**) with gestational age was observed. Hence, normalization to sample volume does not render specifically useful for the between sample comparison of EV lipidomic fingerprints as it depends on the particle count and size of the samples (see results from simulation in **Figure 8.1**).

With strategies B and C, based on particle count, comparisons among different samples in terms of lipid content seem feasible and might be only biased when EVs-subpopulations with broad, multi-modal particle size distributions are present in the study samples (see simulation in **Figure 8.1**). NTA is widely used among laboratories interested in the field of EVs. However, certain inherent limitations to this technique should be taken into

consideration [342]. Unlike the ExoView platform, NTA assumes that all detected particles are vesicles, and therefore, protein aggregates, large crystal of salts and other components (such as milk fat globules and casein micelles in the case of HM) might be mistakenly being accounted for as vesicles, which results in an overestimation of total EVs. Besides, some EVs might be below the resolution limit of NTA and, as a consequence, small EVs might be overestimated in their size. Regarding the ExoView platform, one of its main advantages is that CD63, CD81, and CD9 positive EVs down to 50 nm can be detected. However, an absolute value of total particles cannot be provided since colocalization of the different tetraspanin markers takes place, what hinders the employment of this platform for normalization purposes. In addition, when the purity of EVs isolates is compromised, these two strategies are expected to provide different findings. In fact, in this study, total particle counts of captured CD63, CD81, and CD9 positive EVs with the ExoView platform were three orders of magnitude lower than those obtained by NTA. Apart from differences arising from possible impurities in the sample, differences linked to the detection approaches themselves might also play a role as it has been previously reported for particle count as well as for size [342,343].

Regarding strategy D (i.e., protein content normalization), it assumes a positive correlation between the EVs abundance and the total protein content in the sample, which agrees with results depicted in Figure AI.4. With this strategy, similar issues with respect to purity of EV isolates arises and hence, this method will only render suitable for between-sample comparison when protein contamination can be excluded (especially caseins in HM samples) or can be assumed to be constant among all study samples.

Similarly, in the case of lipidomics studies focusing on the comparison of the relative content of lipids among samples, strategy E might be suitable as any potential lipid contamination will affect both the lipidomic fingerprints and the total lipid content. The main advantage of this approach is that it does not require additional characterization of the sample, since comparisons between samples are assessed in terms of relative abundances (see **Figure 8.1**). However, if the samples differ substantially in their lipid content, a potential bias due to differences in the ionization of lipids during MS analysis can arise.

The choice of the most adequate normalization strategy in lipidomics studies of EVs will depend therefore on the scientific question to be answered as well as on the purity or type of contaminations expected in isolated EVs since depending on the normalization strategy adopted, the biological significance of the findings involving the lipid content differ substantially.

As an example, the study of the effect of pasteurization on HM-EVs can be used. Previously, the resistance of HM-EVs to Holder pasteurization has been described [344]. However, in the present study a considerable rise in the number of particles in the post vs. pre-pasteurization sample was observed determined by NTA, while the ExoView platform provided evidence of the opposite. It has been reported that the heating process (i.e., pasteurization) very likely causes the disruption of the fat globule membrane with a concomitant significant decrease in fat globule size [299]. Hence, we hypothesized that, during EVs isolation of pasteurized milk, small milk fat globules were co-isolated, what would explain higher NTA-particle counts. However, this should have been reflected in an increase in total lipids that was not observed (**Figure 8.3F**). Interestingly, both NTA and the ExoView platform show a decreasing trend in EV size derived from the post- vs. pre-

pasteurized DHM (**Figure 8.3B** and **8.3D**). As for milk fat globules, a potential rearrangement of membrane lipids due to the thermal process might be responsible for the observed phenomenon.

On the contrary, the total protein content of the DHM-post pasteurization sample was outstandingly high (**Figure 8.3E**), while specific EV surface protein markers remained largely unaffected by pasteurization (**Figure 8.4**). Literature reports state that casein micelles have a size comparable to milk EVs and can thus be co-isolated with the latter. In this sense, some approaches to break down or precipitate casein micelles prior to EV isolation (e.g., acidification [345], addition of chelating agents [346], addition of sodium citrate [347]) have been described. However, these approaches were tested in bovine milk samples that have higher casein contents than HM and it remains unexplored whether these findings are transferable across milk types and species. On the other hand, de Oliveira et al. observed that β-casein proteolysis in HM was affected by Holder pasteurization [348], what they attributed to previously described heat-induced structural changes encountered in the protein fraction of milk, that consisted of the aggregation of whey proteins and further formation of complexes with caseins on the surface of casein micelles [349]. According to this, we hypothesize that the higher protein and NTA-particle count observed in isolated EVs from pasteurized DHM can be attributed to the joint effect of pasteurization and co-isolation of casein micelles. In this sense, the altered lipidic fingerprints encountered with normalization strategies B and D might most likely be artifacts due to casein co-isolation with HM-EVs after DHM pasteurization. According to strategies A, C and E, only a small number of lipids was altered upon pasteurization, which supports the findings by Miyake et al. [350] with respect to the therapeutic effect of pasteurized HM-EVs when compared with their isolated raw counterparts.

8.6. Conclusions

In this study, EVs obtained from HM were characterized and subjected to lipidomic analysis using LC-MS. The presented dataset highlights the significance of employing a diverse range of complementary characterization methods for all study samples. The isolation of pure EVs isolates in this study was hindered by the co-isolation of milk fat globules and/or caseins. The choice of data normalization approaches influenced the comparability of lipidomic fingerprints between samples. Results found show that, when normalizing lipidomic fingerprints of EVs, it is crucial to carefully select an appropriate procedure considering the specific scientific question, as well as the purity and size distribution of isolated EVs. The presented findings emphasize the necessity for standardization in the isolation and characterization of HM-EVs. Additionally, there is a specific need for the development of guidelines regarding data processing in untargeted, LC-MS-based lipidomics studies. This standardization and guidance will facilitate future investigations into the clinical significance of the lipid composition of HM-EVs and its impact on newborn outcomes.

Chapter 9. ATR-FTIR spectroscopy for the routine quality control of exosome isolations

9.1. Abstract

Exosomes are nanosized vesicles containing specific cargos of DNA, RNA, proteins, metabolites, and intracellular and membrane lipids. Exosome isolation needs to be optimized carefully depending on the type of biofluid and tissue and the retrieved exosomes need to be characterized. The main objective of this study was to determine the feasibility of a multimodal analysis of Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) spectroscopy and UPLC–QqTOF-MSMS for the development of a routine quality control tool of isolated exosomes and the rapid characterization of their lipid profiles and total protein content. Using human milk as model example, exosomes were isolated by multi-stage ultracentrifugation. After single-phase extraction, lipidomic analysis was carried out by UPLC–QqTOF-MSMS with automated MSMS-based annotation using HMDB, METLIN, LipidBlast and MSDIAL databases. The classes with the largest number of annotated features were glycerophospholipids, sphingolipids, and glycerolipids. Then, dry films of 2 μL exosomes were directly analysed by ATR-FTIR. Multivariate analysis showed significant associations between ATR-FTIR specific regions and the concentrations of different lipid classes. Principal component analysis and Hierarchical Cluster Analysis of IR and lipidomic data showed that ATR−FTIR renders valuable qualitative descriptors of the lipid content of isolated exosomes. Total LC-MS lipid and total protein contents could also be quantified by using the area of CHs and C=O stretching bands as well as the amide I band. As a conclusion, results

obtained show that multimodal analysis of ATR-FTIR and UPLC-MS data is a useful tool for the development of spectroscopic methods. ATR-FTIR provided both, qualitative and quantitative chemical descriptors of isolated exosomes, enabling a fast and direct quantification of total protein and lipid contents.

9.2. Introduction

Exosomes are nanosized membrane vesicles discovered in 1983, released by fusion of an organelle of the endocytic pathway, the multivesicular body, with the plasma membrane [351]. Initially, exosomes were proposed to represent cellular waste vesicles to maintain homeostasis within the cell but nowadays it is acknowledged that exosomes are associated with physiological and pathological functions [351], contributing to different aspects of physiology and disease, including intercellular communication. Besides, exosomes may have clinical applications and their use as source for diagnostic biomarkers and as drug-delivery vectors for therapeutic applications is a very active field of research [352].

Isolation techniques include differential centrifugation, ultracentrifugation, density gradient centrifugation, ultrafiltration, immunoprecipitation, and size exclusion chromatography providing different levels of recovery, purity, and sample throughput. Due to the increasing potential for the use of exosomes in clinical applications and research, there is a need to develop analytical methods to support the technical standardization of their characterization and the quality control (QC) of their isolation [312]. The characterization of exosome subpopulations and the QC of exosome

isolations is a complex analytical challenge due to the diversity, heterogeneity, and complexity of their composition. Indeed, exosomes contain specific cargos of DNA or RNA including single-stranded DNA, double stranded-DNA, mitochondrial DNA, and miRNA, as well as proteins, metabolites and intracellular and membrane lipids [353], with structural and biological activity. Common techniques for their characterization include transmission electron microscopy, scanning electron microscopy, cryogenic electron microscopy, and atomic force microscopy for visualization; analysis of transmembrane proteins (i.e., tetraspanins such as CD9, CD63, and CD81) by Western blotting for identification; nanoparticle tracking analysis, asymmetric field-flow fractionation, resistance pulse sensing and protein quantification (e.g., bicinchoninic acid (BCA) assay); and the ExoView platform (NanoView Biosciences, MA, USA) for surface marker detection.

Lipid analysis of exosomes is a relatively unexplored field of research fostered by technical and methodological advances in high resolution hyphenated liquid chromatography – mass spectrometry (LC-MS). Due to their mechanisms of formation, the distribution of membrane lipids of exosomes is expected to be associated to the composition of the plasma membrane [330], including phospholipids, sphingolipids and cholesterol. Despite its sensitivity and detection range, LC-MS metabolomics requires highly skilled personal and bulky and expensive instrumentation, as well as careful sample preparation, and complex data processing and analysis steps, that limit its application for a routine QC of exosome isolations.

Attenuated Total Fourier Transform Infrared (ATR-FTIR) spectroscopy is an emerging tool in the bio-medical field for the analysis and characterization of biological samples. The technique normally relies on the fast and simple acquisition of the infrared spectrum from the untreated sample

using cost-effective instrumentation and no expensive reagents or consumables. The spectra contain partially overlapped bands representative of its main components, including proteins [354], lipids [355], carbohydrates and DNA [356]. Previous results have shown that, due to its speed, simplicity, and capability for finger-printing the major components of complex samples, ATR-FTIR can be used to assess extraction procedures and sample preprocessing of biofluids in metabolomics [357]. In the field of exosome analysis, ATR-FTIR has been scarcely employed and a very limited number of applications have been proposed, including the quantification of the total protein content [358], the protein-to-lipid ratio [359], the evaluation of their isolation [360], [361] as well as for the investigation of changes in the composition of saliva exosomes caused by oral cancer [362].

In this work, we aimed to assess the advantages and limitations of a multimodal analysis of ATR-FTIR spectroscopy and UPLC–MS data for the development of a routine QC tool of isolated exosomes. To this end, we employed isolated exosomes from human milk (HM) as model example and used lipidomic profiles analyzed by UPLC-QqTOF-MS, and total protein contents determined by the BCA assay as reference methodologies. To our knowledge, no study has combined ATR-FTIR and untargeted UPLC-MS-based lipidomics for the analysis of the lipid composition of isolated exosomes. Results obtained indicated that that the joint analysis of IR and lipidomic information could be a powerful strategy for sensor development with potential to foster innovative cross-disciplinary research. Besides, results confirmed the capabilities of ATR-FTIR spectroscopy for a fast and direct determination of lipids and proteins, and also that this technique can be used for a rapid evaluation and characterization of the exosomes in terms of lipid composition, thus supporting its use for the QC of the exosome isolation.

9.3. Materials and methods

9.3.1. Reagents and materials

LC-MS grade acetonitrile (CH_3CN), isopropanol (IPA), and methanol (CH_3OH) were obtained from Scharlau (Barcelona, Spain); formic acid (HCOOH) (≥95%), albumin from bovine serum (≥98%), BCA Kit for Protein Determination, phosphate buffered saline (PBS) and ammonium acetate (CH_3COONH_4) (≥98%) from Sigma-Aldrich Química SL (Madrid, Spain); prostaglandin $F_{2\alpha}$-D_4 from Cayman Chemical Company (Michigan, United States); oleic acid-D_9 from Avanti Polar Lipids Inc. (Alabama, United States) and tert-butyl methyl ether (MTBE) (≥99%) from Fisher Scientific SL (Madrid, Spain). Ultra-pure water was generated employing a Milli-Q Integral Water Purification System from Merck Millipore (Darmstadt, Germany).

9.3.2. Collection of HM samples

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry # 2019-289-1 and all methods were performed in accordance with relevant guidelines and regulations. Written informed consents were obtained from lactating mothers prior to sample collection and analysis of demographics and clinical information.

Ten HM samples were collected from mothers of preterm infants (< 32 weeks of gestation) and term infants (> 37 weeks of gestation) after establishing full enteral nutrition (i.e., stable intake of >150 ml/kg/day) using electric breast pumps following the instructions of the hospital staff. Milk was

collected from full expression of one breast between 7 and 10 AM and preferably a minimum of 3 h after the last feed or extraction. Extracted milk was stored immediately at 4 ℃ and within six h transported to the laboratory on ice until further processing. Parameters describing the study population are shown in **Table 9.1**. In addition, two aliquots of pooled HM samples were provided by the Human Milk Bank of the University and Polytechnic Hospital La Fe.

Table 9.1. Parameters of the study population.

Parameters	**Median (1st-3rd quartile)**
Mother's age [years]	35 (33 - 42)
Birth Weight [g]	1775 (1325 - 3078)
Gestational age [weeks]	34 ± 6 (30 ± 0 - 38 ± 6)
HM intake [mL/kg]	153 (150 - 170)
Infant's age [days]	14 (8 - 49)

9.3.3. Exosome isolation

After gentle manual shaking during 30 s, 25 mL of milk were centrifuged for removal of milk fat globules in two consecutive centrifugations at 3000 x g for 10 min at 4 ℃ using an Eppendorf 5804 benchtop centrifuge with an A-4-62 rotor (Hamburg, Germany). The upper fat layer was discarded, and the supernatant was syringe-filtered (0.40 μm) prior to a third centrifugation step (3000 x g, 10 min, 4 ℃) to pellet proteins. The supernatant was collected and ultracentrifuged twice at 10000 rpm for 1 h, at 4°C using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti

rotor (Indianapolis, United States) to pellet proteins. The supernatant was syringe-filtered (0.40 μm). Then, three ultracentrifugation steps at 30000 rpm for 2 h, at 4 ℃ to pellet HM exosomes were performed. Between ultracentrifugation steps, supernatants were discarded, and pellets were washed with 25 mL of PBS. After the last ultracentrifugation step, supernatants were aspirated, and the isolated HM exosome pellets were suspended in 200 μL PBS and stored at -80 ℃.

9.3.4. Exosome characterization

Exosome size distribution and quantification of vesicles were analyzed using the ExoView platform (NanoView Biosciences, MA, USA). Isolated exosomes were diluted 1:25x106 in 0.22 μm pre-filtered PBS and then incubated on ExoView Human Tetraspanin chips prior to counter-staining with CD9, CD63, and CD81 fluorescent antibodies. Protein concentration was determined following the BCA Kit for Protein Determination (Sigma-Aldrich Química SL) assay based on the reduction of alkaline Cu(II) to Cu(I) by proteins in a concentration-dependent manner. BCA is a specific chromogenic reagent for Cu(I), forming a complex with an absorbance maximum at 562 nm.

9.3.5. ATR-FTIR spectroscopy

Infrared spectra in the 4000 to 400 cm^{-1} range were acquired using an Alpha II (Bruker Optics GmbH, Ettlingen, Germany) spectrometer equipped with a platinum-ATR with monolithic diamond measurement interface

element (single reflection), a CenterGlowTM IR-source and a temperature stabilized DTGS detector with no purging system required. OPUS 8.5 software (Bruker Optics GmbH) was used to control the instrument. 2 µL of the exosome suspension were dropped onto the ATR crystal using a micropipette and then dried at room temperature for 20 s using an intermittent air stream. Spectra were collected by co-adding 32 scans with a resolution of 4 cm^{-1} using a previously recorded spectrum of air with the same instrumental conditions as background, with an acquisition time of 20 s. Between samples, the ATR interface was cleaned using a cotton swab and H_2O, IPA, and CH_3OH until recovery of the baseline signal. Spectral acquisition order was randomized but sample and spectral replicates of each exosome isolation were acquired in a row to avoid unnecessary freeze and thaw cycles that might affect their composition. All spectra were subjected to two initial quality tests. First, to guarantee that there was a significant amount of sample dried onto the ATR crystal, a minimum absorbance intensity threshold of 75 mAU was established for the amide I band. Secondly, to ensure that the contribution of water vapor to the spectra was negligible, a cut-off value of 10 was set for the ratio between the absorbance of the amide I band at 1642 cm^{-1} and the root mean square value of the mean centered absorbance in the 1820-1800 cm^{-1} interval.

9.3.6. Lipid extraction and LC-MS analysis

Lipids and other polar metabolites were extracted from exosomes using a single-phase extraction procedure [103,280,313]. 45 µL of isolated HM exosomes suspension in PBS were mixed with 5 µL of IS solution

containing oleic acid-D_9 and prostaglandin $F_{2\alpha}$-D_4, 80 µM and 39 µM respectively. After 2 min of sonication, 175 µL of methanol followed by 175 µL of MTBE were added followed by mixing on a Vortex® mixer during 30 s for protein precipitation and compound extraction and sonication during 2 min to assist the release of metabolites from exosomes during extraction. After centrifugation at 4000 x g for 15 min at 4 °C, 100 µL of supernatant containing the extracted lipids and metabolites were dried using a miVac centrifugal vacuum concentrator (Genevac LTD, Ipswich, UK) and dissolved in 100 µL of initial mobile phase (98% of mobile phase A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) and 2% mobile phase B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH)). A pooled QC sample was prepared by mixing 5 µL of each sample extract. In addition, a calibration blank (i.e., water instead of isolated exosomes) and a procedural blank (i.e., PBS supernatant from the last step of exosome isolation) were prepared, both containing IS.

Lipidomic analysis was carried out employing a 1290 Infinity HPLC system from Agilent Technologies (CA, USA) equipped with a UPLC BEH C18 column (50 x 2.1 mm, 1.7 µm, Waters, Wexford, Ireland). A binary mobile phase gradient was employed starting at 98% of mobile phase A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) during 0.5 min followed by a linear gradient from 2 to 20% of mobile phase B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) during 3.5 min and from 20 to 95% v/v of mobile phase B in 4 min; 95% v/v of mobile phase B was maintained during 1 min; return to initial conditions was achieved in 0.25 min and were maintained for a total run time of 13 min. Column and autosampler were kept at 55 and 4 °C, respectively, the injection volume was 2 µL, and the flow rate was set to 400 µL min^{-1}. An Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS system working in the ESI+ and ESI- modes was

used for MS detection. Full scan MS data in the range between 70 and 1500 m/z were acquired at a scan frequency of 5 Hz using the following parameters: gas T, 200 °C; drying gas, 14 L/min; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11 L min^{-1}. Mass reference standards were introduced into the source for on-the-fly automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (phtalic anhydride), 121.050873 (purine), and 922.009798 (HP-0921) m/z in ESI+, and 119.036 (purine) and 980.0163 (HP-0921, $[M^{-}H^{+}CH_3COOH]^{-}$) m/z in ESI-, as references. ESI+ and ESI- analysis were carried out in independent batches and between them, the instrument was cleaned and calibrated according to manufacturer guidelines. QCs were used to monitor the instrument performance and correct within-batch effects [288], [314]. A set of 9 QCs were injected at the beginning of each batch for system conditioning and MS/MS data acquisition. MS/MS spectra were acquired using the auto MS/MS method with the following inclusion m/z precursor ranges: 70–200, 200–350, 350–500, 500–650, 650–800, 800–950, 950–1100, 1100–1200, and from 70-1200 using, in all replicates, a rate of 5 spectra/s in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, an automated selection of five precursor ions per cycle and an exclusion window of 0.15 min after two consecutive selections of the same precursor. During the remaining batch sequence, a QC replicate was injected after every 5 study samples. Sample extracts were analyzed in random order. Three blank extracts were injected at the beginning and end of each batch for identifying unreliable, background, and carry-over features as described elsewhere [282].

9.3.7. MS data pre-processing and metabolite annotation

Peak table generation was carried out using XCMS software [315]. The centWave method was used for peak detection with the following parameters: mass accuracy, 20 ppm; peak width, (3,15); snthresh, 12; prefilter, (5,3000). A minimum difference in m/z of 7.5 mDa was selected for overlapping peaks. Intensity weighted m/z values of each feature were calculated using the wMean function. Peak limits used for integration were found through descent on the Mexican hat filtered data. Grouping before and after RT correction was carried out using the nearest method and 9 s as rtCheck argument. Finally, missing data points were filled by reintegrating the raw data files in the regions of the missing peaks using the fillPeaks method. The CAMERA package [315] was used for the identification of pseudospectra based on peak shape analysis, isotopic information and intensity correlation across samples [285]. Each dataset was processed with the following CAMERA functions: xsAnnotate, groupFWHM, findIsotopes, groupCorr and findAdducts using standard arguments. Identification and elimination of uninformative features was carried for ESI+ and ESI- data sets independently.

Metabolite annotation (Level ID: 2, putatively annotated compounds without matching to data for chemical standards acquired under the same experimental conditions) was carried out by matching experimentally acquired MS/MS spectra with the experimental HMDB, METLIN, and MSDIAL MS/MS databases in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [317], [103]. Metabolite annotation using LipidBlast [318] was carried out using LipiDex [286] with 0.01 Da tolerances in both MS (precursor) and MS2 (fragment) data and the ‘LipidBlast Acetate’ library.

Within batch effect was removed using the non-parametric QC-Supported vector regression (SVR) correction approach employing a Radial Basis Function kernel [288,314]. The selection of the tolerance threshold (ε), the penalty term applied to margin slack values (C) and the kernel width (γ) was carried out using a pre-selection of C and optimization of ε and γ using a grid search, leave-one-out cross validation and the root mean square error of cross validation (RMSECV) as target function. C was selected for each LC-MS feature as the median value of the intensities observed in QC replicates. The ε search range was selected based on the expected instrumental precision (4-10% of the median value of the intensities observed for the whole set of QC replicates). The γ search interval selected was [1, 105]. Variables with more than 2 missing values in QCs, those with RSD(QC)>20% after QC-SVRC, and for those for which the ratio between the median peak area values in QCs and blanks was lower than 6 were classified as unreliable and removed from further analysis.

9.3.8. Software

UPLC-TOFMS data acquisition and manual integration was carried out employing MassHunter Workstation (version B.07.00) from Agilent. Raw data was converted into mzXML format using ProteoWizard (http://proteowizard.sourceforge.net/). Peak detection, integration, deconvolution, alignment and pseudospectra identification were carried out using XCMS and CAMERA in R 3.6.1. Metabolite annotation and data analysis were carried out in MATLAB 2017b (Mathworks Inc., Natick, MA, USA) using in-house written scripts and the PLS Toolbox 8.7 (Eigenvector Research Inc., Wenatchee, USA) as described elsewhere [103].

Statistical heterospectroscopy (SHY) [363] was employed for the joint analysis of the lipidomic and spectroscopic data acquired from the set of isolated exosome samples. SHY involved the estimation of the covariance between signal intensities measured by the two techniques across the set of samples. The SHY plot represents the correlations of IR absorbances at each wavenumber with the UPLC-MS intensities of each annotated lipid. To facilitate the extraction of information, a correlation threshold was applied. A lipid class was not represented if >50% of the correlations of its metabolites failed to reject the null hypothesis considering a 97.5% confidence level (t-test, unequal variances).

MATLAB scripts used datasets generated and analyzed during the current study are available in the Zenodo repository (zenodo.org/deposit/5148582).

9.4. Results

9.4.1. Characteristics of isolated HM exosomes

Figure 9.1 (left) shows captured HM exosomes based on the expression of the ubiquitous tetraspanin markers CD9, CD63, and CD81 from the analysis of a representative HM exosome isolate. The marker colocalization analysis (see **Figure 9.1**, right) shows the distribution of different tetraspanins across the membranes of the extracted HM exosomes. Most of the isolated HM exosomes expressed CD9 and CD81 while only a small fraction expressed all three tetraspanins. The employed multistep ultracentrifugation protocol allowed the extraction of HM exosomes with a median size of 61 (3 interquartile range, IQR) nm and a median number of 8

x 1014 particles mL^{-1} (3 x 1015 IQR), as determined by the ExoView platform. Protein concentrations in HM exosomes isolates determined by the BCA assay ranged between 2 and 6 g L^{-1}. Regarding the lipidomic profiles recorded from HM exosomes, **Figure 9.2** summarizes the main classes of the 377 features annotated in the UPLC-MS data set (208 in ESI+ and 169 in ESI-) after data pre-processing and clean-up. The classes with the largest numbers of annotated features were glycerophospholipids (50%) [glycerophosphocholines (60), glycerophosphoethanolamines (35), glycerophosphoglycerophosphoglycerols (3), glycerophosphoserines (22), glycerophosphoinositols (13), plasmenyl PCs (15), plasmenyl PEs (24), and phosphatidylinositols (11)], sphingolipids (31%) [phosphoshingolipids (71), ceramides (22), glycosphingolipids (6), hexosylceramides (17), Glc Ceramides (2)], and glycerolipids (17%) [triradylglycerols (64)].

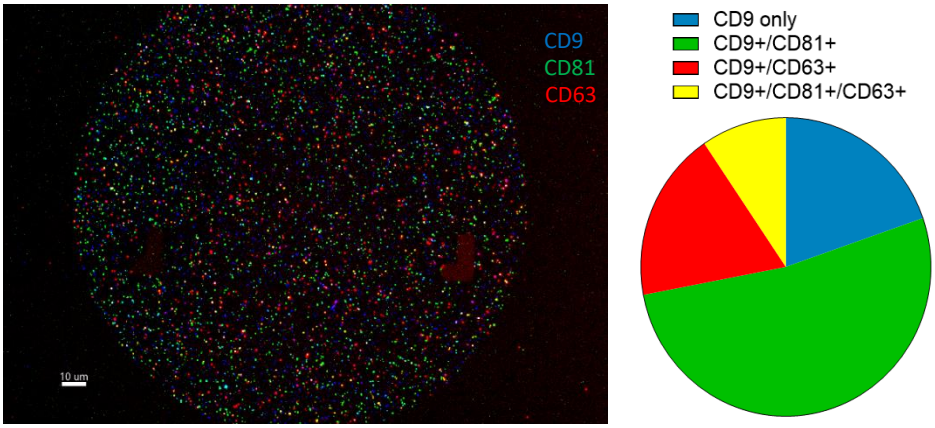


Figure 9.1. Tetraspanin fluorescent staining (left) and marker colocalization analysis (right). Note: CD-9 capture.

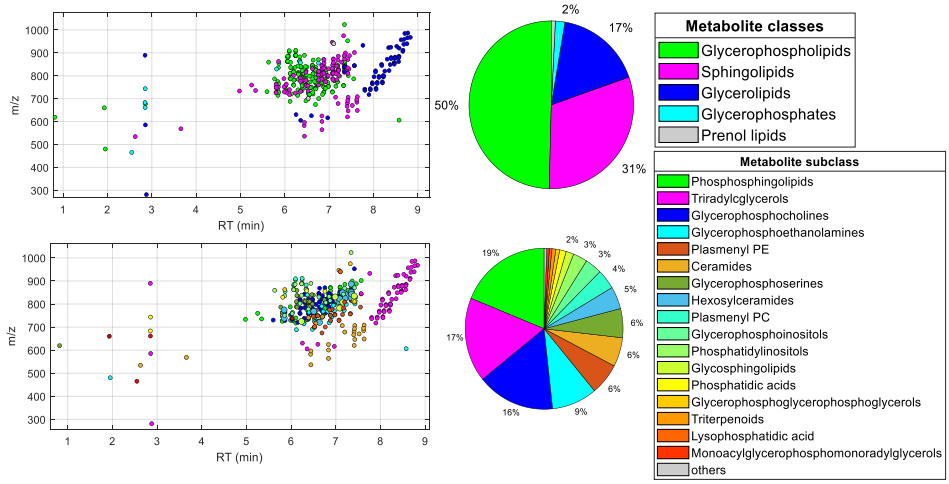


Figure 9.2. Distribution of annotated UPLC−MS features detected in HM exosomes indicating their class (top) or subclass (bottom).

Figure 9.3 shows ATR-FTIR spectra and the first and second derivative spectra in the 3450-800 cm^{-1} range of dry residues from 2 µL of exosomes isolated from four different HM samples. The figure also shows the spectra of a blank extract with PBS bands at 1125, 1069, 978, and 851 cm^{-1}. Despite the overlap between the exosome signal and PBS bands in the 1125-850 cm^{-1} region, the use of the first or second derivative spectra enabled the resolution of overlapping spectral bands in the exosome extracts at slightly different wavenumbers (note that the minimum in a second derivative spectrum corresponds to the band apex in the raw ATR-FTIR spectrum). Spectral contributions from proteins including the amide I, II, and III were detected in exosome spectra. The amide I band governed by the stretching vibrations of C=O (70-85%) and C-N groups (10-20%) of the protein peptide backbone was detected in the 1600-1700 cm^{-1} range. The amide II band arising from in-plane N-H bending (40-60%), and C-N (18-40%) and C-C (10%) stretching vibrations was detected in the 1510-1580 cm^{-1} region. The amide III band at 1250–1350 cm^{-1} is a complex band arising from a combination of several coordinate displacements. The broad, intense band at 3285 cm^{-1} was associated to N-H stretching vibrations of peptide protein groups.

Besides, spectral contributions from carbohydrates, phosphate bands from phospholipids, DNA and RNA, and lipids were observed (see **Table 9.2**). Among them, intense bands were observed in the lipid frequency domains including the $-CH_2$ and $-CH_3$ symmetric and antisymmetric stretching (3000-2800 cm^{-1}). The characteristic C=O stretching band at 1745 cm^{-1} has been associated to the ester groups of glycerophospholipids (e.g., glycerophosphocholines, glycerophosphoethanolamines), glycerolipids (e.g., triradylglycerols), and cholesterol esters, among other lipid classes.

The differences observed in the intensity of protein and lipid bands indicate changes in the protein and lipid composition across exosome isolations. Moreover, the marked differences in band shapes in the amide I and II region could also indicate changes in the higher order structure of proteins.

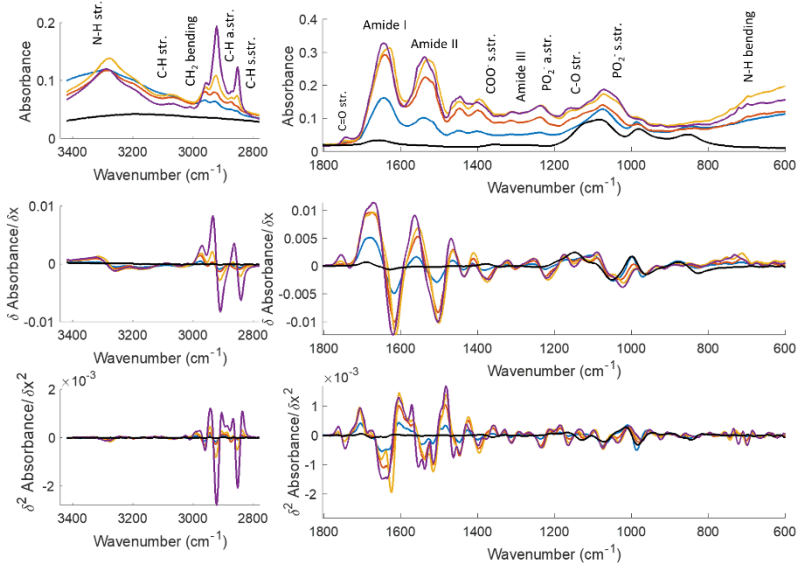


Figure 9.3. ATR-FTIR raw (top), first derivative (middle) and second derivative (bottom) spectra in the 3420-2780 cm^{-1} (left) and 1800-800 cm^{-1} range (right) of dry residues obtained from 2 µL of the set of exosome extracts in PBS using a spectrum of air as background. Note: black line corresponds to a PBS blank spectrum; Colored lines: spectra of isolated exosomes from HM samples.

Table 9.2. Assignments of main bands observed in the ATR-FTIR spectra of HM exosomes.

Wavenumber (cm^{-1})	**Spectral assignment**
3285	N-H stretching (proteins)
2959	CH_3 asymmetric stretching (lipids, proteins)
2921	CH_2 anti-symmetric stretching (lipids)
2872	CH_3 symmetric stretching (lipids and proteins)
2851	CH_2 symmetric stretching (lipids)
1745	Saturated ester C=O stretch (lipids, cholesterol, phospholipids, cholesterol esters)
1646	Amide I (proteins)
1537	Amide II (proteins)
1448	CH_2 bending of lipidic acyl chains (lipids, proteins)
1402	COO^- symmetric stretch (fatty acids, aminoacids)
1314	Amide III (proteins)
1236	PO_2^- antisymmetric stretch (phospholipids, nucleic acids)
1156	$CO{-}O{-}C$ antisymmetric stretching (glycogen, nucleic acids) C-O stretching from alcohol groups (glycogen, lipids)
1080	PO_2^- symmetric stretch (phospholipids, nucleic acids)

9.4.2. Multimodal qualitative analysis of ATR-FTIR spectra and lipidomic profiles

SHY [363] was employed to get further insight into the correlation among IR bands and UPLC-MS lipid profile. SHY enables the co-analysis of multi-source (i.e., lipidomic and spectroscopic) data sets by estimating the covariance between signal intensities measured by two techniques across the set of isolated exosome samples. Accordingly, SHY involved the calculation

of a series of linear models between the ATR-FTIR absorbance for each vibrational feature and the intensities of each of the features annotated as a glycerophospholipid, sphingolipid, or glycerolipid. Results depicted **in Figure 9.4** show statistically significant associations (p-value<0.025) between specific overlapping regions of the raw (**Figure 9.4**, left) and second derivative (**Figure 9.4**, right) ATR-FTIR spectra and the different lipid classes. This result supports the potential of SHY to explore associations between lipidomic profiles and spectroscopic data and revealed relationships between the distribution of specific annotated lipid classes and the IR spectra of exosomes.

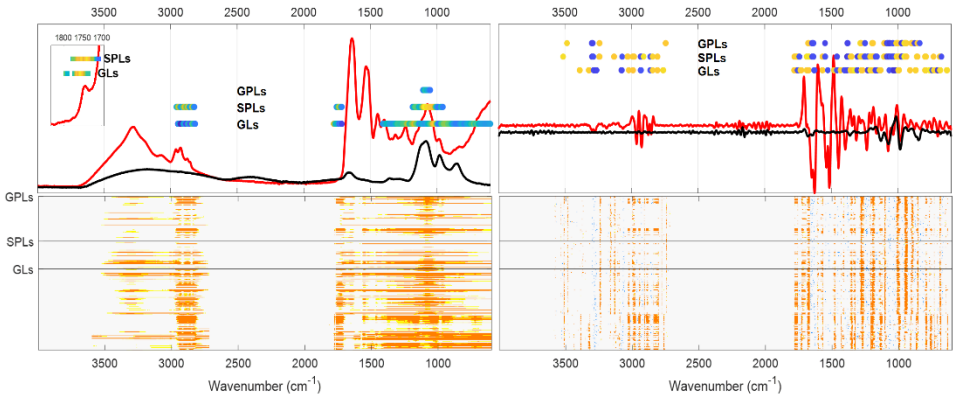


Figure 9.4. Top: Median value of the correlation between annotated UPLC-MS features clustered according to their lipid classes and ATR-FTIR data (raw: left; second derivative: right) determined by SHY using the slope coefficient of a linear model. Here, only those wavenumbers for which >50% of the features of each class showed a significant correlation (p-value<0.025) are depicted. The plot overlays ATR-FTIR spectra from isolated exosomes (red) and a blank extract (black) for a better interpretation of the results. Bottom: Correlation among annotated UPLC-MS features clustered according to their lipid classes and ATR-FTIR data determined by SHY using the slope coefficient of a linear model. A correlation cutoff was applied (linear model p-value<0.025) for better visualization. Note: GPLs: glycerophospholipids; SPLs: sphingolipids, GLs: glycerolipids. Red line = HM exosome spectrum and black line = PBS blank spectrum.

Principal component analysis (PCA) and Hierarchical Cluster Analysis (HCA) were used to identify samples with similar lipidomic profiles,

using autoscaling as data preprocessing, the Pearson correlation coefficient as distance measure, and the Ward's algorithm for clustering**. Figure 9.5a** (left) shows the scores plot of the 2 PCs model explaining 82% of the total variance in the lipidomic UPLC-MS data set. Samples were clustered mainly along PC1 in three groups including samples #7, 11, and 12 (cluster 1); #4, 5, and 9 (cluster 2); and # 1, 2, 3, 6, 8, and 10 (cluster 3). PC2 enabled the discrimination of sample 7 from 11 and 12 in cluster 1. **Figure 9.5a** (center) depicts the scores plot of the 2 PCs model explaining 57% of the total variance in the autoscaled second-derivative ATR-FTIR data, where the samples were labeled according to the clusters selected in the lipidomic data set. Using the distances among samples as similarity criteria, results showed a similar trend: samples included in cluster 1 differ from those included in cluster 3 and were more similar to those included in cluster 2. The similarity between the trends observed in both scores plots was assessed by the Mantel test [364] which evaluated the statistical significance of the correlation between the two pairwise Euclidean distance matrices. An empirical *p*-value was estimated using a permutation test where a reference null distribution is obtained by repeatedly random shuffling the objects (here, n=104). After each permutation, the correlation between the obtained two pairwise Euclidean distance matrices is calculated, and the *p*-value is estimated as the proportion of permuted estimates for which the absolute correlation value was equal to or greater than the correlation estimate calculated using the original ordering. Results obtained depicted in **Figure 9.5a** (right) (p-value<0.005) indicated a statistically significant correlation between the distribution of samples in the PCA scores spaces.

Figure 9.5b shows the distribution of glycerophospholipids, sphingolipids, and glycerolipids in the RT-m/z space, as well as the boxplots representing the distributions of sum of intensities of the three classes of lipids

in the three selected HCA clusters. Cluster 1 showed lower concentrations of the three lipid classes than samples included in clusters 2 and 3. Besides, although clusters 2 and 3 showed similar levels of glycerophospholipids, cluster 3 showed significantly higher relative levels of sphingolipids. Results also indicated a lower ratio of glycerophospholipids/sphingolipids in cluster 3, being cluster 2 the group of exosomes isolates with the highest ratio of glycerophospholipids/sphingolipids. **Figure 9.5c** (top) shows the second derivative spectral regions with a statistically significant correlation with the levels of >50% of the UPLC-MS features annotated as glycerophospholipids, sphingolipids, or glycerolipids in the sample set. The boxplots representing the distributions of the sum of absolute values in the ATR-FTIR second derivative spectra associated to the three classes of lipids in the three selected HCA clusters showed a similar pattern as those found using LC-MS data.

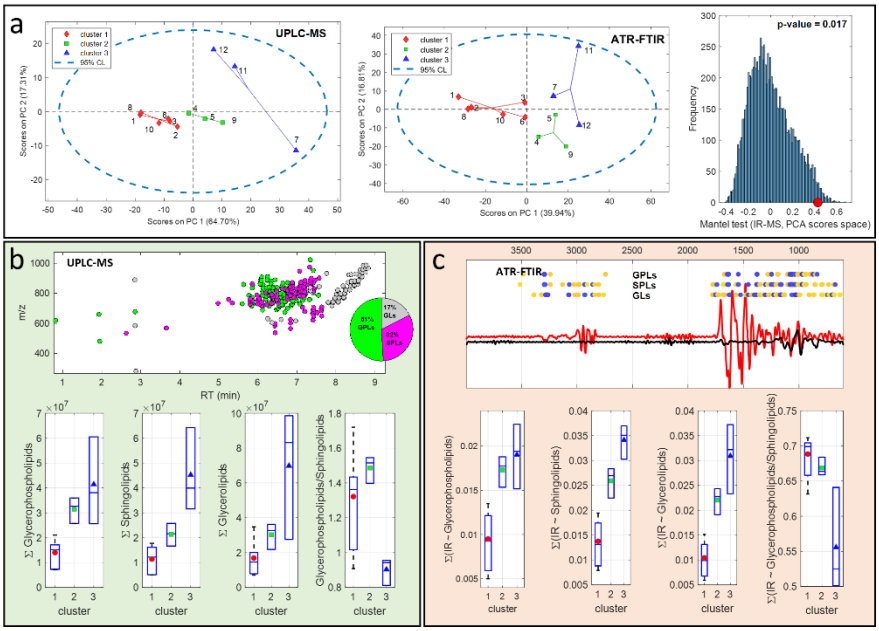


Figure 9.5. a) PC1 vs PC2 scores plots from PCA of UPLC-MS lipid profiles (left), and ATR-FTIR spectra (middle). Sample classes were assigned according to the results of HCA based on UPLC-MS data. The statistical significance of the correlation between the distribution of samples in both PC1 vs PC2 scores spaces was assessed by the Mantel test (p-value<0.001) (right). b) Distribution of GLPs, GLs and SPLs features in the m/z vs RT space (top) and box plots showing the sum of intensities of each of these lipid classes in the three clusters (bottom). c) Second derivative ATR-FTIR spectra showing the spectral regions associated to GLs, SPLs, and GPLs (top) and sum of the intensities of the spectral regions associated to each of these lipid classes in each of the three sample clusters (bottom). Note: GPLs: glycerophospholipids; SPLs: sphingolipids, GLs: glycerolipids.

9.4.3. Quantification of proteins and lipids by ATR-FTIR spectroscopy

The absorbance in the 1724-1591 cm^{-1} region showed a statistically significant correlation with the protein content determined in the isolated exosomes by the BCA assay ($A=(-0.46 \pm 0.34)+(1.09 \pm 0.06)$[protein](g/L), $R^2 = 0.79$, p-value$<10^{-5}$), in agreement with previous reports [359], [365] (see **Figure 9.6**, top left). This result suggests the use of an external calibration using a model protein (e.g., albumin) for a fast quantification of total protein content in exosomes. As shown in **Figure 9.6** (top, right), the exosome protein contents determined using the area of the amide I band of an external calibration line and a serial dilution of albumin standards showed a statistically significant correlation with protein concentrations determined by BCA (slope = (0.80 ± 0.05), $R^2 = 0.79$, p-value$<10^{-5}$).

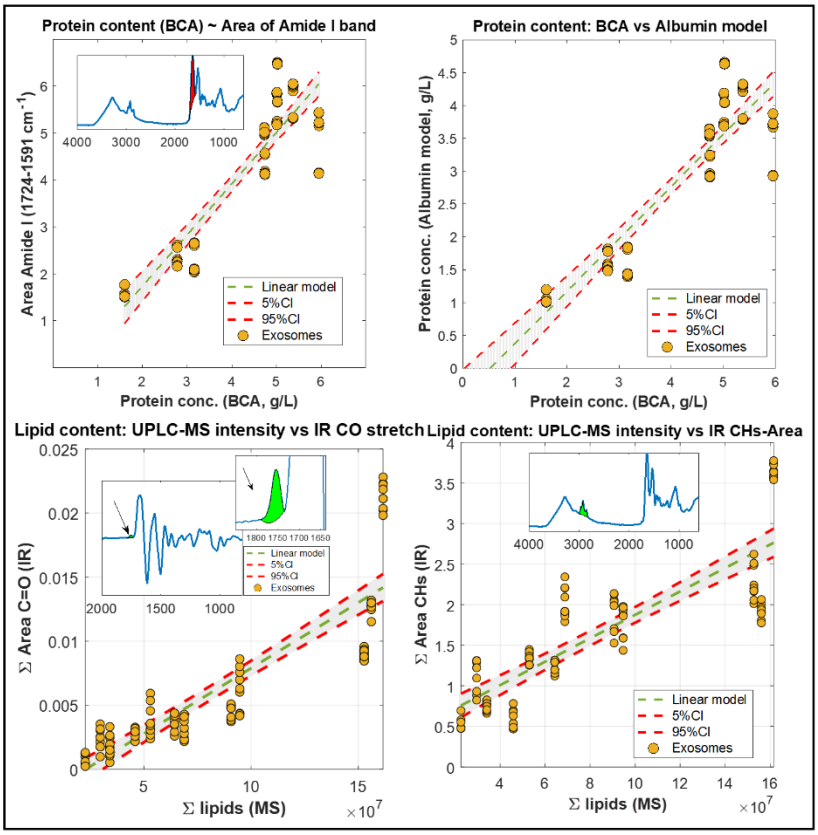


Figure 9.6. Correlation between the protein content determined by the BCA assays and the amide I area in the set of isolated exosomes (top, left) and the protein content determined by BCA assays and the protein concentration determined using an external calibration line from a serial dilution of albumin solutions and amide I area (top, right). Correlation between the lipid content determined as the sum of intensities of LC-MS annotated features and the C=O (bottom, left) and CH (bottom, right) spectral area in the set of isolated exosomes.

The use of ATR-FTIR spectra for estimating the lipid content of exosomes was also explored. Using the sum of UPLC-MS intensities of annotated lipids as surrogate of the total lipid content in samples, the IR spectral intensity between 1790-1733 cm^{-1} in the first derivative spectra associated with the C=O stretching band (~1740 cm^{-1}) correlated with the lipid content ($A=(-0.002 \pm 0.005)+(1.02 \pm 0.05)10^{-10}$[LCMS intensity](AU), R^2 = 0.75, *p*-value<10^{-5}) (see **Figure 9.6**, bottom-left). Similarly, the IR spectral area in the 3000-2800 cm^{-1} range also correlated with the lipid content ($A=(0.43 \pm 0.09)+(1.44 \pm 0.09)10^{-8}$[LCMS intensity](AU), R^2 = 0.70, *p*-value<10^{-6}) (see **Figure 9.6**, bottom-right).

The repeatability of ATR-FTIR determinations is a relevant parameter for assessing its usefulness for quantitative applications. To address this concern, the RSD% observed for each wavenumber in the 3710-2717 cm^{-1} and 1840-800 cm^{-1} ranges was calculated for each sample. In this study, the set of 12 exosome samples were analyzed by replicate (n=3), also acquiring 3 spectra/replicate. The distribution of RSD% median values obtained for each sample varied in the 3.3-11.4 RSD% range (mean value=5 ± 1, median value=5.4). Although these figures show acceptable reproducibility, automation of the dry-film generation for monitoring of exosome isolation would likely result in higher precision and sample throughput levels [366].

9.5. Discussion

We used a standard multi-step ultracentrifugation procedure for the isolation of HM exosomes from twelve HM samples. The detection of ubiquitous tetraspanins confirmed that the isolated particles are HM exosomes

rather than other co-isolated contaminants. As the results from the ExoView platform, isolated exosomes showed varying particle counts, which was in good agreement with varying protein concentrations determined by the BCA assay. This reinforces the importance of quick and reproducible approaches for characterizing exosomes for routine QC. The use of a LC-MS-based lipidomic approach allowed to gain detailed insight into the composition of HM isolated exosomes (**Figure 9.2**). The observed distribution of lipid classes agreed with the expected lipid composition of exosomes. Although their lipid profiles can be linked to their cellular origin, exosomes are bilayered proteolipids comprising mainly plasma membrane lipids (e.g., phospholipids, sphingolipids, and cholesterol) [330] enriched with bioactive cargo captured from the cytosol during the formation of the intraluminal vesicles. Previous results have observed that, compared to the cell of origin, exosome membranes are often enriched with phosphatidylserine, phosphatidylethanolamines, phosphatidylcholines, sphingolipids, and cholesterol. Besides, reported triradylglycerols could be indicative of the co-isolation of extracellular lipids during the exosome isolation, present at high concentrations in HM.

ATR-FTIR spectra showed spectral contributions from proteins, carbohydrates, phosphate bands from phospholipids, DNA and RNA, and lipids (see **Figure 9.3** and **Table 9.2**). Moreover, the joint analysis of LC-Ms lipidomic profiles and ATR-FTIR data (**Figures 9.4** and **9.5**) showed that several spectral features were associated with different lipid classes present in HM exosomes, even though the level of detail provided by this technique is lower. However, we observed specific patterns in the ATR-FTIR spectra that were indicating the presence of the three major lipid classes (i.e., glycerophospholipids, sphingolipids, and glycerolipids) identified by LC-MS. Glycerophospholipids are glycerol-based phospholipids and spectral regions

associated with >50% of glycerophospholipids included the 1100-1045 cm^{-1} region where the characteristic frequencies of the P-O-C asymmetric and symmetric stretching bands of the phosphate group can be observed. Sphingolipids are a complex family of compounds sharing a sphingoid base backbone. They were linearly correlated with ATR-FTIR absorption in the CH_2 and CH_3 symmetric and antisymmetric regions (2963-2814 cm^{-1}), C=O stretching region (1776-1708 cm^{-1}), and in the 1180-995 cm^{-1} region including characteristic frequencies associated to phosphate groups such as vC-O-PO^-_2 (1060-1070 cm^{-1}) and vPO^-_2 sym. (1093 cm^{-1}) [359]. Glycerolipids are also a structurally heterogeneous group of lipids composed of mono-, di-, and tri-substituted glycerols that have at least one hydrophobic chain linked to a glycerol backbone in an ester or ether linkage [367]. Spectral regions associated with this lipid class included bands at 2938-2930, 2866-2858, and 2841-2825 cm^{-1} linked to CH_2 and CH_3 symmetric and antisymmetric vibrational modes, as well as the regions 1772-1735 (C=O stretching), 1226-1141, 1053-881 cm^{-1} (C-O stretching from alcohols), and 739-698 cm^{-1}. The partial overlap between the spectral regions associated to each of the lipid classes agreed with their structural similarity, and included bands from the acyl chains, and the glycerol and phospholipid headgroups.

Furthermore, we could show that ATR-FTIR can be used for quantification. Specific lipid bands in exosomes, namely the 3000-2800 cm^{-1} range (CH stretching vibrations) and the 1790-1733 cm^{-1} (C=O stretching) showed a linear correlation with lipid contents determined by LC-MS (**Figure 9.6**). We could also show that the same ATR-FTIR spectrum can be used for quantitation of total proteins in HM exosomes. With a simple external calibration line from serial dilutions of an albumin solution, protein concentrations determined by ATR-FTIR were found to be comparable to

reference values from the BCA assay (**Figure 9.6**). Differences observed between concentrations obtained by both techniques might be linked to the different detection principles. The amide I band is sensitive to changes in the protein structure and hence, the use of a model protein to be used as reference might bias results. On the other hand, the accuracy of the BCA assay is limited by the presence of sample reducing agents, and by the presence of cysteine, tyrosine, and tryptophan residues. Besides, the presence of reducing sugars, lipids, and phospholipids in the sample can also affect the accuracy of the BCA. Importantly, this is the first literature report on the simultaneous quantitative analysis of proteins and lipids, and qualitative lipid analysis using ATR-FTIR spectra of exosomes. In the ATR system used, the sample is deposited onto the ATR element and an IR beam is directed through an internal reflection element (IRE). Then, the evanescent wave interacts with the sample in direct contact with the IRE providing a penetration depth in the 1-3 μm range. Therefore, the repeatability of the ATR spectral acquisition depends on the distribution of the sample on top of the ATR surface, which in turn depends on the process of dry-film generation. Here, samples were manually deposited and air-dried, achieving RSD% from replicate measurements of 5 ± 1, which is within the general acceptance criteria of 15% [2].

9.6. Conclusions

Results obtained indicated that the joint analysis of IR and lipidomic information could be a powerful strategy for sensor development with potential to foster innovative cross-disciplinary research. Results show that ATR-FTIR spectroscopy provides both, qualitative and quantitative chemical

descriptors of isolated exosomes through changes in bands associated to specific functional groups. ATR-FTIR spectroscopy enables a fast, direct quantification of total protein and lipid contents and, simultaneously, provides biochemical information useful for a routine control of changes in the relative composition of isolated exosomes in very small sample volumes (2 μL). Thus, the use of ATR-FTIR could be considered as a cost-effective alternative, to the use of standard colorimetric assays (e.g., BCA) for QC of exosome isolations. Also, whereas ATR-FTIR cannot replace the level of detailed information provided by MS in terms of specificity or sensitivity, the spectral information could be used to rapidly assess repeatability or reproducibility among exosome isolations, providing at the same time information on differences in their lipid contents.

Conclusions and Outlook

In this PhD thesis, the development of different analytical methods has been carried out and its application in the clinical field has been demonstrated through the analysis of biological samples from infants and their mothers. The conclusions derived are summarized below:

- Determination of nutrition and microbial activity biomarkers in urine samples from lactating mothers can be performed by LC-MS/MS. Likewise, its use as a complementary tool to traditional 24-hour dietary recall questionnaires is possible, presenting great potential for the comprehensive evaluation of nutrition patterns and their effect on the microbiome

- Simultaneous analysis of SCFAs and BCAAs in urine and faeces samples can be carried out by GC-MS. The application of this method allows the precise quantification of these biomarkers through a single derivatization procedure, providing information that allows distinguishing between the different study groups, in this case PIs, TIs and their mothers. In addition, a very significant contamination of acetic acid has been revealed in the gauze and cotton pads that are typically used for newborns urine sample collection that should be considered in future clinical studies

- An LC-MS/MS method has been validated for the quantitative analysis of BAs in human urine and murine faeces samples, demonstrating its applicability in the (pre)clinical field. In addition, a comparison has been obtained between the metabolic pathways and excretion mechanisms of human urinary and murine fecal BAs, according to the differences between species

- Untargeted lipidomic and metabolomic analysis and targeted analysis of the FA profile of DHM by LC-MS/MS and GC-MS, respectively, have shown that HoP significantly affects both lipid composition and physicochemical properties and functionality of HM

- The evaluation of different normalization techniques in the lipidomic analysis of HM-EVs has allowed delving into their correct isolation and characterization, as well as the potential implications derived from the selection of the normalization technique in question.

- Rapid and direct QC of the HM-EVs isolation procedure is possible using total lipids and proteins derived from the spectral profile determined by ATR-FTIR

In addition to the development and application of the different analytical methods presented, a descriptive article of the clinical study of the NUTRISHIELD project, a bibliographic review on the methodologies for the analysis of untargeted metabolomics in HM and a book chapter on the

procedure for the correct isolation and untargeted lipidomic analysis of HM-EVs have been carried out. Finally, after the development of this PhD thesis, new aspects that are currently being carried out by our research group have emerged to study, which are described below:

- The development of an algorithm capable of providing personalized nutrition advice based on the results of the nutrition biomarkers and the microbial activity obtained, together with other clinical variables and the composition of the gut microbiota and HM. This algorithm would be applied in medical consultations to give nutritional advice to lactating mothers in order to improve the composition of the HM, and thus improve the clinical results of PIs in the short and long term

- The analysis of BAs as biomarkers of microbial activity in urine and faeces samples in the neonatal context, to study the correlations between both matrices and determine the levels in this population. In this way, it will be possible to determine if it is possible to use the levels in urine samples as biomarkers of microbial activity, thus facilitating the process of non-invasive sample collection and treatment

- The study of the biological relevance and the impact of the changes observed in the composition and functionality of the components of HM after its pasteurization, as well as promoting the development of technological alternatives to HoP for the microbiological stabilization of HM

- Develop and optimize an alternative protocol to ultracentrifugation for the isolation of HM-EVs by size exclusion chromatography, to reduce the amount of sample needed, impurities, and time required for isolation. Likewise, to evaluate the properties of HM-EVs in order to implement their use as a fortifier in PIs nutrition

- Establish ATR-FTIR analysis for the routine characterization of protein and total lipids in EVs isolated from different matrices

Bibliografía

[1] J. van der Greef, A.K. Smilde, Symbiosis of chemometrics and metabolomics: past, present, and future, Journal of Chemometrics. 19 (2005) 376–386. https://doi.org/10.1002/cem.941.

[2] Food and Drug Administration (FDA). Guidance for Industry: Bioanalytical Method Validation. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine, (2018) 44.

[3] European Medicines Agency, Guideline on Bioanalytical Method Validation, (2011).

[4] X. Domingo-Almenara, J.R. Montenegro-Burke, H.P. Benton, G. Siuzdak, Annotation: a computational solution for streamlining metabolomics analysis, Anal Chem. 90 (2018) 480–489. https://doi.org/10.1021/acs.analchem.7b03929.

[5] H. Tsugawa, Advances in computational metabolomics and databases deepen the understanding of metabolisms, Current Opinion in Biotechnology. 54 (2018) 10–17. https://doi.org/10.1016/j.copbio.2018.01.008.

[6] J.C. Lindon, J.K. Nicholson, Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics, Annual Review of Analytical Chemistry. 1 (2008) 45–69. https://doi.org/10.1146/annurev.anchem.1.031207.113026.

[7] J.A. Kirwan, H. Gika, R.D. Beger, D. Bearden, W.B. Dunn, R. Goodacre, G. Theodoridis, M. Witting, L.-R. Yu, I.D. Wilson, the metabolomics Quality Assurance and Quality Control Consortium (mQACC), Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management, Metabolomics. 18 (2022) 70. https://doi.org/10.1007/s11306-022-01926-3.

[8] G. Liebisch, E. Fahy, J. Aoki, E.A. Dennis, T. Durand, C.S. Ejsing, M. Fedorova, I. Feussner, W.J. Griffiths, H. Köfeler, A.H. Merrill, R.C. Murphy, V.B. O'Donnell, O. Oskolkova, S. Subramaniam, M.J.O. Wakelam, F. Spener, Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures, Journal of Lipid Research. 61 (2020) 1539–1555. https://doi.org/10.1194/jlr.S120001025.

[9] L.W. Sumner, T. Samuel, R. Noble, S.D. Gmbh, D. Barrett, M.H. Beale, N. Hardy, Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards

Initiative (MSI), Metabolomics. 3 (2007) 211–221. https://doi.org/10.1007/s11306-007-0082-2.Proposed.

[10] F. García-Muñoz Rodrigo, A.L. Díez Recinos, A. García-Alix Pérez, J. Figueras Aloy, M. Vento Torres, Changes in perinatal care and outcomes in newborns at the limit of viability in Spain: the EPI-SEN Study, Neonatology. 107 (2015) 120–129. https://doi.org/10.1159/000368881.

[11] H. Blencowe, S. Cousens, M.Z. Oestergaard, D. Chou, A.-B. Moller, R. Narwal, A. Adler, C.V. Garcia, S. Rohde, L. Say, J.E. Lawn, National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications, The Lancet. 379 (2012) 2162–2172. https://doi.org/10.1016/S0140-6736(12)60820-4.

[12] J. Perin, A. Mulick, D. Yeung, F. Villavicencio, G. Lopez, K.L. Strong, D. Prieto-Merino, S. Cousens, R.E. Black, L. Liu, Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals, The Lancet Child & Adolescent Health. 6 (2022) 106–115. https://doi.org/10.1016/S2352-4642(21)00311-4.

[13] B.R. Vohr, B.B. Poindexter, A.M. Dusick, L.T. McKinley, R.D. Higgins, J.C. Langer, W.K. Poole, National Institute of Child Health and Human Development National Research Network, Persistent beneficial effects of breast milk ingested in the neonatal intensive care unit on outcomes of extremely low birth weight infants at 30 months of age, Pediatrics. 120 (2007) e953-959. https://doi.org/10.1542/peds.2006-3227.

[14] J. Meinzen-Derr, B. Poindexter, L. Wrage, A.L. Morrow, B. Stoll, E.F. Donovan, Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death, J Perinatol. 29 (2009) 57–62. https://doi.org/10.1038/jp.2008.117.

[15] M.C. Collado, M. Cernada, C. Baüerl, M. Vento, G. Pérez-Martínez, Microbial ecology and host-microbiota interactions during early life stages, Gut Microbes. 3 (2012) 352–365. https://doi.org/10.4161/gmic.21215.

[16] M.B. Belfort, P.J. Anderson, V.A. Nowak, K.J. Lee, C. Molesworth, D.K. Thompson, L.W. Doyle, T.E. Inder, Breast Milk Feeding, Brain Development, and Neurocognitive Outcomes: A 7-Year Longitudinal Study in Infants Born at Less Than 30 Weeks' Gestation, The Journal of Pediatrics. 177 (2016) 133-139.e1. https://doi.org/10.1016/j.jpeds.2016.06.045.

[17] O. Ballard, A.L. Morrow, Human Milk Composition: Nutrients and Bioactive Factors, Pediatr Clin North Am. 60 (2013) 49–74. https://doi.org/10.1016/j.pcl.2012.10.002.
[18] M.S. Kramer, R. Kakuma, Optimal duration of exclusive breastfeeding, Cochrane Database Syst Rev. (2012) CD003517. https://doi.org/10.1002/14651858.CD003517.pub2.
[19] J.Y. Meek, L. Noble, Section on Breastfeeding, Policy Statement: Breastfeeding and the Use of Human Milk, Pediatrics. 150 (2022) e2022057988. https://doi.org/10.1542/peds.2022-057988.
[20] M. Boland, Exclusive breastfeeding should continue to six months, Paediatr Child Health. 10 (2005) 148.
[21] C. Martin, M. Patel, S. Williams, H. Arora, B. Sims, Human breast milk-derived exosomes attenuate cell death in intestinal epithelial cells, Innate Immun. 24 (2018) 278–284. https://doi.org/10.1177/1753425918785715.
[22] A. Hock, H. Miyake, B. Li, C. Lee, L. Ermini, Y. Koike, Y. Chen, P. Määttänen, A. Zani, A. Pierro, Breast milk-derived exosomes promote intestinal epithelial cell growth, J Pediatr Surg. 52 (2017) 755–759. https://doi.org/10.1016/j.jpedsurg.2017.01.032.
[23] C. Corrado, S. Raimondo, A. Chiesi, F. Ciccia, G. De Leo, R. Alessandro, Exosomes as intercellular signaling organelles involved in health and disease: basic science and clinical applications, Int J Mol Sci. 14 (2013) 5338–5366. https://doi.org/10.3390/ijms14035338.
[24] J. Zempleni, A. Aguilar-Lozano, M. Sadri, S. Sukreet, S. Manca, D. Wu, F. Zhou, E. Mutai, Biological Activities of Extracellular Vesicles and Their Cargos from Bovine and Human Milk in Humans and Implications for Infants, The Journal of Nutrition. 147 (2017) 3–10. https://doi.org/10.3945/jn.116.238949.
[25] Q. Pham, P. Patel, B. Baban, J. Yu, J. Bhatia, Factors Affecting the Composition of Expressed Fresh Human Milk, Breastfeeding Medicine. 15 (2020) 551–558. https://doi.org/10.1089/bfm.2020.0195.
[26] American Academ of Pediatrics, Breastfeeding and the Use of Human Milk, Pediatrics. 129 (2012) e827–e841. https://doi.org/10.1542/peds.2011-3552.
[27] D. Poulimeneas, E. Bathrellou, G. Antonogeorgos, E. Mamalaki, M. Kouvari, J. Kuligowski, M. Gormaz, D.B. Panagiotakos, M. Yannakoulia, NUTRISHIELD Consortium, Feeding the preterm infant: an overview of the evidence, Int J Food Sci Nutr. 72 (2021) 4–13. https://doi.org/10.1080/09637486.2020.1754352.
[28] T.T. Colaizy, Effects of milk banking procedures on nutritional and bioactive components of donor human milk, Semin Perinatol. 45 (2021) 151382. https://doi.org/10.1016/j.semperi.2020.151382.

[29] S.M. Innis, Human milk and formula fatty acids, J Pediatr. 120 (1992) S56-61. https://doi.org/10.1016/s0022-3476(05)81237-5.

[30] C. Gao, J. Miller, P.F. Middleton, Y.-C. Huang, A.J. McPhee, R.A. Gibson, Changes to breast milk fatty acid composition during storage, handling and processing: A systematic review, Prostaglandins Leukot. Essent. Fatty Acids. 146 (2019) 1–10. https://doi.org/10.1016/j.plefa.2019.04.008.

[31] I. Nessel, M. Khashu, S.C. Dyall, The effects of storage conditions on long-chain polyunsaturated fatty acids, lipid mediators, and antioxidants in donor human milk - A review, Prostaglandins Leukot Essent Fatty Acids. 149 (2019) 8–17. https://doi.org/10.1016/j.plefa.2019.07.009.

[32] A. Wesolowska, E. Sinkiewicz-Darol, O. Barbarska, U. Bernatowicz-Lojko, M.K. Borszewska-Kornacka, J.B. van Goudoever, Innovative Techniques of Processing Human Milk to Preserve Key Components, Nutrients. 11 (2019). https://doi.org/10.3390/nu11051169.

[33] M. Keikha, M. Bahreynian, M. Saleki, R. Kelishadi, Macro- and Micronutrients of Human Milk Composition: Are They Related to Maternal Diet? A Comprehensive Systematic Review, Breastfeed Med. 12 (2017) 517–527. https://doi.org/10.1089/bfm.2017.0048.

[34] F. Bravi, F. Wiens, A. Decarli, A. Dal Pont, C. Agostoni, M. Ferraroni, Impact of maternal nutrition on breast-milk composition: a systematic review, The American Journal of Clinical Nutrition. 104 (2016) 646–662. https://doi.org/10.3945/ajcn.115.120881.

[35] M.L. Neuhouser, L. Tinker, P.A. Shaw, D. Schoeller, S.A. Bingham, L.V. Horn, S.A.A. Beresford, B. Caan, C. Thomson, S. Satterfield, L. Kuller, G. Heiss, E. Smit, G. Sarto, J. Ockene, M.L. Stefanick, A. Assaf, S. Runswick, R.L. Prentice, Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative, Am J Epidemiol. 167 (2008) 1247–1259. https://doi.org/10.1093/aje/kwn026.

[36] L. Bode, [Researchers discover a unique human milk oligosaccharide], Kinderkrankenschwester. 31 (2012) 214.

[37] K.M. Meyer, M. Mohammad, J. Ma, D. Chu, M. Haymond, K. Aagaard, 66: Maternal diet alters the breast milk microbiome and microbial gene content, American Journal of Obstetrics and Gynecology. 214 (2016) S47–S48. https://doi.org/10.1016/j.ajog.2015.10.084.

[38] G.D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y.-Y. Chen, S.A. Keilbaugh, M. Bewtra, D. Knights, W.A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L. Nessel, H. Li, F.D. Bushman, J.D. Lewis, Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes, Science. 334 (2011) 105–108. https://doi.org/10.1126/science.1208344.

[39] N. Redondo-Useros, E. Nova, N. González-Zancada, L.E. Díaz, S. Gómez-Martínez, A. Marcos, Microbiota and Lifestyle: A Special Focus on Diet, Nutrients. 12 (2020) 1776. https://doi.org/10.3390/nu12061776.
[40] L. Moles, D. Otaegui, The Impact of Diet on Microbiota Evolution and Human Health. Is Diet an Adequate Tool for Microbiota Modulation?, Nutrients. 12 (2020) 1654. https://doi.org/10.3390/nu12061654.
[41] K. Berding, K. Vlckova, W. Marx, H. Schellekens, C. Stanton, G. Clarke, F. Jacka, T.G. Dinan, J.F. Cryan, Diet and the Microbiota–Gut–Brain Axis: Sowing the Seeds of Good Mental Health, Adv Nutr. 12 (2021) 1239–1285. https://doi.org/10.1093/advances/nmaa181.
[42] W.J. Dahl, D. Rivero Mendoza, J.M. Lambert, Diet, nutrients and the microbiome, Prog Mol Biol Transl Sci. 171 (2020) 237–263. https://doi.org/10.1016/bs.pmbts.2020.04.006.
[43] S. Bibbò, G. Ianiro, V. Giorgio, F. Scaldaferri, L. Masucci, A. Gasbarrini, G. Cammarota, The role of diet on gut microbiota composition, (2016).
[44] A.A. Kolodziejczyk, D. Zheng, E. Elinav, Diet-microbiota interactions and personalized nutrition, Nat Rev Microbiol. 17 (2019) 742–753. https://doi.org/10.1038/s41579-019-0256-8.
[45] B. Chassaing, A.T. Gewirtz, Gut Microbiome and Metabolism, in: Physiology of the Gastrointestinal Tract, Elsevier, 2018: pp. 775–793. https://doi.org/10.1016/B978-0-12-809954-4.00035-9.
[46] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, J. Xia, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, Nucleic Acids Res. 46 (2018) W486–W494. https://doi.org/10.1093/nar/gky310.
[47] M.R. Molenaar, A. Jeucken, T.A. Wassenaar, C.H.A. van de Lest, J.F. Brouwers, J.B. Helms, LION/web: a web-based ontology enrichment tool for lipidomic data analysis, GigaScience. 8 (2019) giz061. https://doi.org/10.1093/gigascience/giz061.
[48] L. Pirhaji, P. Milani, M. Leidl, T. Curran, J. Avila-Pacheco, C.B. Clish, F.M. White, A. Saghatelian, E. Fraenkel, Revealing disease-associated pathways by network integration of untargeted metabolomics, Nat Methods. 13 (2016) 770–776. https://doi.org/10.1038/nmeth.3940.
[49] C.G. Victora, R. Bahl, A.J.D. Barros, G.V.A. França, S. Horton, J. Krasevec, S. Murch, M.J. Sankar, N. Walker, N.C. Rollins, Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect, The Lancet. 387 (2016) 475–490. https://doi.org/10.1016/S0140-6736(15)01024-7.
[50] M. Toscano, R. De Grandi, E. Grossi, L. Drago, Role of the Human Breast Milk-Associated Microbiota on the Newborns' Immune System:

A Mini Review, Front. Microbiol. 8 (2017) 2100. https://doi.org/10.3389/fmicb.2017.02100.

[51] N.J. Andreas, B. Kampmann, K. Mehring Le-Doare, Human breast milk: A review on its composition and bioactivity, Early Hum Dev. 91 (2015) 629–635. https://doi.org/10.1016/j.earlhumdev.2015.08.013.

[52] C. Agostoni, G. Buonocore, V.P. Carnielli, M. De Curtis, D. Darmaun, T. Decsi, M. Domellöf, N.D. Embleton, C. Fusch, O. Genzel-Boroviczeny, O. Goulet, S.C. Kalhan, S. Kolacek, B. Koletzko, A. Lapillonne, W. Mihatsch, L. Moreno, J. Neu, B. Poindexter, J. Puntis, G. Putet, J. Rigo, A. Riskin, B. Salle, P. Sauer, R. Shamir, H. Szajewska, P. Thureen, D. Turck, J.B. van Goudoever, E.E. Ziegler, ESPGHAN Committee on Nutrition, Enteral nutrient supply for preterm infants: commentary from the European Society of Paediatric Gastroenterology, Hepatology and Nutrition Committee on Nutrition, J Pediatr Gastroenterol Nutr. 50 (2010) 85–91. https://doi.org/10.1097/MPG.0b013e3181adaee0.

[53] B. Lönnerdal, P. Erdmann, S.K. Thakkar, J. Sauser, F. Destaillats, Longitudinal evolution of true protein, amino acids and bioactive proteins in breast milk: a developmental perspective, J Nutr Biochem. 41 (2017) 1–11. https://doi.org/10.1016/j.jnutbio.2016.06.001.

[54] B. Koletzko, M. Rodriguez-Palmero, H. Demmelmair, N. Fidler, R. Jensen, T. Sauerwald, Physiological aspects of human milk lipids, Early Human Development. 65 (2001) S3–S18. https://doi.org/10.1016/S0378-3782(01)00204-3.

[55] L. Drago, M. Toscano, R. De Grandi, E. Grossi, E.M. Padovani, D.G. Peroni, Microbiota network and mathematic microbe mutualism in colostrum and mature milk collected in two different geographic areas: Italy versus Burundi, ISME J. 11 (2017) 875–884. https://doi.org/10.1038/ismej.2016.183.

[56] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J.E. Lawn, S. Cousens, C. Mathers, R.E. Black, Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals, Lancet. 388 (2016) 3027–3035. https://doi.org/10.1016/S0140-6736(16)31593-8.

[57] R. Schanler, R. Shulman, C. Lau, Feeding Strategies for Premature Infants: Beneficial Outcomes of Feeding Fortified Human Milk Versus Preterm Formula, Pediatrics. 103 (1999) 1150–7. https://doi.org/10.1542/peds.103.6.1150.

[58] P.M. Sisk, C.A. Lovelady, R.G. Dillard, K.J. Gruber, T.M. O'Shea, Early human milk feeding is associated with a lower risk of necrotizing enterocolitis in very low birth weight infants, J Perinatol. 27 (2007) 428–433. https://doi.org/10.1038/sj.jp.7211758.

[59] A. Maayan-Metzger, S. Avivi, I. Schushan-Eisen, J. Kuint, Human milk versus formula feeding among preterm infants: short-term outcomes, Am J Perinatol. 29 (2012) 121–126. https://doi.org/10.1055/s-0031-1295652.
[60] B.R. Vohr, B.B. Poindexter, A.M. Dusick, L.T. McKinley, L.L. Wright, J.C. Langer, W.K. Poole, NICHD Neonatal Research Network, Beneficial effects of breast milk in the neonatal intensive care unit on the developmental outcome of extremely low birth weight infants at 18 months of age, Pediatrics. 118 (2006) e115-123. https://doi.org/10.1542/peds.2005-2382.
[61] E.B. Isaacs, B.R. Fischl, B.T. Quinn, W.K. Chong, D.G. Gadian, A. Lucas, Impact of breast milk on intelligence quotient, brain size, and white matter development, Pediatr Res. 67 (2010) 357–362. https://doi.org/10.1203/PDR.0b013e3181d026da.
[62] T. Okamoto, M. Shirai, M. Kokubo, S. Takahashi, M. Kajino, M. Takase, H. Sakata, J. Oki, Human milk reduces the risk of retinal detachment in extremely low-birthweight infants, Pediatr Int. 49 (2007) 894–897. https://doi.org/10.1111/j.1442-200X.2007.02483.x.
[63] M.B. Belfort, S.L. Rifas-Shiman, T. Sullivan, C.T. Collins, A.J. McPhee, P. Ryan, K.P. Kleinman, M.W. Gillman, R.A. Gibson, M. Makrides, Infant growth before and after term: effects on neurodevelopment in preterm infants, Pediatrics. 128 (2011) e899-906. https://doi.org/10.1542/peds.2011-0282.
[64] P.P. Meier, A.L. Patel, A. Esquerra-Zwiers, Donor Human Milk Update: Evidence, Mechanisms and Priorities for Research and Practice, J Pediatr. 180 (2017) 15–21. https://doi.org/10.1016/j.jpeds.2016.09.027.
[65] J. Calvo, N.R. García Lara, M. Gormaz, M. Peña, M.J. Martínez Lorenzo, P. Ortiz Murillo, J.M. Brull Sabaté, C.M. Samaniego, A. Gayà, Recomendaciones para la creación y el funcionamiento de los bancos de leche materna en España, An Pediatr (Barc). 89 (2018) 65.e1-65.e6. https://doi.org/10.1016/j.anpedi.2018.01.010.
[66] R.J. Schanler, C. Lau, N.M. Hurst, E.O. Smith, Randomized trial of donor human milk versus preterm formula as substitutes for mothers' own milk in the feeding of extremely premature infants, Pediatrics. 116 (2005) 400–406. https://doi.org/10.1542/peds.2004-1974.
[67] M. Quigley, N.D. Embleton, W. McGuire, Formula versus donor breast milk for feeding preterm or low birth weight infants, Cochrane Database Syst Rev. 2019 (2019) CD002971. https://doi.org/10.1002/14651858.CD002971.pub5.
[68] J. Bitman, L. Wood, M. Hamosh, P. Hamosh, N.R. Mehta, Comparison of the lipid composition of breast milk from mothers of term and preterm

infants, The American Journal of Clinical Nutrition. 38 (1983) 300–312. https://doi.org/10.1093/ajcn/38.2.300.

[69] T. Velonà, L. Abbiati, B. Beretta, A. Gaiaschi, U. Flaúto, P. Tagliabue, C.L. Galli, P. Restani, Protein Profiles in Breast Milk from Mothers Delivering Term and Preterm Babies, Pediatr Res. 45 (1999) 658–663. https://doi.org/10.1203/00006450-199905010-00008.

[70] Z. Zhang, A.S. Adelman, D. Rai, J. Boettcher, B. Lőnnerdal, Amino Acid Profiles in Term and Preterm Human Milk through Lactation: A Systematic Review, Nutrients. 5 (2013) 4800–4821. https://doi.org/10.3390/nu5124800.

[71] S. Austin, C.A. De Castro, N. Sprenger, A. Binia, M. Affolter, C.L. Garcia-Rodenas, L. Beauport, J.-F. Tolsa, C.J.F. Fumeaux, Human Milk Oligosaccharides in the Milk of Mothers Delivering Term versus Preterm Infants, Nutrients. 11 (2019) 1282. https://doi.org/10.3390/nu11061282.

[72] M.A. Underwood, Human milk for the premature infant, Pediatr Clin North Am. 60 (2013) 189–207. https://doi.org/10.1016/j.pcl.2012.09.008.

[73] C.A. Butts, D.I. Hedderley, T.D. Herath, G. Paturi, S. Glyn-Jones, F. Wiens, B. Stahl, P. Gopal, Human Milk Composition and Dietary Intakes of Breastfeeding Women of Different Ethnicity from the Manawatu-Wanganui Region of New Zealand, Nutrients. 10 (2018) 1231. https://doi.org/10.3390/nu10091231.

[74] J.T. Brenna, B. Varamini, R.G. Jensen, D.A. Diersen-Schade, J.A. Boettcher, L.M. Arterburn, Docosahexaenoic and arachidonic acid concentrations in human breast milk worldwide, Am J Clin Nutr. 85 (2007) 1457–1464. https://doi.org/10.1093/ajcn/85.6.1457.

[75] Q. Xi, W. Liu, T. Zeng, X. Chen, T. Luo, Z. Deng, Effect of Different Dietary Patterns on Macronutrient Composition in Human Breast Milk: A Systematic Review and Meta-Analysis, Nutrients. 15 (2023) 485. https://doi.org/10.3390/nu15030485.

[76] S. Donovan, K. Dewey, R. Novotny, J. Stang, E. Taveras, R. Kleinman, R. Raghavan, J. Nevins, S. Scinto-Madonich, J.H. Kim, N. Terry, G. Butera, J. Obbagy, Dietary Patterns during Lactation and Human Milk Composition and Quantity: A Systematic Review, USDA Nutrition Evidence Systematic Review, Alexandria (VA), 2020. http://www.ncbi.nlm.nih.gov/books/NBK578493/ (accessed March 1, 2023).

[77] D.M. Chu, K.M. Meyer, A.L. Prince, K.M. Aagaard, Impact of maternal nutrition in pregnancy and lactation on offspring gut microbial composition and function, Gut Microbes. 7 (2016) 459–470. https://doi.org/10.1080/19490976.2016.1241357.

[78] J.M. Conway, L.A. Ingwersen, A.J. Moshfegh, Accuracy of dietary recall using the USDA five-step multiple-pass method in men: An observational validation study, Journal of the American Dietetic Association. 104 (2004) 595–603. https://doi.org/10.1016/j.jada.2004.01.007.
[79] D.B. Panagiotakos, C. Pitsavos, C. Stefanadis, Dietary patterns: A Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk, Nutrition, Metabolism and Cardiovascular Diseases. 16 (2006) 559–568. https://doi.org/10.1016/j.numecd.2005.08.006.
[80] G.D. Zimet, N.W. Dahlem, S.G. Zimet, G.K. Farley, The Multidimensional Scale of Perceived Social Support, Journal of Personality Assessment. 52 (1988) 30–41. https://doi.org/10.1207/s15327752jpa5201_2.
[81] D. Olsen, J. Portner, Y. Lavee, Family adaptability and cohesion evaluation scales (FACES-II), University of Minnesota, Minneapolis. 32 (1985) 10–22.
[82] E.-H. Lee, Review of the psychometric evidence of the perceived stress scale, Asian Nurs Res (Korean Soc Nurs Sci). 6 (2012) 121–127. https://doi.org/10.1016/j.anr.2012.08.004.
[83] C. Spielberger, R. Gorsuch, R. Lushene, P. Vagg, G. Jacobs, Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2), 1983.
[84] J.L. Cox, J.M. Holden, R. Sagovsky, Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale, Br J Psychiatry. 150 (1987) 782–786. https://doi.org/10.1192/bjp.150.6.782.
[85] A.T. Beck, W.Y. Rial, K. Rickels, Short form of depression inventory: cross-validation, Psychol Rep. 34 (1974) 1184–1186.
[86] J. Davidson, R. Smith, Traumatic experiences in psychiatric outpatients, Journal of Traumatic Stress. 3 (1990) 459–475. https://doi.org/10.1002/jts.2490030314.
[87] J. Bobes, A. Calcedo-Barba, M. García, M. François, F. Rico-Villademoros, M.P. González, M.T. Bascarán, M. Bousoño, Grupo Espanol de Trabajo para el Estudio del Trastornostres Postraumático, Evaluation of the psychometric properties of the Spanish version of 5 questionnaires for the evaluation of post-traumatic stress syndrome, Actas Esp Psiquiatr. 28 (2000) 207–218.
[88] I.F. Brockington, C. Fraser, D. Wilson, The Postpartum Bonding Questionnaire: a validation, Arch Womens Ment Health. 9 (2006) 233–242. https://doi.org/10.1007/s00737-006-0132-1.
[89] G.R. Rivas, I. Arruabarrena, J. de Paúl, Parenting Stress Index-Short Form: psychometric properties of the Spanish version in mothers of

children aged 0 to 8 years, Psy. Intervention. 30 (2020) 27–34. https://doi.org/10.5093/pi2020a14.

[90] J. Squires, E. Twombly, D. Bricker, L. Potter, Technical report. ASQ-3^{TM} user's guide, Brookes Publishing, 2009.

[91] S.P. Putnam, A.L. Helbig, M.A. Gartstein, M.K. Rothbart, E. Leerkes, Development and assessment of short and very short forms of the infant behavior questionnaire-revised, J Pers Assess. 96 (2014) 445–458. https://doi.org/10.1080/00223891.2013.841171.

[92] G.H. Roid, J.L. Sampers, Merrill-Palmer-Revised Scales of Development (MPR) [Database record], APA PsycTests. (2004). https://doi.org/10.1037/t06030-000.

[93] Á. Jácome, R. Jiménez, Validation of the Iowa Infant Feeding Attitude Scale, Pediatria. 47 (2014) 77–82. https://doi.org/10.1016/S0120-4912(15)30143-9.

[94] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, P.C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses, The American Journal of Human Genetics. 81 (2007) 559–575. https://doi.org/10.1086/519795.

[95] C. Bycroft, C. Freeman, D. Petkova, G. Band, L.T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data, Nature. 562 (2018) 203–209. https://doi.org/10.1038/s41586-018-0579-z.

[96] S. Kouhounde, K. Adéoti, M. Mounir, A. Giusti, P. Refinetti, A. Otu, E. Effa, B. Ebenso, V.O. Adetimirin, J.M. Barceló, O. Thiare, H.N. Rabetafika, H.L. Razafindralambo, Applications of Probiotic-Based Multi-Components to Human, Animal and Ecosystem Health: Concepts, Methodologies, and Action Mechanisms, Microorganisms. 10 (2022) 1700. https://doi.org/10.3390/microorganisms10091700.

[97] P.O. Ekstrøm, D.J. Warren, W.G. Thilly, Separation principles of cycling temperature capillary electrophoresis, Electrophoresis. 33 (2012) 1162–1168. https://doi.org/10.1002/elps.201100550.

[98] V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, Á. Solaz-García, I. Lara-Cantón, A. Pinilla-Gonzalez, M. Gormaz, M. Vento, J. Kuligowski, G. Quintás, GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers, Clinica Chimica Acta. 532 (2022) 172–180. https://doi.org/10.1016/j.cca.2022.05.005.

[99] X. Zheng, Y. Qiu, W. Zhong, S. Baxter, M. Su, Q. Li, G. Xie, B.M. Ore, S. Qiao, M.D. Spencer, S.H. Zeisel, Z. Zhou, A. Zhao, W. Jia, A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids, Metabolomics. 9 (2013) 818–827. https://doi.org/10.1007/s11306-013-0500-6.

[100] M.H. Sarafian, M.R. Lewis, A. Pechlivanis, S. Ralphs, M.J.W. McPhail, V.C. Patel, M.-E. Dumas, E. Holmes, J.K. Nicholson, Bile Acid Profiling and Quantification in Biofluids Using Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry, Analytical Chemistry. 87 (2015) 9662–9670. https://doi.org/10.1021/acs.analchem.5b01556.

[101] R. González-Domínguez, M. Urpi-Sarda, O. Jáuregui, P.W. Needs, P.A. Kroon, C. Andrés-Lacueva, Quantitative Dietary Fingerprinting (QDF)—A Novel Tool for Comprehensive Dietary Assessment Based on Urinary Nutrimetabolomics, Journal of Agricultural and Food Chemistry. (2019). https://doi.org/10.1021/acs.jafc.8b07023.

[102] V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, M. Gormaz, M. Vento, M. Karipidou, D. Poulimeneas, E. Mamalaki, E. Bathrellou, J. Kuligowski, Joint Microbiota Activity and Dietary Assessment through Urinary Biomarkers by LC-MS/MS, Nutrients. 15 (2023) 1894. https://doi.org/10.3390/nu15081894.

[103] I. Ten-Doménech, T. Martínez-Sena, M. Moreno-Torres, J.D. Sanjuan-Herráez, J.V. Castell, A. Parra-Llorca, M. Vento, G. Quintás, J. Kuligowski, Comparing Targeted vs. Untargeted MS2 Data-Dependent Acquisition for Peak Annotation in LC-MS Metabolomics, Metabolites. 10 (2020) 126. https://doi.org/10.3390/metabo10040126.

[104] Miris. "HMA User Manual"., (n.d.). http://www.mirissolutions.com/support/user-manuals (accessed June 27, 2022).

[105] C. Cruz-Hernandez, S. Goeuriot, F. Giuffrida, S.K. Thakkar, F. Destaillats, Direct quantification of fatty acids in human milk by gas chromatography, Journal of Chromatography A. 1284 (2013) 174–179. https://doi.org/10.1016/j.chroma.2013.01.094.

[106] I. Ten-Doménech, V. Ramos-Garcia, M. Moreno-Torres, A. Parra-Llorca, M. Gormaz, M. Vento, J. Kuligowski, G. Quintás, The effect of Holder pasteurization on the lipid and metabolite composition of human milk, Food Chemistry. 384 (2022) 132581. https://doi.org/10.1016/j.foodchem.2022.132581.

[107] J. Lu, E.L. Frank, Rapid HPLC measurement of thiamine and its phosphate esters in whole blood, Clin Chem. 54 (2008) 901–906. https://doi.org/10.1373/clinchem.2007.099077.

[108] J. Yang, G.E. Cleland, K.L. Organtini, Determination of Vitamin D and Previtamin D in Food Products, Waters Application Note. (n.d.).

[109] J.M. Oberson, S. Bénet, K. Redeuil, E. Campos-Giménez, Quantitative analysis of vitamin D and its main metabolites in human milk by supercritical fluid chromatography coupled to tandem mass spectrometry, Anal Bioanal Chem. 412 (2020) 365–375. https://doi.org/10.1007/s00216-019-02248-5.

[110] A. Levêques, J.-M. Oberson, E.A. Tissot, K. Redeuil, S.K. Thakkar, E. Campos-Giménez, Quantification of Vitamins A, E, and K and Carotenoids in Submilliliter Volumes of Human Milk, J AOAC Int. 102 (2019) 1059–1068. https://doi.org/10.5740/jaoacint.19-0016.

[111] E. Manoury, K. Jourdon, P. Boyaval, P. Fourcassié, Quantitative measurement of vitamin K2 (menaquinones) in various fermented dairy products using a reliable high-performance liquid chromatography method, Journal of Dairy Science. 96 (2013) 1335–1346. https://doi.org/10.3168/jds.2012-5494.

[112] Y. Xue, E. Campos-Giménez, K.M. Redeuil, A. Lévèques, L. Actis-Goretta, G. Vinyes-Pares, Y. Zhang, P. Wang, S.K. Thakkar, Concentrations of Carotenoids and Tocopherols in Breast Milk from Urban Chinese Mothers and Their Associations with Maternal Characteristics: A Cross-Sectional Study, Nutrients. 9 (2017) 1229. https://doi.org/10.3390/nu9111229.

[113] C.A. Remoroza, T.D. Mak, M.L.A. De Leoz, Y.A. Mirokhin, S.E. Stein, Creating a Mass Spectral Reference Library for Oligosaccharides in Human Milk, Anal Chem. 90 (2018) 8977–8988. https://doi.org/10.1021/acs.analchem.8b01176.

[114] V. Navarro-Esteve, M.M. Cascant-Vilaplana, A. Zöchner, V. Ramos-Garcia, M. Roca, A. Parra-Llorca, A. Moreno-Giménez, L. Campos-Berga, M. Vaya, M. Vento, P. Sáenz-González, M. Gormaz, I. Ten-Doménech, J. Kuligowski, G. Quintás, A Hybrid Liquid Chromatography – Mass Spectrometry method for comprehensive Human Milk Oligosaccharide Screening, (Under revision). (n.d.).

[115] Hyuck Ho Son, Wan Soo Yun, Sung-Hee Cho, Development and validation of an LC-MS/MS method for profiling 39 urinary steroids (estrogens, androgens, corticoids, and progestins), Biomedical Chromatography. 34 (2020) e4723–e4739. https://doi.org/10.1002/bmc.4723.

[116] A. Gomez-Gomez, J. Miranda, G. Feixas, A. Arranz Betegon, F. Crispi, E. Gratacós, O.J. Pozo, Determination of the steroid profile in alternative matrices by liquid chromatography tandem mass spectrometry, The Journal of Steroid Biochemistry and Molecular Biology. 197 (2020) 105520. https://doi.org/10.1016/j.jsbmb.2019.105520.

[117] C.K. Akhgar, G. Ramer, M. Żbik, A. Trajnerowicz, J. Pawluczyk, A. Schwaighofer, B. Lendl, The Next Generation of IR Spectroscopy: EC-QCL-Based Mid-IR Transmission Spectroscopy of Proteins with Balanced Detection, Anal. Chem. 92 (2020) 9901–9907. https://doi.org/10.1021/acs.analchem.0c01406.
[118] A. Schwaighofer, C.K. Akhgar, B. Lendl, Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 253 (2021) 119563. https://doi.org/10.1016/j.saa.2021.119563.
[119] C.K. Akhgar, V. Nürnberger, M. Nadvornik, V. Ramos-Garcia, I. Ten-Doménech, J. Kuligowski, A. Schwaighofer, E. Rosenberg, B. Lendl, Fatty Acid Determination in Human Milk Using Attenuated Total Reflection Infrared Spectroscopy and Solvent-Free Lipid Separation, Appl Spectrosc. 76 (2022) 730–736. https://doi.org/10.1177/00037028211065502.
[120] C.K. Akhgar, V. Ramos-Garcia, V. Nürnberger, A. Moreno-Giménez, J. Kuligowski, E. Rosenberg, A. Schwaighofer, B. Lendl, Solvent-Free Lipid Separation and Attenuated Total Reflectance Infrared Spectroscopy for Fast and Green Fatty Acid Profiling of Human Milk, Foods. 11 (2022) 3906. https://doi.org/10.3390/foods11233906.
[121] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society: Series B (Methodological). 57 (1995) 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[122] L. Brennan, 7 - Metabotyping: moving towards personalised nutrition, in: J.-L. Sébédio, L. Brennan (Eds.), Metabolomics as a Tool in Nutrition Research, Woodhead Publishing, 2015: pp. 137–144. https://doi.org/10.1016/B978-1-78242-084-2.00007-1.
[123] L. Lafay, L. Mennen, A. Basdevant, M.A. Charles, J.M. Borys, E. Eschwège, M. Romon, Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé (FLVS) study, Int J Obes Relat Metab Disord. 24 (2000) 1500–1506. https://doi.org/10.1038/sj.ijo.0801392.
[124] M. Jenab, N. Slimani, M. Bictash, P. Ferrari, S.A. Bingham, Biomarkers in nutritional epidemiology: applications, needs and new horizons, Hum Genet. 125 (2009) 507–525. https://doi.org/10.1007/s00439-009-0662-5.
[125] A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L.O. Dragsted, J. Draper, S.M. Rappaport, J.J.J. van der Hooft, D.S. Wishart, The food metabolome: a window over dietary exposure, Am J Clin Nutr. 99 (2014) 1286–1308. https://doi.org/10.3945/ajcn.113.076133.

[126] P. Maruvada, J.W. Lampe, D.S. Wishart, D. Barupal, D.N. Chester, D. Dodd, Y. Djoumbou-Feunang, P.C. Dorrestein, L.O. Dragsted, J. Draper, L.C. Duffy, J.T. Dwyer, N.J. Emenaker, O. Fiehn, R.E. Gerszten, F. B Hu, R.W. Karp, D.M. Klurfeld, M.R. Laughlin, A.R. Little, C.J. Lynch, S.C. Moore, H.L. Nicastro, D.M. O'Brien, J.M. Ordovás, S.K. Osganian, M. Playdon, R. Prentice, D. Raftery, N. Reisdorph, H.M. Roche, S.A. Ross, S. Sang, A. Scalbert, P.R. Srinivas, S.H. Zeisel, Perspective: Dietary Biomarkers of Intake and Exposure—Exploration with Omics Approaches, Advances in Nutrition. (2019) nmz075. https://doi.org/10.1093/advances/nmz075.

[127] M.M. Ulaszewska, C.H. Weinert, A. Trimigno, R. Portmann, C. Andres Lacueva, R. Badertscher, L. Brennan, C. Brunius, A. Bub, F. Capozzi, M. Cialiè Rosso, C.E. Cordero, H. Daniel, S. Durand, B. Egert, P.G. Ferrario, E.J.M. Feskens, P. Franceschi, M. Garcia-Aloy, F. Giacomoni, P. Giesbertz, R. González-Domínguez, K. Hanhineva, L.Y. Hemeryck, J. Kopka, S.E. Kulling, R. Llorach, C. Manach, F. Mattivi, C. Migné, L.H. Münger, B. Ott, G. Picone, G. Pimentel, E. Pujos-Guillot, S. Riccadonna, M.J. Rist, C. Rombouts, J. Rubert, T. Skurk, P.S.C. Sri Harsha, L. Van Meulebroek, L. Vanhaecke, R. Vázquez-Fresno, D. Wishart, G. Vergères, Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies, Molecular Nutrition & Food Research. 63 (2019) 1800384. https://doi.org/10.1002/mnfr.201800384.

[128] R. González-Domínguez, ed., Mass Spectrometry for Metabolomics, Springer US, New York, NY, 2023. https://doi.org/10.1007/978-1-0716-2699-3.

[129] E.D. Clarke, M.E. Rollo, K. Pezdirc, C.E. Collins, R.L. Haslam, Urinary biomarkers of dietary intake: a review, Nutrition Reviews. 78 (2020) 364–381. https://doi.org/10.1093/nutrit/nuz048.

[130] N. Bar, T. Korem, O. Weissbrod, D. Zeevi, D. Rothschild, S. Leviatan, N. Kosower, M. Lotan-Pompan, A. Weinberger, C.I. Le Roy, C. Menni, A. Visconti, M. Falchi, T.D. Spector, J. Adamski, P.W. Franks, O. Pedersen, E. Segal, A reference map of potential determinants for the human serum metabolome, Nature. 588 (2020) 135–140. https://doi.org/10.1038/s41586-020-2896-2.

[131] L. Chen, D.V. Zhernakova, A. Kurilshikov, S. Andreu-Sánchez, D. Wang, H.E. Augustijn, A. Vich Vila, Lifelines Cohort Study, R.K. Weersma, M.H. Medema, M.G. Netea, F. Kuipers, C. Wijmenga, A. Zhernakova, J. Fu, Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome, Nat Med. (2022). https://doi.org/10.1038/s41591-022-02014-8.

[132] V.E. Hedrick, A.M. Dietrich, P.A. Estabrooks, J. Savla, E. Serrano, B.M. Davy, Dietary biomarkers: advances, limitations and future directions, Nutr J. 11 (2012) 109. https://doi.org/10.1186/1475-2891-11-109.
[133] M.C. Playdon, J.N. Sampson, A.J. Cross, R. Sinha, K.A. Guertin, K.A. Moy, N. Rothman, M.L. Irwin, S.T. Mayne, R. Stolzenberg-Solomon, S.C. Moore, Comparing metabolite profiles of habitual diet in serum and urine, Am J Clin Nutr. 104 (2016) 776–789. https://doi.org/10.3945/ajcn.116.135301.
[134] J.-L. Sébédio, Metabolomics, Nutrition, and Potential Biomarkers of Food Quality, Intake, and Health Status, in: Advances in Food and Nutrition Research, Elsevier, 2017: pp. 83–116. https://doi.org/10.1016/bs.afnr.2017.01.001.
[135] E.M. Brouwer-Brolsma, L. Brennan, C.A. Drevon, H. van Kranen, C. Manach, L.O. Dragsted, H.M. Roche, C. Andres-Lacueva, S.J.L. Bakker, J. Bouwman, F. Capozzi, S. De Saeger, T.E. Gundersen, M. Kolehmainen, S.E. Kulling, R. Landberg, J. Linseisen, F. Mattivi, R.P. Mensink, C. Scaccini, T. Skurk, I. Tetens, G. Vergeres, D.S. Wishart, A. Scalbert, E.J.M. Feskens, Combining traditional dietary assessment methods with novel metabolomics techniques: Nutrition-Society Summer Meeting / Conference on New Technology in Nutrition Research and Practice / Symposium 2 on Use of Biomarkers in Dietary Assessment and Dietary Exposure, Proceedings of the Nutrition Society. 76 (2017) 619–627. https://doi.org/10.1017/S0029665117003949.
[136] L. Gu, M. Laly, H.C. Chang, R.L. Prior, N. Fang, M.J.J. Ronis, T.M. Badger, Isoflavone conjugates are underestimated in tissues using enzymatic hydrolysis, J Agric Food Chem. 53 (2005) 6858–6863. https://doi.org/10.1021/jf050802j.
[137] P. Quifer-Rada, M. Martínez-Huélamo, R.M. Lamuela-Raventos, Is enzymatic hydrolysis a reliable analytical strategy to quantify glucuronidated and sulfated polyphenol metabolites in human fluids?, Food Funct. 8 (2017) 2419–2424. https://doi.org/10.1039/c7fo00558j.
[138] R.P. Feliciano, E. Mecha, M.R. Bronze, A. Rodriguez-Mateos, Development and validation of a high-throughput micro solid-phase extraction method coupled with ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry for rapid identification and quantification of phenolic metabolites in human plasma and urine, Journal of Chromatography A. 1464 (2016) 21–31. https://doi.org/10.1016/j.chroma.2016.08.027.
[139] J.P.E. Spencer, M.M. Abd El Mohsen, A.-M. Minihane, J.C. Mathers, Biomarkers of the intake of dietary polyphenols: strengths, limitations

and application in nutrition research, Br J Nutr. 99 (2008) 12–22. https://doi.org/10.1017/S0007114507798938.

[140] V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, L. Campos-Berga, A. Parra-Llorca, A. Ramón-Beltrán, M.J. Vaya, F. Mohareb, C. Molitor, P. Refinetti, A. Silva, L.A. Rodrigues, S. Rezzi, A.C.C. Hodgson, S. Canarelli, E. Bathrellou, E. Mamalaki, M. Karipidou, D. Poulimeneas, M. Yannakoulia, C.K. Akhgar, A. Schwaighofer, B. Lendl, J. Karrer, D. Migliorelli, S. Generelli, M. Gormaz, M. Vasileiadis, J. Kuligowski, M. Vento, Fact-based nutrition for infants and lactating mothers—The NUTRISHIELD study, Front. Pediatr. 11 (2023) 1130179. https://doi.org/10.3389/fped.2023.1130179.

[141] D.J. Hinojosa Nogueira, D. Romero Molina, M.J. Giménez Asensio, B. González Alzaga, I. López Flores, S. Pastoriza de la Cueva, J.Á. Rufián Henares, A.F. Hernández Jérez, M. Lacasaña, Validity and Reproducibility of a Food Frequency Questionnaire to Assess Nutrients Intake of Pregnant Women in the South-East of Spain, (2021). https://doi.org/10.3390/nu13093032.

[142] Healthy eating during pregnancy and breastfeeding - booklet for mothers, World Health Organization. (2021).

[143] A.J. Lloyd, M. Beckmann, G. Favé, J.C. Mathers, J. Draper, Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption, British Journal of Nutrition. 106 (2011) 812–824. https://doi.org/10.1017/S0007114511001164.

[144] W. Cheung, P. Keski-Rahkonen, N. Assi, P. Ferrari, H. Freisling, S. Rinaldi, N. Slimani, R. Zamora-Ros, M. Rundle, G. Frost, H. Gibbons, E. Carr, L. Brennan, A.J. Cross, V. Pala, S. Panico, C. Sacerdote, D. Palli, R. Tumino, T. Kühn, R. Kaaks, H. Boeing, A. Floegel, F. Mancini, M.-C. Boutron-Ruault, L. Baglietto, A. Trichopoulou, A. Naska, P. Orfanos, A. Scalbert, A metabolomic study of biomarkers of meat and fish intake, The American Journal of Clinical Nutrition. 105 (2017) 600–608. https://doi.org/10.3945/ajcn.116.146639.

[145] K.S. Krogholm, L. Bredsdorff, S. Alinia, T. Christensen, S.E. Rasmussen, L.O. Dragsted, Free fruit at workplace intervention increases total fruit intake: a validation study using 24 h dietary recall and urinary flavonoid excretion, European Journal of Clinical Nutrition. 64 (2010) 1222–1228. https://doi.org/10.1038/ejcn.2010.130.

[146] H. Gibbons, B.A. McNulty, A.P. Nugent, J. Walton, A. Flynn, M.J. Gibney, L. Brennan, A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake, The American Journal of Clinical Nutrition. 101 (2015) 471–477. https://doi.org/10.3945/ajcn.114.095604.

[147] A.J. Cross, J.M. Major, R. Sinha, Urinary biomarkers of meat consumption, Cancer Epidemiol Biomarkers Prev. 20 (2011) 1107–1111. https://doi.org/10.1158/1055-9965.EPI-11-0048.
[148] L.G. Rasmussen, H. Winning, F. Savorani, H. Toft, T.M. Larsen, L.O. Dragsted, A. Astrup, S.B. Engelsen, Assessment of the Effect of High or Low Protein Diet on the Human Urine Metabolome as Measured by NMR, Nutrients. 4 (2012) 112–131. https://doi.org/10.3390/nu4020112.
[149] J.G. Jacobsen, L.H. Smith, Biochemistry and physiology of taurine and taurine derivatives., Physiological Reviews. 48 (1968) 424–511. https://doi.org/10.1152/physrev.1968.48.2.424.
[150] M. Garcia-Aloy, M. Rabassa, P. Casas-Agustench, N. Hidalgo-Liberona, R. Llorach, C. Andres-Lacueva, 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. 69 (2017) 220–229. https://doi.org/10.1016/j.tifs.2017.04.013.
[151] L. Brennan, Metabolomics: a tool to aid dietary assessment in nutrition, Current Opinion in Food Science. 16 (2017) 96–99. https://doi.org/10.1016/j.cofs.2017.09.003.
[152] G. Praticò, Q. Gao, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, S.H. Pedapati, L.A. Afman, D.S. Wishart, R. Vázquez-Fresno, C. Andres-Lacueva, M. Garcia-Aloy, H. Verhagen, E.J.M. Feskens, L.O. Dragsted, Guidelines for Biomarker of Food Intake Reviews (BFIRev): how to conduct an extensive literature search for biomarker of food intake discovery, Genes Nutr. 13 (2018) 3. https://doi.org/10.1186/s12263-018-0592-8.
[153] S.C. Moore, C.E. Matthews, J.N. Sampson, R.Z. Stolzenberg-Solomon, W. Zheng, Q. Cai, Y.T. Tan, W.-H. Chow, B.-T. Ji, D.K. Liu, Q. Xiao, S.M. Boca, M.F. Leitzmann, G. Yang, Y.B. Xiang, R. Sinha, X.O. Shu, A.J. Cross, Human metabolic correlates of body mass index, Metabolomics. 10 (2014) 259–269. https://doi.org/10.1007/s11306-013-0574-1.
[154] B.A.M. Schutte, E.B. van den Akker, J. Deelen, O. van de Rest, D. van Heemst, E.J.M. Feskens, M. Beekman, P.E. Slagboom, The effect of standardized food intake on the association between BMI and 1H-NMR metabolites, Sci Rep. 6 (2016) 38980. https://doi.org/10.1038/srep38980.
[155] P.J. Turnbaugh, V.K. Ridaura, J.J. Faith, F.E. Rey, R. Knight, J.I. Gordon, The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice, Sci Transl Med. 1 (2009) 6ra14. https://doi.org/10.1126/scitranslmed.3000322.

[156] E. Neis, C. Dejong, S. Rensen, The Role of Microbial Amino Acid Metabolism in Host Metabolism, Nutrients. 7 (2015) 2930–2946. https://doi.org/10.3390/nu7042930.
[157] C.L. Johnson, J. Versalovic, The Human Microbiome and Its Potential Importance to Pediatrics, Pediatrics. 129 (2012) 950–960. https://doi.org/10.1542/peds.2011-2736.
[158] W.M. de Vos, H. Tilg, M. Van Hul, P.D. Cani, Gut microbiome and health: mechanistic insights, Gut. (2022) gutjnl-2021-326789. https://doi.org/10.1136/gutjnl-2021-326789.
[159] Y.M.K. Lei, L. Nair, M.-L. Alegre, The Interplay between the Intestinal Microbiota and the Immune System, Clin Res Hepatol Gastroenterol. 39 (2015) 9–19. https://doi.org/10.1016/j.clinre.2014.10.008.
[160] K.H. McLeod, J.L. Richards, Y.A. Yap, E. Mariño, Dietary Short Chain Fatty Acids: How the Gut Microbiota Fight Against Autoimmune and Inflammatory Diseases, in: Bioactive Food as Dietary Interventions for Arthritis and Related Inflammatory Diseases, Elsevier, 2019: pp. 139–159. https://doi.org/10.1016/B978-0-12-813820-5.00007-6.
[161] N.C. Roy, S.A. Bassett, W. Young, C. Thum, W.C. McNabb, Chapter 18 - The Importance of Microbiota and Host Interactions Throughout Life, in: M. Boland, M. Golding, H. Singh (Eds.), Food Structures, Digestion and Health, Academic Press, San Diego, 2014: pp. 489–511. https://doi.org/10.1016/B978-0-12-404610-8.00018-9.
[162] G. den Besten, K. van Eunen, A.K. Groen, K. Venema, D.-J. Reijngoud, B.M. Bakker, The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism, Journal of Lipid Research. 54 (2013) 2325–2340. https://doi.org/10.1194/jlr.R036012.
[163] M. Ziętek, Z. Celewicz, M. Szczuko, Short-Chain Fatty Acids, Maternal Microbiota and Metabolism in Pregnancy, Nutrients. 13 (2021) 1244. https://doi.org/10.3390/nu13041244.
[164] C.T. Pekmez, M.W. Larsson, M.V. Lind, N. Vazquez Manjarrez, C. Yonemitsu, A. Larnkjær, L. Bode, C. Mølgaard, K.F. Michaelsen, L.O. Dragsted, Breastmilk Lipids and Oligosaccharides Influence Branched Short-Chain Fatty Acid Concentrations in Infants with Excessive Weight Gain, Mol. Nutr. Food Res. 64 (2020) 1900977. https://doi.org/10.1002/mnfr.201900977.
[165] S. Zhang, X. Zeng, M. Ren, X. Mao, S. Qiao, Novel metabolic and physiological functions of branched chain amino acids: a review, Journal of Animal Science and Biotechnology. 8 (2017) 10. https://doi.org/10.1186/s40104-016-0139-z.
[166] R. García-Villalba, J.A. Giménez-Bastida, M.T. García-Conesa, F.A. Tomás-Barberán, J.C. Espín, M. Larrosa, Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids

in faecal samples, Journal of Separation Science. 35 (2012) 1906–1913. https://doi.org/10.1002/jssc.201101121.
[167] H.M.H. van Eijk, J.G. Bloemen, C.H.C. Dejong, Application of liquid chromatography–mass spectrometry to measure short chain fatty acids in blood, Journal of Chromatography B. 877 (2009) 719–724. https://doi.org/10.1016/j.jchromb.2009.01.039.
[168] J.C.Y. Chan, D.Y.Q. Kioh, G.C. Yap, B.W. Lee, E.C.Y. Chan, A novel LCMSMS method for quantitative measurement of short-chain fatty acids in human stool derivatized with 12C- and 13C-labelled aniline, Journal of Pharmaceutical and Biomedical Analysis. 138 (2017) 43–53. https://doi.org/10.1016/j.jpba.2017.01.044.
[169] M. Zeng, H. Cao, Fast quantification of short chain fatty acids and ketone bodies by liquid chromatography-tandem mass spectrometry after facile derivatization coupled with liquid-liquid extraction, Journal of Chromatography B. 1083 (2018) 137–145. https://doi.org/10.1016/j.jchromb.2018.02.040.
[170] H.E. Song, H.Y. Lee, S.J. Kim, S.H. Back, H.J. Yoo, A Facile Profiling Method of Short Chain Fatty Acids Using Liquid Chromatography-Mass Spectrometry, Metabolites. 9 (2019) 173. https://doi.org/10.3390/metabo9090173.
[171] X. Gao, E. Pujos-Guillot, J.-F. Martin, P. Galan, C. Juste, W. Jia, J.-L. Sebedio, Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization, Analytical Biochemistry. 393 (2009) 163–175. https://doi.org/10.1016/j.ab.2009.06.036.
[172] H.F.N. Kvitvang, T. Andreassen, T. Adam, S.G. Villas-Bôas, P. Bruheim, Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids, Anal. Chem. 83 (2011) 2705–2711. https://doi.org/10.1021/ac103245b.
[173] S. Zhang, H. Wang, M.-J. Zhu, A sensitive GC/MS detection method for analyzing microbial metabolites short chain fatty acids in fecal and serum samples, Talanta. 196 (2019) 249–254. https://doi.org/10.1016/j.talanta.2018.12.049.
[174] R. Menon, Spontaneous preterm birth, a clinical dilemma: Etiologic, pathophysiologic and genetic heterogeneities and racial disparity, Acta Obstetricia et Gynecologica Scandinavica. 87 (2008) 590–600. https://doi.org/10.1080/00016340802005126.
[175] K.E. Gregory, C.E. DeForge, K.M. Natale, M. Phillips, L.J. Van Marter, Necrotizing Enterocolitis in the Premature Infant, Adv Neonatal Care. 11 (2011) 155–166. https://doi.org/10.1097/ANC.0b013e31821baaf4.

[176] V. Apgar, A Proposal for a New Method of Evaluation of the Newborn Infant, Anesthesia & Analgesia. 120 (2015) 1056–1059. https://doi.org/10.1213/ANE.0b013e31829bdc5c.
[177] National Collaborating Centre for Women's and Children's Health (UK), Urinary Tract Infection in Children: Diagnosis, Treatment and Long-term Management, RCOG Press, London, 2007. http://www.ncbi.nlm.nih.gov/books/NBK50606/ (accessed February 9, 2022).
[178] J. Kaufman, P. Fitzpatrick, S. Tosif, S.M. Hopper, S.M. Donath, P.A. Bryant, F.E. Babl, Faster clean catch urine collection (Quick-Wee method) from infants: randomised controlled trial, BMJ. 357 (2017) j1341. https://doi.org/10.1136/bmj.j1341.
[179] M.L. Herreros Fernandez, N. Gonzalez Merino, A. Tagarro Garcia, B. Perez Seoane, M. de la Serna Martinez, M.T. Contreras Abad, A. Garcia-Pose, A new technique for fast and safe collection of urine in newborns, Archives of Disease in Childhood. 98 (2013) 27–29. https://doi.org/10.1136/archdischild-2012-301872.
[180] N. Altuntas, A. Celebi Tayfur, M. Kocak, H.C. Razi, S. Akkurt, Midstream clean-catch urine collection in newborns: a randomized controlled study, Eur J Pediatr. 174 (2015) 577–582. https://doi.org/10.1007/s00431-014-2434-z.
[181] I. Alothaim, D.R. Gaya, D.G. Watson, Development of a Sensitive Liquid Chromatography Mass Spectrometry Method for the Analysis of Short Chain Fatty Acids in Urine from Patients with Ulcerative Colitis, Current Metabolomics. 6 (2018) 124–130.
[182] F.M. Vaz, S. Ferdinandusse, Bile acid analysis in human disorders of bile acid biosynthesis, Mol Aspects Med. 56 (2017) 10–24. https://doi.org/10.1016/j.mam.2017.03.003.
[183] J.M. Ridlon, D.-J. Kang, P.B. Hylemon, Bile salt biotransformations by human intestinal bacteria, Journal of Lipid Research. 47 (2006) 241–259. https://doi.org/10.1194/jlr.R500013-JLR200.
[184] T. Vasavan, E. Ferraro, E. Ibrahim, P. Dixon, J. Gorelik, C. Williamson, Heart and bile acids – Clinical consequences of altered bile acid metabolism, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 1864 (2018) 1345–1355. https://doi.org/10.1016/j.bbadis.2017.12.039.
[185] S. Li, C. Li, W. Wang, Bile acid signaling in renal water regulation, American Journal of Physiology-Renal Physiology. 317 (2019) F73–F76. https://doi.org/10.1152/ajprenal.00563.2018.
[186] S. Straniero, A. Laskar, C. Savva, J. Härdfeldt, B. Angelin, M. Rudling, Of mice and men: murine bile acids explain species differences in the

regulation of bile acid and cholesterol metabolism, J Lipid Res. 61 (2020) 480–491. https://doi.org/10.1194/jlr.RA119000307.
[187] Q. Lin, X. Tan, W. Wang, W. Zeng, L. Gui, M. Su, C. Liu, W. Jia, L. Xu, K. Lan, Species Differences of Bile Acid Redox Metabolism: Tertiary Oxidation of Deoxycholate is Conserved in Preclinical Animals, Drug Metab Dispos. 48 (2020) 499–507. https://doi.org/10.1124/dmd.120.090464.
[188] D.W. Russell, The enzymes, regulation, and genetics of bile acid synthesis, Annu Rev Biochem. 72 (2003) 137–174. https://doi.org/10.1146/annurev.biochem.72.121801.161712.
[189] P.A. Dawson, Bile Formation and the Enterohepatic Circulation, in: Physiology of the Gastrointestinal Tract, Elsevier, 2018: pp. 931–956. https://doi.org/10.1016/B978-0-12-809954-4.00041-4.
[190] Y. Alnouti, Bile Acid Sulfation: A Pathway of Bile Acid Elimination and Detoxification, Toxicological Sciences. 108 (2009) 225–246. https://doi.org/10.1093/toxsci/kfn268.
[191] O. Barbier, J. Trottier, J. Kaeding, P. Caron, M. Verreault, Lipid-activated transcription factors control bile acid glucuronidation, Mol Cell Biochem. 326 (2009) 3–8. https://doi.org/10.1007/s11010-008-0001-5.
[192] R. Durník, L. Šindlerová, P. Babica, O. Jurček, Bile Acids Transporters of Enterohepatic Circulation for Targeted Drug Delivery, Molecules. 27 (2022) 2961. https://doi.org/10.3390/molecules27092961.
[193] B.L. Zwicker, L.B. Agellon, Transport and biological activities of bile acids, The International Journal of Biochemistry & Cell Biology. 45 (2013) 1389–1398. https://doi.org/10.1016/j.biocel.2013.04.012.
[194] S.P.R. Bathena, R. Thakare, N. Gautam, S. Mukherjee, M. Olivera, J. Meza, Y. Alnouti, Urinary Bile Acids as Biomarkers for Liver Diseases I. Stability of the Baseline Profile in Healthy Subjects, Toxicological Sciences. 143 (2015) 296–307. https://doi.org/10.1093/toxsci/kfu227.
[195] V. Dosedělová, P. Itterheimová, P. Kubáň, Analysis of bile acids in human biological samples by microcolumn separation techniques: A review, ELECTROPHORESIS. 42 (2021) 68–85. https://doi.org/10.1002/elps.202000139.
[196] X. Zheng, F. Huang, A. Zhao, S. Lei, Y. Zhang, G. Xie, T. Chen, C. Qu, C. Rajani, B. Dong, D. Li, W. Jia, Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice, BMC Biol. 15 (2017) 120. https://doi.org/10.1186/s12915-017-0462-7.
[197] C. Steiner, A. von Eckardstein, K.M. Rentsch, Quantification of the 15 major human bile acids and their precursor 7α-hydroxy-4-cholesten-3-one in serum by liquid chromatography-tandem mass spectrometry, J

Chromatogr B Analyt Technol Biomed Life Sci. 878 (2010) 2870–2880. https://doi.org/10.1016/j.jchromb.2010.08.045.

[198] S. Perwaiz, B. Tuchweber, D. Mignault, T. Gilat, I.M. Yousef, Determination of bile acids in biological fluids by liquid chromatography-electrospray tandem mass spectrometry, Journal of Lipid Research. 42 (2001) 114–119. https://doi.org/10.1016/S0022-2275(20)32342-7.

[199] J.C. García-Cañaveras, M.T. Donato, J.V. Castell, A. Lahoz, Targeted profiling of circulating and hepatic bile acids in human, mouse, and rat using a UPLC-MRM-MS-validated method, J Lipid Res. 53 (2012) 2231–2241. https://doi.org/10.1194/jlr.D028803.

[200] L. Humbert, M.A. Maubert, C. Wolf, H. Duboc, M. Mahé, D. Farabos, P. Seksik, J.M. Mallet, G. Trugnan, J. Masliah, D. Rainteau, Bile acid profiling in human biological samples: Comparison of extraction procedures and application to normal and cholestatic patients, Journal of Chromatography B. 899 (2012) 135–145. https://doi.org/10.1016/j.jchromb.2012.05.015.

[201] K. Ushijima, A. Kimura, T. Inokuchi, Y. Yamato, K. Maeda, Y. Yamashita, E. Nakashima, H. Kato, Placental Transport of Bile Acids: Analysis of Bile Acids in Maternal Serum and Urine, Umbilical Cord Blood, and Amniotic Fluid., Kurume Med. J. 48 (2001) 87–91. https://doi.org/10.2739/kurumemedj.48.87.

[202] B.S. Kumar, B.C. Chung, Y.-J. Lee, H.J. Yi, B.-H. Lee, B.H. Jung, Gas chromatography-mass spectrometry-based simultaneous quantitative analytical method for urinary oxysterols and bile acids in rats, Anal Biochem. 408 (2011) 242–252. https://doi.org/10.1016/j.ab.2010.09.031.

[203] S. Matysik, G. Schmitz, Application of gas chromatography–triple quadrupole mass spectrometry to the determination of sterol components in biological samples in consideration of the ionization mode, Biochimie. 95 (2013) 489–495. https://doi.org/10.1016/j.biochi.2012.09.015.

[204] European Medicines Agency (EMA), ICH M10 on bioanalytical method validation and sample analysis - Scientific guideline, (2022).

[205] S. Takahashi, T. Fukami, Y. Masuo, C.N. Brocker, C. Xie, K.W. Krausz, C.R. Wolf, C.J. Henderson, F.J. Gonzalez, Cyp2c70 is responsible for the species difference in bile acid metabolism between mice and humans, J Lipid Res. 57 (2016) 2130–2137. https://doi.org/10.1194/jlr.M071183.

[206] World Health Organization, Breastfeeding, (n.d.). https://www.who.int/topics/breastfeeding/en/ (accessed July 2, 2019).

[207] D. Geddes, S. Perrella, Breastfeeding and human lactation, Nutrients. 11 (2019) 802–806. https://doi.org/10.1177/089033449401000132.
[208] C.G. Owen, R.M. Martin, P.H. Whincup, G. Davey Smith, D.G. Cook, Does breastfeeding influence risk of type 2 diabetes in later life? A quantitative analysis of published evidence, American Journal of Clinical Nutrition. 84 (2006) 1043–1054. https://doi.org/10.3945/ajcn.111.033035.
[209] G. Der, G.D. Batty, I.J. Deary, Effect of breast feeding on intelligence in children: Prospective study, sibling pairs analysis, and meta-analysis, British Medical Journal. 333 (2006) 945–948. https://doi.org/10.1136/bmj.38978.699583.55.
[210] J.C. Rozé, D. Darmaun, C.Y. Boquien, C. Flamant, J.C. Picaud, C. Savagner, O. Claris, A. Lapillonne, D. Mitanchez, B. Branger, U. Simeoni, M. Kaminski, P.Y. Ancel, The apparent breastfeeding paradox in very preterm infants: Relationship between breast feeding, early weight gain and neurodevelopment based on results from two cohorts, EPIPAGE and LIFT, BMJ Open. 2 (2012) 1–9. https://doi.org/10.1136/bmjopen-2012-000834.
[211] B.L. Horta, C. Loret De Mola, C.G. Victora, Breastfeeding and intelligence: A systematic review and meta-analysis, Acta Paediatrica, International Journal of Paediatrics. 104 (2015) 14–19. https://doi.org/10.1111/apa.13139.
[212] B. Lönnerdal, Bioactive proteins in human milk: mechanisms of action, The Journal of Pediatrics. 156 (2010) S26–S30. https://doi.org/10.1016/j.jpeds.2009.11.017.
[213] S. Musilova, V. Rada, E. Vlkova, V. Bunesova, Beneficial effects of human milk oligosaccharides on gut microbiota, Beneficial Microbes. 5 (2014) 273–283. https://doi.org/10.3920/BM2013.0080.
[214] M. Alsaweed, P.E. Hartmann, D.T. Geddes, F. Kakulas, MicroRNAs in breastmilk and the lactating breast: Potential immunoprotectors and developmental regulators for the infant and the mother, 2015. https://doi.org/10.3390/ijerph121113981.
[215] M. van den Berg, K. Kypke, A. Kotz, A. Tritscher, S.Y. Lee, K. Magulova, H. Fiedler, R. Malisch, WHO/UNEP global surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit–risk evaluation of breastfeeding, Archives of Toxicology. 91 (2017) 83–96. https://doi.org/10.1007/s00204-016-1802-z.
[216] D. Garwolińska, J. Namieśnik, A. Kot-Wasik, W. Hewelt-Belka, State of the art in sample preparation for human breast milk metabolomics—merits and limitations, TrAC - Trends in Analytical Chemistry. 114 (2019) 1–10. https://doi.org/10.1016/j.trac.2019.02.014.

[217] F.C. Marincola, A. Noto, P. Caboni, A. Reali, L. Barberini, M. Lussu, F. Murgia, M.L. Santoru, L. Atzori, V. Fanos, A metabolomic study of preterm human and formula milk by high resolution NMR and GC/MS analysis: preliminary results, The Journal of Maternal-Fetal an Neonatal Medicine. 25 (2012) 62–67. https://doi.org/10.3109/14767058.2012.715436.

[218] J.T. Smilowitz, A.O.Õ. Sullivan, D. Barile, J.B. German, B. Lo, The human milk metabolome reveals diverse oligosaccharide profiles, The Journal of Nutrition. 143 (2013) 1709–1718. https://doi.org/10.3945/jn.113.178772.1709.

[219] K. Li, J. Jiang, H. Xiao, K. Wu, C. Qi, J. Sund, D. Li, Changes in metabolites profile of breast milk over lactation stages and their relationship with dietary intake in Chinese: HPLC-QTOFMS based metabolomic analysis, Food & Function. 9 (2018) 5189–5197. https://doi.org/10.1039/C8FO01005F.

[220] M. Longini, M.L. Tataranno, F. Proietti, M. Tortoriello, E. Belvisi, A. Vivi, M. Tassini, S. Perrone, G. Buonocore, A metabolomic study of preterm and term human and formula milk by proton MRS analysis: preliminary results, The Journal of Maternal-Fetal and Neonatal Medicine. 7058 (2014) 27–33. https://doi.org/10.3109/14767058.2014.955958.

[221] A. Murgia, P. Scano, M. Contu, I. Ibba, M. Altea, M. Demuru, A. Porcu, P. Caboni, Characterization of donkey milk and metabolite profile comparison with human milk and formula milk, LWT - Food Science and Technology. 74 (2016) 427–433. https://doi.org/10.1016/j.lwt.2016.07.070.

[222] L. Qian, A. Zhao, Y. Zhang, T. Chen, S.H. Zeisel, W. Jia, W. Cai, Metabolomic approaches to explore chemical diversity of human breast-milk, formula milk and bovine milk, International Journal of Molecular Sciences. 17 (2016) 2128–2143. https://doi.org/10.3390/ijms17122128.

[223] P. Scano, A. Murgia, M. Demuru, R. Consonni, P. Caboni, Metabolite profiles of formula milk compared to breast milk, Food Research International. 87 (2016) 76–82. https://doi.org/10.1016/j.foodres.2016.06.024.

[224] T.I.B. Lopes, M.C. Cañedo, F.M.P. Oliveira, G.B. Ancantara, Toward precision nutrition: commercial infant formulas and human milk compared for stereospecific distribution of fatty acids using metabolomics, OMICS A Journal of Integrative Biology. 22 (2018) 484–492. https://doi.org/10.1089/omi.2018.0064.

[225] A. O'Sullivan, X. He, E.M.S. McNiven, K. Hinde, N.W. Haggarty, B. Lönnerdal, C.M. Slupsky, Metabolomic phenotyping validates the infant rhesus monkey as a model of human infant metabolism, Journal

of Pediatric Gastroenterology and Nutrition. 56 (2013) 355–363. https://doi.org/10.1097/MPG.0b013e31827e1f07.
[226] A.R. Spevacek, J.T. Smilowitz, E.L. Chin, M.A. Underwood, J.B. German, C.M. Slupsky, Infant maturity at birth reveals minor differences in the maternal milk metabolome in the first month of lactation, The Journal of Nutrition. 145 (2015) 1698–1708. https://doi.org/10.3945/jn.115.210252.compared.
[227] U.K. Sundekilde, E. Downey, J.A.O. Mahony, C.O. Shea, C.A. Ryan, A.L. Kelly, H.C. Bertram, The effect of gestational and lactational age on the human milk metabolome, Nutrients. 8 (2016) 304–318. https://doi.org/10.3390/nu8050304.
[228] M.-C. Alexandre-Gouabau, T. Moyon, A. David-Sochard, F. Fenaille, S. Cholet, A.-L. Royer, Y. Guitton, H. Billard, D. Darmaun, J.-C. Rozé, C.-Y. Boquien, Comprehensive preterm breast milk metabotype associated with optimal infant early growth pattern, Nutrients. 11 (2019) 528–553. https://doi.org/10.3390/nu11030528.
[229] M.C. Alexandre-Gouabau, T. Moyon, V. Cariou, J.P. Antignac, E.M. Qannari, M. Croyal, M. Soumah, Y. Guitton, A. David-Sochard, H. Billard, A. Legrand, C. Boscher, D. Darmaun, J.C. Rozé, C.Y. Boquien, Breast milk lipidome is associated with early growth trajectory in preterm infants, Nutrients. 10 (2018) 164–192. https://doi.org/10.3390/nu10020164.
[230] A. Dessì, D. Briana, S. Corbu, S. Gavrili, F.C. Marincola, S. Georgantzi, R. Pintus, V. Fanos, A. Malamitsi-Puchner, Metabolomics of breast milk: The importance of phenotypes, Metabolites. 8 (2018) 79–88. https://doi.org/10.3390/metabo8040079.
[231] A. Villaseñor, I. Garcia-Perez, A. Garcia, J.M. Posma, M. Fernández-Lópes, A.J. Nicholas, N. Modi, E. Holmes, C. Barbas, Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach∇, Anal. Chem. 86 (2014) 8245–8252. https://doi.org/10.1021/ac501853d.
[232] N.J. Andreas, M.J. Hyde, M. Gomez-romero, M.A. Lopez-Gonzalvez, A. Villaseñor, A. Wijeyesekera, C. Barbas, N. Modi, E. Holmes, I. Garcia-Perez, Multiplatform characterization of dynamic changes in breast milk during lactation, Electrophoresis. 36 (2015) 2269–2285. https://doi.org/10.1002/elps.201500011.
[233] J. Wu, M. Domellöf, A.M. Zivkovic, G. Larsson, A. Öhman, M.L. Nording, NMR-based metabolite profiling of human milk: A pilot study of methods for investigating compositional changes during lactation, Biochemical and Biophysical Research Communications. 469 (2016) 626–632. https://doi.org/10.1016/j.bbrc.2015.11.114.

[234] E. Isganaitis, S. Venditti, T.J. Matthews, C. Lerin, E.W. Demerath, D.A. Fields, Maternal obesity and the human milk metabolome: associations with infant body composition and postnatal weight gain, American Journal of Clinical Nutrition. 110 (2019) 111–120. https://doi.org/10.1093/ajcn/nqy334.

[235] K. Dangat, D. Upadhyay, A. Kilari, U. Sharma, N. Kemse, S. Mehendale, S. Lalwani, G. Wagh, S. Joshi, N.R. Jagannathan, Altered breast milk components in preeclampsia; An in-vitro proton NMR spectroscopy study, Clinica Chimica Acta. 463 (2016) 75–83. https://doi.org/10.1016/j.cca.2016.10.015.

[236] G. Praticò, G. Capuani, A. Tomassini, E. Baldassarre, M. Delfini, A. Miccheli, Exploring human breast milk composition by NMR-based metabolomics, Natural Product Research. 28 (2013) 95–101. https://doi.org/10.1080/14786419.2013.843180.

[237] M.C.L. Gay, P.T. Koleva, C.M. Slupsky, E. Toit, M. Eggesbo, C.C. Johnson, G. Wegienka, N. Shimojo, Worldwide variation in human milk metabolome: Indicators of breast physiology and maternal lifestyle?, Nutrients. 10 (2018) 1151–1162. https://doi.org/10.3390/nu10091151.

[238] C. Gómez-Gallego, J.M. Morales, D. Monleón, E. du Toit, H. Kumar, K.M. Linderborg, Y. Zhang, B. Yang, E. Isolauri, S. Salminen, M.C. Collado, Human breast milk NMR metabolomic profile across specific geographical locations and its association with the milk microbiota, Nutrients. 10 (2018) 1355–1375. https://doi.org/10.3390/nu10101355.

[239] W. Hewelt-Belka, D. Garwolińska, M. Belka, T. Bączek, J. Namieśnik, A. Kot-Wasik, A new dilution-enrichment sample preparation strategy for expanded metabolome monitoring of human breast milk that overcomes the simultaneous presence of low- and high-abundance lipid species, Food Chemistry. 288 (2019) 154–161. https://doi.org/10.1016/j.foodchem.2019.03.001.

[240] C. Urbaniak, A. Mcmillan, M. Angelini, G.B. Gloor, M. Sumarah, J.P. Burton, G. Reid, Effect of chemotherapy on the microbiota and metabolome of human milk, a case report, Microbiome. 2 (2014) 24–35. https://doi.org/10.1186/2049-2618-2-24.

[241] H. Demmelmair, B. Koletzko, Variation of metabolite and hormone contents in human milk, Clinics in Perinatology. 44 (2017) 151–164. https://doi.org/10.1016/j.clp.2016.11.007.

[242] F.C. Marincola, A. Dessì, S. Corbu, A. Reali, V. Fanos, Clinical impact of human breast milk metabolomics, Clinica Chimica Acta. 451 (2015) 103–106. https://doi.org/10.1016/j.cca.2015.02.021.

[243] C.M. Slupsky, Metabolomics in human milk research, Nestle Nutrition Institute Workshop Series. 90 (2019) 179–190. https://doi.org/10.1159/000490305.

[244] F. Bardanzellu, V. Fanos, A. Reali, “Omics” in human colostrum and mature milk: Looking to old data with new eyes, Nutrients. 9 (2017) 843–867. https://doi.org/10.3390/nu9080843.
[245] A.D. George, M.C.L. Gay, R.D. Trengove, D.T. Geddes, Human milk lipidomics: Current techniques and methodologies, Nutrients. 10 (2018) 1169–1179. https://doi.org/10.3390/nu10091169.
[246] D.A. Gidrewicz, T.R. Fenton, A systematic review and meta-analysis of the nutrient content of preterm and term breast milk, BMC Pediatrics. 14 (2014) 216–230. https://doi.org/10.1186/2046-1682-4-13.
[247] D. Hampel, S. Shahab-Ferdows, M.M. Islam, J.M. Peerson, L.H. Allen, Vitamin concentrations in human milk vary with time within feed, circadian rhythm, and single-dose supplementation, The Journal of Nutrition. 147 (2017) 603–611. https://doi.org/10.3945/jn.116.242941.
[248] D. Mung, L. Li, Development of chemical isotope labeling LC-MS for milk metabolomics: Comprehensive and quantitative profiling of the amine/phenol submetabolome, Anal. Chem. 89 (2017) 4435–4443. https://doi.org/10.1021/acs.analchem.6b03737.
[249] D. Mung, L. Li, Applying quantitative metabolomics based on chemical isotope labeling LC-MS for detecting potential milk adulterant in human milk, Analytica Chimica Acta. 1001 (2018) 78–85. https://doi.org/10.1016/j.aca.2017.11.019.
[250] C. Garcia, R.D. Duan, V. Brévaut-Malaty, C. Gire, V. Millet, U. Simeoni, M. Bernard, M. Armand, G.G. C., D. R.D., B.-M. V., G.G. C., M. V., S. U., B. M., A. M., Bioactive compounds in human milk and intestinal health and maturity in preterm newborn: An overview, Cellular and Molecular Biology. 59 (2013) 108–131. https://doi.org/10.1170/T952.
[251] D. Garwolińska, J. Namieśnik, A. Kot-Wasik, W. Hewelt-Belka, Chemistry of human breast milk - A comprehensive review of the composition and role of milk metabolites in child development, Journal of Agricultural and Food Chemistry. 66 (2018) 11881–11896. https://doi.org/10.1021/acs.jafc.8b04031.
[252] S.M. Totten, A.M. Zivkovic, S. Wu, U. Ngyuen, S.L. Freeman, L.R. Ruhaak, M.K. Darboe, J.B. German, A.M. Prentice, C.B. Lebrilla, Comprehensive profiles of human milk oligosaccharides yield highly sensitive and specific markers for determining secretor status in lactating mothers, Journal of Proteome Research. 11 (2012) 6124–6133. https://doi.org/10.1021/pr300769g.
[253] R. Lubetzky, Y. Littner, F.B. Mimouni, S. Dollberg, D. Mandel, R. Lubetzky, R. Lubetzky, Y. Littner, F.B. Mimouni, S. Dollberg, D. Mandel, Circadian variations in fat content of expressed breast milk from mothers of preterm infants, Journal of the American College of

Nutrition. 25 (2006) 151–154. https://doi.org/10.1080/07315724.2006.10719526.
[254] T. Saarela, J. Kokkonen, M. Koivisto, Macronutrient and energy contents of human milk fractions during the first six months of lactation, Acta Paediatrica (Oslo, Norway : 1992). 94 (2005) 1176–1181. https://doi.org/10.1080/08035250510036499.
[255] R.G. Jensen, C.J. Lammi-Keefe, B. Koletzko, Representative sampling of human milk and the extraction of fat for analysis of environmental lipophilic contaminants, Toxicological and Environmental Chemistry. 62 (1997) 229–247. https://doi.org/10.1080/02772249709358510.
[256] E. Bertino, C. Peila, F. Cresi, E. Maggiora, S. Sottemano, D. Gazzolo, S. Arslanoglu, A. Coscia, Donor human milk: Effects of storage and heat treatment on oxidative stress markers, Frontiers in Pediatrics. 6 (2018) 1–5. https://doi.org/10.3389/fped.2018.00253.
[257] C. Peila, G.E. Moro, E. Bertino, L. Cavallarin, M. Giribaldi, F. Giuliani, F. Cresi, A. Coscia, The effect of holder pasteurization on nutrients and biologically-active components in donor human milk: A review, Nutrients. 8 (2016) 1–19. https://doi.org/10.3390/nu8080477.
[258] N.R. García-Lara, D.E. Vieco, J. De La Cruz-Bértolo, D. Lora-Pablos, N.U. Velasco, C.R. Pallás-Alonso, Effect of holder pasteurization and frozen storage on macronutrients and energy content of breast milk, Journal of Pediatric Gastroenterology and Nutrition. 57 (2013) 377–382. https://doi.org/10.1097/MPG.0b013e31829d4f82.
[259] J. Folch, M. Lees, G.H. Sloane Stantley, A simple method for the isolation and purification of total lipides from animal tissues, Journal of Biological Chemistry. 266 (1957) 497–509.
[260] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Canadian Journal of Biochemistry and Physiology. 37 (1959) 911–917. https://doi.org/dx.doi.org/10,1139/cjm2014-0700.
[261] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W.B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, Metabolomics. 14 (2018) 72. https://doi.org/10.1007/s11306-018-1367-3.
[262] M. Vinaixa, E.L. Schymanski, S. Neumann, M. Navarro, R.M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects, TrAC - Trends in Analytical Chemistry. 78 (2016) 23–35. https://doi.org/10.1016/j.trac.2015.09.005.

[263] O. Fiehn, D.K. Barupal, T. Kind, Extending biochemical databases by metabolomic surveys, Journal of Biological Chemistry. 286 (2011) 23637–23643. https://doi.org/10.1074/jbc.R110.173617.
[264] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: The human metabolome database for 2018, Nucleic Acids Research. 46 (2018) D608–D617. https://doi.org/10.1093/nar/gkx1089.
[265] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: A metabolite mass spectral database, Therapeutic Drug Monitoring. 27 (2005) 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39.
[266] T. Kind, G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, FiehnLib – mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry, Analytical Chemistry. 81 (2009) 10038–10048. https://doi.org/10.1021/ac9019522.
[267] Cardiff University, Babraham Institute, S.D. University of California, LIPID MAPS Lipidomics Gateway, (n.d.). http://www.lipidmaps.org/ (accessed November 8, 2019).
[268] A. Foroutan, A.C. Guo, R. Vazquez-fresno, M. Lipfert, L. Zhang, J. Zheng, H. Badran, Z. Budinski, R. Mandal, B.N. Ametaj, D.S. Wishart, Chemical composition of commercial cow 's milk, Journal of Agricultural and Food Chemistry. 67 (2019) 4897–4914. https://doi.org/10.1021/acs.jafc.9b00204.
[269] Milk Composition Database, (n.d.). http://www.mcdb.ca/ (accessed November 5, 2019).
[270] KEGG PATHWAY Database, (n.d.). https://www.kegg.jp/kegg/pathway.html (accessed November 8, 2019).
[271] L. Li, R. Li, J. Zhou, A. Zuniga, A.E. Stanislaus, Y. Wu, T. Huan, J. Zheng, Y. Shi, D.S. Wishart, G. Lin, MyCompoundID: Using an evidence-based metabolome library for metabolite identification, Analytical Chemistry. 85 (2013) 3401–3408. https://doi.org/10.1021/ac400099b.
[272] A. Gil de la Fuente, J. Godzien, S. Saugar, R. Garcia-Carmona, H. Badran, D.S. Wishart, C. Barbas, A. Otero, CEU Mass Mediator 3.0: A Metabolite Annotation Tool, Journal of Proteome Research. 18 (2019) 797–802.
[273] CEU Mass Mediator, (n.d.). http://ceumass.eps.uspceu.es/mediator/ (accessed November 5, 2019).

[274] J.C. Oliveros, Venny. An interactive tool for comparing lists with Venn's diagrams, (2007). http://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed November 4, 2019).
[275] World Health Organization, Guideline: Protecting, promoting and supporting breastfeeding in facilities providing maternity and newborn services., 2017. http://www.ncbi.nlm.nih.gov/books/NBK487819/ (accessed March 8, 2019).
[276] C.-Y. Boquien, Human Milk: An Ideal Food for Nutrition of Preterm Newborn, Front Pediatr. 6 (2018). https://doi.org/10.3389/fped.2018.00295.
[277] J. Miller, E. Tonkin, R.A. Damarell, A.J. McPhee, M. Suganuma, H. Suganuma, P.F. Middleton, M. Makrides, C.T. Collins, A Systematic Review and Meta-Analysis of Human Milk Feeding and Morbidity in Very Low Birth Weight Infants, Nutrients. 10 (2018) 707. https://doi.org/10.3390/nu10060707.
[278] A. Parra-Llorca, M. Gormaz, C. Alcántara, M. Cernada, A. Nuñez-Ramiro, M. Vento, M.C. Collado, Preterm Gut Microbiome Depending on Feeding Type: Significance of Donor Human Milk, Front Microbiol. 9 (2018). https://doi.org/10.3389/fmicb.2018.01376.
[279] A. Parra-Llorca, M. Gormaz, Á. Sánchez-Illana, J.D. Piñeiro-Ramos, M.C. Collado, E. Serna, M. Cernada, A. Nuñez-Ramiro, A. Ramón-Beltrán, C. Oger, J.-M. Galano, C. Vigor, T. Durand, J. Kuligowski, M. Vento, Does pasteurized donor human milk efficiently protect preterm infants against oxidative stress?, Antioxid. Redox Signal. (2019). https://doi.org/10.1089/ars.2019.7821.
[280] J.D. Piñeiro-Ramos, A. Parra-Llorca, I. Ten-Doménech, M. Gormaz, A. Ramón-Beltrán, M. Cernada, G. Quintás, M.C. Collado, J. Kuligowski, M. Vento, Effect of donor human milk on host-gut microbiota and metabolic interactions in preterm infants, Clin Nutr. 40 (2021) 1296–1309. https://doi.org/10.1016/j.clnu.2020.08.013.
[281] D. Horta, M. Moreno-Torres, M.J. Ramírez-Lázaro, S. Lario, J. Kuligowski, J.D. Sanjuan-Herráez, G. Quintas, A. Villoria, X. Calvet, Analysis of the Association between Fatigue and the Plasma Lipidomic Profile of Inflammatory Bowel Disease Patients, J. Proteome Res. (2020). https://doi.org/10.1021/acs.jproteome.0c00462.
[282] T. Martínez-Sena, G. Luongo, D. Sanjuan-Herráez, J.V. Castell, M. Vento, G. Quintás, J. Kuligowski, Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis, Sci Rep. 9 (2019) 1–9. https://doi.org/10.1038/s41598-019-46371-w.
[283] D. Kessner, M. Chambers, R. Burke, D. Agus, P. Mallick, ProteoWizard: open source software for rapid proteomics tools

development, Bioinformatics. 24 (2008) 2534–2536. https://doi.org/10.1093/bioinformatics/btn323.
[284] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data, Anal. Chem. 84 (2012) 5035–5039.
[285] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets, Anal. Chem. 84 (2012) 283–289. https://doi.org/10.1021/ac202450g.
[286] P.D. Hutchins, J.D. Russell, J.J. Coon, LipiDex: An Integrated Software Package for High-Confidence Lipid Identification, Cell Systems. 6 (2018) 621-625.e5. https://doi.org/10.1016/j.cels.2018.03.011.
[287] X. Shen, R. Wang, X. Xiong, Y. Yin, Y. Cai, Z. Ma, N. Liu, Z.-J. Zhu, Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics, Nat Commun. 10 (2019) 1516. https://doi.org/10.1038/s41467-019-09550-x.
[288] J. Kuligowski, Á. Sánchez-Illana, D. Sanjuán-Herráez, M. Vento, G. Quintás, Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC), Analyst. 140 (2015) 7810–7817. https://doi.org/10.1039/c5an01638j.
[289] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1-27:27. https://doi.org/10.1145/1961189.1961199.
[290] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C.A. Fulcher, I.M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L.A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D.S. Weaver, P.D. Karp, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, Nucleic Acids Res. 44 (2016) D471–D480. https://doi.org/10.1093/nar/gkv1164.
[291] S. Li, Y. Park, S. Duraisingham, F.H. Strobel, N. Khan, Q.A. Soltow, D.P. Jones, B. Pulendran, Predicting Network Activity from High Throughput Metabolomics, PLoS Comput Biol. 9 (2013). https://doi.org/10.1371/journal.pcbi.1003123.
[292] B. van der Voorn, M. de Waard, L.R. Dijkstra, A.C. Heijboer, J. Rotteveel, J.B. van Goudoever, M.J.J. Finken, Stability of Cortisol and Cortisone in Human Breast Milk During Holder Pasteurization, J Pediatr Gastroenterol Nutr. 65 (2017) 658–660. https://doi.org/10.1097/MPG.0000000000001678.
[293] P. Piemontese, D. Mallardi, N. Liotto, C. Tabasso, C. Menis, M. Perrone, P. Roggero, F. Mosca, Macronutrient content of pooled donor

human milk before and after Holder pasteurization, BMC Pediatr. 19 (2019) 58. https://doi.org/10.1186/s12887-019-1427-5.

[294] M. Vincent, O. Ménard, J. Etienne, J. Ossemond, A. Durand, R. Buffin, E. Loizon, E. Meugnier, A. Deglaire, D. Dupont, J.-C. Picaud, C. Knibbe, M.-C. Michalski, A. Penhoat, Human milk pasteurisation reduces pre-lipolysis but not digestive lipolysis and moderately decreases intestinal lipid uptake in a combination of preterm infant in vitro models, Food Chemistry. 329 (2020) 126927. https://doi.org/10.1016/j.foodchem.2020.126927.

[295] N. Fidler, T.U. Sauerwald, B. Koletzko, H. Demmelmair, Effects of Human Milk Pasteurization and Sterilization on Available Fat Content and Fatty Acid Composition, Journal of Pediatric Gastroenterology and Nutrition. 27 (1998) 317–322.

[296] M. Mateos-Vivas, E. Rodríguez-Gonzalo, J. Domínguez-Álvarez, D. García-Gómez, R. Ramírez-Bernabé, R. Carabias-Martínez, Analysis of free nucleotide monophosphates in human milk and effect of pasteurisation or high-pressure processing on their contents by capillary electrophoresis coupled to mass spectrometry, Food Chem. 174 (2015) 348–355. https://doi.org/10.1016/j.foodchem.2014.11.051.

[297] W.-H. Hahn, J. Kim, S. Song, S. Park, N.M. Kang, The human milk oligosaccharides are not affected by pasteurization and freeze-drying, J Matern Fetal Neonatal Med. 32 (2019) 985–991. https://doi.org/10.1080/14767058.2017.1397122.

[298] K.S.B. Bergstrom, L. Xia, Mucin-type O-glycans and their roles in intestinal homeostasis, Glycobiology. 23 (2013) 1026–1037. https://doi.org/10.1093/glycob/cwt045.

[299] C. Lopez, C. Cauty, F. Guyomarc'h, Organization of lipids in milks, infant milk formulas and various dairy products: role of technological processes and potential impacts, Dairy Sci. & Technol. 95 (2015) 863–893. https://doi.org/10.1007/s13594-015-0263-0.

[300] M.A. Pitino, S.M. Alashmali, K.E. Hopperton, S. Unger, Y. Pouliot, A. Doyen, D.L. O'Connor, R.P. Bazinet, Oxylipin concentration, but not fatty acid composition, is altered in human donor milk pasteurised using both thermal and non-thermal techniques, Br J Nutr. 122 (2019) 47–55. https://doi.org/10.1017/S0007114519000916.

[301] R. Kalluri, V.S. LeBleu, The biology, function, and biomedical applications of exosomes, Science. 367 (2020). https://doi.org/10.1126/science.aau6977.

[302] T. Altadill, I. Campoy, L. Lanau, K. Gill, M. Rigau, A. Gil-Moreno, J. Reventos, S. Byers, E. Colas, A.K. Cheema, Enabling Metabolomics Based Biomarker Discovery Studies Using Molecular Phenotyping of

Exosome-Like Vesicles, PLOS ONE. 11 (2016) e0151339. https://doi.org/10.1371/journal.pone.0151339.
[303] Y. Liao, X. Du, J. Li, B. Lönnerdal, Human milk exosomes and their microRNAs survive digestion in vitro and are taken up by human intestinal cells, Molecular Nutrition & Food Research. 61 (2017) 1700082. https://doi.org/10.1002/mnfr.201700082.
[304] S. Kahn, Y. Liao, X. Du, W. Xu, J. Li, B. Lönnerdal, Exosomal MicroRNAs in Milk from Mothers Delivering Preterm Infants Survive in Vitro Digestion and Are Taken Up by Human Intestinal Cells, Molecular Nutrition & Food Research. 62 (2018) 1701050. https://doi.org/10.1002/mnfr.201701050.
[305] M. Yang, D. Song, X. Cao, R. Wu, B. Liu, W. Ye, J. Wu, X. Yue, Comparative proteomic analysis of milk-derived exosomes in human and bovine colostrum and mature milk samples by iTRAQ-coupled LC-MS/MS, Food Research International. 92 (2017) 17–25. https://doi.org/10.1016/j.foodres.2016.11.041.
[306] C. de la Torre Gomez, R.V. Goreham, J.J. Bech Serra, T. Nann, M. Kussmann, "Exosomics"—A Review of Biophysics, Biology and Biochemistry of Exosomes With a Focus on Human Breast Milk, Frontiers in Genetics. 9 (2018) 92. https://doi.org/10.3389/fgene.2018.00092.
[307] F. Royo, D. Gil-Carton, E. Gonzalez, J. Mleczko, L. Palomo, M. Perez-Cormenzana, R. Mayo, C. Alonso, J.M. Falcon-Perez, Differences in the metabolite composition and mechanical properties of extracellular vesicles secreted by hepatic cellular models, Journal of Extracellular Vesicles. 8 (2019) 1575678. https://doi.org/10.1080/20013078.2019.1575678.
[308] M. Clos-Garcia, A. Loizaga-Iriarte, P. Zuñiga-Garcia, P. Sánchez-Mosquera, A. Rosa Cortazar, E. González, V. Torrano, C. Alonso, M. Pérez-Cormenzana, A. Ugalde-Olano, I. Lacasa-Viscasillas, A. Castro, F. Royo, M. Unda, A. Carracedo, J.M. Falcón-Pérez, Metabolic alterations in urine extracellular vesicles are associated to prostate cancer pathogenesis and progression, Journal of Extracellular Vesicles. 7 (2018) 1470442. https://doi.org/10.1080/20013078.2018.1470442.
[309] R.A. Haraszti, M.-C. Didiot, E. Sapp, J. Leszyk, S.A. Shaffer, H.E. Rockwell, F. Gao, N.R. Narain, M. DiFiglia, M.A. Kiebish, N. Aronin, A. Khvorova, High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources, Journal of Extracellular Vesicles. 5 (2016) 32570. https://doi.org/10.3402/jev.v5.32570.
[310] V. Ramos-Garcia, I. Ten-Doménech, A. Moreno-Giménez, M. Gormaz, A. Parra-Llorca, A.P. Shephard, P. Sepúlveda, D. Pérez-Guaita, M.

Vento, B. Lendl, G. Quintás, J. Kuligowski, ATR-FTIR spectroscopy for the routine quality control of exosome isolations, Chemometrics and Intelligent Laboratory Systems. (2021) 104401. https://doi.org/10.1016/j.chemolab.2021.104401.

[311] M.I. Zonneveld, M.J.C. Herwijnen, M.M. Fernandez-Gutierrez, A. Giovanazzi, A.M. Groot, M. Kleinjan, T.M.M. Capel, A.J.A.M. Sijts, L.S. Taams, J. Garssen, E.C. Jong, M. Kleerebezem, E.N.M. Nolte-'t Hoen, F.A. Redegeld, M.H.M. Wauben, Human milk extracellular vesicles target nodes in interconnected signalling pathways that enhance oral epithelial barrier function and dampen immune responses, Journal of Extracellular Vesicles. 10 (2021). https://doi.org/10.1002/jev2.12071.

[312] K.W. Witwer, E.I. Buzás, L.T. Bemis, A. Bora, C. Lässer, J. Lötvall, E.N. Nolte-'t Hoen, M.G. Piper, S. Sivaraman, J. Skog, C. Théry, M.H. Wauben, F. Hochberg, Standardization of sample collection, isolation and analysis methods in extracellular vesicle research, Journal of Extracellular Vesicles. 2 (2013) 20360. https://doi.org/10.3402/jev.v2i0.20360.

[313] I. Ten-Doménech, V. Ramos-Garcia, J.D. Piñeiro-Ramos, M. Gormaz, A. Parra-Llorca, M. Vento, J. Kuligowski, G. Quintás, Current Practice in Untargeted Human Milk Metabolomics, Metabolites. 10 (2020) 43. https://doi.org/10.3390/metabo10020043.

[314] Á. Sánchez-Illana, D. Pérez-Guaita, D. Cuesta-García, J.D. Sanjuan-Herráez, M. Vento, J.L. Ruiz-Cerdá, G. Quintás, J. Kuligowski, Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control - Support vector regression, Anal Chim Acta. 1026 (2018) 62–68. https://doi.org/10.1016/j.aca.2018.04.055.

[315] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, Anal. Chem. 78 (2006) 779–787. https://doi.org/10.1021/ac051437y.

[316] G. Libiseller, M. Dvorzak, U. Kleb, E. Gander, T. Eisenberg, F. Madeo, S. Neumann, G. Trausinger, F. Sinner, T. Pieber, C. Magnes, IPO: a tool for automated optimization of XCMS parameters, BMC Bioinformatics. 16 (2015) 118. https://doi.org/10.1186/s12859-015-0562-8.

[317] R.M. Salek, C. Steinbeck, M.R. Viant, R. Goodacre, W.B. Dunn, The role of reporting standards for metabolite annotation and identification in metabolomic studies, GigaScience. 2 (2013). https://doi.org/10.1186/2047-217X-2-13.

[318] T. Kind, K.-H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid

identification, Nat Methods. 10 (2013) 755–758. https://doi.org/10.1038/nmeth.2551.
[319] Á. Sánchez-Illana, J.D. Piñeiro-Ramos, J.D. Sanjuan-Herráez, M. Vento, G. Quintás, J. Kuligowski, Evaluation of batch effect elimination using quality control replicates in LC-MS metabolite profiling, Anal. Chim. Acta. 1019 (2018) 38–48. https://doi.org/10.1016/j.aca.2018.02.053.
[320] C. Admyre, S.M. Johansson, K.R. Qazi, J.-J. Filén, R. Lahesmaa, M. Norman, E.P.A. Neve, A. Scheynius, S. Gabrielsson, Exosomes with Immune Modulatory Features Are Present in Human Breast Milk, J Immunol. 179 (2007) 1969–1978. https://doi.org/10.4049/jimmunol.179.3.1969.
[321] S. Wijenayake, S. Eisha, Z. Tawhidi, M.A. Pitino, M.A. Steele, A.S. Fleming, P.O. McGowan, Comparison of methods for pre-processing, exosome isolation, and RNA extraction in unpasteurized bovine and human milk, PLoS ONE. 16 (2021) e0257633. https://doi.org/10.1371/journal.pone.0257633.
[322] M.I. Zonneveld, A.R. Brisson, M.J.C. van Herwijnen, S. Tan, C.H.A. van de Lest, F.A. Redegeld, J. Garssen, M.H.M. Wauben, E.N.M. Nolte-’t Hoen, Recovery of extracellular vesicles from human breast milk is influenced by sample collection and vesicle isolation procedures, Journal of Extracellular Vesicles. 3 (2014) 24215. https://doi.org/10.3402/jev.v3.24215.
[323] W. Chen, X. Chen, Y. Qian, X. Wang, Y. Zhou, X. Yan, B. Yu, S. Yao, Z. Yu, J. Zhu, S. Han, Lipidomic Profiling of Human Milk Derived Exosomes and Their Emerging Roles in the Prevention of Necrotizing Enterocolitis, Mol. Nutr. Food Res. 65 (2021) 2000845. https://doi.org/10.1002/mnfr.202000845.
[324] B. Lönnerdal, Human Milk MicroRNAs/Exosomes: Composition and Biological Effects, in: S.M. Donovan, J.B. German, B. Lönnerdal, A. Lucas (Eds.), Nestlé Nutrition Institute Workshop Series, S. Karger AG, 2019: pp. 83–92. https://doi.org/10.1159/000490297.
[325] S. Chutipongtanate, A.L. Morrow, D.S. Newburg, Human Milk Extracellular Vesicles: A Biological System with Clinical Implications, Cells. 11 (2022) 2345. https://doi.org/10.3390/cells11152345.
[326] J.D. Galley, G.E. Besner, The Therapeutic Potential of Breast Milk-Derived Extracellular Vesicles, Nutrients. 12 (2020) 745. https://doi.org/10.3390/nu12030745.
[327] Y. Hu, J. Thaler, R. Nieuwland, Extracellular Vesicles in Human Milk, Pharmaceuticals. 14 (2021) 1050. https://doi.org/10.3390/ph14101050.
[328] A. Kupsco, D. Prada, D. Valvi, L. Hu, M.S. Petersen, B. Coull, P. Grandjean, P. Weihe, A.A. Baccarelli, Human milk extracellular vesicle

miRNA expression and associations with maternal characteristics in a population-based cohort from the Faroe Islands, Sci Rep. 11 (2021) 5840. https://doi.org/10.1038/s41598-021-84809-2.

[329] J. Donoso-Quezada, S. Ayala-Mar, J. González-Valdez, The role of lipids in exosome biology and intercellular communication: Function, analytics and applications, Traffic. 22 (2021) 204–220. https://doi.org/10.1111/tra.12803.

[330] T. Skotland, K. Sagini, K. Sandvig, A. Llorente, An emerging focus on lipids in extracellular vesicles, Advanced Drug Delivery Reviews. 159 (2020) 308–321. https://doi.org/10.1016/j.addr.2020.03.002.

[331] A.M.D. Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J.A. Simpson, T.P. Speed, Statistical Methods for Handling Unwanted Variation in Metabolomics Data, Anal. Chem. 87 (2015) 3606–3615. https://doi.org/10.1021/ac502439y.

[332] P. Cuevas-Delgado, D. Dudzik, V. Miguel, S. Lamas, C. Barbas, Data-dependent normalization strategies for untargeted metabolomics—a case study, Anal Bioanal Chem. 412 (2020) 6391–6405. https://doi.org/10.1007/s00216-020-02594-9.

[333] C. Théry, K.W. Witwer, E. Aikawa, M.J. Alcaraz, J.D. Anderson, R. Andriantsitohaina, A. Antoniou, T. Arab, F. Archer, G.K. Atkin-Smith, D.C. Ayre, J.-M. Bach, D. Bachurski, H. Baharvand, L. Balaj, S. Baldacchino, N.N. Bauer, A.A. Baxter, M. Bebawy, C. Beckham, A. Bedina Zavec, A. Benmoussa, A.C. Berardi, P. Bergese, E. Bielska, C. Blenkiron, S. Bobis-Wozowicz, E. Boilard, W. Boireau, A. Bongiovanni, F.E. Borràs, S. Bosch, C.M. Boulanger, X. Breakefield, A.M. Breglio, M.Á. Brennan, D.R. Brigstock, A. Brisson, M.L. Broekman, J.F. Bromberg, P. Bryl-Górecka, S. Buch, A.H. Buck, D. Burger, S. Busatto, D. Buschmann, B. Bussolati, E.I. Buzás, J.B. Byrd, G. Camussi, D.R. Carter, S. Caruso, L.W. Chamley, Y.-T. Chang, C. Chen, S. Chen, L. Cheng, A.R. Chin, A. Clayton, S.P. Clerici, A. Cocks, E. Cocucci, R.J. Coffey, A. Cordeiro-da-Silva, Y. Couch, F.A. Coumans, B. Coyle, R. Crescitelli, M.F. Criado, C. D'Souza-Schorey, S. Das, A. Datta Chaudhuri, P. de Candia, E.F. De Santana, O. De Wever, H.A. del Portillo, T. Demaret, S. Deville, A. Devitt, B. Dhondt, D. Di Vizio, L.C. Dieterich, V. Dolo, A.P. Dominguez Rubio, M. Dominici, M.R. Dourado, T.A. Driedonks, F.V. Duarte, H.M. Duncan, R.M. Eichenberger, K. Ekström, S. EL Andaloussi, C. Elie-Caille, U. Erdbrügger, J.M. Falcón-Pérez, F. Fatima, J.E. Fish, M. Flores-Bellver, A. Försönits, A. Frelet-Barrand, F. Fricke, G. Fuhrmann, S. Gabrielsson, A. Gámez-Valero, C. Gardiner, K. Gärtner, R. Gaudin, Y.S. Gho, B. Giebel, C. Gilbert, M. Gimona, I. Giusti, D.C. Goberdhan, A. Görgens, S.M. Gorski, D.W. Greening, J.C. Gross, A. Gualerzi, G.N. Gupta, D.

Gustafson, A. Handberg, R.A. Haraszti, P. Harrison, H. Hegyesi, A. Hendrix, A.F. Hill, F.H. Hochberg, K.F. Hoffmann, B. Holder, H. Holthofer, B. Hosseinkhani, G. Hu, Y. Huang, V. Huber, S. Hunt, A.G.-E. Ibrahim, T. Ikezu, J.M. Inal, M. Isin, A. Ivanova, H.K. Jackson, S. Jacobsen, S.M. Jay, M. Jayachandran, G. Jenster, L. Jiang, S.M. Johnson, J.C. Jones, A. Jong, T. Jovanovic-Talisman, S. Jung, R. Kalluri, S. Kano, S. Kaur, Y. Kawamura, E.T. Keller, D. Khamari, E. Khomyakova, A. Khvorova, P. Kierulf, K.P. Kim, T. Kislinger, M. Klingeborn, D.J. Klinke, M. Kornek, M.M. Kosanović, Á.F. Kovács, E.-M. Krämer-Albers, S. Krasemann, M. Krause, I.V. Kurochkin, G.D. Kusuma, S. Kuypers, S. Laitinen, S.M. Langevin, L.R. Languino, J. Lannigan, C. Lässer, L.C. Laurent, G. Lavieu, E. Lázaro-Ibáñez, S. Le Lay, M.-S. Lee, Y.X.F. Lee, D.S. Lemos, M. Lenassi, A. Leszczynska, I.T. Li, K. Liao, S.F. Libregts, E. Ligeti, R. Lim, S.K. Lim, A. Linē, K. Linnemannstöns, A. Llorente, C.A. Lombard, M.J. Lorenowicz, Á.M. Lörincz, J. Lötvall, J. Lovett, M.C. Lowry, X. Loyer, Q. Lu, B. Lukomska, T.R. Lunavat, S.L. Maas, H. Malhi, A. Marcilla, J. Mariani, J. Mariscal, E.S. Martens-Uzunova, L. Martin-Jaular, M.C. Martinez, V.R. Martins, M. Mathieu, S. Mathivanan, M. Maugeri, L.K. McGinnis, M.J. McVey, D.G. Meckes, K.L. Meehan, I. Mertens, V.R. Minciacchi, A. Möller, M. Møller Jørgensen, A. Morales-Kastresana, J. Morhayim, F. Mullier, M. Muraca, L. Musante, V. Mussack, D.C. Muth, K.H. Myburgh, T. Najrana, M. Nawaz, I. Nazarenko, P. Nejsum, C. Neri, T. Neri, R. Nieuwland, L. Nimrichter, J.P. Nolan, E.N. Nolte-'t Hoen, N. Noren Hooten, L. O'Driscoll, T. O'Grady, A. O'Loghlen, T. Ochiya, M. Olivier, A. Ortiz, L.A. Ortiz, X. Osteikoetxea, O. Østergaard, M. Ostrowski, J. Park, D.M. Pegtel, H. Peinado, F. Perut, M.W. Pfaffl, D.G. Phinney, B.C. Pieters, R.C. Pink, D.S. Pisetsky, E. Pogge von Strandmann, I. Polakovicova, I.K. Poon, B.H. Powell, I. Prada, L. Pulliam, P. Quesenberry, A. Radeghieri, R.L. Raffai, S. Raimondo, J. Rak, M.I. Ramirez, G. Raposo, M.S. Rayyan, N. Regev-Rudzki, F.L. Ricklefs, P.D. Robbins, D.D. Roberts, S.C. Rodrigues, E. Rohde, S. Rome, K.M. Rouschop, A. Rughetti, A.E. Russell, P. Saá, S. Sahoo, E. Salas-Huenuleo, C. Sánchez, J.A. Saugstad, M.J. Saul, R.M. Schiffelers, R. Schneider, T.H. Schøyen, A. Scott, E. Shahaj, S. Sharma, O. Shatnyeva, F. Shekari, G.V. Shelke, A.K. Shetty, K. Shiba, P.R.-M. Siljander, A.M. Silva, A. Skowronek, O.L. Snyder, R.P. Soares, B.W. Sódar, C. Soekmadji, J. Sotillo, P.D. Stahl, W. Stoorvogel, S.L. Stott, E.F. Strasser, S. Swift, H. Tahara, M. Tewari, K. Timms, S. Tiwari, R. Tixeira, M. Tkach, W.S. Toh, R. Tomasini, A.C. Torrecilhas, J.P. Tosar, V. Toxavidis, L. Urbanelli, P. Vader, B.W. van Balkom, S.G. van der Grein, J. Van Deun, M.J. van Herwijnen, K. Van Keuren-Jensen, G. van

Niel, M.E. van Royen, A.J. van Wijnen, M.H. Vasconcelos, I.J. Vechetti, T.D. Veit, L.J. Vella, É. Velot, F.J. Verweij, B. Vestad, J.L. Viñas, T. Visnovitz, K.V. Vukman, J. Wahlgren, D.C. Watson, M.H. Wauben, A. Weaver, J.P. Webber, V. Weber, A.M. Wehman, D.J. Weiss, J.A. Welsh, S. Wendt, A.M. Wheelock, Z. Wiener, L. Witte, J. Wolfram, A. Xagorari, P. Xander, J. Xu, X. Yan, M. Yáñez-Mó, H. Yin, Y. Yuana, V. Zappulli, J. Zarubova, V. Žėkas, J. Zhang, Z. Zhao, L. Zheng, A.R. Zheutlin, A.M. Zickler, P. Zimmermann, A.M. Zivkovic, D. Zocco, E.K. Zuba-Surma, Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines, Journal of Extracellular Vesicles. 7 (2018) 1535750. https://doi.org/10.1080/20013078.2018.1535750.

[334] V. Ramos-Garcia, I. Ten-Doménech, A. Albiach-Delgado, M. Gómez-Ferrer, P. Sepúlveda, A. Parra-Llorca, L. Campos-Berga, A. Moreno-Giménez, G. Quintás, J. Kuligowski, Isolation and Lipidomic Screening of Human Milk Extracellular Vesicles, in: R. González-Domínguez (Ed.), Mass Spectrometry for Metabolomics, Springer US, New York, NY, 2023: pp. 177–188. https://doi.org/10.1007/978-1-0716-2699-3_18.

[335] M. Gómez-Ferrer, E. Villanueva-Badenas, R. Sánchez-Sánchez, C.M. Sánchez-López, M.C. Baquero, P. Sepúlveda, A. Dorronsoro, HIF-1α and Pro-Inflammatory Signaling Improves the Immunomodulatory Activity of MSC-Derived Extracellular Vesicles, Int J Mol Sci. 22 (2021) 3416. https://doi.org/10.3390/ijms22073416.

[336] EV-TRACK Consortium, J. Van Deun, P. Mestdagh, P. Agostinis, Ö. Akay, S. Anand, J. Anckaert, Z.A. Martinez, T. Baetens, E. Beghein, L. Bertier, G. Berx, J. Boere, S. Boukouris, M. Bremer, D. Buschmann, J.B. Byrd, C. Casert, L. Cheng, A. Cmoch, D. Daveloose, E. De Smedt, S. Demirsoy, V. Depoorter, B. Dhondt, T.A.P. Driedonks, A. Dudek, A. Elsharawy, I. Floris, A.D. Foers, K. Gärtner, A.D. Garg, E. Geeurickx, J. Gettemans, F. Ghazavi, B. Giebel, T.G. Kormelink, G. Hancock, H. Helsmoortel, A.F. Hill, V. Hyenne, H. Kalra, D. Kim, J. Kowal, S. Kraemer, P. Leidinger, C. Leonelli, Y. Liang, L. Lippens, S. Liu, A. Lo Cicero, S. Martin, S. Mathivanan, P. Mathiyalagan, T. Matusek, G. Milani, M. Monguió-Tortajada, L.M. Mus, D.C. Muth, A. Németh, E.N.M. Nolte-'t Hoen, L. O'Driscoll, R. Palmulli, M.W. Pfaffl, B. Primdal-Bengtson, E. Romano, Q. Rousseau, S. Sahoo, N. Sampaio, M. Samuel, B. Scicluna, B. Soen, A. Steels, J.V. Swinnen, M. Takatalo, S. Thaminy, C. Théry, J. Tulkens, I. Van Audenhove, S. van der Grein, A. Van Goethem, M.J. van Herwijnen, G. Van Niel, N. Van Roy, A.R. Van Vliet, N. Vandamme, S. Vanhauwaert, G. Vergauwen, F. Verweij, A.

Wallaert, M. Wauben, K.W. Witwer, M.I. Zonneveld, O. De Wever, J. Vandesompele, A. Hendrix, EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research, Nat Methods. 14 (2017) 228–232. https://doi.org/10.1038/nmeth.4185.
[337] H. Heberle, G.V. Meirelles, F.R. da Silva, G.P. Telles, R. Minghim, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams, BMC Bioinformatics. 16 (2015) 169. https://doi.org/10.1186/s12859-015-0611-3.
[338] K. Blans, M.S. Hansen, L.V. Sørensen, M.L. Hvam, K.A. Howard, A. Möller, L. Wiking, L.B. Larsen, J.T. Rasmussen, Pellet-free isolation of human and bovine milk extracellular vesicles by size-exclusion chromatography, Journal of Extracellular Vesicles. 6 (2017) 1294340. https://doi.org/10.1080/20013078.2017.1294340.
[339] I.E. Ingvordsen Lindahl, V.M. Artegoitia, E. Downey, J.A. O'Mahony, C.-A. O'Shea, C.A. Ryan, A.L. Kelly, H.C. Bertram, U.K. Sundekilde, Quantification of Human Milk Phospholipids: the Effect of Gestational and Lactational Age on Phospholipid Composition, Nutrients. 11 (2019) 222. https://doi.org/10.3390/nu11020222.
[340] A. Pérez-Gálvez, M.V. Calvo, J. Megino-Tello, J. Aguayo-Maldonado, R. Jiménez-Flores, J. Fontecha, Effect of gestational age (preterm or full term) on lipid composition of the milk fat globule and its membrane in human colostrum, Journal of Dairy Science. 103 (2020) 7742–7751. https://doi.org/10.3168/jds.2020-18428.
[341] M. Wójcik, H. Mojska, Human milk metabolome: Impact of gestational age, lactational stage and maternal diet, Rocz Panstw Zakl Hig. 73 (2022) 139–145. https://doi.org/10.32394/rpzh.2022.0214.
[342] D. Bachurski, M. Schuldner, P.-H. Nguyen, A. Malz, K.S. Reiners, P.C. Grenzi, F. Babatz, A.C. Schauss, H.P. Hansen, M. Hallek, E. Pogge von Strandmann, Extracellular vesicle measurements with nanoparticle tracking analysis – An accuracy and repeatability comparison between NanoSight NS300 and ZetaView, Journal of Extracellular Vesicles. 8 (2019) 1596016. https://doi.org/10.1080/20013078.2019.1596016.
[343] H.H. Jung, J.-Y. Kim, J.E. Lim, Y.-H. Im, Cytokine profiling in serum-derived exosomes isolated by different methods, Sci Rep. 10 (2020) 14069. https://doi.org/10.1038/s41598-020-70584-z.
[344] U. Smyczynska, M.A. Bartlomiejczyk, M.M. Stanczak, P. Sztromwasser, A. Wesolowska, O. Barbarska, E. Pawlikowska, W. Fendler, Impact of processing method on donated human breast milk microRNA content, PLoS ONE. 15 (2020) e0236126. https://doi.org/10.1371/journal.pone.0236126.
[345] M. Somiya, Y. Yoshioka, T. Ochiya, Biocompatibility of highly purified bovine milk-derived extracellular vesicles, Journal of Extracellular

Vesicles. 7 (2018) 1440132. https://doi.org/10.1080/20013078.2018.1440132.
[346] T. Wolf, S.R. Baier, J. Zempleni, The Intestinal Transport of Bovine Milk Exosomes Is Mediated by Endocytosis in Human Colon Carcinoma Caco-2 Cells and Rat Small Intestinal IEC-6 Cells, The Journal of Nutrition. 145 (2015) 2201–2206. https://doi.org/10.3945/jn.115.218586.
[347] A. Benmoussa, S. Michel, C. Gilbert, P. Provost, Isolating Multiple Extracellular Vesicles Subsets, Including Exosomes and Membrane Vesicles, from Bovine Milk Using Sodium Citrate and Differential Ultracentrifugation, BIO-PROTOCOL. 10 (2020). https://doi.org/10.21769/BioProtoc.3636.
[348] S.C. de Oliveira, A. Deglaire, O. Ménard, A. Bellanger, F. Rousseau, G. Henry, E. Dirson, F. Carrière, D. Dupont, C. Bourlieu, Holder pasteurization impacts the proteolysis, lipolysis and disintegration of human milk under in vitro dynamic term newborn digestion, Food Research International. 88 (2016) 263–275. https://doi.org/10.1016/j.foodres.2015.11.022.
[349] F. Guyomarc'h, A.J.R. Law, D.G. Dalgleish, Formation of Soluble and Micelle-Bound Protein Aggregates in Heated Milk, J. Agric. Food Chem. 51 (2003) 4652–4660. https://doi.org/10.1021/jf0211783.
[350] H. Miyake, C. Lee, S. Chusilp, M. Bhalla, B. Li, M. Pitino, S. Seo, D.L. O'Connor, A. Pierro, Human breast milk exosomes attenuate intestinal damage, Pediatr Surg Int. 36 (2020) 155–163. https://doi.org/10.1007/s00383-019-04599-7.
[351] N.P. Hessvik, A. Llorente, Current knowledge on exosome biogenesis and release, Cell Mol Life Sci. 75 (2018) 193–208. https://doi.org/10.1007/s00018-017-2595-9.
[352] M. Zhou, S.R. Weber, Y. Zhao, H. Chen, J.M. Sundstrom, Chapter 2 - Methods for exosome isolation and characterization, in: L. Edelstein, J. Smythies, P. Quesenberry, D. Noble (Eds.), Exosomes, Academic Press, 2020: pp. 23–38. https://doi.org/10.1016/B978-0-12-816053-4.00002-X.
[353] Exosomes: A Rising Star in Failing Hearts, (n.d.). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5508217/ (accessed January 25, 2021).
[354] D. Perez-Guaita, Z. Richardson, P. Heraud, B. Wood, Quantification and Identification of Microproteinuria Using Ultrafiltration and ATR-FTIR Spectroscopy, Anal. Chem. 92 (2020) 2409–2416. https://doi.org/10.1021/acs.analchem.9b03081.
[355] S. Yoshida, Y. Okazaki, T. Yamashita, H. Ueda, R. Ghadimi, A. Hosono, T. Tanaka, K. Kuriki, S. Suzuki, S. Tokudome, Analysis of

human oral mucosa ex vivo for fatty acid compositions using Fourier-transform infrared spectroscopy, Lipids. 43 (2008) 361–372. https://doi.org/10.1007/s11745-007-3147-0.
[356] B.R. Wood, The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues, Chem Soc Rev. 45 (2016) 1980–1998. https://doi.org/10.1039/c5cs00511f.
[357] J. Kuligowski, D. Pérez-Guaita, J. Escobar, I. Lliso, M. de la Guardia, B. Lendl, M. Vento, G. Quintás, Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics, Talanta. 127 (2014) 181–190. https://doi.org/10.1016/j.talanta.2014.04.009.
[358] V. Szentirmai, A. Wacha, C. Németh, D. Kitka, A. Rácz, K. Héberger, J. Mihály, Z. Varga, Reagent-free total protein quantification of intact extracellular vesicles by attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy, Anal Bioanal Chem. 412 (2020) 4619–4628. https://doi.org/10.1007/s00216-020-02711-8.
[359] J. Mihály, R. Deák, I.C. Szigyártó, A. Bóta, T. Beke-Somfai, Z. Varga, Characterization of extracellular vesicles by IR spectroscopy: Fast and simple classification based on amide and CH stretching vibrations, Biochim Biophys Acta Biomembr. 1859 (2017) 459–466. https://doi.org/10.1016/j.bbamem.2016.12.005.
[360] A. Drożdż, A. Kamińska, M. Surman, A. Gonet-Surówka, R. Jach, H. Huras, M. Przybyło, E.Ł. Stępień, Low-Vacuum Filtration as an Alternative Extracellular Vesicle Concentration Method: A Comparison with Ultracentrifugation and Differential Centrifugation, Pharmaceutics. 12 (2020) 872. https://doi.org/10.3390/pharmaceutics12090872.
[361] M. Dash, K. Palaniyandi, S. Ramalingam, S. Sahabudeen, N.S. Raja, Exosomes isolated from two different cell lines using three different isolation techniques show variation in physical and molecular characteristics, Biochim Biophys Acta Biomembr. 1863 (2021) 183490. https://doi.org/10.1016/j.bbamem.2020.183490.
[362] A. Zlotogorski-Hurvitz, B.Z. Dekel, D. Malonek, R. Yahalom, M. Vered, FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer, J Cancer Res Clin Oncol. 145 (2019) 685–694. https://doi.org/10.1007/s00432-018-02827-6.
[363] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zirah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies, Anal. Chem. 78 (2006) 363–371. https://doi.org/10.1021/ac051444m.

[364] N. Mantel, The detection of disease clustering and a generalized regression approach, Cancer Res. 27 (1967) 209–220.
[365] S. Navarro, D. Borchman, E. Bicknell-Brown, Lipid-protein ratios by infrared spectroscopy, Anal Biochem. 136 (1984) 382–389. https://doi.org/10.1016/0003-2697(84)90233-1.
[366] J. Ollesch, S.L. Drees, H.M. Heise, T. Behrens, T. Brüning, K. Gerwert, FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification, Analyst. 138 (2013) 4092–4102. https://doi.org/10.1039/c3an00337j.
[367] R. Bittman, Glycerolipids: Chemistry, in: G.C.K. Roberts (Ed.), Encyclopedia of Biophysics, Springer, Berlin, Heidelberg, 2013: pp. 907–914. https://doi.org/10.1007/978-3-642-16712-6_527.

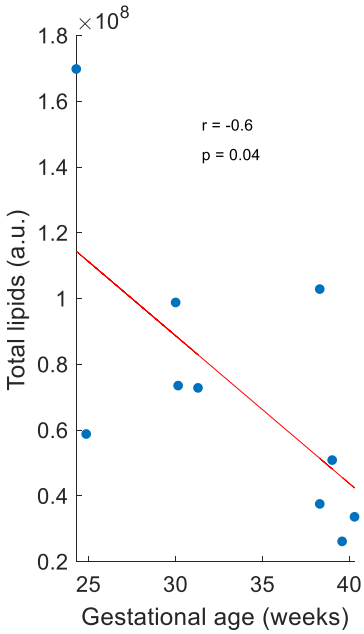
Annex I. Supplementary figures and tables

AI.1. Parameters employed for the simulation of EVs in the different samples

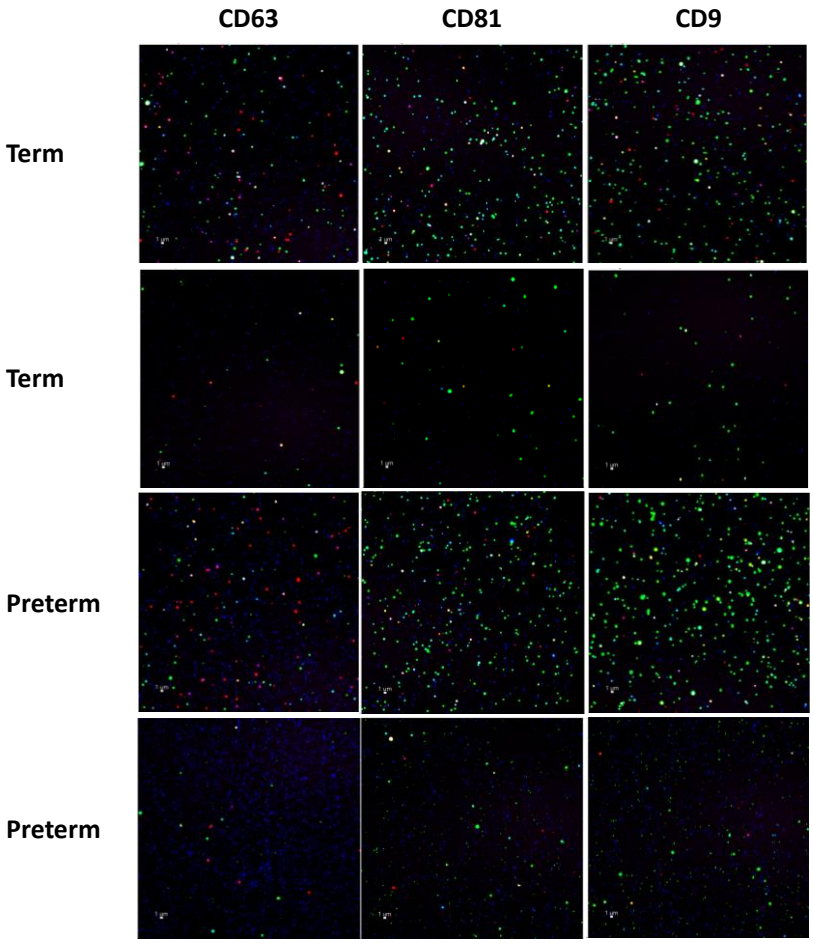
	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**
N	100	100	200	200
r (nm)	50	100	50	100
α_p	1/750	1/750	1/750	1/750
β_p	1/1000	1/1000	1/1000	1/1000
α_{L1}	1/100000	1/100000	1/100000	1/100000
B_{L1}	1/100	1/100	1/100	1/100
α_{L2}	1/200000	1/200000	1/200000	1/200000
β_{L2}	1/200	1/200	1/200	1/200

Note: N, number of particles; r, radius; α and β, number of proteins (p), lipid 1 (L1) and lipid 2 (L2) per surface and volume in the EVs, respectively. For the radius, a normal distribution with a mean value as indicated and σ value of 15 was considered.

AI.2. Pearson's correlation of total lipids of HM-EVs isolates from preterm and term samples with gestational age

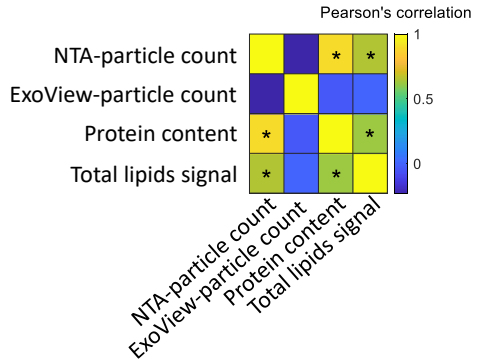


AI.3. Tetraspanin fluorescent staining analysis of EVs isolated from term and preterm HM samples with the different capture probes



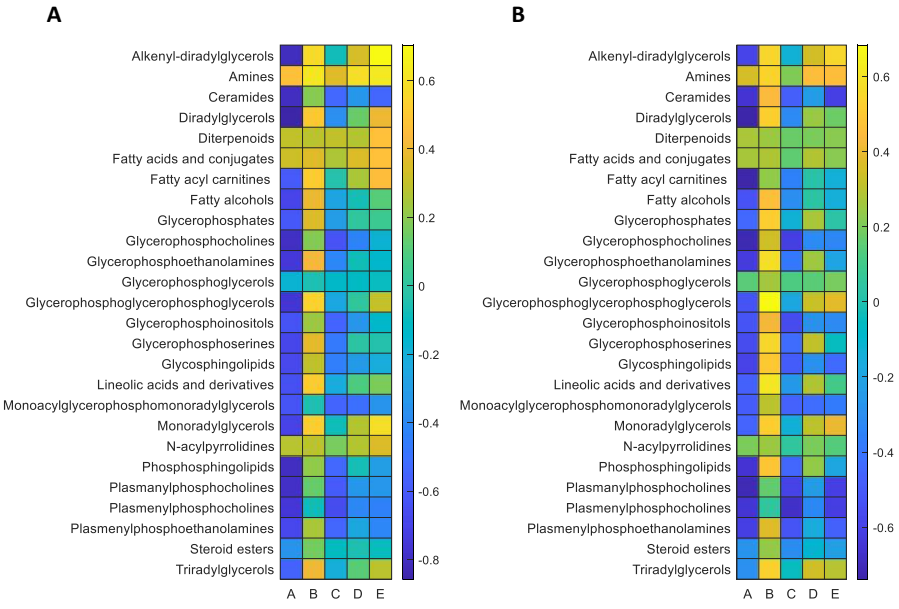
Note: scale bar, 1 μm; color code for fluorescent single positive particles: CD81+ (green), CD63+ (red) and CD9+ (blue); color code for colocalized fluorescent positive particles: CD81+/CD63+ (yellow), CD81+/CD9+ (cyan), CD9+/CD63+ (magenta), and CD81+/CD63+/CD9+ (white).

AI.4. Pearson's correlation between the different characteristics of HM-EVs



Note: **p*-value < 0.05.

AI.5. Spearman's correlation of lipid sub-classes with different parameters



Note: Correlation with gestational age, corrected by postnatal age (A) and vice versa (B) for the different normalization strategies: sample volume (A), particle count determined by NTA analysis (B), particle count determined by ExoView platform (C), protein content determined by BCA assay (D), and total lipids determined by LC-MS analysis (E). **p*-value < 0.05.

Annex II. Articles included in the compendium

frontiers | Frontiers in Pediatrics

TYPE Study Protocol
PUBLISHED 18 April 2023
DOI 10.3389/fped.2023.1130179

OPEN ACCESS

EDITED BY
Giovanna Verlato,
University Hospital of Padua, Italy

REVIEWED BY
Clair-Yves Boquien,
INRA Centre Angers-Nantes Pays de la Loire, France
Aleksandra Maria Wesolowska,
Medical University of Warsaw, Poland

*CORRESPONDENCE
Julia Kuligowski
julia.kuligowski@uv.es

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Neonatology, a section of the journal Frontiers in Pediatrics

RECEIVED 22 December 2022
ACCEPTED 31 March 2023
PUBLISHED 18 April 2023

CITATION
Ramos-Garcia V, Ten-Doménech I, Moreno-Giménez A, Campos-Berga L, Parra-Llorca A, Ramón-Beltrán A, Vaya MJ, Mohareb F, Molitor C, Refinetti P, Silva A, Rodrigues LA, Rezzi S, Hodgson ACC, Canarelli S, Bathrellou E, Mamalaki E, Karipidou M, Poulimeneas D, Yannakoulia M, Akhgar CK, Schwaighofer A, Lendl B, Karrer J, Migliorelli D, Generelli S, Gormaz M, Vasileiadis M, Kuligowski J and Vento M (2023) Fact-based nutrition for infants and lactating mothers—The NUTRISHIELD study.
Front. Pediatr. 11:1130179.
doi: 10.3389/fped.2023.1130179

Fact-based nutrition for infants and lactating mothers—The NUTRISHIELD study

Victoria Ramos-Garcia[1†], Isabel Ten-Doménech[1†], Alba Moreno-Giménez[1], Laura Campos-Berga[1], Anna Parra-Llorca[1], Amparo Ramón-Beltrán[2], María J. Vaya[3], Fady Mohareb[4], Corentin Molitor[4], Paulo Refinetti[5], Andrei Silva[5], Luis A. Rodrigues[5], Serge Rezzi[6], Andrew C. C. Hodgson[6], Stéphane Canarelli[6], Eirini Bathrellou[7], Eirini Mamalaki[7], Melina Karipidou[7], Dimitrios Poulimeneas[7], Mary Yannakoulia[7], Christopher K. Akhgar[8], Andreas Schwaighofer[8], Bernhard Lendl[8], Jennifer Karrer[9], Davide Migliorelli[10], Silvia Generelli[10], María Gormaz[1,2], Miltiadis Vasileiadis[11], Julia Kuligowski[1*] and Máximo Vento[1,2] on behalf of the NUTRISHIELD team

[1]Neonatal Research Group, Health Research Institute La Fe (IISLAFE), Valencia, Spain, [2]Division of Neonatology, University & Polytechnic Hospital La Fe (HULAFE), Valencia, Spain, [3]Blood Transfusion Center from the Valencian Community, Valencia, Spain, [4]The Bioinformatics Group, Cranfield Soil and Agrifood Institute, Cranfield University, Bedford, United Kingdom, [5]REM Analytics S.A., Monthey, Switzerland, [6]Swiss Nutrition and Health Foundation, Epalinges, Switzerland, [7]Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece, [8]Research Division of Environmental Analytics, Process Analytics and Sensors, Institute of Chemical Technologies and Analytics, Technische Universität Wien, Vienna, Austria, [9]Quantared Technologies GmbH, Vienna, Austria, [10]Swiss Center for Electronics and Microtechnology (CSEM), Landquart, Suiza, [11]ALPES Lasers S.A., St Blaise, Switzerland

Background: Human milk (HM) is the ideal source of nutrients for infants. Its composition is highly variable according to the infant's needs. When not enough own mother's milk (OMM) is available, the administration of pasteurized donor human milk (DHM) is considered a suitable alternative for preterm infants. This study protocol describes the NUTRISHIELD clinical study. The main objective of this study is to compare the % weight gain/month in preterm and term infants exclusively receiving either OMM or DHM. Other secondary aims comprise the evaluation of the influence of diet, lifestyle habits, psychological stress, and pasteurization on the milk composition, and how it modulates infant's growth, health, and development.

Methods and design: NUTRISHIELD is a prospective mother-infant birth cohort in the Spanish-Mediterranean area including three groups: preterm infants <32 weeks of gestation (i) exclusively receiving (i.e., >80% of total intake) OMM, and (ii) exclusively receiving DHM, and (iii) term infants exclusively receiving OMM, as well as their mothers. Biological samples and nutritional, clinical, and anthropometric characteristics are collected at six time points covering the period from birth and until six months of infant's age. The genotype, metabolome, and microbiota as well as the HM composition are characterized. Portable sensor prototypes for the analysis of HM and urine are benchmarked. Additionally, maternal psychosocial status is measured at the beginning of the study and at month six. Mother-infant postpartum bonding and parental stress are also examined. At six months, infant neurodevelopment scales are applied. Mother's concerns and attitudes to breastfeeding are registered through a specific questionnaire.

Discussion: NUTRISHIELD provides an in-depth longitudinal study of the mother-infant-microbiota triad combining multiple biological matrices, newly developed analytical methods, and *ad-hoc* designed sensor prototypes with a wide range of clinical outcome measures. Data obtained from this study will be used to train a machine-learning algorithm for providing dietary advice to lactating mothers and will be implemented in a user-friendly platform based on a combination of user-provided information and biomarker analysis. A better understanding of the factors affecting milk's composition, together with the health implications for infants plays an important role in developing improved strategies of nutraceutical management in infant care.
Clinical trial registration: https://register.clinicaltrials.gov, identifier: NCT05646940.

KEYWORDS

preterm infants, lactation, human milk, breastfeeding, nutrition, donor human milk

1. Introduction

Breastfeeding (BF) is the optimal feeding practice for all infants, associated with numerous health benefits for the mother–infant dyad, as well as, being an ecologic practice (1). The World Health Organization highly recommends exclusive BF for all infants up to six months, born either full-term (2, 3) or prematurely (4). Human Milk (HM) is a dynamic fluid that meets the offspring's needs (5–8), and is ample in nutrients and bio-active compounds intended to enhance immunity and growth (9). HM also hosts a complex ecosystem of microbiota (10), which is of paramount importance for the offspring's immunity (5).

During the last decades, the incidence of preterm deliveries (<37 weeks of gestation) and survival rate of preterm infants (PI) have been steadily increasing (11), hand in hand with the interest in preterm and early infant nutrition. Associated birth complications are the leading cause of death among children under five years of age, responsible for ~1 million global deaths per year (12). Progress in medical interventions has allowed to enhance survival of an increasing proportion of extremely low gestational age newborns and low birth-weight infants. Early infant nutrition has become a major player in improving clinical outcomes of survivors (13). HM is recommended for PI based on an impressive array of benefits provided to this highly vulnerable population, including ameliorated immunological and gastrointestinal outcomes (14–20). Optimal growth in the postnatal period is challenging dure to increased nutritional demands and metabolic and digestive immaturity of PI. Greater weight gain is associated with better neurodevelopment outcomes (21).

In situations where mothers are unable to produce sufficient milk quantities to exclusively or partially breastfeed, pasteurized donor human milk (DHM) is a viable option to avoid formula feeding, especially for low birth-weight children (22). DHM is a valuable but limited resource and Human Milk Banks prioritize its distribution to the patients with highest risk of necrotizing enterocolitis, generally PI <32 weeks of gestation or with a birthweight <1,500 g (3, 23). To date, most studies focus on the benefits of using DHM over formula in PI, when own mother's milk (OMM) is limited or unavailable. There is clear evidence of the superiority of OMM against formula (24); however, although DHM apparently protects against necrotizing enterocolitis as compared to formula, results are not conclusive (25) and there are only few reports directly comparing DHM and OMM. While providing some bioactive agents, DHM consumption is associated with slower growth rates in comparison to the administration of OMM or formula (24). This finding may be attributable to several factors. Most DHM is provided by women who have delivered at term and donate their milk in later stages of lactation up to several months after delivery. The effect of gestational age on HM composition has been reported to affect a wide array of compounds including lipids (26), lactoferrin (27), amino acids (28), and HM oligosaccharides (29). In comparison to preterm milk during the first weeks after delivery, studies on the mean composition of DHM show a lower content in total protein, fat, and other bioactive molecules (30). Its composition is also affected by the processing of expressed milk, including stringent protocols applied in HM banks, i.e., pasteurization, freezing, and storage (31, 32), necessary to reduce the potential risk to transmit infectious agents.

Beyond the physiological adaptations, research has focused on the potential effects of the maternal diet on the HM composition. In an earlier report, HM composition was found rather different among mothers of diverse ethnic backgrounds, with different dietary habits (33). The fatty acid profile of HM varies in relation to maternal diet, particularly, in the long chain polyunsaturated fatty acids (LCPUFAs) (31, 32, 34, 35). Additionally, in a recent comprehensive systematic review of 104 observational and interventional studies, HM composition was found to relate to the maternal consumption of fatty acids, fat-soluble vitamins, and vitamin B1 and C intake (36, 37). No similar relationships were found for dietary intake of iron, folate, calcium, selenium, or proteins (36).On the contrary, another recent systematic review determined that information regarding the relationship between dietary patterns during lactation and HM content is insufficient for total fat, vitamins B and C, and choline content and inexistent for total proteins, macronutrient distribution, human milk oligosaccharides (HMOs), vitamins A, D, E and K, iodine, and selenium (38). Hence, more studies are needed to conclude for several nutrients and little is known on how the maternal diet may also affect non-nutritive constituents of HM, e.g., prebiotics content (39) or microbiota (40, 41). Finally, existing evidence

comes mainly from studies on full-term infants (TI); the impact of maternal nutrition on the composition of HM for the preterm offspring has not been adequately documented.

Therefore, the main objective of this study is to compare the % weight gain/month in PI and TI exclusively receiving either OMM or DHM. In addition, secondary objectives are (i) to evaluate associations between the mother's diet, physical, and psychosocial status (e.g., lifestyle, perceived stress, anxiety, and depression) and HM composition in PI and TI, (ii) to evaluate the effect of pasteurization/storage on DHM composition, (iii) to assess the interplay of microbiota and microbiota activity and HM composition, (iv) to assess the interplay of microbiota and microbiota activity detected in HM and PIs and TIs, (v) to evaluate the impact of microbiota and microbiota activity on the vitamin status of PIs and TIs, (vi) to test the performance of novel sensor devices developed within this project, and (vii) the development of a personalized nutrition algorithm for lactating mothers.

2. Methods and analysis

2.1. Study design

The NUTRISHIELD study is a parallel group, non-randomized, observational study performed at the Division of Neonatology of the University & Polytechnic Hospital La Fe (HULAFE), including TI and PI and their mothers, covering the period from birth to 24 months of infant's age. In addition, mothers providing DHM to study participants are also included.

The study protocol has been approved by the Scientific and Ethics Committee for Biomedical Research (CEIm) of the HULAFE (#2019-289-1). All methods have been performed in accordance with the relevant guidelines and regulations and written permission has been obtained from mothers or legal representatives by signing an informed consent form.

Confidentiality of subjects is maintained during the study. All participants have been assigned a code and data allowing personal identification will not be shared at any time. Participants may withdraw consent for participating in the study at any time.

2.2. Study population and recruitment

Participants were recruited at the end of pregnancy (when admitted to the obstetric ward) or within one week after birth in case of hospitalized infants at HULAFE. Recruitment started in October 2020 and was completed in July 2022. Three mother-infant groups were enrolled, i.e., PI fed with OMM (PI-OMM), PI fed with pasteurized DHM (PI-DHM), and TI receiving OMM (TI-OMM), and their mothers.

The inclusion criteria were (i) acceptance of the mother to participate and sign an informed consent form, (ii) a gestational age <32 weeks for the group of PI and >37 weeks for the group of TI, and (iii) exclusive consumption (i.e., >80% of total intake) of either OMM or DHM at time point CEN (complete enteral nutrition, 150 mL/Kg/day) for PI or RBW (recovery of birth weight) for TI. Although the type of feeding of PIs included in either group (PI-OMM or PI-DHM) may have changed at later time points, they remained in the same study group, so the effect of feeding of each type of milk at CEN and its long-term effects can be studied. Likewise, the use of infant formula and the initiation of weaning at later time points was recorded without affecting follow-up assessments. The exclusion criteria were (i) non-compliance with any of the inclusion criteria, (ii) the requirement of a special diet for the mother (e.g., celiac disease, diabetes) or maternal consumption of probiotics, (iii) the need of intestinal surgery, severe congenital malformations, or chromosomopathies of the infant, (iv) mother's residence outside the Valencian Community and (v) severe language barriers hampering the collection of necessary data from mothers not speaking Spanish and/or English.

Sample size estimation was performed for the primary outcome (i.e., % weight gain/month from birth until hospital discharge of PI fed with DHM and OMM) and was based on result of a previous study (42). Considering the median % weight gain/month (interquartile range, IQR) of 52% (IQR: 30) in the PI-DHM group and 63% (IQR: 21) in the PI-OMM, 18 infants per group were needed to achieve a power of 90% with an alpha error of 5% between the % weight gain/month of PI exclusively receiving either OMM or DHM at CEN.

On the other hand, HM donors were recruited during their regular visits at the hospital's HM Bank of the Valencian Community. All participants met ordinary criteria for HM donation (e.g., negative screening results of a series of transmissible diseases, toxic habits, or some chronic medication) and accepted to sign an informed consent form and participate in the study.

2.3. Assessment points and biological samples

Samples from lactating mothers and their infants are collected at different timepoints, as shown in **Figure 1**. At birth, cord blood, and urine, and faeces from both infants and mothers are collected when possible. When PIs achieve CEN and TIs RBW, as well as one, two, three, and six months after delivery, urine and faeces from mothers and infants, as well as, HM from lactating mothers, are collected. Infants' and mothers' buccal swabs are collected at month six. During hospitalization of participants, samples are collected at the hospital by healthcare providers. After discharge, samples are collected at home and either transported to the hospital by study participants or staff of the NUTRISHIELD study. Mothers receive Standard Operation Procedures (SOPs) with detailed instructions and all the required material for sample collection at home. Additionally, hospital's staff instructs mothers on how to collect and store samples correctly.

Samples collected at home are stored at 4°C during a maximum of 24 h before their transport to the hospital on ice, where they are registered, aliquoted, labelled, and stored at −80°C until analysis. A HM and faeces aliquot is directly stored in a tube containing a

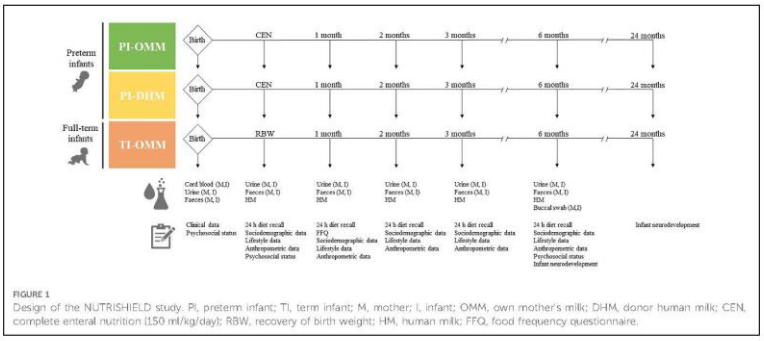


FIGURE 1
Design of the NUTRISHIELD study. PI, preterm infant; TI, term infant; M, mother; I, infant; OMM, own mother's milk; DHM, donor human milk; CEN, complete enteral nutrition (150 ml/kg/day); RBW, recovery of birth weight; HM, human milk; FFQ, food frequency questionnaire.

microbiome preservative solution. For DHM, collection times respective to delivery are heterogeneous, since from each donor several aliquots are collected over time, pooled, and then subjected to Holder pasteurization in batches (i.e., 62.5°C during 30 min followed by rapid cooling to 4°C). The lactation time was estimated as the mean (standard deviation, SD) of time elapsed during the collection of different aliquots from one DHM batch, being 16 (SD 12) days.

Mothers' and infants' buccal swab samples are collected by thoroughly rubbing a flocked sterile swab up and down the inner side of both cheeks for 30 s each. Swabs are cut and stored in a sterile tube at −80°C until further analysis. Discarded arterial and venous blood from the umbilical cord is collected in EDTA blood sample tubes after the placenta is delivered and separated from the baby. Blood is centrifuged (1,300 × *g* for 10 min at 20°C) and the upper plasma layer is stored in opaque vials at −80°C until analysis. Mother's first morning urine is collected in a polypropylene container, and infant's urine is collected by placing sterile cotton pads in the diaper and, after urinating, squeezing them with a sterile plastic syringe into sterile tubes (see **Figure 2**). Mother's faeces are collected in a polypropylene container, and infant's faeces are collected directly from the diaper into sterile tubes using sterile tweezers.

Mothers express milk using breast milk pumps following an SOP employed routinely in the hospital and the HM bank. The removable parts of the breast milk pump as well as the collection bottles are sterilized before their use. HM is collected at least three hours after breastfeeding. Mothers are required to wash their hands with soap and water and clean nipples with water. HM should be preferably expressed between 7 and 10 a.m., full expression of one breast is required (a minimum volume of 50 ml is recommended), and details regarding the extraction method (e.g., breast pump brand), date, and extracted volume are registered. For DHM samples, aliquots of each pooled DHM sample are collected before and after Holder pasteurization.

2.4. Dietary and psychological status assessments

For all three study groups, information obtained from clinical records is collected. Anthropometric data (e.g., weight, height,

FIGURE 2
Infants' urine sample collection procedure. Sterile cotton pads are placed in the diaper (left); cotton pads soaked with urine are collected using sterile tweezers (middle); cotton pads are squeezed with a sterile syringe to collect the urine sample (right).

and head circumference), sociodemographic, dietary, other relevant information related to lifestyle (e.g., smoking habits, sleep hours, and physical activity), and psychosocial status are provided through questionnaires by participating mothers as summarized in **Table 1**. As shown in **Figure 1**, 24 h dietary recalls are recorded at all timepoints except at birth, food frequency questionnaire (FFQ) at month 1, maternal psychosocial status at CEN/RBW and month six (corrected age for PI groups), and infant neurodevelopment at months six and twenty-four. The questionnaires are administered online, on paper, or through direct interviews with the mothers.

Regarding the 24 h recall, trained researchers ask for all foods and beverages participants consumed during the previous day, using the multiple-pass method (43). Recall data are analyzed in terms of nutrients using the dietary analysis software Nutritionist Pro™ (2007, Axxya Systems, Texas, USA). Additionally, dietary intake is grouped into food groups, (i.e., fruits, vegetables, bread/starch, meat/high fat, meat/medium fat, meat/low fat, meat/very low fat, milk/non-fat, milk/low fat, milk/full fat, eggs, fish, soy, dairy products, soft drinks, and other carbohydrate-rich foods).

TABLE 1 Recorded parameters from lactating mothers and infants.

Assessment	Parameter
Nutrition	FFQ
	24-hour dietary recall
Clinical data	Maternal conditions
	Pregnancy complications
	Delivery type
	Medical information
	Maternal medication
	IIFAS
	Infant's gestational age
	Infant sex
	Infant diagnosis
	Perinatal complications
Sociodemographic data	Origin
	Education
	Household income
	Employment status
Lifestyle	Smoking habits
	Physical activity
	Sleeping habits
Anthropometric data	Maternal weight
	Infant height
	Infant head circumference
Psychosocial status	MSPSS
	FACES
	PSS-10
	STAI
	EPDS
	BDI
	TEQ
	PBQ
	PSI
Infant neurodevelopment	ASQ-3
	IBQ
	Merril-Palmer revised scales

FFQ, Food Frequency Questionnaire; IIFAS, Iowa Infant Feeding Attitude Scale; BMI, Body Mass Index; MSPSS, Multidimensional Scale of Perceived Social Support; FACES, Family Adaptability and Cohesion Evaluation Scale; PSS-10, Perceived Stress Scale-10; STAI, State-Trait Anxiety Inventory; EPDS, Edinburgh Postnatal Depression Scale; BDI, Beck Depression Inventory; TEQ, Traumatic Experience Questionnaire; PBQ, Postpartum Bonding Questionnaire; PSI, Parenting Stress Index; ASQ-3, Ages & Stages Questionnaires; IBQ, Infant Rothbart's Temperament Questionnaire.

The FFQ is administered by trained personnel, and it comprises 142 questions on the consumption of foods that are commonly eaten by the Spanish population throughout a year, including dairy products, cereals, fruits, vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants, and sweets. Using a 9-grade scale ("never or less than 1 time/month", "1–3 times/month", "1 time/week", "3–4 times/week", "5–6 times/week", "1 time/day", "2–3 times/day", "4–5 times/day", "≥6 times/day") participants are required to indicate the absolute frequency of consuming a certain amount of food, expressed in grams, milliliters or in other common measures, such as slice, tablespoon, or cup, depending on the food. The previous month is set as the timeframe. The FFQ is an easy-to-use questionnaire and is not expected to increase the burden of lactating mothers.

Based on the FFQ-responses, adherence to the Mediterranean diet is evaluated by using the MedDietScore, a composite score calculated for each participant (44). For food groups presumed to be part of the Mediterranean pattern (i.e., those with a recommended intake of 4 servings per week or more, such as non-refined cereals, fruits, vegetables, legumes, olive oil, fish, and potatoes) higher scores are assigned when the consumption is according to the rationale of the Mediterranean pattern, while lower scores are assigned when participants report no, rare, or moderate consumption. For the consumption of foods presumed to be eaten less frequently within the Mediterranean diet (i.e., consumption of meat and meat products, poultry, and full fat dairy products), scores are assigned on a reverse scale. As this study is focused on lactating mothers, the original score was modified by removing the alcohol consumption component. Thus, the range of this modified MedDietScore is between 0 and 50, with higher values of the score indicating greater adherence to the Mediterranean diet.

Maternal psychosocial status is measured at recruitment and at six months, including social support [Multidimensional Scale of Perceived Social Support, MSPSS (45)], family functioning [Family Adaptability and Cohesion Evaluation Scale, FACES (46)], perceived stress [Perceived Stress Scale -10, PSS-10 (47)], anxiety [State-Trait Anxiety Inventory, STAI (48)], depression [Edinburgh Postnatal Depression Scale, EPDS (49), Beck Depression Inventory (BDI) (50)] symptoms, and traumatic life events [Traumatic Experience Questionnaire, TEQ (51, 52)]. Mother-infant postpartum bonding [Postpartum Bonding Questionnaire, PBQ (53)] and parental stress [Parenting Stress Index, PSI (54)] are also examined.

As for the infants, at six and twenty-four months (corrected age for PIs) assessment of neurodevelopment is conducted during a hospital visit by a trained psychologist and psychiatrist from the team. Parents complete the Ages & Stages Questionnaires, Third Edition (ASQ-3) (55) for psychomotor development evaluation and the Infant Rothbart's Temperament Questionnaire (IBQ) (56) for infant temperament. The ASQ is a well standardized

instrument used in clinical and research practice to examine children's psychomotor development gathering items in five domains: communication, gross motor, fine motor, problem solving, and personal-social. Clinical observation and Merrill-Palmer revised scales (57) for development are applied during the session.

Mother's concerns and attitudes to breastfeeding are also registered through the Iowa Infant Feeding Attitude Scale (IIFAS) (58) at CEN/RBW and six months.

Milk donors are asked to provide sociodemographic, medical, dietary, and other lifestyle data. PSS-10 is evaluated when entering the study in order to correlate with hormones and other compounds present in the milk. In particular, each time a batch of milk is prepared, a FFQ is administered (representing the period of time that the batch of milk covered) and the donors are asked to provide medical, anthropometric, and lifestyle data.

2.5. Analysis of biological samples employing laboratory methods

A range of state-of-the-art as well as novel laboratory methods are employed for the analysis of the collected biological samples, as summarized in **Table 2**.

2.5.1. Genome sequencing

For genome sequencing analysis, DNA extraction from buccal swab samples takes place according to the protocol developed by the sequencing provider facility. Polymerase chain reaction (PCR) quality control (QC) is performed before samples are shipped to the sequencing service provider, and again upon receival by the sequencing facility. QC and library preparation follow the Illumina NovaSeq protocol (Paired End, 150 bp). Three polygenic risk scores (PRS) models are developed for body mass index (BMI), diabetes type 2, and lactose intolerance, calculated by Plink software (59), and the resulting genotype files, one per sample, are then used as separate input to the PRS models.

This score can then be compared to the scores obtained from the UK Biobank set (60), to estimate the relative risk for this individual compared to the rest of the UK Biobank individuals (~370.000 individuals), which are split in 10 quantiles according to their genetic risk.

2.5.2. Microbiota analysis

Advanced Testing for Genetic Composition (ATGC) is used for microbiota analysis of faeces and HM samples (61). It is a targeted measurement technique that provides greater precision, specificity, and versatility than rtPCR, while being just as cost-effective and fast. It has been developed to translate discovery data from untargeted analysis (e.g., 16S or shotgun meta-genomics) into tests for routine use.

The Bifidobacterium (Bb) assay used is specifically designed to analyze the main species and subspecies varieties in HM and newborn gut microbiomes. For sample processing, the HM lipid layer is solubilized using a detergent, while for faeces, a lysis step using a lysis buffer with bead beating for 20 s is required. Then, an in-house DNA extraction procedure and a PCR are carried out. Cycling temperature capillary electrophoresis (CTCE) is made for each primer on an independent capillary. It is performed on a MegaBace 1,000 instrument (General Electrics, Boston MA, USA) as described earlier (62).

TABLE 2 Analysis of biological samples employed.

Analysis	Parameters	Technique	Matrix
Genome	Whole genome sequencing	Illumina NovaSeq	Buccal swab
Microbiota	Bb species	ATGC	Faeces
			HM
Microbiota activity	SCFAs BCAAs	GC-MS	Faeces
			Urine
			HM
	Bas	LC-MS/MS	Urine
Nutrition	Flavonoids	LC-MS/MS	Urine
	Isoflavones		
	Arylglycines		
	Amino acids		
Metabolome	Metabolomic fingerprinting	LC-HRMS	Urine
Macronutrients	Fat	Miris HM analyzer	HM
	Carbohydrates		
	Crude & true protein		
	Total solids		
	Energy		
Fatty acids	34 fatty acids (C6-C24)	GC-MS	HM
	SAT	HM fatty acid sensor	
	MONOs		
	PUFAs		
	UNSAT		
	SCFAs		
	MCFAs		
	LCFAs		
Vitamins	Vitamin B, D and K groups	LC-MS/MS	HM Umbilical cord blood
	Retinol forms	LC-UV	
	Carotenoids and vitamin E groups	LC-UV/ Fluorescence	
HMOs	HMO screening	LC-HRMS	HM
Steroids	19 steroid hormones	LC-MS/MS	HM
			Urine
Proteins	Total protein	HM protein sensor	HM
	Alpha-lactalbumin		
	Lactoferrin		
	Casein		
pH	pH	Urine pH sensor	Urine
Phosphate and creatinine	Phosphate and creatinine	Phosphate and creatinine sensor	Urine

ATGC, Advanced Testing for Genetic Composition; Bb, Bifidobacterium; GC, Gas Chromatography; (HR)MS, (High Resolution) Mass Spectrometry; SCFAs, Short Chain Fatty Acids; BCAAs, Branched Chain Amino Acids; Bas, Bile Acids; SAT, saturated fatty acids; MONOs, monounsaturated fatty acids; PUFAs, polyunsaturated fatty acids; UNSAT, unsaturated fatty acids; MCFAs, medium-chain fatty acids; LCFAs, long-chain fatty acids; LC, Liquid Chromatography; UV, Ultraviolet; HMOs, Human Milk Oligosaccharides.

2.5.3. Microbiota activity biomarkers analysis

The microbiota activity is evaluated with the following methods:

(i) Short chain fatty acids (SCFAs) and branched chain amino acids (BCAAs) in faeces, urine and HM samples

SCFAs (i.e., acetic, propionic, isobutyric, butyric, 2-methylbutyric, isovaleric, valeric, caproic and heptanoic acids) and BCAAs (i.e., valine, leucine and isoleucine) are determined by targeted gas chromatography coupled to mass spectrometry (GC-MS), as described elsewhere (63, 64), using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector (Agilent Technologies, Santa Clara, CA, USA).

(ii) Bile acids (BAs) in urine samples

In total, 34 BAs (6 primary, 7 primary conjugated, 5 secondary, 10 secondary conjugated and 6 sulphated) are determined in urine samples by liquid chromatography coupled to tandem MS (LC-MS/MS), as described elsewhere (65) with minor modifications, using an ACQUITY LC chromatograph (Waters Ltd, Elstree, UK) coupled to a Xevo TQ-S MS detector (Waters, Manchester, UK).

2.5.4. Nutrition biomarkers analysis

Quantification and semi-quantification of 20 and 205 urinary nutrition biomarkers (e.g., flavonoids, isoflavones, and arylglycines), respectively, related to nine food groups such as fruits, vegetables, meat, fish, dairy products, milk, seeds, coffee, and soft drinks is carried out (66, 67). Additionally, seven microbiota activity biomarkers are determined using this method (i.e., phenylpropionylglycine, L-kynurenine, L-tyrosine, hippuric acid, 3-indolepropionic acid, ferulic acid sulphate and 3-indoleacetic acid). LC-MS/MS analysis is conducted using a Sciex QTRAP 6500 + system (Sciex, Framingham, Massachusetts, USA).

2.5.5. Untargeted metabolomic fingerprinting

Untargeted metabolomic analysis is carried out in urine samples by LC-MS/MS, employing an Agilent 1,290 Infinity HPLC system coupled to an Agilent 6,550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS detector (Agilent Technologies, Santa Clara, CA, USA), as previously described (68).

2.5.6. Macronutrients analysis

Direct measurement of HM macronutrients including fat, carbohydrates, crude and true proteins, total solids, and energy are determined using a Miris HM analyzer (Miris AB, Uppsala, Sweden). Determinations, QC, and instrument calibration are performed following the SOP provided by the manufacturer (69).

2.5.7. Fatty acid profile analysis

The targeted analysis of 36 fatty acids in HM samples is carried out by GC-MS as described elsewhere (70, 71), using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector (Agilent Technologies, Santa Clara, CA, USA).

2.5.8. Vitamin analysis

Quantification of water-soluble (group of vitamin B: thiamine, thiamine monophosphate, riboflavin, flavin adenine dinucleotide, nicotinamide, pyridoxal and pyridoxal phosphate) and lipid-soluble vitamins (group of vitamin A: retinol forms, β-carotene, β -cryptoxanthin, lutein, lycopene and zeaxanthin; group of vitamin E: α-tocopherol and γ-tocopherol; group of vitamin K: phylloquinone and menaquinone-4; group of vitamin D: cholecalciferol and calcifediol) in HM and umbilical cord blood samples is carried out.

The analysis of B, D and K vitamins are determined as described somewhere (72–76), using an Acquity LC chromatographic system (Waters AG, Switzerland) hyphenated to a Xevo TQ-S mass spectrometer (Waters AG, Switzerland).

The retinol forms are analyzed as previously described (75), using a LC using a quaternary Flexar chromatographic system (Perkin Elmer, Switzerland) with UV detection (325 nm).

Carotenoids and vitamin E analysis is carried out as previously described (77), using a binary Acquity LC chromatographic system (Waters AG, Switzerland) with multiple UV detection (295 nm, 450 nm and 472 nm) and fluorescence detection (exc. 296 nm/em. 330 nm).

2.5.9. Oligosaccharides analysis

HMOs analysis is carried out by LC-MS in accordance with a previously described protocol (78, 79) with some modifications, employing a Vanquish LC Binary Pump coupled to an Orbitrap QExactive Plus MS detector (ThermoFisher, Waltham, MA, USA).

2.5.10. Steroid analysis

Steroid analysis of HM and urine samples is carried out on an Acquity UPLC system (Waters Ltd, Elstree, UK) coupled to a Waters Xevo TQ-S MS detector (Waters, Manchester, UK) as previously described (80, 81). A panel of 19 steroids is targeted (i.e., cortisol, 5β-tetrahydrocortisol, 6β-hydrocortisol, 20α-dihydrocortisol, 20β-dihydrocortisol, corticosterone, aldosterone, estrone, androstenedione, dehydroepiandrosterone, progesterone, 17-hydorxy-progesterone, pregnenolone, cortisone, 20α-dihydrocortisone, 20β-dihydrocortisone, 6α-hydroxycortisone, testosterone, and 5α-dihydrotestosterone).

2.6. Sensor prototypes

Due to the special requirements of the study, three sensor prototypes enabling the determination of complementary parameters in HM and urine samples are being developed and benchmarked within the NUTRISHIELD study.

2.6.1. HM protein sensor

Direct quantification of the most abundant proteins (i.e., casein, α-lactalbumin, and lactoferrin) in HM milk is performed using quantum cascade laser-based mid-infrared (QCL-IR) spectroscopy (82, 83). Laser-based IR spectroscopy enables more robust and sensitive analysis in the spectral region of IR signatures of proteins compared to conventional Fourier-transform infrared (FTIR) spectrometers (82). Reference protein analysis is carried out using specific Kjeldahl and HPLC analysis. Within the study, a dedicated QCL-based protein analyzer for HM is developed and benchmarked.

2.6.2. HM fatty acids sensor

HM fatty acid profiling is performed based on mid-IR spectroscopy of an extracted lipid HM fraction. For analysis, 15 µl of the pure fat are transferred onto a diamond single bounce attenuated total reflection (ATR) accessory (Platinum ATR, Bruker, Ettlingen) connected to a Tensor 37 (Bruker, Ettlingen) FTIR spectrometer (84, 85).

2.6.3. Urine pH sensor

A custom-made portable system based on potentiometric measurements and screen-printed electrode technology is developed to measure pH in urine. The system works with small sample volumes (i.e., 50 µl) and no dilution is needed. An additive is spiked into samples prior to measurement in order to preserve and regenerate the sensor's surface, which extends sensor lifetime to multiple uses with no carry-over between urine samples.

2.6.4. Urine phosphate and creatinine sensor

Quantification of phosphate and creatinine in urine samples is performed using a method based on FTIR transmission spectroscopy. Within the study, a dedicated QCL-based urine analyzer is developed. Reference values for phosphate are obtained by a colorimetric determination after reaction of inorganic phosphorus with ammonium molybdate, while creatinine reference values are obteined following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA).

2.7. Statistics and data integration

The Student's *t*-test, Wilcoxon ranksum test or X^2 test with $\alpha = 0.05$ will be used for between-group comparisons and fold changes (FC) will be calculated as the ratio of means or medians between groups in accordance to the underlying distribution of the data. Pearson's linear or Spearman rank correlation coefficients will be determined between continuous variables. If necessary, partial correlation coefficients adjusting for known confounding factors will be computed. The false discovery rate (FDR) from the *p*-values of multiple-hypothesis testing will be estimated using the Benjamini and Hochberg procedure (86) and adjusted *p*-values <0.05 will be considered statistically significant.

One of NUTRISHIELD secondary objectives is the development of a personalized nutrition system based on measured biomarkers. The system is to be made available to medical personnel, as well as mothers, to improve infant-mother dyads' health and wellbeing. The results and the associations revealed from the study will be used to train a personalized nutritional algorithm with data obtained from clinical settings. This will be used to build and validate the NUTRISHIELD platform, describing how a theoretical framework designed and fed by the patients' data translates to clinical practice. Toward that goal, a comprehensive data integration system has been developed as the Clinical Trial App (CTA). The CTA is a system intended for personalized medicine/nutrition, implementing the following main features: (i) data acquisition from questionnaires, (ii) data acquisition from laboratory biomarker analysis, (iii) processing data to train machine learning algorithms, and (iv) deploying trained algorithms to produce reports for final users.

Figure 3 shows the data flow and functionalities of the CTA. Questionnaires data are collected by medical staff directly using the phone app. Alternatively, questionnaires that have been collected on paper or on spreadsheets, can be uploaded as well. All samples can be identified with a unique QR code and, at the moment of collection, the phone app can be used to scan the QR code and input sample data. This way the samples can travel between labs with the QR code as only identification. The QR code can be scanned again to associate the results with the

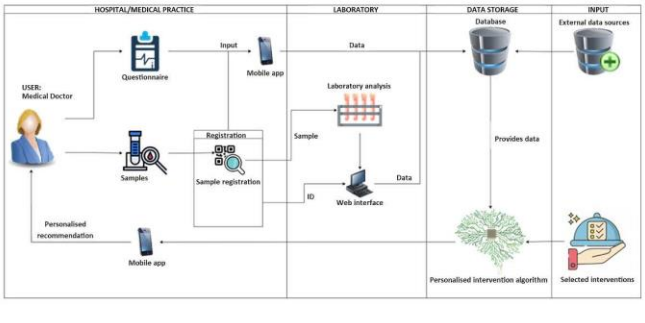


FIGURE 3
Data flow and functionalities of the Clinical Trial App (CTA).

original sample when uploading it through the web portal. The collected data is stored on a database and used to feed the personalized recommendation algorithm. This algorithm produces a report and feeds it back to the phone app.

Using the CTA, all study data is consolidated in a structured database, of which an anonymized version can be later stored on public repositories for re-use in further research.

3. Results and discussion

Figure 4 summarizes the recruitment process, including the eligible PI ($n = 104$), excluded ($n = 49$), lost to follow up ($n = 10$) and the number of PI that were finally included in the study ($n = 45$). In addition, a control group comprising 31 full term OMM infant-mother dyads was recruited, with five participants being lost during follow-up. Altogether, a total of 71 infant-mother dyads were included and completed the study (28 PI-OMM, 17 PI-DHM and 26 TI-OMM), after losing 15 infant participants during follow-up. In total 662 urine, 594 faeces, 134 buccal swab, 33 venous and arterial cord blood, and 234 HM samples, as well as aliquots from 147 DHM batches before and after pasteurization have been collected. An ample array of laboratory methods for biomarker analysis and screening of biological samples and three *ad-hoc* designed sensors were developed. Currently, the analysis of all collected study samples is in progress at the different participating institutions.

The results obtained from HM analysis will be used to study how macronutrients, proteins, vitamins, HMOs, steroids, fatty acids, and microbiota are affected by maternal nutrition, psychosocial status (e.g., social support, family functioning, and perceived stress), other clinical variables (e.g., gestational age, BMI, and type of delivery), and pasteurization, in the case of DHM. Additionally, it is assessed how HM composition affects growth and other health parameters in TI and PI. The results obtained from urine samples analysis will also be used to study how urinary nutrition biomarkers and the metabolic profile are affected by nutrition, psychosocial status, the PRS obtained from genome sequencing and other clinical variables. In addition, HM and faeces microbiome analysis will be used to study its correlation with microbiota biomarker analysis (i.e., SCFAs, BCAAs, and BAs). HM and urine samples will also be used to test the pH, protein, fatty acids, and phosphate and creatinine sensors developed within the study.

This study presents several strengths and limitations. During the study design, special emphasis was put on the comfort of participants. Hence, the collected biological samples, including faeces, urine, HM, buccal swap, and umbilical cord blood samples, were obtained by non-invasive procedures that had been tested and employed in earlier studies carried out at HULAFE. SOPs for sample collection and handling were designed and reviewed by the study team ahead of time. Sample collection kits with detailed instructions and pictures were prepared for sample collection at home. Finally, all procedures were tested in a small pilot study conducted before the initiation of the main study, allowing for minor amendments in the protocols and leading to a homogeneous process conducted all through the main study. During follow-up, some visits were carried out at home avoiding the additional effort for participants to travel to the hospital. These efforts helped to encourage mothers to participate in the study and enhanced adherence during the follow-up period.

One of the most outstanding strengths of this study is the rich array of information obtained from participants during the first six months of life, a pivotal period for health programming. Direct access to participants through trained staff of the research team at

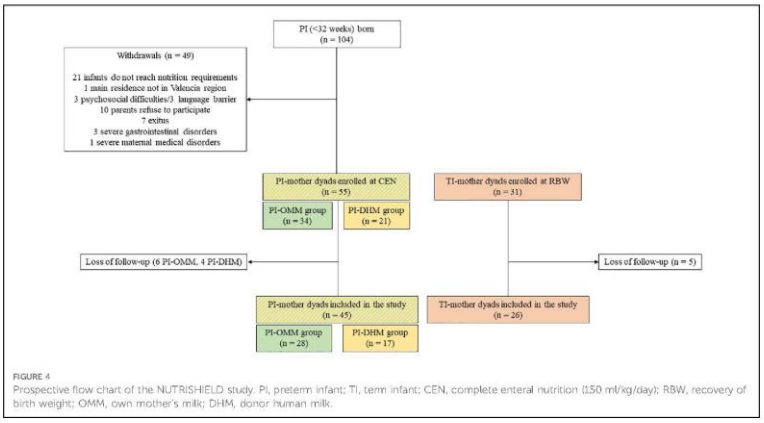


FIGURE 4
Prospective flow chart of the NUTRISHIELD study. PI, preterm infant; TI, term infant; CEN, complete enteral nutrition (150 ml/kg/day); RBW, recovery of birth weight; OMM, own mother's milk; DHM, donor human milk.

the hospital greatly aided to reduce the rate of missing data. The promotion of good relationships with participants could facilitate to further extending the life of the study cohort. Nevertheless, the follow-up period presented challenges. We would like to highlight the loss of some samples due to technical issues when collected at home as well as the temporary loss of some sampling time points, e.g., due to illness of participants or overburdening of lactating mothers. Our preliminary results need to be confirmed in a greater cohort of infants and in different European settings.

This study will provide a better understanding of the impact of maternal nutrition in HM composition, and the interplay of HM composition, microbiota, and newborn physiology, especially in PI's development. More knowledge and a better understanding could allow for a personalized nutrition or even individualized care at an early stage of the neonatal period to improve the outcome of these vulnerable newborns. Additionally, data obtained from this study will be used to train a machine-learning algorithm for providing dietary advice to lactating mothers and will be implemented in a user-friendly platform based on a combination of user-provided information and biomarker analysis.

4. Conclusion

A better understanding of the factors affecting the milk's composition, together with the health implications for infants plays an important role in developing improved strategies of nutraceutical management in infant care. The development of a user-friendly platform based on a combination of user-provided information and biomarker analysis for providing dietary advice would make latest scientific advances directly accessible to a wide range of lactating mothers. This development is anticipated to be of special importance for health, growth, and development of prematurely born infants.

Ethics statement

The studies involving human participants were reviewed and approved by Scientific and Ethics Committee for Biomedical Research (CEIm) of the HUiP La Fe. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

VR-G performed the microbiota activity, nutrition biomarkers, untargeted, macronutrients and fatty acid analysis and wrote the first draft of the manuscript. IT-D performed the oligosaccharides and steroid analysis and wrote the first draft of the manuscript. AM-G, LC-B and AP-L performed the participants recruitment, sample collection and data curation. AR-B, MJV and MG performed the data curation and supervision. FM performed the genome sequencing analysis, funding acquisition and project administration. CM performed the genome sequencing analysis. PR performed the microbiota analysis, funding acquisition and project administration. ASi and LR performed the microbiota analysis. SR performed the vitamin analysis, funding acquisition and project administration. AH and SC performed the vitamin analysis. EB and EM performed the dietary analysis, funding acquisition and project administration. MK, DP and MY performed the dietary analysis. CA and ASc performed the development of the IR-based methods for creatinine, phosphate, proteins and fatty acids. JKa performed the sensor development, funding acquisition and project administration. DM and SG performed the sensor development, funding acquisition, and project administration. MVa performed the laser development, supervision, conceptualization, funding acquisition, and project administration. BL, JKu and MVe performed the supervision, conceptualization, funding acquisition and project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110], the Instituto de Salud Carlos III (Ministry of Science and Innovation, Spain) and co-funded by the European Union (grant numbers CD19/00176, CM20/00143, and CPII21/00003), and the Generalitat Valenciana [project number GV/2021/186].

Acknowledgments

The authors would like to thank all parents and their infants who agreed to participate in this study, as well as, the nurses and medical staff involved, for their effort and commitment. Technical support from the Central Support Service for Experimental Research (SCSIE) at the University of Valencia is acknowledged.

Conflict of interest

PR, ASi and LR are employed by REM Analytics S.A.; author JKa was employed by Quantared Technologies GmbH; and MVa is employed by ALPES Lasers. S.A. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet*. (2016) 387(10017):475–90. doi: 10.1016/S0140-6736(15)01024-7

2. Kramer MS, Kakuma R. Optimal duration of exclusive breastfeeding. *Cochrane Database Syst Rev*. (2012) 2012(8):CD003517. doi: 10.1002/14651858.CD003517.pub2

3. Meek JY, Noble L. Section on breastfeeding. Policy statement: breastfeeding and the use of human milk. *Pediatrics*. (2022) 150(1):e2022057988. doi: 10.1542/peds.2022-057988

4. Boland M. Exclusive breastfeeding should continue to six months. *Paediatr Child Health*. (2005) 10(3):148. doi: 10.1093/pch/10.3.148

5. Toscano M, De Grandi R, Grossi E, Drago L. Role of the human breast milk-associated microbiota on the newborns' immune system: a mini review. *Front Microbiol*. (2017) 8:2100. doi: 10.3389/fmicb.2017.02100

6. Andreas NJ, Kampmann B, Mehring Le-Doare K. Human breast milk: a review on its composition and bioactivity. *Early Hum Dev*. (2015) 91(11):629–35. doi: 10.1016/j.earlhumdev.2015.08.013

7. Agostoni C, Buonocore G, Carnielli VP, De Curtis M, Darmaun D, Decsi T, et al. Enteral nutrient supply for preterm infants: commentary from the European society of paediatric gastroenterology, hepatology and nutrition committee on nutrition. *J Pediatr Gastroenterol Nutr*. (2010) 50(1):85–91. doi: 10.1097/MPG.0b013e3181adaee0

8. Lönnerdal B, Erdmann P, Thakkar SK, Sauser J, Destaillats F. Longitudinal evolution of true protein, amino acids and bioactive proteins in breast milk: a developmental perspective. *J Nutr Biochem*. (2017) 41:1–11. doi: 10.1016/j.jnutbio.2016.06.001

9. Koletzko B, Rodriguez-Palmero M, Demmelmair H, Fidler N, Jensen R, Sauerwald T. Physiological aspects of human milk lipids. *Early Hum Dev*. (2001) 65:S3–18. doi: 10.1016/S0378-3782(01)00204-3

10. Drago L, Toscano M, De Grandi R, Grossi E, Padovani EM, Peroni DG. Microbiota network and mathematic microbe mutualism in colostrum and mature milk collected in two different geographic areas: italy versus Burundi. *ISME J*. (2017) 11(4):875–84. doi: 10.1038/ismej.2016.183

11. García-Muñoz RF, Díez Recinos AL, García-Alix PA, Figueras AJ, Vento TM. Changes in perinatal care and outcomes in newborns at the limit of viability in Spain: the EPI-SEN study. *Neonatology*. (2015) 107(2):120–9. doi: 10.1159/000368881

12. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. *Lancet*. (2016) 388(10063):3027–35. doi: 10.1016/S0140-6736(16)31593-8

13. Vohr BR, Poindexter BB, Dusick AM, McKinley LT, Higgins RD, Langer JC, et al. Persistent beneficial effects of breast milk ingested in the neonatal intensive care unit on outcomes of extremely low birth weight infants at 30 months of age. *Pediatrics*. (2007) 120(4):e953–959. doi: 10.1542/peds.2006-3227

14. Schanler R, Shulman R, Lau C. Feeding strategies for premature infants: beneficial outcomes of feeding fortified human milk versus preterm formula. *Pediatrics*. (1999) 103:1150–7. doi: 10.1542/peds.103.6.1150

15. Sisk PM, Lovelady CA, Dillard RG, Gruber KJ, O'Shea TM. Early human milk feeding is associated with a lower risk of necrotizing enterocolitis in very low birth weight infants. *J Perinatol*. (2007) 27(7):428–33. doi: 10.1038/sj.jp.7211758

16. Meinzen-Derr J, Poindexter B, Wrage L, Morrow AL, Stoll B, Donovan EF. Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. *J Perinatol*. (2009) 29(1):57–62. doi: 10.1038/jp.2008.117

17. Maayan-Metzger A, Avivi S, Schushan-Eisen I, Kuint J. Human milk versus formula feeding among preterm infants: short-term outcomes. *Am J Perinatol*. (2012) 29(2):121–6. doi: 10.1055/s-0031-1295652

18. Vohr BR, Poindexter BB, Dusick AM, McKinley LT, Wright LL, Langer JC, et al. Beneficial effects of breast milk in the neonatal intensive care unit on the developmental outcome of extremely low birth weight infants at 18 months of age. *Pediatrics*. (2006) 118(1):e115–123. doi: 10.1542/peds.2005-2382

19. Isaacs EB, Fischl BR, Quinn BT, Chong WK, Gadian DG, Lucas A. Impact of breast milk on intelligence quotient, brain size, and white matter development. *Pediatr Res*. (2010) 67(4):357–62. doi: 10.1203/PDR.0b013e3181d026da

20. Okamoto T, Shirai M, Kokubo M, Takahashi S, Kajino M, Takase M, et al. Human milk reduces the risk of retinal detachment in extremely low-birthweight infants. *Pediatr Int*. (2007) 49(6):894–7. doi: 10.1111/j.1442-200X.2007.02483.x

21. Belfort MB, Rifas-Shiman SL, Sullivan T, Collins CT, McPhee AJ, Ryan P, et al. Infant growth before and after term: effects on neurodevelopment in preterm infants. *Pediatrics*. (2011) 128(4):e899–906. doi: 10.1542/peds.2011-0282

22. Meier PP, Patel AL, Esquerra-Zwiers A. Donor human milk update: evidence, mechanisms and priorities for research and practice. *J Pediatr*. (2017) 180:15–21. doi: 10.1016/j.jpeds.2016.09.027

23. Calvo J, García Lara NR, Gormaz M, Peña M, Martínez Lorenzo MJ, Ortiz Murillo P, et al. Recomendaciones para la creación y el funcionamiento de los bancos de leche materna en españa. *An Pediatr (Barc)*. (2018) 89(1):65.e1–6. doi: 10.1016/j.anpedi.2018.01.010

24. Schanler RJ, Lau C, Hurst NM, Smith EO. Randomized trial of donor human milk versus preterm formula as substitutes for mothers' own milk in the feeding of extremely premature infants. *Pediatrics*. (2005) 116(2):400–6. doi: 10.1542/peds.2004-1974

25. Quigley M, Embleton ND, McGuire W. Formula versus donor breast milk for feeding preterm or low birth weight infants. *Cochrane Database Syst Rev*. (2019) 2019(7):CD002971. doi: 10.1002/14651858.CD002971.pub5

26. Bitman J, Wood L, Hamosh M, Hamosh P, Mehta NR. Comparison of the lipid composition of breast milk from mothers of term and preterm infants. *Am J Clin Nutr*. (1983) 38(2):300–12. doi: 10.1093/ajcn/38.2.300

27. Velonà T, Abbiati L, Beretta B, Gaiaschi A, Flaùto U, Tagliabue P, et al. Protein profiles in breast milk from mothers delivering term and preterm babies. *Pediatr Res*. (1999) 45(5):658–63. doi: 10.1203/00006450-199905010-00008

28. Zhang Z, Adelman AS, Rai D, Boettcher J, Lönnerdal B. Amino acid profiles in term and preterm human milk through lactation: a systematic review. *Nutrients*. (2013) 5(12):4800–21. doi: 10.3390/nu5124800

29. Austin S, De Castro CA, Sprenger N, Binia A, Affolter M, Garcia-Rodenas CL, et al. Human milk oligosaccharides in the milk of mothers delivering term versus preterm infants. *Nutrients*. (2019) 11(6):1282. doi: 10.3390/nu11061282

30. Underwood MA. Human milk for the premature infant. *Pediatr Clin North Am*. (2013) 60(1):189–207. doi: 10.1016/j.pcl.2012.09.008

31. Ballard O, Morrow AL. Human milk composition: nutrients and bioactive factors. *Pediatr Clin North Am*. (2013) 60(1):49–74. doi: 10.1016/j.pcl.2012.10.002

32. Innis SM. Human milk and formula fatty acids. *J Pediatr*. (1992) 120(4 Pt 2):S56–61. doi: 10.1016/S0022-3476(05)81237-5

33. Butts CA, Hedderley DI, Herath TD, Paturi G, Glyn-Jones S, Wiens F, et al. Human milk composition and dietary intakes of breastfeeding women of different ethnicity from the manawatu-wanganui region of New Zealand. *Nutrients*. (2018) 10(9):1231. doi: 10.3390/nu10091231

34. Brenna JT, Varamini B, Jensen RG, Diersen-Schade DA, Boettcher JA, Arterburn LM. Docosahexaenoic and arachidonic acid concentrations in human breast milk worldwide. *Am J Clin Nutr*. (2007) 85(6):1457–64. doi: 10.1093/ajcn/85.6.1457

35. Xi Q, Liu W, Zeng T, Chen X, Luo T, Deng Z. Effect of different dietary patterns on macronutrient composition in human breast milk: a systematic review and meta-analysis. *Nutrients*. (2023) 15(3):485. doi: 10.3390/nu15030485

36. Keikha M, Bahreynian M, Saleki M, Kelishadi R. Macro- and micronutrients of human milk composition: are they related to maternal diet? A comprehensive systematic review. *Breastfeed Med*. (2017) 12(9):517–27. doi: 10.1089/bfm.2017.0048

37. Bravi F, Wiens F, Decarli A, Dal Pont A, Agostoni C, Ferraroni M. Impact of maternal nutrition on breast-milk composition: a systematic review. *Am J Clin Nutr*. (2016) 104(3):646–62. doi: 10.3945/ajcn.115.120881

38. Donovan S, Dewey K, Novotny R, Stang J, Taveras E, Kleinman R, et al. *Dietary patterns during lactation and human milk composition and quantity: A systematic review*. Alexandria (VA): USDA Nutrition Evidence Systematic Review (2020). Available at: http://www.ncbi.nlm.nih.gov/books/NBK578493/ (Cited March 1, 2023).

39. Bode L. Researchers discover a unique human milk oligosaccharide. *Kinderkrankenschwester*. (2012) 31(5):214. PMID: 22685944

40. Chu DM, Meyer KM, Prince AL, Aagaard KM. Impact of maternal nutrition in pregnancy and lactation on offspring gut microbial composition and function. *Gut Microbes*. (2016) 7(6):459–70. doi: 10.1080/19490976.2016.1241357

41. Meyer KM, Mohammad M, Ma J, Chu D, Haymond M, Aagaard K. 66: maternal diet alters the breast milk microbiome and microbial gene content. *Am J Obstet Gynecol*. (2016) 214(1):S47–8. doi: 10.1016/j.ajog.2015.10.084

42. Parra-Llorca A, Gormaz M, Sánchez-Illana Á, Piñeiro-Ramos JD, Collado MC, Serna E, et al. Does pasteurized donor human milk efficiently protect preterm infants against oxidative stress? *Antioxid Redox Signaling*. (2019) 31(11):791–9. doi: 10.1089/ars.2019.7821

43. Conway JM, Ingwersen LA, Moshfegh AJ. Accuracy of dietary recall using the USDA five-step multiple-pass method in men: an observational validation study. *J Am Diet Assoc*. (2004) 104(4):595–603. doi: 10.1016/j.jada.2004.01.007

44. Panagiotakos DB, Pitsavos C, Stefanadis C. Dietary patterns: a Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. *Nutr Metab Cardiovasc Dis*. (2006) 16(8):559–68. doi: 10.1016/j.numecd.2005.08.006

45. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. *J Pers Assess*. (1988) 52(1):30–41. doi: 10.1207/s15327752jpa5201_2

46. Olsen D, Portner J, Lavee Y. *Family adaptability and cohesion evaluation scales (FACES-II)*. Minneapolis: University of Minnesota (1985). Vol. 32, No. 2, p. 10–22.

47. Lee EH. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res (Korean Soc Nurs Sci)*. (2012) 6(4):121–7. doi: 10.1016/j.anr.2012.08.004

48. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. *Manual for the state-trait anxiety inventory (form Y1–Y2)*. Palo Alto, CA: Consulting Psychologists Press (1983). Vol. IV.

49. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. *Br J Psychiatry*. (1987) 150:782–6. doi: 10.1192/bjp.150.6.782

50. Beck AT, Rial WY, Rickels K. Short form of depression inventory: cross-validation. *Psychol Rep*. (1974) 34(3):1184–6. doi: 10.1177/003329417403403s01

51. Davidson J, Smith R. Traumatic experiences in psychiatric outpatients. *J Trauma Stress*. (1990) 3(3):459–75. doi: 10.1002/jts.2490030314

52. Bobes J, Calcedo-Barba A, García M, François M, Rico-Villademoros F, González MP, et al. Evaluation of the psychometric properties of the spanish version of 5 questionnaires for the evaluation of post-traumatic stress syndrome. *Actas Esp Psiquiatr*. (2000) 28(4):207–18. PMID: 11116791

53. Brockington IF, Fraser C, Wilson D. The postpartum bonding questionnaire: a validation. *Arch Womens Ment Health*. (2006) 9(5):233–42. doi: 10.1007/s00737-006-0132-1

54. Rivas GR, Arruabarrena I, Paúl JD. Parenting stress index-short form: psychometric properties of the spanish version in mothers of children aged 0 to 8 years. *Psychosoc Interv*. (2020) 30(1):27–34. doi: 10.5093/pi2020a14

55. Squires J, Twombly E, Bricker D, Potter L. *Technical report. ASQ-3™ user's guide*. Baltimore: Brookes Publishing (2009).

56. Putnam SP, Helbig AL, Gartstein MA, Rothbart MK, Leerkes E. Development and assessment of short and very short forms of the infant behavior questionnaire-revised. *J Pers Assess*. (2014) 96(4):445–58. doi: 10.1080/00223891.2013.841171

57. Roid GH, Sampers JL. *Merrill-Palmer-Revised scales of development (MPR) [Database record]*. Illinois: APA PsycTests (2004).

58. Jácome Á, Jiménez R. Validation of the Iowa infant feeding attitude scale. *Padiatria (Bucur)*. (2014) 47(4):77–82. doi: 10.1016/S0120-4912(15)30143-9

59. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. (2007) 81(3):559–75. doi: 10.1086/519795

60. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. (2018) 562 (7726):203–9. doi: 10.1038/s41586-018-0579-z

61. Kouhounde S, Adéoti K, Mounir M, Giusti A, Refinetti P, Otu A, et al. Applications of probiotic-based multi-components to human, animal and ecosystem health: concepts, methodologies, and action mechanisms. *Microorganisms*. (2022) 10 (9):1700. doi: 10.3390/microorganisms10091700

62. Ekström PO, Warren DJ, Thilly WG. Separation principles of cycling temperature capillary electrophoresis. *Electrophoresis*. (2012) 33(7):1162–8. doi: 10.1002/elps.201100550

63. Ramos-Garcia V, Ten-Doménech I, Moreno-Giménez A, Campos-Berga L, Parra-Llorca A, Solaz-García Á, et al. GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers. *Clin Chim Acta*. (2022) 532:172–80. doi: 10.1016/j.cca.2022.05.005

64. Zheng X, Qiu Y, Zhong W, Baxter S, Su M, Li Q, et al. A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids. *Metabolomics*. (2013) 9(4):818–27. doi: 10.1007/s11306-013-0500-6

65. Sarafian MH, Lewis MR, Pechlivanis A, Ralphs S, McPhail MJW, Patel VC, et al. Bile acid profiling and quantification in biofluids using ultra-performance liquid chromatography tandem mass spectrometry. *Anal Chem*. (2015) 87(19):9662–70. doi: 10.1021/acs.analchem.5b01556

66. González-Domínguez R, Urpi-Sarda M, Jáuregui O, Needs PW, Kroon PA, Andrés-Lacueva C. Quantitative dietary fingerprinting (QDF)—a novel tool for comprehensive dietary assessment based on urinary nutrimetabolomics. *J Agric Food Chem*. (2019) 68(7):1851–61. doi: 10.1021/acs.jafc.8b07023

67. Ramos-Garcia V, Moreno-Giménez A, Campos-Berga L, Parra-Llorca A, Gormaz M, Vento M, et al. Joint microbiota activity and dietary assessment through urinary biomarkers by LC-MS/MS. (Under revision).

68. Ten-Doménech I, Martínez-Sena T, Moreno-Torres M, Sanjuan-Herráez JD, Castell JV, Parra-Llorca A, et al. Comparing targeted vs. Untargeted MS2 data-dependent acquisition for peak annotation in LC-MS metabolomics. *Metabolites*. (2020) 10(4):126. doi: 10.3390/metabo10040126

69. Miris. "HMA User Manual". Available at: http://www.mirissolutions.com/support/user-manuals (Cited June 27, 2022).

70. Cruz-Hernandez C, Goeuriot S, Giuffrida F, Thakkar SK, Destaillats F. Direct quantification of fatty acids in human milk by gas chromatography. *J Chromatogr, A*. (2013) 1284:174–9. doi: 10.1016/j.chroma.2013.01.094

71. Ten-Doménech I, Ramos-García V, Moreno-Torres M, Parra-Llorca A, Gormaz M, Vento M, et al. The effect of holder pasteurization on the lipid and metabolite composition of human milk. *Food Chem*. (2022) 384:132581. doi: 10.1016/j.foodchem.2022.132581

72. Lu J, Frank EL. Rapid HPLC measurement of thiamine and its phosphate esters in whole blood. *Clin Chem*. (2008) 54(5):901–6. doi: 10.1373/clinchem.2007.099077

73. Yang J, Cleland GE, Organtini KL. Determination of vitamin D and previtamin D in food products. *Waters Appl Note*. (2017).

74. Oberson JM, Bénet S, Redeuil K, Campos-Giménez E. Quantitative analysis of vitamin D and its main metabolites in human milk by supercritical fluid chromatography coupled to tandem mass spectrometry. *Anal Bioanal Chem*. (2020) 412(2):365–75. doi: 10.1007/s00216-019-02248-5

75. Levêques A, Oberson JM, Tissot EA, Redeuil K, Thakkar SK, Campos-Giménez E. Quantification of vitamins A, E, and K and carotenoids in submilliliter volumes of human milk. *J AOAC Int*. (2019) 102(4):1059–68. doi: 10.5740/jaoacint.19-0016

76. Manoury E, Jourdon K, Boyaval P, Fourcassié P. Quantitative measurement of vitamin K2 (menaquinones) in various fermented dairy products using a reliable high-performance liquid chromatography method. *J Dairy Sci*. (2013) 96 (3):1335–46. doi: 10.3168/jds.2012-5494

77. Xue Y, Campos-Giménez E, Redeuil KM, Lévêques A, Actis-Goretta L, Vinyes-Pares G, et al. Concentrations of carotenoids and tocopherols in breast milk from urban Chinese mothers and their associations with maternal characteristics: a cross-sectional study. *Nutrients*. (2017) 9(11):1229. doi: 10.3390/nu9111229

78. Remoroza CA, Mak TD, De Leoz MLA, Mirokhin YA, Stein SE. Creating a mass spectral reference library for oligosaccharides in human milk. *Anal Chem*. (2018) 90 (15):8977–88. doi: 10.1021/acs.analchem.8b01176

79. Navarro-Esteve V, Cascant-Vilaplana MM, Zöchner A, Ramos-Garcia V, Roca M, Parra-Llorca A, et al. A Hybrid Liquid Chromatography—Mass Spectrometry method for comprehensive Human Milk Oligosaccharide Screening. (Under revision).

80. Son HH, Yun WS, Cho S-H. Development and validation of an LC-MS/MS method for profiling 39 urinary steroids (estrogens, androgens, corticoids, and progestins). *Biomed Chromatogr*. (2020) 34:e4723–39. doi: 10.1002/bmc.4723

81. Gomez-Gomez A, Miranda J, Feixas G, Arranz Betegon A, Crispi F, Gratacós E, et al. Determination of the steroid profile in alternative matrices by liquid chromatography tandem mass spectrometry. *J Steroid Biochem Mol Biol*. (2020) 197:105520. doi: 10.1016/j.jsbmb.2019.105520

82. Akhgar CK, Ramer G, Zbik M, Trajnerowicz A, Pawluczyk J, Schwaighofer A, et al. The next generation of IR spectroscopy: EC-QCL-based mid-IR transmission spectroscopy of proteins with balanced detection. *Anal Chem*. (2020) 92(14):9901–7. doi: 10.1021/acs.analchem.0c01406

83. Schwaighofer A, Akhgar CK, Lendl B. Broadband laser-based mid-IR spectroscopy for analysis of proteins and monitoring of enzyme activity. *Spectrochim Acta, Part A*. (2021) 253:119563. doi: 10.1016/j.saa.2021.119563

84. Akhgar CK, Nürnberger V, Nadvornik M, Ramos-Garcia V, Ten-Doménech I, Kuligowski J, et al. Fatty acid determination in human milk using attenuated total reflection infrared spectroscopy and solvent-free lipid separation. *Appl Spectrosc*. (2022) 76(6):730–6. doi: 10.1177/00037028211065502

85. Akhgar CK, Ramos-García V, Nürnberger V, Moreno-Giménez A, Kuligowski J, Rosenberg E, et al. Solvent-Free lipid separation and attenuated total reflectance infrared spectroscopy for fast and green fatty acid profiling of human milk. *Foods*. (2022) 11(23):3906. doi: 10.3390/foods11233906

86. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*(1995) 57(1):289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

nutrients

MDPI

Article

Joint Microbiota Activity and Dietary Assessment through Urinary Biomarkers by LC-MS/MS

Victoria Ramos-Garcia [1], Isabel Ten-Doménech [1], Alba Moreno-Giménez [1], Laura Campos-Berga [1], Anna Parra-Llorca [1], María Gormaz [2], Máximo Vento [1,2], Melina Karipidou [3], Dimitrios Poulimeneas [3], Eirini Mamalaki [3], Eirini Bathrellou [3] and Julia Kuligowski [1,*]

[1] Neonatal Research Unit, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain; victoria_ramos@iislafe.es (V.R.-G.); isabel_ten@iislafe.es (I.T.-D.); albamorenoo@hotmail.com (A.M.-G.); lauracamposberga@gmail.com (L.C.-B.); annaparrallorca@gmail.com (A.P.-L.); maximo.vento@uv.es (M.V.)

[2] Division of Neonatology, University & Polytechnic Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain; gormaz_mar@gva.es

[3] Department of Nutrition and Dietetics, Harokopio University of Athens, El. Venizelou 70, 17676 Kallithea, Greece; mkarip@hua.gr (M.K.); dpoul@hua.gr (D.P.); eir.mamalaki@gmail.com (E.M.); ebathrellou@hua.gr (E.B.)

* Correspondence: julia.kuligowski@uv.es; Tel.: +34-961246661

Abstract: Accurate dietary assessment in nutritional research is a huge challenge, but essential. Due to the subjective nature of self-reporting methods, the development of analytical methods for food intake and microbiota biomarkers determination is needed. This work presents an ultra-high performance liquid chromatography coupled to tandem mass spectrometry (UHPLC-MS/MS) method for the quantification and semi quantification of 20 and 201 food intake biomarkers (BFIs), respectively, as well as 7 microbiota biomarkers applied to 208 urine samples from lactating mothers (M) ($N = 59$). Dietary intake was assessed through a 24 h dietary recall (R24h). BFI analysis identified three distinct clusters among samples: samples from clusters 1 and 3 presented higher concentrations of most biomarkers than those from cluster 2, with dairy products and milk biomarkers being more concentrated in cluster 1, and seeds, garlic and onion in cluster 3. Significant correlations were observed between three BFIs (fruits, meat, and fish) and R24h data ($r > 0.2$, p-values < 0.01, Spearman correlation). Microbiota activity biomarkers were simultaneously evaluated and the subgroup patterns detected were compared to clusters from dietary assessment. These results evidence the feasibility, usefulness, and complementary nature of the determination of BFIs, R24h, and microbiota activity biomarkers in observational nutrition cohort studies.

Keywords: nutrition biomarkers; microbiota biomarkers; food-intake; dietary assessment; lactating mothers; urine

Citation: Ramos-García, V.; Ten-Doménech, I.; Moreno-Giménez, A.; Campos-Berga, L.; Parra-Llorca, A.; Gormaz, M.; Vento, M.; Karipidou, M.; Poulimeneas, D.; Mamalaki, E.; et al. Joint Microbiota Activity and Dietary Assessment through Urinary Biomarkers by LC-MS/MS. *Nutrients* **2023**, *15*, 1894. https://doi.org/10.3390/nu15081894

Academic Editor: Emile Levy

Received: 13 March 2023
Revised: 3 April 2023
Accepted: 12 April 2023
Published: 14 April 2023

1. Introduction

The measurement of diet exposure is crucial for determining associations between food intake and health status. Additionally, optimizing nutritional advice to specific population groups (e.g., chronic diseases patients, lactating mothers, etc.) has recently become a major challenge [1]. The incidence of preterm deliveries (<37 weeks of gestation) and the survival rate of preterm infants have been steadily increasing over recent decades [2] and early infant nutrition is key for improving clinical outcomes. As human milk (HM) is recommended as the gold standard for infant nutrition [3] and the impact of maternal nutrition on the composition of HM has been evidenced [4–6], it is important to develop tools that can help to evaluate maternal dietary patterns with the aim of providing dietary advice that could enhance preterm infants' growth rates.

Dietary intake data is most commonly collected using tools based on self-reporting, such as food frequency questionnaires (FFQ) for the assessment of regular consumption and

food diaries or 24 h recalls (R24h) for the assessment of short-term consumption. This food intake data is then translated into quantitative information regarding specific nutrients or food groups using food composition databases. However, such methods are prone to errors due to their subjective nature [7]. In particular, foods perceived as being unhealthy (e.g., processed foods and high fats) are typically under-reported [8], while foods perceived as being healthy (e.g., fruits and vegetables) are often over-reported. Therefore, the generation of robust data on regular dietary intake is essential to improve the accuracy of dietary assessment. The measurement of metabolomic food intake biomarkers (BFIs) in biological samples [9,10] has emerged in recent years as a method of supporting a more accurate assessment of nutritional intake [11,12]. Metabolomic fingerprinting has opened new opportunities for the discovery of specific BFIs in body fluids [13]. For example, flavonoids, methyl histidine, isoflavones, trimethylamine N-oxide (TMAO), and arylglycines have been described recently as biomarkers of fruits and vegetables, meat, seeds, fish, and dairy products, respectively [10,14].

After ingestion, these compounds are metabolized by phase I and II enzymes, yielding glucuronidated, sulfated, and methylated metabolites [11]. Although recent studies illustrated that an individual's dietary habits can predict the levels of specific metabolites present in plasma [15] and research has identified hundreds of mediation linkages that provide insight into diet—microbiome interactions [16], urine is the most practical and feasible biofluid used for BFI identification since many metabolic byproducts are excreted through it [17]. Compared to the collection of other biofluids, such as blood and plasma, the collection of urine is easier, cheaper, allows quicker collection of large volumes, and is less burdensome and invasive for the participants [18].

Due to the different nature of BFIs, high sensitivity and wide coverage analytical methods are needed to assess the metabolic fingerprint of food intake. Liquid (LC) or gas (GC) chromatography coupled with mass spectrometry (MS) and H^1 nuclear magnetic resonance (NMR) spectroscopy are the most commonly used techniques for the analysis of BFIs [19] and targeted assays that enable the quantification of metabolites are the preferred strategy [20]. The vast majority of previously published studies about BFI fingerprinting required complex and time-consuming sample preparation, such as enzymatic hydrolysis [21,22] or solid phase extraction [23,24], resulting in incomplete hydrolysis and low recoveries, respectively. Moreover, the lack of commercial standards of metabolic byproducts entangle the targeted approach [25] and therefore an attractive strategy to overcome this drawback could be the semi-quantification of these urinary metabolic byproducts.

In addition, diet is one of the key factors involved in shaping the gut microbiota, affecting the microbial diversity, as well as the abundance of specific microbes [26–28]. Metabolic end products of dietary micro and macronutrients, such as bile acids (e.g., cholic and lithocholic acids), short chain fatty acids (e.g., acetic and butyric acid), indole and polyphenyl derivatives (e.g., 3-indolepropionic acid (3-IPA) and hippuric acid), and phenolic acids (e.g., gallic acid and ferulic acid) are used to assess the activity of gut microbiota [29–33]. A joint assessment of food intake and microbiota activity biomarkers would be of interest.

The aim of this work was to develop an ultra-high performance LC coupled with the tandem MS (UHPLC-MS/MS) method for the quantification and semi-quantification of 20 and 201 BFIs, respectively, of different food groups (e.g., fruits and vegetables, meat, etc.) and the simultaneous determination of 7 microbiota biomarkers. The applicability of the method was evaluated by the analysis of 208 urine samples from lactating mothers (N = 59). The performance of the determined BFIs as a complementary tool to self-reported R24h dietary assessment was evaluated. Furthermore, nutrition patterns were compared to microbiota activity biomarker diversity clusters.

2. Materials and Methods

2.1. Study Design, Population, and Sample Collection

Samples were collected under the framework of the NUTRISHIELD project (https://nutrishield-project.eu/ (accessed on 13 April 2023)), in a prospective, observational, cohort study (NCT05646940) including infant–mother dyads and performed at the Division of Neonatology of the University and Polytechnic Hospital La Fe (Valencia, Spain). Urine samples from lactating mothers were collected at six time points, covering the period from birth to six months of age. R24h were recorded at all timepoints except at birth, while FFQs were recorded at month one. During data analysis, urine samples and R24h from each time point were treated individually, as maternal food intake varied between time points. Demographic and clinical characteristics of participants are shown in Table 1. The mothers' first morning urine was collected in sterile polypropylene PP containers. In total, 208 urine samples from 59 lactating mothers were collected. For additional information on the NUTRISHIELD study, the reader is referred to Ramos-Garcia et al. [34].

Table 1. Demographic and clinical characteristics of participants at time point 'Recovery of Birth Weight' or 'Complete Enteral Nutrition' for mothers of term and preterm infants, respectively.

Parameters	Mothers (*N* = 59)
Age (years), mean (SD)	36 (5)
Weight (kg), mean (SD)	62 (14)
BMI (kg/m^2), mean (SD)	24 (5)
C-Section delivery, N (%)	28 (48)
Antibiotic therapy, N (%)	3 (5)
Dietary supplements, N (%)	52 (88)
MedDietScore *, mean (SD)	30 (5)

Note: * derived from FFQ recorded at month one; SD = standard deviation; BMI = body mass index; C-Section = caesarean section; MedDietScore = adherence to the Mediterranean diet.

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry #2019-289-1 and all methods were performed in accordance with the relevant guidelines and regulations. Written informed consents were obtained from lactating mothers prior to sample collection and analysis of clinical and demographic information.

2.2. Dietary Assessment

Regarding the dietary assessment methods, a R24h and a validated FFQ [35] were performed. For the R24h collection, trained researchers asked for all foods and beverages participants consumed the previous day, using the multiple-pass method [36]. Recall data were analyzed in terms of nutrients using the dietary analysis software Nutritionist Pro™ (2007, Axxya Systems, Redmond, WA, USA; https://nutritionistpro.com/). Additionally, dietary intake was grouped into food groups, namely fruits, vegetables, meat, fish, egg, bread/starch, seeds, milk, dairy products, fat, other carbohydrates, and soft drinks.

Regarding the FFQ, it comprises 142 questions on the consumption of foods that are commonly eaten by the Spanish population throughout the year, including dairy products, cereals, fruits, vegetables, meat, fish, legumes, added fats, alcoholic beverages, stimulants, and sweets. Using a 9-grade scale ("never or less than 1 time/month", "1–3 times/month", "1 time/week", "3–4 times/week", "5–6 times/week", "1 time/day", "2–3 times/day", "4–5 times/day", "≥6 times/day") participants were required to indicate the absolute frequency of consuming a certain amount of food, expressed in g, milliliters or in other common measures, such as slice, tablespoon or cup, depending on the food. The previous month was set as the timeframe.

Based on the FFQ-responses, adherence to the Mediterranean diet was evaluated by using the MedDietScore, a composite score calculated for each participant [37]. For food groups presumed to be part of the Mediterranean pattern (i.e., those with a recommended

intake of 4 servings per week or more, such as non-refined cereals, fruits, vegetables, legumes, olive oil, fish, and potatoes), higher scores are assigned when the consumption is according to the rationale of the Mediterranean pattern, while lower scores are assigned when participants report no, rare, or moderate consumption. For the consumption of foods presumed to be eaten less frequently within the Mediterranean diet (i.e., consumption of meat and meat products, poultry, and full fat dairy products), scores are assigned on a reverse scale. As the sample of the study is lactating mothers, the original score was modified by removing the component regarding alcohol consumption. Thus, the range of this modified MedDietScore is between 0 and 50, with higher values of the score indicating greater adherence to the Mediterranean diet.

2.3. Standards and Reagents

HPLC grade acetonitrile (ACN) (≥99.9%), ammonium formate (≥99.0%), formic acid (≥95%), phenylpropionylglycine (PPG) (≥99%), 3-indolepropionic acid (3-IPA) (≥99%), L-kynurenine (≥98%), 3-indoleacetic acid (3-IAA) (≥98%), hippuric acid (≥98%), ferulic acid sulphate (≥99%), proline betaine (≥99%), hesperetin (≥95%), phloretin (≥99%), quercetin (≥95%), kaempferol (≥97%), O-desmethylangolensin (O-DMA) (≥97%), daidzein (≥98%), equol (≥99%), glycitein (≥97%), genistein (≥98%), trimethylamine N-oxide (TMAO) (≥95%), isovalerylglycine (≥97%), isobutyrylglycine (≥95%), galactitol (≥99%), gallic acid (≥98%), and the "Amino Acid Standards, physiological" solution containing the amino acids L-tyrosine, 1-methylhistidine, 3-methylhistidine, anserine, citrulline, and taurine, as well as the isotopically labelled internal standards (IS) caffeine-D_9 (≥99%), phenylalanine-D_5 (≥99%), and taxifolin (≥95%) were purchased from Sigma-Aldrich Química SL (Madrid, Spain). Betaine-D_{11} (≥99%) was purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA). Standard solutions were prepared in ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany).

2.4. UHPLC-MS/MS Determination of Nutrition and Microbiota Biomarkers

The UHPLC-MS/MS method for the quantification of 20 nutrition biomarkers (i.e., proline betaine, hesperetin, phloretin, quercetin, kaempferol, O-DMA, daidzein, equol, glycitein, genistein, 1-methylhistidine, 3-methylhistidine, anserine, TMAO, isovalerylglycine, isobutyrylglycine, galactitol, gallic acid, citrulline, and taurine) and 7 microbiota biomarkers (i.e., PPG, 3-IPA, L-kynurenine, 3-IAA, L-tyrosine, hippuric acid, and ferulic acid sulphate) and semi-quantification of 201 nutrition biomarkers of fruits and vegetables ($N = 105$), meat ($N = 8$), fish ($N = 3$), seeds ($N = 17$), olive oil ($N = 4$), coffee ($N = 10$), curcuma ($N = 2$), garlic and onion ($N = 1$), grains ($N = 9$), soft drinks ($N = 3$), alcoholic beverages ($N = 13$), and other groups (i.e., potato, cocoa, mushrooms, legumes, and nuts; $N = 26$) (see Supplementary Table S1) was developed based on previous results [23]. A total of 110 µL of 1:20 (*v*/*v*) diluted urine samples in deionized water were added to a 96-well plate and mixed with 10 µL of IS (i.e., caffeine-D_9, phenylalanine-D_5, betaine-D_{11}, and taxifolin at 15 µM each).

UHPLC-MS/MS analysis was conducted using a Sciex QTRAP 6500+ system (Sciex, Framingham, MA, USA) operating in positive and negative ionization modes (ESI+/−). Separations were performed using a Luna Omega Polar C18 column (100 mm × 2.1 mm, 1.6 µm) equipped with a fully porous polar C18 security guard cartridge (Phenomenex, Torrance, CA, USA). Conditions were as follows: column temperature, 40 °C; autosampler temperature, 10 °C; injection volume, 10 µL. Additionally, 0.1% formic acid and 10 mM ammonium formate in water and pure ACN were used as aqueous (A) and organic (B) mobile phases (MP), respectively. The gradient program was as follows: 0–8 min, 5–20% B; 8–10 min, 20–100% B; 10–12 min, 100% B; 12–12.1 min, 100–5% B; and 12.1–14 min, 5% B. MS detection was performed using the multiple reaction monitoring (MRM) mode (see Supplementary Table S2). The mass spectrometer operated using the following parameters: ion spray voltage, (±)4500 V; source temperature, 600 °C; curtain gas, 35 psi; ion source gases 1 and 2, 60 psi each; and entrance potential, (±)10 V. When available, the transitions

were optimized by infusing 5 μM individual solutions of commercial standards dissolved in MP into the mass spectrometer. For the semi-quantification of nutrition biomarkers identified in urine samples for which authentic standards were not available, their corresponding transitions, as reported in the literature [23], were included (see Supplementary Table S1) and the semi-quantification was carried out using the relative peak areas. Quantification was carried out using the relative peak areas with an external calibration line using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. For monitoring the instrument's performance, a quality control (QC) sample (5 μL of each sample pooled) was analyzed every 20 samples in the randomized analytical batch. The batch acceptance criterion was QC relative standard deviation (RSD) < 25%. In addition, calibration blanks (i.e., H_2O) and a process blank (i.e., processing of H_2O as described for urine samples) were injected at the beginning of the sequence for system suitability testing.

Urinary biomarker concentrations were normalized to creatinine concentrations quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA). Samples were diluted with deionized water prior to measurements employing a 1:20 (*v*/*v*) dilution.

2.5. Method Validation

Method validation was based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [38], including the following bioanalytical parameters: accuracy, precision, linearity range, carryover, selectivity, specificity, and stability.

Replicates ($N = 3$) of standards at three concentration levels and replicates ($N = 3$) of spiked samples at three concentration levels (low, medium, and high) on three validation days were analyzed to assess accuracy. The RSD of replicate standards/samples within one validation batch (intra-day) and between validation batches (inter-day) was used to estimate precision. The calibration curves included a zero calibrator (i.e., blank with IS) and, at least six standards covering the selected concentration ranges ensuring linearity. Carryover between samples was assessed by the analysis of a calibration blank after the injection of the standards. The analysis of calibration and process blank samples from multiple ($N = 6$) sources were used to demonstrate selectivity and specificity. Analytes' freeze-thaw and long-term stabilities were tested by comparing concentrations observed in a freshly prepared sample to sample extracts after three freeze-thaw cycles and to a sample stored for six months (−80 °C).

2.6. Data Availability and Statistical Analysis

UHPLC-MS/MS data were acquired and processed using SCIEX OS Software (Sciex, Framingham, MA, USA). Data analysis was carried out in MATLAB R2019b (MathWorks, Natick, MA, USA) and using the PLS Toolbox 8.9 (Eigenvector Research Inc., Manson, WA, USA). Biomarker levels normalized to creatinine in urine samples determined in this work are available in Supplementary Table S3. Continuous variables were expressed as mean ± standard deviation or medians with interquartile range, depending on underlying data distribution. Student's *t*-test ($\alpha = 0.05$) or Wilcoxon rank-sum test ($\alpha = 0.05$) were used for inter-group comparison. Pearson's and Spearman's correlation coefficients were used for assessing paired associations among metabolite concentrations, and among metabolite concentrations and R24h food groups, respectively. Hierarchical clustering analysis was carried out using autoscaled data.

Diversity of microbiota biomarkers within a sample was assessed with Simpson's Index of Diversity (1−D), where *D* was defined as:

$$D = \frac{\sum n(n-1)}{N(N-1)} \quad (1)$$

n being the concentration of a particular microbiota activity biomarker in the sample; and N the total concentration of microbiota activity biomarkers. Thus, the higher the value for this index (1−D), the higher the microbiota activity biomarker diversity.

3. Results and Discussion

3.1. UHPLC-MS/MS Method Validation

The recommended guidelines were used to perform the validation of the analytical method [38]. Linearity of response was assessed covering up to four orders of magnitude with limits of detection (LODs) and limits of quantification (LOQs) in the 0.03–8 µM range (see Supplementary Table S2) and carryover did not exceed 20% of LOQ. Regarding accuracy and precision, intra- and inter-day recoveries in standard solutions spiked at the different concentration levels were between 82 and 119%, with an average precision of 7% (min-max precision of 1–20%) (see Supplementary Table S4). Similarly, in spiked urine samples, intra-day and inter-day recoveries and precisions ranged between 80 and 120% and between 1 and 20% (average precision of 10%), respectively, with the exception of galactitol and genistein spiked at the low concentration level that presented an inter-day recovery of 123% and an inter-day accuracy %RSD of 21. This evidences an adequate method performance for all quantified metabolites. Additionally, all compounds were stable after freezing stock solutions for one year at −80 °C, and after three freeze-thaw cycles (−80 °C) (t-test, p-values > 0.05). Process blanks using sterile PP containers were analyzed and compared to calibration blanks to assess selectivity and specificity, as well as to test the compatibility of the sample collection procedure with the analysis of these metabolites, and no contaminations were detected.

3.2. R24h Results

As shown in Figure 1, results from R24h evidence that mothers' intake of fruits, vegetables, meat, milk, egg, fish, dairy, and other carbohydrates ranged between 1 and 5 portions/day, while bread/starch and fat were the food groups of higher intakes, with 5 to 10 portions/day. However, seeds, and soft drinks were rarely present in their diets. The majority of the mothers participating in the study reported lower intake of fruits and vegetables, meat and fish, bread/starch, and milk, and dairy products than the recommended intake according to the World Health Organization diet guidelines for lactating mothers (5, 2, 8.5, and 3 portions/day, respectively) [39]. Additionally, results from FFQ indicate that 51 of the 59 mothers (86%) included in the study presented adherence to the Mediterranean diet (i.e., MedDietScore between 25 and 35), while 8 women showed lower adherence (7 and 1 participants < 35 and <25, respectively).

3.3. Biomarker Profiles of Lactating Mothers

The concentrations of the nutrition and microbiota biomarkers determined in urine samples, after normalization to creatinine using the validated UHPLC-MS/MS method, are shown in Table 2.

Table 2. Microbiota and nutrition biomarker concentrations in urine samples from lactating mothers.

Category	Metabolite	Range (µmol/g Creat)	Median (µmol/g Creat)	IQR (25–75)	Detection Frecuency (%)
Microbiota	Phenylpropionylglycine	0.007–58	0.03	0.4	72
	3-IPA	0.012–8	2	3	3
	L-Kynurenine	0.009–21	0.5	2	80
	3-IAA	0.012–176	1.3	7	99
	L-Tyrosine	0.014–2829	16	43	98
	Hippuric Acid	11–8496	255	598	100
	Gallic Acid	0.013–10	0.4	0.4	62
	Ferullic Acid Sulphate	0.02–126	0.4	3	74
Fruits	Proline betaine	1.4–2568	88	315	100
	Hesperetin	0.02–11	0.05	0.13	64
	Phloretin	0.004–2	0.02	0.10	29

Table 2. *Cont.*

Category	Metabolite	Range (µmol/g Creat)	Median (µmol/g Creat)	IQR (25–75)	Detection Frecuency (%)
Vegetables	Quercetin	0.03–76	0.4	0.2	48
	Kaempferol	0.02–2	0.2	0.6	93
Seeds	O-DMA	0.007–5	0.02	0.3	61
	Daidzein	0.005–3	0.4	0.9	96
	Equol	0.012–10	0.3	0.05	34
	Glycitein	0.006–1	0.05	0.2	25
	Genistein	0.4–2	0.03	0.02	2
Meat	1-Methylhistidine	4–1784	48	95	100
	3-Methylhistidine	0.2–4555	31	87	100
	Anserine	0.6–304	2	10	88
Fish	TMAO	3–3368	142	473	100
Dairy products	Isovalerylglycine	0.003–91	3	13	87
	Isobutyrylglycine	0.7–125	3	9	93
Milk	Galactitol	1.2–2424	22	67	92
Soft drinks	Citrulline	0.09–60	7	11	95
	Taurine	0.2–2341	55	207	100

Note: 3-IPA = 3-indolepropionic acid; 3-IAA = indole-3-acetic acid; O-DMA = O-desmethylangolensin; TMAO = trimethylamine N-oxide; IQR = interquartile range.

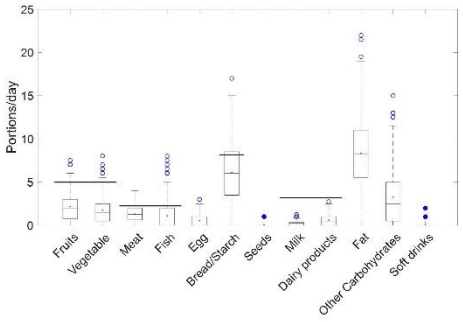


Figure 1. R24h food groups intake in lactating mothers. Note: Horizontal black line represents recommended intake according to the World Health Organization diet guidelines for lactating mothers. (Red dot = median; blue circle (open) = standard outlier (1.5 - 3.0 × IQR); blue circle (closed) = extreme outlier (≥ 3.0 × IQR)).

Results show that, overall, median concentrations of microbiota biomarkers were relatively low, with the exception of hippuric acid and L-tyrosine which were higher than the rest. As observed, a high variability (two to four orders of magnitude) was encountered for these two microbiota biomarkers. It is noteworthy that 3-IPA, an end product of tryptophan metabolism, was only detected in 3% of the samples, whereas the other microbiota biomarkers were detected in over 62% of the samples. In respect of the targeted analysis of nutrition biomarkers, the highest concentrations were found for TMAO, proline betaine, taurine, and 1- and 3-methylhistidine, with phloretin, O-DMA, genistein, hesperetin, and glycitein being the least abundant BFIs. As with the microbiota biomarkers, broad concentration ranges for the more the prevalent metabolites were observed, while lower detection frequencies were found for certain less-abundant biomarkers (e.g., genistein, phloretin, and glycitein) were found.

Figure 2 depicts the significant paired correlations among BFIs (p-value < 0.05). Some metabolites that are biomarkers of the same food group are significantly correlated among each other, such as 1- and 3-methylhistidine (meat); isovalerylglycine and isobutyrylglycine (dairy products); quercetin, phloretin and kaempferol (fruits and vegetables); and genistein, equol and glycitein (seeds). However, additional correlations between biomarkers from different food groups were also significant, such as meat BFI anserine with fruits and vegetables BFIs (i.e., quercetin, phloretin and kaempferol), soft drink BFI taurine with phloretin, and some meat BFIs (i.e., 1- and 3-methylhistidine) with quercetin. This could be due to the paired intake of both food groups or chance correlations.

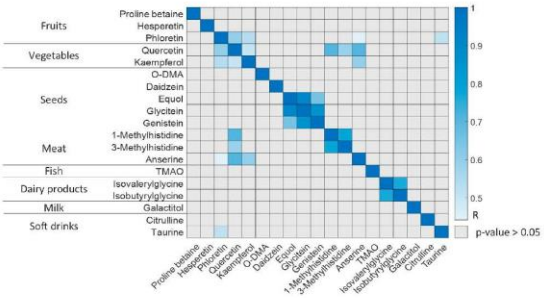


Figure 2. Significant paired correlations among BFIs (Pearson's correlation, p-value < 0.05).

Regarding the semi-quantitative analysis, a higher number of detected metabolites belonged to fruits and vegetables, followed by other groups (i.e., potato, cocoa, mushrooms, legumes, and nuts), coffee, meat, and seeds, as shown in Figure 3.

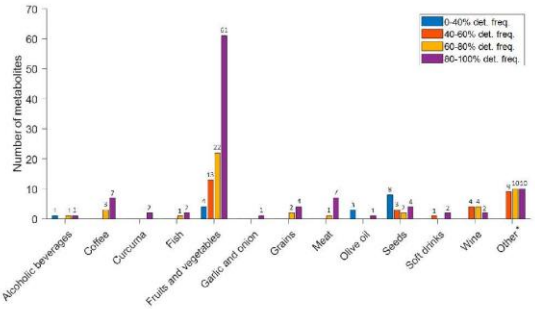


Figure 3. Number of metabolites and detection frequencies of the semi-quantitative analysis. Note: (*) Potato, cocoa, mushrooms, legumes and nuts.

3.4. R24h and BFIs in Lactating Mothers

In Figure 4 (top, left), BFI patterns in mothers' urine samples detected by hierarchical clustering analysis (HCA) are shown. Three clusters were identified as clusters 1, 2, and 3,

and were depicted in the Principal Component Analysis (PCA) scores plot in Figure 4 (top, right). Samples from clusters 1 and 3 presented higher concentrations of all biomarkers than samples from cluster 2, while dairy product and milk biomarkers were more concentrated in cluster 1, and seeds and garlic and onion BFIs in cluster 3 as shown in the PCA loadings plot in Figure 4 (bottom, left).

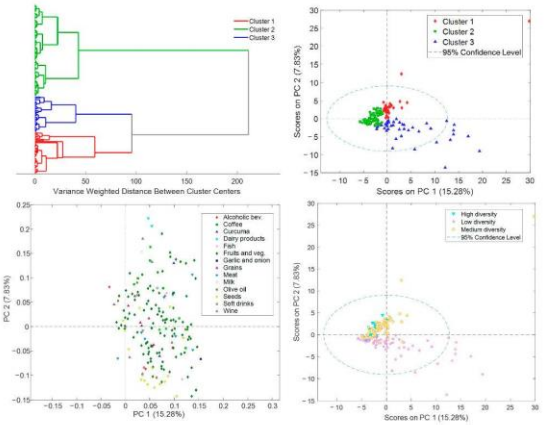


Figure 4. BFIs patterns in mothers' urine samples. **Top**: Hierarchical clustering analysis revealing three sub-groups within the study samples (**left**) and PCA scores plot (**right**). **Bottom**: loadings plot (**left**) and PCA scores plot with Simpson's Index of Diversity classes (**right**).

Figure 5 depicts the mean R24h food group values for each of the three clusters detected by analysing BFI patterns in urine samples. It shows that mothers from cluster 1 had a higher intake of some food groups (i.e., vegetables, meat, egg, and other carbohydrates) than mothers from cluster 3, who presented a lower intake, evidencing that they followed a different dietary pattern. Conversely, mothers in cluster 3 had highest intakes of fruits, fish, bread/starch, seeds, milk, dairy products, fat, and soft drinks. Mothers from cluster 2 showed a mean intake in between mothers from clusters 1 and 3 for all food classes, except for other carbohydrates, egg, bread/starch, and seeds.

In concordance with the results from the HCA, clusters 1 and 3 present higher intakes of the majority of food groups and cluster 3 is characterized by a high intake of seeds. However, R24h results show that the class with lowest consumption of milk and dairy products is cluster 1, contradictory to the HCA results. These divergences could be due to the small absolute difference in the mean R24h food group values between clusters, as the majority of the mothers follow a similar Mediterranean diet (p-value > 0.05, Wilcoxon rank-sum test), and/or errors due to the subjective nature of the R24h recording. Additionally, we would like to highlight the complementary information provided by R24h food groups and BFIs, as the information provided by BFIs is more detailed and reports food groups which are not specified in R24h (e.g., garlic and onion, alcoholic beverages, coffee, curcuma, olive oil, wine). Furthermore, these differences hamper the comparison of the information accessible through both R24h and BFIs.

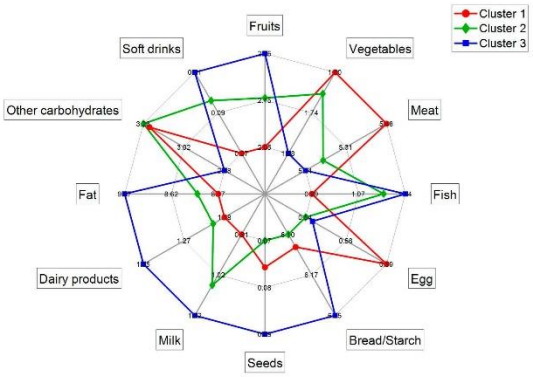


Figure 5. Mean R24h food group values for clusters 1 to 3.

Correlations between urinary BFIs and the food groups obtained in the R24h are depicted in Figure 6, and show significant correlations between three biomarkers (i.e., proline betaine, anserine, and TMAO) and their corresponding food groups (i.e., fruits, meat, and fish) from the R24h, with a p-value < 0.01 for fruits, and p-values < 0.001 for meat and fish ($r > 0.2$, Spearman's correlation). These results strengthen the findings of Lloyd et al., reporting urinary excretion of proline betaine as a potentially useful biomarker of habitual citrus consumption [40], and the conclusions reached by Cheung et al. in which anserine and TMAO are presented as potential biomarkers of chicken and fish intake, respectively [41]. Additionally, there are three other metabolites that are significantly correlated to meat (i.e., phloretin, kaempferol, and taurine), which are described as fruit, vegetable and soft drink BFIs [42,43]. This could be due to a paired intake of vegetables and meat (see Figure 5) and/or due to the correlation of taurine levels with meat intake reported in other studies, as ingestion of foods of animal origin can be another source of taurine [44–46]. It should be remarked that not all food groups recorded in the R24h have their corresponding urinary BFI (i.e., egg, bread/starch, fat and other carbohydrates), due to the difficulty in finding specific BFIs [47].

On the other hand, the specificity of most of the already proposed BFIs is limited, as many dietary compounds are present in different types of foods and similar compounds from different food sources can result in the same metabolite [47]. There are additional concerns, such as the lack of knowledge about the dose–response relationships, the limited quantitative data available, the interindividual variation that can lead to different BFIs levels with the same intake levels [47,48], and the fact that some metabolites are derived not only from diet, but also from other parallel endogenous pathways affected by intake [49]. However, self-reported surveys are known to be prone to misreporting issues and, therefore, can lead to uncertain research findings. Consequently, due to the high complexity of accurate dietary assessments, the implementation of a combined strategy bringing together complementary tools might allow enhanced performance to be achieved.

Despite the fact that the mothers included in the study are similar ages, the same diet could lead to different metabolite by-products due to differences in metabolism between individuals, as recent studies have found significant associations between food intake metabolites and body mass index (BMI) [50,51]. However, in this study we did not find any significant correlation between mothers' BMI and their food intake (p-value > 0.05, Spearman's correlation). In this study, repeated sampling from the same individual was

performed, corresponding to the collection of urine samples and R24h at different study time points, but it has to be taken into consideration that lactating mothers are a particular population, as the body readjusts after giving birth, usually resulting in a weight reduction.

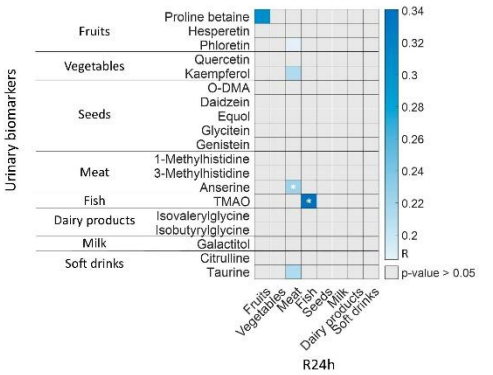


Figure 6. Significant correlations among BFIs and R24h food groups (Spearman's correlation, p-value < 0.05). Note: (*) p-value < 0.001.

3.5. *Microbiota Activity Biomarkers in Lactating Mothers*

The diversity of microbiota activity biomarkers was estimated through the Simpson's Index of Diversity (Equation (1)) and ranged between 0 and 0.6 with a median value of 0.08. Samples were classified as low (N = 111), medium (N = 60), and high (N = 37) diversity for Simpson's Index of Diversity if they were under 0.05, between 0.05 and 0.25, and over 0.25, respectively. In Figure 4 (bottom, right), a PCA scores plot of the urinary BFIs concentrations with the microbiota activity biomarker diversity classes assigned is shown. Samples classified as high and medium diversity presented different BFI levels from the samples classified as low diversity, reflecting a nexus between dietary patterns and microbiota activity. This corroborates findings reported by Turnbaugh et al. indicating that diet has a significant impact on sculpting the microbial communities in the gut, and changes in dietary patterns can directly influence the composition and functionality of the gut microbiota through the availability of macro- and micronutrients in the gut [52]. More specifically, the high number of samples from dietary cluster 3 that show a low microbiota activity diversity (43%) is noticeable. However, BFIs and microbiome analysis should be assessed in more detail in further studies on the impact of diet on the gut microbiome.

4. Conclusions

A UHPLC-MS/MS method for the analysis of BFIs and microbiota activity biomarkers was developed and validated. Levels of microbiota activity biomarkers were reported and a panel of both quantitative and semi-quantitative BFIs was determined in an observational study involving lactating mothers. Correlations between some BFIs in the same food group were found (i.e., meat, dairy products, fruits and vegetables, and seeds). Additionally, the BFI profile allowed three clearly distinct dietary patterns present in the study population to be discerned, evidencing the capability of urinary food intake assessment as a complementary tool to traditionally employed questionnaires. The presence of some highly specific BFIs in urine samples (e.g., curcuma and garlic and onion) highlights its complementary nature. Correlations were found between fruit, meat, and fish biomarker concentrations

and R24h results. This evidences that the use of some biomarker measurements for the assessment of some food group intakes is possible, even though more studies are needed in order to expand the number of food groups that can be assessed by BFIs. Furthermore, the diversity of microbiome activity biomarkers reflected dietary patterns detected in lactating mothers. In future studies, biomarkers and microbiome analysis will be integrated for a joint assessment of diet, microbiome, and, ultimately, health.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/nu15081894/s1, Table S1: Measurement parameters and main figures of merit of the semi-quantitative LC-MS method for BFI determination; Table S2: Measurement parameters and main figures of merit of the targeted LC-MS method for nutrition and microbiota biomarker determination; Table S3: Nutrition and microbiota biomarker levels in urine samples from lactating mothers; and Table S4: Calculated intra- and inter- day accuracy (i.e., recovery) and precision (i.e., RSD) of the LC-MS method in standard solutions and spiked urine samples.

Author Contributions: Conceptualization, J.K., M.G. and E.B.; methodology, V.R.-G., I.T.-D., A.M.-G., L.C.-B., A.P.-L. and E.M.; formal analysis, V.R.-G., E.M., M.K. and D.P.; data curation, V.R.-G., E.M., M.K. and D.P.; writing—original draft preparation, V.R.-G., J.K. and E.M.; writing—review and editing, all authors; supervision, J.K., M.V. and E.B.; project administration, J.K. and E.B.; funding acquisition, J.K. and E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110], the Instituto de Salud Carlos III (Ministry of Science and Innovation, Spain) and co-funded by the European Union (grant numbers CD19/00176, CM20/00143, and CPII21/00003).

Institutional Review Board Statement: The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry #2019-289-1.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available in supplementary material here.

Acknowledgments: The authors would like to thank all the mothers who agreed to participate in this study, as well as the nurses and medical staff involved, for their effort and commitment. Technical support from the Central Support Service for Experimental Research (SCSIE) at the University of Valencia is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brennan, L. 7—Metabotyping: Moving towards personalised nutrition. In *Metabolomics as a Tool in Nutrition Research*; Sébédio, J.-L., Brennan, L., Eds.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Woodhead Publishing: Sawston, UK, 2015; pp. 137–144. ISBN 978-1-78242-084-2.
2. García-Muñoz Rodrigo, F.; Díez Recinos, A.L.; García-Alix Pérez, A.; Figueras Aloy, J.; Vento Torres, M. Changes in perinatal care and outcomes in newborns at the limit of viability in Spain: The EPI-SEN Study. *Neonatology* **2015**, *107*, 120–129. [CrossRef]
3. Vohr, B.R.; Poindexter, B.B.; Dusick, A.M.; McKinley, L.T.; Higgins, R.D.; Langer, J.C.; Poole, W.K.; National Institute of Child Health and Human Development National Research Network. Persistent beneficial effects of breast milk ingested in the neonatal intensive care unit on outcomes of extremely low birth weight infants at 30 months of age. *Pediatrics* **2007**, *120*, e953–e959. [CrossRef] [PubMed]
4. Ballard, O.; Morrow, A.L. Human Milk Composition: Nutrients and Bioactive Factors. *Pediatr. Clin. N. Am.* **2013**, *60*, 49–74. [CrossRef] [PubMed]
5. Innis, S.M. Human milk and formula fatty acids. *J. Pediatr.* **1992**, *120*, S56–S61. [CrossRef] [PubMed]
6. Brenna, J.T.; Varamini, B.; Jensen, R.G.; Diersen-Schade, D.A.; Boettcher, J.A.; Arterburn, L.M. Docosahexaenoic and arachidonic acid concentrations in human breast milk worldwide. *Am. J. Clin. Nutr.* **2007**, *85*, 1457–1464. [CrossRef]
7. Neuhouser, M.L.; Tinker, L.; Shaw, P.A.; Schoeller, D.; Bingham, S.A.; Horn, L.V.; Beresford, S.A.A.; Caan, B.; Thomson, C.; Satterfield, S.; et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am. J. Epidemiol.* **2008**, *167*, 1247–1259. [CrossRef]

8. Lafay, L.; Mennen, L.; Basdevant, A.; Charles, M.A.; Borys, J.M.; Eschwège, E.; Romon, M. Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Santé (FLVS) study. *Int. J. Obes. Relat. Metab. Disord.* **2000**, *24*, 1500–1506. [CrossRef]
9. Jenab, M.; Slimani, N.; Bictash, M.; Ferrari, P.; Bingham, S.A. Biomarkers in nutritional epidemiology: Applications, needs and new horizons. *Hum. Genet.* **2009**, *125*, 507–525. [CrossRef]
10. 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*, 1286–1308. [CrossRef] [PubMed]
11. Maruvada, P.; Lampe, J.W.; Wishart, D.S.; Barupal, D.; Chester, D.N.; Dodd, D.; Djoumbou-Feunang, Y.; Dorrestein, P.C.; Dragsted, L.O.; Draper, J.; et al. Perspective: Dietary Biomarkers of Intake and Exposure—Exploration with Omics Approaches. *Adv. Nutr.* **2019**, *11*, 200–215. [CrossRef] [PubMed]
12. Ulaszewska, M.M.; Weinert, C.H.; Trimigno, A.; Portmann, R.; Andres Lacueva, C.; Badertscher, R.; Brennan, L.; Brunius, C.; Bub, A.; Capozzi, F.; et al. Nutrimetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol. Nutr. Food Res.* **2019**, *63*, 1800384. [CrossRef]
13. González-Domínguez, R. (Ed.) *Mass Spectrometry for Metabolomics*; Methods in Molecular Biology; Springer US: New York, NY, USA, 2023; Volume 2571, ISBN 978-1-07-162698-6.
14. Clarke, E.D.; Rollo, M.E.; Pezdirc, K.; Collins, C.E.; Haslam, R.L. Urinary biomarkers of dietary intake: A review. *Nutr. Rev.* **2020**, *78*, 364–381. [CrossRef]
15. Bar, N.; Korem, T.; Weissbrod, O.; Zeevi, D.; Rothschild, D.; Leviatan, S.; Kosower, N.; Lotan-Pompan, M.; Weinberger, A.; Le Roy, C.I.; et al. A reference map of potential determinants for the human serum metabolome. *Nature* **2020**, *588*, 135–140. [CrossRef] [PubMed]
16. Chen, L.; Zhernakova, D.V.; Kurilshikov, A.; Andreu-Sánchez, S.; Wang, D.; Augustijn, H.E.; Vich Vila, A.; Lifelines Cohort Study; Weersma, R.K.; Medema, M.H.; et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat. Med.* **2022**, *28*, 2333–2343. [CrossRef]
17. Hedrick, V.E.; Dietrich, A.M.; Estabrooks, P.A.; Savla, J.; Serrano, E.; Davy, B.M. Dietary biomarkers: Advances, limitations and future directions. *Nutr. J.* **2012**, *11*, 109. [CrossRef]
18. Playdon, M.C.; Sampson, J.N.; Cross, A.J.; Sinha, R.; Guertin, K.A.; Moy, K.A.; Rothman, N.; Irwin, M.L.; Mayne, S.T.; Stolzenberg-Solomon, R.; et al. Comparing metabolite profiles of habitual diet in serum and urine. *Am. J. Clin. Nutr.* **2016**, *104*, 776–789. [CrossRef] [PubMed]
19. Sébédio, J.-L. Metabolomics, Nutrition, and Potential Biomarkers of Food Quality, Intake, and Health Status. In *Advances in Food and Nutrition Research*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 82, pp. 83–116. ISBN 978-0-12-812633-2.
20. Brouwer-Brolsma, E.M.; Brennan, L.; Drevon, C.A.; van Kranen, H.; Manach, C.; Dragsted, L.O.; Roche, H.M.; Andres-Lacueva, C.; Bakker, S.J.L.; Bouwman, J.; et al. Combining traditional dietary assessment methods with novel metabolomics techniques: Nutrition-Society Summer Meeting/Conference on New Technology in Nutrition Research and Practice / Symposium 2 on Use of Biomarkers in Dietary Assessment and Dietary Exposure. *Proc. Nutr. Soc.* **2017**, *76*, 619–627. [CrossRef] [PubMed]
21. Gu, L.; Laly, M.; Chang, H.C.; Prior, R.L.; Fang, N.; Ronis, M.J.J.; Badger, T.M. Isoflavone conjugates are underestimated in tissues using enzymatic hydrolysis. *J. Agric. Food Chem.* **2005**, *53*, 6858–6863. [CrossRef]
22. Quifer-Rada, P.; Martínez-Huélamo, M.; Lamuela-Raventos, R.M. Is enzymatic hydrolysis a reliable analytical strategy to quantify glucuronidated and sulfated polyphenol metabolites in human fluids? *Food Funct.* **2017**, *8*, 2419–2424. [CrossRef]
23. González-Domínguez, R.; Urpi-Sarda, M.; Jáuregui, O.; Needs, P.W.; Kroon, P.A.; Andrés-Lacueva, C. Quantitative Dietary Fingerprinting (QDF)—A Novel Tool for Comprehensive Dietary Assessment Based on Urinary Nutrimetabolomics. *J. Agric. Food Chem.* **2019**, *68*, 1851–1861. [CrossRef]
24. Feliciano, R.P.; Mecha, E.; Bronze, M.R.; Rodriguez-Mateos, A. Development and validation of a high-throughput micro solid-phase extraction method coupled with ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry for rapid identification and quantification of phenolic metabolites in human plasma and urine. *J. Chromatogr. A* **2016**, *1464*, 21–31. [CrossRef] [PubMed]
25. Spencer, J.P.E.; Abd El Mohsen, M.M.; Minihane, A.-M.; Mathers, J.C. Biomarkers of the intake of dietary polyphenols: Strengths, limitations and application in nutrition research. *Br. J. Nutr.* **2008**, *99*, 12–22. [CrossRef]
26. Wu, G.D.; Chen, J.; Hoffmann, C.; Bittinger, K.; Chen, Y.-Y.; Keilbaugh, S.A.; Bewtra, M.; Knights, D.; Walters, W.A.; Knight, R.; et al. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science* **2011**, *334*, 105–108. [CrossRef]
27. Redondo-Useros, N.; Nova, E.; González-Zancada, N.; Díaz, L.E.; Gómez-Martínez, S.; Marcos, A. Microbiota and Lifestyle: A Special Focus on Diet. *Nutrients* **2020**, *12*, 1776. [CrossRef] [PubMed]
28. Moles, L.; Otaegui, D. The Impact of Diet on Microbiota Evolution and Human Health. Is Diet an Adequate Tool for Microbiota Modulation? *Nutrients* **2020**, *12*, 1654. [CrossRef]
29. Berding, K.; Vlckova, K.; Marx, W.; Schellekens, H.; Stanton, C.; Clarke, G.; Jacka, F.; Dinan, T.G.; Cryan, J.F. Diet and the Microbiota–Gut–Brain Axis: Sowing the Seeds of Good Mental Health. *Adv. Nutr.* **2021**, *12*, 1239–1285. [CrossRef] [PubMed]
30. Dahl, W.J.; Rivero Mendoza, D.; Lambert, J.M. Diet, nutrients and the microbiome. *Prog. Mol. Biol. Transl. Sci.* **2020**, *171*, 237–263. [CrossRef]

31. Bibbò, S.; Ianiro, G.; Giorgio, V.; Scaldaferri, F.; Masucci, L.; Gasbarrini, A.; Cammarota, G. The role of diet on gut microbiota composition. *Eur. Rev. Med. Pharmacol. Sci.* **2016**, *20*, 4742–4749.
32. Kolodziejczyk, A.A.; Zheng, D.; Elinav, E. Diet-microbiota interactions and personalized nutrition. *Nat. Rev. Microbiol.* **2019**, *17*, 742–753. [CrossRef]
33. Chassaing, B.; Gewirtz, A.T. Gut Microbiome and Metabolism. In *Physiology of the Gastrointestinal Tract*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 775–793. ISBN 978-0-12-809954-4.
34. Ramos-Garcia, V.; Ten-Doménech, I.; Moreno-Giménez, A.; Campos-Berga, L.; Parra-Llorca, A.; Ramón, A.; Vaya, M.J.; Mohareb, F.; Molitor, C.; Refinetti, P.; et al. Fact-based nutrition for infants and lactating mothers—The Nutrishield study. *Front. Pediatr.* 2023; *Accepted for publication*. [CrossRef]
35. Hinojosa Nogueira, D.J.; Romero Molina, D.; Giménez Asensio, M.J.; González Alzaga, B.; López Flores, I.; Pastoriza de la Cueva, S.; Rufián Henares, J.Á.; Hernández Jérez, A.F.; Lacasaña, M. Validity and Reproducibility of a Food Frequency Questionnaire to Assess Nutrients Intake of Pregnant Women in the South-East of Spain. *Nutrients* **2021**, *13*, 3032. [CrossRef]
36. Conway, J.M.; Ingwersen, L.A.; Moshfegh, A.J. Accuracy of dietary recall using the USDA five-step multiple-pass method in men: An observational validation study. *J. Am. Diet. Assoc.* **2004**, *104*, 595–603. [CrossRef]
37. Panagiotakos, D.B.; Pitsavos, C.; Stefanadis, C. Dietary patterns: A Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. *Nutr. Metab. Cardiovasc. Dis.* **2006**, *16*, 559–568. [CrossRef]
38. FDA. *Guidance for Industry: Bioanalytical Method Validation*; Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine: Rockville, MD, USA, 2018; Volume 44.
39. *Healthy Eating during Pregnancy and Breastfeeding—Booklet for Mothers*; World Health Organization: Geneva, Switzerland, 2021.
40. Lloyd, A.J.; Beckmann, M.; Favé, G.; Mathers, J.C.; Draper, J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br. J. Nutr.* **2011**, *106*, 812–824. [CrossRef]
41. Cheung, W.; Keski-Rahkonen, P.; Assi, N.; Ferrari, P.; Freisling, H.; Rinaldi, S.; Slimani, N.; Zamora-Ros, R.; Rundle, M.; Frost, G.; et al. A metabolomic study of biomarkers of meat and fish intake. *Am. J. Clin. Nutr.* **2017**, *105*, 600–608. [CrossRef] [PubMed]
42. Krogholm, K.S.; Bredsdorff, L.; Alinia, S.; Christensen, T.; Rasmussen, S.E.; Dragsted, L.O. Free fruit at workplace intervention increases total fruit intake: A validation study using 24 h dietary recall and urinary flavonoid excretion. *Eur. J. Clin. Nutr.* **2010**, *64*, 1222–1228. [CrossRef]
43. Gibbons, H.; McNulty, B.A.; Nugent, A.P.; Walton, J.; Flynn, A.; Gibney, M.J.; Brennan, L. A metabolomics approach to the identification of biomarkers of sugar-sweetened beverage intake. *Am. J. Clin. Nutr.* **2015**, *101*, 471–477. [CrossRef]
44. Cross, A.J.; Major, J.M.; Sinha, R. Urinary biomarkers of meat consumption. *Cancer Epidemiol. Biomark. Prev.* **2011**, *20*, 1107–1111. [CrossRef] [PubMed]
45. Rasmussen, L.G.; Winning, H.; Savorani, F.; Toft, H.; Larsen, T.M.; Dragsted, L.O.; Astrup, A.; Engelsen, S.B. Assessment of the Effect of High or Low Protein Diet on the Human Urine Metabolome as Measured by NMR. *Nutrients* **2012**, *4*, 112–131. [CrossRef] [PubMed]
46. Jacobsen, J.G.; Smith, L.H. Biochemistry and physiology of taurine and taurine derivatives. *Physiol. Rev.* **1968**, *48*, 424–511. [CrossRef]
47. Garcia-Aloy, M.; Rabassa, M.; Casas-Agustench, P.; Hidalgo-Liberona, N.; Llorach, R.; Andres-Lacueva, C. 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**, *69*, 220–229. [CrossRef]
48. Brennan, L. Metabolomics: A tool to aid dietary assessment in nutrition. *Curr. Opin. Food Sci.* **2017**, *16*, 96–99. [CrossRef]
49. Praticò, G.; Gao, Q.; Scalbert, A.; Vergères, G.; Kolehmainen, M.; Manach, C.; Brennan, L.; Pedapati, S.H.; Afman, L.A.; Wishart, D.S.; et al. Guidelines for Biomarker of Food Intake Reviews (BFIRev): How to conduct an extensive literature search for biomarker of food intake discovery. *Genes Nutr.* **2018**, *13*, 3. [CrossRef] [PubMed]
50. Moore, S.C.; Matthews, C.E.; Sampson, J.N.; Stolzenberg-Solomon, R.Z.; Zheng, W.; Cai, Q.; Tan, Y.T.; Chow, W.-H.; Ji, B.-T.; Liu, D.K.; et al. Human metabolic correlates of body mass index. *Metabolomics* **2014**, *10*, 259–269. [CrossRef]
51. Schutte, B.A.M.; van den Akker, E.B.; Deelen, J.; van de Rest, O.; van Heemst, D.; Feskens, E.J.M.; Beekman, M.; Slagboom, P.E. The effect of standardized food intake on the association between BMI and 1H-NMR metabolites. *Sci. Rep.* **2016**, *6*, 38980. [CrossRef] [PubMed]
52. Turnbaugh, P.J.; Ridaura, V.K.; Faith, J.J.; Rey, F.E.; Knight, R.; Gordon, J.I. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Sci. Transl. Med.* **2009**, *1*, 6ra14. [CrossRef]

Clinica Chimica Acta 532 (2022) 172–180

Contents lists available at ScienceDirect

Clinica Chimica Acta

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

GC-MS analysis of short chain fatty acids and branched chain amino acids in urine and faeces samples from newborns and lactating mothers

Victoria Ramos-Garcia [a], Isabel Ten-Doménech [a], Alba Moreno-Giménez [a], Laura Campos-Berga [a], Anna Parra-Llorca [b], Álvaro Solaz-García [a], Inmaculada Lara-Cantón [a], Alejandro Pinilla-Gonzalez [a], María Gormaz [b], Máximo Vento [a,b], Julia Kuligowski [a,*], Guillermo Quintás [c,d]

[a] Neonatal Research Unit, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[b] División of Neonatology, University & Polytechnic Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[c] Health and Biomedicine, Leitat Technological Center, Carrer de la Innovació, 2, 08225 Terrassa, Spain
[d] Analytical Unit, Health Research Institute La Fe, Avda Fernando, Abril Martorell 106, 46026 Valencia, Spain

ARTICLE INFO

Keywords:
Preterm infants
Urine
Faeces
Gas-Chromatography – Mass Spectrometry (GC-MS)
Short-Chain Fatty Acids (SCFAs)
Branched-Chain Amino Acids (BCAAs)

ABSTRACT

Background: Short chain fatty acids (SCFAs) and branched chain amino acids (BCAAs) are frequently determined in faeces, and widely used as biomarkers of gut-microbiota activity. However, collection of faeces samples from neonates is not straightforward, and to date levels of these metabolites in newborn's faeces and urine samples have not been described.
Methods: A targeted gas chromatography – mass spectrometry (GC–MS) method for the determination of SCFAs and BCAAs in both faeces and urine samples has been validated. The analysis of 210 urine and 137 faeces samples collected from preterm (PI), term infants (TI) and their mothers was used to report faecal and urinary SCFA and BCAA levels in adult and neonatal populations.
Results: A significant correlation among five SCFAs and BCAAs in faeces and urine samples was observed. Reference ranges of SCFAs and BCAAs in mothers, PI and TI were reported showing infant's lower concentrations in faeces and higher concentrations in urine.
Conclusion: This method presents a non-invasive approach for the simultaneous assessment of SCFAs and BCAAs in faecal and urine samples and the results will serve as a knowledge base for future experiments that will focus on the study of the impact of nutrition on the microbiome of lactating mothers and their infants.

1. Introduction

The host-gut microbiota crosstalk is vital for the development and maturation of the immune system [1–3] and it has been widely acknowledged that this crosstalk is regulated by nutrition [4]. Gut microbiota provides essential nutrients, such as vitamins, and helps catabolizing dietary fibres, human milk oligosaccharides, and other carbohydrates to produce short chain fatty acids (SCFAs), which are an important energy source for the intestinal epithelium, and modulate epithelial integrity [5]. These compounds provide general protection against pathogen gut colonization [6]; moreover, taken up by the host, they stimulate the development of the immune system and are used as energy sources or regulators [7], being particularly relevant in the perinatal period.

The characterization of SCFA profiles in faeces is widely used to study the activity of gut microbiota in clinical studies. The most abundant SCFAs are acetate, propionate, and butyrate produced by anaerobic fermentation in the intestine. Valerate, caproate, heptanoate, and branched chain SCFAs (i.e., 2-methyl-butyrate, isobutyrate, and isovalerate) are also produced by intestinal microbiota, but in considerably lower amounts. Branched chain SCFAs are derived from the bacterial metabolism of dietary proteins, human milk oligosaccharides and branched chain amino acids (BCAAs) [1,7–9]. BCAAs (i.e., valine, leucine, and isoleucine) are considered essential amino acids since they that cannot be synthesized endogenously and must be provided by the diet. They serve as substrates for protein synthesis or energy production

* Corresponding author.
E-mail address: julia.kuligowski@uv.es (J. Kuligowski).

https://doi.org/10.1016/j.cca.2022.05.005
Received 22 March 2022; Received in revised form 29 April 2022; Accepted 3 May 2022
Available online 8 May 2022

and are of key importance in several metabolic and signalling functions, e.g. the activation of the mammalian target of rapamycin (mTOR) signalling pathway [10].

Mass spectrometry (MS) hyphenated with separation techniques, such as gas chromatography (GC) or liquid chromatography (LC), is frequently used for the metabolic analysis of biological samples due to its sensitivity and selectivity. LC-MS requires minimal sample preparation for SCFA and BCAA analysis. However, harsh experimental conditions, such as a very acidic aqueous mobile phases, need to be employed in LC-MS for the quantitation of SCFAs without chemical derivatization [11]. In addition, poor chromatographic separation, and insufficient ionization due to the hydrophilicity of SCFAs are an additional burden [12]. To overcome these problems, several chemical derivatization methods have been developed for SCFA and BCAA quantification with LC-MS [13–15], requiring long reaction times or specific reaction conditions. Alternatively, SCFAs and BCAAs have been derivatized by methyl-, ethyl-, and propyl-chloroformate, as well as trimethylsilylation, and determined by GC–MS [16–19].

Preterm birth is a serious event affecting 5–13% of births with an increasing tendency in recent years [20]. Prematurely born infants are characterized by an immature immune system and gut barrier, and hence, they are more vulnerable to suffer infections and inflammatory processes such as necrotizing enterocolitis, associated with high mortality rates [21]. In studies targeting the newborn population, the collection and handling of stool samples can be troublesome, and the determination of absolute concentrations typically requires the normalization of the detected concentrations to wet or dry weight of faecal samples. Hence, alternative procedures supporting straightforward sample collection and enabling the quantitative analysis of these biomarkers are warranted. The assessment of newborns' gut microbiota activity using metabolic profiles of urine samples as surrogates could enable an effective simplification of sample collection protocols in clinical studies, which would be especially useful in neonatology.

In the present study we show the applicability of a novel GC–MS method for the determination of SFCAs and BCAAs in both faecal and urine samples, specifically tailored to the very low sample volumes typically available in studies involving newborns. The method encompasses a 1-step derivatization of SCFAs and BCAAs to propyl-esters and provides appropriate sensitivity, linearity, and accuracy for their quantification. The applicability of the method for the determination of SCFA and BCAA concentration levels is assessed through the analysis of a set of 137 faecal and 210 urine samples from a clinical cohort comprising lactating mothers and term (TI) and preterm infants (PI), providing for the first time reference concentration ranges of these metabolites, and an initial study of the association between the levels found in faeces and urine in paired samples. Results obtained during method development include the pre-screening of sterile polypropylene (PP) bags, cotton pads, and gauzes widely used for urine collection from the newborn for assessing their suitability with respect to the determination of SCFAs and BCAAs in future clinical studies.

2. Material and methods

2.1. Standards and reagents

Certified ACS grade sodium hydroxide, HPLC grade propanol (PrOH), pyridine (Py), propyl-chloroformate (PCF), and *n*-hexane, and SCFAs (acetic acid (≥99.8%), propionic acid (≥99.5%), isobutyric acid (≥99%), butyric acid (≥99%), 2-methyl butyric acid (≥98%), isovaleric acid (≥99%), valeric acid (≥99%), caproic acid (≥99%), and heptanoic acid (≥99%)), and BCAAs (valine (≥99.5%), leucine (≥99.5%), and isoleucine (≥99%)) standards as well as the isotopically labelled internal standard (IS) caproic acid-D_3 (≥99%) were purchased from Sigma-Aldrich Química SL (Madrid, Spain). Standard solutions were prepared in ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany).

Table 1
Characteristics of the population and general confounders.

Parameter	M (N = 45)	TI (N = 9)	PI (N = 50)	*p*-value
Maternal age (years)/ Gestational age (weeks), mean (SD)/ median (5–95% CI)	35 (5)	37 (1.2)	30 (0.9)	<0.001
Maternal weight (kg)/ infant birth weight (g), mean (SD)	67 (12)	2486 (776)	1420 (445)	<0.001
BMI (kg/m^2), mean (SD)	26 (7)	–	–	–
Sex male, n (%)	0 (0)	8 (89)	29 (58)	0.09
C-Section delivery, n (%)	22 (48)	4 (44)	38 (76)	0.06
Apgar[1] 1 min, median (5–95% CI)	–	9 (1.4)	7 (0.8)	0.6
Apgar[1] 5 min, median (5–95% CI)	–	10 (1.2)	9 (0.4)	0.8
Antibiotic therapy, n (%)	4 (9)	3 (38)	2 (5)	0.006

Note: Body mass index (BMI) weight/height^2; SD standard deviation; CI confident interval; M mothers; TI term infants; PI preterm infants; [1][22].

2.2. Study design, population, and sample collection

Samples were collected in the frame of the Nutrishield project (https://nutrishield-project.eu/), in a prospective, observational, cohort study performed at the Division of Neonatology of the University and Polytechnic Hospital La Fe (HUiP La Fe), including three groups of infants: i) PI (<32 weeks of gestation) exclusively (i.e., >80% v/v) receiving own mother's milk (N = 30), ii) PI exclusively receiving donor human milk (N = 7) and iii) TI (>37 weeks of gestation) exclusively receiving own mother's milk (N = 7), as well as mothers of infants of all three groups (N = 45). Urine and faeces samples from infants and mothers were collected at six time points, covering the period from achieving complete enteral nutrition, i.e., a stable daily intake of 150 mL/kg, in PI and recovery of birth weight in TI, followed by sampling at one, two, three and six months of infant's age. Additional faeces and matched urine samples from neonates (N = 15) during their stay at the hospital were collected. A total of 45 mothers, 50 PI and 9 TI were included in the study. Demographic, clinical, and perinatal characteristics from participants were recorded as shown in Table 1. Mother's first morning urine was collected in sterile PP containers and infants' urine was collected using sterile cotton pads placed in the diaper. Cotton pads were collected after 1 h and squeezed with a sterile PP syringe. The process was repeated until collecting a minimum of 1 mL. Urine samples were aliquoted to avoid freeze-thawing cycles and stored at −80 °C until analysis. Faeces samples were collected in sterile tubes using sterile tweezers. In total, 210 urine and 137 faeces samples collected were employed for method development and for the comparison between SCFA and BCAA concentrations in faeces and urine.

The study protocol was approved by the Scientific and Ethics Committee for Biomedical Research (CEIm) of the HUiP La Fe (study approvals #2019–289-1 and #2019/0312). All methods were performed in accordance with the relevant guidelines and regulations and written permission was obtained from mothers or legal representatives by signing an informed consent form.

2.3. GC–MS determination of SCFAs and BCAAs

The GC–MS method for the determination of SCFAs (i.e., acetic acid, propionic acid, butyric acid, valeric acid, caproic acid, heptanoic acid, isobutyric acid, 2-methylbutyric acid, and isovaleric acid) and BCAAs (i.e. valine, leucine, and isoleucine) was developed based on previous results [18]. 1000 μL of 5 mM aqueous NaOH containing IS (5 μL mL^{-1} caproic acid-D_3) was added to 100 mg of wet faeces, then the sample was homogenized for 10 min and centrifuged (13200 × *g* for 20 min at 4 °C). 500 μL of faecal supernatant was transferred into a 15 mL falcon tube, and 300 μL of water were added. For urine, 300 μL of sample and 500 μL of 5 mM aqueous NaOH containing IS (5 μL mL^{-1} caproic acid-D_3) were mixed in a 15 mL falcon tube. Both faecal and urine samples were derivatized as follows. An aliquot of 500 μL PrOH/Py solvent mixture

Table 2
Measurement parameters and main figures of merit of the GC–MS method.

				Faeces				Urine			
Compound class	Metabolite	m/z	RT ± s (min)	Calibration range (μM)	R^2	LOD (μM)	LOQ (μM)	Calibration range (μM)	R^2	LOD (μM)	LOQ (μM)
SCFA	Acetic acid	61	2.75 ± 0.02	2–8593	0.9996	0.3	1.0	0.7–655	0.997	0.5	2
	Propionic acid	75	3.98 ± 0.02	0.5–8927	0.998	0.10	0.2	0.6–547	0.994	0.6	2
	Isobutyric acid	89	4.763 ± 0.012	0.5–8149	0.998	0.12	0.4	0.6–609	0.996	0.12	0.4
	Butyric acid	89	5.59 ± 0.02	0.10–1619	0.999	0.05	0.2	0.2–171	0.997	0.110	0.4
	2-Methylbutyric acid	103	6.733 ± 0.014	0.08–1332	0.998	0.07	0.2	0.13–121	0.996	0.10	0.3
	Isovaleric acid	85	6.83 ± 0.03	0.05–783	0.9997	0.02	0.07	0.07–65	0.998	0.05	0.2
	Valeric acid	103	8.24 ± 0.03	0.07–1077	0.9996	0.08	0.3	0.10–98	0.997	0.05	0.2
	Caproic acid	117	11.22 ± 0.02	0.08–1309	0.9998	0.010	0.04	0.10–98	0.998	0.04	0.13
	Caproic acid-d3 (IS)	120	11.281 ± 0.014	–	–	–	–	–	–	–	–
	Heptanoic acid	131	14.31 ± 0.03	0.05–896	0.998	0.012	0.04	0.05–52	0.998	0.02	0.05
BCAA	Valine	158	17.843 ± 0.012	0.14–2328	0.996	0.2	0.8	1.0–908	0.998	0.5	2
	Leucine	172	18.164 ± 0.012	0.13–2165	0.997	0.112	0.4	0.9–844	0.997	0.5	2
	Isoleucine	172	18.240 ± 0.015	0.12–1926	0.9995	0.110	0.2	0.8–751	0.998	0.4	1.3

Note: RT retention time; R coefficient of determination; Limit of quantification (LOQ) concentration of analyte that can be measured with an imprecision of <20% and a deviation from target of <20%, taking into account the preconcentration factor achieved during sample processing; Limit of detection (LOD) 3/10*LOQ.

(v/v = 3:2) and 100 μL of PCF were added and vortexed briefly. Derivatization was carried out during 1 min in an ultrasonic water bath prior to a two-step extraction by adding 300 and 200 μL of *n*-hexane, respectively, followed by centrifugation (2000 × *g* for 5 min at 25 °C). The upper *n*-hexane layers containing the extracted derivatives were collected and pooled followed by thorough mixing during 3 s prior to analysis.

GC–MS analysis was conducted using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole mass spectrometric detector (Agilent Technologies, Santa Clara, CA, USA) operating in selected ion monitoring (SIM) mode. Separations were performed using an HP-5 MS capillary column coated with 5% phenyl-95% methylpolysiloxane (30 m × 250 μm i.d., 0.25 μm film thickness, Agilent J & W Scientific, Folsom, CA, USA). One microliter of derivatized sample extracts was injected in split mode with a ratio of 10:1, and the solvent delay time was set to 2.36 min. The initial oven temperature was held at 50 °C for 2 min, ramped to 70 °C at a rate of 10 °C min^{-1}, to 85 °C at a rate of 3 °C min^{-1}, to 110 °C at a rate of 5 °C min^{-1}, to 290 °C at a rate of 30 °C min^{-1}, and finally held at 290 °C for 8 min. Helium was used as a carrier gas at a constant flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source were set at 260, 290, and 230 °C, respectively, and the electron energy was −70 eV.

The measurement parameters used for the studied analytes are summarized in Table 2. For quantification, an external calibration was carried out using standard solutions obtained from serial dilutions of a working solution containing mixtures of pure analytical standards in ultrapure water. Aliquots of a quality control (QC) sample of each matrix were analysed every 10 samples in the randomized analytical batch for monitoring the instrument's performance. QC RSD < 25% was the batch acceptance criterion. In addition, calibration blanks (i.e., addition of H_2O instead of sample into the tubes) and a process blank (addition of H_2O instead of sample to a PP bag/container, cotton or gauze pad and squeezed before adding it to the tubes) were injected.

SCFA and BCAA concentrations in faecal and urine samples were normalized by dividing by sample weight and creatinine concentration, respectively. Creatinine was quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA). Samples were diluted with deionized water prior to measurements employing a 1:20 and a 1:4 dilution for adult and infant urine, respectively.

2.4. Method validation

The method validation was based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [23], including the following bioanalytical parameters: linearity range, accuracy, precision, recovery, selectivity, specificity, carry over, and stability. The ranges of the analytical method were selected to enable the quantification of the metabolites in faeces and urine samples with adequate precision, accuracy, and linearity. Table 2 summarizes the employed working concentration ranges, which were chosen considering the expected values and a wide inter- and intra-individual variability. The calibration curves included a zero calibrator (i.e., blank with IS) and, at least, 6 standards covering the selected concentration ranges. Accuracy, precision, and recovery were assessed by replicate (n = 3) analysis of standards at three concentration levels and replicate (n = 3) analysis of spiked samples at three concentration levels (low, medium, and high) on three validation days. Precision was estimated as the percentage of relative standard deviation (RSD) of replicate standards within one validation batch (intra-day) and between validation batches (inter-day). Selectivity and specificity were demonstrated by analyzing calibration and process blank samples from multiple (n = 6) sources. Carry-over between samples was assessed by the analysis of zero-injections after the analysis of high concentrated standards and spiked samples (n = 3), and by the analysis of a blank without analytes or IS after the injection of the standards. Autosampler sample stability was assessed by comparing concentrations observed in a freshly prepared sample and in the same processed sample after 20 h stored in the autosampler (sealed vial, 25 °C). Analytes' freeze–thaw stability and long-term stability were established by comparing concentrations observed in sample extracts after three freeze–thaw cycles and in a sample stored for one year (-80 °C), respectively, to a freshly prepared sample.

2.5. Data availability and statistical analysis

GC–MS data were acquired and processed using MassHunter B.07.01 (Agilent Technologies, Santa Clara, CA). Data analysis was carried out in MATLAB R2019b (MathWorks, Natick, MA, USA) and using the PLS Toolbox 8.9 (Eigenvector Research Inc., Manson, WA, USA). SCFA and BCAA normalized levels in faeces and urine samples determined in this work are available as Supplementary Tables 1 and 2. Categorical variables were compared using Pearson's χ^2 test ($\alpha = 0.05$). Continuous variables were expressed as mean ± standard deviation or medians with interquartile range, depending on underlying data distribution and Student's *t*-test ($\alpha = 0.05$) or Wilcoxon rank-sum test ($\alpha = 0.05$) were used for inter-group comparisons, respectively. The Pearson's correlation coefficient was used for assessing paired associations among metabolite concentrations. Principal Component Analysis (PCA) was carried out using autoscaled, normalized data using creatinine and wet

Table 3
Calculated intra- and inter-day accuracy (i.e., recovery) and precision (i.e., RSD) of the method in standard solutions and spiked urine and faeces samples.

Matrix	Compound class	Metabolite	Concentration levels (μM)		Accuracy ± RSD											
					Standard solutions						Spiked samples					
					Intra-day (N = 3)			Inter-day (N = 3)			Intra-day (N = 3)			Inter-day (N = 3)		
			Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Urine	SCFA	Acetic acid	15	101	150	99 ± 3	100 ± 7	120 ± 2	116 ± 3	113 ± 8	105 ± 6	105 ± 11	106 ± 2	100 ± 5	102 ± 17	100 ± 7	106 ± 8
		Propionic acid	13	84	126	112 ± 5	99 ± 11	115 ± 9	115 ± 6	115 ± 2	115 ± 4	114 ± 3	100 ± 7	102 ± 6	106 ± 4	102 ± 3	109 ± 7
		Isobutyric acid	14	94	140	108 ± 2	96 ± 3	116 ± 5	102 ± 3	108 ± 5	103 ± 9	93 ± 4	88 ± 3	94 ± 6	92 ± 11	83 ± 3	93 ± 5
		Butyric acid	4	26	39	116 ± 7	118 ± 6	112 ± 5	118 ± 2	96 ± 1	117 ± 8	107 ± 4	106 ± 3	106 ± 4	94 ± 6	98 ± 7	100 ± 5
		2-Methylbutyric acid	3	19	28	115 ± 6	119 ± 4	104 ± 8	99 ± 5	109 ± 2	110 ± 5	110 ± 4	97 ± 3	102 ± 5	99 ± 4	100 ± 7	97 ± 4
		Isovaleric acid	2	10	15	112 ± 9	103 ± 5	118 ± 2	95 ± 4	99 ± 7	116 ± 4	114 ± 10	110 ± 6	112 ± 4	94 ± 12	93 ± 17	95 ± 16
		Valeric acid	2	15	23	100 ± 4	116 ± 11	112 ± 4	120 ± 10	107 ± 8	100 ± 6	102 ± 3	97 ± 4	99 ± 1	96 ± 10	96 ± 5	98 ± 3
		Caproic acid	2	15	23	100 ± 11	117 ± 2	115 ± 9	122 ± 7	110 ± 3	110 ± 7	102 ± 8	102 ± 2	100 ± 1	90 ± 10	97 ± 4	98 ± 2
		Heptanoic acid	1.2	8	12	95 ± 4	98 ± 5	110 ± 2	100 ± 11	97 ± 5	97 ± 9	93 ± 5	99 ± 2	99 ± 3	92 ± 11	97 ± 4	100 ± 3
	BCAA	Valine	21	140	209	107 ± 3	107 ± 9	113 ± 11	112 ± 4	104 ± 3	112 ± 10	110 ± 10	103 ± 8	107 ± 4	101 ± 6	103 ± 5	102 ± 8
		Leucine	19	130	195	103 ± 2	102 ± 1	97 ± 10	103 ± 7	95 ± 3	102 ± 7	110 ± 9	105 ± 8	106 ± 5	86 ± 12	96 ± 11	88 ± 12
		Isoleucine	17	116	173	97 ± 7	97 ± 3	107 ± 3	95 ± 7	114 ± 3	108 ± 7	103 ± 11	94 ± 10	102 ± 9	99 ± 8	96 ± 12	92 ± 9
		Acetic acid	26	103	155	93 ± 7	90 ± 5	85 ± 4	99 ± 8	90 ± 5	105 ± 5	93 ± 5	109 ± 4	99 ± 4	99 ± 4	105 ± 4	112 ± 9
		Propionic acid	27	107	162	116 ± 10	109 ± 4	116 ± 1	103 ± 9	110 ± 3	93 ± 4	92 ± 3	112 ± 2	95 ± 6	106 ± 7	108 ± 2	102 ± 6
		Isobutyric acid	24	98	147	89 ± 4	108 ± 4	105 ± 1	93 ± 1	116 ± 7	104 ± 2	110 ± 4	102 ± 1	93 ± 3	99 ± 3	104 ± 8	96 ± 10
		Butyric acid	5	19	29	117 ± 3	95 ± 5	95 ± 5	99 ± 7	88 ± 2	98 ± 7	100 ± 5	113 ± 10	118 ± 7	108 ± 5	110 ± 5	110 ± 7
	SCFA	2-Methylbutyric acid	4	16	24	105 ± 8	103 ± 10	116 ± 1	97 ± 5	89 ± 1	98 ± 1	110 ± 5	114 ± 5	116 ± 6	108 ± 1	108 ± 6	114 ± 4
Faeces		Isovaleric acid	2	9	14	106 ± 7	86 ± 1	112 ± 5	110 ± 8	109 ± 2	117 ± 9	95 ± 4	107 ± 4	95 ± 4	100 ± 5	108 ± 4	107 ± 6
		Valeric acid	3	13	19	104 ± 11	88 ± 4	106 ± 1	97 ± 2	104 ± 6	98 ± 8	106 ± 10	103 ± 2	94 ± 6	106 ± 10	103 ± 1	94 ± 9
		Caproic acid	4	16	24	110 ± 13	92 ± 9	102 ± 3	110 ± 5	90 ± 8	119 ± 4	108 ± 6	99 ± 3	97 ± 6	108 ± 5	99 ± 9	97 ± 9
		Heptanoic acid	3	11.2	16	119 ± 9	100 ± 10	107 ± 1	115 ± 9	102 ± 3	108 ± 5	112 ± 5	110 ± 5	92 ± 4	112 ± 7	110 ± 9	92 ± 1
	BCAA	Valine	35	140	209	100 ± 5	86 ± 4	95 ± 5	118 ± 3	116 ± 10	112 ± 1	105 ± 6	115 ± 4	113 ± 9	106 ± 8	103 ± 10	94 ± 2
		Leucine	32	130	195	115 ± 9	104 ± 2	95 ± 4	117 ± 3	116 ± 1	93 ± 9	103 ± 5	108 ± 7	112 ± 5	108 ± 3	99 ± 3	97 ± 8
		Isoleucine	29	116	173	103 ± 11	103 ± 2	103 ± 11	90 ± 8	116 ± 3	90 ± 5	106 ± 2	103 ± 3	94 ± 6	112 ± 3	110 ± 10	92 ± 10

Table 4
SCFA concentrations in faecal and urine samples.

Compound class	Metabolite	Faeces (N = 137), μmol/g faeces				Urine (N = 210), μmol/g creatinine			
		Range	Median	IQR (25–75)	Detection frequency (%)	Range	Median	IQR (25–75)	Detection frequency (%)
SCFA	Acetic acid	0.09–72	21	18	100	23–4935	76	60	100
	Propionic acid	0.08–18	0.6	2	98	0.08–3761	5	47	72
	Isobutyric acid	0.012–84	13	25	99	2–47841	1534	4004	100
	Butyric acid	0.013–17	3	5	100	7–844	42	72	96
	2-Methylbutyric acid	0.012–5	0.8	2	96	0.12–67	3	15	88
	Isovaleric acid	0.014–3	0.5	0.6	96	0.02–15	0.2	3	42
	Valeric acid	0.03–3	0.6	1.0	93	0.02–252	2	7	80
	Caproic acid	0.012–2	0.02	0.2	95	0.03–193	2	13	84
	Heptanoic acid	0.02–0.3	0.0012	0.03	90	0.02–19	0.8	2	93
BCAA	Valine	0.03–8	0.3	0.3	100	4–1132	88	137	100
	Leucine	0.013–1.3	0.4	0.4	99	5–1955	154	266	100
	Isoleucine	0.02–2	0.3	0.2	100	2–678	53	108	100

Note: IQR interquartile range.

weight as normalization factor for urine and faeces, respectively.

3. Results and discussion

3.1. Method validation

Analytical method validation was performed following recommended guidelines [23]. Linearity of response was assessed covering up to four orders of magnitude with LODs and LOQs in the 0.02–0.6 and 0.05–2 μM ranges in urine, and in the 0.01–0.3 and 0.04–1.0 μM ranges in faeces, respectively (see Table 2). Appropriate accuracies with recoveries between 95 and 122% and 85 and 119% at the LOQ and precisions ranging between 1 and 11 %RSD and 1 and 13 %RSD at the LOQ were observed in standard solutions in urine and faeces, respectively, as shown in Table 3. Stock solutions from all compounds showed good stability after 1 year at −80 °C, and after three freeze–thaw cycles (-80 °C), and derivatized compounds were stable for 20 h in autosampler storage conditions (*t*-test, *p*-values > 0.05). Extraction efficiency and matrix effects were assessed in spiked samples. Data summarized in Table 3 displayed excellent intra- and inter-day accuracy and precision for all compounds in spiked faeces and urine samples, evidencing an adequate method performance for all nine SCFAs and three BCAAs.

The impact of different sample collection procedures on the retrieved SCFA and BCAA profiles was evaluated. Several options are available for the collection of urine samples and their selection largely depends on the study population. Typically, sterile PP containers are used for the collection of urine samples from adults. Likewise, for faeces samples PP containers are also typically employed. However, collecting urine samples from newborns can be a very difficult task as they are obviously non-toilet trained. The recommended procedure for clean catch urine collection involves waiting for a nappy-free child to void spontaneously [24]. To overcome the limitations of this approach, different methods to stimulate voiding were recently proposed, involving gently rubbing the lower abdomen in circular motions with a piece of gauze soaked in cold liquid [25] and bladder stimulation followed by paravertebral lumbar massage, to trigger urination [26,27]. However, the efficiency of this method has not been demonstrated in preterm infants and so, either sterile plastic bags or cotton pads and gauzes that are placed in the diaper are typically used in this population. To test the compatibility of the sample collection procedure with the determination of SCFAs and BCAAs, process blanks using sterile PP bags, cotton pads, and gauzes were analysed and compared to calibration blanks. Using PP materials (i.e., either containers or bags), no SCFAs or BCAAs were detected in blanks. Remarkably, an intense background concentration of acetic acid was observed when cotton pads or gauzes were used (520 ± 3 and 850 ± 20 μM, respectively). Hence, in case of urine samples collected with cotton pads or gauzes, as it is typically the case for urine samples from newborns, acetic acid concentrations must be excluded from further data analysis. The observed peak area values of the other considered metabolites and the IS were within the acceptability criteria (i.e., <1/10 of the signal at LOQ).

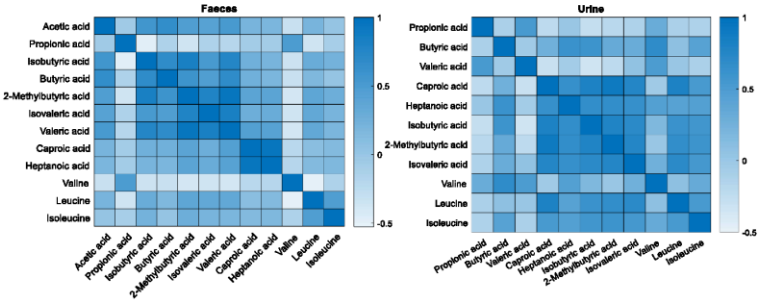


Fig. 1. Pearson correlation coefficients of SCFAs and BCAAs in faecal supernatants and urine samples.

Table 5
Linear regression parameters among faecal and urinary SCFA and BCAA concentrations in paired samples (N 80).

Compound class	Metabolite	Slope	*p*-value	R^2
SCFA	Acetic acid	−0.6 ± 2.0	0.8	0.003
	Propionic acid	45 ± 12	< 0.001	0.2
	Isobutyric acid	−8 ± 10	0.4	0.009
	Butyric acid	−5 ± 3	0.04	0.05
	2-Methylbutyric acid	−4.3 ± 1.0	< 0.001	0.2
	Isovaleric acid	−2.3 ± 0.6	< 0.001	0.2
	Valeric acid	−4 ± 3	0.2	0.03
	Caproic acid	−8 ± 5	0.12	0.03
	Heptanoic acid	−3 ± 2	0.12	0.03
BCAA	Valine	44 ± 12	< 0.001	0.2
	Leucine	−360 ± 83	< 0.001	0.2
	Isoleucine	−42 ± 30	0.2	0.02

3.2. Analysis of SCFAs and BCAAs in faeces and urine

Table 4 summarizes the concentrations of SCFAs determined in faeces and urine samples, after normalization to sample weight and to creatinine, respectively, using the validated GC–MS method. Results showed that acetic acid is the most concentrated SCFA present in faeces and together with isobutyric acid represents 86% of SCFAs molarity in this matrix. With respect to urine, isobutyric acid is the most abundant SCFA followed by acetic and butyric acids, and they account for 99% of total SCFAs measured in this matrix. Unfortunately, although acetic acid contributes noticeably to total SCFAs in faeces, the levels of this metabolite in newborn's urine were excluded due to the above-mentioned contamination during sample collection, and therefore its contribution could not be studied. The lowest relative concentrations among the considered metabolites were found for caproic acid in faeces, isovaleric acid in urine and heptanoic acid in both matrices. While all metabolites presented a detection frequency higher than 93% in faeces samples, propionic, 2-methylbutyric, isovaleric, valeric, and caproic

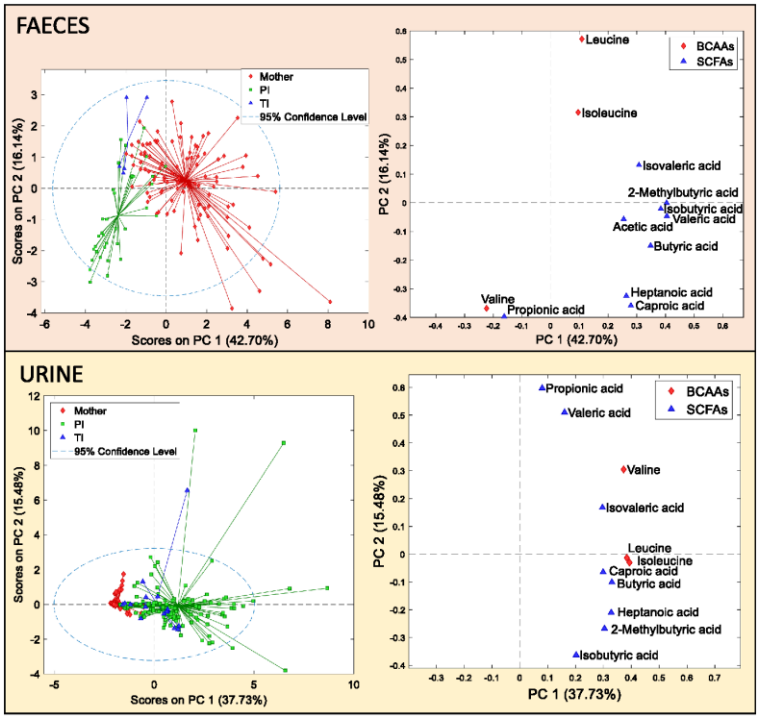


Fig. 2. PCA scores (left) and loadings plots (right) in faeces (top) and urine (bottom) samples.

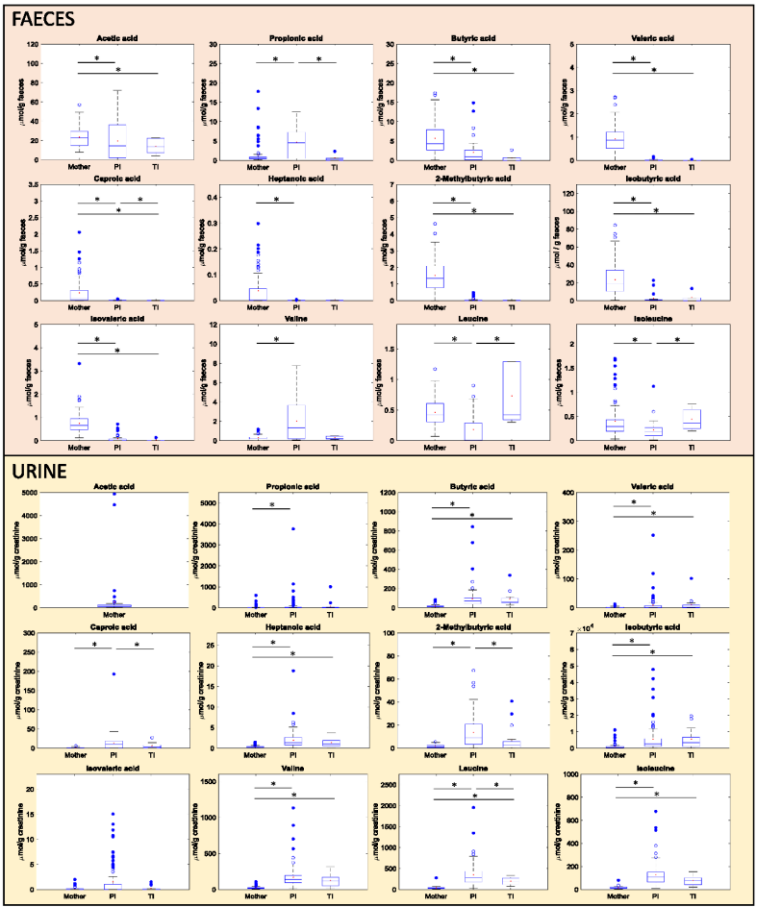


Fig. 3. Faecal and urinary SCFA and BCAA levels in mothers and their term (TI) and preterm infants (PI). Note: * indicates significant differences (Wilcoxon rank-sum test, p-value < 0.05).

acids were detected in 42 to 88% of urine samples.

Fig. 1 depicts the paired correlations among SCFAs and BCAAs found in faecal and urine samples separately, and Table 5 summarizes results from a regression analysis between faecal and urinary SCFA and BCAA concentrations. As a general trend, it can be observed that SCFAs in faecal supernatants are significantly correlated among each other. Interestingly, similar patterns of significant correlations among SCFAs were observed in urine samples, except in the case of propionic, butyric and valeric acids. BCAAs in both faeces and urine samples followed a similar pattern, being the correlation between leucine and isoleucine

higher than with valine. Between both matrices, i.e., faeces and urine, significant positive correlations were found for propionic acid and valine and significant negative correlations were found for butyric, 2-methylbutyric, and isovaleric acid as well as leucine.

3.3. SCFAs and BCAAs in samples from newborn infants and their mothers

As shown in Table 1, demographic, clinical, and perinatal characteristics from participants were recorded and no differences in prenatal demographic characteristics or confounders during the hospitalization and antibiotic therapy were found between both infant groups (p-value > 0.05), except from gestational age and birth weight, according to the definition of PI and TI. Metabolite levels in both matrices were compared between lactating mothers, TI, and PI. Fig. 2 (top, left) shows PCA scores plots of faeces samples from mothers and infants. Three clusters for mothers, PI and TI can be observed when considering faecal SCFA and BCAA concentrations. For urinary SCFA and BCAA concentrations (Fig. 2 bottom, left), a clear separation of the tightly clustered group of urine samples collected from lactating mothers from those of infants is evidenced, although SCFA and BCAA profiles of infants were overlapping in the PC1-PC2 scores space summarizing 53% of the variance. The joint analysis of the scores and loadings plot (Fig. 2, top) shows that faecal SCFAs and BCAAs are present at higher concentrations in maternal samples, and a negative correlation of valine and propionic acid with isoleucine and leucine, linked to the clustering of PI samples in the scores plot. On the other hand, the analysis of the scores and loadings plots from the analysis of urinary SCFAs and BCAAs (Fig. 2, bottom) shows a high positive correlation between isoleucine and leucine, and higher levels of the metabolites in infant (PI and TI) samples. Hence, the determination of both SCFAs and BCAAs in urine and faeces samples using the developed GC–MS method provides information that allows to distinguish between groups in the present data set.

Fig. 3 shows the biomarker concentration ranges in the three study groups. In faeces, higher concentration values were detected in mothers in comparison to PI and TI, except from propionic acid and BCAAs (Wilcoxon rank-sum test, $\alpha = 0.05$). In urine, higher concentration values of some SCFAs (i.e., butyric, valeric, heptanoic and isobutyric acids) and all BCAAs were detected when comparing M to PI and TI, respectively. Regarding PI vs. TI, slightly higher faecal concentration values in PI were identified in two SCFAs (i.e., propionic and caproic acids) and lower faecal concentration values in PI were identified in two BCAAs (i.e., leucine and isoleucine). In urine, higher concentration values of caproic and 2-methylbutyric acids and leucine were found in PI. No statistically significant correlations between biomarker concentrations of mothers and their infants or between biomarker concentrations and gestational or postnatal age were found in this study (Pearson's χ^2 test, $\alpha = 0.05$). Likewise, no statistically significant differences were observed between male and female infants in any of the assessed metabolites (Wilcoxon rank-sum test, $\alpha = 0.05$).

Further studies are needed to gain a deeper understanding of the biological meaning of SCFA and BCAA levels in faeces and urine samples and the effect of possible confounders such as, e.g., gestational, and postnatal age, type of nutrition. The biochemical interpretation of the results is hampered by the difficulty to assess the origin of SCFAs and BCAAs in urine due to the possibility of their generation by microbial activity in other parts of the body. Nonetheless, a previous work showed an association between the SCFA urinary profiles and the disease status of patients with ulcerative colitis, where higher butyrate levels in patients at remission were observed, and it was suggested that the urinary output could reflect the levels of butyrate produced in the gut [28].

4. Conclusions

A method for the analysis of nine SCFAs and three BCAAs in faeces and urine samples has been reported. The method enabled an accurate quantification of these biomarkers in urine and faeces using a single derivatization procedure. It was successfully validated including the entire test system, from sample collection to analysis, revealing a very significant contamination with acetic acid from cotton and gauze pads typically employed for newborn urine samples' collection that should be considered in future clinical studies. The suitability of the developed method to the analysis of clinical samples has been demonstrated through the analysis of a set of 137 faecal and 210 urine samples from a broad clinical cohort comprising adults, term and preterm infants. A significant correlation among SCFAs and BCAAs in faecal supernatants and urine samples was observed and a total of five metabolites showed a significant correlation across matrices. Furthermore, reference ranges of SCFAs and BCAAs in mothers, PI and TI were reported and different patterns were found among groups. Infants showed lower concentrations of SCFAs and BCAAs than mothers in faeces and higher concentrations in urine. Slightly higher concentration values in PI compared to TI were identified, except for leucine and isoleucine in faecal samples that were higher in TI. The presented results will serve as a knowledge base for future studies that will examine the interplay of human milk composition, microbiota, newborn physiology, and the study of serious gastrointestinal diseases that mostly affect preterm babies such as necrotizing enterocolitis.

CRediT authorship contribution statement

Victoria Ramos-Garcia: Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Isabel Ten-Doménech:** Methodology, Writing – review & editing. **Alba Moreno-Giménez:** Data curation, Investigation. **Laura Campos-Berga:** Data curation, Investigation. **Anna Parra-Llorca:** Data curation, Investigation. **Álvaro Solaz-García:** Data curation, Investigation. **Inmaculada Lara-Cantón:** Data curation, Investigation. **Alejandro Pinilla-González:** Data curation, Investigation. **María Gormaz:** Data curation, Supervision. **Máximo Vento:** Conceptualization, Resources, Writing – review & editing. **Julia Kuligowski:** Methodology, Conceptualization, Funding acquisition, Project administration, Investigation, Resources, Supervision, Visualization, Writing – review & editing. **Guillermo Quintás:** Conceptualization, Investigation, Resources, Supervision, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to parents and their infants who agreed to participate in this study. Technical support from the Central Support Service for Experimental Research (*SCSIE*) at the University of Valencia is acknowledged. I.T.-D., L.C.-B., A.P.-L., LL.-C., and J.K. received salary support by the *Instituto de Salud Carlos III* (Ministry of Science and Innovation, Spain) and co-funded by the European Union, grant numbers CD19/00176, CM20/00143, JR21/00055, CM20/00187, and CP16/00034, respectively; and A.P.G. from *Fundación para la Investigación del Hospital Universitario La Fe de Valencia* (Health Research Institute La Fe, Spain), grant number 2019-050-1_CRC. This work was also supported by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2022.05.005.

References

[1] E. Neis, C. Dejong, S. Rensen, The Role of Microbial Amino Acid Metabolism in Host Metabolism, Nutrients. 7 (2015) 2930–2946, https://doi.org/10.3390/nu7042930.

[2] C.L. Johnson, J. Versalovic, The Human Microbiome and Its Potential Importance to Pediatrics, Pediatrics 129 (2012) 950–960, https://doi.org/10.1542/peds.2011-2736.

[3] W.M. de Vos, H. Tilg, M. Van Hul, P.D. Cani, Gut microbiome and health: mechanistic insights, Gut. (2022) gutjnl-2021-326789. https://doi.org/10.1136/gutjnl-2021-326789.

[4] Y.M.K. Lei, L. Nuit, M.-L. Alegre, The Interplay between the Intestinal Microbiota and the Immune System, Clin. Res. Hepatol. Gastroenterol. 39 (2015) 9–19, https://doi.org/10.1016/j.clinre.2014.10.008.

[5] K.H. McLeod, J.L. Richards, Y.A. Yap, E. Mariño, Dietary Short Chain Fatty Acids: How the Gut Microbiota Fight Against Autoimmune and Inflammatory Diseases, in: Bioactive Food as Dietary Interventions for Arthritis and Related Inflammatory Diseases, Elsevier, 2019, pp. 139–159, https://doi.org/10.1016/B978-0-12-813820-5.00007-6.

[6] N.C. Roy, S.A. Bassett, W. Young, C. Thum, W.C. McNabb, Chapter 18 - The Importance of Microbiota and Host Interactions Throughout Life, in: M. Boland, M. Golding, H. Singh (Eds.), Food Structures, Digestion and Health, Academic Press, San Diego, 2014, pp. 489–511, https://doi.org/10.1016/B978-0-12-404610-8.00018-9.

[7] G. den Besten, K. van Eunen, A.K. Groen, K. Venema, D.-J. Reijngoud, B.M. Bakker, The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism, J. Lipid Res. 54 (2013) 2325–2340, https://doi.org/10.1194/jlr.R036012.

[8] M. Ziętek, Z. Celewicz, M. Szczuko, Short-Chain Fatty Acids, Maternal Microbiota and Metabolism in Pregnancy, Nutrients. 13 (2021) 1244, https://doi.org/10.3390/nu13041244.

[9] C.T. Pekmez, M.W. Larsson, M.V. Lind, N. Vazquez Manjarrez, C. Yonemitsu, A. Larnkjær, L. Bode, C. Mølgaard, K.F. Michaelsen, L.O. Dragsted, Breastmilk Lipids and Oligosaccharides Influence Branched Short-Chain Fatty Acid Concentrations in Infants with Excessive Weight Gain, Mol. Nutr. Food Res. 64 (2020) 1900977, https://doi.org/10.1002/mnfr.201900977.

[10] S. Zhang, X. Zeng, M. Ren, X. Mao, S. Qiao, Novel metabolic and physiological functions of branched chain amino acids: a review, J. Anim. Sci. Biotechnol. 8 (2017) 10, https://doi.org/10.1186/s40104-016-0139-z.

[11] R. García-Villalba, J.A. Giménez-Bastida, M.T. García-Conesa, F.A. Tomás-Barberán, J.C. Espín, M. Larrosa, Alternative method for gas chromatography-mass spectrometry analysis of short-chain fatty acids in faecal samples, J. Sep. Sci. 35 (2012) 1906–1913, https://doi.org/10.1002/jssc.201101121.

[12] H.M.H. van Eijk, J.G. Bloemen, C.H.C. Dejong, Application of liquid chromatography–mass spectrometry to measure short chain fatty acids in blood, J. Chromatogr. B 877 (2009) 719–724, https://doi.org/10.1016/j.jchromb.2009.01.039.

[13] J.C.Y. Chan, D.Y.Q. Kioh, G.C. Yap, B.W. Lee, E.C.Y. Chan, A novel LCMSMS method for quantitative measurement of short-chain fatty acids in human stool derivatized with 12C- and 13C-labelled aniline, J. Pharm. Biomed. Anal. 138 (2017) 43–53, https://doi.org/10.1016/j.jpba.2017.01.044.

[14] M. Zeng, H. Cao, Fast quantification of short chain fatty acids and ketone bodies by liquid chromatography-tandem mass spectrometry after facile derivatization coupled with liquid-liquid extraction, J. Chromatogr. B 1083 (2018) 137–145, https://doi.org/10.1016/j.jchromb.2018.02.040.

[15] H.E. Song, H.Y. Lee, S.J. Kim, S.H. Baek, H.J. Yoo, A Facile Profiling Method of Short Chain Fatty Acids Using Liquid Chromatography-Mass Spectrometry, Metabolites. 9 (2019) 173, https://doi.org/10.3390/metabo9090173.

[16] X. Gao, E. Pujos-Guillot, J.-F. Martin, P. Galan, C. Juste, W. Jia, J.-L. Sebedio, Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization, Anal. Biochem. 393 (2009) 163–175, https://doi.org/10.1016/j.ab.2009.06.036.

[17] H.F.N. Kvitvang, T. Andreassen, T. Adam, S.G. Villas-Bôas, P. Bruheim, Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids, Anal. Chem. 83 (2011) 2705–2711, https://doi.org/10.1021/ac103245b.

[18] X. Zheng, Y. Qiu, W. Zhong, S. Baxter, M. Su, Q. Li, G. Xie, B.M. Ore, S. Qiao, M. D. Spencer, S.H. Zeisel, Z. Zhou, A. Zhao, W. Jia, A targeted metabolomic protocol for short-chain fatty acids and branched-chain amino acids, Metabolomics 9 (2013) 818–827, https://doi.org/10.1007/s11306-013-0500-6.

[19] S. Zhang, H. Wang, M.-J. Zhu, A sensitive GC/MS detection method for analyzing microbial metabolites short chain fatty acids in fecal and serum samples, Talanta 196 (2019) 249–254, https://doi.org/10.1016/j.talanta.2018.12.049.

[20] R. Menon, Spontaneous preterm birth, a clinical dilemma: Etiologic, pathophysiologic and genetic heterogeneities and racial disparity, Acta Obstet. Gynecol. Scand. 87 (2008) 590–600, https://doi.org/10.1080/00016340802005126.

[21] K.E. Gregory, C.E. DeForge, K.M. Natale, M. Phillips, L.J. Van Marter, Necrotizing Enterocolitis in the Premature Infant, Adv Neonatal Care. 11 (2011) 155–166, https://doi.org/10.1097/ANC.0b013e31821baaf4.

[22] V. Apgar, A Proposal for a New Method of Evaluation of the Newborn Infant, Anesth. Analg. 120 (2015) 1056–1059, https://doi.org/10.1213/ANE.0b013e31629bde5c.

[23] Food and Drug Administration (FDA). Guidance for Industry: Bioanalytical Method Validation. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine, (2018) 44.

[24] National Collaborating Centre for Women's and Children's Health (UK), Urinary Tract Infection in Children: Diagnosis, Treatment and Long-term Management, RCOG Press, London, 2007. http://www.ncbi.nlm.nih.gov/books/NBK50606/ (accessed February 9, 2022).

[25] J. Kaufman, P. Fitzpatrick, S. Tosif, S.M. Hopper, S.M. Donath, P.A. Bryant, F. E. Babl, Faster clean catch urine collection (Quick-Wee method) from infants: randomised controlled trial, BMJ 357 (2017), j1341, https://doi.org/10.1136/bmj.j1341.

[26] M.L. Herreros Fernandez, N. Gonzalez Merino, A. Tagarro Garcia, B. Perez-Seoane, M. de la Serna Martinez, M.T. Contreras Abad, A. Garcia-Pose, A new technique for fast and safe collection of urine in newborns, Arch. Dis. Child. 98 (2013) 27–29, https://doi.org/10.1136/archdischild-2012-301872.

[27] N. Altuntas, A. Celebi Tayfur, M. Kocak, H.C. Razi, S. Akkurt, Midstream clean-catch urine collection in newborns: a randomized controlled study, Eur. J. Pediatr. 174 (2015) 577–582, https://doi.org/10.1007/s00431-014-2434-z.

[28] I. Alothaim, D.R. Gaya, D.G. Watson, Development of a Sensitive Liquid Chromatography Mass Spectrometry Method for the Analysis of Short Chain Fatty Acids in Urine from Patients with Ulcerative Colitis, Current, Metabolomics 6 (2) (2018) 124–130.

Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-023-04802-8

RESEARCH PAPER

Fast profiling of primary, secondary, conjugated, and sulfated bile acids in human urine and murine feces samples

Victoria Ramos-Garcia[1] · Isabel Ten-Doménech[1] · Máximo Vento[1,2] · Clara Bullich-Vilarrubias[3] · Marina Romaní-Pérez[3] · Yolanda Sanz[3] · Angelica Nobili[4] · Marika Falcone[4] · Marina Di Stefano[5] · Guillermo Quintás[6] · Julia Kuligowski[1]

Received: 12 May 2023 / Revised: 5 June 2023 / Accepted: 7 June 2023

Abstract
Bile acids (BAs) are a complex class of metabolites that have been described as specific biomarkers of gut microbiota activity. The development of analytical methods allowing the quantification of an ample spectrum of BAs in different biological matrices is needed to enable a wider implementation of BAs as complementary measures in studies investigating the functional role of the gut microbiota. This work presents results from the validation of a targeted ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) method for the determination of 28 BAs and six sulfated BAs, covering primary, secondary, and conjugated BAs. The analysis of 73 urine and 20 feces samples was used to test the applicability of the method. Concentrations of BAs in human urine and murine feces were reported, ranging from 0.5 to 50 nmol/g creatinine and from 0.012 to 332 nmol/g, respectively. Seventy-nine percent of BAs present in human urine samples corresponded to secondary conjugated BAs, while 69% of BAs present in murine feces corresponded to primary conjugated BAs. Glycocholic acid sulfate (GCA-S) was the most abundant BA in human urine samples, while taurolithocholic acid was the lowest concentrated compound detected. In murine feces, the most abundant BAs were α-murocholic, deoxycholic, dehydrocholic, and β-murocholic acids, while GCA-S was the lowest concentrated BA. The presented method is a non-invasive approach for the simultaneous assessment of BAs and sulfated BAs in urine and feces samples, and the results will serve as a knowledge base for future translational studies focusing on the role of the microbiota in health.

Keywords Bile acids · Human urine · Murine feces · Microbiota

Introduction

Bile acids (BAs) are a group of molecules from the family of steroids with a hydrophobic core structure of four fused carbon rings with hydroxyl and methyl groups and a carbon side chain ending in a carboxylic acid group. Primary BAs (e.g., cholic (CA) and chenodeoxycholic acids (CDCA)) are produced from cholesterol in the liver through a series of enzymatic reactions and are modified thereafter through the addition of glycine or taurine to form conjugated BAs (e.g., glycocholic (GCA) and taurocholic (TCA) acids) before being released to the bile. BAs are then secreted into the intestine through the gall bladder. During their transit through the intestine, BAs undergo modifications due to the action of gut microbiota (e.g., deconjugation, dehydroxylation, hydroxyl group oxidation, and epimerization), producing secondary BAs (e.g., lithocholic (LCA) and deoxycholic acids (DCA)), making BAs more hydrophobic and,

✉ Julia Kuligowski
julia.kuligowski@uv.es

[1] Neonatal Research Unit, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain

[2] Division of Neonatology, University & Polytechnic Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain

[3] Institute of Agrochemistry and Food Technology, Spanish National Research Council (IATA-CSIC), Valencia, Spain

[4] Autoimmune Pathogenesis Unit, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

[5] Maternal and Child Health Area, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

[6] Health and Biomedicine, Leitat Technological Center, Carrer de la Innovació, 2, 08225 Terrassa, Spain

Published online: 20 June 2023

consequently, less dangerous to bacteria [1]. Both primary and secondary BAs are reabsorbed in the terminal ileum and large intestine before being transported back to the liver, where they are recycled [2]. This process, known as enterohepatic circulation, is important for the efficient use of BAs as it helps to maintain a balance of these substances in the body and plays an important role in various physiological processes, such as emulsifying ingested lipids [3]. Additionally, BAs can be found in small amounts distributed across other parts of the body, as they can function as signaling molecules in other physiological processes apart from the digestive system, such as cardiac system or renal water regulation [4, 5].

The BAs metabolism is known to differ between species (e.g., humans and murine animals) [6]. In humans, the classical pathway for BA metabolism (80%) is by the downstream metabolism of CA and DCA, while in murine animals, BAs metabolism is dominated by the production of α-muricholic (α-MCA) and β-muricholic acid (β-MCA) from CDCA, LCA, and ursodeoxycholic acid (UDCA) [7]. Only a small portion of BAs, approximately 5% of the total pool, is not reabsorbed but excreted [8]. BAs excreted though feces have been reported to be mostly present in the unconjugated form. However, urine is the primary route for BA elimination, and BAs not recycled from the liver or renal cells are excreted through it in the main detoxification pathways (i.e., glucuronidation and sulfation) [9–11], mostly in the conjugated and/or sulfated form [12, 13].

BAs and gut microbiota are considered a dynamic equilibrium, as both have a clear impact on one another. Changes in BA levels or types in biofluids may be indicative of certain health conditions or response to treatment. In addition, the measurement of BAs can help to identify potential biomarkers for diseases such as liver disease, gastrointestinal disorders, and metabolic disorders [14]. Overall, the analysis of BAs provides valuable information for clinical monitoring.

There are various methods that are currently used to identify and quantify BAs in biological samples [15]. Liquid (LC) or gas (GC) chromatography coupled to mass spectrometry (MS) have been the most commonly used techniques for the analysis of BAs in human and animal biofluids in recent years [16–24]. Despite GC–MS providing a very high peak resolution, its main disadvantage is the time-consuming sample preparation that requires group fractionation, enzymatic or chemical hydrolysis of conjugates, and preparation of volatile methyl ester trimethylsilyl ether derivatives [15]. In LC–MS, BAs are typically resolved using reversed-phase chromatography columns and gradient elution programs. LC–MS offers excellent sensitivity and specificity, which may reduce the need for extensive sample preparation and ultimately improve analytical throughput. However, the majority of the studies regarding BA analysis by LC–MS involve solid phase extraction and/or preconcentration steps. Furthermore, previously reported LC–MS methods for BA quantification do not include sulfated BAs [17–19].

The aim of this work was to develop a straightforward ultra-high performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) method for the determination of 28 BAs and six sulfated BAs, covering a wide range of primary, secondary, conjugated, and sulfated BAs. The applicability of the method in the clinical field and in animal models was evaluated by the analysis of 73 human urine samples and 20 murine feces samples.

Material and methods

Standards and reagents

LC–MS grade methanol, propanol, acetic acid, ammonium acetate, and acetonitrile were obtained from Merck Life Science S.L.U. (Madrid, Spain). Bile acid standards (CA, CDCA, LCA, DCA, GCA, TCA, α-MCA, β-MCA, UDCA, hyodeoxycholic (HDCA), murocholic (MCA), 3,7,12-dehydrocholic (DHCA), hyocholic (HCA), glycolithocholic (GLCA), glycoursodeoxycholic (GUDCA), glycohyodeoxycholic (GHDCA), glycochenodeoxycholic (GCDCA), glycodeoxycholic (GDCA), 3,7,12-glycodehydrocholic (GDHCA), glycohyocholic (GHCA), taurolithocholic (TLCA), tauroursodeoxycholic (TUDCA), taurohyodeoxycholic (THDCA), taurochenodeoxycholic (TCDCA), taurodeoxycholic (TDCA), 3,7,12-taurodehydrocholic (TDHCA), taurohyocholic (THCA), tauro-α-muricholic (T-α-MCA) acids, and deuterated internal standards (IS) (lithocholic-D_4 (LCA-D_4), cholic-D_4 (CA-D_4), glycochenodeoxycholic-D_4 (GCDCA-D_4), glycocholic-D_4 (GCA-D_4) acids) were purchased from Steraloids (Newport, RI, USA). Sulfated BAs (lithocholic (LCA-S), chenodeoxycholic (CDCA-S), ursodeoxycholic (UDCA-S), deoxycholic (DCA-S), cholic (CA-S), glycocholic (GCA-S) acid sulfates), and lithocholic-D_4 acid sulfate (LCA-S-D4) as IS were obtained from Qmx Laboratories Ltd. (Essex, UK). Individual 5 mM standard solutions were prepared in methanol.

Study design, population, and sample collection

Urine samples were collected from 73 children, i.e., 44 male and 29 female individuals, aged 13 ± 3 years (mean ± standard deviation) with a range between 8 and 18 years recruited at the Pediatric Department of the San Raffaele Hospital in Milan (Italy). During the enrolment phase, pediatricians informed the parents about the purpose of this study, the lack of reported risks related to the collection of samples, the effort required to take part in this study, and their right to deny or withdraw their consent at any time. Parents accepted to participate in the study and signed an informed consent

form. The study was approved by the Institutional Ethical Committee of the IRCCS San Raffaele Scientific Institute (Protocol: NUTRI-T1D, 2019). Individual identifiable private information was protected according to the EU General Data Protection Regulation (EU-GDPR) with the supervision of the Institutional Data Protection Officer. The Clinical Research Investigator assigned a code to each patient and identifiers that link to protected information. Each subject collected urine samples at home and kept them at 4 °C until transported to the hospital on the same day. Sterile plastic containers were used for sample collection and a minimum of 10 mL were required. Samples were aliquoted, labeled, and stored at − 80 °C until analysis.

Feces samples were collected from CC57BL/6 J male mice ($n = 20$, 7 weeks old) provided by Charles River Laboratories (Wilmington, USA). Mice were collectively housed ($n = 5$/cage) and maintained under constant conditions of humidity and temperature (i.e., 23 ± 2 °C) and regular 12-h light–dark cycles. Mice fed with control diet (CD, D12450K, Research Diets, New Brunswick, NJ) were fasted overnight to collect feces 2 h after the administration of a mixed solution of lipids (intralipid, 0.06% per mouse, Merck KGaA) by oral gavage. Collected samples were immediately frozen in liquid nitrogen until further analysis. This experimental procedure using animals was in accordance with European Union 2010/63/UE and Spanish RD53/2013 guidelines and approved by the ethics committee of the University of Valencia (Animal Production Section, SCSIE, University of Valencia, Spain) and authorized by *Dirección General de Agricultura, Ganadería y Pesca (Generalitat Valenciana)* (approval ID 2021/VSC/PEA/0273).

UHPLC-MS/MS determination of BAs

The UHPLC-MS/MS method for the quantification of 34 BAs was developed based on previously reported method [21] with modifications. For feces, 1 mL of ultrapure water was added to 100 mg of wet samples, and then the mixture was homogenized for 10 min and centrifuged at 13,200 × *g* for 20 min at room temperature (RT). Urine samples were centrifuged at 3500 × *g* for 15 min at RT. Fifty microliter of fecal or urinary supernatants was mixed with 10 µL of IS solution containing LCA-D_4, CA-D_4, GCDCA-D_4, GCA-D_4, and LCA-S-D_4 at 2.1 µM each, and 150 µL of ice-cold methanol was added for protein precipitation. The mixture was then agitated for 15 min at 4 °C, incubated for 20 min at − 20 °C, and centrifuged at 3500 × *g* for 15 min at 4 °C. Urine samples were diluted 1:20 (v/v) with ultrapure H_2O prior to analysis. Supernatants were transferred to a 96-well plate for analysis.

UHPLC-MS/MS analysis was conducted using an ACQUITY UPLC instrument (Waters Ltd, Elstree, UK) coupled to a Xevo TQ-S MS (Waters, Manchester, UK) operating in negative electrospray ionization mode (ESI^-). Chromatographic separation was performed using an ACQUITY BEH C8 column (100 mm × 2.1 mm, 1.7 µm) (Waters) at an operating temperature of 60 °C and a flow rate of 0.6 mL min^{-1}. Mobile phase component A consisted of 100 mL of acetonitrile added to 1 mM ammonium acetate in 1 L ultrapure water, adjusting the pH to 4.15 with acetic acid. Mobile phase component B consisted of an acetonitrile:propanol (1:1, v/v) solution. The gradient separation was performed as follows: 0–0.1 min, 10% B, 0.1–9.25 min, 10–35% B; 9.25–11.5 min, 35–85% B; 11.5–11.8 min, 85–100% B; 11.8–14.3 min, 100–10% B; and 14.3–16 min, 10% B. The injection volume was 5 µL. MS/MS detection was performed using the multiple reaction monitoring (MRM) mode and the following ionization conditions: ion spray voltage, 1.5 kV; cone voltage, 60 V; source temperature, 150 °C; desolvation temperature, 600 °C; desolvation gas flow, 1000 L/h; and cone gas flow, 150 L/h. The MRM transitions for each BA and IS summarized in Table 1 were optimized by infusing individual solutions of commercial standards.

Standard solutions were prepared by serial dilution of a working solution containing mixtures of pure analytical standards in methanol. For monitoring instrument performance, aliquots of quality control (QC) samples obtained by pooling 5 µL of each urine or fecal supernatant, respectively, were analyzed after every 20 samples in each analytical batch. The batch acceptance criterion for each analyte was a QC RSD < 25%. In addition, calibration (i.e., ultrapure H_2O) and process blanks (generated replacing the sample by H_2O) were included in each batch to assess method specificity and the lack of carry over and cross-contamination.

Fecal BA concentrations were normalized to sample weight, while urinary BA concentrations were normalized to creatinine concentrations quantified following the manufacturer's instructions of the modified Jaffe's method implemented in the DetectX® urinary creatinine detection kit from Arbor Assays (Ann Arbor, MI, USA).

Method validation

The method validation included the determination of linearity range, accuracy, precision, selectivity, specificity, carry over, and analyte stability, based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation [25] and the International Council for Harmonisation (ICH) M10 on bioanalytical method validation — scientific guideline [26]. The range of concentrations employed ranged from 1.2 nM to 1.25 µM, selected to accommodate for the anticipated values in real samples and a significant inter- and intra-individual variability, thereby enabling precise, accurate, and linear quantification of the metabolites. At least six standards covering the concentrations ranges, as well as a process blank, were included in the calibration curves.

Table 1 Measurement parameters and main figures of merit of the LC–MS method for BA determination

Class	Metabolite	MRM transition	CE (V)	Dwell (s)	RT ± s (min)	R^2	Urine samples (nM)		Feces samples (nM)	
							LOD	LOQ	LOD	LOQ
Primary	MCA	391.3→391.3	10	0.03	9.489 ± 0.002	0.997	0.5	1.7	0.8	3
	CDCA	391.3→391.3	10	0.03	10.760 ± 0.014	0.997	0.5	1.7	0.6	2
	CA	407.3→407.3	10	0.02	10.155 ± 0.002	0.998	0.2	0.7	0.3	1.1
	α-MCA	407.3→407.3	10	0.02	8.782 ± 0.015	0.997	0.2	0.7	0.5	1.6
	β-MCA	407.3→407.3	10	0.02	8.883 ± 0.012	0.997	0.3	1.0	0.2	0.7
	HCA	407.3→407.3	10	0.02	9.857 ± 0.014	0.998	0.3	1.0	0.3	0.8
Primary conjugated	GCDCA	448.4→74.0	40	0.02	9.347 ± 0.014	0.996	1.0	3	0.002	0.012
	GHCA	464.4→74.0	40	0.02	6.820 ± 0.002	0.996	0.2	0.7	0.13	0.4
	GCA	464.4→74.0	40	0.02	7.573 ± 0.012	0.996	0.3	1.0	0.013	0.04
	TCDCA	498.4→80.0	60	0.02	8.721 ± 0.004	0.9993	0.5	1.7	0.02	0.06
	TCA	514.3→80.0	60	0.02	7.047 ± 0.013	0.998	0.2	0.7	0.02	0.07
	T-α-MCA	514.3→80.0	60	0.02	5.238 ± 0.002	0.998	0.2	0.7	0.03	0.08
	THCA	514.3→80.0	60	0.02	6.308 ± 0.014	0.9992	0.2	0.7	0.03	0.10
Secondary	LCA	375.3→375.3	10	0.03	11.140 ± 0.002	0.998	1.0	3	0.6	1.8
	DCA	391.3→391.3	10	0.03	10.812 ± 0.003	0.998	1.0	3	0.4	1.3
	UDCA	391.3→391.3	10	0.03	10.060 ± 0.012	0.996	0.2	0.7	0.3	1.1
	HDCA	391.3→391.3	10	0.03	10.283 ± 0.012	0.998	0.2	0.7	0.2	0.8
	DHCA	468.4→74.0	10	0.02	5.440 ± 0.003	0.996	0.2	0.7	0.3	0.8
Secondary conjugated	GLCA	432.4→74.0	40	0.03	10.444 ± 0.003	0.9992	1.0	3	0.010	0.03
	GDCA	448.4→74.0	40	0.02	9.772 ± 0.002	0.998	0.3	1.0	0.02	0.03
	GUDCA	448.4→74.0	40	0.02	7.122 ± 0.016	0.9992	0.2	0.7	0.003	0.013
	GHDCA	448.4→74.0	40	0.02	7.490 ± 0.004	0.998	1.0	3	0.013	0.04
	GDHCA	458.3→74.0	40	0.04	2.831 ± 0.002	0.997	0.2	0.7	0.02	0.06
	TLCA	466.3→80.0	40	0.03	10.220 ± 0.003	0.998	0.2	0.7	0.02	0.07
	TDCA	498.4→80.0	60	0.02	9.123 ± 0.016	0.9995	0.3	1.0	0.03	0.08
	TUDCA	498.4→80.0	60	0.02	6.640 ± 0.015	0.998	0.5	1.7	0.13	0.4
	THDCA	498.4→80.0	60	0.02	6.945 ± 0.002	0.998	0.3	1.0	0.04	0.13
	TDHCA	508.3→80.0	60	0.08	2.464 ± 0.013	0.997	0.2	0.7	0.010	0.03
Sulfated	GCA-S	566.3→97.0	40	0.02	6.931 ± 0.014	0.996	0.3	1.0	0.005	0.02
	UDCA-S	471.2→97.0	40	0.02	7.320 ± 0.012	0.996	0.2	0.7	0.008	0.03
	LCA-S	455.2→97.0	40	0.03	10.320 ± 0.003	0.997	0.2	0.7	0.006	0.02
	DCA-S	471.3→97.0	40	0.02	9.288 ± 0.002	0.998	0.2	0.7	0.08	0.3
	CDCA-S	471.2→97.0	40	0.02	9.488 ± 0.014	0.998	0.2	0.7	0.13	0.4
	CA-S	487.2→97.0	40	0.02	7.571 ± 0.012	0.996	0.3	1.0	0.4	1.4
IS	LCA-d4	379.6→97.0	10	0.03	11.143 ± 0.002	-	-	-	-	-
	CA-d4	411.6→97.0	10	0.03	10.151 ± 0.002	-	-	-	-	-
	GCDCA-d4	451.5→74.0	40	0.03	9.330 ± 0.002	-	-	-	-	-
	GCA-d4	468.6→74.0	40	0.02	7.572 ± 0.016	-	-	-	-	-
	LCA-S-d4	459.2→97.0	10	0.03	10.312 ± 0.013	-	-	-	-	-

IS internal standard, *CE* collision energy, *RT* retention time; *R* coefficient of determination. Limit of quantification (LOQ) = concentration of analyte that can be measured with an imprecision of less than 20% and a deviation from target of less than 20% and taking into account the pre-concentration factor achieved during sample processing; limit of detection (LOD) = 3/10*LOQ

Accuracy and precision in standards were evaluated, respectively, by comparing absolute concentrations values and calculating the relative standard deviation (%RSD) of replicates of standards within a single validation batch (intra-day) and between different validation batches (inter-day). To assess the accuracy and precision in the sample matrix, three replicates of both standard and spiked samples at three different concentrations (i.e., low, medium, and high) were analyzed

on three separate validation days, and recoveries as well as %RSD were calculated. The analysis of calibration and process blank samples from six different sources was used to assess selectivity and specificity of the analytical method. Additionally, the analysis of a process blank after the injection of the high concentrated standard mixture was used for evaluating carry-over and cross contamination between consecutive samples. The freeze–thaw and long-term stability of the analytes were assessed by comparing the concentrations in a sample that had undergone three freeze–thaw cycles and a sample that had been stored for 6 months at − 80 °C, respectively, to a freshly prepared sample.

Data availability and statistical analysis

UHPLC-MS/MS data were acquired and processed using MassLynx software version 4.0 (Waters Ltd, Elstree, UK). Data analysis was carried out in MATLAB R2018b (MathWorks, Natick, MA, USA). Biomarker normalized levels in urine and feces samples determined in this work are available in Tables S1 and S2. Continuous variables were expressed as mean ± standard deviation or medians with interquartile (IQR) ranges.

Results and discussion

Method validation

The chromatographic peaks of the different BAs are depicted in Fig. 1. Primary, secondary, and sulfated BAs presented different chromatographic retention clusters. Primary BAs RTs ranged between 8.7 and 10.8 min, and secondary BAs eluted between 10.1 and 11.2 min, except DHCA at 5.4 min. Sulfated BAs eluted in the 6.9–10.3 min range. Due to their structural and physicochemical heterogeneity,

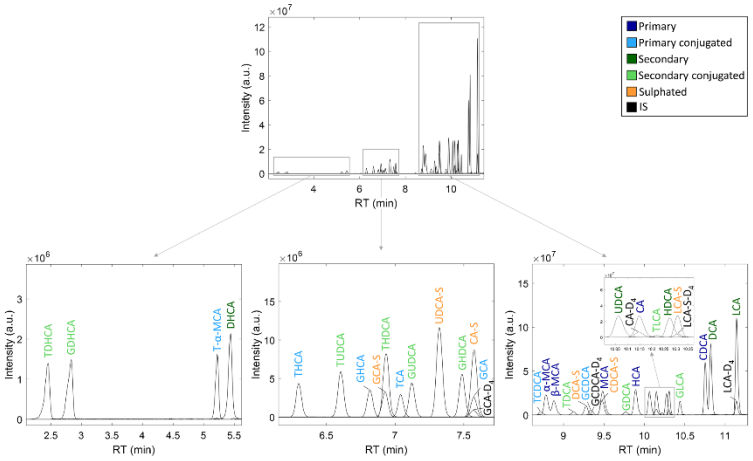


Fig. 1 Separation of BAs in the developed UPLC-MS/MS method. Note: *TDHCA*, 3,7,12-taurodehydrocholic acid; **GDHCA**, 3,7,12-glycodehydrocholic acid; **T-α-MCA**, tauro-α-muricholic acid; **DHCA**, 3,7,12-dehydrocholic acid; **THCA**, taurohyocholic acid; **TUDCA**, tauroursodeoxycholic acid; **GHCA**, glycohyocholic acid; **GCA-S**, glycocholic acid sulfate; **THDCA**, taurohyodeoxycholic acid; **TCA**, taurocholic acid; **GUDCA**, glycoursodeoxycholic acid; **UDCA-S**, ursodeoxycholic acid sulfate; **GHDCA**, glycohyodeoxycholic acid; **CA-S**, cholic acid sulfate; **GCA**, glycocholic acid; **GCA-D_4**, deuterated glycocholic acid; **TCDCA**, taurochenodeoxycholic acid; **α-MCA**, α-muricholic acid; **β-MCA**, β-muricholic acid; **TDCA**, taurodeoxycholic acid; **DCA-S**, deoxycholic acid sulfate; **GCDCA**, glycochenodeoxycholic acid; **GCDCA-D_4**, deuterated glycochenodeoxycholic acid; **MCA**, muricholic acid; **CDCA-S**, chenodeoxycholic acid sulfate; **GDCA**, glycodeoxycholic acid; **HCA**, hyocholic acid; **UDCA**, ursodeoxycholic acid; **CA**, cholic acid; **CA-D_4**, deuterated cholic acid; **TLCA**, taurolithocholic acid; **HDCA**, hyodeoxycholic acid; **LCA-S**, lithocholic acid sulfate; **LCA-S-D_4**, deuterated lithocholic acid sulfate; **GLCA**, glycolithocholic acid; **CDCA**, chenodeoxycholic acid; **DCA**, deoxycholic acid; **LCA**, lithocholic acid; **LCA-D_4**, deuterated lithocholic acid

Table 2 Calculated intra- and inter- day accuracies (i.e., recovery) and precisions (i.e., %RSD) of the LC–MS method in standard solutions and spiked urine and feces samples

		Accuracy ± RSD																	
		Standard solutions						Spiked urine samples						Spiked feces samples					
		Intra day (N = 5)			Inter day (N = 3)			Intra day (N = 5)			Inter day (N = 3)			Intra day (N = 5)			Inter day (N = 3)		
Class	Metabolite	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Primary	MCA	87±13	109±2	114±9	102±12	100±12	112±2	109±15	90±2	107±12	111±5	97±7	106±5	109±5	93±15	112±2	104±15	101±11	107±10
	CDCA	89±11	97±5	103±6	100±11	93±6	105±3	101±6	92±12	102±9	110±7	105±15	103±1	98±1	95±15	106±3	96±3	93±1	103±4
	CA	105±9	101±11	115±7	107±15	102±10	108±6	112±8	92±7	102±5	114±2	98±5	102±7	118±18	85±3	96±12	100±16	93±8	98±8
	α-MCA	110±11	101±15	104±11	111±6	91±10	107±7	110±10	94±14	107±2	107±5	105±3	106±1	96±15	104±5	86±5	95±4	103±5	98±15
	β-MCA	84±13	111±6	87±9	93±15	107±13	98±14	85±18	86±14	98±9	98±18	101±13	103±5	81±15	105±7	94±5	96±13	106±2	97±3
	HCA	113±17	87±5	109±13	111±7	99±13	106±11	85±3	85±13	104±5	94±16	105±17	100±12	117±4	107±7	92±5	99±18	94±13	92±1
Primary conjugated	GCDCA	112±8	96±12	92±12	98±13	91±4	94±5	86±12	114±7	114±12	91±5	104±15	106±10	102±18	92±9	105±12	95±7	97±7	100±5
	GHCA	117±8	110±9	102±8	102±16	108±5	108±4	112±4	90±12	85±5	98±13	91±2	95±11	104±15	110±9	96±11	101±7	100±12	98±9
	GCA	111±5	99±13	115±1	104±5	98±1	106±9	113±14	98±14	106±3	110±7	100±3	110±5	115±5	101±12	113±8	101±24	102±2	98±13
	TCDCA	114±19	90±7	108±11	102±12	102±11	99±10	113±7	88±2	99±5	98±15	100±21	102±4	117±7	96±2	107±10	101±23	95±1	100±5
	TCA	93±3	93±3	108±5	109±11	95±5	99±3	93±11	85±15	96±5	96±10	100±15	97±11	108±14	95±5	111±7	104±8	99±5	105±8
	T-α-MCA	108±3	94±5	111±4	107±4	100±10	107±5	89±9	88±9	89±10	102±15	95±8	98±13	80±14	90±15	83±13	100±10	100±11	94±5
	THCA	104±12	115±7	117±7	101±12	100±14	107±5	95±14	95±9	96±8	102±10	101±11	97±10	100±4	98±2	105±11	98±7	99±14	104±12
Secondary	LCA	101±5	89±13	105±12	90±13	99±10	96±13	101±19	113±6	103±8	103±7	112±9	106±3	87±10	95±10	94±9	103±14	105±9	103±11
	DCA	106±14	96±5	102±7	104±13	92±4	103±11	92±6	91±10	113±11	99±13	102±12	106±11	95±12	111±11	103±4	105±11	105±10	109±5
	UDCA	105±4	114±15	107±2	107±5	95±15	102±5	100±4	109±12	94±1	106±5	107±6	103±8	100±10	95±4	101±6	109±9	103±9	105±7
	HDCA	112±4	100±5	104±10	109±11	95±8	98±6	112±14	112±12	100±10	101±11	113±5	104±6	115±17	105±15	111±8	107±9	100±15	107±7
	DHCA	91±8	90±12	101±6	99±14	93±10	109±7	110±8	105±9	109±8	111±9	101±14	105±6	105±12	105±4	114±9	107±13	104±3	102±12
Secondary conjugated	GLCA	112±9	100±9	109±10	101±10	99±12	105±6	98±10	93±2	104±10	105±9	101±9	102±2	101±13	93±15	108±12	104±15	106±11	106±7
	GDCA	109±9	106±12	87±3	96±12	101±10	109±14	105±3	102±7	97±6	105±14	105±5	100±12	100±15	97±4	113±6	102±15	100±11	106±10
	GUDCA	104±7	94±3	107±9	101±6	104±10	103±6	97±13	101±3	105±13	111±12	103±11	108±5	95±9	85±2	104±5	105±10	115±15	109±5
	GHDCA	115±5	96±9	114±6	105±10	95±3	107±9	95±2	104±12	101±10	98±10	105±11	103±7	108±18	114±13	104±4	110±3	110±5	108±4
	GDHCA	84±4	85±9	104±11	98±13	85±2	106±9	109±5	100±10	112±12	108±14	106±3	101±11	115±7	95±5	111±8	107±12	99±10	107±15
	TLCA	87±8	100±4	99±9	95±11	95±6	101±17	99±12	90±4	107±5	101±17	96±7	100±8	106±19	103±2	113±7	106±12	106±5	105±14
	TDCA	119±9	86±5	105±11	103±15	90±5	105±6	105±11	108±2	95±11	104±15	100±10	100±7	96±5	96±3	112±7	109±12	103±7	106±10
	TUDCA	87±13	114±3	110±7	94±7	108±7	104±7	94±5	94±2	110±8	105±11	95±7	109±4	102±14	104±11	111±8	107±5	102±15	105±5
	THDCA	90±4	105±7	97±5	95±10	100±14	102±5	109±18	104±15	113±1	110±7	101±9	112±1	112±11	109±8	112±10	104±13	106±4	107±8
	TDHCA	95±16	94±11	87±2	93±8	99±4	88±3	98±3	97±14	115±4	104±10	105±5	106±15	117±17	105±8	113±2	113±9	100±11	103±12
Sulfated	GCA-S	111±13	105±7	106±3	113±7	103±5	96±10	102±16	113±1	103±4	101±5	103±9	109±5	98±7	99±3	89±16	105±7	101±4	102±13
	UDCA-S	91±12	95±10	101±1	95±3	117±8	97±8	86±5	107±9	101±4	101±12	107±3	105±6	91±6	111±12	111±12	100±11	105±5	107±4
	LCA-S	104±19	108±10	89±3	108±4	100±7	91±2	96±2	111±10	104±8	102±13	104±14	105±4	109±7	91±7	103±9	104±6	100±9	104±8
	DCA-S	107±15	112±10	114±10	94±12	104±3	110±4	91±16	89±6	105±4	100±8	101±11	106±3	90±16	105±5	101±5	96±11	99±6	104±10
	CDCA-S	120±4	87±15	102±4	111±9	95±16	106±8	81±11	103±6	107±2	91±10	98±11	105±9	109±19	107±9	104±4	107±6	108±1	103±11
	CA-S	112±10	111±4	95±1	104±15	101±12	94±2	115±4	94±1	92±5	109±12	102±7	106±13	102±6	95±11	95±1	95±6	100±11	99±4

Table 3 BA concentrations in human urine and murine feces samples, respectively

Class	Metabolite	Urine ($N=73$), nmol/g creatinine				Feces ($N=20$), nmol/g feces			
		Range	Median	IQR (25–75)	Detection frequency (%)	Range	Median	IQR (25–75)	Detection frequency (%)
Primary	MCA	0.8–40	1.8	1.3	57	2–26	5	8	100
	CDCA	0.8–30	2	2	41	4–72	17	19	100
	CA	1.2–500	14	49	99	5–376	38	78	100
	α-MCA	0.3–13	1.2	2	81	134–791	332	290	100
	β-MCA	0.4–20	3	5	77	35–610	103	207	100
	HCA	0.6–20	1.7	1.2	18	0.6–162	3	4	95
Primary conjugated	GCDCA	2–30	4	6	70	0.0005–1.3	0.06	0.19	100
	GHCA	0.7–20	3	3	82	0.3–2	0.4	0.6	60
	GCA	0.8–16	5	5	82	0.003–12	0.2	0.4	100
	TCDCA	0.6–80	2	2	43	0.004–1.6	0.05	0.2	80
	TCA	0.3–13	0.8	0.7	77	0.005–6	0.17	0.5	90
	T-α-MCA	0.4–70	1.8	3	88	0.006–26	0.19	4	90
	THCA	0.3–20	0.7	0.7	28	0.007–1.6	0.02	0.14	55
Secondary	LCA	1.9–30	4	1.8	100	1.3–14	4	3	100
	DCA	2–40	5	6	49	35–461	182	120	100
	UDCA	0.3–14	0.7	0.6	50	35–244	66	52	100
	HDCA	0.3–12	0.7	0.7	53	21–126	61	46	100
	DHCA	0.4–30	1.3	1.2	72	19–644	117	301	95
Secondary conjugated	GLCA	3–3	3	0	1.4	0.002–1.12	0.02	0.16	80
	GDCA	0.6–20	2	3	68	0.02–3	0.12	0.17	100
	GUDCA	0.7–45,000	4	11	73	0.0006–12	0.07	0.18	100
	GHDCA	1.2–340	7	15	11	0.003–1.1	0.02	0.16	75
	GDHCA	0.3–9	0.7	0.9	9	0.004–1	0.02	0.10	80
	TLCA	0.5–0.5	0.7	0	1.4	<LOD	<LOD	<LOD	0
	TDCA	1.2–20	1.5	12	4	0.006–1.09	0.05	0.2	80
	TUDCA	1.2–20	2	1.2	8	0.3–2	1.0	2	60
	THDCA	0.8–30	2	9	7	0.09–1.5	0.2	0.2	100
	TDHCA	0.3–12	0.8	3	7	0.002–1.14	0.13	0.06	45
Sulfated	GCA-S	0.5–200	50	50	100	0.013–1.05	0.012	0.07	45
	UDCA-S	0.3–20	4	5	92	0.02–1.4	0.8	1.0	90
	LCA-S	0.3–6	0.7	0.5	54	0.014–1.09	0.03	0.2	70
	DCA-S	0.3–600	1.4	4	81	0.2–23	5	6	100
	CDCA-S	0.3–30	1.6	3	92	0.3–22	5	6	100
	CA-S	0.7–5	3	5	3	1–46	11	13	100

IQR interquartile range, *LOD* limit of detection

primary and secondary conjugated BAs eluted in a wider time ranging between 2.4 and 10.4 min. Results depicted in Fig. 1 indicated an adequate chromatographic and/or spectral resolution among BAs that enabled their simultaneous quantification.

Analytical method validation was performed following recommended FDA guidelines for bioanalytical method validation [25]. The evaluation of the linearity of response was carried out by covering up to four orders of magnitude with LODs and LOQs in the 0.2–3 and 0.002–3 nM range in urine and feces, respectively (see Table 1). As shown in Table 2, appropriate accuracies with recoveries between 84 and 120% and precisions ranging between 1 and 19%RSD at the three levels were observed in standard solutions, except for the UDCA-S and CDCA-S inter-day accuracy and precision at the medium concentration level with a recovery of 117% and an RSD of 16%, respectively. In spiked urine and feces samples, recoveries between 80 and 119%, 85 and 115%, and 85 and 115% and precisions between 2 and 19%, 1 and 17%, and 1 and 16% at low, medium, and high levels were found,

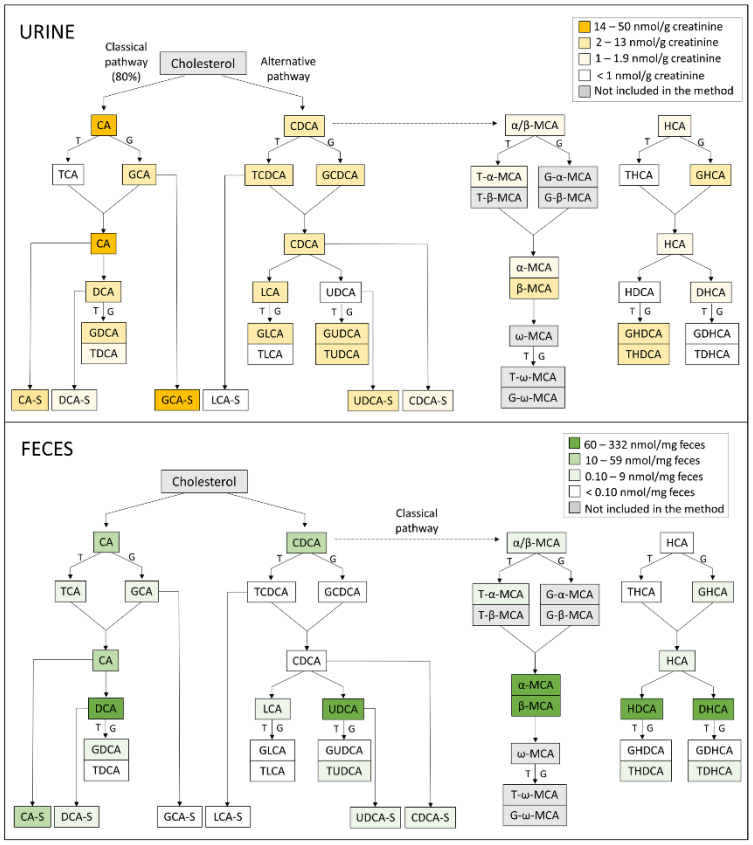


Fig. 2 BA pathways and concentration ranges in human urine and murine feces samples. Note: T, taurine; G, glycine

respectively, except for the inter-day accuracy of the medium concentration level of HDCA (118%) in urine and the precision of HCA (17%, medium level, urine), GCA (24%, low level, feces), and TCDCA (23%, low level, feces). After six months at − 80 °C and after undergoing three freeze–thaw cycles (− 80 °C), all compounds remained stable in samples (*t*-test, *p*-values > 0.05). The compatibility of the sample collection procedure, as well as the selectivity, specificity, and carry over were assessed, and no interfering contamination was detected (< 1/10 of the signal at LOQ). The analysis of

blanks and process blanks confirmed the lack of cross-contamination or carry over in the assessed chromatographic conditions. The presented data show adequate analytical method performance for all compounds in spiked urine samples and fecal supernatants, supporting the suitability of the method for the analysis of BAs in the two considered matrices.

BAs in urine and fecal samples

In order to test the applicability of the validated UHPLC-MS/MS method, a set of 73 human urine samples and 20 murine fecal samples were analyzed. Table 3 summarizes the results obtained from the quantification of BAs in human urine and murine fecal samples, after normalization to creatinine concentration and sample weight, respectively. Results show that GCA-S and CA are the most abundant BAs in human urine samples, while TLCA, THCA, UDCA, HDCA, GDHCA, and LCA-S were the lowest concentrated compounds detected. In murine feces samples, the most abundant metabolites were the murine-specific BAs α-MCA and β-MCA and the secondary BAs DCA and DHCA, while the lowest concentrated BAs were TLCA and GCA-S. Twenty and 31 BAs were detected in more than 50% of the urine and feces samples, respectively, being TLCA the less frequently detected in both matrices and GLCA in urine.

Differences in concentration profiles of BAs in human urine and murine feces samples were observed.

Figure 2 depicts the BA metabolism pathways in conjunction with BA concentration ranges found in the studied samples according to species and matrix. It shows that the most concentrated metabolites in human urine belong to the CA metabolism (the classical pathway for BA metabolism in humans) and that the most concentrated metabolites in murine feces belong to the CDCA and UDCA metabolism (the classical pathway in murine species). These findings are in agreement with Lin et al. results, where CA metabolism prevails over CDCA metabolism in humans, whereas CDCA metabolism is significantly enhanced in murine species [7]. Additionally, most of BAs present in human urine samples correspond to secondary conjugated BAs (e.g., GLCA and TUDCA), followed by sulfated BAs (e.g., GCA-S and CA-S), being those classes the less concentrated in murine feces samples. These results are in concordance with the literature [15], where most concentrated urinary BAs were conjugated and sulfated, confirming the urinary BA excretion mechanisms in humans. Thus, differences in BA profiles observed between human urine and murine feces samples demonstrate the different BA excretion mechanisms [6, 7, 27], what must be taken into consideration in studies focused on the interrelation of gut microbiota and BA metabolism.

Additionally, two clusters were identified after a hierarchical clustering analysis (HCA) in both urine and feces samples and projected on a principal component analysis (PCA) scores plot, as shown in Fig. 3. As observed in the loadings plot, urine samples corresponding to cluster 1 presented higher concentrations of primary, sulfated and secondary conjugated BAs, while samples from cluster 2 were more concentrated in secondary and primary conjugated BAs. This can also be evidenced in the BA class abundance distribution, as exemplified in the pie charts from four of the samples in each cluster. Regarding feces, samples included in cluster 2 present higher concentrations in primary and secondary BAs, while samples from cluster 1 were less concentrated in those BA classes, as evidenced in the sample pie charts.

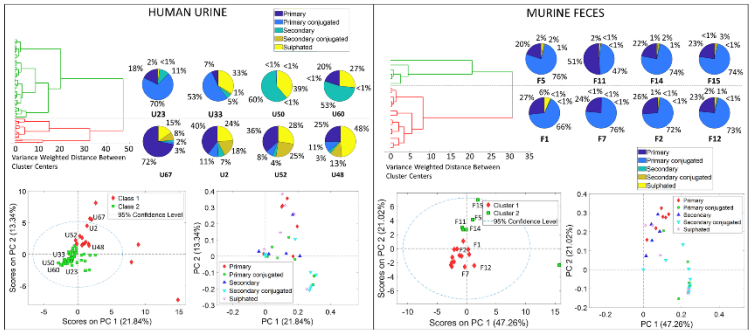


Fig. 3 BA patterns in human urine and murine feces samples. Top: hierarchical clustering analysis revealing two sub-groups within the study samples (left) and pie charts of four samples from each cluster as examples (right). Bottom: PCA scores (left) and loadings (right) plots

Conclusions

An UHPLC-MS/MS method suitable for the quantitative analysis of BAs in urine and feces samples has been validated. Levels of human urinary BAs and murine fecal samples were reported showcasing the applicability of the method in the (pre-)clinical and clinical fields. Furthermore, a comparison between human urinary and murine fecal BA metabolic pathways and excretion mechanisms was provided. Additionally, subgroups with distinct BA profiles within both urine and feces were observed. The novel approach of BA analysis in both human urine and murine feces samples provides valuable insights into the metabolic processes, allowing for a comprehensive understanding of BA metabolism across different species. In future studies, integrating BAs and microbiota analysis could provide a more comprehensive understanding of the relationship between diet, microbiome, health, and disease.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00216-023-04802-8.

Acknowledgements The authors would like to thank all study participants, as well as the nurses and medical staff involved, for their effort and commitment.

Author contribution Conceptualization: Julia Kuligowski and Guillermo Quintás. Methodology: Victoria Ramos-Garcia. Formal analysis: Victoria Ramos-Garcia, Guillermo Quintás. Data curation: Victoria Ramos-Garcia, Guillermo Quintás. Writing—original draft preparation: Victoria Ramos-Garcia, Clara Bullich-Vilarrubias, Marina Romaní-Pérez, Yolanda Sanz, Angelica Nobili, Marika Falcone, Marina Di Stefano. Writing—review and editing: all authors. Supervision: Julia Kuligowski, Isabel Ten-Doménech, and Máximo Vento. Project administration: Julia Kuligowski. Funding acquisition: Julia Kuligowski. All authors have read and agreed to the published version of the manuscript.

Funding This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110] and the Carlos III Health Institute (Ministry of Science and Innovation, Spain) and co-funded by the European Union [grant numbers CD19/00176, and CPII21/00003], grant Despega IATA from IATA-CSIC, the Spanish Ministry of Science and Innovation [grant number PID2020-119536RB-I00], and MCIN/AEI/https://doi.org/10.13039/501100011033 and by ERDF A way of making Europe, by the "European Union" [grant number PID2021-125573OB-I00].

Declarations

Ethics approval During the enrolment phase, pediatricians informed the parents about the purpose of this study, the lack of reported risks related to the collection of samples, the effort required to take part in this study, and their right to deny or withdraw their consent at any time. Parents accepted to participate in the study and signed an informed consent form. The study was approved by the Institutional Ethical Committee of the IRCCS San Raffaele Scientific Institute (Protocol: NUTRI-T1D, 2019). The experimental procedure using animals was in accordance with European Union 2010/63/UE and Spanish RD53/2013 guidelines and approved by the ethics committee of the University of Valencia (Animal Production Section, SCSIE, University of Valencia, Spain) and authorized by *Dirección General de Agricultura, Ganadería y Pesca (Generalitat Valenciana)* (approval ID 2021/VSC/PEA/0273).

Conflict of interest The authors declare no competing interests.

References

1. Vaz FM, Ferdinandusse S. Bile acid analysis in human disorders of bile acid biosynthesis. Mol Aspects Med. 2017;56:10–24. https://doi.org/10.1016/j.mam.2017.03.003.
2. Chassaing B, Gewirtz AT. Gut microbiome and metabolism, in: Physiology of the Gastrointestinal Tract, Elsevier. 2018: pp. 775–793. https://doi.org/10.1016/B978-0-12-809954-4.00035-9.
3. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res. 2006;47:241–59. https://doi.org/10.1194/jlr.R500013-JLR200.
4. Vasavan T, Ferraro E, Ibrahim E, Dixon P, Gorelik J, Williamson C. Heart and bile acids – Clinical consequences of altered bile acid metabolism. Biochimica et Biophysica Acta (BBA) - Mol Basis Dis. 2018:1864;1345–55. https://doi.org/10.1016/j.bbadis.2017.12.039.
5. Li S, Li C, Wang W. Bile acid signaling in renal water regulation. Am J Physiol Renal Physiol. 2019;317:F73–6. https://doi.org/10.1152/ajprenal.00563.2018.
6. Straniero S, Laskar A, Savva C, Härdfeldt J, Angelin B, Rudling M. Of mice and men: murine bile acids explain species differences in the regulation of bile acid and cholesterol metabolism. J Lipid Res. 2020;61:480–91. https://doi.org/10.1194/jlr.RA119000307.
7. Lin Q, Tan X, Wang W, Zeng W, Gui L, Su M, Liu C, Jia W, Xu L, Lan K. Species differences of bile acid redox metabolism: tertiary oxidation of deoxycholate is conserved in preclinical animals. Drug Metab Dispos. 2020;48:499–507. https://doi.org/10.1124/dmd.120.090464.
8. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. Annu Rev Biochem. 2003;72:137–74. https://doi.org/10.1146/annurev.biochem.72.121801.161712.
9. Dawson PA. Bile formation and the enterohepatic circulation, in: Physiology of the Gastrointestinal Tract, Elsevier. 2018: pp. 931–956. https://doi.org/10.1016/B978-0-12-809954-4.00041-4.
10 Alnouti Y. Bile acid sulfation: a pathway of bile acid elimination and detoxification. Toxicol Sci. 2009;108:225–46. https://doi.org/10.1093/toxsci/kfn268.
11. Barbier O, Trottier J, Kaeding J, Caron P, Verreault M. Lipid-activated transcription factors control bile acid glucuronidation. Mol Cell Biochem. 2009;326:3–8. https://doi.org/10.1007/s11010-008-0001-5.
12. Durník R, Šindlerová L, Babica P, Jurček O. Bile acids transporters of enterohepatic circulation for targeted drug delivery. Molecules. 2022;27:2961. https://doi.org/10.3390/molecules27092961.
13. Zwicker BL, Agellon LB. Transport and biological activities of bile acids. Int J Biochem Cell Biol. 2013;45:1389–98. https://doi.org/10.1016/j.biocel.2013.04.012.
14. Bathena SPR, Thakare R, Gautam N, Mukherjee S, Olivera M, Meza J, Alnouti Y. Urinary bile acids as biomarkers for liver diseases I. Stability of the Baseline Profile in Healthy Subjects. Toxicol Sci. 2015;143:296–307. https://doi.org/10.1093/toxsci/kfu227.
15. Dosedělová V, Itterheimová P, Kubáň P. Analysis of bile acids in human biological samples by microcolumn separation techniques: a review. Electrophoresis. 2021;42:68–85. https://doi.org/10.1002/elps.202000139.
16. Zheng X, Huang F, Zhao A, Lei S, Zhang Y, Xie G, Chen T, Qu C, Rajani C, Dong B, Li D, Jia W. Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice. BMC Biol. 2017;15:120. https://doi.org/10.1186/s12915-017-0462-7.

17. Steiner C, von Eckardstein A, Rentsch KM. Quantification of the 15 major human bile acids and their precursor 7α-hydroxy-4-cholesten-3-one in serum by liquid chromatography-tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. 2010;878:2870–80. https://doi.org/10.1016/j.jchromb.2010.08.045.
18. Perwaiz S, Tuchweber B, Mignault D, Gilat T, Yousef IM. Determination of bile acids in biological fluids by liquid chromatography-electrospray tandem mass spectrometry. J Lipid Res. 2001;42:114–9. https://doi.org/10.1016/S0022-2275(20)32342-7.
19. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. Targeted profiling of circulating and hepatic bile acids in human, mouse, and rat using a UPLC-MRM-MS-validated method. J Lipid Res. 2012;53:2231–41. https://doi.org/10.1194/jlr.D028803.
20. Humbert L, Maubert MA, Wolf C, Duboc H, Mahé M, Farabos D, Seksik P, Mallet JM, Trugnan G, Masliah J, Rainteau D. Bile acid profiling in human biological samples: comparison of extraction procedures and application to normal and cholestatic patients. J Chromatogr B. 2012;899:135–45. https://doi.org/10.1016/j.jchromb.2012.05.015.
21. Sarafian MH, Lewis MR, Pechlivanis A, Ralphs S, McPhail MJW, Patel VC, Dumas M-E, Holmes E, Nicholson JK. Bile acid profiling and quantification in biofluids using ultra-performance liquid chromatography tandem mass spectrometry. Anal Chem. 2015;87:9662–70. https://doi.org/10.1021/acs.analchem.5b01556.
22. Ushijima K, Kimura A, Inokuchi T, Yamato Y, Maeda K, Yamashita Y, Nakashima E, Kato H. Placental transport of bile acids: analysis of bile acids in maternal serum and urine, umbilical cord blood, and amniotic fluid., Kurume. Med J. 2001;48:87–91. https://doi.org/10.2739/kurumemedj.48.87.
23. Kumar BS, Chung BC, Lee Y-J, Yi HJ, Lee B-H, Jung BH. Gas chromatography-mass spectrometry-based simultaneous quantitative analytical method for urinary oxysterols and bile acids in rats. Anal Biochem. 2011;408:242–52. https://doi.org/10.1016/j.ab.2010.09.031.
24. Matysik S, Schmitz G. Application of gas chromatography–triple quadrupole mass spectrometry to the determination of sterol components in biological samples in consideration of the ionization mode. Biochimie. 2013;95:489–95. https://doi.org/10.1016/j.biochi.2012.09.015.
25. Food and Drug Administration (FDA). US Department of Health and Human Services. Bioanalytical method validation. Guidance for industry. 2018.
26. Committee for Medicinal Products for Human Use (CHMP). Guideline on validation of bioanalytical methods. 2022.
27. Takahashi S, Fukami T, Masuo Y, Brocker CN, Xie C, Krausz KW, Wolf CR, Henderson CJ, Gonzalez FJ. Cyp2c70 is responsible for the species difference in bile acid metabolism between mice and humans. J Lipid Res. 2016;57:2130–7. https://doi.org/10.1194/jlr.M071183.

metabolites

MDPI

Review

Current Practice in Untargeted Human Milk Metabolomics

Isabel Ten-Doménech [1,†], Victoria Ramos-Garcia [1,†], José David Piñeiro-Ramos [1], María Gormaz [1,2], Anna Parra-Llorca [1], Máximo Vento [1,2], Julia Kuligowski [1,*] and Guillermo Quintás [3,4]

1 Neonatal Research Unit, Health Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain; isabel_ten@iislafe.es (I.T.-D.); victoria_ramos@iislafe.es (V.R.-G.); jose_pineiro@iislafe.es (J.D.P.-R.); gormaz_mar@gva.es (M.G.); annaparrallorca@gmail.com (A.P.-L.); maximo.vento@uv.es (M.V.)

2 Division of Neonatology, University & Polytechnic Hospital La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

3 Health and Biomedicine, Leitat Technological Center, Carrer de la Innovació, 2, 08225 Terrassa, Spain; gquintas@leitat.org

4 Unidad Analítica, Health Research Institute La Fe, Avenida Fernando Abril Martorell 106, 46026 Valencia, Spain

* Correspondence: julia.kuligowski@uv.es; Tel.: +34-961-246-661

† Both authors contributed equally.

Received: 29 November 2019; Accepted: 19 January 2020; Published: 22 January 2020

check for updates

Abstract: Human milk (HM) is considered the gold standard for infant nutrition. HM contains macro- and micronutrients, as well as a range of bioactive compounds (hormones, growth factors, cell debris, etc.). The analysis of the complex and dynamic composition of HM has been a permanent challenge for researchers. The use of novel, cutting-edge techniques involving different metabolomics platforms has permitted to expand knowledge on the variable composition of HM. This review aims to present the state-of-the-art in untargeted metabolomic studies of HM, with emphasis on sampling, extraction and analysis steps. Workflows available from the literature have been critically revised and compared, including a comprehensive assessment of the achievable metabolome coverage. Based on the scientific evidence available, recommendations for future untargeted HM metabolomics studies are included.

Keywords: human milk; metabolome; sampling; extraction; liquid chromatography–mass spectrometry (LC-MS); nuclear magnetic resonance (NMR); gas chromatography–mass spectrometry (GC-MS); capillary electrophoresis—mass spectrometry (CE-MS)

1. Introduction

Human milk (HM) has been markedly established as the optimal way of providing infants with the necessary nutrients and bioactive factors for their early development. Many health associations and organisms, including World Health Organization, recommend exclusive breastfeeding for the first six months of life [1]. Health benefits of HM for infants include reduced mortality and morbidity, including sepsis, respiratory diseases, otitis media, gastroenteritis, and urinary tract infections, among others [2]. In addition, studies reporting on long-term benefits of HM consumption such as lower risk of suffering from type 1 diabetes and inflammatory bowel disease or overweight in adulthood emerged [3]. HM may also be associated with a slightly improved neurological outcome as cohort studies report [4], especially in preterm infants [5], although potential confounders must be accounted for [6].

HM composition is dynamic and influenced by several factors including genetics, gestational and infant's age, circadian rhythm, maternal nutrition, or ethnicity. It provides a series of nutrients

such as lipids, proteins, carbohydrates, and vitamins, jointly with a number of bioactive factors that contribute to several physiological activities in the newborn infant as well as to short- and long-term outcomes [7,8]. Living cells including stem cells, hormones, growth factors, enzymes, microbiota, and even genetic material are part of this vast array of HM components with impact in early development, particularly the immune system [9]. In addition, HM appears to be one of the richest sources of microRNAs [10]. On the other hand, because of the maternal environmental exposure and lifestyle, the presence of some contaminants such as persistent organic pollutants or pharmacologically active substances in HM has been described [11,12].

Due to its complex composition, the analysis of HM is not straightforward. While the advent of "omics" approaches has offered valuable insights into the composition of this unique biofluid, untargeted metabolomic and lipidomic studies have only recently been applied to HM [13]. The comprehensive study of the HM metabolome, which includes the intermediate and end products of metabolism, can shed light on maternal status or phenotype [14,15]. The generation, analysis, and integration of large and complex data sets obtained in metabolomic studies go hand in hand with the following challenges: (i) the intrinsic complexity of the sample: a rich variety of jointly present, structurally heterogeneous compounds at concentrations that strongly vary covering several orders of magnitude; (ii) pre-analytical steps related to sampling, storage, and pre-processing (e.g., extraction, clean-up); and (iii) the diversity of platforms currently available including nuclear magnetic resonance (NMR), as well as gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) coupled to mass spectrometry (MS). The analysis of the HM metabolome has been approached employing a variety of extraction and analytical techniques to respond to a spectrum of clinically relevant questions. Several studies have compared HM metabolome with formula milk [13,16–20] or with milk from other mammalian species including monkey [21], donkey [17], and cow [18], whereas others have made efforts in defining the metabolome of preterm milk [13,16,22–26] and the evaluation of the HM metabolome during the course of lactation [15,23,27–30]. Furthermore, the influence of maternal diet [14,15,31], phenotype [14,32], obesity [30], or atopy status [33], as well as geographical location [33,34], time of the day [29,35], chemotherapy [36], or preeclampsia during pregnancy [31] on the HM metabolome have been reported.

Recent review articles that address the HM metabolome or lipidome [12,37–41] are available. For information on the compounds and compound families typically found in HM and their function the reader is referred to [37–40]. Readers with a particular interest in HM lipidomics are referred to a recent compilation study [41]. Technical aspects of HM analysis when performing metabolomics studies in HM have been recently described [12]. This review article gathers recent literature available on metabolomic analysis of HM, particularly focusing on untargeted approaches as indicated in Figure 1, to provide an up-to-date overview of the key factors that may influence HM metabolome coverage. Based on the information provided within the available literature, recommendations to guide study design and analytical method development of untargeted HM metabolomics assays were developed.

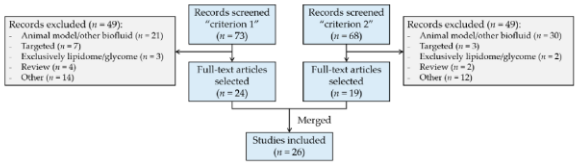


Figure 1. Flow diagram of literature selection and review process. Search "criterion 1": term ("human milk" OR "breast milk"), AND "metabolom*", AND "infant"; only articles. Search "criterion 2": term ("human milk" OR "breast milk"), AND "metabolom*", AND ("GC" OR "LC" OR "NMR" OR "CE"); only articles. Web of Science database was employed for literature search.

2. Considerations Regarding the Study Design

HM is a biofluid characterized by a dynamically varying composition according to several factors including lactation time, time of the day, throughout each feed, maternal status, and the environmental exposure. Although compositional variations have been mainly studied regarding the protein content of HM [42], changes of other compound classes such as fat or vitamins have been also reported [43,44]. Considering the intrinsic variability of HM, the complexity of obtaining representative HM samples is not negligible. Sources of variation related to sample manipulation and compositional variation can be minimized using standard operational procedures (SOPs). SOPs are fundamental to maintain quality assurance (QA) and quality control (QC) process and facilitate repeatable and reproducible research within and across laboratories. However, biologically meaningful results across studies will only be obtained if several key factors during the sample collection process are successfully controlled. This is of special importance in untargeted approaches, where the interpretation of results is especially challenging, and confounding factors introduced by a non-exhaustive sampling protocol can be wrongly attributed to differences between subjects of a studied population. Conversely, biologically meaningful information can be missed or remain unnoticed due to unwanted bias introduced during sample collection.

The information regarding study design provided in HM metabolomics studies varies considerably [13–24,26–32,34–36,45–47] as shown in Figure 2. Repeatedly reported factors have been grouped into three categories and are discussed in detail in the following sections: (1) maternal-infant-related factors (blue bars), (2) time-related factors (green bars), and (3) HM collection-related factors (orange bars). It should be noted that, although the importance of each factor might vary with the scientific question of each study, the authors encourage (i) the use of SOPs employed during sample collection to assure homogenous and representative sampling and (ii) the reporting of all documented factors in order to enhance comparability between results of metabolomic studies on HM. In case of HM, samples are typically collected, handled and sometimes temporary stored and transported by the mothers and not, such as it is the case for other biofluids (e.g., plasma or serum), by health professionals. During study design it is therefore very important to assure that mothers receive detailed instructions and/or training for the correct handling of collected samples. In addition, one should keep in mind that sampling protocols should neither interfere with infant feeding nor negatively impact on the mother-baby bonding. Hence, the collection of transitional and mature milk is usually preferred over colostrum, especially in studies involving mothers of preterm infants, where colostrum is usually kept exclusively for the infant's supply.

2.1. *Maternal-Infant-Related Factors*

In HM metabolomic studies, gestational age is frequently reported (see Figure 2), although the impact of this factor on the HM metabolome has not been fully characterized. Studies focused on preterm milk showed that, analogously to full-term milk, its composition is dynamic throughout the first month of lactation [13,16,22]. However, after 5–7 weeks, metabolite composition of HM from mothers of preterm infants resembled that collected from mothers of full-term infants [23]. On the other hand, Marincola et al. [13] observed that HM from mothers of early preterm infants (26 weeks of gestation) differentiated from milk samples from term infants. However, the low number of samples involved in the study (n = 20 and n = 3 mothers of preterm and term infants, respectively) hindered the assessment of the statistical significance of the impact of gestational age on the milk metabolite composition. Sundekilde et al. [23] carried out a longitudinal study on milk from mothers of preterm and full-term infants covering similar lactation periods (3–14 weeks and 3–26 weeks after birth, respectively) and showed that some metabolites were present at significantly different levels in full-term milk compared to preterm milk. On the contrary, Longini et al. [16] did not observe significant differences between preterm and full-term milk within the first week after delivery, only being able to discriminate milk samples from early preterm infants (<29 weeks of gestation). It is worth noting that the effect of gestational age on the HM metabolome has been mainly studied employing NMR

platforms [13,16,22,23], in which metabolite coverage is limited (see Figure 5) and some metabolite classes (e.g., lipids) are barely accessible. For this reason, and in order to further evaluate the impact of gestational age on the milk metabolome, we warrant more comprehensive metabolomic studies employing complementary analytical platforms.

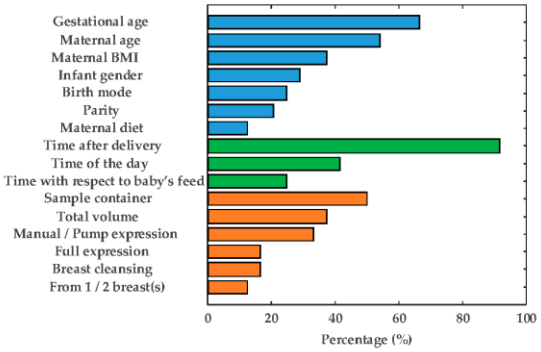


Figure 2. Reporting frequency of factors relevant to the human milk (HM) sampling process: Maternal-infant-related factors (blue bars), time-related factors (green bars), and HM collection-related factors (orange bars). Note: BMI = body mass index.

Other potentially relevant, miscellaneous information about the studied population of mother-infant pairs such as infant gender, parity, and birth mode, have been frequently reported in metabolomic studies (see Figure 2), although these characteristics often remain in the background since the studies focus on other aspects. The influence of these factors on the metabolite composition of HM has not been elucidated yet, and this might be addressed in forthcoming studies. An additional factor that is not typically reported in HM metabolomics studies is maternal secretor status. Significant differences in the oligosaccharides profile of milk between so-called secretors (Se^+), which are those mothers that provide a functional *FUT2* gene, and non-secretors (Se^-) have been reported [48]. Secretor status is mainly established based on the presence (Se^+) or absence (Se^-) of 2′-fucosyllactose, with a prevalence rate of approximately 80% of secretors over non-secretors [14,22,23,26,32,49]. Maternal secretor status is therefore usually determined *a posteriori* during data processing and analysis. Oligosaccharides are polar compounds that are present at concentrations in the mM range that will likely be preserved during sample extraction procedures employed for metabolomics studies. As their presence/absence might potentially affect clustering of milk based on maternal secretor status [24], to provide this information, when available, might be of interest.

2.2. *Time-Related Factors*

HM undergoes significant changes over time, having established three differentiated lactation stages: colostrum, transitional milk and mature milk. As can be seen in Figure 2, lactation time is reported in the vast majority of metabolomic studies. In particular, several studies have focused on the HM metabolome throughout lactation [15,23,27–30], all of them concluding that significant differences in the metabolic profile over time exist. Therefore, it seems reasonable to report this factor. On the other hand, although it has been demonstrated that diurnal variation affects HM fat content [50], its effect on the overall metabolite composition is a controversial issue which has not yet been adequately

addressed in the available literature. Whereas no significant changes in some lipids and small polar metabolites have been observed [29,35], differences in some micronutrients (e.g., vitamins) could be evidenced [44]. The use of a pool of a 24-h expression of HM should compensate for changes due to diurnal variation, thus, obtaining more representative samples [13,24,25] in longitudinal studies. However, the drawback is that this practice is incompatible with breastfeeding of the infant, which, in turn, might raise severe ethical concerns. A feasible compromise for ameliorating diurnal variations is the use of pooled morning and evening samples [29,35].

Regarding time of collection with respect to baby's feed, the influence of this variable has not been studied to date, but given the differences found between fore- and hindmilk [51], it seems reasonable to assume that this factor might be potentially relevant.

2.3. HM Collection-Related Factors

Any uncontrolled variable within an experiment can result in a potential source of bias. In this sense, although less attention has been payed to other factors related to the expression and storage of HM (see Figure 2, orange bars), they may be relevant to the outcomes of metabolomic studies. As can be seen, the type of sample container is indicated in 50% of the studies, whereas other specifications regarding HM expression are included scarcely. The latter factor deserves some special attention, since differences in the milk fat content between foremilk (initial milk of a feed) and hindmilk (last milk of a feed) have been reported [51]. Therefore, full expression of breast(s) is desirable in order to obtain a representative HM aliquot [52]. The influence of all other factors, to date, remains unstudied.

2.4. Pasteurization and Storage

HM banks rely on stringent protocols in which pasteurization, indispensable for minimizing the potential to transmit infectious agents, as well as freezing and long-term storage procedures are established. The pasteurization process affects some of the nutritional and biological properties of HM [53–55]. In this review, three studies that use milk from HM banks are included [16,20,23], but only one specifies whether or not HM has undergone pasteurization [23]. Variability of the metabolite profile of HM caused by pasteurization has not been comprehensively explored to date. Future studies focused on the systematic exploration of the effect of thermal treatment on HM are warranted.

HM is usually stored frozen employing −20 °C and −80 °C for short- and long-term storage, respectively. However, duration of storage and the effect of repeated freeze-thaw cycles are identified as additional factors with potential impact on HM composition that are missing in most published studies. In lipidomic studies, the integrity of HM samples is preserved by subjecting HM to inactivation of endogenous enzymes such as lipases in order to minimize lipolysis and lipogenesis. In this sense, immediate storage at −80 °C is advisable [41]. Particularly for metabolite composition analysis, storage at −80 °C is widespread [13,15,18,19,24–26,28,31,45,46], sometimes with a prior short-term storage at −20 °C [14,21,22,27,29,33,34]. Wu et al. [29] investigated the effect of storage conditions by keeping samples for different times at −20 °C and then transferring them to −80 °C versus storing samples directly at −80 °C. Variations in duration of storage at −20 °C versus −80 °C showed no detectable effect on the metabolites considered (e.g., lactose and other carbohydrates, choline and its derivatives, and a variety of amino acids) by visual inspection of sample clusters in principal component analysis scores plots. However, analysis of variance evidenced differences in butyrate, caprate, and acetate contents. However, time of storage considered in this study was limited to two weeks, which is not representative for conditions employed in clinical studies or standard home routines. It is therefore clear that further studies are required in this regard.

On the other hand, HM employed for research studies is typically stored in small aliquots. When working with raw milk, this procedure might introduce bias due to phase separation prior to the preparation of the aliquots. Hence, the homogenization of HM with a disruptor, resulting in a stable emulsion with reduced size of milk fat globules [56], as employed prior to the quantitation of macronutrients with HM analyzers, might be advisable.

3. Metabolite Extraction from HM

For metabolite extraction from HM, an array of methods has been reported. An overview of the employed approaches is shown in Figure 3. The selection of the extraction method is conditioned by the study objective and the subsequent analysis method. As in other untargeted metabolomics workflows, for HM metabolomics, the selected sample preparation approach should enable a high degree of metabolome coverage while making the sample matrix compatible with the analytical platform. Other considerations might include the available amount of sample volume and the use of one sample extraction procedure for subsequent analysis by multiple, complementary analytical platforms [13,27,28].

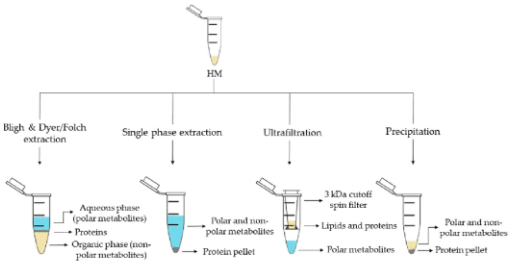


Figure 3. Sample preparation approaches employed in human milk (HM) metabolomics.

Liquid-liquid extraction (LLE) is the classical extraction method employed in metabolomics and lipidomics. This method, developed by Folch et al. [57] in 1957, uses a chloroform-methanol mixture (2:1, *v/v*), which results in two differentiate phases: an upper phase containing polar metabolites and a lower phase containing nonpolar metabolites. Subsequently, in 1959 Bligh and Dyer [58] developed a modified method using a miscible chloroform-methanol-water mixture and later separated into two phases by adding chloroform or water. Both approaches enable the separation of polar and nonpolar metabolites, thus, allowing the analysis of a wide range of metabolites and making them compatible with several analytical platforms. While the use of Bligh and Dyer LLE is widely extended for HM metabolomics studies (see Table 1) [13,16–19,24,25,29,32], only Andreas et al. [28] used a modified Folch extraction protocol for processing HM samples.

Methyl tert-butyl ether (MTBE) in combination with methanol has recently been proposed for single-phase extraction [27]. MTBE is a nontoxic and noncarcinogenic solvent and it is therefore considered a safe and environmentally friendly alternative to harmful solvents employed in traditional LLE methods, such as chloroform, which is a suspected human carcinogen. In this extraction method, a unique phase containing both, polar and nonpolar metabolites is obtained with a protein pellet at the bottom (see Figure 3). Thus, the simultaneous analysis of lipidome and metabolome in a very small amount of biological sample is achievable. This method has been successfully employed to determine polar metabolites and fatty acids (FAs) in HM by GC-MS [27,28], as well as lipids and polar metabolites by LC-MS [15,27,28], thus, increasing the metabolome coverage by the combined use of complementary analytical platforms.

Table 1. Sample preparation steps and platforms employed in untargeted analysis of HM metabolome.

Sample Preparation (1st. step)	Sample Preparation (2nd. step)	Compound Class	Platform	Column/Capillary	References
Bligh & Dyer extraction	Deuterated solvent addition to aqueous phase	Polar metabolites	^{1}H-NMR	-	[13,16,29,32]
	Derivatization of aqueous phase: methoximation and silylation	Polar metabolites and FAs	GC-MS	DB-5ms	[17–19]
	Derivatization of organic phase: methylation	FAs	GC-MS	DB-5ms	[13]
	Direct injection of aqueous phase	Polar metabolites	LC-QTOF-MS (+)	HILIC	[33]
	Redissolution of aqueous phase in H_2O:ACN (95:5)	Polar metabolites	LC-Orbitrap-MS (+, –)	C18	[24]
	Redissolution of organic phase in (ACN:IPA:H_2O (65:30:5)	Lipidic metabolites	LC-Orbitrap-MS (+,–)	C18	[25]
Folch extraction	Deuterated solvent addition to aqueous and organic phases	Hydrophobic and polar metabolites	^{1}H-NMR	-	[28]
	Redissolution of aqueous phase in formic acid and centrifugation	Polar metabolites (amino acids)	CE-TOF-MS (+)	60 m × 50 µm I.D.	
	Redissolution of organic phase in IPA:H_2O:ACN (2:1:1) and centrifugation	Lipidic metabolites	UPLC-QTOF-MS (+,–)	C18	
Single phase extraction	Derivatization: methoximation and silylation	Polar metabolites and FAs	GC-MS	DB-5ms	[27,28]
	Direct injection	Lipidic (and polar) metabolites	LC-QTOF-MS (+,–)	C8	[27,28]
			UPLC-QTOF-MS (+)	C18	[15]
Fat extraction with n-hexane/IPA	Deuterated solvent addition	TGs	^{13}C-NMR; ^{1}H-NMR	-	[20]
Filtration 3 kDa cutoff spin filter	Deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[14,21,22,29,33]

Table 1. *Cont.*

Sample Preparation (1st. step)	Sample Preparation (2nd. step)	Compound Class	Platform	Column/Capillary	References
Protein precipitation	Derivatization: methoximation and silylation	Polar metabolites	GC-MS	DB-5ms	[36]
	Hybrid SPE-Phospholipid extraction and redissolution in diluted organic phase of Bligh & Dyer extraction	Lipidic metabolites	LC-QTOF-MS (+)	C8	[35]
	Fat removal with CH_2Cl_2 and dansylation of aqueous phase	Polar metabolites (amine/phenol submetabolome)	Chemical isotope labelling LC-QTOF-MS (+)	C18	[45,46]
	Direct injection	Polar metabolites and FAs	UPLC-QTOF-MS (+,–)	C18	[18]
Fat removal by centrifugation	Two additional centrifugations and deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[34]
	Filtration 10 kDa cutoff spin filter and deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[23,26]
Homogenization	Deuterated solvent addition	Polar metabolites	^{1}H-NMR	-	[31]
H_2O-dilution	$NaBH_4$-reduction and PGC cartridge	Oligosaccharides	UPLC-TQD-MS (+)	Hypercarb®	[24]

CE, capillary electrophoresis; FAs, fatty acids; GC, gas chromatography; HILIC, hydrophilic interaction liquid chromatography; IPA, 2-propanol; I.D., inner diameter; LC, liquid chromatography; MS, mass spectrometry; ^{13}C-NMR, carbon-13 nuclear magnetic resonance; ^{1}H-NMR, proton nuclear magnetic resonance; PGC, porous graphitic carbon; QTOF, quadrupole time of flight; TGs, triacylglycerols; TQD, triple quadrupole; UPLC, ultraperformance liquid chromatography; +, positive ionization mode; -, negative ionization mode.

Ultrafiltration makes use of centrifugal molecular weight cutoff filters. Different molecular weight cut-off filters are commercially available for this purpose and repeated centrifugation steps might be employed to remove proteins and lipids (see Table 1). Unlike single-phase extraction, ultrafiltration allows to separate polar metabolites from the HM without dilution [14,21,22,29], however this method does not have the capacity to study the global metabolome of HM. At present, this extraction method has only been used in combination with NMR analyses [14,21,22,29].

Precipitation with organic solvents separates the polar and nonpolar metabolites of the proteins that settle at the bottom of the tube which can then be easily removed by centrifugation. This simple method has been employed for the analysis of polar metabolites by GC-MS after derivatization [36] as well as for the analysis of polar and nonpolar metabolites by LC-MS without further pre-processing [18]. Furthermore, this approach has been implemented in more sophisticated workflows as recently shown by Hewelt-Belka et al. [35]. Here, the authors combined LLE and a protein precipitation and solid-phase extraction (SPE) procedure to prepare HM samples, thereby, enabling the detection of high- and low-abundant lipid species (e.g., glycerolipids and phospholipids) in one LC-MS run.

4. Analytical Platforms Employed in HM Metabolomics

As reflected in Table 1, the use of all analytical platforms that are commonly employed in untargeted metabolomics studies, such as LC-MS, GC-MS, NMR, and, to a lesser extent, CE-MS, has been reported for performing HM metabolomics. ^{1}H-NMR is the most frequently used technique [13,14, 16,20–22,26,28,29,31–33] for both, the analysis of polar and hydrophobic metabolites in HM. ^{13}C-NMR has been reported for the detection of triacylglycerols [20]. NMR is a highly reproducible technique that allows a straightforward library matching after spectral alignment, while at the same time supporting structural elucidation of detected metabolites. However, it presents lower sensitivity, and hence, the achievable metabolome coverage is low in comparison to other analytical platforms.

All other analytical platforms rely on the use of MS detection. LC-MS provides high sensitivity and is characterized by a huge versatility due to the availability of (i) a large selection of chromatographic columns with a variety of stationary phases, that in combination with appropriate mobile phases achieve compound separation based on different retention mechanisms and (ii) an array of different instrumental configurations (i.e. different ion sources and mass analyzers). For example, reversed phase (C8, C18) LC-quadrupole time of flight MS (LC-QTOF-MS) [15,18,27,28,35,46] and LC-Orbitrap-MS [24,25] have been reported for the detection of both, polar and lipidic metabolites in HM; and hydrophilic interaction LC (HILIC)-QTOF-MS for polar metabolite detection [35]. On the other hand, for the successful screening of HM oligosaccharides, a porous graphitic carbon column installed on a LC-triple quadrupole (TQD)-MS instrument was used [24].

GC-MS is the most suitable platform for measuring volatile compounds, while other non-volatile compounds must be derivatized prior to analysis. For HM analysis, methoximation followed by silylation or methylation are commonly employed (see Table 1). The most frequently used column is the DB-5ms column for both polar and FA detection [13,17–19], in some cases with an integrated 10 m pre-column (deactivated fused silica) [27,28,36].

Regarding CE-MS, only one study has been reported for polar metabolite detection in HM [28]. CE provides a series of advantages over other techniques, mainly due to the small sample volumes employed and the efficient separation of polar compounds that is difficult to be achieved by LC columns. However, issues with poor reproducibility, matrix effects and sensitivity may be hindering a widely extended use of this technique for the analysis of complex biological samples such as HM.

Due to the diversified composition of HM, no single analytic technique can resolve the entire HM metabolome. Only multiplatform approaches enable a comprehensive characterization providing a high metabolome coverage including polar and nonpolar metabolites present in HM. In this sense, two studies performing a multiplatform approach were found in the literature combining LC-MS and GC-MS [18,27], and only one study that performed analysis using four different techniques (LC-MS, GC-MS, NMR, and CE-MS) was reported [28].

The use of high-end analytical platforms requires the implementation of QA and QC processes to improve data quality, repeatability, and reproducibility, especially in untargeted metabolomics. For practical guidelines on the use of QC measures in untargeted, MS-based assays the reader is referred to [59]. Pooled QC samples are prepared by mixing small aliquots of the study samples, and therefore, they are considered representative in terms of matrix composition and concentration ranges of the metabolites present in the study samples. QC samples are analyzed repeatedly throughout the analytical sequence alongside the study samples. The signal of each feature detected in QC samples can be used to model and correct systematic changes in the instrument response during the analytical sequence. Additionally, the obtained data can be used to perform intra-study reproducibility assessments and to correct for systematic variation across batches. In HM metabolomics, Smilowitz et al. [14], Andreas et al. [28], and Gay et al. [33] used QC samples for NMR studies, while Villaseñor et al. [27], Mung et al. [46], Hewelt-Belka et al. [35], and Alexandre-Gouabau et al. [24,25] used pooled HM samples for QC purposes in LC-MS-based assays. Considering the highly complex sample matrix of HM, the authors strongly recommend the implementation of QC measures, including the analysis of QC samples, to increase reproducibility and facilitate the joint analysis of data from different studies.

5. The HM Metabolome: Compound Annotation and Coverage

As in other areas of metabolomic research, compound identification is still a major bottleneck in data analysis and interpretation. The Metabolomics Standards Initiative's (MSI) defines four levels of metabolite identification, which include: identified metabolites (level 1); putatively annotated compounds (level 2); putatively annotated compound classes (level 3); and unknown compounds (level 4) [60]. Due to the limited availability of pure analytical standards required to reach level 1, biological databanks and spectral databases are the most important resources for metabolite annotation (levels 2 and 3). A large number of databases are available today, providing different levels of information and complementary data on chemical structures, physicochemical properties, biological functions, and pathway mapping of metabolites [61]. The metabolomics community classifies these resources in several categories: (i) chemical databases; (ii) spectral libraries; (iii) pathway databases; (iv) knowledge databases; and (v) references repositories [62].

Regarding HM metabolomics, the most frequently used databases and libraries are: Human Metabolome Database (HMDB) [63], Metabolite and Chemical Entity Database (METLIN) [64], National Institute of Science and Technology (NIST) library, Fiehn RTL Library [65], LipidMAPS Structure Database (LMSD) [66], Milk Metabolome Database (MCDB) [67,68], Kyoto Encyclopedia of Genes and Genomes (KEGG) [69], MycompoundID with the evidence-based metabolome library (EML) [70], Chenomx NMR Suite Profiles and other online university databases, such as CEU-mass mediator [71,72].

Metabolite assignment in NMR spectra has been performed based on literature data and commercial resonance databases, such as Chenomx NMR Suite Profiles. Metabolite annotation was contrasted with in-house libraries containing pure compound spectra. Some of the proposed assignments were confirmed by two-dimensional NMR spectra, such as Correlation Spectroscopy (COSY) [13,29,31,32], Homonuclear Correlation Spectroscopy (TOCSY) [13,31,32,34], Diffusion-Ordered Spectroscopy (DOSY) [32], Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) [32,34], and Heteronuclear Multiple Bond Correlation (HMBC) [32].

In LC-MS and CE-MS-based studies of the HM metabolome, tentative metabolite annotation has been carried out by matching of accurate masses, isotopic profiles, and/or fragmentation patterns to candidate metabolites in online databases such as KEGG, METLIN, LipidMAPS, and HMDB [18,24,25,27,28,35]. In-house built databases generated by the analysis of commercial standards are also commonly employed [24,25]. In GC-MS, retention index (RI) corrections are made by analyzing a fatty acid methyl ester (FAME) mixture standard solution and assigning a match score between the experimental FAME mixture and theoretical RI values based on the values contained in the Fiehn RTL library. Furthermore, metabolites were complementarily annotated by comparing their mass fragmentation patterns with those available in Fiehn RTL and NIST libraries [13,17–19,27,28,36].

A comprehensive list of annotated and/or identified metabolites in HM from untargeted metabolomics studies [14,15,17–19,21–29,31–36] is reported in Table S1. This table contains information about the metabolites reported in each reference, such as their molecular formula, IDs (LipidMAPS and/or HMDB IDs), the extraction procedure performed, the analytical platform used, and the detected metabolite class. Readers can select metabolites dynamically by filtering data according to the latter information. A total of 1187, 111, and 128 metabolites were reported using LC-MS, GC-MS, and NMR, respectively (see Figure 4). As shown in the Venn diagram, LC-MS and GC-MS allowed the detection of 36 common metabolites (mainly carbohydrates and FAs); a total of 29 metabolites overlapped between LC-MS and NMR (principally oligosaccharides); and 21 metabolites (predominantly amino acids and organic acids) were commonly reported in GC-MS and NMR based studies. Only 13 metabolites were reported by all three platforms, i.e., creatine, tyrosine, arabinose, galactose, glucose, lactose, maltose, capric acid/caprate, caprylic acid/ caprylate, citric acid/citrate, pyruvic acid/pyruvate, hippuric acid/hippurate, and myo-inositol. These metabolites were assigned to different classes including amino acids, carbohydrates, FAs, and organic acids.

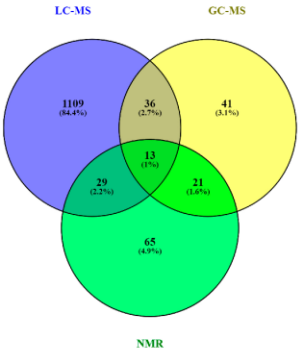


Figure 4. Venn diagram of metabolites reported in human milk (HM) according to the technique in [73]. Note: GC-MS, gas chromatography—mass spectrometry; LC-MS, liquid chromatography—mass spectrometry; NMR, nuclear magnetic resonance.

Based on the available data from the literature, the distribution of metabolite classes present in HM according to each technique was assessed. As can be seen in Figure 5, the difference in detected metabolite classes as observed by LC-MS in comparison to GC-MS and NMR is evident. Using GC-MS and NMR, carbohydrates are the most reported metabolites in HM, followed by amino acids, organic acids, organooxygen compounds, and organoheterocyclic compounds, with all these metabolite classes being certainly less abundant in LC-MS studies. In the case of NMR, organonitrogen compounds have also been reported, as well as nucleosides and nucleotides on a smaller scale. In the case of lipid classes, fatty acyls have been identified by LC-MS and GC-MS with similar incidence and in lesser extent by NMR. It is indubitable that lipid classes are more comprehensively studied by LC-MS assays, where glycerophospholipids, glycerolipids, and fatty acyls are detected at relatively high abundances, followed by sphingolipids, sterol lipids, and, to a lesser extent, prenol lipids.

Table 2 shows a list of metabolites reported in > 80% of studies employing either LC-MS, GC-MS, or NMR-based assays. This table is intended to aid method development of future untargeted metabolomics workflows tailored to the study of the HM metabolome, as it shows a shortlist of

metabolites that should be detected by each platform regardless of the instrumental settings employed. It should be noted that due to the high versatility of LC-MS, there is a greater variation in metabolites recorded and in return, the list of consistently reported metabolites in HM across studies is shorter than for NMR and GC-MS, where differences in experimental conditions and variations between the employed detection parameters and instruments are smaller. Again, this table represents the high orthogonality between the detected metabolites using NMR and LC-MS. While the use of LC-MS is clearly of advantage for the measurement of different lipids, NMR provides information on amino acids and small organic acids. Metabolome coverage provided by GC-MS falls in-between the other two platforms, consistently providing information on lipids, sugars, amino acids, and organic acids.

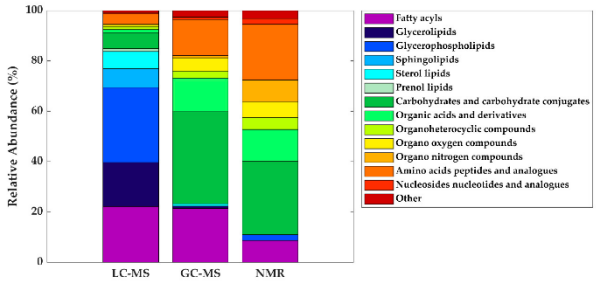


Figure 5. Distribution of metabolite classes annotated and/or identified in HM according to technique. Note: GC-MS, gas chromatography—mass spectrometry; LC-MS, liquid chromatography—mass spectrometry; NMR, nuclear magnetic resonance.

Table 2. Most frequently reported metabolites (>80% of studies) according to technique.

Metabolite class	LC-MS	GC-MS	NMR
Fatty acyls	Linoleic acid (C18:2) Oleic acid (C18:1) Palmitoleic acid (C16:1)	Oleic acid (C18:1) Palmitic acid (C16:0) Stearic acid (C18:0)	-
Glycerolipids	DG (36:1)	-	-
Glycerophospholipids	LysoPC (16:0)	-	-
Carbohydrates and carbohydrate conjugates	-	Fructose Fucose Ribose	Lactose
Organic acids and derivatives	-	Malic acid Urea	Acetate Citrate Lactate
Organo nitrogen compounds	-	-	Choline
Amino acids, peptides, and analogues	-	Alanine Glutamate Glycine Pyroglutamic acid Serine Valine	Alanine Creatine Glutamate Glutamine Isoleucine Leucine Tyrosine Valine

GC-MS, gas chromatography—mass spectrometry; LC-MS, liquid chromatography—mass spectrometry; NMR, nuclear magnetic resonance; DG, diacylglycerol; PC, phosphatidylcholine.

6. Conclusions and Future Perspectives

In less than a decade, 26 research papers have been published trying to shed light on the complex and dynamic composition of HM and the feasibility of different options for sample extraction and metabolite detection has been demonstrated. Due to the many factors that influence HM composition, a thorough study design including SOPs for milk extraction, collection, and storage is indispensable for obtaining biologically meaningful results. Multi-platform approaches are encouraged for providing adequate metabolome coverage, as the diversity of compounds contained in HM will not be properly reflected using one single assay. In line with metabolomics workflows tailored to other sample types, the reproducibility of HM metabolomics studies will benefit from the implementation of QA/QC procedures. Automated metabolite annotation and identification with pure chemical standards is warranted and the authors encourage the use of publicly accessible platforms for enabling the exchange of raw data for comparison between studies.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-1989/10/2/43/s1, Table S1: List of metabolites annotated and/or identified in HM metabolomic studies.

Author Contributions: Conceptualization, G.Q., J.K., M.G., and M.V.; methodology, A.P.-L., I.T.-D., J.D.P.-R. and V.R.-G.; software, G.Q., I.T.-D. and J.K., and validation, G.Q. and J.K.; formal analysis, I.T.-D. and V.R.-G.; investigation, I.T.-D. and V.R.-G.; resources, J.K.; data curation, I.T.-D. and V.R.-G.; writing—original draft preparation, A.P.-L., I.T.-D., J.D.P.-R. and V.R.-G.; writing—review and editing, all authors; visualization, I.T.-D. and V.R.-G.; supervision, J.K. and M.V; project administration, G.Q. and J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: A.P.-L., I.T.-D. and J.K. received salary support by the *Instituto de Salud Carlos III* (Ministry of Economy and Competitiveness, Spain), grant numbers CD19/00176, CM18/00165, and CP16/00034, respectively. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 818110.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization Breastfeeding. Available online: https://www.who.int/topics/breastfeeding/en/ (accessed on 2 July 2019).
2. Geddes, D.; Perrella, S. Breastfeeding and human lactation. *Nutrients* **2019**, *11*, 802–806. [CrossRef]
3. Owen, C.G.; Martin, R.M.; Whincup, P.H.; Davey Smith, G.; Cook, D.G. Does breastfeeding influence risk of type 2 diabetes in later life? A quantitative analysis of published evidence. *Am. J. Clin. Nutr.* **2006**, *84*, 1043–1054. [CrossRef]
4. Der, G.; Batty, G.D.; Deary, I.J. Effect of breast feeding on intelligence in children: Prospective study, sibling pairs analysis, and meta-analysis. *Br. Med. J.* **2006**, *333*, 945–948. [CrossRef] [PubMed]
5. Rozé, J.C.; Darmaun, D.; Boquien, C.Y.; Flamant, C.; Picaud, J.C.; Savagner, C.; Claris, O.; Lapillonne, A.; Mitanchez, D.; Branger, B.; et al. The apparent breastfeeding paradox in very preterm infants: Relationship between breast feeding, early weight gain and neurodevelopment based on results from two cohorts, EPIPAGE and LIFT. *BMJ Open* **2012**, *2*, 1–9. [CrossRef] [PubMed]
6. Horta, B.L.; Loret De Mola, C.; Victora, C.G. Breastfeeding and intelligence: A systematic review and meta-analysis. *Acta Paediatr. Int. J. Paediatr.* **2015**, *104*, 14–19. [CrossRef] [PubMed]
7. Lönnerdal, B. Bioactive proteins in human milk: Mechanisms of action. *J. Pediatr.* **2010**, *156*, S26–S30. [CrossRef]
8. Musilova, S.; Rada, V.; Vlkova, E.; Bunesova, V. Beneficial effects of human milk oligosaccharides on gut microbiota. *Benef. Microbes* **2014**, *5*, 273–283. [CrossRef]
9. Ballard, O.; Morrow, A.L. Human milk composition: Nutrients and bioactive factors. *Pediatr. Clin. N. Am.* **2013**, *60*, 49–74. [CrossRef]
10. Alsaweed, M.; Hartmann, P.E.; Geddes, D.T.; Kakulas, F. MicroRNAs in Breastmilk and the Lactating Breast: Potential Immunoprotectors and Developmental Regulators for the Infant and the Mother. *Int. J. Environ. Res. Public Health* **2015**, *12*, 13981–14020. [CrossRef]

11. Van den Berg, M.; Kypke, K.; Kotz, A.; Tritscher, A.; Lee, S.Y.; Magulova, K.; Fiedler, H.; Malisch, R. WHO/UNEP global surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit–risk evaluation of breastfeeding. *Arch. Toxicol.* **2017**, *91*, 83–96. [CrossRef]
12. Garwolińska, D.; Namieśnik, J.; Kot-Wasik, A.; Hewelt-Belka, W. State of the art in sample preparation for human breast milk metabolomics—Merits and limitations. *TrAC Trends Anal. Chem.* **2019**, *114*, 1–10. [CrossRef]
13. Marincola, F.C.; Noto, A.; Caboni, P.; Reali, A.; Barberini, L.; Lussu, M.; Murgia, F.; Santoru, M.L.; Atzori, L.; Fanos, V. A metabolomic study of preterm human and formula milk by high resolution NMR and GC/MS analysis: Preliminary results. *J. Matern.-Fetal Neonatal Med.* **2012**, *25*, 62–67. [CrossRef] [PubMed]
14. Smilowitz, J.T.; Sullivan, A.O.Ö.; Barile, D.; German, J.B.; Lo, B. The human milk metabolome reveals diverse oligosaccharide profiles. *J. Nutr.* **2013**, *143*, 1709–1718. [CrossRef] [PubMed]
15. Li, K.; Jiang, J.; Xiao, H.; Wu, K.; Qi, C.; Sund, J.; Li, D. Changes in metabolites profile of breast milk over lactation stages and their relationship with dietary intake in Chinese: HPLC-QTOFMS based metabolomic analysis. *Food Funct.* **2018**, *9*, 5189–5197. [CrossRef] [PubMed]
16. Longini, M.; Tataranno, M.L.; Proietti, F.; Tortoriello, M.; Belvisi, E.; Vivi, A.; Tassini, M.; Perrone, S.; Buonocore, G. A metabolomic study of preterm and term human and formula milk by proton MRS analysis: Preliminary results. *J. Matern.-Fetal Neonatal Med.* **2014**, *7058*, 27–33. [CrossRef]
17. Murgia, A.; Scano, P.; Contu, M.; Ibba, I.; Altea, M.; Demuru, M.; Porcu, A.; Caboni, P. Characterization of donkey milk and metabolite profile comparison with human milk and formula milk. *LWT* **2016**, *74*, 427–433. [CrossRef]
18. Qian, L.; Zhao, A.; Zhang, Y.; Chen, T.; Zeisel, S.H.; Jia, W.; Cai, W. Metabolomic approaches to explore chemical diversity of human breast-milk, formula milk and bovine milk. *Int. J. Mol. Sci.* **2016**, *17*, 2128–2143. [CrossRef]
19. Scano, P.; Murgia, A.; Demuru, M.; Consonni, R.; Caboni, P. Metabolite profiles of formula milk compared to breast milk. *Food Res. Int.* **2016**, *87*, 76–82. [CrossRef]
20. Lopes, T.I.B.; Cañedo, M.C.; Oliveira, F.M.P.; Ancantara, G.B. Toward precision nutrition: Commercial infant formulas and human milk compared for stereospecific distribution of fatty acids using metabolomics. *Omics J. Integr. Biol.* **2018**, *22*, 484–492. [CrossRef]
21. O'Sullivan, A.; He, X.; McNiven, E.M.S.; Hinde, K.; Haggarty, N.W.; Lönnerdal, B.; Slupsky, C.M. Metabolomic phenotyping validates the infant rhesus monkey as a model of human infant metabolism. *J. Pediatr. Gastroenterol. Nutr.* **2013**, *56*, 355–363. [CrossRef]
22. Spevacek, A.R.; Smilowitz, J.T.; Chin, E.L.; Underwood, M.A.; German, J.B.; Slupsky, C.M. Infant maturity at birth reveals minor differences in the maternal milk metabolome in the first month of lactation. *J. Nutr.* **2015**, *145*, 1698–1708. [CrossRef] [PubMed]
23. Sundekilde, U.K.; Downey, E.; Mahony, J.A.O.; Shea, C.O.; Ryan, C.A.; Kelly, A.L.; Bertram, H.C. The effect of gestational and lactational age on the human milk metabolome. *Nutrients* **2016**, *8*, 304–318. [CrossRef] [PubMed]
24. Alexandre-Gouabau, M.-C.; Moyon, T.; David-Sochard, A.; Fenaille, F.; Cholet, S.; Royer, A.-L.; Guitton, Y.; Billard, H.; Darmaun, D.; Rozé, J.-C.; et al. Comprehensive preterm breast milk metabotype associated with optimal infant early growth pattern. *Nutrients* **2019**, *11*, 528–553. [CrossRef] [PubMed]
25. Alexandre-Gouabau, M.C.; Moyon, T.; Cariou, V.; Antignac, J.P.; Qannari, E.M.; Croyal, M.; Soumah, M.; Guitton, Y.; David-Sochard, A.; Billard, H.; et al. Breast milk lipidome is associated with early growth trajectory in preterm infants. *Nutrients* **2018**, *10*, 164–192. [CrossRef]
26. Dessì, A.; Briana, D.; Corbu, S.; Gavrili, S.; Marincola, F.C.; Georgantzi, S.; Pintus, R.; Fanos, V.; Malamitsi-Puchner, A. Metabolomics of breast milk: The importance of phenotypes. *Metabolites* **2018**, *8*, 79–88. [CrossRef]
27. Villaseñor, A.; Garcia-Perez, I.; Garcia, A.; Posma, J.M.; Fernández-Lópes, M.; Nicholas, A.J.; Modi, N.; Holmes, E.; Barbas, C. Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach. *Anal. Chem.* **2014**, *86*, 8245–8252. [CrossRef]
28. Andreas, N.J.; Hyde, M.J.; Gomez-romero, M.; Lopez-Gonzalvez, M.A.; Villaseñor, A.; Wijeyesekera, A.; Barbas, C.; Modi, N.; Holmes, E.; Garcia-Perez, I. Multiplatform characterization of dynamic changes in breast milk during lactation. *Electrophoresis* **2015**, *36*, 2269–2285. [CrossRef]

29. Wu, J.; Domellöf, M.; Zivkovic, A.M.; Larsson, G.; Öhman, A.; Nording, M.L. NMR-based metabolite profiling of human milk: A pilot study of methods for investigating compositional changes during lactation. *Biochem. Biophys. Res. Commun.* **2016**, *469*, 626–632. [CrossRef]
30. Isganaitis, E.; Venditti, S.; Matthews, T.J.; Lerin, C.; Demerath, E.W.; Fields, D.A. Maternal obesity and the human milk metabolome: Associations with infant body composition and postnatal weight gain. *Am. J. Clin. Nutr.* **2019**, *110*, 111–120. [CrossRef]
31. Dangat, K.; Upadhyay, D.; Kilari, A.; Sharma, U.; Kemse, N.; Mehendale, S.; Lalwani, S.; Wagh, G.; Joshi, S.; Jagannathan, N.R. Altered breast milk components in preeclampsia; An in-vitro proton NMR spectroscopy study. *Clin. Chim. Acta* **2016**, *463*, 75–83. [CrossRef]
32. Praticò, G.; Capuani, G.; Tomassini, A.; Baldassarre, E.; Delfini, M.; Miccheli, A. Exploring human breast milk composition by NMR-based metabolomics. *Nat. Prod. Res.* **2013**, *28*, 95–101. [CrossRef] [PubMed]
33. Gay, M.C.L.; Koleva, P.T.; Slupsky, C.M.; Toit, E.; Eggesbo, M.; Johnson, C.C.; Wegienka, G.; Shimojo, N. Worldwide variation in human milk metabolome: Indicators of breast physiology and maternal lifestyle? *Nutrients* **2018**, *10*, 1151–1162. [CrossRef] [PubMed]
34. Gómez-Gallego, C.; Morales, J.M.; Monleón, D.; du Toit, E.; Kumar, H.; Linderborg, K.M.; Zhang, Y.; Yang, B.; Isolauri, E.; Salminen, S.; et al. Human breast milk NMR metabolomic profile across specific geographical locations and its association with the milk microbiota. *Nutrients* **2018**, *10*, 1355–1375. [CrossRef] [PubMed]
35. Hewelt-Belka, W.; Garwolińska, D.; Belka, M.; Bączek, T.; Namieśnik, J.; Kot-Wasik, A. A new dilution-enrichment sample preparation strategy for expanded metabolome monitoring of human breast milk that overcomes the simultaneous presence of low- and high-abundance lipid species. *Food Chem.* **2019**, *288*, 154–161. [CrossRef]
36. Urbaniak, C.; Mcmillan, A.; Angelini, M.; Gloor, G.B.; Sumarah, M.; Burton, J.P.; Reid, G. Effect of chemotherapy on the microbiota and metabolome of human milk, a case report. *Microbiome* **2014**, *2*, 24–35. [CrossRef]
37. Demmelmair, H.; Koletzko, B. Variation of metabolite and hormone contents in human milk. *Clin. Perinatol.* **2017**, *44*, 151–164. [CrossRef]
38. Marincola, F.C.; Dessì, A.; Corbu, S.; Reali, A.; Fanos, V. Clinical impact of human breast milk metabolomics. *Clin. Chim. Acta* **2015**, *451*, 103–106. [CrossRef]
39. Slupsky, C.M. Metabolomics in human milk research. *Nestle Nutr. Inst. Workshop Ser.* **2019**, *90*, 179–190.
40. Bardanzellu, F.; Fanos, V.; Reali, A. "Omics" in human colostrum and mature milk: Looking to old data with new eyes. *Nutrients* **2017**, *9*, 843–867. [CrossRef]
41. George, A.D.; Gay, M.C.L.; Trengove, R.D.; Geddes, D.T. Human milk lipidomics: Current techniques and methodologies. *Nutrients* **2018**, *10*, 1169–1179. [CrossRef]
42. Lönnerdal, B.; Erdmann, P.; Thakkar, S.K.; Sauser, J.; Destaillats, F. Longitudinal evolution of true protein, amino acids and bioactive proteins in breast milk: A developmental perspective. *J. Nutr. Biochem.* **2017**, *41*, 1–11. [CrossRef] [PubMed]
43. Gidrewicz, D.A.; Fenton, T.R. A systematic review and meta-analysis of the nutrient content of preterm and term breast milk. *BMC Pediatr.* **2014**, *14*, 216–230. [CrossRef] [PubMed]
44. Hampel, D.; Shahab-Ferdows, S.; Islam, M.M.; Peerson, J.M.; Allen, L.H. Vitamin concentrations in human milk vary with time within feed, circadian rhythm, and single-dose supplementation. *J. Nutr.* **2017**, *147*, 603–611. [CrossRef] [PubMed]
45. Mung, D.; Li, L. Development of chemical isotope labeling LC-MS for milk metabolomics: Comprehensive and quantitative profiling of the amine/phenol submetabolome. *Anal. Chem.* **2017**, *89*, 4435–4443. [CrossRef] [PubMed]
46. Mung, D.; Li, L. Applying quantitative metabolomics based on chemical isotope labeling LC-MS for detecting potential milk adulterant in human milk. *Anal. Chim. Acta* **2018**, *1001*, 78–85. [CrossRef] [PubMed]
47. Garcia, C.; Duan, R.D.; Brévaut-Malaty, V.; Gire, C.; Millet, V.; Simeoni, U.; Bernard, M.; Armand, M. Bioactive compounds in human milk and intestinal health and maturity in preterm newborn: An overview. *Cell. Mol. Biol.* **2013**, *59*, 108–131.
48. Garwolińska, D.; Namieśnik, J.; Kot-Wasik, A.; Hewelt-Belka, W. Chemistry of human breast milk—A comprehensive review of the composition and role of milk metabolites in child development. *J. Agric. Food Chem.* **2018**, *66*, 11881–11896. [CrossRef]

49. Totten, S.M.; Zivkovic, A.M.; Wu, S.; Ngyuen, U.; Freeman, S.L.; Ruhaak, L.R.; Darboe, M.K.; German, J.B.; Prentice, A.M.; Lebrilla, C.B. Comprehensive profiles of human milk oligosaccharides yield highly sensitive and specific markers for determining secretor status in lactating mothers. *J. Proteome Res.* **2012**, *11*, 6124–6133. [CrossRef]
50. Lubetzky, R.; Littner, Y.; Mimouni, F.B.; Dollberg, S.; Mandel, D.; Lubetzky, R.; Lubetzky, R.; Littner, Y.; Mimouni, F.B.; Dollberg, S.; et al. Circadian variations in fat content of expressed breast milk from mothers of preterm infants. *J. Am. Coll. Nutr.* **2006**, *25*, 151–154. [CrossRef]
51. Saarela, T.; Kokkonen, J.; Koivisto, M. Macronutrient and energy contents of human milk fractions during the first six months of lactation. *Acta Paediatr.* **2005**, *94*, 1176–1181. [CrossRef]
52. Jensen, R.G.; Lammi-Keefe, C.J.; Koletzko, B. Representative sampling of human milk and the extraction of fat for analysis of environmental lipophilic contaminants. *Toxicol. Environ. Chem.* **1997**, *62*, 229–247. [CrossRef]
53. Bertino, E.; Peila, C.; Cresi, F.; Maggiora, E.; Sottemano, S.; Gazzolo, D.; Arslanoglu, S.; Coscia, A. Donor human milk: Effects of storage and heat treatment on oxidative stress markers. *Front. Pediatr.* **2018**, *6*, 1–5. [CrossRef] [PubMed]
54. Peila, C.; Moro, G.E.; Bertino, E.; Cavallarin, L.; Giribaldi, M.; Giuliani, F.; Cresi, F.; Coscia, A. The effect of holder pasteurization on nutrients and biologically-active components in donor human milk: A review. *Nutrients* **2016**, *8*, 447. [CrossRef] [PubMed]
55. García-Lara, N.R.; Vieco, D.E.; De La Cruz-Bértolo, J.; Lora-Pablos, D.; Velasco, N.U.; Pallás-Alonso, C.R. Effect of holder pasteurization and frozen storage on macronutrients and energy content of breast milk. *J. Pediatr. Gastroenterol. Nutr.* **2013**, *57*, 377–382. [CrossRef]
56. Billard, H.; Simon, L.; Desnots, E.; Sochard, A.; Boscher, C.; Riaublanc, A.; Alexandre-Gouabau, M.C.; Boquien, C.Y. Calibration adjustment of the mid-infrared analyzer for an accurate determination of the macronutrient composition of human milk. *J. Hum. Lact.* **2016**, *32*, NP19–NP27. [CrossRef]
57. Folch, J.; Lees, M.; Sloane Stantley, G.H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **1957**, *266*, 497–509.
58. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917. [CrossRef]
59. Broadhurst, D.; Goodacre, R.; Reinke, S.N.; Kuligowski, J.; Wilson, I.D.; Lewis, M.R.; Dunn, W.B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018**, *14*, 72–89. [CrossRef]
60. Sumner, L.W.; Samuel, T.; Noble, R.; Gmbh, S.D.; Barrett, D.; Beale, M.H.; Hardy, N. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211–221. [CrossRef]
61. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. [CrossRef]
62. Fiehn, O.; Barupal, D.K.; Kind, T. Extending biochemical databases by metabolomic surveys. *J. Biol. Chem.* **2011**, *286*, 23637–23643. [CrossRef] [PubMed]
63. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef] [PubMed]
64. Smith, C.A.; O'Maille, G.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN: A metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747–751. [CrossRef] [PubMed]
65. Kind, T.; Wohlgemuth, G.; Lee, D.Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib—Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* **2009**, *81*, 10038–10048. [CrossRef]
66. Cardiff University; Babraham Institute; University of California, S.D. LIPID MAPS Lipidomics Gateway. Available online: http://www.lipidmaps.org/ (accessed on 8 November 2019).
67. Foroutan, A.; Guo, A.C.; Vazquez-fresno, R.; Lipfert, M.; Zhang, L.; Zheng, J.; Badran, H.; Budinski, Z.; Mandal, R.; Ametaj, B.N.; et al. Chemical composition of commercial cow's milk. *J. Agric. Food Chem.* **2019**, *67*, 4897–4914. [CrossRef]

68. Milk Composition Database. Available online: http://www.mcdb.ca/ (accessed on 5 November 2019).
69. KEGG PATHWAY Database. Available online: https://www.kegg.jp/kegg/pathway.html (accessed on 8 November 2019).
70. Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A.E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D.S.; et al. MyCompoundID: Using an evidence-based metabolome library for metabolite identification. *Anal. Chem.* **2013**, *85*, 3401–3408. [CrossRef]
71. Gil de la Fuente, A.; Godzien, J.; Saugar, S.; Garcia-Carmona, R.; Badran, H.; Wishart, D.S.; Barbas, C.; Otero, A. CEU Mass Mediator 3.0: A Metabolite Annotation Tool. *J. Proteome Res.* **2019**, *18*, 797–802. [CrossRef]
72. CEU Mass Mediator. Available online: http://ceumass.eps.uspceu.es/mediator/ (accessed on 5 November 2019).
73. Oliveros, J.C. Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. Available online: http://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed on 4 November 2019).

Food Chemistry 384 (2022) 132581

Contents lists available at ScienceDirect

Food Chemistry

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

The effect of Holder pasteurization on the lipid and metabolite composition of human milk

Isabel Ten-Doménech [a,1], Victoria Ramos-Garcia [a,1], Marta Moreno-Torres [b], Anna Parra-Llorca [a], María Gormaz [a,c], Máximo Vento [a,c], Julia Kuligowski [a,*], Guillermo Quintás [d,e]

[a] Neonatal Research Group, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[b] Unidad de Hepatología Experimental, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[c] Division of Neonatology, University & Polytechnic Hospital La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain
[d] Health and Biomedicine, Leitat Technological Center, Carrer de la Innovació 2, 08225 Terrassa, Spain
[e] Analytical Unit, Health Research Institute La Fe, Avda Fernando Abril Martorell 106, 46026 Valencia, Spain

ARTICLE INFO

Keywords:
Donor human milk
Holder pasteurization
Preterm infant
Metabolomics
Lipidomics
Network analysis

ABSTRACT

Human milk (HM) is the gold standard for newborn nutrition. When own mother's milk is not sufficiently available, pasteurized donor human milk becomes a valuable alternative. In this study we analyzed the impact of Holder pasteurization (HoP) on the metabolic and lipidomic composition of HM. Metabolomic and lipidomic profiles of twelve paired HM samples were analysed before and after HoP by liquid chromatography mass spectrometry (MS) and gas chromatography-MS. Lipidomic analysis enabled the annotation of 786 features in HM out of which 289 were significantly altered upon pasteurization. Fatty acid analysis showed a significant decrease of 22 out of 29 detectable fatty acids. The observed changes were associated to five metabolic pathways. Lipid ontology enrichment analysis provided insight into the effect of pasteurization on physical and chemical properties, cellular components, and functions. Future research should focus on nutritional and/or developmental consequences of these changes.

1. Introduction

Human milk (HM) is the gold standard for infant nutrition and provides the optimum composition of nutritional elements needed for growth and development. Consequently, HM offers numerous short and long-term benefits to the infant-mother dyad (World Health Organization, 2017). Of note, the survival of extremely low gestational age newborns has consistently improved in the last decades (Boquien, 2018). In this scenario, early infant nutrition has become a major player in improving clinical outcomes of survivors and based on an impressive array of benefits, HM is the first choice for feeding preterm infants (Miller et al., 2018).

When own mother's milk (OMM) renders insufficient, pasteurized donor human milk (DHM) rather than preterm infant formula is the preferred alternative for preterm infants (Poulimeneas et al., 2021). The use of DHM in comparison to the use of formula milk might reduce the incidence of necrotizing enterocolitis (Miller et al., 2018; Quigley, Embleton, & McGuire, 2019), and although it has been linked to lower rates of weight gain, linear growth, and head growth during hospital admission, no differences in long term growth have been described (Quigley et al., 2019). The exclusive feeding of pasteurized DHM undoubtedly has an impact on different aspects of preterm biology. Hence, DHM modifies the gut-microbiota composition of preterm infants as compared to the use of OMM (Parra-Llorca et al., 2018), although it does not compromise the protection against oxidative stress (Parra-Llorca et al., 2019). Furthermore, preterm infants exclusively receiving DHM exhibited different urinary steroid hormone levels when compared to infants receiving OMM. These findings might be potentially linked to steroid hormone concentrations present in fresh and pasteurized HM (Pineiro-Ramos et al., 2021).

The composition of HM is affected by processing and handling of expressed HM following stringent protocols applied in HM banks involving pasteurization, necessary for minimizing the potential to transmit infectious agents, as well as freezing, long-term storage, and multiple container passages (Colaizy, 2021). Low-temperature (62.5 °C) long-time (30 min) pasteurization known as "Holder" pasteurization

* Corresponding author.
E-mail address: Julia.kuligowski@uv.es (J. Kuligowski).
[1] Both authors contributed equally.

https://doi.org/10.1016/j.foodchem.2022.132581
Received 14 October 2021; Received in revised form 4 February 2022; Accepted 25 February 2022
Available online 26 February 2022

(HoP) is routinely employed at HM banks to avoid newborn infection through transmission of pathogens that might be present in DHM as it destroys vegetative forms of bacteria and most viruses including human immunodeficiency virus, herpes, and cytomegalovirus (Colaizy, 2021). The use of softer alternatives (e.g., high-temperature short-time, high pressure processing, ultraviolet and microwave irradiation, and thermosonic processing (Wesolowska et al., 2019) has been in the spotlight of on-going investigations. However, to our knowledge, none of these techniques has been yet implemented for the routine processing of small volume milk batches (approximately 3–6 L) at HM banks.

While total carbohydrates and proteins contents remain relatively stable upon HoP, there is controversy regarding the degree of loss of total fat, as percentages between −6.2 and −25% have been reported in studies replicating current milk bank procedures (Colaizy, 2021). Furthermore, several bioactive components, such as enzymes, immunoglobulins, cytokines, microRNAs, and immune cells are inactivated or destroyed (Wesolowska et al., 2019). Scientific evidence documenting the effects of HoP on changes of specific lipid and metabolite classes is scant. The present study aims at the characterization of the impact of HoP on the global HM metabolome and lipidome. Metabolomic and lipidomic profiles of twelve paired HM samples collected before and after HoP were analyzed by liquid chromatography–high resolution mass spectrometry (LC-HRMS) and gas chromatography-MS (GC–MS). Results obtained enabled the identification of potentially relevant alterations at the metabolic pathway level. Finally, the potential implications of the observed compositional changes in the properties and functionality of HM are discussed.

2. Material and methods

2.1. Human milk samples

The study was conducted in accordance with relevant guidelines and regulations including the Declaration of Helsinki. The Ethics Committee for Biomedical Research of the Health Research Institute La Fe (Valencia, Spain) approved the study protocol (approval number 2014/0247), and mothers gave their written consent to participate. DHM (N = 12) samples were provided by healthy HM donors admitted after routine screening and interview to the HM bank at the University and Polytechnic Hospital La Fe (Valencia, Spain). The median gestational age at which milk donors had given birth was 40 (interquartile range, IQR = 4) weeks. Milk expression was accomplished with breast milk pumps following the standard operating procedure routinely employed at the hospital and the HM bank. Prior to extraction, removable parts of the breast milk pump and collection bottles were sterilized. In addition, mothers washed their hands with soap and water, and their nipples with water. After extraction, bottles were stored at −20 °C. Milk bottles from the same mother were defrosted and pooled together to form a batch of approximately 2 L and after gentle shaking, a 1 mL aliquot of it was collected in a dry, 1.5 mL microcentrifuge tube before and after HoP (30 min at 62.5 °C followed by fast cooling to 4 °C). The median elapsed time between the first and last expression of a pooled DHM sample was 25 (IQR = 36) days. Time of collection in relation to the infants' age was established from this median value and ranged between 21 and 164 days after delivery with a median value of 87 days (IQR = 83).

2.2. Lipidomic and metabolomic analyses of HM samples

Standards and reagents. LC-MS grade acetonitrile (CH_3CN), methanol (CH_3OH), *n*-hexane (>98%), and isopropanol (IPA), reagent grade methyl *tert*-butyl ether (MTBE), ammonium acetate (≥98%), formic acid (HCOOH, >98%), methanolic HCl (3 N), Supelco 37-component fatty acid methyl ester (FAME) mix, lauric acid-D_{23} (≥98%), and nonadecanoic acid (≥98%), were obtained from Merck Life Science S.L.U. (Madrid, Spain). Ultra-pure water (>18.2 MΩ) was generated using a Milli-Q Water Purification System (Merck Millipore, Darmstadt, Germany). (15,15,16,16,17,17,18,18,18-D_9) oleic acid-D_9 (>99%) was purchased from Avanti Polar Lipids (Alabaster, AL, USA) and prostaglandin $F_{2\alpha}$-D_4 (≥98 %, deuterated incorporation ≥ 99%) was purchased from Cayman Chemical Company (Ann Arbor, MI, USA).

HM untargeted metabolomic and lipidomic analyses. HM samples were thawed at room temperature and then heated in a water bath for 10 min at 33 °C. 5 μL of an internal standard (IS) solution containing oleic acid-D_9 (80 μM) and prostaglandin F2α-D_4 (39 μM) in H_2O were added to 45 μL of HM. A single-phase extraction procedure adding 175 μL of CH_3OH and 175 μL of MTBE (Villaseñor et al., 2014) to each HM sample followed by a 4 fold dilution of the supernatant with a CH_3OH:MTBE (1:1, v/v) solution was applied. The resulting extract was used for both lipidomic and metabolomic analyses. In addition, a blank extract was prepared following the steps described for HM samples but replacing HM with water. A pooled quality control (QC) sample was prepared by mixing 20 μL of each HM sample extract.

For untargeted lipidomic and metabolomic analysis, a 1290 Infinity ultra-performance LC (UPLC) system coupled to a 6550 Spectrometer iFunnel quadrupole time-of-flight (QqTOF) MS system from Agilent Technologies (Santa Clara, CA, USA) was used. A detailed description of the parameters used for lipidomic fingerprinting of HM samples can be found elsewhere (Horta et al., 2020; Ten-Doménech et al., 2020). In short, an Acquity BEH C18 column (50 × 2.1 mm, 1.7 μm) from Waters (Milford, MA, USA) running a binary mobile phase gradient with (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) and (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) as mobile phase components was used. Column and autosampler were kept at 55 and 4 °C, respectively, the flow rate was 0.4 mL min^{-1}, and the injection volume was 2 μL. For untargeted metabolomics screening of HM extracts, a Synergi™ Hydro-RP 80 Å LC C18 column (150 × 2 mm, 4 μm) from Phenomenex (Torrance, CA, USA) employing a stepwise gradient with solvent A (H_2O, 0.1% v/v HCOOH) and solvent B (CH_3CN, 0.1% v/v/HCOOH) as mobile phase components was used as follows: 1% B was held for 2 min followed by a linear gradient from 1 to 80% B in 8 min and from 80 to 98% B in 0.1 min; 98% B were held for 1.9 min before returning to initial conditions in 0.1 min and column equilibration with 1% B during 2.9 min. The flow rate was set to 0.4 mL min^{-1}, column and autosampler to 40 and 4 °C, respectively, and the injection volume was 3 μL.

MS detection was carried out in the ESI^+ mode (lipidomics) and the ESI^+ and ESI^- modes (metabolomics). Full scan MS data was acquired between 70 and 1500 *m/z* using the following ionization parameters: gas T, 200 °C; drying gas, 14 L min^{-1}; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11 L min^{-1}. Mass reference standards were introduced into the source for automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (background contaminant), 121.050873 (purine), and 922.009798 (HP-0921) *m/z* as references in ESI^+ and 119.036 (purine) and 980.0163 ([HP-0921 + CH_3COOH-H]-) *m/z* in ESI^-. MS^2 data were acquired using data dependent acquisition methods as explained elsewhere (Ten-Doménech et al., 2020) using centroid mode at a rate of 5 Hz in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, medium isolation window (~4 amu), MS^2 fragmentation with automated selection of five precursor ions per cycle, and an exclusion window of 0.15 min after two consecutive selections of the same precursor. UPLC-QqTOF-MS data acquisition was carried out employing MassHunter Workstation (version B.07.00) from Agilent.

Before UPLC-QqTOF-MS experiments, a system suitability check was carried out by analyzing a 2 μM IS solution. Then, 2 blanks and several QCs (9 and 20 replicates for metabolomics and lipidomic analysis, respectively) were injected at the beginning of the sequence for system conditioning and MS^2 data acquisition. HM sample extracts were injected in random order. The QC was injected every six samples and twice at the beginning and end of the batch for assessment and correction of instrumental performance (Broadhurst et al., 2018). The blank extract was injected twice at the end of the measurement sequence to

identify signals from other than biological origin, and possible carry-over (Martínez-Sena et al., 2019).

Analysis of fatty acid methyl esters (FAMEs). The determination of 36 FAMEs was performed employing a previously described GC–MS method (Cruz-Hernandez, Goeuriot, Giuffrida, Thakkar, & Destaillats, 2013) with slight modifications. Briefly, a HM aliquot was defrosted on ice and gently shaken to avoid phase separation. Then, 250 μL of HM and 600 μL of *n*-hexane containing ISs (12 μM lauric acid-D_{23} and 26 μM nonadecanoic acid) were mixed in a 15 mL test tube equipped with Teflon-lined screw caps. An aliquot of 2 mL of CH_3OH, 2 mL of CH_3OH /HCl (3 N), and 1 mL of *n*-hexane were added and vortexed vigorously. Derivatization was carried out in a water bath at 90 °C for 60 min, with occasional additional shaking. After cooling down to room temperature, 2 mL of water were added and shaken vigorously prior to centrifugation (1200 × *g* for 5 min at 4 °C). The upper hexane layer containing the extracted derivatives was transferred into GC–MS vials. GC–MS analysis was conducted using an Agilent 7890B GC system coupled to an Agilent 5977A quadrupole MS detector operating in selected ion monitoring (SIM) mode. Separations were performed using a Zebron™ ZB-WAX-plus™ column (30 m × 250 μm i.d., 0.25 μm film thickness, Phenomenex, Torrance, CA, USA). Two microliters of derivatives were injected in split mode with a ratio of 40:1, and the solvent delay time was set to 2.6 min. The initial oven temperature was held at 60 °C for 2 min, ramped to 150 °C at a rate of 13 °C min^{-1} and held for 15 min, and to 240 °C at a rate of 2 °C min^{-1} and held for 2 min. Helium was used as a carrier gas at a constant flow rate of 1 mL min^{-1} through the column. The temperatures of the front inlet, transfer line, and electron impact (EI) ion source were set at 250, 290, and 230 °C, respectively, and the electron energy was −70 eV. Further measurement parameters used for the studied analytes are summarized in Supplementary Table S1. For quantification, an external calibration line was employed using standard solutions obtained from different volumes of the Supelco 37-component FAME mix after evaporation and reconstitution in *n*-hexane containing derivatized IS compounds. This procedure was used to remove the 37-component FAME mix solvent (i.e., dichloromethane) and consequently, the most volatile FAME (i.e., FAME of butyric acid) was lost and could no longer be quantified.

The method was validated based on the US Food and Drug Administration (FDA) guidelines for bioanalytical method validation (Food and Drug Administration (FDA). Guidance for Industry: Bioanalytical Method Validation. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine, 2018) including the following bioanalytical parameters: linearity range, selectivity and specificity, sensitivity, carry-over, accuracy, precision, recovery, and stability. The linear range was selected according to the expected concentrations ranges in HM samples. Calibration curves included a blank without analytes nor IS, a zero calibrator (i.e., blank with IS) and, at least, six standards covering the selected concentration ranges. Accuracy, precision, specificity, and recovery of the method were established from the replicate analysis (N = 3) of standards and spiked samples at three concentration levels (low, medium, and high) conducted on three different days to ensure that the extraction of the metabolites was efficient and reproducible. Selectivity and specificity were demonstrated by analyzing multiple blanks. Carry-over was assessed by the analysis of zero-injections after the analysis of high concentrated standards and spiked samples (N = 3). Autosampler and processed sample stability were assessed by comparing concentrations observed in a freshly prepared sample and in the same processed sample after 20 h stored in the autosampler (sealed vial, 25 °C). Analytes' freeze–thaw stability was established by comparing concentrations observed in a sample after three freeze–thaw cycles (−80 °C) to a freshly prepared sample.

2.3. Data processing and statistics

Lipidomics and metabolomics data pre-processing. Centroid UPLC-QqTOF-MS raw data were converted to mzXML format employing ProteoWizard (Kessner, Chambers, Burke, Agus, & Mallick, 2008) (http://proteowizard.sourceforge.net). XCMS software (http://metlin.scripps.edu/xcms/) (Tautenhahn, Patti, Rinehart, & Siuzdak, 2012) and CAMERA (Kuhl, Tautenhahn, Böttcher, Larson, & Neumann, 2012) in R 3.6.1 were employed for the generation of peak tables. Peak table extraction for lipidomics is described elsewhere (Ten-Doménech et al., 2020). The selection of parameters for peak table extraction and alignment for metabolomics was based on the observed variation of RT and *m/z* values of ISs. The centWave method with the following settings was used for peak detection: *m/z* range = 70–1500, ppm = 15, peakwidth = (5 and 20), snthr = 6, prefilter = (3, 100). A minimum difference in *m/z* of 0.01 Da was selected for overlapping peaks. Intensity-weighted *m/z* values of each feature were calculated using the wMean function. Peak limits used for integration were found through descent on the Mexican hat filtered data. Peak grouping was carried out using the "density" method using mzwid = 0.015 and bw = 5. RT correction was carried out using the "obiwarp" method. After peak grouping, the fillPeaks method with the default parameters was applied to fill missing peak data. Automatic integration was assessed by comparison to manual integration using IS signals. A total of 18401 (lipidomic analysis), 1826 (metabolomic analysis, ESI^+) and 893 (metabolomic analysis, ESI^-) features were initially detected after peak detection, integration, chromatographic deconvolution, and alignment in HM samples.

Further data processing and statistical analysis were carried out in MATLAB 2019b (Mathworks Inc., Natick, MA, USA) using in-house written scripts available from the authors, and the PLS Toolbox 8.7 (Eigenvector Research Inc., Wenatchee, USA). Data were annotated by automatic matching of experimental MS^2 spectra to publicly available databases (i.e., HMDB, MS-DIAL, and LipidBlast) as described elsewhere (Ten-Doménech et al., 2020). In addition, Lipidex (Hutchins, Russell, & Coon, 2018) was used for the annotation of lipidomic data (mass error 20 ppm). Enlargement of metabolite annotation was achieved with the Metabolic reaction network-based recursive algorithm (MetDNA) (Shen et al., 2019).

Intra-batch effect correction was performed using the Quality Control-Support Vector Regression algorithm employing a Radial Basis Function kernel (Kuligowski, Sánchez-Illana, Sanjuán-Herráez, Vento, & Quintás, 2015) and the LIBSVM library (Chang & Lin, 2011) with the following parameters: ε–range = 2 to 5%; γ-range = 1 to 10^6; C = 90%. Features with RSD (QC) > 20% after QC-SVRC, and those for which the ratio between the median peak area values in QCs and blanks was lower than a threshold value (i.e., 9 and 3 in metabolomic and lipidomic experiments, respectively) were classified as unreliable and removed from further analysis.

Network-based analysis. Differences between HM before and after pasteurization on the pathway level were studied employing the "Functional Analysis" tool (version 2.0) available in MetaboAnalyst 5.0 (Chong et al., 2018) using a mass accuracy of 10 ppm, the *mummichog* algorithm with a *p*-value cut-off of 0.005, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) *Homo Sapiens* pathway library (Caspi et al., 2016). *Mummichog* analysis was carried out using the *m/z* and RT values of each feature and the FDR-corrected *p*-values from a *t*-test to test whether the unknown population means of the two groups were equal, accounting for unequal variances. Using the results of the *mummichog* algorithm, matched compounds (i.e., hits) were assigned to their corresponding *m/z*-RT features of the dataset. Then, KEGG compounds and KEGG glycans of each significantly altered pathway were searched in the matched *mummichog* compound list and assigned to unique features in the dataset. Fold changes (FC) were calculated as the ratio of medians between groups. The pasteurization effect on the HM lipidome was studied with the web-based ontology enrichment tool for lipidomic analysis: Lipid Ontology analysis - LION (www.lipidontology.com) (Molenaar et al., 2019). As an input, peak areas of annotated features in HM samples were used. The "ranking mode" with $\log_2$FC values as local statistic and a two-tailed alternative hypothesis testing was employed.

To uncover molecular alterations caused by HoP, an integrative analysis of untargeted lipidomic and metabolomic data was carried out with the network-based Prize-collecting Steiner forest algorithm for integrative analysis of untargeted metabolomics (PIUMet) (http://fraenkel-nsf.csbi.mit.edu/piumet2/) (Pirhaji et al., 2016). As an input, *m/z* features with *p*-values < 0.01 from a Wilcoxon signed-rank test of peak areas between pre- and post-pasteurized samples were introduced either as metabolomic or lipidomic peaks, also using the -log(*p*-value) as prize to each input data point.

Data availability. Peak tables extracted from HM UPLC-QqTOF-MS analysis are accessible via the Mendeley Data repository at https://data.mendeley.com/datasets/fnzbxmkv83/1 (lipidomics) and https://data.mendeley.com/datasets/jymtst88jm/1 (metabolomics). Supplementary Table S2 summarizes FAME concentrations in all samples before and after pasteurization.

3. Results & discussion

3.1. Compositional and functional alteration of pasteurized vs raw DHM samples

Lipidomic profiles of DHM samples were acquired by UPLC-QqTOF-MS retrieving a total of 7109 metabolic features after peak detection, deconvolution, integration, alignment, within-batch effect correction, and clean-up. In total, 786 (11%) of all features were annotated. The classes with the largest numbers of annotated lipids were triradylglycerols (TGs) and diradylglycerols (DGs) followed by alkenyl-DGs and ceramides (see Fig. 1, left). 3259 features (45.8% of the total) had significantly different mean values (*t*-test, FDR-adjusted *p*-value < 0.05) and $|\log_2(FC)|>1$, including 289 (9%) annotated features using spectral libraries and MetDNA. The sub-class distribution depicted in Fig. 1 (right) shows that lipid levels across multiple lipid classes were affected by pasteurization, including DGs and TGs, alkenyl-DGs, and linoleic acids and derivatives, among others. Interestingly, 83% of DGs were affected by pasteurization while only 13% of TGs changed. In addition, it must be highlighted that, although the UPLC-QqTOF-MS method was specifically tailored to the detection of lipids, the use of a single-phase extraction procedure in combination with an untargeted detection allowed the detection of compounds from non-lipid classes such as amino acids, peptides, and analogues, and carbohydrates and carbohydrate conjugates.

Lipidomic network analysis was employed for a functional interpretation of the observed changes in the lipidomic profiles within relevant networks using the *mummichog* algorithm (Li et al., 2013). Pathway analysis detected 14 pathways with at least two significantly altered metabolites in DHM samples collected before and after pasteurization. Results summarized in Table 1 indicate a significant impact (*p*-values < 0.05) on metabolites included in the steroid hormone biosynthesis, glycosylphosphatidylinositol (GPI)-anchor biosynthesis, linoleic acid metabolism, biosynthesis of unsaturated fatty acids, and mucin type O-glycan biosynthesis pathways.

Table 1
Altered pathways in DHM linked to the pasteurization process.

Pathway name	Hits (total)	Hits (sig.)	*p*-value
Steroid hormone biosynthesis	6	4	0.02
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	4	3	0.03
Linoleic acid metabolism	7	4	0.03
Biosynthesis of unsaturated fatty acids	17	7	0.04
Mucin type O-glycan biosynthesis	2	2	0.04

Note: *p*-values from Fisher's exact *t*-test; only detected pathways with at least 2 significantly altered features are reported.

Between group comparison of the features' intensities found in the dataset, that corresponded to KEGG compounds and KEGG glycans of each altered pathway (i.e., pre-pasteurization *vs* post-pasteurization) revealed significantly different mean values (FDR-adjusted *p*-value < 0.05) and $|\log_2(FC)|$ >1. Remarkably, for all features in all altered pathways, higher intensities were detected in DHM samples before pasteurization compared with DHM samples after pasteurization, except for (4Z,7Z,10Z,13Z,16Z,19Z)-Docosahexaenoyl-CoA (KEGG ID C16169, *m/z*-RT: 1146.3436–7.36) in the biosynthesis of unsaturated fatty acids pathway. The results are summarized in Supplementary Table S3 and Supplementary Figs. S1-S5.

Metabolic analysis of DHM samples retrieved a total of 466 and 379 metabolic features in the ESI^+ and ESI mode, respectively, after peak detection, deconvolution, integration, alignment, within-batch effect correction, and clean-up.

Fig. 2 shows the resulting network for the joint analysis of lipidomic (1033 features with *p*-value < 0.01; Wilcoxon signed-rank test) and metabolomic (35 features with *p*-value < 0.01; Wilcoxon signed-rank test) data (Pirhaji et al., 2016), in which the relation between detected features and hidden metabolites and proteins is displayed. Several features associated to the steroid hormone biosynthesis (i.e., tetrahydrocortisol, tetrahydrocorticosterone, 7a-hydroxydehydroepiandrosterone, tetrahydrodeoxycorticosterone) and the biosynthesis of unsaturated fatty acids (i.e., palmitic acid, eicosapentaenoic acid) were identified, in agreement with results shown in Table 1 and Supplementary Table S3. In addition, several metabolites associated to the steroid biosynthesis (i.e., 4,4-dimethylcholesta-8,14,24-trienol, 5-dehydroavenasterol, 7-dehydrodesmosterol, 7-dehydrocholesterol, lathosterol) and purine metabolism (i.e., guanosine, adenine, hypoxanthine, deoxyguanosine) pathways were significantly altered.

A recent study of the impact of the type of nutrition on the urinary metabolome of preterm infants has shown a significant alteration of the steroid hormone biosynthesis pathway associated to the intake of fresh

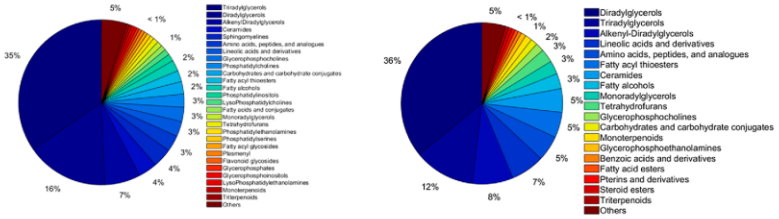


Fig. 1. Distribution by sub-classes of the metabolites annotated using HMDB/METLIN or LipidBlast spectral libraries and MetDNA detected in HM (left) and with significantly different mean values (*t*-test, FDR-adjusted p-value < 0.05) and |log2(FC)| >1 before and after pasteurization (right).

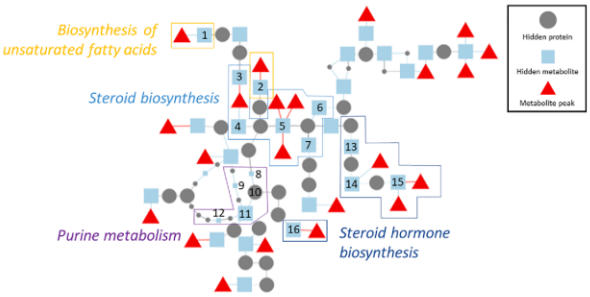


Fig. 2. PIUMet network showing metabolic pathways altered by pasteurization. Note: Pathways were highlighted if three or more hidden metabolites from the same pathway were interconnected as well as pathways those that were identified earlier by pathway analysis shown in Table 1 (biosynthesis of unsaturated fatty acids). Hidden metabolite / hidden protein #1: Palmitic acid; #2: Eicosapentaenoic acid; #3: 4,4-Dimethylcholesta-8,14,24-trienal; #4: 5-Dehydroavenasterol; #5: 7-Dehydrodesmosterol; #6: 7-Dehydrocholesterol; #7: Lathosterol; #8: Deoxyguanosine; #9: Guanosine; #10: purine nucleoside phosphorylase (PNP) protein; #11: Adenine; #11: Hypoxanthine; #12: 7a-Hydroxydehydroepiandrosterone; #13: Tetrahydrocorticosterone; #14: Tetrahydrocortisol; #15: Tetrahydrodeoxy corticosterone.

OMM or pasteurized DHM (Piñeiro-Ramos et al., 2021). This result suggested either that those steroid hormones present in HM may have a significant influence in the activity of the steroid hormone biosynthesis pathway in preterm infants, either directly or via the modification of gut-microbiota crosstalk, or that ingested levels of those compounds differ between pasteurized and fresh HM. In the present study, we focused on the analysis of DHM samples before and after HoP. Previous results found no significant differences of the levels of cortisol and cortisone in HM before and after HoP (van der Voorn et al., 2017). However, in the present study, the alteration of structurally closely related compounds included in the steroid and steroid hormone biosynthesis pathways upon pasteurization was identified (see Table 1 and Fig. 2).

The variation of total fat and fatty acid composition upon pasteurization has been studied but there is considerable variation in the reported effects (Nessel, Khashu, & Dyall, 2019). A recent review on this topic concluded that the total fat content and total fatty acid composition of expressed HM was not generally influenced by storage and handling process as most changes were found below 10% and within the expected random methodological variation (Gao et al., 2019). On the contrary, in a recent study with more than four hundred DHM samples, Piemontese et al. (Piemontese et al., 2019) confirmed that pasteurization reduced macronutrient composition, especially in terms of lipids and protein. Vincent et al. (Vincent et al., 2020) attributed these differences to the adherence of disrupted milk fat globules to container surfaces and to whether thermal treatment took place in a laboratory environment or in industrial pasteurizers routinely used in HM banks. Thus, Fidler *et al.* (Fidler, Sauerwald, Koletzko, & Demmelmair, 1998) did not find significant changes of fatty acids in HM after HoP, but a trend toward slightly lower percentage values of some fatty acids, including eicosapentaenoic acid was found. This observation was in agreement with the results from PIUMet network analysis presented in this study (see Fig. 2).

Furthermore, alterations of the purine metabolism were detected. A significant increase in nucleotide monophosphates associated to HoP of HM, which could be produced from the break-down of polymeric nucleotides and/or nucleotide adducts was reported (Mateos-Vivas et al., 2015). Here, we observed changes in guanosine and deoxyguanosine, which participate as building blocks in the synthesis of oligonucleotides. In addition, in the network-based approach, the purine nucleoside phosphorylase (PNP) protein, which catalyses the phosphorolytic breakdown of the *N*-glycosidic bond in the beta-(deoxy)ribonucleoside molecules, with the formation of the corresponding free purine bases and pentose-1-phosphate, appeared altered (see Fig. 2).

Finally, no literature reports regarding changes of glycans in pasteurized HM have been found. Although HM oligosaccharides (HMOs) are reported to be unaffected by HoP (Hahn, Kim, Song, Park, & Kang, 2019), in this work, mucin type *O*-glycans appeared altered upon pasteurization. As for HMOs, it has been hypothesized that mucin type *O*-glycans may have a certain role in mucus barrier function by promoting mutualism with host microbiota (Bergstrom & Xia, 2013). Hence, additional studies on specific glycans are needed.

In the lipid ontology enrichment analysis performed in pursuit of functional alterations of the lipidomic profile of HM after pasteurization, 634 out of 786 (80.66%) annotated features were matched to LION entries. The LION enrichment graph with the four major branches (lipid classification, chemical and physical properties, function, and subcellular component) is shown in Fig. 3.

The terms "lipid droplet" and "lipid storage", belonging to the category "sub-cellular component" and "function", respectively, were enriched in pasteurized DHM. Inspection of the lipid species responsible of these LION-terms revealed the presence of TGs. It has been described that the heating process very likely causes the disruption of the fat globule membranes. In fact, a significant decreased in fat globule size and, hence, an increase in the overall surface area upon pasteurization have been previously reported (Lopez, Cauty, & Guyomarc'h, 2015). This surface increase of the TG/water interface jointly with potential changes in the interface composition might play an important role with respect to chemical and enzymatic reactions that take place at the interface. In this work, the terms "membrane component" and "lipid-mediated signaling", belonging to the ontology root "function", were down-regulated by pasteurization. The first term is associated with lipids primarily regarded as structural components of lipid bilayers (i.e., DGs, glycerophosphocholines, phosphosphingolipids, and lyso-glycerophospholipids (GPs)) indicating a decay of membrane lipids during pasteurization. In this study, a total of 258 identifiers were ascribed to this LION-term, distributed as DGs (59%), GPs (24%), sphingomyelins (12%), and lyso-GPs (5%). The term "lipid-mediated signaling", associated with lipids implicated in signaling, such as DGs, monoradylglycerols, and lyso-GPs, was also down-regulated because of processing. Changes in the oxylipin composition during HoP have been previously observed (Pitino et al., 2019). This potential functional alteration should be addressed in future studies, as it could potentially be important for cellular processes, and especially relevant for the brain development of preterm infants. Lipids related to the term "negative intrinsic curvature" were downregulated in pasteurized DHM. LION assumes curvature to be predominantly head-group-dependent and neglects fatty acid composition. In this work, 184 identifiers, distributed in

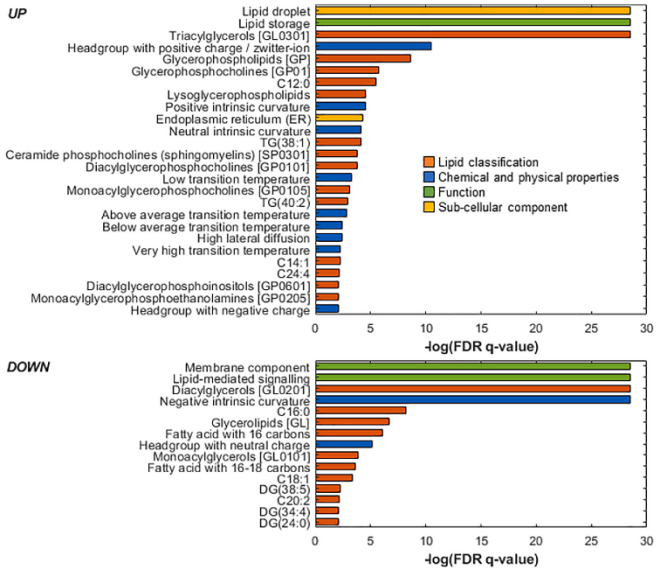


Fig. 3. Lipid ontology enrichment analysis of the pasteurization process performed in the "ranking mode" (DHM-Post *vs* DHM-Pre). Only the 40 most enriched LION-term have been represented.

DGs (80%), ceramides (15%) and GPs (4%), were ascribed to this term indicating a change in physico-chemical properties of HM upon pasteurization.

Lipid ontology enrichment analysis underpinned the alteration of different lipid classes upon pasteurization as discussed earlier (see also Fig. 1). In addition, it allowed to draw conclusions regarding the consequences of those changes that were associated to the physical and chemical properties (i.e., negative intrinsic curvature), sub-cellular components (i.e., lipid droplet) and function (i.e., lipid storage, membrane component, and lipid-mediated signaling) of altered lipids.

3.2. Fatty acid analysis

A method for the quantification of derivatized free fatty acids after acidic hydrolysis of lipids in HM was developed and validated following the FDA guidelines for bioanalytical method validation. Supplementary Table S4 summarizes the employed concentration intervals, which were selected considering the wide expected inter- and intra-individual variability. Intra and inter-day accuracy and precision of the method varied between 80% and 121%, and 1.0% and 20%, respectively, in standards and between 80% and 122%, and 1.0% and 14%, respectively, in spiked samples. Selectivity was confirmed and no carry-over was detected.

A total of 29 fatty acids from the 36 fatty acids included in the analytical method were found at higher levels than the lower limit of quantification and were successfully quantified in HM samples. Mean concentrations of five fatty acids were unaffected by pasteurization, two fatty acids showed an increase, and 22 fatty acids showed a statistically significant decrease (median decrease 10% (14% IQR)) (see Table 2).

Besides, the relative decrease in the concentration was higher for saturated fatty acids (SFAs) (25%), followed by polyunsaturated fatty acids (PUFAs) (18%), long-chain fatty acids (LCFAs) (15%), and monounsaturated fatty acids (MUFAs) (12%) (see Fig. 4).

The observed reduction in the fatty acids concentrations was in agreement with previous results (Gao et al., 2019). The small decrease in concentration found in this work might be, at least partially, responsible for the controversy in literature reports (Nessel et al., 2019) on the effect of pasteurization on fatty acid concentration in HM and underlines the importance of method validation for enabling the detection of subtle changes between groups.

4. Conclusions

This study provides the first comprehensive assessment of the impact of pasteurization on the HM lipidome and metabolome. Results identified affected metabolites that were related to steroids (i.e., steroid

Table 2
Individual and total content of fatty acids in DHM samples before (DHM-Pre) and after (DHM-Post) pasteurization.

Fatty acid	Median (IQR) (mM)		*p*-value
	DHM-Pre	DHM-Post	
Hexanoate (C6:0)	56 (20–95)	35 (18–75)	0.0002
Octanoate (C8:0)	41 (19–55)	29 (11–47)	0.0005
Decanoate (C10:0)	42 (22–53)	35 (15–45)	0.0002
Undecanoate (C11)	< LOQ	< LOQ	–
Laurate (C12:0)	14 (10–18)	10 (9–16)	0.02
Tridecanoate	< LOQ	< LOQ	–
Myristate (C14:0)	4 (3–4)	3 (2–4)	0.013
Myristoleate (C14:1)	0.10 (0.06–0.15)	0.08 (0.05–0.14)	0.0005
Pentadecanoate (C15:0)	0.10 (0.08–0.14)	0.09 (0.06–0.13)	0.0007
cis 10 Pentadecenoic (C15:1)	< LOQ	< LOQ	–
Palmitate (C16:0)	9 (7–10)	8 (5–10)	0.0005
Palmitoleate (C16:1)	0.9 (0.6–1.2)	0.9 (0.5–1.0)	0.0012
Heptadecanoate (C17:0)	0.3 (0.2–0.3)	0.3 (0.2–0.3)	0.003
cis 10 Heptadecenoic (C17:1)	0.2 (0.15–0.3)	0.2 (0.14–0.2)	0.0002
Stearate (C18:0)	3 (2–5)	3 (3–4)	0.0002
Oleic (C18:1n9c)	30 (26–35)	26 (18–30)	0.002
Elaidic (C18:1n9t)	0.8 (0.7–1.0)	0.8 (0.5–0.9)	0.013
Linoleic (C18:2n6c)	6 (5–8)	5 (4–7)	0.0002
Linolelaidic (C18:2n6t)	5 (1.1–8)	4 (0.8–7)	0.0007
gamma-Linolenic (C18:3n6)	0.2 (0.15–0.3)	0.2 (0.15–0.2)	0.0012
Linolenic (C18:3n3)	0.4 (0.4–0.5)	0.4 (0.3–0.4)	0.0002
Eicosanoic (C20:0)	0.3 (0.3–0.3)	0.3 (0.3–0.3)	0.02
cis-11-Eicosenoic (C20:1)	0.2 (0.13–0.2)	0.15 (0.13–0.2)	0.02
cis 11,14 Eicosadienoic (C20:2)	0.09 (0.05–0.2)	0.07 (0.04–0.2)	0.02
cis 8,11,14 Eicosatrienoic (C20:3n6)	0.4 (0.3–0.5)	0.4 (0.3–0.4)	0.05
Arachidonic (C20:4n6)	0.3 (0.3–0.4)	0.3 (0.3–0.4)	0.002
Henericosanoate (C21:0)	< LOQ	< LOQ	–
cis 11,14,17 Eicosatrienoic (C20:3n3)	0.15 (0.11–0.2)	0.15 (0.11–0.2)	0.13
cis 5,8,11,14,17 Eicospentaenoic (C20:5n3)	< LOQ	< LOQ	–
Docosanoate (C22:0)	0.5 (0.4–0.5)	0.5 (0.4–0.5)	0.02
Erucic acid (C22:1)	0.006 (0.005–0.13)	0.007 (0.004–0.13)	0.13
cis 13,16 Docosadienoic (C22:2)	0.3 (0.2–0.3)	0.3 (0.2–0.3)	0.5
Tricosanoate (C23:0)	< LOQ	< LOQ	–
cis-4,7,10,13,16,19-Docosahexaenoic (C22:6n3)	0.5 (0.4–0.7)	0.5 (0.4–0.6)	0.008
Lignocerate (C24:0)	0.2 (0.2–0.3)	0.2 (0.2–0.3)	0.05
Nervonic acid (C24:1)	< LOQ	< LOQ	–
TOTAL	230 (120–290)	169 (104–252)	0.0002

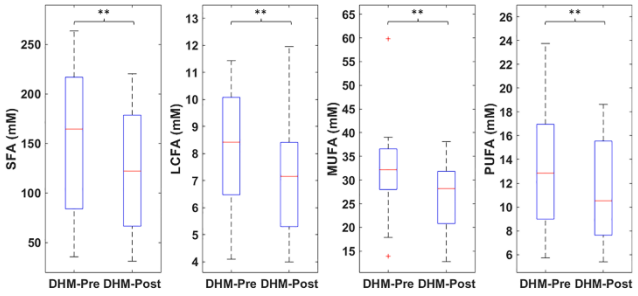


Fig. 4. Total saturated fatty acids (SFA), long chain fatty acids (LCFA), monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA) of DHM-samples before (DHM-Pre) and after (DHM-Post) pasteurization. **p-value < 0.01, one-tailed Wilcoxon signed-rank test.

biosynthesis and steroid hormone biosynthesis) as well as fatty acids (i.e., biosynthesis of unsaturated fatty acids and linoleic acid metabolism) metabolic pathways. Earlier studies solely focusing on cortisone and cortisol did not find alterations due to HoP. In this work, however, the levels of structurally related metabolites were affected by the thermal treatment. The present work furthermore demonstrates that HoP has an impact not only on the composition, but also on the physico-chemical properties, cellular components, and the functionality of lipids. Finally, the concentrations of the 76% of the analyzed fatty acids were altered after pasteurization, with a median decrease in the relative concentration of 10%. Results obtained encourage further research into the analysis of the biological relevance and impact of the observed changes in composition and functionality of HM components.

5. Funding sources

This work was supported by the *Instituto de Salud Carlos III*, Spain [grant numbers CD19/00176 and CP16/00034]; the Ministry of Science and Innovation, Spain [grant number LIC2018-036209-I], *Generalitat Valenciana* [project number GV/2021/186], and the European Union's Horizon 2020 Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110].

CRediT authorship contribution statement

Isabel Ten-Doménech: Conceptualization, Supervision, Funding acquisition, Visualization, Writing – original draft, Software, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – review & editing. **Victoria Ramos-Garcia:** Visualization, Writing – original draft, Software, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – review & editing. **Marta Moreno-Torres:** Investigation, Methodology, Writing – review & editing. **Anna Parra-Llorca:** Software, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – review & editing. **María Gormaz:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing – review & editing. **Máximo Vento:** Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing – review & editing. **Julia Kuligowski:** Conceptualization, Supervision, Funding acquisition, Visualization, Writing – original draft, Software, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – review & editing. **Guillermo Quintás:** Conceptualization, Supervision, Funding acquisition, Software, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.132581.

References

Bergstrom, K. S. B., & Xia, L. (2013). Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology, 23*(9), 1026–1037. https://doi.org/10.1093/glycob/cwt045

Boquien, C.-Y. (2018). Human Milk: An Ideal Food for Nutrition of Preterm Newborn. *Frontiers in Pediatrics, 6*. https://doi.org/10.3389/fped.2018.00295

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics, 14*(6), 72. https://doi.org/10.1007/s11306-018-1367-3

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., … Karp, P. D. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 44*(D1), D471–D480. https://doi.org/10.1093/nar/gkv1164

Chang, C. -C., & Lin, C. -J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology 2*(3), 27:1-27:27. doi:10.1145/1961189.1961199.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., … Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research, 46*(W1), W486–W494. https://doi.org/10.1093/nar/gky310

Colaizy, T. T. (2021). Effects of milk banking procedures on nutritional and bioactive components of donor human milk. *Seminars in Perinatology, 45*(2), Article 151382. https://doi.org/10.1016/j.semperi.2020.151382

Cruz-Hernandez, C., Goeuriot, S., Giuffrida, F., Thakkar, S. K., & Destaillats, F. (2013). Direct quantification of fatty acids in human milk by gas chromatography. *Journal of Chromatography A, 1284*, 174–179. https://doi.org/10.1016/j.chroma.2013.01.094

Fidler, N., Sauerwald, T. U., Koletzko, B., & Demmelmair, H. (1998). Effects of Human Milk Pasteurization and Sterilization on Available Fat Content and Fatty Acid Composition. *Journal of Pediatric Gastroenterology and Nutrition, 27*(3), 317–322.

Food and Drug Administration (FDA). Guidance for Industry: Bioanalytical Method Validation. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine. (2018). 44.

Gao, C., Miller, J., Middleton, P. F., Huang, Y.-C., McPhee, A. J., & Gibson, R. A. (2019). Changes to breast milk fatty acid composition during storage, handling and processing: A systematic review. *Prostaglandins, Leukotrienes, and Essential Fatty Acids, 146*, 1–10. https://doi.org/10.1016/j.plefa.2019.04.008

Hahn, W.-H., Kim, J., Song, S., Park, S., & Kang, N. M. (2019). The human milk oligosaccharides are not affected by pasteurization and freeze-drying. *The Journal of Maternal-Fetal & Neonatal Medicine: The Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians, 32*(6), 985–991. https://doi.org/10.1080/14767058.2017.1397122

Horta, D., Moreno-Torres, M., Ramírez-Lázaro, M. J., Lario, S., Kuligowski, J., Sanjuan-Herráez, J. D., … Calvet, X. (2020). Analysis of the Association between Fatigue and the Plasma Lipidomic Profile of Inflammatory Bowel Disease Patients. *Journal of Proteome Research*. https://doi.org/10.1021/acs.jproteome.0c00462

Hutchins, P. D., Russell, J. D., & Coon, J. J. (2018). LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. *Cell Systems, 6*(5), 621–625.e5. https://doi.org/10.1016/j.cels.2018.03.011

Kessner, D., Chambers, M., Burke, R., Agus, D., & Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics (Oxford, England), 24*(21), 2534–2536. https://doi.org/10.1093/bioinformatics/btn323

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Analytical Chemistry, 84*(1), 283–289. https://doi.org/10.1021/ac202450g

Kuligowski, J., Sánchez-Illana, Á., Sanjuán-Herráez, D., Vento, M., & Quintás, G. (2015). Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC). *The Analyst, 140*(22), 7810–7817. https://doi.org/10.1039/c5an01638j

Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., … Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. *PLoS Computational Biology, 9*(7). https://doi.org/10.1371/journal.pcbi.1003123

Lopez, C., Cauty, C., & Guyomarc'h, F. (2015). Organization of lipids in milks, infant milk formulas and various dairy products: Role of technological processes and potential impacts. *Dairy Science & Technology, 95*(6), 863–893. https://doi.org/10.1007/s13594-015-0263-0

Martínez-Sena, T., Luongo, G., Sanjuan-Herráez, D., Castell, J. V., Vento, M., Quintás, G., & Kuligowski, J. (2019). Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis. *Scientific Reports, 9*(1), 1–9. https://doi.org/10.1038/s41598-019-46371-w

Mateos-Vivas, M., Rodríguez-Gonzalo, E., Domínguez-Álvarez, J., García-Gómez, D., Ramírez-Bernabé, R., & Carabias-Martínez, R. (2015). Analysis of free nucleotide monophosphates in human milk and effect of pasteurisation or high-pressure processing on their contents by capillary electrophoresis coupled to mass spectrometry. *Food Chemistry, 174*, 348–355. https://doi.org/10.1016/j.foodchem.2014.11.051

Miller, J., Tonkin, E., Damarell, R. A., McPhee, A. J., Suganuma, M., Suganuma, H., … Collins, C. T. (2018). A Systematic Review and Meta-Analysis of Human Milk Feeding and Morbidity in Very Low Birth Weight Infants. *Nutrients, 10*(6), 707. https://doi.org/10.3390/nu10060707

Molenaar, M. R., Jeucken, A., Wassenaar, T. A., van de Lest, C. H. A., Brouwers, J. F., & Helms, J. B. (2019). LION/web: A web-based ontology enrichment tool for lipidomic data analysis. *GigaScience, 8*(6). https://doi.org/10.1093/gigascience/giz061

Nessel, I., Khasbiu, M., & Dyall, S. G. (2019). The effects of storage conditions on long-chain polyunsaturated fatty acids, lipid mediators, and antioxidants in donor human milk—A review. *Prostaglandins, Leukotrienes, and Essential Fatty Acids, 149*, 8–17. https://doi.org/10.1016/j.plefa.2019.07.009

Parra-Llorca, A., Gormaz, M., Alcántara, C., Cernada, M., Nuñez-Ramiro, A., Vento, M., & Collado, M. C. (2018). Preterm Gut Microbiome Depending on Feeding Type: Significance of Donor Human Milk. *Frontiers in Microbiology*, 9. https://doi.org/10.3389/fmicb.2018.01376

Parra-Llorca, A., Gormaz, M., Sánchez-Illana, Á., Piñeiro-Ramos, J. D., Collado, M. C., Serna, E., … Vento, M. (2019). Does pasteurized donor human milk efficiently protect preterm infants against oxidative stress? *Antioxidants & Redox Signaling*. https://doi.org/10.1089/ars.2019.7821

Piemontese, P., Mallardi, D., Liotto, N., Tabasso, C., Menis, C., Perrone, M., … Mosca, F. (2019). Macronutrient content of pooled donor human milk before and after Holder pasteurization. *BMC Pediatrics, 19*(1), 58. https://doi.org/10.1186/s12887-019-1427-5

Piñeiro-Ramos, J. D., Parra-Llorca, A., Ten-Doménech, I., Gormaz, M., Ramón-Beltrán, A., Cernada, M., … Vento, M. (2021). Effect of donor human milk on host-gut microbiota and metabolic interactions in preterm infants. *Clinical Nutrition (Edinburgh, Scotland), 40*(3), 1296–1309. https://doi.org/10.1016/j.clnu.2020.08.013

Pirhaji, L., Milani, P., Leidl, M., Curran, T., Avila-Pacheco, J., Clish, C. B., … Fraenkel, E. (2016). Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nature Methods, 13*(9), 770–776. https://doi.org/10.1038/nmeth.3940

Pitino, M. A., Alashmali, S. M., Hopperton, K. E., Unger, S., Pouliot, Y., Doyen, A., … Bazinet, R. P. (2019). Oxylipin concentration, but not fatty acid composition, is altered in human donor milk pasteurised using both thermal and non-thermal techniques. *The British Journal of Nutrition, 122*(1), 47–55. https://doi.org/10.1017/S0007114519000916

Poulimeneas, D., Bathrellou, E., Antonogeorgos, G., Mamalaki, E., Kouvari, M., Kuligowski, J., … NUTRISHIELD Consortium. (2021). Feeding the preterm infant: An overview of the evidence. *International Journal of Food Sciences and Nutrition, 72*(1), 4–13. https://doi.org/10.1080/09637486.2020.1754352

Quigley, M., Embleton, N. D., & McGuire, W. (2019). Formula versus donor breast milk for feeding preterm or low birth weight infants. *The. Cochrane Database of Systematic Reviews, 7*, CD002971. https://doi.org/10.1002/14651858.CD002971.pub5

Shen, X., Wang, R., Xiong, X., Yin, Y., Cai, Y., Ma, Z., … Zhu, Z.-J. (2019). Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. *Nature Communications, 10*(1), 1516. https://doi.org/10.1038/s41467-019-09550-x

Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem., 84*(11), 5035–5039.

Ten-Doménech, I., Martínez-Sena, T., Moreno-Torres, M., Sanjuan-Herráez, J. D., Castell, J. V., Parra-Llorca, A., … Kuligowski, J. (2020). Comparing Targeted vs. Untargeted MS2 Data-Dependent Acquisition for Peak Annotation in LC-MS Metabolomics. *Metabolites, 10*(4), 126. https://doi.org/10.3390/metabo10040126

van der Voorn, B., de Waard, M., Dijkstra, I. R., Heijboer, A. C., Rotteveel, J., van Goudoever, J. B., & Finken, M. J. J. (2017). Stability of Cortisol and Cortisone in Human Breast Milk During Holder Pasteurization. *Journal of Pediatric Gastroenterology and Nutrition, 65*(6), 658–660. https://doi.org/10.1097/MPG.0000000000001678

Villaseñor, A., García-Pérez, I., García, A., Posma, J. M., Fernández-López, M., Nicholas, A. J., … Barbas, C. (2014). Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach. *Analytical Chemistry, 86*(16), 8245–8252. https://doi.org/10.1021/ac5018534

Vincent, M., Ménard, O., Etienne, J., Ossemond, J., Durand, A., Buffin, R., … Penhoat, A. (2020). Human milk pasteurisation reduces pre-lipolysis but not digestive lipolysis and moderately decreases intestinal lipid uptake in a combination of preterm infant in vitro models. *Food Chemistry, 329*, Article 126927. https://doi.org/10.1016/j.foodchem.2020.126927

Wesolowska, A., Sinkiewicz-Darol, E., Barbarska, O., Bernatowicz-Łojko, U., Borszewska-Kornacka, M. K., & van Goudoever, J. B. (2019). Innovative Techniques of Processing Human Milk to Preserve Key Components. *Nutrients, 11*(5). https://doi.org/10.3390/nu11051169

World Health Organization. (2017). *Guideline: Protecting, promoting and supporting breastfeeding in facilities providing maternity and newborn services*. http://www.ncbi.nlm.nih.gov/books/NBK487819/.

Chapter 18

Isolation and Lipidomic Screening of Human Milk Extracellular Vesicles

Victoria Ramos-Garcia, Isabel Ten-Doménech, Abel Albiach-Delgado, Marta Gómez-Ferrer, Pilar Sepúlveda, Anna Parra-Llorca, Laura Campos-Berga, Alba Moreno-Giménez, Guillermo Quintás, and Julia Kuligowski

Abstract

Extracellular vesicles (EVs) are secreted by cells and can be found in biological fluids (e.g., blood, saliva, urine, cerebrospinal fluid, and milk). EV isolation needs to be optimized carefully depending on the type of biofluid and tissue. Human milk (HM) is known to be a rich source of EVs, and they are thought to be partially responsible for the benefits associated with breastfeeding. Here, a workflow for the isolation and lipidomic analysis of HM-EVs is described. The procedure encompasses initial steps such as sample collection and storage, a detailed description for HM-EV isolation by multistage ultracentrifugation, metabolite extraction, and analysis by liquid chromatography coupled to mass spectrometry, as well as data analysis and curation.

Key words Extracellular vesicles, Exosomes, Human milk (HM), Lipidomics, Liquid chromatography–mass spectrometry (LC–MS)

1 Introduction

Eukaryotic and prokaryotic cells release a variety of nano- and micron-sized membrane-containing vesicles into their extracellular environment, which are collectively referred to as extracellular vesicles (EVs). EVs can be harvested from cell culture supernatants and from all body fluids including plasma, saliva, urine, cerebrospinal fluid, and human milk (HM). HM is known to be a rich source of EVs, with early milk containing a greater EV concentration compared to mature milk.

The physiological purpose of generating EVs remains largely unknown, and questions surrounding the function of EVs are mostly focused on understanding the fate of their constituents

Raúl González-Domínguez (ed.), *Mass Spectrometry for Metabolomics*,
Methods in Molecular Biology, vol. 2571, https://doi.org/10.1007/978-1-0716-2699-3_18,

and the phenotypic and molecular alterations that they induce in recipient cells. The effects of EVs on recipient cells can vary due to the expression of different cell surface receptors, resulting in an array of possible biological functions including the induction of cell survival, apoptosis, and immunomodulation. In addition, recent studies indicate a functional, targeted, mechanism-driven accumulation of specific cellular components such as RNAs, small RNAs, nucleic acids, lipids, proteins, and metabolites in EVs, suggesting that they have a role in regulating intercellular communication [1]. Furthermore, encapsulation in EVs confers protection against enzymatic and nonenzymatic degradation of cargos and provides a pathway for cellular uptake of cargos by endocytosis of EVs [2, 3].

The composition of HM-EVs is still an open question. The richness of microRNAs (miRNAs) within HM-EVs was first discovered, and the dynamics during lactation stages was surveyed [4, 5]. Proteins have been studied in HM-EVs [6, 7], showing an enrichment of pathways involved in early-life immunity, as well as intestinal cell proliferation and migration. Although metabolites are known to be part of the EV cargo [8–10], literature reports providing insight into the metabolite cargo or lipid composition of HM-EVs are scarce. Hence, we developed a pipeline for the isolation and liquid chromatography–mass spectrometry (LC–MS)-based lipidomic screening of HM-EVs [11].

2 Materials

Prepare all solutions using ultrapure water (Q-POD® system, Merck KGaA, Darmstadt, Germany) and analytical grade reagents. Prepare all reagents at room temperature.

1. Phosphate-buffered saline (PBS) solution: 10 mM phosphate buffer, 2.7 mM potassium chloride, and 137 mM sodium chloride solution, pH 7.4. Dissolve one commercially available phosphate-buffered saline tablet in 200 mL of ultrapure water at 25 °C (*see* **Note 1**). Filter this solution with a 0.40 μm syringe filter. Store at 4 °C.
2. Internal standard (IS) mixture: Prepare individual 5 mM stock solutions of internal standards by weighing, e.g., 1.46 mg of oleic acid-D_9 and 1.49 mg of prostaglandin $F_{2\alpha}$-D_4 and dissolving them in 1 mL of ultrapure water. Put 976 μL of ultrapure water into a microcentrifuge tube. Add 16 and 7.8 μL of oleic acid-D_9 and prostaglandin $F_{2\alpha}$-D_4 5 mM stock solutions, respectively (resulting concentrations: 80 and 39 μM, respectively) (*see* **Note 2**). Aliquot to avoid freeze–thaw cycles and store at −20 °C.
3. Methanol (MeOH). Store at room temperature.

4. Methyl tert-butyl ether (MTBE). Store at room temperature.
5. Solution A: Isopropanol (IPA):MeOH:H_2O (5:1:4, v:v:v), 5 mM CH_3COONH_4, 0.1% formic acid (FA). Add 500 mL of IPA and 100 mL of MeOH into a 1-L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 20 mL of ultrapure water, and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with ultrapure water. Store at 4 °C.
6. Solution B: IPA:H_2O (99:1, v:v), 5 mM CH_3COONH_4, 0.1% FA. Add 900 mL of IPA into a 1-L volumetric flask. Do not fill to the calibration mark yet. Weigh 0.39 g of CH_3COONH_4, dissolve it in 9 mL of ultrapure water, and add this solution into the volumetric flask. Add 1 mL of pure FA into the volumetric flask. Fill to the calibration mark with IPA. Store at 4 °C.
7. Solution A:B (1:1, v:v): mix 10 mL of solution A with 10 mL of solution B. Store at 4 °C.

For the performance of untargeted screening experiments, an LC (preferably: ultra-performance LC)-high-resolution MS instrument and a reversed-phase chromatographic column should be employed. The following protocol was developed using a 1290 Infinity HPLC system from Agilent Technologies (CA, USA) equipped with a UPLC BEH C18 column (50 × 2.1 mm, 1.7 μm, Waters, Wexford, Ireland) coupled to an Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 HM Sample Collection and Storage (See Note 3)

1. Clean the hands with water and soap for at least 15 s followed by drying with a clean towel.
2. Sterilize hands with hand sanitizer.
3. Clean and sterilize the milk pump (*see* **Note 4**).
4. Clean the skin area that gets into contact with the milk pump with water and dry it with sterile gauzes (*see* **Note 5**).
5. Extract the milk with the milk pump into clean sterile bottles following the manufacturer's instructions (*see* **Note 6**).
6. Manual shake gently during 30 s for sample homogenization.
7. Put the sample into 50-mL plastic Falcon tubes using a 25-mL plastic pipette (*see* Fig. 1a).

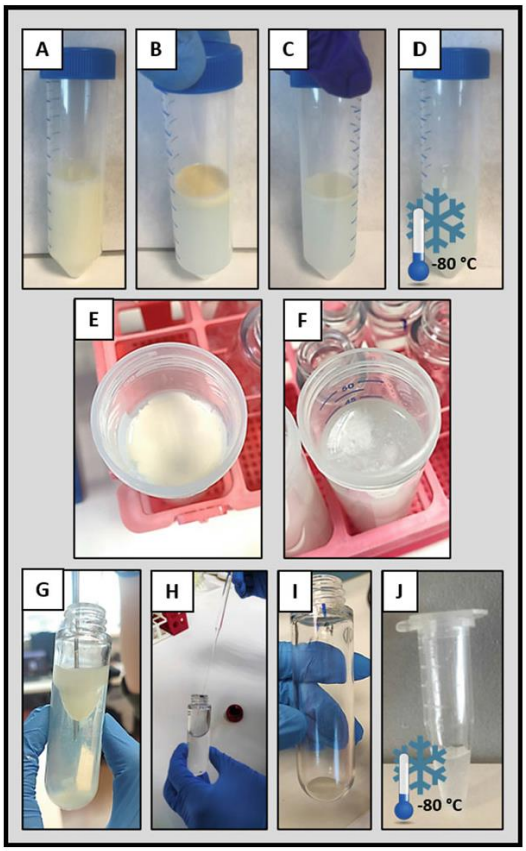


Fig. 1 Isolation of extracellular vesicles (EVs) from human milk (HM). (**a**) Raw HM sample. (**b** and **c**) HM sample after consecutive centrifugation steps with fat layer on the top. (**d**) Partially defatted HM sample for storage at −80 °C prior to EV isolation. (**e**) HM sample after first centrifugation step with remaining fat layer on the top. (**f**) Skimmed HM sample. (**g**) Supernatant aspiration after the first ultracentrifugation (10,000 rpm, 1 h, 4 °C) (protein pellet is discarded). (**h**) Supernatant removal after the second ultracentrifugation (30,000 rpm, 2 h, 4°) with the pellet containing EVs at the bottom. (**i**) Pellet containing EVs. (**j**) Reconstituted EVs in phosphate-buffered saline for storage at −80 °C until use

8. Centrifuge at 3000 × *g* for 10 min at 22 °C for cream, cell, and platelet removal [12, 13].
9. Remove and discard the cream layer with a spatula (*see* **Note 7** and Fig. 1b).
10. Transfer the liquid into a new 50-mL plastic Falcon tube.
11. Repeat **steps 8** and **9** (*see* Fig. 1c) and store milk at −80 °C until further processing (*see* Fig. 1d).

3.2 Isolation of EVs from HM

1. Thaw HM samples at 4 °C overnight by putting them in the refrigerator.
2. Centrifuge at 3000 × *g* for 10 min at 4 °C to remove the remaining milk fat and milk fat globules.
3. Remove the upper fat layer with a spatula (*see* **Note 7**) and discard (*see* Fig. 1e and f).
4. Transfer the liquid into a 25-mL polycarbonate bottle appropriate for ultracentrifugation.
5. Continue filling the ultracentrifuge tube with PBS until an estimated volume of 25 mL is reached and make sure all tubes have the same weight (including the caps) (*see* **Note 8**). This step is crucial as any ultracentrifuge can easily become unbalanced if equal masses are not located opposite to each other in the rotor.
6. Ultracentrifuge at 10,000 rpm for 1 h at 4 °C to pellet cellular debris, large size vesicles, and protein aggregates using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti rotor (Indianapolis, United States) or similar.
7. Collect the supernatant with a 25-mL plastic pipette or a syringe with a needle (*see* **Note 9**) and transfer it to a new 25-mL ultracentrifuge tube (*see* Fig. 1g).
8. Repeat **steps 5** and **6**.
9. Collect the supernatant into a different 25-mL ultracentrifuge tube, filtering it with a 0.45-μm syringe filter (*see* **Note 10**).
10. Repeat **step 5**.
11. Ultracentrifuge at 30,000 rpm for 2 h at 4 °C to pellet HM EVs using the same centrifuge rotor as in **step 6**.
12. Slowly aspirate and discard the supernatant using, e.g., a pipette (*see* Fig. 1h).
13. Wash and reconstitute the remaining pellet with 25 mL of PBS (*see* **Note 11**).
14. Repeat **steps 5** and **11**–**13** for a second ultracentrifugation.
15. Repeat **steps 5** and **11** for a third ultracentrifugation.
16. Slowly aspirate the PBS supernatant and save it to be later used as a blank. Store at −80 °C.

17. Reconstitute the pellet containing the isolated EVs (*see* Fig. 1i) with 200 µL of PBS (*see* **Note 12**).
18. Aliquot the isolated EVs in microtubes for quality control (QC) assays such as bicinchoninic acid (BCA) assay for protein quantification, nanoparticle tracking analysis (NTA) for vesicle concentration and size determination, etc., and for LC–MS analysis. Store at −80 °C (*see* Fig. 1j).

3.3 Extraction of the HM-EV Lipid Fraction

Lipids and other polar metabolites are extracted from HM-EVs using a single-phase extraction procedure [14–16].

1. Thaw the isolated HM-EV suspension in PBS obtained in Subheading 3.2 on ice and vortex for sample homogenization.
2. Mix 45 µL of isolated EVs with 5 µL of IS mixture.
3. Sonicate for 2 min.
4. Add 175 µL of MeOH followed by 175 µL of MTBE (*see* **Note 13**).
5. Vortex for 30 s for protein precipitation and compound extraction.
6. Sonicate for 2 min to assist the release of metabolites from EVs during extraction.
7. Centrifuge at 4000 × *g* for 15 min at 4 °C.
8. Transfer 100 µL of supernatant containing the extracted lipids and metabolites to a different microtube.
9. Dry at 35 °C using a centrifugal vacuum concentrator (*see* **Note 14**).
10. Reconstitute in 100 µL of solution A:B (1:1).
11. Prepare a pooled QC sample by mixing 5 µL of each reconstituted sample extract.
12. Prepare a calibration blank by repeating **steps 1–10**, replacing the initial 45 µL of the sample with 45 µL of ultrapure water.
13. Prepare a procedural blank by repeating **steps 1–10**, replacing the initial 45 µL of the sample with 45 µL of aspirated PBS supernatant from **step 18** in Subheading 3.2.
14. Analyze by LC–MS or, alternatively, store at −80 °C for further analysis.

3.4 LC–MS Lipidomics Method (See Note 15)

1. Perform a system suitability check of the MS (e.g., resolution, accuracy, and sensitivity) and UPLC performances (e.g., retention time (RT) of standards, resolution, lack of contamination of the analytical system) using blanks, standard mixtures, and QC samples (*see* **Note 15**).
2. Thaw the samples and place them in the autosampler.

3. Equilibrate the system by injecting the QC sample repeatedly. MS/MS data for peak annotation is typically acquired at the beginning or end of the batch [16] (*see* **Note 16**).
4. Run sample sequence including QC samples, study samples, and blanks, acquiring data using a suitable ionization interface (e.g., positive and/or negative electrospray ionization) (*see* **Notes 17** and **18**).

3.5 Data Processing and Analysis

1. Perform an initial data quality check through the manual integration of IS and several (e.g., 5–10) endogenous compounds with different intensities and RTs, which are expected to be detected in the sample under study. Assess the stability of peak areas, *m/z* accuracy, RT, and chromatographic parameters (e.g., peak width, resolution) throughout the analytical sequence.
2. Convert raw data into mzXML format using, e.g., ProteoWizard (http://proteowizard.sourceforge.net/) (optional) and generate a peak table (*see* **Notes 19** and **20**).
3. Annotate lipids using MS/MS information and available spectral and/or in-house LC–MS libraries (*see* **Note 21** and Figs. 2a–c).
4. Identify and correct the within-batch effect (*see* **Note 22** and Fig. 2d).
5. Perform data cleanup: (i) filtering of features found in blanks and (ii) filtering of features with high relative standard deviation (RSD) detected in QC samples (*see* **Note 23**).
6. Check data quality (*see* **Note 24** and Fig. 2e).

4 Notes

1. Prepare this solution fresh each time.
2. It is convenient to immerse the pipette tip in the water when adding small volumes.
3. HM samples are usually collected by lactating mothers and not by the hospital staff. Hence, the staff must provide detailed instructions to the mothers to avoid sampling bias and procedural differences between the different HM samples collected within a study. We recommend establishing a Standard Operation Procedure for the collection of milk samples and their processing to avoid experimental bias between samples. Details regarding the sample collection procedure need to be indicated, including, e.g., extraction of foremilk, hindmilk, or full expression of one or both breasts; time of the day; time since the last feeding; and extraction method (i.e., manual expression, milk pump).

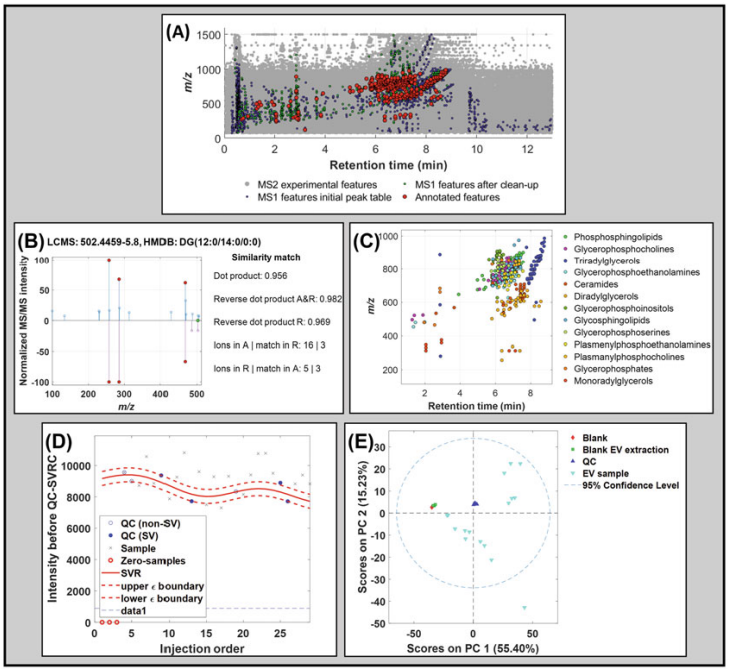


Fig. 2 LC–MS lipidomic data processing. (**a**) Distribution of MS1 (blue dots) and MS2 (gray dots) experimental features measured in extracellular vesicles (EVs) from human milk (HM) samples in the *m/z*-retention time space and MS1 experimental features and annotated features after applying cleanup steps (green and red dots, respectively). (**b**) Similarity match of a feature (*m/z*-retention time: 502.4489–5.8) annotated as DG (12:0/14:0/0:0) from the HMDB database. MS2 experimental spectrum (up). MS2 database spectrum (down). (**c**) Distribution in the *m/z*-retention time space of annotated features by sub-classes in EVs from HM samples. Only sub-classes containing at least five features are represented. (**d**) Intensity of a selected feature as a function of the injection order before intra-batch effect correction with the quality control-supported vector regression correction (QC-SVRC) approach. (**e**) Score plot of the principal component analysis model using the preprocessed data set. Note: *DG* diradylglycerol, *QC* quality control

4. We recommend the use of electric milk pumps as indicated in this protocol; however, if participants prefer manual expression, this could also be accommodated within the protocol. Milk pumps can be sterilized in the microwave by using dedicated plastic bags or in a water bath (15 min at 100 °C).

All parts need to be dried with sterile gauzes. The milk pump should be cleaned with hot water and soap immediately after every use.

5. No ointments should be applied to the skin before HM extraction. If they had been applied, clean the skin carefully with water and soap.
6. Take note of date, time, and extracted HM volume. When HM samples are manipulated, gloves must be used in order to avoid sample contamination.
7. Remove the upper fat layer with a spatula and aspirate the supernatant with a pipette. Be careful not to aspirate the remaining fat.
8. If the tubes do not have the same weight, compensate by adding PBS.
9. Pellet should not be aspirated.
10. More than one filter may be needed due to obturation.
11. Reconstitute the pellet in 1 mL of PBS and add the remaining 24 mL when the pellet is dissolved.
12. Do it gently to avoid foam formation.
13. The order of the solvents' addition is important.
14. This step usually takes 3–4 h.
15. Basic considerations for untargeted, LC–MS-based lipidomics and metabolomics studies should be implemented throughout the pipeline. For further information, the reader is referred to [17].
16. A set of approximately 5–10 QCs should be injected at the beginning of each batch for system conditioning and MS/MS data acquisition. MS/MS spectra need to be acquired for subsequent annotation purposes as described elsewhere [16]. Depending on the employed MS system, parameters need to be adjusted. For this example, a rate of 5 spectra/s in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, an automated selection of five precursor ions per cycle, and an exclusion window of 0.15 min after two consecutive selections of the same precursor were used.
17. Samples should be injected in randomized order and a QC sample should be intercalated every 5–10 samples for subsequent correction of the batch effect [17–19]. In our pipeline, we use a binary mobile phase gradient starting at 98% of solution A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) for 0.5 min followed by a linear gradient from 2 to 20% of solution B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v FA) for 3.5 min and from 20 to 95% v/v of solution B in 4 min; 95% v/v of solution B are

maintained for 1 min; return to initial conditions in 0.25 min and maintain for a total run time of 13 min. Column and autosampler are kept at 55 and 4 °C, respectively, an injection volume of 2 µL is used, and the flow rate is set to 400 µL/min. Full scan MS data in the range between 70 and 1500 *m/z* are acquired at a scan frequency of 5 Hz using the following parameters: gas T, 200 °C; drying gas, 14 L/min; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11/min. Mass reference standards are introduced into the source for on-the-fly automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (phthalic anhydride), 121.050873 (purine), and 922.009798 (HP-0921) *m/z* in ESI+, and 119.036 (purine) and 980.0163 (HP-0921, [M-H + CH_3COOH]-) *m/z* in ESI-, as references. ESI+ and ESI- analyses were carried out in independent batches, and between them, the instrument was cleaned and calibrated according to the manufacturer's guidelines.

18. Blank injections can affect the performance of the LC–MS system and hence, the position of the blanks within the analytical sequence should be chosen carefully [20].
19. We use MassHunter Workstation (version B.07.00) from Agilent for UPLC-TOFMS data acquisition and manual integration. In our workflow, peak detection, integration, deconvolution, alignment, and pseudospectra identification are carried out using XCMS [21] and CAMERA [22] in R 3.6.1. See reference [11] for an example of XCMS and CAMERA settings. Employed parameters depend on the LC–MS system and the method's performance and might need to be optimized in advance using, e.g., Isotopologue Parameter Optimization (IPO) [23]. Alternatively, peak tables can be generated using vendor and other available software such as MZmine2 or MS-DIAL. We recommend comparing manual vs. automatic integration for IS and endogenous metabolites considered in **step 1** to assess the accuracy of the automated integration.
20. If ESI+ and ESI- data were acquired in separate injections, perform data cleanup and filtering independently.
21. Metabolite annotation (Level ID: 2, putatively annotated compounds without matching to data for chemical standards acquired under the same experimental conditions) can be performed by matching experimentally acquired MS/MS spectra with the experimental MS/MS databases (e.g., HMDB, METLIN) in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [24]. Metabolite annotation can also exploit in-silico databases (e.g., LipidBlast [25]). Annotation and subsequent correction and cleanup steps can

be programmed in different programming languages for data science such as MATLAB (Mathworks Inc., Natick, MA, USA), R, or Python. Besides, metabolite annotation can also be performed using vendor software and other widely used tools such as MZMine2, MS-DIAL, GNPS (Global Natural Products Social Molecular Networking), or LipidDex [16, 26].

22. Features detected in QC samples can efficiently be used to monitor the instrument performance and correct within-batch effects [17]. We developed a nonparametric approach based on the QC-Supported vector regression correction (SVRC) approach employing a Radial Basis Function kernel for within batch effect removal [18, 19]. The selection of the tolerance threshold (ε), the penalty term applied to margin slack values (C), and the kernel width (γ) can be carried out using a grid search. QC-SVRC is robust to hyperparameter selection and a pre-selection of C and optimization of ε and γ using a grid search by leave-one-out cross-validation, and using the root mean square error of cross-validation (RMSECV) as target function allows a significant reduction in computation time. C can be selected for each LC–MS feature as the median value of the intensities observed in QC replicates. The ε search range is chosen based on the expected instrumental precision (e.g., 5–10% of the signal intensity). The γ search interval can be set to $[1, 10^5]$.
23. It is important to remove variables that are likely originating from background contaminants or cross-contamination, as well as those from unstable features to enhance performance of subsequent data analysis steps. We typically remove features with more than 20% missing values in QCs, those with RSD (QC)>20% after within-batch effect correction, and those for which the ratio between the median peak area values in QCs and blanks is lower than 6.
24. Different qualitative and quantitative tools are available for checking data quality [27], including principal component analysis, which is the most widely employed tool for this purpose. We recommend comparing the data quality of raw data and preprocessed data.

References

1. Kalluri R, LeBleu VS (2020) The biology, function, and biomedical applications of exosomes. Science 367
2. Altadill T, Campoy I, Lanau L et al (2016) Enabling metabolomics based biomarker discovery studies using molecular phenotyping of exosome-like vesicles. PLoS One 11: e0151339
3. Zempleni J, Aguilar-Lozano A, Sadri M et al (2017) Biological activities of extracellular vesicles and their cargos from bovine and human milk in humans and implications for infants. J Nutr 147:3–10
4. Liao Y, Du X, Li J et al (2017) Human milk exosomes and their microRNAs survive

digestion in vitro and are taken up by human intestinal cells. Mol Nutr Food Res 61: 1700082

5. Kahn S, Liao Y, Du X et al (2018) Exosomal MicroRNAs in milk from mothers delivering preterm infants survive in vitro digestion and are taken up by human intestinal cells. Mol Nutr Food Res 62:1701050
6. Yang M, Song D, Cao X et al (2017) Comparative proteomic analysis of milk-derived exosomes in human and bovine colostrum and mature milk samples by iTRAQ-coupled LC-MS/MS. Food Res Int 92:17–25
7. de la Torre Gomez C, Goreham RV, Bech Serra JJ et al (2018) "Exosomics"—a review of biophysics, biology and biochemistry of exosomes with a focus on human breast Milk. Front Genet 9:92
8. Royo F, Gil-Carton D, Gonzalez E et al (2019) Differences in the metabolite composition and mechanical properties of extracellular vesicles secreted by hepatic cellular models. J Extracell Vesicles 8:1575678
9. Clos-Garcia M, Loizaga-Iriarte A, Zuñiga-Garcia P et al (2018) Metabolic alterations in urine extracellular vesicles are associated to prostate cancer pathogenesis and progression. J Extracell Vesicles 7:1470442
10. Haraszti RA, Didiot M-C, Sapp E et al (2016) High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. J Extracell Vesicles 5:32570
11. Ramos-Garcia V, Ten-Doménech I, Moreno-Giménez A et al (2021) ATR-FTIR spectroscopy for the routine quality control of exosome isolations. Chemom Intell Lab Syst 104401
12. Zonneveld MI, Herwijnen MJC, Fernandez-Gutierrez MM et al (2021) Human milk extracellular vesicles target nodes in interconnected signalling pathways that enhance oral epithelial barrier function and dampen immune responses. J Extracell Vesicles:10
13. Witwer KW, Buzás EI, Bemis LT et al (2013) Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. J Extracell Vesicles 2:20360
14. Ten-Doménech I, Ramos-Garcia V, Piñeiro-Ramos JD et al (2020) Current practice in untargeted human milk metabolomics. Meta 10:43
15. Piñeiro-Ramos JD, Parra-Llorca A, Ten-Doménech I et al (2020) Effect of donor human milk on host-gut microbiota and metabolic interactions in preterm infants. Clin Nutr
16. Ten-Doménech I, Martínez-Sena T, Moreno-Torres M et al (2020) Comparing targeted vs. untargeted MS2 data-dependent acquisition for peak annotation in LC-MS metabolomics. Metabolites 10:126
17. Broadhurst D, Goodacre R, Reinke SN et al (2018) Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metab Off J Metab Soc 14:72
18. Kuligowski J, Sánchez-Illana Á, Sanjuán-Herráez D et al (2015) Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC). Analyst 140:7810–7817
19. Sánchez-Illana Á, Pérez-Guaita D, Cuesta-García D et al (2018) Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control – support vector regression. Anal Chim Acta 1026:62–68
20. Martínez-Sena T, Luongo G, Sanjuan-Herráez D et al (2019) Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis. Sci Rep 9: 1–9
21. Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787
22. Kuhl C, Tautenhahn R, Böttcher C et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84:283–289
23. Libiseller G, Dvorzak M, Kleb U et al (2015) IPO: a tool for automated optimization of XCMS parameters. BMC Bioinf 16:118
24. Salek RM, Steinbeck C, Viant MR et al (2013) The role of reporting standards for metabolite annotation and identification in metabolomic studies. GigaScience 2
25. Kind T, Liu K-H, Lee DY et al (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. Nat Methods 10: 755–758
26. Hutchins PD, Russell JD, Coon JJ (2018) LipiDex: an integrated software package for high-confidence lipid identification. Cell Syst 6:621-625.e5
27. Sánchez-Illana Á, Piñeiro-Ramos JD, Sanjuan-Herráez JD et al (2018) Evaluation of batch effect elimination using quality control replicates in LC-MS metabolite profiling. Anal Chim Acta 1019:38–48

Chemometrics and Intelligent Laboratory Systems 217 (2021) 104401

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

ATR-FTIR spectroscopy for the routine quality control of exosome isolations

Victoria Ramos-Garcia [a,1], Isabel Ten-Doménech [a,1], Alba Moreno-Giménez [a], María Gormaz [a,b], Anna Parra-Llorca [a], Alex P. Shephard [c], Pilar Sepúlveda [d], David Pérez-Guaita [e], Máximo Vento [a,b], Bernhard Lendl [f], Guillermo Quintás [g,h,*], Julia Kuligowski [a]

[a] Neonatal Research Group, Health Research Institute La Fe, Avda Fernando Abril Martorell 106, 46026, Valencia, Spain
[b] División of Neonatology, University & Polytechnic Hospital La Fe, Avda Fernando Abril Martorell 106, 46026, Valencia, Spain
[c] Nanoview Biosciences, Malvern Hills Science Park, Geraldine Road, WR14 3SZ, Malvern, United Kingdom
[d] Regenerative Medicine and Heart Transplantation Unit, Health Research Institute Hospital La Fe, Avda Fernando Abril Martorell 106, 46026, Valencia, Spain
[e] Department of Analytical Chemistry, University of Valencia, 50 Dr. Moliner Street, 46100, Burjassot, Valencia, Spain
[f] Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164, A 1060, Vienna, Austria
[g] Health and Biomedicine, Leitat Technological Center, Carrer de La Innovació, 2, 08225, Terrassa, Spain
[h] Analytical Unit, Health Research Institute La Fe, Avda Fernando Abril Martorell 106, 46026, Valencia, Spain

ARTICLE INFO

Keywords:
Attenuated total reflectance
Fourier transform infrared (ATR-FTIR)
Lipidomics
Exosomes
Human milk
Omics
Extracellular vesicles

ABSTRACT

Exosomes are nanosized vesicles containing specific cargos of DNA, RNA, proteins, metabolites, and intracellular and membrane lipids. Exosome isolation needs to be optimized carefully depending on the type of biofluid and tissue and the retrieved exosomes need to be characterized. The main objective of this study was to determine the feasibility of a multimodal analysis of Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) spectroscopy and UPLC-QqTOF-MSMS for the development of a routine quality control tool of isolated exosomes and the rapid characterization of their lipid profiles and total protein content. Using human milk as model example, exosomes were isolated by multi-stage ultracentrifugation. After single-phase extraction, lipidomic analysis was carried out by UPLC QqTOF-MSMS with automated MSMS-based annotation using HMDB, METLIN, LipidBlast and MSDIAL databases. The classes with the largest number of annotated features were glycerophospholipids, sphingolipids, and glycerolipids. Then, dry films of 2 μl exosomes were directly analyzed by ATR-FTIR. Multivariate analysis showed significant associations between ATR-FTIR specific regions and the concentrations of different lipid classes. Principal component analysis and Hierarchical Cluster Analysis of IR and lipidomic data showed that ATR–FTIR renders valuable qualitative descriptors of the lipid content of isolated exosomes. Total LC-MS lipid and total protein contents could also be quantified by using the area of CHs and C=O stretching bands as well as the amide I band. As a conclusion, results obtained show that multimodal analysis of ATR-FTIR and UPLC-MS data is a useful tool for the development of spectroscopic methods. ATR-FTIR provided both, qualitative and quantitative chemical descriptors of isolated exosomes, enabling a fast and direct quantification of total protein and lipid contents.

1. Introduction

Exosomes are nanosized membrane vesicles discovered in 1983, released by fusion of an organelle of the endocytic pathway, the multivesicular body, with the plasma membrane [1]. Initially, exosomes were proposed to represent cellular waste vesicles to maintain homeostasis within the cell but nowadays it is acknowledged that exosomes are associated with physiological and pathological functions [1], contributing to different aspects of physiology and disease, including intercellular communication. Besides, exosomes may have clinical applications and their use as source for diagnostic biomarkers and as drug-delivery vectors for therapeutic applications is a very active field of research [2].

Isolation techniques include differential centrifugation, ultracentrifugation, density gradient centrifugation, ultrafiltration, immunoprecipitation, and size exclusion chromatography providing different levels

* Corresponding author. Health and Biomedicine, Leitat Technological Center, Carrer de la Innovació, 2, 08225, Terrassa, Spain.
E-mail address: gquintas@leitat.org (G. Quintás).
[1] Both authors contributed equally to the study.

https://doi.org/10.1016/j.chemolab.2021.104401
Received 28 June 2021; Received in revised form 30 July 2021; Accepted 10 August 2021
Available online 18 August 2021
0169-7439/

of recovery, purity, and sample throughput. Due to the increasing potential for the use of exosomes in clinical applications and research, there is a need to develop analytical methods to support the technical standardization of their characterization and the quality control (QC) of their isolation [3]. The characterization of exosome subpopulations and the QC of exosome isolations is a complex analytical challenge due to the diversity, heterogeneity, and complexity of their composition. Indeed, exosomes contain specific cargos of DNA or RNA including single-stranded DNA, double stranded-DNA, mitochondrial DNA, and miRNA, as well as proteins, metabolites and intracellular and membrane lipids [4], with structural and biological activity. Common techniques for their characterization include transmission electron microscopy, scanning electron microscopy, cryogenic electron microscopy, and atomic force microscopy for visualization; analysis of transmembrane proteins (i.e., tetraspanins such as CD9, CD63, and CD81) by Western blotting for identification; nanoparticle tracking analysis, asymmetric field-flow fractionation, resistance pulse sensing and protein quantification (e.g., bicinchoninic acid (BCA) assay); and the ExoView platform (NanoView Biosciences, MA, USA) for surface marker detection.

Lipid analysis of exosomes is a relatively unexplored field of research fostered by technical and methodological advances in high resolution hyphenated liquid chromatography – mass spectrometry (LC-MS). Due to their mechanisms of formation, the distribution of membrane lipids of exosomes is expected to be associated to the composition of the plasma membrane [5], including phospholipids, sphingolipids and cholesterol. Despite its sensitivity and detection range, LC-MS metabolomics requires highly skilled personal and bulky and expensive instrumentation, as well as careful sample preparation, and complex data processing and analysis steps, that limit its application for a routine QC of exosome isolations.

Attenuated Total Fourier Transform Infrared (ATR-FTIR) spectroscopy is an emerging tool in the bio-medical field for the analysis and characterization of biological samples. The technique normally relies on the fast and simple acquisition of the infrared spectrum from the untreated sample using cost-effective instrumentation and no expensive reagents or consumables. The spectra contain partially overlapped bands representative of its main components, including proteins [6], lipids [7], carbohydrates and DNA [8]. Previous results have shown that, due to its speed, simplicity, and capability for finger-printing the major components of complex samples, ATR-FTIR can be used to assess extraction procedures and sample preprocessing of biofluids in metabolomics [9]. In the field of exosome analysis, ATR-FTIR has been scarcely employed and a very limited number of applications have been proposed, including the quantification of the total protein content [10], the protein-to-lipid ratio [11], the evaluation of their isolation [12,13] as well as for the investigation of changes in the composition of saliva exosomes caused by oral cancer [14].

In this work, we aimed to assess the advantages and limitations of a multimodal analysis of ATR-FTIR spectroscopy and UPLC–MS data for the development of a routine QC tool of isolated exosomes. To this end, we employed isolated exosomes from human milk (HM) as model example and used lipidomic profiles analyzed by UPLC-QqTOF-MS, and total protein contents determined by the BCA assay as reference methodologies. To our knowledge, no study has combined ATR-FTIR and untargeted UPLC-MS-based lipidomics for the analysis of the lipid composition of isolated exosomes. Results obtained indicated that the joint analysis of IR and lipidomic information could be a powerful strategy for sensor development with potential to foster innovative cross-disciplinary research. Besides, results confirmed the capabilities of ATR-FTIR spectroscopy for a fast and direct determination of lipids and proteins, and also that this technique can be used for a rapid evaluation and characterization of the exosomes in terms of lipid composition, thus supporting its use for the QC of the exosome isolation.

2. Materials and methods

2.1. Reagents and materials

LC-MS grade acetonitrile (CH_3CN), isopropanol (IPA), and methanol (CH_3OH) were obtained from Scharlau (Barcelona, Spain); formic acid (HCOOH) (≥95%), albumin from bovine serum (≥98%), BCA Kit for Protein Determination, phosphate buffered saline (PBS) and ammonium acetate (CH_3COONH_4) (≥98%) from Sigma-Aldrich Química SL (Madrid, Spain); prostaglandin $F_{2\alpha}$-D_4 from Cayman Chemical Company (Michigan, United States); oleic acid-D_9 from Avanti Polar Lipids Inc. (Alabama, United States) and *tert*-butyl methyl ether (MTBE) (≥99%) from Fisher Scientific SL (Madrid, Spain). Ultra-pure water was generated employing a Milli-Q Integral Water Purification System from Merck Millipore (Darmstadt, Germany).

2.2. Collection of HM samples

The study was approved by the Ethics Committee for Biomedical Research of the Health Research Institute La Fe, University and Polytechnic Hospital La Fe (Valencia, Spain) with registry # 2019-289-1 and all methods were performed in accordance with relevant guidelines and regulations. Written informed consents were obtained from lactating mothers prior to sample collection and analysis of demographics and clinical information.

Ten HM samples were collected from mothers of preterm infants (<32 weeks of gestation) and term infants (>37 weeks of gestation) after establishing full enteral nutrition (i.e., stable intake of >150 ml/kg/day) using electric breast pumps following the instructions of the hospital staff. Milk was collected from full expression of one breast between 7 and 10 a.m. and preferably a minimum of 3 h after the last feed or extraction. Extracted milk was stored immediately at 4 °C and within 6 h transported to the laboratory on ice until further processing. Parameters describing the study population are shown in Table 1. In addition, two aliquots of pooled HM samples were provided by the Human Milk Bank of the University and Polytechnic Hospital La Fe.

2.3. Exosome isolation

After gentle manual shaking during 30 s, 25 mL of milk were centrifuged for removal of milk fat globules in two consecutive centrifugations at 3000×*g* for 10 min at 4 °C using an Eppendorf 5804 benchtop centrifuge with an A-4-62 rotor (Hamburg, Germany). The upper fat layer was discarded, and the supernatant was syringe-filtered (0.40 μm) prior to a third centrifugation step (3000×*g*, 10 min, 4 °C) to pellet proteins. The supernatant was collected and ultracentrifuged twice at 10000 rpm for 1 h, at 4 °C using a Hitachi CP100NX centrifuge with a Beckman Coulter 50.2 Ti rotor (Indianapolis, United States) to pellet proteins. The supernatant was syringe-filtered (0.40 μm). Then, three ultracentrifugation steps at 30000 rpm for 2 h, at 4 °C to pellet HM exosomes were performed. Between ultracentrifugation steps, supernatants were discarded, and pellets were washed with 25 mL of PBS. After the last ultracentrifugation step, supernatants were aspirated, and the isolated HM exosome pellets were suspended in 200 μL PBS and stored at −80 °C.

Table 1
Parameters of the study population.

Parameters	Median (1st-3rd quartile)
Mother's age [years]	35 (33 42)
Birth Weight [g]	1775 (1325 3078)
Gestational age [weeks]	34 ± 6 (30 ± 0 38 ± 6)
HM intake [mL/kg]	153 (150 170)
Infant's age [days]	14 (8 49)

2.4. Exosome characterization

Exosome size distribution and quantification of vesicles were analyzed using the ExoView platform (NanoView Biosciences, MA, USA). Isolated exosomes were diluted 1:25 × 10^6 in 0.22 μm pre-filtered PBS and then incubated on ExoView Human Tetraspanin chips prior to counter-staining with CD9, CD63, and CD81 fluorescent antibodies. Protein concentration was determined following the BCA Kit for Protein Determination (Sigma-Aldrich Química SL) assay based on the reduction of alkaline Cu(II) to Cu(I) by proteins in a concentration-dependent manner. BCA is a specific chromogenic reagent for Cu(I), forming a complex with an absorbance maximum at 562 nm.

2.5. ATR-FTIR spectroscopy

Infrared spectra in the 4000 to 400 cm^{-1} range were acquired using an Alpha II (Bruker Optics GmbH, Ettlingen, Germany) spectrometer equipped with a platinum-ATR with monolithic diamond measurement interface element (single reflection), a CenterGlow™ IR-source and a temperature stabilized DTGS detector with no purging system required. OPUS 8.5 software (Bruker Optics GmbH) was used to control the instrument. 2 μL of the exosome suspension were dropped onto the ATR crystal using a micropipette and then dried at room temperature for 20 s using an intermittent air stream. Spectra were collected by co-adding 32 scans with a resolution of 4 cm^{-1} using a previously recorded spectrum of air with the same instrumental conditions as background, with an acquisition time of 20 s. Between samples, the ATR interface was cleaned using a cotton swab and H_2O, IPA, and CH_3OH until recovery of the baseline signal. Spectral acquisition order was randomized but sample and spectral replicates of each exosome isolation were acquired in a row to avoid unnecessary freeze and thaw cycles that might affect their composition. All spectra were subjected to two initial quality tests. First, to guarantee that there was a significant amount of sample dried onto the ATR crystal, a minimum absorbance intensity threshold of 75 mAU was established for the amide I band. Secondly, to ensure that the contribution of water vapor to the spectra was negligible, a cut-off value of 10 was set for the ratio between the absorbance of the amide I band at 1642 cm^{-1} and the root mean square value of the mean centered absorbance in the 1820–1800 cm^{-1} interval.

2.6. Lipid extraction and LC-MS analysis

Lipids and other polar metabolites were extracted from exosomes using a single-phase extraction procedure [15] [–] [17]. 45 μL of isolated HM exosomes suspension in PBS were mixed with 5 μL of IS solution containing oleic acid-D_9 and prostaglandin $F_{2\alpha}$-D_4, 80 μM and 39 μM respectively. After 2 min of sonication, 175 μL of methanol followed by 175 μL of MTBE were added followed by mixing on a Vortex® mixer during 30 s for protein precipitation and compound extraction and sonication during 2 min to assist the release of metabolites from exosomes during extraction. After centrifugation at 4000×*g* for 15 min at 4 °C, 100 μL of supernatant containing the extracted lipids and metabolites were dried using a miVac centrifugal vacuum concentrator (Genevac LTD, Ipswich, UK) and dissolved in 100 μL of initial mobile phase (98% of mobile phase A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) and 2% mobile phase B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH)). A pooled QC sample was prepared by mixing 5 μL of each sample extract. In addition, a calibration blank (i.e., water instead of isolated exosomes) and a procedural blank (i.e., PBS supernatant from the last step of exosome isolation) were prepared, both containing IS.

Lipidomic analysis was carried out employing a 1290 Infinity HPLC system from Agilent Technologies (CA, USA) equipped with a UPLC BEH C18 column (50 × 2.1 mm, 1.7 μm, Waters, Wexford, Ireland). A binary mobile phase gradient was employed starting at 98% of mobile phase A (5:1:4 IPA:CH_3OH:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) during 0.5 min followed by a linear gradient from 2 to 20% of mobile phase B (99:1 IPA:H_2O 5 mM CH_3COONH_4, 0.1% v/v HCOOH) during 3.5 min and from 20 to 95% v/v of mobile phase B in 4 min; 95% v/v of mobile phase B was maintained during 1 min; return to initial conditions was achieved in 0.25 min and were maintained for a total run time of 13 min. Column and autosampler were kept at 55 and 4 °C, respectively, the injection volume was 2 μL, and the flow rate was set to 400 μL min^{-1}. An Agilent 6550 Spectrometer iFunnel quadrupole time-of-flight (QTOF) MS system working in the ESI+ and ESI- modes was used for MS detection. Full scan MS data in the range between 70 and 1500 *m/z* were acquired at a scan frequency of 5 Hz using the following parameters: gas T, 200 °C; drying gas, 14 L min^{-1}; nebulizer, 37 psi; sheath gas T, 350 °C; sheath gas flow, 11 L min^{-1}. Mass reference standards were introduced into the source for on-the-fly automatic MS spectra recalibration during analysis via a reference sprayer valve using the 149.02332 (phthalic anhydride), 121.050873 (purine), and 922.009798 (HP-0921) *m/z* in ESI+, and 119.036 (purine) and 980.0163 (HP-0921, $[M\text{-}H + CH_3COOH]^-$) *m/z* in ESI-, as references. ESI+ and ESI- analysis were carried out in independent batches and between them, the instrument was cleaned and calibrated according to manufacturer guidelines. QCs were used to monitor the instrument performance and correct within-batch effects [18,19]. A set of 9 QCs were injected at the beginning of each batch for system conditioning and MS/MS data acquisition. MS/MS spectra were acquired using the auto MS/MS method with the following inclusion *m/z* precursor ranges: 70–200, 200–350, 350–500, 500–650, 650–800, 800–950, 950–1100, 1100–1200, and from 70 to 1200 using, in all replicates, a rate of 5 spectra/s in the extended dynamic range mode (2 GHz), a collision energy set to 20 V, an automated selection of five precursor ions per cycle and an exclusion window of 0.15 min after two consecutive selections of the same precursor. During the remaining batch sequence, a QC replicate was injected after every 5 study samples. Sample extracts were analyzed in random order. Three blank extracts were injected at the beginning and end of each batch for identifying unreliable, background, and carry-over features as described elsewhere [20].

2.7. MS data pre-processing and metabolite annotation

Peak table generation was carried out using XCMS software [21]. The *centWave* method was used for peak detection with the following parameters: mass accuracy, 20 ppm; peak width, (3,15); snthresh, 12; prefilter, (5,3000). A minimum difference in *m/z* of 7.5 mDa was selected for overlapping peaks. Intensity weighted *m/z* values of each feature were calculated using the *wMean* function. Peak limits used for integration were found through descent on the Mexican hat filtered data. Grouping before and after RT correction was carried out using the *nearest* method and 9 s as *rtCheck* argument. Finally, missing data points were filled by reintegrating the raw data files in the regions of the missing peaks using the *fillPeaks* method. The CAMERA package [22] was used for the identification of pseudospectra based on peak shape analysis, isotopic information and intensity correlation across samples [23]. Each dataset was processed with the following CAMERA functions: *xsAnnotate*, *groupFWHM*, *findIsotopes*, *groupCorr* and *findAdducts* using standard arguments. Identification and elimination of uninformative features was carried for ESI+ and ESI- data sets independently.

Metabolite annotation (Level lD: 2, putatively annotated compounds without matching to data for chemical standards acquired under the same experimental conditions) was carried out by matching experimentally acquired MS/MS spectra with the experimental HMDB, METLIN, and MSDIAL MS/MS databases in accordance with the Metabolomics Standards Initiative (MSI) reporting standards [17,24]. Metabolite annotation using LipidBlast [25] was carried out using LipiDex [26] with 0.01 Da tolerances in both MS (precursor) and MS^2 (fragment) data and the 'LipidBlast Acetate' library.

Within batch effect was removed using the non-parametric QC-Supported vector regression (SVR) correction approach employing a Radial Basis Function kernel [18,19]. The selection of the tolerance threshold (ε), the penalty term applied to margin slack values (C) and the kernel

width (γ) was carried out using a pre-selection of C and optimization of ε and γ using a grid search, leave-one-out cross validation and the root mean square error of cross validation (RMSECV) as target function. C was selected for each LC-MS feature as the median value of the intensities observed in QC replicates. The ε search range was selected based on the expected instrumental precision (4–10% of the median value of the intensities observed for the whole set of QC replicates). The γ search interval selected was $[1, 10^5]$. Variables with more than 2 missing values in QCs, those with RSD(QC) > 20% after QC-SVRC, and for those for which the ratio between the median peak area values in QCs and blanks was lower than 6 were classified as unreliable and removed from further analysis.

2.8. Software

UPLC-TOFMS data acquisition and manual integration was carried out employing MassHunter Workstation (version B.07.00) from Agilent. Raw data was converted into mzXML format using ProteoWizard (http://proteowizard.sourceforge.net/). Peak detection, integration, deconvolution, alignment and pseudospectra identification were carried out using XCMS and CAMERA in R 3.6.1. Metabolite annotation and data analysis were carried out in MATLAB 2017b (Mathworks Inc., Natick, MA, USA) using in-house written scripts and the PLS Toolbox 8.7 (Eigenvector Research Inc., Wenatchee, USA) as described elsewhere [17].

Statistical heterospectroscopy (SHY) [27] was employed for the joint analysis of the lipidomic and spectroscopic data acquired from the set of isolated exosome samples. SHY involved the estimation of the covariance between signal intensities measured by the two techniques across the set of samples. The SHY plot represents the correlations of IR absorbances at each wavenumber with the UPLC-MS intensities of each annotated lipid. To facilitate the extraction of information, a correlation threshold was applied. A lipid class was not represented if >50% of the correlations of its metabolites failed to reject the null hypothesis considering a 97.5% confidence level (t-test, unequal variances).

MATLAB scripts used datasets generated and analyzed during the current study are available in the Zenodo repository (zenodo.org/deposit/5148582).

3. Results

3.1. Characteristics of isolated HM exosomes

Fig. 1 (left) shows captured HM exosomes based on the expression of the ubiquitous tetraspanin markers CD9, CD63, and CD81 from the analysis of a representative HM exosome isolate. The marker colocalization analysis (see Fig. 1, right) shows the distribution of different tetraspanins across the membranes of the extracted HM exosomes. Most of the isolated HM exosomes expressed CD9 and CD81 while only a small fraction expressed all three tetraspanins. The employed multistep ultracentrifugation protocol allowed the extraction of HM exosomes with a median size of 61 (3 interquartile range, IQR) nm and a median number of 8×10^{14} particles mL^{-1} (3×10^{15} IQR), as determined by the ExoView platform. Protein concentrations in HM exosomes isolates determined by the BCA assay ranged between 2 and 6 g L^{-1}. Regarding the lipidomic profiles recorded from HM exosomes, Fig. 2 summarizes the main classes of the 377 features annotated in the UPLC-MS data set (208 in ESI+ and 169 in ESI-) after data pre-processing and clean-up. The classes with the largest numbers of annotated features were glycerophospholipids (50%) [glycerophosphocholines (60), glycerophosphoethanolamines (35), glycerophosphoglycerophosphoglycerols (3), glycerophosphoserines (22), glycerophosphoinositols (13), plasmenyl PCs (15), plasmenyl PEs (24), and phosphatidylinositols (11)], sphingolipids (31%) [phosphoshingolipids (71), ceramides (22), glycosphingolipids (6), hexosylceramides (17), Glc Ceramides (2)], and glycerolipids (17%) [triradylglycerols (64)].

Fig. 3 shows ATR-FTIR spectra and the first and second derivative spectra in the 3450-800 cm^{-1} range of dry residues from 2 μL of exosomes isolated from four different HM samples. The figure also shows the spectra of a blank extract with PBS bands at 1125, 1069, 978, and 851 cm^{-1}. Despite the overlap between the exosome signal and PBS bands in the 1125-850 cm^{-1} region, the use of the first or second derivative spectra enabled the resolution of overlapping spectral bands in the exosome extracts at slightly different wavenumbers (note that the minimum in a second derivative spectrum corresponds to the band apex in the raw ATR-FTIR spectrum). Spectral contributions from proteins including the amide I, II, and III were detected in exosome spectra. The amide I band governed by the stretching vibrations of C—O (70–85%) and C—N groups (10–20%) of the protein peptide backbone was detected in the 1600-1700 cm^{-1} range. The amide II band arising from in-plane N–H bending (40–60%), and C–N (18–40%) and C–C (10%) stretching vibrations was detected in the 1510-1580 cm^{-1} region. The amide III band at 1250–1350 cm^{-1} is a complex band arising from a combination of several coordinate displacements. The broad, intense band at 3285 cm^{-1} was associated to N–H stretching vibrations of peptide protein groups.

Besides, spectral contributions from carbohydrates, phosphate bands from phospholipids, DNA and RNA, and lipids were observed (see Table 2). Among them, intense bands were observed in the lipid frequency domains including the $-CH_2$ and $-CH_3$ symmetric and antisymmetric stretching (3000-2800 cm^{-1}). The characteristic C—O stretching band at 1745 cm^{-1} has been associated to the ester groups of glycerophospholipids (e.g., glycerophosphocholines, glycerophosphoethanolamines), glycerolipids (e.g., triradylglycerols), and cholesterol esters, among other lipid classes.

The differences observed in the intensity of protein and lipid bands indicate changes in the protein and lipid composition across exosome

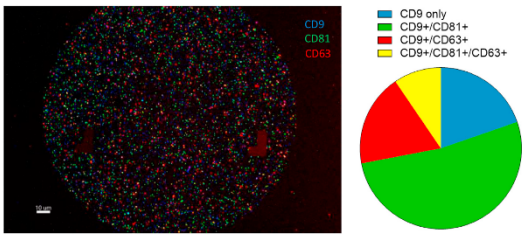


Fig. 1. Tetraspanin fluorescent staining (left) and marker colocalization analysis (right). Note: CD-9 capture.

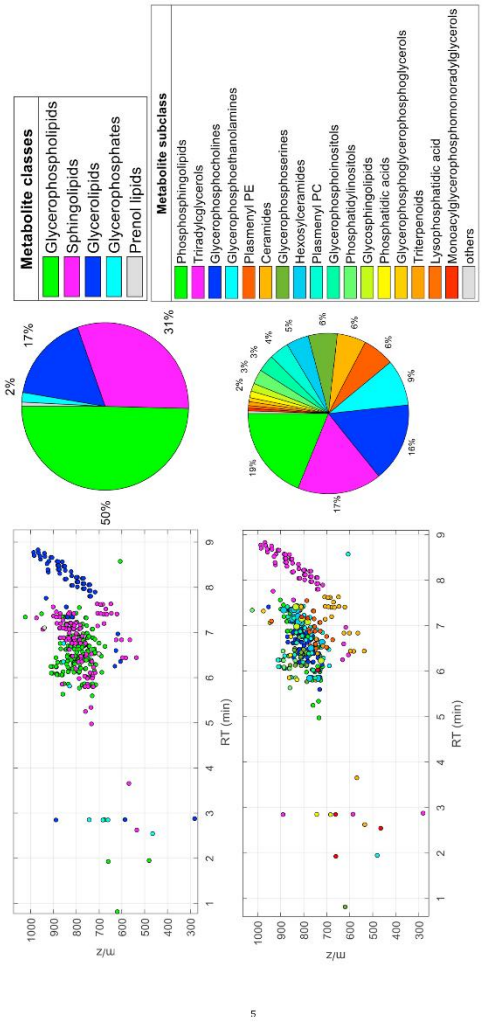


Fig. 2. Distribution of annotated UPLC–MS features detected in HM exosomes indicating their class (top) or subclass (bottom).

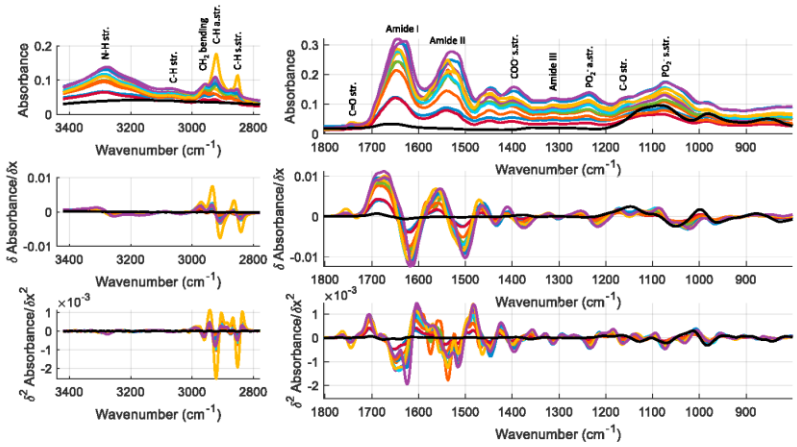


Fig. 3. ATR-FTIR raw (top), first derivative (middle) and second derivative (bottom) spectra in the 3420-2780 cm^{-1} (left) and 1800-800 cm^{-1} range (right) of dry residues obtained from 2 μL of the set of exosome extracts in PBS using a spectrum of air as background. Note: black line corresponds to a PBS blank spectrum; Colored lines: spectra of isolated exosomes from HM samples.

Table 2
Assignments of main bands observed in the ATR-FTIR spectra of HM exosomes.

Wavenumber (cm $^{-1}$)	Spectral assignment
3285	N H stretching (proteins)
2959	CH_3 asymmetric stretching (lipids, proteins)
2921	CH_2 anti-symmetric stretching (lipids)
2872	CH_3 symmetric stretching (lipids and proteins)
2851	CH_2 symmetric stretching (lipids)
1745	Saturated ester C O stretch (lipids, cholesterol, phospholipids, cholesterol esters)
1646	Amide I (proteins)
1537	Amide II (proteins)
1448	CH_2 bending of lipidic acyl chains (lipids, proteins)
1402	COO– symmetric stretch (fatty acids, aminoacids)
1314	Amide III (proteins)
1236	PO_2 antisymmetric stretch (phospholipids, nucleic acids)
1156	CO O C antisymmetric stretching (glycogen, nucleic acids) C O stretching from alcohol groups (glycogen, lipids)
1080	PO_2 symmetric stretch (phospholipids, nucleic acids)

isolations. Moreover, the marked differences in band shapes in the amide I and II region could also indicate changes in the higher order structure of proteins.

3.2. Multimodal qualitative analysis of ATR-FTIR spectra and lipidomic profiles

SHY [27] was employed to get further insight into the correlation among IR bands and UPLC-MS lipid profile. SHY enables the co-analysis of multi-source (i.e., lipidomic and spectroscopic) data sets by estimating the covariance between signal intensities measured by two techniques across the set of isolated exosome samples. Accordingly, SHY involved the calculation of a series of linear models between the ATR-FTIR absorbance for each vibrational feature and the intensities of each of the features annotated as a glycerophospholipid, sphingolipid, or glycerolipid. Results depicted in Fig. 4 show statistically significant associations (p-value<0.025) between specific overlapping regions of the raw (Fig. 4, left) and second derivative (Fig. 4, right) ATR-FTIR spectra and the different lipid classes. This result supports the potential of SHY to explore associations between lipidomic profiles and spectroscopic data and revealed relationships between the distribution of specific annotated lipid classes and the IR spectra of exosomes.

Principal component analysis (PCA) and Hierarchical Cluster Analysis (HCA) were used to identify samples with similar lipidomic profiles, using autoscaling as data preprocessing, the Pearson correlation coefficient as distance measure, and the Ward's algorithm for clustering. Fig. 5a (left) shows the scores plot of the 2 PCs model explaining 82% of the total variance in the lipidomic UPLC-MS data set. Samples were clustered mainly along PC1 in three groups including samples #7, 11, and 12 (cluster 1); #4, 5, and 9 (cluster 2); and # 1, 2, 3, 6, 8, and 10 (cluster 3). PC2 enabled the discrimination of sample 7 from 11 to 12 in cluster 1. Fig. 5a (center) depicts the scores plot of the 2 PCs model explaining 57% of the total variance in the autoscaled second-derivative ATR-FTIR data, where the samples were labeled according to the clusters selected in the lipidomic data set. Using the distances among samples as similarity criteria, results showed a similar trend: samples included in cluster 1 differ from those included in cluster 3 and were more similar to those included in cluster 2. The similarity between the trends observed in both scores plots was assessed by the Mantel test [28] which evaluated the statistical significance of the correlation between the two pairwise Euclidean distance matrices. An empirical *p*-value was estimated using a permutation test where a reference null distribution is obtained by repeatedly random shuffling the objects (here, n = 10^6). After each permutation, the correlation between the obtained two pairwise Euclidean distance matrices is calculated, and the *p*-value is estimated as the proportion of permuted estimates for which the absolute correlation value was equal to or greater than the correlation estimate calculated

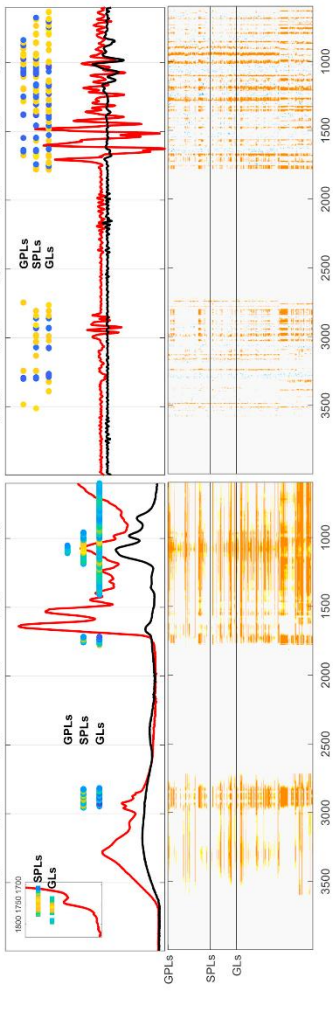


Fig. 4. Top: Median value of the correlation between annotated UPLC-MS features clustered according to their lipid classes and ATR-FTIR data (raw: left; second derivative: right) determined by STIY using the slope coefficient of a linear model. Here, only those wavenumbers for which >50% of the features of each class showed a significant correlation (p-value<0.025) are depicted. The plot overlays ATR-FTIR spectra from isolated exosomes (red) and a blank extract (black) for a better interpretation of the results. Bottom: Correlation among annotated UPLC-MS features clustered according to their lipid classes and ATR-FTIR data determined by STIY using the slope coefficient of a linear model. A correlation cutoff was applied (linear model p-value<0.025) for better visualization. Note: GPLs: glycerophospholipids; SPLs: sphingolipids; GLs: glycerolipids. Red line – IM exosome spectrum and black line – PBS blank spectrum. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

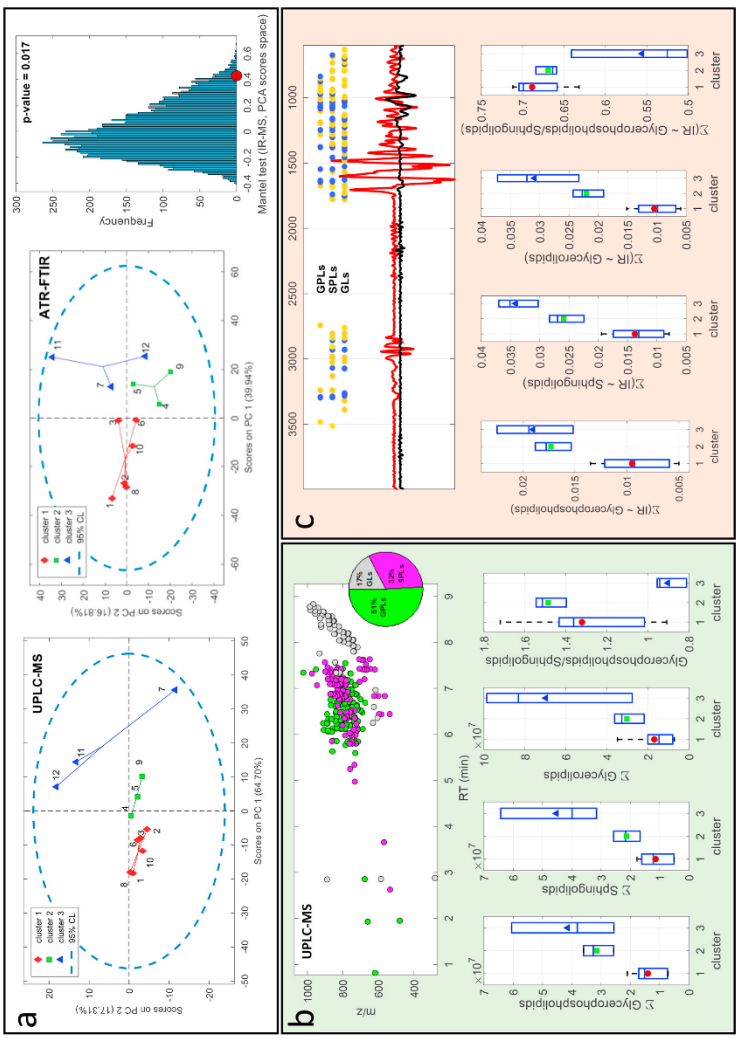


Fig. 5. a) PC1 vs PC2 scores plots from PCA of UPLC-MS lipid profiles (left), and ATR-FTIR spectra (middle). Sample classes were assigned according to the results of HCA based on UPLC-MS data. The statistical significance of the correlation between the distribution of samples in both PC1 vs PC2 scores spaces was assessed by the Mantel test (p-value<0.001) (right). b) Distribution of GLPs, GLs and SPLs features in the *m/z* vs RT space (top) and box plots showing the sum of intensities of each of these lipid classes in the three clusters (bottom). c) Second derivative ATR-FTIR spectra showing the spectral regions associated to GLs, SPLs, and GPLs (top) and sum of the intensities of the spectral regions associated to each of these lipid classes in each of the three sample clusters (bottom).
Note: GPLs: glycerophospholipids; SPLs: sphingolipids; GLs: glycerolipids.

using the original ordering. Results obtained depicted in Fig. 5a (right) (p-value<0.005) indicated a statistically significant correlation between the distribution of samples in the PCA scores spaces.

Fig. 5b shows the distribution of glycerophospholipids, sphingolipids, and glycerolipids in the RT-m/z space, as well as the boxplots representing the distributions of sum of intensities of the three classes of lipids in the three selected HCA clusters. Cluster 1 showed lower concentrations of the three lipid classes than samples included in clusters 2 and 3. Besides, although clusters 2 and 3 showed similar levels of glycerophospholipids, cluster 3 showed significantly higher relative levels of sphingolipids. Results also indicated a lower ratio of glycerophospholipids/sphingolipids in cluster 3, being cluster 2 the group of exosomes isolates with the highest ratio of glycerophospholipids/sphingolipids. Fig. 5c (top) shows the second derivative spectral regions with a statistically significant correlation with the levels of >50% of the UPLC-MS features annotated as glycerophospholipids, sphingolipids, or glycerolipids in the sample set. The boxplots representing the distributions of the sum of absolute values in the ATR-FTIR second derivative spectra associated to the three classes of lipids in the three selected HCA clusters showed a similar pattern as those found using LC-MS data.

3.3. Quantification of proteins and lipids by ATR-FTIR spectroscopy

The absorbance in the 1724-1591 cm $^{-1}$ region showed a statistically significant correlation with the protein content determined in the isolated exosomes by the BCA assay (A=(-0.46 ± 0.34)+(1.09 ± 0.06) [protein](g/L), $R^2 = 0.79$, p-value<10^{-5}), in agreement with previous reports [11,29] (see Fig. 6, top left). This result suggests the use of an external calibration using a model protein (e.g., albumin) for a fast quantification of total protein content in exosomes. As shown in Fig. 6 (top, right), the exosome protein contents determined using the area of the amide I band of an external calibration line and a serial dilution of albumin standards showed a statistically significant correlation with protein concentrations determined by BCA (slope = (0.80 ± 0.05),

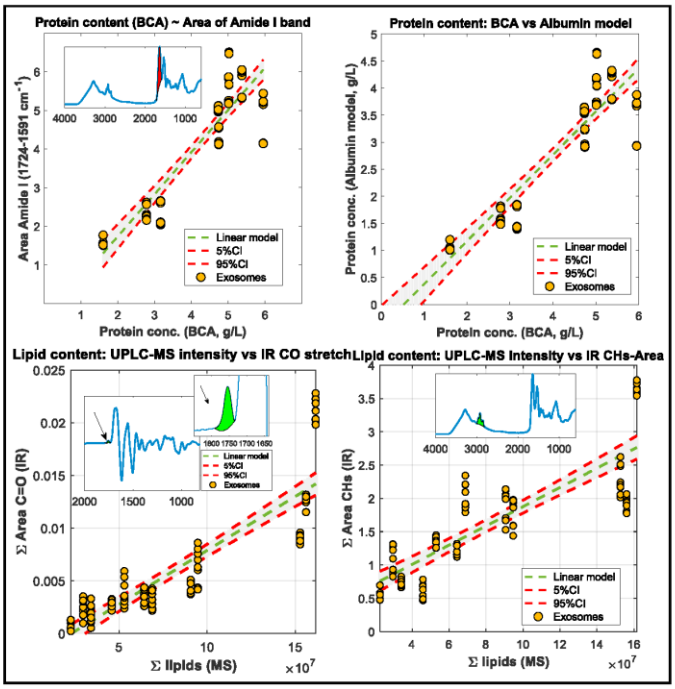


Fig. 6. Correlation between the protein content determined by the BCA assays and the amide I area in the set of isolated exosomes (top, left) and the protein content determined by BCA assays and the protein concentration determined using an external calibration line from a serial dilution of albumin solutions and amide I area (top, right). Correlation between the lipid content determined as the sum of intensities of LC-MS annotated features and the C=O (bottom, left) and CH (bottom, right) spectral area in the set of isolated exosomes.

$R^2 = 0.79$, p-value$<10^{-5}$).

The use of ATR-FTIR spectra for estimating the lipid content of exosomes was also explored. Using the sum of UPLC-MS intensities of annotated lipids as surrogate of the total lipid content in samples, the IR spectral intensity between 1790 and 1733 cm^{-1} in the first derivative spectra associated with the C=O stretching band (~1740 cm^{-1}) correlated with the lipid content ($A=(-0.002 \pm 0.005)+(1.02 \pm 0.05)\ 10^{-10}$[LCMS intensity](AU), $R^2 = 0.75$, p-value$<10^{-5}$) (see Fig. 6, bottom-left). Similarly, the IR spectral area in the 3000-2800 cm^{-1} range also correlated with the lipid content ($A=(0.43 \pm 0.09)+(1.44 \pm 0.09)\ 10^{-8}$[LCMS intensity](AU), $R^2 = 0.70$, p-value$<10^{-5}$) (see Fig. 6, bottom-right).

The repeatability of ATR-FTIR determinations is a relevant parameter for assessing its usefulness for quantitative applications. To address this concern, the RSD% observed for each wavenumber in the 3710-2717 cm^{-1} and 1840-800 cm^{-1} ranges was calculated for each sample. In this study, the set of 12 exosome samples were analyzed by replicate ($n = 3$), also acquiring 3 spectra/replicate. The distribution of RSD% median values obtained for each sample varied in the 3.3–11.4 RSD% range (mean value $= 5 \pm 1$, median value $= 5.4$). Although these figures show acceptable reproducibility, automation of the dry-film generation for monitoring of exosome isolation would likely result in higher precision and sample throughput levels [30].

4. Discussion

We used a standard multi-step ultracentrifugation procedure for the isolation of HM exosomes from twelve HM samples. The detection of ubiquitous tetraspanins confirmed that the isolated particles are HM exosomes rather than other co-isolated contaminants. As the results from the ExoView platform, isolated exosomes showed varying particle counts, which was in good agreement with varying protein concentrations determined by the BCA assay. This reinforces the importance of quick and reproducible approaches for characterizing exosomes for routine QC. The use of a LC-MS-based lipidomic approach allowed to gain detailed insight into the composition of HM isolated exosomes (Fig. 2). The observed distribution of lipid classes agreed with the expected lipid composition of exosomes. Although their lipid profiles can be linked to their cellular origin, exosomes are bilayered proteolipids comprising mainly plasma membrane lipids (e.g., phospholipids, sphingolipids, and cholesterol) [5] enriched with bioactive cargo captured from the cytosol during the formation of the intraluminal vesicles. Previous results have observed that, compared to the cell of origin, exosome membranes are often enriched with phosphatidylserine, phosphatidylethanolamines, phosphatidylcholines, sphingolipids, and cholesterol. Besides, reported triradylglycerols could be indicative of the co-isolation of extracellular lipids during the exosome isolation, present at high concentrations in HM.

ATR-FTIR spectra showed spectral contributions from proteins, carbohydrates, phosphate bands from phospholipids, DNA and RNA, and lipids (see Fig. 3 and Table 2). Moreover, the joint analysis of LC-Ms lipidomic profiles and ATR-FTIR data (Figs. 4 and 5) showed that several spectral features were associated with different lipid classes present in HM exosomes, even though the level of detail provided by this technique is lower. However, we observed specific patterns in the ATR-FTIR spectra that were indicating the presence of the three major lipid classes (i.e., glycerophospholipids, sphingolipids, and glycerolipids) identified by LC-MS. Glycerophospholipids are glycerol-based phospholipids and spectral regions associated with >50% of glycerophospholipids included the 1100-1045 cm^{-1} region where the characteristic frequencies of the P–O–C asymmetric and symmetric stretching bands of the phosphate group can be observed. Sphingolipids are a complex family of compounds sharing a sphingoid base backbone. They were linearly correlated with ATR-FTIR absorption in the CH_2 and CH_3 symmetric and antisymmetric regions (2963-2814 cm^{-1}), C=O stretching region (1776-1708 cm^{-1}), and in the 1180-995 cm^{-1} region including characteristic frequencies associated to phosphate groups such as $\nu C\text{-}O\text{-}PO_2^-$ (1060-1070 cm^{-1}) and νPO_2^- sym. (1093 cm^{-1}) [11]. Glycerolipids are also a structurally heterogeneous group of lipids composed of mono-, di-, and tri-substituted glycerols that have at least one hydrophobic chain linked to a glycerol backbone in an ester or ether linkage [31]. Spectral regions associated with this lipid class included bands at 2938–2930, 2866–2858, and 2841-2825 cm^{-1} linked to CH_2 and CH_3 symmetric and antisymmetric vibrational modes, as well as the regions 1772–1735 (C=O stretching), 1226–1141, 1053-881 cm^{-1} (C–O stretching from alcohols), and 739-698 cm^{-1}. The partial overlap between the spectral regions associated to each of the lipid classes agreed with their structural similarity, and included bands from the acyl chains, and the glycerol and phospholipid headgroups.

Furthermore, we could show that ATR-FTIR can be used for quantification. Specific lipid bands in exosomes, namely the 3000-2800 cm^{-1} range (CH stretching vibrations) and the 1790-1733 cm^{-1} (C=O stretching) showed a linear correlation with lipid contents determined by LC-MS (Fig. 6). We could also show that the same ATR-FTIR spectrum can be used for quantitation of total proteins in HM exosomes. With a simple external calibration line from serial dilutions of an albumin solution, protein concentrations determined by ATR-FTIR were found to be comparable to reference values from the BCA assay (Fig. 6). Differences observed between concentrations obtained by both techniques might be linked to the different detection principles. The amide I band is sensitive to changes in the protein structure and hence, the use of a model protein to be used as reference might bias results. On the other hand, the accuracy of the BCA assay is limited by the presence of sample reducing agents, and by the presence of cysteine, tyrosine, and tryptophan residues. Besides, the presence of reducing sugars, lipids, and phospholipids in the sample can also affect the accuracy of the BCA. Importantly, this is the first literature report on the simultaneous quantitative analysis of proteins and lipids, and qualitative lipid analysis using ATR-FTIR spectra of exosomes. In the ATR system used, the sample is deposited onto the ATR element and an IR beam is directed through an internal reflection element (IRE). Then, the evanescent wave interacts with the sample in direct contact with the IRE providing a penetration depth in the 1–3 μm range. Therefore, the repeatability of the ATR spectral acquisition depends on the distribution of the sample on top of the ATR surface, which in turn depends on the process of dry-film generation. Here, samples were manually deposited and air-dried, achieving RSD% from replicate measurements of 5 ± 1, which is within the general acceptance criteria of 15% [32].

5. Conclusions

Results obtained indicated that the joint analysis of IR and lipidomic information could be a powerful strategy for sensor development with potential to foster innovative cross-disciplinary research. Results show that ATR-FTIR spectroscopy provides both, qualitative and quantitative chemical descriptors of isolated exosomes through changes in bands associated to specific functional groups. ATR-FTIR spectroscopy enables a fast, direct quantification of total protein and lipid contents and, simultaneously, provides biochemical information useful for a routine control of changes in the relative composition of isolated exosomes in very small sample volumes (2 μL). Thus, the use of ATR-FTIR could be considered as a cost-effective alternative, to the use of standard colorimetric assays (e.g., BCA) for QC of exosome isolations. Also, whereas ATR-FTIR cannot replace the level of detailed information provided by MS in terms of specificity or sensitivity, the spectral information could be used to rapidly assess repeatability or reproducibility among exosome isolations, providing at the same time information on differences in their lipid contents.

Funding

This work was supported by the European Union's Horizon 2020

Research and Innovation Programme through the Nutrishield project (https://nutrishield-project.eu/) [Grant Agreement No 818110] and the Ayudas Intramurales 2021 para actuaciones de innovación en el ámbito de la UCIE IIS La Fe – Línea Nominativa Agencia Valenciana de Innovación [2021-071-1]. JK and ITD acknowledge support received from Instituto de Salud Carlos III (Spain) [grant numbers CP16/00034 and CD19/00176, respectively]. DPG acknowledges financial support from the Ramón y Cajal programme [grant number RYC2019-026556-I] provided by the Ministerio de Ciencia e Innovación.

The funding organizations were not involved in the collection, analysis, and interpretation of data, the writing of the report and in the decision to submit the article for publication.

CRediT authorship contribution statement

Victoria Ramos-Garcia: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Isabel Ten-Doménech:** Formal analysis, Investigation, Writing – review & editing. **Alba Moreno-Giménez:** Data curation, Investigation. **María Gormaz:** Data curation, Resources, Supervision. **Anna Parra-Llorca:** Data curation, Investigation. **Alex P. Shephard:** Formal analysis, Investigation, Methodology, Resources, Visualization, Writing – review & editing. **Pilar Sepúlveda:** Methodology, Investigation, Resources. **David Pérez-Guaita:** Formal analysis, Investigation, Software. **Máximo Vento:** Conceptualization, Resources. **Bernhard Lendl:** Conceptualization, Resources. **Guillermo Quintás:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. **Julia Kuligowski:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Investigation, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to lactating mothers who agreed to participate in this study and to Amparo Ramón and Antonia Gálvez from the Human Milk Bank at the University and Polytechnic Hospital La Fe for their support and to NanoView Bioscience for technical support during exosome characterization.

References

[1] N.P. Hessvik, A. Llorente, Current knowledge on exosome biogenesis and release, Cell. Mol. Life Sci. CMLS 75 (2018) 193 208, https://doi.org/10.1007/s00018-017-2595-9.

[2] M. Zhou, S.R. Weber, Y. Zhao, H. Chen, J.M. Sundstrom, Chapter 2 - methods for exosome isolation and characterization, in: L. Edelstein, J. Smythies, P. Quesenberry, D. Noble (Eds.), Exosomes, Academic Press, 2020, pp. 23 38, https://doi.org/10.1016/B978-0-12-816053-4.00002-X.

[3] K.W. Witwer, E.I. Buzás, L.T. Bemis, A. Bora, C. Lässer, J. Lötvall, E.N.N. 't Hoen, M.G. Piper, S. Sivaraman, J. Skog, C. Théry, M.H. Wauben, F. Hochberg, Standardization of sample collection, isolation and analysis methods in extracellular vesicle research, J. Extracell. Vesicles 2 (2013) 20360, https://doi.org/10.3402/jev.v2i0.20360.

[4] Exosomes: A Rising Star in Failing Hearts, (n.d.). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5508217/(accessed January 25, 2021).

[5] T. Skotland, K. Sagini, K. Sandvig, A. Llorente, An emerging focus on lipids in extracellular vesicles, Adv. Drug Deliv. Rev. 159 (2020) 308 321, https://doi.org/10.1016/j.addr.2020.03.002.

[6] D. Perez-Guaita, Z. Richardson, P. Heraud, B. Wood, Quantification and identification of microproteinuria using ultrafiltration and ATR-FTIR spectroscopy, Anal. Chem. 92 (2020) 2409 2416, https://doi.org/10.1021/acs.analchem.9b03081.

[7] S. Yoshida, Y. Okazaki, T. Yamashita, H. Ueda, R. Ghadimi, A. Hosono, T. Tanaka, K. Kuriki, S. Suzuki, S. Tokudome, Analysis of human oral mucosa ex vivo for fatty acid compositions using Fourier-transform infrared spectroscopy, Lipids 43 (2008) 361 372, https://doi.org/10.1007/s11745-007-3147-0.

[8] B.R. Wood, The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues, Chem. Soc. Rev. 45 (2016) 1980 1998, https://doi.org/10.1039/c5cs00511f.

[9] J. Kuligowski, D. Pérez-Guaita, J. Escobar, I. Lliso, M. de la Guardia, B. Lendl, M. Vento, G. Quintás, Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics, Talanta 127 (2014) 181 190, https://doi.org/10.1016/j.talanta.2014.04.009.

[10] V. Szentirmai, A. Wacha, C. Németh, D. Kitka, A. Rácz, K. Héberger, J. Mihály, Z. Varga, Reagent-free total protein quantification of intact extracellular vesicles by attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy, Anal. Bioanal. Chem. 412 (2020) 4619 4628, https://doi.org/10.1007/s00216-020-02711-8.

[11] J. Mihály, R. Deák, I.C. Szigyártó, A. Bóta, T. Beke-Somfai, Z. Varga, Characterization of extracellular vesicles by IR spectroscopy: fast and simple classification based on amide and CH stretching vibrations, Biochim. Biophys. Acta Biomembr. 1859 (2017) 459 466, https://doi.org/10.1016/j.bbamem.2016.12.005.

[12] A. Drożdż, A. Kamińska, M. Surman, A. Gonet-Surówka, R. Jach, H. Huras, M. Przybyło, Ł.Ł. Stępień, Low-vacuum filtration as an alternative extracellular vesicle concentration method: a comparison with ultracentrifugation and differential centrifugation, Pharmaceutics 12 (2020) 872, https://doi.org/10.3390/pharmaceutics12090872.

[13] M. Dash, K. Palanisyandi, S. Ramalingam, S. Sahabadeen, N.S. Raja, Exosomes isolated from two different cell lines using three different isolation techniques show variation in physical and molecular characteristics, Biochim. Biophys. Acta Biomembr. 1863 (2021) 183490, https://doi.org/10.1016/j.bbamem.2020.183490.

[14] A. Zlotogorski-Hurvitz, B.Z. Dekel, D. Malonek, R. Yahalom, M. Vered, FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer, J. Canc. Res. Clin. Oncol. 145 (2019) 685 694, https://doi.org/10.1007/s00432-018-02827-6.

[15] I. Ten-Doménech, V. Ramos-García, J.D. Piñeiro-Ramos, M. Gormaz, A. Parra-Llorca, M. Vento, J. Kuligowski, G. Quintás, Current practice in untargeted human milk metabolomics, Metabolites 10 (2020) 43, https://doi.org/10.3390/metabo10020043.

[16] J.D. Piñeiro-Ramos, A. Parra-Llorca, I. Ten-Doménech, M. Gormaz, A. Ramón-Beltrán, M. Cernada, G. Quintás, M.C. Collado, J. Kuligowski, M. Vento, Effect of donor human milk on host-gut microbiota and metabolic interactions in preterm infants, Clin. Nutr. 40 (3) (2021) 1296 1309, https://doi.org/10.1016/j.clnu.2020.08.013, 0.

[17] I. Ten-Doménech, T. Martínez-Sena, M. Moreno-Torres, J.D. Sanjuan-Herráez, J.V. Castell, A. Parra-Llorca, M. Vento, G. Quintás, J. Kuligowski, Comparing targeted vs. Untargeted MS2 data-dependent acquisition for peak annotation in LC-MS metabolomics, Metabolites 10 (2020) 126, https://doi.org/10.3390/metabo10040126.

[18] J. Kuligowski, Á. Sánchez-Illana, D. Sanjuán-Herráez, M. Vento, G. Quintás, Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC), Analyst 140 (2015) 7810 7817, https://doi.org/10.1039/c5an01638j.

[19] Á. Sánchez-Illana, D. Pérez-Guaita, D. Cuesta-García, J.D. Sanjuan-Herráez, M. Vento, J.L. Ruiz-Cerdá, G. Quintás, J. Kuligowski, Model selection for within-batch effect correction in UPLC-MS metabolomics using quality control - support vector regression, Anal. Chim. Acta 5 (2018) 62 68.

[20] T. Martínez-Sena, G. Monge, D. Sanjuan-Herráez, J.V. Castell, M. Vento, G. Quintás, J. Kuligowski, Monitoring of system conditioning after blank injections in untargeted UPLC-MS metabolomic analysis, Sci. Rep. 9 (2019) 9822, https://doi.org/10.1038/s41598-019-46371-w.

[21] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779 787, https://doi.org/10.1021/ac051437y.

[22] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779 787, https://doi.org/10.1021/ac051437y.

[23] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: an integrated strategy for compound spectra extraction and annotation of LC/MS data sets, Anal. Chem. 84 (2012) 283 289, https://doi.org/10.1021/ac2024509.

[24] R.M. Salek, C. Steinbeck, M.R. Viant, R. Goodacre, W.B. Dunn, The role of reporting standards for metabolite annotation and identification in metabolomic studies, GigaScience 2 (2013) 13, https://doi.org/10.1186/2047-217X-2-13.

[25] T. Kind, K.-H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, Nat. Methods 10 (2013) 755 758, https://doi.org/10.1038/nmeth.2551.

[26] P.D. Hutchins, J.D. Russell, J.J. Coon, LipiDex: an integrated software package for high-confidence lipid identification, Cell Syst. 6 (2018) 621 625, https://doi.org/10.1016/j.cels.2018.03.011, e5.

[27] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zirah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data Sets: application in metabonomic toxicology studies, Anal. Chem. 78 (2006) 363 371, https://doi.org/10.1021/ac051444m.

[28] N. Mantel, The detection of disease clustering and a generalized regression approach, Canc. Res. 27 (1967) 209 220.

[29] S. Navarro, D. Borchman, E. Bicknell-Brown, Lipid-protein ratios by infrared spectroscopy, Anal. Biochem. 136 (1984) 382–389, https://doi.org/10.1016/0003-2697(84)90233-1.

[30] J. Ollesch, S.L. Drees, H.M. Heise, T. Behrens, T. Brüning, K. Gerwert, FTIR spectroscopy of biofluids revisited: an automated approach to spectral biomarker identification, Analyst 138 (2013) 4092–4102, https://doi.org/10.1039/c3an00337j.

[31] R. Bittman, Glycerolipids: chemistry, in: G.C.K. Roberts (Ed.), Encycl. Biophys., Springer, Berlin, Heidelberg, 2013, pp. 907–914, https://doi.org/10.1007/978-3-642-16712-6_527.

[32] Food and Drug Administration (FDA), Guidance for Industry: Bioanalytical Method Validation. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Veterinary Medicine, 2018, p. 44.