



UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Análisis del aire exhalado *off-line* para la búsqueda de biomarcadores de enfermedades respiratorias en la clínica pediátrica

Dña. Rosa Alba Sola Martínez
2022

UNIVERSIDAD DE MURCIA

Facultad de Química

Departamento de Bioquímica y Biología Molecular B e
Inmunología

**Análisis del aire exhalado *off-line* para la búsqueda
de biomarcadores de enfermedades respiratorias en
la clínica pediátrica**

TESIS DOCTORAL

Rosa Alba Sola Martínez

DIRECTORES

Dr. Manuel Cánovas Díaz

Dra. Teresa de Diego Puente

Murcia, 2022

Durante su formación predoctoral, Rosa Alba Sola Martínez ha sido beneficiaria de una beca para la Formación de Profesorado Universitario (FPU) concedida por el Ministerio de Ciencia, Innovación y Universidades (MCIU) (FPU18/00545) y una ayuda para estancias breves en la Universidad de Basilea (Suiza) concedida por el Ministerio de Universidades (EST21/00489).

Esta Tesis Doctoral ha sido financiada por los siguientes proyectos:

- “Unravelling in utero determinants predicting lung function in infants: a step for prenatal of asthma”, (PIE15/00051 EXCELENCIA) concedido por el Ministerio de Economía y Competitividad - Instituto de Salud Carlos III.
- “Biología de Sistemas y Sintética de la acetilación-desacetilación del Proteoma de *Escherichia coli*”, (BIO2014-54411-C2-1-R) concedido por el Ministerio de Economía y Competitividad y Fondo Europeo de Desarrollo Regional (FEDER).
- “Desvelando la dinámica del acetiloma para biología sintética en *E.coli*”, (RTI2018-094393-B-C21) concedido por MCIU, Agencia Estatal de Investigación (AEI) y FEDER.
- “Biotecnología de sistemas para la mejora de bioprocesos de producción de proteínas recombinantes en *E.coli*: Integración de la regulación transcripcional y post-traducciona del metabolismo de carbono y nitrógeno”, (20786/PI/18) concedido por la Fundación Séneca.

A lo largo de su formación predoctoral, Rosa Alba Sola Martínez ha contribuido en las siguientes publicaciones:

ARTÍCULOS CIENTÍFICOS

- Lozano-Terol, G., Gallego-Jara, J., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. Engineering protein production by rationally choosing a carbon and nitrogen source using *E. coli* BL21 acetate metabolism knockout strains. *Microb. Cell Fact.* **18**, 151 (2019). DOI: 10.1186/s12934-019-1202-1
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Yanes, O., Cánovas, M., de Diego, T., Vinaixa, M. Exhaled volatile organic compounds analysis in clinical pediatrics: A Systematic Review. *Pediatr. Res.* **89**, 1352-1363 (2021). DOI: 10.1038/s41390-020-01116-8
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Lozano-Terol, G., Gallego-Jara, J., García-Marcos, L., Cánovas, M., de Diego, T. Data preprocessing workflow for exhaled breath analysis by GC/MS using open sources. *Sci. Rep.* **10**, 22008 (2020). DOI: 10.1038/s41598-020-79014-6
- Gallego-Jara, J., Lozano-Terol, G., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. A Comprehensive Review about Taxol®: History and Future challenges. *Molecules.* **25**, 5986 (2020). DOI: 10.3390/molecules25245986
- Lozano-Terol, G., Gallego-Jara, J., **Sola-Martínez, R.A.**, Martínez Vivancos, A., Cánovas, M., de Diego, T. Impact of the expression system on recombinant protein production in *Escherichia coli* BL21. *Front. Microbiol.* **12**, 682001 (2021). DOI: 10.3389/fmicb.2021.682001
- **Sola-Martínez, R.A.**, Lozano-Terol, G., Gallego-Jara, J., Morales, E., Cantero Cano, E., Sanchez-Solís, M., García-Marcos, L., Jiménez-Guerrero, P., Noguera Velasco, J.A., Cánovas, M., de Diego, T. Exhaled volatilome analysis as a useful tool to discriminate asthma with other coexisting atopic diseases in women of childbearing age. *Sci. Rep.* **11**, 13823 (2021). DOI: 10.1038/s41598-021-92933-2
- Gallego-Jara, J., Ortega Retuerta, A., Lozano-Terol, G., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. Bacterial Sirtuins Overview: An Open Niche to Explore. *Front. Microbiol.* **12**, 744416 (2021). DOI: 10.3389/fmicb.2021.744416
- Morales, E., Alcántara-López, M.V., Cabezas Herrera, J., de Diego, T., Hernández Caselles, T., Jiménez-Guerrero, P., Larque Daza, E., Lopez-Soler, C., Martínez Gracia, M. C., Martínez Torres, A.E., Martín-Orozco Santiago, M.E., Mendiola, J., Nieto Diaz, A., Noguera Velasco, J.A., Pérez-Fernández, V., Prieto

- Sánchez, M. T., Salvador García, C., Santaella Pascual, M., **Sola-Martínez, R.A.**, Torres Cantero, A., Yagüe-Guirao, G., Zornoza Moreno, M., García-Marcos, L. The nutrition in early life and asthma (NELA) birth cohort study: Rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310-324 (2022). DOI: 10.1111/ppe.12826
- **Sola-Martínez, R.A.**, Sánchez-Solís, M., Lozano-Terol, G., Gallego-Jara, J., García-Marcos, L., Cánovas, M., de Diego, T. Relationship between lung function and exhaled volatile organic compounds in healthy infants. *Pediatr. Pulmonol.* **57**, 1282-1292 (2022). DOI: 10.1002/ppul.25849
 - **Sola-Martínez, R.A.**, Lozano-Terol, G., Gallego-Jara, J., Morales, E., García-Marcos, L., Noguera-Velasco, J.A., Cánovas, M., de Diego, T. Influence of home indoor dampness exposure on volatile organic compounds in exhaled breath of mothers and their infants: the NELA birth cohort. *Indoor Air* (en revisión).
 - **Sola-Martínez, R.A.**, Decrue, F., Zeng, J., Awchi, M., Gisler, A., Richard, M., Arnold, K., Dev Singh, K., Frey, U., Cánovas, M., de Diego, T., Sinues, P. Breath analysis by off-line SESI-HRMS: conservation of breath samples (en preparación).

CAPÍTULOS DE LIBRO

- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Cánovas, M., de Diego, T. Papel de la acetilación de lisinas sobre la proteína CaiC del operón *caiTABCD E* involucrado en la ruta de biotransformación de la L-carnitina en *Escherichia coli*. En *Caminando hacia el futuro: tendencias metodológicas de los jóvenes investigadores* (eds. Lozano Reina, G., Ponce Gea, A.I., Planes Muñoz, D.) 317-337 (Editum, 2019). ISBN: 978-84-09-09149-2
- Gallego-Jara, J., Lozano-Terol, G., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. Engineering of microbial cell factories for plant based natural products production. En *Microbial cell factories engineering for production of biomolecules* (ed. Singh, V.) 381-392 (Elsevier, 2021). ISBN: 978-0-12-821477-0
- **Sola-Martínez, R.A.**, Lozano-Terol, G., Gallego-Jara, J., Cánovas, M., de Diego, T. Off-line breath analysis: standardization of breath sampling and analysis using mass spectrometry and innovative algorithms. En *Breath Analysis - An approach for smart diagnostics* (Springer Nature) (aceptado).

CONTRIBUCIONES A CONGRESOS

- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., de Diego, T., Cánovas, M. The role of lysine acetylation on the CaiC protein of *caiTABCDE* operon involved in the pathway of L-carnitine biotransformation in *Escherichia coli*. En *I Jornadas Interdisciplinares de Jóvenes Investigadores*, Murcia, España, (octubre, 2017). Comunicación Oral.
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., de Diego, T., Cantero Cano, E., García-Marcos, L., Prieto Sánchez, M. T., Martínez-Gracia, C., Gázquez-García, A., Mendiola-Olivares, J., Valero-Guillén, P.L., Fernández-Breis, J.T., Castillo, R. Searching for asthma biomarkers in early childhood: profiling of volatile organic compounds in children's exhaled breath. En *II Jornadas IMIB-Arrixaca*, Murcia, España, (noviembre, 2017). Comunicación Oral.
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Cantero Cano, E., de Diego, T., García-Marcos, L., Prieto-Sánchez, M.T., Martínez-Gracia, C., Gázquez-García, A., Torres Cantero, A., Yagüe-Guirao, G., Cabezas Herrera, J., Esteban-Gil, A., Castillo, R., Ballesteros, C., Cánovas, M. Analysis of volatile organic compounds in asthmatic mothers as tool for finding asthma biomarkers in early childhood. En *IV Jornadas Doctorales de la Universidad de Murcia*, Murcia, España, (mayo, 2018). Comunicación Oral.
- **Sola Martínez, R.A.**, Pastor Hernández, J.M., de Diego, T., Cánovas, M. N-lysine acetylation of the CaiC protein within the *caiTABCDE* operon involved in *Escherichia coli* L-carnitine metabolism. En *II Jornadas Españolas de Biocatálisis*, Oviedo, España, (junio, 2018). Póster. **PREMIO MEJOR PÓSTER PRESENTADO**
- **Sola Martínez, R.A.**, Pastor Hernández, J.M., de Diego, T., Cánovas, M. N-lysine acetylation of the CaiC protein within the *caiTABCDE* operon involved in *Escherichia coli* L-carnitine production. En *12th European Symposium on Biochemical Engineering Sciences (ESBES)*, Lisboa, Portugal, (septiembre, 2018). Póster.
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Cantero Cano, E., Esteban-Gil, A., Cánovas, M., de Diego, T. Potential of profiling of volatile organic compounds in babies' exhaled breath: diagnosis and monitorization of asthma and other respiratory diseases. En *II Congreso Interdisciplinar de Jóvenes Investigadores*, Murcia, España, (octubre, 2018). Comunicación Oral.
- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Cantero Cano, E., Cánovas, M., de Diego, T. Role of R in searching biomarkers for asthma and other respiratory

diseases in early childhood. En *X Jornadas de Usuarios de R*, Murcia, España, (noviembre, 2018). Comunicación Oral.

- **Sola-Martínez, R.A.**, Pastor Hernández, J.M., Cantero Cano, E., Esteban-Gil, A., Cánovas, M., de Diego, T. Profiling of volatile organic compounds in exhaled breath of babies and their mothers of NELA cohort. En *III Jornadas Científicas del IMIB-Arrixaca*, Murcia, España, (noviembre, 2018). Comunicación Oral.
- **Sola-Martínez, R.A.**, Capellades, J., Pastor Hernández, J.M., Vinaixa, M., Yanes, O., Cánovas, M., de Diego, T. Processing of data obtained from analysis of exhaled air of children and their mothers of NELA cohort using TD-GC-MS. En *V Jornadas Doctorales de la Universidad de Murcia*, Murcia, España, (mayo, 2019). Comunicación Oral.
- **Sola-Martínez, R.A.**, Lozano-Terol, G., Gallego-Jara, J., Cánovas, M., de Diego, T. Development of a platform to analyse human exhaled volatile organic compounds: workflow for biomarker discovery. En *I Symposium on Chemical and Physical Sciences for Young Researchers*, Online, (octubre, 2020). Póster.
- Lozano-Terol, G., Gallego-Jara, J., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. Comparison of Acetate Overflow between *E. coli* BL21 and *E. coli* K12 Strains using Different Carbon and Nitrogen Sources. En *International Conference on Cell and Experimental Biology*, Online, (diciembre, 2020). Póster.
- **Sola-Martínez, R.A.**, Lozano-Terol, G., Gallego-Jara, J., Cánovas, M., de Diego, T. Assessment of season of measurement impact on disease biomarker discovery by exhaled volatilome analysis. En *17th Annual Conference of the Metabolomics Society. METABOLOMICS 2021*, Online, (junio, 2021). Póster.
- de Diego, T., Martínez-Vivancos, A., Lozano-Terol, G., **Sola Martínez, R.A.**, Gallego-Jara, J., Cánovas, M. Carbamoyl Phosphate synthetase of *Escherichia coli*: study of its regulation. En *43rd Annual Meeting of the Spanish Society of Biochemistry & Molecular Biology*, Barcelona, España, (julio, 2021). Póster.
- González Pérez, A., Escudero Pastor, A.I., Carbonell Martínez, A., **Sola-Martínez, R.A.** Evolución de parámetros in vitro en pacientes sensibilizados a proteínas transportadoras de lípidos (LTPs) tras 3 años de tratamiento con inmunoterapia específica. En *VI Jornadas Doctorales de la Universidad de Murcia*, Murcia, España, (junio, 2021). Comunicación Oral.
- González Pérez, A., Navarro Garrido, C., Escudero Pastor, A.I., Miralles López, J.C., **Sola-Martínez, R.A.**, Carbonell Martínez, A. Análisis de calidad de vida en pacientes sensibilizados a proteínas transportadoras de lípidos (LTPs) tras 3 años de tratamiento con inmunoterapia específica, frente al grupo control. En *33^o*

Congreso de la Sociedad Española de Alergología e Inmunología Clínica, Zaragoza, España, (octubre, 2021). Comunicación Oral.

- González Pérez, A., Navarro Garrido, C., Escudero Pastor, A.I., Miralles López, J.C., **Sola-Martínez, R.A.**, Carbonell Martínez, A. Evolución de parámetros in vitro en pacientes sensibilizados a proteínas transportadoras de lípidos (L) tras 3 años de tratamiento con inmunoterapia específica, frente al grupo control. En *33º Congreso de la Sociedad Española de Alergología e Inmunología Clínica*, Zaragoza, España, (octubre, 2021). Póster.
- **Sola-Martínez, R.A.**, Decrue, F., Dev Singh, K., Gisler, A., Awchi, M., Zeng, J., Usemann, J., Frey, U., Sinues, P. Development of a method for off-line breath analysis by secondary electrospray ionization-high resolution mass spectrometry (SESI-HRMS). En *SGMS/SMS Meeting*, Beatenberg, Suiza, (noviembre, 2021). Póster.
- Martínez Vivancos, A., Lozano-Terol, G., Gallego-Jara, J., **Sola-Martínez, R.A.**, Cánovas, M., de Diego, T. Study of Carbamoyl phosphate synthetase enzyme in *Escherichia coli*. En *II Symposium on Chemical and Physical Sciences for Young Researchers*, Murcia, España, (marzo, 2022). Póster.

AGRADECIMIENTOS

A mis directores de Tesis, Manuel Cánovas Díaz y Teresa de Diego Puente, por depositar vuestra confianza en mí y ofrecerme la oportunidad de realizar esta Tesis Doctoral e incorporarme a vuestro grupo de investigación. Manolo, gracias por tus valiosos consejos que me han servido de guía a lo largo de todos estos años de formación universitaria desde el Grado en Bioquímica. Teresa, no tengo palabras para agradecerte toda la ayuda que me has prestado, gracias por volcarte para que este proyecto saliera adelante, por apoyarme y escucharme siempre que lo he necesitado.

A Gema, eres una gran amiga y la mejor compañera de trabajo que se puede tener, afrontar todos los problemas que han surgido durante esta Tesis Doctoral ha sido mucho fácil gracias a ti. A Julia, gracias por estar dispuesta a asesorarnos y ayudarnos en todo momento.

A José María, gracias por todas tus enseñanzas en mis primeros años en el Departamento y por iniciar conmigo la andadura en la línea de investigación del análisis del aire exhalado. Eres uno de los responsables de mi vocación científica. A Ester, gracias por toda tu ayuda durante los primeros años de la Tesis Doctoral, me hiciste sentir como en casa cuando llegué al LAIB.

A todos los compañeros con los que tenido la suerte de contar durante estos años tanto en el Departamento como en el LAIB: Ana, Adrián, Álvaro, David, Elena, Rocío, Susana, Pedro, Rebeca, Antonio, Azahara, Trini, Elena, Marien, Paco...gracias por haber hecho que el día a día fuera mucho más agradable.

Al equipo que conforma el estudio NELA, en especial a los doctores Luis García-Marcos, Manuel Sánchez-Solís y Eva Morales, gracias por vuestra colaboración. Es un honor poder ser miembro de la "familia NELA".

Al grupo del Profesor Óscar Yanes de la Universitat Rovira i Virgili (Reus, Tarragona), en especial a Jordi y Mariona. Gracias por permitirme realizar una estancia en vuestro centro, hacerme sentir como una más, y enseñarme tanto sobre metabolómica.

A todo el grupo del Profesor Pablo Sinues de la Universidad de Basilea (Suiza), gracias por darme la oportunidad de trabajar unos meses con vosotros y aprender mucho más sobre el análisis del aire exhalado, y por acogerme con los brazos abiertos.

A mis amigos, Gloria, María, José, Julia, Cris y Sara, gracias por preocuparos tanto por mí y animarme durante estos años en los momentos difíciles.

A mi hermana Reme, gracias por sacarme siempre una sonrisa y creer más en mí que yo misma.

A mis padres, gracias por confiar siempre en mí, por vuestro apoyo incondicional, por inculcarme vuestra pasión por la ciencia, y por encima de todo transmitirme valores como la honestidad, la responsabilidad, la sensatez y el respeto. Sois los mejores padres que se pueden desear.

A Jesús, gracias por estar siempre conmigo en los buenos y malos momentos, durante todos estos años has sido "mi roca". No tengo palabras para agradecer todo lo que ha significado para mí tenerte en mi vida.

RESUMEN

Durante la respiración humana, es posible distinguir una gran variedad de compuestos orgánicos volátiles (VOCs) en el aire exhalado¹. Sin embargo, el conjunto de VOCs, “volatiloma”, presentes en el aire exhalado y su concentración en cada individuo pueden depender de diversos factores, como diferencias en el metabolismo, presencia de patologías o exposición a contaminantes ambientales, entre otros². Por todo ello, en los últimos años, el aire exhalado humano ha sido considerado una fuente prometedora de biomarcadores para el diagnóstico y la monitorización de numerosas patologías como enfermedades respiratorias de alta prevalencia³⁻⁵. Asimismo, también se ha sugerido que el análisis del aire exhalado podría ser útil para controlar las exposiciones a contaminantes ambientales y para el seguimiento del estrés oxidativo, puesto que muchos VOCs son producidos en procesos metabólicos como la peroxidación lipídica^{2,5-7}. Una de las principales ventajas de este enfoque, es que se trata de una metodología no invasiva, lo que es esencial en pacientes pasivos no cooperativos, como la población pediátrica⁸. La estrategia más extendida para la determinación de VOCs en el aliento humano es el análisis del aire exhalado fuera de línea u *off-line* mediante tecnologías basadas en espectrometría de masas⁹. En el análisis del aire exhalado *off-line*, las muestras de aliento son recogidas en contenedores de aire exhalado (ej. bolsas de muestreo de gas) y luego son analizadas principalmente mediante sistemas analíticos *off-line* (ej. cromatografía de gases acoplada a espectrometría de masas), que permiten un paso previo de preconcentración de las muestras, y más recientemente también mediante sistemas analíticos en línea u *on-line* donde las muestras no reciben ningún tratamiento antes del análisis^{10,11}.

El objetivo principal de esta Tesis Doctoral ha sido aplicar y avanzar en el desarrollo del análisis del aire exhalado *off-line* para la búsqueda de biomarcadores de enfermedades respiratorias y/o atópicas en población pediátrica y en mujeres en edad fértil. Gran parte de la presente Tesis Doctoral se enmarca dentro del estudio NELA (*Nutrition in Early Life and Asthma*)¹² que ha reclutado una cohorte de 738 parejas madre/hijo de la Región de Murcia desde la semana 20 de embarazo, con la finalidad de esclarecer el origen del asma durante la infancia y determinar qué factores influyen en su desarrollo. El objetivo principal de esta Tesis Doctoral puede dividirse en los siguientes objetivos particulares: 1) Realizar una revisión sistemática de los estudios previos donde se ha aplicado el análisis del aire exhalado en población pediátrica; 2) Desarrollar un protocolo de recogida de aire exhalado para adultos y niños de corta edad, y optimizar el análisis del aire exhalado *off-line* mediante TD-GC/q-MS (desorción

térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple); 3) Recoger y analizar las muestras de aire exhalado de las parejas de madres e hijos de la cohorte NELA; 4) Establecer un flujo de trabajo de código abierto para el preprocesamiento de datos brutos obtenidos por espectrometría de masas; 5) Identificar VOCs en el aire exhalado de mujeres en edad fértil que permitan discriminar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas; 6) Identificar factores que influyan en el análisis de datos de volatiloma y establecer estrategias para reducir los sesgos introducidos; 7) Estudiar la relación entre los VOCs exhalados y la función pulmonar en lactantes sanos; 8) Determinar la influencia de la exposición a la humedad en los hogares sobre los VOCs en el aire exhalado y 9) Evaluar el efecto del almacenamiento de muestras de aire exhalado antes de su análisis mediante sistemas analíticos *on-line*.

A continuación, se resumen brevemente las principales metodologías empleadas y los resultados más relevantes obtenidos en la presente Tesis Doctoral:

En el Capítulo 3, se llevó a cabo una revisión sistemática cuyo objetivo era determinar el estado actual de la técnica del análisis del aire exhalado para el diagnóstico y/o monitorización terapéutica de enfermedades en niños y adolescentes, así como establecer las pautas para conseguir la traslación clínica de esta metodología. La revisión sistemática se realizó siguiendo las directrices de la declaración PRISMA¹³ y el protocolo de la revisión fue registrado en la base de datos prospectiva internacional de salud PROSPERO (CRD42020151186). La búsqueda de los artículos se realizó empleando las bases de datos Web of Science, SCOPUS y PubMed; y de los artículos seleccionados se extrajo información sobre el diseño experimental, la población de estudio, los protocolos de toma de muestra, las técnicas analíticas empleadas, los métodos utilizados para el análisis de datos y los resultados obtenidos. Además, se realizó un compendio de todos los VOCs presentes en el aire exhalado señalados en los estudios como relevantes para el diagnóstico y/o monitorización para patologías respiratorias, infecciosas y metabólicas. Adicionalmente, con la finalidad de conocer la red metabólica relacionada con los VOCs seleccionados en los estudios se realizó un análisis de enriquecimiento funcional. Por último, se evaluó la calidad de los estudios mediante el sistema de puntuación QUADAS-2¹⁴. En base a toda la información recopilada en esta revisión sistemática, se concluyó que el análisis del aire exhalado en pediatría clínica se encontraba todavía en una fase exploratoria (fase de descubrimiento de biomarcadores), y que la adopción de las prácticas habituales de la metabolómica podría ayudar a conseguir la utilidad clínica del análisis del volatiloma exhalado (grandes

poblaciones de estudio, evaluación de la influencia ambiental, transparencia en la etapa del preprocesamiento de datos, información sobre la identificación de los VOCs, validación externa, etc.), lo que se tuvo en consideración en el resto de trabajos incluidos en la presente Tesis Doctoral. A partir de este riguroso análisis se concluyó que la técnica de determinación de VOCs en aliento mediante cromatografía gases-espectrometría de masas acoplada a desorción térmica era actualmente la más adecuada en esta etapa de descubrimiento de biomarcadores.

En el Capítulo 4 se expone el desarrollo de un flujo de trabajo transparente y reproducible de código abierto para la gestión del preprocesamiento de los datos obtenidos en el análisis químico de muestras de aire exhalado mediante un sistema de desorción térmica acoplado a cromatografía de gases-espectrometría de masas de cuadrupolo simple (TD-GC/q-MS). Este flujo de trabajo combina las dos estrategias más extendidas para el preprocesamiento de datos generados por espectrometría de masas: la detección de los fragmentos iónicos o *features* presentes en las muestras de aire exhalado y la detección de los compuestos de las muestras de aliento. Así, en este flujo de trabajo se han aplicado funciones de tres paquetes gratuitos escritos en lenguaje de programación R (*xcms*¹⁵, *cliqueMS*¹⁶ y *eRah*¹⁷). Gracias a este flujo de trabajo, es posible transformar el gran volumen de datos brutos obtenidos por espectrometría de masas en una matriz de datos útil para su posterior análisis estadístico, y conocer a que compuestos pertenecen. Asimismo, este flujo de trabajo permite la comparación de los compuestos presentes en las muestras de aire exhalado con una biblioteca espectral para la identificación de los mismos. Las muestras de aire exhalado de las madres y los niños de la cohorte de nacimiento NELA, así como las muestras de aire ambiental recogidas en el momento del muestreo fueron utilizadas para el diseño y la optimización del flujo de trabajo de preprocesamiento de datos. Además, en este capítulo se presenta un protocolo de análisis del aire exhalado *off-line* mediante TD-GC/q-MS, y un protocolo para la recogida de forma no invasiva de muestras de aire exhalado en adultos y adaptado para niños menores de 2 años.

En el Capítulo 5 se muestra cómo el aire exhalado humano puede constituir una prometedora fuente de biomarcadores de enfermedades de alta prevalencia como el asma, si se aplican las prácticas consolidadas en el campo de la metabolómica. En base a VOCs detectados en el aire exhalado mediante TD-GC/q-MS de mujeres en edad fértil, fue posible discriminar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas). La población de estudio estuvo conformada por 336 mujeres en edad fértil (madres de la cohorte NELA), que se

dividieron en un conjunto de entrenamiento (n=211) y un conjunto de validación o testeo (n= 125). El conjunto de entrenamiento se empleó para la construcción de modelos predictivos usando un análisis de datos robusto con LASSO (operador de selección y contracción mínima absoluta) combinado con regresión logística. El rendimiento de los modelos generados se evaluó mediante validación cruzada de 5 iteraciones junto una prueba de permutación, y mediante validación sobre el conjunto de testeo. Así, se obtuvieron 15 modelos estadísticamente significativos que discriminaban el asma con otras enfermedades atópicas coexistentes en mujeres en edad reproductiva. En cuanto a los VOCs que se incluyeron en los perfiles discriminantes, la acetona, el 2-etil-1-hexanol y un derivado de la tetrahidroisoquinolina actuaron como discriminantes del asma con otras enfermedades atópicas coexistentes, y el disulfuro de carbono, un derivado de la tetrahidroisoquinolina, el 2-etil-1-hexanol y el decano permitieron la diferenciación de mujeres asmáticas con otras enfermedades atópicas coexistentes de mujeres que no padecían asma pero sí sufrían otros trastornos atópicos. Además, en este capítulo, se analizaron en detalle los datos recopilados por las dos últimas encuestas de salud en España elaboradas por el Instituto Nacional de Estadística (INE) para identificar factores subyacentes relacionados con el asma a través de un análisis de correspondencias múltiples (MCA). Asimismo, también se evaluó el impacto de la época del año cuando se realizan las determinaciones de VOCs, sobre el proceso de descubrimiento de biomarcadores de enfermedades mediante el análisis del aire exhalado *off-line*, y se desarrollaron estrategias como la implementación de ASCA (ANOVA - análisis de componentes simultáneos) para afrontar los posibles sesgos introducidos.

En el Capítulo 6 se estudió la relación del volatilo exhalado y la función pulmonar en lactantes sanos. Concretamente, se realizó un análisis dirigido de los datos y se compararon los niveles de seis VOCs (acetona, isopreno, decano, undecano, tetradecano y pentadecano) en el aire exhalado de 50 niños de la cohorte NELA a los 3 meses de edad, y los valores medidos de los siguientes parámetros de función pulmonar: capacidad vital forzada (FVC), volumen espiratorio forzado a 0,5 s ($FEV_{0,5}$), flujo espiratorio forzado al 75% de la FVC (FEF_{75}), flujo espiratorio forzado al 25%-75% de la FVC (FEF_{25-75}) y el cociente $FEV_{0,5}/FVC$. La medición de la función pulmonar se basó en la técnica de compresión toracoabdominal rápida con volumen pulmonar aumentado, y las determinaciones de VOCs en el aire exhalado se realizaron mediante TD-GC/q-MS siguiendo los protocolos descritos en el Capítulo 4. La evaluación de la relación entre los VOCs exhalados y la función pulmonar se llevó a cabo mediante correlaciones de Pearson, regresión lineal múltiple y análisis de componentes

principales. De este modo, se observó una relación negativamente significativa entre los niveles en el aire exhalado de acetona, tetradecano y pentadecano, y los valores de algunos de los parámetros de función pulmonar (FVC, FEV_{0,5} y FEF₂₅₋₇₅), demostrándose la asociación entre la función pulmonar y VOCs relacionados con el estrés oxidativos en bebés sanos.

En el Capítulo 7, se empleó el análisis del aire exhalado como herramienta para evaluar la exposición ambiental, concretamente, la exposición a la humedad en el hogar. Para ello, se determinaron mediante TD-GC/q-MS los niveles de VOCs exógenos y endógenos (acetona, isopreno, tolueno, p/m-xileno, o-xileno, estireno, benzaldehído, naftaleno y 2-etil-1-hexanol) en el aire exhalado de 337 parejas de madres e hijos de la cohorte NELA. Utilizando pruebas de contraste de hipótesis, se compararon los niveles de estos 9 VOCs entre los sujetos expuestos y no expuestos a humedad en sus residencias habituales, y se comprobó que los niveles de 2-etil-1-hexanol estaban significativamente más elevados en el aliento de las madres que vivían en casas que presentaban humedad en las paredes. Asimismo, también se observó que la exposición a la humedad en el hogar también influyó significativamente en los recuentos en sangre de eosinófilos de las madres.

En el Capítulo 8, se llevó a cabo un estudio enfocado al empleo de sistemas analíticos *on-line* como SESI-HRMS (ionización secundaria por *electrospray*-espectrometría de masas de alta resolución) adaptados al análisis del aire exhalado *off-line*. Con este fin, se ha evaluado el efecto del almacenamiento de las muestras de aire exhalado en bolsas de muestreo de gas antes de su análisis químico. Para ello, utilizando un método recientemente desarrollado por el grupo del Prof. Sinues de la Universidad de Basilea para realizar un análisis del aire exhalado *off-line* mediante SESI-HRMS validado tanto en adultos como en neonatos¹¹, se han recogido y analizado 307 muestras de aire exhalado de adultos sanos que han sido almacenadas durante diferentes periodos de tiempo y a distintas temperaturas (temperatura ambiente y hielo seco). En este sentido, se ha analizado en detalle el efecto del almacenamiento de 27 aldehídos relacionados con la peroxidación lipídica pertenecientes a las familias 2-alquenales, 4-hidroxi-2-alquenales y 4-hidroxi-2,6-alcadienales¹⁸. Tras un almacenamiento de 2 horas a temperatura ambiente la intensidad de todas las señales *m/z* (masa/carga) disminuyó de media menos de un 20% con respecto a las muestras analizadas sin ser almacenadas previamente, lo que indica que es posible conservar muestras de aire exhalado sin pérdidas sustanciales para su análisis posterior. La

conservación de las muestras puede flexibilizar su gestión analítica en el hospital e incluso reclutar pacientes de otros hospitales.

Finalmente, las principales conclusiones que se pueden extraer de la presente Tesis Doctoral son: 1) La introducción de las prácticas típicas del campo de la metabolómica es fundamental para alcanzar la traslación clínica del análisis del aire exhalado; 2) Se ha diseñado un protocolo para la recogida y análisis de muestras de aire exhalado mediante TD-GC/q-MS adecuado para población adulta y pediátrica; 3) Se ha desarrollado un flujo de trabajo de código abierto para el preprocesamiento de datos obtenidos en el análisis del aire exhalado *off-line* mediante TD-GC/q-MS; 4) El análisis del aire exhalado *off-line* mediante TD-GC/q-MS en mujeres en edad fértil permite diferenciar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas); 5) Se ha caracterizado la varianza causada por los factores experimentales por implementación de ASCA, que combina el análisis de varianza (ANOVA) y el análisis de componentes simultáneos y se ha solventado el sesgo causado por la estación del año en la que se recolectaban las muestras de aliento; 6) Existe una asociación entre los VOCs exhalados y la función pulmonar en lactantes sanos; 7) La exposición prolongada a humedad en el hogar puede influir en el volatiloma exhalado, concretamente en los niveles del contaminante 2-etil-1-hexanol en mujeres en edad fértil y 8) Las muestras de aire exhalado pueden ser almacenadas durante horas en bolsas de muestreo de gas sin una pérdida determinante de intensidad de los compuestos, y posteriormente ser analizadas mediante SESI-HRMS.

BIBLIOGRAFÍA

1. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
2. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
3. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
4. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation

- partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
5. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
 6. Ferraro, V., Carraro, S., Bozzetto, S., Zanconato, S. & Baraldi, E. Exhaled biomarkers in childhood asthma: old and new approaches. *Asthma Res. Pract.* **4**, 9 (2018).
 7. Ratcliffe, N. *et al.* A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *J. Breath Res.* **14**, 034001 (2020).
 8. Neerincx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
 9. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
 10. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).
 11. Decrue, F. *et al.* Combination of Exhaled Breath Analysis with Parallel Lung Function and FeNO Measurements in Infants. *Anal. Chem.* **93**, 15579–15583 (2021).
 12. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).
 13. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **6**, e1000097 (2009).
 14. Whiting, P. F. *et al.* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* **155**, 529 (2011).
 15. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
 16. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution

- similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
17. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
 18. García-Gómez, D. *et al.* Identification of 2-alkenals, 4-hydroxy-2-alkenals, and 4-hydroxy-2,6-alkadienals in exhaled breath condensate by UHPLC-HRMS and in breath by real-time HRMS. *Anal. Chem.* **87**, 3087–3093 (2015).

SUMMARY

During human exhalation, a wide range of volatile organic compounds (VOCs) can be distinguished in exhaled breath¹. However, the whole set of VOCs, "volatilome", detected in exhaled breath and their concentration in each subject depends on several factors, such as differences in metabolism, presence of pathologies or exposure to environmental pollutants, among others². For all these reasons, in recent years, human exhaled breath has been considered as a promising source of biomarkers for the diagnosis and monitoring of numerous disorders such as prevalent respiratory diseases³⁻⁵. Furthermore, it has also been suggested that breath analysis could be useful for environmental exposure and oxidative stress monitoring, since many VOCs are produced in metabolic processes such as lipid peroxidation^{2,5-7}. One of the main advantages of this approach is that it is a noninvasive methodology, which is essential in passive non-cooperative patients, such as pediatric population⁸. The most widespread strategy for VOC determination in exhaled breath is off-line breath analysis by mass spectrometry-based technologies⁹. In off-line breath analysis, breath samples are collected in breath containers (e.g. gas sampling bags) and then analyzed mainly by off-line analytical systems (e.g. gas chromatography coupled to mass spectrometry), which allow a sample preconcentration step, and more recently also by on-line analytical systems where the samples do not receive any treatment prior to analysis^{10,11}.

The main objective of this thesis has been to apply and advance in the development of off-line breath analysis for searching for biomarkers of respiratory and/or atopic diseases in the pediatric population and in women of childbearing age. The present thesis is in the context of the NELA study (*Nutrition in Early Life and Asthma*)¹² which has recruited a large part of a cohort of 738 mother/child pairs in the Region of Murcia from the 20th week of pregnancy, in order to unravel the asthma origin during childhood and to determine which factors influence in its development. The main objective of this thesis can be divided into the following particular objectives: 1) To perform a systematic review of previous studies where exhaled breath analysis has been applied in pediatric population; 2) To develop a protocol for exhaled breath collection for adults and young children, and to optimize off-line breath analysis by TD-GC/q-MS (thermal desorption coupled to gas chromatography-single quadrupole mass spectrometry); 3) To collect and analyze exhaled breath samples from mother-infant pairs of the NELA cohort; 4) To establish an open-source workflow for preprocessing raw data obtained by mass spectrometry; 5) To identify VOCs in the exhaled breath of women of childbearing age which discriminate between asthmatics with other coexisting

atopic diseases and non-asthmatics; 6) To identify factors that affect the analysis of volatilomics data and to establish strategies to reduce the biases introduced; 7) To study the relationship between exhaled VOCs and lung function in healthy infants; 8) To determine the influence of exposure to household dampness on VOCs in exhaled breath; and 9) To evaluate the effect of exhaled breath storage prior to analysis using on-line analytical systems.

The main methodologies employed and the most relevant results obtained in this thesis are briefly summarized below:

In Chapter 3, a systematic review was carried out with the aim to determine the current state of the art of breath analysis for diagnosis and/or therapeutic monitoring of diseases in children and adolescents, as well as to establish the guidelines to achieve the clinical translation of this methodology. The systematic review was performed following the guidelines of the PRISMA¹³ statement and the review protocol was registered in the international prospective health database PROSPERO (CRD42020151186). The search for articles was performed on the Web of Science, SCOPUS and PubMed databases; and information on the experimental design, study population, sampling protocols, analytical techniques, methods for data analysis and results obtained was extracted from the selected articles. In addition, a summary was compiled of all the VOCs present in exhaled breath that were identified in the studies as relevant for the diagnosis and/or monitoring of respiratory, infectious and metabolic pathologies. Additionally, in order to find out the metabolic network related to the VOCs selected in the studies, a functional enrichment analysis was performed. Finally, the quality of the studies was assessed using the scoring system QUADAS-2¹⁴. Based on all the information gathered in this systematic review, it was concluded that breath analysis in clinical pediatrics was still in an exploratory phase (biomarker discovery phase), and that the adoption of well-established metabolomics practices could help to achieve clinical utility of exhaled volatilome analysis (large study populations, assessment of environmental influence, transparency at the data preprocessing stage, reporting on VOCs identification, external validation, etc.), which was taken into consideration in the rest of the studies included in this thesis. Based on this rigorous analysis, it was concluded that the technique for the determination of VOCs in exhaled breath by gas chromatography-mass spectrometry coupled to thermal desorption was currently the most appropriate at the biomarker discovery stage.

Chapter 4 shows the development of a transparent and reproducible open-source workflow for the preprocessing of data obtained in the chemical analysis of exhaled

breath samples using a thermal desorption system coupled to gas chromatography-single quadrupole mass spectrometry (TD-GC/q-MS). This workflow combines the two most widespread approaches for the preprocessing of data generated by mass spectrometry: detection of ion fragments or features from exhaled breath samples and compound detection from breath samples. Thus, functions from three free packages written in R programming language (*xcms*¹⁵, *cliqueMS*¹⁶ y *eRah*¹⁷) were used in this workflow. Thanks to this workflow, it is possible to transform the large volume of raw data obtained by mass spectrometry into a data matrix useful for further statistical analysis, and to determine to which compounds they belong. Furthermore, this workflow enables the matching of the compounds from exhaled breath samples with a spectral library for VOCs identification. Exhaled breath samples from the mothers and children of the NELA birth cohort, as well as ambient air samples collected at the time of breath sampling were used for the design and the optimization of our data preprocessing workflow. In addition, this chapter presents a protocol for off-line breath analysis by TD-GC/q-MS, and a protocol for noninvasive collection of exhaled breath samples in adults and adapted for children under 2 years of age.

Chapter 5 shows how human exhaled breath can be a promising source of biomarkers of prevalent diseases such as asthma, if well-established practices in the field of metabolomics are applied. Based on VOCs detected in exhaled breath by TD-GC/q-MS from women of childbearing age, it was possible to discriminate between asthmatics with other coexisting atopic diseases and non-asthmatics (with or without atopic diseases). The study population consisted of 336 women of childbearing age (mothers of the NELA cohort), who were divided into a training set (n=211) and a validation or test set (n= 125). The training set was used for predictive model building using robust data analysis with LASSO (least absolute shrinkage and selection operator) combined with logistic regression. The performance of the generated models was assessed by 5-fold cross-validation together with a permutation test, and by validation on the testing set. Thus, 15 statistically significant models that discriminated asthma with other coexisting atopic diseases in women of reproductive age were obtained. Regarding the VOCs that were included in the discriminant profiles, acetone, 2-ethyl-1-hexanol and a tetrahydroisoquinoline derivative were found as discriminants of asthma with other coexisting atopic diseases, and carbon disulfide, a tetrahydroisoquinoline derivative, 2-ethyl-1-hexanol and decane enabled the differentiation of asthmatic women with other coexisting atopic diseases from women who did not suffer from asthma but have other atopic disorders. In addition, in this chapter, the data collected by the last two health surveys in Spain elaborated by the National Institute of Statistics (INE) were analyzed in

detail to identify factors related to asthma by means of multiple correspondence analysis (MCA). Furthermore, the impact of the season of the year when VOC determinations are performed on the process of disease biomarker discovery by means of off-line breath analysis was also evaluated, and strategies such as the implementation of ASCA (ANOVA - simultaneous components analysis) were implemented to deal with the possible biases introduced.

Chapter 6 studied the relationship of exhaled volatilome and lung function in healthy infants. Specifically, a targeted data analysis was performed and the levels of six VOCs (acetone, isoprene, decane, undecane, tetradecane, and pentadecane) in the exhaled breath of 50 infants from the NELA cohort at 3 months of age, and the measured values of the following lung function parameters were compared: forced vital capacity (FVC), forced expiratory volume at 0.5 s ($FEV_{0.5}$), forced expiratory flow at 75% of FVC (FEF_{75}), forced expiratory flow at 25%-75% of FVC (FEF_{25-75}) and the $FEV_{0.5}/FVC$ ratio. The measurement of lung function was based on the rapid thoracoabdominal compression technique with increased lung volume, and VOC determinations in exhaled breath were performed by TD-GC/q-MS using the protocols described in Chapter 4. The assessment of the relationship between exhaled VOCs and lung function was carried out by Pearson's correlations, multiple linear regression and principal component analysis. Thus, a negatively significant relationship was observed between the levels in exhaled breath of acetone, tetradecane and pentadecane, and the values of some of the lung function parameters (FVC, $FEV_{0.5}$ and FEF_{25-75}), demonstrating the association between lung function and VOCs related to oxidative stress in healthy infants.

In Chapter 7, breath analysis was used as a tool to assess environmental exposure, specifically, exposure to home indoor dampness. For this purpose, levels of exogenous and endogenous VOCs (acetone, isoprene, toluene, p/m-xylene, o-xylene, styrene, benzaldehyde, naphthalene, and 2-ethyl-1-hexanol) were determined by TD-GC/q-MS in the exhaled breath of 337 mother-child pairs from the NELA cohort. Using statistical hypothesis testing, the levels of these 9 VOCs were compared between subjects exposed and not exposed to dampness in their usual residences, and it was found that 2-ethyl-1-hexanol levels were significantly higher in the exhaled breath of mothers who live in houses with damp stains or moulds on the walls. In addition, it was also observed that exposure to home indoor dampness also significantly impact on the eosinophil blood counts of the mothers.

In Chapter 8, a study focused on the use of on-line analytical systems such as SESI-HRMS (secondary electrospray ionization-high resolution mass spectrometry)

adapted to off-line breath analysis. For this purpose, the effect of storage of exhaled breath samples in gas sampling bags prior to chemical analysis has been assessed. So, 307 exhaled breath samples from healthy adults that have been stored for different periods of time and at different temperatures (room temperature and dry ice) were collected and analyzed using a new method developed by the group of Prof. Sinues at the University of Basel to conduct off-line breath analysis by SESI-HRMS validated in both adults and neonates¹¹. In this regard, the effect of storage of 27 aldehydes related to lipid peroxidation from the families 2-alkenals, 4-hydroxy-2-alkenals and 4-hydroxy-2,6-alkadienals¹⁸ has been studied in detail. After storage for 2 hours at room temperature the intensity of all m/z (mass-to-charge ratio) signals decreased on average by less than 20% compared to samples analyzed without prior storage, indicating that it is possible to conserve exhaled breath samples without substantial loss for subsequent analysis. Sample conservation could increase the flexibility of analytical handling in the hospital and even recruit patients from other hospitals.

Finally, the main conclusions of this thesis are: 1) The introduction of typical practices from the field of metabolomics is fundamental to achieve the clinical translation of breath analysis; 2) A protocol for the collection and analysis of exhaled breath samples by TD-GC/q-MS suitable for adult and pediatric population was designed; 3) An open-source workflow has been developed for the preprocessing of data obtained in off-line breath analysis by TD-GC/q-MS; 4) Off-line breath analysis by TD-GC/q-MS in women of childbearing age allows differentiation between asthmatics with other coexisting atopic diseases and non-asthmatics (with or without atopic diseases); 5) Variance caused by experimental factors has been characterized by implementation of ASCA, which combines analysis of variance (ANOVA) and simultaneous components analysis and solved the bias caused by the season of the year when breath samples were collected; 6) There is an association between exhaled VOCs and lung function in healthy infants; 7) Prolonged exposure to indoor dampness can influence the exhaled volatilome, specifically levels of the pollutant 2-ethyl-1-hexanol in women of childbearing age; and 8) Exhaled breath samples can be stored for hours in gas sampling bags without a substantial loss of compound intensity, and subsequently analyzed by SESI-HRMS.

REFERENCES

1. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
2. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in

- exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
3. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
 4. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
 5. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
 6. Ferraro, V., Carraro, S., Bozzetto, S., Zanconato, S. & Baraldi, E. Exhaled biomarkers in childhood asthma: old and new approaches. *Asthma Res. Pract.* **4**, 9 (2018).
 7. Ratcliffe, N. *et al.* A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *J. Breath Res.* **14**, 034001 (2020).
 8. Neerincx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
 9. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
 10. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).
 11. Decrue, F. *et al.* Combination of Exhaled Breath Analysis with Parallel Lung Function and FeNO Measurements in Infants. *Anal. Chem.* **93**, 15579–15583 (2021).
 12. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).
 13. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **6**, e1000097 (2009).

14. Whiting, P. F. *et al.* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* **155**, 529 (2011).
15. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
16. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
17. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
18. García-Gómez, D. *et al.* Identification of 2-alkenals, 4-hydroxy-2-alkenals, and 4-hydroxy-2,6-alkadienals in exhaled breath condensate by UHPLC-HRMS and in breath by real-time HRMS. *Anal. Chem.* **87**, 3087–3093 (2015).

ÍNDICE DE ABREVIATURAS

4-HNE: 4-hidroxi-2-nonal	d-PLS-DA: disimilitud-análisis discriminante de mínimos cuadrados parciales
A-AD: asmáticas con otras enfermedades atópicas coexistentes	DVB: divinilbenceno
AaOx: aminoácido oxidasa.	EBC: aire exhalado condensado
Acox: acil-coenzima A oxidasa	EDC: disruptor endocrino
AMPc: adenosín monofosfato cíclico	EI: ionización electrónica
ANCOVA: análisis de covarianza	EIC: cromatograma de iones extraídos
ANNs: redes neuronales artificiales	E-nose: nariz electrónica
ANOVA: análisis de varianza	Error RI: error del índice de retención
ARDS: síndrome de dificultad respiratoria aguda	ESI: ionización por <i>electrospray</i>
ASCA: ANOVA - análisis de componentes simultáneos	FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC
AUC: área bajo la curva	FEF₇₅: flujo espiratorio forzado al 75% de la FVC
AUC_{cv}: AUC obtenida mediante validación cruzada de 5 iteraciones	FeNO: fracción exhalada de óxido nítrico
AUC_{vs}: AUC obtenida en la validación sobre el conjunto de testeo.	FEV_{0,5}: volumen espiratorio forzado a 0,5 s
BAL: broncoscopia con lavado broncoalveolar	FEV₁: volumen espiratorio forzado a 1 s
BCA: aparato de recogida del aliento	FVC: capacidad vital forzada
CAR: carboxeno	GC: cromatografía de gases
CD: enfermedad de Crohn	GC-MS: cromatografía de gases-espectrometría de masas
CDA: análisis discriminante canónico	GCxGC: cromatografía de gases bidimensional integral
CF: fibrosis quística	HCTR: controles sanos
CI: ionización química	HS-SPME: espacio de cabeza-microextracción en fase sólida
CKD: enfermedad renal crónica	IBD: enfermedad inflamatoria intestinal
CLD: enfermedad hepática crónica	IC 95%: intervalo de confianza del 95%
CMLC: agrupamiento de los compuestos por covarianza local	IMC: índice de masa corporal
COVID-19: enfermedad por coronavirus 2019	IMR: reacción ion-molecular
CTR: controles	IQR: rango intercuartílico
CV: validación cruzada	KTx: trasplante renal
DC: células dendríticas	LASSO: operador de selección y contracción mínima absoluta
DE: desviación estándar	LDA: análisis lineal discriminante
DEHP: di(2-etilhexil)ftalato	Lin's CCC: coeficiente de correlación de concordancia de Lin
Dim: dimensión	LLN: límite inferior de la normalidad

LR: regresión logística

m/z: relación masa-carga

MCA: análisis de correspondencias múltiples

MIR: infrarrojo medio

MLR: regresión logística múltiple

MS: espectrometría de masas

NA: no asmáticas

NA-AD: no asmáticas con otras enfermedades atópicas

NALFD: enfermedad del hígado graso no alcohólico

NA-NAD: no asmáticas sin enfermedades atópicas

Nox: NADPH oxidasa

NTDs: dispositivos de trampa de aguja

OR: *odd ratio*

OSAS: síndrome de apnea obstructiva del sueño

OSD: deconvolución de señales ortogonales

OVM: monitores de vapores orgánicos

PA: poliacrilato

PC: componente principal

PCA: análisis de componentes principales

PCD: discinesia ciliar primaria

PCR: regresión de componentes principales

PDMS: polidimetilsiloxano

PLS: regresión de mínimos cuadrados parciales

PLS-DA: análisis discriminante mínimos cuadrados parciales

PTR: reacción de transferencia de protones

PTR-MS: reacción de transferencia de protones - espectrometría de masas

PUFAs: ácidos grasos poliinsaturados

PVC: policloruro de vinilo

q-MS: espectrómetro de masas de cuadrupolo simple

RF: bosque aleatorio o *random forest*

RI: índice de retención

ROS: especies reactivas de oxígeno

RT: tiempo de retención

RW: sibilancias recurrentes

SBS: síndrome del edificio enfermo

SESI: ionización secundaria por *electrospray*

SESI-HRMS: ionización secundaria por *electrospray* – espectrometría de masas de alta resolución

SESI-MS: ionización secundaria por *electrospray* – espectrometría de masas

SIFT: tubo de selección de iones con flujo

SIFT-MS: tubo de selección de iones con flujo - espectrometría de masas

SIM: monitorización de iones seleccionados

SLR: regresión logística dispersa

s-PLS-DA: análisis discriminante mínimos cuadrados parciales disperso

SPME: microextracción en fase sólida

SVM: máquinas de vector soporte

TD: desorción térmica

TD-GC/q-MS: sistema acoplado de desorción térmica y cromatografía de gases-espectrometría de masas de cuadrupolo simple

TIC: cromatograma de iones total

TOF: tiempo de vuelo

tubos TD: tubos de desorción térmica

UC: colitis ulcerosa

ULN: límite superior de la normalidad

VOCs: compuestos orgánicos volátiles

WBCs: recuento de glóbulos blancos

	ÍNDICE
Resumen	i
<i>Summary</i>	ix
Índice de abreviaturas	xvii
Capítulo 1. Introducción	1
Introducción: análisis del aire exhalado	3
Análisis del aire exhalado <i>off-line</i>	7
Factores que influyen en el análisis del aire exhalado <i>off-line</i>	7
Recogida de muestras de aire exhalado	8
Porciones de muestreo del aire exhalado	8
Contenedores para la recogida de aire exhalado	10
Estandarización del muestreo del aire exhalado	11
Plataformas analíticas basadas en espectrometría de masas	12
Sistemas analíticos <i>off-line</i>	12
Métodos de preconcentración	13
Sistemas analíticos <i>on-line</i>	14
Preprocesamiento de datos	16
Análisis de datos	19
Algoritmos de aprendizaje automático	19
Validación de los modelos	21
Conclusiones y perspectivas	22
Bibliografía	24
Capítulo 2. Objetivos	35
Capítulo 3. Análisis de compuestos orgánicos volátiles exhalados en pediatría clínica: revisión sistemática	39
Resumen	41
Introducción	41
Resultados	43
Discusión	58
Conclusión	62
Materiales y métodos	63
Bibliografía	66
Apéndice	73
Capítulo 4. Flujo de trabajo de código abierto para preprocesamiento de datos en el análisis del aire exhalado <i>off-line</i> mediante GC-MS	83
Resumen	85

Introducción	85
Resultados	88
Discusión	101
Conclusión	104
Materiales y métodos	105
Bibliografía	109
Apéndice	113
<hr/>	
Capítulo 5. Análisis del volatiloema exhalado como herramienta útil para discriminar el asma con otras enfermedades atópicas coexistentes en mujeres en edad fértil	121
<hr/>	
Resumen	123
Introducción	123
Resultados	127
Discusión	139
Conclusión	144
Materiales y métodos	145
Bibliografía	153
Apéndice	161
<hr/>	
Capítulo 6. Relación entre la función pulmonar y los compuestos orgánicos volátiles exhalados en lactantes sanos	175
<hr/>	
Resumen	177
Introducción	177
Resultados	179
Discusión	186
Conclusión	188
Materiales y métodos	188
Bibliografía	193
Apéndice	197
<hr/>	
Capítulo 7. Influencia de la exposición a la humedad en el interior del hogar sobre los niveles de compuestos orgánicos volátiles en el aire exhalado	203
<hr/>	
Resumen	205
Introducción	205
Resultados	207
Discusión	218
Conclusión	220
Materiales y métodos	221

Bibliografía	225
Apéndice	231
Capítulo 8. Análisis del aire exhalado mediante <i>off-line</i> SESI-HRMS:	
conservación de muestras de aliento	235
Resumen	237
Introducción	237
Resultados	239
Discusión	251
Conclusión	254
Materiales y métodos	254
Bibliografía	257
Apéndice	261
Capítulo 9. Conclusiones/Conclusions	263

Capítulo 1

Introducción

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, Gema Lozano-Terol, Julia Gallego-Jara, Manuel Cánovas Díaz y Teresa de Diego Puente

Off-line breath analysis: standardization of breath sampling and analysis using mass spectrometry and innovative algorithms

En *Breath Analysis - An approach for smart diagnostics* (Springer Nature)
(aceptado).

INTRODUCCIÓN: ANÁLISIS DEL AIRE EXHALADO

En el siglo XXI, uno de los principales retos de la medicina es el establecimiento de metodologías no invasivas y rentables para el diagnóstico precoz y el seguimiento de enfermedades crónicas de alta prevalencia, así como la detección rápida de enfermedades infecciosas. Gracias a ello, se podría mejorar el pronóstico de los pacientes, reducir la carga médica nacional y disminuir la sobresaturación de los servicios sanitarios. En este sentido, el análisis de compuestos orgánicos volátiles (VOCs) en el aire exhalado ha generado una gran expectación en los últimos años como posible herramienta de diagnóstico^{1,2}.

Aunque el análisis del aliento humano se ha empleado con fines médicos desde la Edad Antigua³, la aparición del actual campo del análisis del aire exhalado se remonta a la década de 1970 gracias a las investigaciones del premio Nobel Linus Pauling, que fue capaz de detectar unos 250 compuestos en el aire exhalado⁴. Desde entonces, gracias a los avances en las técnicas analíticas, se ha demostrado que existe un gran número de VOCs de diferentes grupos químicos (hidrocarburos, aldehídos, cetonas, alcoholes, etc.) presentes en el aire exhalado humano. En este sentido, en 2014, Lacy Costello *et al.* recopilaron un listado de al menos 872 VOCs diferentes presentes en el aliento humano identificados en la literatura⁵. Sin embargo, estos compuestos se encuentran en bajas concentraciones y representan un porcentaje relativamente pequeño del aire exhalado, cuyos componentes principales son nitrógeno (N₂), oxígeno (O₂), dióxido de carbono (CO₂), vapor de agua y gases inertes^{2,6,7}. Además, las muestras de aire exhalado también contienen compuestos no volátiles (por ejemplo, citoquinas, isoprostanos, leucotrienos, etc.), que pueden determinarse mediante el análisis del aire exhalado condensado (EBC)^{8,9}.

Los VOCs detectados en el aire exhalado pueden proceder de diferentes fuentes. De hecho, en el aire exhalado es posible distinguir VOCs derivados del tabaco, de la combustión de combustibles fósiles u otros contaminantes ambientales. Por lo tanto, la medida de estos VOCs permite evaluar la exposición humana a los contaminantes atmosféricos y los posibles riesgos para la salud^{10,11}. Por el contrario, gran número de VOCs encontrados en el aliento tienen un origen endógeno y provienen tanto del metabolismo humano como de la propia microbiota^{3,8}. En este sentido, muchos de los hidrocarburos del aire exhalado proceden de la peroxidación lipídica, que aumenta en los procesos inflamatorios y el estrés oxidativo. En la inflamación, típica de muchos trastornos, se genera un aumento de la producción de especies reactivas de oxígeno (ROS), lo que potencia la oxidación de los ácidos grasos insaturados de las membranas

celulares y la generación de subproductos como los hidrocarburos^{12,13}. Del mismo modo, las ROS producidas durante los procesos inflamatorios y el estrés oxidativo también pueden dañar el ADN, las proteínas y los carbohidratos, y generar subproductos^{14,15}. Por todas estas razones, el análisis del aire exhalado ha sido propuesto como una alternativa no invasiva muy atractiva para monitorizar el estrés oxidativo y la inflamación de las vías respiratorias en las enfermedades respiratorias, dado que las metodologías actuales de referencia, como la broncoscopia con lavado broncoalveolar (BAL) y la biopsia, son muy intrusivas (Figura 1)^{12,16,17}.



Figura 1. Técnicas de monitorización de la inflamación de las vías respiratorias y del estrés oxidativo. BAL: broncoscopia con lavado broncoalveolar; EBC: aire exhalado condensado. Adaptado de van de Kant *et al.*¹² y Ferraro *et al.*¹⁶

Además, entre los VOCs detectados en el aire exhalado, también figuran subproductos de diferentes rutas metabólicas como la biosíntesis del colesterol (ej. isopreno), el metabolismo del etanol (ej. acetaldehído) o la descarboxilación del acetoacetato y el acetil-CoA (ej. acetona)^{3,8}. Por lo tanto, el análisis de todo el conjunto de VOCs, o también llamado "volatiloma", en el aire exhalado podría ser de utilidad para conocer los procesos metabólicos del organismo tanto en condiciones saludables como patológicas^{10,18}. En consecuencia, en los últimos años se han realizado un gran número de estudios centrados en la búsqueda de biomarcadores en el aire exhalado que permitan discriminar entre controles sanos y pacientes con diferentes trastornos (ej.: asma^{19,20}, diabetes mellitus^{21,22}, cáncer de pulmón²³, enfermedad inflamatoria intestinal²⁴, enfermedades infecciosas²⁵, etc.), para distinguir entre diferentes estadios de una patología (ej.: asmáticos estables y asmáticos con exacerbaciones²⁶⁻²⁸), y para diferenciar entre pacientes con síntomas similares que padecen una enfermedad diferente (ej.: pacientes con ARDS (síndrome de dificultad respiratoria aguda) por COVID-19 (enfermedad por coronavirus 2019) y pacientes con ARDS sin COVID-19²⁹).

Sin embargo, el campo del análisis del aire exhalado se encuentra todavía en sus inicios. Hasta la fecha, no se ha logrado ningún avance más allá de la fase de descubrimiento de biomarcadores y todavía no se ha producido su implementación en la práctica clínica diaria²⁰. Después de la fase de descubrimiento de biomarcadores, debe llevarse a cabo tanto una validación analítica como una validación clínica antes de

que los nuevos biomarcadores puedan utilizarse en la práctica clínica diaria³⁰. La fase de descubrimiento de biomarcadores incluye los siguientes pasos: diseño del estudio, toma de muestras de aire exhalado, análisis del aire exhalado, preprocesamiento de datos, identificación de los VOCs, análisis de datos y validación del biomarcador putativo en un estudio de cohorte independiente (Figura 2). La falta de estandarización, especialmente en el paso de muestreo del aliento, ha sido tradicionalmente señalada como la principal causa del estancamiento de este enfoque^{31,32}. Sin embargo, también es esencial la integración de las prácticas consolidadas en el campo de la metabolómica, como el reclutamiento de grandes cohortes de sujetos, la incorporación de flujos de trabajo reproducibles y transparentes para el preprocesamiento y el análisis de datos, y la validación externa, entre otros³³. Por consiguiente, tanto la selección adecuada de estrategias como la correcta implementación de cada una de las etapas del descubrimiento de biomarcadores son imprescindibles para un progreso exitoso del análisis del aire exhalado³¹.

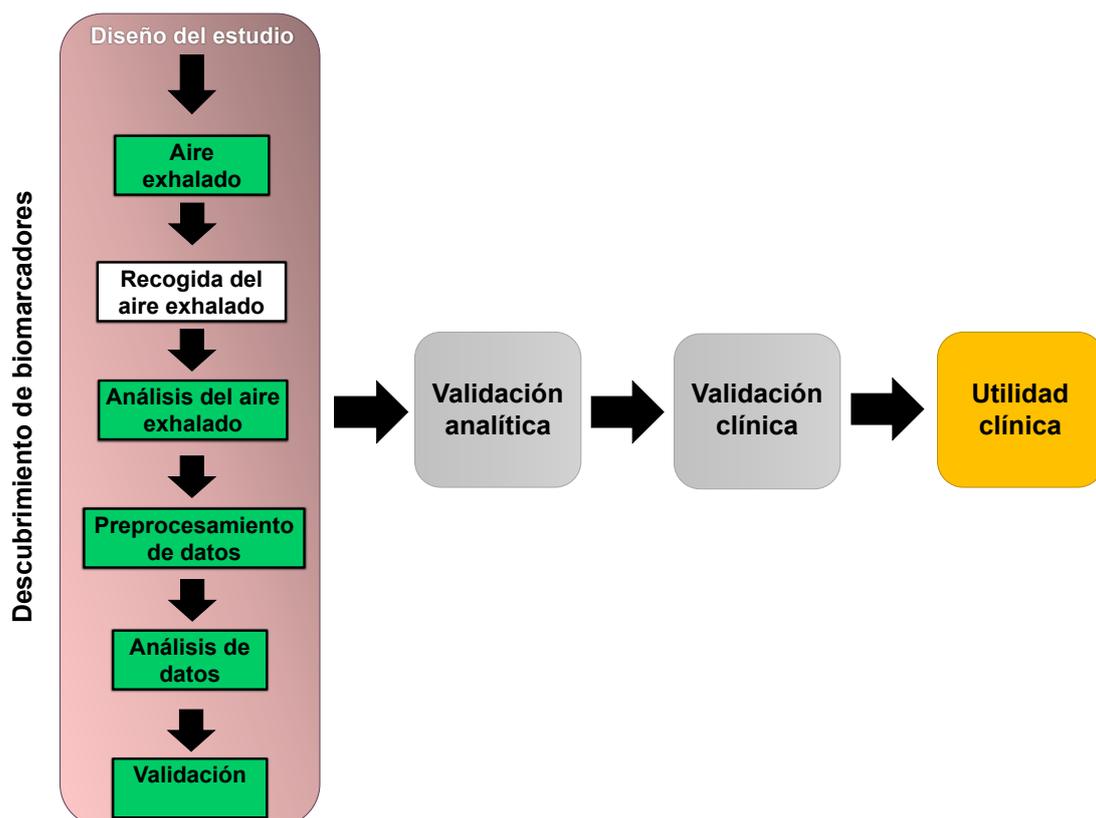


Figura 2. Proceso de establecimiento de un biomarcador hacia la aplicación clínica: fase de descubrimiento del biomarcador, fase de validación analítica, fase de validación clínica y fase de utilidad clínica.

Otra importante limitación en el proceso de estandarización es la gran diversidad de compuestos orgánicos volátiles en concentraciones muy variadas, que pueden requerir un protocolo específico para su muestreo y posterior análisis, que a menudo a

priori se desconocen³⁴. Las técnicas analíticas más utilizadas para el análisis del aire exhalado son las tecnologías basadas en la espectrometría de masas (MS) y en matrices de sensores como la nariz electrónica (E-nose)³⁵. Además, aunque no tan ampliamente, también se han empleado otras metodologías, por ejemplo, las tecnologías basadas en espectroscopia óptica³⁶. En este sentido, los recientes avances en la espectroscopia láser del infrarrojo medio (MIR) han permitido augurar la aparición de instrumentos ópticos compactos para la detección de moléculas ligeras³⁷. Por su parte, las matrices de sensores poseen algunos puntos fuertes para el entorno clínico cotidiano, como el bajo coste, la sencillez y la portabilidad. Sin embargo, mientras que en la actualidad existen multitud de plataformas de espectrometría de masas, por ejemplo, la cromatografía de gases-espectrometría de masas, que son capaces de identificar individualmente los VOCs en una mezcla de gases compleja, las matrices de sensores sólo pueden distinguir patrones o "huellas de aliento". Por lo tanto, las plataformas analíticas basadas en espectrometría de masas son realmente útiles para la búsqueda no dirigida de biomarcadores de enfermedades, ya que son capaces de determinar la identidad de los VOCs discriminantes y las rutas metabólicas relacionadas con ellos³². Estas cualidades permiten que las tecnologías basadas en espectrometría de masas sean más apropiadas en esta fase del desarrollo del método hasta lograr la identificación de biomarcadores específicos.

En cuanto al procedimiento de recogida de muestras de aliento, existen dos tipos de análisis del aire exhalado (Figura 3): análisis del aire exhalado fuera de línea u *off-line* (la muestra se recoge en un contenedor de aire exhalado y es almacenada antes del análisis) y análisis del aire exhalado en línea u *on-line* a tiempo real (el aire exhalado es analizado directamente)^{10,38}. En los últimos años, el análisis del aire exhalado *on-line* a tiempo real ha ganado popularidad, ya que ofrece numerosas ventajas para la implementación del análisis del aliento en la práctica clínica diaria, por ejemplo, la mínima manipulación de la muestra antes del análisis, la monitorización continua del volatilo de los pacientes o la rapidez en la disponibilidad de resultados. A pesar de estas ventajas, el análisis del aire exhalado *off-line* sigue siendo el método más extendido en el análisis de aire exhalado^{38,39}. En este sentido, en el análisis del aire exhalado *off-line*, las muestras de aliento pueden almacenarse y transportarse fácilmente. Como resultado, las muestras pueden recogerse en diferentes lugares y los análisis mediante MS pueden centralizarse en un único laboratorio, sin que el paciente tenga que desplazarse a la ubicación de la plataforma analítica. Esto es especialmente atractivo para pacientes pasivos no cooperativos (ej. pacientes de cuidados intensivos,

ancianos, neonatos, bebés, etc.)^{39,40}. También es preferible para los analitos que requieren una etapa de preconcentración para su correcta detección.

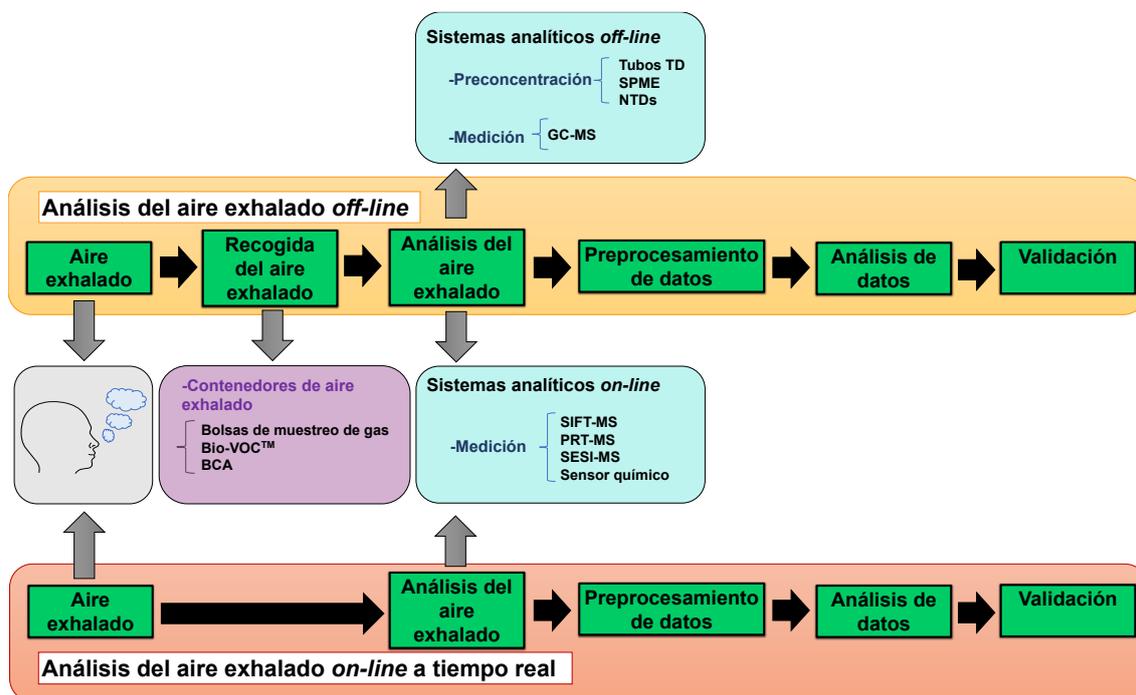


Figura 3. Principales métodos de análisis del aire exhalado: análisis del aire exhalado *off-line* y análisis del aire exhalado *on-line* a tiempo real. BCA: aparato de recogida del aliento; GC-MS: cromatografía de gases-espectrometría de masas; tubos TD: tubos de desorción térmica; SPME: microextracción en fase sólida; NTDs: dispositivos de trampa de aguja; SIFT-MS: tubo de selección de iones con flujo - espectrometría de masas; PTR-MS: reacción de transferencia de protones - espectrometría de masas; SESI-MS: ionización secundaria por *electrospray* - espectrometría de masas.

La presente Tesis Doctoral se ha centrado en la aplicación y optimización del análisis del aire exhalado *off-line* utilizando tecnologías basadas en espectrometría de masas para la detección de compuestos orgánicos volátiles en el aliento de población pediátrica y en mujeres en edad fértil. En consecuencia, en este capítulo se muestra una revisión exhaustiva de este enfoque, haciendo especial hincapié en los protocolos, los retos para la estandarización del muestreo del aliento, las técnicas analíticas, los métodos para el preprocesamiento y el análisis de datos, y las perspectivas futuras del análisis del aliento fuera de línea.

ANÁLISIS DEL AIRE EXHALADO OFF-LINE

Factores que influyen en el análisis del aire exhalado *off-line*

Diversos factores pueden influir en los estudios de volatilómica, desde las características demográficas de los sujetos hasta las metodologías empleadas que pueden condicionar la presencia de factores de confusión. En lo que respecta a la

población de estudio, diferentes aspectos como la edad, el sexo, las comorbilidades, el consumo de tabaco, la actividad física o la medicación de los sujetos se deben tener en cuenta en el diseño del estudio y en la interpretación de los resultados^{7,19,23,41}. De hecho, Blanchet *et al.* determinaron que la edad, el sexo, el tabaquismo y algunos fármacos afectaban a la composición de VOCs en el aire exhalado⁴². Por este motivo, algunos estudios se han centrado en grupos de población específicos (niños¹⁹, mujeres⁴³, obesos⁴⁴, etc.) o han intentado estimar la influencia real de algunos de estos factores en los niveles de VOCs en el aire exhalado. Por ejemplo, Capone *et al.* comunicaron la existencia de diferencias en los niveles de algunos VOCs entre fumadores y no fumadores⁴⁵.

Por otra parte, los protocolos de muestreo del aliento, los contenedores utilizados para la recogida del aire exhalado, el tiempo y la forma de almacenamiento de las muestras o los protocolos de limpieza del material son también factores que desempeñan un papel muy importante en el análisis del aire exhalado *off-line*. En este sentido, las mediciones cuantitativas del aire exhalado van a estar condicionadas por la técnica de muestreo utilizada que debe ser adecuada y reproducible, por lo que las muestras pueden no ser comparables entre sí, si se han recogido con protocolos diferentes. Además, la elección de la plataforma de análisis y los métodos de preprocesamiento y análisis de datos tienen un gran impacto en el proceso de búsqueda de biomarcadores, aumentando su complejidad^{7,12,23,32,46}.

Recogida de muestras de aire exhalado

El muestreo del aire exhalado es un aspecto distintivo del análisis del aliento *off-line*, por lo que debe llevarse a cabo correctamente para evitar la introducción de sesgos en la estabilidad de la muestra y en el análisis del aire exhalado. En general, las principales etapas son la selección del tipo de aire exhalado recogido y el contenedor de aire exhalado utilizado. Además, otro aspecto a tener en cuenta es el número de exhalaciones recogidas².

Porciones de muestreo del aire exhalado

El ciclo respiratorio humano puede monitorizarse mediante capnografía por la medición continua del CO₂ exhalado. La Figura 4 muestra un capnograma de las diferentes fases de la exhalación: La fase I incluye el aire del espacio muerto o el aire de las vías respiratorias sin intercambio gaseoso con los alvéolos, en la que se observan en el capnograma bajos niveles de CO₂, similares a los ambientales. La fase II consiste

en aire procedente tanto de las vías respiratorias como de los alvéolos, registrándose un aumento casi lineal de los niveles de CO₂ en el capnograma. La fase III incluye aire procedente de los alvéolos, reflejando un aplanamiento de la curva de CO₂ en el capnograma llamado meseta alveolar. Por último, se observa de nuevo un descenso del CO₂ en la inspiración⁴⁷⁻⁵⁰.

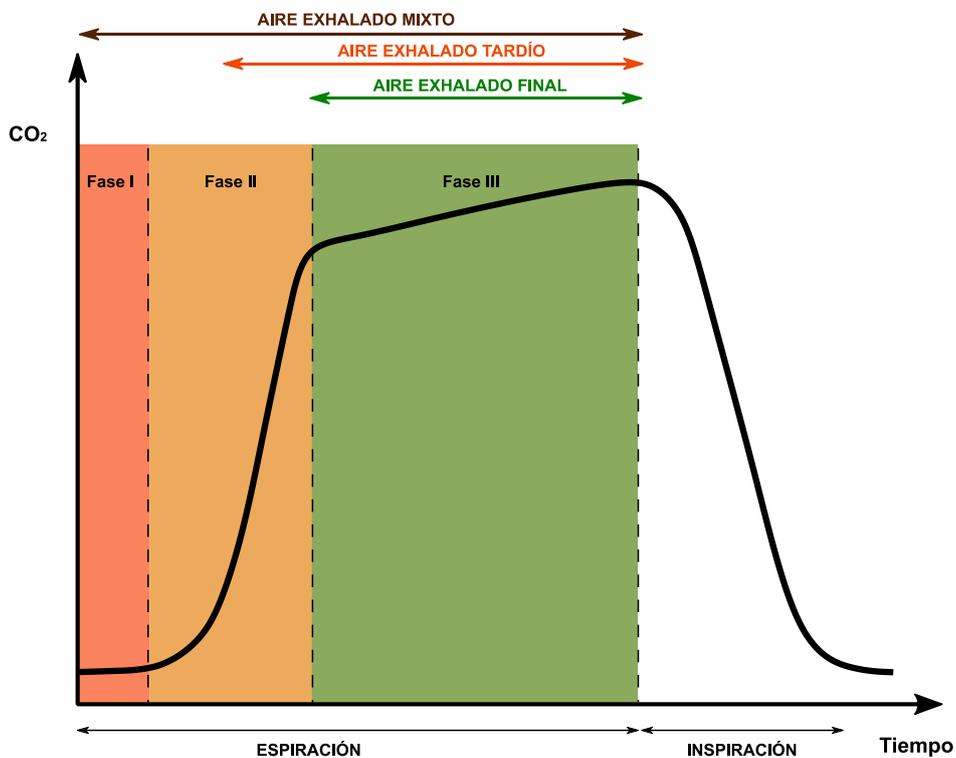


Figura 4. Representación esquemática de un capnograma del ciclo respiratorio humano. Fases de la exhalación (Fase I, Fase II y Fase III) y porciones del aire exhalado (Aire exhalado mixto, Aire exhalado tardío y Aire exhalado final). Adaptado de Lawal *et al.*⁴⁹ y Beauchamp y Miekisch⁵⁰.

En función de las fases de exhalación recogidas para su posterior análisis, las porciones de muestreo del aire exhalado se clasifican en tres tipos⁴⁹:

-Aire exhalado mixto. Todo el aire exhalado por el sujeto es recogido sin excluir ninguna de las fases de la exhalación. Por lo tanto, se incluye tanto el aire del espacio muerto como el aire alveolar. El aire exhalado mixto es ampliamente utilizado porque su proceso de muestreo es sencillo y fácil, siendo una alternativa útil para los pacientes pasivos no cooperativos. Sin embargo, es más susceptible a los contaminantes ambientales. Por ello, si se elige esta opción, debe realizarse un análisis del aire de la habitación para eliminar la influencia del fondo ambiental⁴⁹.

-Aire exhalado tardío. El aire del espacio muerto es descartado utilizando diferentes estrategias, como excluir los primeros segundos de la espiración o recoger el aire del espacio muerto en otra bolsa de muestreo. El aire exhalado tardío incluye aire

tanto de la fase III como de parte de la fase II, ya que las estrategias empleadas no pueden determinar con exactitud el momento de inicio de la fase III⁴⁹.

-Aire exhalado final. El aire de la fase III es recogido utilizando los niveles de un VOC endógeno como referencia, en general, la monitorización del CO₂ exhalado es la estrategia más frecuente. Así, el muestreo de aire exhalado comienza una vez que el capnograma empieza a aplanarse tras la subida inicial, evitando la recogida de aire de la fase I y la fase II. Por lo tanto, éste es el enfoque más preciso y reproducible para recoger el aire alveolar⁴⁹.

Contenedores para la recogida de aire exhalado

Desde un punto de vista clínico, el colector ideal debe ser barato y fácil de usar, inerte y, lo que es más importante, no permitir la entrada y salida de los VOCs. Los contenedores de aliento más utilizados en el análisis de aire exhalado *off-line* son las bolsas de muestreo de gas y otros dispositivos de muestreo como el muestreador Bio-VOCTM o los aparatos de recogida de aliento^{32,49,51}.

-Bolsas de muestreo de gas. Las bolsas de muestreo de gas son los colectores de aliento más extendidos debido a su simplicidad y a su bajo coste. Las bolsas de polímeros, como las Tedlar® (fluoruro de polivinilo) y las Mylar® aluminizadas, son las más utilizadas^{32,49,52}. Son inertes y pueden incorporarse fácilmente a los sistemas de análisis del aire exhalado. Las bolsas Tedlar® pueden ser reutilizadas más de una vez. Sin embargo, las bolsas Tedlar® tienen dos contaminantes inherentes (fenol y N,N-dimetilacetamida), por lo que deben limpiarse con gas nitrógeno puro antes de su uso. También hay que tener en cuenta la difusión a través de las paredes de las bolsas de muestreo y los problemas de condensación que pueden comprometer la estabilidad de las muestras de aire exhalado durante largos períodos de almacenamiento. Por lo tanto, se recomienda utilizar estos contenedores para recoger el aliento exhalado y utilizar matrices absorbentes para el almacenamiento de las muestras. Otras bolsas de polímeros empleadas son las bolsas de Teflon® (politetrafluoroetileno) y las bolsas Nalophan® (tereftalato de polietileno)^{32,52-54}.

-Muestreador Bio-VOCTM. El muestreador Bio-VOCTM permite descartar el aire del espacio muerto, es inerte y económico. El muestreador Bio-VOCTM consta de tres partes: una boquilla, un muestreador volumétrico (un pequeño recipiente con una capacidad de 100-150 mL) y un émbolo. El muestreador Bio-VOCTM tiene un lado abierto para que el aire se desplace a medida que avanza la espiración, con el fin de recoger

sólo el aire exhalado tardío. Una de las principales limitaciones del muestreador Bio-VOC™ es que no es capaz de recoger un gran volumen de aire exhalado^{32,49,55}.

-Aparato de recogida de aliento (BCA). El sujeto respira en una estructura tubular con una boquilla. Los VOCs son retenidos en el extremo del tubo por resinas absorbentes. Dispone de dos trampas absorbentes diferentes, una suele utilizarse para recoger el aire exhalado tardío y la otra para captar el aire ambiental. Además, posee una banda calefactora para reducir los problemas de condensación^{32,49,56-58}.

White *et al.*⁵⁹ han descrito otros dispositivos de recogida de aire exhalado, como el RTubeVOC o el Alveosampler, cuyo uso está menos extendido.

Estandarización del muestreo del aire exhalado

En los últimos años, se ha enfatizado la importancia de la estandarización del muestreo del aire exhalado, ya que el éxito del descubrimiento de potenciales biomarcadores en el aliento depende en gran medida de este primer paso. En primer lugar, es esencial determinar la porción de aire exhalado que debe recogerse en función del objetivo del estudio, y establecer un protocolo de muestreo reproducible. Aunque el aire de la fase III (sólo el aire alveolar) proporciona la información más significativa para la investigación del metabolismo y las enfermedades humanas, la recogida de porciones de aire exhalado que también incluyen el aire del espacio muerto y el aire de las vías respiratorias podría ser interesante. Por ejemplo, el aire del espacio muerto puede ser útil para controlar la exposición a contaminantes ambientales y para estudios sobre la salud bucodental, y la recogida de aire de las vías respiratorias podría ser beneficiosa en estudios centrados en el asma, por la sintomatología de esta enfermedad. Otros factores que se deben tener en cuenta son los hábitos de tabaquismo o la medicación diaria que toman los sujetos para evaluar si tienen un impacto en el volatilo. En este sentido, también deberían tenerse en cuenta las horas de ayuno previas a la recogida de muestras de aire exhalado^{1,7,60}.

Por otra parte, es muy recomendable recoger muestras de aire ambiental durante el muestreo del aliento, especialmente si no se excluyen el aire del espacio muerto y el aire de las vías respiratorias. Así, es posible identificar qué variaciones observadas en las muestras de aire exhalado se deben exclusivamente al aire ambiental, evitando la determinación de biomarcadores incorrectos. Otro enfoque alternativo para minimizar el impacto de los VOCs del aire ambiental es el uso de filtros de VOCs durante la toma de muestras de aire exhalado^{1,7,23}. Además, otro factor a tener en cuenta es la estación del

año en la que se realizó el muestreo y el análisis del aire exhalado, ya que podría influir en la composición de VOCs del aliento¹².

En resumen, la variabilidad de las mediciones de VOCs en el aliento puede verse afectada por muchos factores durante el muestreo⁴⁹, incluidas las propiedades físico-químicas de los compuestos, que se deben tener en cuenta a fin de mejorar la reproducibilidad del método aplicado⁶¹.

Plataformas analíticas basadas en espectrometría de masas

La amplia gama de plataformas analíticas basadas en espectrometría de masas puede dividirse en dos grupos principales: sistemas analíticos *off-line* (las muestras son preconcentradas antes del análisis) y sistemas analíticos *on-line* (las muestras no necesitan un tratamiento previo antes del análisis)⁵⁰.

Sistemas analíticos *off-line*

La mayoría de los estudios centrados en el análisis del aire exhalado *off-line* emplean la cromatografía de gases acoplada a la espectrometría de masas (GC-MS). La GC-MS es un típico sistema analítico *off-line*, ya que requiere una preconcentración previa de las muestras de aire exhalado^{2,32,49}. Hoy en día, la GC-MS se considera la técnica de referencia en el campo del análisis del aire exhalado. Aunque presenta algunos inconvenientes, como un mayor consumo de tiempo que los sistemas analíticos *on-line*, aun así, la GC-MS ofrece una alta sensibilidad, robustez y reproducibilidad. En este sentido, esta metodología permite la identificación y cuantificación de diferentes VOCs individuales a partir de muestras complejas de aire exhalado, incluyendo compuestos desconocidos^{31,32,35,62,63}.

La cromatografía de gases separa los VOCs del aire exhalado en función del tiempo de retención (tiempo necesario para atravesar la columna capilar de GC), que depende de la interacción entre los VOCs y la fase estacionaria que recubre la columna^{1,31}. A este respecto, en algunos estudios se utiliza la cromatografía de gases bidimensional integral (GCxGC), donde se emplea una columna de cromatografía de gases adicional para mejorar la separación de los compuestos coeluyentes^{57,64}. Los principales inconvenientes de esta estrategia es que su precio es más elevado y que el análisis de datos es más complejo⁶⁵.

Una vez que los metabolitos exhalados son separados por GC, estos compuestos llegan al espectrómetro de masas para ser analizados. Los principales

componentes de un espectrómetro de masas son la fuente de iones, el analizador de masas y el detector. En la etapa de la fuente de iones, normalmente, los compuestos son ionizados y fragmentados mediante ionización electrónica (EI). Así, en función de los compuestos analizados, en este proceso se generan diferentes fragmentos cargados. El analizador de masas se utiliza para separar estos fragmentos por su relación masa-carga (m/z). El tipo de analizador de masas determina las características del espectrómetro de masas (resolución, sensibilidad, precisión y rango de masas)^{62,66}. En el análisis del aire exhalado, se emplean analizadores de masas de baja resolución con resolución de masa unitaria (precisión de 1 Da) como el cuádrupolo simple^{67,68}, y de alta resolución (precisión de al menos 10000 Da) como el tiempo de vuelo (TOF)^{31,69,70}. La GC-MS tiene la capacidad de detectar VOCs en un rango de partes por billón, o inferior, con buena reproducibilidad y linealidad. Finalmente, las señales de los fragmentos separados son registrados en el detector, dando lugar a un gráfico denominado espectro de masas que ilustra la intensidad de las señales de los fragmentos m/z obtenidos. En este sentido, el cotejo del espectro de masas experimental con las bibliotecas de espectros de masas es útil y necesario para la identificación de compuestos^{32,62,66}.

Métodos de preconcentración

Los métodos de preconcentración suelen implicar dos etapas, en las que los VOCs son retenidos en trampas adsorbentes y luego son liberados mediante desorción térmica (TD). Actualmente, las trampas adsorbentes más utilizadas en el análisis del aire exhalado *off-line* son los tubos de desorción térmica, la microextracción en fase sólida (SPME) y los dispositivos de trampa de aguja (NTDs). La preconcentración es un paso crítico porque algunos VOCs podrían diluirse durante el muestreo del aire exhalado, especialmente cuando se recoge el aire exhalado mixto, y permite la eliminación del vapor de agua de las muestras^{3,49,50}.

-Tubos de desorción térmica (tubos TD). Actualmente, los tubos de desorción térmica son los dispositivos que más se utilizan para la preconcentración en el análisis del aire exhalado. El aire exhalado recogido en bolsas de muestreo de gas puede transferirse a los tubos de desorción térmica, donde los VOCs quedan retenidos y pueden almacenarse durante un período de tiempo más largo. Además, es posible llenar los tubos de desorción térmica con aire ambiental, bien por difusión o por captación activa utilizando una jeringa, lo que resulta útil para controlar la influencia ambiental sobre la muestra biológica exhalada. Otra ventaja del uso de tubos de desorción térmica es que las muestras pueden transportarse fácilmente entre laboratorios sin ser

alteradas. Los tubos de desorción térmica suelen ser de acero inoxidable o de vidrio y están rellenos de materiales sorbentes como polímeros orgánicos, carbón grafitizado o tamices moleculares de carbón. Mientras que los polímeros orgánicos (por ejemplo, Tenax TA) presentan una baja sensibilidad a la humedad y están indicados para capturar hidrocarburos de tamaño medio, los tubos de carbón grafitizado (por ejemplo, Carbopack X) permiten atrapar VOCs con un bajo número de carbonos. Por este motivo, se recomienda la combinación de diferentes lechos absorbentes. Teniendo en cuenta el alto porcentaje de agua en el aire exhalado, el uso de polímeros orgánicos como sorbentes es de gran interés para que su efecto sea despreciable^{3,49,50,71,72}.

-Microextracción en fase sólida (SPME). Es una metodología comúnmente utilizada en la preconcentración y el almacenamiento de muestras de aire exhalado. La extracción de compuestos orgánicos volátiles se basa en la técnica del espacio de cabeza. Así, una fibra o barra de sílice recubierta con diferentes materiales sorbentes se expone a la muestra de aire exhalado hasta que se alcanza el equilibrio químico entre ambos. El tipo y el grosor de la fibra van a influir en el tiempo necesario para alcanzar el equilibrio^{49,73,74}. Caldeira *et al.*⁵³ probaron fibras con diferentes recubrimientos (DVB/CAR/PDMS, PDMS/DVB, PDMS y PA), obteniendo una mejor eficiencia de extracción con DVB/CAR/PDMS (DVB: divinilbenceno; CAR: carboxeno; PDMS: polidimetilsiloxano; PA: poliacrilato). Es importante señalar que tanto la temperatura como el tiempo de extracción tienen un gran impacto en el proceso SPME.

-Dispositivos de trampa de agujas (NTDs). Los dispositivos de trampa de aguja no se emplean tan frecuentemente como los tubos TD y la SPME en el análisis del aire exhalado, pero ofrecen algunas ventajas de ambos. Consisten en dispositivos en forma de aguja con materiales absorbentes en su interior, donde el aire exhalado se introduce a través de una jeringa para ser capturado. Al igual que los tubos TD, el almacenamiento y el transporte de las muestras de aire exhalado son sencillos⁴⁹.

Sistemas analíticos on-line

Los sistemas analíticos *on-line* basados en la espectrometría de masas más extendidos para la medición de VOCs en el aire exhalado son SIFT-MS (tubo de selección de iones con flujo - espectrometría de masas), PTR-MS (reacción de transferencia de protones - espectrometría de masas) y SESI-MS (ionización secundaria por *electrospray* - espectrometría de masas). Aunque los sistemas analíticos *on-line* son utilizados en el análisis del aire exhalado *on-line* a tiempo real, estas plataformas analíticas también se aplican en el análisis del aire exhalado *off-*

line^{38,75}. La Tabla 1 muestra algunos ejemplos en los que los sistemas analíticos *on-line* han sido utilizados con éxito en el análisis del aire exhalado *off-line*. En este caso, se utilizaron contenedores de aire exhalado para la recogida de muestras antes del análisis del aliento mediante los sistemas analíticos *on-line*.

Tabla 1. Ejemplos de análisis del aire exhalado *off-line* mediante sistemas analíticos *on-line*.

Referencias	Contenedores de aliento	Sistemas analíticos <i>on-line</i>
Alkhoury <i>et al.</i> ⁷⁶	Bolsas Mylar®	SIFT-MS
Boshier <i>et al.</i> ⁷⁷	Bolsas Nalophan®	SIFT-MS
Dryahina <i>et al.</i> ⁷⁸	Bolsas Nalophan®	SIFT-MS
Wang <i>et al.</i> ⁷⁹	Bolsas Tedlar®	SIFT-MS
Alkhoury <i>et al.</i> ⁴⁴	Bolsas Mylar®	SIFT-MS
Eng <i>et al.</i> ⁸⁰	Bolsas Mylar®	SIFT-MS
Greiter <i>et al.</i> ⁸¹	Bolsas Teflon®	PTR-MS
Decrue <i>et al.</i> ⁴⁰	Bolsas Nalophan®	SESI-MS

SIFT-MS: tubo de selección de iones con flujo - espectrometría de masas; PTR-MS: reacción de transferencia de protones – espectrometría de masas; SESI-MS: ionización secundaria por electrospray – espectrometría de masas.

Tanto para SIFT-MS como para PTR-MS, los VOCs del aliento exhalado se ionizan mediante ionización química (CI). Mientras que los iones reactivos que se suelen utilizar en SIFT-MS son H_3O^+ , NO^+ y O_2^+ , el H_3O^+ es el ion reactivo más popular en PTR-MS. En SIFT-MS, la reacción entre los iones reactivos, producidos en un plasma de microondas y tamizados en un filtro cuadrupolar, y los metabolitos exhalados se lleva a cabo en un tubo de deriva. En general, se utiliza helio como gas portador. Sin embargo, el PTR-MS no requiere una selección previa de los iones reactivos antes de la reacción en el tubo de deriva, por lo que no posee un filtro cuadrupolar. En PTR-MS, los iones reactivos H_3O^+ se producen a partir de vapor de agua utilizando una fuente de iones de cátodo hueco. Además, PTR-MS es más sensible que SIFT-MS, ya que la muestra de aire exhalado no se diluye porque el aire de la muestra se utiliza como gas portador. Por otro lado, la resolución de PTR-MS y SIFT-MS viene determinada por el espectrómetro de masas conectado al instrumento. Tanto PTR-MS como SIFT-MS permiten la cuantificación de metabolitos. En particular, la cuantificación de los VOCs de las muestras de aire exhalado en SIFT-MS, se basa en estudios cinéticos de las reacciones químicas entre estos compuestos y los iones reactivos, y en la comparación con bibliotecas cinéticas. En este sentido, SIFT-MS se utiliza principalmente en los análisis dirigidos, centrándose en compuestos cuyas reacciones químicas con los iones reactivos están bien documentadas. Para el análisis del aire exhalado, la humedad

presente en las muestras de aliento puede interferir en el cálculo de las concentraciones de los VOCs debido a la generación de hidratos^{38,75,82}.

En los últimos años, SESI-MS ha generado un considerable entusiasmo en el campo del análisis del aire exhalado. La ionización secundaria por *electrospray* (SESI) es una variante innovadora de la ionización por *electrospray* (ESI) que permite el análisis de muestras gaseosas. Así, en SESI-MS, los VOCs de las muestras de aire exhalado se ionizan a presión ambiente al reaccionar con iones primarios formados por una fuente de ionización por *electrospray* a partir de un disolvente puro y un aditivo que mejora la conductividad. Por lo tanto, el proceso de ionización es eficiente y suave, obteniendo muy a menudo el ion molecular. Otro punto fuerte de esta técnica es que puede acoplarse a cualquier analizador de masas, incluidos los espectrómetros de masas de alta resolución^{38,83,84}.

Preprocesamiento de datos

Los datos brutos obtenidos en el análisis del aire exhalado mediante tecnologías basadas en espectrometría de masas deben ser procesados para obtener una matriz útil para el análisis estadístico posterior. Un preprocesamiento adecuado de los datos es clave para minimizar los posibles artefactos analíticos y obtener resultados precisos en el análisis de los mismos^{85,86}. Sin embargo, a pesar de la gran relevancia de este paso en el proceso de descubrimiento de biomarcadores, sólo unos pocos estudios proporcionan una descripción exhaustiva de los flujos de trabajo de preprocesamiento de datos realizados²⁰. En cambio, algunos estudios recientes centrados en el análisis del aire exhalado⁸⁷⁻⁹⁰ han implementado herramientas de código abierto típicas de la metabolómica, como el paquete *xcms* en el entorno de programación del lenguaje R⁹¹. El uso de herramientas de código abierto es esencial para superar los retos de los estudios de metabolómica y volatilómica, ya que fomenta la colaboración dentro de la comunidad científica y el desarrollo de protocolos reproducibles y transparentes para el preprocesamiento de datos⁹².

Los datos brutos obtenidos y los flujos de trabajo de preprocesamiento de datos que deben realizarse dependen de la elección de la plataforma analítica. En el caso de GC-MS por ionización EI, los datos brutos constan de tres dimensiones (tiempo de retención, valor de m/z e intensidad) que se representan a través de cromatogramas que registran el tiempo de retención en el eje x y la intensidad en el eje y, y espectros de masas con el valor de m/z en el eje x y la intensidad en el eje y (Figura 5). El TIC o cromatograma de iones total se define como la suma de las intensidades de todas las

señales m/z o picos de iones en el espectro de masas por unidad de tiempo o por barrido. El cromatograma de gases es el TIC total de los escaneos realizados durante el análisis del aire exhalado o cromatograma TIC. Por otro lado, los EIC (cromatograma de iones extraídos) muestran las intensidades de una señal m/z concreta frente al tiempo de retención. Por lo tanto, cada pico del cromatograma de gases, que podría corresponder a uno o más compuestos orgánicos volátiles, tiene asociado un espectro de masas con varios *features* o picos de iones caracterizados por un tiempo de retención y un valor de m/z específicos asociados a diferentes intensidades^{66,85,86,91,93}.

En este sentido, existen dos estrategias principales para el preprocesamiento de datos brutos de GC-MS. Mientras que la estrategia 1 se basa en el *peak-picking* o en la determinación de picos iónicos en los espectros obtenidos de las muestras de aire exhalado, la estrategia 2 se centra en la detección de los compuestos presentes en las muestras de aliento mediante la deconvolución espectral⁹³. Ambos tienen ventajas y limitaciones⁹³⁻⁹⁵, por lo que una estrategia interesante podría ser la integración de ambos enfoques. Algunos ejemplos de herramientas computacionales son *xcms*⁹¹ o *MetAlign*⁹⁶ para la estrategia 1, y *eRah*^{93,97} para la estrategia 2. Independientemente de la estrategia empleada, hay algunos pasos comunes en el preprocesamiento de los datos, como la eliminación de ruido, la corrección de la línea de base, la corrección del tiempo de retención o alineamiento, la recuperación de los compuestos/*features* perdidos y la normalización^{32,85,91,93}.

Una tarea importante en el preprocesamiento de datos, especialmente en el análisis no dirigido, es la identificación de compuestos. En GC-MS, que utiliza la ionización EI, la identificación de los compuestos se basa principalmente en la similitud de los tiempos de retención y la coincidencia de los espectros de masas con una biblioteca espectral. Por lo tanto, se recomienda encarecidamente analizar estándares químicos para el cálculo del índice de retención (RI) y así, confirmar la identidad de los compuestos^{10,33,66,93}.

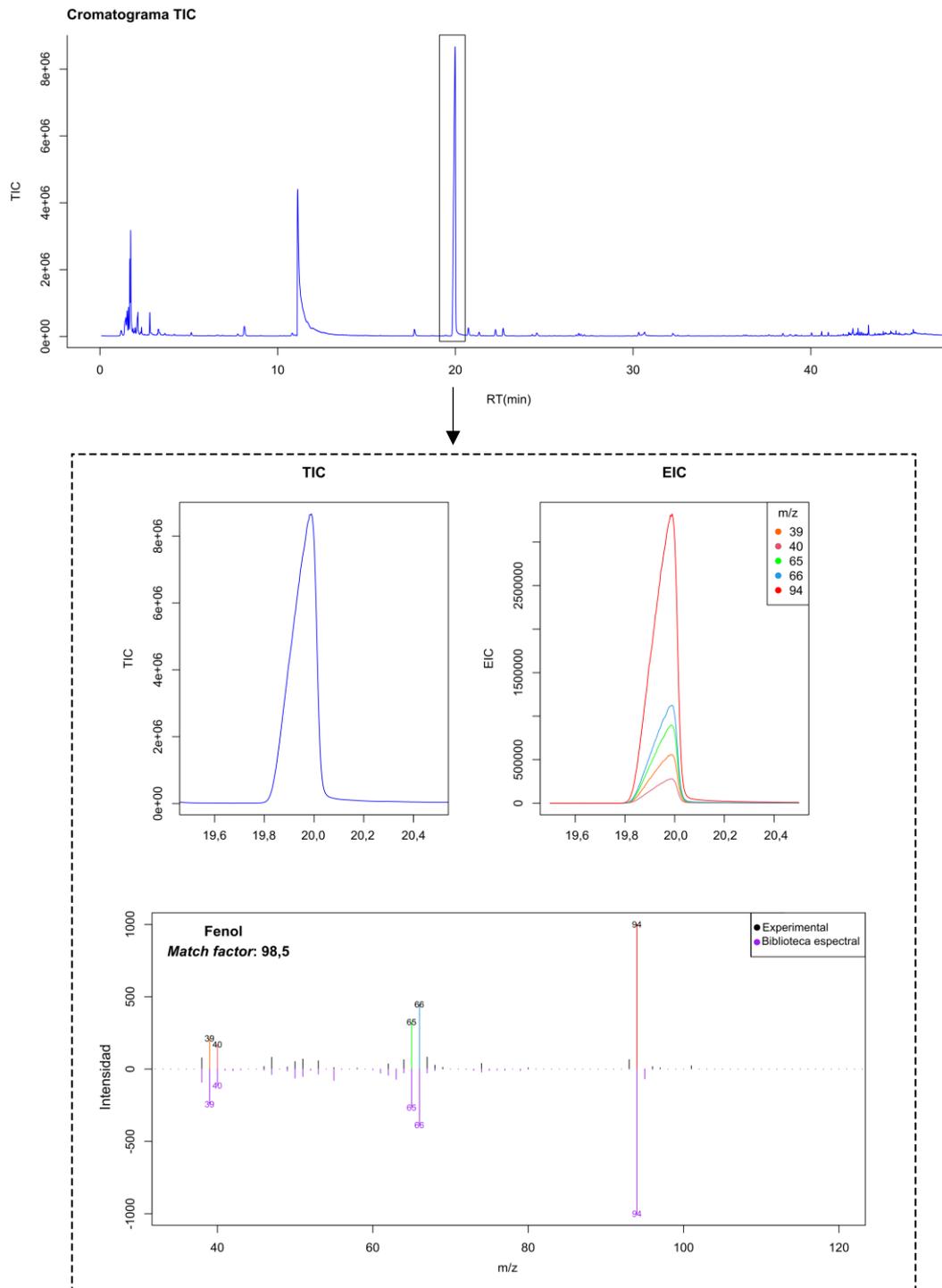


Figura 5. Resultados del análisis del aire exhalado mediante cromatografía de gases-espectrometría de masas. TIC: cromatograma de iones totales, EIC: cromatograma de iones extraídos. Los gráficos de esta figura se generaron mediante las funciones *plotChr* y *plotSpectra* del paquete *eRah* en R.

Análisis de datos

Una vez que los datos brutos se han transformado en una matriz de datos mediante el preprocesamiento de estos, el siguiente paso en el descubrimiento de biomarcadores es el análisis de los mismos³¹. Un solo biomarcador no suele ser suficiente para realizar un óptimo diagnóstico y seguimiento de enfermedades complejas y heterogéneas. Por este motivo, en los últimos años ha aumentado el interés por encontrar un conjunto de biomarcadores para desentrañar los procesos metabólicos relacionados con condiciones patológicas. Esta tendencia también se ha extendido al campo del análisis del aire exhalado y varios estudios se centran en la búsqueda de perfiles de VOCs en lugar de un único VOC discriminante. En este sentido, el análisis estadístico multivariante está considerado como el enfoque más adecuado para la determinación de perfiles de VOCs. De hecho, actualmente, existe una amplia gama de algoritmos de aprendizaje automático que puede aplicarse en el análisis de datos extraídos del aire exhalado^{2,33,98}.

Algoritmos de aprendizaje automático

Los datos obtenidos en el análisis del aire exhalado presentan una alta dimensionalidad. En consecuencia, los datos de volatiloma exhalado se caracterizan por presentar un mayor número de VOCs detectados que el de muestras disponibles, una baja prevalencia de cada VOC en todas las muestras y una alta probabilidad de distribución no normal. Además, otro problema típico en los estudios de perfiles de VOCs es la multicolinealidad, que consiste en correlaciones aleatorias de las variables^{85,99}. Los métodos estadísticos multivariantes suelen clasificarse en: *métodos de aprendizaje no supervisado* (ej. análisis de componentes principales (PCA)) y *métodos de aprendizaje supervisado* (ej. análisis lineal discriminante (LDA), regresión logística (LR), bosque aleatorio o *random forest* (RF), máquinas de vector soporte (SVM) o redes neuronales artificiales (ANNs)). Mientras que los métodos de aprendizaje no supervisado se utilizan para realizar un análisis exploratorio de los datos sin suposiciones previas ni datos etiquetados, los métodos de aprendizaje supervisado se aplican para obtener un modelo discriminante y predictivo mediante datos etiquetados e información previa (Figura 6)^{85,94,100,101}.

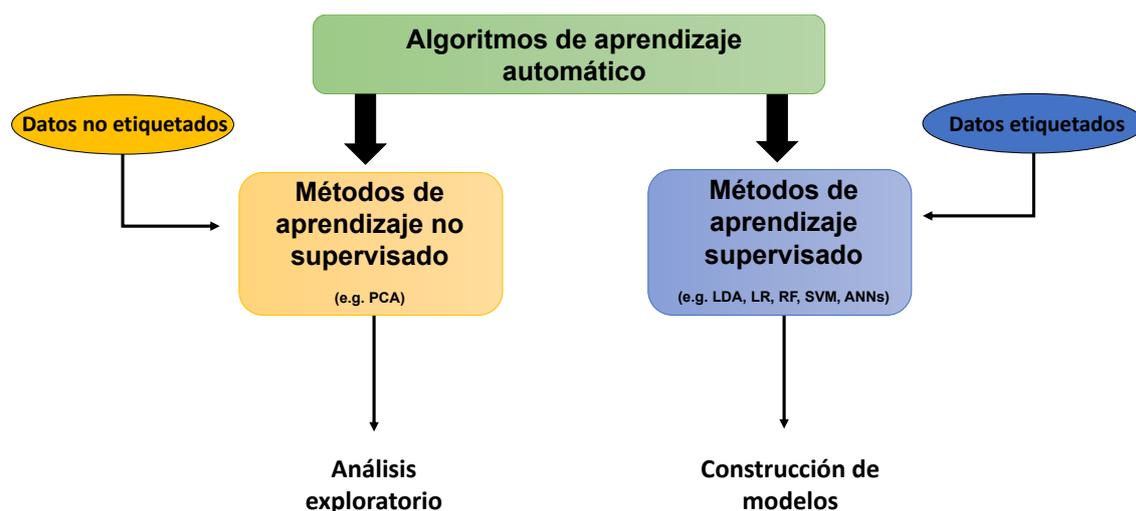


Figura 6. Algoritmos de aprendizaje automático. PCA: análisis de componentes principales; LDA: análisis discriminante lineal; LR: regresión logística; RF: *random forest* o bosque aleatorio; SVM: máquinas de vector soporte; ANNs: redes neuronales artificiales.

En general, los métodos no supervisados son utilizados como primer paso en el análisis de datos. En este sentido, el PCA es sin duda la técnica más utilizada en los estudios centrados en el análisis del aire exhalado. Principalmente, el PCA se lleva a cabo para realizar un análisis exploratorio con el fin de observar grupos o tendencias en los datos^{85,102}. A continuación, el siguiente paso suele ser la implementación de métodos de aprendizaje supervisado para la construcción de modelos. Los métodos de aprendizaje supervisado son capaces de predecir los valores de salida o variables dependientes, basándose en los valores de entrada o variables independientes^{100,103}. Por ello, el uso de métodos de aprendizaje supervisado es una excelente estrategia para la determinación de biomarcadores⁹⁴. Sin embargo, cualquier método de aprendizaje supervisado no es adecuado para los datos de volatiloma, debido principalmente a la presencia de multicolinealidad. En este sentido, el LDA y la regresión logística no son la mejor opción para la elaboración de perfiles de VOCs sin una selección previa de las variables^{85,104}. El problema de la multicolinealidad puede abordarse principalmente desde tres perspectivas diferentes: 1) utilizando un método de aprendizaje no supervisado antes de un método de aprendizaje supervisado (por ejemplo, PCA + LDA)¹⁰⁵, 2) pretratamiento de los datos mediante técnicas específicas como los métodos de contracción (*ridge*, LASSO (operador de selección y contracción mínima absoluta), etc.) o métodos que utilizan direcciones de entrada derivadas (PCR (regresión de componentes principales), PLS (regresión de mínimos cuadrados parciales))¹⁰⁰ y un método de aprendizaje supervisado, o 3) métodos de aprendizaje supervisado con tolerancia a los datos de alta dimensionalidad (por ejemplo, *random forest*, SVM, etc.)¹⁰⁶⁻¹⁰⁸. La Tabla 2 muestra algunos ejemplos de estudios centrados en el análisis del aire

exhalado mediante tecnologías basadas en espectrometría de masas que utilizaron alguna de las estrategias mencionadas para obtener un modelo predictivo.

Tabla 2. Ejemplos de algoritmos de aprendizaje automático utilizados en el análisis del aire exhalado mediante tecnologías basadas en espectrometría de masas para obtener un modelo predictivo.

Algoritmos de aprendizaje automático	Referencias
PCA+LDA	Zhou <i>et al.</i> ¹⁰⁹
LASSO+LR	Wang <i>et al.</i> ⁷⁹ , Monasta <i>et al.</i> ¹¹⁰
PLS-DA	Caldeira <i>et al.</i> ⁵³ , Van Oort <i>et al.</i> ⁸⁸ , Beccaria <i>et al.</i> ¹¹¹
SVM	Robroeks <i>et al.</i> ²⁸ , Van Berkel <i>et al.</i> ¹¹² , Beccaria <i>et al.</i> ¹¹¹
<i>Random forest</i>	Beccaria <i>et al.</i> ¹¹¹ , Gashimova <i>et al.</i> ¹¹³
ANNs	Rudnicka <i>et al.</i> ¹¹⁴

PCA: análisis de componentes principales; LDA: análisis discriminante lineal; LASSO: operador de selección y contracción mínima absoluta; LR: regresión logística; PLS-DA: análisis discriminante mínimos cuadrados parciales; SVM: máquinas de vector soporte; ANNs: redes neuronales artificiales.

PLS-DA es un algoritmo que combina el método de regresión PLS y un análisis discriminante para tratar el problema de la multicolinealidad^{85,115,116}. Así, PLS-DA es un método de aprendizaje supervisado muy utilizado en el descubrimiento de biomarcadores^{85,116}. Sin embargo, Brereton *et al.* afirman que es posible conseguir resultados similares con la primera estrategia (por ejemplo, PCA + LDA)¹⁰⁵. No obstante, un problema de esta estrategia es que información importante puede que sea ignorada si no es recopilada por los primeros componentes principales⁸⁵. Por otro lado, varios estudios centrados en el análisis del aire exhalado han optado por la tercera estrategia y han implementado algoritmos innovadores como *random forest*, SVM o ANNs. Sin embargo, estos métodos tienen una alta complejidad computacional y sus resultados no suelen ser fáciles de entender e interpretar⁸⁵.

Validación de los modelos

La validación de los modelos es un paso esencial en el análisis de datos para evaluar el rendimiento de los modelos, la aplicabilidad real de éstos sobre datos desconocidos y la idoneidad de los perfiles de VOCs seleccionados. Sin embargo, si la construcción y la validación de los modelos se realizan sobre el mismo conjunto de datos, lo que desgraciadamente es muy frecuente en los estudios de volatilómica, existe un alto riesgo de sobreajuste. En consecuencia, la validación externa de los modelos es la mejor opción. En este enfoque, el modelo se construye a partir de un conjunto de entrenamiento, y la validación del modelo se realiza sobre otro conjunto independiente

(conjunto de validación). En este sentido, es frecuente dividir aleatoriamente las muestras de aliento exhalado de los sujetos en un conjunto de entrenamiento y un conjunto de validación. Sin embargo, es más deseable que las muestras de aire exhalado del conjunto de validación pertenezcan a un grupo de nuevos sujetos, reclutados en un momento o lugar diferentes a los del conjunto de entrenamiento (verdadera validación externa). Otro factor determinante es el tamaño de la muestra en los estudios centrados en el análisis del aire exhalado, que a veces es insuficiente para proporcionar dos conjuntos independientes. Por este motivo, un enfoque interesante es la validación interna del modelo o técnicas de remuestreo, como la validación cruzada (CV) o el *bootstrapping*. Así, las estrategias de validación cruzada más populares son la validación cruzada *leave-one-out* y *n-fold*^{33,85,94,101}. El uso de métodos de remuestreo suele generar una gran controversia, ya que se ha sugerido que estas técnicas podrían sobreestimar el poder predictivo de los modelos. En este sentido, la prueba de permutación puede utilizarse junto con las técnicas de remuestreo para resolver estos problemas. La prueba de permutación evalúa si los rendimientos de los modelos construidos son significativamente mejores que los rendimientos de los modelos predictivos construidos con datos etiquetados aleatoriamente^{85,86}. Aunque la prueba de permutación todavía no se utiliza ampliamente en los estudios de volatilómica, algunos estudios recientes han introducido esta prueba para la evaluación de la significación estadística del rendimiento del modelo propuesto⁶⁷.

CONCLUSIONES Y PERSPECTIVAS

En la presente Tesis Doctoral, se ha empleado el análisis del aire exhalado *off-line* mediante tecnologías basadas en espectrometría de masas para la detección de VOCs en población pediátrica y en mujeres en edad fértil. Desde hace años, el aire exhalado ha sido considerado una fuente prometedora de biomarcadores. En particular, este enfoque no invasivo podría ser especialmente beneficioso para los pacientes pasivos no cooperativos, como los pacientes de pediatría. En este sentido, el análisis de VOCs podría ser muy útil en el futuro para el diagnóstico y el seguimiento de enfermedades prevalentes en los niños, como es el asma. Además, también podría desempeñar un papel importante en el seguimiento del estrés oxidativo y la exposición a contaminantes ambientales. En concreto, el análisis del aire exhalado *off-line* permite almacenar y transportar fácilmente las muestras de aliento, evitando así el desplazamiento innecesario de los pacientes hasta la ubicación de la plataforma de análisis. Otra fortaleza de las técnicas analíticas típicas utilizadas en este enfoque es que ofrecen una alta sensibilidad, robustez y reproducibilidad.

En este capítulo se han revisado los principales pasos implicados en el descubrimiento de biomarcadores mediante el análisis del aire exhalado *off-line*. Se ha hecho especial hincapié en los protocolos y en el proceso de estandarización del muestreo del aire exhalado, ya que los datos adquiridos se ven considerablemente afectados por el diseño experimental y los parámetros instrumentales. Así, la elección de la fracción de aire exhalado muestreado se debe basar en el objetivo principal del estudio, la elección del contenedor de muestreo se considerará en función del tipo de paciente y plataforma analítica utilizada y, además, es esencial comprobar la influencia del aire ambiental de la sala donde se realiza el muestreo del aliento. Por otra parte, esta revisión ha abordado diferentes plataformas analíticas basadas en la espectrometría de masas que se han utilizado en el análisis del aire exhalado *off-line*. De hecho, los sistemas analíticos *off-line*, como la GC-MS con métodos de preconcentración, son claramente los más utilizados para el análisis del aire exhalado *off-line*. Sin embargo, esta revisión también explora el uso emergente de sistemas analíticos *on-line* para el enfoque *off-line*. Por último, se han expuesto qué estrategias y algoritmos innovadores son recomendados para el preprocesamiento y el análisis de datos en el campo del análisis del aire exhalado. En efecto, los datos de volatiloma tienen una alta dimensionalidad y a menudo presentan problemas de multicolinealidad. Por lo tanto, es crucial seleccionar técnicas robustas que permitan hacer frente a estos problemas. Mediante el uso de algoritmos de aprendizaje automático, específicamente métodos de aprendizaje supervisado, es posible obtener modelos predictivos y perfiles de VOCs discriminantes. Finalmente, la validación de estos modelos se debe llevar a cabo de forma rigurosa para evitar el sobreajuste y garantizar su uso universal.

En resumen, el análisis del aire exhalado es actualmente un área de intensa investigación que requiere un esfuerzo colectivo para superar su fase exploratoria, ya que tiene un inmenso potencial clínico en la atención sanitaria en un futuro próximo.

BIBLIOGRAFÍA

1. Ibrahim, W. *et al.* Breathomics for the Clinician: The use of volatile organic compounds in respiratory diseases. *Thorax* **76**, 514–521 (2021).
2. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
3. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
4. Pauling, L., Robinson, A. B., Teranishi, R. & Cary, P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 2374–2376 (1971).
5. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
6. Bos, L. D. J. Diagnosis of acute respiratory distress syndrome by exhaled breath analysis. *Ann. Transl. Med.* **6**, 33–33 (2018).
7. Davis, M. D., Fowler, S. J. & Montpetit, A. J. Exhaled breath testing – A tool for the clinician and researcher. *Paediatr. Respir. Rev.* **29**, 37–41 (2019).
8. Buszewski, B., Keszy, M., Ligor, T. & Amann, A. Human exhaled air analytics: Biomarkers of diseases. *Biomed. Chromatogr.* **21**, 553–566 (2007).
9. Montuschi, P. Analysis of exhaled breath condensate in respiratory medicine: methodological aspects and potential clinical applications. *Ther. Adv. Respir. Dis.* **1**, 5–23 (2007).
10. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
11. Tang, Z., Liu, Y. & Duan, Y. Breath analysis: technical developments and challenges in the monitoring of human exposure to volatile organic compounds. *J. Chromatogr. B* **1002**, 285–299 (2015).
12. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
13. Ratcliffe, N. *et al.* A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of

- volatile organic compounds from the human body. *J. Breath Res.* **14**, 034001 (2020).
14. Zuo, L., Otenbaker, N. P., Rose, B. A. & Salisbury, K. S. Molecular mechanisms of reactive oxygen species-related pulmonary inflammation and asthma. *Molecular Immunology* **56**, 57–63 (2013).
 15. Calenic, B. *et al.* Oxidative stress and volatile organic compounds: Interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* **13**, 1020–1030 (2015).
 16. Ferraro, V., Carraro, S., Bozzetto, S., Zanconato, S. & Baraldi, E. Exhaled biomarkers in childhood asthma: old and new approaches. *Asthma Res. Pract.* **4**, 9 (2018).
 17. Tenero, L., Zaffanello, M., Piazza, M. & Piacentini, G. Measuring airway inflammation in asthmatic children. *Front. Pediatr.* **6**, 196 (2018).
 18. Kim, K. H., Jahan, S. A. & Kabir, E. A review of breath analysis for diagnosis of human health. *TrAC Trends Anal. Chem.* **33**, 1–8 (2012).
 19. Neerinx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
 20. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
 21. Wang, W. *et al.* Accuracy of breath test for diabetes mellitus diagnosis: a systematic review and meta-analysis. *BMJ Open Diabetes Res. Care* **9**, e002174 (2021).
 22. Dixit, K., Fardindoost, S., Ravishankara, A., Tasnim, N. & Hoorfar, M. Exhaled Breath Analysis for Diabetes Diagnosis and Monitoring: Relevance, Challenges and Possibilities. *Biosensors* **11**, 476 (2021).
 23. Marzorati, D. *et al.* A review of exhaled breath: a key role in lung cancer diagnosis. *J. Breath Res.* **13**, 034001 (2019).
 24. Kurada, S., Alkhoury, N., Fiocchi, C., Dweik, R. & Rieder, F. Review article: breath analysis in inflammatory bowel diseases. *Aliment. Pharmacol. Ther.* **41**, 329–341 (2015).
 25. Ahmed, W. M., Lawal, O., Nijssen, T. M., Goodacre, R. & Fowler, S. J. Exhaled Volatile Organic Compounds of Infection: A Systematic Review. *ACS Infect. Dis.* **3**, 695–710 (2017).

26. Brinkman, P. *et al.* Exhaled breath profiles in the monitoring of loss of control and clinical recovery in asthma. *Clin. Exp. Allergy* **47**, 1159–1169 (2017).
27. van Vliet, D. *et al.* Can exhaled volatile organic compounds predict asthma exacerbations in children? *J. Breath Res.* **11**, 016016 (2017).
28. Robroeks, C. M. *et al.* Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study. *Eur. Respir. J.* **42**, 98–106 (2013).
29. Grassin-Delyle, S. *et al.* Metabolomics of exhaled breath in critically ill COVID-19 patients: A pilot study. *EBioMedicine* **63**, 103154 (2021).
30. Mazzone, P. J. *et al.* Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American Thoracic Society Policy Statement. *Am. J. Respir. Crit. Care Med.* **196**, e15–e29 (2017).
31. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res.* **6**, 027108 (2012).
32. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
33. Boots, A. W., Bos, L. D., van der Schee, M. P., van Schooten, F. J. & Sterk, P. J. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. *Trends Mol. Med.* **21**, 633–644 (2015).
34. Herbig, J. & Beauchamp, J. Towards standardization in the analysis of breath gas volatiles. *J. Breath Res.* **8**, 037101 (2014).
35. Kaloumenou, M., Skotadis, E., Lagopati, N., Efstathopoulos, E. & Tsoukalas, D. Breath Analysis: A Promising Tool for Disease Diagnosis-The Role of Sensors. *Sensors* **22**, 1238 (2022).
36. Metsälä, M. Optical techniques for breath analysis: from single to multi-species detection. *J. Breath Res.* **12**, 027104 (2018).
37. Selvaraj, R., Vasa, N. J., Nagendra, S. M. S. & Mizaikoff, B. Advances in Mid-Infrared Spectroscopy-Based Sensing Techniques for Exhaled Breath Diagnostics. *Molecules* **25**, 2227 (2020).
38. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).

39. Zhang, J. *et al.* Breath volatile organic compound analysis: an emerging method for gastric cancer detection. *J. Breath Res.* **15**, (2021).
40. Decrue, F. *et al.* Combination of Exhaled Breath Analysis with Parallel Lung Function and FeNO Measurements in Infants. *Anal. Chem.* **93**, 15579–15583 (2021).
41. Skelly, A. C., Dettori, J. R. & Brodt, E. D. Assessing bias: the importance of considering confounding. *Evid. Based. Spine. Care. J.* **3**, 9 (2012).
42. Blanchet, L. *et al.* Factors that influence the volatile organic compound content in human breath. *J. Breath Res.* **11**, 016013 (2017).
43. Sukul, P., Schubert, J. K., Trefz, P. & Miekisch, W. Natural menstrual rhythm and oral contraception diversely affect exhaled breath compositions. *Sci. Rep.* **8**, 10838 (2018).
44. Alkhouri, N. *et al.* Breathprints of childhood obesity: changes in volatile organic compounds in obese children compared with lean controls. *Pediatr. Obes.* **10**, 23–29 (2015).
45. Capone, S. *et al.* Chromatographic analysis of VOC patterns in exhaled breath from smokers and nonsmokers. *Biomed. Chromatogr.* **32**, (2018).
46. Van Der Schee, M. P. *et al.* Breathomics in lung disease. *Chest* **147**, 224–231 (2015).
47. Verscheure, S., Massion, P. B., Verschuren, F., Damas, P. & Magder, S. Volumetric capnography: lessons from the past and current clinical applications. *Crit. Care* **20**, 184 (2016).
48. Pham, Y. L. & Beauchamp, J. Breath Biomarkers in Diagnostic Applications. *Molecules* **26**, 5514 (2021).
49. Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R. & Fowler, S. J. Exhaled breath analysis: a review of ‘breath-taking’ methods for off-line analysis. *Metabolomics* **13**, 110 (2017).
50. Beauchamp, J. D. & Miekisch, W. Breath sampling and standardization. in *Breathborne Biomarkers and the Human Volatilome* (eds. Beauchamp, J., Davis, C. & Pleil, J.) 23–40 (Elsevier, 2020).
51. Zhou, J., Huang, Z. A., Kumar, U. & Chen, D. D. Y. Review of recent developments in determining volatile organic compounds in exhaled breath as biomarkers for lung cancer diagnosis. *Anal. Chim. Acta* **996**, 1–9 (2017).

52. Mochalski, P., Wzorek, B., Śliwka, I. & Amann, A. Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. *J. Chromatogr. B* **877**, 189–196 (2009).
53. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–80 (2011).
54. Beauchamp, J., Herbig, J., Gutmann, R. & Hansel, A. On the use of Tedlar® bags for breath-gas sampling and analysis. *J. Breath Res.* **2**, 046001 (2008).
55. Kwak, J. *et al.* Evaluation of Bio-VOC Sampler for Analysis of Volatile Organic Compounds in Exhaled Breath. *Metabolites* **4**, 879 (2014).
56. Phillips, M. *et al.* Detection of volatile biomarkers of therapeutic radiation in breath. *J. Breath Res.* **7**, 036002 (2013).
57. Phillips, M. *et al.* Detection of an Extended Human Volatome with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry. *PLoS One* **8**, e75274 (2013).
58. Phillips, M. Method for the collection and assay of volatile organic compounds in breath. *Anal. Biochem.* **247**, 272–278 (1997).
59. White, I. R. & Fowler, S. J. Capturing and Storing Exhaled Breath for Offline Analysis. in *Breath Analysis* (eds. Pennazza, G. & Santonico, M.) 13–31 (Academic Press, 2019).
60. Ratiu, I. A., Ligor, T., Bocos-Bintintan, V., Mayhew, C. A. & Buszewski, B. Volatile Organic Compounds in Exhaled Breath as Fingerprints of Lung Cancer, Asthma and COPD. *J. Clin. Med.* **10**, 1–41 (2020).
61. Gilio, A. Di *et al.* Breath Analysis: Comparison among Methodological Approaches for Breath Sampling. *Molecules* **25**, 5823 (2020).
62. Belizário, J. E., Faintuch, J. & Malpartida, M. G. Breath Biopsy and Discovery of Exclusive Volatile Organic Compounds for Diagnosis of Infectious Diseases. *Front. Cell. Infect. Microbiol.* **10**, 564194 (2021).
63. Shende, P., Vaidya, J., Kulkarni, Y. A. & Gaud, R. S. Systematic approaches for biodiagnostics using exhaled air. *J. Control. Release* **268**, 282–295 (2017).
64. Caldeira, M. *et al.* Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **1254**, 87–97 (2012).

65. Fortunato de Carvalho Rocha, W., Schantz, M. M., Sheen, D. A., Chu, P. M. & Lippa, K. A. Unsupervised classification of petroleum Certified Reference Materials and other fuels by chemometric analysis of gas chromatography-mass spectrometry data. *Fuel* **197**, 248–258 (2017).
66. Smith, J. S. & Thakur, R. A. Mass Spectrometry. in *Food Analysis* (ed. Suzanne Nielsen, S.) 457–470 (Springer, 2010).
67. Neerincx, A. H. *et al.* Detection of *Staphylococcus aureus* in cystic fibrosis patients using breath VOC profiles. *J. Breath Res.* **10**, 046014 (2016).
68. Grabowska-Polanowska, B., Skowron, M., Miarka, P., Pietrzycka, A. & Śliwka, I. The application of chromatographic breath analysis in the search of volatile biomarkers of chronic kidney disease and coexisting type 2 diabetes mellitus. *J. Chromatogr. B* **1060**, 103–110 (2017).
69. Pizzini, A. *et al.* Analysis of volatile organic compounds in the breath of patients with stable or acute exacerbation of chronic obstructive pulmonary disease. *J. Breath Res.* **12**, 036002 (2018).
70. Smolinska, A. *et al.* Profiling of Volatile Organic Compounds in Exhaled Breath As a Strategy to Find Early Predictive Signatures of Asthma in Children. *PLoS One* **9**, e95668 (2014).
71. Woolfenden, E. Thermal desorption gas chromatography. in *Gas Chromatography* (ed. Poole, C. F.) 267–323 (Elsevier, 2021).
72. Dharmawardana, N., Woods, C., Watson, D. I., Yazbeck, R. & Ooi, E. H. A review of breath analysis techniques in head and neck cancer. *Oral Oncol.* **104**, 104654 (2020).
73. Sun, X., Shao, K. & Wang, T. Detection of volatile organic compounds (VOCs) from exhaled breath as noninvasive methods for cancer diagnosis. *Anal. Bioanal. Chem.* **408**, 2759–2780 (2016).
74. Nerín, C., Salafranca, J., Aznar, M. & Batlle, R. Critical review on recent developments in solventless techniques for extraction of analytes. *Anal. Bioanal. Chem.* **393**, 809–833 (2009).
75. Španěl, P. & Smith, D. Quantification of volatile metabolites in exhaled breath by selected ion flow tube mass spectrometry, SIFT-MS. *Clin. Mass Spectrom.* **16**, 18 (2020).
76. Alkhouri, N. *et al.* Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children. *Eur. J. Gastroenterol.*

- Hepatol.* **26**, 82–7 (2014).
77. Boshier, P. R. *et al.* Breath metabolite response to major upper gastrointestinal surgery. *J. Surg. Res.* **193**, 704–712 (2015).
 78. Dryahina, K. *et al.* Pentane and other volatile organic compounds, including carboxylic acids, in the exhaled breath of patients with Crohn’s disease and ulcerative colitis. *J. Breath Res.* **12**, 016002 (2017).
 79. Wang, M. H. *et al.* Use of a least absolute shrinkage and selection operator (LASSO) model to selected ion flow tube mass spectrometry (SIFT-MS) analysis of exhaled breath to predict the efficacy of dialysis: a pilot study. *J. Breath Res.* **10**, 046004 (2016).
 80. Eng, K. *et al.* Analysis of breath volatile organic compounds in children with chronic liver disease compared to healthy controls. *J. Breath Res.* **9**, 026002 (2015).
 81. Greiter, M. B. *et al.* Differences in Exhaled Gas Profiles Between Patients with Type 2 Diabetes and Healthy Controls. *Diabetes Technol. Ther.* **12**, 455–463 (2010).
 82. Smith, D., Španěl, P., Herbig, J. & Beauchamp, J. Mass spectrometry for real-time quantitative breath analysis. *J. Breath Res.* **8**, 027101 (2014).
 83. Singh, K. D. *et al.* Translating secondary electrospray ionization-high-resolution mass spectrometry to the clinical environment. *J. Breath Res.* **12**, 027113 (2018).
 84. Casas-Ferreira, A. M., Nogal-Sánchez, M. del, Pérez-Pavón, J. L. & Moreno-Cordero, B. Non-separative mass spectrometry methods for non-invasive medical diagnostics based on volatile organic compounds: A review. *Anal. Chim. Acta* **1045**, 10–22 (2019).
 85. Smolinska, A. *et al.* Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J. Breath Res.* **8**, 027105 (2014).
 86. Stavropoulos, G., Salman, D., Alkhalifah, Y., van Schooten, F.-J. & Smolinska, A. Preprocessing and analysis of volatilome data. in *Breathborne Biomarkers and the Human Volatilome* (eds. Beauchamp, J., Davis, C. & Pleil, J.) 633–647 (Elsevier, 2020).
 87. Bos, L. D. J. *et al.* A simple breath sampling method in intubated and mechanically ventilated critically ill patients. *Respir. Physiol. Neurobiol.* **191**, 67–74 (2014).
 88. Van Oort, P. M. P. *et al.* Exhaled breath metabolomics for the diagnosis of

- pneumonia in intubated and mechanically-ventilated intensive care unit (ICU)-patients. *Int. J. Mol. Sci.* **18**, 449 (2017).
89. Koureas, M. *et al.* Comparison of Targeted and Untargeted Approaches in Breath Analysis for the Discrimination of Lung Cancer from Benign Pulmonary Diseases and Healthy Persons. *Molecules* **26**, 2609 (2021).
 90. Phillips, M., Cataneo, R. N., Lebauer, C., Mundada, M. & Saunders, C. Breath mass ion biomarkers of breast cancer. *J. Breath Res.* **11**, 016004 (2017).
 91. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
 92. Barnes, S. Overview of Experimental Methods and Study Design in Metabolomics, and Statistical and Pathway Considerations. *Methods Mol. Biol.* **2104**, 1–10 (2020).
 93. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
 94. Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D. & Lu, L. J. Computational and statistical analysis of metabolomics data. *Metabolomics* **11**, 1492–1513 (2015).
 95. Domingo-Almenara, X. & Siuzdak, G. Metabolomics Data Processing Using XCMS. *Methods Mol. Biol.* **2104**, 11–24 (2020).
 96. Lommen, A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal. Chem.* **81**, 3079–3086 (2009).
 97. Domingo-Almenara, X. An introduction to erah package. *J. Chromatogr. A* **1409**, 226–233 (2015).
 98. van Mastrigt, E., de Jongste, J. C. & Pijnenburg, M. W. The analysis of volatile organic compounds in exhaled breath and biomarkers in exhaled breath condensate in children – clinical tools or scientific toys? *Clin. Exp. Allergy* **45**, 1170–1188 (2015).
 99. Broadhurst, D. I. & Kell, D. B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171–196 (2007).
 100. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction.* (Springer, 2009).

101. Leopold, J. H. *et al.* Comparison of classification methods in breath analysis by electronic nose. *J. Breath Res.* **9**, 046002 (2015).
102. Paleczek, A. & Rydosz, A. Review of the algorithms used in exhaled breath analysis for the detection of diabetes. *J. Breath Res.* **16**, 026003 (2022).
103. Peng, J., Jury, E. C., Dönnies, P. & Ciurtin, C. Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges. *Front. Pharmacol.* **12**, 2667 (2021).
104. Huang, J., Ma, S. & Zhang, C.-H. *The Iterated Lasso for High-Dimensional Logistic Regression.* (2008).
105. Brereton, R. G. & Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *J. Chemom.* **28**, 213–225 (2014).
106. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
107. Amato, F. *et al.* Artificial neural networks in medical diagnosis. *J. Appl. Biomed.* **11**, 47–58 (2013).
108. Lê Cao, K.-A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253 (2011).
109. Zhou, M. *et al.* Rapid breath analysis for acute respiratory distress syndrome diagnostics using a portable two-dimensional gas chromatography device. *Anal. Bioanal. Chem.* **411**, 6435 (2019).
110. Monasta, L. *et al.* Inflammatory bowel disease and patterns of volatile organic compounds in the exhaled breath of children: A case-control study using Ion Molecule Reaction-Mass Spectrometry. *PLoS One* **12**, e0184118 (2017).
111. Beccaria, M. *et al.* Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *J. Breath Res.* **13**, 016005 (2018).
112. Van Berkel, J. J. B. N. *et al.* Development of accurate classification method based on the analysis of volatile organic compounds from human exhaled air. *J. Chromatogr. B* **861**, 101–107 (2008).
113. Gashimova, E. *et al.* Exhaled breath analysis using GC-MS and an electronic nose for lung cancer diagnostics. *Anal. Methods* **13**, 4793–4804 (2021).
114. Rudnicka, J., Kowalkowski, T. & Buszewski, B. Searching for selected VOCs in human breath samples as potential markers of lung cancer. *Lung Cancer* **135**,

- 123–129 (2019).
115. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16 (2012).
116. Lee, L. C., Liong, C.-Y. & Jemain, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst* **143**, 3526–3539 (2018).

Capítulo 2

Objetivos

El objetivo principal de esta Tesis Doctoral ha sido desarrollar y aplicar el análisis del aire exhalado fuera de línea u *off-line* utilizando tecnologías basadas en espectrometría de masas para la detección de compuestos orgánicos volátiles (VOCs) en el aire exhalado de población pediátrica y en mujeres en edad fértil, e identificar biomarcadores de enfermedades respiratorias y/o atópicas de alta prevalencia como el asma utilizando métodos de análisis de datos robustos y adecuados a datos con alta dimensionalidad.

A su vez, el objetivo global puede desglosarse en los siguientes objetivos particulares:

1. Determinar el estado actual de la técnica del análisis del aire exhalado. Realizar una rigurosa revisión sistemática de los estudios previos donde se han llevado a cabo determinaciones de VOCs en aire exhalado con el propósito de discriminar la presencia de una enfermedad o de monitorizar patologías en población pediátrica (**Capítulo 3**).
2. Diseñar y optimizar el protocolo de recogida de aire exhalado de forma no invasiva en adultos. Adaptar el protocolo anterior dirigido a población pediátrica de corta edad no colaborativa. Optimizar el análisis del aire exhalado *off-line* mediante TD-GC/q-MS (desorción térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple) (**Capítulo 4**).
3. Recoger las muestras de aire exhalado de las parejas de madres e hijos de la cohorte de nacimiento NELA (*Nutrition in Early Life and Asthma*) a los tres meses de edad del niño, y analizarlas mediante TD-GC/q-MS.
4. Establecer un flujo de trabajo de código abierto para el preprocesamiento de los datos masivos en bruto generados por cromatografía de gases-espectrometría de masas para obtener una matriz útil para su análisis estadístico posterior que se realizara en cada capítulo según las diferentes hipótesis de trabajo planteadas (**Capítulo 4**).
5. Hallar perfiles de VOCs en el aire exhalado de mujeres en edad fértil que permitan discriminar entre asmáticas con otras enfermedades atópicas coexistentes (rinitis alérgica, dermatitis atópica, conjuntivitis alérgica, alergia alimentaria o alergia a medicamentos) y no asmáticas (con o sin enfermedades atópicas). Identificar factores asociados a la enfermedad del asma (**Capítulo 5**).
6. Comparar los niveles de VOCs en el aire exhalado humano con los de las muestras del aire ambiental, e identificar factores que deban tenerse en cuenta para el análisis de datos (ej. época del año en la se recogen y analizan las

- muestras, etc.). Establecer estrategias para reducir los sesgos derivados de dichos factores **(Capítulo 5)**.
7. Estudiar la relación entre los valores obtenidos en los parámetros de función pulmonar, y los niveles de VOCs relacionados con el estrés oxidativo en el aire exhalado en los niños de la cohorte NELA a los 3 meses de edad **(Capítulo 6)**.
 8. Evaluar la influencia de la presencia de humedad en los hogares en los niveles de VOCs exógenos y endógenos en el aire exhalado de las madres y los niños de la cohorte NELA **(Capítulo 7)**.
 9. Adaptar el uso de sistemas analíticos en línea u *on-line* como SESI-HRMS (ionización secundaria por *electrospray*-espectrometría de masas de alta resolución) mediante un protocolo de recogida de muestras *off-line*. Evaluar el efecto del almacenamiento a diferentes condiciones de muestras de aire exhalado sobre los niveles de los VOCs determinados mediante SESI-HRMS **(Capítulo 8)**.

Capítulo 3

Análisis de compuestos orgánicos volátiles exhalados en pediatría clínica: revisión sistemática

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, José María Pastor Hernández, Óscar Yanes Torrado,
Manuel Cánovas Díaz, Teresa de Diego Puente y María Vinaixa Crevillent

**Exhaled volatile organic compounds analysis in clinical pediatrics: a
systematic review**

Pediatric Research **89**, 1352-1363 (2021). doi: 10.1038/s41390-020-01116-8.

RESUMEN

Desde hace tiempo se reivindica que la determinación de los compuestos orgánicos volátiles (VOCs) presentes en el aire exhalado o volatiloema exhalado, puede constituir una fuente potencial de biomarcadores, aplicables de manera no invasiva en la clínica rutinaria. Sin embargo, todavía no se ha podido demostrar la viabilidad de la aplicación del volatiloema exhalado en la práctica clínica, especialmente en pediatría, donde es más urgente disponer de mejores métodos de diagnóstico y seguimiento no invasivos. Este capítulo se presenta la primera evaluación formal del potencial clínico del volatiloema exhalado en la población pediátrica. Para ello, se llevó a cabo una rigurosa revisión sistemática de los artículos recogidos en las bases de datos Web of Science, SCOPUS y PubMed, siguiendo las directrices de la declaración PRISMA. Dos revisores independientes determinaron que 22 de los 229 registros encontrados inicialmente, satisfacían los criterios de inclusión. La calidad de los estudios seleccionados fue evaluada utilizando el sistema de puntuación QUADAS-2 (*Quality Assessment of Diagnostic Accuracy Studies 2*). Asimismo, en este trabajo se muestra una recopilación de los VOCs presentes en el aire exhalado que han sido considerados relevantes para varias patologías respiratorias, infecciosas y metabólicas. Además, se llevó a cabo un análisis de enriquecimiento funcional para evaluar la cobertura metabólica asociada. Nuestros resultados indican que la investigación actual permanece estancada en un entorno exploratorio preclínico, y que el diseño de experimentos preliminares basados en las prácticas propias de la metabolómica debería impulsar la traslación clínica del análisis de VOCs en el aire exhalado.

INTRODUCCIÓN

Una amplia gama de compuestos orgánicos volátiles (VOCs) está presente en diferentes matrices del cuerpo humano, tal es el caso de las emanaciones de la piel, orina, sangre, saliva, heces y aire exhalado¹. La composición de los VOCs presentes en el aire exhalado puede proporcionar información sobre el estado metabólico de un individuo y, por lo tanto, el análisis del volatiloema exhalado podría ser una estrategia prometedora para el diagnóstico clínico y la monitorización terapéutica²⁻⁴. El análisis de VOCs en el aire exhalado se ha empleado en población adulta para la búsqueda de biomarcadores en patologías como el asma⁵, la enfermedad pulmonar obstructiva crónica⁶, la fibrosis quística (CF)^{7,8}, el cáncer de pulmón^{9,10}, el cáncer colorrectal¹¹, el carcinoma gástrico¹², el cáncer de tiroides¹³, la tuberculosis¹⁴, la cirrosis hepática¹⁵ y la diabetes mellitus tipo 2¹⁶, entre otras. Además, también se han utilizado para distinguir estadios de una enfermedad o exacerbaciones¹⁷⁻¹⁹. Cabe destacar que el volatiloema

exhalado puede obtenerse directamente en el centro asistencial de una manera bastante sencilla y directa, con una disponibilidad ilimitada de muestra y utilizando métodos de muestreo no invasivos⁴. Este hecho sitúa al aire exhalado como el tipo de muestra biológica ideal para trabajar en la población pediátrica, donde el análisis de VOCs ha despertado un gran interés para el diagnóstico precoz y la vigilancia de las patologías infantiles, ya que la mayoría se convierten en enfermedades crónicas que suponen una gran carga socioeconómica²⁰⁻²². En particular, el análisis de VOCs en el aire exhalado podría contribuir al diagnóstico del asma, que sigue siendo un reto debido a la baja sensibilidad intrínseca de la espirometría con broncodilatación²³. Por lo tanto, se espera que el análisis de volatilo exhalado se generalice para el diagnóstico en niños muy pequeños no colaboradores²⁴.

Sin embargo, a pesar de todos estos prometedores resultados iniciales, el análisis del volatilo en el aire exhalado ha experimentado hasta la fecha poco o ningún progreso desde el laboratorio hasta el entorno clínico. Esto se ha atribuido en gran medida a la falta de estandarización en la recogida del aire exhalado, en las plataformas de detección de perfiles de VOCs y en el análisis robusto de los datos^{20,25}. No obstante, la comunidad científica se esfuerza con firmeza en convertir el análisis del aire exhalado en una herramienta clínica de rutina, mediante la implementación de nuevas tecnologías y apostando por el consenso comunitario para la estandarización del proceso²⁶. Por otro lado, la comunidad médica se pregunta si los actuales métodos analíticos, técnicamente complejos, para medir los VOCs en el aire exhalado son rentables en términos de viabilidad y reproducibilidad, y se cuestiona su traslación a la práctica clínica²⁷. El análisis de VOCs en el aire exhalado en pediatría sigue estando poco estudiado y las conclusiones sobre su viabilidad clínica siguen siendo inciertas y contradictorias. En el caso del asma infantil, por ejemplo, mientras que en algunos estudios las mediciones de VOCs han presentado un valor clínico limitado, en otros han mostrado una precisión de predicción de moderada a buena para el diagnóstico del asma pediátrico y cierto potencial para predecir las exacerbaciones del asma²⁸⁻³⁰. Por lo tanto, actualmente, el potencial clínico de las mediciones de VOCs en el aire exhalado es objeto de debate.

Con el propósito de aclarar si el conjunto de pruebas disponibles en la actualidad es lo suficientemente fiable como para respaldar el potencial clínico del análisis de VOCs en el aire exhalado en el campo de la pediatría, aquí presentamos la primera revisión sistemática sobre su aplicación en el diagnóstico clínico o en el seguimiento terapéutico de enfermedades infantiles.

RESULTADOS

Selección de estudios

En total, 229 registros fueron encontrados siguiendo nuestra estrategia de búsqueda en las bases de datos Web of Science, SCOPUS y PubMed. Después de eliminar las entradas duplicadas, 125 artículos fueron examinados. Entre ellos, sólo 43 registros presentaron un resumen que cumplía con los criterios de la revisión y, entonces, fueron examinados íntegramente. Se excluyó un artículo que tenía como objetivo mejorar el diagnóstico del asma combinando los VOCs exhalados, la expresión génica y las mediciones de la función pulmonar puesto que las mediciones de los VOCs habían sido reutilizadas de un estudio anterior³¹. La Figura 1 muestra el diagrama de flujo seguido para la búsqueda y la selección de registros basado en la guía PRISMA. Finalmente, 22 artículos cumplieron nuestros criterios de inclusión/exclusión y fueron incluidos en la síntesis cualitativa^{30,32-52}.

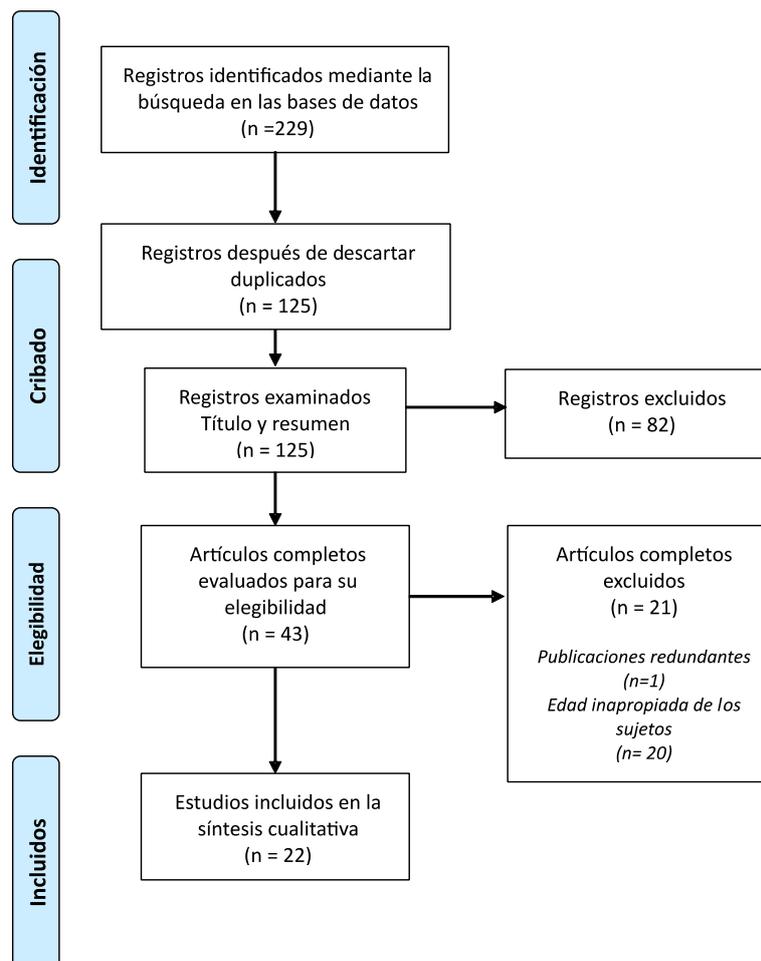


Figura 1. Selección e inclusión de estudios. Diagrama de flujo de la búsqueda sistemática guiada por la declaración PRISMA⁵³ y selección de registros de la literatura.

Características de los estudios

Los datos de los 22 estudios incluidos en la síntesis cualitativa fueron revisados exhaustivamente y la información extraída está resumida en la Tabla 1. La población de estudio comprendía grupos de niños y adolescentes con una edad media o mediana menor de 18 años. Estos estudios abordaban enfermedades respiratorias siendo el asma la más estudiada (11 de los 22 registros) y en menor medida otras enfermedades respiratorias como la fibrosis quística (CF), el síndrome de apnea obstructiva del sueño (OSAS), o la discinesia ciliar primaria (PCD). Sin embargo, esta revisión sistemática no se limitó exclusivamente a enfermedades respiratorias, sino también a otros trastornos en los que el análisis de VOCs fue utilizado como herramienta de diagnóstico o de monitorización. Entre ellas, se incluyeron tanto enfermedades infecciosas (malaria, infección por rinovirus, e infecciones bacterianas en pacientes con enfermedades respiratorias) y enfermedades metabólicas, como la enfermedad del hígado graso no alcohólico (NALFD), la enfermedad hepática crónica (CLD), la enfermedad renal crónica (CKD) y la enfermedad inflamatoria intestinal (IBD).

Para los tres principales grupos de patologías (enfermedades respiratorias, enfermedades infecciosas y enfermedades metabólicas), el análisis de VOCs se empleó para distinguir entre los pacientes con una determinada enfermedad y los sujetos control, para predecir diferentes estadios de la enfermedad o exacerbaciones, y para discriminar entre enfermedades con síntomas similares y etiologías diferentes. La mayoría de los estudios presentaban un diseño transversal (73%) y el resto eran estudios longitudinales.

Tabla 1. Información obtenida de los estudios incluidos en la síntesis cualitativa.

Referencia	Enfermedades diana y población de estudio			Recogida de muestras de aire exhalado		Análisis del aire exhalado	Análisis de datos	Resultados de los estudios
	Enfermedad	Tamaño población estudio / Diseño estudio	Edad (años)	Muestra de aire ambiental / filtro de VOCs / muestreo en la misma habitación	Porción de aire exhalado muestreado / contenedor de aire exhalado	Métodos analíticos	Métodos estadísticos	n° VOCs seleccionados / tasa de clasificación / sensibilidad / especificidad
Tecnologías basadas en sensores portátiles								
32	Asma	A) 20 asmáticos vs. 22 HCTR	Asmáticos: 12,7± 3,1 CF: 14,4± 4,2 HCTR: 9,7± 2,9	No / Sí / Sí	Aire exhalado mixto / -	Matrices de sensores -eNose: tres sensores de óxido metálico	ANNs	_ / AUC = 0,79 / 74% / 91%
	CF	B) 13 CF vs. 22 HCTR						_ / AUC = 0,87 / 85% / 77%
	Asma vs. CF	C) 20 asmáticos vs. 13 CF						_ / AUC = 0,90 / 89% / 77%
47	CF	A) 25 CF (9 con exacerbación pulmonar) vs. 23 HCTR	CF: 11,4[7,7, 17,9] PCD: 10,7[7,1, 14,5] HCTR: 9,3[5,4, 12,6]	No / Sí / No	Aire exhalado mixto / <i>Spacer Babyhaler</i>	Matrices de sensores -Cyranose 320: matriz de 32 nanosensores de nanocompuestos poliméricos	PCR+CDA	_ / AUC = 0,76 / 84% / 65%
	PCD	B) 25 PCD (4 con exacerbación pulmonar) vs. 23 HCTR						_ / AUC = 0,80 / 88% / 52%
	CF con exacerbaciones pulmonares	C) 9 CF con exacerbación pulmonar vs. 16 CF estables						_ / AUC = 0,76 / 89% / 56%
	PDC con exacerbaciones pulmonares	D) 4 PCD con exacerbación pulmonar vs. 19 PCD estables						_ / AUC = 0,9 / 100% / 90%
	CF vs. PCD	E) 25 CF (9 con exacerbación pulmonar) vs. 25 PCD (4 con exacerbación pulmonar)						_ / AUC = 0,77 / 84% / 60%
	CF estable vs. PCD estable	F) 16 CF estables vs. 19 PCD estables						_ / AUC = 0,77 / 95% / 63%
	<i>S.aureus</i> en CF	G) 13 CF infectados con <i>S.aureus</i> vs. 12 CF sin infección de <i>S.aureus</i>						X / _ / _ / _
	<i>H. influenzae</i> en PCD	H) 6 PCD infectados con <i>H. influenzae</i> vs. 19 PCD sin infección de <i>H. influenzae</i>						X / _ / _ / _

Tabla 1 (Continuación I). Información obtenida de los estudios incluidos en la síntesis cualitativa.

Referencia	Enfermedades diana y población de estudio			Recogida de muestras de aire exhalado		Análisis del aire exhalado	Análisis de datos	Resultados de los estudios
	Enfermedad	Tamaño población estudio / Diseño estudio	Edad (años)	Muestra de aire ambiental / filtro de VOCs / muestreo en la misma habitación	Porción de aire exhalado muestreado / contenedor de aire exhalado	Métodos analíticos	Métodos estadísticos	n° VOCs seleccionados / tasa de clasificación / sensibilidad / especificidad
50	sibilancias inducidas por rinovirus	A) 37 con sibilancias (rinovirus positivo) vs. 26 asintomáticos (rinovirus positivo)	1,8±0,8	Sí / Sí / No	Aire exhalado mixto / Spacer Babyhaler	Matrices de sensores -Cyranose 320: matriz de 32 nanosensores de nanocompuestos poliméricos	PCA, prueba t de Student y CDA	_ / AUC = 0,77 / 73% / 96%
		B) 37 ex-sintomáticos (rinovirus positivo) vs. 26 asintomáticos (rinovirus positivo)						_ / AUC = 0,84 / 84% / 75,9%
	sibilancias no inducidas por rinovirus	C) 44 con sibilancias (rinovirus negativo) vs. 71 asintomáticos (rinovirus negativo)						_ / AUC = 0,81 / 75,6% / 71,4%
		D) 44 ex-sintomáticos (rinovirus negativo) vs. 71 asintomáticos (rinovirus negativo)						_ / AUC = 0,67 / 62,9% / 65,6%
51	OSAS	A) 18 OSAS vs. 10 CTR roncadores	OSAS: 8,2±1,9 CTR roncadores: 9,1±1,6	Sí / Sí / Sí	Aire exhalado tardío / Bolsa Mylar® (recubierta de teflón)	Matrices de sensores -Cyranose 320: matriz de 32 nanosensores de nanocompuestos poliméricos	PCA+LR	4 PCs (PC 3 mejor discriminante) / AUC = 0,84 / 78% / 70%
Tecnologías basadas en espectrometría de masas								
52	Asma	A) 63 asmáticos vs. 57 HCTR	5-16	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (5L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	Análisis discriminante por pasos	8 VOCs / 92% / 85% / 95%
	Asma atópica	B) 42 con asma atópica vs. 57 HCTR						7 VOCs* / _ / 90% / 89%
		C) 42 con asma atópica vs. 21 con asma no atópica						7 VOCs* / _ / 84% / 95%
33	Asma alérgica	A) 35 con asma alérgica (13 con rinitis alérgica) vs. 15 HCTR	4-13	Sí / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (1L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : HS-SPME(DVB/CAR/PDMS)-GC-MS (cuadrupolo simple)	PLS-DA	28 VOCs / 88% / _ / _
34	Asma alérgica	A) 32 asmáticos (10 con rinitis alérgica) vs. 27 HCTR	Asmáticos: 4-16 HCTR: 3-6	Sí / No / No	Aire exhalado mixto / Bolsa Tedlar® (1L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : HS-SPME(DVB/CAR/PDMS)-GCxGC-MS(TOF)	PLS-DA	9 VOCs / 96% / 98% / 93%

Tabla 1 (Continuación II). Información obtenida de los estudios incluidos en la síntesis cualitativa.

Referencia	Enfermedades diana y población de estudio			Recogida de muestras de aire exhalado		Análisis del aire exhalado	Análisis de datos	Resultados de los estudios
	Enfermedad	Tamaño población estudio / Diseño estudio	Edad (años)	Muestra de aire ambiental / filtro de VOCs / toma de muestra en la misma habitación	Porción de aire exhalado muestreado / contenedor de aire exhalado	Métodos analíticos	Métodos estadísticos	n° VOCs seleccionados / tasa de clasificación / sensibilidad / especificidad
36	RW	A) 202 con síntomas de sibilancias recurrentes vs. 50 sin síntomas de sibilancias	2-4	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (1L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	ASCA + SLR	28 VOCs / 73% / 79% / 50%
37	Asma	A) 76 asmáticos vs. 49 HCTR	2-4 hasta 6	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (1L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	RF	12 VOCs* / 85,84% / 81,5% / 74,2%
		B) 76 asmáticos vs. 121 con sibilancias transitorias						12 VOCs* / 77,8% / 74,6% / 70%
			2-4				d-PLS-DA	17 VOCs / 80% / _ / _
35	Asma	A) 11 asmáticos vs. 12 HCTR	8-16	Sí / No / No	Aire exhalado tardío / _	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Tenax®/Carbotrap)-GC-MS(trampa iónica)	PLS-DA	8 VOCs / 51% / _ / _
38	Exacerbaciones asmáticas (intersubjecto)	A) 40 asmáticos (29 con asma atópica) - 16 de los 40 asmáticos experimentaron exacerbaciones	6-16	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (5L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	SVM	7 VOCs / 91% / 79% / 100%
	Exacerbaciones asmáticas (intrasubjecto)	B) 16 con exacerbaciones asmáticas						6 VOCs / 96% / 100% / 93%
39	Exacerbaciones asmáticas	A) 9 con asma persistente no controlada vs. 34 con asma persistente controlada	6-18	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (5L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	RF	15 VOCs / AUC = 0,86 / 81% / 67%
30	Exacerbaciones asmáticas	A) 16 asmáticos estables vs. 16 asmáticos con una exacerbación 14 días después del muestreo	6-18	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (5L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	RF	7 VOCs / AUC = 0,90 / 88% / 75%
40	CF	A) 48 CF vs. 57 HCTR	CF: 13,0±0,6 HCTR: 9,9±0,4	No / No / Sí	Aire exhalado mixto / Bolsa Tedlar® (5L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Carbograph 1TD/Carbopack X)-GC-MS(TOF)	Análisis discriminante por pasos	10 VOCs / AUC = 0,962 / _ / _
	<i>P. aeruginosa</i> en CF	B) 23 CF infectados con <i>P. aeruginosa</i> vs. 17 CF sin infección con <i>P. aeruginosa</i>						14 VOCs* / 100% / _ / _

Tabla 1 (Continuación III). Información obtenida de los estudios incluidos en la síntesis cualitativa.

Referencia	Enfermedades diana y población de estudio			Recogida de muestras de aire exhalado		Análisis del aire exhalado	Análisis de datos	Resultados de los estudios
	Enfermedad	Tamaño población estudio / Diseño estudio	Edad (años)	Muestra de aire ambiental / filtro de VOCs / toma de muestra en la misma habitación	Porción de aire exhalado muestreado / contenedor de aire exhalado	Métodos analíticos	Métodos estadísticos	n° VOCs seleccionados / tasa de clasificación / sensibilidad / especificidad
41	<i>S. aureus</i> en CF	A) 13 CF infectados con <i>S. aureus</i> vs. 5 CF sin infección con <i>S. aureus</i>	CF infectados con <i>S. aureus</i> : 13 [9, 20] CF sin infección con <i>S. aureus</i> : 7 [7, 13]	Sí / Sí / No	Aire exhalado final / Bolsa Tedlar® (3L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Tenax® TA) – GC-MS (cuadrupolo simple)	s-PLS-DA	9 VOCs / _ / 100% / 80%
42	Malaria	A) 17 sujetos febriles con infección de malaria vs. 18 sujetos febriles sin infección de malaria	3-15	Sí / No / No	Aire exhalado tardío / Bolsa SamplePro® FlexFilm (3 L)	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : TD(Tenax® 60/80, Carbograph 1 60/80 y Carboxen 1003 40/60)-GCxGC-MS(TOF)	Algoritmo de clasificación de abundancia acumulada y media más cercana (clasificación binaria)	6 VOCs / 83% / 71% / 94%
							Prueba t de Student	1 VOC / 69% / _ / _
								1 VOC / 77% / _ / _
44	NAFLD	A) 37 obesos con NAFLD vs. 20 obesos con hígado normal	6-18	No / Sí / No	Aire exhalado tardío / Bolsa Mylar®	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : SIFT-MS	CDA y selección de variables por pasos	4 picos iónicos / AUC = 0,763 / _ / _
							ANCOVA	5 VOCs / _ / _ / _
45	CLD	A) 49 CLD (20 con fibrosis avanzada, 20 sin fibrosis avanzada y 9 con patología no disponible) vs. 55 HCTR	CLD: 12,1±3,7 HCTR: 13,6±4,2	No / Sí / No	Aire exhalado tardío / Bolsa Mylar®	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : SIFT-MS	ANCOVA	7 VOCs / _ / _ / _
							Selección de variables por pasos, LDA y MLR	5 VOCs / AUC = 0,97 / _ / _
43	CDK	A) 48 CKD y 8 con un KTx vs. 60 HCTR	4-18	Sí / No / No	Aire exhalado final / -	Espectrometría de masas -Análisis de aire exhalado <i>on-line</i> : PTR-TOF-MS	Mapa de calor, PCA y prueba de Kruskal-Wallis seguida de la prueba de Dunn	6 VOCs / _ / _ / _
46	IBD	A) 62 IBD (11 con UC y 51 con CD) vs. 55 HCTR	IBD: 12,1±3,0 HCTR: 15,7±3,3 CD: 12,4±3,5 UC: 13,2 ±3,7	No / Sí / No	Aire exhalado tardío / Bolsa Mylar®	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : SIFT-MS	LDA, MLR y ANCOVA	3 VOCs / AUC = 0,96 / _ / _
		B) 51 CD vs. 11 UC						X / _ / _ / _

Tabla 1 (Continuación IV). Información obtenida de los estudios incluidos en la síntesis cualitativa.

Referencia	Enfermedades diana y población de estudio			Recogida de muestras de aire exhalado		Análisis del aire exhalado	Análisis de datos	Resultados de los estudios
	Enfermedad	Tamaño población estudio / Diseño estudio	Edad (años)	Muestra de aire ambiental / filtro de VOCs / toma de muestra en la misma habitación	Porción de aire exhalado muestreado / contenedor de aire exhalado	Métodos analíticos	Métodos estadísticos	nº VOCs seleccionados / tasa de clasificación / sensibilidad / especificidad
48	IBD	A) 67 IBD (34 con CD y 33 con UC) vs. 167 CTR (65 CTR gastroenterológico y 102 HCTR)	10-17	Sí / No / No	Aire exhalado mixto / Muestreador Bio-VOC™	Espectrometría de masas -Análisis de aire exhalado <i>off-line</i> : IMR-MS	LASSO+LR	18 VOCs / AUC = 0,925 / 95% / 69%
		B) 34 CD vs. 33 UC						13 VOCs / AUC = 0,934 / 94% / 76%
		C) IBD (34 con CD y 33 con UC) vs. 65 CTR gastroenterológico						15 VOCs / AUC = 0,918 / 94% / 65%
		D) 67 IBD (34 con CD y 33 con UC) vs. 167 CTR (65 CTR gastroenterológico y 102 HCTR) sólo con los VOCs directa o indirectamente calibrados.						12 VOCs / AUC = 0,888 / 94% / 71%
Otras tecnologías								
49	Asma	A) 39 asmáticos vs. 35 HCTR	6-18	No / No / No	Aire exhalado mixto / Bolsa Tedlar® (3L)	Espectroscopia cuántica de banda ancha basada en láser	funciones del paquete <i>lima</i> (paquete para análisis de datos, modelos lineales y expresión diferencial para datos de <i>microarrays</i>)	23 VOCs / _ / _ / _
	CF	B) 15 CF vs. 35 HCTR						12 VOCs / _ / _ / _

Enfermedades respiratorias. Enfermedades infecciosas. Enfermedades metabólicas.

CTR: controles; HCTR: controles sanos; CF: fibrosis quística; PCD: discinesia ciliar primaria; OSAS: síndrome de apnea obstructiva del sueño; RW: sibilancias recurrentes; NAFLD: enfermedad del hígado graso no alcohólico; CLD: enfermedad hepática crónica; CKD: enfermedad renal crónica; KTx: trasplante renal; IBD: enfermedad inflamatoria intestinal; UC: colitis ulcerosa; CD: enfermedad de Crohn; TD: desorción térmica; GC: cromatografía de gases; MS: espectrometría de masas; TOF: tiempo de vuelo; HS-SPME: espacio de cabeza-microextracción en fase sólida; GCxGC: cromatografía de gases bidimensional Integral; SIFT: tubo de selección de iones con flujo; IMR: reacción ion-molecular; PTR: reacción de transferencia de protones; PDMS: polidimetilsiloxano; DVB: divinilbenceno; CAR: carboxeno; ANNs: redes neuronales artificiales; PCR: regresión de componentes principales; CDA: análisis discriminante canónico; PCA: análisis de componentes principales; LR: regresión logística; PLS-DA: análisis discriminante mínimos cuadrados parciales; s-PLS-DA: análisis discriminante mínimos cuadrados parciales disperso; d-PLS-DA: disimilitud-análisis discriminante de mínimos cuadrados parciales; ASCA: ANOVA- análisis de componentes simultáneos; SLR: regresión logística dispersa; RF: *Random Forest* o bosque aleatorio; SVM: máquinas de vector soporte; ANCOVA: análisis de covarianza; LDA: análisis discriminante lineal; MLR: regresión logística múltiple; LASSO: operador de selección y contracción mínima absoluta; AUC: área bajo la curva. * indica que no se ha notificado la identidad de los VOCs. X indica que no se han encontrado VOCs o "huella de aliento" discriminantes. Los datos de edad se presentan como rango de años, media de años ± DE (desviación estándar) o mediana de años [P25, P75].

Porción de aire exhalado muestreado y contenedores de aire exhalado

Los estudios incluidos en nuestra revisión sistemática difieren en cuanto a la fracción de aire exhalado muestreado y los contenedores de recogida del aire exhalado (Tabla 1). La porción de aire exhalado suele clasificarse en aire exhalado tardío, aire exhalado final y aire exhalado mixto⁵⁴. La porción de aire exhalado mixto consiste en la recogida completa del aire liberado durante la respiración, incluido el aire del espacio muerto, mientras que el aire exhalado tardío y el aire exhalado final comprenden la última porción de la exhalación y ambos persiguen la recogida del aire alveolar. Más de la mitad de los estudios incluidos en esta revisión sistemática recogieron aire exhalado mixto como fracción de aire exhalado muestreado. Probablemente esto se deba a que la simplicidad de la toma de muestra de aire exhalado es prioritaria en el caso de pacientes pasivos no colaboradores, como los niños pequeños. En cuanto a las otras dos porciones de aire exhalado, la porción de aire exhalado tardío se recogió en 6 de los 22 estudios incluidos, mientras que en sólo 2 de ellos se recogió aire exhalado final. La principal diferencia entre el aire exhalado tardío y el aire exhalado final es la forma en que se selecciona la fracción objetivo. Así pues, en la recogida del aire exhalado tardío, la primera parte del aire exhalado es descartado, mientras que, en el aire exhalado final, los niveles de CO₂⁴¹ o acetona⁴³ son empleados como indicadores de la porción recogida. El aire exhalado mixto es la porción más sencilla de recoger, pero está muy influenciada por contaminantes ambientales exógenos⁵⁴. Por este motivo, la recogida del aire exhalado mixto suele ir acompañada de diferentes estrategias para reducir la influencia de los VOCs derivados de fuentes exógenas. En este sentido, el muestreo de aire ambiental es de gran utilidad en el análisis del aire exhalado mixto para comprobar minuciosamente la presencia de compuestos exógenos y del entorno. Así, para controlar la calidad del proceso de recogida de aliento, los compuestos del entorno y los compuestos endógenos, pueden monitorizarse en las muestras de aire ambiental y sus niveles pueden compararse con los obtenidos en las muestras de aliento. En este sentido, Schaber *et al.*⁴² controlaron los niveles de acetona e isopreno, dos de los VOCs endógenos más comunes y abundantes en el aire exhalado, y observaron que la abundancia de ambos compuestos era al menos el doble del nivel observado en el aire de la habitación, lo que confirmó la recolección exitosa del aire exhalado. En conclusión, la recogida de aire exhalado mixto combinada con la toma de muestra de aire ambiental es un procedimiento prometedor para la recogida de muestras en lactantes menores de 2 años no colaboradores, siempre que se controle la influencia del aire del entorno.

En cuanto a los colectores de recogida del aliento, las bolsas de polímeros (Tedlar®) seguidas de las bolsas aluminizadas (Mylar®) fueron las más utilizadas (Tabla 1). El lavado con un gas inerte, como el nitrógeno, de las bolsas Tedlar® antes del muestreo demostró ser una herramienta eficaz para reducir los niveles de fenol y N,N-dimetilacetamida (dos contaminantes intrínsecos de las bolsas Tedlar®) y evitar la contaminación cruzada en bolsas reutilizadas³³. También se utilizaron, en menor medida, otros colectores como Bio-VOC™³³. Además de los VOCs del aire exhalado, la fracción con menor volatilidad contenida en el aire exhalado condensado (EBC) también demostró ser una matriz adecuada para ayudar a discriminar el asma persistente en niños y adolescentes, superando la sensibilidad de la espirometría tradicional con prueba de broncodilatación²³.

Plataformas analíticas empleadas en la medición de VOCs en aire exhalado

En los estudios revisados, el análisis de VOCs en el aire exhalado fue llevado a cabo con diferentes tecnologías. La espectrometría de masas (MS) se utilizó en 17 de los 22 estudios, seguida de lejos por las matrices de sensores y la espectroscopia basada en láser, utilizados en sólo 4 y 1 estudios, respectivamente (Tabla 1). La espectroscopia cuántica de banda ancha basada en láser sigue siendo una tecnología no consolidada con escasa repetibilidad⁴⁹. Por otro lado, el coste reducido, la portabilidad y la facilidad de manejo de las tecnologías basadas en matrices de sensores de gases las convierten en una tecnología adecuada para la traslación clínica del análisis del aire exhalado^{23,55}. Sin embargo, la identificación de los VOCs es intrínsecamente imposible con las matrices de sensores, y su uso queda relegado a la obtención de una “huella de aliento” o un patrón de respuestas de las señales del sensor a partir de una mezcla específica de VOCs⁵⁶. Las plataformas analíticas basadas en la MS ofrecen la mayor sensibilidad y selectividad y permiten la identificación y cuantificación de los VOCs en el aire exhalado, normalmente a nivel de traza^{57,58}. Además, tanto la sensibilidad como la selectividad dependen de la configuración particular de cada plataforma de espectrometría de masas. Con esta finalidad, la configuración más común (12 de los 22 estudios incluidos en la síntesis cualitativa) consistió en utilizar métodos de preconcentración fuera de línea u *off-line* junto con la GC-MS (cromatografía de gases acoplada a MS de alta o baja resolución). Los métodos de preconcentración fuera de línea suelen implicar un procedimiento de dos pasos consecutivos. En primer lugar, los VOCs del aire exhalado quedan atrapados en un adsorbente activo (es decir, una fibra de microextracción en fase sólida (SPME) o tubos empaquetados (por ejemplo, Tenax)) y, posteriormente, se desorben térmicamente para

liberar los VOCs atrapados y concentrarlos en una trampa fría, previa al puerto de inyección del sistema de GC-MS.

Por otra parte, 5 estudios utilizaron una tecnología con prestaciones inherentes para el análisis del aire exhalado en línea u *on-line*, como SIFT-MS (tubo de selección de iones con flujo - espectrometría de masas), IMR-MS (Reacción ion-molecular - Espectrometría de Masas) o PTR-MS (reacción de transferencia de protones – espectrometría de masas). Sin embargo, sólo el estudio que empleó PTR-MS llevó a cabo un análisis del aire exhalado en línea u *on-line* a tiempo real⁴³ (Tabla 1). En cambio, los 4 restantes poseían una configuración fuera de línea u *off-line*, porque todos ellos contaron con un paso de recogida y almacenamiento del aire exhalado previo al análisis de VOCs (Tabla 1). En este sentido, hay una serie de ventajas asociadas al uso de métodos de preconcentración *off-line* como la posibilidad de realizar muestreos en lugares alejados a la plataforma de análisis o de adaptar diferentes adsorbentes a familias específicas de VOCs y la capacidad de preconcentrar VOCs, entre otras^{3,54,58}. Además, estos equipos tienen un precio más asequible que los utilizados en el análisis del aire exhalado *on-line*. Sin embargo, el análisis *on-line* a tiempo real no requiere ningún paso de preconcentración, lo que simplifica los procedimientos operativos y, por lo tanto, es más apto para el diagnóstico clínico diario^{57,59}. De hecho, el análisis *on-line* a tiempo real de VOCs en el aire exhalado es hoy en día una realidad gracias al empleo de tecnologías emergentes como SESI-HRMS (ionización secundaria por *electrospray* – espectrometría de masas de alta resolución), o PTR-MS^{24,60–62}. Cabe destacar que SESI-HRMS es una de las tecnologías mejor posicionadas y con mayor potencial para la traslación clínica del análisis del aliento. Los recientes avances en los procedimientos de estandarización con esta tecnología permitirán hacer realidad las primeras comparaciones interlaboratorios, un primer paso obligatorio para su introducción en la práctica clínica de rutina en pediatría^{24,63,64}. Sin embargo, el fin de esta revisión no es describir en detalle las tecnologías de análisis del aire exhalado *on-line*. No obstante, Bruderer *et al.* llevaron a cabo un exhaustivo resumen de estas técnicas²⁴.

Análisis de datos

El primer paso en el análisis de datos de los VOCs en el aire exhalado, es la conversión de los datos brutos obtenidos por espectrometría de masas en una matriz de picos adecuada para el posterior análisis estadístico (preprocesamiento de datos). Sin embargo, este paso apenas es descrito en ninguno de los estudios incluidos en esta síntesis cualitativa. En la mayoría de los casos, se recurre a soluciones sencillas para el usuario mediante un *software* privativo que dificulta la reproducibilidad y las

comparaciones entre laboratorios, donde el usuario no siempre conoce la complejidad del proceso de análisis. Después de este paso de preprocesamiento de datos, se utilizó una amplia gama de diferentes algoritmos de aprendizaje supervisado que incluían modelos lineales (ej. PLS-DA (análisis discriminante mínimos cuadrados parciales)^{33,34}, o no lineales (ej. *Random Forest* (bosque aleatorio), Máquinas de Vector Soporte (SVM)⁶⁵⁻⁶⁷). En la mayoría de casos, la evaluación del rendimiento de los modelos se realizó sobre el mismo conjunto de datos usado para su creación, utilizando *bootstrapping* o cualquier forma de validación cruzada (*n-fold*, *leave-one-out*, etc.) (Tabla 1). Tres de ellos dividieron su población de estudio en un conjunto de entrenamiento y un conjunto de validación o testeo^{37,44,50}. Sin embargo, ninguno de ellos incluyó una validación externa con un nuevo grupo de sujetos (verdadera validación externa). Hay que tener en cuenta que los conjuntos de datos del volatiloма suelen presentar una alta dimensionalidad (mayor número de VOCs o fragmentos iónicos en comparación con el número de sujetos del estudio), existiendo múltiples variables correlacionadas, con una representación bastante baja de los VOCs en todas las muestras, lo que conlleva matrices dispersas. Esto plantea problemas como la multicolinealidad y la problemática de la alta dimensionalidad, entre otros^{68,69}, que dificultan su análisis estadístico.

Por último, la identificación de VOCs se realizó principalmente a través de la comparación espectral con la biblioteca NIST (*The National Institute of Standards and Technology*). Únicamente 5 estudios indicaron que habían usado estándares comerciales para el cálculo de los índices de retención (RI) para la identificación. Además, ninguno de los estudios revisados comunicó el nivel de confianza utilizado para la identificación de los VOCs⁷⁰.

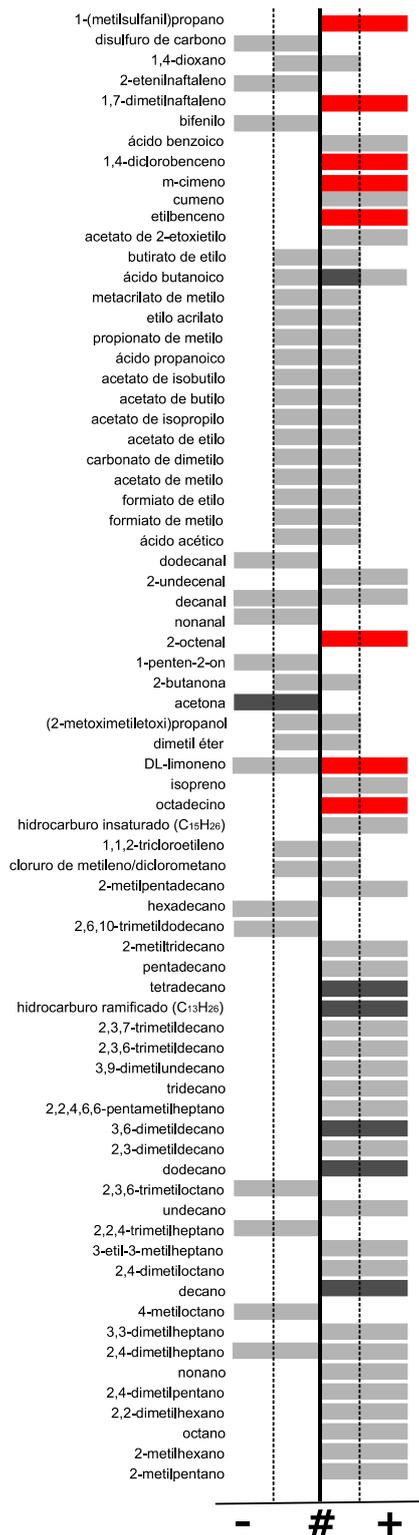
VOCs en el aire exhalado relacionados con enfermedades pediátricas

En el Apéndice 1 se muestra una lista completa de los VOCs discriminantes, clasificados según las familias químicas, que se consideraron relevantes en cualquiera de los estudios incluidos en la síntesis cualitativa. De este modo, en contraste con la población adulta¹⁸, se observó que los niveles de acetona estaban asociados negativamente con el asma infantil^{33,37}. Por el contrario, los niveles de alcanos estaban asociados positivamente con el asma en la población pediátrica¹⁷, lo que coincide con lo observado previamente en la población adulta. Muchos de los alcanos detectados en el aire exhalado han sido descritos como subproductos de la peroxidación lipídica, que es un proceso normalmente desencadenado por la inflamación que, a su vez, está típicamente asociada al asma³. Otros productos derivados de la peroxidación lipídica,

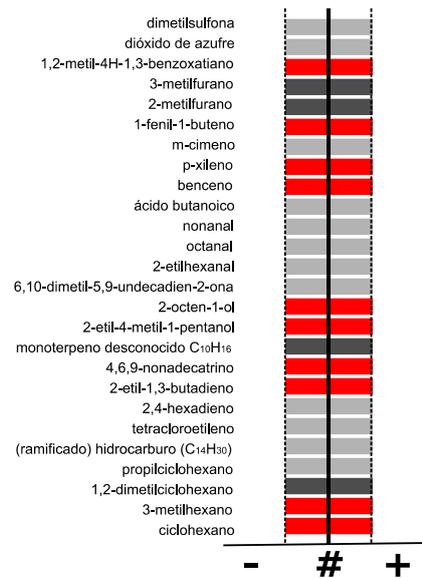
como los aldehídos^{3,71}, estaban desregulados en el asma infantil^{33–35,37}, tendencia que también se observa en los adultos¹⁷. Cabe destacar que la acetona, los alcanos y los aldehídos también estaban desregulados en otras enfermedades infantiles distintas del asma, como la enfermedad renal crónica (CKD), la enfermedad del hígado graso no alcohólico (NALFD), la enfermedad inflamatoria intestinal (IBD) o las infecciones bacterianas^{41,43,44,48}. Por otra parte, se notificaron diferencias en el isopreno, entre controles obesos y pacientes con NAFLD⁴⁴ y entre pacientes con CKD y los controles sanos⁴³, que como se ha mencionado, junto con acetona es un compuesto frecuente en el aliento humano¹. Por último, en el aire exhalado de sujetos con problemas de hígado, intestino o riñón se detectaron niveles elevados de amoníaco^{43,44,48}. En conjunto, los resultados del Apéndice 1 sugieren que la mayoría de los VOCs del aliento descritos tienen una baja especificidad diagnóstica cuando son considerados individualmente como un único biomarcador. A modo de ejemplo, se puede señalar el caso de la 2-butanona, que se ha asociado al asma, a la infección por *Staphylococcus aureus* en pacientes con CF y a la IBD. Ahora bien, una combinación determinada de VOCs en el aire exhalado, un patrón que constituye el análisis del aire exhalado completo, está vinculada a una patología en particular⁷². A este respecto, en la Figura 2 se muestran los perfiles de VOCs exhalados notificados para el asma y la CF y para las exacerbaciones del asma, estudiados en la población pediátrica. Llama la atención que, a pesar de compartir algunos puntos en común, los perfiles de asma y CF son totalmente diferentes.

Por otro lado, la utilidad de los biomarcadores en el aire exhalado ha sido cuestionada si no se dispone del conocimiento de los mecanismos bioquímicos implicados en la producción de éstos⁷³. Estos mecanismos bioquímicos pueden estudiarse mediante el análisis funcional. Sin embargo, hay que tener en cuenta que, para interpretar correctamente los resultados de estos análisis funcionales, la cobertura del metabolismo que puede alcanzarse basándose únicamente en los VOCs es bastante baja. Para explorar con más detalle la parte de la red metabólica humana que abarca el análisis de VOCs en el aliento en las enfermedades pediátricas, se efectuó un análisis de enriquecimiento de rutas metabólicas mediante *FELLA*⁷⁴, utilizando los compuestos del Apéndice 1 como datos de entrada. En base a esto, *FELLA* construyó una representación jerárquica del metabolismo humano utilizando la Enciclopedia de Genes y Genomas de Kioto (KEGG) como base de datos^{75,76}, y aplicó un proceso difusivo nulo derivando una subred relevante (Figura 3) con una lista de rutas metabólicas y elementos intermedios (enzimas y reacciones) afectados.

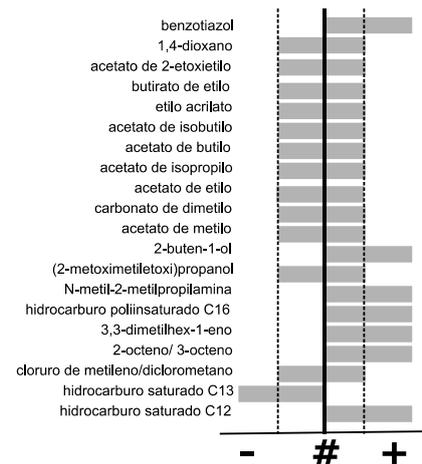
(A) Asma vs. HCTR



(B) Exacerbaciones asmáticas vs. Asmáticos



(C) CF vs. HCTR



Leyenda

- Compuesto seleccionado como discriminante en un estudio
- Compuesto seleccionado como discriminante en más de un estudio
- Diferencias significativas entre ambos grupos
- +
-
- #

Asociación positiva con la enfermedad
 Asociación negativa con la enfermedad
 Discriminante de la enfermedad

Figura 2. Perfiles discriminantes obtenidos a partir de los VOCs seleccionados por los estudios incluidos en la síntesis cualitativa. (A) Pacientes con asma o asma alérgica vs. controles sanos (HCTR). (B) Asmáticos con exacerbaciones vs. pacientes estables. (C) Fibrosis quística (CF) vs. HCTR.

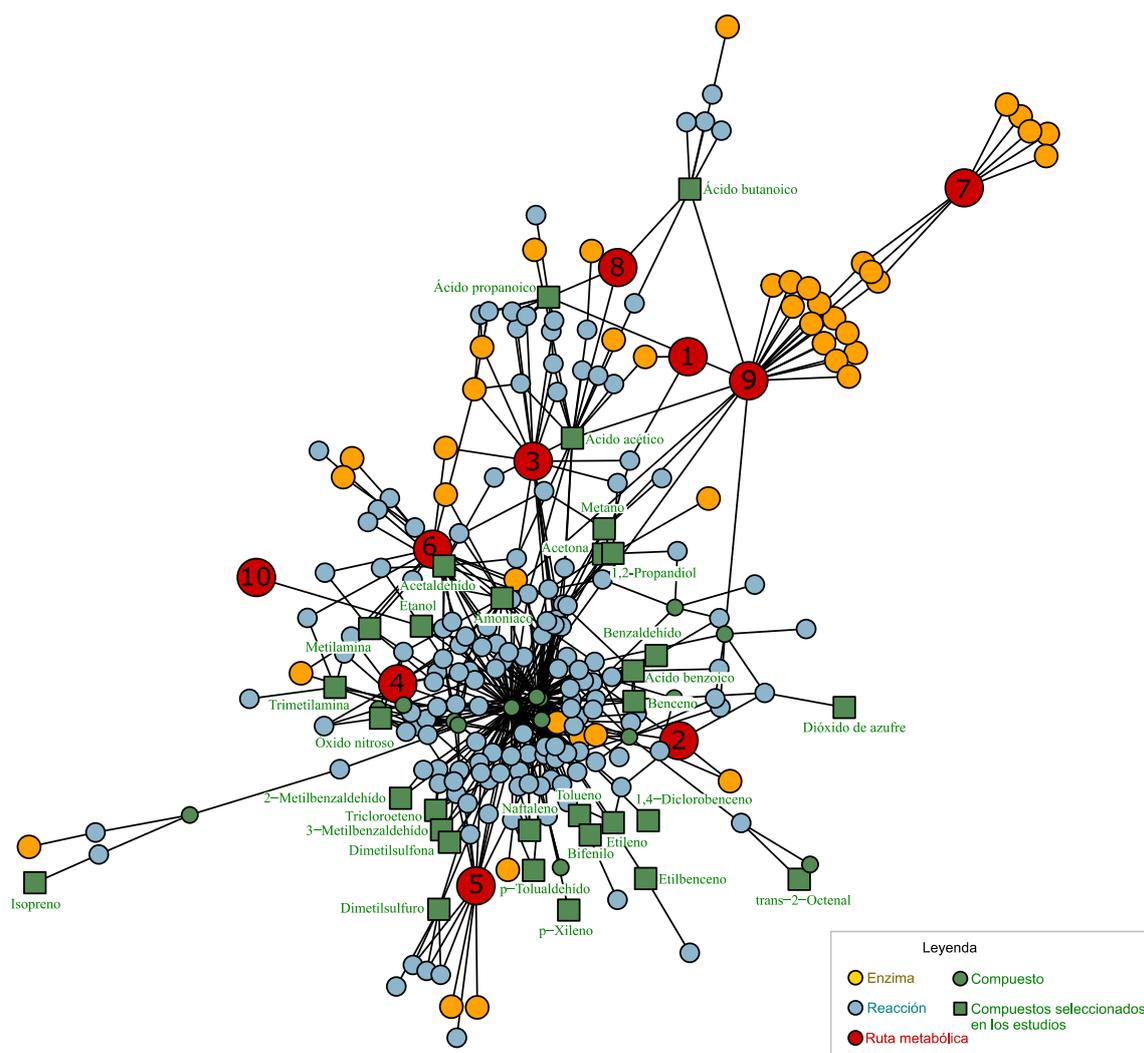


Figura 3. Subred derivada del análisis con *FELLA*⁷⁴ que muestra los nodos con mayor puntuación teniendo en cuenta los VOCs discriminantes señalados en los estudios seleccionados (Apéndice 1) que están disponibles en la base de datos KEGG como datos de entrada para el algoritmo de difusión. Las 10 rutas metabólicas más destacadas fueron las siguientes: 1) Síntesis y degradación de cuerpos cetónicos (KEGG ID: hsa00072), 2) Metabolismo de la cisteína y la metionina (KEGG ID: hsa00270), 3) Metabolismo del propionato (KEGG ID: hsa00640), 4) Metabolismo del nitrógeno (KEGG ID: hsa00910), 5) Metabolismo del azufre (KEGG ID: hsa00920), 6) Metabolismo del carbono (KEGG ID: hsa01200), 7) Sistema renina-angiotensina (KEGG ID: hsa04614), 8) Digestión y absorción de carbohidratos (KEGG ID: hsa04973), 9) Digestión y absorción de proteínas (KEGG ID: hsa04974), y 10) Alcoholismo (KEGG ID: hsa05034).

Evaluación de la calidad de los estudios

Los resultados de la evaluación de la calidad de los estudios seleccionados mediante el sistema de puntuación QUADAS-2⁷⁷ se resumen en la Figura 4. El análisis QUADAS-2 evidenció problemas en varios dominios, siendo la prueba índice y la selección de pacientes los dominios que presentaron el mayor riesgo de sesgo (Figura 4, panel A y B) y, además, existió una gran concordancia entre las evaluaciones de los dos revisores ($k = 0,81$ (IC 95% 0,75-0,88)) (Apéndice 2 y Apéndice 3). La fuente de

sesgo más importante estaba relacionada con el análisis de VOCs (prueba índice), ya que ninguno de los estudios seleccionados realizó una verdadera validación externa, utilizando nuevas cohortes independientes. Además, se consideró que los estudios que utilizaban matrices de sensores de gases tenían un alto riesgo de sesgo, porque la identificación de los VOCs no se puede llevar a cabo con estas tecnologías^{32,47,50,51}. Por otra parte, la selección de los pacientes también se consideró que podía tener un alto riesgo de sesgo en casi todos los casos, ya que la mayoría de los estudios tenían un diseño de tipo caso-control. Por lo tanto, en muchos estudios se excluyeron los casos que presentaban dificultades de diagnóstico, lo que puede conllevar a resultados demasiado optimistas y a una interpretación excesivamente favorable del rendimiento diagnóstico de los VOCs. No obstante, en los estudios de las cohortes ADEM^{36,37,78}, RASTER^{30,39}, EUROPA⁵⁰ y FLAME³⁸ y en el de Neerincx *et al.*⁴¹ se empleó una estrategia de selección de pacientes más realista. Por otro lado, Benedek *et al.*⁵¹, Schaber *et al.*⁴² y Alkhouri *et al.*⁴⁴ incluyeron como grupo control a pacientes con condiciones favorables para desarrollar la patología objetivo en lugar de utilizar controles sanos. Además, Monasta *et al.*⁴⁸ utilizaron tanto controles con síntomas gastrointestinales como controles sanos. Otra debilidad de los estudios es que se consideró que el dominio de flujo y tiempos poseía un alto riesgo de sesgo. Esto se debe a que en los diseños de tipo caso-control, los controles sanos no se sometían a la misma prueba de referencia que los casos. En cambio, el intervalo de tiempo entre la prueba de referencia y la prueba índice era adecuado debido a la propia naturaleza crónica de las enfermedades consideradas en muchos estudios. Por último, el dominio de la prueba de referencia no suscitó preocupación a excepción de dos estudios en los que los propios autores cuestionaron la metodología empleada para el diagnóstico de la enfermedad: Alkhouri *et al.*⁴⁴, que utilizaron una ecografía hepática convencional en lugar de una biopsia para el diagnóstico de NAFLD, y van de Kant *et al.*³⁶ que consideraron sólo dos episodios registrados en el cuestionario ISAAC (*International Study of Asthma and Allergy in Childhood*) para diagnosticar las sibilancias recurrentes. Para concluir, la preocupación por la aplicabilidad de los estudios seleccionados con respecto a la pregunta de nuestra revisión fue, en general, baja (Figura 4, panel C). Sólo hubo un estudio en el que la aplicabilidad relativa a la prueba de referencia no estaba clara, ya que abordaba las sibilancias recurrentes, que se definen como un síntoma y no como una enfermedad³⁶.

(A)

Estudio	RIESGO DE SESGO				PREOCUPACIÓN POR LA APLICABILIDAD		
	SELECCIÓN DE PACIENTES	PRUEBA ÍNDICE	PRUEBA DE REFERENCIA	FLUJO Y TIEMPOS	SELECCIÓN DE PACIENTES	PRUEBA ÍNDICE	PRUEBA DE REFERENCIA
Bannier <i>et al.</i> (2019)	☹️	☹️	😊	☹️	😊	😊	😊
Paff <i>et al.</i> (2013)	☹️	☹️	😊	☹️	😊	😊	😊
van der Schee <i>et al.</i> (2015)	☹️	☹️	😊	☹️	😊	😊	😊
Benedek <i>et al.</i> (2013)	😊	☹️	😊	😊	😊	😊	😊
Dallinga <i>et al.</i> (2010)	☹️	☹️	😊	☹️	😊	😊	😊
Caldeira <i>et al.</i> (2011)	☹️	☹️	😊	☹️	😊	😊	😊
Caldeira <i>et al.</i> (2012)	☹️	☹️	😊	☹️	😊	😊	😊
van de Kant <i>et al.</i> (2013)	☹️	☹️	☹️	☹️	😊	😊	☹️
Smolinska <i>et al.</i>	😊	☹️	😊	😊	😊	😊	😊
Gahleitner (2013)	☹️	☹️	😊	☹️	😊	😊	😊
Robroeks <i>et al.</i> (2013)	☹️	☹️	😊	😊	😊	😊	😊
van Vliet <i>et al.</i> (2016)	😊	☹️	😊	☹️	😊	😊	😊
van Vliet <i>et al.</i> (2017)	😊	☹️	😊	☹️	😊	😊	😊
Robroeks <i>et al.</i> (2010)	☹️	☹️	😊	☹️	😊	😊	😊
Neerincx <i>et al.</i> (2016)	😊	☹️	😊	😊	😊	😊	😊
Schaber <i>et al.</i> (2018)	☹️	☹️	😊	?	😊	😊	😊
Alkhoury <i>et al.</i> (2014)	😊	☹️	☹️	?	😊	😊	😊
Eng <i>et al.</i> (2015)	☹️	☹️	😊	☹️	😊	😊	😊
Obermeier <i>et al.</i> (2017)	☹️	☹️	😊	☹️	😊	😊	😊
Patel <i>et al.</i> (2014)	☹️	☹️	😊	☹️	😊	😊	😊
Monasta <i>et al.</i> (2017)	☹️	☹️	😊	☹️	😊	😊	😊
van Mastriigt <i>et al.</i> (2016)	☹️	☹️	😊	☹️	😊	😊	😊

😊 Riesgo bajo ☹️ Riesgo alto ? Riesgo incierto

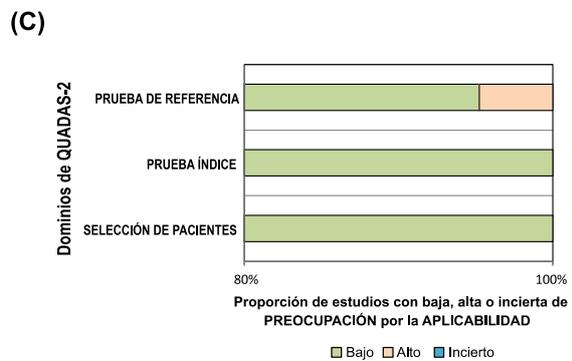
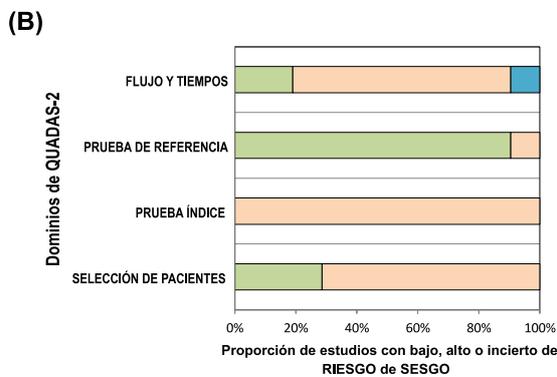


Figura 4. Evaluación de la calidad de los estudios incluidos en la síntesis cualitativa mediante el sistema de puntuación QUADAS-2⁷⁷. (A) Resumen de estudios por separado. (B) Representación del riesgo de sesgo en los diferentes ámbitos. (C) Representación de la preocupación por la aplicabilidad en los diferentes ámbitos.

DISCUSIÓN

Análisis de VOCs en la clínica pediátrica: ¿Se ha alcanzado ya esa meta?

A pesar de haber encontrado un importante conjunto de pruebas, se consideró que el riesgo de sesgo era alto, lo que demuestra que es poco probable que el análisis de los perfiles de VOCs en aire exhalado llegue a la práctica clínica en pediatría en su

estado actual. Todos los estudios incluidos en esta síntesis cualitativa permanecen estancados en una fase muy preliminar de descubrimiento preclínico, ya que la mayoría de ellos persigue el cribado no dirigido de VOCs al que puedan asociarse posteriormente diferentes estados de enfermedad. Esta fase de descubrimiento es muy anterior a la validación del método analítico y, finalmente, de cualquier validación clínica aceptada por la comunidad médica. Cabe destacar que para que una prueba basada en VOCs en el aire exhalado llegue a la práctica clínica, debe pasar por un proceso de tres fases que incluye el descubrimiento de biomarcadores, la validación de la prueba y la evaluación de la utilidad clínica^{79,80}. Es importante afrontar las fases de descubrimiento/exploración y de validación de la prueba, teniendo en cuenta los criterios de la lista de comprobación que deben abordarse para determinar la preparación de una prueba ómica para su uso en un ensayo clínico prospectivo⁸¹. La fase de descubrimiento, idealmente debería considerarse como la parte del proceso destinada a evaluar si la prueba basada en el análisis de VOCs tiene una posibilidad razonable de demostrar su validez clínica. Por lo tanto, la finalización completa de esta fase de descubrimiento implica los siguientes cuatro pasos: i) datos de controles de calidad; ii) desarrollo de modelos computacionales y validación interna; iii) validación externa; y iv) distribución de los datos, el código y los procedimientos computacionales en su totalidad a la comunidad científica⁸². Cabe mencionar que ninguno de los estudios de esta síntesis cualitativa incluye estos cuatro pasos. En consecuencia, a día de hoy, sus resultados difícilmente se pueden trasladar a la práctica clínica.

Análisis de VOCs en la clínica pediátrica: El camino hacia la clínica

El diseño de experimentos exploratorios, de acuerdo con los cuatro pasos obligatorios mencionados anteriormente, debería hacer avanzar la investigación sobre la traslación clínica del análisis de VOCs en el aire exhalado. En este sentido, la comunidad dedicada al análisis del aire exhalado debería inspirarse en la comunidad de la metabolómica, con la que comparte múltiples aspectos. La adopción de las prácticas habituales de la metabolómica en la fase de descubrimiento, garantizaría una transición fluida hacia la fase de validación de las pruebas, lo que allanaría el camino hacia una futura traslación clínica. En este sentido, la comunidad del aliento exhalado debería acordar en primer lugar un conjunto claro de directrices y normas mínimas de presentación de la información, como ha hecho hasta ahora la comunidad de la metabolómica^{70,82,83}. Esto repercutiría, por ejemplo, en la forma en que se informa de la identificación de los VOCs en la actualidad, y promovería la incorporación de los niveles de confianza en la identificación, aumentando la transparencia. Hay que señalar que

estos niveles no se han mencionado en ninguno de los estudios revisados. También es aconsejable incluir otros identificadores para los VOCs además de su nombre. Atenerse sólo a los nombres puede introducir ambigüedad, especialmente si se utilizan nombres triviales en lugar de sistemáticos. Por lo tanto, se recomienda informar de otros identificadores de VOCs además de su nombre. Tomemos como ejemplo el compuesto denominado como 1-penten-2-on por Dallinga *et al.*⁵², un estudio incluido en nuestra revisión sistemática, sin ninguna otra identificación, por lo que asignar este nombre a una estructura es imposible, ya que el sufijo -on no es indicativo de un grupo funcional. La suposición más sencilla es que podría tratarse de un error tipográfico, pero sin más información es sólo una suposición que se habría evitado si se hubieran incluido otros identificadores. A este respecto, los identificadores reproducibles para los VOCs, como el número de registro CAS (*Chemical Abstracts Service*), PubChem CID, InChI (*International Chemical Identifier*) o SMILES (*Simplified Molecular-Input Line-Entry System*), desempeñan un papel fundamental en el acceso a los datos de los compuestos y permiten garantizar las referencias cruzadas entre estudios y bases de datos como la HMDB⁸⁴.

A lo largo de un estudio de VOCs en el aire exhalado se debe comprobar regularmente la calidad de los datos medidos. Esto suele hacerse en metabolómica mediante la comprobación periódica del rendimiento analítico del instrumento, y el uso de muestras de controles de calidad que tengan en cuenta las fuentes de variabilidad no relacionadas con los fenómenos biológicos en estudio. Las muestras de controles de calidad deberían consistir, idealmente, en un conjunto de muestras que formen parte del estudio y que se midan regularmente a lo largo de la secuencia temporal de este. La preparación de estos conjuntos de muestras de controles de calidad es sencilla en metabolómica, pero podría resultar extremadamente complicada para el análisis de los VOCs en el aire exhalado. Una solución sería utilizar índices de retención basados en estándares externos (por ejemplo, n-alcanos) con un doble objetivo: dar cuenta de las desviaciones de los tiempos de retención, de acuerdo con su propósito original, y estimar la variabilidad del instrumento a lo largo de esta secuencia temporal.

Por otra parte, la calidad de los datos se ve afectada en gran medida por los procedimientos de preprocesamiento de datos de MS (ej.: conversión a modo centroide, eliminación de ruido, corrección de la línea de base, alineación, detección de picos, etc.), un paso que se pasa por alto en la mayoría de los estudios incluidos en nuestra síntesis cualitativa. Ni una verdadera validación externa, ni la transparencia del análisis de datos han sido contempladas en ninguno de los estudios revisados. La falta de validación

externa es una de las principales causas a la que se puede atribuir el hecho de que la investigación centrada en el análisis de VOCs en el aliento exhalado en la población pediátrica permanezca estancada en una fase exploratoria preliminar, ya que aumenta el riesgo de sobreajuste y los resultados deben interpretarse con cuidado, ya que es probable que se sobrestimen los verdaderos resultados. Ambos aspectos deben abordarse urgentemente si el análisis de los VOCs en el aire exhalado se destina a un posterior desarrollo clínico. Un análisis de datos reproducible y transparente implica que los datos y los metadatos estén disponibles a través de repositorios especializados de acceso público. Esto debería ser factible hoy en día utilizando, por ejemplo, Metabolomics Workbench o Metabolights⁸⁵, por nombrar sólo algunos de ellos. Es importante que los datos estén disponibles en formatos estándar abiertos y aceptados por la comunidad. En el caso de los datos obtenidos por espectrometría de masas, algunos ejemplos de formatos estándar serían mzXML o mzML⁸⁶. Asimismo, un análisis de datos transparente implica que los procedimientos computacionales reproducibles y su código asociado también estén disponibles. El uso de soluciones de código abierto basadas en lenguajes de programación (R, Python, Matlab) para el análisis de datos es una buena opción para garantizar la reproducibilidad del código informático. Además, alojar dicho código en plataformas destinadas al desarrollo colaborativo de *software*, como GitHub, permitirá un acceso totalmente viable al mismo.

El resultado de una fase de descubrimiento correctamente diseñada debería ser, idealmente, un panel de VOCs detectados en el aliento correctamente identificados, y señalados como potenciales biomarcadores putativos. El paso siguiente para convertirlos en biomarcadores clínicamente fiables es la fase de validación de la prueba. En esta fase de validación, se optimiza un método analítico dirigido para la detección del panel de VOCs exhalados seleccionados como biomarcadores⁸⁷, lo que permite fijar la mayoría de los aspectos analíticos de la determinación de VOCs en el aliento exhalado. A continuación, el rendimiento del ensayo de este método dirigido puede calcularse utilizando métricas analíticas establecidas, como la exactitud, la precisión, el coeficiente de variación, la sensibilidad, la especificidad, el rango lineal, el límite de detección y el límite de cuantificación, si procede. Es sólo en esta etapa donde las tan discutidas necesidades de estandarización⁷³ cobran verdadero sentido y pueden aplicarse de forma fiable. De hecho, existe un viejo debate en la comunidad dedicada al análisis del aire exhalado sobre la necesidad de estandarizar el muestreo y el análisis de VOCs en el aire exhalado⁷³. Durante mucho tiempo se ha culpado a esta falta de estandarización del hecho de que no se haya progresado del ámbito del laboratorio al contexto clínico de rutina. A pesar de que estamos de acuerdo en la necesidad de la

estandarización, consideramos que es necesario redefinirla y clarificarla. De hecho, intentar estandarizar la fase de descubrimiento de los estudios de VOCs del aliento podría resultar contraproducente y demasiado restrictivo. En esta fase, la composición del perfil de VOCs aún no se ha caracterizado y, por lo tanto, el aire exhalado es perfilado de manera químicamente no selectiva. Por lo tanto, se necesita un amplio margen de maniobra experimental y se suele utilizar una combinación de diferentes métodos de muestreo y análisis para garantizar la cobertura de los VOCs. Sin embargo, esto no exime de cumplir con las normas mínimas de notificación relativas a las condiciones experimentales, el análisis de datos y la identificación de VOCs en el aire exhalado. De hecho, estas normas de presentación de la información son fundamentales para la validación posterior del método y para una posible validación clínica de la determinación de VOCs en el aliento.

Limitaciones

Nuestra estrategia de búsqueda fue intencionadamente amplia e inclusiva para captar todos los estudios que utilizaran el análisis de VOCs exhalados en población pediátrica, independientemente del diseño del estudio, la tecnología de medición y las condiciones de la enfermedad. Esta estrategia podría influir negativamente en el riesgo de sesgo en algunos dominios de QUADAS-2. La realización de un meta-análisis se consideró inviable debido a la elevada heterogeneidad de los estudios incluidos en la síntesis cualitativa. Esta heterogeneidad se debió a: i) la consideración deliberada de enfermedades totalmente diferentes; ii) la definición desigual de los grupos control en los distintos estudios (controles sanos, pacientes estables, sujetos que presentaban determinados síntomas, etc.); iii) el uso de diferentes tecnologías (matrices de sensores de gases, GC-MS); iv) un panel de VOCs poco consistente en los distintos estudios incluidos en la síntesis cualitativa. Además, el número de estudios publicados que incorporan el análisis de VOCs en el aire exhalado en la población pediátrica es relativamente bajo, lo que limita la comparación en el caso de algunas enfermedades.

CONCLUSIÓN

Presentamos la primera revisión sistemática sobre el uso del análisis de VOCs exhalados en pediatría clínica. Concluimos que la tan anunciada potencialidad clínica del análisis del volátiloma en el aire exhalado en pediatría clínica está aún por demostrar formalmente. No obstante, nuestra síntesis descriptiva muestra un número considerable de estudios que han empleado el análisis del volátiloma en la población pediátrica durante los últimos 10 años. A pesar de ello, las conclusiones sobre la potencialidad

clínica de los VOCs en el aliento siguen siendo poco claras. Esto no sólo es atribuible a la heterogeneidad de los estudios, que dificulta la agrupación de datos para realizar un meta-análisis adecuado, sino también al hecho de que todos los estudios revisados permanecen estancados en una fase exploratoria muy preliminar, sin que haya ningún avance posterior. En esta fase, la determinación de VOCs exhalados está lejos de ser una herramienta de diagnóstico consolidada y, por lo tanto, afirmar su potencial clínico es demasiado precipitado. Nuestro punto de vista es que la mejora de los diseños experimentales de estos estudios exploratorios adoptando las prácticas más recomendables de la metabolómica, tales como la comunicación de los niveles de confianza seguidos para la identificación de los VOCs, la divulgación del código y de los datos experimentales a través de repositorios públicos, debería ayudar a impulsar la investigación de los VOCs en el aliento exhalado, allanando el camino para una futura traducción clínica.

MATERIALES Y MÉTODOS

Protocolo y criterios de elegibilidad

Esta revisión sistemática se llevó a cabo de acuerdo con las directrices de la declaración PRISMA⁵³. El protocolo de este trabajo fue enviado el 17 de septiembre de 2019 al registro prospectivo internacional de revisiones sistemáticas PROSPERO y fue registrada el 28 de abril de 2020 (CRD42020151186). La pregunta principal de la revisión tenía como objetivo aclarar si actualmente es factible el uso del análisis de VOCs en el aire exhalado para el diagnóstico y/o la monitorización terapéutica de enfermedades en niños y adolescentes. Únicamente se consideraron los artículos originales publicados en inglés antes de septiembre de 2019. Los estudios fueron considerados elegibles si medían los VOCs exhalados con el fin de discriminar la presencia de una enfermedad o de monitorizar patologías, y si la edad media (o mediana) de la población del estudio era inferior a 18 años. Las publicaciones duplicadas o redundantes fueron excluidas.

Estrategia de búsqueda

Las bases de datos donde se llevó a cabo la búsqueda fueron Web of Science, SCOPUS y PubMed. La estrategia de búsqueda en PubMed consistió en la combinación de las siguientes palabras clave: (("exhaled") AND ("VOC" OR "VOCs" OR "volatile organic compound" OR "volatile organic compounds" OR "breathomics" OR "volatilome" OR "exhaled metabolites")) AND ("children" OR "childhood" OR "baby" OR "babies" OR

"teenager" OR "teenagers" OR "adolescent" OR "adolescents" OR "child" OR "infant")) AND ("disease" OR "diseases" OR "pathology" OR "pathologies" OR "patient" OR "patients" OR "pathophysiology" OR "diagnosis" OR "diagnostic" OR "biomarker" OR "biomarkers" OR "marker" OR "markers" OR "markers" OR "monitoring" OR "monitorization"). En el caso de las bases de datos Web of Science y SCOPUS, aplicó esta estrategia de búsqueda, pero con pequeños ajustes (Apéndice 4).

Selección de estudios

La selección de los estudios se realizó según el diagrama de flujo PRISMA⁵³ y fue realizada de forma independiente por dos revisores (Rosa Alba Sola Martínez y José María Pastor Hernández). Los desacuerdos se resolvieron por consenso.

Extracción de datos

Se extrajeron sistemáticamente los siguientes datos de cada estudio: i) diseño del estudio (seguimiento de la enfermedad y/o presencia de la enfermedad); ii) población del estudio (patología, tamaño de la muestra y edad); iii) recogida de aire exhalado (tipo de contenedor de aire exhalado, porción de aire exhalado muestreado y estrategias utilizadas para minimizar la influencia de los VOCs ambientales); iv) tipo de plataforma analítica; v) análisis estadístico; vi) resultados (en términos de número de VOCs seleccionados, y resultados de la tasa de clasificación, sensibilidad y especificidad); vii) VOCs identificados que fueron señalados como relevantes para el diagnóstico y/o la monitorización de las enfermedades.

Síntesis

La realización de un meta-análisis se consideró inviable debido a la considerable heterogeneidad entre las poblaciones de estudio y los perfiles de VOCs. En su lugar, se realizó una síntesis descriptiva y narrativa de las evidencias. Esto incluyó una descripción exhaustiva de los VOCs seleccionados por los estudios y un análisis de enriquecimiento funcional mediante el paquete de R *FELLA*⁷⁴.

Evaluación de la calidad

Dos revisores independientes (Rosa Alba Sola Martínez y José María Pastor Hernández) evaluaron la calidad de los estudios incluidos mediante la herramienta QUADAS-2 (*Quality Assessment of Diagnostic Accuracy Studies 2*). QUADAS-2 incluye 4 dominios (selección de pacientes, prueba índice, prueba de referencia, y flujo y

tiempos), y se lleva a cabo en cuatro fases: 1) definición de la pregunta de la revisión, 2) adaptación las preguntas orientadoras a las características de la revisión, 3) diseño de un diagrama de flujo, y 4) evaluación de los sesgos y la aplicabilidad. Para la evaluación del riesgo de sesgo, los revisores leyeron cuidadosamente todos los artículos seleccionados y respondieron a una serie de preguntas orientadoras con “sí”, “no” o “incierto”. Dependiendo de las respuestas a las preguntas orientadoras, el riesgo de sesgo se calificó como “alto”, “bajo” o “incierto”. Las preguntas orientadoras referidas al Dominio 2 (Prueba de Índice) fueron modificadas durante la Fase 2 (Adaptación específica a la revisión) respecto a la guía original de QUADAS-2. Además, sólo las dos primeras preguntas orientadoras fueron tenidas en cuenta, el Dominio 1 (Selección de pacientes), ya que fueron consideradas las más adecuadas para responder a la pregunta principal de la revisión. El resto de los dominios se mantuvieron tal como se describen en la guía original de QUADAS-2, conservando las mismas preguntas orientadoras. El Apéndice 5 resume las preguntas orientadoras utilizadas en esta revisión. La preocupación por la aplicabilidad de los diferentes dominios y se calificó como “baja”, “alta” o “incierto”⁷⁷. Se utilizó el coeficiente kappa de Cohen calculado con el paquete *psych* para comprobar el nivel de concordancia en las respuestas de ambos revisores. En el caso de que los dos revisores aportaran respuestas diferentes, la decisión final fue tomada por consenso.

BIBLIOGRAFÍA

1. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
2. Blanchet, L. *et al.* Factors that influence the volatile organic compound content in human breath. *J. Breath Res.* **11**, 016013 (2017).
3. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. . Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
4. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
5. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
6. Fens, N. *et al.* Exhaled Breath Profiling Enables Discrimination of Chronic Obstructive Pulmonary Disease and Asthma. *Am. J. Respir. Crit. Care Med.* **180**, 1076–1082 (2009).
7. Barker, M. *et al.* Volatile organic compounds in the exhaled breath of young patients with cystic fibrosis. *Eur. Respir. J.* **27**, 929–936 (2006).
8. Bennett, L. *et al.* A chemometric study on human breath mass spectra for biomarker identification in cystic fibrosis. *J. Breath Res.* **3**, 046002 (2009).
9. Poli, D. *et al.* Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respir. Res.* **6**, 71 (2005).
10. Ulanowska, A., Kowalkowski, T., Trawińska, E. & Buszewski, B. The application of statistical methods using VOCs to identify patients with lung cancer. *J. Breath Res.* **5**, 046008 (2011).
11. Altomare, D. F. *et al.* Exhaled volatile organic compounds identify patients with colorectal cancer. *Br. J. Surg.* **100**, 144–150 (2013).
12. Tong, H. *et al.* Volatile organic metabolites identify patients with gastric carcinoma, gastric ulcer, or gastritis and control patients. *Cancer Cell Int.* **17**, 108 (2017).
13. Guo, L. *et al.* Exhaled breath volatile biomarker analysis for thyroid cancer. *Transl. Res.* **166**, 188–195 (2015).

14. Beccaria, M. *et al.* Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *J. Breath Res.* **13**, 016005 (2018).
15. Dadamio, J. *et al.* Breath biomarkers of liver cirrhosis. *J. Chromatogr. B* **905**, 17–22 (2012).
16. Greiter, M. B. *et al.* Differences in Exhaled Gas Profiles Between Patients with Type 2 Diabetes and Healthy Controls. *Diabetes Technol. Ther.* **12**, 455–463 (2010).
17. Ibrahim, B. *et al.* Non-invasive phenotyping using exhaled volatile organic compounds in asthma. *Thorax* **66**, 804–809 (2011).
18. Dragonieri, S. *et al.* An electronic nose in the discrimination of patients with asthma and controls. *J. Allergy Clin. Immunol.* **120**, 856–862 (2007).
19. Pizzini, A. *et al.* Analysis of volatile organic compounds in the breath of patients with stable or acute exacerbation of chronic obstructive pulmonary disease. *J. Breath Res.* **12**, 036002 (2018).
20. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
21. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
22. Moschino, L., Zanconato, S., Bozzetto, S., Baraldi, E. & Carraro, S. Childhood asthma biomarkers: present knowledge and future steps. *Paediatr. Respir. Rev.* **16**, 205–212 (2015).
23. Cavaleiro Rufo, J. *et al.* Exhaled breath condensate volatilome allows sensitive diagnosis of persistent asthma. *Allergy Eur. J. Allergy Clin. Immunol.* **74**, 527–534 (2019).
24. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).
25. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res.* **6**, 027108 (2012).
26. Horváth, I. *et al.* A European Respiratory Society technical standard: exhaled biomarkers in lung disease. *Eur. Respir. J.* **49**, 1600965 (2017).

27. van Mastrigt, E., de Jongste, J. C. & Pijnenburg, M. W. The analysis of volatile organic compounds in exhaled breath and biomarkers in exhaled breath condensate in children – clinical tools or scientific toys? *Clin. Exp. Allergy* **45**, 1170–1188 (2015).
28. Fuglsbjerg, M. G. *et al.* Limited clinical value of exhaled volatile organic compound measurements in childhood asthma. *ERJ Open Res.* **4**, 00026–02018 (2018).
29. Neerincx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
30. van Vliet, D. *et al.* Can exhaled volatile organic compounds predict asthma exacerbations in children? *J. Breath Res.* **11**, 016016 (2017).
31. Klaassen, E. M. M. *et al.* Exhaled Biomarkers and Gene Expression at Preschool Age Improve Asthma Prediction at 6 Years of Age. *Am. J. Respir. Crit. Care Med.* **191**, 201–207 (2015).
32. Bannier, M. A. G. E., van de Kant, K. D. G., Jöbssis, Q. & Dompeling, E. Feasibility and diagnostic accuracy of an electronic nose in children with asthma and cystic fibrosis. *J. Breath Res.* **13**, 036009 (2019).
33. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–80 (2011).
34. Caldeira, M. *et al.* Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **1254**, 87–97 (2012).
35. Gahleitner, F., Guallar-Hoyas, C., Beardsmore, C. S., Pandya, H. C. & Thomas, C. P. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis* **5**, 2239–2247 (2013).
36. van de Kant, K. D. G. *et al.* Exhaled breath profiling in diagnosing wheezy preschool children. *Eur. Respir. J.* **41**, 183–188 (2013).
37. Smolinska, A. *et al.* Profiling of Volatile Organic Compounds in Exhaled Breath As a Strategy to Find Early Predictive Signatures of Asthma in Children. *PLoS One* **9**, e95668 (2014).
38. Robroeks, C. M. *et al.* Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study. *Eur. Respir. J.* **42**, 98–106 (2013).
39. Van Vliet, D. *et al.* Association between exhaled inflammatory markers and

- asthma control in children. *J. Breath Res.* **10**, 016014 (2016).
40. Robroeks, C. M. H. H. T. *et al.* Metabolomics of Volatile Organic Compounds in Cystic Fibrosis Patients and Controls. *Pediatr. Res.* **68**, 75–80 (2010).
 41. Neerincx, A. H. *et al.* Detection of *Staphylococcus aureus* in cystic fibrosis patients using breath VOC profiles. *J. Breath Res.* **10**, 046014 (2016).
 42. Schaber, C. L. *et al.* Breathprinting Reveals Malaria-Associated Biomarkers and Mosquito Attractants. *J. Infect. Dis.* **217**, 1553–1560 (2018).
 43. Obermeier, J. *et al.* Exhaled volatile substances mirror clinical conditions in pediatric chronic kidney disease. *PLoS One* **12**, e0178745 (2017).
 44. Alkhouri, N. *et al.* Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children. *Eur. J. Gastroenterol. Hepatol.* **26**, 82–7 (2014).
 45. Eng, K. *et al.* Analysis of breath volatile organic compounds in children with chronic liver disease compared to healthy controls. *J. Breath Res.* **9**, 026002 (2015).
 46. Patel, N. *et al.* Metabolomic analysis of breath volatile organic compounds reveals unique breathprints in children with inflammatory bowel disease: a pilot study. *Aliment. Pharmacol. Ther.* **40**, 498–507 (2014).
 47. Paff, T. *et al.* Exhaled molecular profiles in the assessment of cystic fibrosis and primary ciliary dyskinesia. *J. Cyst. Fibros.* **12**, 454–60 (2013).
 48. Monasta, L. *et al.* Inflammatory bowel disease and patterns of volatile organic compounds in the exhaled breath of children: A case-control study using Ion Molecule Reaction-Mass Spectrometry. *PLoS One* **12**, e0184118 (2017).
 49. van Mastrigt, E. *et al.* Exhaled breath profiling using broadband quantum cascade laser-based spectroscopy in healthy children and children with asthma and cystic fibrosis. *J. Breath Res.* **10**, 026003 (2016).
 50. Van Der Schee, M. P. *et al.* Altered exhaled biomarker profiles in children during and after rhinovirus-induced wheeze. *Eur. Respir. J.* **45**, 440–448 (2015).
 51. Benedek, P. *et al.* Exhaled biomarker pattern is altered in children with obstructive sleep apnoea syndrome. *Int. J. Pediatr. Otorhinolaryngol.* **77**, 1244–7 (2013).
 52. Dallinga, J. W. *et al.* Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clin. Exp. Allergy* **40**, 68–76 (2009).
 53. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & PRISMA Group. Preferred

- Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* **6**, e1000097 (2009).
54. Lawal, O., Ahmed, W. M., Nijssen, T. M. E., Goodacre, R. & Fowler, S. J. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics* **13**, 110 (2017).
 55. Farraia, M. V. *et al.* The electronic nose technology in clinical diagnosis. *Porto Biomed. J.* **4**, e42 (2019).
 56. de Heer, K. *et al.* Electronic Nose Technology for Detection of Invasive Pulmonary Aspergillosis in Prolonged Chemotherapy-Induced Neutropenia: a Proof-of-Principle Study. *J. Clin. Microbiol.* **51**, 1490–1495 (2013).
 57. Hauschild, A.-C. *et al.* Computational Methods for Metabolomic Data Analysis of Ion Mobility Spectrometry Data—Reviewing the State of the Art. *Metabolites* **2**, 733–755 (2012).
 58. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
 59. Prince, B. J., Milligan, D. B. & McEwan, M. J. Application of selected ion flow tube mass spectrometry to real-time atmospheric monitoring. *Rapid Commun. Mass Spectrom.* **24**, 1763–9 (2010).
 60. Bregy, L. *et al.* Real-time mass spectrometric identification of metabolites characteristic of chronic obstructive pulmonary disease in exhaled breath. *Clin. Mass Spectrom.* **7**, 29–35 (2018).
 61. Alkhouri, N. *et al.* Breathprints of childhood obesity: changes in volatile organic compounds in obese children compared with lean controls. *Pediatr. Obes.* **10**, 23–29 (2015).
 62. Netzer, M. *et al.* A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry. *Bioinformatics* **25**, 941–947 (2009).
 63. Singh, K. D. *et al.* Translating secondary electrospray ionization-high-resolution mass spectrometry to the clinical environment. *J. Breath Res.* **12**, 027113 (2018).
 64. Singh, K. D. *et al.* Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **411**, 4883–4898 (2019).
 65. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

66. Amato, F. *et al.* Artificial neural networks in medical diagnosis. *J. Appl. Biomed.* **11**, 47–58 (2013).
67. Lê Cao, K.-A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253 (2011).
68. Smolinska, A. *et al.* Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J. Breath Res.* **8**, 027105 (2014).
69. Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D. & Lu, L. J. Computational and statistical analysis of metabolomics data. *Metabolomics* **11**, 1492–1513 (2015).
70. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221 (2007).
71. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
72. Boots, A. W., Bos, L. D., van der Schee, M. P., van Schooten, F. J. & Sterk, P. J. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. *Trends Mol. Med.* **21**, 633–644 (2015).
73. Beauchamp, J. D. & Pleil, J. D. Simply breath-taking? Developing a strategy for consistent breath sampling. *J. Breath Res.* **7**, 042001 (2013).
74. Picart-Armada, S., Fernández-Albert, F., Vinaixa, M., Yanes, O. & Perera-Lluna, A. FELLA: an R package to enrich metabolomics data. *BMC Bioinformatics* **19**, 538 (2018).
75. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **27**, 29–34 (1999).
76. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
77. Whiting, P. F. *et al.* QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* **155**, 529 (2011).
78. van de Kant, K. D. *et al.* Early diagnosis of asthma in young children by using non-invasive biomarkers of airway inflammation and early lung function measurements: study protocol of a case-control study. *BMC Public Health* **9**, 210 (2009).

79. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy & Institute of Medicine. In *Evolution of Translational Omics: Lessons Learned and the Path Forward* (eds. Micheel, C. M., Nass, S. J. & Omenn, G.S.) (National Academies Press (US), Washington (DC), 2012)
80. Mazzone, P. J. *et al.* Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American Thoracic Society Policy Statement. *Am. J. Respir. Crit. Care Med.* **196**, e15–e29 (2017).
81. McShane, L. M. *et al.* Criteria for the use of omics-based predictors in clinical trials. *Nature* **502**, 317–320 (2013).
82. MSI Board Members *et al.* The Metabolomics Standards Initiative. *Nat. Biotechnol.* **25**, 846–848 (2007).
83. Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241 (2007).
84. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
85. Haug, K. *et al.* MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz1019
86. Martens, L. *et al.* mzML--a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133 (2011).
87. Hanna, G. B., Boshier, P. R., Markar, S. R. & Romano, A. Accuracy and Methodologic Challenges of Volatile Organic Compound–Based Exhaled Breath Tests for Cancer Diagnosis: A Systematic Review and Meta-analysis. *JAMA Oncol.* **5**, (2019).

APÉNDICE 1. Compuestos orgánicos volátiles (VOCs) que fueron seleccionados por los estudios.

PubChem / KEGG	Compuestos	Enfermedades respiratorias												Enfermedades infecciosas		Enfermedades metabólicas								
		ASMA				ASMA ALÉRGICA		RW	EXACERBACIONES ASMÁTICAS				CF		S. aureus EN CF	MALARIA	NAFLD	CLD	CKD	IBD				
		52 ^A	37 ^{A-B}	35 ^A	49 ^A	33 ^A	34 ^A	36 ^A	38 ^A	38 ^B	39 ^A	30 ^A	49 ^B	40 ^A	41 ^A	42 ^A	44 ^A	45 ^A	43 ^A	46 ^A	48 ^A	48 ^B	48 ^C	48 ^D
	Alcanos																							
297 / C01438	metano																							
6324 / _	etano																							
8003 / _	pentano																							
8078 / C11249	ciclohexano								#															
7892 / _	2-metilpentano		+																					
7282 / _	3-metilpentano									#														
8900 / _	n-heptano																							
11582 / _	2-metilhexano		+																					
11507 / _	3-metilhexano																							
356 / C01387	octano		+																					
11551 / _	2,2-dimetilhexano						+																	
11416 / _	1,2-dimetilciclohexano												#	#										
7907 / _	2,4-dimetilpentano		+																					
8141 / _	nonano									+														
_ / _	trimetilhexano																							
15505 / _	propilciclohexano																							
16656 / _	2,4-dimetilheptano		+				-																	
520991 / _	3,3-dimetilheptano						+																	
16665 / _	4-metiloctano						-																	
15600 / _	decano									+														
92978 / _	2,4-dimetiloctano						+																	
140213 / _	3-etil-3-metilheptano						-																	
26839 / _	2,2,4-trimetilheptano		-																					
14257 / _	undecano						+																	
112465 / _	2,3,6-trimetiloctano		-											+										
8182 / C08374	dodecano						+			+														
_ / _	hidrocarburo saturado C12												+											
_ / _	metilundecano																							
_ / _	dimetildecano																							
86544 / _	2,3-dimetildecano						+																	
519395 / _	3,6-dimetildecano						+			+														
26058 / _	2,2,4,6,6-pentametilheptano									+														
12388 / C13834	tridecano						+																	
519388 / _	3,9-dimetilundecano						+																	
537330 / _	2,3,6-trimetildecano						+																	
537331 / _	2,3,7-trimetildecano						+																	
537327 / _	2,4,6-trimetildecano						+																	
_ / _	hidrocarburo ramificado (C ₁₃ H ₂₈)		+																					
_ / _	hidrocarburo ramificado (C ₁₃ H ₂₈)		+																					
_ / _	hidrocarburo saturado C13																							
12389 / _	tetradecano						+			+														
_ / _	(ramificado) hidrocarburo (C ₁₄ H ₃₀)																							
12391 / C08388	pentadecano						+																	
15269 / _	2-metiltridecano						+																	
19773 / _	2,6,10-trimetildodecano		-																					
11006 / _	hexadecano						-																	
15267 / _	2-metilpentadecano						+																	
	Compuestos Halogenados																							
6344 / C02271	cloruro de metileno/diclorometano						#						#											
6575 / C06790	1,1,2-tricloroetileno						#																	
31373 / _	tetracloroetileno																							

APÉNDICE 1 (Continuación I). Compuestos orgánicos volátiles (VOCs) que fueron seleccionados por los estudios.

PubChem / KEGG	Compuestos	Enfermedades respiratorias												Enfermedades infecciosas		Enfermedades metabólicas									
		ASMA				ASMA ALÉRGICA		RW	EXACERBACIONES ASMÁTICAS				CF		S. aureus EN CF	MALARIA	NAFLD	CLD	CKD	IBD					
		52 ^A	37 ^{A-B}	35 ^A	49 ^A	33 ^A	34 ^A	36 ^A	38 ^A	38 ^B	39 ^A	30 ^A	49 ^B	40 ^A	41 ^A	42 ^A	44 ^A	45 ^A	43 ^A	46 ^A	48 ^A	48 ^B	48 ^C	48 ^D	
	Nitrocompuestos																								
14285 / _	nitrociclohexano							+																	
	Nitrilos																								
6342 / _	Acetonitrilo																						-	-	-
	Hidrocarburos insaturados																								
6325 / C06547 8252 / C11505 11240 / _ 11239 / _ 11587 / _ 638701 / _ 18850 / _ 11610 / _ 8125 / _ 5364448 o 32935 / _ 137924 / _ 7868 / _ 5367588 / _ 31285 / _ 5364590 / _ 123385 / _ 123385 / _ 13381 / _ 518719 / _ _ / _ _ / _ 69425 / _ _ / _	etileno propeno 2-metil-1-buteno 3-metil-1-buteno 1,4-pentadieno 2,4-hexadieno 2-etil-1,3-butadieno 1-hepteno 1-octeno hidrocarburo C ₈ H ₁₆ (2-octeno, 3-octeno) 3,3-dimetilhex-1-eno 2,4,4-trimetil-1-penteno 2,4-octadieno 1-noneno (E)-2-noneno 2,4-dimetil-1-hepteno 2,4-dimetil-1-hepteno \$ 1-deceno 4-metil-1-deceno hidrocarburo insaturado (C ₁₅ H ₂₆) hidrocarburo poliinsaturado C16 octadecino 4,6,9-nonadecatrieno																								
	Isopreno e isoprenoides																								
6557 / C16521 _ / _ _ / _ 22311 / C06078 26049 / C09839 6554 / C09880	isopreno monoterpeno desconocido C ₁₀ H ₁₆ monoterpeno desconocido C ₁₀ H ₁₆ DL-Limoneno 3-careno alfa-pineno						+																		
	Aminas																								
6329 / C00218 1146 / C00565 8471 / C14691 _ / _	metilamina trimetilamina trietilamina N-metil-2-metilpropilamina																								

APÉNDICE 1 (Continuación II). Compuestos orgánicos volátiles (VOCs) que fueron seleccionados por los estudios.

PubChem / KEGG	Compuestos	Enfermedades respiratorias												Enfermedades infecciosas		Enfermedades metabólicas								
		ASMA				ASMA ALÉRGICA		RW	EXACERBACIONES ASMÁTICAS				CF		S. aureus EN CF	MALARIA	NAFLD	CLD	CKD	IBD				
		52 ^A	37 ^{A-B}	35 ^A	49 ^A	33 ^A	34 ^A	36 ^A	38 ^A	38 ^B	39 ^A	30 ^A	49 ^B	40 ^A	41 ^A	42 ^A	44 ^A	45 ^A	43 ^A	46 ^A	48 ^A	48 ^B	48 ^C	48 ^D
	Compuestos sulfúricos																							
1119 / C09306 6348 / C19033 1068 / C00580 6213 / C11142 19754 / _	dióxido de azufre disulfuro de carbono dimetilsulfuro dimetilsulfona 1-(metilsulfanil)propano	+								#														
	Compuestos nitrogenados																							
222 / C00014 24539 / _ 948 / C00887	amoníaco ácido nitroso óxido nitroso																		+					
	Alcoholes y éteres																							
702 / C00469 8254 / C11144 25484 / _ 7858 / C02001 1030 / C00583 637922 / _ 7821 / _ 29060 / _ 85779 / _	etanol dimetil éter (2-metoximetiletoxi)propanol 2-propen-1-ol 1,2-propanediol 2-buten-1-ol 2-etil-4-metil-1-pentanol 2-octen-1-ol hexadecan-2-ol				# #										+									
	Cetonas																							
180 / C00207 6569 / C02845 179 / C00466 _ / _ 13572 / _ 19633 / _	acetona 2-butanona 3-hydroxi-2-butanona 1-penten-2-on 6-metil-2-heptanona 6,10-dimetil-5,9-undecadien-2-ona		-		#	-																		
	Aldehídos																							
177 / C00084 8063 / _ 6184 / _ 591654 / _ 31241 / _ 454 / C01545 25610 / C21138 31289 / _ 8175 / C12307 62447 / _ 8194 / C02278	acetaldehído pentanal hexanal 4-metil-2-pentenal 2-etilhexanal octanal 2-octenal nonanal decanal 2-undecenal dodecanal																							

APÉNDICE 1 (Continuación IV). Compuestos orgánicos volátiles (VOCs) que fueron seleccionados por los estudios.

PubChem / KEGG	Compuestos	Enfermedades respiratorias												Enfermedades infecciosas		Enfermedades metabólicas								
		ASMA				ASMA ALÉRGICA		RW	EXACERBACIONES ASMÁTICAS				CF		S. aureus EN CF	MALARIA	NAFLD	CLD	CKD	IBD				
		52 ^A	37 ^{A-B}	35 ^A	49 ^A	33 ^A	34 ^A	36 ^A	38 ^A	38 ^B	39 ^A	30 ^A	49 ^B	40 ^A	41 ^A	42 ^A	44 ^A	45 ^A	43 ^A	46 ^A	48 ^A	48 ^B	48 ^C	48 ^D
	Desconocidos																							
-/-	M67																							
-/-	M69																							
-/-	M70																							
-/-	M74																							
-/-	M75																							
-/-	M76																							
-/-	M77																							
-/-	M79																							
-/-	M81																							
-/-	M89																							
-/-	M90																							
-/-	M91																							
-/-	M99																							
-/-	M105																							
-/-	M107																							
-/-	M114																							
-/-	M115																							
-/-	M118																							

+ indica que los niveles de COV están asociados a esa patología. - indica que los niveles de VOCs están asociados con el grupo control. # representa que los VOCs son discriminantes, pero no se indica si la asociación con la enfermedad es positiva o negativa. Las letras rojas indican diferencias estadísticamente significativas en los niveles de VOCs entre los dos grupos. La letra junto a la referencia indica el grupo de pacientes analizado en cada estudio.

APÉNDICE 2. QUADAS-2 Revisor 1 (Rosa Alba Sola Martínez).

Referencia	A. RIESGO de SESGO (Sí (S) / N (N) / Incierto (I))											B. PREOCUPACIÓN por la APLICABILIDAD (Alta (A) / Baja (B) / Incierta (I))		
	D1: SELECCIÓN DE PACIENTES		D2: PRUEBA ÍNDICE			D3: PRUEBA DE REFERENCIA		D4: FLUJO Y TIEMPOS				D1: SELECCIÓN DE PACIENTES	D2: PRUEBA ÍNDICE	D3: PRUEBA DE REFERENCIA
	PO1	PO2	PO1	PO2	PO3	PO1	PO2	PO1	PO2	PO3	PO4			
32	I	N	N	N	N	S	S	S	N	N	N	B	B	B
47	S	N	N	N	N	S	S	I	N	N	S	B	B	B
50	S	N	S	N	N	S	S	S	S	S	N	B	B	B
51	S	S	N	N	N	S	S	S	S	S	S	B	B	B
52	I	N	N	N	S	S	S	S	N	S	S	B	B	B
33	I	N	N	N	S	S	S	S	N	S	S	B	B	B
34	I	N	N	N	S	S	S	S	N	S	S	B	B	B
36	S	N	N	N	S	N	S	S	N	I	S	B	B	A
37	S	S	S	N	S	S	S	S	S	S	S	B	B	B
35	I	N	N	N	S	S	S	S	S	S	N	B	B	B
38	N	S	N	N	S	S	S	S	S	S	S	B	B	B
39	S	S	N	N	S	S	S	S	S	S	N	B	B	B
30	S	S	N	N	S	S	S	S	S	S	N	B	B	B
40	I	N	N	N	S	S	S	I	N	S	S	B	B	B
41	S	S	N	N	S	S	S	S	S	S	S	B	B	B
42	S	N	N	N	S	S	S	I	S	S	S	B	B	B
44	I	S	S	N	S	N	S	S	S	S	I	B	B	B
45	I	N	N	N	S	S	S	S	N	N	S	B	B	B
43	I	N	N	N	S	S	S	S	N	S	S	B	B	B
46	S	N	N	N	S	S	S	S	N	S	S	B	B	B
48	S	N	N	N	S	S	S	I	N	S	S	B	B	B
49	S	N	N	N	S	S	S	I	S	N	S	B	B	B

PO: Pregunta orientadora.

APÉNDICE 3. QUADAS-2 Revisor 2 (José María Pastor Hernández).

Referencia	A. RIESGO de SESGO (Sí (S) / N (N) / Incierto (I))											B. PREOCUPACIÓN por la APLICABILIDAD (Alta (A) / Baja (B) / Incierta (I))		
	D1: SELECCIÓN DE PACIENTES		D2: PRUEBA ÍNDICE			D3: PRUEBA DE REFERENCIA		D4: FLUJO Y TIEMPOS				D1: SELECCIÓN DE PACIENTES	D2: PRUEBA ÍNDICE	D3: PRUEBA DE REFERENCIA
	PO1	PO2	PO1	PO2	PO3	PO1	PO2	PO1	PO2	PO3	PO4			
32	I	N	N	N	N	S	S	S	N	N	I	B	B	B
47	I	N	N	N	N	S	S	I	N	S	S	B	B	B
50	N	N	S	N	N	S	S	S	S	S	N	B	B	B
51	S	S	N	N	N	S	S	I	S	S	S	B	B	B
52	I	N	N	N	S	S	S	S	N	S	S	B	B	B
33	I	N	N	N	S	S	S	S	N	S	S	B	B	B
34	I	N	N	N	S	S	S	S	N	I	I	B	B	B
36	S	I	N	N	S	N	S	S	S	S	N	B	B	A
37	S	I	S	N	S	S	S	S	S	S	S	B	B	B
35	I	I	N	N	S	S	S	S	S	S	N	B	B	B
38	I	S	N	N	S	S	S	S	S	S	S	B	B	B
39	S	S	N	N	S	S	S	S	S	S	N	B	B	B
30	S	S	N	N	S	S	S	S	S	S	N	B	B	B
40	I	N	N	N	S	S	S	S	N	S	S	B	B	B
41	S	S	N	N	S	S	S	S	S	S	S	B	B	B
42	I	N	N	N	S	I	S	I	S	S	S	B	B	B
44	S	S	S	N	S	N	S	S	S	S	I	B	B	B
45	I	N	N	N	S	S	S	S	N	N	S	B	B	B
43	I	N	N	N	S	S	S	S	N	S	S	B	B	B
46	I	N	N	N	S	S	S	I	N	S	S	B	B	B
48	I	N	N	N	S	S	S	S	N	I	S	B	B	B
49	I	N	N	N	S	S	S	S	N	S	N	B	B	B

PO: Pregunta orientadora.

APÉNDICE 4. Estrategia de búsqueda.

WEB OF SCIENCE (Web of Science Core Collection)
Advanced Search: (TS=((("exhaled") AND ("VOC" OR "VOCs" OR "volatile organic compound" OR "volatile organic compounds" OR "breathomics" OR "volatilome" OR "exhaled metabolites"))) AND ("children" OR "childhood" OR "baby" OR "babies" OR "teenager" OR "teenagers" OR "adolescent" OR "adolescents" OR "child" OR "infant")) AND TS= ("disease" OR "diseases" OR "pathology" OR "pathologies" OR "patient" OR "patients" OR "pathophysiology" OR "diagnosis" OR "diagnostic" OR "biomarker" OR "biomarkers" OR "marker" OR "markers" OR "monitoring" OR "monitorization")) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)
SCOPUS
Advanced Search: (TITLE-ABS-KEY ("exhaled") AND TITLE-ABS-KEY ("VOC" OR "VOCs" OR "volatile organic compound" OR "volatile organic compounds" OR "breathomics" OR "volatilome" OR "exhaled metabolites") AND TITLE-ABS-KEY ("children" OR "childhood" OR "baby" OR "babies" OR "teenager" OR "teenagers" OR "adolescent" OR "adolescents" OR "child" OR "infant") AND TITLE-ABS-KEY ("disease" OR "diseases" OR "pathology" OR "pathologies" OR "patient" OR "patients" OR "pathophysiology" OR "diagnosis" OR "diagnostic" OR "biomarker" OR "biomarkers" OR "marker" OR "markers" OR "monitoring")) AND DOCTYPE (ar) AND (LIMIT-TO (LANGUAGE , "English"))
PubMed
((("exhaled") AND ("VOC" OR "VOCs" OR "volatile organic compound" OR "volatile organic compounds" OR "breathomics" OR "volatilome" OR "exhaled metabolites"))) AND ("children" OR "childhood" OR "baby" OR "babies" OR "teenager" OR "teenagers" OR "adolescent" OR "adolescents" OR "child" OR "infant")) AND ("disease" OR "diseases" OR "pathology" OR "pathologies" OR "patient" OR "patients" OR "pathophysiology" OR "diagnosis" OR "diagnostic" OR "biomarker" OR "biomarkers" OR "marker" OR "markers" OR "monitoring" OR "monitorization") NOT (review[ptyp]) AND English

APÉNDICE 5. Preguntas orientadoras utilizadas para juzgar el riesgo de sesgo y la aplicabilidad en cada uno de los estudios seleccionados.

RIESGO DE SESGO		
DOMINIO 1: Selección de Pacientes	<p>¿Podría la selección de pacientes haber introducido sesgo? -Pregunta orientadora 1: ¿Se enroló una muestra consecutiva o aleatoria de pacientes? -Pregunta orientadora 2: ¿Se evitó un diseño de casos y controles?</p>	Riesgo: Bajo/Alto/Incierto Sí/No/Incierto Sí/No/Incierto
DOMINIO 2: Prueba Índice	<p>¿Podría la realización o la interpretación de la prueba índice haber introducido sesgo? -Pregunta orientadora 1: ¿Se ha realizado una validación externa del modelo para obtener un valor de la tasa de clasificación? -Pregunta orientadora 2: ¿Se probaron los VOCs o la huella de aliento discriminantes seleccionados en cohortes de pacientes totalmente independientes? -Pregunta orientadora 3: ¿Se utilizaron los VOCs discriminantes seleccionados e identificados en lugar de una huella de aliento?</p>	Riesgo: Bajo/Alto/Incierto Sí/No/Incierto Sí/No/Incierto Sí/No/Incierto
DOMINIO 3: Prueba de Referencia	<p>¿Podría la prueba de referencia o su interpretación haber introducido sesgos? -Pregunta orientadora 1: ¿Es probable que la prueba de referencia valore correctamente la condición diana? -Pregunta orientadora 2: ¿Fueron interpretados los resultados de la prueba de referencia sin conocimiento de los resultados de la prueba índice?</p>	Riesgo: Bajo/Alto/Incierto Sí/No/Incierto Sí/No/Incierto
DOMINIO 4: Flujos y Tiempos	<p>¿Podría el flujo de pacientes haber introducido sesgos? -Pregunta orientadora 1: ¿Hubo un intervalo apropiado entre la prueba índice y la prueba de referencia? -Pregunta orientadora 2: ¿Fue aplicada en todos los individuos de la misma prueba de referencia? -Pregunta orientadora 3: ¿Recibieron los pacientes el mismo estándar de referencia? -Pregunta orientadora 4: ¿Fueron incluidos todos los pacientes en el análisis?</p>	Riesgo: Bajo/Alto/Incierto Sí/No/Incierto Sí/No/Incierto Sí/No/Incierto Sí/No/Incierto
PREOCUPACIÓN POR LA APLICABILIDAD		
DOMINIO 1: Selección de Pacientes	<p>¿Existe preocupación de que los pacientes incluidos no coincidan con la pregunta de la revisión?</p>	Preocupación: Baja/Alta/Incierta
DOMINIO 2: Prueba Índice	<p>¿Existe preocupación de que la prueba índice, su realización o interpretación difieran de la pregunta de la revisión?</p>	Preocupación: Baja/Alta/Incierta
DOMINIO 3: Prueba de Referencia	<p>¿Existe preocupación de que la condición definida por la prueba de referencia no coincida con la pregunta de la revisión?</p>	Preocupación: Baja/Alta/Incierta

Capítulo 4

Flujo de trabajo de código abierto para preprocesamiento de datos en el análisis del aire exhalado off-line mediante GC-MS

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, José María Pastor Hernández, Gema Lozano-Terol,
Julia Gallego-Jara, Luis García-Marcos, Manuel Cánovas Díaz y Teresa de
Diego Puente

**Data preprocessing workflow for exhaled breath analysis by GC/MS using
open sources**

Scientific Reports **10**, 22008 (2020). doi: 10.1038/s41598-020-79014-6.

RESUMEN

Actualmente, un objetivo prioritario en el ámbito médico es la búsqueda de metodologías no invasivas para el diagnóstico y el seguimiento de enfermedades de alta prevalencia como enfermedades cardiovasculares, distintos tipos de cáncer y enfermedades respiratorias crónicas. En este sentido, el análisis de los compuestos orgánicos volátiles (VOCs) ha sido señalado como una herramienta no invasiva con alto potencial clínico para el diagnóstico y monitorización de varias enfermedades. A pesar de las ventajas de este enfoque, esta herramienta se encuentra en una fase exploratoria y no está preparada todavía para su instauración en la clínica diaria. Uno de los principales obstáculos para alcanzar este propósito es la falta de protocolos reproducibles para cada una de las etapas del proceso de descubrimiento de biomarcadores, especialmente en la etapa de preprocesamiento de datos. Por lo tanto, en este trabajo se presenta un flujo de trabajo de código abierto para el preprocesamiento de los datos obtenidos por el análisis de muestras de aire exhalado mediante cromatografía de gases acoplada a espectrometría de masas (GC-MS) de cuadrupolo simple. Este flujo de trabajo se basa en la unión de las dos principales estrategias para transformar los datos brutos extraídos del MS en una matriz útil para el análisis estadístico. Además, este flujo de trabajo incluye el cotejo de los compuestos de las muestras de aire exhalado con una biblioteca espectral. Para todo ello, se han empleado tres paquetes gratuitos escritos en el lenguaje de programación R (*xcms*, *cliqueMS* y *eRah*). Al mismo tiempo, en este trabajo se presenta un protocolo para la recogida de muestras de aire exhalado de niños menores de 2 años para su posterior análisis mediante GC-MS.

INTRODUCCIÓN

En las últimas décadas, el análisis de los compuestos orgánicos volátiles (VOCs) en el aire exhalado ha sido propuesto como un enfoque prometedor en la búsqueda de biomarcadores para el diagnóstico y el seguimiento de diferentes enfermedades, y ha sido utilizado para evaluar la exposición a la contaminación del aire ambiental o al tabaco¹⁻⁷. Una de las principales ventajas de esta estrategia es que se basa en un procedimiento no invasivo¹. Esto es especialmente importante para poblaciones como los niños y los ancianos, así como para enfermedades para las que las pruebas diagnósticas de referencia son técnicas invasivas como las biopsias o las broncoscopias⁸⁻¹⁰. Sin embargo, este enfoque todavía se encuentra en la fase de descubrimiento de biomarcadores y aún no se ha implantado en la clínica^{11,12}. En este sentido, es necesario superar varias etapas para lograr la utilidad clínica de los

biomarcadores en aire exhalado. En general, el proceso para alcanzar la utilidad clínica de un biomarcador comprende las siguientes etapas: descubrimiento de biomarcadores, validación analítica y validación clínica¹³. Por otro lado, el flujo de trabajo de la fase de descubrimiento de biomarcadores suele incluir los siguientes pasos: diseño del estudio, toma de muestras de aire exhalado, análisis del aire exhalado, preprocesamiento de datos, identificación de los VOCs, análisis de datos, la interpretación de los resultados y la validación del biomarcador putativo sobre una cohorte independiente¹⁴.

Actualmente, el análisis del aire exhalado a tiempo real en línea u *on-line* está en auge gracias al empleo de tecnologías basadas en la espectrometría de masas (MS), como SESI-MS (ionización secundaria por *electrospray* – espectrometría de masas), SIFT-MS (tubo de selección de iones con flujo - espectrometría de masas) o PTR-MS (reacción de transferencia de protones – espectrometría de masas)¹⁵. No obstante, a día de hoy, la tecnología más utilizada para el análisis de VOCs en el aire exhalado sigue siendo la espectrometría de masas acoplada a cromatografía de gases (GC-MS), puesto que es una técnica muy sensible y fiable. Esta técnica se utiliza principalmente en el análisis del aire exhalado fuera de línea u *off-line*^{11,16,17} y generalmente precisa de métodos de preconcentración de los analitos, como los tubos de desorción térmica (tubos TD) y la microextracción en fase sólida (SPME)¹⁸. En el análisis mediante GC-MS, los compuestos presentes en el aire exhalado se fragmentan en iones con diferentes relaciones masa/carga (m/z)¹⁹. Aunque diversas técnicas de ionización pueden ser empleadas en GC-MS, la ionización electrónica (EI) a 70 eV es claramente la más popular. La EI es una metodología de ionización fuerte que posee una alta reproducibilidad. Por lo tanto, es posible realizar comparaciones con bibliotecas de espectros de masas y con otros laboratorios²⁰. Por otro lado, existen dos tipos de adquisición de datos en GC-MS: adquisición de todos los iones dentro de un determinado rango de m/z (modo de barrido completo) y adquisición de un único ion (modo de monitorización de iones seleccionados (SIM)). El modo SIM se utiliza con frecuencia en los análisis dirigidos, que consisten en la determinación de unos pocos compuestos conocidos. Por el contrario, el modo de barrido completo se utiliza generalmente para el análisis no dirigido, en el que se determina cualquier compuesto presente en una muestra^{21,22}. Por ello, es frecuente recurrir al análisis no dirigido en la fase de descubrimiento de biomarcadores²³.

Antes de realizar los análisis estadísticos, el preprocesamiento del gran volumen de datos brutos obtenidos tras el análisis de las muestras del aire exhalado a través de GC-MS es esencial para la correcta cuantificación e identificación de los VOCs

presentes en las muestras^{24,25}. Sin embargo, a menudo, este paso no es descrito con suficiente detalle en los estudios centrados en el análisis de VOCs en el aire exhalado²⁶. Por lo tanto, el desarrollo de protocolos reproducibles y comparables para el preprocesamiento de datos desempeña un papel crucial para superar la fase de descubrimiento de biomarcadores^{12,26}. La comunidad científica sigue estancada en el uso de *software* privativo, a pesar de los grandes avances de los últimos años en el campo de la bioinformática, especialmente en las propuestas de código abierto. De hecho, actualmente, para el desarrollo de esta tarea existen varias herramientas computacionales y paquetes escritos en lenguaje de programación R que están disponibles de forma gratuita^{23,27-30}.

Los problemas más típicos del preprocesamiento de datos suelen ser: el gran volumen de datos brutos (especialmente en los análisis no dirigidos), el ruido de fondo, las variaciones en los tiempos de retención (RT) de los VOCs entre las muestras, las variaciones en el perfil del espectro de masas de los VOCs entre las muestras (variaciones en los valores de las intensidades relativas de los fragmentos iónicos o *features* dentro de un compuesto), o el solapamiento de los VOCs con tiempos de retención muy similares, entre otros^{24,27,31,32}. En la actualidad, las estrategias para llevar a cabo el preprocesamiento de los datos de volatilómica se clasifican en dos grandes grupos. Básicamente, la primera estrategia se centra, principalmente, en la búsqueda de los picos de iones detectados en un espectro de EI-MS (*features*) o el llamado *peak-picking*, y la segunda estrategia tiene como objetivo determinar los compuestos presentes en muestras de aire exhalado mediante deconvolución espectral. Por un lado, la dificultad para determinar la identidad de los compuestos es un punto débil de la primera estrategia²⁷ y, por otro, el uso de la deconvolución es cuestionado porque puede añadir errores y aumentar la variación²⁵. Por lo tanto, en este trabajo se describe un procedimiento nuevo y sencillo para el preprocesamiento de datos tras el análisis del aire exhalado mediante un sistema acoplado de desorción térmica y cromatografía de gases-espectrometría de masas de cuadrupolo simple (TD-GC/q-MS) que integra las dos estrategias. De este modo, es posible beneficiarse de las ventajas de ambos enfoques, incluida la doble validación entre las dos estrategias para reducir posibles errores y evitar duplicaciones de compuestos (detección de un único compuesto como dos compuestos). Para ello, se han implementado las funciones de tres paquetes escritos en lenguaje de programación R: *xcms*²³, *cliqueMS*²⁸ y *eRah*²⁷.

RESULTADOS

Flujo de trabajo del preprocesamiento de datos

Las etapas del proceso de adquisición de datos brutos realizado antes del preprocesamiento de datos se muestran en la Figura 1. Para desarrollar el flujo de trabajo del preprocesamiento de datos se emplearon los datos brutos adquiridos tras el análisis de las muestras de aire exhalado de las parejas de madres e hijos de la cohorte de nacimiento NELA (*Nutrition in Early Life and Asthma*)³³ (véase la sección de Materiales y métodos para más detalles). Como se puede observar en la Figura 1, el paso previo al preprocesamiento de datos fue la conversión de los datos brutos a un formato estándar y abierto como mzXML que es independiente del *software* de la instrumentación empleada. Para ello, se utilizó el programa de libre uso *MConvert* de *Proteowizard*^{34,35}. A continuación, las muestras se dividieron aleatoriamente en dos grupos (Grupo 1 y Grupo 2). En cada grupo, las muestras de aire exhalado de las madres, de los bebés y del contenido del aire de la habitación donde se tomaban las muestras fueron preprocesadas por separado.

El flujo de trabajo seguido para el preprocesamiento de los datos obtenidos por TD-GC/q-MS se muestra en la Figura 2. En este flujo de trabajo, la búsqueda de los compuestos en las muestras de aire exhalado se llevó a cabo mediante las dos estrategias más relevantes mencionadas previamente en la introducción, es decir, detección de los picos iónicos (*features*) y detección de los compuestos. Para la primera estrategia, se utilizaron dos paquetes del *software* estadístico gratuito R (*xcms*²³ y *cliqueMS*²⁸), mientras que, para la segunda estrategia, se empleó otro paquete de R, *eRah*^{27,36}. Finalmente, las dos estrategias se fusionaron en el último paso. En el Apéndice 1 se muestra un resumen esquemático, en forma de tutorial, del flujo de trabajo y el código de R de las funciones más importantes.



Figura 1. Etapas de la fase de descubrimiento de biomarcadores realizadas antes del preprocesamiento de datos. TD-GC/q-MS: Sistema acoplado de desorción térmica y cromatografía de gases-espectrometría de masas de cuadrupolo simple.

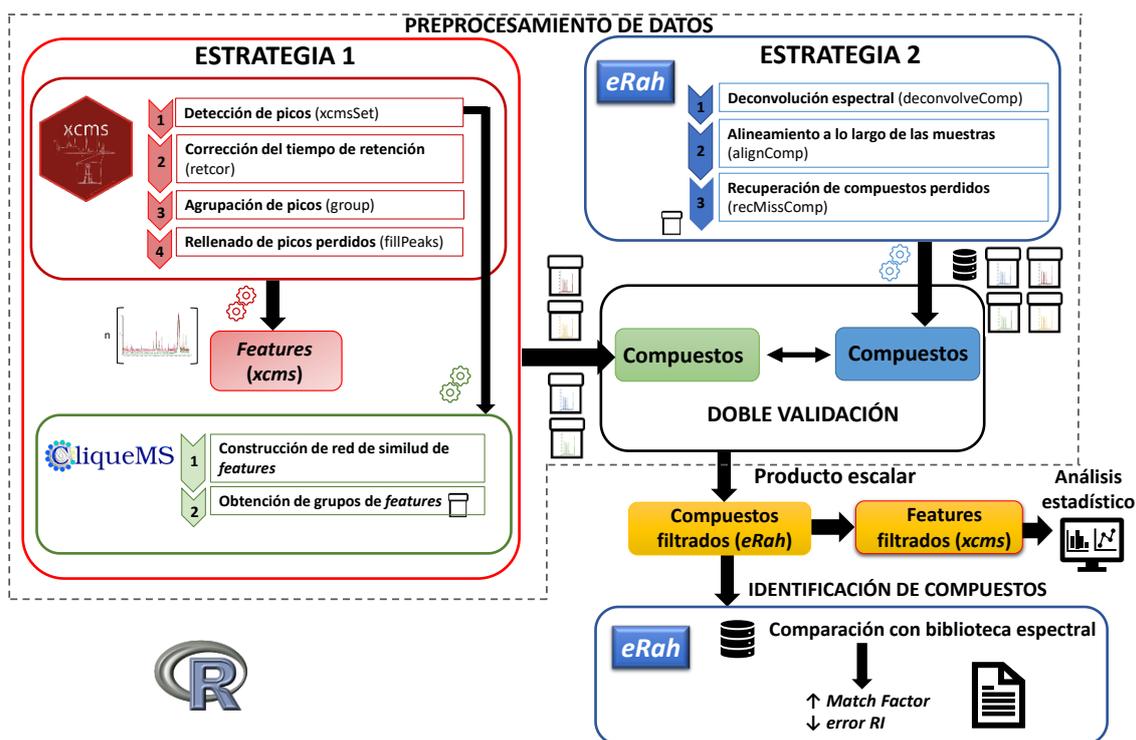


Figura 2. Flujo de trabajo para el preprocesamiento de datos una vez realizado el análisis del aire exhalado mediante TD-GC/q-MS (desorción térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple).

El procedimiento seguido fue el siguiente: en primer lugar, se utilizaron las muestras de aire exhalado del Grupo 1 para establecer los valores de los parámetros de cada función de los paquetes y para el diseño del flujo de trabajo; a continuación, este método se implementó sobre las muestras de aire exhalado del Grupo 2. La única diferencia entre ambos grupos fue que las muestras del segundo grupo se analizaron con una columna de cromatografía de gases diferente a la utilizada para el Grupo 1. Aunque ambas columnas eran de la misma marca y del mismo modelo, se observaron ligeras diferencias entre los tiempos de retención de los *features* y los compuestos de las muestras de ambos grupos. La Figura 3 muestra las diferencias entre los tiempos de retención de un pico iónico característico del tolueno (compuesto omnipresente en las muestras de aire ambiental) para ambos grupos. Como el objetivo era verificar la reproducibilidad de este protocolo en otro conjunto de muestras, no se realizaron cambios en los parámetros de análisis para disminuir las desviaciones entre los tiempos de retención. Además, se identificaron varios artefactos en las muestras, como por ejemplo, N,N-dimetilacetamida y fenol, que son contaminantes habituales y ampliamente documentados de las bolsas Tedlar^{®37} (Figura 4). Por lo tanto, ambos compuestos estaban presentes en todas las muestras de aire exhalado. También se identificaron otros contaminantes analíticos conocidos y persistentes, como los polidimetilsiloxanos.

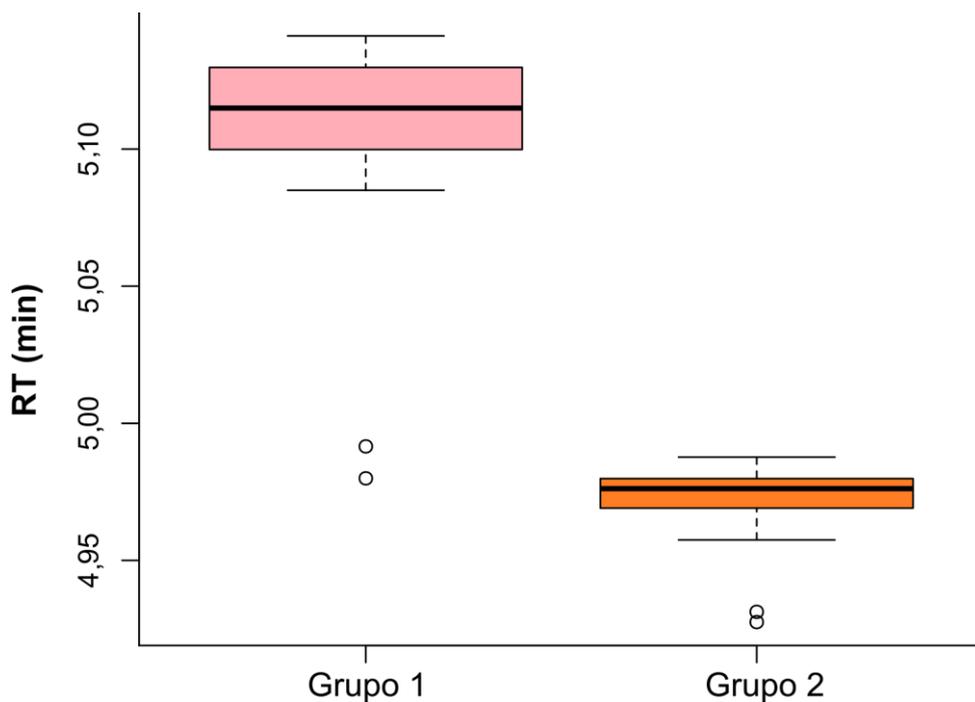


Figura 3. Diferencias en el tiempo de retención (RT) entre ambos grupos de muestras. En el gráfico se muestran los tiempos de retención de un *feature* ($m/z = 91$) del tolueno presente en las 337 muestras de aire ambiental (210 muestras del grupo 1) y (127 muestras del grupo 2). Se observaron diferencias estadísticamente significativas (p -valor = $2,2e-16$) entre los tiempos de retención de ambos grupos mediante la prueba U de Mann-Whitney.

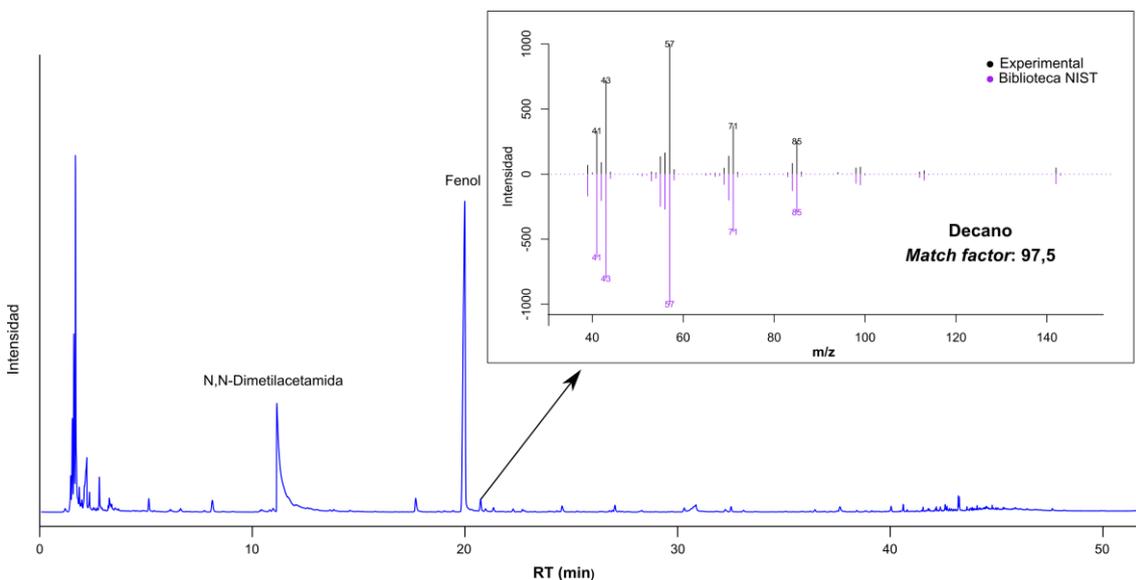


Figura 4. Cromatograma de iones totales (TIC) obtenido al analizar muestra de aire exhalado de una madre del grupo 1. Se pueden observar los contaminantes de las bolsas Tedlar® (Fenol y N,N-Dimetilacetamida). También se muestra el espectro de masas de un compuesto (Decano). El espectro extraído fue comparado con la biblioteca NIST para conocer la identidad del compuesto. Los gráficos de esta figura fueron generados empleando las funciones *plotChr* y *plotSpectra* del paquete *eRah* en R.

El tiempo empleado para el preprocesamiento de datos y la identificación de compuestos usando este flujo de trabajo fue de aproximadamente 2 minutos por muestra. En el Apéndice 2 se detalla el tiempo empleado en cada etapa del flujo de trabajo para cada subconjunto de muestras de ambos grupos.

Preprocesamiento de datos: Estrategia 1

Como se ha comentado, esta estrategia utiliza los paquetes *xcms* y *CliqueMS*. Las cuatro principales etapas de *xcms* fueron aplicadas sobre nuestros datos utilizando las siguientes funciones: *xcmsSet* (detección de picos), *retcor* (alineación del tiempo de retención), *group* (agrupamiento de picos) y *fillPeaks* (relleno de picos perdidos) (Figura 2). Antes de la detección de picos, se llevó a cabo una eliminación de ruido o *denoising* y una corrección de la línea de base, optimizando los parámetros *fwhm* (*Full Width at Half Maximum*) y *snthresh* (umbral de la relación señal/ruido) de la función *xcmsSet*²³. *fwhm* hace referencia al ancho total espectral a la mitad del máximo de altura para aumentar la capacidad de diferenciar señales estrechamente relacionadas y por tanto la resolución. En la detección de picos (etapa 1), el objetivo era determinar el mayor número de picos en cada muestra sin duplicar los picos incorrectamente, y separando los picos superpuestos. Un par de algoritmos (*MatchedFilter* y *Centwave*) pueden ser empleados en la detección de picos mediante *xcms*. El algoritmo *Centwave* reduce el coste computacional del preprocesamiento y el tamaño de los datos, por lo que es ideal para datos generados por MS de alta resolución en modo centroide. De esta manera, los datos en modo *profile* o continuo pueden transformarse en datos en modo centroide. Sin embargo, como nuestros datos brutos eran datos en modo *profile* adquiridos por un espectrómetro de masas de baja resolución (cuadrupolo simple), se eligió el algoritmo *matchedFilter* para la extracción de los picos iónicos. En la etapa 2, se empleó el método *retcor* para corregir las desviaciones de los tiempos de retención de las distintas muestras y así garantizar que los tiempos de retención estuvieran alineados en todas las muestras de cada subconjunto. A continuación, la tercera etapa, se basa en el agrupamiento por similitud en todas las muestras de los picos que han sido detectados previamente en cada una de ellas. Así, los picos con características similares (mismo *m/z* y tiempo de retención similar) se consideraron como el mismo *feature*. Tras realizar las etapas 2 y 3, se obtuvo una matriz de datos con las intensidades de los *features* de cada muestra.

Finalmente, el propósito de la última etapa era rellenar los valores faltantes o "huecos" en la matriz de datos generada (etapa 4). Por lo tanto, los valores que faltaban

de las intensidades de los *features* más frecuentes en las muestras se buscaron entre el inicio y el final de los tiempos de retención alineados (etapa 2) de cada *feature*^{23,38}.

Los parámetros seleccionados para el preprocesamiento de datos empleando *xcms* se especifican en el Apéndice 3. Se utilizó la intensidad máxima o *maxo* como valor de intensidad para cada *feature*. La Tabla 1 muestra el número total de *features* detectados en las muestras de ambos grupos mediante el preprocesamiento con *xcms*.

Tabla 1. Resultados obtenidos con las dos estrategias.

Muestras	ESTRATEGIA 1				ESTRATEGIA 2	
	<i>xcms</i>		<i>cliqueMS</i>		<i>eRah</i>	
	Nº de <i>features</i>		Nº de compuestos		Nº de compuestos	
	Grupo 1	Grupo 2	Grupo 1	Grupo 2	Grupo 1	Grupo 2
Aire exhalado de las madres	542	524	1613	1241	835	554
Aire exhalado de los niños	467	543	1242	1105	802	569
Aire ambiental	867	937	2927	2223	913	610

Posteriormente, el paquete *CliqueMS* fue utilizado para identificar que *features* pertenecían al mismo compuesto. Los resultados obtenidos después de aplicar la función *xcmsSet* (función de *xcms* para la detección de picos) fueron empleados como datos de entrada. El flujo de trabajo del paquete *CliqueMS* ha sido ampliamente descrito previamente por Senan *et al.*²⁸. La agrupación de los *features* mediante *CliqueMS* se basa en la construcción de redes de similitud e implica dos etapas: 1) construcción de una red de similitud de *features* y 2) obtención de grupos de *features*. Una desventaja de este paquete es que anota las muestras una a una. Para poder aplicarlo a todas las muestras, se programó un bucle a partir de los *features* determinados en el primer paso de *xcms*. El número total de compuestos diferentes identificados en todas las muestras de ambos grupos se detalla en la Tabla 1, y los parámetros de *cliqueMS* empleados se recopilan en el Apéndice 3.

Preprocesamiento de datos: Estrategia 2

En la segunda estrategia, se utilizaron tres funciones del paquete *eRah*, *deconvolveComp*, *alignComp* y *recMissComp*, para realizar tres etapas: 1) deconvolución espectral, 2) alineamiento a lo largo de las muestras y 3) recuperación de compuestos perdidos (Figura 2). La eliminación de ruido y la corrección de la línea base se realizaron automáticamente a través del parámetro *threshold* de la función *deconvolveComp*. La deconvolución espectral (etapa 1) de *eRah* incluye dos pasos: i) agrupamiento de los compuestos por covarianza local (CMLC) y ii) deconvolución de señales ortogonales (OSD). La CMLC facilita la identificación de compuestos en los cromatogramas de las muestras, y la OSD permite la recuperación del espectro de los compuestos a partir del perfil de elución de éstos²⁷. De este modo, gracias a la deconvolución espectral (etapa 1), se obtuvo una lista con las señales *m/z* y las intensidades relativas de los iones que formaban parte de los compuestos. Así pues, el ion de mayor intensidad (pico base) tenía asignada una abundancia de 1000. Posteriormente, se llevó a cabo la alineación de los tiempos de retención de los compuestos a lo largo de todas las muestras (etapa 2) y la determinación de los compuestos que faltaban (etapa 3). Aunque estas dos etapas (2 y 3) tenían el mismo objetivo que las de etapas de la estrategia 1, la metodología empleada en *eRah* se basa en otros principios matemáticos y en datos de entrada diferentes. En la Figura 5, se muestra los perfiles de elución de dos compuestos ubicuos en las muestras de contenido de aire ambiente (tolueno y limoneno) antes y después del alineamiento.

Los parámetros seleccionados en cada función en esta segunda estrategia se muestran en el Apéndice 3. Dado que el análisis por TD-GC/q-MS se realizó utilizando tres rampas de temperatura diferentes, el preprocesamiento de datos de las muestras por *eRah* se ejecutó de forma independiente en tres partes (Apéndice 4).

Por último, la identificación de los compuestos por cotejo con una biblioteca espectral se llevó a cabo mediante otra función de *eRah* y se consideró que cumplían los criterios de coincidencia con los compuestos de la biblioteca espectral NIST (*National Institute of Standards and Technology*) sólo aquellos compuestos con un factor de coincidencia o *match factor* (puntuación de similitud espectral) superior al 80%. La Tabla 1 muestra el número total de compuestos seleccionados mediante la segunda estrategia aplicando este filtro.

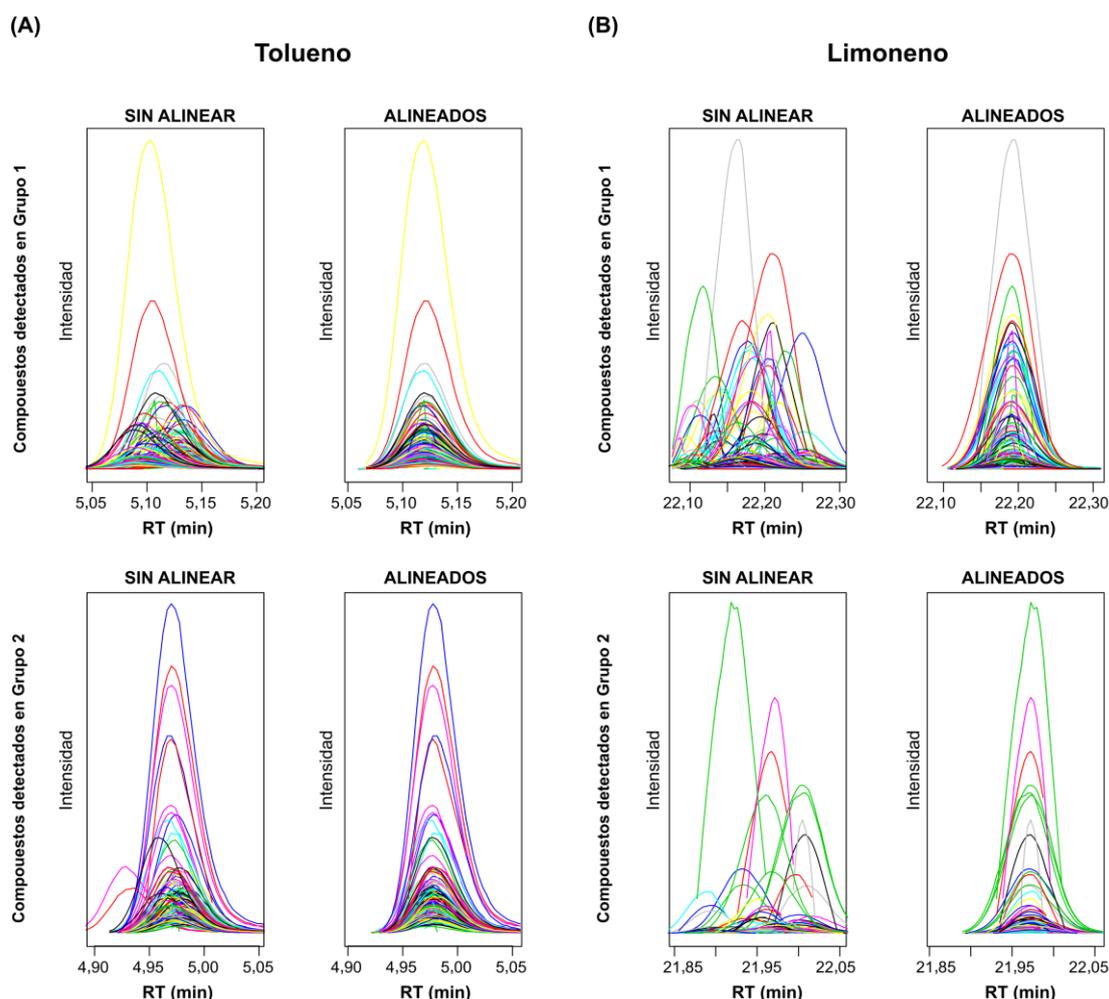


Figura 5. Alineamiento espectral usando *eRah*. (A) Perfil de elución del tolueno en las muestras de aire ambiental de ambos grupos de muestras (Grupo 1 y Grupo 2) antes y después del alineamiento. (B) Perfil de elución del limoneno en las muestras de aire ambiental de ambos grupos de muestras (Grupo 1 y Grupo 2) antes y después del alineamiento. Los gráficos de esta figura se generaron con la función *plotAlign* del paquete *eRah* en R.

Conexión entre ambas estrategias

Ambas estrategias se combinaron para verificar los resultados obtenidos, evitar los duplicados y generar una lista única de *features* filtrados y compuestos filtrados. Una vez que los compuestos de las muestras fueron detectados por las dos estrategias, se realizaron dos pasos adicionales previos para permitir la comparación entre los compuestos. En primer lugar, dado que la segunda estrategia permite determinar el perfil de cada compuesto, se determinaron los *features* detectados con *xcms* que coincidían con los iones de los compuestos encontrados con *eRah*. Para ello, un *feature* obtenido por *xcms* y un ion de un compuesto obtenido por *eRah* debían tener la misma señal *m/z* y el mismo tiempo de retención $\pm 0,05$ min para ser considerados iguales. En segundo lugar, se determinó la intensidad relativa de cada *feature* dentro de un

compuesto obtenido mediante *cliqueMS*. Para ello, se siguió un protocolo de tres pasos: 1) Se calculó la intensidad media de cada *feature* en todo el conjunto de muestras, 2) Se asignó un valor de 1000 como valor de intensidad relativa del *feature* con la mayor intensidad en el compuesto (pico base), 3) El valor de la intensidad relativa de los restantes *features* se calculó de la siguiente manera (intensidad media del *feature*/intensidad media del *feature* con mayor intensidad) * 1000.

Posteriormente, cada compuesto obtenido mediante *xcms-cliqueMS* se comparó con todos los compuestos detectados con *eRah*. Para ello, se construyeron dos vectores (un vector para un compuesto detectado con la primera estrategia y otro para un compuesto detectado con la segunda estrategia) con 412 posiciones, donde cada posición correspondía a una señal *m/z* dentro del rango de *m/z* 38-450. Estos vectores se rellenaron con las intensidades relativas (que iban de 0 a 1000) de los fragmentos iónicos de cada compuesto. Un valor de 0 indicaba que la señal *m/z* no se había detectado en el compuesto. El producto escalar entre ambos vectores era calculado, obteniendo un valor único entre 0-1. Un valor cercano a 1 indicaba un alto nivel de similitud entre los compuestos. Por lo tanto, fue posible determinar qué compuesto detectado con *eRah* era el más semejante al compuesto determinado con *xcms-cliqueMS* (compuesto con el valor del producto escalar más elevado). Este proceso es similar a la comparación del espectro de un compuesto con los de una biblioteca³⁹. Como resultado, los compuestos determinados por ambas estrategias se emparejaron a través del cálculo de un producto escalar. A continuación, únicamente se seleccionaron los compuestos detectados por *eRah* cuyo resultado del producto escalar era el mayor para algunos de los compuestos de *cliqueMS* con un valor superior a 0,7 (compuestos filtrados). Adicionalmente, se aplicó un segundo filtro. Sólo se seleccionaron los *features* detectados con *xcms* que habían sido previamente emparejados con los compuestos filtrados de la segunda estrategia (*features* filtrados). La Tabla 2 muestra el número total de *features* filtrados y de compuestos filtrados de ambos grupos. Asimismo, los valores de intensidad de los *features* se normalizaron en todas las muestras de cada subconjunto. Para ello, los valores de intensidad fueron previamente transformados empleando logaritmos de base 10. A continuación, los valores transformados de intensidad de cada *feature* en una muestra fueron divididos por la suma de los valores transformados de intensidad de todos los *features* presentes en esa muestra, y multiplicados por 1000.

Tabla 2. *Features* filtrados y compuestos filtrados.

Muestras	Nº de <i>features</i> filtrados		Nº de compuestos filtrados	
	Grupo 1	Grupo 2	Grupo 1	Grupo 2
Aire exhalado de las madres	494	452	194	150
Aire exhalado de los niños	377	483	155	158
Aire ambiental	824	877	231	193

Entre los compuestos determinados por *eRah* y *xcms-cliqueMS*, se encontraron muchos posibles compuestos duplicados. Sin embargo, tras la integración de ambas estrategias, el porcentaje de posibles compuestos duplicados se redujo significativamente (Figura 6). Todos los compuestos detectados en cada subconjunto se compararon entre sí mediante el cálculo del producto escalar de los vectores de intensidades relativas de los fragmentos iónicos de cada compuesto. Se consideró una posible duplicación en los compuestos detectados en *eRah* si dos compuestos tenían un valor de producto escalar superior a 0,8, el mismo tiempo de retención $\pm 0,05$ min y sus áreas obtenidas por deconvolución espectral estaban correlacionadas (coeficiente de correlación de Pearson $> 0,75$). Las duplicaciones en los compuestos obtenidos con *xcms-cliqueMS* se determinaron en función del tiempo de retención y del valor del producto escalar.

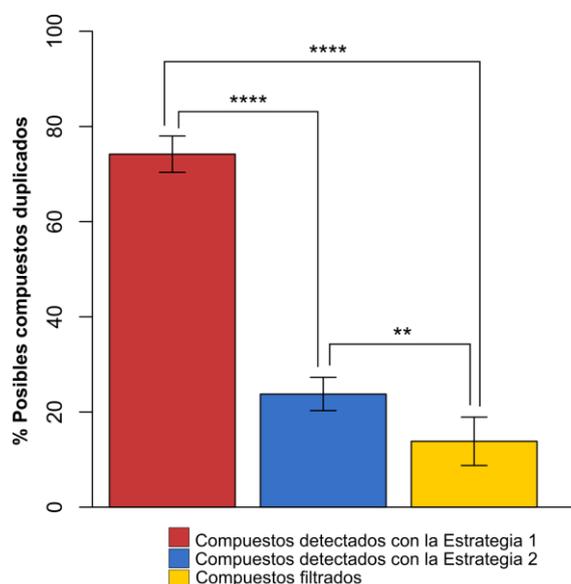


Figura 6. Posibles compuestos duplicados. Las comparaciones estadísticas entre los porcentajes de cada subconjunto de posibles compuestos duplicados en los compuestos detectados por la primera estrategia, los compuestos detectados por la segunda estrategia y los compuestos filtrados se realizaron con una prueba ANOVA de seguida del método de comparaciones múltiples o *post-hoc* de Bonferroni (p-valor $< 0,01$ (**), p-valor $< 0,0001$ (****)).

Identificación de los compuestos filtrados

El paquete *eRah* permite la comparación de los compuestos filtrados con los compuestos registrados en una biblioteca espectral. *eRah* calcula dos factores para indicar la similitud entre dos compuestos: el factor de coincidencia (*match factor*) y el error del índice de retención (*error RI*). Mientras que el primero está relacionado con la semejanza entre los espectros de masas, el otro hace referencia a la similitud de los tiempos de retención de los compuestos. Ambos valores se expresan en porcentajes. En este sentido, si dos compuestos son idénticos, tendrán un factor de coincidencia del 100% (alta similitud entre los espectros) y un error del índice de retención del 0% (mínima diferencia entre los tiempos de retención). En este flujo de trabajo de preprocesamiento de datos, la biblioteca NIST fue seleccionada como biblioteca de espectros de masas para llevar a cabo la comparación. El *error RI* de cada compuesto se calculó utilizando compuestos de referencia (con un tiempo de retención y un valor de RI) y se comparó con el RI de los compuestos de la biblioteca. El *error RI* de los compuestos filtrados también se determinó en tres partes independientes, ya que el análisis por TD-GC/q-MS se llevó a cabo utilizando tres rampas de temperatura (Apéndice 4). Un estándar comercial de calibración de VOCs que contenía 54 compuestos individuales se empleó como referencia de tiempo de retención para el cálculo del *error RI*. Sólo 10 de los compuestos incluidos en la mezcla del estándar comercial tenían un RI que estaba disponible en la biblioteca NIST (Apéndice 5). Asimismo, ninguno de estos 10 compuestos tenía un tiempo de retención dentro del intervalo de 38,67-46,12 min (tercera rampa de temperatura). No obstante, aunque la biblioteca NIST es la mayor biblioteca de espectros que conocemos, su base de datos de índices de retención sigue siendo muy limitada.

En consecuencia, para aumentar la cantidad de valores de índices de retención de los compuestos de referencia, se utilizó un estándar comercial de n-alcános saturados C7-C30. Los tiempos de retención de los alcános lineales observados en cada columna se muestran en el Apéndice 6. Además, se determinaron los tiempos de retención de los alcános lineales en las muestras de aire exhalado de ambos grupos comparando los tiempos de retención con los de los estándares conocidos. Para ello, se seleccionaron diez muestras de aire exhalado de las madres del Grupo 1 y se analizaron siguiendo la segunda estrategia. Así, se llevó a cabo una comparación de los compuestos detectados por *eRah* con exclusivamente las entradas de la biblioteca NIST de los alcános lineales de 7-24 carbonos. En primer lugar, únicamente los compuestos detectados con un *match factor* superior al 95% (alta similitud espectral) fueron

seleccionados. Así, entre todos los compuestos detectados por *eRah* en las muestras, 38 de ellos mostraron este grado de similitud espectral con algún alcano lineal. A continuación, estos 38 compuestos se emparejaron independientemente con cada uno de los alcanos lineales. Entonces, se seleccionaron los compuestos con el mayor valor de *match factor* para un alcano lineal concreto o aquellos con hasta 5 puntos menos que el mayor *match factor* considerado. Además, los compuestos en los que el *m/z* equivalente al peso molecular del alcano lineal no estaba incluido en el perfil determinado por *eRah* fueron descartados. De esta manera, se obtuvieron 82 combinaciones de posibles alcanos lineales, ya que cada uno de los 38 compuestos seleccionados anteriormente podía ser emparejado con varios alcanos lineales. También se realizó una búsqueda de alcanos lineales empleando *xcms* con el fin de reducir el número de combinaciones de posibles alcanos lineales. De este modo, 76 *features* con *m/z* iguales al peso molecular de los alcanos lineales fueron seleccionados. A continuación, las dos estrategias se volvieron a unir. Por consiguiente, solamente fueron elegidas las combinaciones de posibles alcanos lineales que poseían un $RT \pm 0,05$ min en comparación con el RT del *feature* con *m/z* equivalente al peso molecular del alcano lineal. De este modo, se redujo sustancialmente las posibles combinaciones de alcanos, disminuyendo de 82 posibles combinaciones hasta sólo 14. Posteriormente, se seleccionó el compuesto con el RT más bajo cada alcano lineal. Asimismo, después de representar los tiempos de retención de los compuestos seleccionados frente al número de carbonos de los alcanos lineales, se eliminaron tres valores atípicos. Por tanto, finalmente, se consiguió determinar los tiempos de retención de 8 alcanos lineales (Figura 7). Adicionalmente, este proceso se repitió también para 10 muestras de las madres del Grupo 2.

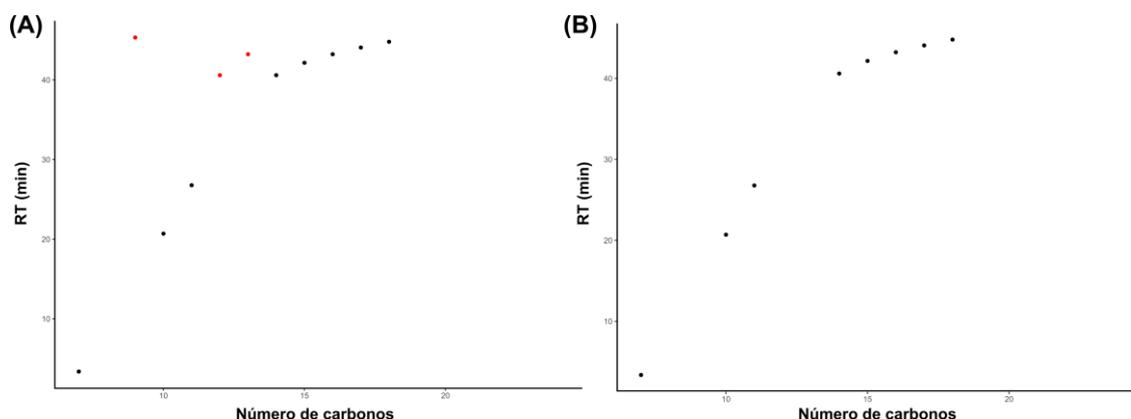


Figura 7. Tiempos de retención de alcanos lineales determinados en muestras de aire exhalado. (A) Tiempos de retención de posibles alcanos lineales en muestras de aire exhalado del Grupo 1. Los valores atípicos están indicados en rojo. (B) Tiempos de retención de alcanos lineales en las muestras de aire exhalado del Grupo 1 después de eliminar los valores atípicos.

Como se puede observar en la Figura 8, los tiempos de retención de los alcanos lineales determinados para las muestras de aire exhalado eran muy próximos a los tiempos de retención observados al analizar el estándar comercial con una desviación de $\pm 0,02$ min.

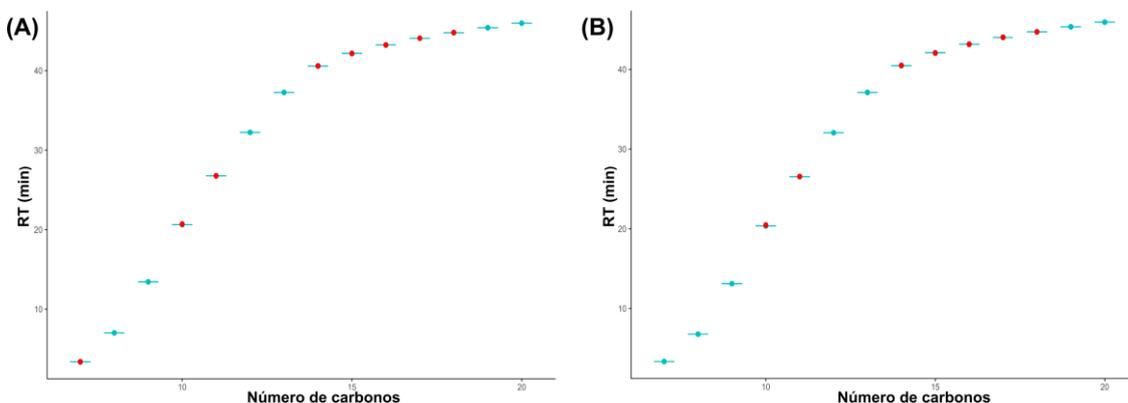


Figura 8. Tiempos de retención de los alcanos lineales en muestras de aire exhalado y en un estándar comercial. (A) Tiempos de retención de alcanos lineales determinados en muestras de aire exhalado (rojo) y tiempos de retención de alcanos lineales observados en el estándar comercial (azul) en la columna 1. (B) Tiempos de retención de alcanos lineales obtenidos en muestras de aire exhalado (rojo) y tiempos de retención de alcanos lineales observados en el estándar comercial (azul) en la columna 2.

Comparación con la biblioteca NIST

Por último, se llevó a cabo un protocolo para comparar cada compuesto filtrado con la biblioteca NIST. En primer lugar, se estableció un umbral por encima del 80% de *match factor* para seleccionar las entradas de la biblioteca NIST. A continuación, se seleccionaron todas aquellas entradas que tenían un *match factor* hasta un 5% menos del mayor hallado para ese compuesto filtrado (siempre superior al 80%), extrayéndose un máximo de 450 entradas para cada compuesto filtrado. A continuación, se calculó el RI de cada compuesto filtrado mediante *eRah* utilizando el RT de los estándares de referencia (10 compuestos del estándar de calibración de VOCs y los alcanos lineales). Además, se calculó el *error RI* para todas las entradas extraídas. Además, se descartaron las entradas con un *error RI* superior al 20%. Finalmente, se consideró la entrada con mayor grado de semejanza con un compuesto filtrado aquella con el menor *error RI* y el mayor *factor match* para ese compuesto filtrado. Sin embargo, si no se disponía de ningún valor de *error RI* inferior al 20%, se consideró que la entrada con mayor similitud era aquella que poseía un mayor *match factor*. Como se ha mencionado anteriormente, el RI de todas las entradas de la biblioteca NIST no está disponible. En

el Apéndice 7 se muestra que porcentaje de compuestos filtrados para los que se llevó a cabo el emparejamiento con la biblioteca NIST en base a ambos parámetros (*error RI* y *match factor*).

Análisis exploratorio de los *features* filtrados

Tras la aplicación del flujo de trabajo diseñado sobre los datos obtenidos, se llevó a cabo un análisis exploratorio de los *features* filtrados realizando un análisis de componentes principales (PCA) en ambos grupos de muestras (Grupo 1 y 2). En la Figura 9 se muestran los gráficos de puntuaciones del PCA (PC-1 vs. PC-2).

El componente que representa el máximo porcentaje de la varianza total (PC1), discriminaba principalmente las muestras de aire exhalado humano y las muestras del aire ambiental en ambos grupos de muestras (Grupo 1 y Grupo 2). Además, el PCA también se aplicó sobre todos los *features* detectados por *xcms* (Figura 10).

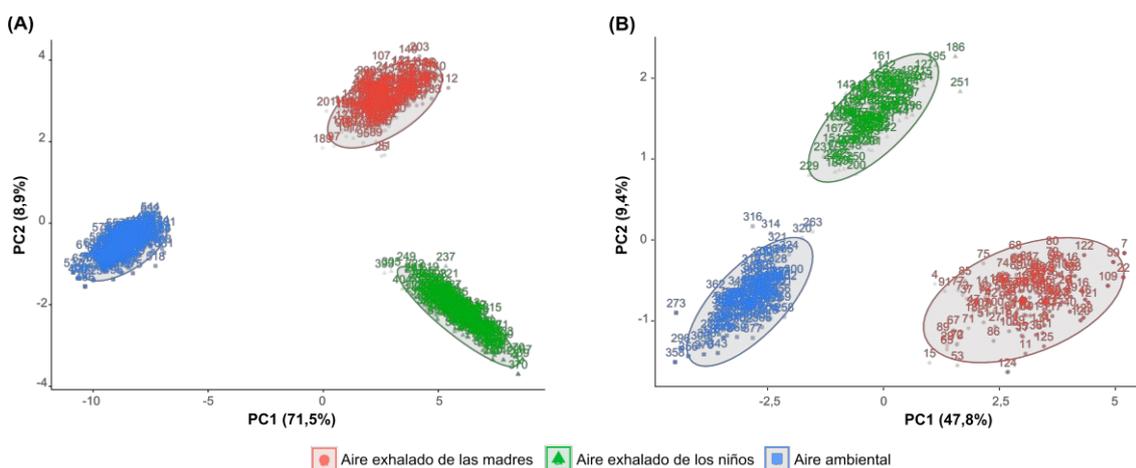


Figura 9. Análisis exploratorio de los *features* filtrados mediante el análisis de componentes principales (PCA). (A) Gráfico de puntuaciones PCA (PC-1 vs. PC-2) de las muestras de aire exhalado del Grupo 1. (B) Gráfico de puntuaciones PCA (PC-1 vs. PC-2) de las muestras de aire exhalado del Grupo 2. PC: componente principal.

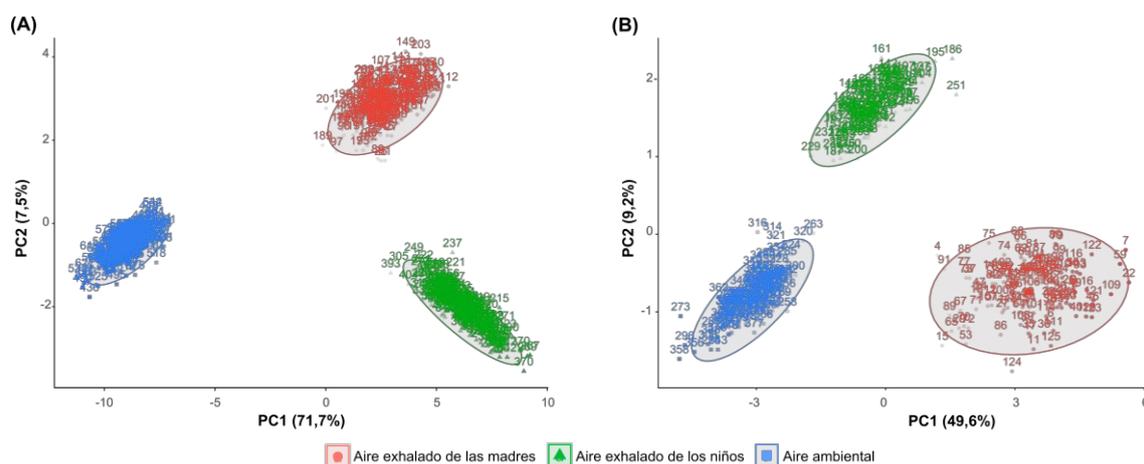


Figura 10. Análisis exploratorio de los *features* obtenidos por la estrategia 1 (paquete *xcms*) mediante el análisis de componentes principales (PCA). (A) Gráfico de puntuaciones PCA (PC-1 vs. PC-2) de las muestras de aire exhalado del Grupo 1. (B) Gráfico de puntuaciones PCA (PC-1 vs. PC-2) de las muestras de aire exhalado del Grupo 2. PC: componente principal.

DISCUSIÓN

El preprocesamiento de los datos obtenidos por espectrometría de masas constituye un cuello de botella del análisis del volatilo exhalado, que es preciso abordar con éxito en la fase de descubrimiento de biomarcadores. De hecho, la estandarización de la transformación digital de los datos obtenidos tras el análisis del aire exhalado es esencial para hacer progresar este campo^{12,14,26,40}. Por lo tanto, este trabajo presenta, por primera vez, un flujo de trabajo para el preprocesamiento de los datos obtenidos en el análisis de muestras de aire exhalado humano utilizando TD-GC/q-MS, que integra las dos estrategias más importantes en la actualidad. Aunque ambas estrategias pretenden alcanzar el mismo objetivo, es decir, identificar los compuestos presentes en las muestras, cada estrategia sigue un procedimiento y una secuencia temporal diferentes. Mientras que en la primera estrategia se determinan primero los *features* y luego éstos se agrupan en compuestos, en la segunda estrategia se buscan primero los compuestos, y luego se determina el perfil iónico de los mismos.

Además, en la mayoría de los estudios anteriores se ha recurrido a *software* privativo en el que el código no se encuentra disponible y el usuario dispone de escasas opciones de configuración. Sin embargo, para conseguir un procedimiento reproducible es imprescindible contar con una descripción detallada de los flujos de trabajo empleados^{41,42}. En consecuencia, otra ventaja del flujo de trabajo aquí descrito es el hecho de que la búsqueda de los compuestos presentes en las muestras de aire exhalado se efectúa mediante soluciones de código abierto. Por tanto, se usaron tres

paquetes basados en el lenguaje de programación R, así, para la primera estrategia se empleó *xcms*²³ y *cliqueMS*²⁸ y para la segunda se utilizó *eRah*²⁷. Además, *cliqueMS* es un paquete recientemente desarrollado que fue especialmente diseñado para datos obtenidos mediante LC/MS (cromatografía líquida-espectrometría de masas)²⁸ y este trabajo demuestra, por primera vez, que *cliqueMS* también puede utilizarse con éxito sobre datos obtenidos por TD-GC/q-MS.

Un reto importante cuando se trabaja con un gran volumen de datos es la correcta diferenciación entre los verdaderos compuestos y los compuestos duplicados para evitar errores de identificación. En este sentido, se observó un alto porcentaje de posibles compuestos duplicados en los resultados de cada estrategia (Figura 6). La duplicación fue especialmente elevada en la primera estrategia, puesto que *cliqueMS* agrupa los *features* muestra por muestra. Por lo tanto, se llevó a cabo la combinación de las dos estrategias mediante una doble validación basada en el cálculo de un producto escalar. Únicamente se seleccionaron los compuestos detectados por ambas estrategias (producto escalar superior a 0,7) y de este modo, se consiguió reducir los compuestos duplicados y la aparición de errores durante la identificación. Además, la detección de un compuesto por ambas estrategias es una evidencia sólida que constata su presencia en una muestra de aire exhalado, de hecho, el porcentaje de posibles compuestos duplicados disminuyó significativamente cuando se unieron las dos estrategias (Figure 6). Este proceso demostró que la detección de compuestos también mediante *cliqueMS* fue útil para lograr una lista filtrada de compuestos obtenidos por *eRah*. Además, la conexión entre métodos permitió la determinación con gran precisión de la intensidad de los *features* empleando *xcms*, y la comparación de los compuestos detectados con una biblioteca espectral usando *eRah*. En resumen, la integración de las dos estrategias de preprocesamiento de datos refuerza la validez de los datos ya que, en esencia, efectúa comprobaciones cruzadas de los datos brutos y reduce los errores al eliminar los datos menos fiables.

Como resultado de este flujo de trabajo, se obtuvo una matriz con las intensidades normalizadas de los *features* filtrados. Esta matriz puede utilizarse como elemento de entrada para el análisis de datos. Como puede observarse, el PCA realizado sobre los *features* obtenidos con *xcms* (Figura 10) y el PCA con los *features* filtrados (Figura 9) fueron muy similares. Por lo tanto, la filtración de los *features* no conlleva la pérdida de datos valiosos. Además, gracias a la unión de ambas estrategias, fue posible determinar los *features* que pertenecían a cada compuesto.

Asimismo, se propuso un procedimiento para la identificación de compuestos filtrados. Estos compuestos fueron comparados con la biblioteca espectral NIST y se calcularon dos valores: *match factor* (proporciona información sobre la similitud entre los espectros) y el *error RI* (proporciona información sobre la similitud entre los tiempos de retención). Los tiempos de retención de los compuestos de los estándares comerciales de calibración de VOCs y de alcanos lineales se utilizaron como compuestos de referencia para calcular el valor de *error RI*. Adicionalmente, se pudo estudiar la evolución de los tiempos de retención a lo largo del muestreo del aire exhalado mediante la determinación del tiempo de retención de compuestos ubicuos como los alcanos lineales. De este modo, hemos demostrado, que estos compuestos ubicuos de las muestras de aire exhalado pueden ser usados como estándares internos para el cálculo del RI.

Por otra parte, hasta donde sabemos, es la primera vez que se recogen muestras de aire exhalado de niños menores de 2 años para su posterior análisis mediante GC-MS. La simplicidad es fundamental en pacientes pasivos como la población pediátrica. Por este motivo, se apostó por el muestreo del aire exhalado mixto junto con la recogida de muestras de aire ambiental. En el muestreo del aire exhalado mixto, se recoge todo el aire exhalado durante la respiración (aire del espacio muerto, aire de las vías respiratorias y aire alveolar) sin excluir ninguna fase de la respiración. Por este motivo, el muestreo del aire exhalado mixto ha sido tradicionalmente cuestionado por su posible contaminación por fuentes exógenas¹⁸. Sin embargo, para solventar esta limitación, se recogió, al mismo tiempo, una muestra del aire de la habitación donde se tomaban las muestras de aire exhalado. Este protocolo permitió identificar los VOCs presentes en el ambiente, y minimizar la interferencia del entorno comparando las muestras de aliento humano y las muestras de aire ambiental. Además, los resultados obtenidos del PCA muestran que existían diferencias entre los tres subconjuntos de muestras (aire exhalado de las madres, aire exhalado de los niños y aire ambiental) y que la diferencia era mayor entre las muestras de aire exhalado humano y el aire procedente del ambiente que entre las muestras de aire exhalado humano (Figure 9). Por lo tanto, las muestras de los sujetos incluían VOCs emanados del aire alveolar y no sólo VOCs del aire ambiental.

Limitaciones

Lamentablemente, a día de hoy, la biblioteca NIST cuenta con muy pocas entradas con RI disponibles. En consecuencia, el valor de *error RI* no puede ser calculado para muchos compuestos. Además, varios compuestos filtrados muestran

valores de *match factor* muy próximos para un gran número de compuestos de la biblioteca del NIST. Por lo tanto, hay más de 450 entradas en la biblioteca del NIST con un valor de *match factor* entre el valor máximo de *match factor* encontrado para ese compuesto y el valor máximo de *match factor* encontrado para ese compuesto menos el 5%. Por otra parte, en la mayoría de los compuestos filtrados sólo se ha podido realizar una anotación putativa, en vez de una identificación verdadera⁴². En consecuencia, después de seleccionar los compuestos de interés en el análisis estadístico posterior, es muy aconsejable confirmar las identidades de estos analitos utilizando un estándar de referencia.

CONCLUSIÓN

En los estudios de análisis del volatiloма exhalado se obtiene un gran volumen de datos destinados a la búsqueda de biomarcadores. Los datos brutos deben ser preprocesados correctamente para poder aplicar un análisis estadístico con el fin de generar conclusiones útiles y contundentes. A pesar de la importancia del preprocesamiento de datos, en la mayoría de los estudios se siguen utilizando programas privativos, que a veces tienen poco interés para determinadas aplicaciones porque no permiten su adaptación ni manipulación.

En este trabajo se muestra por primera vez un flujo de trabajo para el preprocesamiento de datos brutos obtenidos por TD-GC/q-MS. En él se implementan y cruzan tres paquetes escritos en lenguaje de programación R de fuente libre: *xcms*, *cliqueMS* y *eRah*, en los que se han optimizado los parámetros determinantes para cada uno de ellos. Este flujo de trabajo combina de forma adecuada y sencilla dos estrategias (detección de los *features* de la muestra y detección de los compuestos de la muestra) para obtener una matriz útil para el análisis de datos con intensidades normalizadas de los *features* de las muestras de aire exhalado. En este sentido, la integración de ambas estrategias permite reducir la duplicación de compuestos y obtener una lista única de *features* filtrados y compuestos filtrados en poco tiempo (2 min por muestra). Además, en este trabajo se demuestra claramente, gracias a la integración de las dos estrategias, que es posible identificar qué *features* pertenecen a cada compuesto, y cotejar estos compuestos con los compuestos de una biblioteca espectral, como la biblioteca NIST. En el futuro, este proceso podría simplificarse con el avance de herramientas bioinformáticas gratuitas y con el aumento de los valores disponibles de RI de los compuestos de las colecciones espectrales.

Por todo ello, el flujo de trabajo aquí presentado representa una herramienta muy útil para la implementación del análisis del volatiloма como un método no invasivo para la búsqueda de biomarcadores. Ello será de vital importancia para el desarrollo de estrategias de diagnóstico y seguimiento de enfermedades como el asma, las alergias, el cáncer, etc.

Por último, en este trabajo se muestra un protocolo reproducible para la toma de muestras de aire exhalado adaptado para bebés de 3 meses.

MATERIALES Y MÉTODOS

Población del estudio

Esta investigación se enmarca en el estudio de cohorte prospectiva de nacimiento NELA (*Nutrition in Early Life and Asthma*)³³, una cohorte de nacimiento creada en 2015 en España, donde se reclutaron parejas madre-hijo en la semana 20 de embarazo. El estudio fue aprobado por el Comité Ético de Investigación Clínica (CEIC) del Hospital Universitario Virgen de la Arrixaca de Murcia (España). Todos los métodos se llevaron a cabo de acuerdo con las directrices y regulaciones pertinentes. Se obtuvo el consentimiento informado de todos los sujetos mayores de 18 años, y de un progenitor y/o tutor legal en el caso de los sujetos menores de edad. Los criterios de inclusión fueron: residencia habitual en el área de salud I y en algunos distritos de las áreas VI y VII de la Región de Murcia (España); origen caucásico; edad de la madre entre 18 y 45 años; embarazo único; concepción no asistida; y ecografía normal a las 20 semanas de gestación (sin malformaciones importantes). El estudio NELA pretende llevar a cabo un seguimiento de los sujetos reclutados en las siguientes etapas: 1) a las 20 semanas de embarazo, 2) a las 32 semanas de embarazo, 3) en el parto, 4) a los 3 meses de edad del niño, 5) a los 18 meses, 6) a los 5 años y 7) a los 7 años. Durante la cuarta vista de seguimiento, se recogieron muestras del aire exhalado de niños de 3 meses y de sus madres. Los sujetos fueron distribuidos en dos grupos: Grupo 1 (211 parejas madre-hijo) y Grupo 2 (126 parejas madre-hijo). Las muestras de aire exhalado se recogieron en el Grupo 1 desde mayo de 2017 hasta febrero de 2018, y en el Grupo 2 desde marzo de 2018 hasta octubre de 2018.

Recogida de muestras de aire exhalado

Las muestras de aliento exhalado de los niños de 3 meses y de las madres del estudio NELA fueron recogidas utilizando un protocolo no invasivo inspirado en la

metodología utilizada por Van der Kant *et al.*⁴³. En concreto, se llevó a cabo un muestreo del aire exhalado mixto (aire del espacio muerto, aire de las vías respiratorias y aire alveolar)¹⁸ tanto en las madres como en los niños. Los bebés, al ser sujetos pasivos no pueden llevar a cabo una espiración forzada durante la toma de muestra para recoger sólo el aire alveolar. Por lo tanto, el muestreo de aire exhalado de las madres también se realizó sin espiración forzada. En ambos casos, las muestras de aire exhalado son una mezcla de gases (aire exhalado mixto). En el caso de las madres, las muestras de aire exhalado fueron recogidas en bolsas de muestreo de gas Tedlar® de 1 L. Para los niños, sin embargo, el proceso se realizó en dos pasos: en primer lugar, utilizando una mascarilla facial con almohadilla, se llenaron dos bolsas de muestreo de gas QuinTron® de 400 mL, las cuales presentan una menor resistencia al llenado; y luego el aire de estas muestras se transfirió a una bolsa Tedlar®. Finalmente, el aire contenido en las bolsas de muestreo se transfirió a tubos de desorción térmica (Tenax TA/carbograph 5td, Markes International) para retener y preconcentrar los VOCs. Además, se recogió también una muestra del aire de la sala para evaluar los posibles contaminantes ambientales, utilizando una jeringa Easy-VOCS (Markes International) que permitía tomar muestras de aire directamente en un tercer tubo de desorción térmica. Todos los tubos TD se sellaron inmediatamente con tapones de latón provistos de unas férulas de PTFE y se almacenaron hasta su análisis. Si el período de almacenamiento era inferior a 24 horas, los tubos TD se almacenaban a temperatura ambiente (25 °C). En una investigación reciente se recomienda que las muestras de aliento exhalado se conserven a baja temperatura (4 °C) para lograr la máxima estabilidad durante el almacenamiento⁴⁴. Por lo tanto, si el período de almacenamiento era superior a 24 h, los tubos se almacenaron a 4 °C. El tiempo transcurrido desde la toma de muestra hasta la posterior desorción térmica fue siempre inferior a una semana. Las bolsas Tedlar® fueron reutilizadas 10 veces, limpiándose después de cada uso lavándolas 5 veces con gas nitrógeno (99,9% de pureza). Sin embargo, las bolsas QuinTron® no fueron reutilizadas y se desecharon tras su uso. Cualquier material que entrara en contacto con el aire exhalado de la madre o el niño, como los tubos de plástico o la mascarilla, era limpiado rociándolo con una solución de etanol al 70%.

Análisis del aire exhalado mediante TD-GC/q-MS

El análisis de las muestras de aire exhalado se llevó a cabo mediante un sistema de desorción térmica acoplado a la cromatografía de gases-espectrometría de masas, utilizando helio como gas portador. Los tubos que contenían las muestras se sometieron a un proceso de desorción térmica (TD) de dos pasos (UltraTD Multi-Tube Autosampler

y Unity Thermal Desorber with Cold Trap, Markes Int. Ltd). En el UltraTD, los tubos TD se calentaron a 300 °C durante 10 minutos para desorber todo el aire almacenado en su interior. Los VOCs de las muestras fueron arrastrados por el gas portador que fluía a 50 mL/min y retenidos en la trampa fría del Unity (-10 °C). A continuación, la trampa fría era calentada desde -10 °C hasta 300 °C a 100 °C/s, y la temperatura final se mantenía durante 10 minutos. Posteriormente, los VOCs desorbidos fueron transportados por el gas helio e inyectados directamente en el GC/MS. El sistema utilizado para el análisis de la muestra fue un espectrómetro de masas de cuadrupolo simple 5977B (q-MS) acoplado a un cromatógrafo de gases (GC) 7890B de Agilent Technologies con una columna capilar ultra inerte HP-5 ms (30 m, 0,25 mm de diámetro interior, 0,25 µm de espesor de película). Como fuente de iones se empleó un sistema de ionización de electrones (70 eV). La adquisición de datos de MS se realizó utilizando el modo de barrido completo (rango de barrido de 35-350 AMU). La separación de los compuestos en el GC se realizó programando tres rampas de temperatura: 1) 35 °C durante 10 min, 2) aumento de 3 °C/min hasta 121 °C, 3) aumento de 20 °C/min hasta 270 °C, y finalmente 270 °C durante 5 min y aumento de 30 °C/min hasta 300 °C. Los tubos TD utilizados se sometieron a un proceso de limpieza, previamente optimizado, que consistió en pasar helio a un flujo de 20 mL/min durante 1 min, y luego los tubos se calentaron a 335 °C en UltraTD durante 25 min. Este proceso se repitió dos veces. Aunque las muestras de aire exhalado de los grupos 1 y 2 se midieron con la misma metodología, se utilizaron diferentes columnas del mismo modelo para analizar cada grupo. Dos tipos de estándares químicos de Sigma-Aldrich (n-alcanos saturados C7-C30 y calibración de VOCs) fueron cargados en tubos de desorción térmica y analizados utilizando el mismo método.

Análisis estadístico

Tanto el preprocesamiento de los datos como el análisis estadístico se realizaron con la versión 3.5.2 de R, en un ordenador Intel Core i9 de 3,6 GHz. Se emplearon las pruebas de Shapiro-Wilk y Lilliefors (paquete *nortest*) para comprobar la distribución normal de los datos. Dependiendo de la distribución de los datos, se realizaron pruebas paramétricas (prueba T de Student o prueba ANOVA seguida de pruebas *post-hoc* de Bonferroni) o pruebas no paramétricas (prueba U de Mann-Whitney o prueba de Kruskal Wallis) para comprobar si había diferencias significativas (p-valor < 0,05) en las variables continuas. Además, se realizó un análisis exploratorio mediante el análisis de componentes principales (PCA) utilizando el paquete *FactoMineR*⁴⁵ para observar las diferencias entre las muestras del aire exhalado humano y las muestras de aire

ambiental en ambos grupos de muestras (Grupo 1 y Grupo 2). Los valores de intensidad relativa de los *features* filtrados de los *features* totales obtenidos por *xcms* se utilizaron como datos de entrada para el PCA. Para evitar la introducción de sesgos sistemáticos debidos a la metodología de muestreo de aire exhalado, los *features* de los contaminantes de las bolsas de muestreo (*features* del N,N-dimetilacetamida y del fenol) no fueron tenidos en cuenta en el PCA.

BIBLIOGRAFÍA

1. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
2. Blanchet, L. *et al.* Factors that influence the volatile organic compound content in human breath. *J. Breath Res.* **11**, 016013 (2017).
3. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
4. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
5. Scheepers, P. T. J. *et al.* Determination of exposure to benzene, toluene and xylenes in Turkish primary school children by analysis of breath and by environmental passive sampling. *Sci. Total Environ.* **408**, 4863–4870 (2010).
6. Chen, X. *et al.* Association of Smoking with Metabolic Volatile Organic Compounds in Exhaled Breath. *Int. J. Mol. Sci.* **18**, 2235 (2017).
7. Capone, S. *et al.* Chromatographic analysis of VOC patterns in exhaled breath from smokers and nonsmokers. *Biomed. Chromatogr.* **32**, (2018).
8. Neerincx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
9. Alkhouri, N. *et al.* Analysis of breath volatile organic compounds as a noninvasive tool to diagnose nonalcoholic fatty liver disease in children. *Eur. J. Gastroenterol. Hepatol.* **26**, 82–7 (2014).
10. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
11. Boots, A. W., Bos, L. D., van der Schee, M. P., van Schooten, F. J. & Sterk, P. J. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. *Trends Mol. Med.* **21**, 633–644 (2015).
12. Sola Martínez, R. A. *et al.* Exhaled volatile organic compounds analysis in clinical pediatrics: a systematic review. *Pediatr. Res.* **89**, 1352–1363 (2021).
13. Mazzone, P. J. *et al.* Evaluating Molecular Biomarkers for the Early Detection of Lung Cancer: When Is a Biomarker Ready for Clinical Use? An Official American

- Thoracic Society Policy Statement. *Am. J. Respir. Crit. Care Med.* **196**, e15–e29 (2017).
14. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res.* **6**, 027108 (2012).
 15. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).
 16. Hauschild, A.-C. *et al.* Computational Methods for Metabolomic Data Analysis of Ion Mobility Spectrometry Data—Reviewing the State of the Art. *Metabolites* **2**, 733–755 (2012).
 17. Brekelmans, M. P. *et al.* Smelling the Diagnosis: The Electronic Nose as Diagnostic Tool in Inflammatory Arthritis. A Case-Reference Study. *PLoS One* **11**, e0151715 (2016).
 18. Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R. & Fowler, S. J. Exhaled breath analysis: a review of ‘breath-taking’ methods for off-line analysis. *Metabolomics* **13**, 110 (2017).
 19. Matthiesen, R. & Bunkenborg, J. Introduction to Mass Spectrometry-Based Proteomics. in *Mass Spectrometry Data Analysis in Proteomics* (ed. Matthiesen, R.) **2051**, 1–58 (Humana, New York, NY, 2020).
 20. Beale, D. J. *et al.* Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics* **14**, 152 (2018).
 21. Li, Y. *et al.* A novel approach to transforming a non-targeted metabolic profiling method to a pseudo-targeted method using the retention time locking gas chromatography/mass spectrometry-selected ions monitoring. *J. Chromatogr. A* **1255**, 228–236 (2012).
 22. Morimoto, J. *et al.* Untargeted/targeted 2D gas chromatography/mass spectrometry detection of the total volatile tea metabolome. *Molecules* **24**, (2019).
 23. Smith, C. A., Want, E. J., O’Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
 24. Smolinska, A. *et al.* Current breathomics—a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J. Breath Res.* **8**, 027105 (2014).
 25. Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D. & Lu, L. J. Computational and statistical analysis of metabolomics data. *Metabolomics* **11**, 1492–1513

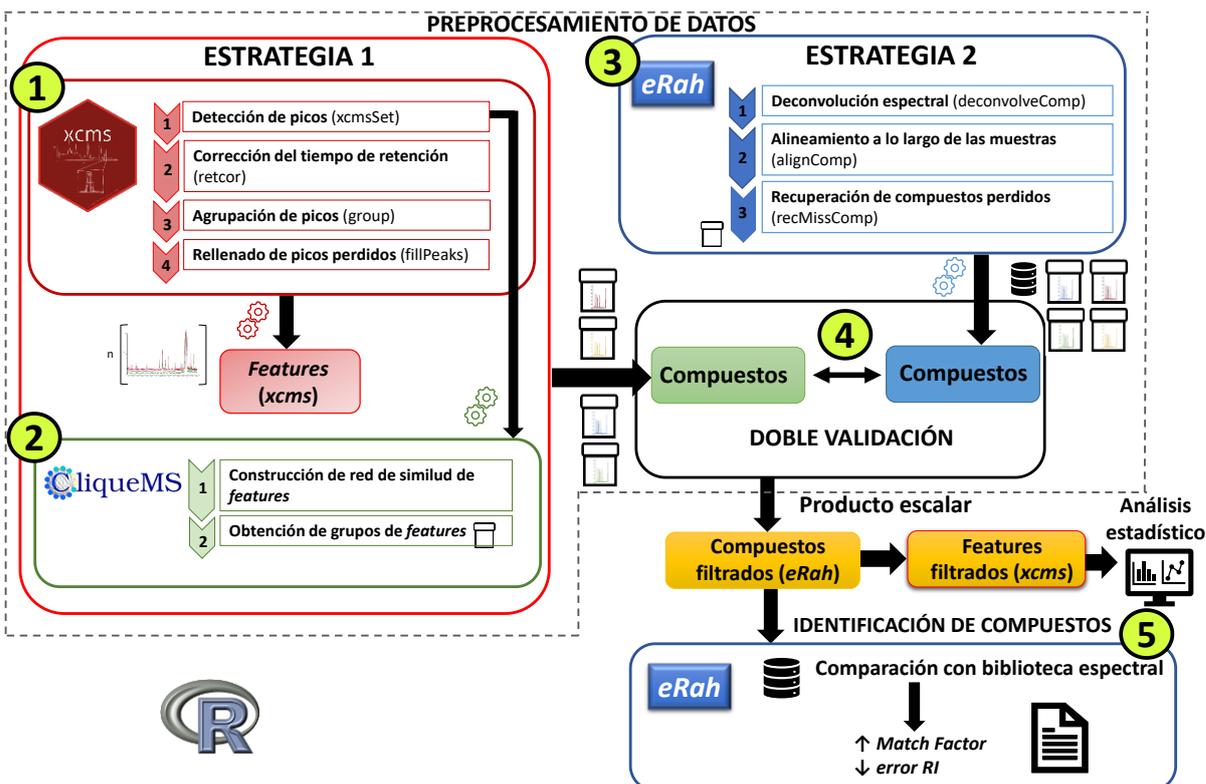
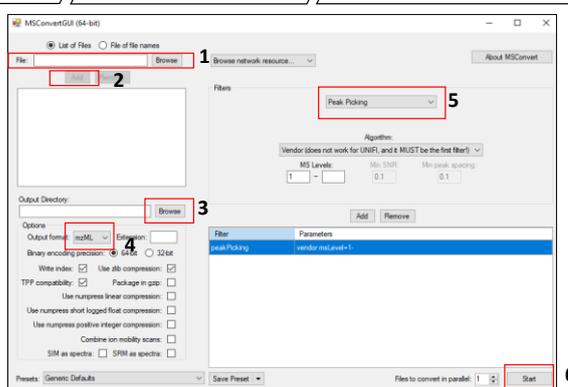
- (2015).
26. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
 27. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
 28. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
 29. Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R. & Neumann, S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* **84**, 283–289 (2012).
 30. Smirnov, A. *et al.* ADAP-GC 4.0: Application of Clustering-Assisted Multivariate Curve Resolution to Spectral Deconvolution of Gas Chromatography-Mass Spectrometry Metabolomics Data. *Anal. Chem.* **91**, 9069–9077 (2019).
 31. Stavropoulos, G., Salman, D., Alkhalifah, Y., van Schooten, F.-J. & Smolinska, A. Preprocessing and analysis of volatilome data. in *Breathborne Biomarkers and the Human Volatilome* (eds. Beauchamp, J., Davis, C. & Pleil, J.) 633–647 (Elsevier, 2020).
 32. Alkhalifah, Y. *et al.* VOCCluster: Untargeted Metabolomics Feature Clustering Approach for Clinical Breath Gas Chromatography/Mass Spectrometry Data. *Anal. Chem.* **92**, 2937–2945 (2020).
 33. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).
 34. Martens, L. *et al.* mzML--a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133 (2011).
 35. Adusumilli, R. & Mallick, P. Data conversion with proteoWizard msConvert. in *Proteomics: Methods and Protocols* (eds. Comai, U., Katz, J. E. & Mallick, P.) **1550**, 339–368 (Humana, New York, NY, 2017).
 36. Domingo-Almenara, X. An introduction to erah package. *J. Chromatogr. A* **1409**, 226–233 (2015).
 37. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a

- headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–80 (2011).
38. Domingo-Almenara, X. & Siuzdak, G. Metabolomics Data Processing Using XCMS. *Methods Mol. Biol.* **2104**, 11–24 (2020).
 39. Ogawa, T. *et al.* Exploration of polar lipid accumulation profiles in *Euglena gracilis* using LipidBlast, an MS/MS spectral library constructed *in silico*. *Biosci. Biotechnol. Biochem.* **78**, 14–18 (2014).
 40. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
 41. Goodacre, R. *et al.* Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**, 231–241 (2007).
 42. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221 (2007).
 43. van de Kant, K. D. G. *et al.* Exhaled breath profiling in diagnosing wheezy preschool children. *Eur. Respir. J.* **41**, 183–188 (2013).
 44. Harshman, S. W. *et al.* Storage stability of exhaled breath on Tenax TA. *J. Breath Res.* **10**, (2016).
 45. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

APÉNDICE 1. Breve tutorial del flujo de trabajo de preprocesamiento de datos para el análisis de VOCs por TD-GC/q-MS (desorción térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple). Código R de las principales funciones.



0 Conversión de datos en formato estándar unificado (mzXML)



1 ESTRATEGIA 1 (paquete *xcms*)



1 Detección de picos (*xcmsSet*)

```
bpparam <- BiocParallel::SnowParam(workers = 5)
xcms.out1 <- xcmsSet(files = datos de entrada, method = "matchedFilter", fwh=3,
                    max = 1000, snthresh = 5, step = 0.5, steps = 1, mzdiff= -0.15,
                    BPPARAM = bpparam)
```

2 Corrección del tiempo de retención (*retcor*)

```
xcms.out2 <- retcor(xcms.out1, method="obiwarp",profStep=1)
```

3 Agrupación de picos (*group*)

```
xcms.out3 <- group(xcms.out2,mzwid=0.15,minsamp=10,bw=#)
```

4 Rellenado de picos perdidos (*fillPeaks*)

```
xcms.out4 <- fillPeaks(xcms.out3,BPPARAM = bpparam)
```

Buscar el valor de cada parámetro en el Apéndice 2

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).

2 ESTRATEGIA 1 (paquete *cliqueMS*)



1 Construcción de red de similitud de *features*

2 Obtención de grupos de *features*

```
cliqueMS.out1 <- getCliques(xcms.out1[,i], filter = TRUE, mzeror=0.1,
                          rtdiff=0.1)
```

```
cliqueMS.out2 <- getIsotopes(cliqueMS.out1, ppm = 500)
```

```
cliqueMS.out3 <- cliqueMS.out2$peaklist
```

cliqueMS realiza la anotación muestra a muestra

Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).

3 ESTRATEGIA 2 (paquete *eRah*)



```
erah.out <- newExp(instrumental = "inst.csv", phenotype = "pheno.csv")
```

1 Deconvolución espectral (*deconvolveComp*)

```
dec.par <- setDecPar(min.peak.width = #, avoid.processing.mz = c(0),
                    analysis.time = #)
```

```
erah.out <- deconvolveComp(erah.out, dec.par)
```

2 Alineamiento a lo largo de las muestras (*alignComp*)

```
al.par <- setAlPar(min.spectra.cor = 0.85, max.time.dist = #,mz.range = c(38:400))
```

```
erah.out <- alignComp(erah.out, alParameters = al.par, blocks.size = 55)
```

3 Recuperación de compuestos perdidos (*recMissComp*)

```
erah.out <- recMissComp(erah.out, min.samples = 10)
```

Buscar el valor de cada parámetro en el Apéndice 2

Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).

Domingo-Almenara, X. *An introduction to erah package.* <https://cran.r-project.org/web/packages/erah/vignettes/eRahManual.pdf> (2020).

4 DOBLE VALIDACIÓN

```
dotproduct <- function(x,y, from=38, to = 450){
  mzrange <- seq(from =from, to = to, by=1)
  x2 <- sapply(mzrange,function(i){
    dif <- abs(round(x[,1])-i)
    res <- 0
    if(any(dif<0.5)){
      res <- x[which(dif==min(dif)),2]
    }
    return(res)
  })
  y2 <- sapply(mzrange,function(i){
    dif <- abs(round(y[,1])-i)
    res <- 0
    if(any(dif<0.5)){
      res <- y[ which(dif==min(dif)) ,2]
    }
    return(res)
  })
  as.vector((x2 %*% y2) / (sqrt(sum(x2^2)) * sqrt(sum(y2^2))))
}
```

x: vector con las intensidades
relativas de los compuestos obtenidos
con *xcms-cliqueMS*

y: vector con las intensidades
relativas de los compuestos obtenidos
con *eRah*

5 IDENTIFICACIÓN DE COMPUESTOS (paquete *eRah*)



Comparación con la biblioteca espectral y cálculo del *Match factor*

```
ident.out <- identifyComp(erah.out, id.database = NIST_library, n.putative = 450)
```

Cálculo del *error RI*

```
errorRI.out <- computeRIerror(ident.out, id.database = NIST_library, reference.list =  
referenceCompounds, ri.error.type = "relative", plot.results = F)
```

Los alcanos lineales y los compuestos del estándar comercial de VOCs han sido empleados como compuestos de referencia (Apéndice 5 y Apéndice 6)

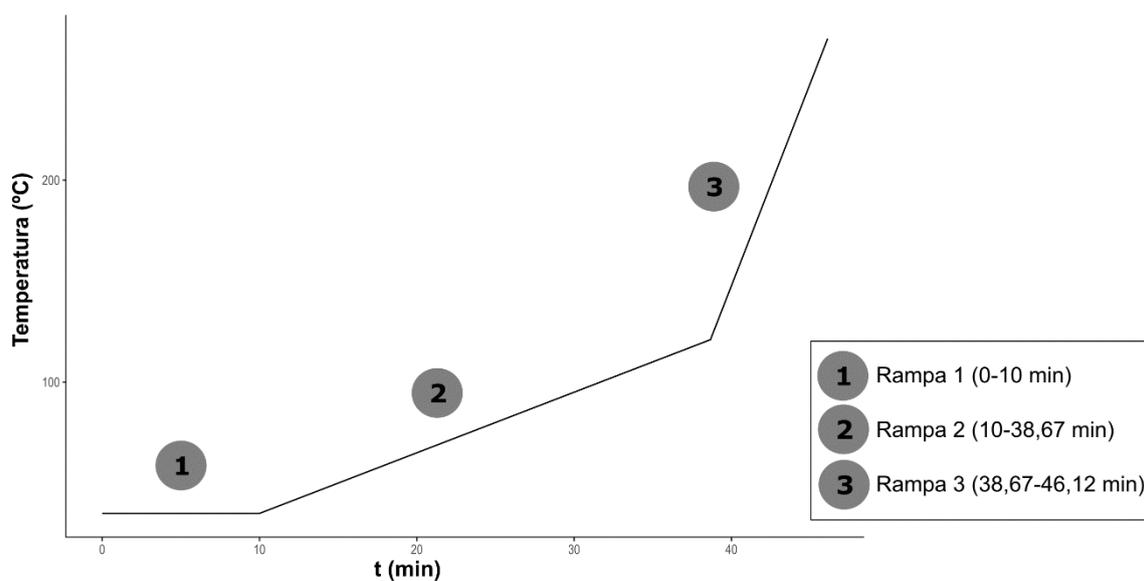
Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
Domingo-Almenara, X. *An introduction to erah package.* <https://cran.r-project.org/web/packages/erah/vignettes/eRahManual.pdf> (2020).

APÉNDICE 2. Tiempo empleado en el preprocesamiento de los datos y la identificación de los compuestos de las muestras de aire exhalado del Grupo 1 y del Grupo 2

Muestras	Fases del flujo de trabajo	Tiempo empleado en cada fase	Tiempo total empleado
Muestra de aire exhalado de las madres del Grupo 1	PREPROCESAMIENTO DE DATOS	6 h 42 min	7 h 37 min (2 min 10 s / muestra)
	ESTRATEGIA 1	55 min (51 min <i>xcms</i> y 4 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	5 h 54 min	
	DOBLE VALIDACIÓN	2 min	
	IDENTIFICACIÓN DE COMPUESTOS	55 min	
Muestras de aire exhalado de los niños del Grupo 1	PREPROCESAMIENTO DE DATOS	6 h 17 min	7 h 09 min (2 min 01 s / muestra)
	ESTRATEGIA 1	1 h (55 min <i>xcms</i> y 5 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	5 h 16 min	
	DOBLE VALIDACIÓN	1 min	
	IDENTIFICACIÓN DE COMPUESTOS	52 min	
Muestras de aire Ambiental del Grupo 1	PREPROCESAMIENTO DE DATOS	7 h 13 min	8 h 10 min (2 min 20 s / muestra)
	ESTRATEGIA 1	1 h (54 min <i>xcms</i> y 6 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	6 h 9 min	
	DOBLE VALIDACIÓN	4 min	
	IDENTIFICACIÓN DE COMPUESTOS	57 min	
Muestra de aire exhalado de las madres del Grupo 2	PREPROCESAMIENTO DE DATOS	3 h 7 min	3 h 58 min (1 min 53 s / muestra)
	ESTRATEGIA 1	38 min (33 min <i>xcms</i> y 5 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	2 h 28 min	
	DOBLE VALIDACIÓN	1 min	
	IDENTIFICACIÓN DE COMPUESTOS	51 min	
Muestras de aire exhalado de los niños del Grupo 2	PREPROCESAMIENTO DE DATOS	2 h 57 min	3 h 47 min (1 min 49 s / muestra)
	ESTRATEGIA 1	38 min (34 min <i>xcms</i> y 4 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	2 h 18 min	
	DOBLE VALIDACIÓN	1 min	
	IDENTIFICACIÓN DE COMPUESTOS	50 min	
Muestras de aire Ambiental del Grupo 2	PREPROCESAMIENTO DE DATOS	3 h 13 min	4 h 5 min (1 min 56 s / muestra)
	ESTRATEGIA 1	38 min (33 min <i>xcms</i> y 5 min <i>cliqueMS</i>)	
	ESTRATEGIA 2	2 h 32 min	
	DOBLE VALIDACIÓN	2 min	
	IDENTIFICACIÓN DE COMPUESTOS	52 min	

APÉNDICE 3. Parámetros de las funciones de los paquetes empleados en el preprocesamiento de datos.

	Parámetros en <i>xcms</i>	Parámetros en <i>cliqueMS</i>	Parámetros en <i>eRah</i>
Aire exhalado de las madres	<p>xcmsSet: method="matchedFilter", fwhm=3, max=1000, snthresh=5, step=0.5, steps=1, mzdiff= -0.15</p> <p>retcor: method="obiwarp", profStep=1</p> <p>group: mzwid=0.25, minsamp=10; bw=12 (first part) / 5 (second part)</p> <p>fillPeaks</p>	<p>getCliques: filter=TRUE, mzerror=0.1, rtdiff=0.1</p>	<p>Temperature Ramp 1</p> <p>deconvolveComp: min.peak.width=0.653, avoid.processing.mz=c(0), analysis.time=c(1,10), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=29.9, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 2</p> <p>deconvolveComp: min.peak.width=1.1, avoid.processing.mz=c(0), analysis.time=c(10,38.67) , noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 3</p> <p>deconvolveComp: min.peak.width=1, avoid.processing.mz=c(0), analysis.time=c(38.67,46.12), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p>
Aire exhalado de los niños	<p>xcmsSet: method="matchedFilter", fwhm=3, max=1000, snthresh=5, step=0.5, steps=1, mzdiff= -0.15</p> <p>retcor: method="obiwarp"; profStep=1</p> <p>group: mzwid=0.25, minsamp=10, bw=10 (first part) / 4 (second part)</p> <p>fillPeaks</p>	<p>getCliques: filter=TRUE, mzerror=0.1, rtdiff=0.1</p>	<p>Temperature Ramp 1</p> <p>deconvolveComp: min.peak.width=0.653, avoid.processing.mz=c(0), analysis.time=c(1,10), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=29.9, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 2</p> <p>deconvolveComp: min.peak.width=1.1, avoid.processing.mz=c(0), analysis.time=c(10,38.67), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 3</p> <p>deconvolveComp: min.peak.width=1, avoid.processing.mz=c(0), analysis.time=c(38.67,46.12), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p>
Aire ambiental	<p>xcmsSet: method="matchedFilter", fwhm=3, max=1000; snthresh=5, step=0.5, steps=1, mzdiff= -0.15</p> <p>retcor: method="obiwarp", profStep=1</p> <p>group: mzwid=0.25, minsamp=10; bw=10 (first part) / 4 (second part)</p> <p>fillPeaks</p>	<p>getCliques: filter=TRUE, mzerror=0.1, rtdiff=0.1</p>	<p>Temperature Ramp 1</p> <p>deconvolveComp: min.peak.width=0.653, avoid.processing.mz=c(0), analysis.time=c(1,10), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=29.9, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 2</p> <p>deconvolveComp: min.peak.width=1.1, avoid.processing.mz=c(0), analysis.time=c(10,38.67), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p> <p>Temperature Ramp 3</p> <p>deconvolveComp: min.peak.width=1, avoid.processing.mz=c(0), analysis.time=c(38.67,46.12), noise.threshold=500</p> <p>alignComp: min.spectra.cor=0.85, max.time.dist=7, mz.range=c(38:400)</p> <p>recMissComp: min.samples=10</p>



APÉNDICE 4. Rampas de temperatura aplicadas en el análisis de las muestras de aire exhalado mediante TD-GC/q-MS (desorción térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple).

APÉNDICE 5. Compuestos del estándar comercial de calibración VOCs con índice de retención registrado en la biblioteca NIST.

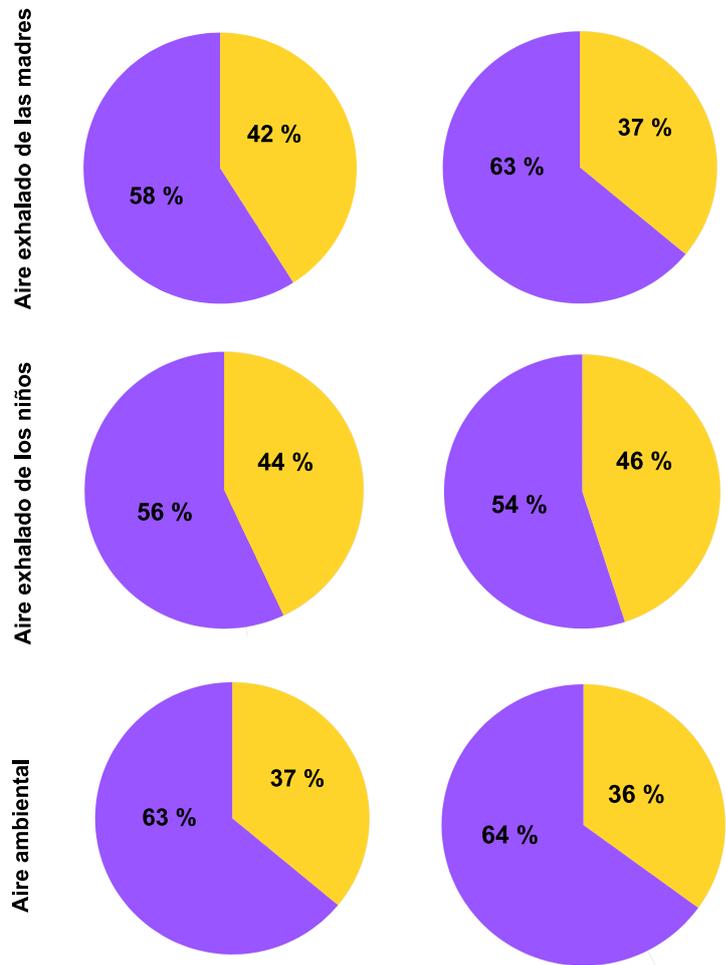
Nombre del compuesto	PubChem CID	RI disponible en biblioteca NIST	Columna 1. RT (min)	Columna 2. RT (min)
Cloruro de metileno	6344	391,00	1,84	1,79
Bromoclorometano	6333	835,00	3,49	3,41
Tolueno	1140	764,04	5,13	4,99
Tetracloroetano	31373	808,00	6,88	6,67
Clorobenceno	7964	844,00	9,08	8,80
Estireno	7501	892,00	12,66	12,50
o-Xileno	7237	891,64	12,80	12,33
n-Propilbenceno	7668	952,30	17,38	17,07
n-Butilbenceno	7705	1055,98	24,04	23,77
Naftaleno	931	1180,70	31,02	30,71

APÉNDICE 6. Tiempos de retención de alcanos lineales observados en cada columna tras el análisis del estándar comercial de n-alcanos saturados.

Alcanos lineales	PubChem CID	RI*	Columna 1. RT (min)	Columna 2. RT (min)
Heptano	8900	700	3,39	3,36
Octano	356	800	7,02	6,79
Nonano	8141	900	13,44	13,12
Decano	15600	1000	20,63	20,36
Undecano	14257	1100	26,78	26,55
Dodecano	8182	1200	32,23	32,05
Tridecano	12388	1300	37,26	37,11
Tetradecano	12389	1400	40,58	40,47
Pentadecano	12391	1500	42,18	42,09
Hexadecano	11006	1600	43,24	43,16
Heptadecano	12398	1700	44,08	44,01
Octadecano	11635	1800	44,77	44,71
Nonadecano	12401	1900	45,40	45,34
Eicosano	8222	2000	45,97	45,92

* Índice de retención de Kovats

Compuestos filtrados del Grupo 1 **Compuestos filtrados del Grupo 2**



■ Compuestos cotejados con la biblioteca NIST en base a los valores de *match factor* y *error RI*
■ Compuestos cotejados con la biblioteca NIST solamente en base del valor de *match factor*

APÉNDICE 7. Comparación de los compuestos filtrados y la biblioteca NIST.

Capítulo 5

Análisis del volatiloema exhalado como herramienta útil para discriminar el asma con otras enfermedades atópicas coexistentes en mujeres en edad fértil

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, Gema Lozano-Terol, Julia Gallego-Jara, Eva Morales, Esther Cantero-Cano, Manuel Sánchez-Solís, Luis García-Marcos, Pedro Jiménez-Guerrero, José A. Noguera-Velasco, Manuel Cánovas Díaz y Teresa de Diego Puente

Exhaled volatiloema analysis as a useful tool to discriminate asthma with other coexisting atopic diseases in women of childbearing age

Scientific Reports 11, 13823 (2021). doi: 10.1038/s41598-021-92933-2.

RESUMEN

La prevalencia del asma entre las mujeres en edad fértil es considerablemente elevada y la mayoría de las mujeres asmáticas suelen tener también otros trastornos atópicos. Por lo tanto, la diferenciación entre pacientes con enfermedades atópicas sin asma y asmáticas con enfermedades coexistentes es esencial para evitar el infradiagnóstico del asma, contribuir a diseñar estrategias para reducir la gravedad de los síntomas y mejorar la calidad de vida de las pacientes. Por ello, este trabajo tiene como objetivo aplicar, por primera vez, el análisis de los compuestos orgánicos volátiles en el aire exhalado de mujeres en edad fértil con el fin de discriminar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas), puesto que, en un futuro próximo, este abordaje podría ser una herramienta útil no invasiva para una detección y un seguimiento precisos del asma. En este estudio, las muestras de aire exhalado de 336 mujeres (conjunto de entrenamiento (n = 211) y conjunto de validación (n = 125)) fueron analizadas mediante un sistema de desorción térmica acoplado a cromatografía de gases-espectrometría de masas. Para el análisis de datos, se emplearon los algoritmos ASCA (ANOVA (análisis de varianza) - análisis de componentes simultáneos) y LASSO (operador de selección y contracción mínima absoluta) combinado con LR (regresión logística). De este modo, se generaron 15 modelos estadísticamente significativos (p-valor < 0,05 en las pruebas de permutación) que discriminaban el asma con otras enfermedades atópicas coexistentes en mujeres en edad fértil. En ellos, la acetona, el 2-etil-1-hexanol y un derivado de la tetrahydroisoquinolina fueron seleccionados como discriminantes del asma con otras enfermedades atópicas coexistentes. Además, se observó que el disulfuro de carbono, un derivado de la tetrahydroisoquinolina, el 2-etil-1-hexanol y el decano permitían discriminar la enfermedad del asma entre los pacientes con otros trastornos atópicos. Por todo ello, los resultados de este estudio muestran que el desarrollo de un sofisticado análisis metabolómico del aire exhalado permite discriminar el asma con otras enfermedades atópicas coexistentes en mujeres en edad reproductiva.

INTRODUCCIÓN

El asma es una enfermedad crónica que supone un enorme coste económico para los sistemas sanitarios de los países. Además, la prevalencia del asma en todo el mundo ha aumentado considerablemente en los últimos años¹. Estudios anteriores han sugerido que factores asociados a los cambios en el estilo de vida y a la exposición a contaminantes ambientales durante el embarazo influyen en el riesgo de padecer asma u otras afecciones atópicas durante la infancia^{2,3}. Las afecciones atópicas suelen incluir

enfermedades como la dermatitis atópica, el asma, la alergia alimentaria o la rinitis alérgica. Todas estas enfermedades están fuertemente asociadas entre sí, y la manifestación de una suele implicar la aparición de las otras^{4,5}. Por eso, el asma suele coexistir con otras enfermedades atópicas⁶. Hasta la fecha, muchos estudios han demostrado la progresión de los trastornos atópicos desde la dermatitis atópica en los lactantes hasta la rinitis alérgica y el asma en los niños⁷⁻⁹. Además, la prevalencia del asma difiere entre hombres y mujeres, así, mientras que durante la infancia el asma es más frecuente en los hombres, en edad adulta, la prevalencia es mayor en las mujeres. En concreto, la prevalencia es bastante elevada entre las mujeres en edad fértil, y los síntomas suelen ser más graves^{10,11}.

La inflamación de las vías respiratorias es la característica más típica del asma. El aumento del estrés oxidativo desempeña un papel importante en la inflamación de las vías respiratorias¹² y suele estar relacionado con el aumento de la producción de especies reactivas de oxígeno (ROS) o el mal funcionamiento de las defensas antioxidantes^{13,14}. La generación de ROS puede afectar al ADN, los lípidos, las proteínas y los hidratos de carbono¹⁵. Las ROS pueden proceder de fuentes endógenas, como consecuencia del metabolismo en orgánulos celulares como, mitocondrias, peroxisomas o retículo endoplásmico, principalmente. Igualmente, los alérgenos o los contaminantes ambientales también pueden promover la producción de ROS^{12,14,16}. En el asma alérgica, que es el fenotipo clínico más prevalente, la presencia de alérgenos genera ROS a través de un complejo mecanismo de activación que implica principalmente la activación de células dendríticas (DC), la activación de células T CD4+, la producción de interleucinas, la producción de IgE por parte de las células B y la activación de mastocitos y eosinófilos¹² (Figura 1).

Actualmente, los procedimientos más fiables para el diagnóstico del asma son el análisis de la historia clínica y la espirometría^{17,18}. Sin embargo, estos procedimientos no permiten evaluar la inflamación de las vías respiratorias. Aunque se ha sugerido que el recuento de eosinófilos en sangre es una herramienta útil para el diagnóstico del asma, los niveles de éstos pueden verse influidos por varios factores^{19,20}. Técnicas como el esputo inducido, la broncoscopia con lavado broncoalveolar (BAL) y la biopsia son adecuadas para evaluar la inflamación de las vías respiratorias y el estrés oxidativo, sin embargo, únicamente el esputo inducido es una técnica no invasiva^{21,22}. En este sentido, la comunidad científica está buscando incansablemente nuevas técnicas alternativas no invasivas o biomarcadores para la identificación de fenotipos de asma, la monitorización de su tratamiento, la predicción de exacerbaciones, el diagnóstico diferencial de otras

patologías con síntomas similares y el diagnóstico personalizado cuando los pacientes sufren comorbilidades como la rinosinusitis u otros trastornos atópicos^{6,22,23}.

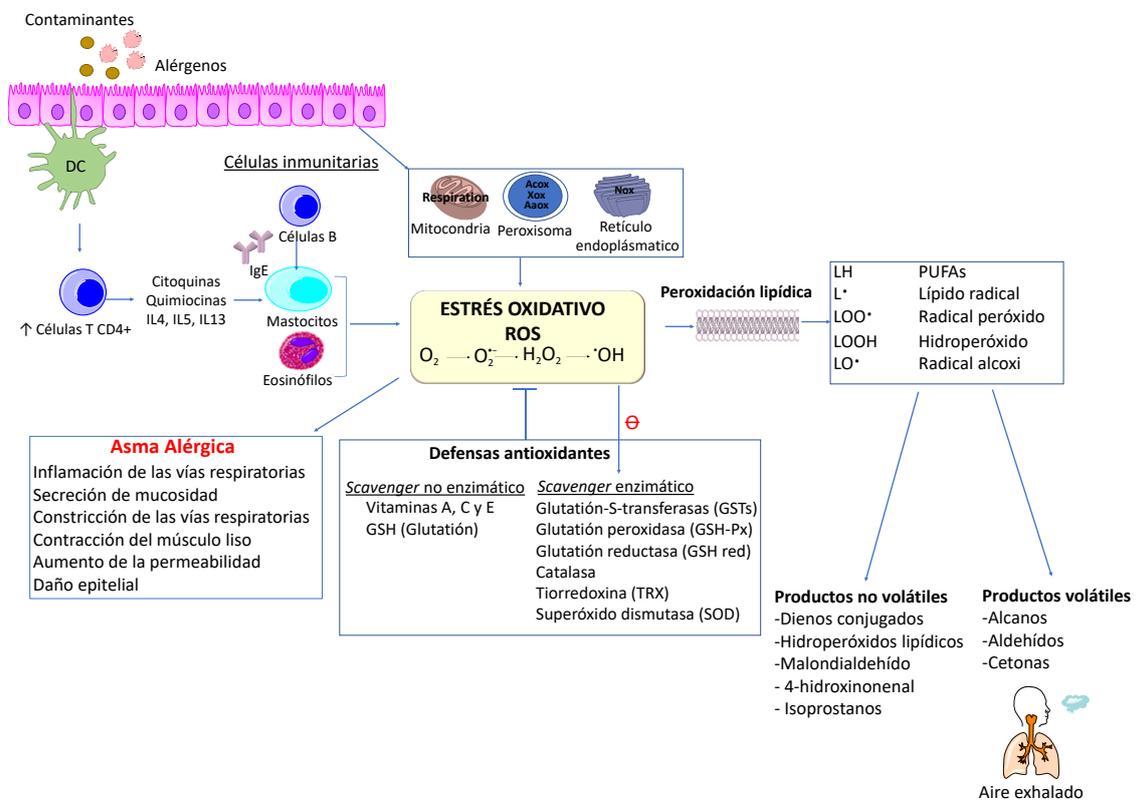


Figura 1. Peroxidación lipídica y producción de compuestos orgánicos volátiles en el asma alérgica. PUFAs: ácidos grasos poliinsaturados; ROS: especies reactivas de oxígeno; Nox: NADPH oxidasa; Acox: Acil-coenzima A-oxidasa; AaOx: aminoácido oxidasa.

El análisis del aire exhalado es un método no invasivo para evaluar la inflamación de las vías respiratorias y el estrés oxidativo²¹, ya que, por ejemplo, como consecuencia de la peroxidación lipídica se producen numerosos compuestos orgánicos volátiles (VOCs) que se exhalan en el aliento²⁴. En esta línea, varios estudios han empleado la medición de la fracción exhalada de óxido nítrico (FeNO) para obtener información adicional sobre la inflamación de las vías respiratoria, que pueda ser de utilidad para el diagnóstico del asma atópica. Sin embargo, esta metodología no es útil para el asma no alérgica, ya que los cambios en los niveles de FeNO pueden ser indicativos de varios trastornos distintos del asma^{21,25,26}. Como se ha comentado, es posible detectar cientos de compuestos orgánicos volátiles diferentes en el aliento exhalado²⁷. Principalmente, los VOCs en el aire exhalado pueden ser analizados mediante tecnologías basadas en la espectrometría de masas o matrices de sensores como la nariz electrónica (E-nose)^{23,28}. Actualmente, la cromatografía de gases acoplada a la espectrometría de masas (GC-MS) es considerada una de las técnicas más idóneas para el análisis del aire exhalado porque permite identificar los VOCs con una alta sensibilidad en el rango

de ppb²⁹. Los VOCs medidos en el aliento pueden proceder de la exposición a contaminantes exógenos (exposoma) o provenir de una fuente endógena, ya que pueden ser producidos por el metabolismo humano o incluso por las bacterias intestinales o localizadas en otras partes del organismo³⁰. Por todas estas razones, el análisis de VOCs en el aire exhalado ha sido propuesto como un enfoque novedoso para la predicción, el diagnóstico y el seguimiento del asma^{22,31}.

Desgraciadamente, esta estrategia está todavía en fase exploratoria y es necesario mejorar varios aspectos para que pueda ser introducida en el contexto clínico^{29,32,33}. En este sentido, es fundamental aplicar las mejores prácticas metabolómicas para superar esta fase inicial, como por ejemplo, realizar estudios que incluyan un gran número de sujetos, evaluar las influencias ambientales en las muestras de aliento, implementar flujos de trabajo de preprocesamiento de datos reproducibles y transparentes y realizar un análisis de datos robusto, entre otros^{29,33}. Así pues, la identificación de los VOCs como posibles biomarcadores de enfermedades requiere un adecuado análisis de datos. No obstante, a pesar de que multitud de estudios han invertido esfuerzos en la creación de modelos predictivos para datos clínicos, gran parte de la literatura existente recoge resultados atractivos que no han sido validados en un conjunto independiente para comprobar la interpretabilidad y la universalidad de los modelos generados^{29,33}.

Este trabajo se centra en el análisis de VOCs en el aire exhalado de mujeres en edad fértil que participan en una cohorte de nacimiento con el objetivo de discriminar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas). En este estudio se han aplicado prácticas consolidadas en el campo de la metabolómica. Para ello, hemos llevado a cabo un proceso de modelización con una validación cruzada sobre el conjunto de entrenamiento, y una validación externa utilizando un conjunto de testeo diferente para garantizar la aplicabilidad de los modelos a datos desconocidos. Asimismo, los datos recogidos en la Encuesta Europea de Salud en España de 2014 (EHSS-2014) y en la Encuesta Nacional de Salud de 2017 (ENSE-2017) han sido analizados estadísticamente para identificar factores asociados a la enfermedad del asma.

A través de la identificación de biomarcadores entre los compuestos orgánicos volátiles, el desarrollo metodológico del análisis del aire exhalado puede mejorar en gran medida nuestra capacidad para comprender la heterogeneidad del asma con otras enfermedades atópicas coexistentes, y los trastornos atópicos sin asma.

RESULTADOS

Análisis de los datos recogidos en las encuestas de salud en España en los últimos años

Análisis de la Encuesta Europea de Salud en España de 2014 (EHSS-2014)

La representación del MCA (análisis de correspondencias múltiples) realizado sobre la colección de datos de la EHSS-2014 se muestra en la Figura 2A. Los resultados de las pruebas de chi-cuadrado y las pruebas exactas de Fisher (Apéndice 1) mostraron que la variable "asma" tenía una relación significativa con las siguientes variables: edad, sexo, estado de salud, hipertensión arterial, otros trastornos atópicos (rinitis alérgica, dermatitis atópica, conjuntivitis alérgica, alergia alimentaria u otras alergias (excluyendo el asma)), diabetes, trastornos de la piel, cirrosis, depresión, trastornos de ansiedad, problemas renales, problemas de tiroides, osteoporosis, índice de masa corporal (IMC), actividad física y tabaco. La representación del MCA realizado sobre una cohorte abierta de mujeres de 18 a 45 años construida a partir de los datos recogidos de la EHSS-2014 se muestra en la Figura 2B. El 6,9% de las mujeres en edad fértil eran asmáticas. Los resultados de las pruebas de chi-cuadrado y las pruebas exactas de Fisher sobre los datos de las mujeres en edad fértil (Apéndice 1) mostraron que la variable "asma" tenía una relación significativa con las siguientes variables: nacionalidad, estado de salud, hipertensión arterial, otros trastornos atópicos, trastornos de la piel, trastornos de ansiedad, colesterol, IMC y dieta.

Análisis de la Encuesta Nacional de Salud de España de 2017 (ENSE-2017)

La representación del MCA realizado sobre los datos recogidos en la ENSE-2017 se muestra en la Figura 3A. Los resultados de las pruebas de chi-cuadrado y las pruebas exactas de Fisher (Apéndice 1) mostraron que la variable "asma" tenía una relación significativa con las siguientes variables: edad, sexo, nacionalidad, hipertensión arterial, estado de salud, otros trastornos atópicos, diabetes, trastornos de la piel, colesterol, cirrosis, depresión, trastornos de ansiedad, problemas renales, tumores malignos, problemas de tiroides, osteoporosis, IMC, tabaco y alcohol. La representación del MCA realizado sobre una cohorte abierta de mujeres de 18 a 45 años construida a partir de los datos recogidos en la ENSE-2017 se muestra en la Figura 3B. El 7,1% de las mujeres en edad fértil eran asmáticas. Los resultados de las pruebas de chi-cuadrado y las pruebas exactas de Fisher sobre los datos de las mujeres en edad fértil (Apéndice 1) mostraron que la variable "asma" tenía una relación significativa con las siguientes

variables: estado de salud, hipertensión arterial, otros trastornos atópicos, trastornos de la piel, depresión, trastornos de ansiedad y alcohol.

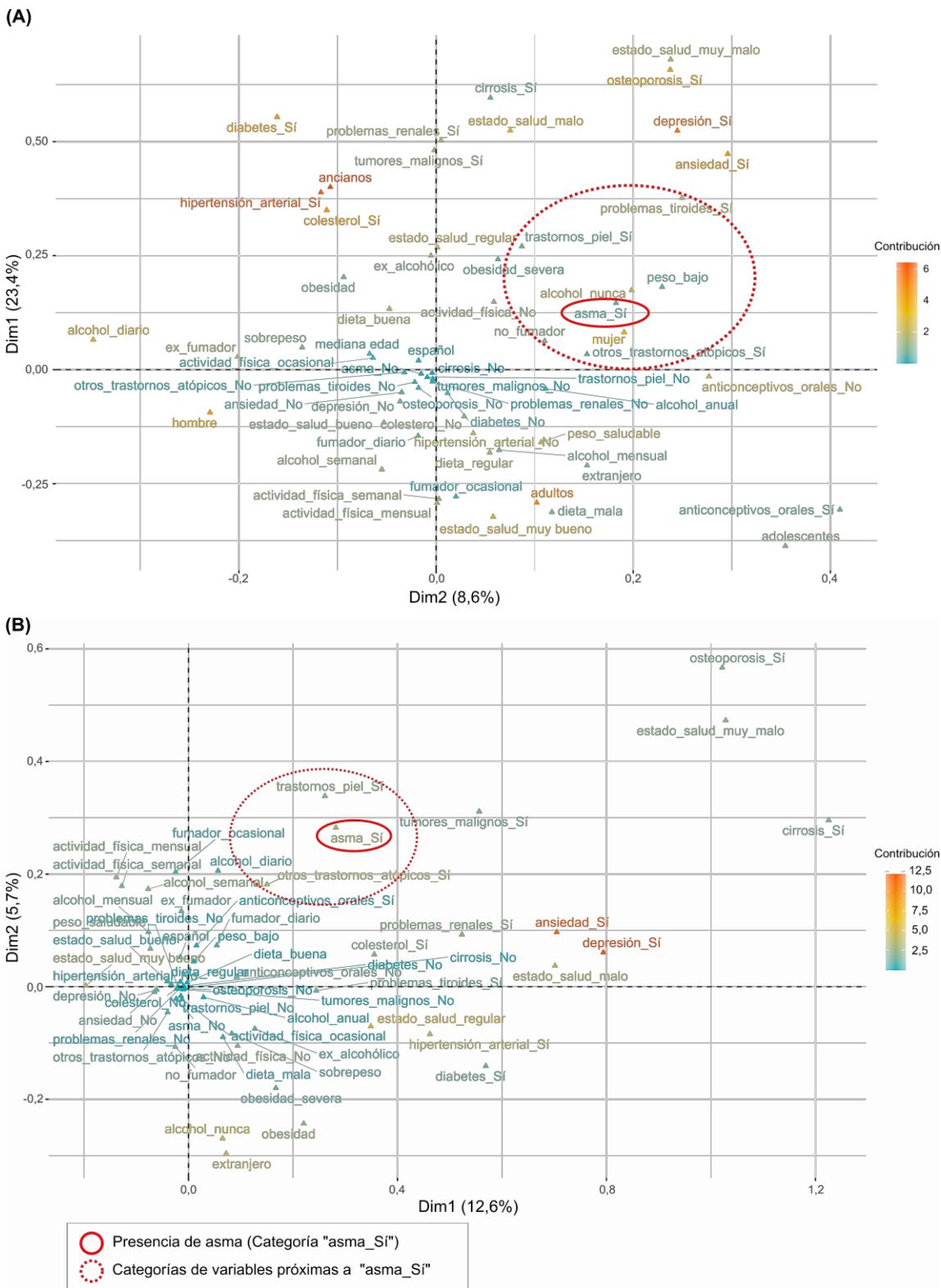


Figura 2. Representación de los resultados del MCA (análisis de correspondencias múltiples) realizado sobre los datos recogidos en la Encuesta Europea de Salud en España 2014 (EHSS-2014). (A) Todos los datos recogidos en la EHSS-2014. (B) Cohorte abierta de mujeres de 18 a 45 años construida a partir de la EHSS-2014. Dim: dimensión.

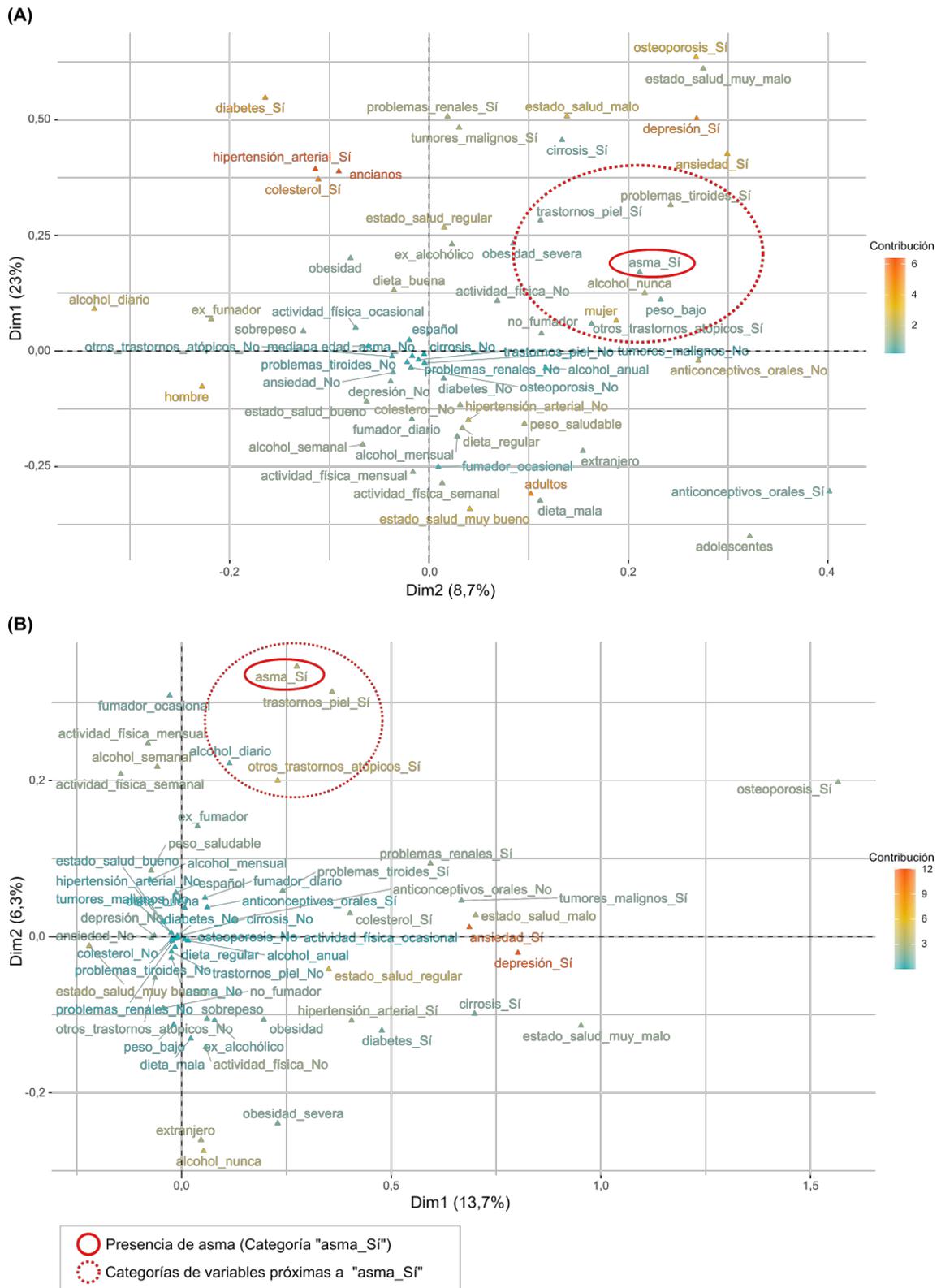


Figura 3. Representación de los resultados del MCA (análisis de correspondencias múltiples) realizado sobre los datos recogidos en la Encuesta Nacional de Salud de España de 2017 (ENSE-2017). (A) Todos los datos recogidos en la ENSE-2017. (B) Cohorte abierta de mujeres de 18 a 45 años construida a partir de la ENSE-2017. Dim: dimensión

Características de la población de estudio: asociaciones entre enfermedades atópicas

Un total de 337 mujeres de la cohorte NELA (*Nutrition in Early Life and Asthma*)³⁴ participaron en el presente estudio. Las mujeres se distribuyeron aleatoriamente en dos grupos: Grupo 1 (n = 211), que fue utilizado como conjunto de entrenamiento, y Grupo 2 (n = 126), que fue empleado como conjunto de validación. En función del diagnóstico previo de asma y de otras enfermedades atópicas, las mujeres de cada grupo se clasificaron en cuatro categorías: asmáticas con otras enfermedades atópicas coexistentes (A-AD), no asmáticas con otras enfermedades atópicas (NA-AD), no asmáticas sin enfermedades atópicas (NA-NAD) y no asmáticas (NA) (esta categoría incluye tanto a las no asmáticas con otras enfermedades atópicas como a las no asmáticas sin enfermedades atópicas). El diagrama de flujo del estudio se muestra en la Figura 4. Una mujer del Grupo 2 fue excluida por ser asmática sin otras enfermedades atópicas. La Figura 5 incluye dos diagramas de Venn que muestran la coexistencia de enfermedades atópicas en las mujeres asmáticas de ambos grupos. La rinitis alérgica y la conjuntivitis alérgica fueron las enfermedades atópicas coexistentes más comunes en las mujeres con asma. Las características demográficas de las mujeres del Grupo 1 y del Grupo 2 se muestran en las Tablas 1 y 2, respectivamente. El porcentaje de mujeres con rinitis alérgica fue mayor en A-AD que en NA-AD, tanto en el Grupo 1 como en el Grupo 2. El número de mujeres del Grupo 1 con antecedentes parentales atópicos (antecedentes parentales de asma y rinitis alérgica) en NA y NA-NAD fue menor que en A-AD. Además, en el Grupo 1, el porcentaje de mujeres con antecedentes parentales de rinitis alérgica fue mayor en NA-AD y NA-NAD. Al mismo tiempo, el recuento de eosinófilos en sangre resultó ser significativamente mayor en A-AD que en NA, NA-AD y NA-NAD, tanto en el Grupo 1 (A-AD vs. NA (p-valor = 4,4e-05), A-AD vs. NA-AD (p-valor = 8,3e-05), y A-AD vs. NA-NAD (p-valor = 0,00014)) y el Grupo 2 (A-AD vs. NA (p-valor = 0,0012), A-AD vs. NA-AD (p-valor = 0,0086), y A-AD vs. NA-NAD (p-valor = 0,00075)) (Apéndice 2).

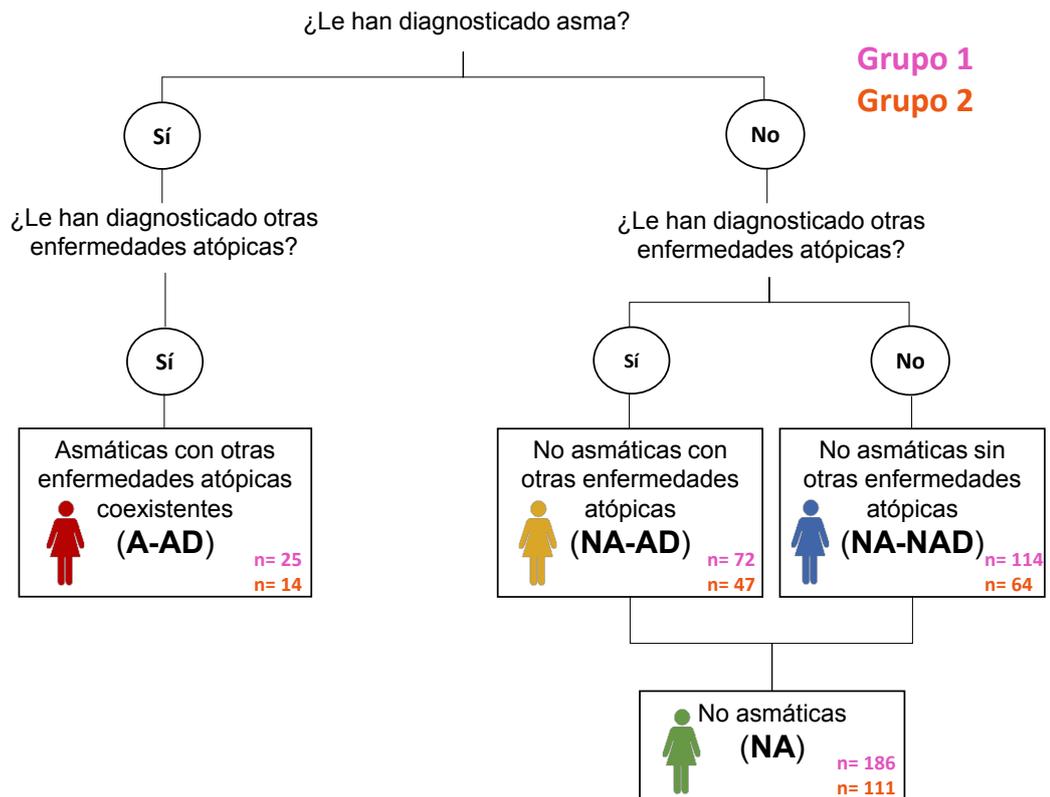


Figura 4. Diagrama de flujo del estudio. Otras enfermedades atópicas incluyen la rinitis alérgica, la dermatitis atópica, la conjuntivitis alérgica, la alergia alimentaria o la alergia a medicamentos.

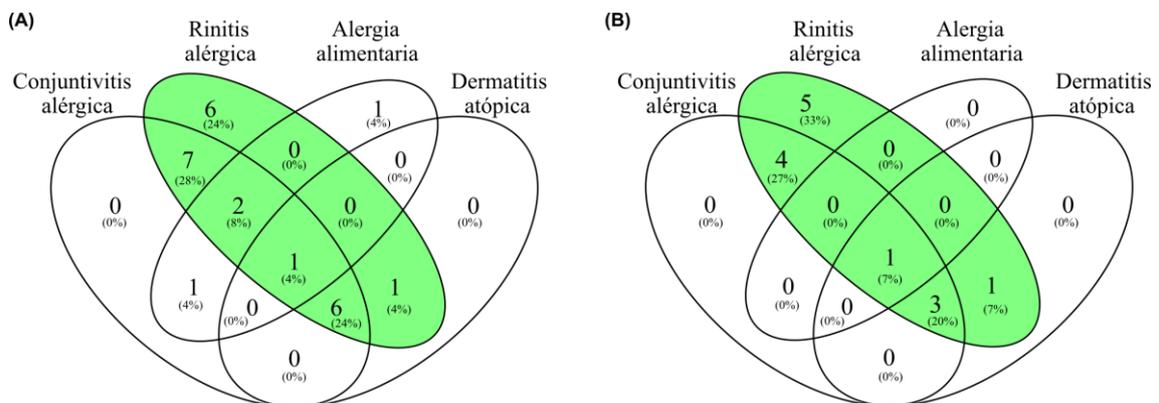


Figura 5. Coexistencia de trastornos atópicos en mujeres con asma. (A) Diagrama de Venn que indica la coexistencia de trastornos atópicos en las mujeres con asma del Grupo 1. (B) Diagrama de Venn que indica la coexistencia de trastornos atópicos en las mujeres con asma del Grupo 2.

Tabla 1. Características de la población de estudio: mujeres en edad fértil (Conjunto de entrenamiento - Grupo 1).

	Grupo 1 (211)			
	A-AD	NA	NA-AD	NA-NAD
Sujetos n	25	186	72	114
Edad (años), media (rango)	32,4 (20-39)	32,9 (20-43)	32,8 (23-41)	33,0 (20-43)
IMC (kg/m²) antes del embarazo, media (rango)	25,23 (19,72-36,81)	24,13 (16,23-42,32)	23,59 (16,65-38,67)	24,48 (16,23-42,32)
Clase social Directivos-técnicos/ trabajadores cualificados/ trabajadores semicualificados-sin cualificar/ desempleados, %	36,0/20,0/20,0/24,0	35,5/18,3/22,0/24,2	34,7/15,2/20,8/29,2	36,0/20,2/22,8/21,1
Nivel educativo Secundaria incompleta o menos/ secundaria completa/ universitaria, %	28,0/12,0/60,0	19,3/25,3/55,4	25,0/27,8/47,2	15,8/23,7/60,5
Tabaquismo durante el embarazo, Sí/No n/n	2/23	30/156	9/63	21/93
Fumadora, Sí/No n/n	3/22	30/156	8/64	22/92
Otros trastornos atópicos, Sí	25 (100%) ^b	72 (39%) ^a	72 (100%)	-
Rinitis alérgica, Sí	23 (92%) ^{b, c}	47 (25%) ^a	47 (65%) ^a	-
Dermatitis atópica, Sí	8 (32%) ^b	12 (6%) ^a	12 (17%)	-
Conjuntivitis alérgica, Sí	17 (68%) ^b	26 (14%) ^a	26 (36%)	-
Alergia alimentaria, Sí	5 (20%) ^b	10 (5%) ^a	10 (14%)	-
Alergia a medicamentos, Sí	5 (20%)	14 (8%)	14 (19%)	-
Antecedentes parentales de asma, Sí	6 (24%) ^{b, d}	11 (6%) ^a	6 (8%)	5 (4%) ^a
Antecedentes parentales de rinitis alérgica, Sí	10 (40%) ^{b, d}	36 (19%) ^a	21 (29%) ^d	15 (13%) ^{a, c}
Antecedentes parentales de dermatitis atópica, Sí	3 (12%)	8 (4%)	2 (3%)	6 (5%)
Consumo de antibióticos, Sí	11 (44%)	57 (31%)	22 (31%)	35 (31%)
Consumo de paracetamol, Sí	18 (72%)	107 (56%)	39 (54%)	68 (60%)
Consumo de corticoides inhalados, Sí	2 (8%)	5 (3%)	3 (4%)	2 (2%)
Consumo de corticoides inyectables, Sí	2 (8%)	7 (4%)	6 (8%) ^d	1 (1%) ^c
Recuento de leucocitos en sangre /μl, media (rango)	9548 (6380-16610)	9247 (5270-16720)	9054 (5270-15560)	9369 (5270-16720)
Recuento de monocitos en sangre /μl, media (rango)	590 (320-1080)	614 (240-1460)	618 (370-1160)	611 (240-1460)
Recuento de linfocitos en sangre /μl, media (rango)	1847 (1160-3130)	2017 (770-4290)	2021 (960-3600)	2015 (770-4290)
Recuento de basófilos en sangre /μl, media (rango)	32 (10-60)	31 (0-100)	28 (10-60)	32 (0-100)
Recuento de eosinófilos en sangre /μl, media (rango)	213 (70-350) ^{b, c, d}	138 (0-500) ^a	131 (0-380) ^a	143 (0-500) ^a
Recuento de neutrófilos en sangre /μl, media (rango)	6866 (3960-12780)	6447 (3380-11360)	6255 (3380-11360)	6568 (3380-10970)
Estación de medición Invierno/primavera/verano/otoño %	24,0/20,0/20,0/36,0	25,3/10,8/22,0/41,9	38,9/11,1/19,4/30,6 ^d	16,7/10,5/23,7/49,1 ^c

^a Diferencias significativas (p-valor < 0,05) con respecto a las asmáticas con otras enfermedades atópicas coexistentes (A-AD)

^b Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas (NA)

^c Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas con otras enfermedades atópicas (NA-AD)

^d Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas sin enfermedades atópicas (NA-NAD)

IMC: índice de masa corporal; A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA-AD: no asmáticas con otras enfermedades atópicas; NA-NAD: no asmáticas sin enfermedades atópicas; NA: no asmáticas.

Tabla 2. Características de la población de estudio: mujeres en edad fértil (Conjunto de validación - Grupo 2).

	Grupo 2 (n=126) [†]			
	A-AD	NA	NA-AD	NA-NAD
Sujetos n	14	111	47	64
Edad (años), media (rango)	33,0 (25-39)	33,3 (18-42)	33,0 (18-42)	33,5 (22-41)
IMC (kg/m²) antes del embarazo, media (rango)	23,50 (18,46-34,37)	23,28 (15,94-39,91)	23,35 (17,63-37,29)	23,23 (15,94-39,91)
Clase social Directivos-técnicos/ trabajadores cualificados/ trabajadores semicualificados-sin cualificar/ desempleados, %	28,6/50,0/14,3/7,1	39,6/20,7/18,0/21,6	29,8/19,1/23,4/27,7	46,9/21,9/14,1/17,2
Nivel educativo Secundaria incompleta o menos/ secundaria completa/ universitaria, %	7,1/35,7/57,1	11,7/28,8/59,5	14,9/31,9/53,2	9,4/26,6/64,1
Tabaquismo durante el embarazo, Sí/No n/n	2/12	14/97	4/43	10/54
Fumadora, Sí/No n/n	3/11	13/98	6/41	7/57
Otros trastornos atópicos, Sí	14 (100%) ^b	47 (42%) ^a	47 (100%)	-
Rinitis alérgica, Sí	14 (100%) ^{b, c}	34 (31%) ^a	34 (72%) ^a	-
Dermatitis atópica, Sí	5 (36%) ^b	13 (12%) ^a	13 (28%)	-
Conjuntivitis alérgica, Sí	8 (57%) ^b	20 (18%) ^a	20 (43%)	-
Alergia alimentaria, Sí	1 (7%)	7 (6%)	7 (15%)	-
Alergia a medicamentos, Sí	3 (21%)	8 (7%)	8 (17%)	-
Antecedentes parentales de asma, Sí	3 (21%)	12 (11%)	5 (11%)	7 (11%)
Antecedentes parentales de rinitis alérgica, Sí	7 (50%)	34 (31%)	19 (40%)	15 (23%)
Antecedentes parentales de dermatitis atópica, Sí	2 (14%)	11 (10%)	6 (13%)	5 (8%)
Consumo de antibióticos, Sí	1 (7%)	32 (29%)	12 (26%)	20 (31%)
Consumo de paracetamol, Sí	8 (57%)	67 (60%)	31 (66%)	36 (56%)
Consumo de corticoides inhalados, Sí	2 (14%) ^{b, d}	1 (1%) ^a	1 (2%)	0 (0%) ^a
Consumo de corticoides inyectables, Sí	0 (0%)	2 (2%)	1 (2%)	1 (2%)
Recuento de leucocitos en sangre /μl, media (rango)	8537 (7050-11570)	9268 (5180-15310)	9258 (5760-14090)	9276 (5180-15310)
Recuento de monocitos en sangre /μl, media (rango)	630 (450-960)	595 (320-1280)	601 (360-1030)	590 (320-1280)
Recuento de linfocitos en sangre /μl, media (rango)	2035 (1270-2650)	2002 (990-3790)	1988 (1110-3790)	2013 (990-3700)
Recuento de basófilos en sangre /μl, media (rango)	34 (20-60)	34 (0-90)	37 (10-80)	33 (0-90)
Recuento de eosinófilos en sangre /μl, media (rango)	234 (90-510) ^{b, c, d}	146 (20-430) ^a	159 (20-430) ^a	135 (30-430) ^a
Recuento de neutrófilos en sangre /μl, media (rango)	5604 (4390-8320)	6491 (3470-10990)	6473 (3510-9790)	6504 (3470-10990)
Estación de medición Invierno/primavera/verano/otoño %	0,0/35,7/35,7/28,6	3,6/32,4/41,4/22,5	2,1/36,2/38,3/23,4	4,7/29,7/43,8/21,9

^a Diferencias significativas (p-valor < 0,05) con respecto a las asmáticas con otras enfermedades atópicas coexistentes (A-AD)

^b Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas (NA)

^c Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas con otras enfermedades atópicas (NA-AD)

^d Diferencias significativas (p-valor < 0,05) con respecto a las no asmáticas sin enfermedades atópicas (NA-NAD)

[†] Una mujer no fue incluida en ninguna de las cuatro categorías porque era asmática sin otras enfermedades atópicas. IMC: índice de masa corporal; A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA-AD: no asmáticas con otras enfermedades atópicas; NA-NAD: no asmáticas sin enfermedades atópicas; NA: no asmáticas.

Creación de modelos predictivos para identificar VOCs discriminantes en el análisis del aire exhalado

En el análisis exploratorio de los datos mediante un análisis de componentes principales (PCA), no se observó ninguna agrupación debida a la presencia de asma (Apéndice 3). Sin embargo, sí que se observó en todos los tipos de muestras (muestras de aire exhalado humano y muestras de aire ambiental) una agrupación debida a la estación del año en la que se realizó el muestreo (Apéndice 4). Por lo tanto, se aplicó ASCA (ANOVA (análisis de varianza) - análisis de componentes simultáneos) para evitar las influencias de la estación climática de medición. Los factores seleccionados para la aplicación de ASCA fueron la estación del año de medición y la zona de residencia habitual definida por la modelización de la calidad del aire mediante el sistema de modelización *Weather Research and Forecasting* (WRF) + CHIMERE, considerando los niveles de ozono (O₃), dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y partículas^{35,36}. Tal y como se puede observar en el Apéndice 4, tras la implementación de ASCA no se observó ningún sesgo causado por la estación de muestreo y medición.

Las propiedades y las curvas ROC (*receiver operator characteristic*) de los modelos construidos se muestran en la Figura 6. Las fórmulas de los modelos están disponibles en el Apéndice 5, y las características de los *features* (fragmentos iónicos de los VOCs detectados mediante GC-MS) seleccionados por los modelos se detallan en el Apéndice 6. El compuesto que incluía el *feature* F62 tenía un patrón de fragmentación iónica característico de varios compuestos de la familia de las tetrahydroisoquinolinas; por lo tanto, a este compuesto se le denominó "derivado de tetrahydroisoquinolina" y no se le asignó un nombre específico. La identificación de los compuestos a los que pertenecían todos los *features* seleccionados se llevó a cabo en base al cotejo de los espectros de masas y los tiempos de retención con la biblioteca NIST (*National Institute of Standard and Technology*) y con los estándares comerciales (cálculo del factor de coincidencia o *match factor* y del valor de *error RI* (error del índice de retención)), excepto en el caso de tres compuestos (isopreno, derivado de tetrahydroisoquinolina y ácido E-3-(2-hidroxifenil)prop-2-enoico, cuya identificación únicamente se basó en el cotejo con la biblioteca de espectros de masas. El rendimiento de los modelos construidos y las variables utilizadas en cada modelo se resumen en la Tabla 3.

Sujetos	Hipótesis	Modelos	Variables
 I) A-AD vs. NA	Diferenciación de las asmáticas con otras enfermedades atópicas coexistentes entre la población general	I.A	VOCs
		I.B	VOCs + Antecedentes parentales de atopía
		I.C	VOCs + WBCs
		I.D	VOCs + Antecedentes parentales de atopía + WBCs
		I.E	VOCs + Other atopic disorders
		I.F	VOCs + Antecedentes parentales de atopía + Otros trastornos atópicos
		I.G	VOCs + Antecedentes parentales de atopía + WBCs + Otros trastornos atópicos
 II) A-AD vs. NA-NAD	Discriminación entre las asmáticas con otras enfermedades atópicas coexistentes y la población sana	II.A	VOCs
		II.B	VOCs + Antecedentes parentales de atopía
		II.C	VOCs + WBCs
		II.D	VOCs + Antecedentes parentales de atopía + WBCs
 III) A-AD vs. NA-AD	Diferenciación de las asmáticas con otras enfermedades atópicas coexistentes entre pacientes con otros trastornos atópicos	III.A	VOCs
		III.B	VOCs + Antecedentes parentales de atopía
		III.C	VOCs + WBCs
		III.D	VOCs + Antecedentes parentales de atopía + WBCs
 IV) NA-AD vs. NA-NAD	Discriminación entre las no asmáticas con otras enfermedades atópicas y la población sana	IV.A	VOCs

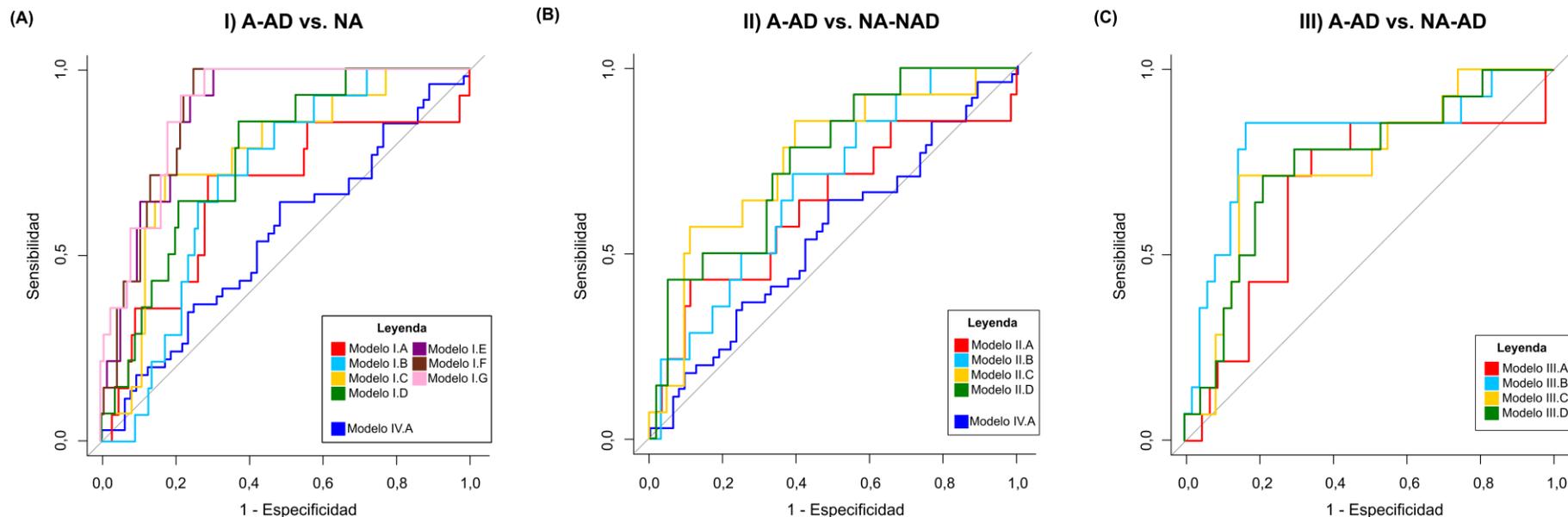


Figura 6. Modelos construidos. Panel superior. Características de los modelos construidos. (WBCs: recuento de glóbulos blancos). Panel inferior. Curvas ROC (*receiver operating characteristic curves*) de los modelos construidos. (A) Curvas ROC para discriminar entre A-AD (asmáticas con otras enfermedades atópicas coexistentes) y NA (no asmáticas). (B) Curvas ROC para discriminar entre A-AD y NA-NAD (no asmáticas sin enfermedades atópicas). (C) Curvas ROC para discriminar entre A-AD y NA-AD (no asmáticas con otras enfermedades atópicas). VOCs: Compuestos orgánicos volátiles.

Tabla 3. Rendimiento de los modelos construidos y las variables seleccionadas en cada modelo (VOCs y otras variables).

Modelos	AUC _{CV}	p-valor	AUC _{VS}	Sensibilidad	Especificidad	VOCs	Otras variables
I) A-AD vs NA							
I.A	0,70	0,017	0,66	71 %	62 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	
I.B	0,73	0,003	0,70	71 %	60 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de rinitis - Antecedentes parentales de dermatitis
I.C	0,79	0,001	0,77	71 %	65 %	- Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Recuento de eosinófilos en sangre
I.D	0,80	0,001	0,76	86 %	63 %	- Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de rinitis - Recuento de eosinófilos en sangre
I.E	0,88	0,001	0,88	100 %	69 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Rinitis alérgica - Dermatitis atópica - Conjuntivitis alérgica - Alergia alimentaria - Alergia a medicamentos
I.F	0,90	0,001	0,89	100 %	70 %	- Acetona (Código CAS: 67-64-1) - Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de dermatitis - Rinitis alérgica - Dermatitis atópica - Conjuntivitis alérgica - Alergia alimentaria - Alergia a medicamentos
I.G	0,91	0,001	0,90	100 %	68 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de dermatitis - Recuento de eosinófilos en sangre - Rinitis alérgica - Dermatitis atópica - Conjuntivitis alérgica - Alergia alimentaria - Alergia a medicamentos
II) A-AD vs NA-NAD							
II.A	0,68	0,031	0,63	64 %	55 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	
II.B	0,73	0,008	0,68	71 %	58 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de rinitis

VOCs: compuestos orgánicos volátiles; A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA-AD: no asmáticas con otras enfermedades atópicas; NA-NAD: no asmáticas sin enfermedades atópicas; NA: no asmáticas; AUC: área bajo la curva ROC (*receiver operating characteristic curve*); AUC_{CV}: AUC obtenida mediante validación cruzada de 5 iteraciones; AUC_{VS}: AUC obtenida en la validación sobre el conjunto de testeo.

Tabla 3 (Continuación). Rendimiento de los modelos construidos y las variables seleccionadas en cada modelo (VOCs y otras variables).

Modelos	AUC _{CV}	p-valor	AUC _{VS}	Sensibilidad	Especificidad	VOCs	Otras variables
II) A-AD vs NA-NAD							
II.C	0,77	0,001	0,75	64 %	68 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Recuento de eosinófilos en sangre
II.D	0,79	0,002	0,75	71 %	62 %	- Acetona (Código CAS: 67-64-1) - Derivado de Tetrahydroisoquinolina - 2-Etil-1-hexanol (Código CAS: 104-76-7)	- Antecedentes parentales de asma - Antecedentes parentales de rinitis - Recuento de eosinófilos en sangre
III) A-AD vs NA-AD							
III.A	0,70	0,025	0,68	71 %	72 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - Decano (Código CAS: 124-18-5) - 2-Etil-1-hexanol (Código CAS: 104-76-7) - 2,2,4-trimetil-1,3-pentanodiol diisobutirato (TXIB) (Código CAS: 6846-50-0)	
III.B	0,67	0,042	0,81	86 %	60 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - Decano (Código CAS: 124-18-5) - 2-Etil-1-hexanol (Código CAS: 104-76-7) - 2,2,4-trimetil-1,3-pentanodiol diisobutirato (TXIB) (Código CAS: 6846-50-0)	- Antecedentes parentales de asma - Antecedentes parentales de dermatitis
III.C	0,80	0,001	0,74	71%	60 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - Decano (Código CAS: 124-18-5) - 2-Etil-1-hexanol (Código CAS: 104-76-7) - Ftalato de diisobutilo (Código CAS: 84-69-5)	- Recuento de eosinófilos en sangre - Recuento de linfocitos en sangre
III.D	0,82	0,001	0,75	79 %	70 %	- Disulfuro de carbono (Código CAS: 75-15-0) - Derivado de Tetrahydroisoquinolina - Decano (Código CAS: 124-18-5) - 2-Etil-1-hexanol (Código CAS: 104-76-7) - Ftalato de diisobutilo (Código CAS: 84-69-5)	- Antecedentes parentales de asma - Antecedentes parentales de dermatitis - Recuento de eosinófilos en sangre - Recuento de linfocitos en sangre
IV) NA-AD vs NA-NAD							
IV.A	0,55	0,236	0,54	46 %	62 %	- Isopreno (Código CAS: 78-79-5) - Ácido E-3-(2-hidroxifenil)prop-2-enoico (Código CAS: 614-60-8) - 2,2,4-trimetil-1,3-pentanodiol diisobutirato (TXIB) (Código CAS: 6846-50-0)	

VOCs: compuestos orgánicos volátiles; A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA-AD: no asmáticas con otras enfermedades atópicas; NA-NAD: no asmáticas sin enfermedades atópicas; NA: no asmáticas; AUC: área bajo la curva ROC (*receiver operating characteristic curve*); AUC_{CV}: AUC obtenida mediante validación cruzada de 5 iteraciones; AUC_{VS}: AUC obtenida en la validación sobre el conjunto de testeo.

Comparación I: Asmáticas con otras enfermedades atópicas coexistentes (A-AD) vs No asmáticas (NA)

Siete modelos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F y modelo I.G) fueron generados para distinguir entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas. Todos los modelos fueron significativos (p -valor $< 0,05$ en las pruebas de permutación), y los *features* seleccionados de 2-etil-1-hexanol y del derivado de tetrahydroisoquinolina se comportaron como discriminantes en todos los modelos. *Features* pertenecientes a la acetona, el 2-etil-1-hexanol y un derivado de la tetrahydroisoquinolina fueron seleccionados como discriminantes en el modelo I.A, que sólo tenía en cuenta las medidas de los VOCs en el aire exhalado. El valor de AUC (área bajo la curva ROC) de este modelo sobre el conjunto de validación fue de 0,67 (71% de sensibilidad y 63% de especificidad). El modelo I.B, que también incluía los antecedentes parentales atópicos, seleccionó el mismo conjunto de *features* que el modelo I.A, así como variables relacionadas con los antecedentes atópicos (antecedentes parentales de asma, antecedentes parentales de rinitis alérgica y antecedentes parentales de dermatitis atópica).

Comparación II: Asmáticas con otras enfermedades atópicas coexistentes (A-AD) vs. No asmáticas sin enfermedades atópicas (NA-NAD)

Se construyeron cuatro modelos para distinguir entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas sin enfermedades atópicas. En los cuatro modelos se seleccionaron los *features* de la acetona, del 2-etil-1-hexanol y de un derivado de la tetrahydroisoquinolina. Todos los modelos resultaron ser significativos (p -valor $< 0,05$ en las pruebas de permutación).

Comparación III: Asmáticas con otras enfermedades atópicas coexistentes (A-AD) vs. No asmáticas con otras enfermedades atópicas (NA-AD)

Se construyeron cuatro modelos (modelo III.A, modelo III.B, modelo III.C y modelo III.D)) para discriminar la presencia de asma entre las mujeres que padecen alguna enfermedad atópica. Todos los modelos mostraron significación estadística (p -valor $< 0,05$ en las pruebas de permutación). En los cuatro modelos se seleccionaron los *features* del disulfuro de carbono, de un derivado de la tetrahydroisoquinolina, del decano y del 2-etil-1-hexanol.

Comparación IV: No asmáticas con otras enfermedades atópicas (NA-AD) vs. No asmáticas sin enfermedades atópicas (NA-NAD)

El modelo IV.A no permitió discriminar entre no asmáticas con otras enfermedades atópicas y no asmáticas sin enfermedades atópicas. Este modelo no fue significativo (p-valor = 0,236), y el grado de acierto sobre el conjunto de validación fue pobre (AUC = 0,54, 46% de sensibilidad y 62% de especificidad).

Los niveles en el aire exhalado de los *features* seleccionados por los quince modelos significativos no se vieron influidos por los hábitos de consumo de tabaco ni tampoco por el consumo de fármacos declarado por las mujeres (antibióticos, paracetamol, corticosteroides inhalados y corticosteroides inyectables), con excepción de los niveles del 2,2,4-trimetil-1,3-pentanediol diisobutirato ($m/z = 99$) que fueron significativamente mayores en las mujeres que consumían paracetamol (p-valor = 0,013) (Apéndices 7-11). Sin embargo, no se observaron diferencias significativas en el consumo de paracetamol entre las cuatro categorías (A-AD, NA, NA-AD y NA-NAD) (Tabla 1 y Tabla 2). Asimismo, las intensidades de los *features* de los VOCs discriminantes fueron significativamente mayores en las muestras de aire exhalado humano que en las muestras ambientales (Apéndice 12). Además, el 2,2,4-trimetil-1,3-pentanodiol diisobutirato ($m/z = 99$) no fue detectado en las muestras ambientales.

DISCUSIÓN

Los resultados del análisis de datos, disponibles hasta la fecha, de las dos últimas encuestas de salud en España (EHSS-2014 y ENSE-2017) elaboradas por el INE (Instituto Nacional de Estadística de España) indicaron un sesgo de género en la enfermedad del asma en España, debido a su alta prevalencia en las mujeres. En los gráficos de los resultados del MCA de ambas encuestas (Figura 2A y Figura 3A), la edad de los sujetos fue la variable con mayor contribución a la Dimensión 1. La categoría "mujer" de la variable de género se encontraba muy cerca de la categoría "asma_Sí" en ambas dimensiones (Dimensión 1 y Dimensión 2). Además, la enfermedad del asma estaba muy próxima en la Dimensión 2 de otras enfermedades como los problemas de tiroides, otros trastornos atópicos (cualquier enfermedad atópica excluyendo el asma) o trastornos de la piel. Asimismo, se observó una relación significativa entre el asma y otras enfermedades como los trastornos de la piel, otros trastornos atópicos, los problemas de tiroides, la diabetes, los problemas renales, la hipertensión arterial o la osteoporosis (Apéndice 1). Se ha demostrado que el asma influye en algunas enfermedades crónicas como la cardiopatía coronaria, la diabetes mellitus y la

hipertensión, pero aún no se ha comprobado su impacto en otras patologías como la enfermedad renal crónica. Sin embargo, los pacientes con asma bronquial pueden tener un mayor riesgo de desarrollar enfermedad renal crónica³⁷. También, se ha observado una relación entre el asma y los problemas de tiroides^{38,39}, pero aún no están claros los efectos que puedan ejercer las hormonas tiroideas sobre la contractilidad de las vías respiratorias, aunque la prevalencia de los problemas tiroideos es mayor en las mujeres. En este sentido, se ha señalado que el sesgo de sexo podría deberse a que la regulación hormonal puede desempeñar un papel relevante en los problemas tiroideos^{10,40}. No obstante, no se observó ninguna asociación significativa entre el asma y los problemas de tiroides, la diabetes, los problemas renales y la osteoporosis cuando únicamente se seleccionaron los datos de las mujeres de 18 a 45 años en ambas encuestas de salud (Apéndice 1). Por otra parte, los resultados del análisis de los datos de las cohortes abiertas de mujeres de 18 a 45 años confirmaron que la enfermedad del asma está asociada a los trastornos de la piel y otros trastornos atópicos. La categoría "asma_Sí" estaba muy cerca de las categorías "trastornos piel_Sí" y "otros trastornos atópicos_Sí" en la dimensión 1 en los gráficos del MCA de ambas encuestas (Figura 2B y Figura 3B). En consecuencia, los factores más asociados a la enfermedad del asma en adultos fueron el género, otros trastornos atópicos y los trastornos de la piel.

Este trabajo muestra el primer estudio centrado en la discriminación del asma con otras enfermedades atópicas coexistentes en mujeres en edad fértil, utilizando una técnica no invasiva como es el análisis de VOCs en el aire exhalado. Aunque el análisis de VOCs del aire exhalado se ha propuesto como una estrategia potencial para el diagnóstico y la monitorización del asma, su implementación en la práctica clínica no ha sido posible hasta la fecha. En este sentido, es crucial la implementación de prácticas bien establecidas en la metabolómica para superar las limitaciones actuales que impiden su integración en la práctica clínica diaria^{29,33}. Por lo tanto, en este estudio se adoptó una perspectiva metabolómica. Para ello, se reclutó una gran cohorte de sujetos, y se recogió una muestra del aire de la habitación de muestreo por cada muestra de las participantes, con el fin de evaluar la posible contaminación ambiental. Además, se ha hecho especial hincapié en la etapa de preprocesamiento de datos, que es un clásico cuello de botella en el análisis del aire exhalado por GC-MS. El mayor reto para el desarrollo de la técnica ha sido precisamente abordar el preprocesamiento de datos para transformarlos en una matriz útil para el análisis de datos. De hecho, en el presente estudio se ha aplicado por primera vez un flujo de trabajo reproducible y transparente desarrollado recientemente por nuestro grupo para el preprocesamiento de datos utilizando fuentes abiertas para el descubrimiento de biomarcadores (Capítulo 4)⁴¹. La

identificación de los VOCs se realizó basándose en la similitud espectral y los tiempos de retención.

Una vez finalizado el preprocesamiento de datos, es necesario un análisis de datos robusto con una validación adecuada para medir el rendimiento predictivo de un modelo estadístico con predicciones fiables sobre casos desconocidos. En este sentido, se aplicó una novedosa combinación de técnicas robustas (ASCA y LASSO (operador de selección y contracción mínima absoluta) + LR (regresión logística)) para el análisis de datos. Aunque estas técnicas se han utilizado previamente de forma independiente para el análisis de VOCs del aire exhalado (ej. ASCA fue utilizado por van de Kant *et al.*⁴² y LASSO + LR por Monasta *et al.*⁴³), hasta donde sabemos nunca se han aplicado de manera conjunta. ASCA es una herramienta útil para grandes cohortes longitudinales, ya que muchos factores (por ejemplo, la estación del año en la que se llevan a cabo las medidas) pueden influir indirectamente en los resultados⁴⁴. Por otra parte, LASSO + LR permite el análisis de datos de alta dimensión no siendo una "caja negra", como lo son otras técnicas de aprendizaje supervisado, como las Máquinas de Vector Soporte (SVM) cuyos resultados son difíciles de interpretar. De hecho, la interpretación de LASSO+LR es tan sencilla como la de la regresión logística convencional, lo que resulta esencial en el ámbito médico. Otra ventaja de LASSO es su capacidad para seleccionar variables e identificar *features* discriminantes⁴⁵. En este caso, los modelos elaborados se validaron mediante validación cruzada y mediante validación sobre un conjunto independiente que no había participado en el proceso de elaboración del modelo. Hasta la fecha, pocos estudios llevan a cabo este tratamiento estadístico. Por supuesto, se trata de un proceso costoso desde el punto de vista computacional, pero merece la pena, para evitar el sesgo introducido por una estimación incorrecta del rendimiento, que puede ser considerable. Además, éste es el primer estudio centrado en la aplicación del análisis del aire exhalado mediante GC-MS en el estudio de la enfermedad del asma en el cual se evaluó la significación estadística de cada modelo mediante pruebas de permutación (la prueba de permutación permite calcular un p-valor)^{29,33}. No obstante, en el futuro sería interesante la validación del modelo en otro conjunto de sujetos que no pertenezcan a la cohorte NELA.

Este estudio ha demostrado que la discriminación entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas) mediante el análisis del aire exhalado es factible cuando se aplican las mejores prácticas típicas de la metabolómica. Todos los modelos construidos que incluían sujetos con asma fueron significativos (p-valor < 0,05 en las pruebas de permutación). Además, el

sobreajuste de los modelos fue bajo, ya que el rendimiento observado tanto en la validación cruzada como en la validación sobre el conjunto de testeo fue muy similar. Esto garantiza un buen rendimiento del modelo, extrapolable a nuevas muestras desconocidas. Además, aunque la precisión de los modelos que también incluían otras variables relacionadas con los antecedentes parentales atópicos y el recuento de glóbulos blancos (WBCs) era mayor que la de los modelos que sólo incluían variables relativas a los VOCs, se seleccionaron los mismos *features* en todos los modelos. De hecho, los *features* de la acetona, del derivado de la tetrahydroisoquinolina y del 2-etil-1-hexanol fueron seleccionados como discriminantes del asma con otras enfermedades atópicas coexistentes. En consecuencia, este sólido proceso de modelización confirmó que el análisis de VOCs junto con los antecedentes parentales atópicos permite discriminar el asma con otras enfermedades atópicas coexistentes.

Respecto al origen de los VOCs discriminantes, la acetona es un producto secundario de la peroxidación lipídica de los ácidos grasos poliinsaturados (PUFAs)^{46,47}, y previamente se ha descrito una asociación positiva entre el asma y los niveles de acetona en el aliento⁴⁸. Por su parte, el 2-etil-1-hexanol es un contaminante del aire de interiores, ya que es el principal metabolito del di(2-etilhexil)ftalato (DEHP), que es un plastificante frecuente del policloruro de vinilo (PVC)⁴⁹⁻⁵¹. Asimismo, los niveles de este compuesto aumentan con el incremento de la humedad relativa en los hogares⁵². Sin embargo, se comprobó que la intensidad del 2-etil-1-hexanol en las bolsas Tedlar® era insignificante en comparación con las muestras de aire exhalado (Apéndice 13). Además, aparte de ser un contaminante del aire de interiores, también está bien documentado que el 2-etil-1-hexanol es disruptor endocrino (EDC)⁵³⁻⁵⁵. Existen evidencias que sostienen que los EDCs están relacionados con el aumento del estrés oxidativo y que modulan la respuesta inmunológica⁵⁶. Estudios previos indican que la exposición al 2-etil-1-hexanol aumenta la activación de las células T CD4+ y la prevalencia del asma^{53,57}. Además, se ha observado un aumento significativo del 2-etil-1-hexanol en el aire exhalado de pacientes con cáncer de pulmón⁵⁸, así como, en líneas celulares cancerígenas de diferentes orígenes histológicos⁵⁹. Asimismo, el 2-etil-1-hexanol es también considerado una sustancia exógena que induce la proliferación de peroxisomas en el hígado⁶⁰. Por otro lado, derivados de la tetrahydroisoquinolina que contienen 5-fenil-2-furano presentan una considerable actividad inhibitoria de las fosfodiesterasas PD4, aumentando la concentración intracelular del segundo mensajero adenosín monofosfato cíclico (AMPc)⁶¹. De hecho, estos compuestos se han estudiado ampliamente como fármacos antiinflamatorios⁶¹. El esqueleto de la tetrahydroisoquinolina se encuentra habitualmente en los fármacos, sobre todo en los

relajantes musculares de amonio cuaternario. Los derivados de la tetrahydroisoquinolina pueden formarse en el organismo como metabolitos de algunos fármacos que suelen localizarse en la membrana celular, y su neurotoxicidad, entre otros aspectos, depende de su predisposición a formar radicales libres⁶². La producción endógena de derivados neurotóxicos de la tetrahydroisoquinolina, como el norsalsolinol, sigue siendo estudiada como posible causa de algunas afecciones, como la enfermedad de Parkinson⁶³. Sin embargo, no se observó ninguna relación significativa entre los fármacos notificados por las mujeres incluidas en nuestro estudio y los niveles del derivado de la tetrahydroisoquinolina (Apéndices 8-11). Por otra parte, aunque estudios anteriores demuestran que varios compuestos, como las hidroquinonas, pueden ser productos de degradación del relleno de los tubos Tenax⁶⁴, la intensidad observada del derivado de la tetrahydroisoquinolina en los tubos de desorción térmica reacondicionados fue marginal en comparación con la de las muestras de aire exhalado (Apéndice 13).

La mayoría de los estudios de volatilómica se centran exclusivamente en la comparación de pacientes con asma y controles sanos. Sin embargo, también es importante comparar el aire exhalado de los pacientes con asma y los que presentan síntomas similares³³. Además, dado que la mayoría de los asmáticos suelen presentar otras enfermedades coexistentes^{6,23}, es crucial tenerlas en cuenta para el correcto estudio de esta enfermedad. Por este motivo, en este estudio se compararon los perfiles de aire exhalado de A-AD y NA-AD, discriminando con éxito entre ambas categorías (se generaron cuatro modelos significativos). A este respecto, Dragonieri *et al.*⁶⁵ consiguieron distinguir entre pacientes con asma y rinitis alérgica, y pacientes sin asma pero con rinitis alérgica. Sin embargo, no pudieron determinar la identidad de los VOCs discriminantes, y sólo se obtuvo una "huella de aliento" porque en dicho estudio se empleó E-nose en lugar de tecnologías basadas en la espectrometría de masas. En nuestro estudio, los VOCs con *features* seleccionados como discriminantes de la enfermedad asmática entre los pacientes con otros trastornos atópicos en los modelos III.A, III.B, III.C y III.D fueron el disulfuro de carbono, un derivado de la tetrahydroisoquinolina, el 2-etil-1-hexanol y el decano. El disulfuro de carbono es un contaminante ambiental clasificado como neurotóxico que ha sido seleccionado previamente como discriminante del asma⁶⁶. Por su parte, el origen endógeno de los alcanos en el aliento exhalado es objeto de debate, ya que también pueden proceder de fuentes exógenas⁶⁷. En cualquier caso, el decano ha sido destacado como discriminante del asma alérgica por estudios anteriores^{68,69}. Tradicionalmente, los alcanos se han identificado como posibles biomarcadores para el diagnóstico del asma⁷⁰ porque algunos de ellos son producidos durante la peroxidación lipídica^{24,71,72} (Figura 1).

En este estudio, aparte de los VOCs, otras variables como los antecedentes parentales atópicos o los recuentos de glóbulos blancos, fueron incluidas en el proceso de construcción de los modelos con resultados satisfactorios. Todo ello está en consonancia con los resultados de estudios anteriores que también han demostrado que el análisis de VOCs en el aire exhalado puede ser compatible con otras estrategias ya establecidas⁷³⁻⁷⁵. En este sentido, los médicos y los científicos reclaman con urgencia un cambio radical de concepto en el diagnóstico del asma. Por consiguiente, los protocolos de diagnóstico deben basarse en una combinación de técnicas preferentemente no invasivas y que no dependan de un único método.

Limitaciones

El presente estudio tiene varias limitaciones. De hecho, no ha sido posible distinguir entre fenotipos de asma. Todos las asmáticas incluidas en el estudio, excepto una, presentaban otras enfermedades atópicas. Por lo tanto, no fue posible diferenciar entre asmáticas con otras enfermedades atópicas coexistentes y asmáticas sin otras enfermedades atópicas. Además, es posible que otros tratamientos farmacológicos no incluidos en los cuestionarios tengan un efecto de confusión que podría afectar a las muestras de aire exhalado.

CONCLUSIÓN

Los resultados del presente estudio muestran fundamentalmente que la distinción entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas) utilizando el análisis de VOCs en el aire exhalado es viable cuando se emplean las prácticas recomendadas típicas del campo de la metabolómica, es decir: se reclutó una gran cohorte de sujetos, se evaluó la influencia ambiental, se utilizó un flujo de trabajo reproducible para la etapa de preprocesamiento de datos, se llevó a cabo un análisis de datos robusto para la construcción de modelos, se evaluó el rendimiento de cada modelo mediante dos enfoques (validación cruzada de 5 iteraciones y validación sobre el conjunto de testeo), y se evaluó la significancia de los modelos mediante pruebas de permutación. Además, el resultado del proceso de modelización confirma que el análisis de VOCs junto con los antecedentes parentales atópicos del sujeto, puede ser una buena estrategia para el cribado del asma con otras enfermedades atópicas coexistentes. Asimismo, los resultados de este estudio confirman que el análisis de VOCs (por sí mismo o junto con otras técnicas establecidas) es útil para identificar la presencia de asma entre pacientes con otras enfermedades atópicas.

MATERIALES Y MÉTODOS

Análisis de los datos recogidos en la Encuesta Europea de Salud en España de 2014 (EHSS-2014) y en la Encuesta Nacional de Salud de España de 2017 (ENSE-2017)

La información recopilada tanto en la Encuesta Europea de Salud en España de 2014 (EHSS-2014) como en la Encuesta Nacional de Salud de España de 2017 (ENSE-2017), elaboradas por el Instituto Nacional de Estadística (INE) fue analizada utilizando R (versión 3.6.1). En concreto, se realizó un análisis de correspondencias múltiples (MCA) empleando el paquete *FactoMineR*⁷⁶. Además, para comprobar si existían diferencias significativas (p -valor $< 0,05$) en las variables entre asmáticos y no asmáticos se emplearon la prueba de chi-cuadrado o la prueba exacta de Fisher⁷⁷. En primer lugar, el análisis se realizó sobre los datos de todos los sujetos y, en segundo lugar, se realizó sólo sobre los datos de las mujeres en edad fértil (mujeres de 18 a 45 años).

Diseño del estudio y participantes

Los datos utilizados proceden de sujetos del estudio NELA (*Nutrition in Early Life and Asthma*), una cohorte prospectiva de nacimiento establecida en Murcia (España)^{34,41}. El protocolo del estudio fue revisado y aprobado por el Comité de Ética del Hospital Clínico Universitario Virgen de la Arrixaca (HCUVA) de acuerdo con las directrices de la Declaración de Helsinki. El consentimiento informado por escrito de las participantes fue obtenido en el momento del reclutamiento.

El reclutamiento de las embarazadas se realizó durante 36 meses (marzo 2015-abril 2018) durante el control ecográfico a las 20 semanas de gestación en la Unidad Materno-Fetal del HCUVA. Se realizaron varios controles a los sujetos reclutados: visita de seguimiento 1 (a las 20-24 semanas de gestación), visita de seguimiento 2 (a las 32 semanas de gestación), visita de seguimiento 3 (en el momento del parto), visita de seguimiento 4 (3 meses después del parto) (Apéndice 14). Los criterios de inclusión fueron: residencia habitual en el Área de Salud I y en algunos distritos de las Áreas de Salud VI y VII de la Región de Murcia; intención de vivir en el área de estudio durante al menos 2 años; intención de dar a luz en el hospital de referencia; origen caucásico español; 18-45 años de edad; embarazo único; concepción no asistida; y ecografía normal a las 20 semanas de gestación (sin malformaciones mayores). Los criterios de exclusión fueron: existencia de otras enfermedades crónicas; complicaciones del

embarazo (excepto diabetes gestacional y trastornos hipertensivos); y no tener intención de dar a luz en el hospital de referencia.

De las 1.350 mujeres invitadas a participar, 738 se inscribieron finalmente en el estudio NELA. La recogida del aire exhalado se llevó a cabo en la visita de seguimiento 4 (3 meses después del parto) entre mayo de 2017 y octubre de 2018. Durante ese periodo, se recogió el aire exhalado de 337 mujeres que fueron incluidas en el presente estudio (Apéndice 14).

Información sobre el asma y otros trastornos atópicos

Se recopiló información sobre las mujeres en edad reproductiva con antecedentes médicos de asma y otras manifestaciones atópicas mediante un cuestionario estructurado administrado en persona por entrevistadores entrenados a las 20-24 semanas de gestación (visita de seguimiento 1). Se consideró mujeres asmáticas a aquellas que respondieron afirmativamente a la siguiente pregunta "¿Le han diagnosticado alguna vez asma?". Se consideró mujeres con otros trastornos atópicos a aquellas que tuvieron una respuesta positiva a la siguiente pregunta "¿Le han diagnosticado alguna vez rinitis alérgica, dermatitis atópica, conjuntivitis alérgica, alergia a los alimentos o alergia a los medicamentos?". De esta manera, las mujeres se dividieron en cuatro categorías: asmáticas con otras enfermedades atópicas coexistentes (A-AD), no asmáticas con otras enfermedades atópicas (NA-AD), no asmáticas sin enfermedades atópicas (NA-NAD) y no asmáticas (NA) (esta categoría incluye tanto NA-AD como NA-NAD) (Figura 4). También se recabó información por medio de cuestionarios sobre los antecedentes paternos de asma (sí/no), rinitis alérgica (sí/no) y dermatitis atópica (sí/no).

Otras variables empleadas en el estudio

A través de cuestionarios realizados personalmente a las mujeres durante el embarazo, se obtuvo información sobre las siguientes variables: edad; clase social (definida como la ocupación durante el embarazo en función de la clase social más alta, utilizando una adaptación española muy extendida del sistema de codificación internacional ISCO88 : I-II, directivos/técnicos; III, trabajadores cualificados; IV-V, trabajadores semicualificados/sin cualificar; y desempleados)⁷⁸; nivel educativo (secundaria incompleta o menos, secundaria completa y universitaria); tabaquismo durante el embarazo (sí/no); consumo de antibióticos (sí/no); consumo de paracetamol (sí/no), consumo de corticoides inhalados (sí/no); y consumo de corticoides inyectables

(sí/no). Se calculó el índice de masa corporal (IMC) en base la altura y el peso pregestacional (Kg/m^2) indicado por las propias mujeres del estudio. El recuento de glóbulos blancos (WBCs) en las muestras de sangre de las embarazadas se determinó utilizando un analizador hematológico Sysmex® XN9000 (Sysmex Corporation, Kobe, Japón) que combina la dispersión de luz y la detección óptica con la impedancia eléctrica. Además, la información referida a los hábitos de consumo de tabaco (fumador (sí/no)) en el momento de la toma de muestras del aire exhalado se recopiló mediante cuestionarios realizados en persona, tres meses después de la gestación (visita de seguimiento 4). También se tuvo en cuenta la estación del año en la que se recogieron las muestras del aire exhalado (invierno, diciembre-febrero; primavera, marzo-mayo; verano, junio-agosto; y otoño, septiembre-noviembre).

Recogida de muestras de aire exhalado

La recogida de muestras de aire exhalado se realizó a los 3 meses del parto (visita de seguimiento 4) siguiendo el protocolo descrito en el Capítulo 4⁴¹. Brevemente, se recogió el aire exhalado en bolsas de muestreo de gas Tedlar® de 1 L. En concreto, se recogió una muestra de aire exhalado mixto (aire del espacio muerto, aire de las vías respiratorias y aire alveolar). A continuación, el aire exhalado contenido en las bolsas de muestreo se transfirió inmediatamente a tubos de desorción térmica (Tenax TA/carbograph 5td, Markes International) para evitar la difusión a través de la pared de la bolsa. Además, por cada muestra de aire exhalado, se recogió también una muestra de aire ambiental directamente en un tubo de desorción térmica utilizando una jeringa Easy-VOC (Markes International) para controlar las condiciones ambientales en el momento del muestreo. Casi todas las muestras se analizaron el mismo día que fueron recogidas, siendo el periodo máximo de almacenamiento inferior a una semana. Los tubos TD se almacenaron a 4 °C en caso de que el período de almacenamiento fuera superior a un día. Después de cada uso, los tubos TD eran calentados a 335°C durante 25 minutos para su reacondicionamiento. Las bolsas Tedlar® eran limpiadas con 10 lavados de nitrógeno (99,9% de pureza de nitrógeno) antes de su uso. De este modo, los niveles de artefactos de las bolsas de muestro (N,N-dimetilacetamida y fenol) se redujeron considerablemente (Apéndice 13).

Análisis del aire exhalado y preprocesamiento de datos

El análisis del aire exhalado y el preprocesamiento de datos se realizaron siguiendo un protocolo detallado previamente en el Capítulo 4⁴¹. Las muestras de aire exhalado se analizaron mediante un sistema de desorción térmica acoplado a la

cromatografía de gases-espectrometría de masas de cuadrupolo simple (TD-GC/q-MS). En el análisis por GC-MS, se obtuvieron cientos de *features* (picos de iones con un tiempo de retención y una señal m/z característica) a partir de la fragmentación de los compuestos del aire exhalado en espectros de masas⁷⁹. A continuación, los datos en brutos se convirtieron a formato unificado mzXML mediante *MsConvert* de *Proteowizard*^{80,81}. Posteriormente, se aplicó un flujo de trabajo de código abierto que empleaba las funciones de tres paquetes (*xcms*⁸², *cliqueMS*⁸³ y *eRah*⁷⁹) escritos en el lenguaje R para el preprocesamiento de los datos. Este flujo de trabajo permite la integración de las dos principales estrategias para el preprocesamiento de datos del análisis GC-MS (detección de *features* y detección de compuestos). De este modo, se obtuvo una matriz con las intensidades relativas de los *features* de las muestras de aire exhalado. Además, se pudo determinar qué *feature* detectado pertenecía a cada compuesto. Por otra parte, la identificación de compuestos se llevó a cabo mediante el cotejo con la biblioteca espectral NIST (*National Institute of Standard and Technology*) y el cálculo de dos factores (*match factor* y *error RI*) utilizando el paquete *eRah*⁷⁹. Para el cálculo del *error RI*, se utilizaron los tiempos de retención de dos estándares comerciales (estándar de alcanos saturados C7-C30 y estándar de calibración de VOCs, Sigma-Aldrich) y los índices de retención de los compuestos registrados en la biblioteca NIST.

Análisis de datos

Características de la población de estudio. Asociaciones entre las condiciones atópicas

Los sujetos del estudio (n = 336) se dividieron aleatoriamente en dos conjuntos según la fecha de muestreo: Grupo 1 (n = 211) y Grupo 2 (n = 125). Las muestras de aire exhalado recogidas entre mayo de 2017 y febrero de 2018 constituyeron el Grupo 1, y las muestras de aire exhalado recogidas entre marzo de 2018 y octubre de 2018 constituyeron el Grupo 2. Las características de la población del estudio se analizaron con la versión 3.6.1 de R para identificar las diferencias entre las asmáticas y no asmáticas y las asociaciones con otras condiciones atópicas. Se emplearon las pruebas de Shapiro-Wilk y Lilliefors (paquete *nortest*) para evaluar la distribución normal de los datos. Según la distribución de los datos, se realizaron pruebas paramétricas (prueba t de Student o prueba ANOVA) o no paramétricas (prueba U de Mann-Whitney o prueba de Kruskal-Wallis) para confirmar las diferencias estadísticamente significativas en las variables continuas (p -valor < 0,05). Además, se aplicó la prueba de chi-cuadrado o la prueba exacta de Fisher para evaluar las diferencias entre las variables categóricas.

Creación de modelos predictivos para la identificación de VOCs discriminantes en el análisis del aliento exhalado

Los modelos fueron construidos utilizando la versión 3.6.1 de R. El flujo de trabajo empleado para la generación de modelos predictivos se muestra en la Figura 7. Para el análisis de datos se emplearon las matrices con las intensidades relativas de los *features* filtrados obtenidos tras el preprocesamiento de datos. Además, se descartaron los *features* procedentes de los contaminantes de las bolsas de muestreo (N,N-dimetilacetamida y fenol), de los aceites de las bombas y de los siloxanos de las columnas de GC/MS. Para reducir la influencia de los contaminantes exógenos del aire ambiente, los valores de las intensidades no normalizadas de los *features* de las muestras humanas y de los *features* del aire ambiental fueron comparadas. Para ello, se realizaron pruebas de rango con signo de Wilcoxon para comparar las intensidades de los *features* entre las muestras ambientales y las muestras de aire exhalado humano de la cohorte NELA, eliminándose también los *features* con intensidades significativamente más altas en las muestras de aire ambiental. Los *features* elegidos tras este proceso de selección recibieron el nombre de *features* hiperfiltrados.

A continuación, se realizó un análisis exploratorio inicial de los *features* hiperfiltrados mediante un PCA (paquete *FactoMineR*⁷⁶) para buscar tendencias, agrupaciones y valores atípicos. Posteriormente, se aplicó ASCA (ANOVA- análisis de componentes simultáneos) (paquete *MetStaT*⁸⁴) sobre los *features* hiperfiltrados para reducir las posibles variaciones inducidas por diferentes factores experimentales. Este método permite separar las contribuciones de diferentes factores de la variación en la matriz de datos original para obtener una matriz residual libre de esa influencia^{42,44}.

En el proceso de modelización, las muestras del Grupo 1 se utilizaron como conjunto de entrenamiento; las muestras del Grupo 2 se usaron como conjunto de validación, como se muestra en la Figura 7. Los modelos se generaron mediante LASSO (operador de selección y contracción mínima absoluta) y regresión logística (LR) utilizando el paquete R *glmnet*⁸⁵.

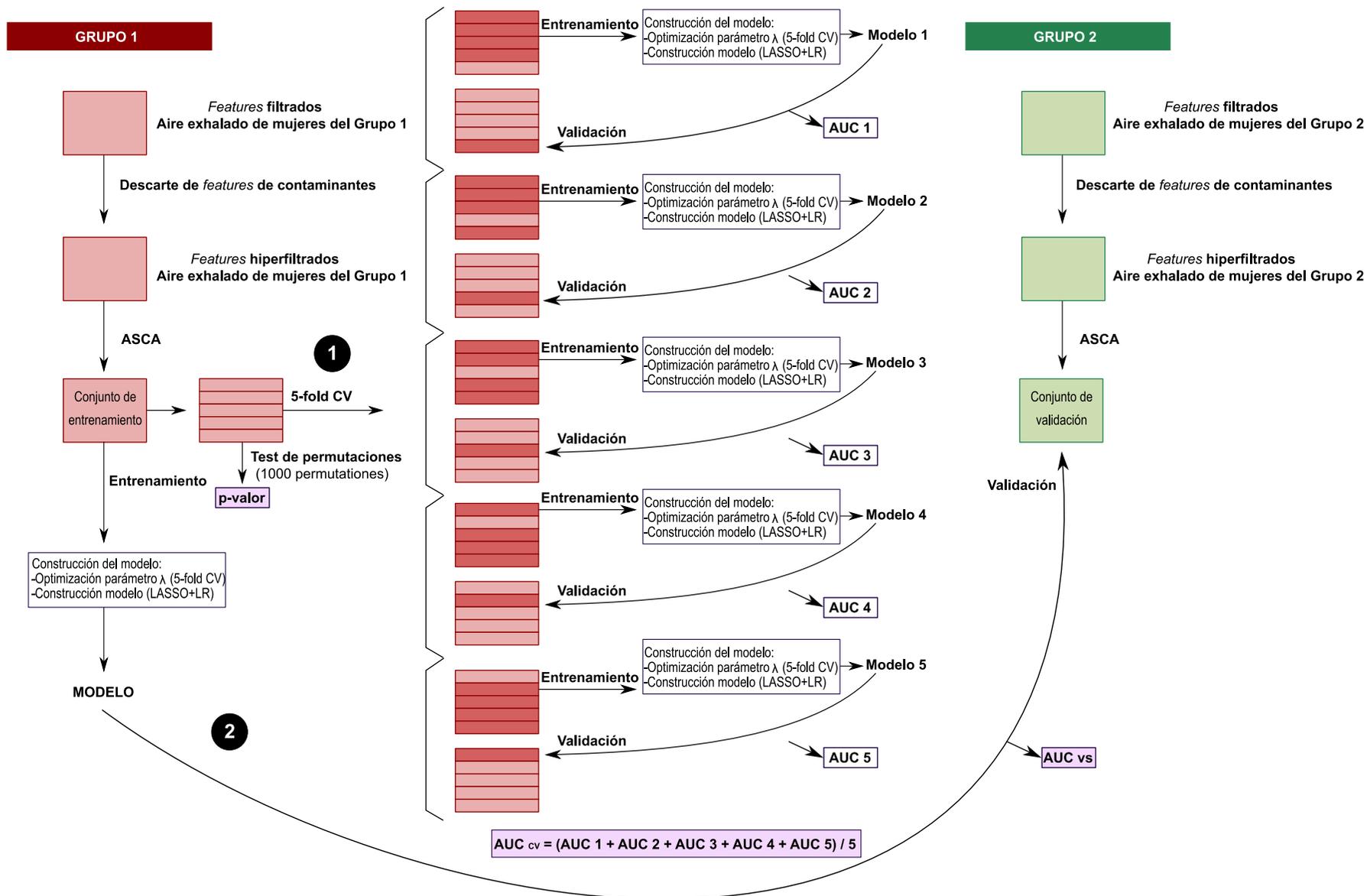


Figura 7. Flujo de trabajo para la elaboración de modelos. El rendimiento de los modelos se evaluó mediante dos enfoques: (1) validación cruzada de 5 iteraciones y (2) validación sobre el conjunto de testeo. 5-fold CV: validación cruzada de 5 iteraciones; ASCA: ANOVA-análisis de componentes simultáneos; LASSO: operador de selección y contracción mínima absoluta; LR: regresión logística; AUC: área bajo la curva ROC (*receiver operating characteristic curve*); AUC_{cv} : AUC obtenida en la validación cruzada de 5 iteraciones, AUC_{vs} : AUC obtenida en la validación en el conjunto de testeo.

El proceso de construcción del modelo implica dos pasos: (1) optimización de los parámetros y (2) construcción del modelo. LASSO + LR es un método lineal basado en una combinación de un método de contracción y una técnica de aprendizaje supervisado. LASSO + LR puede interpretarse como una regresión logística, ya que permite analizar la relación entre las variables y calcular los valores de las *odds ratio*. Sin embargo, el cálculo de los coeficientes es diferente al de los métodos lineales convencionales. Debido a los problemas de multicolinealidad y alta dimensionalidad, los coeficientes de los modelos obtenidos por los métodos lineales convencionales pueden alcanzar valores extremadamente altos. LASSO incorpora una penalización en la maximización de la verosimilitud mediante el parámetro λ durante el cálculo de los coeficientes. Dado que muchos coeficientes obtienen un valor de 0 tras la penalización, éstos son excluidos. De este modo, mediante la selección de variables es posible evitar o reducir el sobreajuste^{45,86}. El parámetro λ se optimizó mediante validación cruzada de 5 iteraciones.

Finalmente, el rendimiento del modelo se evaluó mediante dos enfoques. En primer lugar, el modelo construido se validó mediante una validación cruzada de cinco iteraciones. Las muestras del Grupo 1 se dividieron aleatoriamente en 5 subgrupos. Cuatro de ellos se utilizaron para el entrenamiento y el proceso de construcción del modelo. El subgrupo restante se empleó para la validación, y se determinaron curvas ROC con el paquete *pROC*⁸⁷ para obtener las AUCs. Los valores de corte de las curvas ROC se calcularon automáticamente en base al balance de casos/controles en el conjunto de entrenamiento de cada modelo. Este proceso se repitió 5 veces para cada modelo, de forma que todos los subgrupos fueron empleados como conjunto de validación o testeo. El valor de AUC obtenido en la validación cruzada, AUC_{CV} , se calculó como la media de los valores de AUC de los 5 submodelos. Además, se realizaron pruebas de permutación con 1000 permutaciones para obtener la significación estadística de los modelos construidos. En las pruebas de permutación, el p-valor se obtiene comparando el rendimiento de los modelos construidos y el de modelos predictivos generados con las etiquetas de las categorías permutadas aleatoriamente. La segunda estrategia consiste en evaluar el rendimiento del modelo construido en las muestras exhaladas del Grupo 2 (conjunto de validación) y obtener el AUC_{VS} mediante la validación sobre el conjunto de testeo.

De este modo, dieciséis modelos predictivos fueron construidos utilizando las matrices residuales, y sus características se pueden observar en la Figura 6 (panel superior). Aparte de los VOCs, variables como los antecedentes parentales atópicos, el

recuento de glóbulos blancos u otros trastornos atópicos fueron también incluidos en los modelos correspondientes.

BIBLIOGRAFÍA

1. Holgate, S. T. *et al.* Asthma. *Nat. Rev. Dis. Prim.* **1**, 15025 (2015).
2. Beckhaus, A. A. *et al.* Maternal nutrition during pregnancy and risk of asthma, wheeze, and atopic diseases during childhood: A systematic review and meta-analysis. *Allergy* **70**, 1588–1604 (2015).
3. Korten, I., Ramsey, K. & Latzin, P. Air pollution during pregnancy and lung development in the child. *Paediatr. Respir. Rev.* **21**, 38–46 (2017).
4. Hill, D. A. & Spergel, J. M. The Atopic March: Critical Evidence and Clinical Relevance. *Ann. Allergy, Asthma Immunol.* **120**, 131–137 (2018).
5. Amat, F. *et al.* New insights into the phenotypes of atopic dermatitis linked with allergies and asthma in children: An overview. *Clin. Exp. Allergy* **48**, 919–934 (2018).
6. Heck, S. *et al.* High probability of comorbidities in bronchial asthma in Germany. *npj Prim. Care Respir. Med.* **27**, 28 (2017).
7. Spergel, J. M. & Paller, A. S. Atopic dermatitis and the atopic march. *J. Allergy Clin. Immunol.* **112**, S118–S127 (2003).
8. Saunes, M. *et al.* Early eczema and the risk of childhood asthma: A prospective, population-based study. *BMC Pediatr.* **12**, 168 (2012).
9. Bantz, S. K., Zhu, Z. & Zheng, T. The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. *J. Clin. Cell. Immunol.* **5**, 202 (2014).
10. Pignataro, F. S., Bonini, M., Forgiione, A., Melandri, S. & Usmani, O. S. Asthma and gender: The female lung. *Pharmacol. Res.* **119**, 384–390 (2017).
11. Shah, R. & Newcomb, D. C. Sex Bias in Asthma Prevalence and Pathogenesis. *Front. Immunol.* **9**, 2997 (2018).
12. Qu, J., Li, Y., Zhong, W., Gao, P. & Hu, C. Recent developments in the role of reactive oxygen species in allergic asthma. *Journal of Thoracic Disease* **9**, E32–E43 (2017).
13. Kirkham, P. & Rahman, I. Oxidative stress in asthma and COPD: Antioxidants as a therapeutic strategy. *Pharmacology and Therapeutics* **111**, 476–494 (2006).
14. Jesenak, M., Zelieskova, M. & Babusikova, E. Oxidative stress and bronchial asthma in children-causes or consequences? *Frontiers in Pediatrics* **5**, 162 (2017).

15. Zuo, L., Otenbaker, N. P., Rose, B. A. & Salisbury, K. S. Molecular mechanisms of reactive oxygen species-related pulmonary inflammation and asthma. *Molecular Immunology* **56**, 57–63 (2013).
16. Yoboue, E. D., Sitia, R. & Simmen, T. Redox crosstalk at endoplasmic reticulum (ER) membrane contact sites (MCS) uses toxic waste to deliver messages. *Cell Death and Disease* **9**, 331 (2018).
17. Brigham, E. P. & West, N. E. Diagnosis of asthma: Diagnostic testing. *Int. Forum Allergy Rhinol.* **5**, S27–S30 (2015).
18. Tenero, L., Zaffanello, M., Piazza, M. & Piacentini, G. Measuring Airway Inflammation in Asthmatic Children. *Front. Pediatr.* **6**, 196 (2018).
19. Price, D. B. *et al.* Blood eosinophil count and prospective annual asthma disease burden: A UK cohort study. *Lancet Respir. Med.* **3**, 849–858 (2015).
20. Gibson, P. G. Variability of blood eosinophils as a biomarker in asthma and COPD. *Respirology* **23**, 12–13 (2018).
21. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
22. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
23. Broza, Y. Y., Braverman, I. & Haick, H. Breath volatolomics for diagnosing chronic rhinosinusitis. *Int. J. Nanomedicine* **13**, 4661–4670 (2018).
24. Calenic, B. *et al.* Oxidative stress and volatile organic compounds: Interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* **13**, 1020–1030 (2015).
25. Pijnenburg, M. W. The role of FeNO in predicting asthma. *Front. Pediatr.* **7**, 41 (2019).
26. Zhang, D. *et al.* Measurement of fractional exhaled nitric oxide and nasal nitric oxide in male patients with obstructive sleep apnea. *Sleep Breath.* **23**, 785–793 (2018).
27. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
28. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking

- your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
29. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
 30. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
 31. Cavaleiro Rufo, J., Madureira, J., Oliveira Fernandes, E. & Moreira, A. Volatile organic compounds in asthma diagnosis: A systematic review and meta-analysis. *Allergy: European Journal of Allergy and Clinical Immunology* **71**, 175–188 (2016).
 32. Horváth, I. *et al.* A european respiratory society technical standard: Exhaled biomarkers in lung disease. *European Respiratory Journal* **49**, 1600965 (2017).
 33. Sola Martínez, R. A. *et al.* Exhaled volatile organic compounds analysis in clinical pediatrics: a systematic review. *Pediatr. Res.* **89**, 1352–1363 (2021).
 34. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).
 35. García-Serna, A. M. *et al.* Air pollution from traffic during pregnancy impairs newborn's cord blood immune cells: The NELA cohort. *Environ. Res.* **198**, 110468 (2021).
 36. Skamarock, W. C. *et al.* *A Description of the Advanced Research WRF Version 3. Technical Report* (2008).
 37. Huang, H. L. *et al.* Bronchial asthma is associated with increased risk of chronic kidney disease. *BMC Pulm. Med.* **14**, 1–8 (2014).
 38. Gong, P. H. *et al.* Acute severe asthma with thyroid crisis and myasthenia: a case report and literature review. *Clinical Respiratory Journal* **11**, 671–676 (2017).
 39. Bingyan, Z. & Dong, W. Impact of thyroid hormones on asthma in older adults. *J. Int. Med. Res.* **47**, 4114–4125 (2019).
 40. Krysiak, R., Kowalcze, K. & Okopień, B. The effect of testosterone on thyroid autoimmunity in euthyroid men with Hashimoto's thyroiditis and low testosterone levels. *J. Clin. Pharm. Ther.* **44**, 742–749 (2019).
 41. Sola Martínez, R. A. *et al.* Data preprocessing workflow for exhaled breath analysis by GC/MS using open sources. *Sci. Rep.* **10**, 22008 (2020).

42. van de Kant, K. D. G. *et al.* Exhaled breath profiling in diagnosing wheezy preschool children. *Eur. Respir. J.* **41**, 183–188 (2013).
43. Monasta, L. *et al.* Inflammatory bowel disease and patterns of volatile organic compounds in the exhaled breath of children: A case-control study using Ion Molecule Reaction-Mass Spectrometry. *PLoS One* **12**, e0184118 (2017).
44. Smilde, A. K. *et al.* ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **21**, 3043–3048 (2005).
45. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction.* (Springer, 2009).
46. Silva, C. L., Perestrelo, R., Silva, P., Tomás, H. & Câmara, J. S. Volatile metabolomic signature of human breast cancer cell lines. *Sci. Rep.* **7**, 43969 (2017).
47. Scholpp, J., Schubert, J. K., Miekisch, W. & Geiger, K. Breath markers and soluble lipid peroxidation markers in critically ill patients. *Clin. Chem. Lab. Med.* **40**, 587–594 (2002).
48. Dragonieri, S. *et al.* An electronic nose in the discrimination of patients with asthma and controls. *J. Allergy Clin. Immunol.* **120**, 856–862 (2007).
49. Veenaas, C., Ripszam, M. & Haglund, P. Analysis of volatile organic compounds in indoor environments using thermal desorption with comprehensive two-dimensional gas chromatography and high-resolution time-of-flight mass spectrometry. *J. Sep. Sci.* **43**, 1489–1498 (2020).
50. Araki, A. *et al.* Validation of diffusive mini-samplers for aldehyde and VOC and its feasibility for measuring the exposure levels of elementary school children. *Journal of Environmental Monitoring* **14**, 368–374 (2012).
51. Bahrmann, H., Hahn, H.-D. & Mayer, D. 2-Ethylhexanol. in *Ullmann's Encyclopedia of Industrial Chemistry* 579–584 (Wiley-VCH Verlag GmbH & Co. KGaA, 2013).
52. Markowicz, P. & Larsson, L. Influence of relative humidity on VOC concentrations in indoor air. *Environ. Sci. Pollut. Res.* **22**, 5772–5779 (2015).
53. Paciência, I. *et al.* Exposure to indoor endocrine-disrupting chemicals and childhood asthma and obesity. *Allergy* **74**, 1277–1291 (2019).
54. Zhang, H. *et al.* Bioremediation of di-(2-ethylhexyl) phthalate contaminated red soil by *Gordonia terrae* RL-JC02: Characterization, metabolic pathway and kinetics. *Sci. Total Environ.* **733**, 139138 (2020).

55. Wakayama, T. *et al.* Comprehensive review of 2-ethyl-1-hexanol as an indoor air pollutant. *Journal of Occupational Health* **61**, 19–35 (2019).
56. Sur, U., Erkekoglu, P., Bulus, A. D., Andiran, N. & Kocer-Gumusel, B. Oxidative stress markers, trace elements, and endocrine disrupting chemicals in children with Hashimoto's thyroiditis. *Toxicol. Mech. Methods* **29**, 633–643 (2019).
57. Yoshida, Y. *et al.* The indoor air pollutant 2-ethyl-hexanol activates CD4 cells. *Chem. Biol. Interact.* **177**, 137–141 (2009).
58. Peng, G. *et al.* Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat. Nanotechnol.* **4**, 669–673 (2009).
59. Jia, Z. *et al.* Detection of Lung Cancer: Concomitant Volatile Organic Compounds and Metabolomic Profiling of Six Cancer Cell Lines of Different Histological Origins. *ACS Omega* **3**, 5131–5140 (2018).
60. Jiang, M. & Yang, N. Peroxisome Proliferators. in *Encyclopedia of Toxicology: Third Edition* (ed. Wexler, P.) 815–819 (Elsevier, 2014).
61. Li, Y. S. *et al.* Tetrahydroquinoline and tetrahydroisoquinoline derivatives as potential selective PDE4B inhibitors. *Bioorganic Med. Chem. Lett.* **28**, 3271–3275 (2018).
62. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
63. Scholz, J., Klingemann, I. & Moser, A. Increased systemic levels of norsalsolinol derivatives are induced by levodopa treatment and do not represent biological markers of Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **75**, 634–636 (2004).
64. Klenø, J. G., Wolkoff, P., Clausen, P. A., Wilkins, C. K. & Pedersen, T. Degradation of the adsorbent tenax TA by nitrogen oxides, ozone, hydrogen peroxide, OH radical, and limonene oxidation products. *Environ. Sci. Technol.* **36**, 4121–4126 (2002).
65. Dragonieri, S., Quaranta, V. N., Carratu, P., Ranieri, T. & Resta, O. Exhaled breath profiling by electronic nose enabled discrimination of allergic rhinitis and extrinsic asthma. *Biomarkers* **24**, 70–75 (2019).
66. Dallinga, J. W. *et al.* Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clin. Exp. Allergy* **40**, 68–76 (2009).
67. Mitsui, T., Kondo, T., Phillips, M., Cataneo, R. N. & Greenberg, J. Inadequacy of theoretical basis of breath methylated alkane contour for assessing oxidative

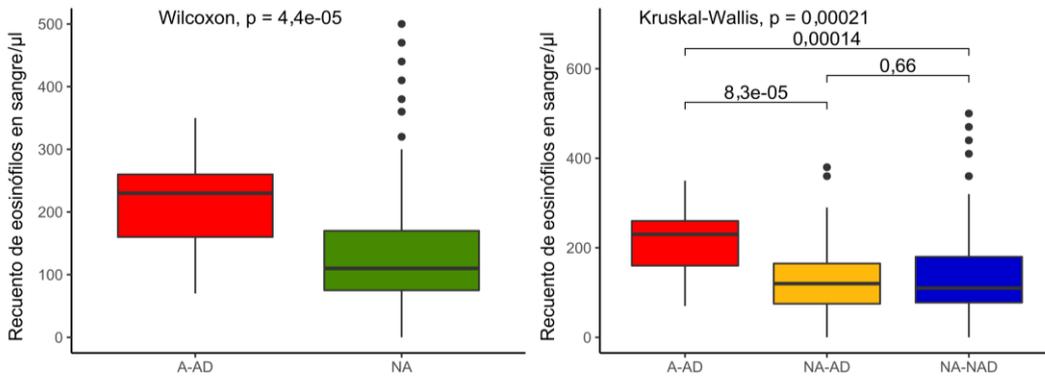
- stress. *Clinica Chimica Acta* **333**, 91–94 (2003).
68. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–80 (2011).
 69. Caldeira, M. *et al.* Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **1254**, 87–97 (2012).
 70. Ibrahim, B. *et al.* Non-invasive phenotyping using exhaled volatile organic compounds in asthma. *Thorax* **66**, 804–809 (2011).
 71. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. . Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
 72. Phillips, M. *et al.* Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress. *J. Lab. Clin. Med.* **136**, 243–249 (2000).
 73. Drake, S. M., Simpson, A. & Fowler, S. J. Asthma Diagnosis: The Changing Face of Guidelines. *Pulm. Ther.* **5**, 103–115 (2019).
 74. Klaassen, E. M. M. *et al.* Exhaled Biomarkers and Gene Expression at Preschool Age Improve Asthma Prediction at 6 Years of Age. *Am. J. Respir. Crit. Care Med.* **191**, 201–207 (2015).
 75. Schleich, F. N. *et al.* Exhaled Volatile Organic Compounds Are Able to Discriminate between Neutrophilic and Eosinophilic Asthma. *Am. J. Respir. Crit. Care Med.* **200**, 444–453 (2019).
 76. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
 77. Kim, H.-Y. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor. Dent. Endod.* **42**, 152 (2017).
 78. Domingo-Salvany, A., Regidor, E., Alonso, J. & Alvarez-Dardet, C. Proposal for a social class measure. Working Group of the Spanish Society of Epidemiology and the Spanish Society of Family and Community Medicine. *Aten. Primaria* **25**, 350–363 (2000).
 79. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).

80. Martens, L. *et al.* mzML--a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133 (2011).
81. Adusumilli, R. & Mallick, P. Data conversion with proteoWizard msConvert. in *Proteomics: Methods and Protocols* (eds. Comai, U., Katz, J. E. & Mallick, P.) **1550**, 339–368 (Humana, New York, NY, 2017).
82. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
83. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
84. Bertinetto, C., Engel, J. & Jansen, J. ANOVA simultaneous component analysis: A tutorial review. *Analytica Chimica Acta: X* **6**, 100061 (2020).
85. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
86. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
87. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, (2011).

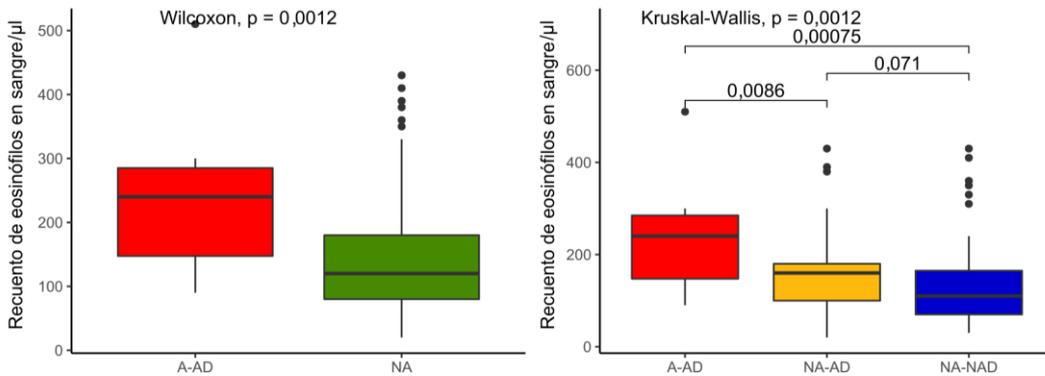
APÉNDICE 1. Análisis de asociaciones entre la presencia de asma y otras variables recogidas en la Encuesta Europea de Salud en España de 2014 (EHSS-2014) y en la Encuesta Nacional de Salud de España de 2017 (ENSE-2017).

Variables	Categorías de las variables	EHSS-2014		ENSE-2017	
		Todos los sujetos	Mujeres en edad fértil	Todos los sujetos	Mujeres en edad fértil
		p-valor	p-valor	p-valor	p-valor
Edad	-Adolescentes (15-18 años) -Adultos (18-45 años) -Mediana edad (45-65 años) -Ancianos (>65 años)	8,827e-06	-	1,689e-06	-
Sexo	-Hombre -Mujer	1,291e-05	-	3,041e-07	-
Nacionalidad	-Español -Extranjero	0,072	0,037	0,026	0,078
Estado de salud	-Estado Salud_Muy bueno -Estado Salud_Bueno -Estado Salud_Regular -Estado Salud_Malo -Estado Salud_Muy malo	< 2,2e-16	6,344e-12	< 2,2e-16	1,587e-09
Hipertensión arterial	-Hipertensión arterial_Sí -Hipertensión arterial_No	1,472e-05	0,024	0,002	6,081e-04
Otros trastornos atópicos (excluyendo el asma)	-Otros trastornos atópicos_Sí -Otros trastornos atópicos_No	< 2,2e-16	< 2,2e-16	< 2,2e-16	< 2,2e-16
Diabetes	-Diabetes_Sí -Diabetes_No	8,113e-05	0,792	5,024e-07	0,301
Colesterol	-Colesterol_Sí -Colesterol_No	0,126	0,010	0,002	0,375
Trastornos de la piel	-Trastornos piel_Sí -Trastornos piel_No	1,056e-14	5,586e-06	< 2,2e-16	5,299e-06
Cirrosis	-Cirrosis_Sí -Cirrosis_No	0,008	0,122	0,041	0,321
Depresión	-Depresión_Sí -Depresión_No	< 2,2e-16	0,075	< 2,2e-16	0,041
Trastornos de ansiedad	-Ansiedad_Sí -Ansiedad_No	< 2,2e-16	6,547e-05	1,961e-14	0,024
Problemas renales	-Problemas riñón_Sí -Problemas riñón_No	6,601e-06	0,135	7,152e-08	0,387
Tumores malignos	-Tumores malignos_Sí -Tumores malignos_No	0,122	0,104	1,118e-04	0,424
Problemas de tiroides	-Problemas tiroides_Sí -Problemas tiroides_No	5,288e-07	0,146	0,001	0,812
Osteoporosis	-Osteoporosis_Sí -Osteoporosis_No	5,217e-08	1	2,216e-12	0,262
Anticonceptivos orales	-Anticonceptivos orales_Sí -Anticonceptivos orales_No	0,205	0,476	0,841	0,917
IMC	-Peso bajo -Peso saludable -Sobrepeso -Obesidad -Obesidad severa	5,963e-08	3,231e-04	2,197e-11	0,506
Actividad física	-Actividad física_No -Actividad física_ocasional -Actividad física_mensual -Actividad física_semanal	6,965e-05	0,942	0,095	0,093
Dieta	-Dieta mala -Dieta regular -Dieta buena	0,281	0,037	0,240	0,693
Tabaco	-Fumador_diario -Fumador_ocasional -Ex-fumador -No fumador	0,006	0,630	8,206e-05	0,342
Alcohol	-Alcohol_diario -Alcohol_semanal -Alcohol_mensual -Alcohol_anual -Ex-alcohólico -Alcohol_nunca	0,092	0,122	0,003	0,004

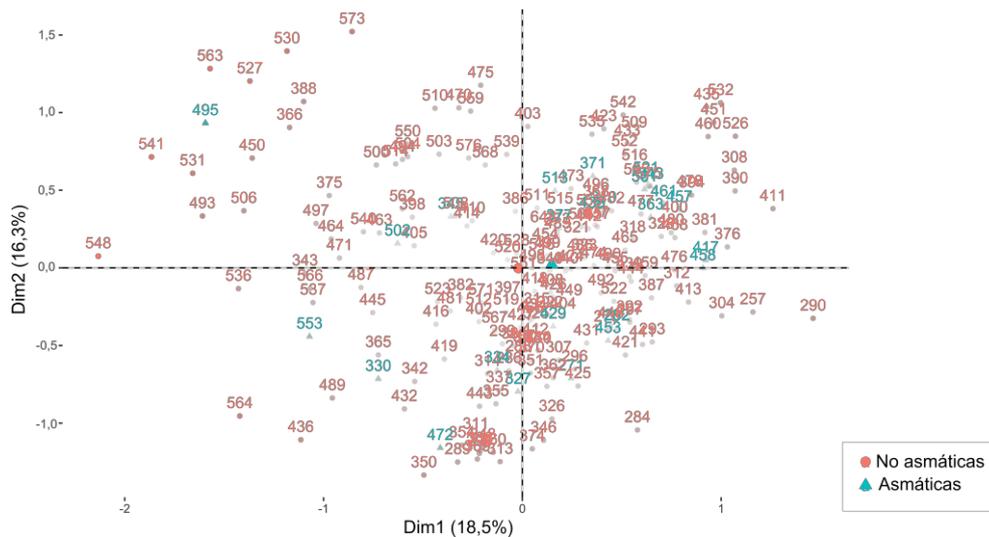
(A) Recuento de eosinófilos en sangre en las mujeres del Grupo 1



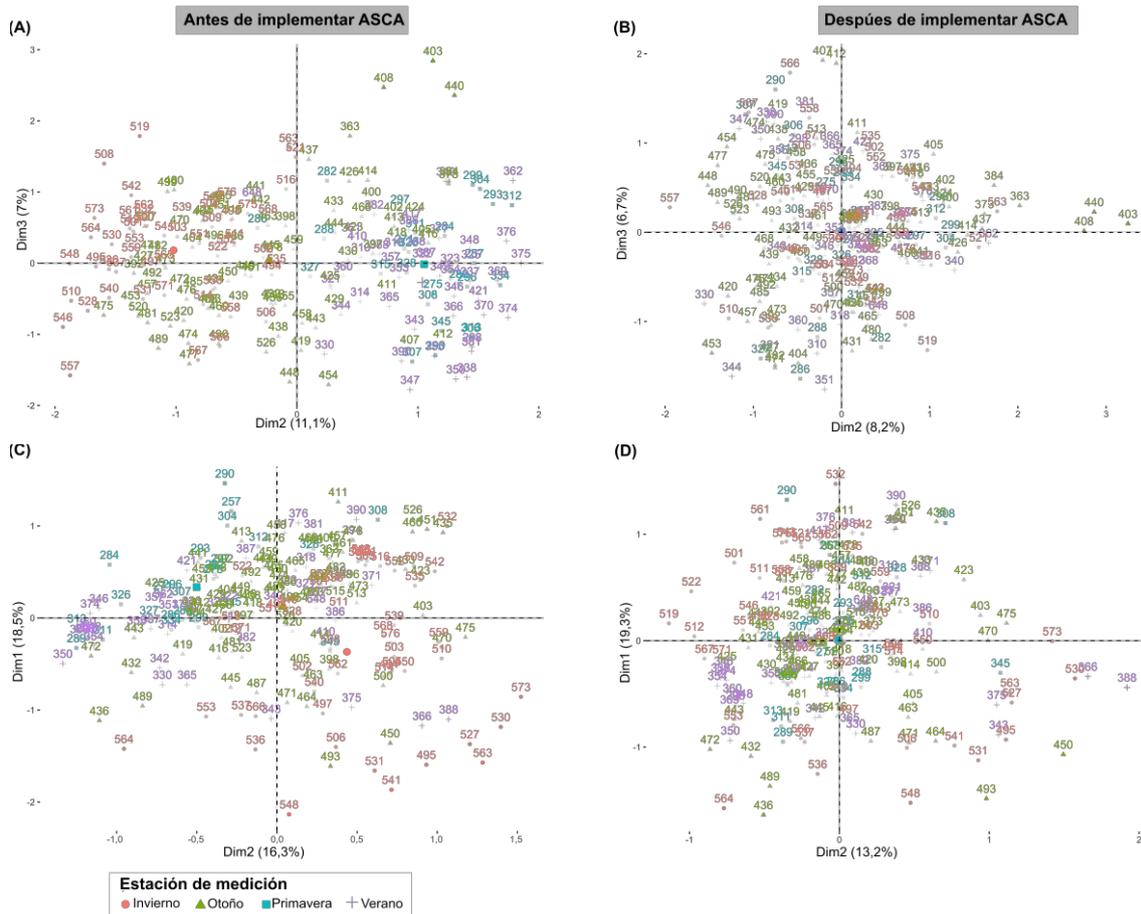
(B) Recuento de eosinófilos en sangre en las mujeres del Grupo 2



APÉNDICE 2. Recuento de eosinófilos en sangre en mujeres en edad fértil. (A) Recuento de eosinófilos en sangre en las mujeres del Grupo 1. (B) Recuento de eosinófilos en sangre en las mujeres del Grupo 2. A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA: no asmáticas; NA-NAD: no asmáticas sin enfermedades atópicas; NA-AD: no asmáticas con otras enfermedades atópicas.



APÉNDICE 3. Análisis exploratorio de la influencia de la enfermedad del asma en el aire exhalado de las mujeres. Representación de los resultados del PCA (análisis de componentes principales) realizado sobre los *features* filtrados del aire exhalado de las mujeres del Grupo 1 tras descartar los *features* de contaminantes exógenos (*features* hiperfiltrados).



APÉNDICE 4. Análisis exploratorio de la variación estacional en las muestras de aire exhalado humano y las muestras de aire ambiental. (A) Representación del PCA (Análisis de Componentes Principales) realizado sobre las muestras de aire ambiental del Grupo 1 antes de la implementación de ASCA (ANOVA- análisis de componentes simultáneos). (B) Representación del PCA realizado sobre las muestras de aire ambiental del Grupo 1 después de la implementación de ASCA. (C) Representación del PCA realizado sobre las muestras de aire exhalado de las mujeres del Grupo 1 antes de la implementación de ASCA. (D) Representación del PCA realizado sobre las muestras de aire exhalado de las mujeres del Grupo 1 después de la implementación de ASCA.

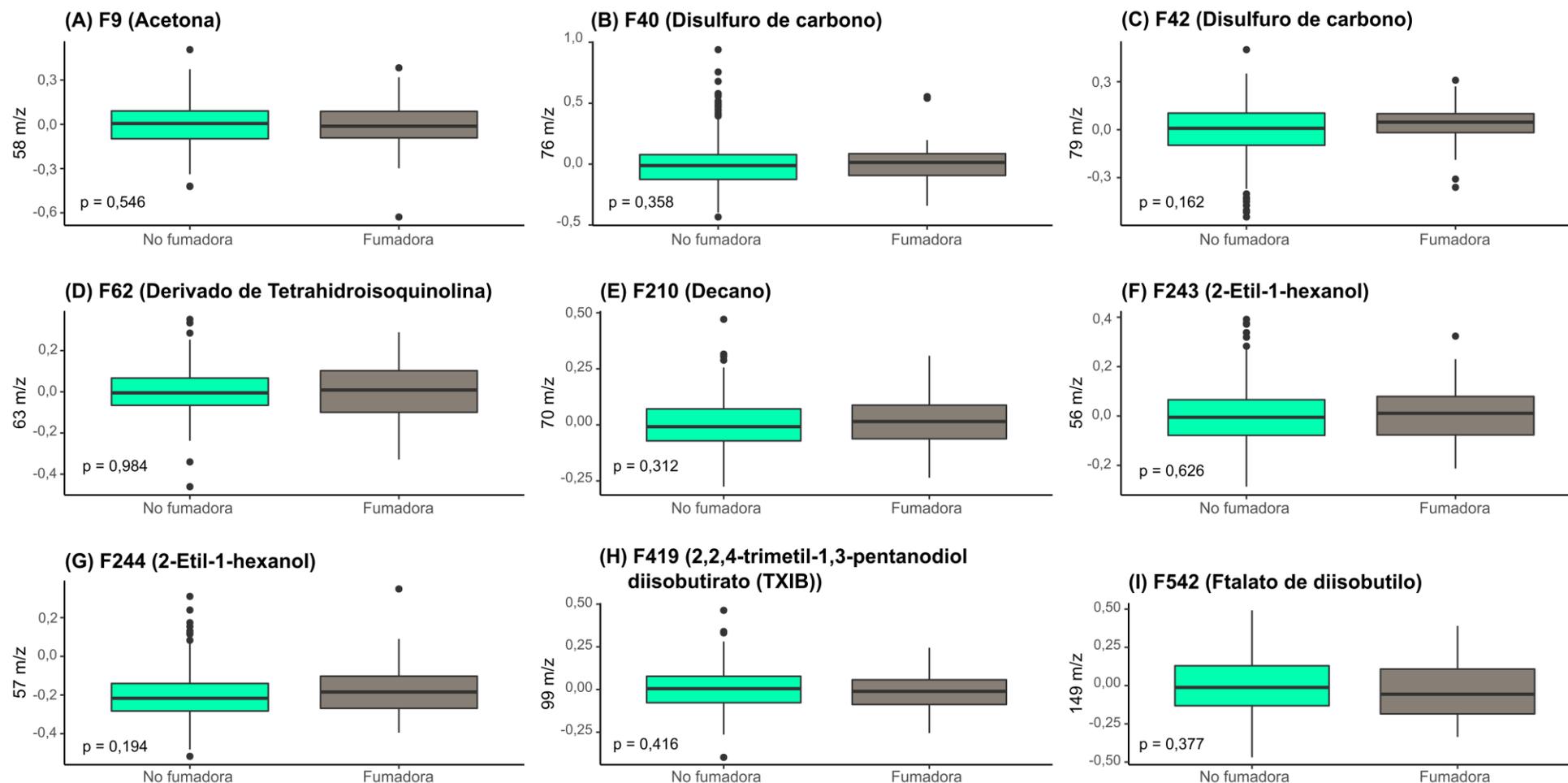
APÉNDICE 5. Fórmulas de los modelos generados.

Modelos	Probabilidad predicha (y)
I) A-AD vs NA	
I.A	$y = 1 / (1 + \exp(- (0,014 * F9 + 1,232 * F62 + 1,894 * F243 - 2,042)))$
I.B	$y = 1 / (1 + \exp(- (0,388 * F9 + 1,526 * F62 + 1,861 * F243 + 0,840 * \text{Antecedentes parentales de asma} + 0,261 * \text{Antecedentes parentales de rinitis} + 0,164 * \text{Antecedentes parentales de dermatitis} - 2,217)))$
I.C	$y = 1 / (1 + \exp(- (0,138 * F62 + 0,845 * F243 + 3,171 * \text{Recuento de eosinófilos en sangre} - 2,498)))$
I.D	$y = 1 / (1 + \exp(- (0,687 * F62 + 1,200 * F243 + 3,478 * \text{Recuento de eosinófilos en sangre} + 0,552 * \text{Antecedentes parentales de asma} + 0,160 * \text{Antecedentes parentales de rinitis} - 2,660)))$
I.E	$y = 1 / (1 + \exp(- (0,038 * F42 + 0,724 * F62 + 1,060 * F243 + 1,777 * \text{Rinitis Alérgica} + 0,653 * \text{Dermatitis Atópica} + 0,646 * \text{Conjuntivitis alérgica} + 0,337 * \text{Alergia alimentaria} + 0,348 * \text{Alergia a medicamentos} - 3,362)))$
I.F	$y = 1 / (1 + \exp(- (0,216 * F9 - 0,296 * F40 + 0,231 * F42 + 1,134 * F62 + 1,388 * F243 + 0,579 * \text{Antecedentes parentales de asma} + 0,272 * \text{Antecedentes parentales de dermatitis} + 1,872 * \text{Rinitis Alérgica} + 0,675 * \text{Dermatitis Atópica} + 0,652 * \text{Conjuntivitis alérgica} + 0,473 * \text{Alergia alimentaria} + 0,530 * \text{Alergia a medicamentos} - 3,582)))$
I.G	$y = 1 / (1 + \exp(- (0,356 * F40 + 0,981 * F62 + 0,881 * F243 + 0,421 * \text{Antecedentes parentales de asma} + 0,126 * \text{Antecedentes parentales de dermatitis} + 4,287 * \text{Recuento de eosinófilos en sangre} + 1,892 * \text{Rinitis Alérgica} + 0,609 * \text{Dermatitis Atópica} + 0,599 * \text{Conjuntivitis alérgica} + 0,703 * \text{Alergia alimentaria} + 0,541 * \text{Alergia a medicamentos} - 4,232)))$
II) A-AD vs NA-NAD	
II.A	$y = 1 / (1 + \exp(- (0,282 * F9 + 0,789 * F62 + 1,292 * F243 - 1,544)))$
II.B	$y = 1 / (1 + \exp(- (0,648 * F9 + 1,292 * F62 + 1,754 * F243 + 0,809 * \text{Antecedentes parentales de asma} + 0,694 * \text{Antecedentes parentales de rinitis} - 1,799)))$
II.C	$y = 1 / (1 + \exp(- (0,036 * F9 + 0,503 * F62 + 1,157 * F243 + 2,660 * \text{Recuento de eosinófilos en sangre} - 1,955)))$
II.D	$y = 1 / (1 + \exp(- (0,0004 * F9 + 0,582 * F62 + 1,233 * F243 + 2,431 * \text{Recuento de eosinófilos en sangre} + 0,562 * \text{Antecedentes parentales de asma} + 0,605 * \text{Antecedentes parentales de rinitis} - 2,101)))$
III) A-AD vs NA-AD	
III.A	$y = 1 / (1 + \exp(- (0,301 * F40 + 0,267 * F42 + 0,683 * F62 + 0,287 * F210 + 1,232 * F243 + 0,144 * F419 - 1,077)))$
III.B	$y = 1 / (1 + \exp(- (0,829 * F40 + 0,472 * F42 + 1,245 * F62 + 0,589 * F210 + 1,760 * F243 + 0,100 * F419 + 0,104 * \text{Antecedentes parentales de asma} + 0,146 * \text{Antecedentes parentales de dermatitis} - 1,099)))$
III.C	$y = 1 / (1 + \exp(- (0,827 * F40 + 1,148 * F62 + 0,420 * F210 + 0,925 * F243 + 0,124 * F542 + 6,886 * \text{Recuento de eosinófilos en sangre} - 0,155 * \text{Recuento de linfocitos en sangre} - 1,885)))$
III.D	$y = 1 / (1 + \exp(- (1,379 * F40 + 1,613 * F62 + 0,700 * F210 + 1,439 * F243 + 0,008 * F244 + 0,433 * F542 + 7,614 * \text{Recuento de eosinófilos en sangre} - 0,264 * \text{Recuento de linfocitos en sangre} + 0,034 * \text{Antecedentes parentales de asma} + 0,398 * \text{Antecedentes parentales de dermatitis} - 1,846)))$
IV) NA-AD vs NA-NAD	
IV.A	$y = 1 / (1 + \exp(- (0,201 * F23 + 0,461 * F360 - 0,65 * F421 - 0,461)))$

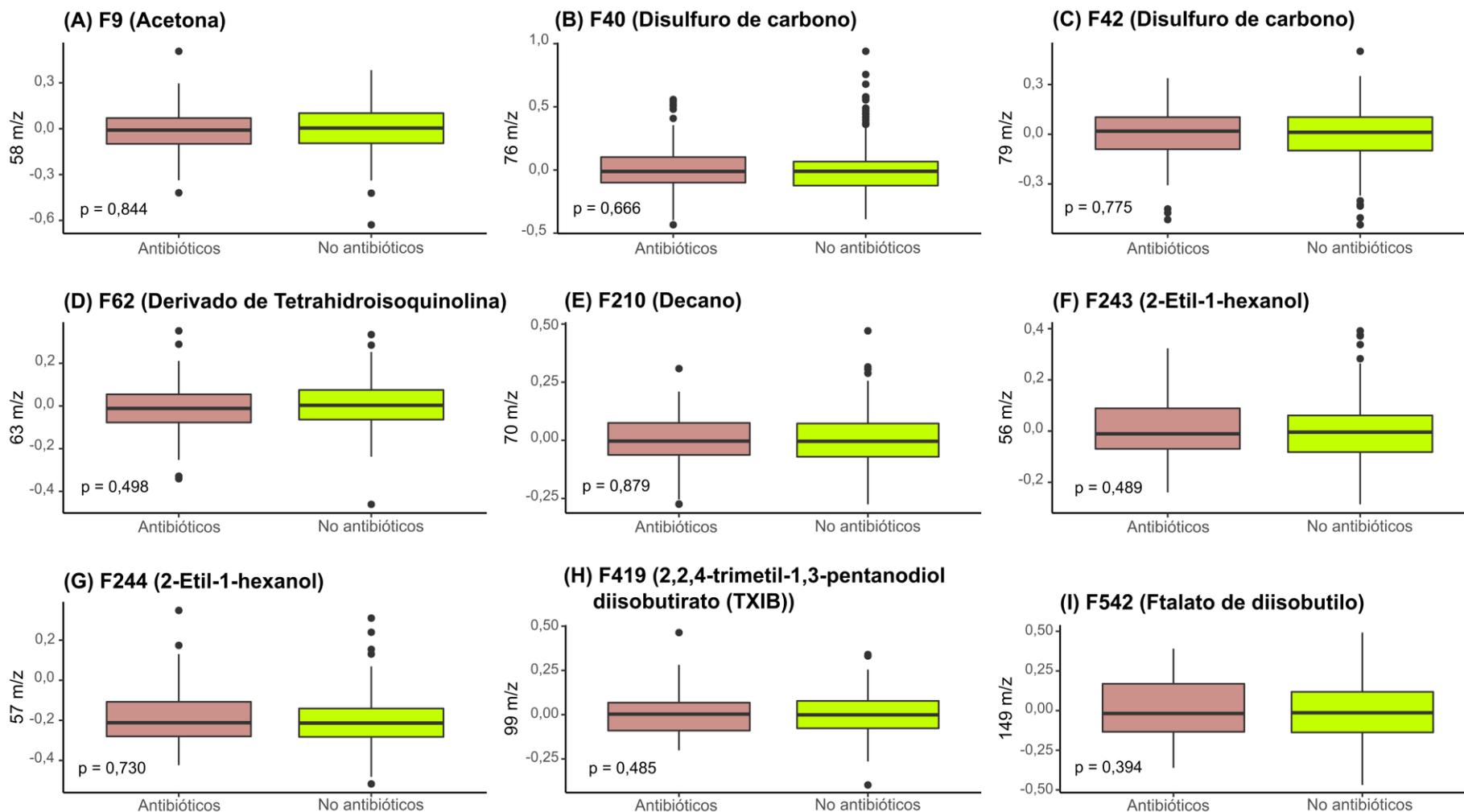
A-AD: asmáticas con otras enfermedades atópicas coexistentes; NA-AD: no asmáticas con otras enfermedades atópicas; NA-NAD: no asmáticas sin enfermedades atópicas; NA: no asmáticas.

APÉNDICE 6. Características de los *features* seleccionados por los modelos generados.

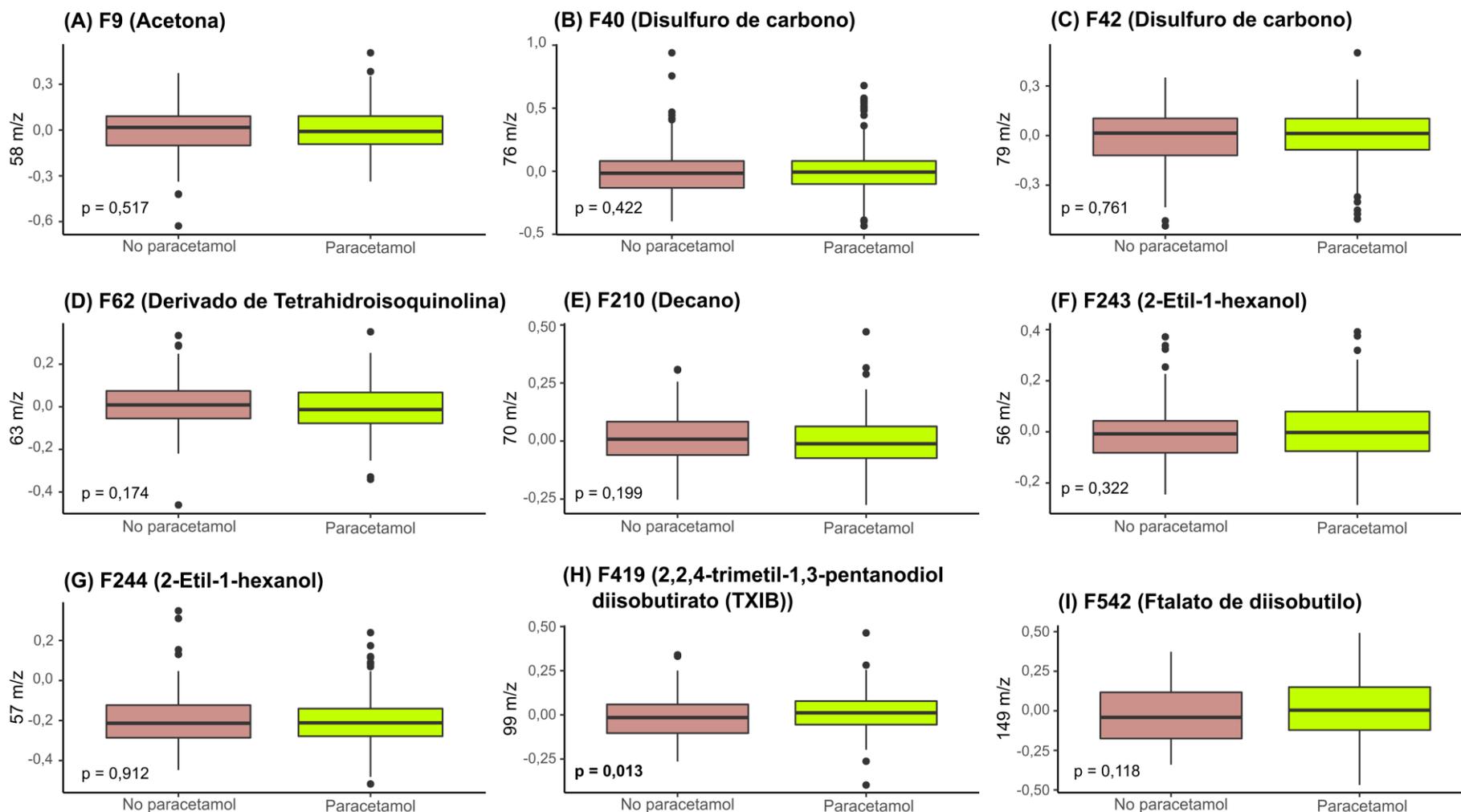
<i>Features</i>	<i>m/z</i>	Modelos	Compuestos Orgánicos Volátiles (VOCs)	Código CAS	Match factor y error RI	Solamente Match factor
F9	58	I.A, I.B, I.F, II.A, II.B, II.C, II.D	Acetona	67-64-1	X	
F23	50	IV.A	Isopreno	78-79-5		X
F40	76	I.F, I.G, III.A, III.B, III.C, III.D	Disulfuro de carbono	75-15-0	X	
F42	79	I.E, III.A, III.B				
F62	63	I.A, I.B, I.C, I.D, I.E, I.F, I.G, II.A, II.B, II.C, II.D, III.A, III.B, III.C, III.D	Derivado de Tetrahydroisoquinolina			X
F210	70	III.A, III.B, III.C, III.D	Decano	124-18-5	X	
F243	56	I.A, I.B, I.C, I.D, I.E, I.F, I.G, II.A, II.B, II.C, II.D, III.A, III.B, III.C, III.D	2-Etil-1-hexanol	104-76-7	X	
F244	57	III.D				
F360	118	IV.A	Ácido E-3-(2-hidroxifenil)prop-2-enoico	614-60-8		X
F419	99	III.A, III.B	2,2,4-Trimetil-1,3-pentanodiol diisobutirato (TXIB)	6846-50-0	X	
F421	57	IV.A				
F542	149	III.C, III.D	Ftalato de diisobutilo	84-69-5	X	



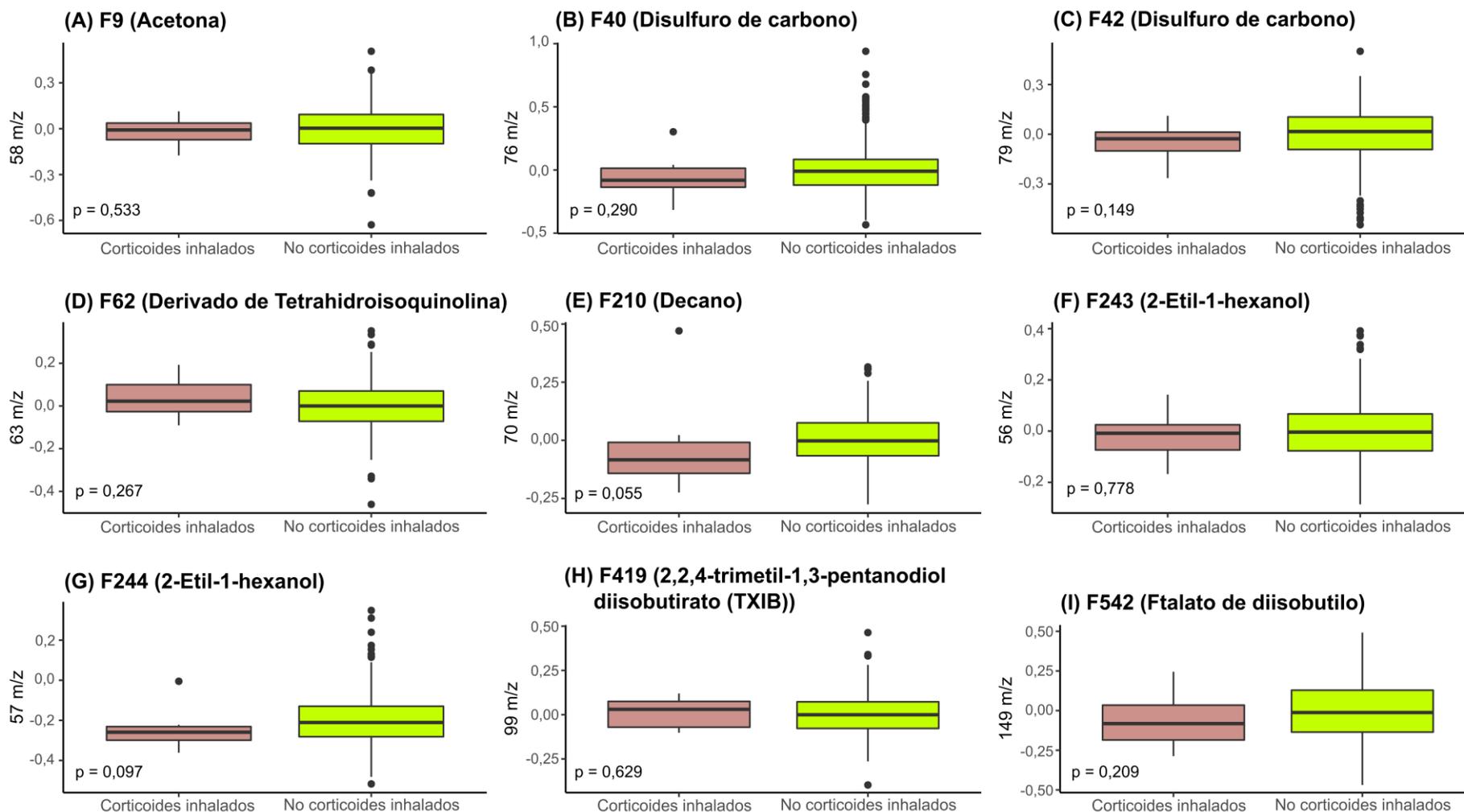
APÉNDICE 7. Evaluación de la influencia del consumo de tabaco en los niveles en aire exhalado de los *features* seleccionados como discriminantes en los modelos significativos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F, modelo I.G, modelo II.A, modelo II.B, modelo II.C, modelo II.D, modelo III.A, modelo III.B, modelo III.C y modelo III.D). (A) Impacto del consumo de tabaco en los niveles de acetona (F9) en el aire exhalado. (B) Impacto del consumo de tabaco en los niveles de disulfuro de carbono (F40) en el aire exhalado. (C) Impacto del consumo de tabaco en los niveles de disulfuro de carbono (F42) en el aire exhalado. (D) Impacto del consumo de tabaco en los niveles del derivado de tetrahydroisoquinolina (F62) en el aire exhalado. (E) Impacto del consumo de tabaco en los niveles de decano (F210) en el aire exhalado. (F) Impacto del consumo de tabaco en los niveles de 2-etil-1-hexanol (F243) en el aire exhalado. (G) Impacto del consumo de tabaco en los niveles de 2-etil-1-hexanol (F244) en el aire exhalado. (H) Impacto del consumo de tabaco en los niveles de 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) en el aire exhalado. (I) Impacto del consumo de tabaco en los niveles de Ftalato de diisobutilo (F542) en el aire exhalado.



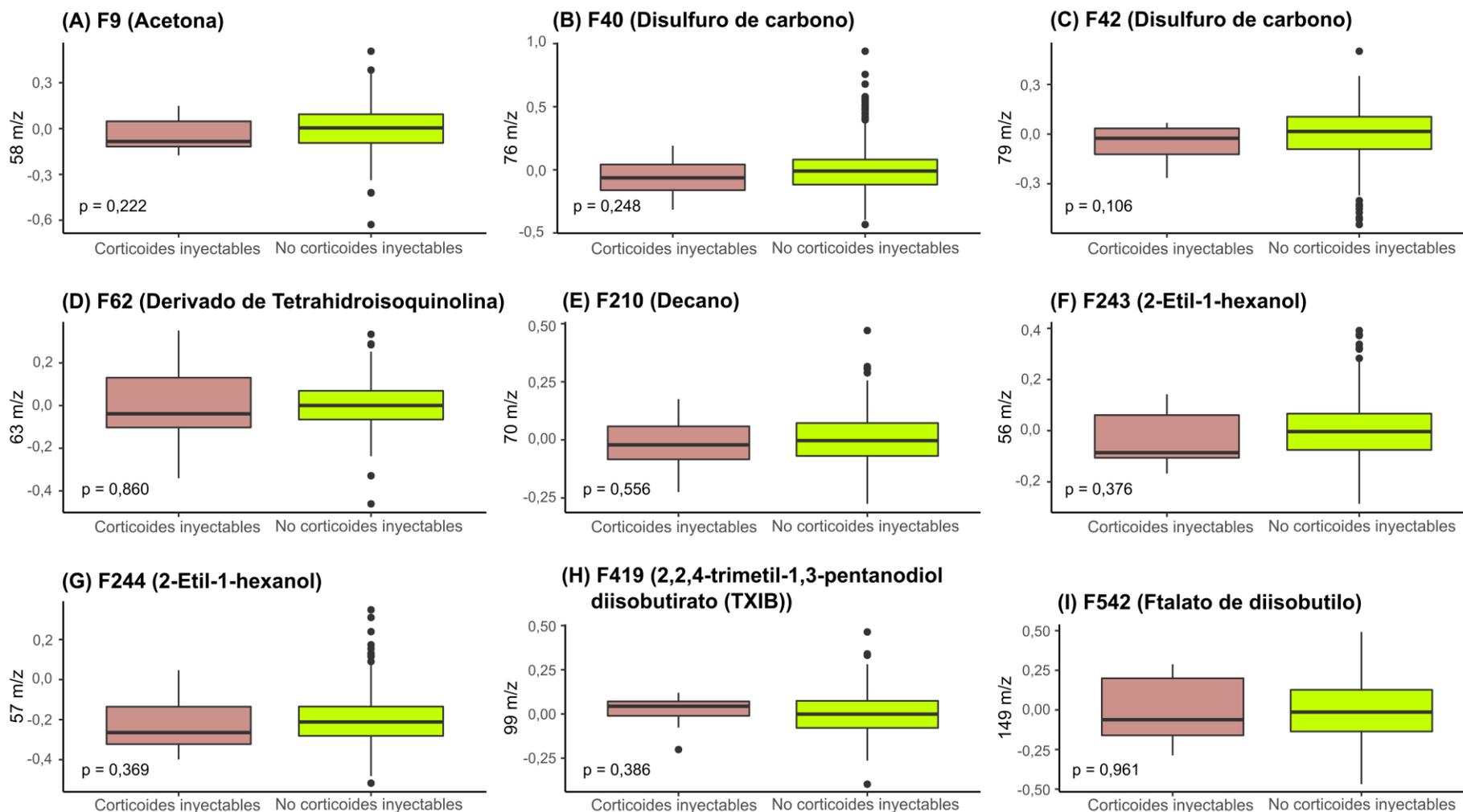
APÉNDICE 8. Evaluación de la influencia del consumo de antibióticos en los niveles en el aire exhalado de los *features* seleccionados como discriminantes en los modelos significativos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F, modelo I.G, modelo II.A, modelo II.B, modelo II.C, modelo II.D, modelo III.A, modelo III.B, modelo III.C y modelo III.D). (A) Impacto del consumo de antibióticos en los niveles de acetona (F9) en el aire exhalado. (B) Impacto del consumo de antibióticos en los niveles de disulfuro de carbono (F40) en el aire exhalado. (C) Impacto del consumo de antibióticos en los niveles de disulfuro de carbono (F42) en el aire exhalado. (D) Impacto del consumo de antibióticos en los niveles del derivado de tetrahydroisoquinolina (F62) en el aire exhalado. (E) Impacto del consumo de antibióticos en los niveles de decano (F210) en el aire exhalado. (F) Impacto del consumo de antibióticos en los niveles de 2-etil-1-hexanol (F243) en el aire exhalado. (G) Impacto del consumo de antibióticos en los niveles de 2-etil-1-hexanol (F244) en el aire exhalado. (H) Impacto del consumo de antibióticos en los niveles de 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) en el aire exhalado. (I) Impacto del consumo de antibióticos en los niveles de Ftalato de diisobutilo (F542) en el aire exhalado.



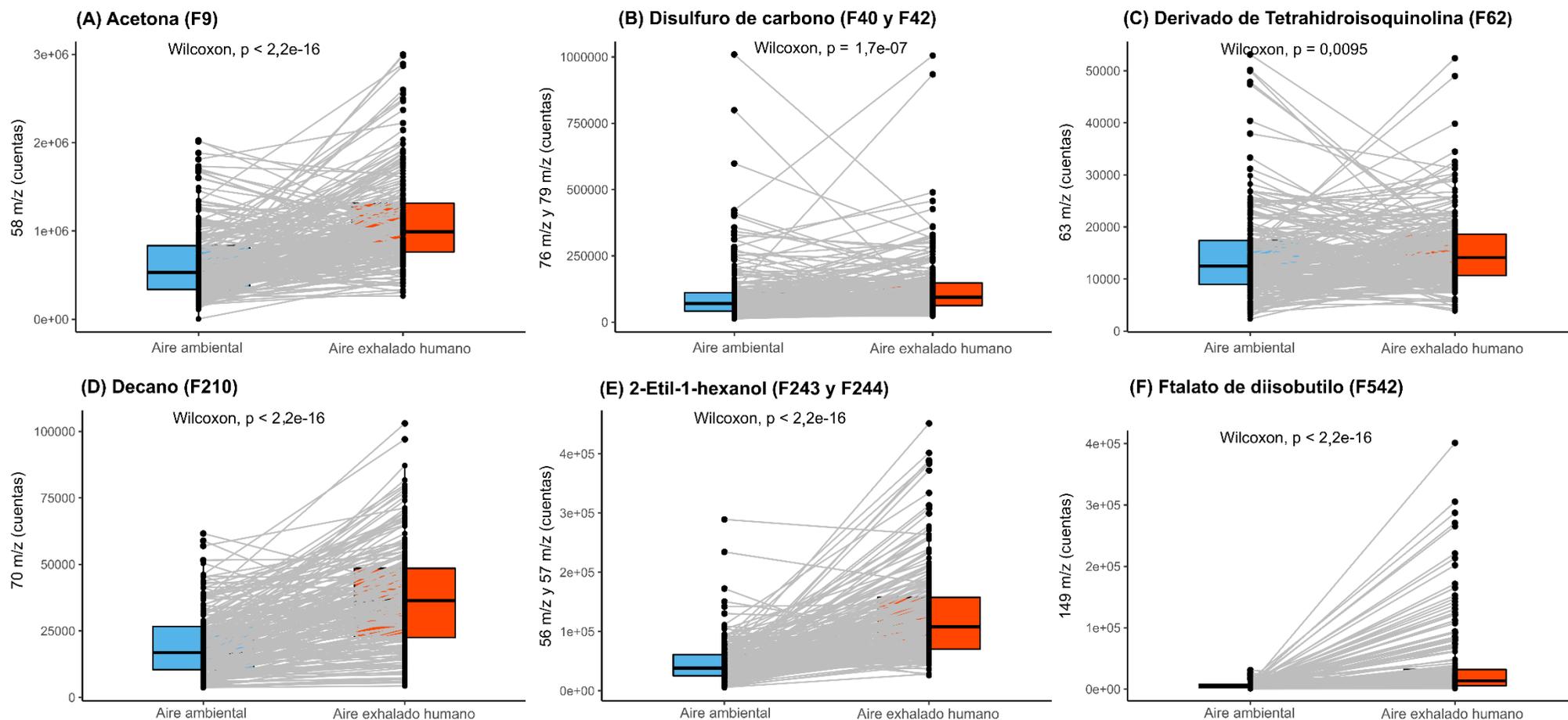
APÉNDICE 9. Evaluación de la influencia del consumo de paracetamol en los niveles en el aire exhalado de los *features* seleccionados como discriminantes en los modelos significativos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F, modelo I.G, modelo II.A, modelo II.B, modelo II.C, modelo II.D, modelo III.A, modelo III.B, modelo III.C y modelo III.D). (A) Impacto del consumo de paracetamol en los niveles de acetona (F9) en el aire exhalado. (B) Impacto del consumo de paracetamol en los niveles de disulfuro de carbono (F40) en el aire exhalado. (C) Impacto del consumo de paracetamol en los niveles de disulfuro de carbono (F42) en el aire exhalado. (D) Impacto del consumo de paracetamol en los niveles del derivado de tetrahydroisoquinolina (F62) en el aire exhalado. (E) Impacto del consumo de paracetamol en los niveles de decano (F210) en el aire exhalado. (F) Impacto del consumo de paracetamol en los niveles de 2-etil-1-hexanol (F243) en el aire exhalado. (G) Impacto del consumo de paracetamol en los niveles de 2-etil-1-hexanol (F244) en el aire exhalado. (H) Impacto del consumo de paracetamol en los niveles de 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) en el aire exhalado. (I) Impacto del consumo de paracetamol en los niveles de Ftalato de diisobutilo (F542) en el aire exhalado.



APÉNDICE 10. Evaluación de la influencia del consumo de corticoides inhalados en los niveles en el aire exhalado de los *features* seleccionados como discriminantes en los modelos significativos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F, modelo I.G, modelo II.A, modelo II.B, modelo II.C, modelo II.D, modelo III.A, modelo III.B, modelo III.C y modelo III.D). (A) Impacto del consumo de corticoides inhalados en los niveles de acetona (F9) en el aire exhalado. (B) Impacto del consumo de corticoides inhalados en los niveles de disulfuro de carbono (F40) en el aire exhalado. (C) Impacto del consumo de corticoides inhalados en los niveles de disulfuro de carbono (F42) en el aire exhalado. (D) Impacto del consumo de corticoides inhalados en los niveles del derivado de tetrahydroisoquinolina (F62) en el aire exhalado. (E) Impacto del consumo de corticoides inhalados en los niveles de decano (F210) en el aire exhalado. (F) Impacto del consumo de corticoides inhalados en los niveles de 2-etil-1-hexanol (F243) en el aire exhalado. (G) Impacto del consumo de corticoides inhalados en los niveles de 2-etil-1-hexanol (F244) en el aire exhalado. (H) Impacto del consumo de corticoides inhalados en los niveles de 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) en el aire exhalado. (I) Impacto del consumo de corticoides inhalados en los niveles de Ftalato de diisobutilo (F542) en el aire exhalado.



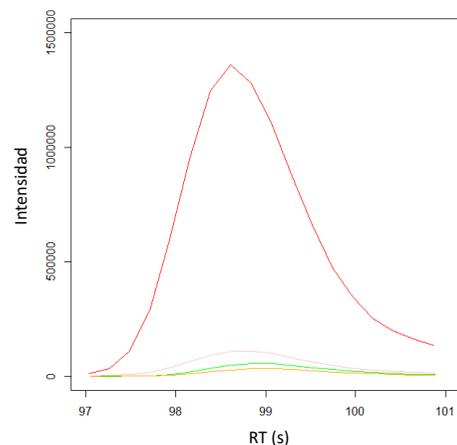
APÉNDICE 11. Evaluación de la influencia del consumo de corticoides inyectables en los niveles en el aire exhalado de los *features* seleccionados como discriminantes en los modelos significativos (modelo I.A, modelo I.B, modelo I.C, modelo I.D, modelo I.E, modelo I.F, modelo I.G, modelo II.A, modelo II.B, modelo II.C, modelo II.D, modelo III.A, modelo III.B, modelo III.C y modelo III.D). (A) Impacto del consumo de corticoides inyectables en los niveles de acetona (F9) en el aire exhalado. (B) Impacto del consumo de corticoides inyectables en los niveles de disulfuro de carbono (F40) en el aire exhalado. (C) Impacto del consumo de corticoides inyectables en los niveles de disulfuro de carbono (F42) en el aire exhalado. (D) Impacto del consumo de corticoides inyectables en los niveles del derivado de tetrahidroisoquinolina (F62) en el aire exhalado. (E) Impacto del consumo de corticoides inyectables en los niveles de decano (F210) en el aire exhalado. (F) Impacto del consumo de corticoides inyectables en los niveles de 2-etil-1-hexanol (F243) en el aire exhalado. (G) Impacto del consumo de corticoides inyectables en los niveles de 2-etil-1-hexanol (F244) en el aire exhalado. (H) Impacto del consumo de corticoides inyectables en los niveles de 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) en el aire exhalado. (I) Impacto del consumo de corticoides inyectables en los niveles de Ftalato de diisobutilo (F542) en el aire exhalado.



APÉNDICE 12. Comparación de las intensidades de los *features* de los VOCs discriminantes en las muestras ambientales y en las muestras de aire exhalado humano de la cohorte NELA. El 2,2,4-Trimetil-1,3-pentanediol diisobutirato (TXIB) (F419) no se detectó en las muestras de aire ambiental. (A) Intensidad de la acetona (F9) en el aire ambiente y en el aire humano exhalado. (B) Intensidad del disulfuro de carbono (F40-F42) en el aire ambiente y en el aire exhalado humano. (C) Intensidad del derivado de tetrahydroisoquinolina (F62) en el aire ambiente y en el aire exhalado humano. (D) Intensidad del decano (F210) en el aire ambiente y en el aire exhalado humano. (E) Intensidad del 2-etil-1-hexanol (F243) en el aire ambiente y en el aire exhalado humano. (F) Intensidad del Ftalato de diisobutilo (F542) en el aire ambiente y en el aire exhalado humano.

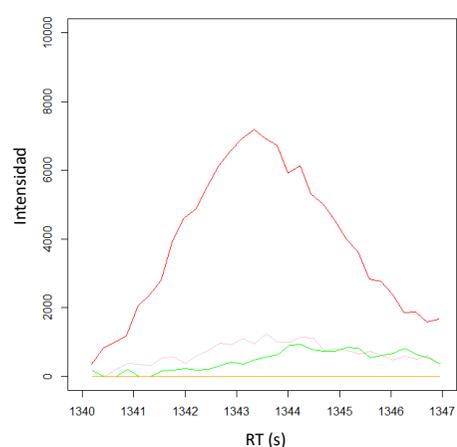
(A) Acetona (F9)

Cromatograma de Ion Extraído m/z 57,5 – 58,5



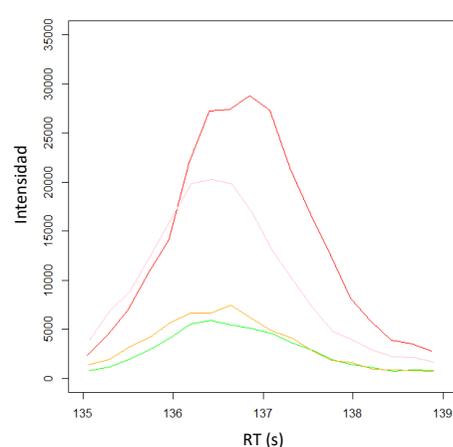
(B) 2-Etil-1-hexanol (F243)

Cromatograma de Ion Extraído m/z 55,5 – 56,5



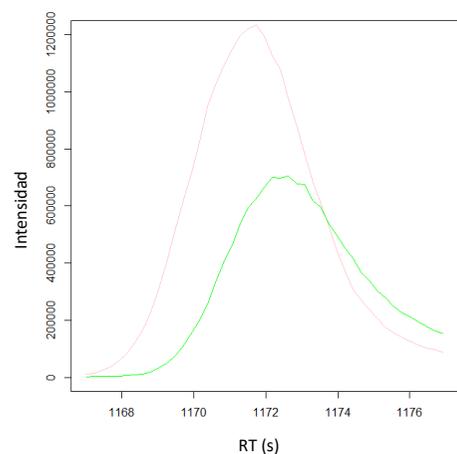
(C) Derivado de Tetrahydroisoquinolina (F62)

Cromatograma de Ion Extraído m/z 62,5 – 63,5



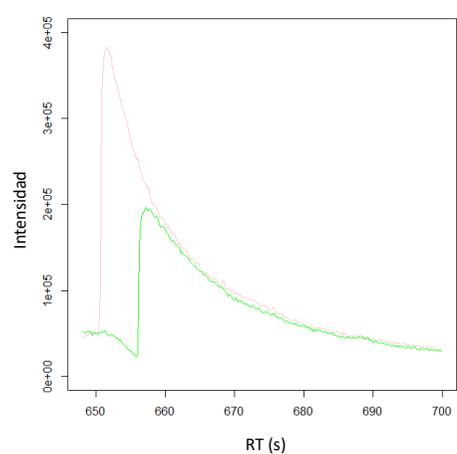
(D) Fenol

Cromatograma de Ion Extraído m/z 35,02 – 348,17



(E) N,N-dimetilacetamida

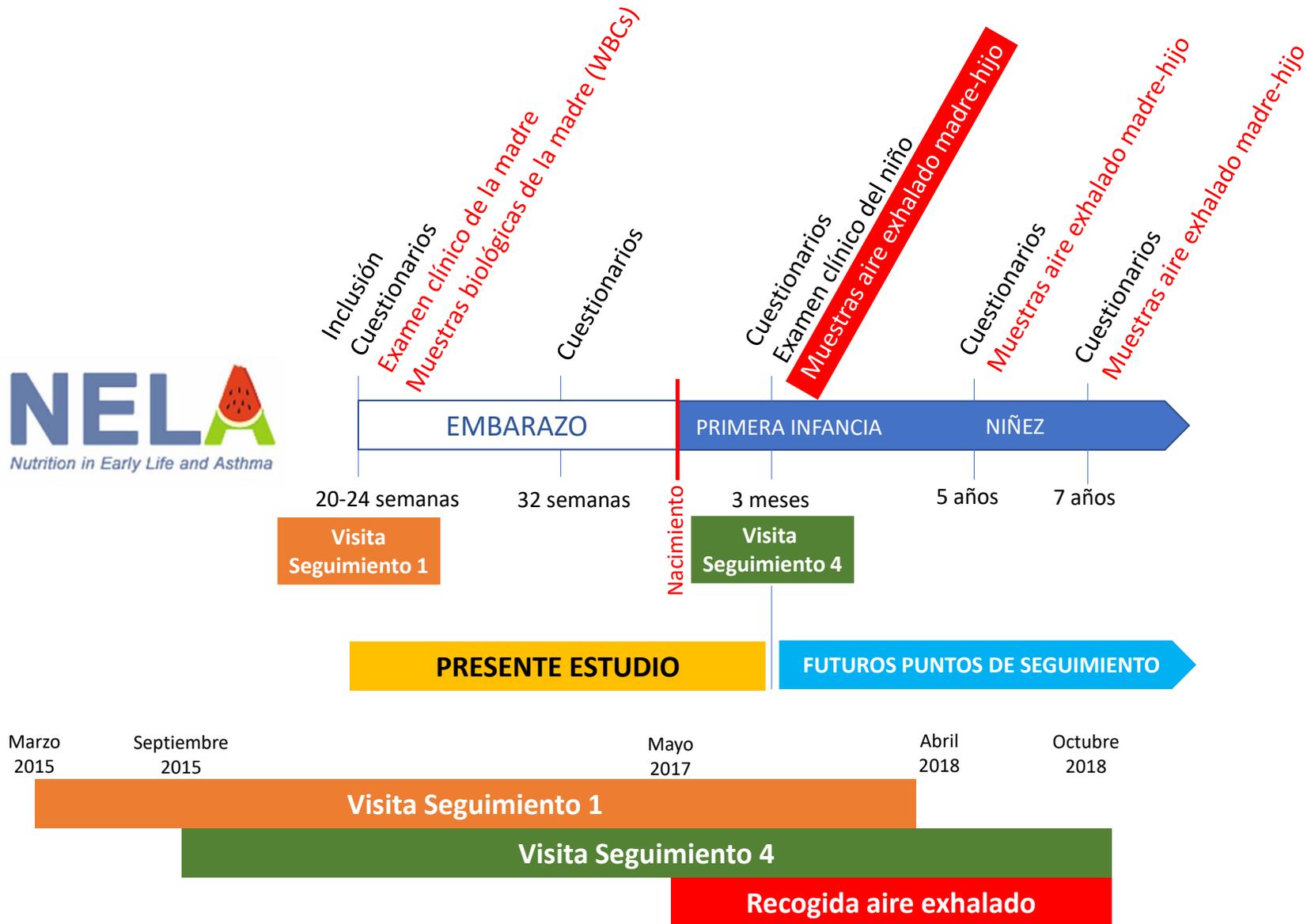
Cromatograma de Ion Extraído m/z 35,02 – 348,17



Leyenda

- Aire exhalado
- Bolsa Tedlar® (Sin lavados con gas nitrógeno)
- Bolsa Tedlar® (10 lavados con gas nitrógeno)
- Tubo de desorción térmica

APÉNDICE 13. Influencia de los contaminantes de las bolsas Tedlar® y de los productos de degradación de los tubos de desorción térmica (TD) en las intensidades de los *features* de los VOCs seleccionados como discriminantes del asma con otras enfermedades atópicas coexistentes (acetona, 2-etil-1-hexanol y un derivado de tetrahydroisoquinolina). (A) Intensidad del *feature* F9 (acetona) en el aire exhalado, en la bolsa Tedlar® y en un tubo TD reacondicionado. (B) Intensidad del *feature* F243 (2-etil-1-hexanol) en el aire exhalado, en la bolsa Tedlar® y en un tubo TD reacondicionado. (C) Intensidad del *feature* F62 (tetrahydroisoquinolina) en el aire exhalado, en la bolsa Tedlar® y en un tubo TD reacondicionado. (D) Intensidad del fenol en la bolsa Tedlar® antes y después del lavado con nitrógeno. (E) Intensidad de la N,N-dimetilacetamida en la bolsa Tedlar® antes y después de los lavados con nitrógeno.



APÉNDICE 14. Diagrama de las visitas de seguimiento de los pacientes reclutados en la cohorte NELA (*Nutrition in Early Life and Asthma*). El presente estudio está enmarcado en el estudio NELA.

Capítulo 6

Relación entre la función pulmonar y los compuestos orgánicos volátiles exhalados en lactantes sanos

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, Manuel Sánchez-Solís, Gema Lozano-Terol, Julia Gallego-Jara, Luis García-Marcos, Manuel Cánovas Díaz y Teresa de Diego
Puente

**Relationship between lung function and exhaled volatile organic
compounds in healthy infants**

Pediatric Pulmonology **57**, 1282-1292 (2022) doi: 10.1002/ppul.25849.

RESUMEN

La exploración de la función pulmonar es clave en el diagnóstico de enfermedades respiratorias crónicas. Por otro lado, el análisis de compuestos orgánicos volátiles (VOCs) en el aire exhalado es una técnica no invasiva que permite la monitorización del estrés oxidativo característico de algunos tipos de inflamación de las vías respiratorias. En el presente trabajo, se evalúa, por primera vez, la relación del volatilo exhalado y la función pulmonar en lactantes sanos, lo que puede ser de gran ayuda para la detección precoz de ciertas enfermedades respiratorias. Por consiguiente, se examinó la función pulmonar de 50 lactantes sanos de 3 a 8 meses de edad y se determinaron los siguientes parámetros: capacidad vital forzada (FVC), volumen espiratorio forzado a 0,5 s ($FEV_{0,5}$), flujo espiratorio forzado al 75% de la FVC (FEF_{75}), flujo espiratorio forzado al 25%-75% de la FVC (FEF_{25-75}) y el cociente $FEV_{0,5}/FVC$. La función pulmonar fue medida según la técnica de compresión toracoabdominal rápida con volumen pulmonar aumentado. Además, se llevó a cabo un análisis dirigido centrado en el estudio de seis VOCs (acetona, isopreno, decano, undecano, tetradecano y pentadecano) en el aire exhalado de los bebés, mediante un sistema de desorción térmica acoplado a cromatografía de gases -espectrometría de masas de cuadrupolo simple (TD-GC/q-MS). A este respecto, se ha observado una relación negativamente significativa entre los niveles de acetona, tetradecano y pentadecano en el aire exhalado, y varios de los parámetros de función pulmonar. Los niveles de acetona ($m/z = 58$) se asociaron significativamente de forma negativa con la FVC y el $FEV_{0,5}$, los niveles de tetradecano ($m/z = 71$) con el $FEV_{0,5}$, y los niveles de pentadecano ($m/z = 71$) con el $FEV_{0,5}$ y el FEF_{25-75} . Por lo tanto, los resultados de este estudio ponen de manifiesto que existe una asociación significativa entre VOCs relacionados con el estrés oxidativo y la función pulmonar en lactantes sanos.

INTRODUCCIÓN

La incidencia de enfermedades respiratorias crónicas como el asma, la enfermedad pulmonar obstructiva crónica o la fibrosis pulmonar ha aumentado de forma alarmante en los últimos 20 años, lo que supone un elevado desembolso económico directo e indirecto para la sociedad y los servicios sanitarios de los países desarrollados¹. Actualmente, las pruebas de función pulmonar y la historia clínica desempeñan un papel esencial en el diagnóstico de las enfermedades respiratorias². Los parámetros medidos en las pruebas de función pulmonar en adultos y niños de mayor edad, como el volumen espiratorio forzado a 1 s (FEV_1) y el cociente FEV_1 /capacidad vital forzada (FVC), son difíciles de obtener en lactantes, por lo que en

el caso de bebés son preferibles el $FEV_{0,5}$ y el cociente $FEV_{0,5}/FVC$, ya que el tiempo necesario para el vaciado pulmonar completo es menor a una edad más temprana³. Asimismo, la exploración de la función pulmonar suele también incluir la medición de otros parámetros como la FVC, el flujo espiratorio forzado al 75% de la FVC (FEF_{75}), y el flujo espiratorio forzado al 25%-75% de la FVC (FEF_{25-75})⁴. Sin embargo, los parámetros de la función pulmonar no permiten obtener información sobre la gravedad o el tipo de inflamación que podría estar causando una función pulmonar baja, más allá del grado de obstrucción que pueda provocar la inflamación. Por lo tanto, sería muy interesante disponer de información complementaria sobre la evolución de la inflamación y sobre los parámetros de función pulmonar⁵⁻⁷.

En este sentido, aunque existen varias técnicas para monitorizar el estrés oxidativo y la inflamación de las vías respiratorias (ej. broncoscopia con lavado broncoalveolar, biopsias, etc.), estas pruebas son difíciles de aplicar dado su alto grado de invasividad. De este modo, en la última década, ha aumentado el interés por el desarrollo de métodos no invasivos, que sean poco o nada lesivos para los pacientes. En este sentido, el análisis de compuestos orgánicos volátiles (VOCs) en el aire exhalado ha suscitado una gran expectación, ya que se trata de un procedimiento no invasivo y que resulta más amigable para el paciente que otras estrategias⁷. En el aire exhalado humano, pueden determinarse VOCs procedentes de fuentes exógenas, como los contaminantes ambientales, y VOCs endógenos, como los producidos por el metabolismo celular humano o por la microbiota. Además, las concentraciones relativas de los VOCs pueden cambiar en respuesta a procesos patológicos. Precisamente, la acetona y el isopreno son los dos VOCs endógenos más abundantes en el aire exhalado humano^{8,9}. En los últimos años, esta técnica ha sido considerada potencialmente valiosa para el diagnóstico y el seguimiento de las enfermedades respiratorias, así como una herramienta útil para la detección de cambios metabólicos que implican estrés oxidativo⁹⁻¹¹. Por ejemplo, los niveles elevados de alcanos, como el pentano y el etano en el aire exhalado, han sido asociados previamente con la presencia de asma, ya que estos compuestos pueden ser generados durante la peroxidación lipídica debido a un desequilibrio oxidativo^{12,13}. Por lo tanto, es de gran importancia determinar la relación entre el perfil de VOCs en el aire exhalado humano y los parámetros de función pulmonar para evaluar la conexión entre el estrés oxidativo y los trastornos pulmonares en la primera infancia. Esta estrategia podría servir para diagnosticar el asma preclínica en los niños, ya que los cambios inflamatorios subyacentes podrían estar presentes en el pulmón mucho antes que las alteraciones de la función pulmonar. En los últimos años, varios estudios han encontrado una asociación entre la función pulmonar y la

inflamación, utilizando técnicas menos específicas que el análisis de VOCs en el aire exhalado, como la determinación de la fracción exhalada de óxido nítrico (FeNO) y el recuento de eosinófilos en sangre¹⁴. Fuglsbjerg *et al.*¹⁵ analizaron los VOCs en el aire exhalado y midieron el FEV₁ de niños asmáticos mayores de 2 años, pero, no obtuvieron resultados concluyentes, y tampoco analizaron la función pulmonar en niños sanos para excluir el asma latente.

En el presente trabajo, se ha evaluado por primera vez la relación existente entre los parámetros de las pruebas de función pulmonar y el volatilo exhalado en lactantes sanos. Se compararon la FVC, el FEV_{0,5}, el FEF₇₅, el FEF₂₅₋₇₅ y el FEV_{0,5}/FVC con los niveles en el aire exhalado de seis VOCs (acetona, isopreno, decano, undecano, tetradecano y pentadecano) en lactantes sanos. Estos compuestos se seleccionaron en función de su gran abundancia en el aire exhalado humano o de la evidencia previa en la literatura (Capítulo 3) de una asociación con la presencia de enfermedades respiratorias/atópicas en la población pediátrica¹³.

RESULTADOS

Relación entre la función pulmonar o el volatilo exhalado y las características demográficas de los sujetos

La población de estudio de este trabajo consistió en 50 lactantes sanos (nacidos a término) de la cohorte prospectiva de nacimiento NELA (*Nutrition in Early Life and Asthma*)¹⁶. Las características demográficas de los lactantes sanos incluidos en este estudio se muestran en la Tabla 1 y las de sus padres en el Apéndice 1.

Tabla 1. Características demográficas de los lactantes incluidos en este estudio.

	Niños (n = 23)	Niñas (n = 37)	p-valor
Peso (Kg), media (DE)	7,25 (1,01)	6,93 (0,77)	0,216
Talla (cm), media (DE)	66,8 (3,5)	65,3 (2,6)	0,090
Edad (semanas), media (DE)	22,02 (3,81)	22,05 (4,87)	0,977
Edad gestacional al nacimiento (semanas), media (DE)	39,43 (1,47)	40,15 (1,26)	0,084
FVC (mL), mediana (IQR)	230 (51)	215 (35)	0,137
FEV _{0,5} (mL), mediana (IQR)	178 (38)	175 (23)	0,984
FEV _{0,5} /FVC (mL), mediana (IQR)	0,79 (0,15)	0,82 (0,08)	0,038
FEF ₇₅ (mL), mediana (IQR)	185 (60)	193 (51)	0,104
FEF ₂₅₋₇₅ (mL), mediana (IQR)	314 (91)	316 (67)	0,298

En negrita se muestran los resultados estadísticamente significativos (p-valor < 0,05). IQR: Rango intercuartílico; FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

El cociente $FEV_{0,5}/FVC$ fue el único parámetro de la función pulmonar en el que influyó significativamente el sexo del bebé. Así pues, las niñas mostraron valores más altos del cociente $FEV_{0,5}/FVC$ que los niños (Tabla 1). El peso, la estatura y la edad de los lactantes estaban correlacionados significativamente con muchos de los parámetros de función pulmonar (Tabla 2 y Figura 1). De hecho, los valores de FVC y $FEV_{0,5}$ fueron significativamente mayores en los niños con mayor peso, altura y edad. Por el contrario, el peso de los niños se asoció negativamente con los valores del cociente $FEV_{0,5}/FVC$.

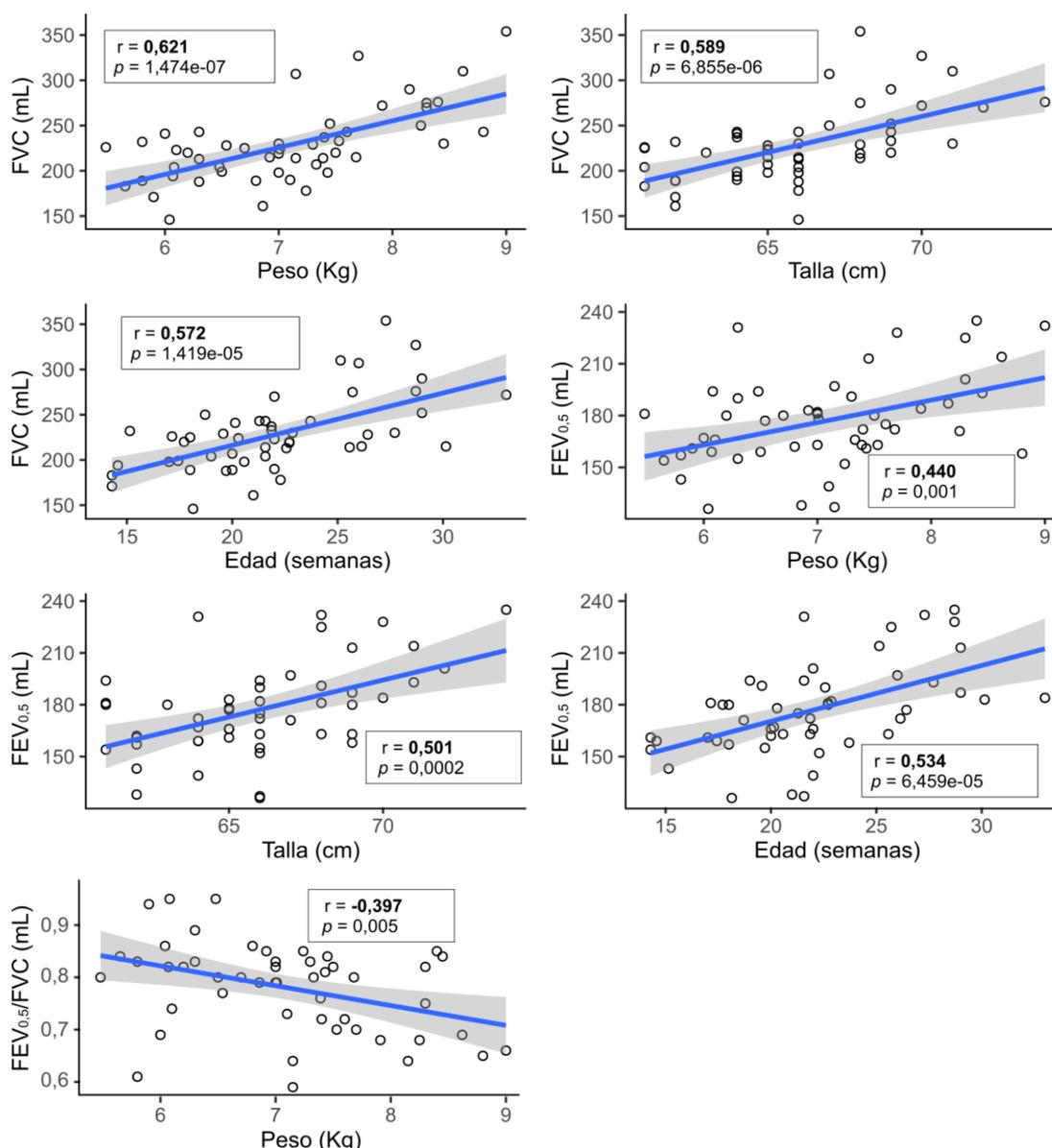


Figura 1. Correlaciones significativas entre los parámetros de función pulmonar y las características demográficas de los lactantes. FVC: capacidad vital forzada; $FEV_{0,5}$: volumen espiratorio forzado a 0,5 s; FEF_{75} : flujo espiratorio forzado al 75% de la FVC; FEF_{25-75} : flujo espiratorio forzado al 25%-75% de la FVC.

Por otra parte, no se encontró ninguna relación significativa entre ninguno de los *features* o fragmentos iónicos (señales masa/carga o m/z) de los VOCs seleccionados,

y el peso, la talla, la edad o el sexo de los lactantes (Tabla 2 y Apéndice 2). Asimismo, la estación del año en que se tomaron las muestras de aliento no tuvo una influencia significativa en la función pulmonar de los lactantes (Apéndice 3).

Tabla 2. Coeficientes de correlación entre las características de los lactantes de este estudio (peso, talla y edad) y los parámetros de función pulmonar o los *features* de los compuestos orgánicos volátiles del aire exhalado de los niños.

	Peso	Talla	Edad
FVC	0,621	0,589	0,572
FEV_{0,5}	0,440	0,501	0,534
FEV_{0,5}/FVC	-0,397	-0,256	-0,255
FEF ₇₅	0,096	0,133	0,245
FEF ₂₅₋₇₅	0,011	0,109	0,192
Acetona (m/z = 58)	-0,079	0,024	-0,118
Isopreno (m/z = 67)	0,127	-0,087	0,037
Decano (m/z = 71)	-0,034	0,007	0,056
Undecano (m/z = 57)	-0,097	-0,122	-0,021
Tetradecano (m/z = 71)	-0,240	-0,184	-0,193
Pentadecano (m/z = 71)	-0,117	-0,074	-0,015

En negrita se muestran los resultados estadísticamente significativos (p-valor < 0,05). FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

La Figura 2 muestra los z-scores de todos los parámetros de función pulmonar medidos en los lactantes incluidos en el presente estudio.

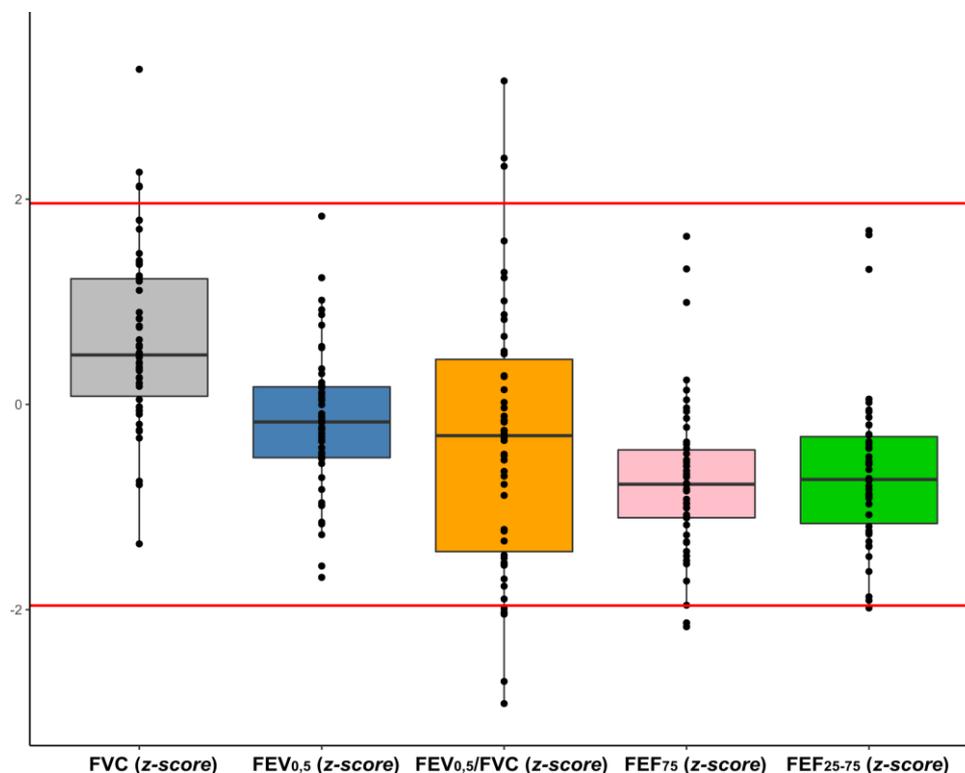


Figura 2. Z-scores de los parámetros de función pulmonar determinados en los lactantes. Las líneas rojas indican el límite superior (ULN) y el límite inferior (LLN) del rango normal ($\pm 1,96$ z-scores). FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

La media (DE) de los *z-scores* para todos los parámetros de función pulmonar fue cercana a 0 (1) (media (DE) de la FVC (*z-score*): -0,02 (0,89); FEV_{0,5} (*z-score*): -0,21 (0,76); FEV_{0,5}/FVC (*z-score*): 0,66 (1,31); FEF₇₅ (*z-score*): -0,49 (0,80); FEF₂₅₋₇₅ (*z-score*): -0,46 (0,80)). Algunos *z-scores* de FEV_{0,5}/FVC estaban fuera del rango normal ($\pm 1,96$ *z-scores*). De este modo, tres lactantes mostraron *z-scores* de FEV_{0,5}/FVC por encima del límite superior de la normalidad ($\geq 1,96$ *z-scores*), y cinco lactantes por debajo del límite inferior de la normalidad ($\leq -1,96$ *z-scores*). Sin embargo, más del 95% de los *z-scores* de FEV_{0,5}, FEF₇₅ y FEF₂₅₋₇₅ que se determinaron en los lactantes estaban dentro del rango normal ($\pm 1,96$ *z-scores*), y menos del 2,5% de los *z-scores* de FVC, FEV_{0,5}, FEF₇₅ y FEF₂₅₋₇₅ estaban por debajo del límite inferior de la normalidad ($\leq -1,96$ *z-scores*).

Relación entre la función pulmonar y el volatilo exhalado en lactantes sanos

Tres VOCs (acetona, tetradecano y pentadecano) estaban correlacionados significativamente con varios parámetros de la función pulmonar (Figura 3). La Tabla 3 muestra los coeficientes de correlación de Pearson entre todos los *features* seleccionados medidos en el aire exhalado y los parámetros de función pulmonar (*z-score*). El fragmento iónico $m/z = 58$ de la acetona, se correlacionó negativamente de forma significativa con la FVC y el FEV_{0,5}. Además, los niveles en el aire exhalado de un par de alcanos lineales, el tetradecano ($m/z = 71$) y el pentadecano ($m/z = 71$) mostraron una correlación negativa significativa con el FEV_{0,5}. Adicionalmente, el pentadecano ($m/z = 71$) también se correlacionó negativamente con el FEF₂₅₋₇₅ (*z-score*). Asimismo, los análisis de regresión lineal múltiple también mostraron que valores más bajos de la FVC estaban asociados significativamente con niveles más altos de acetona ($m/z = 58$) en el aire exhalado; y que valores más bajos de FEV_{0,5} estaban asociados significativamente de forma negativa con niveles más altos en el aire exhalado de acetona ($m/z = 58$), tetradecano ($m/z = 71$) y pentadecano ($m/z = 71$). Igualmente, también se observó que los valores de FEF₂₅₋₇₅ y los niveles de pentadecano ($m/z = 71$) en el aire exhalado estaban asociados significativamente de forma negativa. En la Tabla 4 se muestran los análisis de regresión lineal múltiple, realizados sobre los valores de los parámetros de función pulmonar sin ser convertidos en *z-scores*. Estos modelos estaban ajustados por el sexo, la talla, el peso y la edad del lactante en el momento de las pruebas de función pulmonar, y por el tabaquismo materno durante el embarazo. Asimismo, en el Apéndice 4 se muestran los modelos realizados sobre los *z-scores* de los parámetros de función pulmonar.

Tabla 3. Coeficientes de correlación entre los z-scores de los parámetros de función pulmonar y los *features* de compuestos orgánicos volátiles (VOCs) seleccionados del aire exhalado de los lactantes

	FVC (z-score)	FEV_{0,5} (z-score)	FEV_{0,5}/FVC (z-score)	FEF₇₅ (z-score)	FEF₂₅₋₇₅ (z-score)
Acetona (m/z = 58)	-0,377	-0,399	-0,011	-0,122	-0,157
Isopreno (m/z = 67)	-0,010	-0,081	-0,076	-0,086	-0,060
Decano (m/z = 71)	-0,135	-0,199	-0,055	-0,119	-0,092
Undecano (m/z = 57)	0,014	-0,127	-0,188	-0,135	-0,158
Tetradecano (m/z = 71)	-0,189	-0,300	-0,103	-0,202	-0,222
Pentadecano (m/z = 71)	-0,124	-0,311	-0,218	-0,190	-0,310

En negrita se muestran los resultados estadísticamente significativos (p -valor $< 0,05$). FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

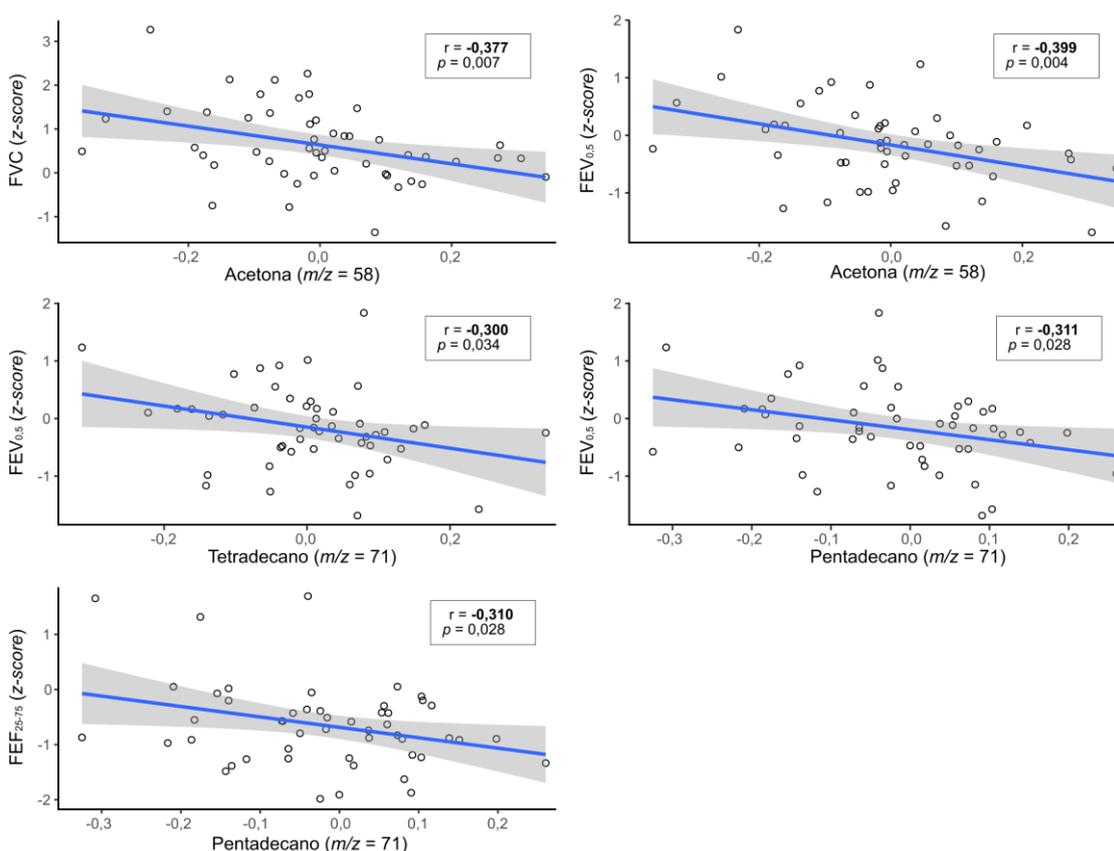


Figura 3. Correlaciones significativas entre los parámetros de función pulmonar (z-score) y los *features* de los compuestos orgánicos volátiles determinados en el aire exhalado de los lactantes. FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

Tabla 4. Análisis de regresión lineal múltiple para determinar la relación entre los parámetros de función pulmonar y los *features* de los compuestos orgánicos volátiles (VOCs) del aire exhalado de los lactantes. Los modelos fueron ajustados por el sexo, la talla, el peso y la edad del lactante en el momento de las pruebas de función pulmonar, así como por el tabaquismo materno durante el embarazo.

		β^1	IC 95% ²	p-valor
FVC	Acetona (<i>m/z</i> = 58)	-69,415	-125,815; -13,015	0,017
	Isopreno (<i>m/z</i> = 67)	-38,022	-106,165; 30,122	0,267
	Decano (<i>m/z</i> = 71)	-27,967	-78,426; 22,492	0,270
	Undecano (<i>m/z</i> = 57)	-10,275	-80,182; 59,631	0,768
	Tetradecano (<i>m/z</i> = 71)	-68,959	-148,135; 10,217	0,086
	Pentadecano (<i>m/z</i> = 71)	-39,909	-111,545; 31,727	0,268
FEV_{0,5}	Acetona (<i>m/z</i> = 58)	-57,285	-98,934; -15,636	0,008
	Isopreno (<i>m/z</i> = 67)	-18,096	-69,643; 33,452	0,483
	Decano (<i>m/z</i> = 71)	-27,438	-64,883; 10,007	0,147
	Undecano (<i>m/z</i> = 57)	-31,040	-82,640; 20,560	0,232
	Tetradecano (<i>m/z</i> = 71)	-74,022	-131,117; -16,927	0,012
	Pentadecano (<i>m/z</i> = 71)	-69,026	-119,223; -18,829	0,008
FEV_{0,5}/FVC	Acetona (<i>m/z</i> = 58)	-0,004	-0,163; 0,156	0,964
	Isopreno (<i>m/z</i> = 67)	0,028	-0,150; 0,205	0,755
	Decano (<i>m/z</i> = 71)	-0,001	-0,134; 0,132	0,991
	Undecano (<i>m/z</i> = 57)	-0,086	-0,268; 0,096	0,345
	Tetradecano (<i>m/z</i> = 71)	-0,114	-0,323; 0,095	0,278
	Pentadecano (<i>m/z</i> = 71)	-0,156	-0,337; 0,025	0,089
FEF₇₅	Acetona (<i>m/z</i> = 58)	-51,187	-157,287; 54,913	0,336
	Isopreno (<i>m/z</i> = 67)	-17,929	-140,781; 104,922	0,770
	Decano (<i>m/z</i> = 71)	-46,586	-136,495; 43,323	0,302
	Undecano (<i>m/z</i> = 57)	-78,426	-200,533; 43,681	0,202
	Tetradecano (<i>m/z</i> = 71)	-123,164	-263,948; 17,621	0,085
	Pentadecano (<i>m/z</i> = 71)	-112,806	-237,336; 11,723	0,075
FEF₂₅₋₇₅	Acetona (<i>m/z</i> = 58)	-87,814	-234,817; 59,188	0,235
	Isopreno (<i>m/z</i> = 67)	-21,096	-192,331; 150,139	0,805
	Decano (<i>m/z</i> = 71)	-54,734	-180,476; 71,007	0,385
	Undecano (<i>m/z</i> = 57)	-119,516	-289,012; 49,981	0,162
	Tetradecano (<i>m/z</i> = 71)	-192,948	-387,239; 1,342	0,052
	Pentadecano (<i>m/z</i> = 71)	-227,382	-393,386; -61,378	0,008

En negrita se muestran los resultados estadísticamente significativos (p-valor < 0,05). FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

¹Coefficiente de regresión

²Intervalo de confianza del 95%

Análisis de componentes principales

Con el objetivo de estudiar la función pulmonar de los lactantes de manera global, se llevó a cabo un análisis de componentes principales (PCA) sobre los parámetros de función pulmonar. Los dos primeros componentes obtenidos en el PCA explicaban más del 80% de la variabilidad de los datos (el componente principal 1 explicaba el 58,5% y el componente principal 2 el 37,3%). En este sentido, casi toda la información proporcionada por los cinco parámetros de función pulmonar puede representarse en un único gráfico (Figura 4A). Los parámetros que más contribuyeron

al Componente Principal 1 fueron el FEV_{0,5}, el FEF₇₅ y el FEF₂₅₋₇₅ (Figura 4B). Por otra parte, como se muestra en la Figura 4C y en la Tabla 5, el Componente Principal 1 estaba significativamente correlacionado de forma negativa con el tetradecano ($m/z = 71$) y el pentadecano ($m/z = 71$).

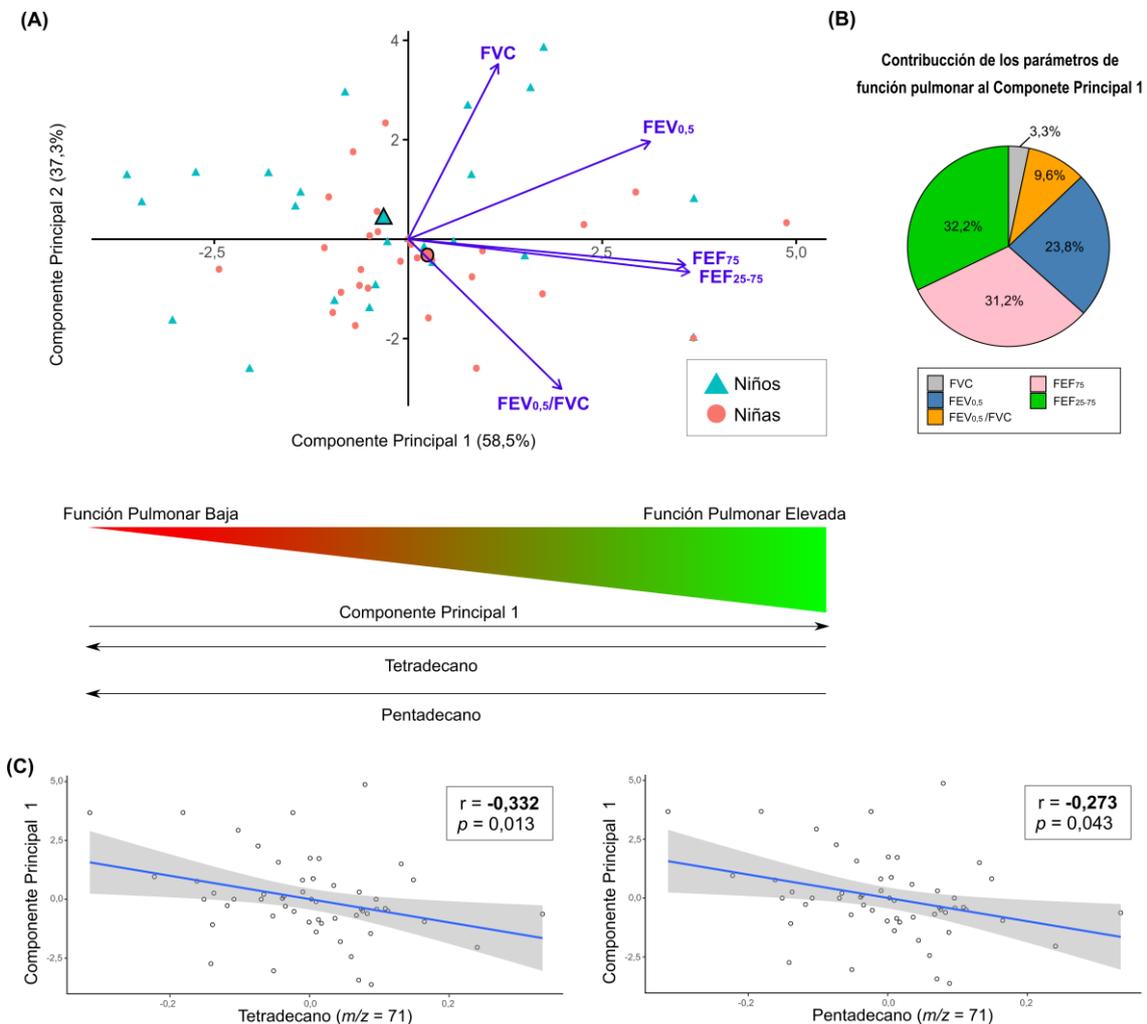


Figura 4. Análisis de componentes principales (PCA) realizado sobre los parámetros de función pulmonar en los lactantes. (A) Representación de los dos primeros componentes. Los vectores de color púrpura corresponden a las variables implicadas en el PCA. (B) Contribución de las variables al primer componente. (C) Correlaciones significativas negativas entre el primer componente y el tetradecano ($m/z = 71$) y el pentadecano ($m/z = 71$) determinados en el aire exhalado de los lactantes. FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

Tabla 5. Relación entre el componente principal 1 obtenido en el análisis de componentes principales realizado sobre los parámetros de función pulmonar en lactantes, y los *features* de los compuestos orgánicos volátiles detectados en el aire exhalado de los niños.

	Componente Principal 1	
	r	p-valor
Acetona (m/z = 58)	-0,261	0,054
Isopreno (m/z = 67)	-0,144	0,294
Decano (m/z = 71)	-0,199	0,145
Undecano (m/z = 57)	-0,244	0,072
Tetradecano (m/z = 71)	-0,332	0,013
Pentadecano (m/z = 71)	-0,273	0,043

En negrita se muestran los resultados estadísticamente significativos (p-valor < 0,05).

DISCUSIÓN

En este estudio se ha demostrado que existe una asociación entre el volatiloma exhalado y la función pulmonar en lactantes sanos. Algunos parámetros de función pulmonar (FVC, FEV_{0,5} y FEF₂₅₋₇₅) estaban correlacionados negativamente con los niveles detectados en el aire exhalado de acetona, tetradecano y pentadecano (Tabla 3 y Figura 3). Aunque las correlaciones encontradas entre los parámetros de la función pulmonar y los niveles de VOCs fueron débiles (r alrededor de -0,3), los hallazgos fueron significativos porque la probabilidad de que existiera una relación por azar era extremadamente baja (p-valor < 0,05). Además, se han obtenido también los mismos resultados mediante otras estrategias como los análisis de regresión lineal múltiple (Tabla 4 y Apéndice 4). Por consiguiente, los resultados de los análisis de regresión lineal múltiple, y de la evaluación de la relación entre el volatiloma exhalado y las características demográficas de los lactantes confirman que la asociación entre estos tres compuestos y la función pulmonar en los lactantes es independiente de factores como el peso, la altura, la edad, el sexo y el tabaquismo materno durante el embarazo (Tablas 2 y 4, Apéndice 4 y Figura 1).

La acetona, que es generada a partir del acetoacetato, es uno de los compuestos endógenos que se encuentra en mayores concentraciones en el aire humano exhalado⁸. Los alcanos lineales pueden provenir de fuentes endógenas, así como de fuentes exógenas¹⁷ (Apéndice 5). Por lo tanto, para una determinación precisa de los VOCs endógenos en el aire exhalado, es esencial analizar el aire ambiental, tal y como se ha llevado a cabo en este estudio (Apéndice 6). Previamente, se ha sugerido que tanto la acetona^{18,19} como varios alcanos lineales^{8,20,21} podrían ser biomarcadores de inflamación ya que son subproductos de la peroxidación lipídica (Apéndice 5). De hecho, la acetona²² y varios alcanos lineales¹³ han sido señalados como posibles biomarcadores en el aire exhalado para enfermedades respiratorias como el asma. Los

niveles de los fragmentos iónicos de la acetona, el tetradecano y el pentadecano fueron mayores en los niños que presentaban una función pulmonar más baja, lo que sugiere que una función pulmonar deficiente podría estar relacionada con procesos inflamatorios. Paredi *et al.*²³ observaron que los niveles de etano en el aire exhalado de adultos se correlacionaban negativamente con los valores de FEV₁.

A los lactantes incluidos en el presente estudio procedentes de la cohorte del NELA se les hará un seguimiento durante su infancia (visita de seguimiento a los 18 meses, 5 años y 7 años de edad)^{16,24}. En este sentido, se realizará una evaluación prospectiva para comprobar la asociación de los resultados de los parámetros de función pulmonar y los VOCs exhalados, con el futuro desarrollo de enfermedades respiratorias/atópicas como el asma. De este modo, será posible determinar si las pruebas de función pulmonar junto con la monitorización del estrés oxidativo, mediante el análisis del aire exhalado en lactantes sanos durante los primeros meses de vida, es una estrategia útil para la detección temprana de enfermedades respiratorias.

La principal limitación de este estudio fue el reducido número de sujetos ($n = 50$). De esta manera, algunas relaciones cercanas a la significancia estadística (p -valor $< 0,10$) entre los parámetros de función pulmonar y algunos VOCs, como el tetradecano o el pentadecano (Tabla 4), podrían ser significativas si se incluyeran más sujetos. En consecuencia, se realizó un PCA para comparar los niveles de los VOCs seleccionados y la función pulmonar de forma global, y para comprobar si las relaciones cercanas a la significancia estadística tenían un mayor impacto cuando se consideran todos los parámetros de función pulmonar en su conjunto. El PCA permite reducir un gran número de variables a un número mucho menor de componentes, lo que nos permite identificar componentes representativos de los datos y estudiar la función pulmonar global de los lactantes. Se ha observado una correlación significativa negativa entre el tetradecano ($m/z = 71$) y el pentadecano ($m/z = 71$) en el aire exhalado, y el primer componente principal obtenido en el PCA realizado sobre los parámetros de función pulmonar (Figura 4C). El componente principal 1 (58,5% de la variabilidad) explica el nivel de función pulmonar de los individuos, puesto que valores más altos de los cinco parámetros de función pulmonar dan lugar a un valor elevado del componente principal 1. Por consiguiente, los lactantes con valores más altos en el Componente Principal 1 (situados a la izquierda del gráfico de la Figura 4A) y niveles más bajos de tetradecano ($m/z = 71$) y pentadecano ($m/z = 71$) tenían una función pulmonar más alta; y los que tenían valores más bajos (situados a la derecha del gráfico de la Figura 4A) y niveles más altos de tetradecano ($m/z = 71$) y pentadecano ($m/z = 71$) presentaban una función pulmonar

más baja. Por lo tanto, estos resultados también constatan que la función pulmonar y el estrés oxidativo están estrechamente asociados. Hasta donde sabemos, no se ha descrito, previamente, ninguna asociación entre los parámetros de función pulmonar y los niveles de tetradecano y pentadecano en el aliento exhalado. Sin embargo, estos compuestos, principalmente el tetradecano, sí que han sido anteriormente identificados como característicos del volatiloema de los niños con asma alérgica^{25,26}.

CONCLUSIÓN

Los hallazgos de este estudio muestran, por primera vez hasta donde sabemos, la relación entre la función pulmonar y los niveles en aire exhalado de VOCs relacionados con el estrés oxidativo en lactantes sanos. Además, nuestros resultados contribuyen a la caracterización de la función pulmonar y el volatiloema de lactantes sanos. Finalmente, este trabajo demuestra que esta técnica analítica tiene una gran importancia en la aplicación clínica hospitalaria no invasiva, incluso en niños lactantes no colaboradores.

MATERIALES Y MÉTODOS

Diseño del estudio y participantes

Los lactantes sanos (nacidos a término) incluidos en este estudio pertenecen a la cohorte prospectiva de nacimiento NELA (*Nutrition in Early Life and Asthma*)^{16,24,27}. El reclutamiento de los lactantes se realizó durante 36 meses (desde marzo de 2015 hasta abril de 2018) y se llevó a cabo en la visita de seguimiento de la semana 20 de gestación en la Unidad Materno fetal del Hospital Universitario Virgen de la Arrixaca (Murcia, España). Los criterios de inclusión fueron los siguientes: edad materna entre 18 y 45 años; residencia habitual en el Área de Salud I, y algunas comarcas de las Áreas de Salud VI y VII de la Región de Murcia; previsión de vivir en el mismo lugar de residencia durante al menos 2 años; intención de dar a luz en el hospital de referencia; origen caucásico español; embarazo único; concepción no asistida; y ecografía normal a las 20 semanas de embarazo (sin malformaciones graves). Los criterios de exclusión incluían: la presencia de enfermedades crónicas en la madre (excepto asma); complicaciones durante el embarazo (excepto diabetes gestacional y trastornos hipertensivos); y que la madre tuviera intención de dar a luz en otro hospital. De las 1.350 mujeres invitadas a participar, 738 fueron finalmente incluidas en el estudio NELA. Las pruebas de función pulmonar se llevaron a cabo cuando los niños tenían entre 3 y 8 meses de edad, y las muestras de aliento se recogieron a los 3 meses de edad en una

submuestra del estudio NELA (niños cuyas madres aceptaron ambas pruebas). Así, un total de 55 niños se sometieron a ambos procedimientos. Ambas pruebas se pudieron realizar con éxito en 50 de los niños (lactantes incluidos en el presente estudio), ya que en cinco niños no se pudo efectuar la prueba de función pulmonar porque no estaban completamente dormidos.

El protocolo del estudio fue revisado y aprobado por el Comité de Ética del Hospital Clínico Universitario Virgen de la Arrixaca de acuerdo con las directrices de la Declaración de Helsinki (CEIC 09/14). Se obtuvo el consentimiento informado por escrito de las madres de los participantes tanto en el momento del reclutamiento como antes de realizar las pruebas de función pulmonar.

Evaluación de la relación entre la función pulmonar o el volatilo exhalado y las características demográficas de los sujetos

La información sobre las características demográficas de los sujetos se obtuvo mediante cuestionarios realizados en persona a las madres de los niños durante la 20^a semana de embarazo. Las variables incluidas en el presente estudio fueron: edad del padre; edad de la madre; clase social del padre (definida como la ocupación durante el embarazo en función de la clase social más alta, utilizando una adaptación española muy extendida del sistema de codificación internacional ISCO88: I-II, directivos/técnicos; III, trabajadores cualificados; IV-V, trabajadores semicualificados/sin cualificar; y desempleados)²⁸; clase social de la madre; nivel educativo del padre (secundaria incompleta o menos, secundaria completa y universitaria); nivel educativo de la madre; asma en el padre (sí/no); asma en la madre (sí/no); tabaquismo paterno durante el embarazo (sí/no); y tabaquismo materno durante el embarazo (sí/no). Además, se midió el peso y la altura de los niños en la Unidad de Neumología Pediátrica del Hospital Universitario Infantil Virgen de la Arrixaca (Murcia, España) antes de realizar las pruebas de función pulmonar. También se tuvo en cuenta la estación del año en la que se recogieron las muestras de aire exhalado para el pretratamiento de los datos (invierno: diciembre-febrero, primavera: marzo-mayo, verano: junio-agosto, y otoño: septiembre-noviembre)

Medición de la función pulmonar

La FVC, el FEV_{0,5}, el FEF₇₅, el FEF₂₅₋₇₅ y el cociente FEV_{0,5}/FVC se obtuvieron en la Unidad de Neumología Pediátrica del Hospital Universitario Infantil Virgen de la Arrixaca a partir de curvas de volumen espiratorio máximo mediante la técnica de

compresión toraco-abdominal rápida de volumen elevado, según las especificaciones de la ATS/ERS²⁹⁻³². Los niños fueron sedados con hidrato de cloral (80-100 mg/Kg), y la saturación de oxígeno fue controlada hasta la recuperación de la sedación. Se empleó un pletismógrafo Master-Screen BabyBody para las mediciones de la función pulmonar, y un reanimador infantil Neopuff (Fisher & Paykel Healthcare) acoplado a una mascarilla para la elevación del volumen pulmonar a una presión estándar de 30 cm H₂O.

Recogida y análisis de muestras de aire exhalado

Los protocolos seguidos para el muestreo y análisis del aire exhalado de los lactantes pueden consultarse en el Capítulo 4²⁷. Se llevó a cabo una recogida no invasiva del aire exhalado mixto de lactantes de tres meses. Las muestras de aire exhalado se recogieron empleando una mascarilla conectada a bolsas de muestreo de gas Quintron® de 400 mL (con una resistencia de llenado mínima), y posteriormente el contenido fue transferido a una bolsa de muestreo de gas Tedlar® de 1 L debido a su mayor hermeticidad. Posteriormente, el aire exhalado se transfirió a tubos de desorción térmica (Tenax TA/carbograph 5td, Markes International). Además, por cada muestra de aire exhalado humano, también se recogió una muestra de aire de la sala de muestreo mediante una jeringa Easy-VOC (Markes International), para evaluar la influencia del aire ambiental. El análisis del aire exhalado y de las muestras ambientales se realizó mediante un sistema de desorción térmica acoplado a la cromatografía de gases-espectrometría de masas de cuadrupolo simple (TD-GC/q-MS). Además, dos estándares químicos fueron analizados para calcular los índices de retención de los VOCs de las muestras (estándar de n-alcenos saturados C7-C30 y estándar de calibración de VOCs, Sigma-Aldrich).

Preprocesamiento de datos y pretratamiento estadístico

Utilizando el flujo de trabajo descrito en el Capítulo 4, a partir del preprocesamiento de los datos brutos del análisis del aire exhalado por TD-GC/q-MS, se obtuvo una matriz con las intensidades normalizadas de los *features* o fragmentos iónicos (señales *m/z*) detectados en las muestras de aire exhalado²⁷. Este flujo de trabajo incluye funciones de tres paquetes de R, *xcms*³³, *cliqueMS*³⁴, y *eRah*³⁵, y permite determinar a qué compuesto pertenece cada fragmento iónico, así como la identidad de estos compuestos. La identificación de los compuestos se realizó mediante el cotejo con los espectros de masas e índices de retención disponibles en la biblioteca NIST (*National Institute of Standard and Technology*), a través del cálculo del factor de coincidencia o *match factor* y del valor de *error RI* (error del índice de retención). Tras el

preprocesamiento de los datos, se realizó un pretratamiento estadístico mediante ASCA (análisis de la varianza (ANOVA)-análisis de componentes simultáneos) (paquete *MetStaT*)^{36,37} para obtener una matriz útil para el análisis estadístico sin la variación causada por factores como la estación del año en la que se tomaron las muestras de aliento y la contaminación de la residencia habitual²⁴ (Capítulo 5). La zona de contaminación de la residencia habitual se definió mediante la modelización de la calidad del aire utilizando el sistema de modelización Weather Research and Forecasting + CHIMERE, bajo la consideración de los niveles de ozono (O₃), dióxido de nitrógeno (NO₂), dióxido de azufre (SO₂) y materia particulada^{38,39}. De la matriz de datos generada, se seleccionaron los *features* más característicos de seis compuestos: acetona (*feature m/z* = 58), isopreno (*feature m/z* = 67), decano (*feature m/z* = 71), undecano (*feature m/z* = 57), tetradecano (*feature m/z* = 71) y pentadecano (*feature m/z* = 71) (Apéndice 7).

Análisis estadístico

La relación entre los parámetros de función pulmonar (FVC, FEV_{0,5}, cociente FEV_{0,5}/FVC, FEF₇₅ y FEF₂₅₋₇₅) y los compuestos seleccionados del aire exhalado (acetona, isopreno, decano, undecano, tetradecano y pentadecano) fue evaluada mediante tres enfoques independientes. En primer lugar, se comprobó la correlación entre los valores de los parámetros de función pulmonar convertidos en *z-scores* y los *features* (fragmentos de iones) de los compuestos seleccionados mediante la correlación de Pearson. Los *z-scores* de la FVC, el FEV_{0,5}, el cociente FEV_{0,5}/FVC, el FEF₇₅ y el FEF₂₅₋₇₅ se calcularon tal y como especifican Lum *et al.*⁴⁰. Se consideró un rango normal de *z-scores* de los parámetros de función pulmonar dentro del intervalo comprendido desde -1,96 (límite inferior de la normalidad (LLN)) a +1,96 (límite superior de la normalidad (ULN))^{40,41}. Posteriormente, se generaron modelos de regresión lineal múltiple para cada uno de los parámetros de función pulmonar y cada uno de los *features* de los VOCs seleccionados (los parámetros de función pulmonar como variables dependientes y *features* como variables independientes), incluyendo las siguientes covariables: sexo, altura, peso y edad en el momento de la prueba de función pulmonar, y tabaquismo materno durante el embarazo. Además, también se construyeron modelos de regresión lineal múltiple utilizando los *z-scores* de los parámetros de función pulmonar y los *features* de los VOCs seleccionados. Estos modelos se ajustaron sólo por el consumo de tabaco de la madre durante el embarazo. Por último, como análisis exploratorio, se realizó un análisis de componentes principales (PCA) (paquete *FactoMineR*⁴²) sobre los parámetros de función pulmonar, y se

comprobó la asociación del primer componente principal con los *features* de los VOCs seleccionados, mediante pruebas de correlación de Spearman. La evaluación de la correlación entre los niveles de los VOCs en el aire exhalado y el primer componente principal se llevó a cabo para estudiar la relación del volatiloema y la función pulmonar global en lactantes sanos, considerando los parámetros de función pulmonar como un todo.

Para evitar posibles sesgos en el análisis de la asociación entre el volatiloema y la función pulmonar infantil, la relación entre las características demográficas de los sujetos y los parámetros de función pulmonar o el nivel de los *features* de los VOCs seleccionados, se evaluó mediante las pruebas de correlación de Pearson o correlación de Spearman (variable continua frente a variable continua), pruebas t de Student o pruebas U de Mann-Whitney (variable continua frente a variable categórica con dos grupos), o pruebas ANOVA o Kruskal-Wallis (variable continua frente a variable categórica con más de dos grupos). La distribución de los datos se evaluó mediante las pruebas de Shapiro-Wilk para la selección de las pruebas estadísticas más adecuadas (pruebas paramétricas o no paramétricas). Todos los análisis estadísticos de este estudio se realizaron en R (versión 4.0.5) y se consideró como umbral de significación un p-valor < 0,05.

BIBLIOGRAFÍA

1. Xie, M., Liu, X., Cao, X., Guo, M. & Li, X. Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017. *Respir. Res.* **21**, 49 (2020).
2. Garcia-Marcos, L. *et al.* Priorities for future research into asthma diagnostic tools: A PAN-EU consensus exercise from the European asthma research innovation partnership (EARIP). *Clin. Exp. Allergy* **48**, 104–120 (2018).
3. Oliva Hernández, C. *et al.* Estudio de la función pulmonar en el paciente colaborador. Parte I. *An. Pediatr.* **66**, 393–406 (2007).
4. Mallol, J. & Aguirre, V. Advances in the study of infant lung function: Forced expiratory maneuvers from an increased lung volume. *Arch. Bronconeumol.* **43**, 233–238 (2007).
5. Tenero, L., Zaffanello, M., Piazza, M. & Piacentini, G. Measuring Airway Inflammation in Asthmatic Children. *Front. Pediatr.* **6**, 196 (2018).
6. Ochs-Balcom, H. M. *et al.* Oxidative stress and pulmonary function in the general population. *Am. J. Epidemiol.* **162**, 1137–1145 (2005).
7. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
8. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. . Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
9. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
10. Smolinska, A. *et al.* Profiling of Volatile Organic Compounds in Exhaled Breath As a Strategy to Find Early Predictive Signatures of Asthma in Children. *PLoS One* **9**, e95668 (2014).
11. Muñoz-Lucas, M. Á. *et al.* Influence of Chronic Obstructive Pulmonary Disease on Volatile Organic Compounds in Patients with Non-small Cell Lung Cancer. *Arch. Bronconeumol.* **56**, 801–805 (2020).
12. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).

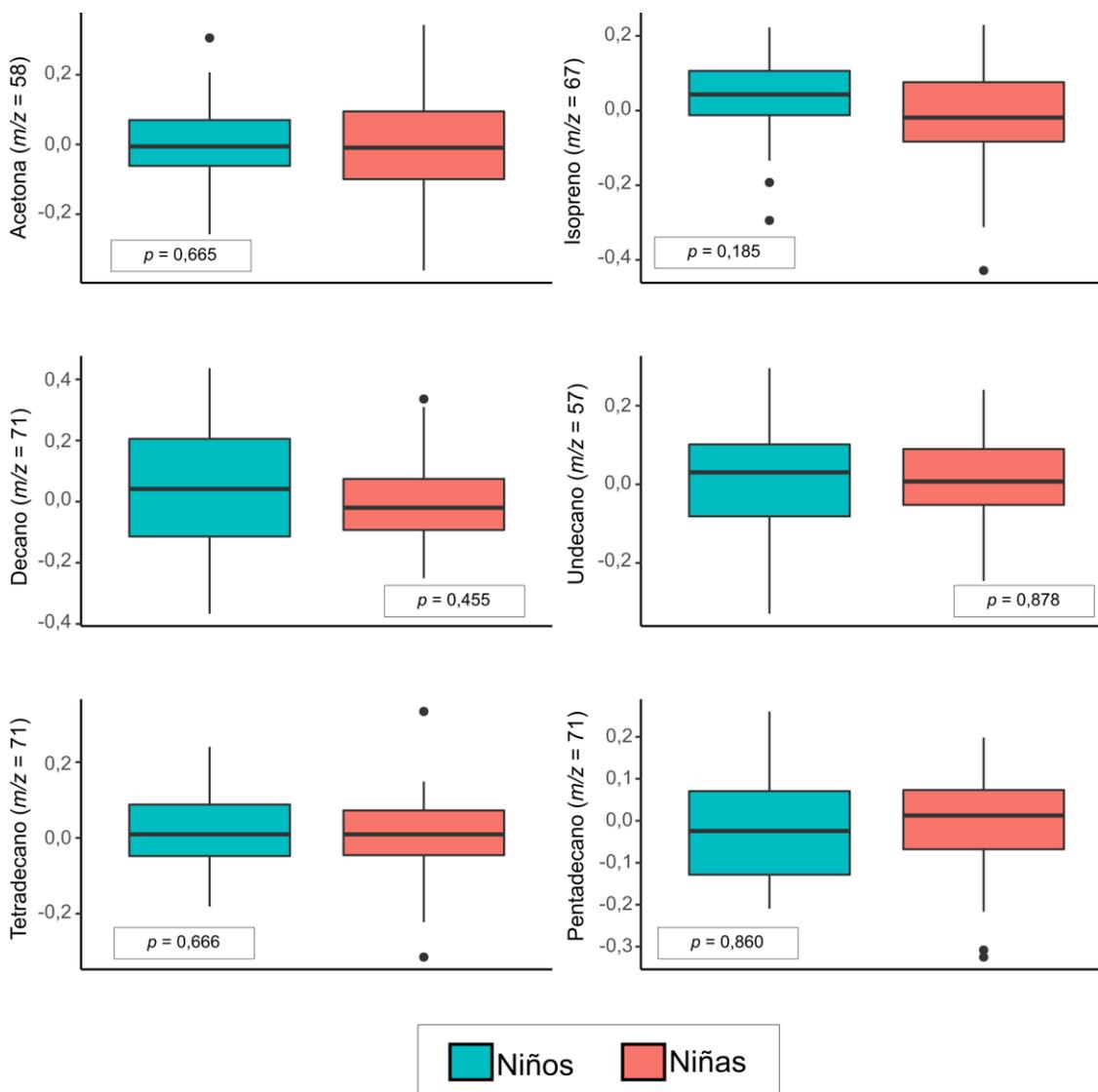
13. Sola Martínez, R. A. *et al.* Exhaled volatile organic compounds analysis in clinical pediatrics: a systematic review. *Pediatr. Res.* **89**, 1352–1363 (2021).
14. Sanchez-Solis, M. & Garcia-Marcos, L. Lung function in wheezing infants. *Frontiers in Bioscience - Elite* **6 E**, 185–197 (2014).
15. Fuglsbjerg, M. G. *et al.* Limited clinical value of exhaled volatile organic compound measurements in childhood asthma. *ERJ Open Res.* **4**, 00026–02018 (2018).
16. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).
17. Mitsui, T., Kondo, T., Phillips, M., Cataneo, R. N. & Greenberg, J. Inadequacy of theoretical basis of breath methylated alkane contour for assessing oxidative stress. *Clinica Chimica Acta* **333**, 91–94 (2003).
18. Silva, C. L., Perestrelo, R., Silva, P., Tomás, H. & Câmara, J. S. Volatile metabolomic signature of human breast cancer cell lines. *Sci. Rep.* **7**, 43969 (2017).
19. Scholpp, J., Schubert, J. K., Miekisch, W. & Geiger, K. Breath markers and soluble lipid peroxidation markers in critically ill patients. *Clin. Chem. Lab. Med.* **40**, 587–594 (2002).
20. Calenic, B. *et al.* Oxidative stress and volatile organic compounds: Interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* **13**, 1020–1030 (2015).
21. Phillips, M. *et al.* Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress. *J. Lab. Clin. Med.* **136**, 243–249 (2000).
22. Dragonieri, S. *et al.* An electronic nose in the discrimination of patients with asthma and controls. *J. Allergy Clin. Immunol.* **120**, 856–862 (2007).
23. Paredi, P. *et al.* Exhaled ethane, a marker of lipid peroxidation, is elevated chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **162**, 369–373 (2000).
24. Sola-Martínez, R. A. *et al.* Exhaled volatilome analysis as a useful tool to discriminate asthma with other coexisting atopic diseases in women of childbearing age. *Sci. Rep.* **11**, 13823 (2021).
25. Caldeira, M. *et al.* Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. *J. Chromatogr. A* **1218**, 3771–80 (2011).

26. Caldeira, M. *et al.* Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* **1254**, 87–97 (2012).
27. Sola Martínez, R. A. *et al.* Data preprocessing workflow for exhaled breath analysis by GC/MS using open sources. *Sci. Rep.* **10**, 22008 (2020).
28. Domingo-Salvany, A., Regidor, E., Alonso, J. & Alvarez-Dardet, C. Proposal for a social class measure. Working Group of the Spanish Society of Epidemiology and the Spanish Society of Family and Community Medicine. *Aten. Primaria* **25**, 350–363 (2000).
29. Culver, B. H. *et al.* Recommendations for a standardized pulmonary function report. An official American Thoracic Society technical statement. *Am. J. Respir. Crit. Care Med.* **196**, 1463–1472 (2017).
30. American Thoracic Society (ATS) and the European Respiratory Society (ERS). ATS/ERS Statement: Raised volume forced expirations in infants - Guidelines for current practice. *Am. J. Respir. Crit. Care Med.* **172**, 1463–1471 (2005).
31. Sanchez-Solis, M., Perez-Fernandez, V., Bosch-Gimenez, V., Quesada, J. & Garcia-Marcos, L. Lung function gain in preterm infants with and without bronchopulmonary dysplasia. *Pediatr. Pulmonol.* **51**, 936–942 (2016).
32. Sanchez-Solis, M., Garcia-Marcos, P. W., Agüera-Arenas, J., Mondejar-Lopez, P. & Garcia-Marcos, L. Impact of early caffeine therapy in preterm newborns on infant lung function. *Pediatr. Pulmonol.* **55**, 102–107 (2020).
33. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
34. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
35. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).
36. Smilde, A. K. *et al.* ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **21**, 3043–3048 (2005).
37. Bertinetto, C., Engel, J. & Jansen, J. ANOVA simultaneous component analysis: A tutorial review. *Analytica Chimica Acta: X* **6**, 100061 (2020).

38. García-Serna, A. M. *et al.* Air pollution from traffic during pregnancy impairs newborn's cord blood immune cells: The NELA cohort. *Environ. Res.* **198**, 110468 (2021).
39. Skamarock, W. C. *et al.* *A Description of the Advanced Research WRF Version 3. Technical Report* (2008).
40. Lum, S. *et al.* New reference ranges for interpreting forced expiratory manoeuvres in infants and implications for clinical interpretation: A multicentre collaboration. *Thorax* **71**, 276–283 (2016).
41. Quanjer, P. H. *et al.* Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur. Respir. J.* **40**, 1324–1343 (2012).
42. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

APÉNDICE 1. Características demográficas de los progenitores de los lactantes incluidos en este estudio.

	Padres de los niños	Madres de los niños
Edad (años), media (DE)	35,0 (5,8)	32,5 (5,3)
Clase social Directivos-técnicos/ trabajadores cualificados/ trabajadores semicualificados-sin cualificar/ desempleados, %	20/14/58/8	24/18/24/34
Nivel educativo Secundaria incompleta o menos/ secundaria completa/ universitaria, %	26/52/22	22/40/38
Tabaquismo durante el embarazo (Sí), %	21 (50%)	7 (14%)
Asma (Sí), %	9 (18%)	5 (10%)



APÉNDICE 2. Influencia del sexo en el volatiloma exhalado en niños.

APÉNDICE 3. Relación entre la estación del año en que se recogieron las muestras de aire exhalado y los parámetros de función pulmonar en lactantes sanos.

	Estación del año de muestreo de aire exhalado			
	Invierno (n = 7)	Primavera (n = 13)	Verano (n = 13)	Otoño (n = 17)
FVC (mL), mediana (IQR)	215 (26)	226 (83)	230 (56)	220 (26)
FEV_{0,5} (mL), mediana (IQR)	171 (19)	162 (28)	181 (42)	172 (21)
FEV_{0,5}/FVC (mL), mediana (IQR)	0,82 (0,12)	0,80 (0,18)	0,79 (0,07)	0,81 (0,10)
FEF₇₅ (mL), mediana (IQR)	176 (54)	174 (63)	194 (41)	188 (36)
FEF₂₅₋₇₅ (mL), mediana (IQR)	310 (98)	304 (46)	320 (54)	314 (63)

En negrita se muestran los resultados estadísticamente significativos (p-valor < 0,05). IQR: Rango intercuartílico; FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

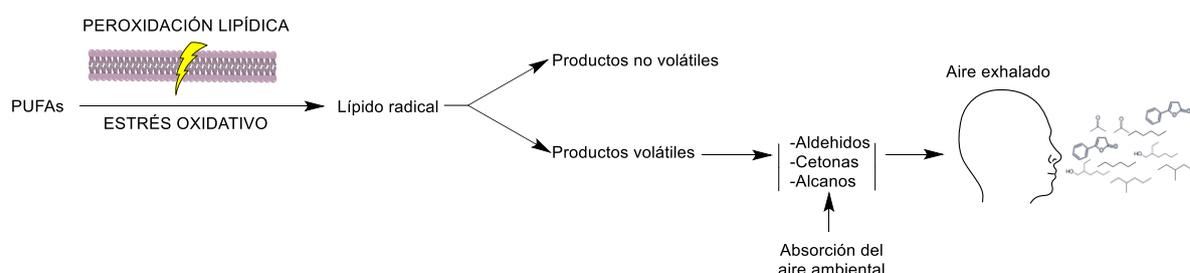
APÉNDICE 4. Análisis de regresión lineal múltiple para determinar la relación entre los parámetros de función pulmonar (*z-score*) y los *features* de los compuestos orgánicos volátiles (VOCs) del aire exhalado de los lactantes. Los modelos fueron ajustados por el tabaquismo materno durante el embarazo.

		β^1	IC 95% ²	p-valor
FVC (<i>z-score</i>)	Acetona (<i>m/z</i> = 58)	-2,053	-3,592; -0,513	0,010
	Isopreno (<i>m/z</i> = 67)	-0,059	-1,868; 1,750	0,948
	Decano (<i>m/z</i> = 71)	-0,651	-2,054; 0,752	0,355
	Undecano (<i>m/z</i> = 57)	-0,002	-1,935; 1,931	0,998
	Tetradecano (<i>m/z</i> = 71)	-1,679	-3,824; 0,466	0,122
	Pentadecano (<i>m/z</i> = 71)	-1,076	-3,064; 0,912	0,282
FEV_{0,5} (<i>z-score</i>)	Acetona (<i>m/z</i> = 58)	-1,802	-3,048; -0,557	0,006
	Isopreno (<i>m/z</i> = 67)	-0,411	-1,887; 1,066	0,579
	Decano (<i>m/z</i> = 71)	-0,784	-1,920; 0,352	0,172
	Undecano (<i>m/z</i> = 57)	-0,734	-2,302; 0,834	0,351
	Tetradecano (<i>m/z</i> = 71)	-1,976	-3,682; -0,270	0,024
	Pentadecano (<i>m/z</i> = 71)	-1,870	-3,424; -0,317	0,019
FEV_{0,5}/FVC (<i>z-score</i>)	Acetona (<i>m/z</i> = 58)	-0,222	-2,721; 2,277	0,859
	Isopreno (<i>m/z</i> = 67)	-0,721	-3,449; 2,007	0,598
	Decano (<i>m/z</i> = 71)	-0,418	-2,556; 1,720	0,696
	Undecano (<i>m/z</i> = 57)	-1,804	-4,679; 1,071	0,213
	Tetradecano (<i>m/z</i> = 71)	-0,974	-4,290; 2,342	0,557
	Pentadecano (<i>m/z</i> = 71)	-2,112	-5,092; 0,868	0,161
FEF₇₅ (<i>z-score</i>)	Acetona (<i>m/z</i> = 58)	-0,608	-2,070; 0,853	0,407
	Isopreno (<i>m/z</i> = 67)	-0,471	-2,077; 1,135	0,558
	Decano (<i>m/z</i> = 71)	-0,510	-1,762; 0,742	0,417
	Undecano (<i>m/z</i> = 57)	-0,801	-2,507; 0,905	0,350
	Tetradecano (<i>m/z</i> = 71)	-1,380	-3,298; 0,538	0,154
	Pentadecano (<i>m/z</i> = 71)	-1,189	-2,947; 0,570	0,180
FEF₂₅₋₇₅ (<i>z-score</i>)	Acetona (<i>m/z</i> = 58)	-0,835	-2,299; 0,629	0,257
	Isopreno (<i>m/z</i> = 67)	-0,337	-1,958; 1,285	0,678
	Decano (<i>m/z</i> = 71)	-0,403	-1,669; 0,862	0,525
	Undecano (<i>m/z</i> = 57)	-0,915	-2,630; 0,800	0,289
	Tetradecano (<i>m/z</i> = 71)	-1,460	-3,388; 0,469	0,135
	Pentadecano (<i>m/z</i> = 71)	-1,881	-3,601; -0,161	0,033

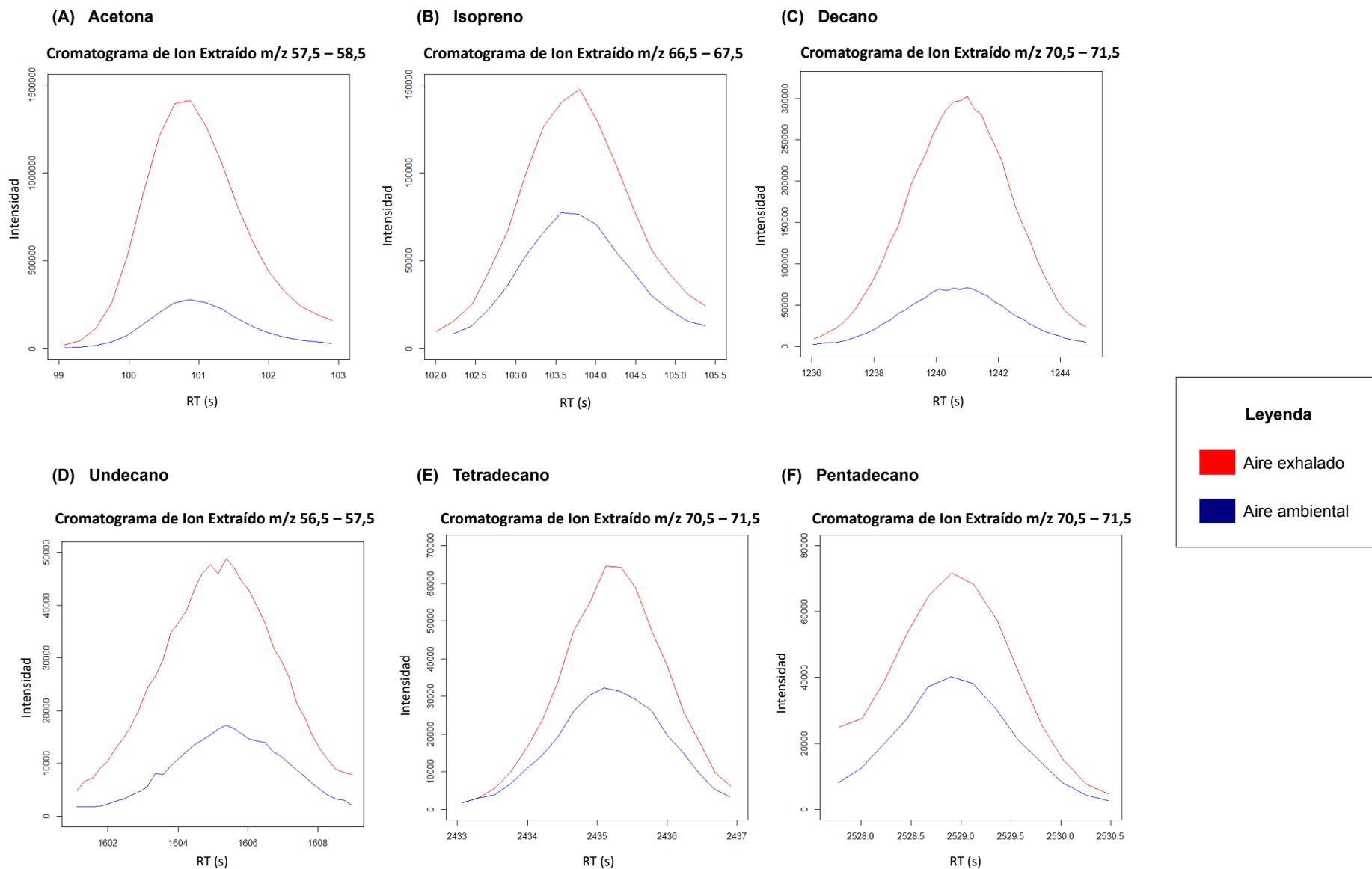
En negrita se muestran los resultados estadísticamente significativos (p -valor < 0,05). FVC: capacidad vital forzada; FEV_{0,5}: volumen espiratorio forzado a 0,5 s; FEF₇₅: flujo espiratorio forzado al 75% de la FVC; FEF₂₅₋₇₅: flujo espiratorio forzado al 25%-75% de la FVC.

¹Coefficiente de regresión

²Intervalo de confianza del 95%



APÉNDICE 5. Monitorización del estrés oxidativo mediante el análisis de compuestos orgánicos volátiles en el aliento exhalado. Este diagrama muestra el efecto del estrés oxidativo sobre los ácidos grasos poliinsaturados (PUFAs).



APÉNDICE 6. Intensidad de los *features* de compuestos orgánicos volátiles (VOCs) seleccionados en una muestra de aire exhalado humano y su correspondiente muestra de aire ambiental. (A) Intensidad de la acetona ($m/z = 58$) en el aire exhalado y en el aire ambiental. (B) Intensidad del isopreno ($m/z = 67$) en el aire exhalado y en el aire ambiental. (C) Intensidad del decano ($m/z = 71$) en el aire exhalado y en el aire ambiental. (D) Intensidad del undecano ($m/z = 57$) en el aire exhalado y en el aire ambiental. (E) Intensidad del tetradecano ($m/z = 71$) en el aire exhalado y en el aire ambiental. (F) Intensidad del pentadecano ($m/z = 71$) en el aire exhalado y en el aire ambiental.

APÉNDICE 7. Compuestos orgánicos volátiles (VOCs) estudiados en el presente estudio.

VOCs	m/z	Código CAS	Match factor y error RI	Solamente Match factor
Acetona	58	67-64-1	X	
Isopreno	67	78-79-5		X
Decano	71	124-18-5	X	
Undecano	57	1120-21-4	X	
Tetradecano	71	629-59-4	X	
Pentadecano	71	629-62-9	X	

Capítulo 7

Influencia de la exposición a la humedad en el interior del hogar sobre los niveles de compuestos orgánicos volátiles en el aire exhalado

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, Gema Lozano-Terol, Julia Gallego-Jara, Eva Morales, Luis García-Marcos, Pedro Jiménez-Guerrero, Manuel Cánovas Díaz y Teresa de Diego Puente

Influence of home indoor dampness exposure on volatile organic compounds in exhaled breath of mothers and their infants: the NELA birth cohort

Indoor Air (en revisión).

RESUMEN

Actualmente, el efecto de la exposición a los contaminantes del aire interior y la presencia de humedad en el hogar sobre la salud respiratoria/atópica preocupa especialmente a los médicos. La determinación de compuestos orgánicos volátiles (VOCs) en el aire exhalado constituye un método útil para controlar la exposición ambiental. Una gran ventaja de esta estrategia es que permite estudiar el impacto de los contaminantes en el metabolismo mediante un método no invasivo. En este trabajo se evaluaron los niveles de nueve VOCs (acetona, isopreno, tolueno, p/m-xileno, o-xileno, estireno, benzaldehído, naftaleno y 2-etil-1-hexanol) en el aire exhalado de sujetos expuestos y no expuestos a humedad en el hogar. Se recogieron muestras de aire exhalado de 337 parejas de madres e hijos de la cohorte prospectiva de nacimiento NELA (*Nutrition in Early Childhood Asthma*), y fueron analizadas mediante cromatografía de gases-espectrometría de masas. Se observó que los niveles de 2-etil-1-hexanol en el aire exhalado de las madres estaban significativamente influidos por la exposición a la humedad en el hogar. En el caso de los lactantes, se observaron diferencias en los niveles de algunos VOCs relacionados con la exposición a la humedad en el hogar; sin embargo, no alcanzaron significación estadística. Además, también se constató que los recuentos en sangre de eosinófilos de las madres expuestas a la humedad en el hogar fueron significativamente elevados en comparación con los de las no expuestas. Hasta donde sabemos, estos resultados muestran por primera vez que la exposición a la humedad en el hogar puede influir en los patrones de VOCs en el aire exhalado.

INTRODUCCIÓN

En los últimos años se ha registrado un aumento considerable del número de personas que padecen asma o enfermedades atópicas como la rinitis y la dermatitis en todo el mundo¹. El problema es especialmente grave en la población pediátrica, donde el asma es actualmente la enfermedad crónica más prevalente². La aparición de eczema atópico y sibilancias recurrentes durante los primeros años de vida está relacionada con el desarrollo de asma en el futuro^{3,4}. En las últimas décadas, la presencia de humedad en el hogar y la exposición a contaminantes ambientales, tanto en el interior como en el exterior, han sido señalados como posibles factores de riesgo para el desarrollo de enfermedades atópicas⁵⁻⁷. En la actualidad, los ciudadanos pasan muchas horas en lugares cerrados (en el hogar, en los lugares de trabajo, en las escuelas, etc.), por ello, el estudio del impacto de los contaminantes del aire interior y la presencia de humedad

en el hogar sobre el desarrollo del asma y otras enfermedades atópicas ha constituido un objetivo prioritario de la comunidad científica en los últimos años^{8,9}.

Los contaminantes del aire interior pueden proceder del humo de tabaco del entorno, de productos para el hogar o de materiales de construcción, entre otros. Varios compuestos orgánicos volátiles (VOCs) pueden ser emitidos por estas fuentes y entrar en las vías respiratorias humanas¹⁰. Además, se ha observado que la emisión de VOCs es más elevada en los edificios de nueva construcción^{11,12}. La exposición a los VOCs contaminantes suele medirse mediante el análisis del aire interior. Además, la exposición personal puede evaluarse mediante muestreadores OVM (Monitores de vapores orgánicos). Estos dispositivos son portados por los sujetos temporalmente y permiten medir los VOCs a los que éstos están expuestos^{13,14}. Los compuestos orgánicos volátiles pueden determinarse con múltiples tipos de plataformas analíticas, pero las tecnologías basadas en la espectrometría de masas, por ejemplo, la cromatografía de gases-espectrometría de masas (GC-MS), son las más extendidas¹⁵. Además, los VOCs también pueden detectarse en diferentes matrices humanas como la orina y el aire exhalado¹⁶⁻¹⁸. De hecho, el análisis de VOCs en el aire exhalado proporciona información relevante sobre la exposición personal a contaminantes ambientales (exposoma). Por otra parte, los VOCs endógenos derivados del metabolismo humano y la flora intestinal también pueden ser detectados en el aire exhalado¹⁹. Por lo tanto, el análisis de VOCs es una estrategia útil para controlar la inflamación o los cambios en el metabolismo humano debidos a enfermedades y/o a la exposición a contaminantes ambientales^{11,20}. De hecho, VOCs relacionados con el estrés oxidativo pueden ser detectados en el aire exhalado²¹.

El efecto de la exposición a la humedad en el desarrollo de los síntomas respiratorios ha sido ampliamente estudiado, y se ha sugerido que la exposición prenatal y postnatal a la humedad en el hogar puede estar involucrada en la aparición de sibilancias, rinitis y asma en la infancia²²⁻²⁴. Además, la humedad en interiores y los niveles de algunos VOCs en el interior de las viviendas, como el tolueno, el etilbenceno o el 2-etil-1-hexanol, también han sido asociados con el síndrome del edificio enfermo (SBS), que causa problemas respiratorios en sus ocupantes, y que, además, puede acabar provocando asma, entre otros síntomas²⁵⁻²⁷. En este sentido, se ha informado que la exposición a los microorganismos del moho y los VOCs liberados en los lugares en los que hay mucha humedad podrían ser las principales causas implicadas en los cambios del estado de salud relacionados con la exposición a la humedad^{28,29}. En consecuencia, el análisis de la influencia de la exposición a la humedad en interiores,

en relación con la salud, sería útil para establecer directrices para reducir el riesgo de desarrollo de enfermedades³⁰. Hasta la fecha, la inflamación inducida por la exposición a la humedad en las residencias y su impacto en el metabolismo humano se ha medido mediante el análisis de la fracción exhalada de óxido nítrico (FeNO) y el recuento de eosinófilos en el lavado nasal y en sangre, respectivamente³¹⁻³⁴.

Por todo ello, en este trabajo se presenta un estudio realizado en un país de clima mediterráneo como España, cuyo objetivo era evaluar los efectos de la exposición a la humedad en el hogar sobre los patrones de VOCs en el aire exhalado y sobre la aparición de trastornos atópicos. Se trata del primer estudio, hasta donde sabemos, en el que se han analizado los niveles de 9 VOCs (acetona, isopreno, tolueno, p/m-xileno, o-xileno, estireno, benzaldehído, naftaleno y 2-etil-1-hexanol) en el aire exhalado de mujeres de edad fértil y sus hijos con este fin. La acetona y el isopreno son dos de los VOCs más frecuentes en el aliento exhalado que proceden del metabolismo humano, mientras que los otros 7 VOCs proceden de contaminantes atmosféricos. Entre ellos, los compuestos aromáticos (tolueno, estireno, etc.) podrían proceder de diversas fuentes, como la combustión de carburantes, el tabaquismo o las reformas domésticas, etc.^{16,27,35}.

RESULTADOS

Características de la población de estudio

Las características de la población de estudio se muestran en la Tabla 1, la Tabla 2 y el Apéndice 1. Más del 25% de los hogares mostraban humedad en las paredes, el 20% de los hogares contaban con uno de los progenitores con asma, el 50% de los hogares con uno de los progenitores con rinitis alérgica y el 17% de los hogares con uno de los progenitores con dermatitis atópica. Durante los tres primeros meses de vida, el 12% de los niños presentó eczema atópico y el 3% sufrió algún episodio de infección respiratoria. No se demostró ninguna asociación significativa entre la existencia de humedad en el hogar y las características demográficas de la población estudiada, como la edad, el peso y la altura, la clase social o el nivel de estudios (Tabla 1 y Tabla 2). Se recogieron y analizaron un total de 674 muestras de aire exhalado de las madres y los niños de 3 meses del estudio NELA (*Nutrition in Early Childhood Asthma*)³⁶. Además, se recogieron 337 muestras de aire ambiental de la habitación donde se realizó el muestreo, una por cada pareja madre-niño, con el fin de confirmar que las diferencias observadas en el aire exhalado humano se debían a la exposición a la humedad en el hogar y no a la interferencia del aire ambiental en la sala de muestreo del aliento. El

tiempo medio mínimo de exposición a la humedad en el hogar de los sujetos fue de 7 meses (4 meses de exposición durante el embarazo y 3 meses después del parto).

Tabla 1. Características demográficas de las madres incluidas en este estudio.

	Sin humedad en el hogar (n=251)	Humedad en el hogar (n=86)	p-valor
Edad (años), media (rango)	33 (18-43)	33 (20-41)	0,896
IMC (Kg/m ²), antes del embarazo, media (rango)	23,8 (15,9-40,4)	24,3 (16,7-42,3)	0,251
Clase social (Directivos-técnicos/ trabajadores cualificados/ trabajadores semicualificados-sin cualificar/ desempleados, %)	38,2/19,9/19,1/22,7	31,4/22,1/24,4/22,1	0,602
Nivel educativo (Secundaria incompleta o menos/ secundaria completa/ universitaria, %)	16,7/25,5/57,8	17,4/27,9/54,7	0,873
Tabaquismo durante el embarazo, %	35 (13,9%)	13 (15,1%)	0,858

Tabla 2. Características demográficas de los niños incluidos en este estudio.

	Sin humedad en el hogar (n=250)	Humedad en el hogar (n=87)	p-valor
Niños / Niñas n/n	119/131	39/48	0,709
Peso al nacer (Kg), media (rango)	3,239 (1,560-4,980)	3,183 (1,940-4,505)	0,366
Talla al nacer (cm), media (rango)	51 (36-56)	50 (46-54)	0,153

Humedad en el hogar y trastornos atópicos

Se notificaron manchas de humedad o moho en la pared de la vivienda habitual del 18% de los padres asmáticos, el 19% de los padres con rinitis alérgica, el 29% de los padres con dermatitis atópica, el 24% de los padres con conjuntivitis alérgica y el 13% de los niños que habían desarrollado una enfermedad atópica o respiratoria durante los tres primeros meses de vida (Apéndice 2). La Figura 1 muestra los *odd ratio* (OR) y sus intervalos de confianza del 95% que asociaban el riesgo de padecer una enfermedad atópica con la presencia de humedad en el hogar. Como puede observarse, la humedad en el interior del hogar aumentó significativamente el riesgo de que uno de los progenitores sufriera rinitis alérgica. Por otro lado, aunque el desarrollo de una enfermedad atópica o respiratoria en los primeros meses de vida del bebé no se asoció significativamente con la exposición a la humedad del hogar, el riesgo estuvo cerca de ser significativo (p-valor = 0,095). De hecho, la frecuencia observada de síntomas respiratorios y atópicos fue superior a la esperada en los niños cuyos hogares presentaban manchas de humedad.

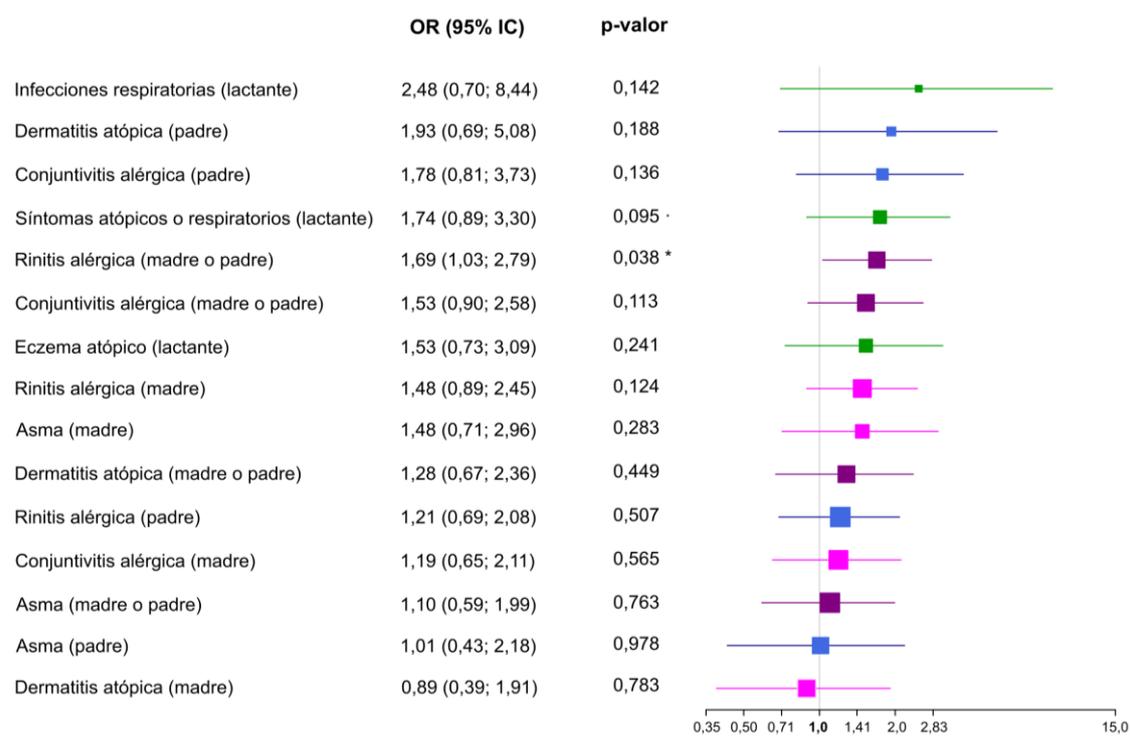


Figura 1. Humedad y trastornos atópicos. *Odd ratios* (OR) y sus intervalos de confianza del 95% (95% IC) que muestran el riesgo de sufrir una enfermedad atópica en relación con la humedad en el interior. · p-valor <0,10, * p-valor<0,05.

Humedad en el hogar y análisis del recuento de glóbulos blancos en sangre

La Tabla 3 muestra el recuento de glóbulos blancos en sangre de las madres que vivían en casas con manchas de humedad y, de las madres sin humedad en las paredes de sus viviendas. Se observaron diferencias estadísticamente significativas en el recuento de eosinófilos, de hecho, el recuento de eosinófilos fue mayor en las madres expuestas a la humedad que en las que no tenían manchas de humedad en su residencia habitual (Figura 2).

Tabla 3. Relación entre el recuento de glóbulos blancos y la exposición a humedad en interiores.

	Sin humedad en el hogar (n=251)	Humedad en el hogar (n=86)	p-valor
Recuento de leucocitos, x 10 ³ /μL	9,31 ± 2,13	9,05 ± 2,14	0,312
Recuento de linfocitos, x 10 ³ /μL	2,02 ± 0,61	1,95 ± 0,49	0,598
Recuento de monocitos, x 10 ³ /μL	0,61 ± 0,18	0,60 ± 0,16	0,879
Recuento de eosinófilos, x 10 ³ /μL	0,14 ± 0,09	0,17 ± 0,10	0,049*
Recuento de neutrófilos, x 10 ³ /μL	6,51 ± 1,67	6,30 ± 1,82	0,199
Recuento de basófilos, x 10 ³ /μL	0,03 ± 0,02	0,03 ± 0,01	0,117

*p-valor<0,05

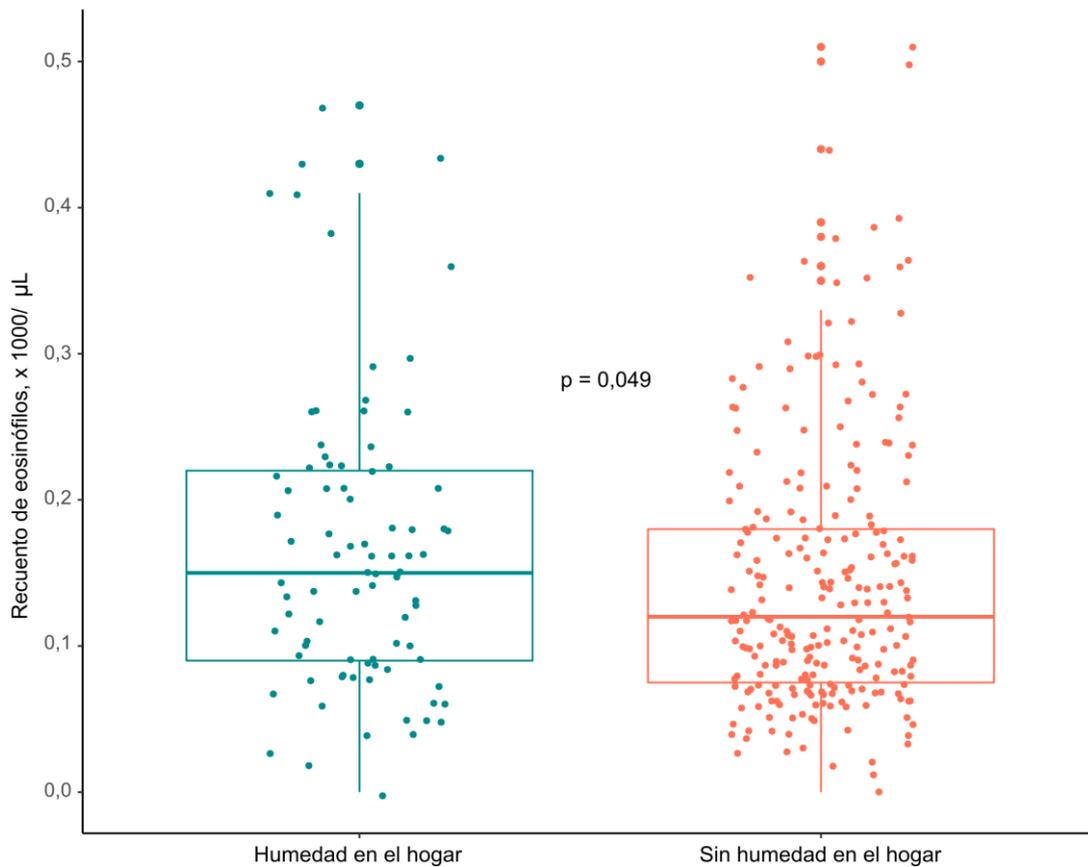


Figura 2. Influencia de la exposición a la humedad en interiores en el recuento en sangre de eosinófilos en las madres de la cohorte NELA.

Humedad en el hogar y aire exhalado de las madres

La Figura 3 muestra el impacto de la humedad en el hogar sobre el aire exhalado de las madres. Los resultados indican que los niveles de 2-etil-1-hexanol y tolueno eran significativamente más altos en el aire exhalado de las mujeres cuyos hogares presentaban humedad en las paredes. Sin embargo, los niveles de tolueno también eran significativamente altos en las muestras de aire ambiente de la habitación donde se recogieron las muestras de aire exhalado de las mujeres (Figura 4). Por lo tanto, sólo los niveles de 2-etil-1-hexanol en el aire exhalado se asociaron significativamente con la presencia de humedad en las casas.

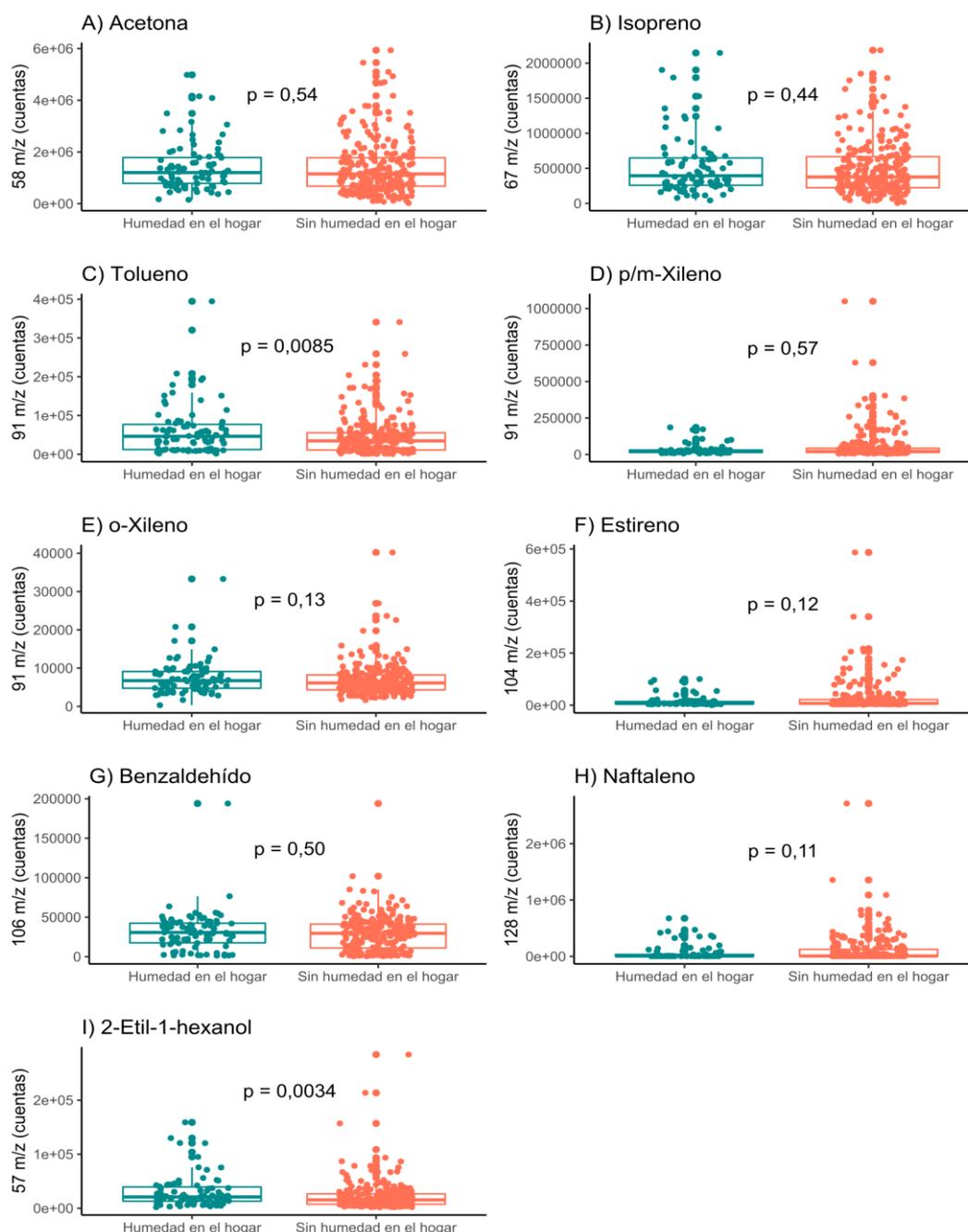


Figura 3. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire exhalado de las madres, considerando los fragmentos iónicos de los compuestos seleccionados. (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto de la exposición a la humedad en interiores sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire exhalado de las madres.

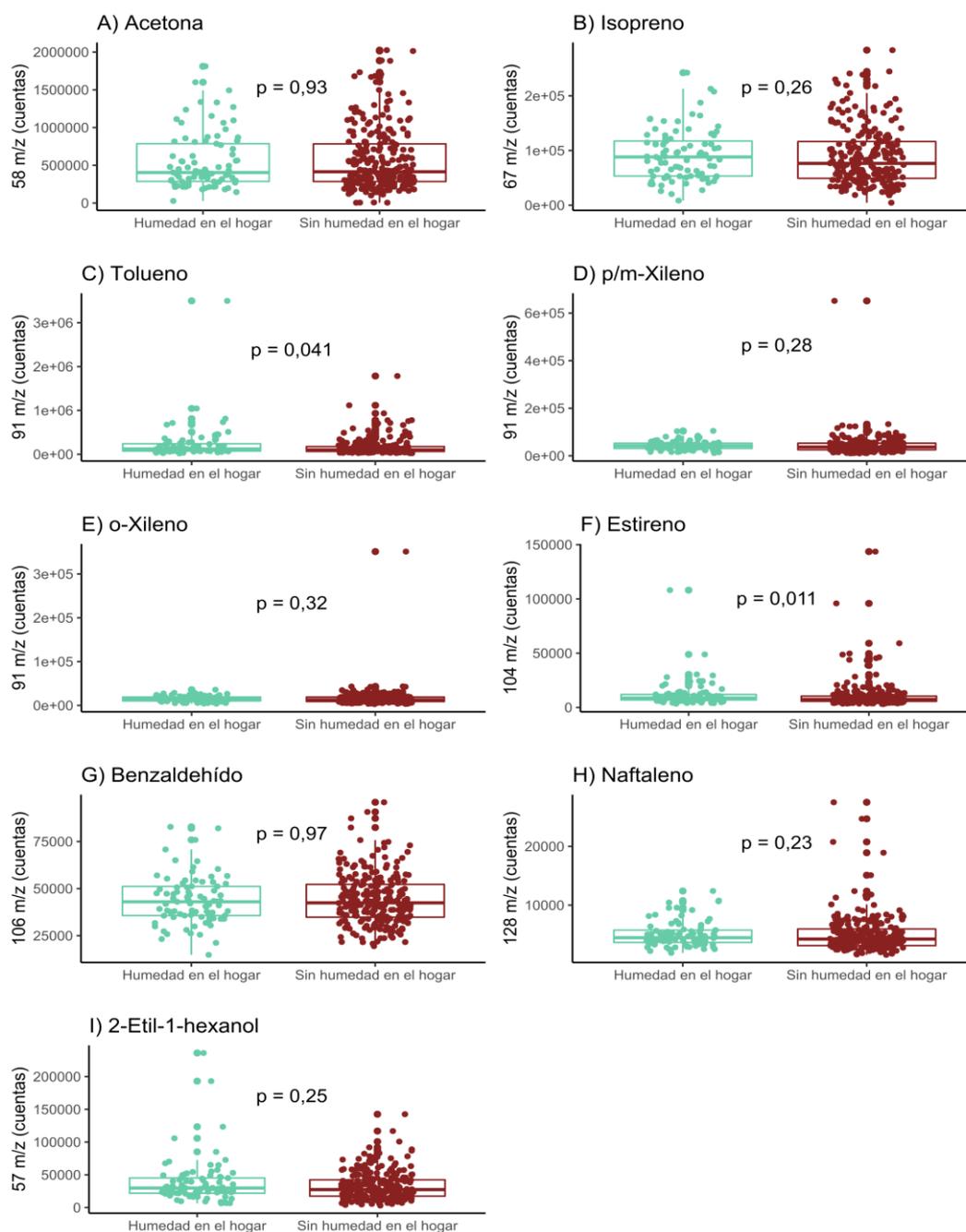


Figura 4. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire ambiental de la sala de muestreo considerando los fragmentos iónicos de los compuestos seleccionados. (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire ambiental de la sala de muestreo.

Humedad en el hogar y aire exhalado de los lactantes

La Figura 5 muestra la influencia de la humedad en el hogar sobre el aire exhalado de los lactantes. Aunque se observaron algunas diferencias, ninguno de los

VOCs seleccionados mostró una asociación estadísticamente significativa con la presencia de humedad en los hogares de los sujetos.

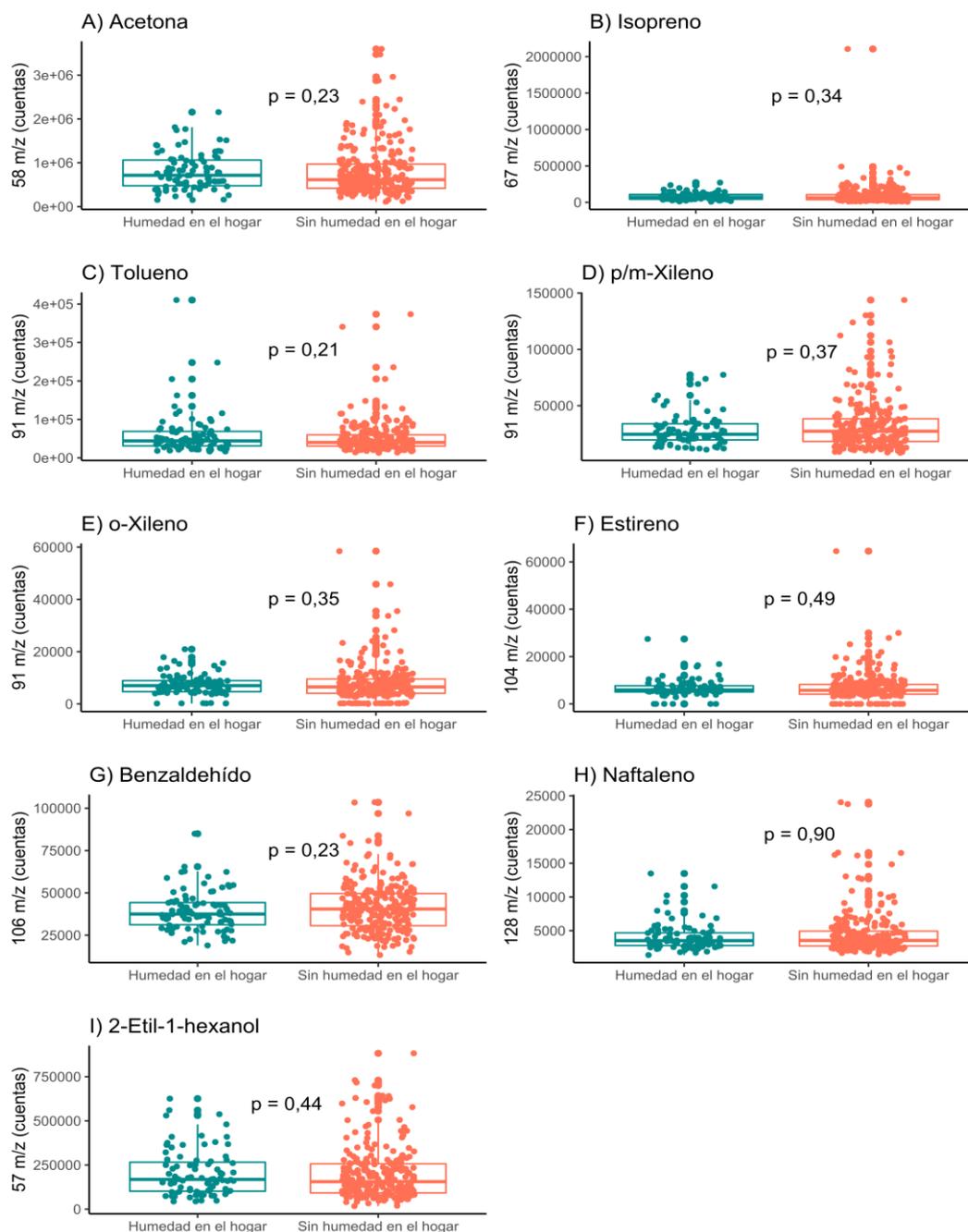


Figura 5. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire exhalado de los lactantes considerando los fragmentos iónicos de los compuestos seleccionados. (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto s sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire exhalado de los lactantes.

La asociación de los niveles de VOCs y la presencia de humedad en el hogar también se analizó estratificada por el sexo del lactante (Figura 6 y Figura 7). En el aire exhalado de los niños, no hubo diferencias significativas en los niveles de los 9 VOCs estudiados (Figura 6).

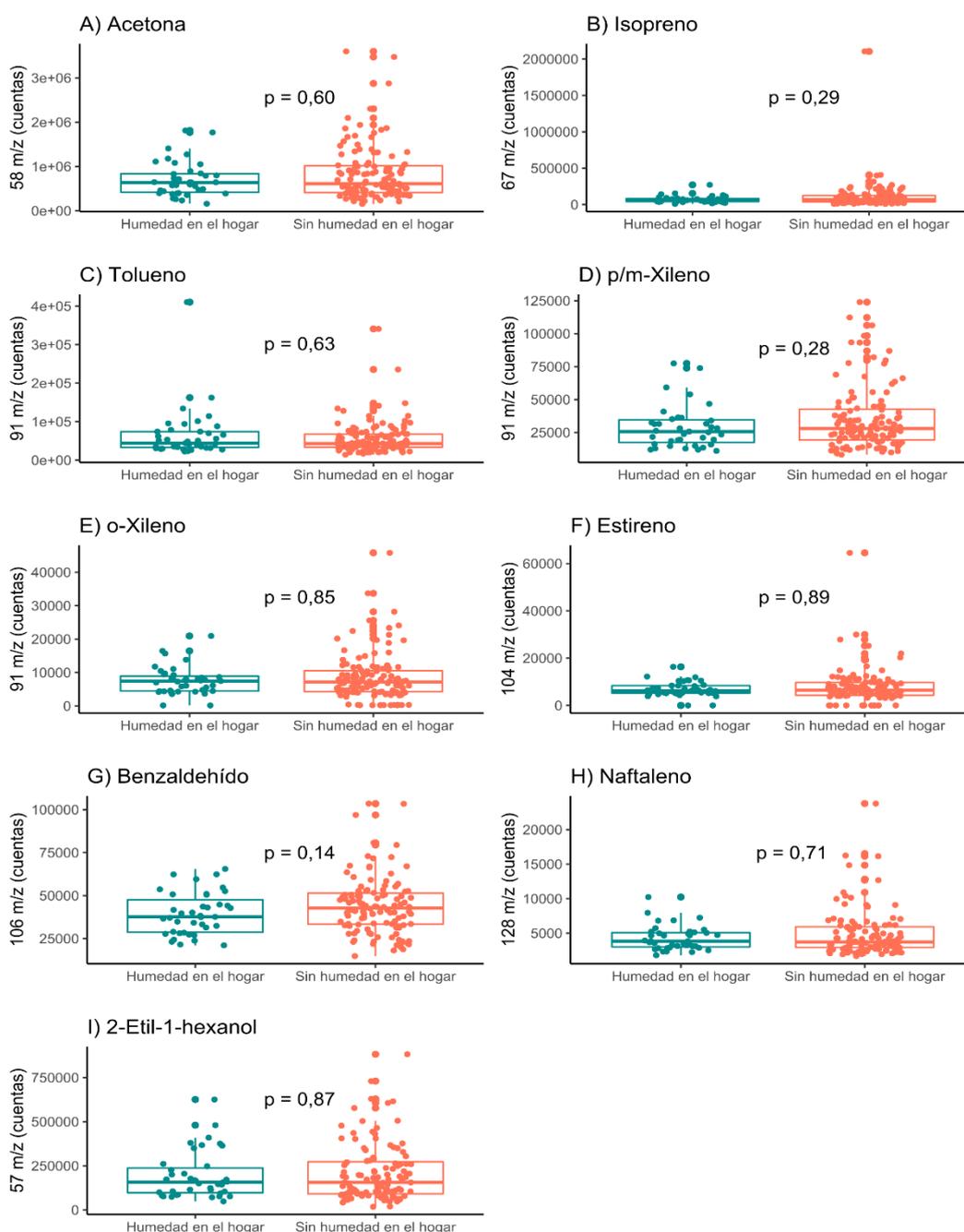


Figura 6. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire exhalado de los niños considerando los fragmentos iónicos de los compuestos seleccionados. (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto s sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire exhalado de los niños.

Sin embargo, en el caso del aire exhalado de las niñas, se observaron diferencias significativas en los niveles de acetona e isopreno (Figura 7).

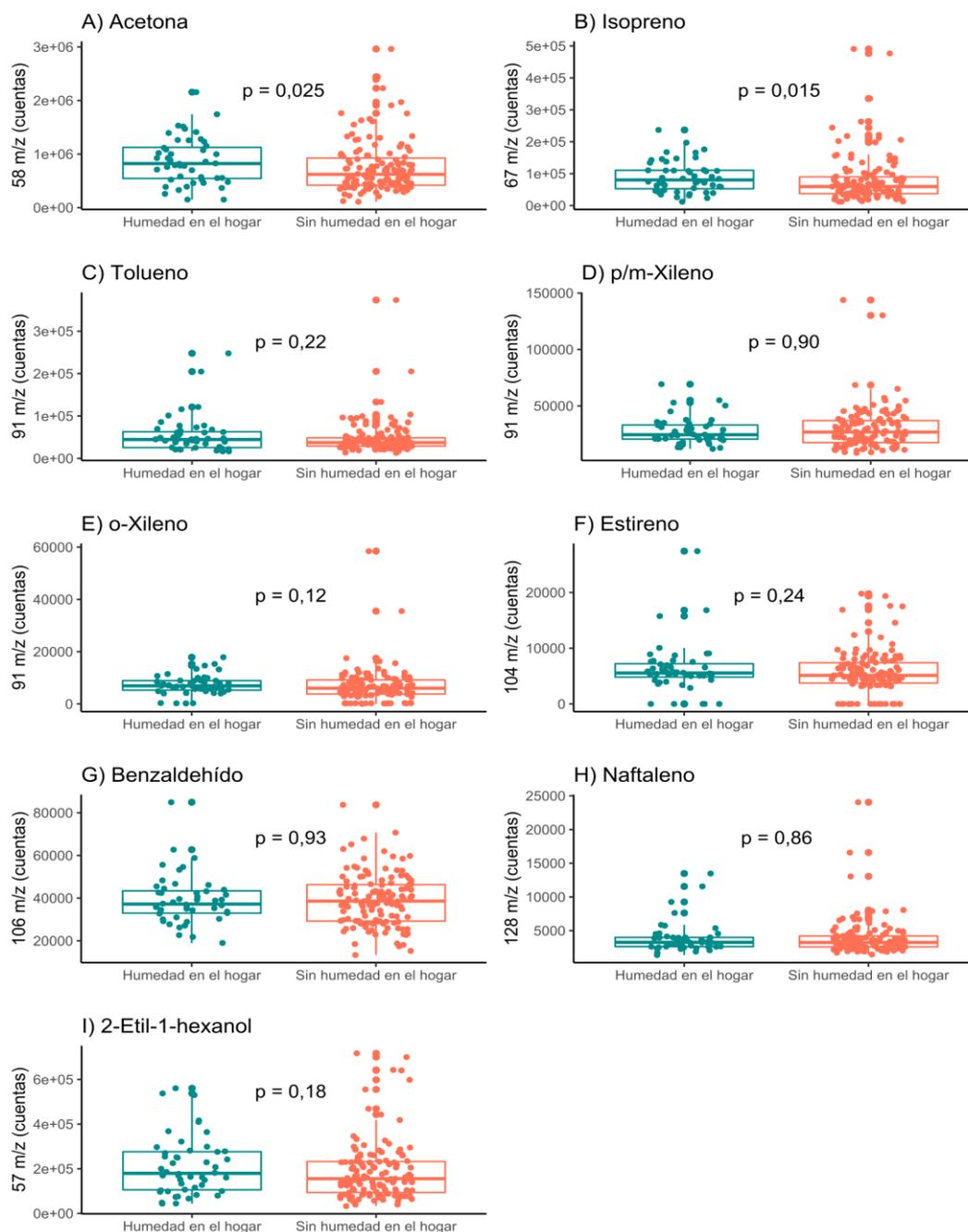


Figura 7. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire exhalado de las niñas considerando los fragmentos iónicos de los compuestos seleccionados. (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto s sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire exhalado de las niñas.

Además, los niveles de los compuestos seleccionados en el aire ambiental de la sala de recogida de muestras también fueron evaluados (Figura 8 y Figura 9).

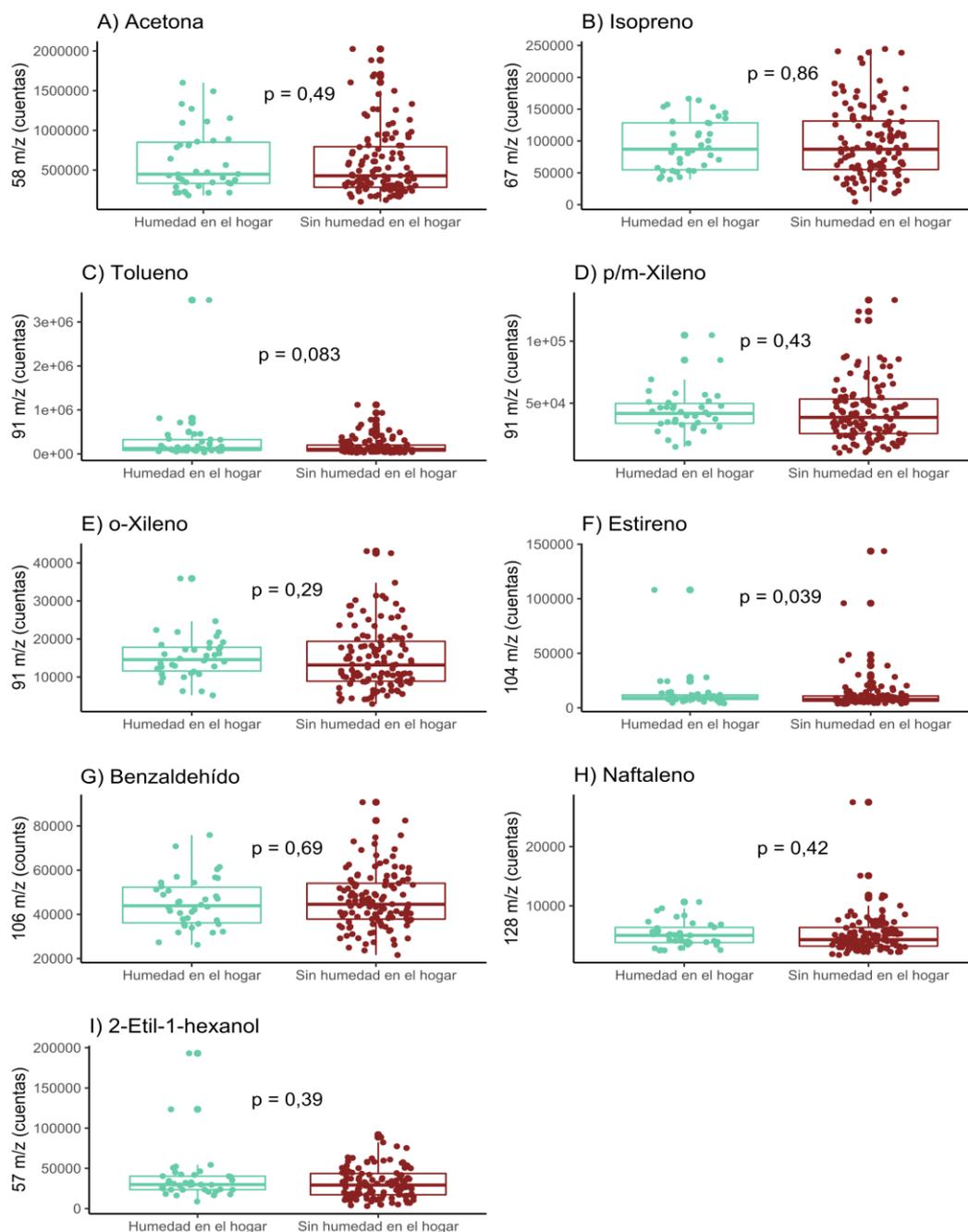


Figura 8. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire ambiental de la sala de muestreo considerando los fragmentos iónicos de los compuestos seleccionados (sólo muestras recogidas cuando los lactantes eran niños). (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire ambiental de la sala de muestreo (sólo muestras recogidas cuando los lactantes eran niños).

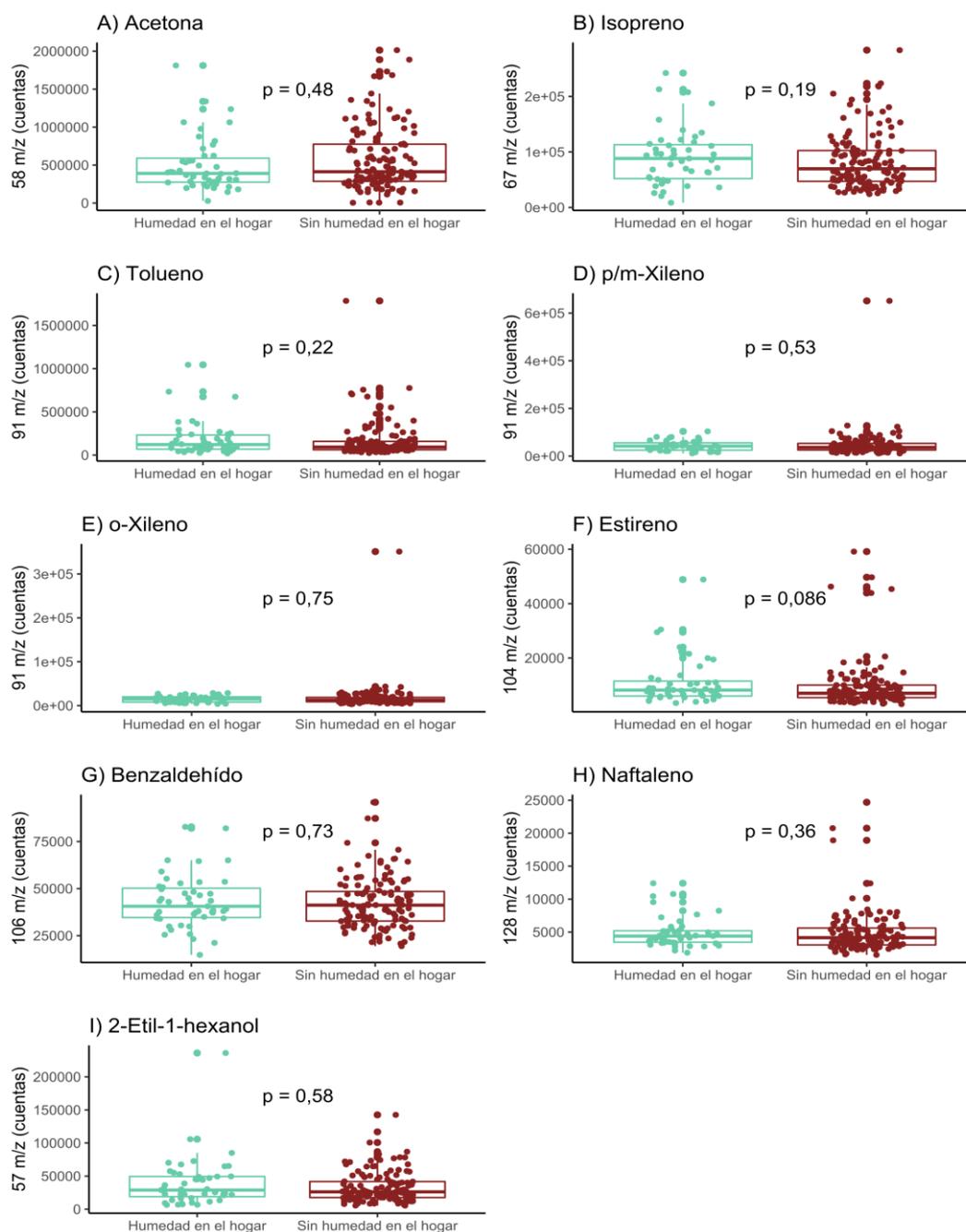


Figura 9. Evaluación de la influencia de la exposición a la humedad en interiores sobre los niveles en el aire ambiental de la sala de muestreo considerando los fragmentos iónicos de los compuestos seleccionados (sólo muestras recogidas cuando los lactantes eran niñas). (A) Impacto sobre los niveles de acetona ($m/z = 58$), (B) Impacto sobre los niveles de isopreno ($m/z = 67$), (C) Impacto sobre los niveles de tolueno ($m/z = 91$), (D) Impacto sobre los niveles de p/m-xileno ($m/z = 91$), (E) Impacto sobre los niveles de o-xileno ($m/z = 91$), (F) Impacto sobre los niveles de estireno ($m/z = 104$), (G) Impacto sobre los niveles de benzaldehído ($m/z = 106$), (H) Impacto sobre los niveles de naftaleno ($m/z = 128$) y (I) Impacto sobre los niveles de 2-etil-1-hexanol ($m/z = 57$) en el aire ambiental de la sala de muestreo (sólo muestras recogidas cuando los lactantes eran niñas).

DISCUSIÓN

Hasta donde sabemos, este es el primer estudio que pretende identificar los cambios en el aire exhalado humano debidos a la exposición a la humedad en interiores. El análisis de VOCs en el aire exhalado es una estrategia emergente para controlar el metabolismo humano^{37,38} y por esta razón, se ha propuesto como un enfoque prometedor para evaluar los efectos causados por la exposición a contaminantes ambientales¹⁹. Nuestros resultados demuestran claramente que el aire exhalado de las madres de la cohorte NELA³⁶ se vio afectado por la presencia de manchas de humedad en las casas. Así, los niveles de 2-etil-1-hexanol en el aire exhalado estaban influidos por la humedad en las residencias. A este respecto, se ha informado previamente de que este compuesto aumenta su concentración debido al aumento de la humedad relativa³⁹. El 2-etil-1-hexanol es un contaminante del aire de interior derivado del di-(2-etilhexil) ftalato (DEHP), que es el plastificante más común del policloruro de vinilo (PVC)²⁷ (Figura 10).

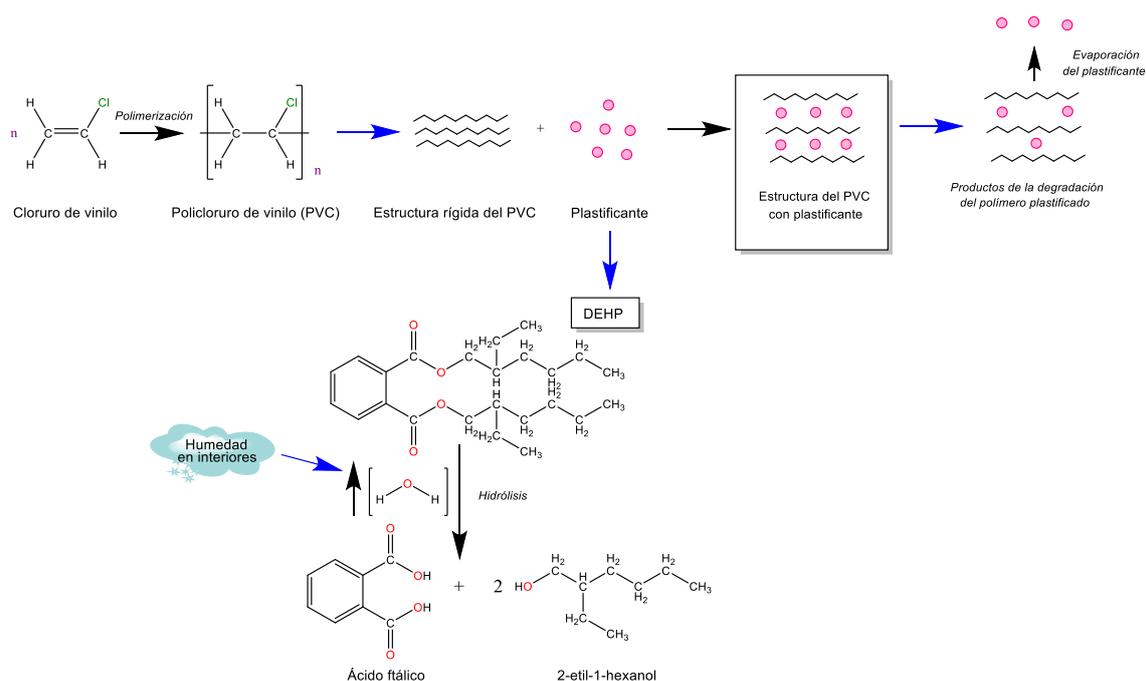


Figura 10. Producción de 2-etil-1-hexanol detectada en el aire exhalado humano en presencia de humedad en las residencias. DEHP: di-(2-etilhexil) ftalato.

Los plastificantes proporcionan nuevas propiedades al PVC, como la flexibilidad, sin embargo, la migración del plastificante es frecuente cuando sufre un deterioro, ya que éstos no están fuertemente unidos a las cadenas de polímeros⁴⁰. El 2-etil-1-hexanol es liberado tras la hidrólisis del DEHP (número de reacción KEGG: R04202), y esta reacción aumenta en presencia de humedad²⁷. Además, estudios anteriores han indicado que la concentración de 2-etil-1-hexanol en el aire de interior era mayor en

presencia de manchas de humedad^{12,33,41,42}. En este sentido, nuestros resultados muestran por primera vez que los niveles de 2-etil-1-hexanol en el aliento humano también fluctúan con la exposición a la humedad en el hogar. Es importante señalar que la exposición a este compuesto ha sido señalada como la principal causa del SBS. Además, el 2-etil-1-hexanol es un disruptor endocrino asociado al desarrollo de varias enfermedades respiratorias como el asma^{39,41-45}. De hecho, recientemente, los niveles elevados de 2-etil-1-hexanol en el aliento exhalado se han relacionado con el asma asociada a enfermedades atópicas coexistentes en las mujeres (Capítulo 5)⁴⁶. Sin embargo, en el presente estudio no se ha observado una relación significativa entre la exposición a la humedad y los niveles en el aire exhalado humano de compuestos aromáticos, que tradicionalmente también se han relacionado con el desarrollo del SBS²⁶.

Respecto a los recuentos de eosinófilos en sangre de las madres de este estudio que vivían en casas con humedad, fueron más altos que en las que no estaban expuestas a la humedad. Este resultado está en consonancia con los resultados de investigaciones anteriores^{33,34}. De hecho, el recuento de eosinófilos tanto en sangre como en el lavado nasal es una estrategia habitual para la detección de la inflamación y se ha sugerido como herramienta para el diagnóstico del asma y otras enfermedades respiratorias^{20,47-49}. Por lo tanto, un recuento elevado de eosinófilos en sangre en madres expuestas a la humedad en interiores es una prueba de que la exposición prolongada podría causar inflamación. En este sentido, aunque no se encontró una asociación significativa entre la exposición a la humedad y el asma en los adultos, sí se encontró una asociación significativa entre la rinitis alérgica y la presencia de humedad en el hogar. En esta línea, Zhang *et al.* han notificado previamente la existencia de una relación significativa entre la rinitis en adultos y la presencia de humedad en las casas⁵⁰.

A pesar de que la asociación entre la presencia de manchas de humedad en las casas y la aparición de enfermedades respiratorias o atópicas en los niños no fue significativa, los niños que estaban expuestos a ellas sufrían más síntomas respiratorios o atópicos que los demás con un riesgo relativo de 1,74. No obstante, es posible que esta relación aún no sea significativa porque los niños sólo tenían unos meses de edad. Es probable que, si se realizaran evaluaciones de riesgo a medida que los niños crezcan, en futuras fases de nuestro estudio, la asociación sería significativa. En esta línea, una revisión sistemática reciente destaca la relevancia de la exposición a la humedad en interiores en el desarrollo del asma en los niños⁵¹.

Con respecto a los niveles de compuestos en el aire exhalado, las diferencias en los niños debidas a la exposición a la humedad en la vivienda fueron menos notorias que en las madres. Sin embargo, algunos compuestos, como el 2-etil-1-hexanol, resultaron ser más elevados en los niños expuestos a la humedad que en los no expuestos. Así pues, las tendencias observadas podrían llegar a ser significativas, como ocurrió en las madres, si se analizara el aire exhalado de los niños en unos pocos meses o años, es decir, en un período de exposición más largo. Además, merece la pena mencionar el hecho de que las diferencias en la exposición a la humedad en interiores en el aire exhalado de las niñas fueron más notables que en las de los niños. Así, los niveles de acetona e isopreno (dos de los compuestos endógenos más típicos en el aliento humano⁵²) fueron significativamente elevados en las niñas expuestas a la humedad en interiores. Este estudio demuestra la necesidad de realizar también análisis estadísticos separando los datos por sexo, ya que las diferencias podrían estar enmascaradas.

Por lo tanto, en vista de la estrecha relación entre la humedad en las residencias y el desarrollo de enfermedades respiratorias/atópicas como el asma o la rinitis, es crucial aplicar estrategias para mantener los hogares libres de humedad y garantizar la mejor calidad posible del aire de interior. De hecho, algunos proyectos de rehabilitación de viviendas han obtenido resultados alentadores para la salud respiratoria^{53,54}.

Limitaciones

La principal limitación de este estudio es que las manchas de humedad en el hogar de las parejas de madres e hijos no han sido comprobadas por un especialista. El indicador cualitativo de exposición a la humedad era comunicado por las madres en los cuestionarios estructurados sobre la presencia de humedad visible en las paredes y el tiempo de exposición.

CONCLUSIÓN

Los resultados de este estudio muestran por primera vez, hasta donde sabemos, que la exposición a la humedad en el hogar puede influir en los niveles de VOCs en el aire exhalado de mujeres y bebés. Se ha encontrado una asociación significativa entre la exposición a la humedad en el interior del hogar y la rinitis alérgica en adultos. Sin embargo, es necesario realizar un seguimiento de los lactantes de la cohorte prospectiva de nacimiento del NELA para confirmar que la exposición a la humedad de interior provoca el desarrollo de enfermedades atópicas en los niños. Asimismo, es esencial

determinar los efectos que esta exposición puede tener en el metabolismo y la salud humana. Además, es imprescindible la búsqueda de las fuentes de los compuestos orgánicos volátiles con niveles elevados asociados a la humedad en las residencias para poder diseñar estrategias de remediación de los hogares.

MATERIALES Y MÉTODOS

Diseño del estudio y participantes

En este estudio participaron parejas de madres e hijos del estudio NELA (*Nutrition in Early Childhood Asthma*), una cohorte prospectiva de nacimiento. El objetivo principal del estudio NELA es desentrañar los orígenes y mecanismos del desarrollo del asma y las alergias. El protocolo del estudio, el tamaño de la muestra, los métodos de reclutamiento y los procesos de recogida de datos han sido descritos previamente³⁶. El estudio fue aprobado por el Comité Ético de Investigación Clínica (CEIC) del Hospital Universitario Virgen de la Arrixaca de Murcia (España) y se obtuvo el consentimiento informado por escrito de todos los sujetos. En el presente estudio se incluyeron las parejas madre/hijo que completaron la visita de seguimiento de tres meses entre mayo de 2017 y octubre de 2018 ⁴⁶.

Información sobre la exposición a la humedad en interiores, hallazgos de salud y otras variables

En el reclutamiento a las 20-24 semanas de gestación, entrevistadores entrenados realizaron un cuestionario estructurado a las madres del estudio. Los sujetos se dividieron en dos grupos ("humedad en el hogar" y "sin humedad en el hogar") en función de si respondían sí o no a la siguiente afirmación "presencia de manchas de humedad o moho en la pared de la vivienda habitual". Además, se obtuvo la siguiente información sobre los datos y el estado de salud de los progenitores: edad de la madre; edad del padre; índice de masa corporal (IMC) de la madre antes del embarazo; clase social de la madre (definida como la ocupación durante el embarazo en función de la clase social más alta, utilizando una adaptación española muy extendida del sistema de codificación internacional ISCO88 : I-II, directivos/técnicos; III, trabajadores cualificados; IV-V, trabajadores semicualificados/sin cualificar; y desempleados)⁵⁵; clase social del padre; nivel educativo de la madre (secundaria incompleta o menos, secundaria completa y universitaria); nivel educativo del padre; tabaquismo materno durante el embarazo (sí/no); asma en la madre (sí/no); asma en el padre (sí/no); rinitis alérgica en la madre (sí/no); rinitis alérgica en el padre (sí/no); dermatitis atópica en la madre (sí/no);

dermatitis atópica en el padre (sí/no); conjuntivitis alérgica en la madre (sí/no); conjuntivitis alérgica en el padre (sí/no). Se consideró que los sujetos tenían una enfermedad atópica si respondían en el cuestionario estructurado que habían sido diagnosticados previamente por un médico. Además, se determinó los recuentos de glóbulos blancos (WBC) de las muestras de sangre de las embarazadas mediante un analizador hematológico Sysmex® XN9000 (Sysmex Corporation, Kobe, Japón)⁴⁶. Tras el parto, también se recopiló la siguiente información sobre los niños: sexo, peso y altura al nacer. Durante la visita de seguimiento de los niños a los tres meses, se realizó otro cuestionario estructurado a las madres, que incluía preguntas relacionadas con la salud atópica y respiratoria de los niños: eczema atópico en los niños (sí/no); infecciones respiratorias en los niños (sí/no). Finalmente, el tiempo mínimo de exposición a la humedad en el hogar se definió como el período comprendido entre la fecha de la visita de seguimiento a las 20-24 semanas de embarazo y la fecha de la visita de seguimiento a los tres meses.

Recogida y análisis del aire exhalado

Los protocolos seguidos para la recogida y el análisis del aire exhalado se han descrito anteriormente en el Capítulo 4⁵⁶. El aire exhalado mixto de los lactantes de 3 meses y sus madres fue recogido en bolsas de muestreo de gas Tedlar® de 1L. Sin embargo, se incluyó un paso previo en la toma de muestras de aliento de los lactantes, ya que son sujetos pasivos y las bolsas Tedlar® mostraron demasiada resistencia para ser llenadas por ellos. Por lo tanto, el aire exhalado de los bebés se recogió previamente mediante una mascarilla conectada a las bolsas de muestreo de gas Quintron® de 400 mL. A continuación, el contenido de aire de las bolsas Quintron® se transfirió a una bolsa Tedlar®. Además, para comprobar la influencia del aire de la habitación en la que se recogieron las muestras de aliento, se bombeó el aire ambiental de forma activa a un tubo de desorción térmica mediante una jeringa Easy-VOC (Markes International). En este estudio, el análisis del aire exhalado se llevó a cabo mediante un sistema de desorción térmica acoplado a la cromatografía de gases - espectrometría de masas de cuadrupolo simple (TD-GC/q-MS). Para ello, tanto las muestras de aire exhalado como las de contenido de aire ambiental se almacenaron en tubos de desorción térmica (Tenax TA/carbograph 5td, Markes International) antes del análisis. Además, se analizaron dos estándares químicos (estándar de n-alcenos saturados C7-C30 y estándar de calibración de VOCs, Sigma-Aldrich) para calcular los índices de retención de los VOCs.

Preprocesamiento de datos

Se implementó el flujo de trabajo de código abierto descrito en el Capítulo 4 para el preprocesamiento de datos⁵⁶. En primer lugar, fue necesario transformar los datos brutos en un formato accesible, como el formato mzXML, mediante *MsConvert* de *Proteowizard*^{57,58}. Este flujo de trabajo integra las funciones de tres paquetes de R (*xcms*⁵⁹, *cliqueMS*⁶⁰ y *eRah*⁶¹) y aún a las dos estrategias más importantes para el preprocesamiento de datos brutos de GC-MS (determinación de fragmentos iónicos o *features* y determinación de compuestos). Mediante este flujo de trabajo, se generó una matriz con las intensidades de los *features* (picos iónicos con un tiempo de retención y una señal masa/carga (m/z) característicos) detectados en las muestras de aire exhalado y se determinaron e identificaron los compuestos a los que correspondían. La identificación de los compuestos se llevó a cabo mediante el cotejo de los espectros de masas y los índices de retención con la biblioteca espectral NIST (*National Institute of Standard and Technology*), a través del cálculo del factor de coincidencia o *match factor* y del valor de *error RI* (error del índice de retención). Para el presente estudio, se seleccionaron nueve compuestos y se evaluaron sus niveles a partir de sus *features* más característicos: acetona ($m/z = 58$), isopreno ($m/z = 67$), tolueno ($m/z = 91$), p/m-xileno ($m/z = 91$), o-xileno ($m/z = 91$), estireno ($m/z = 104$), benzaldehído ($m/z = 106$), naftaleno ($m/z = 128$) y 2-etil-1-hexanol ($m/z = 57$) (Apéndice 3).

Análisis estadístico

El análisis estadístico se realizó con la versión 4.0.5 de R. Se realizó la prueba de chi-cuadrado o la prueba exacta de Fisher para comprobar la relación entre la exposición a la humedad en el interior de los hogares y las variables categóricas. Las asociaciones entre la presencia de humedad en la residencia habitual de las parejas de madres e hijos con las variables continuas, los niveles de los *features* seleccionados del aire exhalado de las madres y los niños o los resultados del análisis del recuento de glóbulos blancos en sangre (WBC) de la madre durante el embarazo, se evaluaron mediante las pruebas t de Student o U de Mann-Whitney y la distribución normal se evaluó mediante pruebas de Lilliefors utilizando el paquete *nortest*. Las diferencias entre ambos grupos ("humedad en el hogar" y "sin humedad en el hogar") se consideraron significativas cuando $p\text{-valor} < 0,05$. Además, también se comprobaron los niveles de los *features* seleccionados en ambos grupos en el aire ambiental de la sala de recogida para diferenciar entre las fluctuaciones del aire exhalado humano influidas por la exposición a la humedad del hogar y las debidas a la contaminación de las muestras de aire exhalado por el aire de la sala de muestreo. También se construyeron modelos de

regresión logística para estimar los valores de *odd ratio* (OR) entre los trastornos atópicos de los padres y los niños con la humedad en las paredes de las casas.

BIBLIOGRAFÍA

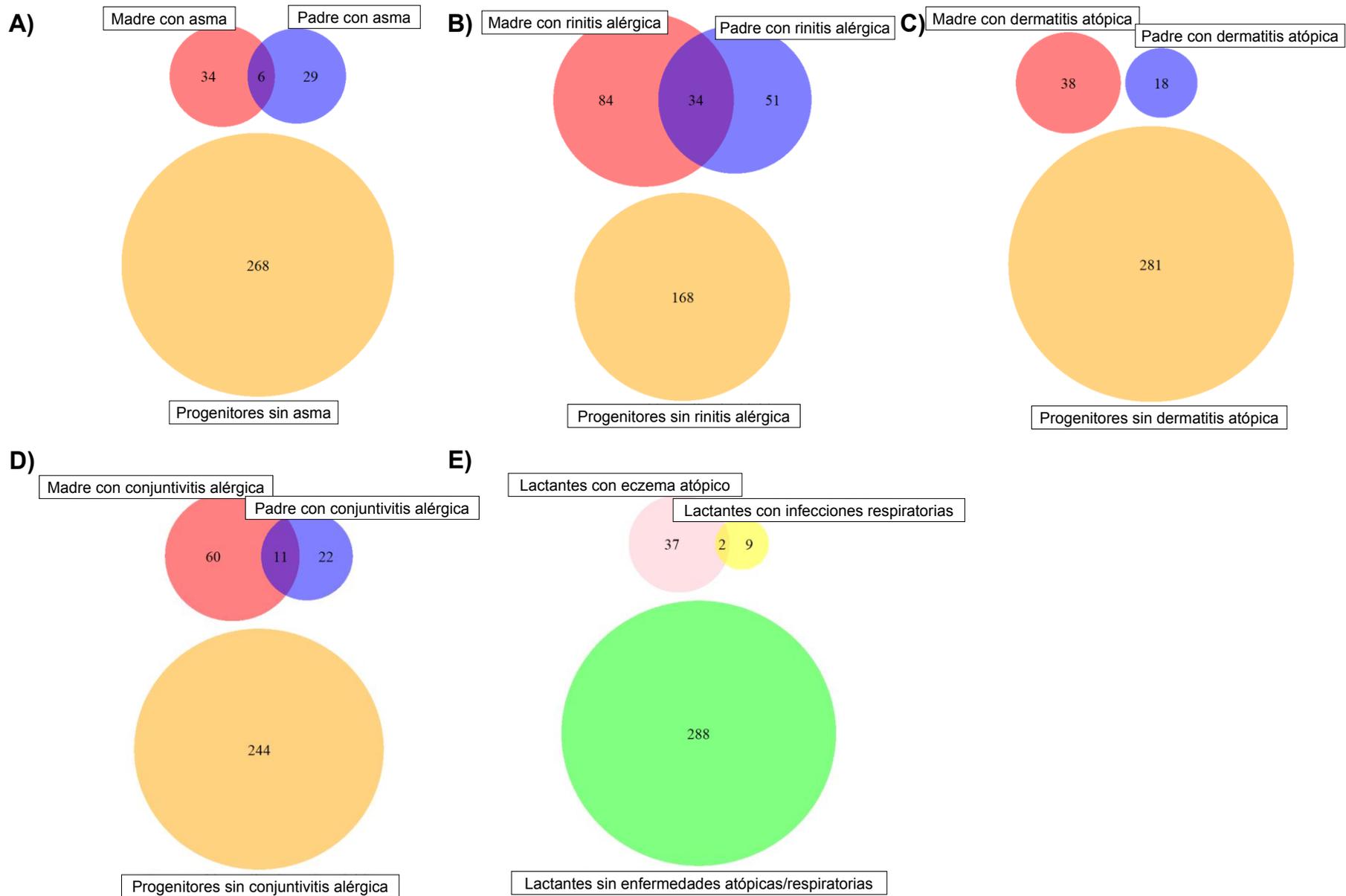
1. Beckhaus, A. A. *et al.* Maternal nutrition during pregnancy and risk of asthma, wheeze, and atopic diseases during childhood: A systematic review and meta-analysis. *Allergy* **70**, 1588–1604 (2015).
2. Kansen, H. M. *et al.* Prevalence and predictors of uncontrolled asthma in children referred for asthma and other atopic diseases. *J. Asthma Allergy* **13**, 67–75 (2020).
3. Amat, F. *et al.* New insights into the phenotypes of atopic dermatitis linked with allergies and asthma in children: An overview. *Clin. Exp. Allergy* **48**, 919–934 (2018).
4. van de Kant, K. D. G. *et al.* Exhaled breath profiling in diagnosing wheezy preschool children. *Eur. Respir. J.* **41**, 183–188 (2013).
5. Adgate, J. L. *et al.* Outdoor, indoor, and personal exposure to VOCs in children. *Environ. Health Perspect.* **112**, 1386–1392 (2004).
6. Sharpe, R. A., Bearman, N., Thornton, C. R., Husk, K. & Osborne, N. J. Indoor fungal diversity and asthma: A meta-analysis and systematic review of risk factors. *J. Allergy Clin. Immunol.* **135**, 110–122 (2015).
7. Testa, D. *et al.* Allergic rhinitis and asthma assessment of risk factors in pediatric patients: A systematic review. *Int. J. Pediatr. Otorhinolaryngol.* **129**, 109759 (2020).
8. Mendell, M. J. & Kumagai, K. Observation-based metrics for residential dampness and mold with dose–response relationships to health: A review. *Indoor Air* **27**, 506–517 (2017).
9. Madureira, J., Paciência, I., Pereira, C., Teixeira, J. P. & Fernandes, E. de O. Indoor air quality in Portuguese schools: levels and sources of pollutants. *Indoor Air* **26**, 526–537 (2016).
10. Billionnet, C., Gay, E., Kirchner, S., Leynaert, B. & Annesi-Maesano, I. Quantitative assessments of indoor air pollution and respiratory health in a population-based sample of French dwellings. *Environ. Res.* **111**, 425–434 (2011).
11. Kwon, J.-W., Park, H.-W., Kim, W. J., Kim, M.-G. & Lee, S.-J. Exposure to volatile organic compounds and airway inflammation. *Environ. Heal.* **17**, 65 (2018).
12. Nielsen, G. D. *et al.* Do indoor chemicals promote development of airway allergy?

- Indoor Air* **17**, 236–255 (2007).
13. Son, B., Breyse, P. & Yang, W. Volatile organic compounds concentrations in residential indoor and outdoor and its personal exposure in Korea. *Environ. Int.* **29**, 79–85 (2003).
 14. Arif, A. A. & Shah, S. M. Association between personal exposure to volatile organic compounds and asthma among US adult population. *Int. Arch. Occup. Environ. Health* **80**, 711–719 (2007).
 15. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
 16. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 014001 (2014).
 17. Caron-Beaudoin, É. *et al.* Gestational exposure to volatile organic compounds (VOCs) in Northeastern British Columbia, Canada: A pilot study. *Environ. Int.* **110**, 131–138 (2018).
 18. Storer, M., Curry, K., Squire, M., Kingham, S. & Epton, M. Breath testing and personal exposure--SIFT-MS detection of breath acetonitrile for exposure monitoring. *J. Breath Res.* **9**, 036006 (2015).
 19. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
 20. van de Kant, K. D. G., van der Sande, L. J. T. M., Jöbsis, Q., van Schayck, O. C. P. & Dompeling, E. Clinical use of exhaled volatile organic compounds in pulmonary diseases: a systematic review. *Respir. Res.* **13**, 117 (2012).
 21. Jalali, M. *et al.* Oxidative Stress Biomarkers in Exhaled Breath of Workers Exposed to Crystalline Silica Dust by SPME-GC-MS. *J. Res. Health Sci.* **16**, 153–161 (2016).
 22. Norbäck, D. *et al.* Onset and remission of childhood wheeze and rhinitis across China — Associations with early life indoor and outdoor air pollution. *Environ. Int.* **123**, 61–69 (2019).
 23. Deng, Q., Lu, C., Ou, C., Chen, L. & Yuan, H. Preconceptional, prenatal and postnatal exposure to outdoor and indoor environmental factors on allergic diseases/symptoms in preschool children. *Chemosphere* **152**, 459–467 (2016).
 24. Lu, C. *et al.* Furry pet-related wheeze and rhinitis in pre-school children across

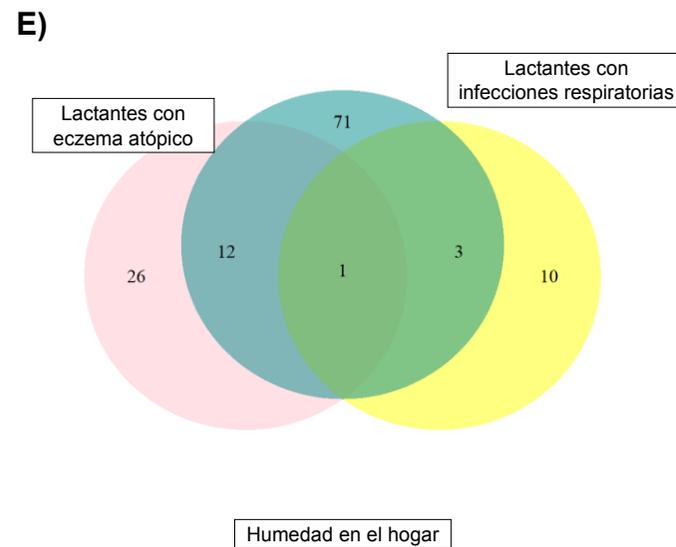
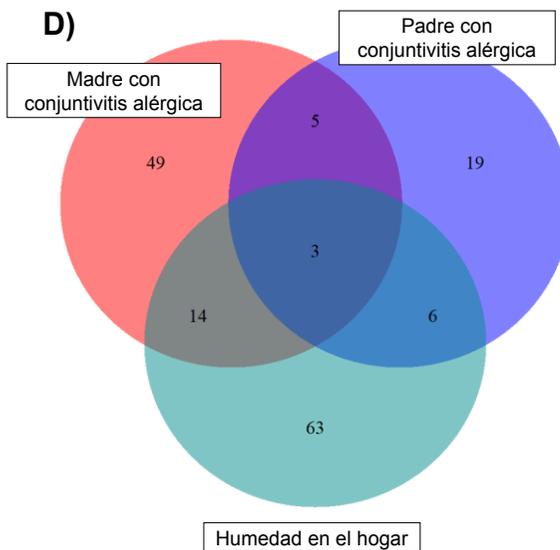
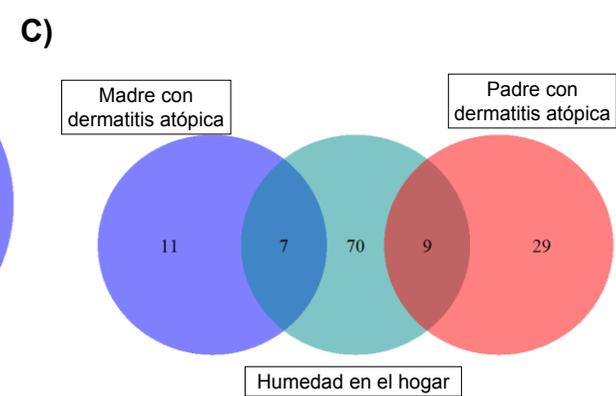
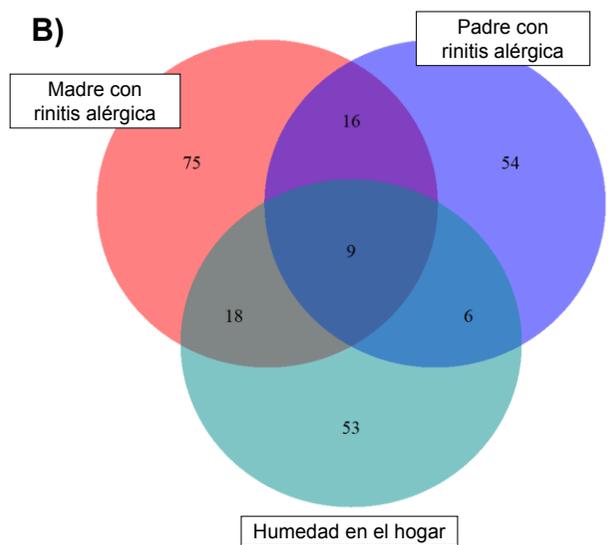
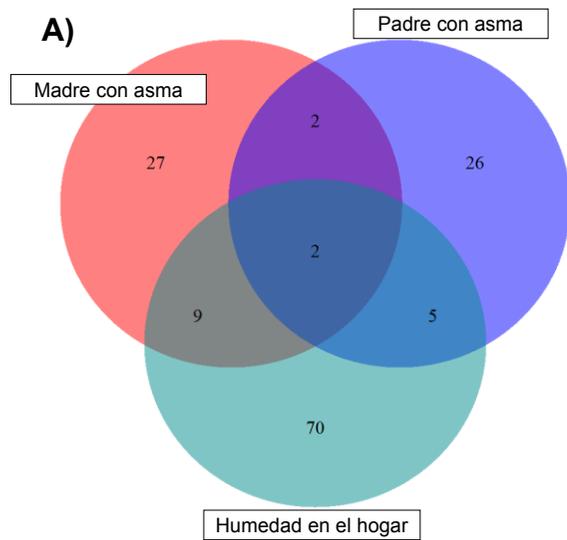
- China: Associations with early life dampness and mould, furry pet keeping, outdoor temperature, PM10 and PM2.5. *Environ. Int.* **144**, 106033 (2020).
25. Valtonen, V. Clinical diagnosis of the dampness and mold hypersensitivity syndrome: Review of the literature and suggested diagnostic criteria. *Front. Immunol.* **8**, 951 (2017).
 26. Saijo, Y. *et al.* Symptoms in relation to chemicals and dampness in newly built dwellings. *Int. Arch. Occup. Environ. Health* **77**, 461–470 (2004).
 27. Wakayama, T. *et al.* Comprehensive review of 2-ethyl-1-hexanol as an indoor air pollutant. *Journal of Occupational Health* **61**, 19–35 (2019).
 28. Fisk, W. J., Lei-Gomez, Q. & Mendell, M. J. Meta-analyses of the associations of respiratory health effects with dampness and mold in homes. *Indoor Air* **17**, 284–296 (2007).
 29. Hope, A. P. & Simon, R. A. Excess dampness and mold growth in homes: An evidence-based review of the aeroirritant effect and its potential causes. *Allergy Asthma Proc.* **28**, 262–270 (2007).
 30. Quansah, R., Jaakkola, M. S., Hugg, T. T., Heikkinen, S. A. M. & Jaakkola, J. J. K. Residential Dampness and Molds and the Risk of Developing Asthma: A Systematic Review and Meta-Analysis. *PLoS One* **7**, 47526 (2012).
 31. Norbäck, D. *et al.* Respiratory symptoms and fractional exhaled nitric oxide (FeNO) among students in Penang, Malaysia in relation to signs of dampness at school and fungal DNA in school dust. *Sci. Total Environ.* **577**, 148–154 (2017).
 32. Rufo, J. C. *et al.* Indoor air quality and atopic sensitization in primary schools: A follow-up study. *Porto Biomed. J.* **1**, 142–146 (2016).
 33. Wålinder, R., Wieslander, G., Norbäck, D., Wessen, B. & Venge, P. Nasal lavage biomarkers: Effects of water damage and microbial growth in an office building. *Arch. Environ. Health* **56**, 30–36 (2001).
 34. Zhang, X. *et al.* Dampness and moulds in workplace buildings: Associations with incidence and remission of sick building syndrome (SBS) and biomarkers of inflammation in a 10year follow-up study. *Sci. Total Environ.* **430**, 75–81 (2012).
 35. Vardoulakis, S. *et al.* Indoor Exposure to Selected Air Pollutants in the Home Environment: A Systematic Review. *Int. J. Environ. Res. Public Heal.* **17**, 8972 (2020).
 36. Morales, E. *et al.* The Nutrition in Early Life and Asthma (NELA) birth cohort study: rationale, design, and methods. *Paediatr. Perinat. Epidemiol.* **36**, 310–324 (2022).

37. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res.* **6**, 027108 (2012).
38. Sola Martínez, R. A. *et al.* Exhaled volatile organic compounds analysis in clinical pediatrics: a systematic review. *Pediatr. Res.* **89**, 1352–1363 (2021).
39. Markowicz, P. & Larsson, L. Influence of relative humidity on VOC concentrations in indoor air. *Environ. Sci. Pollut. Res.* **22**, 5772–5779 (2015).
40. Wei, X.-F., Linde, E. & Hedenqvist, M. S. Plasticiser loss from plastic or rubber products through diffusion and evaporation. *npj Mater. Degrad.* **3**, 1–8 (2019).
41. Wieslander, G., Norbäck, D., Nordström, K., Wålander, R. & Venge, P. Nasal and ocular symptoms, tear film stability and biomarkers in nasal lavage, in relation to building-dampness and building design in hospitals. *Int. Arch. Occup. Environ. Health* **72**, 451–461 (1999).
42. Wieslander, G., Kumlin, A. & Norbäck, D. Dampness and 2-ethyl-1-hexanol in floor construction of rehabilitation center: Health effects in staff. *Arch. Environ. Occup. Heal.* **65**, 3–11 (2010).
43. Paciência, I. *et al.* Exposure to indoor endocrine-disrupting chemicals and childhood asthma and obesity. *Allergy* **74**, 1277–1291 (2019).
44. Sur, U., Erkekoglu, P., Bulus, A. D., Andiran, N. & Kocer-Gumusel, B. Oxidative stress markers, trace elements, and endocrine disrupting chemicals in children with Hashimoto's thyroiditis. *Toxicol. Mech. Methods* **29**, 633–643 (2019).
45. Norback, D., Wieslander, G., Nordstrom, K. & Walinder, R. Asthma symptoms in relation to measured building dampness in upper concrete floor construction, and 2-ethyl-1-hexanol in indoor air. *Int. J. Tuberc. Lung Dis.* **4**, 1016–1025 (2000).
46. Sola-Martínez, R. A. *et al.* Exhaled volatilome analysis as a useful tool to discriminate asthma with other coexisting atopic diseases in women of childbearing age. *Sci. Rep.* **11**, 13823 (2021).
47. Gibson, P. G. Variability of blood eosinophils as a biomarker in asthma and COPD. *Respirology* **23**, 12–13 (2018).
48. Pignatti, P. *et al.* Do blood eosinophils strictly reflect airway inflammation in COPD? Comparison with asthmatic patients. *Respir. Res.* **20**, 145 (2019).
49. Soma, T. *et al.* Implication of fraction of exhaled nitric oxide and blood eosinophil count in severe asthma. *Allergol. Int.* **67**, S3–S11 (2018).
50. Zhang, X. *et al.* Dampness and mold in homes across China: Associations with

- rhinitis, ocular, throat and dermal symptoms, headache and fatigue among adults. *Indoor Air* **29**, 30–42 (2019).
51. Caillaud, D. *et al.* Indoor mould exposure, asthma and rhinitis: Findings from systematic reviews and recent longitudinal studies. *Eur. Respir. Rev.* **27**, 170137 (2018).
 52. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
 53. Kercksmar, C. M. *et al.* Reduction in asthma morbidity in children as a result of home remediation aimed at moisture sources. *Environ. Health Perspect.* **114**, 1574–1580 (2006).
 54. Barnes, C. S. *et al.* Home Assessment and Remediation. *J. Allergy Clin. Immunol. Pract.* **4**, 423-431.e15 (2016).
 55. Domingo-Salvany, A., Regidor, E., Alonso, J. & Alvarez-Dardet, C. Proposal for a social class measure. Working Group of the Spanish Society of Epidemiology and the Spanish Society of Family and Community Medicine. *Aten. Primaria* **25**, 350–363 (2000).
 56. Sola Martínez, R. A. *et al.* Data preprocessing workflow for exhaled breath analysis by GC/MS using open sources. *Sci. Rep.* **10**, 22008 (2020).
 57. Martens, L. *et al.* mzML—a community standard for mass spectrometry data. *Mol. Cell. Proteomics* **10**, R110.000133 (2011).
 58. Adusumilli, R. & Mallick, P. Data conversion with proteoWizard msConvert. in *Proteomics: Methods and Protocols* (eds. Comai, U., Katz, J. E. & Mallick, P.) **1550**, 339–368 (Humana, New York, NY, 2017).
 59. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **78**, 779–787 (2006).
 60. Senan, O. *et al.* CliqueMS: A computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089–4097 (2019).
 61. Domingo-Almenara, X. *et al.* eRah: A Computational Tool Integrating Spectral Deconvolution and Alignment with Quantification and Identification of Metabolites in GC/MS-Based Metabolomics. *Anal. Chem.* **88**, 9821–9829 (2016).



APÉNDICE 1. Distribución de los trastornos atópicos en la población del estudio. (A) Asma en la población de estudio. (B) Rinitis alérgica en la población de estudio. (C) Dermatitis atópica en la población de estudio. (D) Conjuntivitis alérgica en la población de estudio. (E) Enfermedades atópicas/respiratorias en los niños incluidos en este estudio.



APÉNDICE 2. Distribución de la exposición a la humedad en interiores y de los trastornos atópicos en la población del estudio. (A) Asma y exposición a la humedad en interiores en la población del estudio. (B) Rinitis alérgica y exposición a la humedad en interiores en la población del estudio. (C) Dermatitis atópica y exposición a la humedad en interiores en la población del estudio. (D) Conjuntivitis alérgica y exposición a la humedad en interiores en la población del estudio. (E) Enfermedades atópicas/respiratorias y exposición a la humedad en interiores en los niños incluidos en este estudio.

APÉNDICE 3. Compuestos orgánicos volátiles (VOCs) estudiados en el presente estudio.

VOCs	m/z	Código CAS	Match factor y error RI	Solamente Match factor
Acetona	58	67-64-1	X	
Isopreno	67	78-79-5		X
Tolueno	91	108-88-3	X	
p/m-Xileno	91	106-42-3/ 108-38-3	X	
o-Xileno	91	95-47-6	X	
Estireno	104	100-42-5	X	
Benzaldehído	106	100-52-7	X	
Naftaleno	128	91-20-3	X	
2-Etil-1-hexanol	57	104-76-7	X	

Capítulo 8

Análisis del aire exhalado mediante off-line SESI- HRMS: conservación de muestras de aliento

Los resultados presentados en este capítulo están basados en la publicación:

Rosa Alba Sola Martínez, Fabienne Decrue, Jiafa Zeng, Mo Awchi, Amanda
Gisler, Melinda Richard, Kim Arnold, Kapil Dev Singh, Urs Frey, Manuel
Cánovas Díaz, Teresa de Diego Puente y Pablo Sinues

Breath analysis by off-line SESI-HRMS: conservation of breath samples

(en preparación).

RESUMEN

La ionización secundaria por *electrospray* acoplada a espectrometría de masas de alta resolución (SESI-HRMS) ha generado en los últimos años una gran expectación en el campo del análisis del aire exhalado. De hecho, es considerada una de las plataformas analíticas con mayor potencial para lograr la introducción del análisis del aire exhalado en la práctica clínica diaria, debido a sus cortos tiempos de análisis sin renunciar a niveles elevados de sensibilidad y selectividad en las determinaciones. Tradicionalmente, SESI-HRMS sólo se había utilizado para el análisis del aire exhalado *on-line* a tiempo real, lo que implicaba que los sujetos tenían que desplazarse a la localización de la plataforma analítica. Sin embargo, recientemente, el grupo del Prof. Sinues de la Universidad de Basilea, ha desarrollado un método para realizar un análisis del aire exhalado *off-line* empleando SESI-HRMS, recolectando el aliento en bolsas Nalophan® antes de su análisis químico. En el presente estudio, utilizando dicho método se ha llevado a cabo una evaluación exhaustiva del impacto del tiempo de almacenamiento del aire exhalado en las bolsas Nalophan® a distintas temperaturas (temperatura ambiente y hielo seco). Los resultados obtenidos indican que las muestras de aire exhalado pueden conservarse, sin pérdidas decisivas para su análisis posterior, tras un almacenamiento de 2 horas a temperatura ambiente, ya que la media de intensidad relativa de los compuestos se mantuvo superior al 80%, lo que permitirá mayor flexibilidad en la logística de la técnica para su implementación clínica.

INTRODUCCIÓN

En los últimos años, la comunidad médica y científica ha centrado sus esfuerzos en el desarrollo de la medicina personalizada, haciendo hincapié en la búsqueda de métodos de diagnóstico y monitorización terapéutica adaptados a las necesidades de cada paciente para hacer frente a enfermedades heterogéneas. En este contexto, se ha planteado que el aire exhalado humano podría ser una fuente potencial de biomarcadores¹⁻³. No obstante, en el aire exhalado es posible distinguir una gran variedad de compuestos orgánicos volátiles (VOCs) cuya concentración puede cambiar en base al estado de salud y al metabolismo del individuo⁴. En este sentido, a lo largo de los últimos años, se han llevado a cabo multitud de estudios encaminados a la búsqueda en el aire exhalado de biomarcadores de diferentes patologías como el asma^{5,6}, cáncer de pulmón⁷, enfermedad pulmonar obstructiva crónica⁸, diabetes mellitus^{9,10}, fibrosis quística^{11,12}, enfermedades infecciosas¹³, etc. En gran medida, el interés de la implementación en la clínica del análisis del aire exhalado radica en que la disposición del aire exhalado es ilimitada y que se puede obtener de forma no invasiva¹⁴.

El muestreo y la medición del aire exhalado puede efectuarse mediante dos estrategias claramente diferenciadas: análisis del aire exhalado en línea u *on-line* a tiempo real, donde el aire exhalado es analizado directamente, y análisis del aire exhalado fuera de línea u *off-line*, donde se utiliza un contenedor de aire exhalado (ej. bolsas de muestreo de gas) y las muestras pueden ser almacenadas antes del análisis^{4,15,16}. Hoy en día, las tecnologías basadas en espectrometría de masas son las plataformas analíticas de aire exhalado más empleadas en la detección de VOCs, debido a su sensibilidad y selectividad^{1,17}. Así pues, la técnica de referencia del análisis del aire exhalado *off-line* es la cromatografía de gases acoplada a la espectrometría de masas (GC-MS) que requiere un paso previo de preconcentración de las muestras de aliento, empleando tubos de desorción térmica (tubos TD) o microextracción en fase sólida (SPME), entre otros métodos¹⁸. Por otro lado, SIFT-MS (tubo de selección de iones con flujo - espectrometría de masas) y PTR-MS (reacción de transferencia de protones – espectrometría de masas) son metodologías ampliamente utilizadas en el análisis del aire exhalado *on-line*, ya que permiten el análisis directo del aliento sin realizar ningún tratamiento anterior. Sin embargo, la sensibilidad y la selectividad en los resultados analíticos puede verse condicionada, precisamente por no disponer de etapas previas como la preconcentración de las muestras o la separación por cromatografía de gases. Otra técnica con características intrínsecas para el análisis del aire exhalado *on-line* es SESI-MS (ionización secundaria por *electrospray* – espectrometría de masas). Se trata de una opción novedosa, que puede acoplarse a todo tipo de analizadores de masas, inclusive los de alta resolución^{15,19}. Este hecho, junto con los progresos en cuanto a la estandarización de la toma de muestra y del análisis, la convierten en una de las alternativas más prometedoras para conseguir la tan esperada introducción del análisis del aire exhalado en la clínica^{14,15,19,20}. En este sentido, en el marco del proyecto DOPAEx (*Determination of Optimal Procedures for Analysis of Expired Breath by Secondary Electrospray Ionization-Mass Spectrometry*), el grupo de investigación liderado por el Prof. Sinues (Universidad de Basilea, Suiza) ha elaborado una lista de 846 señales masa/carga (m/z), 411 detectadas en modo positivo y 435 en modo negativo, que es posible determinar en el aire exhalado humano (datos no publicados) mediante SESI-HRMS (ionización secundaria por *electrospray* – espectrometría de masas de alta resolución).

El análisis de aire exhalado *on-line* a tiempo real mediante espectrometría de masas presenta una serie de beneficios como tiempos de análisis más cortos, rapidez en la obtención de resultados y poca manipulación de las muestras, lo que reduce la posibilidad de alteración de las muestras durante el muestreo y el almacenamiento de

las mismas^{15,21}. Sin embargo, también tiene inconvenientes como que los sujetos tienen que desplazarse al lugar donde se encuentran las plataformas analíticas. Lamentablemente, muchos pacientes no pueden ser transportados hasta allí (ej. pacientes de cuidados intensivos, neonatos, lactantes, etc.). Por ello, en el caso de pacientes pasivos no cooperativos es fundamental disponer de un contenedor para recoger el aire exhalado y posteriormente transportarlo al laboratorio de análisis²². A este respecto, aunque las técnicas PTR-MS y SIFT-MS han sido concebidas originalmente para el análisis del aire exhalado *on-line* a tiempo real, también se han utilizado satisfactoriamente en varios estudios para analizar el aire exhalado almacenado en bolsas de muestreo de gases de diferentes tipos^{14,15}. Asimismo, el grupo del Prof. Sinues, en su reciente publicación de 2021, Decrue *et al.*²², presentaron el primer protocolo para el análisis del aire exhalado *off-line* mediante SESI-HRMS y fue probado con éxito tanto en adultos como en neonatos. Para ello realizaron una comparación de los resultados obtenidos con el método convencional (análisis del aire exhalado *on-line* a tiempo real) y la medición del aliento por el nuevo método *off-line*, donde las muestras eran recogidas en un dispositivo *off-line* con bolsas Nalophan® y, posteriormente, eran analizadas antes de los 10 minutos tras su recolección. Para realizar la comparación entre ambos enfoques aplicaron el cálculo del Coeficiente de Correlación de Concordancia de Lin (Lin's CCC)²³, identificando 1249 señales *m/z* con un Lin's CCC >0,6. Una vez comprobada la aplicabilidad del método y con el objetivo de optimizar la gestión y manipulación del almacenamiento de las muestras y su efecto en la reproducibilidad y repetibilidad del análisis, en el presente capítulo se ha llevado a cabo un estudio para evaluar la conservación de las muestras de aire exhalado en bolsas Nalophan® a lo largo del tiempo, tanto en hielo seco como a temperatura ambiente, para su posterior análisis con SESI-HRMS. Además, se ha prestado especial interés al efecto del almacenamiento en 27 aldehídos de tres familias diferentes (2-alquénales, 4-hidroxi-2-alquénales y 4-hidroxi-2,6-alcadiénales) relacionados con la peroxidación lipídica y de los que se dispone de una identificación robusta en aire exhalado condensado y aire exhalado mediante SESI-HRMS²⁴.

RESULTADOS

Efecto del tiempo y la temperatura en la conservación del aire exhalado

En la Figura 1 se muestra el diagrama de flujo de trabajo del presente estudio. Los sujetos incluidos en el estudio realizaron un total de 44 visitas en las que se recogieron 7 bolsas Nalophan® de aire exhalado en cada consulta. Una de esas bolsas fue analizada inmediatamente por SESI-HRMS después de ser llenada, identificándola

como bolsa a tiempo cero, tres bolsas se almacenaron a temperatura ambiente y las otras tres se guardaron en hielo seco y se analizaron respectivamente a los 10 min, 60 min y 120 min de la recogida del aire exhalado. Una de las muestras de aliento conservada en hielo seco durante 1 hora tuvo que ser descartada por problemas técnicos. Por ello, tras el preprocesamiento de los datos brutos se obtuvo una matriz de datos con los valores de intensidad de 2681 señales m/z o *features* detectados por muestra para las 307 muestras de aire exhalado recogidas.

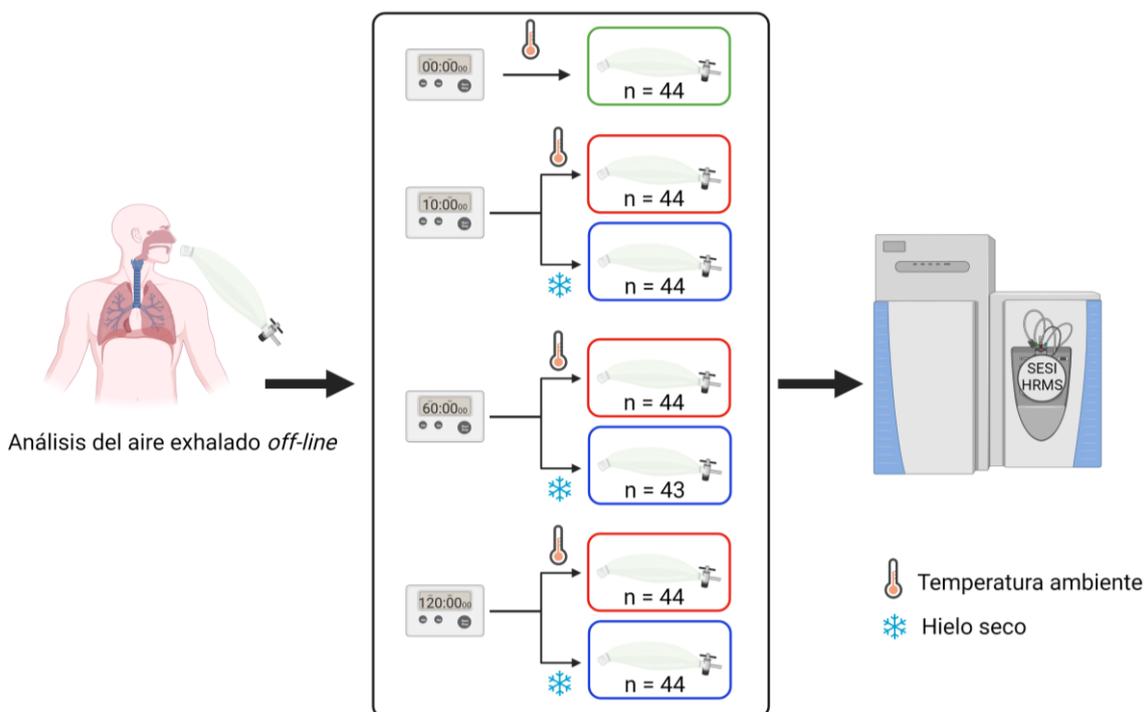


Figura 1. Diagrama de flujo de trabajo del estudio. Estudio de la conservación de las muestras de aire exhalado en bolsas Nalophan® (Dispositivo *off-line* descrito en Decrue *et al.*²²) a lo largo del tiempo en distintas condiciones para su posterior análisis en SESI-HRMS (ionización secundaria por *electrospray* – espectrometría de masas de alta resolución).

Para determinar la pérdida de intensidad de las señales m/z en función del tiempo, se normalizó la intensidad para cada una de las 2681 señales m/z detectadas con respecto a la intensidad obtenida en el análisis del aire exhalado en la bolsa a tiempo cero y se calculó la media entre todas las intensidades relativas para cada tiempo. Como se puede observar en la Figura 2, la media de las intensidades relativas se mantuvo por encima del 80% en las muestras almacenadas en las bolsas durante dos horas, tanto a temperatura ambiente como en hielo seco. Posteriormente, se filtró la matriz de datos obtenida en el presente estudio en base a la comparativa entre los métodos *off-line* y *on-line* realizada por Decrue *et al.*²² y se seleccionaron exclusivamente 870 señales m/z con un valor de concordancia de Lin's de CCC>0,6. En este sentido, se observó que las

intensidades de estas señales m/z a las dos horas se redujeron de media menos de un 25% con respecto a la bolsa a tiempo cero. Por otro lado, se seleccionaron 383 señales m/z detectadas en presente estudio, coincidentes con la lista elaborada por el proyecto DOPAEx, y se observó que las intensidades medias de estos *features* a las dos horas eran muy similares a las medidas a tiempo cero, siendo la media de intensidad relativa superior al 95% en las almacenadas a temperatura ambiente y superior al 85% en las muestras conservadas en hielo seco. Por último, se extrajeron de la matriz de este estudio únicamente las 111 señales m/z del presente estudio incluidas en la lista seleccionada de DOPAEx y que presentaron un Lin's CCC > 0,6 en la comparación entre los métodos *off-line* y *on-line*, y se determinó que la disminución media de sus intensidades relativas a las dos horas con respecto a la bolsa a tiempo cero fue prácticamente nula en las muestras almacenadas a temperatura ambiente, e inferior al 10% en las muestras conservadas en hielo seco. Asimismo, también se evaluó el efecto de los valores de m/z sobre la media de intensidad relativa en distintos periodos de almacenamiento y condiciones (temperatura ambiente y hielo seco) sin observar diferencias determinantes en la tendencia de los datos (Apéndice 1). No obstante, visualmente sí que se observó una mayor dispersión de los datos en unas pocas señales m/z al aumentar el periodo de almacenamiento.

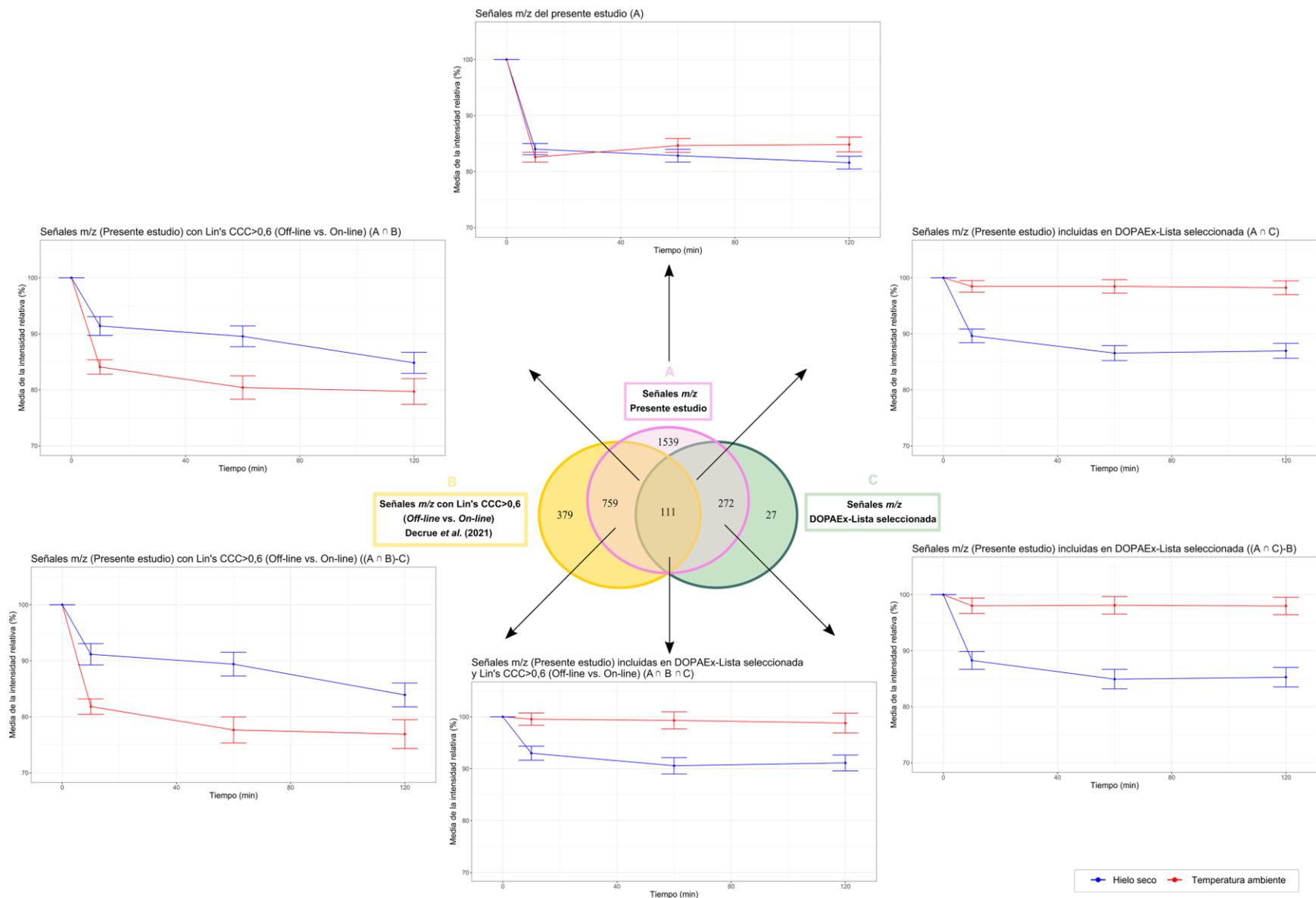


Figura 2. Media de la intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z detectadas en las muestras de aire exhalado del presente estudio en distintas condiciones de almacenamiento. (A) Señales m/z detectadas en el presente estudio. (B) Señales m/z detectadas por Decrue *et al.*²² con un Lin's CCC>0,6 en la comparación entre los métodos *off-line* y *on-line*. (C) Señales m/z detectadas en modo positivo mediante SESI-HRMS incluidas en la lista elaborada en el proyecto DOPAEx (*Determination of Optimal Procedures for Analysis of Expired Breath by Secondary Electrospray Ionization-Mass Spectrometry*). Las barras de error de los gráficos indican el intervalo de confianza del 95%. \cap : intersección de dos o más conjuntos.

Aldehídos en aire exhalado: efecto de las condiciones de almacenamiento

Una vez comprobada la adecuada conservación de las muestras de aliento en las bolsas, nos centramos en el efecto del almacenamiento sobre los aldehídos por su dimensión clínica, ya que son utilizados como indicadores del estrés oxidativo²⁴.

En la Figura 3 se indica la variación en función del tiempo de almacenamiento de la media de intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z de 27 aldehídos pertenecientes a las familias 2-alquenaes, 4-hidroxi-2-alquenaes y 4-hidroxi-2,6-alcadienales. En la Figura 4 se muestra el efecto del tamaño de la cadena de carbonos de los aldehídos en los niveles de intensidad de las señales m/z en las distintas condiciones de conservación del aire exhalado. Además, en la parte superior de la Figura 4 se indican los valores de Lin's CCC $>0,6$ obtenidos por Decrue *et al.*²² en la comparativa entre los enfoques *off-line* y *on-line* para los 27 aldehídos. La estimación de las medias de las intensidades relativas de todos los aldehídos en todas las condiciones fue bastante precisa, con estrechos intervalos de confianza del 95%, salvo en el caso del 4-hidroxi-2-dodecenal, donde se observó una mayor dispersión de los datos y un valor de Lin's CCC inferior al resto de los aldehídos estudiados.

Como se observa en la Figura 3, la media de intensidad relativa de todos los aldehídos identificados se conservó por encima del 80% a las 2 horas de almacenamiento en ambas temperaturas. En la mayoría de los aldehídos, la bajada más brusca de la media de la intensidad relativa se produjo entre los 10 y 60 minutos de almacenamiento, permaneciendo prácticamente estable entre los 60 y 120 minutos. En general, la disminución media de las intensidades relativas fue mayor en los aldehídos con menor número de carbonos de las familias 4-hidroxi-2-alquenaes y 4-hidroxi-2,6-alcadienales.

En cuanto al efecto de la temperatura, todos los aldehídos mantuvieron niveles de intensidad prácticamente iguales tras ser almacenados durante 10 minutos a temperatura ambiente y en las bolsas a tiempo cero mostrando una media de intensidad relativa en torno al 100%. Además, excepto en el 2-hexadecenal y en el 4-hidroxi-2-hexadecenal, las muestras de aliento se conservaron mejor en las bolsas almacenadas a temperatura ambiente, ya que la media de las intensidades relativas de los compuestos fue superior a las detectadas en las bolsas que se guardaron en hielo seco. Esta pérdida, que en ningún caso superó el 15%, era más pronunciada en aldehídos de cadena corta de las dos familias que contaban con un grupo hidroxilo, reduciéndose

estas diferencias al aumentar el número de carbonos, tal y como se observa, en la Figura 3 y en la Figura 4.

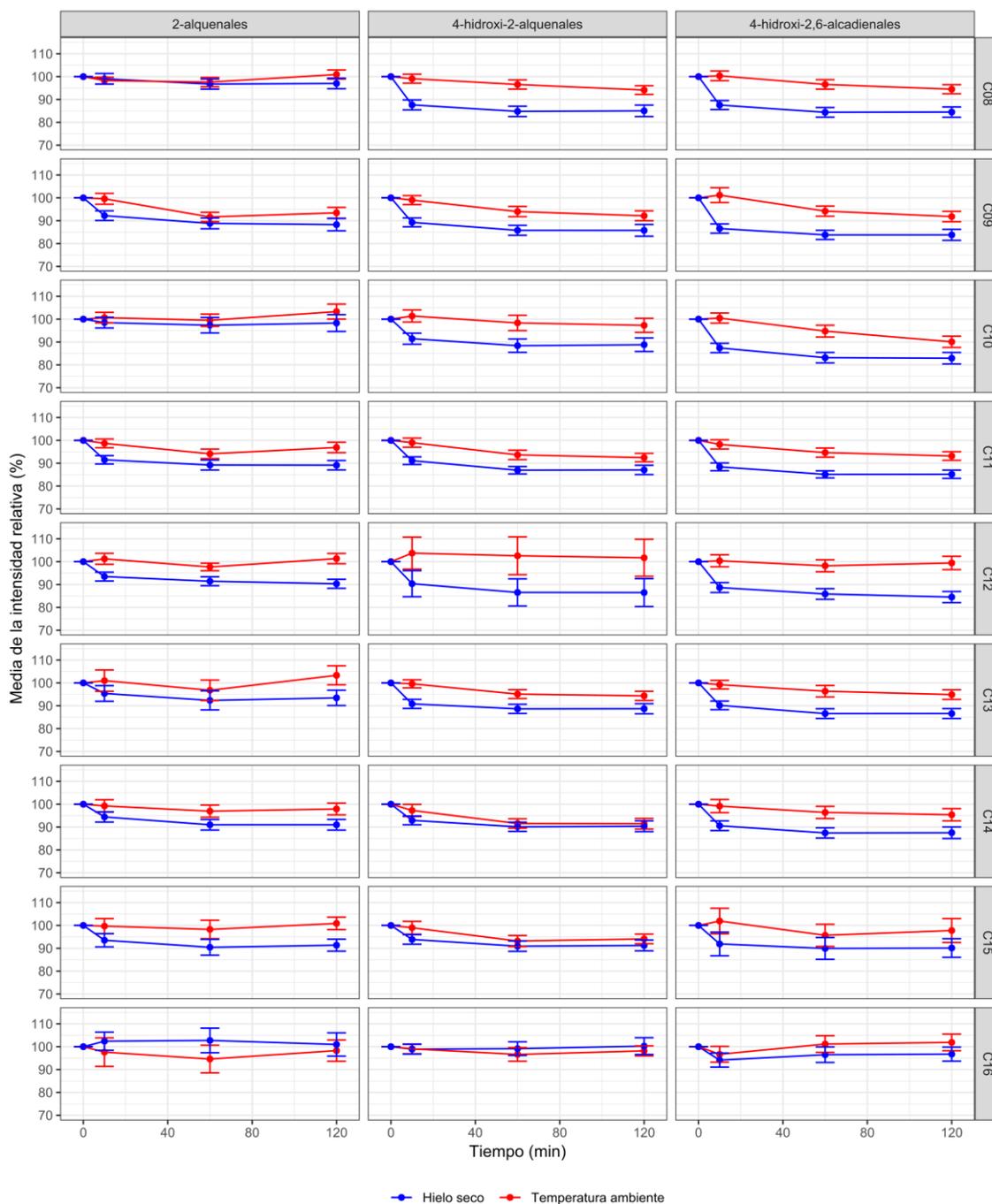


Figura 3. Media de la intensidad relativizada en base a la bolsa a tiempo cero en distintas condiciones de almacenamiento de las señales m/z de 27 aldehídos detectados en las muestras de aire exhalado (C8-C16 2-alquenales, C8-C16 4-hidroxi-2-alquenales y C8-C16 4-hidroxi-2,6-alcadienales). Las barras de error de los gráficos indican el intervalo de confianza del 95%.

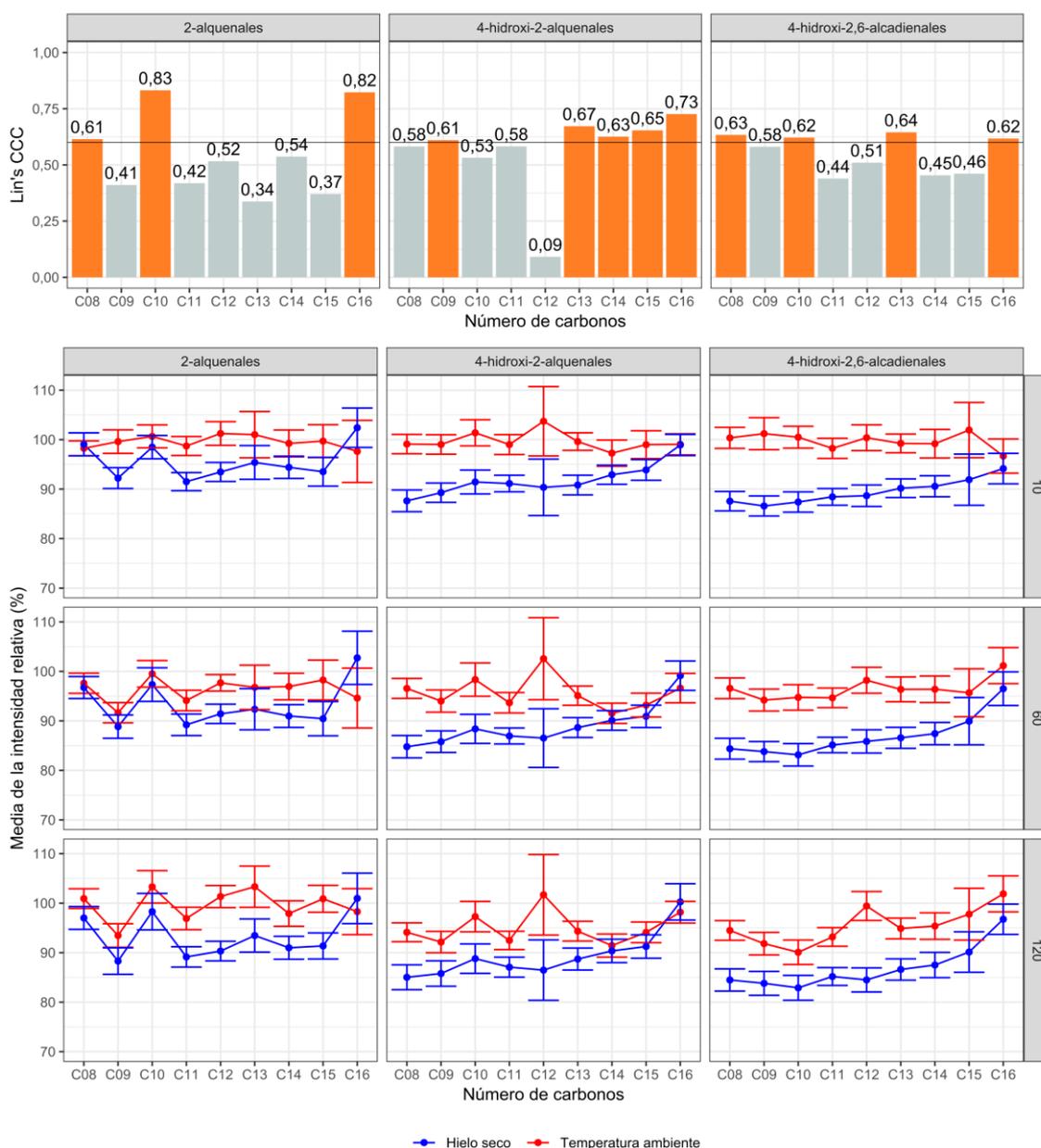


Figura 4. Variación en función del número de carbono de la media de la intensidad relativizada en base a la bolsa a tiempo cero en distintas condiciones de almacenamiento de las señales *m/z* de 2-alquenoales, 4-hidroxi-2-alquenoales y 4-hidroxi-2,6-alcadienoales detectados en las muestras de aire exhalado. Las barras de error de los gráficos indican el intervalo de confianza del 95%. En la parte superior de la figura, se indican los valores de Lin's CCC > 0,6 en la comparación entre las estrategias *off-line* y *on-line* obtenidos por Decrue *et al.*²² para los aldehídos estudiados.

Tras comprobar con éxito que la conservación del aliento en bolsas durante 2 h no afectaba a la determinación adecuada de las señales correspondientes a los aldehídos, se analizó el efecto del almacenamiento durante 24 h. Para ello, en 14 de las 44 visitas, además de las 7 bolsas recogidas en todas las citas, los sujetos llenaron otras dos bolsas Nalophan® de aire exhalado que se almacenaron durante 24 horas, manteniendo una de las bolsas a temperatura ambiente y la otra en hielo seco.

Posteriormente se llevó a cabo el análisis de las muestras mediante SESI-HRMS. En la Figura 5 se muestra la evolución de la media de la intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z de los 27 aldehídos estudiados a lo largo de las 24 horas de almacenamiento.

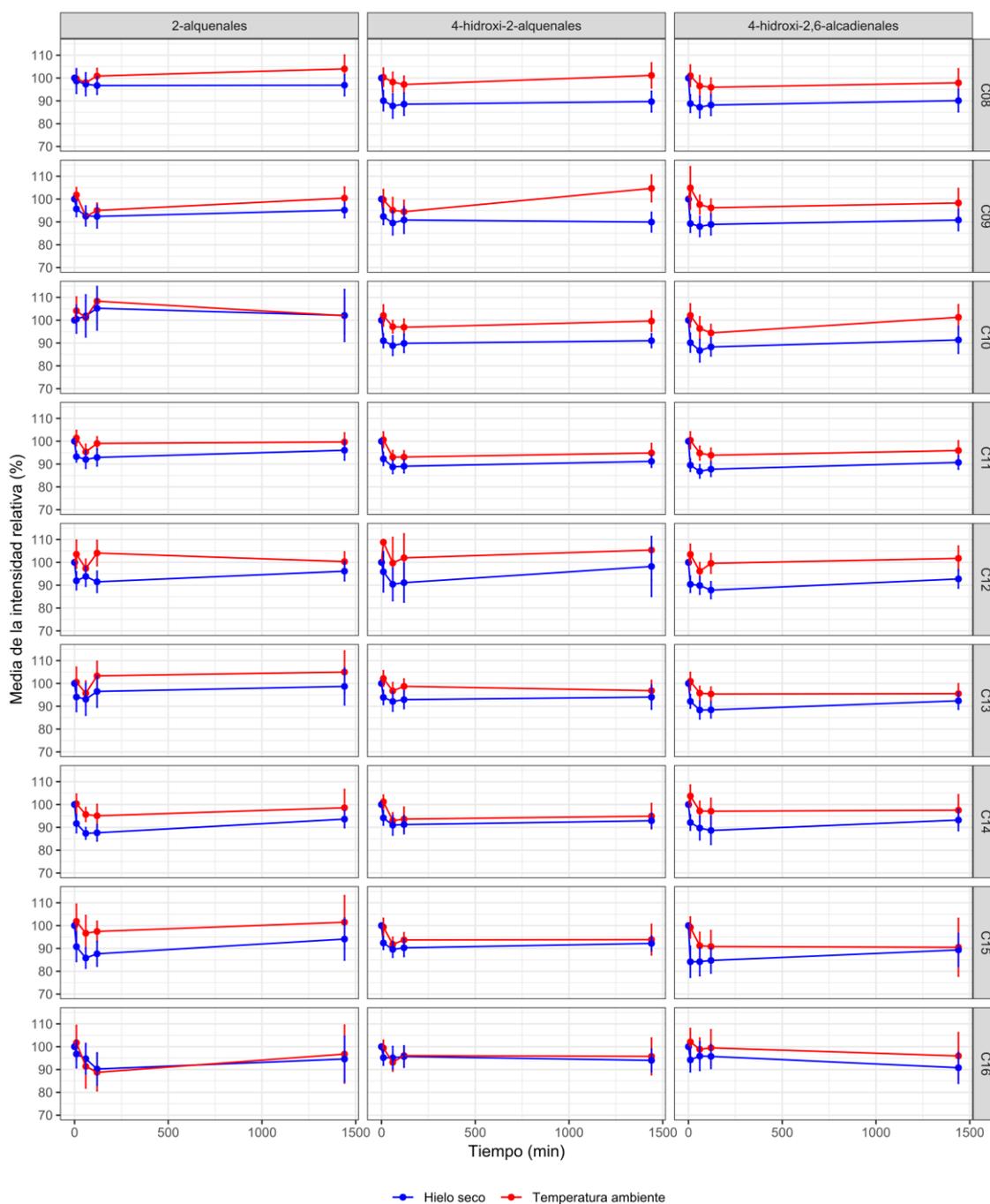


Figura 5. Efecto del almacenado de muestras de aire exhalado durante un periodo de 24 horas en diferentes condiciones sobre la media de la intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z de 27 aldehídos (C8-C16 2-alkenales, C8-C16 4-hidroxi-2-alkenales y C8-C16 4-hidroxi-2,6-alcadienales). Las barras de error de los gráficos indican el intervalo de confianza del 95%.

En todos los aldehídos la media de intensidad relativa se mantuvo por encima del 80% a las 24 horas de almacenamiento, observándose pocos cambios a partir de las 2 horas de almacenamiento. De nuevo, se comprobó, que las medias de las intensidades relativas fueron más elevadas en las muestras almacenadas a temperatura ambiente que las almacenadas en hielo seco, aunque la diferencia, en general, no fue superior al 15%.

En la Figura 6 se muestra la media de las intensidades relativas de las señales m/z de los 12 aldehídos que tenía un valor de Lin's CCC > 0,6 en la comparación entre los métodos *off-line* y *on-line* realizada por Decrue *et al.*²² (Figura 4). Como se puede observar, la media de intensidad relativa se conservó a las 24 de almacenamiento por encima del 95% a temperatura ambiente y por encima del 85% en hielo seco.

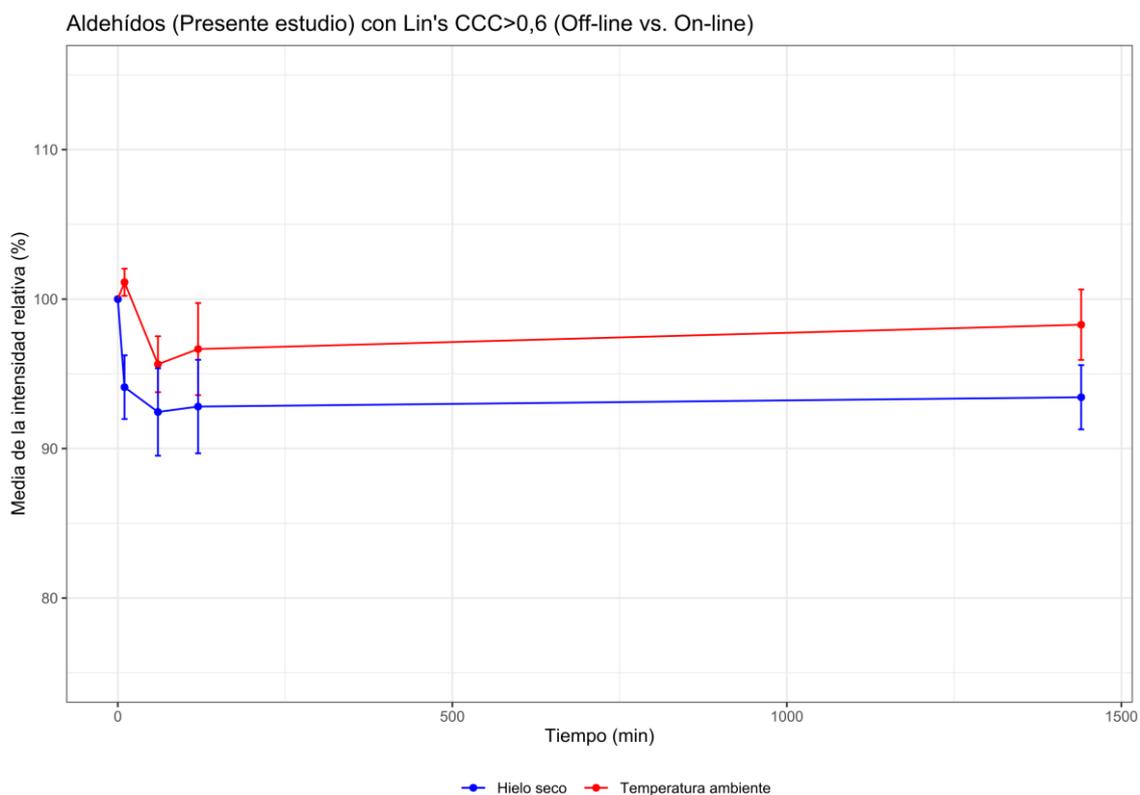


Figura 6. Impacto del almacenamiento del aire exhalado durante un periodo de 24 horas en diferentes condiciones sobre la media de la intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z de los aldehídos que presentaron un valor de Lin's CCC > 0,6 en la comparación entre los métodos *off-line* y *on-line* realizada por Decrue *et al.*²² Las barras de error de los gráficos indican el intervalo de confianza del 95%.

Solubilidad en agua de 2-alquenes, 4-hidroxi-2-alquenes y 4-hidroxi-2,6-alcadienes

Para evaluar la relación entre la naturaleza química de los aldehídos y el tiempo de almacenamiento, se determinó su solubilidad en agua. Entonces, utilizando la herramienta ALOGPS 2.1^{25,26} se obtuvieron los valores predichos de solubilidad en agua expresados como logS de los 27 aldehídos estudiados. Según estos valores, los 4-hidroxi-2-alquenes y los 4-hidroxi-2,6-alcadienes presentaban una mayor solubilidad en agua que los 2-alquenes (Figura 7). Los aldehídos de las dos familias que contaban con un grupo hidroxilo tenían valores de logS muy similares y la solubilidad de los compuestos disminuyó proporcionalmente con el número de carbonos de la cadena alquílica, siendo las pendientes similares en todas las familias de aldehídos. Por otro lado, la mayoría de los 2-alquenes eran insolubles en agua (valores de logS menores de -4), siendo, únicamente, los tres aldehídos de cadena más corta de esta familia ligeramente solubles en agua (valores de logS entre -2 y -4) (Figura 8).

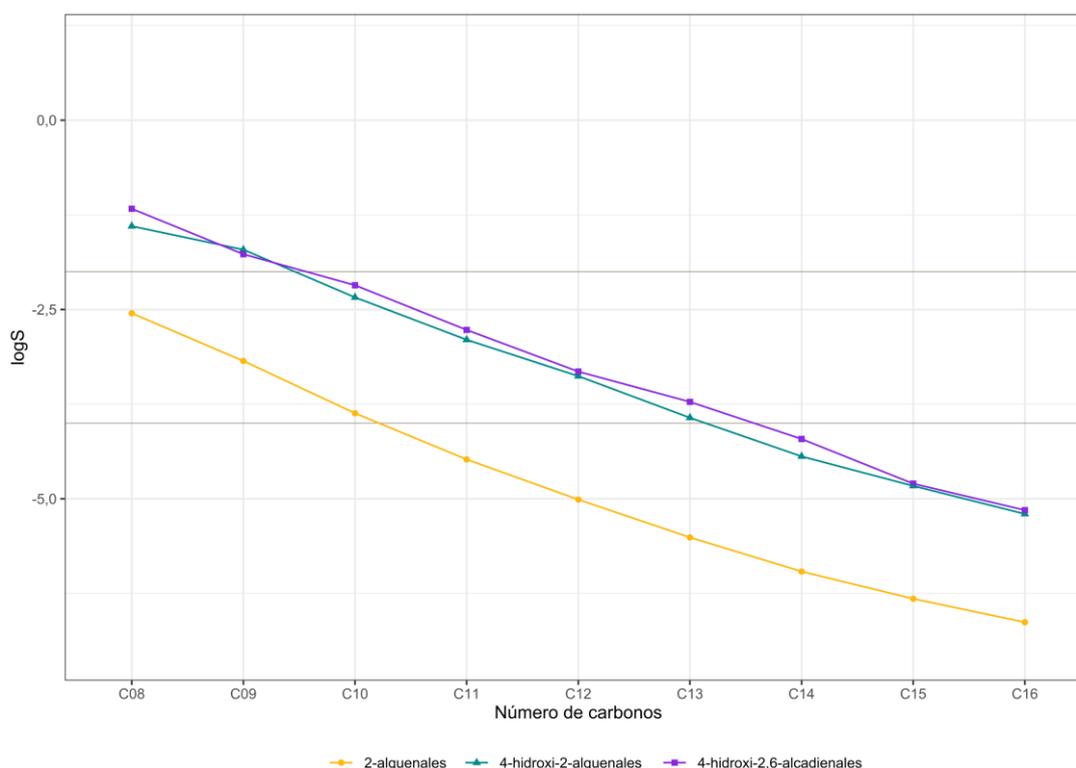


Figura 7. Impacto de la familia química sobre la solubilidad en agua de aldehídos detectados en las muestras de aire exhalado. Valores de logS predichos con la herramienta ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps/>)^{25,26}.

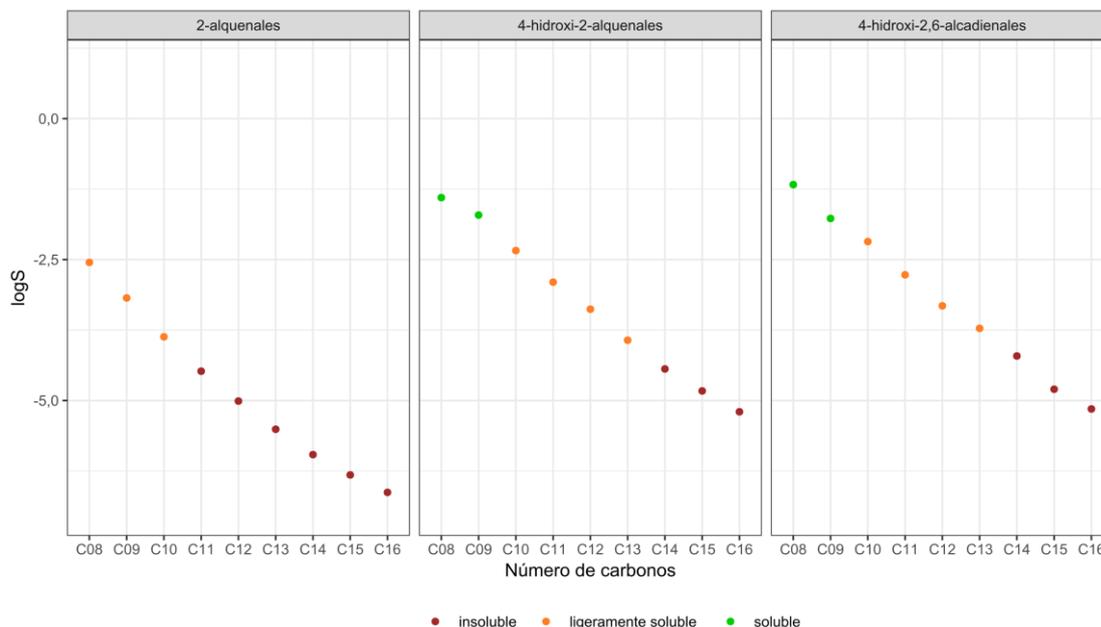


Figura 8. Solubilidad en agua de 2-alkenales, 4-hidroxi-2-alkenales y 4-hidroxi-2,6-alkadienales detectados en las muestras de aire exhalado. Valores de logS predichos con la herramienta ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps/>)^{25,26}. Los aldehídos fueron clasificados en base a su valor de logS en solubles ($\log S > -2$), ligeramente solubles ($\log S < -2$ y $\log S > -4$), e insolubles ($\log S < -4$).

La Figura 9 expone la relación entre los valores de solubilidad de los aldehídos y la variación de la media de la intensidad relativa de los 27 aldehídos tras el almacenamiento de las muestras en diferentes condiciones. Como se observa, las medias de las intensidades relativas y los valores de logS mostraban una asociación lineal. Los compuestos más solubles en agua (valores de logS mayores de -2) sufrieron una mayor pérdida media de intensidad relativa que los que fueron clasificados como insolubles o ligeramente insolubles. Además, esta diferencia se hizo más evidente en hielo seco que a temperatura ambiente, y fue aumentando conforme aumentó el tiempo de almacenamiento antes del análisis de las muestras de aire exhalado.

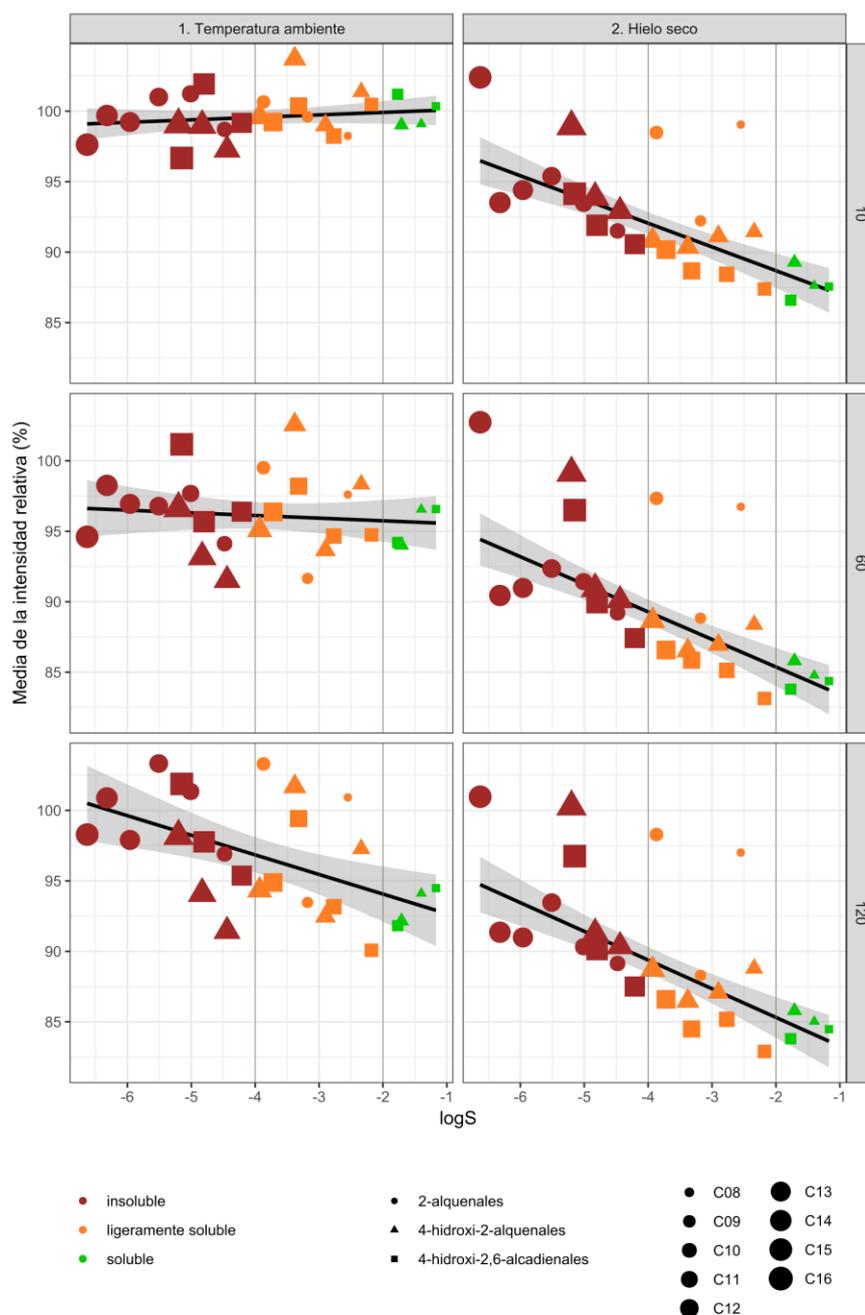


Figura 7. Influencia de la solubilidad en agua en la variación de la media de la intensidad relativizada en base a la bolsa a tiempo cero de las señales m/z de 27 aldehídos (C8-C16 2-alquenes, C8-C16 4-hidroxi-2-alquenes y C8-C16 4-hidroxi-2,6-alcadienes) a causa del almacenamiento del aire exhalado en diferentes condiciones. Valores de $\log S$ predichos con la herramienta ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps/>)^{25,26}. Los aldehídos fueron clasificados en base a su valor de $\log S$ en solubles ($\log S > -2$), ligeramente solubles ($\log S < -2$ y $\log S > -4$), e insolubles ($\log S < -4$). Líneas de tendencia correspondientes a regresión lineal robusta mediante el paquete MASS de R.

DISCUSIÓN

En el campo del análisis del aire exhalado, SESI-MS se ha utilizado tradicionalmente para el análisis del aire exhalado *on-line* a tiempo real¹⁵. Sin embargo, el dispositivo *off-line* desarrollado recientemente por Decrue *et al.*²² ha abierto también la posibilidad del uso de SESI-HRMS en el análisis del aire exhalado *off-line*, probando que se pueden determinar más de un millar de señales *m/z* con un alto grado de semejanza con las medidas realizadas mediante el método convencional de análisis directo del aliento a tiempo real (Lin's CCC > 0,6 en la comparación entre los métodos *off-line* y *on-line*). De esta manera, ya no es imprescindible que los sujetos se desplacen expresamente al laboratorio en el que se localice SESI-HRMS, lo que es muy importante en el caso de pacientes pasivos no cooperativos²². Una vez que se ha comprobado que la recogida previa del aire exhalado en un contenedor, concretamente en una bolsa Nalophan®, no implica una pérdida relevante de la calidad de los resultados en comparación con el análisis directo del aire exhalado²², en el presente capítulo se ha estudiado el impacto que tiene el tiempo de almacenamiento de las muestras de aire exhalado conservadas a dos temperaturas, a temperatura ambiente y en hielo seco. El hecho de garantizar que el almacenamiento de las muestras de aire exhalado no compromete su estabilidad, permitirá tanto poder beneficiarse de las ventajas propias de la estrategia del análisis del aire exhalado *off-line* como la posibilidad de transportar fácilmente las muestras y de centralizar los análisis de aliento en un único laboratorio^{16,27,28}. De este modo, en el presente estudio se ha estudiado la conservación de las señales *m/z* detectas en las muestras de aire exhalado a dos temperaturas (temperatura ambiente y hielo seco) y durante tres periodos de almacenamiento (10 min, 60 min y 120 min) (Figura 1). Como se puede observar en la Figura 2, en el conjunto de las señales *m/z* detectas en este estudio, sólo se produjo una mínima pérdida de intensidad con respecto a la bolsa a tiempo 0, manteniéndose por encima del 80% tras el almacenamiento de 2 horas. Esta reducida disminución en la media de intensidad relativa durante el almacenamiento se pudo ocasionar, bien por la adsorción de los compuestos sobre la superficie de Nalophan®, por la difusión de los VOCs a través de las paredes de la bolsa y/o por la condensación²⁹. Asimismo, las 870 señales *m/z* detectadas con una alta calidad (Lin's CCC>0,6 en la comparativa entre los métodos *off-line* y *on-line* realizada por Decrue *et al.*²²), también conservaron niveles medios de intensidad similares a los de la bolsa a tiempo cero, ya que la disminución de la media de intensidad relativa fue inferior al 25% a las 2 h de almacenamiento. No obstante, las pérdidas de intensidad, a lo largo del almacenamiento, fueron todavía menores al considerar únicamente las 111 señales *m/z* incluidas en la lista elaborada en el proyecto

DOPAEx y con un Lin's CCC > 0,6 en la comparación entre los métodos *off-line* y *on-line*. De hecho, en este subconjunto de señales *m/z*, la media de la intensidad relativa tras 2 horas del almacenamiento fue en torno al 90% en las muestras guardadas en hielo seco y alrededor del 100% en las almacenadas a temperatura ambiente. Por consiguiente, los resultados obtenidos por Decrue *et al.*²² que avalaban la aplicabilidad de un método *off-line* para el análisis de aire exhalado mediante SESI-HRMS a escasos minutos de su muestreo, son extrapolables después de 2 horas de almacenamiento sin una pérdida determinante de la intensidad de las señales *m/z*.

Además, en este estudio, se analizaron en detalle el efecto del almacenamiento de las muestras de aire exhalado hasta las 24 h sobre los *features* de 27 aldehídos de tres familias (2-alquenas, 4-hidroxi-2-alquenas y 4-hidroxi-2,6-alcadienas). Los aldehídos, y en particular el 4-hidroxi-2-nonenal (4-HNE), han sido propuestos en multitud de ocasiones como potenciales biomarcadores del estrés oxidativo relacionado con numerosas patologías, ya que son subproductos de la peroxidación lipídica³⁰. Por ello, estudios recientes han centrado sus esfuerzos en el análisis de aldehídos en aire exhalado mediante diversas técnicas de espectrometría de masas^{31,32}. Tal y como muestran la Figura 4 y la Figura 5, la media de intensidad relativa se mantuvo por encima del 80% en todos los aldehídos después de 2 horas de almacenamiento. En el caso particular del 4-hidroxi-2-docenal, se observó una elevada dispersión de los datos en consonancia con el bajo valor de Lin's CCC obtenido en la comparación con el método *on-line* por Decrue *et al.* (Lin's CCC = 0,09), que podría deberse a la existencia de diferentes isómeros de este compuesto con distintos grados de afinidad por la superficie de Nalophan® de las bolsas²². En general, la disminución de las medias de las intensidades relativas de los aldehídos ocurrió principalmente durante la primera hora de almacenamiento, manteniéndose constantes incluso 24 horas después de la toma de muestra (Figura 5). Además, un almacenamiento de 10 minutos a temperatura ambiente (periodo equivalente al transporte de muestras dentro de una misma instalación o entre edificios colindantes) supuso una variación de la media de la intensidad relativa prácticamente nula en todos los aldehídos (media de intensidad relativa en torno al 100%).

Asimismo, en la mayoría de los aldehídos, las medias de las intensidades relativas se mantuvieron más elevadas en las muestras almacenadas a temperatura ambiente que en las conservadas en hielo seco, independientemente del periodo de almacenamiento (Figura 3 y Figura 4). Este fenómeno también se pudo observar tanto en las señales *m/z* incluidas en la lista elaborada por el proyecto DOPAEx (Figura 2)

como en los *features* de los 12 aldehídos con un Lin's CCC > 0,6 en la comparativa entre el método *off-line* y *on-line* (Figura 6). En este sentido, razonamos que las diferencias observadas entre ambas condiciones, temperatura ambiente y hielo seco, podrían estar relacionadas con la condensación del vapor de agua presente en las muestras de aliento producido por la fluctuación de temperatura³³. El vapor de agua representa un alto porcentaje del aire exhalado y la humedad de las muestras puede influir notablemente en el proceso de su conservación^{28,34}. En este sentido, en el análisis del aire exhalado *off-line* mediante GC-MS, donde es necesario preconcentrar las muestras, existen diversas estrategias para minimizar el efecto de la humedad como el empleo de tubos de desorción térmica que contengan materiales sorbentes poco sensibles a la elevada presencia de agua (ej. Tenax TA)^{35,36}. Sin embargo, en los sistemas analíticos con características inherentes para el análisis del aire exhalado *on-line* que no realizan un paso previo de preconcentración de las muestras¹, es importante la selección de una temperatura elevada de almacenamiento para reducir los problemas relacionados con la condensación del vapor de agua²⁸. Ahora bien, el almacenamiento a alta temperatura puede provocar una mayor liberación de contaminantes procedentes de las bolsas de muestreo de gas, como el 2-metil-1,3-dioxolano y el 1,2-etanodiol en el caso de las bolsas Nalophan®³⁷. En cambio, se ha indicado que el calentamiento de las muestras de aliento hasta la temperatura corporal (37°C) podría mantener el vapor de agua en estado gaseoso y evitar problemas de condensación²⁹, pero también podría ocasionar un aumento de la presencia de artefactos. Por tanto, se debe valorar la sencillez del protocolo de almacenamiento frente a los beneficios y desventajas que se obtienen del calentamiento de las muestras y por ello, en este estudio se seleccionó la conservación de las muestras a temperatura ambiente. Por su parte, la conservación de las muestras en hielo seco produjo un efecto de condensación del vapor de agua por la diferencia de temperatura, pero el impacto de la conservación en hielo seco sobre la media de intensidad relativa no fue igual para todos los 27 aldehídos (Figura 3 y Figura 4). Así, el mayor descenso de la media de intensidad relativa tras el almacenamiento en hielo seco se produjo en los aldehídos de cadena corta de las dos familias que contaban con un grupo hidroxilo. Esto podría deberse a la distinta solubilidad en agua de los aldehídos (Figura 7 y Figura 9), siendo los compuestos solubles en agua más sensibles a los efectos de la condensación del vapor de agua^{29,33}. Como se puede ver en la Figura 9, existe una relación lineal entre las medias de las intensidades relativas de los aldehídos y sus valores de logS. Además, comparando los datos de las muestras conservadas a temperatura ambiente con las conservadas en hielo seco, la pendiente de disminución de la intensidad relativa de los compuestos fue pronunciada en hielo seco, donde la condensación del vapor de agua era más probable.

CONCLUSIÓN

En el presente estudio, se ha demostrado por primera vez que es posible almacenar muestras de aire exhalado humano durante horas en bolsas de muestreo de gas y posteriormente analizarlas mediante SESI-HRMS, sin que afecte gravemente a los niveles de los compuestos determinados (media de intensidad relativa superior 80% tras un almacenamiento de 2 horas). Además, es aconsejable el mantenimiento de las bolsas de aliento a temperatura ambiente para evitar la condensación de agua que afectaría a la conservación de los compuestos, especialmente de los más solubles en agua. La posibilidad de poder almacenar y transportar las muestras de aliento exhalado ofrece una mayor flexibilidad logística en la aplicación de la técnica, permitiendo, por ejemplo, reclutar pacientes de diferentes hospitales.

MATERIALES Y MÉTODOS

Participantes y recogida de muestras de aire exhalado

Para el presente estudio, se reclutaron dos hombres y dos mujeres sanos (edad media 29,8 años, rango 26-38 años). El protocolo del estudio fue aprobado por los Comités de Ética locales (Comité de Ética del Noroeste y Centro de Suiza 2018-01324) de acuerdo con las directrices de la Declaración de Helsinki. El consentimiento informado por escrito de los participantes fue obtenido en el momento del reclutamiento. Durante 2 meses (octubre-diciembre de 2021), los 4 sujetos realizaron diversas visitas al Hospital Infantil Universitario de Basilea (UKBB) (Suiza) (un total de 44 visitas). Para la recogida de muestras de aire exhalado se empleó el dispositivo *off-line* descrito por Decrue *et al.*²² constituido por una bolsa Nalophan® de aproximadamente 2 L y 700 cm² de superficie (Nalophan® NA, 20 µm ± 5 µm de espesor, Kalle) acoplada en un extremo a una boquilla (Hudson RCI®), y en el otro, a un tubo (Rotilabo® PTFE, 6 mm, 8 mm, longitud 90 mm, Carl Roth®) conectado a una válvula (VHK2-08F-08F, SMC Suiza). Los sujetos permanecieron al menos una hora en ayuno antes de las determinaciones del aire exhalado, y llenaron las bolsas Nalophan® con una exhalación profunda a flujo constante, tras contener previamente la respiración durante unos segundos. En cada visita, se recogieron 7 bolsas Nalophan®: una de ellas fue analizada inmediatamente (bolsa a tiempo cero), dos se analizaron 10 minutos después de la toma de muestra (una almacenada a temperatura ambiente (aproximadamente 20°C) y la otra en hielo seco), dos se analizaron 60 minutos después de la toma de muestra (una almacenada a temperatura ambiente y la otra en hielo seco), y otras dos se analizaron 120 minutos después de la toma de muestra (una almacenada a temperatura ambiente

y la otra en hielo seco). Además, en 14 de las visitas se recogieron dos bolsas Nalophan® adicionales que se analizaron a las 24 horas de la toma de muestra (una se almacenó a temperatura ambiente y la otra en hielo seco).

Análisis del aire exhalado

Las muestras de aire exhalado fueron analizadas mediante una fuente de iones SESI (SuperSESI, FIT, España) acoplada a un espectrómetro de masas de alta resolución (Q-Exactive Plus, Thermo Fisher Scientific, Alemania) siguiendo el protocolo descrito previamente^{22,38,39}. Transcurrido el tiempo de almacenamiento, el aire exhalado contenido en las bolsas Nalophan® fue introducido directamente en la fuente de iones. La fuente de iones poseía un controlador de flujo de masa que permitía garantizar la entrada de manera constante del aire exhalado en el ionizador, fijándose a 0,3 L/min. La adquisición de datos se realizó en modo de barrido completo (polaridad positiva y rango de m/z 70-1000) con una resolución de 140 000 (a m/z 200) y se empleó el *software* Q-Exactive Tune. En la formación del *electrospray*, se utilizó un emisor capilar de sílice TaperTip (New Objective, EE.UU.) de 20 μm de diámetro interior y ácido fórmico al 0,1% en agua. Antes de realizar las medidas de aire exhalado, diariamente se analizó un estándar en estado gaseoso de 8 componentes (Dalian Special Gases, Dalian, China), con el fin de confirmar el correcto funcionamiento del SESI-HRMS. Asimismo, se comprobó que la intensidad de señal del α -Terpineno protonado ($m/z = 137$) era superior a 107 u.a. en todas las mediciones.

Preprocesamiento y análisis de datos

El preprocesamiento de datos se llevó a cabo con MATLAB (versión 2020b, MathWorks Inc., EE.UU.), y el posterior análisis de datos con R (versión 4.0.5). El protocolo seguido para el preprocesamiento de datos tras el análisis del aire exhalado *off-line* mediante SESI-HRMS es detallado por Decrue *et al*²². El acceso a los datos brutos obtenidos fue posible gracias a aplicaciones de consola C# basadas en RawFileReader de Thermo Fisher Scientific (versión 5.0.0.38). El área bajo la curva de cada una de las señales m/z o *features* fue determinada y se normalizó en función de la intensidad global de la exhalación. Tras el preprocesamiento de datos se obtuvo una matriz de datos con las intensidades de las señales m/z detectados en las muestras de aire exhalado. Para ajustarse a una distribución normal, las intensidades fueron normalizadas aplicando la raíz 5ª. Una vez obtenida una matriz útil para el análisis de datos, se calcularon las intensidades relativas de cada señal m/z en las distintas condiciones (10 min-Temperatura ambiente, 10 min-Hielo seco, 60 min-Temperatura

ambiente, 60 min-Hielo seco, 120 min-Temperatura ambiente y 120 min-Hielo seco) expresadas como porcentaje con respecto a la bolsa a tiempo cero. De este modo, se definió la media de la intensidad relativa (%) como el promedio de las intensidades normalizadas expresadas en porcentaje de las señales m/z en una determinada condición en base a la bolsa a tiempo cero. Asimismo, en el análisis de datos, también se tuvieron en cuenta el listado de señales m/z que mostraron un Lin's CCC>0,6 en comparación entre las estrategias de análisis de aire exhalado *off-line* y análisis de aire exhalado *on-line* a tiempo mediante SESI-HRMS realizada por Decrue *et al.*²², así como la lista de señales m/z seleccionada por el proyecto DOPAEx (*Determination of Optimal Procedures for Analysis of Expired Breath by Secondary Electrospray Ionization-Mass Spectrometry*) (datos no publicados). Por último, se utilizó la herramienta ALOGPS 2.1 (<http://www.vcclab.org/lab/alogps/>)^{25,26} para la predicción de la solubilidad en agua expresada como logS de 27 aldehídos detectados en el aire exhalado (C8-C16 2-alquenes, C8-C16 4-hidroxi-2-alquenes y C8-C16 4-hidroxi-2,6-alcadienos), a partir de las nomenclaturas SMILES (*Simplified Molecular-Input Line-Entry System*) (Apéndice 2) de dichos compuestos. En función al valor de logS predicho, los aldehídos se clasificaron como solubles si presentaron un logS mayor de -2, ligeramente solubles si mostraron un logS entre -2 y -4, e insolubles si tenían un logS menor de -4⁴⁰.

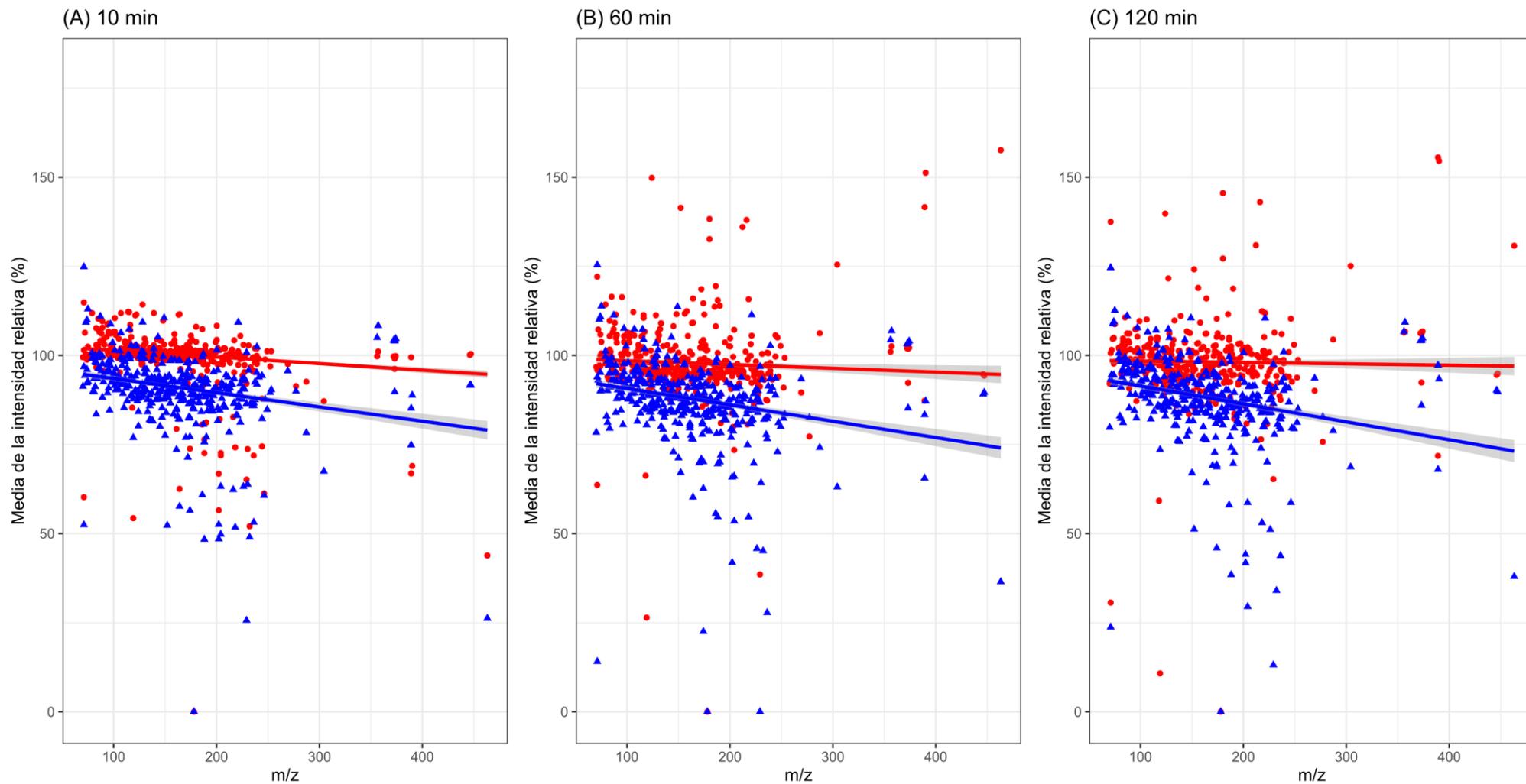
BIBLIOGRAFÍA

1. Rattray, N. J. W., Hamrang, Z., Trivedi, D. K., Goodacre, R. & Fowler, S. J. Taking your breath away: metabolomics breathes life in to personalized medicine. *Trends Biotechnol.* **32**, 538–548 (2014).
2. Boots, A. W. *et al.* The versatile use of exhaled volatile organic compounds in human health and disease. *J. Breath Res.* **6**, 027108 (2012).
3. Boots, A. W., Bos, L. D., van der Schee, M. P., van Schooten, F. J. & Sterk, P. J. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. *Trends Mol. Med.* **21**, 633–644 (2015).
4. Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, 034001 (2014).
5. Azim, A., Barber, C., Dennison, P., Riley, J. & Howarth, P. Exhaled volatile organic compounds in adult asthma: a systematic review. *Eur. Respir. J.* **54**, 1900056 (2019).
6. Neerinx, A. H. *et al.* Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* **52**, 1616–1627 (2017).
7. Marzorati, D. *et al.* A review of exhaled breath: a key role in lung cancer diagnosis. *J. Breath Res.* **13**, 034001 (2019).
8. Fens, N. *et al.* Exhaled Breath Profiling Enables Discrimination of Chronic Obstructive Pulmonary Disease and Asthma. *Am. J. Respir. Crit. Care Med.* **180**, 1076–1082 (2009).
9. Wang, W. *et al.* Accuracy of breath test for diabetes mellitus diagnosis: a systematic review and meta-analysis. *BMJ Open Diabetes Res. Care* **9**, e002174 (2021).
10. Dixit, K., Fardindoost, S., Ravishankara, A., Tasnim, N. & Hoorfar, M. Exhaled Breath Analysis for Diabetes Diagnosis and Monitoring: Relevance, Challenges and Possibilities. *Biosensors* **11**, 476 (2021).
11. Barker, M. *et al.* Volatile organic compounds in the exhaled breath of young patients with cystic fibrosis. *Eur. Respir. J.* **27**, 929–936 (2006).
12. Paff, T. *et al.* Exhaled molecular profiles in the assessment of cystic fibrosis and primary ciliary dyskinesia. *J. Cyst. Fibros.* **12**, 454–60 (2013).
13. Ahmed, W. M., Lawal, O., Nijssen, T. M., Goodacre, R. & Fowler, S. J. Exhaled

- Volatile Organic Compounds of Infection: A Systematic Review. *ACS Infect. Dis.* **3**, 695–710 (2017).
14. Sola Martínez, R. A. *et al.* Exhaled volatile organic compounds analysis in clinical pediatrics: a systematic review. *Pediatr. Res.* **89**, 1352–1363 (2021).
 15. Bruderer, T. *et al.* On-Line Analysis of Exhaled Breath. *Chem. Rev.* **119**, 10803–10828 (2019).
 16. Dharmawardana, N., Woods, C., Watson, D. I., Yazbeck, R. & Ooi, E. H. A review of breath analysis techniques in head and neck cancer. *Oral Oncol.* **104**, 104654 (2020).
 17. Shende, P., Vaidya, J., Kulkarni, Y. A. & Gaud, R. S. Systematic approaches for biodiagnostics using exhaled air. *J. Control. Release* **268**, 282–295 (2017).
 18. Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R. & Fowler, S. J. Exhaled breath analysis: a review of ‘breath-taking’ methods for off-line analysis. *Metabolomics* **13**, 110 (2017).
 19. Singh, K. D. *et al.* Translating secondary electrospray ionization-high-resolution mass spectrometry to the clinical environment. *J. Breath Res.* **12**, 027113 (2018).
 20. Singh, K. D. *et al.* Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **411**, 4883–4898 (2019).
 21. Pereira, J. *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: An Overview. *Metabolites* **5**, 3–55 (2015).
 22. Decrue, F. *et al.* Combination of Exhaled Breath Analysis with Parallel Lung Function and FeNO Measurements in Infants. *Anal. Chem.* **93**, 15579–15583 (2021).
 23. Lin, L. I.-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268 (1989).
 24. García-Gómez, D. *et al.* Identification of 2-alkenals, 4-hydroxy-2-alkenals, and 4-hydroxy-2,6-alkadienals in exhaled breath condensate by UHPLC-HRMS and in breath by real-time HRMS. *Anal. Chem.* **87**, 3087–3093 (2015).
 25. Tetko, I. V., Tanchuk, V. Y., Kasheva, T. N. & Villa, A. E. P. Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *J. Chem. Inf. Comput. Sci.* **41**, 1488–1493 (2001).
 26. Tetko, I. V. & Tanchuk, V. Y. Application of Associative Neural Networks for

- Prediction of Lipophilicity in ALOGPS 2.1 Program. *J. Chem. Inf. Comput. Sci.* **42**, 1136–1145 (2002).
27. Ibrahim, W. *et al.* Breathomics for the Clinician: The use of volatile organic compounds in respiratory diseases. *Thorax* **76**, 514–521 (2021).
 28. Beauchamp, J., Herbig, J., Gutmann, R. & Hansel, A. On the use of Tedlar® bags for breath-gas sampling and analysis. *J. Breath Res.* **2**, 046001 (2008).
 29. Gilchrist, F. J. *et al.* An investigation of suitable bag materials for the collection and storage of breath samples containing hydrogen cyanide. *J. Breath Res.* **6**, 036004 (2012).
 30. Ratcliffe, N. *et al.* A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *J. Breath Res.* **14**, 034001 (2020).
 31. Huang, J., Kumar, S. & Hanna, G. B. Investigation of C3-C10 aldehydes in the exhaled breath of healthy subjects using selected ion flow tube-mass spectrometry (SIFT-MS). *J. Breath Res.* **8**, (2014).
 32. Holz, O. *et al.* Changes of breath volatile organic compounds in healthy volunteers following segmental and inhalation endotoxin challenge. *J. Breath Res.* **16**, 037102 (2022).
 33. Amann, A., Miekisch, W., Pleil, J., Risby, T. & Schubert, J. Methodological issues of sample collection and analysis of exhaled breath. in *Methodological issues for breath analysis* 96–114 (Maney Publishing, 2010).
 34. Ghimenti, S. *et al.* Comparison of sampling bags for the analysis of volatile organic compounds in breath. *J. Breath Res.* **9**, 047110 (2015).
 35. Mochalski, P., Wzorek, B., Śliwka, I. & Amann, A. Improved pre-concentration and detection methods for volatile sulphur breath constituents. *J. Chromatogr. B* **877**, 1856–1866 (2009).
 36. Beghi, S. & Guillot, J. M. Sample water removal method in volatile organic compound analysis based on diffusion through poly(vinyl fluoride) film. *J. Chromatogr. A* **1127**, 1–5 (2006).
 37. Brůhová Michalčíková, R., Dryahina, K., Smith, D. & Španěl, P. Volatile compounds released by Nalophan; implications for selected ion flow tube mass spectrometry and other chemical ionisation mass spectrometry analytical methods. *Rapid Commun. mass Spectrom.* **34**, e8602 (2020).

38. Gisler, A. *et al.* Real-time breath analysis of exhaled compounds upon peppermint oil ingestion by secondary electrospray ionization-high resolution mass spectrometry: technical aspects. *J. Breath Res* **14**, 46001 (2020).
39. Dev Singh, K. *et al.* Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **411**, 4883–4898 (2019).
40. Sorkun, M. C., Khetan, A. & Er, S. AqSolDB, a curated reference set of aqueous solubility and 2D descriptors for a diverse set of compounds. *Sci. Data* **6**, 1–8 (2019).



APÉNDICE 1. Impacto del valor de m/z sobre la media de intensidad relativizada en base a la bolsa a tiempo cero en diferentes condiciones de almacenamiento de las señales m/z del presente estudio incluidas en la lista elaborada en el proyecto DOPAEx (*Determination of Optimal Procedures for Analysis of Expired Breath by Secondary Electrospray Ionization-Mass Spectrometry*).

APÉNDICE 2. Aldehídos estudiados en el presente estudio.

Aldehídos	Familia de aldehídos	m/z	SMILES
2-octenal	2-alquenes	127,11174	CCCCC=CC=O
2-nonenal		141,127392	CCCCC=CC=O
2-decenal		155,143042	CCCCC=CC=O
2-undecenal		169,158692	CCCCC=CC=O
2-dodecenal		183,174342	CCCCC=CC=O
2-tridecenal		197,189992	CCCCC=CC=O
2-tetradecenal		211,205642	CCCCC=CC=O
2-pentadecenal		225,221292	CCCCC=CC=O
2-hexadecenal		239,236942	CCCCC=CC=O
4-hidroxi-2-octenal		4-hidroxi-2-alquenes	143,106656
4-hidroxi-2-nonenal	157,122306		CCCC(C=CC=O)O
4-hidroxi-2-decenal	171,137956		CCCC(C=CC=O)O
4-hidroxi-2-undecenal	185,153606		CCCC(C=CC=O)O
4-hidroxi-2-dodecenal	199,169257		CCCC(C=CC=O)O
4-hidroxi-2-tridecenal	213,184907		CCCC(C=CC=O)O
4-hidroxi-2-tetradecenal	227,200557		CCCC(C=CC=O)O
4-hidroxi-2-pentadecenal	241,216207		CCCC(C=CC=O)O
4-hidroxi-2-hexadecenal	255,231857	CCCC(C=CC=O)O	
4-hidroxi-2,6-octadienal	4-hidroxi-2,6-alcadienales	141,091006	CC=CCC(C=CC=O)O
4-hidroxi-2,6-nonadienal		155,106656	CCC=CCC(C=CC=O)O
4-hidroxi-2,6-decadienal		169,122306	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-undecadienal		183,137956	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-dodecadienal		197,153606	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-tridecadienal		211,169257	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-tetradecadienal		225,184907	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-pentadecadienal		239,200557	CCCC=CCC(C=CC=O)O
4-hidroxi-2,6-hexadecadienal	253,216207	CCCC=CCC(C=CC=O)O	

Capítulo 9

Conclusiones/ Conclusions

CONCLUSIONES

A continuación, se enumeran las principales conclusiones obtenidas a lo largo de esta Tesis doctoral:

1. El análisis de compuestos orgánicos volátiles (VOCs) en aire exhalado se encuentra todavía en una fase exploratoria (fase de descubrimiento de biomarcadores), pero la introducción de las prácticas típicas del campo de la metabolómica ayudará a alcanzar la traslación clínica del análisis del aire exhalado. El estado actual de la técnica refuerza la aplicación del análisis del aire exhalado mediante tecnologías basadas en espectrometría de masas en la fase de descubrimiento de biomarcadores.
2. Se ha diseñado un protocolo para la recogida de muestras de aire exhalado de manera no invasiva aplicable en adultos y lactantes, y se ha optimizado un protocolo de análisis del aire exhalado *off-line* mediante TD-GC/q-MS (desorción térmica acoplada a cromatografía de gases-espectrometría de masas de cuadrupolo simple).
3. Se ha desarrollado y validado un flujo de trabajo de código abierto para el preprocesamiento de datos útil para el análisis del aire exhalado *off-line* mediante TD-GC/q-MS.
4. El análisis del aire exhalado *off-line* mediante TD-GC/q-MS en mujeres en edad fértil permite diferenciar entre asmáticas con otras enfermedades atópicas coexistentes y no asmáticas (con o sin enfermedades atópicas). Se han generado 15 modelos estadísticamente significativos que discriminaban el asma con otras enfermedades atópicas. El sexo, la presencia de otros trastornos atópicos y los antecedentes parentales atópicos son factores que influyen en la presencia de asma.
5. La estación del año en la que se realiza el muestreo y la medición del aire exhalado es un factor que influye en las determinaciones de VOCs. Sin embargo, no se observó un sesgo estacional después de la implementación de ASCA (ANOVA- análisis de componentes simultáneos).
6. Existe una asociación entre el volatiloma exhalado y la función pulmonar en lactantes sanos. Los niveles en el aire exhalado de acetona, tetradecano y pentadecano mostraron una relación negativa significativa con varios parámetros de función pulmonar.
7. La exposición prolongada a humedad en el hogar puede influir en el volatiloma exhalado. Los niveles del contaminante 2-etil-1-hexanol fueron

significativamente elevados en mujeres en edad fértil expuestas a humedad en sus viviendas habituales.

8. Las muestras de aire exhalado humano pueden ser almacenadas durante horas en bolsas de muestreo de gas sin una disminución determinante de los niveles de los compuestos, y posteriormente ser analizadas por sistemas analíticos *on-line* como SESI-HRMS (ionización secundaria por *electrospray*-espectrometría de masas de alta resolución). Es recomendable la conservación de las muestras a temperatura ambiente para reducir los problemas derivados de la condensación del vapor de agua. Este resultado permite flexibilizar la gestión de las muestras en el hospital e incluso reclutar pacientes de otros hospitales.

CONCLUSIONS

The main conclusions obtained during this thesis are listed below:

1. The analysis of volatile organic compounds (VOCs) in exhaled breath is still in an exploratory phase (biomarker discovery phase), but the introduction of typical practices from the field of metabolomics will help to achieve clinical translation of breath analysis. The current state of the art supports the application of breath analysis by mass spectrometry-based technologies in biomarker discovery phase.
2. A protocol for non-invasive collection of exhaled breath samples applicable to adults and infants has been designed and a protocol for off-line breath analysis by TD-GC/q-MS (thermal desorption coupled to gas chromatography-single quadrupole mass spectrometry) has been optimised.
3. An open-source data preprocessing workflow useful for off-line breath analysis by TD-GC/q-MS has been developed and validated.
4. Off-line breath analysis by TD-GC/q-MS in women of childbearing age allows differentiation between asthmatics with other coexisting atopic diseases and non-asthmatics (with or without atopic diseases). Fifteen statistically significant models that discriminate asthma with other atopic diseases have been generated. Gender, presence of other atopic disorders and atopic parental history are factors that have an influence on the presence of asthma.
5. The season of the year in which exhaled breath is collected and measured is a factor that impacts on VOC determinations. However, no seasonal bias was observed after application of ASCA (ANOVA - simultaneous components analysis).
6. There is an association between exhaled volatilome and lung function in healthy infants. Exhaled breath levels of acetone, tetradecane and pentadecane showed a significant negative relationship with several lung function parameters.
7. Prolonged exposure to indoor dampness at home could influence exhaled volatilome. Levels of the pollutant 2-ethyl-1-hexanol were significantly elevated in women of childbearing age exposed to dampness in their usual residences.
8. Human exhaled breath samples can be stored for hours in gas sampling bags without a substantial decrease of compound levels, and subsequently analysed by on-line analytical systems such as SESI-HRMS (secondary electrospray ionisation-high resolution mass spectrometry). It is advisable to store the samples at room temperature to reduce the problems caused by condensation of water

vapour. This result allows flexibility in sample management at the hospital and even recruitment of patients from other hospitals.

