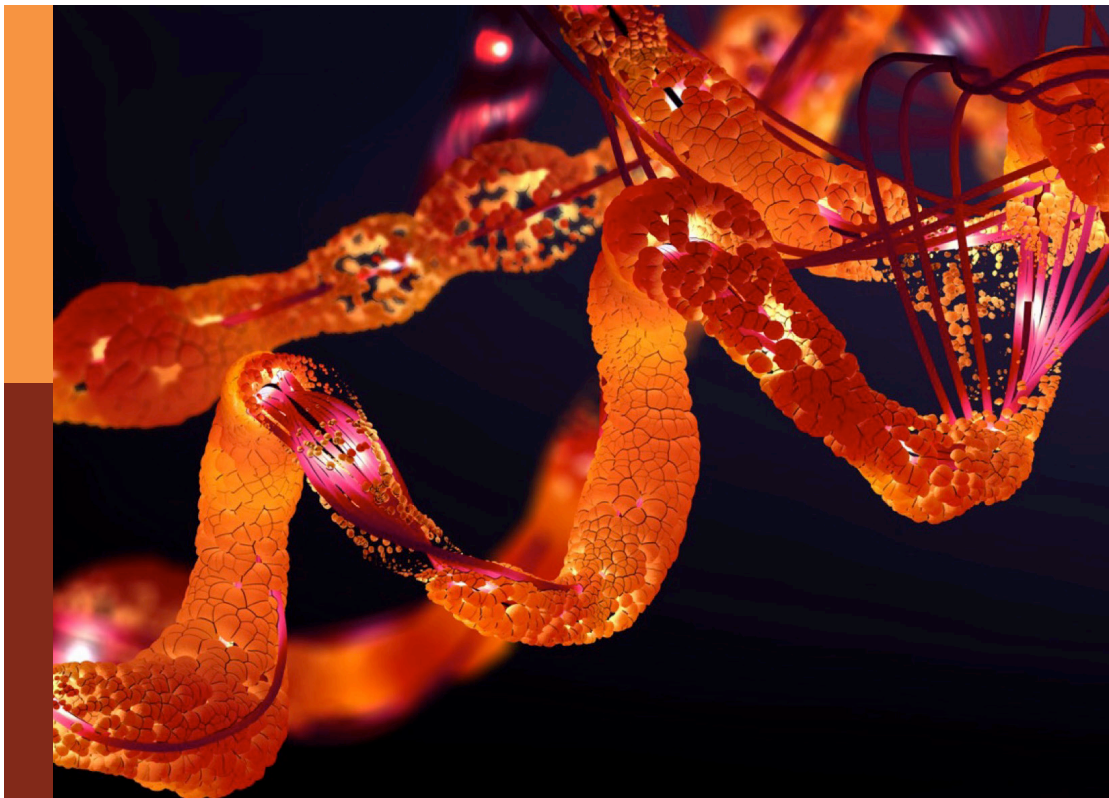
frontiers | Research Topics

Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research

Edited by
Andras Szeitz, Konstantinos Andreas Kouremenos, Steven Hallam and Jens Herbig

Coordinated by
Shane Peter Fitzgerald

Published in
Frontiers in Molecular Biosciences



ISSN 1664-8714
ISBN 978-2-8325-5952-9
DOI 10.3389/978-2-8325-5952-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research

Topic editors

Andras Szeitz — University of British Columbia, Canada
Konstantinos Andreas Kouremenos — NutriPATH Integrative and Functional Pathology Services, Australia
Steven Hallam — University of British Columbia, Canada
Jens Herbig — Ionicon Analytik, Austria

Topic coordinator

Shane Peter Fitzgerald — Imperial College London, United Kingdom

Citation

Szeitz, A., Kouremenos, K. A., Hallam, S., Herbig, J., Fitzgerald, S. P., eds. (2025). *Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5952-9

Table of contents

frontiers | Frontiers in Molecular Biosciences

TYPE Editorial
PUBLISHED 17 January 2025
DOI 10.3389/fmolb.2024.1545016

OPEN ACCESS

EDITED AND REVIEWED BY
Wolfram Weckwerth,
University of Vienna, Austria

*CORRESPONDENCE
Steven J. Hallam,
shallam@mail.ubc.ca

RECEIVED 13 December 2024
ACCEPTED 17 December 2024
PUBLISHED 17 January 2025

CITATION
Szeitz A, Herbig J, Kouremenos KA, Fitzgerald S and Hallam SJ (2025) Editorial: Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research.
Front. Mol. Biosci. 11:1545016.
doi: 10.3389/fmolb.2024.1545016

Editorial: Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research

Andras Szeitz[1,2,3], Jens Herbig[4], Konstantinos A. Kouremenos[5], Shane Fitzgerald[6] and Steven J. Hallam[1,2,3,7,8,9,10]*

[1]Genome Science and Technology Program, University of British Columbia, Vancouver, BC, Canada, [2]Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada, [3]Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada, [4]IONICON Analytik Gesellschaft m.b.H., Innsbruck, Austria, [5]NutriPATH Integrative and Functional Pathology Services, Melbourne, VIC, Australia, [6]Insight Science Foundation Ireland Research Centre for Data Analytics, National Centre for Sensor Research, School of Chemical Sciences, Dublin City University, Dublin 9, Ireland, [7]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada, [8]Department of Biochemistry, Chulalongkorn University, Bangkok, Thailand, [9]Bradshaw Research Institute for Minerals and Mining (BRIMM), University of British Columbia, Vancouver, BC, Canada, [10]ECOSCOPE Training Program, University of British Columbia, Vancouver, BC, Canada

KEYWORDS

volatile organic compounds (VOCs), biomarkers, omics, mass spectrometry, historical review

Editorial on the Research Topic
Metabolomics and transcriptomics in biomarker discovery: mass spectrometric techniques in volatilome research

The central dogma of molecular biology traces a path from genomic sequence information represented in DNA to transcribed messenger RNA decoded at the ribosomal level into proteins that perform a variety of catalytic and regulatory roles within cells (Crick, 1970). The combined activity of these proteins gives rise to pools and fluxes of metabolites within emergent metabolic networks (Durek and Walther, 2008). This genotype to phenotype continuum remains both a fundamental organizing principle for the life sciences and a compelling challenge for systems level thinking in the 21st century (Benfey and Mitchell-Olds, 2008; Costanzo et al., 2019). While our capacity to predict and annotate protein coding potential within genomes and transcriptomes has grown exponentially since the initial introduction of the chain termination method by Sanger and colleagues more than five decade ago (Sanger et al., 1977), it has proven more difficult to constrain the diversity of metabolites over time and space or in relation to health and disease states (Sharon et al., 2014; Sieber and Spradling, 2017).

Nevertheless, researchers have pressed on to profile metabolite pools and fluxes not only within organismal systems, but also within natural and engineered environments, integrating multi-omic (DNA, RNA, protein, and metabolite) data sets to reconstruct emergent metabolic processes (Fondi and Liò, 2015). This has led to the development of new analytical platforms and

methods as well as advanced computational workflows enabling more holistic understanding of metabolic expression at the individual, population, and community levels of biological organization. Nowhere is this more evident than in the rapid development of volatilome research, an extension of analytical chemistry and multi-omics integration focused on detection and quantification of volatile organic compounds (VOCs) produced by interacting cells (Mansurova et al., 2018). This current Research Topic "*Metabolomics and Transcriptomics in Biomarker Discovery: Mass Spectrometric Techniques in Volatilome Research*" presents the readers with a compilation of 10 original research articles and two contemporary reviews exploring advances in metabolomics and VOC detection. The contributions range from human medical applications related to biomarker discovery to biotechnology applications in fungi, plants and microbial cell systems.

Alodiab et al. used a metabolomics approach on dried blood samples to identify two potential biomarkers for galactosemia (GAL), a human genetic disorder that can cause life-threatening side effects related to defective glucose metabolism. Similarly, Sakanaka et al. performed metabolic profiling of saliva and plasma samples and identified a significant association between salivary allantoin and 1,5-anhydroglucitol (1,5-AG) useful in the development of a non-invasive protocol to diagnose atherosclerosis in patients with type 2 diabetes (T2D). Zhou et al. used a higher resolution method to investigate metabolite profiles associated with the treatment of cerebral ischemic stroke with *Danshen Chuanxiongqin* (DSCXQ) preparations, a traditional Chinese medicine. They identified 55 distinct metabolites involved in sphingolipid metabolism between treatment groups that were differentially impacted by DSCXQ consistent with a dampening mode of action on the neuroinflammatory response. Zhang et al. used a similar high-resolution method to profile metabolites associated with clonorchiasis (a parasitic infection caused by the Chinese liver fluke *Clonorchis sinensis*) using a rodent model system of infection. Phan and Blank investigated the fungus *Ustilago maydis* using stable-isotope labeling methods to profile metabolites produced in response to different carbon sources. They compared a range of sample preparation and processing methods to define a process for absolute quantitation in the context of bioprocess development in *Ustilago maydis*.

From the vantage of VOC detection, Furuhashi et al. devised a sampling procedure to extract VOCs from the headspace of lung cell cultures to better understand the interplay between metabolite production and cancer. They identified evidence for increased lipid peroxidation as measured in the increase of trans-2-hexenol and correlated this with changes in the levels of alcohol dehydrogenase 1C gene linking metabolite production to gene expression in cancer cell lines. Fenn et al. used a similar approach to explore the interplay between common respiratory pathogens *Staphylococcus aureus* and *Pseudomonas aeruginosa* grown in isolation or in co-culture with human lung cells. They detected compounds associated with bacterial infection as well as changes in the levels of selected VOCs based on culture conditions relevant to human *in vivo* environments. Issitt et al. measured VOCs produced by breast cancer cell lines exposed to low oxygen conditions representative of the tumor microenvironment to identify potential biomarkers in cell hypoxia. They measured increased uptake of methyl chloride, acetone and n-Hexane from hypoxic cells and concomitant production of styrene, describing volatilomic mechanisms behind this cellular condition for the first time. Expanding on the biomarker theme, Weber et al. used breath analysis combined with a machine learning approach to differentiate between the VOC profiles of children with and without asthma. They were able to discern a wide range of metabolites in the exhaled breath of asthmatic children linked to metabolic processes associated with chronic disease. Farneti et al. used a high-throughput method for rapid phenotyping of raspberry fruit maturation in relation to VOC production to assess fruit quality and to identify cultivars with optimal aroma profiles for future breeding programs.

Finally, two reviews explore methods and applications of volatilome research through a contemporary exploration of the relevant literature. Bajo-Fernández et al. consider the expanding use of exhaled breath analysis as an exciting area of clinical research with the potential to define a new era of non-invasive diagnostic testing. They highlight the need for improved standards of practice related to sample processing, VOC detection, and analysis required to achieve this vision. Szeitz et al. take a more historical approach to mass spectrometry platform innovation in the context of VOC profiling, shining a spotlight on the interrelationship between technological advances and the expansion of volatilome research into increasingly complex matrices. The resulting platform descriptions provide a useful guide intended to assist practitioners and potential end-users performing VOC analysis with advanced instrument performance. Collectively, contributions to this Research Topic capture the current state of the art in metabolite and VOC profiling, reinforcing the potential of metabolomic methods and applications in biomarker discovery and biotechnology innovation. Importantly, this is a rapidly advancing area of research that brings together the peer efforts of scientists and engineers to develop integrated platforms and workflows needed to improve mass resolution and throughput. Effective uptake of these advances will require coordinated and dedicated access to infrastructure and education across multiple training levels and disciplines.

Author contributions

AS: Conceptualization, Investigation, Methodology, Writing–original draft, Writing–review and editing. JH: Methodology, Writing–original draft, Writing–review and editing. KK: Methodology, Writing–original draft, Writing–review and editing. SF: Methodology, Writing–original draft, Writing–review and editing. SH: Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. SH received support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, Genome British Columbia, the Terry Fox New Frontiers Program, and the Canada Foundation for Innovation (CFI).

Conflict of interest

Author JH was employed by company IONICON Analytik Gesellschaft m.b.H. Author SH is a co-founder of Koonkie Inc., a bioinformatics consulting company that designs and provides scalable algorithmic and data analytics solutions in the cloud.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Benfey, P. N., and Mitchell-Olds, T. (2008). From genotype to phenotype: systems biology meets natural vaviation. *Science* 320(5875), 495–497. doi:10.1126/science.1153716

Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C., et al. (2019). Global genetic networks and the genotype-to-phenotype relationship. *Cell.* 177 (1), 85–100. doi:10.1016/j.cell.2019.01.033

Crick, F. (1970). Central dogma of molecular biology. *Nature* 227 (5258), 561–563. doi:10.1038/227561a0

Durek, P., and Walther, D. (2008). The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Syst. Biol.* 2 (1), 100. doi:10.1186/1752-0509-2-100

Fondi, M., and Liò, P. (2015). Multi -omics and metabolic modelling pipelines: challenges and tools for systems microbiology. *Microbiol. Res.* 171, 52–64. doi:10.1016/j.micres.2015.01.003

Mansurova, M., Ebert, B. E., Blank, L. M., and Ibáñez, A. J. (2018). A breath of information: the volatilome. *Curr. Genet.* 64 (4), 959–964. doi:10.1007/s00294-017-0800-x

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 74(12), 5463–5467. doi:10.1073/pnas.74.12.5463

Sharon, G., Garg, N., Debelius, J., Knight, R., Dorrestein, P. C., and Mazmanian, S. K. (2014). Specialized metabolites from the microbiome in health and disease. *Cell. Metab.* 20 (5), 719–730. doi:10.1016/j.cmet.2014.10.016

Sieber, M. H., and Spradling, A. C. (2017). The role of metabolic states in development and disease. *Curr. Opin. Genet. Dev.* 45, 58–68. doi:10.1016/j.gde.2017.03.002

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 30 June 2020
doi: 10.3389/fmolb.2020.00116

Elucidation of Biochemical Pathways Underlying VOCs Production in A549 Cells

Takeshi Furuhashi[1]*, *Ryuga Ishii*[1], *Haruka Onishi*[1] and *Shigenori Ota*[2]

[1] Anicom Specialty Medical Institute Inc., Tokyo, Japan, [2] GL Sciences Inc., Saitama, Japan

Cellular volatile organic compounds (VOCs) are unique compounds whose metabolic pathways remain enigmatic. To elucidate their metabolism, we investigated the VOCs of lung cancer A549 and 2 non-cancer lung cells (HLB; HBEpC). Neutral sugars and lactate in the medium were measured by colorimetric assay. VOCs were enriched by monotrap and profiled by GC-MS. To investigate the enzymes that change VOC metabolism in cells, we conducted ALDH activity assays and qPCR. ROS (reactive oxygen species) assays were conducted to assess oxidation stress. The colorimetric assay showed that especially A549 and HLB took up sugars from the medium and rapidly secreted lactate into the medium. The VOC profile (GC-MS) revealed a *trans*-2-hexenol increase, especially in A549 lung cancer cells. This is a novel lipid peroxidation product from animal cells. Based on the absolute quantification data, *trans*-2-hexenol increased in parallel with number of A549 cancer cells incubated. The qPCR data implies that ADH1c potentially plays an important role in the conversion into *trans*-2-hexenol.

Keywords: cellular volatile, VOC (volatile organic compounds), lipid peroxidation, GC-MS, lung cell

OPEN ACCESS

Edited by:
Martin Giera, Leiden University Medical Center, Netherlands

Reviewed by:
Jean-Marie Galano, UMR5247 Institut des Biomolécules Max Mousseron (IBMM), France
Wojciech Andrzej Filipiak, Nicolaus Copernicus University in Torun, Poland

***Correspondence:**
Takeshi Furuhashi
takeshi.furuhashi@ani-com.com

Specialty section:
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: *15 March 2020*
Accepted: *19 May 2020*
Published: *30 June 2020*

Citation:
Furuhashi T, Ishii R, Onishi H and Ota S (2020) Elucidation of Biochemical Pathways Underlying VOCs Production in A549 Cells. Front. Mol. Biosci. 7:116. doi: 10.3389/fmolb.2020.00116

INTRODUCTION

The potential relationship between VOCs (volatile organic compounds) emitted by the human body and malignant diseases (e.g., cancer) has been investigated since Hippocrates' time (Di Francesco et al., 2005). Recent technological advances have provided new knowledge about the metabolic changes related to VOC production in diseases, and VOC profiles enable comparison between normal and disease conditions (Hakim et al., 2012).

Nonetheless, a gap remains between disease detection and volatile compound analysis. One reason is that the relationship between pathology and the metabolic pathways producing VOC is not fully understood. The mechanism of biological VOC synthesis in cells is enigmatic because many studies on such metabolites are restricted to bacteria, algae, and fungi (Ruffing, 2013). Moreover, the enzymes associated with profiled cell VOCs have not always been identified (e.g., hydrocarbons, alcohols, alkenes). The goal of our study is to elucidate the VOC pathway which can potentially be marker for cancer. VOC analysis by GC-MS profiles (Buszewski et al., 2012) is certainly useful for assessing VOCs as biomarkers, but there is some room for improvement in VOC sampling.

In fact, direct gas injection or headspace analysis (Snow and Slack, 2002) normally show lower VOC values than methods using fiber-based VOC sampling and enrichment. Two types of fiber-based sampling are currently available. One is a fiber attached to a syringe which is exposed to gas by incubation (typical for SPME; solid phase micro extraction). The other is a naked fiber shape (typical for monotrap). In the case of SPME, only the fiber surface can retain VOCs, whereas

in monotrap the VOCs are retained both by the surface and the inner porous part. This opens the door for improving the intensity of VOC detection as well as for introducing disposable VOC-capturing materials.

Monotrap is a disposable sorptive media based on the large surface area of silica monolith. It has been applied in biological research and can be preserved in the sample bottle prior to measurement (Jang et al., 2011; Ma et al., 2013). Nonetheless, despite the large surface for capturing VOCs, monotrap has been used solely for static sampling. This means that improving the preconcentration technique (e.g., NTME; needle trap extraction and ITEX; in tube extraction) requires shorter incubation time but generates better intensity (Mochalski et al., 2018) and should be applicable to single cell studies as well (Serasanambati et al., 2019). For this reason, we developed an active sampling approach for monotrap-based sample preparation which is suitable for cellular VOC analysis.

Moreover, solely a VOC profile is insufficient to elucidate a synthetic pathway. To date, a VOC profile has not been combined with any enzyme assay. In this study, we conducted enzymatic assays to find key enzymes related to marker VOC production.

MATERIALS AND METHODS

In this study, we used A549 (human lung adenocarcinoma), human lung fibroblasts (HLB) cells, and human bronchial epithelial cells (HBEpC) as primary cells. This is because lung cancer is one of the most lethal cancers and also because some VOC analyses have already been reported (de Lacy Costello et al., 2014; Lemjabbar-Alaoui et al., 2015). Cell cultures are normally done in 2-dimensional (2D) form, but this sometimes does not reflect actual physiology (Kapałczynska et al., 2018). Accordingly, we compare flasks with non-adherent plates, which can mimic 3D culture systems.

Based on the VOC profile results, we focus on *trans*-2-hexenol metabolism. ALDH (aldehyde dehydrogenase) assays were applied to assess the conversion of aldehyde groups into carboxylic acid groups. To investigate the speed of sugar uptake (as the main energy source) and to evaluate the glycolysis pathway, we conducted phenol sulfuric acid assays for the medium assay to determine sugar consumption from the medium. Lactate secreted into the medium was calculated using a lactate assay kit; this reflects the balance of reduction and oxidation inside the cell. qPCR (quantitative polymerase chain reaction) was applied to determine whether certain enzymes (i.e., generation of *trans*-2-hexenol) are related to the VOC profile or not. ROS (reactive oxygen species) assays were done to assess whether cells are under oxidation stress.

Abbreviations: A549, Human lung adenocarcinoma; AAR, Acyl ACP reductase; ADC, Aldehyde decarbonylase; ADH, Alcohol dehydrogenase; ALA, Alpha linoleic acid; ALDH, Aldehyde dehydrogenase; ALR, Aldehyde reductase; ATP, Adenosine triphosphate; DMEM, Dulbecco's modified eagle medium; ECM, Extra cellular matrixes; EIC, Extracted ion counting; EI-MS, Electron ionization; FAR, Fatty acyl-CoA reductase; FCS, Fetal bovine serum; GC, Gas chromatography; GAPDH, Glyceraldehyde-3-phosphate dehydrogenase; HBEpC, Human bronchial epithelial cells as primary cells; HLB, Human lung fibroblasts cells; HPL, Hydroperoxide lyase; IR, Infrared spectroscopy; ITEX, In tube extraction; LOX, Lipoxygenase; MS, Mass spectrometry; NAD, Nicotinamide adenine dinucleotide; NTME, Needle trap extraction; PBS, Phosphate-buffered saline; PCA, Principle components analysis; PTR, Proton transfer reaction; PUFA, Polyunsaturated fatty acid; qPCR, Quantitative polymerase chain reaction; ROS, Reactive oxygen species; SCFA, Short chain fatty acid; SIFT, Selected ion flow tube; SPME, Solid phase micro extraction; TD, Thermal desorption; VOC, Volatile organic compounds.

Cell Culture Condition

A549 cells and human lung fibroblasts (HLB) cells were purchased from the American Type Culture Collection. Immortalization was done to HLB. A549 cells were originally isolated from a lung carcinoma of a 58-y-old man, showed epithelial morphology and grew adherent. All cells were grown in DMEM (Dulbecco's modified eagle medium) high-glucose culture medium containing sodium pyruvate (110 mg/L) supplemented with 10% FCS (fetal bovine serum), penicillin (100,000 units/L), streptomycin (100 mg/L), and L-glutamine (293 mg/L). Human bronchial epithelial cells (HBEpC) are primary cells (PromoCell GmbH) isolated from the mucosa of the main bronchi of a 42-y-old male Caucasian. The cells were cultivated in Airway Epithelial Cell Growth Medium (PromoCell GmbH) supplemented with the Airway Epithelial Cell Growth Medium Supplement Pack (PromoCell GmbH) according to the manufacturer's instructions. A T75 cell culture flask (250 mL, 75 cm^2) was used for the culture. For 2-dimensional (2D) and 3-dimensional (3D) cell culture, flasks with a red screw cap (CellStar 618175) and cell-repellent surface flasks with white screw caps (CellStar 658985), respectively, were used. Cultured cells were observed, and photographs were taken using an OLYMPUS IX71 microscope (OLYMPUS) and software AdvanView 3.7 (AdvanVision Co, Ltd).

For all experiments, cells were cultivated under standard conditions at 37°C in a humidified atmosphere with 92.5% air/7.5% CO_2. All cells (1×10^6) were inoculated in 20 mL phenol red-free DMEM high-glucose medium (supplements: 5% FCS, 100,000 units/L penicillin, 100 mg/L streptomycin, 293 mg/L L-glutamine, and 110 mg/L sodium pyruvate). After Days 3 and 4, cells were collected by digestion with 0.25% Trypsin-EDTA (Thermo Fisher) for 5 min at 37°C. After trypsinization, medium (DMEM containing 10% FBS) was added and mixed by pipetting cells. Cells were collected by centrifugation and resuspended in fresh medium. Cell counts were performed on a Cell Counter model R1 (OLYMPUS, USA).

The FCS concentration in DMEM during the experiment was lowered to 5% to reduce the high background of VOCs in the analyzed headspace. Cells were grown in these conditions for 96 h, up to a confluence of 50–60% (around 1.0×10^6 cells/flask). To quantify sugar and lactate in the medium, Days 1, 2, 3, and 4 were sampled. The same days were sampled for the cellular ALDH activity assay and qPCR. Cells were transferred into a new flask 4 days after inoculation. Physiologically, Day 0 cells were the same as Day 4 cells. Considering the accumulation of VOCs inside the cell culture flasks, Days 0, 3, and 4 samples were collected for VOC analysis.

Phenol-Sulfuric Acid Assay

Neutral sugars as cellular energy sources were quantified by the phenol-sulfuric acid assay (modified from a previous study; (DuBois et al., 1956)). At each time point, 5 μL culture medium was taken from the flask and transferred into a 2 mL Eppendorf tube. We then added 195 μL deionized water and 80% 5 μL phenol solution. Thereafter, 500 μL of concentrated sulfuric acid was added by titration (on ice). The solution was mixed by inverting the tube and incubated for 25 min at room temperature. We then transferred 150 μL into 96-well plates and measured absorbency at 490 nm using Enspire (Perkin-Elmer, USA). Each measurement was biologically replicated 4 times. Quantification involved using a linear calibration curve with glucose standard solution (0.5–100 μg).

Lactate Assay

A Lactic acid Assay Kit (10139084035, R-Biopharm AG, Germany) was used to quantify lactate in the medium. At each time point, 1 μL culture medium was taken from the flask and 99 μL deionized water was added. The solution was transferred to 96-well plates, and 100 μL glycylglycine buffer (440 mg/30 mL, pH 10) and 20 μL NAD (nicotinamide adenine dinucleotide) solution (210 mg/6 mL) were added. After that, 2 μL glutamate-pyruvate transaminase suspension (1,100 U) was added. The solution was incubated at room temperature for 5 min. Absorbency at 340 nm was measured by Enspire (A_1). Then, 2 μL lactate dehydrogenase solution (3,800 U) was added and incubated at room temperature for 30 min. Absorbency at 340 nm was measured by Enspire (A_2). The lactate concentration was calculated as A_2-A_1. Each measurement was biologically replicated 4 times. A linear calibration curve with lactate standard solution (10–2,000 ng) was used for quantification.

Monotrap VOC Enrichment-Thermal Desorption (TD)

In monotrap, the mass transfer can be very slow due to a thicker membrane than in SPME. This requires an alternative agitation/convection strategy especially designed for monotrap. In this study, we tested pumping with monotrap, which is active sampling to enrich VOC by continuously pumping air. This enables capturing large volumes of VOCs within 30 min. Graphite and C18 type monotrap were used for this study. At each incubation period (Days 0, 3, and 4), VOCs were collected with a special penetrating vent cap (modified after a previous study; Schallschmidt et al., 2015). Each measurement was done in triplicate. For VOC quantification with different numbers of cells and to determine a limit of detection, a calibration curve was prepared using reference compounds. Three VOCs (i.e., *trans*-2-hexenol, tetradecane, isobutyrate) were chosen from VOC GC-MS profile data. Reference compounds for absolute quantification were purchased from SIGMA and TCI company (*trans*-2-hexenol, SIGMA 132667; tetradecane, TCI T0079; and isobutyrate, SIGMA l1754). For preparing the calibration curve, reference compounds were added into 10 mL deionized water. The solution was put into the cell culture flask, and subsequently the volatilized compounds were trapped by monotrap with 30 min pumping and measured by GC-MS. For comparison, 1 $\times$ 10^6, 2 $\times$ 10^6 A549 cells and cell-less media as negative control were compared. The incubation period was Days 0 and 3.

The experimental set-up used for the cell culture headspace analysis is shown in **Figure S1**. The control medium was obtained by incubating DMEM culture medium in the same conditions as the cell samples, but without seeded cells.

A syringe was connected to the cell culture flask by a penetrating screw cap hole. Two monotraps, graphite type (MonoTrap GC TD, Cat.No. 1050, GL Sciences, Tokyo) and C18 type (MonoTrap RSC18 TD, Cat.No. 1050-73201, GL Sciences, Tokyo) were put into the syringe (TERMO 2.5 mL, SS-02SZ). Pumping (∼1,500 times up and down in total) was repeated for 30 min. After 30 min pumping, monotraps (both graphite type and C18 type) were removed and put tandemly into a Handy-TD glass liner (MonoTrap TD Liner for Handy-TD Cat.No. 1003-75005). As we used a T75 flask (270 mL volume containing 10 mL aqueous solution) and pumped using a 2.5 mL syringe, the total head space volume was 270 + 2.5 – 10 = 262.5 mL.

In preliminary experiments, we determined that 20–30 min was sufficient for the monotrap to capture gases. The liner was inserted into the Handy-TD.

GC-MS Conditions

GC-MS measurements were carried out on a single quadrupole mass spectrometer (5977B-MSD; Agilent Technologies, Santa Clara, CA, USA) equipped with 7890B GC (Agilent Technologies, Santa Clara, CA, USA) and Handy-TD265 (GL science, Iruma city, Japan). The conditions were modified from a previous study (Furuhashi et al., 2018). The initial Handy-TD condition was a constant purging flow of 5 mL/min of helium gas at 40°C for 0.1 min. Predesorb was 70 kPa. For desorb, the ramp rate was 45°C/s, with 1.5 min holding at 250°C. GC-MS measurement was started directly after Handy-TD desorb.

The temperature of the liner was 230°C. The GC column helium gas flow was set at 1 mL/min constant rate. DB-Wax UI 30 m, 0.25 mm, 0.25 μm (122-7032 UI, Agilent, USA) was used as a GC column. The oven temperature gradient for the samples was as follows. After a 5 min, 50°C isothermic period, the oven was programmed to rise to 150°C at a rate of 5°C min^{-1}, then rise to 260°C at a rate of 40°C min^{-1}, held at 260°C for 1 min. The temperature of both the GC-MS ion source and transfer line was set at 250°C. The scan range was between m/z 30 and 600. For quantification, we calculated the peak area of a conventional 70 eV EI mode (Extractor ion source; Agilent Technologies, Santa Clara, CA, USA) extracted ion chromatogram using software (Mass Hunter; Agilent Technologies, Santa Clara, CA, USA). The chosen m/z and retention time for quantification are described in **Table S1**. The sample was measured with split mode. Cellular VOCs were measured with a split ratio of 3. Between batch measurements, C17:0 FAME was measured to check the intensity between different batches as well as the retention time shift.

qPCR Condition

At each time point, culture medium was removed by centrifuging (3,000 g for 5 min). Cells were washed with PBS (phosphate-buffered saline). For 2D culture cells, cells were detached from

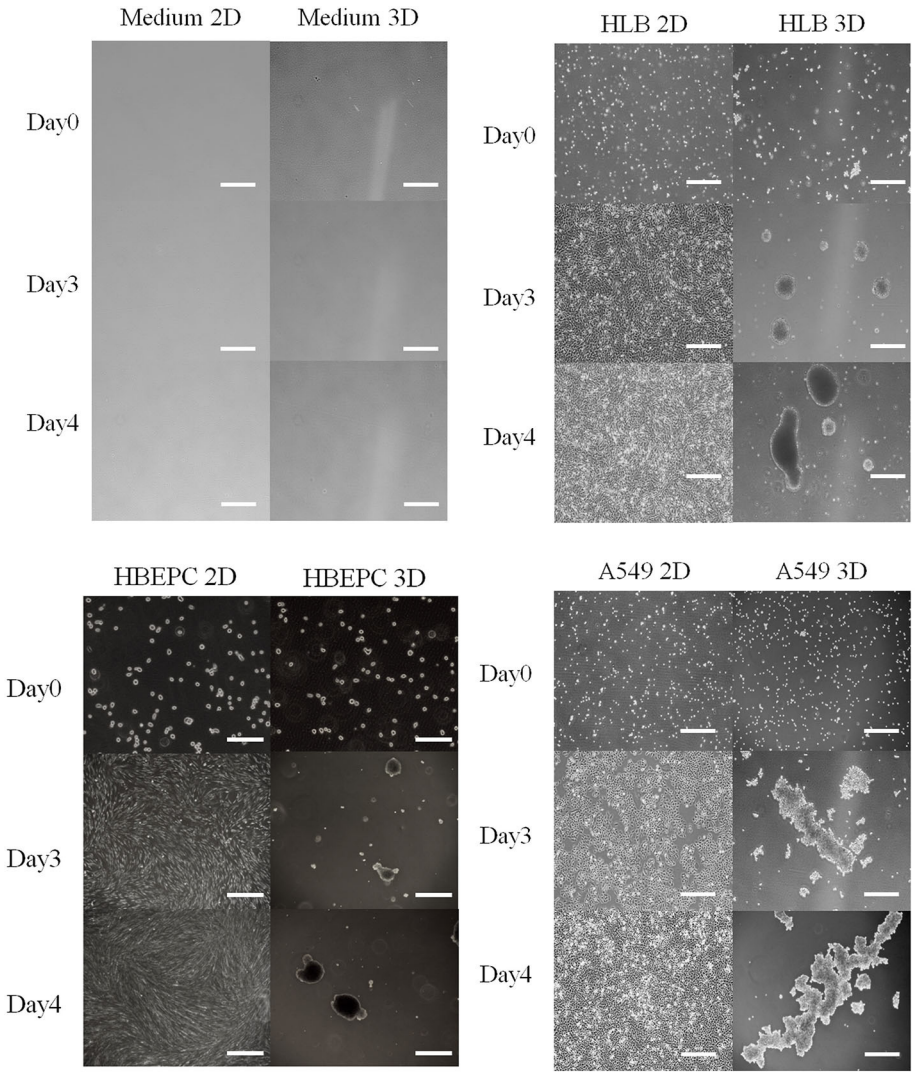


FIGURE 1 | Photos of medium as negative control, HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (Human lung adenocarcinoma) from Days 0 to 4 in both 2D and 3D culture. All cell types in 2D culture spread across the bottom of the flask, whereas those in 3D culture aggregated. Photos were taken at 0, 3, 4 days after inoculation. Scale bar: 500 μm except for HBEpC at Day 0 (200 μm).

the flask by scratching with a scraper. The cell suspension was transferred into 1 mL Eppendorf tubes and centrifuged at 200 g for 5 min. The supernatant was aspirated.

Total RNA was extracted from cells using TRIzol Reagent (Cat: 15596026, Thermo Fisher Scientific, USA) according to the manufacturer's protocol. mRNA was reverse transcribed using an oligo (dT) primer and SuperScript III Reverse Transcriptase (Cat: 18080400, Invitrogen, USA) (Ishii et al., 2012). For the thermal cycle reaction, the cDNA template was amplified by the thermal cycler LightCycler 96 System (Roche Diagnostics, USA) using the FastStart Universal SYBR Green Master (Rox) (Cat: 04 913 914 001, Roche Diagnostics, USA) under the following reaction conditions: 40 cycles of PCR (95°C for 10 s, and 60°C for 1 min) after an initial denaturation (95°C for 10 min). Fluorescence was monitored during every PCR cycle at the annealing step. The authenticity and size of the PCR products were confirmed using a melting curve analysis with the software LightCycler 96 System version SW1.1 (Roche Diagnostics, USA). mRNA levels were normalized using Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a housekeeping gene. Primer sequences are provided in **Table S2**.

ALDH Assay

The ALDH Assay Kit (K731-100, BioVision, USA) was used to quantify cellular ALDH activity. At each time point, culture medium was removed by centrifuging (3000 g for 5 min). Cells were washed with PBS. For 2D culture cells, cells were detached from the flask with a scraper. The cell suspension was transferred into 1 mL Eppendorf tubes and centrifuged at 200 g for 5 min. The supernatant was aspirated and 200 μL ALDH assay buffer

was added. The cell suspension was homogenized by vortex and incubated on ice for 10 min. The suspension was centrifuged at 20,000 g for 3 min and the supernatant transferred to a new tube. The supernatant solution was incubated at room temperature for 10 min. Absorbency at 450 nm was measured by Enspire at 0, 20, 40, 60 min. As enzymatic activity corresponded to NADH generation during the reaction time, the slope of absorbency at 450 nm and time (X axis) was calculated. The obtained slope was normalized with cell number (1×10^5 cells) and then multiplied times 10^4. Each measurement was biologically replicated 4 times. The cell number for the normalization of ALDH activity was determined with an OLYMPUS cell counter model R1 (OLYMPUS).

ROS Assay

The Cellular ROS assay kit (Red) (ab186027; Abcam, Cambridge, UK) was used. The protocol was modified from the ab186027 protocol. 2D or 3D cultured cells (HLB/HBEpC/A549) were prepared (2D or 3D culture for 1–4 days incubation). After trypsin treatment, cells were washed with PBS two times. The cell suspension was prepared as 5×10^4 in 100 μL PBS. After pipetting the cell suspension 3 times, 50 μL suspension was transferred into a microplate (655077, 96 well, F-bottom, black, Fluotrac; Greiner Bio-one, Germany). We added 50 μL ROS red working solution (2 μL ROS red stock solution in 1 mL ROS assay buffer) into cell suspension. We then incubated the microplate at 37°C with 5% CO_2 for 1 h. After incubation, fluorescence was measured as Ext 520 nm Ems 605 nm using Enspire (Perkin-Elmer, USA). Each measurement was biologically replicated 4 times.

RESULTS AND DISCUSSION

All 2D-cultured cells proliferated and spread on the bottom of the flask 3 days after inoculation, and they nearly overflowed 4 days after inoculation (**Figure 1**). The 3D-cultured cells proliferated with aggregation (**Figure 1**). Data obtained by the conventional phenol-sulfuric acid assay indicated that the neutral sugar concentration of all cell culture media gradually decreased over time in the 2D cell culture (**Figure 2A**). The tendency was similar in 3D cell cultures (**Figure 2B**) in HLB and A549. In the HBEpC cell culture medium, the neutral sugar concentration did not decrease significantly (*p*-value of 0–4 days after inoculation: 0.2). Both 2D and 3D culture medium without cells showed no decrease in medium sugar concentration.

Lactate secreted into the medium gradually increased with time in all 2D-cultured cells, (**Figure 2C**) and 3D-cultured cells, except that the lactate increase of HBEpC was moderate (**Figure 2D**). No change was recorded in 2D and 3D culture media without cells. Previous reports suggest that tumor cells under heterogeneous microenvironments can enter into a mutual relationship via lactate transport (Sonveaux et al., 2008). Our lactate assay data showed that both 2D- and 3D-cultured cancer cells secreted lactate into the medium, although the secretion speed in the latter was slower than in the former. Interestingly, compared with other cancer cells, the sugar uptake speeds from the media were slow only in 3D-cultured HBEpC, and the secretion of lactate into the media remained at a lower level. In general, rapid sugar uptake and lactate secretion were conspicuous in cancer cells (A549) and in normal cells that were immortalization treated (HLB), leading us to conclude that such characteristics are common in immortalized cells and are not cancer specific.

From the 81 VOC reference compounds list for this study (**Table S1**) - including hydrocarbons, aldehydes, alcohols, fatty acids, methyl esters, and aromatic groups −30 VOC peaks were detected in the flask based on the data obtained (**Figures S2–S4**). PCA (principle components analysis) based on these 30 VOCs (**Figures 2E,F**) showed that the VOC profiles of the medium were separated from those of cells both in 2D and 3D culture. After 3 days incubation, the difference between solely medium and cells began to increase. In 3D culture, the separation between cells was clearer than in 2D culture after 3 days incubation.

To investigate the VOCs contributing to the group separation in PCA, statistical analysis (*t*-test; $p < 0.05$) (**Table S3**) revealed an increase of benzaldehyde in the medium and decrease in cells (except A549 2D culture Day 4). This tendency is consistent with a previous study on mesenchymal stromal/stem Cellular VOCs (Klemenz et al., 2019). Tetradecane from all cellular samples increased with time. *trans*-2-hexenol conspicuously increased over time particularly in A549 cells in both 2D and 3D culture. The increase of *trans*-2-hexenol between medium and A549 at Day 4 was also significant (**Tables S3, S4**; **Figure S4**). Isobutyrate increased in A549 and HLB showed statistical significance in A549 3D culture and HLB 2D culture. A 3-methyl pentanoate increase was observed as an HBEpC-specific VOC but did not increase in other cells.

Based on the statistical analysis of VOC profile data, we focus on 10 VOCs (e.g., undecanal, tetradecane, nonanal, 2-ethyl-1-hexanol, isobutyrate, *trans*-2-hexenol, tridecane, toluene, acetate, undecane). These increased especially in 2D- or 3D-cultured A549 cancer cells at Day 4. Among these, undecanal, nonanal, 2-ethyl-1-hexanol, acetate, and tridecane showed statistical significance in A549 2D culture. In contrast, the increase was inconspicuous in the cell-less media. The increase of toluene and undecane in A549 Day 4 was statistically significant compared to cell-less media at Day 4, but these two compounds then decreased over time in the A549 incubation. *Trans*-2-hexenol, tetradecane, isobutyrate were positively correlated in cellular samples, showing that these VOCs increased together.

Accordingly, we selected *trans*-2-hexenol, tetradecane, and isobutyrate for quantification. The limits of detection were firstly determined as 5 nmole (*trans*-2-hexenol), 10 nmole (tetradecane), and 50 nmole (isobutyrate) in cell culture flasks with 10 mL aqueous solution. Linearity as a standard curve was valid between 5–500 nmole/10 mL (*trans*-2-hexenol), 20–1,000 nmole/10 mL (tetradecane), and 50–5,000 nmole/10 mL (isobutyrate) (**Figure S5**).

To confirm whether VOCs were derived from cancer cells or not, we compared 1×10^6 and 2×10^6 A549 cells. Both concentrations showed increased *trans*-2-hexenol. The difference was significant between Days 0 and 3 both in 1×10^6 A549

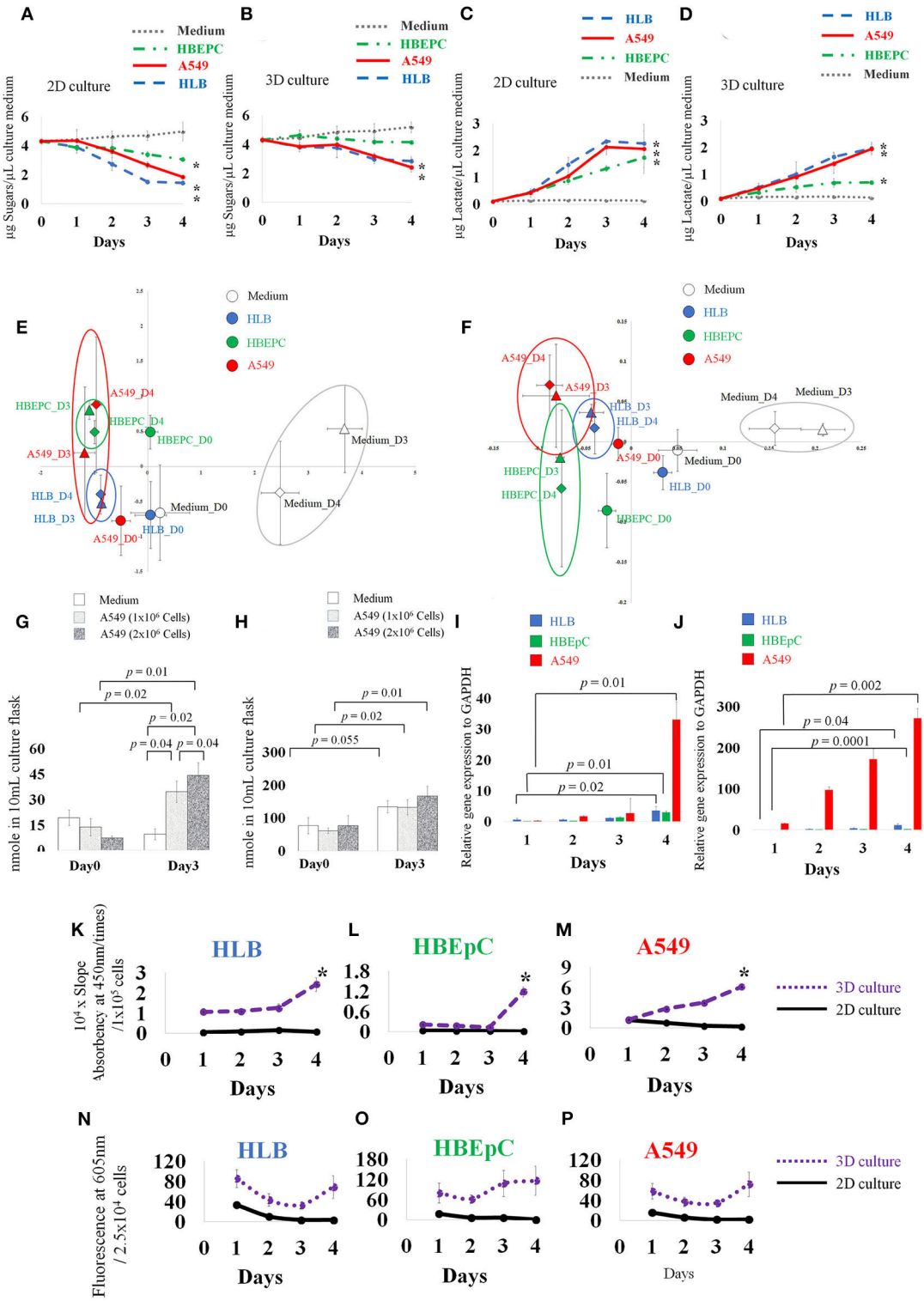


FIGURE 2 | Concentration of medium sugars **(A,B)**, and lactate **(C,D)**. Principle component analysis (PCA) based on 30 VOC GC-MS data of 2D and 3D cultures **(E,F)**, and absolute quantification of *trans*-2-hexenol and isobutyrate in 2D-cultured cell-less media and 1 × 10^6 and 2 × 10^6 A549 cells (human lung adenocarcinoma) A549 cells **(G,H)**. qPCR data **(I,J)**, ALDH (aldehyde dehydrogenase) activity assay over time **(K–M)**, and ROS data **(N–P)**. HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (human lung adenocarcinoma) were compared. Note different scales on y-axis. **(A)** Neutral sugar concentration in 2D-cultured medium. **(B)** Neutral sugar concentration in 3D-cultured medium; Y axis indicates μg Sugars/μL culture medium,

(Continued)

FIGURE 2 | error bars indicate standard deviation. **(C)** Lactate concentration in 2D-cultured medium. **(D)** Lactate concentration in 3D-cultured medium. Y axis indicates μg Lactate/μL culture medium, error bars indicate standard deviation. Medium without cells, HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (Human lung adenocarcinoma) were compared. For the sugar and lactate assays, sampling was done at 0, 1, 2, 3, 4 days after inoculation. Experiments were biologically replicated four times. Asterisk indicates statistical significance ($p < 0.05$) between Day0 and Day4 of each cellular sample. **(A)** The sugar concentration in 2D culture medium fell on day 4, from 4.3 (±0.2) μg/μL medium to 1.4 (±0.1), 3.0 (±0.1), and 1.8 (±0.1) μg/μL medium in HLB, HBEpC, and A549, respectively. The corresponding *p*-values (0 day to 4 days after inoculation) were 0.0000008, 0.00005, and 0.002. **(B)** Sugar concentration in 3D cell cultures in HLB and A549: values 4 days after inoculation were 2.8 (±0.3) and 2.4 (±0.3) μg/μL medium, and *p*-values compared with 0 day were 0.0002 and 0.00006, respectively. **(C)** Lactate secretion into the medium of 2D-cultured cells, 4 days after inoculation, reached 2.2 (±0.4), 1.7 (±0.3), and 2.0 (±0.9) μg lactate/μL medium in HLB, HBEpC, and A549, respectively; corresponding *p*-values (0 day to 4 days after inoculation) 0.002, 0.002, and 0.01. **(D)** In 3D-cultured cells, the lactate concentrations increased in the medium 4 days after inoculation in HLB, HBEpC, and A549: 1.9 (±0.2), 0.6 (±0.05), and 1.9 (±0.1) μg lactate/μL medium, respectively; corresponding *p*-values (0day to 4days after inoculation) 0.0004, 0.000001, and 0.0001. In PCA, X and Y axis indicate PC1 and PC2, respectively **(E,F)**. Error bars: standard deviation of each group ($n = 3$). Each cell type after 3 days and 4 days incubation is circled. In absolute quantification graph **(G,H)**, Y axis indicates nmole VOC in flask. For quantification, extracted ion counting (EIC) of *trans*-2-hexenol fragment (*m/z* 82) and isobutyrate fragment (*m/z* 73) was used. Sampling was done at 0 and 3 days after inoculation. Experiments were biologically triplicated. Error bars: standard deviation. *p*-value showing statistical significance is listed. qPCR data **(I,J)** of ADH1c (alcohol dehydrogenase) in 2D- and 3D-cultured HLB, HBEpC and A549. Y axis indicates relative gene expression to GAPDH. In qPCR, sampling was done at 1, 2, 3, 4 days after inoculation. Experiments were done in triplicate. Error bars: standard deviation. *p*-value between Day1 and Day4 showed statistical significance. In the ALDH assay **(K–M)** and ROS assay **(N–P)**, experiments were biologically replicated four times. Error bars: standard deviation. Sampling was done at 1, 2, 3, 4 days after inoculation. Y axis in ALDH assay indicates 104 times multiplied slope (absorbency at 450 nm over time) which is normalized by cell number (105 cells). In the ROS analysis, Y axis indicates fluorescence at 605 nm/2.5 × 104 cells.

and 2×10^6 A549 cells (*p*-value 0.02 and 0.01, respectively) and and 2×10^6 A549 cells (*p*-value 0.02 and 0.01, respectively) and between Day3 cell-less media and A549 cells (*p*-value 0.04 and 0.02 against 1×10^6 A549 and 2×10^6 A549 cells, respectively) (**Figure 2G**). The results show that the accumulation of *trans*-2-hexenol increased in parallel with the number of cells incubated (*p*-value between 1×10^6 and 2×10^6 A549 cells at Day 3 was 0.04). Our data show that about 20 nmole *trans*-2-hexenol/1 $\times 10^6$ A549 cells was secreted and accumulated in 3 days. In contrast, *trans*-2-hexenol was not accumulated in cell-less media. Isobutyrate increased significantly from Day 0 to Day 3 in A549 cells (*p*-values between Day 0 and Day 3 were 0.02 and 0.01 for 1×10^6 and 2×10^6 A549 cells, respectively) (**Figure 2H**). In cell-less media, isobutyrate also increased from Day 0 to Day 3 (*p*-value 0.055). For tetradecane, the detected EIC (extracted ion counting) peaks in the samples were below the limit of quantification. Based on the quantification results, *trans*-2-hexenol was chosen as the focus for further investigation.

We also observed a *trans*-2-hexenol increase in other cancer cells (Pancreas cell, PaCa2; and Neural cancer cell, YKG1) (unpublished data), indicating that *trans*-2-hexenol might be a commonly secreted VOC in many cancer cells. HLB, i.e., normal cells genetically modified as an immortalization treatment, took up sugars and secreted lactate to the same degree as cancer cells in both 2D and 3D cultures. At the same time, *trans*-2-hexenol did not increase in HLB, indicating that glucose uptake and proliferation alone do not lead to VOC production.

In considering the biosynthesis of *trans*-2-hexenol, lipid peroxidation and ROS are two important issues to understand. Cancer cells are typically heterogeneous and derived from normal cells by accumulating mutations (Aktipis et al., 2015). Such mutations lead to excess ROS inside the cells, and radical species react with cellular metabolites and genes. For instance, an ROS increase in cells leads to membrane degradation followed by lipid peroxidation. Lipid peroxidation, for example, generates toxic compounds such as aldehyde (Ayala et al., 2014). Aldehydes are highly reactive chemical substances whose internal content cells need to reduce.

There are several pathways to convert aldehyde into other compounds (i.e., alcohol, carboxylic acid and hydrocarbon). In the metabolic pathway of mammals, a conversion of aldehydes into hydrocarbons is unlikely because there are no reports that mammalian cells possess such an enzyme. Accordingly, aldehyde conversion probably mainly involves conversion into alcohol and/or carboxylic acid. Regarding genes related to *trans*-2-hexenol production, no information is available about AAR (Acyl ACP reductase) and ADC (Aldehyde decarbonylase) in the human genome. We therefore focused on enzymes that can generate alcohol, i.e., FAR (Fatty acyl-CoA reductase) and ADH (Alcohol dehydrogenase) or ALR (Aldehyde reductase).

Aldehyde conversion into alcohol by ADH (alcohol dehydrogenase = aldehyde reductase) is a reversible chemical reaction, whereas aldehyde oxidation into carboxylic acid is irreversible (Rizzo, 2014). A previous study on rat ADH reported that class I (i.e., ADH1) actively reduces 2-hexenal *in vitro* (Boleda et al., 1993). This suggests that mammalian cells can reduce aldehyde into alcohol and implies that ADH1 is related with the *trans*-2-hexenol increase.

As *trans*-2-hexenol has not been reported in previous lung cell studies (Filipiak et al., 2010, 2016), we further investigated enzymes to confirm this VOC production. From qPCR data, ADH1c of all cells increased 4 days after inoculation in both 2D and 3D culture, and it increased conspicuously especially in A549 (**Figures 2I,J**). The relative expression level to GAPDH was also much lower than that of A549. The mean values of relative expression to GAPDH 4 days after inoculation were 2.95 and 1.8 in 2D and 3D cultures, respectively. At the same time, the respective values of A549 were 33 and 271. Other ADHs (ADH1A, 1B, ADH3B1, ADH5, ADH9A1) showed no clear upregulation over time in either 2D or 3D cultures (unpublished data). We also examined other enzymes such as ALR2 and FAR1, which are related to alcohol production in cells (**Figure S6**), but there were no significant differences over time or according to cell type (HBEpC and A549). Apparently, ALR2 and FAR1 were not related to alcohol production, and *trans*-2-hexenol is produced mainly by ADH class I; moreover,

trans-2-hexenol does not appear to be derived from fatty acyl CoA (**Figure S7**).

Other strategies are available to remove aldehydes, for instance a conversion into carboxylic acid by ALDH. ALDH (19 isoforms) plays a role in tissue repair and pluripotency (Balber, 2011) and across taxa (from bacteria to humans) (Yoshida et al., 1998). Various ALDH isoforms are expressed in different tissues and can metabolize different substrates. ALDH activity remained unchanged in 2D culture cells, whereas it increased in 3D cultures from 2 to 3 days after initial inoculation (**Figures 2K–M**). The activity of 3D-cultured cells was higher in A549 cells than in the other two normal lung cells. Activity started to increase 2 days after A549 inoculation. This result agreed with a previous report stating aldehyde degradation in A549 cells (Filipiak et al., 2010), as well as a report that shows many aldehydes decreased but alcohols increased in cancer cells (Filipiak et al., 2016). In contrast, in the case of HLB and HBEpC, the activity increase occurred 4 days after inoculation. Based on our ALDH results, the increase of the alcohol group was not related to downregulation of ALDH.

Interestingly, our ROS assay data showed that the A549 cancer cells had the lowest values of all tested cell lines. There was a decrease over time in 2D-cultured cells, whereas those in 3D cultures showed no conspicuous increase or decrease (**Figures 2N–P**). Although ROS must be important for VOC production, actual ROS levels inside cultured cells were not always consistent with *trans*-2-hexenol production levels in our study, which partly reflects difficulties in mimicking the tissue microenvironment (Lu et al., 2007).

In this study, we found *trans*-2-hexenol as a novel animal cellular VOC that appears to be a product of the lipid peroxidation pathway. Our study pointed to a relation solely with ADH1c upregulation. Furthermore, simple rapid sugar uptake and lactate production were not directly related to *trans*-2-hexenol increase. As such, the metabolism of *trans*-2-hexenol and its regulation still harbors some unclear points.

A *trans*-2-hexenol increase was detected in both 2D and 3D A549 cell cultures. Although the cell culture types differed somewhat (e.g., ALDH), these differences did not appear to be critical for VOC production. This study did not utilize ECM (extra cellular matrixes) or scaffold structure for 3D culture, implying that solely a 3-dimensional aggregation of cells did not significantly influence VOC production. Improving the culture techniques might uncover hidden differences between 2D and 3D.

Another remaining important point is origin of the aldehyde which would be converted into alcohol. Our data suggested the presence of *trans*-2-hexenol and implied upregulation of ADH1c, so that the *trans*-2-hexenol could be from aldehyde. Nonetheless, the synthesis of aldehyde from PUFAs (polyunsaturated fatty acids) remain unclear. There is a report that the radical species react especially with membrane PUFAs and generate lipid peroxidation products (Liou and Storz, 2010). In the case of plants, some studies show a reaction catalyzed by LOX (lipoxygenase) and HPL (Hydroperoxide lyase). This can generate several aldehydes, e.g., *trans*-2-hexenol and *cis*-3-hexenal, which are known green leaf volatiles (Hoa et al., 2015; Kunishima et al., 2016). Animal cells, in turn, possess LOX and ADH, but there is no evidence that animal cells have HPL. A reaction from PUFA to aldehyde is therefore uncertain. Two possible explanations can be forwarded. One is that radicals (ROS) non-enzymatically produced aldehyde from PUFA (Anderson and Taylor, 2012). Our ROS data, however, did not show a correlation with *trans*-2-hexenol production. Another possibility is that aldehyde, which is a precursor for *trans*-2-hexenol, is produced enzymatically, but the enzyme remains unknown. Some reports focus on enzyme promiscuity (Piedrafita et al., 2015). There are three types of promiscuities, namely substrate, catalytic, and conditional promiscuity. Enzyme conditional promiscuity can act as a reservoir of new functions, and sometimes non-canonical metabolites can be synthesized. This makes it amenable to envisage that enzyme promiscuity in cancer cells contributes to reducing the harmful effects of aldehydes or of any substances produced due to an ROS increase inside a cell.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

TF conducted GC-MS analysis and designed all experiments and wrote paper. HO assisted VOC analysis by GC-MS and enzyme assay. RI cultured all cells in this study and conducted qPCR. SO contributed to develop VOC procedure. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank M. Stachowitsch for improving the English.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00116/full#supplementary-material

Figure S1 | Experimental scheme to capture VOC from cell culture flasks. For cellular VOC sampling, monotrap was placed into a 2.5 mL syringe connected to a cell culture flask cap via a connector. Pumping was conducted at constant speed for 30 min, which corresponds to ∼750 pumping actions.

Figure S2 | 12 VOCs (acetate, benzaldehyde, benzyl alcohol, butyrate, caproate, 1-decanol, decanal, dodecanal, 2-ethyl-1-hexanol, formate, heptadecane, heptanoate) profiles by GC-MS. Y axis indicates extracted ion counting (EIC) of each VOC. sampling was done at 0, 3, and 4 days after inoculation (i.e., D0, D3, and D4). Experiments were biologically triplicated. Error bars: standard deviation. Medium (cell-less media), HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (human lung adenocarcinoma) were compared in both 2D and 3D culture.

Figure S3 | 12 VOC (hexadecane, 4-hydroxy-nonenal, isobutyrate, 3-methyl pentanoate, nonanal, 1- nonanol, octadecane, pentadecane, phenol, propionate,

toluene, tetradecane) profiles by GC-MS. Y axis indicates extracted ion counting (EIC) of each VOC. Sampling was done at 0, 3, and 4 days after inoculation (i.e., D0, D3, and D4). Experiments were biologically triplicated. Error bars: standard deviation. Medium (cell-less media), HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (human lung adenocarcinoma) were compared in both 2D and 3D culture.

Figure S4 | 6 VOC (*trans*-2-hexenol, tridecane, undecane, undecanal, 1-undecanol, valerate) profiles by GC-MS. Y axis indicates extracted ion counting (EIC) of each VOC. Sampling was done at 0, 3, and 4 days after inoculation (i.e., D0, D3, and D4). Experiments were biologically triplicated. Error bars: standard deviation. Medium (cell-less media), HLB (Human lung fibroblasts cells), HBEpC (Human bronchial epithelial cells as primary cells), and A549 (human lung adenocarcinoma) were compared both 2D and 3D culture.

Figure S5 | Calibration curve of VOCs. **(A)** tetradecane (20–1,000 nmole), **(B)** *trans*-2-hexenol (5–500 nmole), and **(C)** isobutyrate (50–5,000 nmole). Reference compounds was added into T75 culture flask (270 mL capacity) with 10 mL water, and volatilized compounds in a flask were trapped by monotrap (RG and RSC18) in a 2.5 mL syringe, the same as used in the enrichment of cellular VOCs (30 min pumping) and measured by GC-MS. VOCs were enriched after sample preparation. X axis indicates nmole reference compounds in 10 mL. Y axis indicates extracted ion counting (EIC) of each VOC. Experiments were triplicated. Error bars: standard deviation.

Figure S6 | qPCR data of ALR2 (Aldehyde reductase) and FAR1 (Fatty acyl-CoA reductase). A549 (human lung adenocarcinoma), and HBEpC (Human bronchial epithelial cells as primary cells) were compared. Sampling was done at 1, 2, 3, 4 days after inoculation. Experiments were biologically triplicated. Y axis indicates relative gene expression to GAPDH. Error bars: standard deviation.

Figure S7 | Metabolic pathway [modified from (Ruffing, 2013)]. Fatty alcohol can be generated either from aldehyde or fatty acyl-CoA (R-SCoA). Fatty aldehyde can be derived from lipid peroxidation or from alcohol (Alcohol dehydrogenase; ADH/Aldehyde reductase; ALR). Compared with single-cell organisms, it is unclear whether metazoans possess enzymes to convert R-SCoA into aldehyde (Acyl ACP reductase; AAR), or aldehyde into hydrocarbon (Aldehyde decarbonylase; ADC).

Table S1 | Metabolite list. Total 30 VOC reference compounds for GC-MS profiling with retention time, chosen fragment *m/z* as well as CAS number.

Table S2 | Primer sequences used in qPCR. F and R indicates forward and reverse primers respectively.

Table S3 | VOC list of p-value between Days 0 and 4 (Cell less media, A549, HLB, and HBEpC). VOC whose *p*-value was below 0.05 were listed here. Red and blue letter indicate increase and decrease from Day 0, respectively. 2D and 3D indicates 2 dimensional and 3 dimensional culture respectively.

Table S4 | VOC list of p-value between cell less media (Day 4) and cell culture (A549, HLB, and HBEpC) (Day 4). VOC whose *p*-value was below 0.05 were listed here. Red and blue letter indicate increase and decrease to cell-less media respectively. In all cell samples, benzaldehyde decrease to media was common characteristics.

REFERENCES

Aktipis, C. A., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., et al. (2015). Cancer across the tree of life: cooperation and cheating in multicellularity. *Philos. Trans. R. Soc. B.* 370:20140219. doi: 10.1098/rstb.2014.0219

Anderson, E. J., and Taylor, D. A. (2012). Stressing the heart of the matter: re-thinking the mechanisms underlying therapeutic effects of n-3 polyunsaturated fatty acids. *Med. Rep.* 4:13. doi: 10.3410/M4-13

Ayala, A., Muñoz, M. F., and Argüelles, S. (2014). Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxid. Med. Cell. Longev.* 2014:360438. doi: 10.1155/2014/360438

Balber, A. E. (2011). Concise review: aldehyde dehydrogenase bright stem and progenitor cell populations from normal tissues: characteristics, activities, and emerging uses in regenerative medicine. *Stem Cells* 29, 570–575. doi: 10.1002/stem.613

Boleda, M. D., Saubi, N., Farrés, J., and Parés, X. (1993). Physiological substrates for rat alcohol dehydrogenase classes: aldehydes of lipid peroxidation, omega-hydroxyfatty acids, and retinoids. *Arch. Biochem. Biophys.* 307, 85–90. doi: 10.1006/abbi.1993.1564

Buszewski, B., Rudnicka, J., Ligor, T., Walczak, M., Jezierski, T., and Amann, A. (2012). Analytical and unconventional methods of cancer detection using odor. *Trends Anal. Chem.* 38, 1–12. doi: 10.1016/j.trac.2012.03.019

de Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., et al. (2014). A review of the volatiles from the healthy human body. *J. Breath Res.* 8:1.014001. doi: 10.1088/1752-7155/8/1/014001

Di Francesco, F., Fuoco, R., Trivella, M. G., and Ceccarini, A. (2005). Breath analysis: trends in techniques and clinical applications. *Microchem. J.* 79, 405–410. doi: 10.1016/j.microc.2004.10.008

DuBois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Anal. Chem.* 28, 350–356. doi: 10.1021/ac60111a017

Filipiak, W., Mochalski, P., Filipiak, A., Ager, C., Cumeras, R., Davis, C. E., et al. (2016). A compendium of volatile organic compounds (VOCs) released by human cell lines. *Curr. Med. Chem.* 23, 2112–2131. doi: 10.2174/0929867323666160510122913

Filipiak, W., Sponring, A., Filipiak, A., Ager, C., Schubert, J., Miekisch, W., et al. (2010). TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells *in vitro*. *Cancer Epidemiol Biomark. Prev.* 19, 182–195. doi: 10.1158/1055-9965.EPI-09-0162

Furuhashi, T., Sugitate, K., Nakai, T., Jikumaru, Y., and Ishihara, G. (2018). Rapid profiling method for mammalian feces short chain fatty acids by GC-MS. *Anal. Biochem.* 543, 51–54. doi: 10.1016/j.ab.2017.12.001

Hakim, M., Broza, Y., Barash, O., Peled, N., Phillips, M., Amann, A., et al. (2012). Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem. Rev.* 112, 5949–5966. doi: 10.1021/cr300174a

Hoa, C. T., Zheng, X., and Li, S. (2015). Tea aroma formation. *Food Sci. Hum. Wellness*. 4, 9–27. doi: 10.1016/j.fshw.2015.04.001

Ishii, R., Kami, D., Toyoda, M., Makino, H., Gojo, S., Ishii, T., et al. (2012). Placenta to cartilage: direct conversion of human placenta to chondrocytes with transformation by defined factors. *Mol. Biol. Cell.* 23, 3511–3521. doi: 10.1091/mbc.e11-10-0869

Jang, H. J., Son, H. H., and Lee, D. S. (2011). Optimization of disk sorptive extraction based on monolithic material for the determination of aroma compounds from Lantana camara L. by gas chromatography-mass spectrometry. *Bull. Korean Chem. Soc.* 32, 4275–4280. doi: 10.5012/bkcs.2011.32.12.4275

Kapałczynska, Z., Kolenda, T., Przybyła, W., Zajaczkowska, M., Teresiak, A., Filas, V., et al. (2018). 2D and 3D cell cultures – a comparison of different types of cancer cell cultures. *Arch. Med. Sci.* 14, 910–919.

Klemenz, A. C., Meyer, J., Ekat, K., Bartels, J., Traxler, S., Schubert, J. K., et al. (2019). Differences in the emission of volatile organic compounds (VOCs) between non-differentiating and adipogenically differentiating mesenchymal stromal/stem cells from human adipose tissue. *Cells.* 8:E697. doi: 10.3390/cells8070697

Kunishima, M., Yamauchi, Y., Mizutani, M., Kuse, M., Takikawa, H., and Sugimoto, Y. (2016). Identification of (*Z*)-3:(*E*)-2-hexenal isomerases essential to the production of the leaf aldehyde in plants. *J. Biol. Chem.* 291, 14023–14033. doi: 10.1074/jbc.M116.726687

Lemjabbar-Alaoui, H., Hassan, O., Yang, Y. W., and Buchanan, P. (2015). Lung cancer: biology and treatment options. *Biochim. Biophys. Acta* 1856, 189–210. doi: 10.1016/j.bbcan.2015.08.002

Liou, G. Y., and Storz, P. (2010). Reactive oxygen species in cancer. *Free Radical. Res.* 44, 479–496. doi: 10.3109/10715761003667554

Lu, W., Ogasawara, M. A., and Huang, P. (2007). Models of reactive oxygen species in cancer. *Drug Discov. Today Dis. Models* 4, 67–73. doi: 10.1016/j.ddmod.2007.10.005

Ma, W., Fu, S., Hashi, Y., and Chen, Z. (2013). Determination of chiral jasmonates in flowers by GC/MS after monolithic material sorptive

extraction. *J. Agric. Food Chem.* 6126, 6288–6292. doi: 10.1021/jf401104g

Mochalski, P., Leja, M., Gasenko, E., Skapars, R., Santare, D., Sivins, A., et al. (2018). *Ex vivo* emission of volatile organic compounds from gastric cancer and non-cancerous tissue. *J. Breath Res.* 12:046005. doi: 10.1088/1752-7163/aacbfb

Piedrafita, G., Keller, M. A., and Ralser, M. (2015). The impact of non-enzymatic reactions and enzyme promiscuity on cellular metabolism during (oxidative) stress conditions. *Biomolecules* 5, 2101–2122. doi: 10.3390/biom5032101

Rizzo, W. B. (2014). Fatty aldehyde and fatty alcohol metabolism: review and importance for epidermal structure and function. *Biochim. Biophys. Acta* 1841, 377–389. doi: 10.1016/j.bbalip.2013.09.001

Ruffing, A. M. (2013). "Metabolic engineering of hydrocarbon biosynthesis for biofuel production," in *Liquid, Gaseous and Solid Biofuels,* ed Z. Fang Chapter 8 (Rijeka: InTech), 3367–3403.

Schallschmidt, K., Becker, R., Zwaka, H., Menzel, R., Johnen, D., Fischer-Tenhagen, C., et al. (2015). *In vitro* cultured lung cancer cells are not suitable for animal-based breath biomarker detection. *J. Breath Res.* 9:027103. doi: 10.1088/1752-7155/9/2/027103

Serasanambati, M., Broza, Y. Y., Marmur, A., and Haick, H. (2019). Profiling single cancer cells with volatolomics approach. *iScience* 11, 178–188. doi: 10.1016/j.isci.2018.12.008

Snow, N. H., and Slack, G. C. (2002). Head-space analysis in modern gas chromatography. *Trends Anal. Chem.* 21, 608–617. doi: 10.1016/S0165-99360200802-6

Sonveaux, P., Végran, F., Schroeder, T., Wergin, M. C., Verrax, J., Rabbani, Z. N., et al. (2008). Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. *J. Clin. Invest.* 118, 3930–3942. doi: 10.1172/JCI36843

Yoshida, A., Rzhetsky, A., Hsu, L. C., and Chang, C. (1998). Human aldehyde dehydrogenase gene family. *Eur. J. Biochem.* 251, 549–557. doi: 10.1046/j.1432-1327.1998.2510549.x

Conflict of Interest: TF, RI, and HO were employed by the company Anicom Specialty Medicinal Institute. SO was employed by the company GL Sciences Inc.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 19 August 2020
doi: 10.3389/fmolb.2020.00211

GC-MS-Based Metabolomics for the Smut Fungus *Ustilago maydis*: A Comprehensive Method Optimization to Quantify Intracellular Metabolites

An N. T. Phan and Lars M. Blank**

Institute of Applied Microbiology – iAMB, Aachen Biology and Biotechnology – ABBt, RWTH Aachen University, Aachen, Germany

Ustilago maydis, a smut fungus, is an appealing model in fundamental research and an upcoming cell factory for industrial biotechnology. The genome of *U. maydis* has been sequenced and some synthesis pathways were biochemically described; however, the operation of the cellular metabolic network is not well-characterized. Thus, we conducted a comprehensive study to optimize the sample preparation procedure for metabolomics of *U. maydis* using GC-MS/MS. Due to the unique characteristics of *U. maydis* cell culture, two quenching solutions, different washing steps, eight extraction methods, and three derivatization conditions have been examined. The optimal method was then applied for stable isotope-assisted quantification of low molecular weight hydrophilic metabolites while *U. maydis* utilized different carbon sources including sucrose, glucose, and fructose. This study is the first report on a methodology for absolute quantification of intracellular metabolites in *U. maydis* central carbon metabolism such as sugars, sugar phosphates, organic acids, amino acids, and nucleotides. For biotechnological use, this method is crucial to exploit the full production potential of this fungus and can also be used to study other fungi of the family *Ustilaginaceae*.

Keywords: metabolomics, GC-MS/MS, *Ustilago maydis*, sample preparation, *Ustilaginaceae*, metabolic engineering

OPEN ACCESS

Edited by:
Zheng-Jiang Zhu,
Shanghai Institute of Organic Chemistry (CAS), China

Reviewed by:
Shuhai Lin,
Xiamen University, China
Edward Emmanuel Kweku Baidoo,
Particle Data Group, United States

****Correspondence:***
An N. T. Phan
an.phan1@rwth-aachen.de
Lars M. Blank
lars.blank@rwth-aachen.de

Specialty section:
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: *03 June 2020*
Accepted: *03 August 2020*
Published: *19 August 2020*

Citation:
Phan ANT and Blank LM (2020) GC-MS-Based Metabolomics for the Smut Fungus Ustilago maydis: A Comprehensive Method Optimization to Quantify Intracellular Metabolites. Front. Mol. Biosci. 7:211. doi: 10.3389/fmolb.2020.00211

INTRODUCTION

Ustilago maydis is a maize pathogen that causes corn smut (Brefort et al., 2009), a plant disease reducing cereal production. *U. maydis* has a long history as a model organism for the study of pathogen-host interactions, fungal mating, DNA recombination, and DNA repair (Bölker, 2001; Martinez-Espinoza et al., 2002; Matei and Doehlemann, 2016). Genome sequencing (Kämper et al., 2006) revealed that approximately 10% of *U. maydis* proteins were highly conserved in humans, which had lower similarity or did not even exist in the yeast *Saccharomyces cerevisiae* (Münsterkötter and Steinberg, 2007). This finding makes *U. maydis* a unique model organism for many fundamental studies that yeast models do not offer such as endocytosis, long-distance mRNA transport, cell signaling, microtubule organization, and polarized growth (Steinberg and Perez-Martin, 2008; Etxebeste and Espeso, 2016; Haag et al., 2019).

In addition, *U. maydis*, which grows in its haploid form non-filamentous (Kubicek et al., 2011), gains attention in biotechnology due to its capability of using sustainable substrates to

produce valuable chemicals of industrial interest such as mannitol, erythritol, mannosylerythritol lipids (MELs), ustilagic acid, itaconic acid, malic acid, and hydroxyparaconic acid (Geiser et al., 2014). A broad range of molecular, genetic, bioinformatic, and cell biological techniques have been developed to take full advantage of this model organism (Steinberg and Perez-Martin, 2008; Khrunyk et al., 2010; Schuster et al., 2016). "Omics" studies including genomics, transcriptomics, proteomics, and metabolomics together can elucidate cell physiology even further. The methodologies for comparative genome analysis (Schirawski et al., 2010; Laurie et al., 2012), transcriptomics (Doyle et al., 2011; Islamovic et al., 2015; Donaldson et al., 2017), and proteomics (Böhmer et al., 2007; Martínez-Salgado et al., 2013) have been established. However, a well-characterized metabolomics method is missing and, indeed, only a few biochemical synthesis pathways with a very limited number of detected metabolites were reported (Hewald et al., 2006; Teichmann et al., 2010; Winterberg et al., 2010; Jonkers et al., 2012). Hence, to study the genotype-phenotype relationship in *U. maydis*, it is crucial to develop a robust protocol for metabolomics.

There is no universal sample preparation in metabolomics that can be applied to all types of cells. The procedure may vary depending on cell properties including cell size, morphology in general, cell wall structure, but also growth conditions, growth rate, and metabolome complexity (Gulik et al., 2013), and not at least on the analytical method of choice. Moreover, *Ustilaginaceae* are well known for the ability to tolerate high substrate concentrations, grow in high cell density, and produce high titer of extracellular products (Zambanini et al., 2016; Becker et al., 2020). Therefore, it is very challenging to isolate the cells from a viscous and complex medium to obtain reliable information on the metabolome.

Here, we focused on metabolomics using gas chromatography coupled with mass spectrometry (GC-MS/MS), one of the analytical methods that covers a wide range of primary metabolites in a single run (Schauer et al., 2005; Oldiges et al., 2007; Fiehn, 2016). Rapid sampling is critical for metabolomics, as especially metabolites from central carbon metabolism have fast turnover rates (Buchholz et al., 2002). There are several methods to stop microbial cell metabolism including fast filtration, and quick quenching with organic solvents or cryoprotectants (Pinu et al., 2017). Additional washing steps using a cold biological buffer such as saline buffer or phosphate-buffered saline can be added to remove analyte-rich culture medium. Notably, glycerol, the most popular cryoprotectant, is not applicable in this case because it intrudes on the derivatization for GC-MS analysis. Next, a proper extraction procedure is necessary to release the metabolites from the cells. Extraction solvents with different polarities (i.e., water, ethanol, methanol, or solvent mixtures) are usually employed in combination with physical forces (i.e., temperature, sonication, vortex or microwave) to enhance extraction efficiency (Pinu et al., 2017). The metabolite compositions obtained after extraction have an impact on sample derivatization (Kanani et al., 2008). For the two-step method using methoximation followed by silylation, there are various conditions with different temperatures and incubation times ranging between 30–90°C and from 0.5–6 h, respectively (Moros et al., 2016). Effective sample preparation methods have been developed for many fungi such as *S. cerevisiae*, *Aspergillus* sp., *Monascus ruber*, and *Penicillium* sp. (Hajjaj et al., 1998; Jernejc, 2004; Canelas et al., 2009; de Jonge et al., 2012; Duportet et al., 2012; Kim et al., 2013; Zheng et al., 2019). However, to our knowledge, no metabolomics studies have been performed using *U. maydis*.

In this study, we aimed to optimize every step of sample preparation including quenching, washing, extraction, and derivatization for metabolomics of *U. maydis*. The method was then applied for absolute quantification of intracellular metabolites using an isotope-assisted approach, which supports deciphering the metabolic network operation in *U. maydis*. Hence, the results from the present study will contribute to the ever-increasing toolbox that makes *U. maydis* a model organism. In addition, the method presented can be used for metabolomics studies of other fungi of the family *Ustilaginaceae*.

MATERIALS AND METHODS

Strain and Culture Conditions

Ustilago maydis strain MB215 (DSM17144) was used for all experiments. MTM medium was used according to Geiser et al. (2014) with 50 g L^{-1} glucose, 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), 0.2 g L^{-1} $MgSO_4 \cdot 7H_2O$, 10 mg L^{-1} $FeSO_4 \cdot 7H_2O$, 0.5 g L^{-1} KH_2PO_4, 0.8 g L^{-1} NH_4Cl, 1 mL L^{-1} vitamin solution and 1 mL L^{-1} trace element solution. For the experiment using different carbon sources, the same concentration (g/L) of fructose or sucrose were utilized instead of glucose. Experiments were performed in 24-deep well plates (*Enzyscreen*, System Duetz®) (Duetz et al., 2000) with 1.5 mL MTM per well, incubated at 30°C, relative air humidity of 80% and shaking speed of 300 rpm (*Infors HT Multitron Pro* shaker).

Sample Collection

Main cultures were started at OD_{600} of 0.5 and OD_{600} was checked over time to determine the collection time. At OD_{600} of 20, corresponding to the mid-exponential phase of cell growth, a volume equal to 10 OD_{600} unit were collected for intracellular metabolite measurement. The same volume of samples was transferred to pre-weighed dry Eppendorf tubes to determine cell dry weight (CDW). The samples were washed twice with water and then dried in an oven at 70°C until a constant weight was achieved.

Quenching and Washing Conditions

Nylon filters with pore sizes of 0.2 and 0.45 μm (hydrophilic, 25 mm diameter, *Millipore*) were used for fast filtration. For quick quenching, two solutions were examined including 4°C Saline Buffer 0.9% NaCl (SB) and −20°C absolute methanol (MeOH). The broth culture was cast into the precooled quenching solution at a ratio of l: 6 and mixed well. The mixture was centrifuged at 12,000 rpm for 30 s. The pellet was washed 1–3 times with 1.5 mL SB. Before determining the optimal extraction conditions, a common extraction method using chloroform/MeOH/water

(2:5:2) was applied (Dempo et al., 2014; Hashim et al., 2014; An et al., 2017).

ATP Measurement and Leakage Calculation

The level of ATP in quenching and washing solution were directly measured using the Molecular Probes® ATP Determination Kit (Invitrogen, Thermo Fisher Scientific). The percentage of ATP leakage was defined by the ratio of extracellular [ATP] to total [ATP] (extracellular + intracellular). We could not determine adenylate energy charge due to the limitation of analytical methods. Instead, we combined the ATP-leakage with peak RSD-values to evaluate intracellular metabolite leakages and method reproducibility.

Extraction Methods

Three extraction solutions were tested including ethanol (EtOH: H_2O; 7:2), MeOH (MeOH: H_2O; 7:2), and chloroform/MeOH/water mixture (2:5:2). In total, 900 µL of extract solution was used for each sample and ribitol was used as an internal standard with the final concentration of 20 µM. Hot extraction (95°C for 5 min and 4°C for 5 min) was applied to EtOH (HE) and MeOH (HM). Cold extraction (vortexed at 4°C for 1 h) was applied to EtOH (CE), MeOH (CM), and chloroform/MeOH/water mixture (CC). Sonication (sonicated for 30 s and set on ice for 2 min, repeat 3 times) was applied to EtOH (SE), MeOH (SM), and chloroform/MeOH/water mixture (SC). Then, all samples were centrifuged (13,000 rpm for 5 min) to collect the supernatant. When using chloroform/MeOH/water mixture to extract metabolites, 900 µL supernatant was mixed with 400 µL water and then centrifuged to separate the polar and non-polar phases. 400 µL of the polar phase were transferred into a fresh 1.5-mL tube. For other methods, 400 µL supernatant was directly transferred to a 1.5-mL microtube. Finally, the solvent was removed using a centrifugal concentrator. In our study, the matrix effects from the various extraction protocols were not examined as matrix effects are generally neglected in GC-MS and we did not encounter problems with the background signal or with peak separation.

Metabolite Derivatization

50 µL of methoxyamine hydrochloride (MeOX) (20 mg/mL in pyridine, *Sigma Aldrich*) were added to each sample for methoximation. Samples were incubated at 37°C (for 90 min), 60°C (for 30 min), or 80°C (for 30 min). Then, silylation has been performed by adding 50 µL of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) (*Sigma Aldrich*) and incubated at 37, 60, or 80°C for 30 min. Before examining the optimal derivatization condition, all samples were derivatized at 37°C.

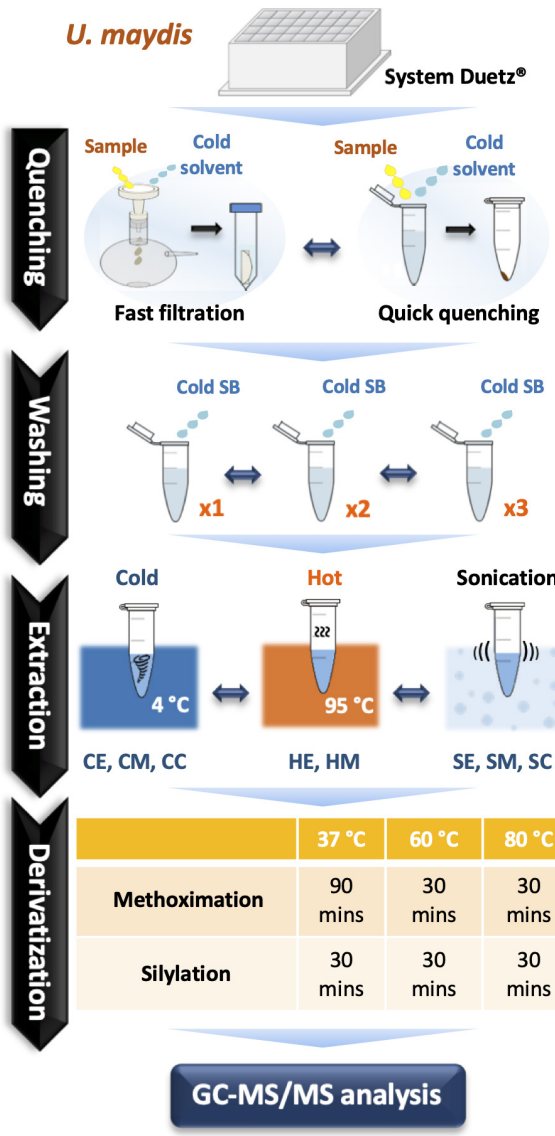


FIGURE 1 | A comprehensive optimization of sample preparation for GC-MS/MS-based metabolomics in *U. maydis*. Cell cultivations were performed in System Duetz® 24-deep well plates at 30°C in minimal media. To instantly stop cellular metabolism, both fast filtration and quick quenching methods were tested. Then, cells were washed one (x1), two (x2), and three (x3) times with cold saline buffer to remove the extracellular components from culture medium. Eight different strategies combining chemical and mechanical disruption methods were tested for extraction efficiency. Finally, metabolites were derivatized at 37°C, 60°C, or 80°C. All experiments were performed with three biological replicates. Abbreviations: SB, Saline Buffer; CE, cold EtOH; CM, cold MeOH; CC, cold chloroform/MeOH/water mixture; HE, hot EtOH; HM, hot MeOH; SE, EtOH with sonication; SM, MeOH with sonication; SC, chloroform/MeOH/water mixture with sonication.

GC-MS/MS Method

The analysis was performed on Trace GC Ultra coupled with TSQ 8000 (*Thermo Fisher Scientific*). Tuning and calibration were done before analysis. A 1 µL aliquot of the derivatives was injected into a VF-5 ms capillary column (30 m × 250 µm i.d., 0.25 µm film thickness) with 10 m EZ-Guard (*Agilent*). Each sample was injected twice with two different split modes including 20:1 (v/v) for low concentration metabolites and 100:1 (v/v) for high concentration metabolites. The injection temperature was 250°C and the helium carrier gas flow was 1 mL/min. The column temperature was held at 100°C for 2 min,

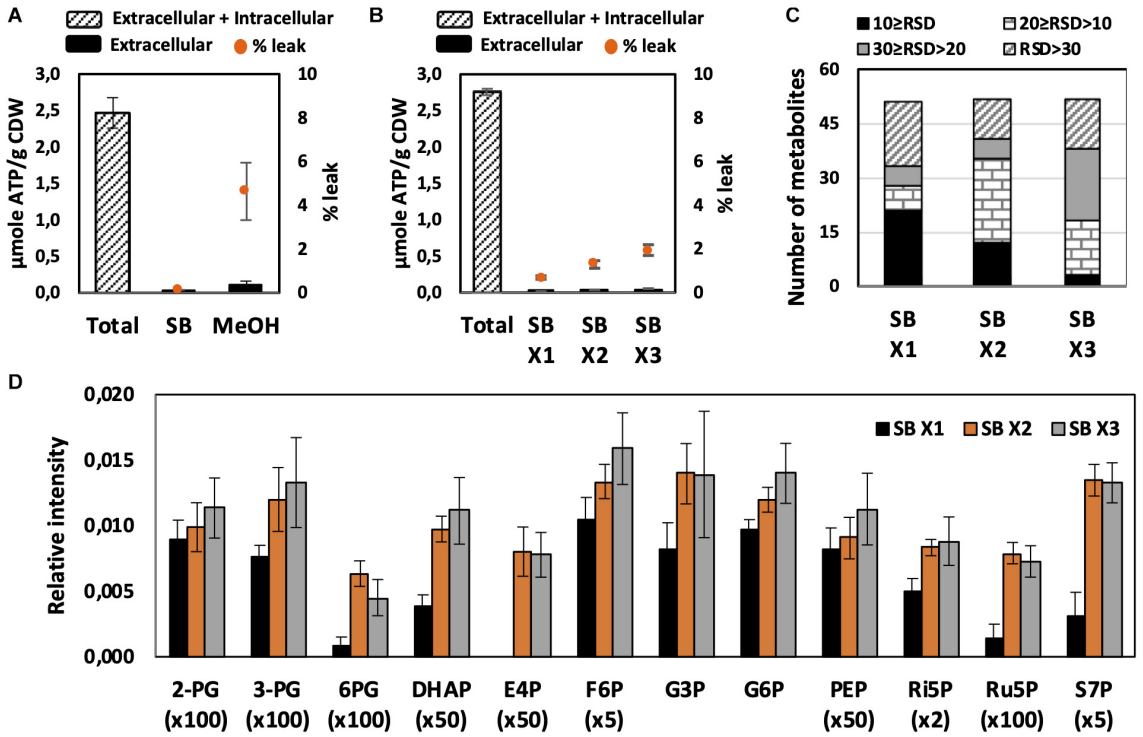


FIGURE 2 | Effect of quenching and washing on *U. maydis* intracellular metabolites. **(A)** Metabolite leakage while quenching with SB and MeOH. **(B)** Metabolite leakage, **(C)** Relative standard deviations (RSD) of all detected peaks, and **(D)** phosphorylated metabolite levels when *U. maydis* cells were washed one (x1), two (x2), and three (x3) times with SB. Each data point represents the average of three biological replication with the error bar indicating the standard deviation. Abbreviations: 2-PG, 2-phosphoglycerate, 3-PG: 3-phosphoglycerate; 6PG, 6-phosphogluconate; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G3P, glycerol-3-phosphate; G6P, glucose-6-phosphate; PEP, phosphoenolpyruvate; Ri5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; and S7P, sedoheptulose-7-phosphate. Normalized peak intensities of some metabolites were multiplied for data visualization.

increased by 15°C/min to 200°C (held for 4 min) and then increased by 20°C/min to 325°C (held for 5 min). The transfer line 280°C and ion source temperatures were 300°C. Ions were generated by electron ionization (EI) at 70 eV and analyzed in positive mode. Argon was utilized as a collision gas.

Data Processing

A Selected Reaction Monitoring (SRM) transition library of metabolites was constructed using authentic standards from *Sigma-Aldrich* (**Supplementary Table S1**). Standard mixtures were injected periodically throughout the analysis to evaluate the stability of the analytical system as well as to support peak identification. All data were processed by using Thermo Xcalibur software version 3.1 (*Thermo Scientific*).

Relative quantification was used during method optimization (**Supplementary Tables S2–S4**). All peak areas obtained from GC-MS/MS analysis were normalized against the ribitol signal.

Absolute quantification was applied to measure intracellular metabolite levels of *U. maydis* while using different carbon sources (**Supplementary Table S5**). The extract from *U. maydis* grown in fully U-^{13}C-labeled glucose (*Sigma-Aldrich*) was used as internal standard. The absolute concentrations of intracellular metabolites in the samples were calculated by using ^{12}C/^{13}C ratio-based calibration curves.

Statistical Analyses

Relative standard deviations (RSD) of all detected peaks were calculated after normalization. Unit Variance (UV) scale was applied for all datasets prior to statistical analyses. Principle Component Analysis (PCA) was performed utilizing SIMCA-P^{+} version 15.0.2 (Umetrics, Umea). Hierarchical Cluster Analysis (HCA), *t*-test, one-way ANOVA test, and pathway enrichment analysis were performed using MetaboAnalyst 4.0 (Chong et al., 2018).

All detected metabolites were used for PCAs. Only metabolites that were significantly different in each condition (p-values < 0.05) were employed for HCAs. HCAs were calculated based on the Euclidean correlation with the Ward clustering algorithm. In *t*-test and one-way ANOVA test, p-value adjustments for multiple metabolites were carried out by using false discovery rate (FDR) adjustments.

RESULTS AND DISCUSSION

Metabolite Leakage During Quenching and Washing

While quenching is crucial to stop cell metabolism quicker than the turnover of metabolites, washing is utilized to remove

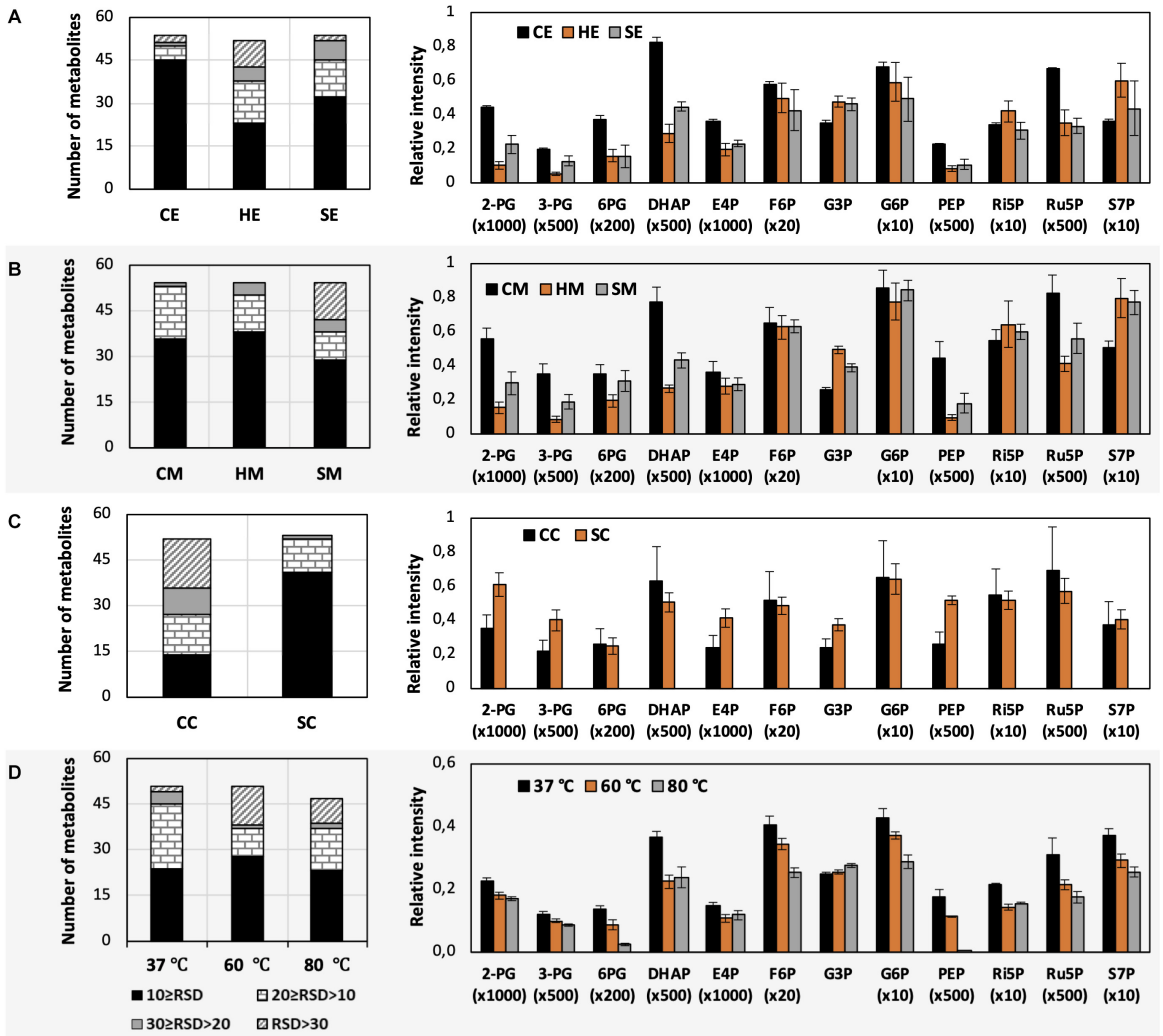


FIGURE 3 | Impacts of different extraction and derivatization conditions on *U. maydis* intracellular metabolites. RSD values and the levels of phosphorylated metabolites extracted with **(A)** EtOH, **(B)** MeOH, and **(C)** chloroform/MeOH/water; as well as **(D)** derivatized under different temperatures. Each data point represents the average of three biological replication with the error bar indicating the standard deviation. Normalized peak intensities of some metabolites were multiplied for data visualization.

extracellular compounds from the culture medium. These steps have to compromise to keep the metabolites lost at a minimum. First, we tested fast filtration using filters with pore sizes of 0.2 and 0.45 μm. Because of sample viscosity, the cells were not able to pass the membranes completely, despite vacuum was applied. This problem is most likely due to the excessive abundance of extracellular lipids, like MELs, which has been reported to cause troubles in the centrifugation efficiency of the liquid medium (Becker et al., 2020).

Quick quenching methods were subsequently examined using SB and MeOH as quenching solutions. Interestingly, *U. maydis* cells tended to accumulate when interacting with MeOH (**Supplementary Figure S1**). Cell accumulation interfered with further washing steps and led to the detection of mainly a saturated glucose peak as well as other extracellular product peaks. Since polysaccharides are essential components of the fungal cell wall and *U. maydis* secretes β-D-glucan (Fonseca-García et al., 2011), the observed phenomenon might be a result of polysaccharide precipitation in organic solvents. Thus, SB with better performance and less leakage was chosen as not only quenching but also washing solution in further experiments (**Figure 2A**).

We evaluated each optimization step based on two main criteria. First, the reproducibility was assessed based on RSD values. A common standard for a good method is an RSD value of detected peak of lower than 30%. Among metabolites in central metabolic pathways, it is always challenging to detect phosphorylated compounds. Therefore, the second priority is for the method that is able to measure this group of metabolites. As predicted, the use of SB, a biological buffer, did not lead to severe metabolite leakage. The leak percentage increased when more washing steps with SB were added but none of them exceeded

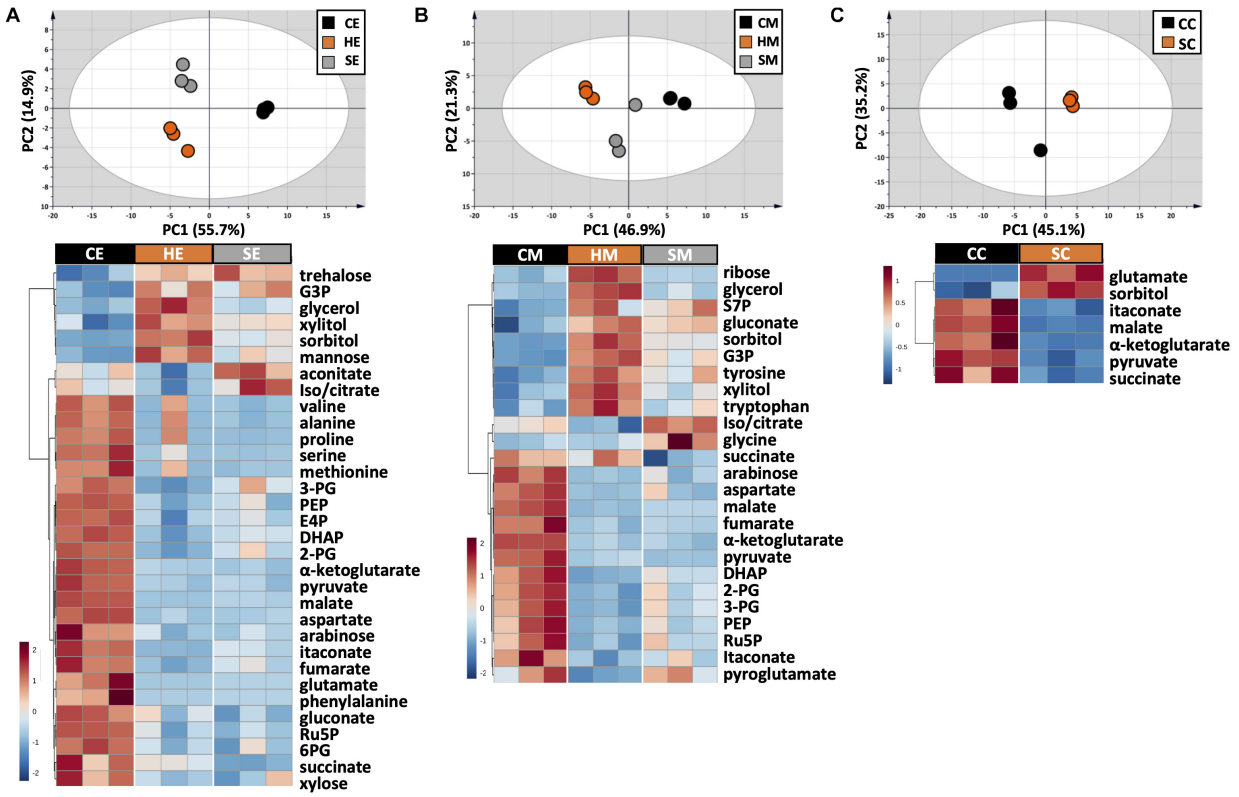


FIGURE 4 | Metabolic profile comparison during extraction optimization. Upper panels are PCA score plots and lower panels are the HCA heatmaps of metabolites extracted with **(A)** EtOH, **(B)** MeOH, and **(C)** a chloroform/MeOH/water mixture. While the whole dataset was used for PCA, only metabolites with adjusted p-value < 0.05 were used for HCA. Each experiment consisted of three biological replications and UV scale was applied prior to statistical analyses. The color scales in HCA indicate relative intensity after normalization and scaling.

3% (**Figure 2B**). The least robust condition was washing only one time with SB (**Figures 2C,D**). Approximately one-third of detected peaks had RSD values higher than 30% and the levels of many phosphorylated metabolites were significantly lower. The lower sensitivity might occur due to the remaining extracellular compounds, which affected the derivatization efficiency. Washing two times with SB appealed to be the best condition because of the high reproducibility of overall metabolite measurements and the determination of phosphorylated compounds.

Extraction Methods for Low Molecular Weight Hydrophilic Metabolites in Central Metabolic Pathways

The choice of extraction method will define the number, type, and abundance of detected metabolites from *U. maydis* cells. A proper extraction method should be able to make the cell envelope permeable to release metabolites completely and denature cellular enzymes to prevent metabolite degradation during sample preparation. Yet, the extraction condition itself should not cause metabolite degradation or transformation. In order to figure out a suitable extraction method, we combined both chemical and mechanical disruption methods into eight different strategies (**Figure 1**). Together with the RSD values and phosphorylated compound levels as mentioned above, PCA was applied to have an overview of the whole dataset among different extraction conditions (**Figure 4**, upper panels). One-way ANOVA test or t-test was performed to figure out the metabolites which had significantly different levels in each condition. The metabolites with adjusted p-values lower than 0.05 were submitted to HCA, a method to cluster the metabolites according to their abundance (**Figure 4**, lower panels).

One critical point during sample preparation was that *U. maydis* cells should not interact directly to organic solvent after centrifugation. The cell pellets had to be mixed with the water portion of extraction solvents first. After the cells were fully resuspended, the organic solvents were added, and extraction started. When the pre-mixed extract solvent was used, *U. maydis* cells accumulated (as in **Supplementary Figure S1**) and were not able to be fully resuspended; thus, the results were not reproducible. This was also the reason why neat organic solvents were not suitable to extract intracellular metabolites from *U. maydis*.

The use of MeOH and EtOH resulted in very comparable results under all tested conditions, which was reasonable as these organic solvents have very close polarity indices of 5.1 and 5.2, respectively (Ramluckan et al., 2014). Cold extraction showed the highest reproducibility. Almost every metabolite had RSD

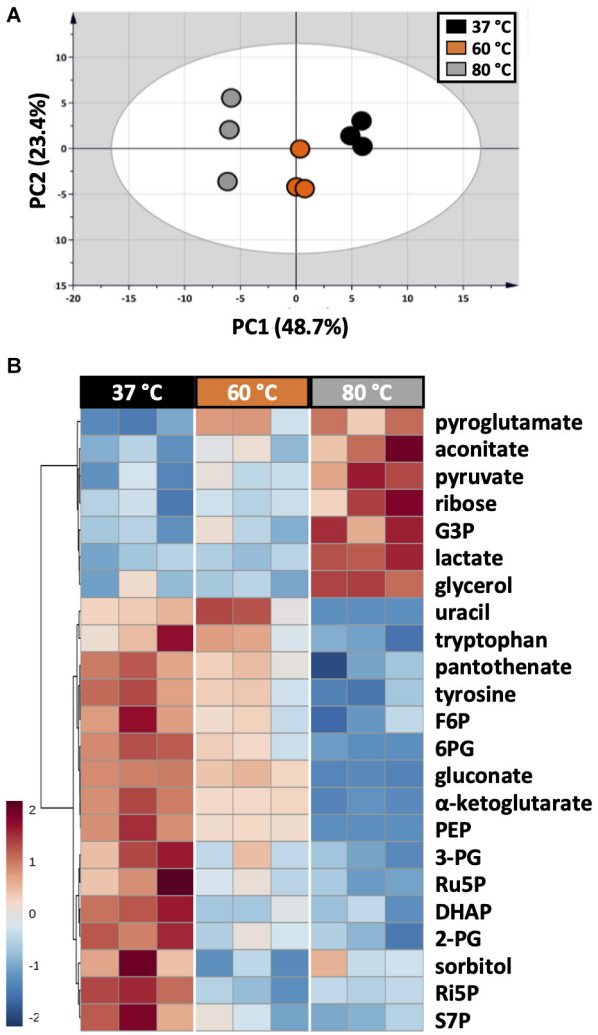


FIGURE 5 | Impacts of different derivatization conditions on *U. maydis* intracellular metabolites. **(A)** PCA score plot acquired from the complete dataset. **(B)** HCA heatmap of metabolites with adjusted *p*-value < 0.05 from one-way ANOVA test. Each experiment had three biological replicates and UV scale was applied before statistical analyses. The color scale in HCA indicates relative intensity after normalization and scaling.

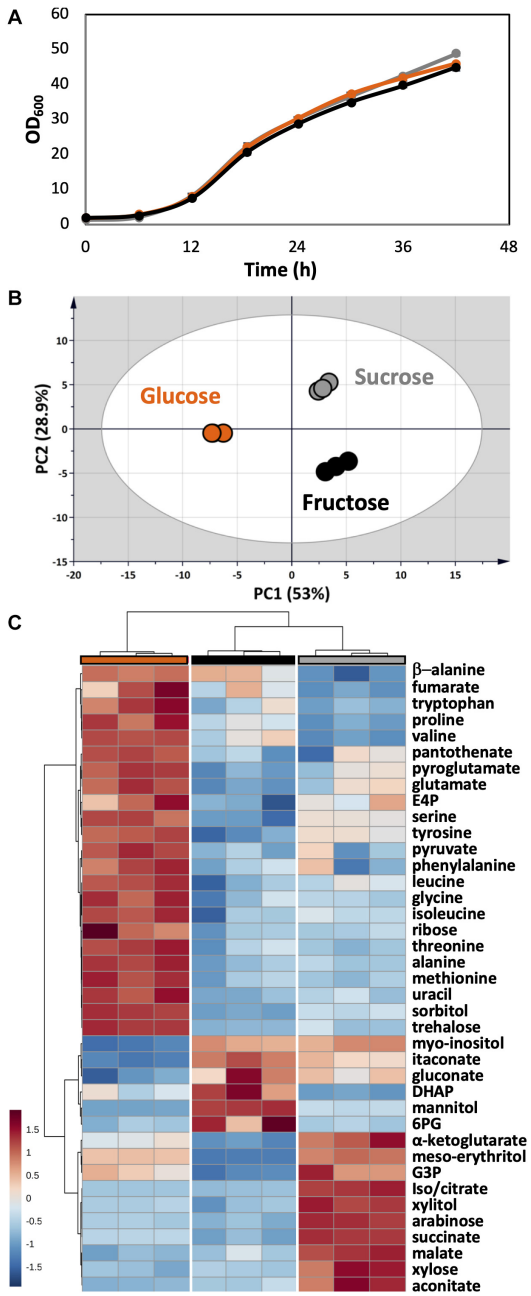


FIGURE 6 | Metabolome analysis of *U. maydis* during growth with glucose, fructose, and sucrose as sole carbon sources. **(A)** Growth curves were determined by measuring OD_{600}. **(B)** PCA was conducted with all detected metabolites. **(C)** The HCA heatmap indicates significant different metabolites among all conditions. The colors were consistent among each group of samples: glucose (orange), fructose (black), and sucrose (gray). All experiments were performed with three biological replicates.

values lower than 30% (**Figures 3A,B**). Changing in extraction conditions from hot to sonicate, and to cold methods showed increased levels of phosphorylated metabolites, especially for low abundant compounds such as 2-PG, 3-PG, 6PG, DHAP, E4P, PEP, and Ru5P (**Figures 3A,B**). Moreover, the organic acids including pyruvate, malate, succinate, α-ketoglutarate, and itaconate could be extracted better with CE and CM (**Figures 4A,B**). Low metabolite levels in hot extraction were most likely due to their thermolability.

The method using chloroform/MeOH/water is very popular for the extraction of both polar and non-polar metabolites. Here, we found that this solvent mixture could extract *U. maydis* metabolites better with sonication (SC) than vortex at low temperature (CC). The overall reproducibility was drastically improved in SC, even though the intensities for most of the metabolites were not significantly different between CC and SC (**Figures 3C**, **4C**). Vice versa, the use of sonication for EtOH and MeOH was not as good as cold extraction. While sonication is more aggressive to cell membranes comparing to vortexing,

alcohols cannot denature protein as good as chloroform (Asakura et al., 1978). Some enzymes might still be partly active during SE and SM; thus, affected metabolite stability.

Generally, the three extraction conditions CE, CM, and SC were all suitable for GC-MS/MS-based metabolomics of *U. maydis*, with an overall good reproducibility (**Supplementary Figure S2**). In addition, the sensitivity was reasonable as all metabolites were in the quantification range of GC-MS/MS, especially for phosphorylated metabolites. For the next steps, CE was chosen because EtOH is environmentally friendly, not carcinogenic, and sample preparation for CE was simple.

Effects of Derivation Conditions on GC-MS/MS Analysis

Almost every metabolite investigated requires derivatization to increase volatility for GC-MS analysis. The two-step derivatization method using methoximation and silylation is well established in the field of metabolomics (Fiehn, 2016). First, MeOX reacts to metabolites with ketone groups. This step is important to open sugar rings, which results in fewer peaks per sugar in GC-MS analysis (Yi et al., 2014). MSTFA is used to derivatize chemical components with an active H, such as –OH, –COOH, $–NH_2$, and –SH. Previous studies have already shown that the formation and stability of derivatives among metabolites differ greatly.

As each organism has unique metabolic profiles, derivatization conditions should be studied prior to applications. Here, three conditions commonly used to derivatize metabolites from central carbon metabolism were examined (**Figure 1**). Throughout derivatization, metabolites were exposed to high temperature for at least 1 h. The observed RSD values indicated that the higher the temperature, the more unstable many metabolites were (**Figure 3D**). The signals of almost all metabolites were highest when derivatized at 37°C (**Figure 5**). Moreover, the peaks of proline and phenylalanine could not be detected after derivatization at 80°C. This result was consistent with previous results from the extraction step, showing that primary metabolites were sensitive to high temperatures. Hence, we recommend performing both methoximation and silylation at 37°C.

Absolute Quantification of Intracellular Metabolites When *U. maydis* Is Grown on Different Carbon Sources

The fungal family *Ustilaginaceae* is well-known for the capability of using a board range of carbon sources to produce molecules of industrial interest such as organic acids, glycolipids, and sugar alcohols (Geiser et al., 2014). The optimized sample preparation method was employed to measure intracellular metabolite levels while *U. maydis* utilizing sucrose, glucose, and fructose as sole carbon source. In order to apply the isotope-assisted approach for absolute quantification, the extract from *U. maydis* grown in fully U-^{13}C-labeled glucose was used as an internal standard.

As a plant pathogen, *U. maydis* had no problem in using fructose and sucrose. The growth of *U. maydis* on these carbon sources was very similar when compared with growth on glucose (**Figure 6A**). Though, cells grown on different carbon sources had distinct metabolic profiles, indicated by three clearly separated groups in the PCA score plot (**Figure 6B**). Samples in the "glucose" group were most distinct from the other samples, shown in the high separation on PC1 (46.6%) in the PCA and the distance in HCA clusters (**Figures 6B,C**). The one-way ANOVA test scored 39 out of 51 metabolite concentrations as significantly different with an adjusted p-value < 0.05.

The concentrations of free amino acids were high when *U. maydis* used glucose. Especially, the high abundances of metabolites related to aromatic amino acid metabolism including erythritol, E4P, phenylalanine, tyrosine, and tryptophan was observed. On the other hand, a high level of mannitol was found when *U. maydis* utilized fructose, most likely because fructose can be converted directly to mannitol by the mannitol dehydrogenase. The concentrations of many organic acids of the TCA cycle were divided into two sets while comparing "glucose" and "fructose" groups. Acotinate and itaconate levels were higher in the "fructose" group; however, pyruvate, malate, fumarate, succinate, and α-ketoglutarate concentrations were higher in the "glucose" group.

From the annotated genome, *U. maydis* has a putative invertase (UMAG_01945), which is potentially capable of hydrolyzing sucrose into glucose and fructose molecules (Kämper et al., 2006). Thus, the metabolite compositions of the "sucrose" group shared the features of both "glucose" and "fructose" groups. While the concentration of intracellular mannitol was higher than cells grown on glucose, the level of erythritol was the highest among all conditions.

Together, the choice of carbon source had substantial impact on intracellular metabolite concentrations. Depending on the target compound, metabolic engineering can be employed to alter the metabolic flux distribution to maximize product synthesis. Erythritol and mannitol are well-known as sweeteners (Grembecka, 2015). The carboxylic acids in the TCA cycle are listed among the value-added chemicals from biomass (Werpy and Petersen, 2004). Amino acids are important dietary supplements of mainly animals, the aromatic amino acids can be used as precursors for industrial and pharmaceutical compounds (i.e., aspartame, L-DOPA, melanin, or phenol) (Parthasarathy et al., 2018; Wynands et al., 2018).

For the first time, the absolute quantification of intracellular metabolites was conducted in *U. maydis*. The present study provides a well-established sample preparation method for GC-MS/MS-based metabolomics, which still can be further adjusted according to research objectives or to a new analytical device (i.e., nuclear magnetic resonance or other MS-based equipment). This method will further support the development of *U. maydis* from a basic science model organism, to a next-generation model organism also fulfilling the requirements for modern biotechnology.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AP performed all the experiments. LB supervised the study. Both authors analyzed the data. Both authors have approved the final version of the manuscript.

FUNDING

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programe under the Marie Skłodowska-Curie grant agreement no. 793158. It was also partially funded by the Alexander von Humboldt Foundation. The laboratory of LB was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Cluster of Excellence FSC 2186 "The Fuel Science Center." The GC-MS/MS was also funded by the DFG under grant agreement no. 233069590.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00211/full#supplementary-material

FIGURE S1 | *Ustilago maydis* cells in quenching solution.

FIGURE S2 | Comparison between CE, CM, and SC extraction method.

TABLE S1 | Selected reaction monitoring (SRM) transitions for GC-MS/MS analysis.

TABLE S2 | Metabolite levels while applying different washing steps.

TABLE S3 | Metabolite levels and results of statistical analyses while applying different washing steps.

TABLE S4 | Metabolite levels and results of statistical analyses in various derivatization conditions.

TABLE S5 | Metabolite concentrations and results of statistical analyses during growth of *U. maydis* on glucose, fructose, or sucrose.

REFERENCES

An, P. N. T., Yamaguchi, M., and Fukusaki, E. (2017). Metabolic profiling of *Drosophila melanogaster* metamorphosis: a new insight into the central metabolic pathways. *Metabolomics* 13:29. doi: 10.1007/s11306-017-1167-1

Asakura, T., Adachi, K., and Schwartz, E. (1978). Stabilizing effect of various organic solvents on protein. *J. Biol. Chem.* 253, 6423–6425.

Becker, J., Hosseinpour Tehrani, H., Gauert, M., Mampel, J., Blank, L. M., and Wierckx, N. (2020). An *Ustilago maydis* chassis for itaconic acid production without by-products. *Microb. Biotechnol.* 13, 350–362. doi: 10.1111/1751-7915.13525

Böhmer, M., Colby, T., Böhmer, C., Bräutigam, A., Schmidt, J., and Bölker, M. (2007). Proteomic analysis of dimorphic transition in the phytopathogenic fungus *Ustilago maydis*. *Proteomics* 7, 675–685. doi: 10.1002/pmic.200600900

Bölker, M. (2001). *Ustilago maydis* – a valuable model system for the study of fungal dimorphism and virulence. *Microbiology* 147, 1395–1401. doi: 10.1099/00221287-147-6-1395

Brefort, T., Doehlemann, G., Mendoza-Mendoza, A., Reissmann, S., Djamei, A., and Kahmann, R. (2009). *Ustilago maydis* as a pathogen. *Annu. Rev. Phytopathol.* 47, 423–445. doi: 10.1146/annurev-phyto-080508-081923

Buchholz, A., Hurlebaus, J., Wandrey, C., and Takors, R. (2002). Metabolomics: quantification of intracellular metabolite dynamics. *Biomol. Eng.* 19, 5–15. doi: 10.1016/S1389-0344(02)00003-5

Canelas, A. B., ten Pierick, A., Ras, C., Seifar, R. M., van Dam, J. C., van Gulik, W. M., et al. (2009). Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal. Chem.* 81, 7379–7389. doi: 10.1021/ac900999t

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 46, W486–W494. doi: 10.1093/nar/gky310

de Jonge, L. P., Douma, R. D., Heijnen, J. J., and van Gulik, W. M. (2012). Optimization of cold methanol quenching for quantitative metabolomics of *Penicillium chrysogenum*. *Metabolomics* 8, 727–735. doi: 10.1007/s11306-011-0367-363

Dempo, Y., Ohta, E., Nakayama, Y., Bamba, T., and Fukusaki, E. (2014). Molar-based targeted metabolic profiling of cyanobacterial strains with potential for biological production. *Metabolites* 4, 499–516. doi: 10.3390/metabo4020499

Donaldson, M. E., Ostrowski, L. A., Goulet, K. M., and Saville, B. J. (2017). Transcriptome analysis of smut fungi reveals widespread intergenic transcription and conserved antisense transcript expression. *BMC Genomics* 18:340. doi: 10.1186/s12864-017-3720-8

Doyle, C. E., Donaldson, M. E., Morrison, E. N., and Saville, B. J. (2011). *Ustilago maydis* transcript features identified through full-length cDNA analysis. *Mol. Genet. Genomics* 286:143. doi: 10.1007/s00438-011-0634-z

Duetz, W. A., Rüedi, L., Hermann, R., O'Connor, K., Büchs, J., and Witholt, B. (2000). Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. *Appl. Environ. Microbiol.* 66, 2641–2646. doi: 10.1128/aem.66.6.2641-2646.2000

Duportet, X., Aggio, R. B. M., Carneiro, S., and Villas-Bôas, S. G. (2012). The biological interpretation of metabolomic data can be misled by the extraction method used. *Metabolomics* 8, 410–421. doi: 10.1007/s11306-011-0324-1

Etxebeste, O., and Espeso, E. A. (2016). Neurons show the path: tip-to-nucleus communication in filamentous fungal development and pathogenesis. *FEMS Microbiol. Rev.* 40, 610–624. doi: 10.1093/femsre/fuw021

Fiehn, O. (2016). Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling. *Curr. Protoc. Mol. Biol.* 114, 30.4.1–30.4.32. doi: 10.1002/0471142727.mb3004s114

Fonseca-García, C., López, M. G., Aréchiga-Carvajal, E. T., and Ruiz-Herrera, J. (2011). A novel polysaccharide secreted by pal/rim mutants of the phytopathogen fungus *Ustilago maydis*. *Carbohydr. Polym.* 86, 1646–1650. doi: 10.1016/j.carbpol.2011.06.077

Geiser, E., Wiebach, V., Wierckx, N., and Blank, L. M. (2014). Prospecting the biodiversity of the fungal family Ustilaginaceae for the production of value-added chemicals. *Fungal Biol. Biotechnol.* 1:2. doi: 10.1186/s40694-014-0002-y

Grembecka, M. (2015). Sugar alcohols—their role in the modern world of sweeteners: a review. *Eur. Food Res. Technol.* 241, 1–14. doi: 10.1007/s00217-015-2437-7

Gulik, W., Canelas, A., Seifar, R., and Heijnen, S. (2013). "The sampling and sample preparation problem in microbial metabolomics," in *Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data*, eds M. Lämmerhofer and W. Weckwerth (Hoboken, NJ: Wiley), 1–19. doi: 10.1002/9783527655861.ch1

Haag, C., Klein, T., and Feldbr, G. M. (2019). ESCRT mutant analysis and imaging of ESCRT components in the model fungus *Ustilago maydis*. *Methods Mol. Biol.* 1998, 251–272.

Hajjaj, H., Blanc, P. J., Goma, G., and François, J. (1998). Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. *FEMS Microbiol. Lett.* 164, 195–200. doi: 10.1111/j.1574-6968.1998.tb13085.x

Hashim, Z., Teoh, S. T., Bamba, T., and Fukusaki, E. (2014). Construction of a metabolome library for transcription factor-related single gene mutants of *Saccharomyces cerevisiae*. *J. Chromatogr. B* 966, 83–92. doi: 10.1016/j.jchromb.2014.05.041

Hewald, S., Linne, U., Scherer, M., Marahiel, M. A., Kämper, J., and Bölker, M. (2006). Identification of a gene cluster for biosynthesis of *Mannosylerythritol* Lipids in the basidiomycetous fungus *Ustilago maydis*. *Appl. Environ. Microbiol.* 72, 5469–5477. doi: 10.1128/aem.00506-06

Islamovic, E., García-Pedrajas, M. D., Chacko, N., Andrews, D. L., Covert, S. F., and Gold, S. E. (2015). Transcriptome analysis of a *Ustilago maydis* ust1 deletion mutant uncovers involvement of Laccase and Polyketide Synthase genes in spore development. *Mol. Plant Microbe Interact.* 28, 42–54. doi: 10.1094/mpmi-05-14-0133-r

Jernejc, K. (2004). Comparison of different methods for metabolite extraction from *Aspergillus niger* mycelium. *Acta Chim. Slov.* 51, 567–578.

Jonkers, W., Rodriguez Estrada, A. E., Lee, K., Breakspear, A., May, G., and Kistler, H. C. (2012). Metabolome and transcriptome of the interaction between *Ustilago maydis* and *Fusarium verticillioides* in vitro. *Appl. Environ. Microbiol.* 78, 3656–3667. doi: 10.1128/AEM.07841-11

Kämper, J., Kahmann, R., Bölker, M., Ma, L.-J., Brefort, T., Saville, B. J., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444, 97–101. doi: 10.1038/nature05248

Kanani, H., Chrysanthopoulos, P. K., and Klapa, M. I. (2008). Standardizing GC-MS metabolomics. *J. Chromatogr. B* 871, 191–201. doi: 10.1016/j.jchromb.2008.04.049

Khrunyk, Y., Münch, K., Schipper, K., Lupas, A. N., and Kahmann, R. (2010). The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus *Ustilago maydis*. *New Phytol.* 187, 957–968. doi: 10.1111/j.1469-8137.2010.03413.x

Kim, S., Lee, D. Y., Wohlgemuth, G., Park, H. S., Fiehn, O., and Kim, K. H. (2013). Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. *Anal. Chem.* 85, 2169–2176. doi: 10.1021/ac302881e

Kubicek, C. P., Punt, P., and Visser, J. (2011). Production of organic acids by filamentous fungi. *Acta Chimica Slov.* 51, 215–234. doi: 10.1007/978-3-642-11458-8_10

Laurie, J. D., Ali, S., Linning, R., Mannhaupt, G., Wong, P., Güldener, U., et al. (2012). Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell* 24, 1733–1745. doi: 10.1105/tpc.112.097261

Martinez-Espinoza, A. D., Garcia-Pedrajas, M. D., and Gold, S. E. (2002). The ustilaginales as plant pests and model systems. *Fungal Genet. Biol.* 35, 1–20. doi: 10.1006/fgbi.2001.1301

Martínez-Salgado, J. L., León-Ramírez, C. G., Pacheco, A. B., Ruiz-Herrera, J., and de la Rosa, A. P. B. (2013). Analysis of the regulation of the *Ustilago maydis* proteome by dimorphism, pH or MAPK and GCN5 genes. *J. Proteomics* 79, 251–262. doi: 10.1016/j.jprot.2012.12.022

Matei, A., and Doehlemann, G. (2016). Cell biology of corn smut disease—*Ustilago maydis* as a model for biotrophic interactions. *Curr. Opin. Microbiol.* 34, 60–66. doi: 10.1016/j.mib.2016.07.020

Moros, G., Chatziioannou, A. C., Gika, H. G., Raikos, N., and Theodoridis, G. (2016). Investigation of the derivatization conditions for GC-MS metabolomics of biological samples. *Bioanalysis* 9, 53–65. doi: 10.4155/bio-2016-0224

Münsterkötter, M., and Steinberg, G. (2007). The fungus *Ustilago maydis* and humans share disease-related proteins that are not found in *Saccharomyces cerevisiae*. *BMC Genomics* 8:473. doi: 10.1186/1471-2164-8-473

Oldiges, M., Lütz, S., Pflug, S., Schroer, K., Stein, N., and Wiendahl, C. (2007). Metabolomics: current state and evolving methodologies and tools. *Appl. Microbiol. Biotechnol.* 76, 495–511. doi: 10.1007/s00253-007-1029-2

Parthasarathy, A., Cross, P. J., Dobson, R. C. J., Adams, L. E., Savka, M. A., and Hudson, A. O. (2018). A three-ring circus: metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Front. Mol. Biosci.* 5:29. doi: 10.3389/fmolb.2018.00029

Pinu, F. R., Villas-Boas, S. G., and Aggio, R. (2017). Analysis of intracellular metabolites from microorganisms: quenching and extraction protocols. *Metabolites* 7:53. doi: 10.3390/metabo7040053

Ramluckan, K., Moodley, K. G., and Bux, F. (2014). An evaluation of the efficacy of using selected solvents for the extraction of lipids from algal biomass by the soxhlet extraction method. *Fuel* 116, 103–108. doi: 10.1016/j.fuel.2013.07.118

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., et al. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* 579, 1332–1337. doi: 10.1016/j.febslet.2005.01.029

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., et al. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330, 1546–1548. doi: 10.1126/science.1195330

Schuster, M., Schweizer, G., Reissmann, S., and Kahmann, R. (2016). Genome editing in *Ustilago maydis* using the CRISPR-Cas system. *Fungal Genet. Biol.* 89, 3–9. doi: 10.1016/j.fgb.2015.09.001

Steinberg, G., and Perez-Martin, J. (2008). *Ustilago maydis*, a new fungal model system for cell biology. *Trends Cell Biol.* 18, 61–67. doi: 10.1016/j.tcb.2007.11.008

Teichmann, B., Liu, L., Schink, K. O., and Bölker, M. (2010). Activation of the ustilagic acid biosynthesis gene cluster in *Ustilago maydis* by the C2H2 Zinc finger transcription factor Rua1. *Appl. Environ. Microbiol.* 76, 2633–2640. doi: 10.1128/aem.02211-09

Werpy, T., and Petersen, G. (2004). *Top Value Added Chemicals From Biomass : Volume I – Results of Screening for Potential Candidates From Sugars and Synthesis Gas*. Report No: DOE/GO-102004-1992. Golden, CO: National Renewable Energy Lab.

Winterberg, B., Uhlmann, S., Linne, U., Lessing, F., Marahiel, M. A., Eichhorn, H., et al. (2010). Elucidation of the complete ferrichrome A biosynthetic pathway in *Ustilago maydis*. *Mol. Microbiol.* 75, 1260–1271. doi: 10.1111/j.1365-2958.2010.07048.x

Wynands, B., Lenzen, C., Otto, M., Koch, F., Blank, L. M., and Wierckx, N. (2018). Metabolic engineering of *Pseudomonas* taiwanensis VLB120 with minimal genomic modifications for high-yield phenol production. *Metab. Eng.* 47, 121–133. doi: 10.1016/j.ymben.2018.03.011

Yi, L., Shi, S., Zhibiao, Y., He, R., Lu, H., and Liang, Y. (2014). MeOx-TMS derivatization for GC-MS metabolic profiling of urine and application in the discrimination between normal C57BL/6J and type 2 diabetic KK-Ay mice. *Anal. Methods* 6:4380. doi: 10.1039/c3ay41522h

Zambanini, T., Kleineberg, W., Sarikaya, E., Buescher, J. M., Meurer, G., Wierckx, N., et al. (2016). Enhanced malic acid production from glycerol with high-cell density *Ustilago trichophora* TZ1 cultivations. *Biotechnol. Biofuels* 9:135. doi: 10.1186/s13068-016-0553-7

Zheng, X., Yu, J., Cairns, T. C., Zhang, L., Zhang, Z., Zhang, Q., et al. (2019). Comprehensive improvement of sample preparation methodologies facilitates dynamic metabolomics of *Aspergillus niger*. *Biotechnol. J.* 14:1800315. doi: 10.1002/biot.201800315

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 06 October 2020
doi: 10.3389/fmolb.2020.561641

Perturbations of Metabolomic Profiling of Spleen From Rats Infected With *Clonorchis sinensis* Determined by LC-MS/MS Method

Xiaoli Zhang[1†], Xinyi Hu[1†], Rui Chen[2], Beibei Sun[1], Yannan Gao[1], Shanshan Duan[1], Liyan Liu[3] and Su Han[1]*

[1] Department of Parasitology, Harbin Medical University, Harbin, China, [2] Department of Orthopaedic Surgery, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, [3] Department of Nutrition and Food Hygiene, School of Public Health, Harbin Medical University, Harbin, China

OPEN ACCESS

Edited by:
Zheng-Jiang Zhu,
Shanghai Institute of Organic Chemistry (CAS), China

Reviewed by:
Yan Huang,
Sun Yat-sen University, China
Chris Zhu,
The Ohio State University, United States

***Correspondence:**
Su Han
hansu1982@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: *15 May 2020*
Accepted: *24 August 2020*
Published: *06 October 2020*

Citation:
Zhang X, Hu X, Chen R, Sun B, Gao Y, Duan S, Liu L and Han S (2020) Perturbations of Metabolomic Profiling of Spleen From Rats Infected With Clonorchis sinensis Determined by LC-MS/MS Method. Front. Mol. Biosci. 7:561641. doi: 10.3389/fmolb.2020.561641

Clonorchiasis is an important zoonotic parasitic disease worldwide. In view of the fact that parasite infection affects host metabolism, and there is an intricate relationship between metabolism and immunity. Metabolic analysis of the spleen could be helpful for understanding the pathophysiological mechanisms in clonorchiasis. A non-targeted ultra high performance liquid tandem chromatography quadrupole time of flight mass spectrometry (UHPLC-QTOF MS) approach was employed to investigate the metabolic profiles of spleen in rats at 4 and 8 weeks post infection with *Clonorchis sinensis* (*C. sinensis*). Then a targeted ultra-high performance liquid chromatography multiple reaction monitoring mass spectrometry (UHPLC-MRM-MS/MS) approach was used to further quantify amino acid metabolism. Multivariate data analysis methods, such as principal components analysis and orthogonal partial least squares discriminant analysis, were used to identify differential metabolites. Finally, a total of 396 and 242 significant differential metabolites were identified in ESI+ and ESI− modes, respectively. These metabolites included amino acids, nucleotides, carboxylic acids, lipids and carbohydrates. There were 38 significantly different metabolites shared in the two infected groups compared with the control group through the Venn diagram. The metabolic pathways analysis revealed that pyrimidine metabolism, aminoacyl-tRNA biosynthesis, purine metabolism and phenylalanine, tyrosine and tryptophan biosynthesis were significantly enriched in differential metabolites, which was speculated to be related to the disease progression of clonorchiasis. Furthermore, 15 amino acids screened using untargeted profiling can be accurately quantified and identifed by targeted metabolomics during clonrochiasis. These results preliminarily revealed the perturbations of spleen metabolism in clonorchiasis. Meanwhile, this present study supplied new insights into the molecular mechanisms of host-parasite interactions.

Keywords: *Clonorchis sinensis*, infection, spleen, non-targeted metabolomics, metabolic pathway

INTRODUCTION

Clonorchiasis caused by *Clonorchis sinensis* (*C. sinensis*) is an important public health problem globally. *C. sinensis* infected at least 15 million people in countries such as China, Vietnam, South Korea, and the Russian Far East (Kim et al., 2016). Eating uncooked freshwater fish containing *C. sinensis* metacercariae is the cause of the infection. The adults of *C. sinensis* parasitize in the intrahepatic bile duct. Infection with *C. sinensis* often leads to chronic hepatobiliary diseases, such as hepatic fibrosis and cholangiocarcinoma. Notably, *C. sinensis* has been classified as a Class I carcinogen by the International Agency for Research on Cancer (Grosse et al., 2009). Our previous research had found that *C. sinensis* infection induced the dysregulation of hepatic microRNA and hepatic apoptosis in rat models (Zhang et al., 2008; Han et al., 2016). However, the molecular pathogenesis of clonorchiasis is still not completely understood.

Compared with transcriptome and proteome analyses, metabolomics measures small molecules in biological samples, which could describe the metabolic phenotype in detail and the relevant metabolic disorders (Murphy, 2020). Metabolomics has been successfully applied to several fields, such as disease diagnosis, biomarker screening, and nutrition research (Kafsack and Llinás, 2010; Khamis et al., 2017). Metabolomics provides globally dynamic changes and contributes to reveal the underlying molecular mechanisms in the diseases (Shinde et al., 2017).

Recently, more and more researchers have combined metabolomics and immunology for scientific research. It is reported that inflammation, disruption of the tricarboxylic acid (TCA) cycle, amino acids metabolism, protein synthesis and oxidative phosphorylation are all related to the immune response during infection (Nguyen et al., 2018). While, immune cells could use pentose phosphate shunt, glutamine breakdown and fatty acid oxidation to meet their metabolic and functional needs (Ganeshan and Chawla, 2014). Interestingly, interaction between immunity and metabolism also plays an important role in the pathogenesis of parasitic diseases. For example, *Toxoplasma gondii* caused metabolic recombination of host cells (Zhou et al., 2016), and the anti-inflammatory and immunoregulatory effects of steroid hormones can affect the host immune responses to infection (Chen X. Q. et al., 2017). The proliferation of B cells in primary lymphoid follicles required amino acids and lipid components to form new cell membranes and organelles in *Fasciola hepatica* infection (Saric et al., 2010). These studies show that metabolites can significantly affect the immune system, and immune inflammatory responses are also related to metabolic changes.

Given the complex relationship between immunity and metabolism, and the importance of the spleen in the immune response, the metabolic profiles of spleen in *C. sinensis*-infected rats were investigated by non-targeted ultra high performance liquid tandem chromatography quadrupole time of flight mass spectrometry (UHPLC-QTOF MS) in this study. Based on the results of untargeted profiling, amino acid metabolism was chosen for further quantifying by a targeted ultra high performance liquid chromatography multiple reaction monitoring mass spectrometry (UHPLC-MRM-MS/MS) approach. The exploration of metabolic disorders and related biochemical pathways could be useful for enhancing our understanding of the pathophysiological mechanisms in clonorchiasis.

MATERIALS AND METHODS

Ethical Approval

This study was reviewed and ethically approved by the Medical Ethics Review Committee of Harbin Medical University. All animal experiments were performed on the basis of the Guide for the Care and Use of Laboratory Animals published by the Ministry of Science and Technology of the People's Republic of China. In the study, we made significant efforts to reduce animal suffering and the number of animals.

Animal Infection

Metacercariae of *C. sinensis* were collected from *Pseudorasbora parva* originating from the Songhuajiang River of Heilongjiang Province. The collection and preparation of metacercariae are described as following. First, we put the fish in an ice box at 0°C and shipped to the laboratory. Second, the fish were washed with tap water, broken up in a Waring Blender, and digested with a pepsin-HCl (0.6%) artificial gastric juice at 37°C for 12 h. Finally, the digested mixture was passed through three sieves with mesh sizes of 1000, 300, and 106 μm in sequence. A large number of pure metacercariae were harvested by centrifugation and stored at 0.1 M phosphate-buffered saline (PBS, pH = 7.4) at 4°C until used. Male wistar rats (5–6 weeks old) were purchased from the Harbin Medical University Laboratory Animal Center. The rats were fed standard laboratory food and drinking water. A total of 20 rats were individually infected orally with 50 metacercariae. Control rats (n = 10) were fed with 50 μl of sterile normal solution.

Tissue Collection and Detection of Infection

Clonorchis sinensis undergoes rapid development and the adults develop matured at 4 weeks post infection (wpi) after infected by metacercariae (Lun et al., 2005). Our previous study also found that the hepatocyte apoptosis index of *C. sinensis*-infected rats with increased from 4 wpi, and reached a peak at 8 wpi (Zhang et al., 2008). Additionally, liver iron deposits were also found apparently at 8 wpi (Han et al., 2017). Furthermore, the same time points were also selected in other *C. sinensis* infection study for pathogenic mechanism (Lee et al., 1987; Uddin et al., 2012). Hence, at 4 and 8 wpi, rats were sacrificed and spleens were rinsed with saline solution (0.9% NaCl w/v), and stored at −80°C until analysis. In addition, control animals were sacrificed at both time points. Feces were collected weekly and microscopically examined by the Kato-Katz method to determine whether the rats had successfully infected with *C. sinensis* (Hong et al., 2003).

Chemicals

Acetonitrile was purchased from Merck. In addition, ammonium acetate (NH_4AC), ammonium hydroxide (NH_4OH), formic acid (FA), and ammonium fluoride (NH_4F) were obtained from Sigma Aldrich.

Metabolites Extraction

A total of 50 mg of spleen sample was taken and placed in a EP tube, then added 1000 μL extraction solvent containing an internal target (V methanol: V acetonitrile: V water = 2:2:1, containing internal standard, 2-Chloro-L-phenylalanine, 2 μg/mL). Samples were homogenized in ball mill for 4 min at 45 Hz, then ultrasound treated for 5 min (incubated in ice water). After homogenization for 3 times and incubation for 1 h at −20°C to precipitate proteins, samples were centrifuged at 12,000 rpm for 15 min at 4°C. Supernatant (825 μL) was transferred into EP tubes. Extracts were dried in a vacuum concentrator without heating, and 200 μL extraction solvent (V acetonitrile: V water = 1:1) reconstitution was added into dried metabolites. Samples were vortexed for 30 s, sonicated for 10 min (4°C water bath) and centrifuged for 15 min at 12,000 rpm, 4°C. Subsequently, the clear supernatant was transferred into a fresh 2 mL LC/MS glass vial for analysis.

UHPLC-QTOF MS Analysis

LC-MS/MS analyses were performed using an UHPLC system (1290, Agilent Technologies) with a UPLC BEH Amide column (1.7 μm 2.1 × 100 mm, Waters) coupled to TripleTOF 6600 (Q-TOF, AB Sciex) in Shanghai Biotree biotech Co., Ltd.

The mobile phase consisted of 25 mM NH_4AC and 25 mM NH_4OH in water (pH = 9.75) (A) and acetonitrile (B) was carried with elution gradient as follows: 0 min, 95% B; 0.5 min, 95% B; 7 min, 65% B; 8 min, 40% B; 9 min, 40% B; 9.1 min, 95% B; 12 min, 95% B, delivered at 0.5 mL min^{-1}. The injection volume were 1 μL. The Triple TOF mass spectrometer was used for its ability to acquire MS/MS spectra on an information-dependent basis (IDA) during an LC/MS experiment. In this mode, the acquisition software (Analyst TF 1.7, AB Sciex) continuously evaluates the full scan survey MS data as it collects and triggers the acquisition of MS/MS spectra depending on preselected criteria. In each cycle, 12 precursor ions whose intensity greater than 100 were chosen for fragmentation at collision energy (CE) of 30 V (15 MS/MS events with product ion accumulation time of 50 ms each). ESI source conditions were set as following: Ion source gas 1 as 60 Psi, Ion source gas 2 as 60 Psi, Curtain gas as 35 Psi, source temperature 600°C, Ion Spray Voltage Floating (ISVF) 5000 or −4000 V in positive or negative modes, respectively.

During mass spectra collection, samples were placed in automatic sampler at 4°C. To monitor the stability and repeatability of the analytical system, quality control (QC) samples were prepared by pooling 10 μL of each sample and injected prior to analysis. And then blank and QC samples were injected every five samples injections throughout the analytical run.

UHPLC-MRM-MS/MS Analysis

The targeted UPHLC-MRM-MS/MS approach was applied to verified amino acids involved in clonorchiasis. Among 4, 8 wpi and control groups, eight randomly selected rats from each group were used for investigation. An aliquot of each individual spleen sample was precisely weighed and then transferred to an Eppendorf tube. After the addition of two little steel balls and 1000 μL of extract solvent (precooled at −20°C acetonitrile-methanol-water, 2:2:1, containing isotopically labeled internal standard mixture), the samples were vortexed for 30 s, homogenized at 40 Hz for 4 min, and sonicated for 5 min in ice-water bath, repeated for three times. Next, the samples were incubated at −40°C for 1 h, followed by centrifugation for 15 min at 12,000 rpm, 4°C. After that, 80 μL supernatant was transferred to an auto-sampler vial for UHPLC-MS/MS analysis.

The LC-MS/MS analyses were also performed in Shanghai Biotree biotech Co., Ltd. The mobile phase A was 1% formic acid in water, and the mobile phase B was 1% formic acid in acetonitrile. The column temperature was set at 35°C. The auto-sampler temperature was set at 4°C and the injection volume was 1 μL. An Agilent 6460 triple quadrupole mass spectrometer (Agilent Technologies), equipped with an AJS electrospray ionization (AJS-ESI) interface, was used for assay development. Typical ion source parameters were: capillary voltage = +4000/−3500 V, Nozzle Voltage = +500/−500 V, gas (N2) temperature = 300°C, gas (N2) flow = 5 L/min, sheath gas (N2) temperature = 250°C, sheath gas flow = 11 L/min, nebulizer = 45 psi. Agilent MassHunter Work Station Software (B.08.00, Agilent Technologies) was employed for MRM data acquisition and processing.

Data Preprocessing and Annotation

MS raw data files (.wiff) were converted to the mzXML format using ProteoWizard, and processed by R package XCMS (version 3.2). The preprocessing results generated a data matrix that consisted of the retention time (RT), massto-charge ratio (m/z) values, and peak intensity. R package CAMERA was used for peak annotation after XCMS data processing. Compound identification of metabolites was performed by comparing the accuracy of m/z values (<25 ppm), and MS/MS spectra were interpreted with an in house MS2 database (Shanghai Biotree biotech Co., Ltd.) established with authentic standards.

Statistical Analysis

After preprocessing the raw data, multivariate statistical analysis (principal component analysis, PCA; orthogonal partial least squares discriminant analysis, OPLS-DA) was performed using SIMCA software (V14.1, Sartorius Stedim Data Analytics AB, Umea, Sweden). The parameters values of R2 and Q2 were verified the fitness and predictive ability of the model. And the OPLS-DA permutation test proves that the original model has excellent stability and there is no over-fitting phenomenon. The P-value of Student's t-test was less than 0.05, and the Variable Importance in the Projection (VIP) of OPLS-DA model was greater than 1, so as to identify the metabolites expressed differently. Log2-fold change (FC) based on metabolite

abundance was used to assess the variation of the metabolites. The box-plots (data range, quartile range, and median values) were used to illustrate the spread and differences of samples from the target analysis.

Heatmaps were applied to describe the unbalanced metabolic profiles among *C. sinensis*-infected and control rats. Euclidean distance algorithm for similarity measure and average linkage clustering algorithm (clustering uses the centroids of the observations) for clustering were selected when performing hierarchical clustering. Based on deferentially expressed metabolite data (log2-scaled), heatmaps were structured by the MultiExperiment Viewer (MeV) v. 4.9 software[1].

Metabolites were identified by comparing the molecular mass data (*m*/*z*) of samples with the KEGG[2] database. According to online Kyoto Encyclopedia of Genes and Genomes (KEGG) database[3], we retrieved metabolites and extracted the corresponding pathways in KEGG. Compared with controls, further screening was performed on the pathway affected by *C. sinensis* infection using MetaboAnalyst 3.0[4].

RESULTS

Metabolic Profiles of Spleen During *C. sinensis* Infection

The spleen metabolites in *C. sinensis*-infected rats was analyzed by the non-targeted UHPLC-QTOF MS system. There were 2270 and 2282 ions detected in ESI+ and ESI− mode, respectively. A series of preparations and collation of the original data were performed, in order to better analyze the data. After deleting low-quality ions [relative standard deviation (RSD) >30%], a total of 2243 and 2274 ions in samples were recognized in ESI+ and ESI− mode, respectively.

In the PCA model, QC samples were successfully separated from the tested samples and clustered together. It was identified that the UHPLC-QTOF MS analysis obtained better stability and reproducibility. However, the PCA score plots could not clearly distinguish the infected group from the control group (**Supplementary Figure S1**). The parameters R2 and Q2 confirmed the validity of the PCA model as follows: ESI+ mode, R2X = 0.58; ESI− mode, R2X = 0.58 (**Supplementary Table S1**). Therefore, the OPLS-DA was used to analyze the metabolites and relationships among different infection groups. The OPLS-DA score scatter plots showed that there were significant separation between the different infection groups and the control group in the ESI+ and ESI− modes, respectively (**Figure 1** and **Supplementary Figure S2**). According to the permutation test results, the OPLS-DA model was proved to have good robustness without over fitting (**Figure 2** and **Supplementary Figure S3**). In addition, the OPLS-DA model parameter are shown in **Supplementary Table S1**.

Differential Metabolites During Different Infection Periods

According to the *P*-value <0.05 and VIP >1, a total of 396 and 242 potential metabolites of spleen involved in clonorchiasis were screened out in ESI+ and ESI− modes, respectively (**Supplementary Table S2**). And the results showed a clear difference between different groups in the volcano plots (**Figure 3** and **Supplementary Figure S4**) and heatmaps (**Figure 4** and **Supplementary Figure S5**). To obtain the significantly differential metabolites, based on FC > 2 or <0.5, and VIP > 1.5, we further identified 82 and 84 metabolites in ESI+ and ESI− modes, respectively. Next, a Venn diagram was constructed to show the metabolites in different infection stages (**Figure 5**). There were 38 significantly different metabolites shared in the two infected groups compared with the control group. These metabolites included amino acids, nucleotides, carboxylic acids, lipids and carbohydrates. Notably, significant differences were observed in the amino acid profiles.

Based on untargeted profiling analysis, targeted UHPLC-MRM-MS/MS quantitative analysis of 25 key amino acids was established and further evaluated in the spleen. Among them, 15 amino acids screened by untargeted profiling can be accurately quantified and mostly identified increased during clonrochiasis, including L-valine, L-asparagine, L-serine, L-methionine, L-phenylalanine, L-histidine, L-threonine, L-tryptophan, L-arginine, L-tyrosine, L-proline, L-alanine, and L-glutamine, expect for L-citrulline decreased at 8 wpi. While

[1]http://mev.tm4.org/
[2]www.genome.jp/kegg/
[3]http://geneontology.org/
[4]http://www.MetaboAnalyst.ca/

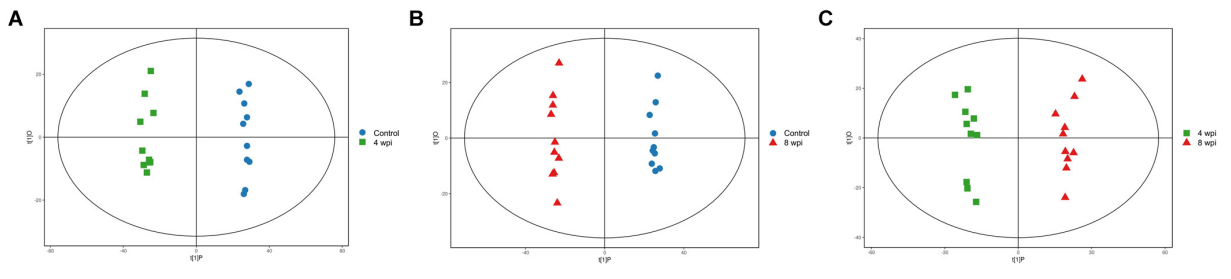


FIGURE 1 | Orthogonal partial least squares discriminant analysis (OPLS-DA) score scatter plots of splenic metabolites during *C. sinensis* infection in ESI+ mode. **(A)** OPLS-DA score scatter plot of 4 wpi vs control in ESI+ mode; **(B)** OPLS-DA score scatter plot of 8 wpi vs control in ESI+ mode; **(C)** OPLS-DA score scatter plot of 8 vs 4 wpi in ESI+ mode. Control, healthy control; 4 wpi, 4 weeks post infection; 8 wpi, 8 weeks post infection.

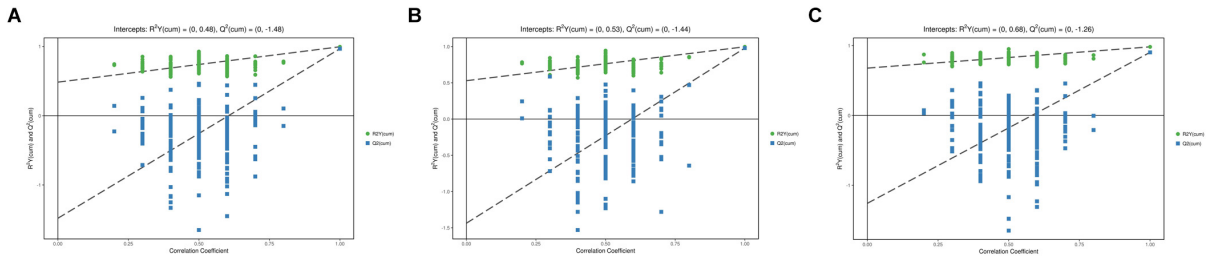


FIGURE 2 | Permutation test of OPLS-DA model showing the stability of the model in ESI+ mode. **(A)** Permutation test of OPLS-DA model of 4 wpi vs control in ESI+ mode; **(B)** Permutation test of OPLS-DA model of 8 wpi vs control in ESI+ mode; **(C)** Permutation test of OPLS-DA model of 8 vs 4 wpi in ESI+ mode. The abscissa indicates the displacement retention of the permutation test, and the ordinate indicates the value of R2Y or Q2. The green dot indicates the R2Y value obtained by the displacement test, the blue square indicates the Q2 value obtained by the permutation test, and the two dotted lines indicate the regression lines of R2Y and Q2, respectively. The point where the displacement retention is 1 is R2Y and Q2 of the original model.

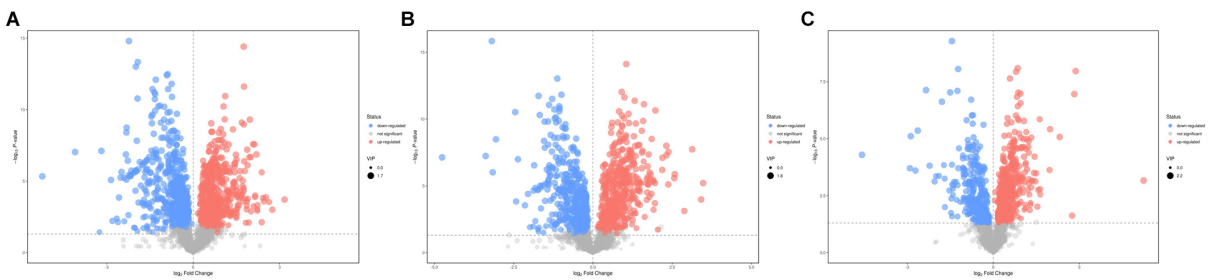


FIGURE 3 | Volcano plot representation of the differential metabolites identified in ESI+ mode. **(A)** 4 wpi vs control; **(B)** 8 wpi vs control; **(C)** 8 vs 4 wpi. Each point in the map represents a metabolite. The size of the scatter represents the VIP value of the OPLS-DA model, and the larger the scatter, the larger the VIP value. Scatter color represents the final screening result, red represents significant up-regulation, blue represents significant down-regulation, and gray represents non-significant difference metabolites.

glycine was identified decreased during clonrochiasis. Several amino acids are represented using boxplots in **Figure 6** and **Supplementary Figure S6**.

Metabolic Pathways Affected by *C. sinensis* Infection

Based on the *P*-value (P-value < 0.05) and pathway impact value, significant differential metabolic pathways involved in clonorchiasis were estimated by KEGG annotation and MetaboAnalyst. As showed in the bubble chart, the differential metabolites were enriched in purine metabolism, aminoacyl-tRNA biosynthesis, pyrimidine metabolism, phenylalanine, tyrosine and tryptophan biosynthesis at 4 wpi; and aminoacyl-tRNA biosynthesis, pyrimidine metabolism, purine metabolism, phenylalanine, tyrosine and tryptophan biosynthesis at 8 wpi (**Figure 7** and **Supplementary Figure S7**). As shown in the **Figure 8** and **Supplementary Tables S3, S4**, purine metabolism, pyrimidine metabolism, glycine, serine and threonine metabolism, and phenylalanine, tyrosine and tryptophan biosynthesis were integrated analyzed in the network.

DISSCUSION

Clonorchiasis is a public health problem worldwide with epidemiological significance (Lun et al., 2005). However, the underlying mechanism of clonorchiasis is not fully clearly. Recently, some studies have evaluated endogenous metabolites and identified perturbed metabolic pathways in parasitic diseases (Besteiro et al., 2010; Notarangelo et al., 2014). However, it is unknown whether *C. sinensis* infection could result in the imbalance of spleen metabolism and related biochemical pathways. In this study, the untargeted and targeted metabolomic methods were carried out to analyze the spleen metabolic profiles and related metabolic pathways in clonorchiasis. These results may help to explore the pathophysiology mechanism, making diagnosis and prevention strategies for clonorchiasis.

A number of animal models have been used to investigate the interaction of host and *C. sinensis*, such as rabbits, mice, hamsters, and rats. With low recovery rates and underdevelopment of worms (the parameter for host susceptibility), mice are not suitable hosts of *C. sinensis* regardless of its strain (Uddin et al., 2012). Because of the limited commercially available antibodies or probes, it is not appropriate for evaluating the pathogenesis or immune responses in hamsters (Uddin et al., 2012). Although the rats develop resistance to reinfection by *C. sinensis* (Quan et al., 2005), with the advantages of small size, low cost, wide source, easy feeding, and suitable for large-scale observation, rats have been broadly used as animal models for exploring pathogenesis of clonrochiasis (Choi et al., 2004; Quan et al., 2005; Sohn et al., 2006; Han et al., 2016, 2017). In addition, we could get better adults recovery rates and development of worms in rats

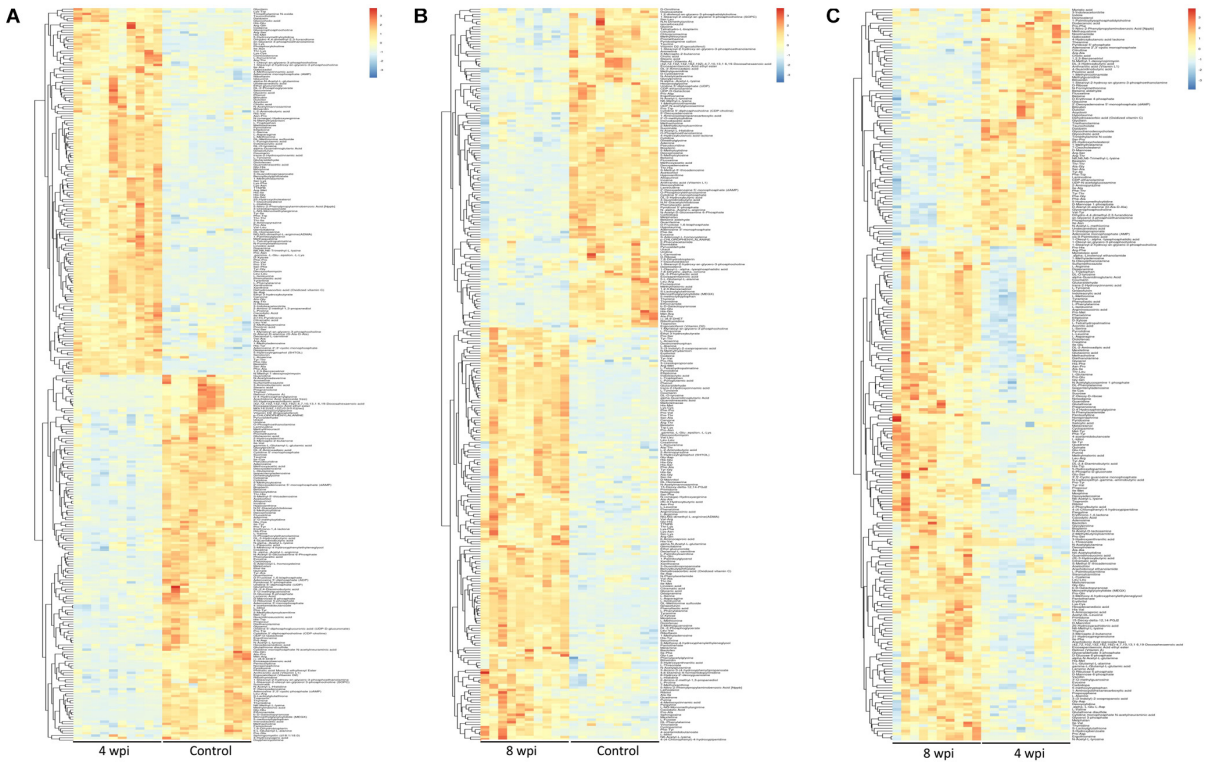


FIGURE 4 | Heatmaps representation of the differential metabolites identified between different infection group vs control group in ESI+ mode. **(A)** 4 wpi vs control; **(B)** 8 wpi vs control; **(C)** 8 vs 4 wpi. The color blocks at different positions represent the relative expression of metabolites at corresponding positions, red represents up-regulated, and blue represents down-regulated.

models. Some studies also had used rats models to investigate the immunity of clonorchiasis (Zhou et al., 2013; He et al., 2014). Thus, the rat models are considered as useful models for analysis the metabolomics of *C. sinensis* infection, and has more reference value for human body. First, it is difficult to collect the samples from clonorchiasis patients, including liver and spleen tissues. Second, some functions of the spleen from animal models are similar with human body. After preliminary studies in rat models, some metabolites are obtained and then could be further verified in clonrochiasis patients. Therefore, it is suited for making rat models for metabolomics analysis of clonorchiasis.

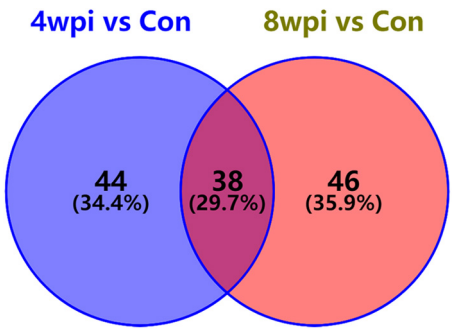


FIGURE 5 | Venn diagram displaying the common and unique metabolites among the two infection groups vs control group. There were 38 significantly differential metabolites were shared in two infected groups. 4 wpi vs con, 4 wpi group vs control group; 8 wpi vs con, 8 wpi group vs control group.

As a major immune organ, the spleen plays a crucial role in innate and adaptive immune responses (Zhao et al., 2015). Innate immunity is important for the immune surveillance of inner and outer threats as well as initial host defense responses (Chen J. et al., 2017). Immune surveillance could protect the host against parasites. Dendritic cells (DCs) have an essential role in immune surveillance (Cabeza-Cabrerizo et al., 2019). In addition, Toll-like receptors (TLRs) are important for against microbial infection (Pulendran et al., 2010). During *C. sinensis* infection, DCs played a key role in immune surveillance through TLR-mediated pathway, with increased levels of IFN-γ, IL-6, TNF-α, and IL-10 in the splenocytes (Hua et al., 2018). In view of the importance of the spleen in immune surveillance during *C. sinensis* infection, we chose the spleen as the research object.

In this study, a total of 396 and 242 significant differential metabolites were identified in ESI+ and ESI− modes, respectively. These metabolites covered amino acids, nucleotides, carboxylic acids, lipids and carbohydrates. The results suggested that *C. sinensis* infection could induce systemic metabolic perturbations in the spleen. In addition, there were 38 significantly different metabolites shared in the two infected groups compared with the control group, according to FC > 2 or <0.5, and VIP > 1.5. These metabolites might be a direct

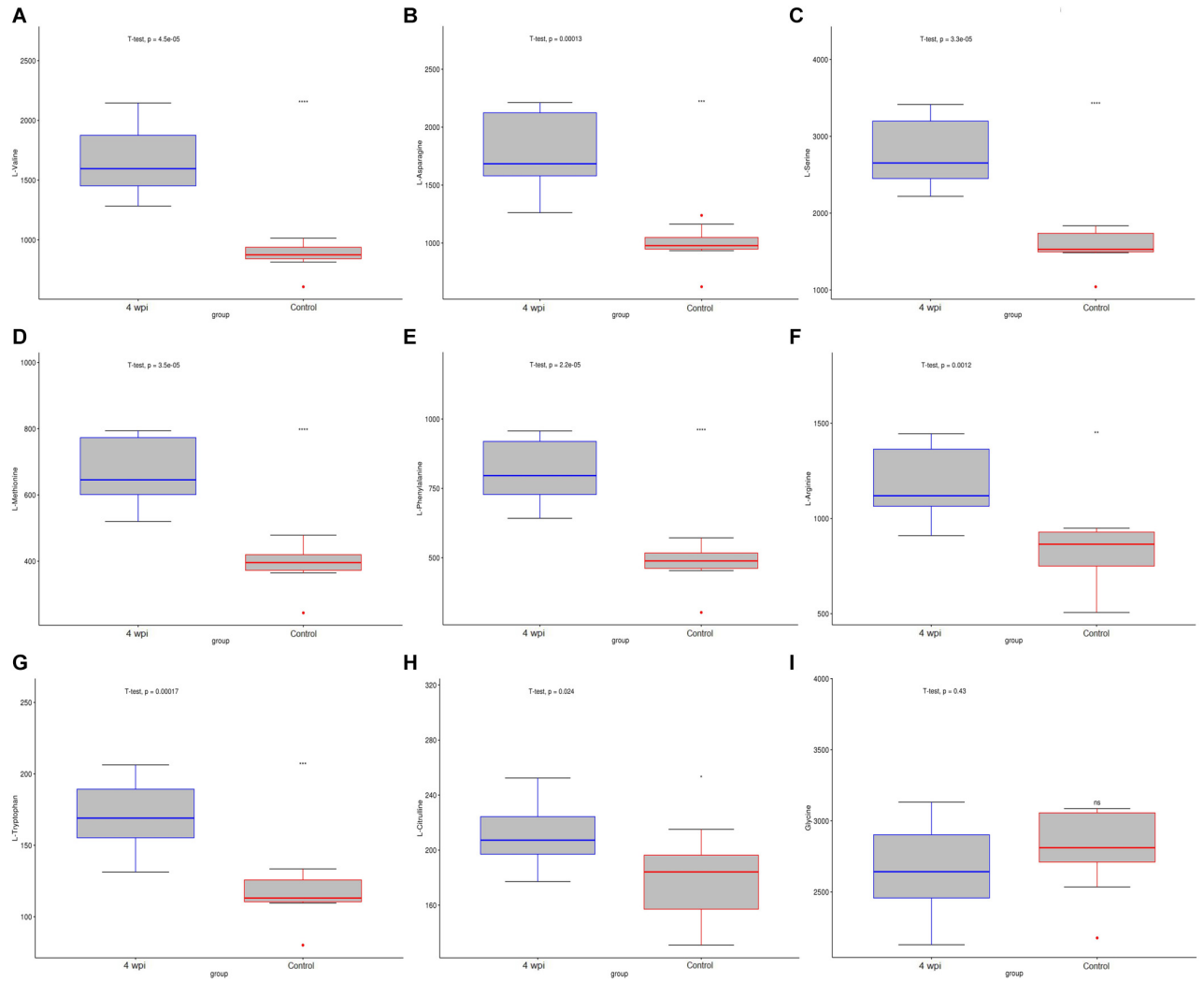


FIGURE 6 | The distribution of amino acids were present by boxplot between 4 wpi vs control group. **(A)** L-Valine; **(B)** L-Asparagine; **(C)** L-Serine; **(D)** L-Methionine; **(E)** L-Phenylalanine; **(F)** L-Arginine; **(G)** L-Tryptophan; **(H)** L-Citrulline; **(I)** Glycine. Boxes represent the interquartile ranges (IQRs) between the first and third quartiles, and the line inside the box represents the median. The amino acid content was compared with the median value between the two groups. Circles represent outliers. $^{*}P < 0.05$, $^{***}P < 0.001$.

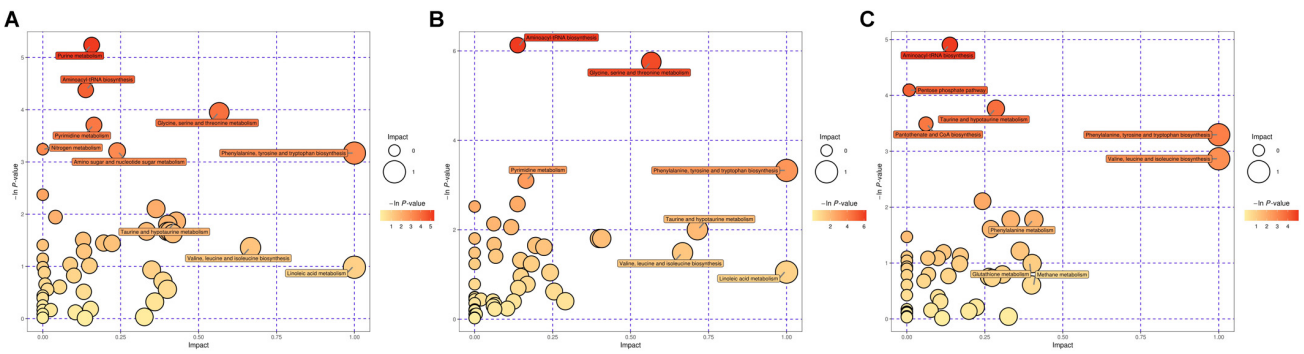


FIGURE 7 | The pathway analysis during *C. sinensis* infection in ESI+ mode. Plots depict the pathway impacts of the key metabolites (*x*-axis) and the computed metabolic pathway as a function of –log (P) (*y*-axis) that different between the 4 wpi vs control **(A)**, 8 wpi vs control **(B)**, 8 vs 4 wpi **(C)**.

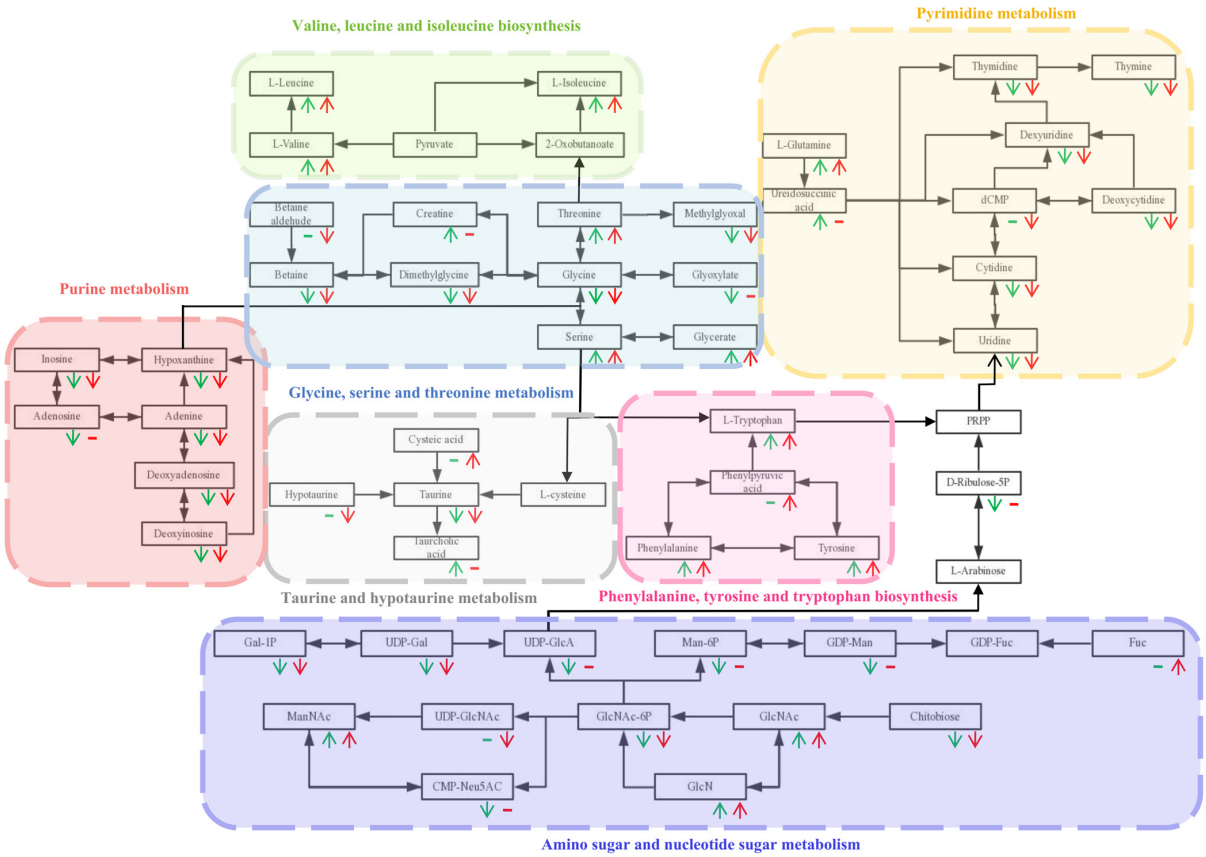


FIGURE 8 | The integrated metabolic pathway shared in 4 and 8 wpi during *C. sinensis* infection. Beneath each metabolite, green and red represent 4 and 8 wpi compared with the control, respectively. The arrows indicate up and down-regulation of metabolites, respectively. Horizontal lines indicate that the metabolites are not significantly different in this infection stage.

signal of *C. sinensis* activity or the consequence of the host response to the parasite. Generally, these dynamic metabolites not only could suggest the interaction between the host and *C. sinensis*, but also might perturb the biochemical profiles of them (Abdelrazig et al., 2017).

Our results found that *C. sinensis* infection induced significant changes in amino acids. In addition, 15 amino acids screened by untargeted profiling can be accurately quantified using targeted analysis during clonrochiasis. These results further corroborate the difference in abundance of amino acids between infection and control groups. Similarly, dysregulation in levels of amino acids metabolism were found and involved in toxoplasmosis and schistosomiasis (Wu et al., 2010; Chen et al., 2018). These amino acids not only affect cell signaling, recruitment and proliferation during infection, but reflect an immune response to infection and/or tissue injury and repair (Chandler et al., 2016). Besides, amino acids are also given the functions, such as energy dissipation, synthesis of basic organic molecules of proteins which could protect the host's innate immune response (Li P. et al., 2009; Lu et al., 2017). For example, L-leucine, L-glutamine, and L-valine have the same trend (up-regulation) at 4 and 8 wpi groups compared with the control, respectively. Moreover, compared with 4 wpi, these three animo acids were also found increased in 8 wpi. Leucine could provide energy when continuous energy consumption (Gironès et al., 2014). Glutamine is an important energy source for mitochondria and could be utilized by lymphocytes (Yaqoob and Calder, 1997). L-valine could alter the function of immune cells (Kakazu et al., 2007). Therefore, it indicates that the host is required to consume more energy, which is consistent with the results from *Schistosoma mansoni* infection (Li J. V. et al., 2009). Except that, we also found L-tryptophan, L-tyrosine, and L-arginine increased in *C. sinensis*-infected rats, compared with controls. These three amino acids were considered as essential amino acids required by the parasite to sustain its own growth (Silva et al., 2002; Marino and Boothroyd, 2017). Energy metabolism plays a key role in facilitating the adaptation of adult flukes to crowded habitat and hostile environment (Li et al., 2020). While, some metabolites were found fluctuated with disease progression. One possible interpretation of this result is that the host makes the regulation of the metabolism level. Therefore, the abnormality of these amino acids suggested that the host needs to maintain the energy requirements, sustain its own growth and regulates the immune system to interact with *C. sinensis* infection.

In addition, inosine and adenine were found down-regulated both in 4 and 8 wpi, compared with the control. Numerous

reports have suggested that inosine could effectively inhibit pro-inflammatory cytokines such as IFN-γ, TNF-α, and IL-12 *in vitro* and *in vivo* (Haskó et al., 2000; Mabley et al., 2003). Moreover the purines play a major role in regulating inflammatory and immune responses during diseases and reducing inflammatory tissue damage (Iregui et al., 2016). Therefore, these results indicate that the host regulates inflammation and immune responses to decrease infection in clonorchiasis.

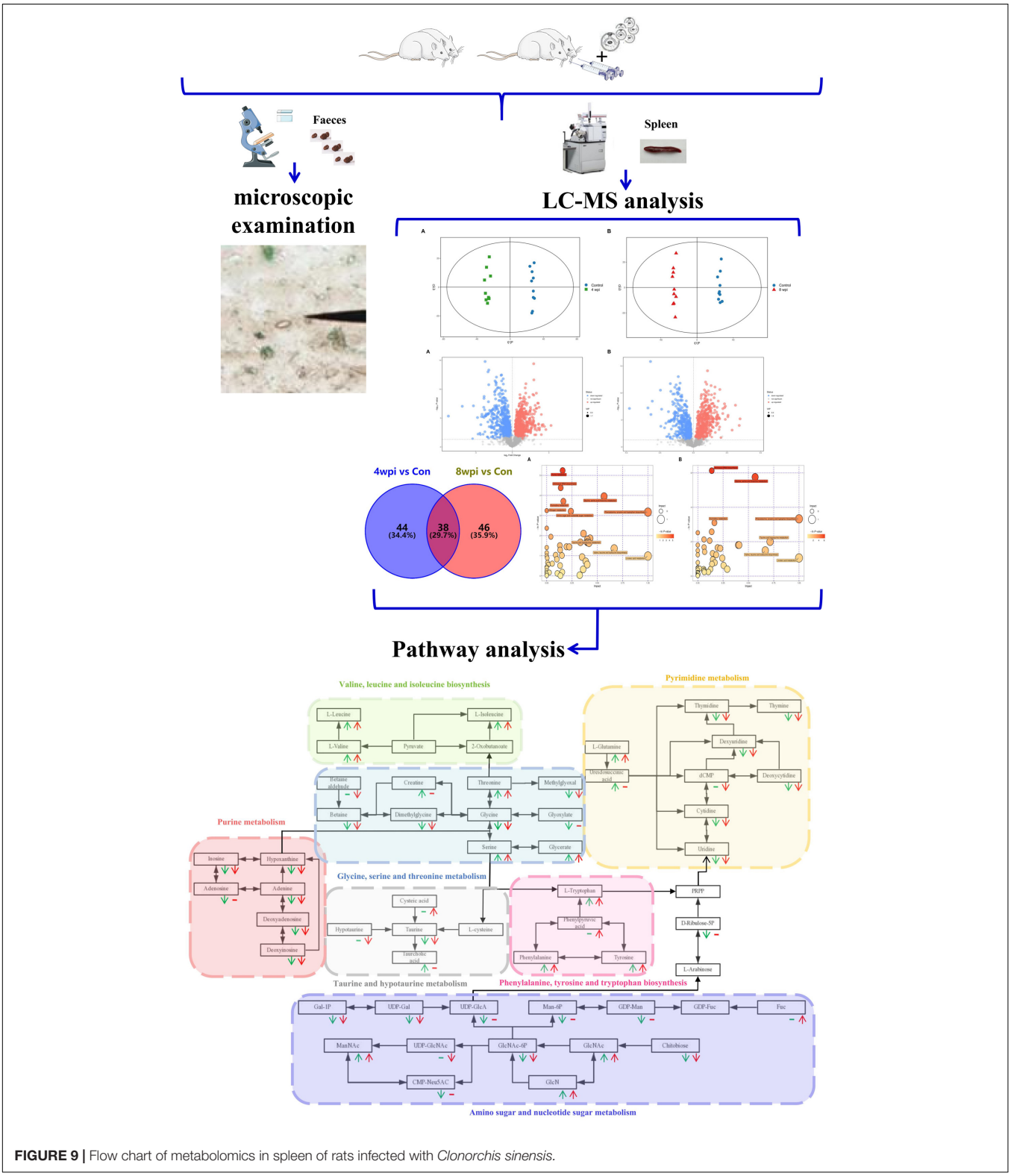


FIGURE 9 | Flow chart of metabolomics in spleen of rats infected with *Clonorchis sinensis*.

Besides, deoxyinosine is involved in the purine metabolism pathway, which could be substituted for glucose as an energy source (Chen et al., 2019). Meanwhile, deoxyinosine could be catabolized into hypoxanthine, regulating energy metabolism (Lee et al., 2018). Thus the decreased deoxyinosine and hypoxanthine in this present study may provide less energy and lead to compromised spleen cell development. Additionally, deoxyadenosine could dampen the function of immune cells by triggering the caspase-3-mediated death of macrophages (Carrera et al., 1990; Thammavongsa et al., 2013), which was down-regulated in this present study. As a precursor of DNA, thymidine was down-regulated, which indicating the degree of lymphocytes proliferation (Wang et al., 2015). Taken together, the metabolites suppressed immunity was significantly decreased, indicating that the host actively regulates the immune response, despite the spleen cells were damaged and the proliferation was inactive.

Generally, our research established the *C. sinensis*-infected rats model to anlayze the spleen metabolomics for the first time, revealing the biochemical characteristics and molecular mechanisms of infection (**Figure 9**). However, the present study also has several limitations. First, the numbers of the samples were low. Thus, larger samples with time-matched control groups should be assessed in future studies. Second, owing to the complex and dynamic cellular heterogeneity of the spleen, the altered metabolite may be the result of differential expression of spleen cell. However, our research serves to deepen the understanding of the molecular mechanisms of clonorchiasis. In the future, an individual cell population in the spleen will be selected for metabolomics or multi-omics joint analysis, and these differential metabolites requires further investigation.

CONCLUSION

The non-targeted UHPLC-QTOF MS method was used to explore the metabolic profiles in the spleen of *C. sinensis*-infected rats in this study. Differential metabolites included amino acids (L-leucine, L-glutamine, and L-valine) and nucleotides (inosine, deoxyadenosine, thymidine, and deoxyinosine). In addition, a total of 15 amino acids screened by untargeted profiling were accurately quantified and identifed by the targeted UHPLC-MRM-MS/MS approach. Several metabolic pathways were associated with the pathogenesis of clonorchiasis, including aminoacyl-tRNA biosynthesis, pyrimidine metabolism, and phenylalanine, tyrosine, and tryptophan biosynthesis. These results could be contributed to understanding the immunoregulatory process of *C. sinensis* infection and providing new insights into the molecular mechanisms of host-parasite interactions.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Medical Ethics Review Committee of Harbin Medical University.

AUTHOR CONTRIBUTIONS

SH and XZ conceived and designed the experiments. SH, XH, and XZ performed the experiments. SH and XH analyzed the data. SH, XH, RC, BS, YG, SD, and LL contributed to reagents, materials, and analysis tools. SH, XH, and XZ wrote the manuscript. All authors edited the manuscript, read, and approved the final version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81971958, 81601785, and 81401684), Natural Science of Heilongjiang province (YQ2020H006), Heilongjiang Province Post-doctoral Science Foundation (LBH-Q17114), Heilongjiang Province College Students Innovative Entrepreneurial Training Program (201810226003 and 201810226027). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2020.561641/full#supplementary-material

Supplementary Figure S1 | Principal component analysis (PCA) score scatter plots of splenic metabolites during *C. sinensis* infection. **(A)** PCA score scatter plot of splenic metabolites in the positive ion mode (ESI+); **(B)** PCA score scatter plot of splenic metabolites in the negative ion mode (ESI−); Control, healthy control; 4 wpi, 4 weeks post infection; 8 wpi, 8 weeks post infection; QC, quality control.

Supplementary Figure S2 | Orthogonal partial least squares discriminant analysis (OPLS-DA) score scatter plots of splenic metabolites during *C. sinensis* infection in ESI− mode. **(A)** OPLS-DA score scatter plot of 4 wpi vs control in ESI− mode; **(B)** OPLS-DA score scatter plot of 8 wpi vs control in ESI− mode; **(C)** OPLS-DA score scatter plot of 8 vs 4 wpi in ESI− mode. Control, healthy control; 4 wpi, 4 weeks post infection; 8 wpi, 8 weeks post infection.

Supplementary Figure S3 | Permutation test of OPLS-DA model showing the stability of the model in ESI− mode. **(A)** Permutation test of OPLS-DA model of 4 wpi vs control in ESI− mode; **(B)** Permutation test of OPLS-DA model of 8 wpi vs control in ESI− mode; **(C)** Permutation test of OPLS-DA model of 8 vs 4 wpi in ESI− mode; The abscissa indicates the displacement retention of the permutation test, and the ordinate indicates the value of R2Y or Q2. The green dot indicates the R2Y value obtained by the displacement test, the blue square indicates the Q2 value obtained by the permutation test, and the two dotted lines indicate the regression lines of R2Y and Q2, respectively. The point where the displacement retention is 1 is R2Y and Q2 of the original model.

Supplementary Figure S4 | Volcano plot representation of the differential metabolites identified in ESI− mode. **(A)** 4 wpi vs control; **(B)** 8 wpi vs control; **(C)** 8 vs 4 wpi; Each point in the map represents a metabolite. The size of the scatter represents the VIP value of the OPLS-DA model, and the larger the scatter, the larger the VIP value. Scatter color represents the final screening result, red

represents significant up-regulation, blue represents significant down-regulation, and gray represents non-significant difference metabolites.

Supplementary Figure S5 | Heatmaps representation of the differential metabolites identified between different infection group vs control group in ESI− mode. **(A)** 4 wpi vs control; **(B)** 8 wpi vs control; **(C)** 8 vs 4 wpi; The color blocks at different positions represent the relative expression of metabolites at corresponding positions, red represents up-regulated, and blue represents down-regulated.

Supplementary Figure S6 | The distribution of amino acids were present by boxplot between 8 wpi vs control group. **(A)** L-Valine; **(B)** L-Asparagine; **(C)** L-Serine; **(D)** L-Methionine; **(E)** L-Phenylalanine; **(F)** L-Arginine; **(G)** L-Tryptophan; **(H)** L-Citrulline; **(I)** Glycine. Boxes represent the interquartile ranges (IQRs) between the first and third quartiles, and the line inside the box represents the median. The amino acid content was compared with the median value between the two groups. Circles represent outliers. $^{*}P < 0.05$, $^{***}P < 0.001$.

Supplementary Figure S7 | The pathway analysis during *C. sinensis* infection in ESI− mode. Plots depict the pathway impacts of the key metabolites (*x*-axis) and the computed metabolic pathway as a function of −log (P) (*y*-axis) that different between the 4 wpi vs control **(A)**, 8 wpi vs control **(B)**, 8 vs 4 wpi **(C)**.

Supplementary Table S1 | Parameters of PCA and OPLS-DA models based on the date from sub-comparisons in the positive ion mode (ESI+) and the negative ion mode (ESI−).

Supplementary Table S2 | List of metabolites identified in the entire infection process.

Supplementary Table S3 | Summary of the top 10 pathway analysis using MetaboAnalyst during different infection groups in ESI+ mode.

Supplementary Table S4 | Summary of the top 10 pathway analysis using MetaboAnalyst during different infection groups in ESI− mode.

REFERENCES

Abdelrazig, S., Ortori, C. A., Davey, G., Deressa, W., Mulleta, D., Barrett, D. A., et al. (2017). A metabolomic analytical approach permits identification of urinary biomarkers for *Plasmodium falciparum* infection: a case-control study. *Malar. J.* 16:229. doi: 10.1186/s12936-017-1875-z

Besteiro, S., Vo Duy, S., Perigaud, C., Lefebvre-Tournier, I., and Vial, H. J. (2010). Exploring metabolomic approaches to analyse phospholipid biosynthetic pathways in Plasmodium. *Parasitology* 137, 1343–1356. doi: 10.1017/S0031182009991934

Cabeza-Cabrerizo, M., van Blijswijk, J., Wienert, S., Heim, D., Jenkins, R. P., Chakravarty, P., et al. (2019). Tissue clonality of dendritic cell subsets and emergency DCpoiesis revealed by multicolor fate mapping of DC progenitors. *Sci. Immunol.* 4:eaaw1941. doi: 10.1126/sciimmunol.aaw1941

Carrera, C. J., Terai, C., Lotz, M., Curd, J. G., Piro, L. D., Beutler, E., et al. (1990). Potent toxicity of 2-chlorodeoxyadenosine toward human monocytes in vitro and in vivo. A novel approach to immunosuppressive therapy. *J. Clin. Invest.* 86, 1480–1488. doi: 10.1172/JCI114865

Chandler, J. D., Hu, X., Ko, E. J., Park, S., Lee, Y. T., Orr, M., et al. (2016). Metabolic pathways of lung inflammation revealed by high-resolution metabolomics (HRM) of H1N1 influenza virus infection in mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 311, R906–R916. doi: 10.1152/ajpregu.00298.2016

Chen, J., Du, H., Cui, S., Liu, T., Yang, G., Sun, H., et al. (2017). *E. fischeriana* root compound dpo activates antiviral innate immunity. *Front. Cell. Infect. Microbiol.* 7:456. doi: 10.3389/fcimb.2017.00456

Chen, M., Zhang, B., Cai, S., Zeng, X., Ye, Q., Mao, X., et al. (2019). Metabolic disorder of amino acids, fatty acids and purines reflects the decreases in oocyte quality and potential in sows. *J. Proteomics* 200, 134–143. doi: 10.1016/j.jprot.2019.03.015

Chen, X. Q., Elsheikha, H. M., Hu, R. S., Hu, G. X., Guo, S. L., Zhou, C. X., et al. (2018). Hepatic metabolomics investigation in acute and chronic murine toxoplasmosis. *Front. Cell. Infect. Microbiol.* 8:189. doi: 10.3389/fcimb.2018.00189

Chen, X. Q., Zhou, C. X., Elsheikha, H. M., He, S., Hu, G. X., and Zhu, X. Q. (2017). Profiling of the perturbed metabolomic state of mouse spleen during acute and chronic toxoplasmosis. *Parasit. Vectors* 10:339. doi: 10.1186/s13071-017-2282-6

Choi, D., Hong, S. T., Li, S., Chung, B. S., Lim, J. H., and Lee, S. H. (2004). Bile duct changes in rats reinfected with *Clonorchis sinensis*. *Korean J. Parasitol.* 42, 7–17. doi: 10.3347/kjp.2004.42.1.7

Ganeshan, K., and Chawla, A. (2014). Metabolic regulation of immune responses. *Annu. Rev. Immunol.* 32, 609–634. doi: 10.1146/annurev-immunol-032713-120236

Gironès, N., Carbajosa, S., Guerrero, N. A., Poveda, C., Chillón-Marinas, C., and Fresno, M. (2014). Global metabolomic profiling of acute myocarditis caused by Trypanosoma cruzi infection. *PLoS Negl. Trop. Dis.* 8:e0003337. doi: 10.1371/journal.pntd.0003337

Grosse, Y., Baan, R., Straif, K., Secretan, B., El Ghissassi, F., Bouvard, V., et al. (2009). A review of human carcinogens–Part A: pharmaceuticals. *Lancet Oncol.* 10, 13–14. doi: 10.1016/s1470-2045(08)70286-9

Han, S., Tang, Q., Chen, R., Li, Y., Shu, J., and Zhang, X. (2017). Hepatic iron overload is associated with hepatocyte apoptosis during *Clonorchis sinensis* infection. *BMC Infect. Dis.* 17:531. doi: 10.1186/s12879-017-2630-3

Han, S., Tang, Q., Lu, X., Chen, R., Li, Y., Shu, J., et al. (2016). Dysregulation of hepatic microRNA expression profiles with *Clonorchis sinensis* infection. *BMC Infect. Dis.* 16:724. doi: 10.1186/s12879-016-2058-1

Haskó, G., Kuhel, D. G., Németh, Z. H., Mabley, J. G., Stachlewitz, R. F., Virág, L., et al. (2000). Inosine inhibits inflammatory cytokine production by a posttranscriptional mechanism and protects against endotoxin-induced shock. *J. Immunol.* 164, 1013–1019. doi: 10.4049/jimmunol.164.2.1013

He, L., Ren, M., Chen, X., Wang, X., Li, S., Lin, J., et al. (2014). Biochemical and immunological characterization of annexin B30 from *Clonorchis sinensis* excretory/secretory products. *Parasitol. Res.* 113, 2743–2755. doi: 10.1007/s00436-014-3935-4

Hong, S. T., Choi, M. H., Kim, C. H., Chung, B. S., and Ji, Z. (2003). The Kato-Katz method is reliable for diagnosis of *Clonorchis sinensis* infection. *Diagn. Microbiol. Infect. Dis.* 47, 345–347. doi: 10.1016/s0732-8893(03)00113-5

Hua, H., Du, Y., Ma, R., Zhang, B. B., Yu, Q., Li, B., et al. (2018). The regulatory roles of toll-like receptor 4 in secretions of type 1/Type 2 relative cytokines by splenocytes and dendritic cells exposed to *Clonorchis sinensis* excretory/secretory products. *Inflammation* 41, 213–220. doi: 10.1007/s10753-017-0679-1

Iregui, C. A., Comas, J., Vásquez, G. M., and Verján, N. (2016). Experimental early pathogenesis of Streptococcus agalactiae infection in red tilapia *Oreochromis* spp. *J. Fish Dis.* 39, 205–215. doi: 10.1111/jfd.12347

Kafsack, B. F., and Llinás, M. (2010). Eating at the table of another: metabolomics of host-parasite interactions. *Cell Host. Microbe* 7, 90–99. doi: 10.1016/j.chom.2010.01.008

Kakazu, E., Kanno, N., Ueno, Y., and Shimosegawa, T. (2007). Extracellular branched-chain amino acids, especially valine, regulate maturation and function of monocyte-derived dendritic cells. *J. Immunol.* 179, 7137–7146. doi: 10.4049/jimmunol.179.10.7137

Khamis, M. M., Adamko, D. J., and El-Aneed, A. (2017). Development of a validated LC- MS/MS method for the quantification of 19 endogenous asthma/COPD potential urinary biomarkers. *Anal. Chim. Acta* 989, 45–58. doi: 10.1016/j.aca.2017.08.007

Kim, T. S., Pak, J. H., Kim, J. B., and Bahk, Y. Y. (2016). *Clonorchis sinensis*, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: a brief review. *BMB Rep.* 49, 590–597. doi: 10.5483/bmbrep.2016.49.11.109

Lee, J. S., Wang, R. X., Alexeev, E. E., Lanis, J. M., Battista, K. D., Glover, L. E., et al. (2018). Hypoxanthine is a checkpoint stress metabolite in colonic epithelial energy modulation and barrier function. *J. Biol. Chem.* 293, 6039–6051. doi: 10.1074/jbc.RA117.000269

Lee, S. H., Hong, S. T., Kim, C. S., Sohn, W. M., Chai, J. Y., and Lee, Y. S. (1987). Histopathological changes of the liver after praziquantel treatment in

Clonorchis sinensis infected rabbits. *Kisaengchunghak Chapchi* 25, 110–122. doi: 10.3347/kjp.1987.25.2.110

Li, J. V., Holmes, E., Saric, J., Keiser, J., Dirnhofer, S., Utzinger, J., et al. (2009). Metabolic profiling of a Schistosoma mansoni infection in mouse tissues using magic angle spinning-nuclear magnetic resonance spectroscopy. *Int. J. Parasitol.* 39, 547–558. doi: 10.1016/j.ijpara.2008.10.010

Li, P., Mai, K., Trushenski, J., and Wu, G. (2009). New developments in fish amino acid nutrition: towards functional and environmentally oriented aquafeeds. *Amino Acids* 37, 43–53. doi: 10.1007/s00726-008-0171-1

Li, S., Chen, X., Zhou, J., Xie, Z., Shang, M., He, L., et al. (2020). Amino acids serve as an important energy source for adult flukes of *Clonorchis sinensis*. *PLoS Negl. Trop. Dis.* 14:e0008287. doi: 10.1371/journal.pntd.0008287

Lu, J., Shi, Y., Cai, S., and Feng, J. (2017). Metabolic responses of Haliotis diversicolor to Vibrio parahaemolyticus infection. *Fish Shellfish Immunol.* 60, 265–274. doi: 10.1016/j.fsi.2016.11.051

Lun, Z. R., Gasser, R. B., Lai, D. H., Li, A. X., Zhu, X. Q., Yu, X. B., et al. (2005). Clonorchiasis: a key foodborne zoonosis in China. *Lancet Infect. Dis.* 5, 31–41. doi: 10.1016/S1473-3099(04)01252-6

Mabley, J. G., Rabinovitch, A., Suarez-Pinzon, W., Haskó, G., Pacher, P., Power, R., et al. (2003). Inosine protects against the development of diabetes in multiple-low-dose streptozotocin and nonobese diabetic mouse models of type 1 diabetes. *Mol. Med.* 9, 96–104. doi: 10.2119/2003-00016.mabley

Marino, N. D., and Boothroyd, J. C. (2017). *Toxoplasma* growth in vitro is dependent on exogenous tyrosine and is independent of AAH2 even in tyrosine-limiting conditions. *Exp. Parasitol.* 176, 52–58. doi: 10.1016/j.exppara.2017.02.018

Murphy, R. (2020). An integrative approach to assessing diet-cancer relationships. *Metabolites* 10:123. doi: 10.3390/metabo10040123

Nguyen, T. V., Alfaro, A. C., Young, T., Ravi, S., and Merien, F. (2018). Metabolomics study of immune responses of New Zealand Greenshell™ mussels (*Perna canaliculus*) infected with pathogenic *Vibrio* sp. *Mar. Biotechnol.* 20, 396–409. doi: 10.1007/s10126-018-9804-x

Notarangelo, F. M., Wilson, E. H., Horning, K. J., Thomas, M. A., Harris, T. H., Fang, Q., et al. (2014). Evaluation of kynurenine pathway metabolism in *Toxoplasma gondii*-infected mice: implications for schizophrenia. *Schizophr. Res.* 152, 261–267. doi: 10.1016/j.schres.2013.11.011

Pulendran, B., Tang, H., and Manicassamy, S. (2010). Programming dendritic cells to induce T(H)2 and tolerogenic responses. *Nat. Immunol.* 11, 647–655. doi: 10.1038/ni.1894

Quan, F. S., Lee, J. B., Bae, J. S., Ohwatari, N., Min, Y. K., and Yang, H. M. (2005). Resistance to reinfection in rats induced by irradiated metacercariae of *Clonorchis sinensis*. *Mem. Inst. Oswaldo Cruz* 100, 549–554. doi: 10.1590/s0074-02762005000500016

Saric, J., Li, J. V., Utzinger, J., Wang, Y., Keiser, J., Dirnhofer, S., et al. (2010). Systems parasitology: effects of *Fasciola hepatica* on the neurochemical profile in the rat brain. *Mol. Syst. Biol.* 6:49. doi: 10.1038/msb.2010.49

Shinde, B. A., Dholakia, B. B., Hussain, K., Panda, S., Meir, S., Rogachev, I., et al. (2017). Dynamic metabolic reprogramming of steroidal glycol-alkaloid and phenylpropanoid biosynthesis may impart early blight resistance in wild tomato (*Solanum arcanum* Peralta). *Plant Mol. Biol.* 95, 411–423. doi: 10.1007/s11103-017-0660-2

Silva, N. M., Rodrigues, C. V., Santoro, M. M., Reis, L. F., Alvarez-Leite, J. I., and Gazzinelli, R. T. (2002). Expression of indoleamine 2,3-dioxygenase, tryptophan degradation, and kynurenine formation during in vivo infection with *Toxoplasma gondii*: induction by endogenous gamma interferon and requirement of interferon regulatory factor 1. *Infect. Immun.* 70, 859–868. doi: 10.1128/iai.70.2.859-868.2002

Sohn, W. M., Zhang, H., Choi, M. H., and Hong, S. T. (2006). Susceptibility of experimental animals to reinfection with *Clonorchis sinensis*. *Korean J. Parasitol.* 44, 163–166. doi: 10.3347/kjp.2006.44.2.163

Thammavongsa, V., Missiakas, D. M., and Schneewind, O. (2013). Staphylococcus aureus degrades neutrophil extracellular traps to promote immune cell death. *Science* 342, 863–866. doi: 10.1126/science.1242255

Uddin, M. H., Li, S., Bae, Y. M., Choi, M. H., and Hong, S. T. (2012). Strain variation in the susceptibility and immune response to *Clonorchis sinensis* infection in mice. *Parasitol. Int.* 61, 118–123. doi: 10.1016/j.parint.2011.07.002

Wang, Y. Y., Chang, X. L., Tao, Z. Y., Wang, X. L., Jiao, Y. M., Chen, Y., et al. (2015). Optimized codon usage enhances the expression and immunogenicity of DNA vaccine encoding *Taenia solium* oncosphere TSOL18 gene. *Mol. Med. Rep.* 12, 281–288. doi: 10.3892/mmr.2015.3387

Wu, J., Xu, W., Ming, Z., Dong, H., Tang, H., and Wang, Y. (2010). Metabolic changes reveal the development of schistosomiasis in mice. *PLoS Negl. Trop. Dis.* 4:e0000807. doi: 10.1371/journal.pntd.0000807

Yaqoob, P., and Calder, P. C. (1997). Glutamine requirement of proliferating T lymphocytes. *Nutrition* 13, 646–651. doi: 10.1016/s0899-9007(97)83008-0

Zhang, X., Jin, Z., Da, R., Dong, Y., Song, W., Chen, X., et al. (2008). Fas/FasL-dependent apoptosis of hepatocytes induced in rat and patients with *Clonorchis sinensis* infection. *Parasitol. Res.* 103, 393–399. doi: 10.1007/s00436-008-0985-5

Zhao, L., Liu, L., Guo, B., and Zhu, B. (2015). Regulation of adaptive immune responses by guiding cell movements in the spleen. *Front. Microbiol.* 6:645. doi: 10.3389/fmicb.2015.00645

Zhou, C., Bian, M., Liao, H., Mao, Q., Li, R., Zhou, J., et al. (2013). Identification and immunological characterization of thioredoxin transmembrane-related protein from *Clonorchis sinensis*. *Parasitol. Res.* 112, 1729–1736. doi: 10.1007/s00436-013-3331-5

Zhou, C. X., Elsheikha, H. M., Zhou, D. H., Liu, Q., Zhu, X. Q., and Suo, X. (2016). Dual identification and analysis of differentially expressed transcripts of porcine PK-15 cells and *Toxoplasma gondii* during in vitro infection. *Front. Microbiol.* 7:721. doi: 10.3389/fmicb.2016.00721

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fmolb.2021.630291

Neuroprotective Effects of Danshen Chuanxiongqin Injection Against Ischemic Stroke: Metabolomic Insights by UHPLC-Q-Orbitrap HRMS Analysis

Peipei Zhou[1†], Lin Zhou[1†], Yingying Shi[1,2], Zhuolun Li[1,2], Liwei Liu[1,2], Lihua Zuo[1,2], Jun Zhang[1,2], Shuhong Liang[1], Jian Kang[1], Shuzhang Du[1], Jing Yang[1], Zhi Sun[1,2] and Xiaojian Zhang[1,2*]*

[1] Pharmaceutical Department, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Precision Clinical Pharmacy Laboratory of Henan Province, Zhengzhou, China

OPEN ACCESS

Edited by:
Francois-Pierre Martin,
Nestlé Institute of Health Sciences,
Nestlé Research, Switzerland

Reviewed by:
Guoxiang Xie,
University of Hawaii Cancer Center,
United States
Gaurav Sharma,
University of Texas Southwestern
Medical Center, United States

***Correspondence:**
Xiaojian Zhang
zhxjyxb18@163.com
Zhi Sun
sunzhi2013@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: 17 November 2020
Accepted: 15 April 2021
Published: 07 May 2021

Citation:
Zhou P, Zhou L, Shi Y, Li Z, Liu L, Zuo L, Zhang J, Liang S, Kang J, Du S, Yang J, Sun Z and Zhang X (2021) Neuroprotective Effects of Danshen Chuanxiongqin Injection Against Ischemic Stroke: Metabolomic Insights by UHPLC-Q-Orbitrap HRMS Analysis. Front. Mol. Biosci. 8:630291. doi: 10.3389/fmolb.2021.630291

The incidence of cerebral ischemic stroke characterized by high mortality is increasing every year. Danshen Chuanxiongqin Injection (DSCXQ), a traditional Chinese medicine (TCM) preparation, is often applied to treat cerebral apoplexy and its related sequelae. However, there is a lack of systematic research on how DSCXQ mediates its protective effects against cerebral ischemia stroke. Metabolomic analysis based on UHPLC-Q-Orbitrap HRMS was employed to explore the potential mechanisms of DSCXQ on ischemic stroke induced by transient middle cerebral artery occlusion (MCAO). Pattern analysis and metabolomic profiling, combined by multivariate analysis disclosed that 55 differential metabolites were identified between Sham group and Model group, involving sphingolipid metabolism, glycerophospholipid metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, primary bile acid biosynthesis, pantothenate and CoA synthesis and valine, leucine and isoleucine biosynthesis pathways. DSCXQ could reverse brain metabolic deviations in stroke by significantly upregulating the levels of L-tryptophan, Lyso (18:0/0:0), LPC (18:2), Indole-3-methyl acetate, and downregulating the levels of sphinganine 1-phosphate, L-threonic acid, glutaconic acid and N6,N6,N6-Trimethyl-L-lysine. In our study, we focused on the neuroprotective effects of DSCXQ against neuroinflammatory responses and neuronal apoptosis on a stroke model based on sphingolipid metabolism. The expressions of Sphk1, S1PR1, CD62P, Bcl-2, Bax, and cleaved Caspase-3 in brain tissue were evaluated. The neurological deficit, cerebral infarct size and behavioral abnormality were estimated. Results showed that DSCXQ intervention significantly reduced cerebral infarct size, ameliorated behavioral abnormality, inhibited the expression of Sphk1, S1PR1, CD62P, Bax, Cleaved Caspase-3, while increased the level of Bcl-2, and prevented neuronal apoptosis. The limitations are that our study mainly focused on the verification of sphingolipid metabolism pathway in stroke, and while other metabolic pathways left unverified. Our study indicates that SphK1-SIP axis may potentiate neuroinflammatory responses and mediate brain

damage through neuronal apoptosis, and DSCXQ could suppress the activity of SphK1-SIP axis to protect brain tissue in cerebral ischemia. In conclusion, this study facilitates our understanding of metabolic changes in ischemia stroke and the underlying mechanisms related to the clinical application of DSCXQ.

Keywords: ischemic stroke, DSCXQ, UHPLC-Q-Orbitrap HRMS, metabolomics, anti-apoptotic

INTRODUCTION

Stroke is a dominant cause of long-term disability worldwide and the second most common cause of mortality after cardiovascular disease, bringing a massive socio-economic burden to the healthcare system and society (Feigin et al., 2016). At present tissue plasminogen activator (t-PA) remains the first choice of treatment in the clinic. However, a major limitation of r-tPA therapy is its narrow therapeutic window, which restrict the use of r-tPA as a long-term therapy (Fonarow et al., 2011). Although great progress has been made in understanding of the pathophysiology of stroke, a great portion of new drug trials have showed disappointing results (Jin et al., 2015; Hu et al., 2018), indicating that it is urgent to search for blood or brain biomarkers for stroke prognosis and new therapeutics.

Danshen Chuanxiongqin injection (CFDA approval # H52020959) is a TCM preparation composed of the extract of *Salvia miltiorrhiza Bunge* and *Conioselinum anthriscoides 'Chuanxiong'*. The injection was listed in 2004 and is currently included in « Guidelines on the Rational Use of Chinese Drugs in Ischemic Stroke » as a recommended medication for the treatment of cerebral infarction and coronary heart disease. Mechanistic studies showed that DSCXQ could inhibit the production of malondialdehyde (MDA), effectively eliminating oxygen free radicals in rats and increasing the resistance of vascular endothelium to thrombosis (Fei et al., 2017; Zhang et al., 2020). However, previous studies did not fully explore the relevant mechanism of DSCXQ therapy in stroke at the level of metabolites. The mechanism of action of TCM drugs may be revealed by traditional pharmacological experiments to a certain extent, while combining metabolomic studies may help uncovering the overall metabolic network, making it a useful combination for discovering multiple interactions among the TCM components (Lyu et al., 2018). Herein we tried to fill such gap by scientifically studying the neuroprotective effects of DSCXQ on cerebral ischemia at metabolomics level. In order to systematically and scientifically study the neuroprotective effect of DSCXQ on cerebral ischemia from the perspective of multi-component, multi-target and multi-pathway, Metabolomics combined with pharmacological approaches were adopted to provide a novel way for the diagnosis and treatment of ischemic stroke.

Abbreviations: DSCXQ, Danshen Chuanxinqin injection; MCAO, middle cerebral artery occlusion; I/R, ischemia reperfusion; Bcl-2, B-cell lymphoma-2; Bax, BCL2-associated X; MDA, malondialdehyde; t-PA, plasminogen activator; S1PR1, sphingosine 1-phosphate receptor-1; SphK1, sphingosine kinase 1; QC, quality control; TCM, traditional chinese medicine; CFDA, china food and drug administration.

In our previous study, A qualitative analytical method of ultraperformance liquid chromatography-quadrupole/orbitrap high resolution mass spectrometry (UHPLC-Q-Orbitrap HRMS) was established for identification and quantification of the constituents of Danshen-Chuanxiong Injection (see **Supplementary Info 1**) (Zhou et al., 2019b), which makes it possible to lay a solid foundation for our follow-up research. In the present study, ultra-high performance liquid chromatography coupled with a Q Exactive hybrid quadrupole-orbitrap high resolution mass spectrometry (UHPLC-Q-Orbitrap HRMS) was employed to profile metabolome of brain tissue in middle cerebral artery occlusion model rats and explore the intervention mechanism of DSCXQ. Immunohistochemistry, western blot, biochemical parameters, behavioral deficit and cognitive impairment were combined by metabolomics analysis to investigate the neuroprotective effects comprehensively, the potential biomarkers connecting with perturbed metabolic pathways were revealed. The limitations are that our study mainly focused on the verification of sphingolipid metabolism pathway in stroke, and while other metabolic pathways left unverified. This study unveil new insights in the understanding of pathological changes of stroke and the dynamic metabolomic profile with pharmacodynamic evaluation of DSCXQ based on metabolomics.

MATERIALS AND METHODS

Chemicals and Reagents

Danshen Chuanxiongqin Injection was provided by the First Affiliated Hospital of Zhengzhou University. Chemical standards for MS/MS analysis of predicted metabolites were obtained from Sigma-Aldrich (St.Louis, MO, United States). HPLC grade acetonitrile and methanol were obtained from Fisher Scientific (Fair Lawn, NJ, United States). HPLC grade formic acid was purchased from Aladdin Industrial Co., Ltd. (Shanghai, China). Ultra-pure water (18.2 M) was prepared daily by a Milli-Q water purification system (Millipore, Shanghai, China). All solutions were filtrated by 0.22 μm pore size filters before use.

Cerebral Ischemia/Reperfusion Model and Drug Administration

Adult male Sprague-Dawley rats (weighing 220-250 g) were obtained from experimental animal center of Zhengzhou university (Zhengzhou, China). Animals were fed with free water and diet under the controlled temperature (25 ± 2°C), humidity (60 ± 5%) and 12 h light/12 h dark cycle for 7 days to adapt to the environment. Rats were fasted overnight before the surgical

operation, all experiments were carried out in adherence with the standard guidelines for the Care and Use of laboratory animals from the National Institute of Health (NIH) and handled strictly according to obligations of the Animals Ethics Committee of Zheng Zhou University.

The MCAO rat model was induced by the intraluminal technique according to the method of originally described by Longa et al. (1989), with little modification. Briefly, the right common carotid artery (CCA), external carotid artery (ECA), and internal carotid artery (ICA) were isolated clearly and exposed with caution on rats which were anesthetized with 3% chloral hydrate (1 ml/100 g) intraperitoneally, then a poly nylon monofilament of 0.24 mm in diameter with tip rounded was inserted through ECA into the ICA to block the origin of MCA until the slightly resistance was felt, nearly 18-20 mm. rats were achieved 2 h cerebral ischemia and then pulled out the filament to complete 24 h reperfusion. Meanwhile, the temperature of rats were kept at 36.5–37.5°C with a thermostat-controlled heating pad. The sham operation was performed the same surgical procedures except for inserting a filament.

Rats were randomly divided into three groups with six in each: sham-operated (sham), model (MCAO) and DSCXQ-treated groups, During the whole experiments process, the research group carried out three batches of animal experiments. Rats in Model and DSCXQ group received ischemia-reperfusion (I/R) surgery, while rats of the Sham group underwent artery exposure and isolation without the insertion. DSCXQ was intravenously administered to rats for 7 days and rats in the Sham and Model groups were intravenously administered with the correspond volume of 0.9% saline in the same way. DSCXQ injection is a mixture of TCM compounds which contains extracts from salvia miltiorrhiza (200 mg/ml) and ligustrazine (see **Supplementary Info 1**). Rats were administered intravenously with DSCXQ (0.16 ml/100 g of body weight) once a day starting immediately post surgery.

Brain Sample Collection and Preparation

After 2 h of ischemia followed by 24 h reperfusion, rats were sacrificed and brains were quickly removed. Samples were collected with the method that 0.5 g brain tissue was homogenized with triple volume saline (w/v), and centrifuged at 3,000 rmp for 10 min, then the supernatant were separated and stored at −80°C for reserve. Sample of 100 ml was transferred into 1.5 ml EP tube and added with 300 μl methanol containing 500 ng/ml ketoprofen and 50 ng/ml 2-chloro-Lphenylalanine as internal standard. After vortexing for 3 min, the mixture was centrifuged at 13,000 rpm for 10 min at 4°C. Lastly, 200 ml of the supernatant was transferred to an autosampler vial to UHPLC-Q-Orbitrap HRMS for metabolomics analysis. To evaluate stability and accuracy of the UPLC-MS/MS analysis system, 20 μl of samples from each group were mixed and generated the pooled quality control (QC) samples which were used the same method to analyze companying with measured samples. Six QC samples were inserted every five injections during the whole measurement process for quality control.

UHPLC-MS/MS System Conditions

Chromatography

Chromatographic experiments were performed on Dionex Ultimate 3000 UHPLC system (Thermo Fisher Scientific, San Jose, CA, United States). An aliquot of 5 μL samples was injected into BEH C18 (2.1 × 100 mm, 1.7 μm) maintained at 4°C and the flow rate was 0.35 mL/min, the mobile phase was formed of solvent A (0.1% formic acid-water, V/V) and solvent B (acetonitrile), the gradient elution was as follows: 0-1 min, 95% A; 1-12 min, 0% A; 12-15 min, 95% A.

Mass Spectrometry

The MS spectrometry was performed on a Q-Orbitrap mass spectrometer with high resolution (Thermo Scientific, San Jose, United States) using a heat electrospray ionization (HESI) ion source, the main parameters were set as follows: capillary temperature of 320°C, Aux gas flow rate of 10 Arb, spray voltage of 3.5 kV (ESI^+)/2.8 kV (ESI^-), sheath gas flow rate of 40 Arb (+)/38 Arb (−), the full scan data acquired ranges from 80-1200 m/z with a resolution of 70,000 and a resolution of 17,500 resolution in MS^2 mode, samples were analyzed at 20, 30, 40 NCE (normalized collisional energy).

Evaluation of Neurological Deficit

Neurobehavioral dysfunction of rats was scored and evaluated by one investigator who was blinded to the experimental design. Neurological deficits were estimated by Zea Longa five-point scale as follows: 0 points indicated no neurobehavioral dysfunction; 1 points showed the nerve function injuries as left forepaw failed to extend fully which represented a mild focal neurologic deficit; 2 points suggested a moderate focal neurologic deficit accompanied by circling to the left; and a score of 3 meant a severe focal deficit as rats were fallen to the affected side and crawled slowly; 4 points represented rats did not walk spontaneously and had a depressed level of consciousness. The inclusion criterion of the model was a neurological function score of 1–3, and excluded a neurological function score of 0 or 4.

Measurement of Cerebral Infract Area

Cerebral infarct area was measured by 2,3,5-triphenyltetrazolium chloride (TTC) staining. TTC staining is conventional used for visualization of hypoxic brain tissue the fast and reliably and for measuring the size of cerebral infarction. After evaluating the neurological function score, rats were sacrificed using the approved protocol. The brain tissues of rats were removed and frozen at −20°C for 20 min, then sectioned into five slices (2 mm thick) along the coronal plane, after that the brain slices were stained with 2% TTC dissolved with 0.9% saline for 20 min at 37°C in the dark, following fixed in 4% paraformaldehyde overnight. Normal cerebral tissue was stained (red) whereas the infarct tissue unstained (white). The infracted tissue areas were analyzed by weighing the infarct area, and infarct area content was calculated as infarct weight/(infarct area + normal area weight) × 100%. The infarct areas were statistically analyzed as percentages of the total slice areas.

Immunohistochemical Determination of CD62P

Paraffin-embedded sections were used to assess the expression of CD62P according to standard histological procedures. The brain tissues were conventionally fixed in 4% paraformaldehyde and embedded in paraffin wax. The samples were serially sectioned at 4 μm by a tissue-slicing machine. After dewaxing and hydration. The blocking of endogenous peroxidases was achieved by incubating the sections in 3% hydrogen peroxide. After that, the tissue antigen was repaired by microwave, and the blocking solution containing goat serum incubated on sections at 37°C for 20 min. Then Rabbit Anti-CD62P antibody (Abcam, Cambridge, United Kingdom) was added to the sections and incubated overnight at 4°C. Wash slides three times and remove excess liquid from around the sections, goat antirabbit immunoglobulin G (IgG) antibody (Proteintech Group Inc., Wuhan, China) labeled with horseradish peroxidase was added to the sections and incubated at 37°C for 30 min. After reacted with DAB solution, the stained tissue sections were observed under a 200 × light microscope in three visual fields of the ischemic cortex region of the infarct.

TUNEL Staining

TUNEL staining was employed to estimate cell apoptosis following standard's instructions. Brain tissues were fixed in 10% formaldehyde in PBS for 24 h and embedded in paraffin blocks. The blocks were cut into 4 μm thickness, by heating the slides for 10 min at 70°C and followed by two 5-min incubations in a fresh xylene bath at room temperature in a dyeing jars. Then the tissue samples were transferred through a graded ethanol series to rehydrate. Nuclear proteins were stripped from the DNA by incubation 20 mg/mL of proteinase K solution for 30 min, and endogenous peroxidase was inactivated with 2% H_2O_2 for 5 min at room temperature. Sections were incubated in a buffer containing TdT in a humidified chamber for 30 min at 37°C. and digoxigenin labeled dUTP followed by digoxigenin-conjugated peroxidase treatment. Stop the reaction by incubating the slides in 2 × SSC for 15min. Washed the slides and incubated in Streptavidin HRP diluted with PBS at the ratio of 1:500 for 30 min at 37°C, DAB (diaminobenzidine) was used as the chromogen. The TUNEL-positive cells stained brown granules were considered apoptotic due to the binding of dUTP enzyme to 3′- OH terminal of broken DNA, which were observed under a 200 × light microscope in three visual fields of the ischemic cortex region.

Western Blotting Analysis

Western blotting analysis were performed according to our previous study (Zhou et al., 2019a). In brief, the total proteins of brains were extracted by ice cold cell lysis buffer and determined the total content by Bio-Rad DC Protein Assay Kit (BIO-RAD, China). Equal quantities of protein samples (40 μg) were loaded into SDS-PAGE and transferred to PVDF membranes. Membranes were probed with primary antibodies against Sphk1 (1:1000), S1PR1 (1:1000) (Affinity Biosciences, OH, United States), Bcl-2, Bax (1:1000), Cleaved Caspase-3 (1:500) (Cell Signaling Technology, Boston, MA, United States) at 4°C overnight. Subsequently, the membrane was washed with Tris-buffer saline containing 0.05% Tween 20 (TBST) buffer three times and probed with secondary antibody conjugated horseradish peroxidase. Protein visualization was achieved by the enhanced chemiluminescence reagents on a gel imaging system (Tannon-5200, Shanghai, China).

Data Processing and Statistical Analysis

The data was acquired and processed by Thermo XcaliburTM software (Version 3.0, Thermo Scientific), and Thermo Scientific Compound Discoverer 3.0 software was employed to pretreat LC-MS raw data. The spectra were chosen from LC-MS data files and retention time alignment was achieved according to mass tolerance and time shift criteria. Preliminary identification of metabolites was fulfilled by searching databases containing ChemSpider, Mass Lists, mzCloud, mzVault, and local database. Then the result data matrix were input into software SIMCA (version 14.0, Umetrics, Umea, Sweden) for multivariate statistical analysis, including the subsequent principal component analysis (PCA) and orthogonal partial least square discrimination analysis (OPLS-DA), The metabolites with variable importance in the projection (VIP) values >1.0 and p values <0.05 for Model versus Sham were screened as potential biomarkers of MCAO. In addition, student's t-test and fold change value were also applied to further screen out the significant variables between different groups.

Quantitative data were expressed as mean ± standard deviation (S.D.) using statistical software SPSS19.0. Differences were assessed by one-way analysis of variance (ANOVA) followed by a least significant difference t-test (LSD). Differences were considered significant at $p < 0.05$.

RESULTS

Assessment of QC Samples

QC samples are used to verify the stability of the experiment. In the process of metabolomics, QC samples were inserted every five injections during the whole measurement process. And the relative standard deviation (RSD) of the peak area in the chromatogram was used to calculate to evaluate the method performance. In the negative and positive mode, more than 90% of RSD is lower than 30%, which indicates that the analysis system is stable and reliable (**Figure 1**). Therefore, the differences of metabolic markers found in this study can truly reflect the differences of biological status among sample groups.

Multivariate Data Analysis for Brain Samples

Based on UPLC-MS/MS, the brain samples of rats were analyzed in positive and negative ion mode. In order to investigate the overall changes of metabolites in rats with cerebral ischemia-reperfusion, PCA method was used to identify the metabolites spectrum data of rats in Sham group and Model group, moreover the sample distribution map which reflect the degree of similarity

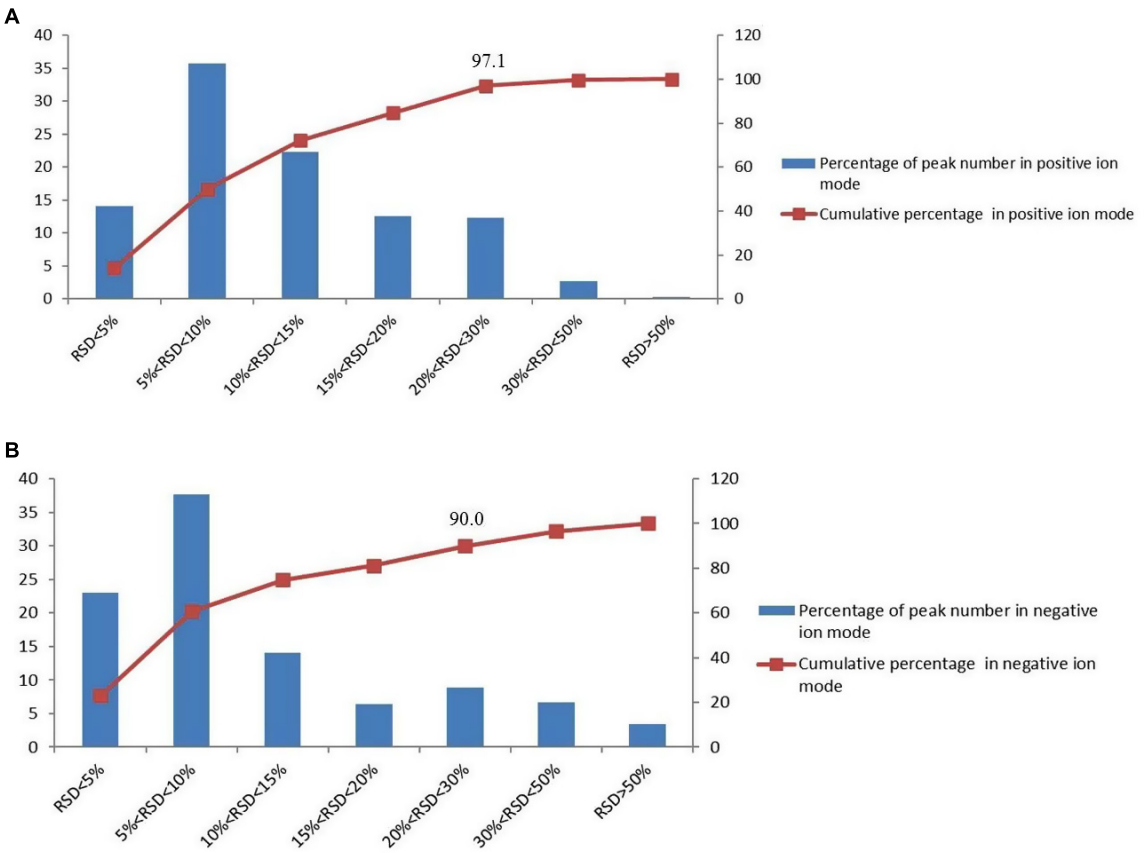


FIGURE 1 | The distributions of the coefficient of variation for each metabolite in QC samples, **(A)** percentage (%) of all detected peaks in positive mode; **(B)** percentage (%) of all detected peaks in negative mode.

and difference between samples was obtained. In the map, samples with small difference in atlas are close to each other, on the contrary, samples with large difference are far apart. In PCA model, Sham group, Model group and DSCXQ group are apparent separated on the scatter plot after automatic fitting (**Figure 2**), to further amplify the differences between groups, the supervised multidimensional analysis method OPLS-DA was adopted (see **Figure 3**). Sham group, Model group and DSCXQ group were well distinguished in PCA, **Figures 3A,B** shows that the Model group and Sham-operated group can be distinguished obviously under the positive and negative ion mode, which indicates the model was successful. Compared with the Sham group, the metabolites in Model group change obviously and abnormal. Significant distinctions were acquired with R^2Y at 0.988 and Q^2 at 0.869for ESI^+ mode (**Figure 3A**), and R^2Y at 0.985 and Q^2 at 0.848 for ESI^- mode (**Figure 3B**), meanwhile, the OPLS-DA model is validated by a permutation of 200 times and the results (R^2 = 0.928, Q^2 = −0.225 for ESI^+ mode; R^2 = 0.897, Q^2 = −0.275 for ESI^- mode). The OPLS-DA model between DSCXQ and Model group were performed to display the neuroprotection of DSCXQ vs. Model group. Notable differentiation was obtained with R^2Y at 0.987and Q^2 at 0.652 for ESI^+ mode (**Figure 3E**), and R^2Y at 0.996 and Q^2 at 0.709 for ESI^- mode (**Figure 3F**). To assess the validity of the OPLS-DA model, permutation test with 200 measurements was performed and the result (R^2 = 0.969, Q^2 = −0.111 for ESI^+ mode; R^2 = 0.96, Q^2 = −0.188 for ESI^- mode) indicated that there was no overfitting of the model.

Biomarkers Identification

In OPLS-DA analysis, each point in S-plot graph of Sham group, Model group and DSCXQ group represents a variable, and the importance of each point for classification is measured by their values, screened according to VIP (variable importance in the projection, VIP). VIP >1 is considered to be a significant variable contributing to the model. The farther away from the center, the greater the contribution of variables to the difference and the more likely they are to become potential characteristic metabolites. Metabolites distributed far from the origin play an important role in the S-plot graph (see **Figures 4A,B**). In addition, the t test and folding changes of the students were calculated to ensure that the metabolites detected were significantly changed in concentration. $p \leq 0.05$ and fold change ≥1.5 (or fold change ≤0.67) were retained in order to obtain more credible and distinct markers. In volcanic maps, the red dots indicate that the p values of metabolites are less than 0.05 and

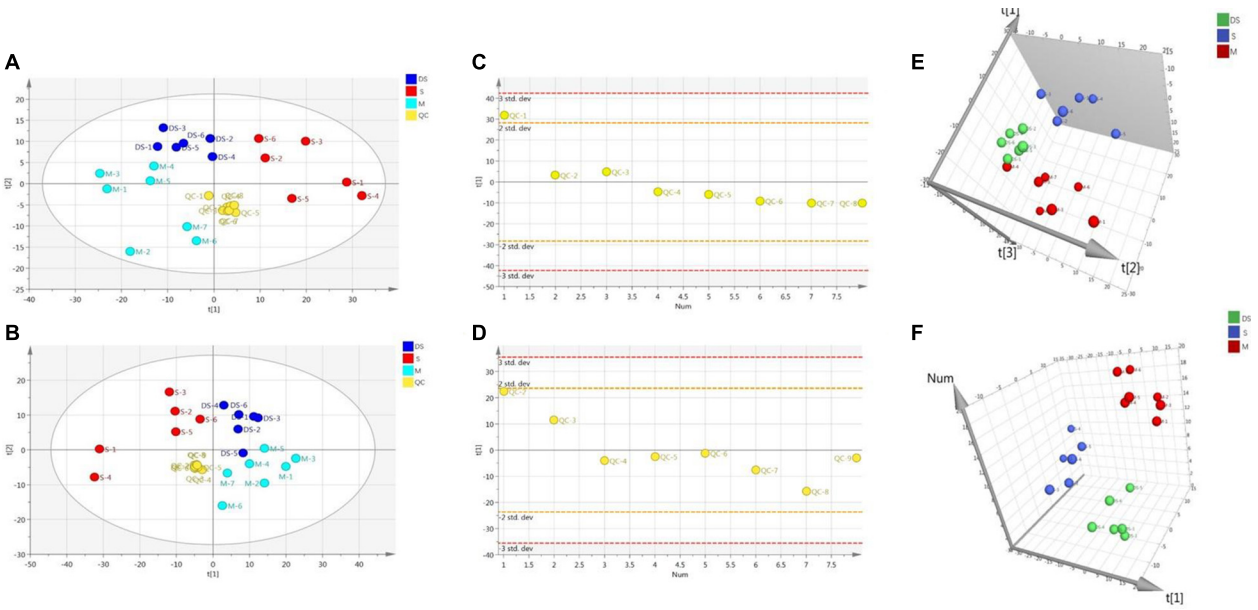


FIGURE 2 | (A) The PCA plot of samples and QC in positive ion mode, **(B)** The PCA plot of samples and QC in negative ion mode, **(C)** The scatter distribution plot in the first principal components in positive ion mode, **(D)** The scatter distribution plot in the first principal components in negative ion mode. (Note: QC represents quality control, S represents Sham group, M represents Model group, DS represents DSCXQ group), **(E)** 3-D PLS-DA score plot of the different groups in positive ion mode, **(F)** 3-D PLS-DA score plot of the different groups in negative ion mode.

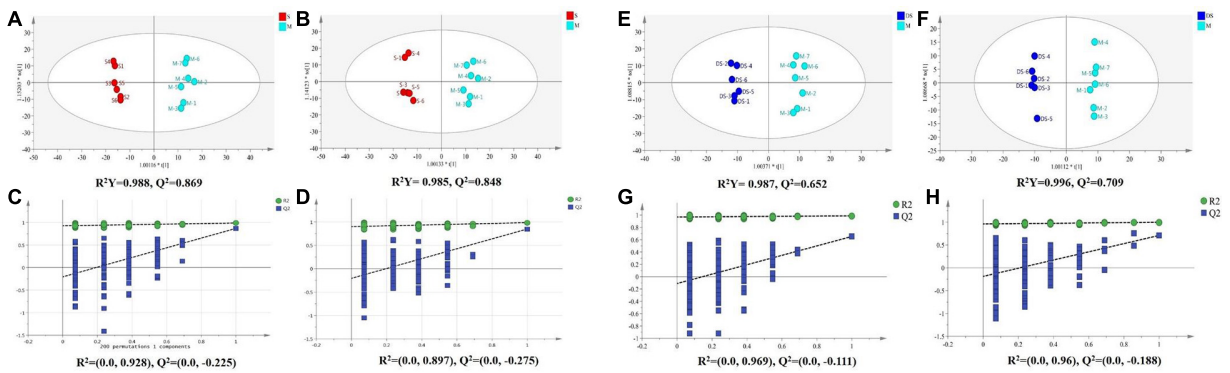


FIGURE 3 | (A) The OPLS-DA plot between Sham and Model group in positive ion mode, **(B)** The OPLS-DA plot between Sham and Model group in negative ion mode, **(C)** The permutations test of Sham vs. Model groups in positive ion mode, **(D)** The permutations test of Sham vs. Model groups in negative ion mode; **(E)** The OPLS-DA plot between Model and DSCXQ group in positive ion mode, **(F)** The OPLS-DA plot between Model and DSCXQ group in negative ion mode. **(G)** The permutations test of Model vs. DSCXQ groups in positive ion mode, **(H)** The permutations test of Model vs. DSCXQ groups in negative ion mode, (Note: S represents Sham group, M represents Model group, DS represents DSCXQ group).

fold change $\geq$1.5 or fold change $\leq$0.67 ($\log_2$FC $\geq$ 0.58 or $\log_2$FC $\leq$ −0.58), which help to determine the potential biomarkers of stroke.

The chemical structures of reserved metabolites were identified by searching accurate molecular mass data and MS/MS fragments in database such as HMDB, METLIN, M/Z cloud, and the database established by ourselves. Finally, in all 55 potential biomarkers that distinguish the difference between the Sham group and Model group were identified in positive and negative mode (**Table 1**). Of the potential biomarkers, it can be revealed that lipids metabolic perturbations play an important role in Model group, including the increased sphingolipids, declined lysophosphatidyl cholines (LysoPCs), lysophosphatidylethanolamines (LysoPEs) and bile acids. Secondly, it is amino acid metabolism, containing the decreased branched amino acids and the imbalance of arginine-NO cycle observed in Model group (shown in **Table 1**).

Metabolic Pathway Analysis

In order to clarify the pathogenesis of stroke, the metabolites were input into MetaboAnalyst 3.0 to construct metabolic pathways discovering the important pathways (**Figure 5**). And a metabolic correlation network (**Figure 6**) was established

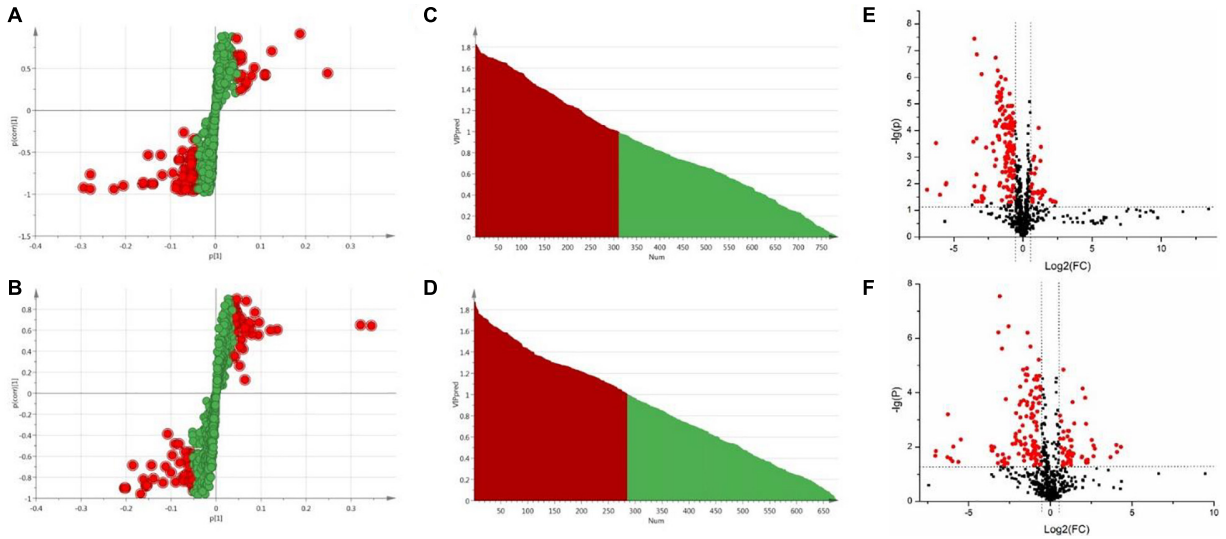


FIGURE 4 | (A) The S-plot of Sham vs. Model groups in positive ion mode, **(B)** The S-plot of Sham vs. Model groups in negative ion mode, **(C)** The VIP plot of Sham versus Model groups in positive ion mode, **(D)** The VIP plot of Sham vs. Model groups in negative ion mode, **(E)** The volcano plot of Sham vs. Model groups in positive ion mode, **(F)** The volcano plot of Sham vs. Model groups in negative ion mode.

by searching online database KEGG[1] and HMDB[2]. Metabolic dysregulations in rats with cerebral ischemia induced by MCAO are mainly related to six metabolic pathways, which involved in sphingolipid metabolism, glycerophospholipid metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, primary bile acid biosynthesis, pantothenate and CoA synthesis and valine, leucine and isoleucine biosynthesis.

In the pathway of phenylalanine, tyrosine and tryptophan biosynthesis, the levels of tyrosine, the metabolite of phenylalanine, decreased in Model compared to Sham group, which suggested the emergence of excitotoxicity induced by glutamate in cerebral ischemia (Kagiyama et al., 2004). A high level of N6,N6,N6-Trimethyl-L-lysine and threonine might correlate with disruption of the nervous system, for the metabolism of glycine, serine and threonine play an essential role in the function of the central nervous system (Tabatabaie et al., 2010; Amelio et al., 2014). The down-regulated tryptophan indicates that apoptosis mechanism was induced in Model group, and studies have revealed that the active degradation of kyn pathway in Trp were closely linked to the severity and long-term prognosis of stroke (Wang et al., 2015). In the pathway of valine, leucine and isoleucine biosynthesis, it was found that neurological dysfunction went hand in hand with the down regulation of catabolism in BCAAs(branched-chain amino acid, BACC), and decreased valine with stroke might be attributed to either an inverse feedback mechanism or the use as an energy source (Kimberly et al., 2013). Lipid metabolites are crucial to the progression of stroke and are likely to be a potential target of interventions for stroke. The levels of lipid-related metabolites, e.g., LysoPCs, LysoPEs, sphinganine, Sphinganine 1-phosphate and fatty acids were significantly altered in Model group. Decreased LysoPCs and LysoPEs levels maybe the inhibition of phospholipase A2 activity by LPCs *in vivo* (Cunningham et al., 2008). While elevated sphinganine and Sphinganine 1-phosphate were closely related to dysregulation of inflammatory responses and apoptosis (Maceyka et al., 2012). The altered levels of 3-hydroxybutyrylcarnitine, propionylcarnitine, Butyrylcarnitine and 2-Methylbutyroylcarnitine, indicated the metabolic dysregulations of fatty acid β oxidation and oxidative stress (Virmani and Binienda, 2004). Other metabolites are mostly basic organic acids from many basal metabolic pathways, such as citric acid, malic acid, 3-hydroxybutyric acid, most of which were TCA pathway intermediates (Gao et al., 2008). In addition, the decreased of bile acids, e.g., glycocholic acid, deoxycholic acid, glycoursodeoxycholic acid, and taurochenodesoxycholic acid, were observed in stroke rats, which implicated the gut microbiome was involved in the progression of stroke (Kasahara and Rey, 2019).

After intervened by DSCXQ, the disorder of metabolites has been improved. DSCXQ reversed brain metabolic deviations in stroke by significantly upregulating the levels of L-tryptophan, Lyso (18:0/0:0), LPC (18:2), Indole-3-methyl acetate, and downregulating the levels of sphinganine 1-phosphate, L-threonic acid, glutaconic acid and N6,N6,N6-Trimethyl-L-lysine. In the research, sphinganine 1-phosphate decreased remarkably after the DSCXQ treatment both in ESI^+ and ESI^- modes. Accordingly, we mainly concentrated on the sphingolipid metabolism pathway.

DSCXQ Injection Ameliorate Neurological Scoring and Reduce Infarct Area in Cerebral Ischemia

The neurological damage was evaluated by an observer who was blinded to the experiments using a four-point scale, as described above. Neurological scores were shown in **Figure 7C**,

[1]https://www.kegg.jp/kegg/pathway.html
[2]http://www.hmdb.ca/

and it was found that the neurological scores in DSCXQ group (score = 2) were lower compared with the Model group (score = 2.83), meaning the neurological symptoms were improved by the treated group. The results of TTC staining showed that the relative infarct area in the model group was increased significantly compared to sham group (**Figure 7A**), indicating our Model group was established successfully. With the treatment of DSCXQ, the area of cerebral infarction 10.86% was notably reduced in contrast with Model group 22.70% (**Figure 7B**, $p < 0.01$).

Effects of DSCXQ on the Expression of SphK1, S1PR1 in Brain Tissues

SphK1 is the dominant kinase for S1P production in the brain and exert a critical role in "sphingolipid rheostat." S1P receptors, S1PR1 is amongst the most abundant subtype of S1P receptors in the brain, which plays a crucial role in sustaining hallmark endothelial functions. To validate the key enzymes and receptors in sphingomyelin metabolism, SphK1and S1PR1 were estimated by western blot (**Figure 8**). The results shows that expression levels of SphK1 and S1PR1 were increased notably following MCAO ($p < 0.05$), and compared to Model group, DSCXQ attenuated the upregulated SphK1 and increased levels S1PR1 ($p < 0.05$). These results verify that SphK1-mediated S1P production is increased in the injured brain after cerebral ischemia, and DSCXQ reduces the SphK1-induced S1P levels by suppressing S1PR1 receptors, which plays a crucial role in endothelial functions and vascular development.

DSCXQ Reduced Immunoreactive Cells of CD62P

Platelets store large amounts of S1P that is only released upon activation (Nitzsche et al., 2021). CD62P, also known as P-selectin, can directly reflect the activation of platelets *in vivo* and interact with leukocytes to potentiate vascular injury. Immunohistochemistry was used to detect the expression of CD62P in cerebral cortex and hippocampus. In **Figure 8C**, there was little expression of CD62P in the sham operation group, however, the amount of CD62P in Model group was heavily activated, with the treatment of DSCXQ, the positive expression of CD62P decreased greatly compared with the Model group. The results, upregulation of CD62P + cells in Model group, further confirmed that the SphK1-S1P axis was activated, worsening the inflammatory response. And its downregulation by DSCXQ, indicating the neuroprotective effects of DSCXQ against neuroinflammatory responses.

Effects of DSCXQ on the Expression of Bcl-2, Bax, Cleaved Caspase-3 in Brain Tissues

SphK1/S1P axis has been involved in various physiological processes, including cell migration, survival, cell death, and it is a critical regulator of the sphingolipid rheostat (Maceyka et al., 2012). To further investigate the possible molecular mechanism of DSCXQ in stroke rats, the expression of Bcl-2, Bax, Cleaved Caspase-3 levels in brain tissue were determined by Western blot (**Figure 9**). The results revealed that the level of Bax and Cleaved Caspase-3 was up-regulated compared to Sham group ($p < 0.05$), while the content of Bcl-2 content in model group were down-regulated and the ratio of Bcl-2/Bax decreased notably compared with Sham group; With the treatment of DSCXQ, the result indicted that DSCXQ decreased the level of Bax and Cleaved Caspase-3, enhanced Bcl-2 greatly caused an increase in the ratio of Bcl-2/Bax significantly. The results, the remarkable downregulation of the ratio of Bcl-2/Bax in Model group, indicated that the balance between apoptosis and survival was broken and confirmed the SphK1-S1P axis was activated. With the treatment of DSCXQ, the downregulation of Bax, Cleaved Caspase-3, and up-regulated of Bcl-2, was achieved, implicating the neuroprotective effects of DSCXQ against neuronal apoptosis on a stroke model.

DSCXQ Prevented Neuronal Apoptosis in Brain Tissues

As shown in **Figure 9**, in Model group, the positive cell shrinkage were brown granules after staining, which was remarkably increased compared with Sham group and was significantly decreased after treatment with DSCXQ. The results indicate that DSCXQ can effectively inhibit neuron apoptosis in cerebral ischemia.

DISCUSSION

In our study, neuroprotective effects of DSCXQ on ischemic stroke and the affected metabolic pathways were revealed by UHPLC-Q-Orbitrap HRMS-based metabolomics approach with multivariate statistical analysis methods. The results demonstrated that DSCXQ mediates its neuroprotective effect against stroke through modifying multiple metabolic pathways involved in lipid metabolism, amino metabolism, oxidative stress, and especially sphingolipid metabolism. our study clearly indicate that SphK1-SIP axis may potentiate neuroinflammatory responses and mediate brain damage in neuronal apoptosis, and DSCXQ suppressed the activity of SphK1-SIP axis to protect brain tissue in cerebral ischemia. Our work paved the way to uncovering the disturbed metabolic pathways, facilitating our understanding on neuroprotective mechanisms of DSCXQ against stroke from metabolomic insights.

Sphingomyelin Metabolism and Apoptosis in Stroke

The metabolites of sphingolipids, such as ceramide, sphingosine and (dihydro)sphingosine-1-phosphate, are all signal lipid molecules and participate in many cell processes (Jin et al., 2003). A large number of documents show that Sph kinase (SphK), which phosphorylates Sph to form S1P, is a key regulator of the sphingolipid rheostat (Spiegel and Milstien, 2003), and SphK1 has been proved as the dominant kinase for S1P production in the brain and exert a critical role in "sphingolipid rheostat" (Maceyka et al., 2002; Blondeau et al., 2007). S1P receptors, S1PR1 is amongst the most abundant subtype of S1P receptors in the brain, which plays a crucial role in sustaining hallmark endothelial functions and could be as a regulator for microglial activation

TABLE 1 | Identified endogenous metabolites between Sham group vs. Model group and adjusted in DSCXQ group in rats brain.

No	m/z	Metabolites	Formula	RT (min)	VIP	Fold change (M/S)	Fold change (DS/M)	Pathway Involved	ppm
1	118.086	L-Valine#	$C_5H_{11}NO_2$	0.84	1.699	0.567	1.067 ↑	Valine, leucine and isoleucine biosynthesis	−3.431
2	147.125	Acetylcholine#	$C_7H_{16}NO_2$	0.828	1.374	1.930	0.747 ↓	Glycerophospholipid metabolism	−2.705
3	153.066	N1-Methyl-2-pyridone-5-carboxamide	$C_7H_8N_2O_2$	1.179	1.579	0.308	0.473	Nicotinate and nicotinamide metabolism	−3.489
4	175.118	D-Arginine#	$C_6H_{14}N_4O_2$	0.756	1.361	1.504	0.822 ↓	D-Arginine and D-ornithine metabolism	−2.868
5	176.103	Citrulline#	$C_6H_{13}N_3O_3$	0.824	1.551	0.649	0.836	Arginine and proline metabolism	−2.713
6	182.081	L-Tyrosine#	$C_9H_{11}NO_3$	1.104	1.559	0.481	1.058 ↑	Phenylalanine, tyrosine and tryptophan biosynthesis	−0.438
7	188.071	Indoleacrylic acid	$C_{11}H_9NO_2$	5.113	1.407	0.248	4.073 ↑**	tryptophan metabolism	−1.144
8	189.160	N6,N6,N6-Trimethyl-L-lysine#	$C_9H_{20}N_2O_2$	0.76	1.597	2.473	0.586 ↓**	Lysine degradation	−1.609
9	204.123	L-Acetylcarnitine#	$C_9H_{17}NO_4$	1.076	1.581	0.487	1.010 ↑	Beta Oxidation of Very Long Chain Fatty Acids	−1.737
10	205.097	L-Tryptophan#	$C_{11}H_{12}N_2O_2$	3.042	1.680	0.541	1.256 ↑*	Tryptophan metabolism	−1.044
11	218.138	Propionylcarnitine	$C_{10}H_{19}NO_4$	1.328	1.102	1.600	0.767 ↓	Oxidation of Branched Chain Fatty Acids	−1.305
12	220.118	Pantothenic acid#	$C_9H_{17}NO_5$	2.269	1.245	2.159	0.762 ↓	Pantothenate and CoA biosynthesis	−1.041
13	232.154	Butyrylcarnitine	$C_{11}H_{21}NO_4$	2.742	1.058	1.638	0.723 ↓	Synthesis and degradation of ketone bodies	−1.355
14	246.170	2-Methylbutyroylcarnitine	$C_{12}H_{23}NO_4$	3.534	1.065	1.642	0.543 ↓*	Synthesis and degradation of ketone bodies	−1.725
15	248.149	(R)-3-hydroxybutyryl carnitine	$C_{11}H_{21}NO_5$	1.075	1.225	0.491	0.530	Synthesis and degradation of ketone bodies	−1.931
16	302.305	Sphinganine#	$C_{18}H_{39}NO_2$	7.434	1.141	2.404	0.764 ↓	Sphingolipid metabolism	−2.402
17	450.321	Chenodeoxycholic acid#	$C_{26}H_{43}NO_5$	6.767	1.253	0.008	1.393 ↑	Primary bile acid biosynthesis	−1.428
18	466.316	Glycocholic acid	$C_{26}H_{43}NO_6$	5.926	1.615	0.013	2.956 ↑	Primary bile acid biosynthesis	−1.155
19	468.308	LysoPC(14:0/0:0)	$C_{22}H_{46}NO_7P$	7.515	1.361	0.591	1.244 ↑	Arachidonic Acid Metabolism	−2.468
20	480.308	LysoPE(18:1/0:0)	$C_{23}H_{46}NO_7P$	8.593	1.498	0.602	0.972	Glycerophospholipids Metabolism	−2.219
21	482.324	LysoPE(18:0/0:0)	$C_{23}H_{48}NO_7P$	7.942	1.782	0.332	1.343 ↑*	Glycerophospholipids Metabolism	−2.313
22	497.347	1-palmitoylglycero phosphocholine	$C_{24}H_{51}NO_7P$	8.203	1.382	0.646	1.030 ↑	Phospholipid metabolism	−1.865
23	508.376	LysoPC(P-18:0)	$C_{26}H_{54}NO_6P$	8.774	1.675	0.407	1.098 ↑	Glycerophospholipids Metabolism	−0.829
24	510.355	LysoPC(17:0)	$C_{25}H_{52}NO_7P$	8.811	1.436	0.442	0.993	Glycerophospholipid metabolism	−0.247
25	520.340	LPC(18:2)	$C_{26}H_{50}NO_7P$	8.004	1.717	0.538	1.378 ↑	Glycerophospholipid metabolism	−1.952
26	522.355	LysoPC(18:1)	$C_{26}H_{52}NO_7P$	8.56	1.559	0.534	1.061 ↑	Glycerophospholipid metabolism	−1.887
27	570.355	LysoPC(22:5)	$C_{30}H_{52}NO_7P$	8.372	1.062	0.350	1.075 ↑	Glycerophospholipid metabolism	1.375

(Continued)

TABLE 1 | Continued

No	m/z	Metabolites	Formula	RT (min)	VIP	Fold change (M/S)	Fold change (DS/M)	Pathway Involved	ppm
28	572.371	LysoPC(22:4)	$C_{30}H_{54}NO_7P$	8.589	1.409	0.359	1.287 ↑	Glycerophospholipid metabolism	−2.194
29	703.575	Palmitoyl sphingomyelin	$C_{39}H_{79}N_2O_6P$	7.987	1.708	0.614	1.001↑	Sphingolipids metabolism	−2.375
30	324.964	Trichloroethanol glucuronide	$C_8H_{11}Cl_3O_7$	3.895	1.215	2.241	0.586 ↓	Glucuronic Acid Derivatives	4.173
31	105.055	3-Hydroxybutyric acid[#]	$C_4H_8O_3$	1.236	1.439	0.340	0.541	Ketone metabolism	7.855
32	131.070	2-Methyl-3-ketovaleric acid	$C_6H_{10}O_3$	3.714	1.556	0.469	1.157 ↑	leucine metabolism	6.659
33	393.300	Deoxycholic acid[#]	$C_{24}H_{40}O_4$	7.57	1.118	0.143	0.756	bile acid metabolism	3.000
34	407.279	3,7-Dihydroxy-12-oxocholanoic acid	$C_{24}H_{38}O_5$	6.437	1.073	0.139	0.408	bile acid metabolism	2.022
35	135.029	Malic acid[#]	$C_4H_6O_5$	0.843	1.169	0.443	0.884	TCA cycle	6.242
36	193.034	Citric acid[#]	$C_6H_8O_7$	0.837	1.080	0.500	0.665	Citrate cycle (TCA cycle)	5.659
37	117.054	Levulinic acid	$C_5H_8O_3$	1.849	1.532	0.507	1.070 ↑	Gamma-ketoacid and derivatives	6.601
38	119.070	2-Hydroxyvaleric acid	$C_5H_{10}O_3$	2.927	1.747	3.927	1.070 ↓	fatty acid metabolism	7.085
39	199.057	4-Methyldibenzothiophene	$C_{13}H_{10}S$	3.484	1.704	0.450	1.292 ↑	Cell membrane	6.102
40	131.034	Glutaconic acid	$C_5H_6O_4$	0.836	1.197	1.843	0.520 ↓*	glutamic acid metabolism	5.696
41	137.044	L-threonic Acid	$C_4H_8O_5$	0.83	1.364	1.688	0.617 ↓*	amino acid metabolism	6.148
42	450.321	Glycoursodeoxycholic acid	$C_{26}H_{43}NO_5$	5.921	1.271	0.008	1.236 ↑	Primary bile acid biosynthesis	2.387
43	500.304	Taurochenodeoxycholic acid	$C_{26}H_{45}NO_6S$	5.768	1.423	0.349	1.119 ↑	Primary bile acid biosynthesis	1.414
44	133.086	2-Hydroxycaproic acid	$C_6H_{12}O_3$	3.941	1.589	2.760	1.420	Fatty acid metabolism	5.869
45	205.097	Tryptophan[#]	$C_{11}H_{12}N_2O_2$	3.039	1.518	0.548	1.218 ↑	Aminoacyl-tRNA biosynthesis	5.839
46	466.316	Glycocholic acid	$C_{26}H_{43}NO_6$	5.936	1.586	0.013	3.001 ↑	Primary bile acid biosynthesis	1.821
47	220.118	Pantothenic acid[#]	$C_9H_{17}NO_5$	2.256	1.105	1.998	0.658 ↓	CoA synthesis	4.177
48	516.299	Taurocholic acid	$C_{26}H_{45}NO_7S$	5.881	1.638	0.322	1.610 ↑	Taurine and hypotaurine metabolism	1.478
49	190.086	Indole-3-methyl acetate	$C_{11}H_{11}NO_2$	5.32	1.575	0.098	4.223 ↑**	tryptophan catabolism	−1.500
50	175.096	Suberic acid	$C_8H_{14}O_4$	4.265	1.164	0.261	1.009 ↑	Fatty acid metabolism	4.418
51	382.272	Sphinganine 1-phosphate	$C_{18}H_{40}NO_5P$	7.636	1.657	2.561	0.532 ↓**	Sphingolipid Metabolism	2.981
52	167.070	D-Phenyllactic acid	$C_9H_{10}O_3$	4.287	1.239	2.360	0.765 ↓	amino acid metabolism	4.661
53	160.097	Isovalerylglycine	$C_7H_{13}NO_3$	3.284	1.273	0.466	0.794	carboxylic acids and derivatives	4.556
54	208.097	N-Acetyl-L-phenylalanine	$C_{11}H_{13}NO_3$	4.315	1.300	2.549	0.815 ↓	phenylalanine metabolism	5.436
55	305.247	Arachidonic acid	$C_{20}H_{32}O_2$	7.874	1.619	0.574	0.890	Arachidonic acid metabolism	3.375

*[#]metabolites were identified by reference standards; *p values < 0.05; **p values <0.01; RT, retention time; VIP, variable importance in the projection obtained from Model group vs. Sham group.*

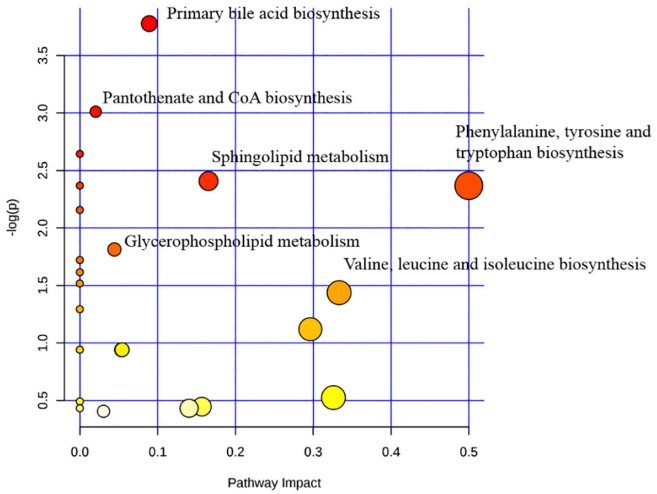


FIGURE 5 | The pathway analysis of the identified metabolites.

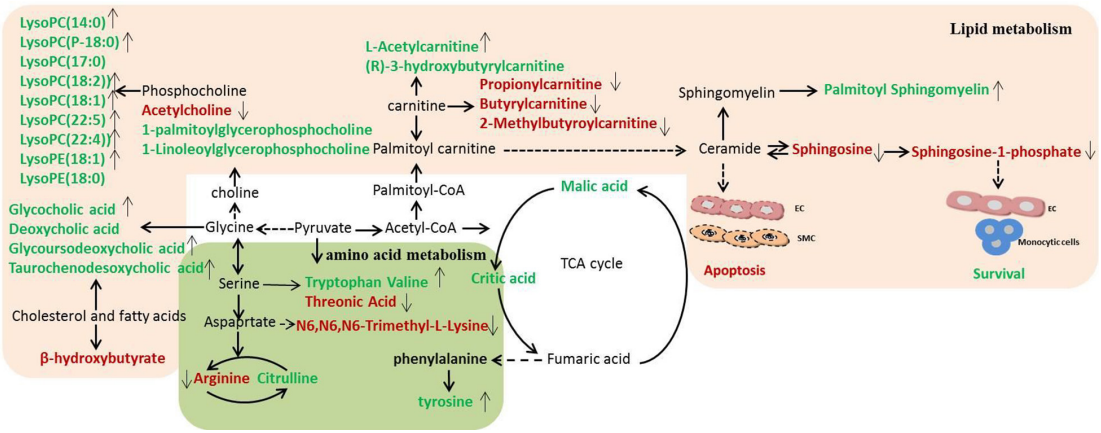


FIGURE 6 | Disordered metabolic pathway network in stroke and the interventional effects of DSCXQ. The names marked in red represent up-regulated metabolites in model rats, and the names marked in green represent down-regulated metabolites in model rats. The names marked in black represent undetected metabolites. The metabolites reversed by DSCXQ are marked with up and down arrows.

following cerebral ischemia (Moon et al., 2015; Nitzsche et al., 2021). S1P can activate platelets, which in turn stimulates the release of S1P into the bloodstream, S1P bind to receptors on the platelet surface, altering cell membrane glycoproteins, exposing fibrinogen receptors, increasing CD62P expression and activation rates and elevating platelet reactivity (Huang et al., 2006). CD62P, as an adhesion molecule on the surface of cell membrane, can mediate the mutual adhesion and aggregation of platelets, vascular endothelial cells, neutrophils and monocytes which play an important role in initiating thrombosis and exert a central role in inflammation and embolism. As platelets are activated, the CD62P is exposed to the surface of the cytosolic membrane and partially released into the blood. Hence, detecting the CD62P of platelet surface can directly reflect the activation degree of platelets in the body and understand the thrombosis process (Danton and Dietrich, 2003). Studies have shown that sphingosine is a lipid metabolite with multiple physiological and immunoregulatory functions, and may regulate apoptosis and necrosis (Radak et al., 2017). In our study, the significant increase of sphingomyelin and sphingosine caused abnormal sphingomyelin metabolism which up-regulated S1P expression, increased CD62P expression, intensified platelet activation, enhanced Bcl-2 level, inhibited cleaved Caspase-3 expression, induced cell necrosis and apoptosis - mechanisms which underlie the pathophysiology of stroke. CD62P immunohistochemistry revealed that number of $CD62P^{+}$ cells was significantly high in Model group compared to that of Sham group, while the rats treated with DSCXQ showed a dramatic reduction in the expression of $CD62P^{+}$ cells. To explore the possible molecular mechanisms related to beneficial effects of DSCXQ against apoptosis in I/R rats, the expression levels of Bax, Bcl-2 and cleaved Caspsae-3 in the rat brain were detected by western blot.

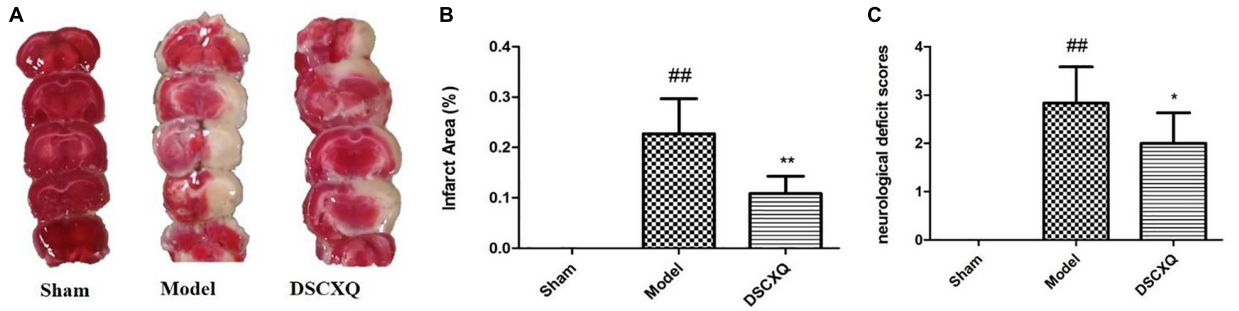


FIGURE 7 | (A) Representative images of TTC staining in different groups of DSCXQ at 24h after I/R. Red areas represent normal tissue, while the white is infarction. **(B)** Quantitative analysis of infarct size in different groups of DSCXQ. **(C)** Effects of DSCXQ on neurologic deficits in different group, the evaluation of neurological deficits was assessed by four-point scale (0–4). Data were expressed with mean values ± standard deviation (SD). ##$p < 0.01$ vs. Sham group, **$p < 0.01$ *$p < 0.05$ vs. Model group.

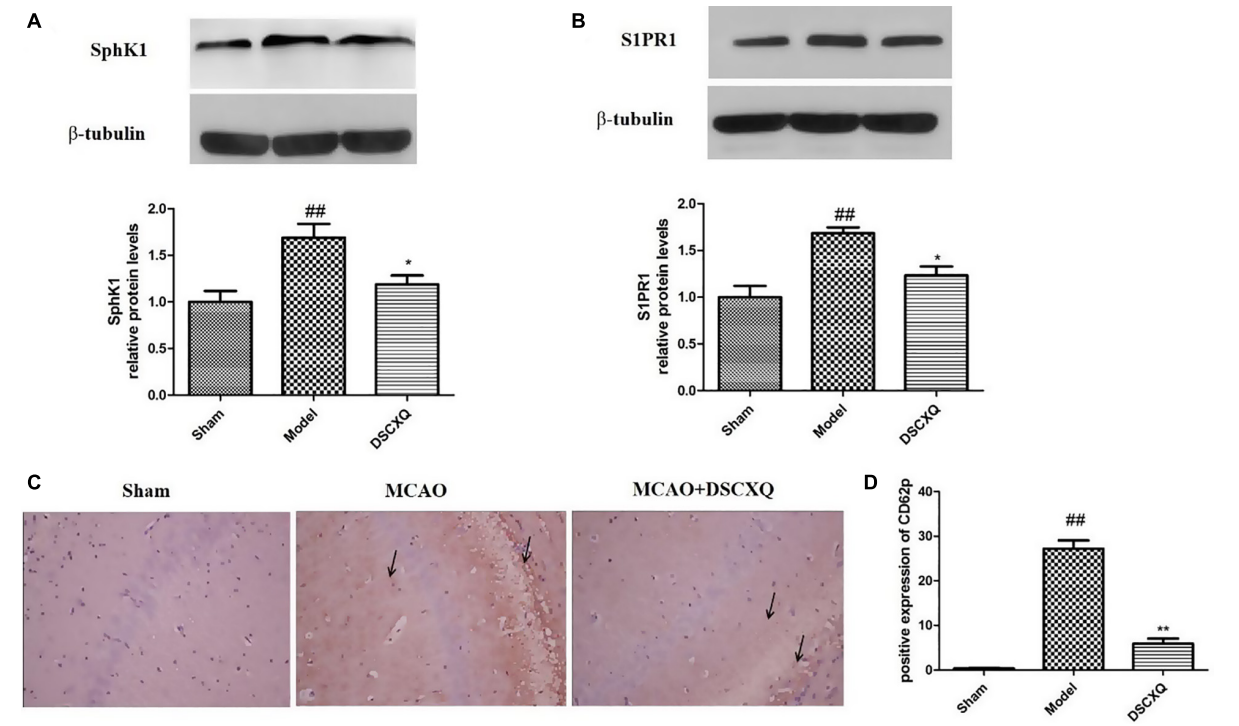


FIGURE 8 | (A,B) Western blot analysis of Sphk1 and S1PR1 in the cerebral cortex and hippocampus of rats at 24h after MCAO. **(C,D)** Representative photographs represented the expression of CD62P in the cerebral cortex and hippocampus at 24h reperfusion after MCAO. The positive products were brown granules after immunostaining, which was remarkably increased followed MCAO and was significantly decreased after treatment with DSCXQ. Representative images of Sphk1 and S1PR1 were presented, and the protein levels were expressed as a ratio of the β-tubulin levels ($n = 3$). Data were expressed with mean values ± standard deviation (SD). ##$p < 0.01$ vs. Sham group, **$p < 0.01$ *$p < 0.05$ vs. Model group.

The results indicate that the expressions of both Bax and cleaved Caspase-3 were markedly up-regulated and Bcl-2 markedly down-regulated compared with that of Sham group. However, DSCXQ administration, inhibited the expression of Bax and cleaved Caspase-3 expressions, while significantly enhanced the expression of Bcl-2 when compared with the non-treated Model group ($p < 0.05$). These results suggest that DSCXQ injection has significant anti-apoptotic effect against stroke.

Ischemic Stroke and Amino Acid Metabolism

Compared with the Sham group, the amino acid metabolism in the MCAO group was obviously disordered, in which tryptophan and valine were negatively correlated with the occurrence of cerebral ischemia. Valine is one of branched amino acids (branched-chain amino acid, BACC) *in vivo*, which are closely related to neurotransmitter synthesis and protein degradation

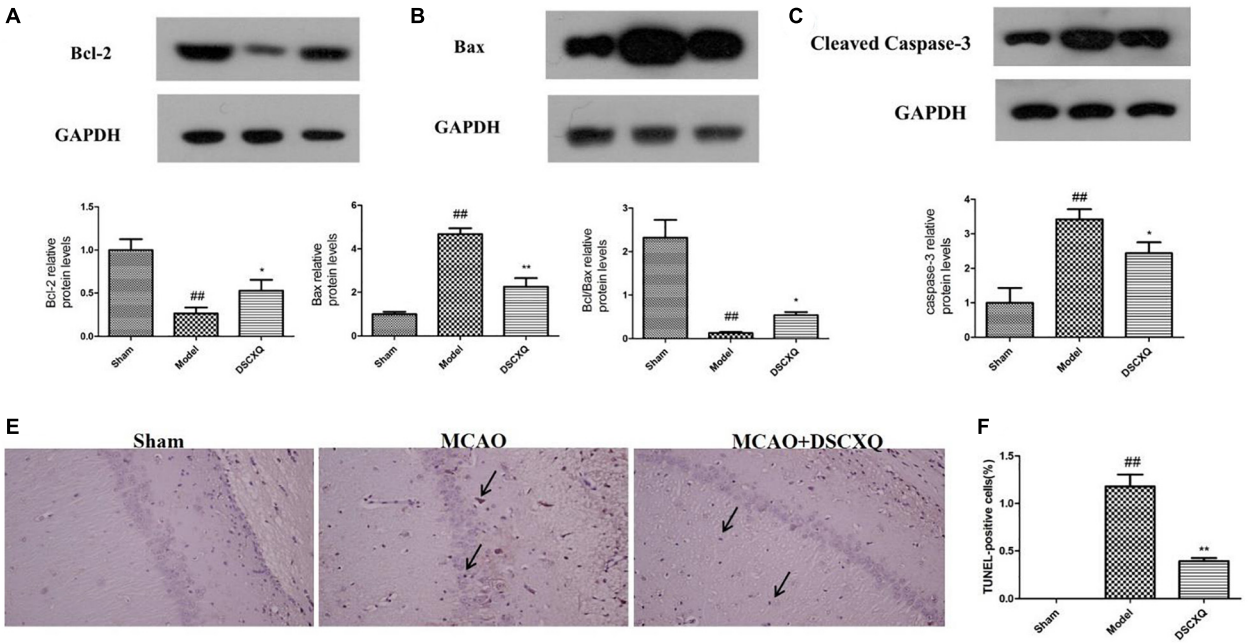


FIGURE 9 | (A-C) Western blot analysis of Bcl-2, Bax, Cleaved Caspase-3 in the cerebral cortex and hippocampus of rats at 24h after MCAO. **(D,E)** Representative TUNEL-staining brain sections in different groups (n = 3) in the cerebral cortex and hippocampus at 24 h reperfusion after MCAO. Representative images of Bcl-2, Bax, Cleaved Caspase-3 protein were presented, and the protein levels were expressed as a ratio of the GAPDH levels (n = 3). Data were expressed with mean values ± standard deviation (SD). ##p < 0.01 vs. Sham group, **p < 0.01 *p < 0.05 vs. Model group.

(Liu et al., 2016). Tryptophan, as one of the essential amino acids, plays an important role in maintaining the activity and increment of immune cells. On the other hand, studies have revealed that the active degradation of kyn pathway in Trp were closely linked to the severity and long-term prognosis of stroke (Cuartero et al., 2016), confirming our study.

Abnormal pathway of arginine metabolism may indicate the abnormal production of NO and the possible oxidative stress injury. As the central nervous system transmitter, the lack of NO may cause the brain information transmission disorder. At the early stage of cerebral ischemia, NO has the neuroprotective effect which can inhibit platelet and blood cell adhesion, researches have showed that NO in patients with acute cerebral infarction decreased significantly (Rashid et al., 2020), consistent with our findings. With the treatment of DSCXQ, the level of arginine was be on a downwards trend, indicating that the production of NO was likely to increase to exert neuroprotective effect.

Lipid Metabolism and Stroke

Lipid metabolism disorder is an important factor causing stroke. Lipid compounds participate in cellular functions and energy storage, as well as a variety of precursor compounds as second messengers. Abnormal changes of lipid metabolism network in brain tissue may be closely related to changes of oxidative stress parameters reflecting apoptosis (Yang et al., 2017; Chen et al., 2019).In our study, the level of LysoPCs and LysoPEs decreased versus Sham group, among them, the level of LysoPE (18:0/0:0), LPC (18:2) increased after DSCXQ treatment which restored toward Sham groups.

Compared with the Sham group, the level of L-acetylcarnitine decreased in the cerebral ischemia model which played an important role in β-oxidation. Since brain is almost completely dependent on the oxidative energy supply of sugar. L-acetylcarnitine that is synthesized by acetyl CoA and carnitine mainly affects the energy metabolism of the brain (Zanelli et al., 2005). In the study, the decreased of 3-hydroxybutyrylcarnitine, the rise of propionylcarnitine, Butyrylcarnitine and 2-Methylbutyroylcarnitine were observed in Model group vs. Sham group, with the treatment of DSCXQ, the level of L-Acetylcarnitine, 2-Methylbutyroylcarnitine, 3-hydroxybutyrylcarnitine showed a trend toward Sham group, indicating that the neuroprotective effects of DSCXQ partially interferences with lipid metabolism in disease.

Ischemic Stroke and Energy Metabolism

Energy metabolism disorder is the leading cause in the course of cerebral ischemia (Kurup et al., 1990). Cerebral ischemia restricts the delivery of oxygen and glucose in brain, and it has been revealed that energy metabolism shifted from tricarboxylic acid cycle to glycolysis. β-hydroxybutyrate, a small lipid-derived molecules which is classified as a type of ketone bodies, can be oxidized and provided energy when brain is in starvation, during glucose is insufficient, fatty acids were transported to liver where converted to ketone bodies, and transferred from blood serum to brain tissue (Newman and Verdin, 2017). In our experiments,

β-hydroxybutyrate significantly decreased in MCAO rats and was raised by DSCXQ which may be alleviate energy supply obstacles and improve fatty acids utilization.

Bile Acids –Gut Microbiome - Brain Axis in Stroke

Bile acids are the metabolites of cholesterol formed by a series of enzymatic reactions in hepatocytes. Meanwhile gut microbiome play a vital role in lipid regulation via promoting the formation of cholesterol oxidase. In our study, the primary bile acids in brain tissue of rats in Model group were significantly lower than those in Sham group. After the treatment of DSCXQ, the level of bile acids was notably higher than that in Model group. The lipids in the liver were cleared by bile acids, but it is interesting that bile acids in hepato-enteric circulation were detected in brain tissue, indicating that bile acid metabolism axis that runs through the liver - gut microbiota - brain is bound up with pathogenesis of stroke.

CONCLUSION

In general, the research provides comprehensive insight into the mechanisms of DSCXQ against stroke by UHPLC-Q-Orbitrap HRMS based on metabolomics, and the results revealed that a series of complex and severe metabolic disorders were caused in MCAO rats after ischemic stroke injury. A total of 55 metabolic differences were detected as potential markers, and DSCXQ ameliorate the perturbed metabolic process mainly involving sphingolipid metabolism, amino acid metabolism, and cerebral homeostasis. Sphingolipid metabolism is closely associated with antiplatelet aggregation and cell survival. The results further verified that DSCXQ improved the cell survival and inhibited apoptosis in ischemia stroke rats. UHPLC-Q-Orbitrap HRMS-based metabolomics approach was applied in this study firstly to explore the neuroprotective effect of DSCXQ on ischemic stroke. By identifying the metabolic pathway and network, we can provide rich information to better understand and unravel the causal mechanisms of ischemic stroke, which has great potential in the future. According to the results of histopathology and metabolomics, the present study renovated our understanding of cerebral I/R injury pathogenesis and the neuroprotective mechanism of DSCXQ, which provided essential basis for clinical application.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by the standard guidelines for the Care and Use of laboratory animals from the National Institute of Health (NIH) and the Animals Ethics Committee of Zheng Zhou University.

AUTHOR CONTRIBUTIONS

PZ contributed to the experimental design, implementation, and article writing. LZ participated in pharmacological experiments, contributed to data statistics, and contributed as a graphic analyst. YS, ZL, LL, and LHZ participated in experiments and differential metabolite identification. JZ participated in pharmacological experiments. JK, SL, JY, and SD partially participated in the experimental design. ZS designed the experiments and checked the project. XZ is the general director, project designer, and the teacher responsible for the funding, writing, and checking the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (Grant Nos. 81703666 and 81873188) and the Foundation of The First Affiliated Hospital of Zhengzhou University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.630291/full#supplementary-material

REFERENCES

Amelio, I., Cutruzzolá, F., Antonov, A., Agostini, M., and Melino, G. (2014). Serine and glycine metabolism in cancer. *Trends Biochem. Sci.* 39, 191–198.

Blondeau, N., Lai, Y., Tyndall, S., Popolo, M., Topalkara, K., Pru, J. K., et al. (2007). Distribution of sphingosine kinase activity and mRNA in rodent brain. *J. Neurochem.* 103, 509–517. doi: 10.1111/j.1471-4159.2007.04755.x

Chen, L., Chao, Y., Cheng, P., Li, N., Zheng, H., and Yang, Y. (2019). UPLC-QTOF/MS-based metabolomics reveals the protective mechanism of hydrogen on mice with ischemic stroke. *Neurochem. Res.* 44, 1950–1963. doi: 10.1007/s11064-019-02829-x

Cuartero, M. I., de la Parra, J., García-Culebras, A., Ballesteros, I., Lizasoain, I., and Moro, M. (2016). The kynurenine pathway in the acute and chronic phases of cerebral ischemia. *Curr. Pharm. Design* 22, 1060–1073. doi: 10.2174/1381612822666151214125950

Cunningham, T. J., Yao, L., and Lucena, A. (2008). Product inhibition of secreted phospholipase A2 may explain lysophosphatidylcholines' unexpected therapeutic properties. *J. Inflamm.* 5:17. doi: 10.1186/1476-9255-5-17

Danton, G. H., and Dietrich, W. D. (2003). Inflammatory mechanisms after ischemia and stroke. *J. Neuropathol. Exp. Neurol.* 62, 127–136. doi: 10.1093/jnen/62.2.127

Fei, Y. X., Wang, S. Q., Yang, L. J., Qiu, Y. Y., Li, Y. Z., Liu, W. Y., et al. (2017). Salvia miltiorrhiza Bunge (Danshen) extract attenuates permanent cerebral ischemia through inhibiting platelet activation in rats. *J. Ethnopharmacol.* 207, 57–66. doi: 10.1016/j.jep.2017.06.023

Feigin, V. L., Roth, G. A., Naghavi, M., Parmar, P., Krishnamurthi, R., Chugh, S., et al. (2016). Global burden of stroke and risk factors in 188 countries, during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet Neurol.* 15, 913–924. doi: 10.1016/s1474-4422(16)30073-4

Fonarow, G. C., Smith, E. E., Saver, J. L., Reeves, M. J., Bhatt, D. L., Grau-Sepulveda, M. V., et al. (2011). Timeliness of tissue-type plasminogen activator therapy in acute ischemic stroke: patient characteristics, hospital factors, and outcomes associated with door-to-needle times within 60 minutes. *Circulation* 123, 750–758. doi: 10.1161/circulationaha.110.974675

Gao, B., Kilic, E., Baumann, C. R., Hermann, D. M., and Bassetti, C. L. (2008). Gamma-hydroxybutyrate accelerates functional recovery after focal cerebral ischemia. *Cerebrovasc. Dis.* 26, 413–419. doi: 10.1159/000151683

Hu, J., Huang, S., Zhu, L., Huang, W., and Zhao, Y. (2018). Tissue plasminogen activator-porous magnetic microrods for targeted thrombolytic therapy after ischemic stroke. *ACS Appl. Mater. Interfaces* 10, 32988–32997. doi: 10.1021/acsami.8b09423

Huang, J., Upadhyay, U. M., and Tamargo, R. J. (2006). Inflammation in stroke and focal cerebral ischemia. *Surg. Neurol.* 66, 232–245. doi: 10.1016/j.surneu.2005.12.028

Jin, K., ZhuGe, Q., Prabhakaran, S., Ruff, I., and Bernstein, R. A. (2015). Acute stroke intervention: a systematic review. *ACS Appl. Mater. Interfaces* 313, 1451–1462. doi: 10.1001/jama.2015.3058

Jin, Y., Knudsen, E., Wang, L., Bryceson, Y., Damaj, B., Gessani, S., et al. (2003). Sphingosine 1-phosphate is a novel inhibitor of T-cell proliferation. *Blood* 101, 4909–4915. doi: 10.1182/blood-2002-09-2962

Kagiyama, T., Glushakov, A. V., Sumners, C., Roose, B., Dennis, D. M., Phillips, M. I., et al. (2004). Neuroprotective action of halogenated derivatives of L-phenylalanine. *Stroke* 35, 1192–1196. doi: 10.1161/01.str.0000125722.10606.07

Kasahara, K., and Rey, F. E. (2019). The emerging role of gut microbial metabolism on cardiovascular disease. *Curr. Opin. Microbiol.* 50, 64–70. doi: 10.1016/j.mib.2019.09.007

Kimberly, W. T., Wang, Y., Pham, L., Furie, K. L., and Gerszten, R. E. (2013). Metabolite profiling identifies a branched chain amino acid signature in acute cardioembolic stroke. *Stroke* 44, 1389–1395. doi: 10.1161/strokeaha.111.000397

Kurup, C. K., Kumaroo, K. K., and Dutka, A. J. (1990). Influence of cerebral ischemia and post-ischemic reperfusion on mitochondrial oxidative phosphorylation. *J. Bioenerget. Biomembr.* 22, 61–80. doi: 10.1007/bf00762846

Liu, M., Tang, L., Liu, X., Fang, J., Zhan, H., Wu, H., et al. (2016). An evidence-based review of related metabolites and metabolic network research on cerebral ischemia. *Oxid. Med. Cell. Longev.* 2016:9162074.

Longa, E. Z., Weinstein, P. R., Carlson, S., and Cummins, R. (1989). Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke* 20, 84–91. doi: 10.1161/01.str.20.1.84

Lyu, S., Yang, S. L., Rao, Y., and Feng, Y. L. (2018). [Application of metabolomics and related technologies in research and development field of traditional Chinese medicine]. *Zhongguo Zhong Yao Za Zhi* 43, 4182–4191.

Maceyka, M., Harikumar, K. B., Milstien, S., and Spiegel, S. (2012). Sphingosine-1-phosphate signaling and its role in disease. *Trends Cell Biol.* 22, 50–60. doi: 10.1016/j.tcb.2011.09.003

Maceyka, M., Payne, S. G., Milstien, S., and Spiegel, S. (2002). Sphingosine kinase, sphingosine-1-phosphate, and apoptosis. *Biochim. Biophys. Acta* 1585, 193–201.

Moon, E., Han, J. E., Jeon, S., Ryu, J. H., Choi, J. W., and Chun, J. (2015). Exogenous S1P exposure potentiates ischemic stroke damage that is reduced possibly by inhibiting S1P receptor signaling. *Mediators Inflamm.* 2015:492659.

Newman, J. C., and Verdin, E. (2017). β-Hydroxybutyrate: a signaling metabolite. *Annu. Rev. Nutr.* 37, 51–76. doi: 10.1146/annurev-nutr-071816-064916

Nitzsche, A., Poittevin, M., Benarab, A., Bonnin, P., Faraco, G., Uchida, H., et al. (2021). Endothelial S1P(1) signaling counteracts infarct expansion in ischemic stroke. *Circ. Res.* 128, 363–382. doi: 10.1161/circresaha.120.316711

Radak, D., Katsiki, N., Resanovic, I., Jovanovic, A., Sudar-Milovanovic, E., Zafirovic, S., et al. (2017). Apoptosis and acute brain ischemia in ischemic stroke. *Curr. Vasc. Pharmacol.* 15, 115–122. doi: 10.2174/1570161115666161104095522

Rashid, J., Kumar, S. S., Job, K. M., Liu, X., Fike, C. D., and Sherwin, C. M. T. (2020). Therapeutic potential of citrulline as an arginine supplement: a clinical pharmacology review. *Paediatr. Drugs* 22, 279–293. doi: 10.1007/s40272-020-00384-5

Spiegel, S., and Milstien, S. (2003). Sphingosine-1-phosphate: an enigmatic signalling lipid. *Nat. Rev. Mol. Cell Biol.* 4, 397–407. doi: 10.1038/nrm1103

Tabatabaie, L., Klomp, L. W., Berger, R., and de Koning, T. J. (2010). L-serine synthesis in the central nervous system: a review on serine deficiency disorders. *Mol. Genet. Metab.* 99, 256–262. doi: 10.1016/j.ymgme.2009.10.012

Virmani, A., and Binienda, Z. (2004). Role of carnitine esters in brain neuropathology. *Mol. Aspects Med.* 25, 533–549. doi: 10.1016/j.mam.2004.06.003

Wang, Q., Liu, D., Song, P., and Zou, M. H. (2015). Tryptophan-kynurenine pathway is dysregulated in inflammation, and immune activation. *Front. Biosci.* 20:1116–1143. doi: 10.2741/4363

Yang, L., Lv, P., Ai, W., Li, L., Shen, S., Nie, H., et al. (2017). Lipidomic analysis of plasma in patients with lacunar infarction using normal-phase/reversed-phase two-dimensional liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 409, 3211–3222. doi: 10.1007/s00216-017-0261-6

Zanelli, S. A., Solenski, N. J., Rosenthal, R. E., and Fiskum, G. (2005). Mechanisms of ischemic neuroprotection by acetyl-L-carnitine. *Ann. N. Y. Acad. Sci.* 1053, 153–161. doi: 10.1196/annals.1344.013

Zhang, Z., Ji, J., Zhang, D., Ma, M., and Sun, L. (2020). Protective effects and potential mechanism of salvianolic acid B on sodium laurate-induced thromboangiitis obliterans in rats. *Phytomedicine* 66:153110. doi: 10.1016/j.phymed.2019.153110

Zhou, P., Du, S., Zhou, L., Sun, Z., Zhuo, L. H., He, G., et al. (2019a). Tetramethylpyrazine-2′O-sodium ferulate provides neuroprotection against neuroinflammation and brain injury in MCAO/R rats by suppressing TLR-4/NF-κB signaling pathway. *Pharmacol. Biochem. Behav.* 176, 33–42. doi: 10.1016/j.pbb.2018.08.010

Zhou, P., Zhou, L., Sun, Z., Li, Z., Zhang, R., Guan, K., et al. (2019b). Qualitative and quantitative analysis of Danshen-chuanxiong injection by using UHPLC-Q-Orbitrap HRMS. *Chin. Pharm. J.* 54, 327–333.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 21 December 2022
DOI 10.3389/fmolb.2022.1074285

OPEN ACCESS

EDITED BY
Xiaoyang Su,
The State University of New Jersey, United States

REVIEWED BY
Christopher Papandreou,
Institut d'Investigació Sanitària Pere Virgili (IISPV), Spain
Tianxia Xiao,
Princeton University, United States

*CORRESPONDENCE
Masae Kuboniwa,
kuboniwa.masae.dent@osaka-u.ac.jp

SPECIALTY SECTION
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 19 October 2022
ACCEPTED 28 November 2022
PUBLISHED 21 December 2022

CITATION
Sakanaka A, Katakami N, Furuno M, Nishizawa H, Omori K, Taya N, Ishikawa A, Mayumi S, Inoue M, Tanaka Isomura E, Amano A, Shimomura I, Fukusaki E and Kuboniwa M (2022), Salivary metabolic signatures of carotid atherosclerosis in patients with type 2 diabetes hospitalized for treatment. *Front. Mol. Biosci.* 9:1074285. doi: 10.3389/fmolb.2022.1074285

Salivary metabolic signatures of carotid atherosclerosis in patients with type 2 diabetes hospitalized for treatment

Akito Sakanaka[1], Naoto Katakami[2], Masahiro Furuno[3], Hitoshi Nishizawa[2], Kazuo Omori[2], Naohiro Taya[2], Asuka Ishikawa[1], Shota Mayumi[1], Moe Inoue[1], Emiko Tanaka Isomura[4], Atsuo Amano[1], Iichiro Shimomura[2], Eiichiro Fukusaki[3] and Masae Kuboniwa[1]*

[1]Department of Preventive Dentistry, Osaka University Graduate School of Dentistry, Suita, Japan, [2]Department of Metabolic Medicine, Osaka University Graduate School of Medicine, Suita, Japan, [3]Department of Biotechnology, Osaka University Graduate School of Engineering, Suita, Japan, [4]First Department of Oral and Maxillofacial Surgery, Osaka University Graduate School of Dentistry, Suita, Japan

Atherosclerosis is a life-threatening disease associated with morbidity and mortality in patients with type 2 diabetes (T2D). This study aimed to characterize a salivary signature of atherosclerosis based on evaluation of carotid intima-media thickness (IMT) to develop a non-invasive predictive tool for diagnosis and disease follow-up. Metabolites in saliva and plasma samples collected at admission and after treatment from 25 T2D patients hospitalized for 2 weeks to undergo medical treatment for diabetes were comprehensively profiled using metabolomic profiling with gas chromatography-mass spectrometry. Orthogonal partial least squares analysis, used to explore the relationships of IMT with clinical markers and plasma and salivary metabolites, showed that the top predictors for IMT included salivary allantoin and 1,5-anhydroglucitol (1,5-AG) at both the baseline examination at admission and after treatment. Furthermore, though treatment induced alterations in salivary levels of allantoin and 1,5-AG, it did not modify the association between IMT and these metabolites ($p_{interaction} > 0.05$), and models with these metabolites combined yielded satisfactory diagnostic accuracy for the high IMT group even after treatment (area under curve = 0.819). Collectively, this salivary metabolite combination may be useful for non-

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUC, area under the curve; BMI, body mass index; BP, blood pressure; Cre, serum creatinine; CV-ANOVA, cross-validation analysis of variance; CVD, cardiovascular disease; FPG, fasting plasma glucose; GA, glycated albumin; GABA, γ-Aminobutyric acid; GC/MS, gas chromatography coupled with mass spectrometry; GlcNAc, *N*-acetylglucosamine; Gln, glutamine; eGFR, estimated glomerular filtration rate; γ-GTP, gamma-glutamyl transferase; HDL-C, high-density lipoprotein cholesterol; hsCRP, high-sensitivity C-reactive protein; IAA, indole acetic acid; LDL-C, low-density lipoprotein cholesterol; LOWESS, locally weighted scatter plot smoothing; MBP, mean blood pressure; OPLS, orthogonal partial least squares; PISA, periodontal inflamed surface area; QC, quality control; ROC, receiver operating characteristic; TG, triglycerides; U-Alb, urine albumin; U-Cre, urine creatinine; 1,5-AG, 1,5-anhydroglucitol.

invasive identification of T2D patients with a higher atherosclerotic burden in clinical settings.

KEYWORDS

saliva, diabetes, atherosclerosis, carotid intima-media thickness (IMT), oxidative stress, metabolomics

Introduction

Atherosclerosis refers to the formation of fibrofatty lesions in artery walls, which can impede blood flow, leading to tissue ischemia. Atherosclerotic cardiovascular disease (CVD) is the leading cause of morbidity and mortality in patients with type 2 diabetes (T2D) (Mach et al., 2019). Furthermore, diabetes itself is an independent risk factor for CVD, and recent findings indicate that its presence increases that risk by approximately two-fold on average, though there are wide variations depending on the subject population and prophylactic treatment received (Sattar, 2013; Olesen et al., 2017). Therefore, long-term management of T2D patients requires early and accurate identification, as well as monitoring of those at high risk for CVD. Carotid ultrasonography is useful for evaluating subclinical atherosclerosis and various measurement results obtained with that method, including carotid intima-media thickness (IMT), have been shown to be predictive of CVD (Lorenz et al., 2007). However, it is not feasible to perform carotid ultrasonography for universal screening in patients with T2D, as specialized machines, skilled technicians, and patient restraint are required. Novel biomarkers that can better predict atherosclerosis burden and CVD risk are anticipated to facilitate early treatment induction, and reduce CVD-related morbidity and mortality associated with T2D.

Metabolites are small molecules known to reflect biological processes and their measurements are often utilized in clinical medicine as biomarkers for diagnosis, prognosis, and treatment efficacy (Wishart, 2019). Recent advancements in high-throughput technologies have allowed for systematic evaluation of a metabolome, a collection of metabolites, with regard to cardiometabolic changes, and several studies using blood metabolome have characterized disease-related metabolic pathways, including amino acid and fatty acid metabolism (Fan and Pedersen, 2021). In addition, state-of-the-art metabolomics data related to the risk of type 2 diabetes and its complications, analyzed from a wide variety of biological samples (plasma, serum, and urine), have been published (Sharma et al., 2013; Morze et al., 2022). Previously, we investigated the association of paired plasma and salivary metabolomic datasets from patients with T2D, and demonstrated the potential utility of salivary metabolites for evaluating systemic metabolic dysfunction (Sakanaka et al., 2021). However, there remains a lack of metabolomic studies that sought to address the complexity of T2D-CVD crosstalk, such as how atherosclerotic burden can alter metabolic profiles in plasma and saliva in T2D patients.

The aim of the present study was to identify multivariate covariation patterns between carotid atherosclerosis, and saliva and plasma metabolomes to develop non-invasive tools for prediction of atherosclerotic burden in T2D patients. To achieve this, comprehensive metabolomic profiling of plasma and saliva obtained from T2D patients was performed, and multivariate covariations with clinical markers of oral and systemic health were investigated. For the analyses, we employed a powerful multivariate method termed orthogonal partial least square (OPLS) to go beyond simple correlations. This method has been reported suitable for analysis of high-dimensional datasets where numerous variables are expected to be highly correlated (Wheelock and Wheelock, 2013). Using results obtained by modeling the association between IMT and clinical and metabolomic parameters with OPLS, this study presents a catalog of plasma and salivary metabolites that potentially reflect atherosclerotic burden. It is considered that the findings presented suggest the potential of salivary metabolites for evaluating cardiovascular risk in T2D patients.

Materials and methods

Study population

The present study was approved by the Osaka University Research Ethics Committee, and performed according to the principles of the Helsinki Declaration and STROBE guidelines for human observational studies. All participants gave written informed consent prior to enrollment and provided samples at Osaka University Medical Hospital. Participants with T2D, diagnosed using the criteria of the Japan Diabetes Society (Haneda et al., 2018), were recruited from November 2017 through March 2019 from among patients who visited the Department of Metabolic Medicine at Osaka University Medical Hospital for intensive diabetes treatment in an inpatient setting. Those with severe renal dysfunction or end-stage renal failure (serum creatinine >2.0 mg/dL), or under 50 years of age were excluded. All enrolled patients received comprehensive diabetes care, including intensive glycemic control, as well as blood pressure, dyslipidemia, and body weight control treatments while under hospitalization. Saliva, fasting blood, and urine samples were also obtained, and vital signs and weight were measured at admission (baseline) and again 2 weeks after treatment. Of 33 patients examined from November 2017 to March 2019, 25 met the criteria and were included in the study.

Blood and urine sample collection, and laboratory measurements

Blood and urine samples were collected after an overnight fast at admission and again 2 weeks after treatment during hospitalization. Those were subjected to biochemical testing such as HbA1c and urine albumin, which were performed according to standard protocols. Furthermore, fasting plasma was collected and kept at 4°C in a freezer (CubeCooler®; Forte Grow Medical, Tochigi, Japan), and then frozen at −80°C, and used for metabolomics. All patients underwent anthropometric measurements and were asked to complete a variety of surveys regarding demographics, current and past medical history, medications, smoking history, and family history. Determinations of hypertension (defined as systolic blood pressure ≥130 mmHg, diastolic blood pressure ≥80 mmHg, or anti-hypertensive medication use), dyslipidemia [defined as serum low-density lipoprotein cholesterol (LDL-C) ≥120 mg/dL, serum triglycerides (TG) ≥150 mg/dL, high-density lipoprotein cholesterol (HDL-C) <40 mg/dL, or lipid-lowering medication use], and obesity (BMI ≥ 25 kg/m^2) were based on the criteria of the Japan Diabetes Society.

Carotid IMT measurement

Details of the carotid ultrasonic examination methods used have been presented (Omori et al., 2020). Briefly, a B-mode ultrasonography examination of the carotid artery was performed with a 7.5-MHz liner transducer. All scanning was conducted by experienced laboratory physicians using the same measuring method, in accordance with the guidelines of the Japan Society of Ultrasonics in Medicine (Terminology and Diagnostic Criteria Committee and Japan Society of Ultrasonics in Medicine, 2009). The thickest point for IMT in both common carotid arteries was separately determined, with the highest value defined as IMT for each individual in this study. This measurement was only performed at admission.

Oral examination and saliva sample collection

All participants were asked to refrain from eating, drinking, or brushing their teeth for at least 1 h before undergoing the following procedures. Four calibrated licensed dentists performed oral examinations using techniques previously described (Sakanaka et al., 2021), and obtained saliva samples on the same day as blood and urine samples. For saliva output, each participant was asked to collect unstimulated whole saliva over a 10-min period in a 50-ml tube (Corning, NY, United States) kept on ice. Four diabetes patients with saliva output of less than 3 ml/10 min were asked to take 3 ml of distilled water (HPLC grade; Sigma-Aldrich, St. Louis, MO, United States) into their mouth and spit it out into a tube. By making the corrections described below regarding metabolomics analysis, we confirmed that exclusion of these four samples did not change the main results. After incubation on ice for 15 min, 1 and 0.1 ml of the aqueous layer were designated as the study sample and the quality control (QC) sample, respectively, which were then separately aliquoted into 2-ml tubes and maintained at 4°C in a CubeCooler®. They were subsequently frozen with liquid nitrogen and stored at −80°C until analysis.

Saliva and plasma metabolomics

Metabolomics analysis was performed as previously described (Sakanaka et al., 2021). Briefly, saliva samples were thawed to 4°C, then vortexed and centrifuged (18,000 × *g*) for 3 min. Next, 0.8 ml of the aqueous layer was aliquoted and weighed, then 0.3 ml of that was transferred into a 2-ml glass vial (Nichiden-Rika Glass, Kobe, Japan) and kept at 4°C in a CubeCooler®. For extraction, 0.3 ml of a deaerated ribitol aqueous solution (0.02 mg/ml) was added as the internal standard. After incubation using an Eppendorf thermomixer (25°C, 1,000 rpm, 10 min), 1.4 ml of deaerated acetonitrile was added. After another incubation (25°C, 1,000 rpm, 10 min) and then centrifugation (4°C, 1800 × *g*) for 3 min, 1.6 ml of the supernatant was transferred to a 2-ml tube and dried in a vacuum concentrator (VC-96R; TAITEC, Koshigaya, Japan) for 30 min, then lyophilized overnight. For derivatization, a methoxyamine hydrochloride solution with pyridine was used at a concentration of 20 mg/ml, followed by silylation treatment with N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA). Analysis using gas chromatography coupled with mass spectrometry (GC/MS) was performed using a GCMS-TQ8040 (Shimadzu, Kyoto, Japan) equipped with an AOC-20i autosampler (Shimadzu), a SKY™ liner (Restek, Bellefonte, PA, United States), and an InertCap 5MS/NP capillary column (0.25 mm × 30 m, 0.25 μm; GL Sciences, Tokyo, Japan), operated in full MS scan mode. QC samples consisting of an equimolar mixture of all saliva samples were injected every five samples to monitor MS signal drift and derivatization efficiency, followed by normalization with locally weighted scatter plot smoothing (LOWESS) in the subsequent data processing steps. GC-MS data were converted into ABF format, then processed using MS-DIAL (version 3.90) to perform feature detection, spectra deconvolution, metabolite identification, and peak alignment (Tsugawa et al., 2015). The acquired peak list was further normalized based on internal standard (ribitol) and sample weight (g/ml), as well as LOWESS algorithm. Metabolites from blanks and those with a coefficient of variation in QC samples above 30% were discarded. A total of 976 salivary metabolites were detected using this metabolomics platform, among which 142 were identified by matching retention time and fragmentation spectra to authentic standards. Plasma samples were prepared and analyzed with a

GC-MS/MS-TQ8040 in multiple reaction monitoring mode, as previously described (Katakami et al., 2020; Taya et al., 2021). The obtained metabolomics data are available at Metabolomics Workbench (Study ID: ST001905 and ST001906).

Statistical analysis

Using the SIMCA-P software package, v. 16 (Umetrics, Umeå, Sweden), an OPLS model was constructed with IMT as the Y response variable and all other parameters as X variables, which were all block-scaled by unit variance prior to analysis so that the influence of blocks of variables could be balanced in relation to their size. A seven-fold cross-validation was performed to avoid model over-fitting. Model quality and performance were assessed using R^2 (goodness of fit) and Q^2 (goodness of prediction) values, cross-validation analysis of variance (CV-ANOVA), and a permutation test (assessment of risk of over-fitting). The OPLS method can distinguish data variations correlated to the Y response variable from those orthogonal to Y response. Results thus obtained can assist with biological interpretation and enables establishment of a link between variations of variables and outcomes while removing information from other sources of variation. Variables reflecting the IMT variation were selected based on variable importance in projection of the predictive component (VIP predictive) of the OPLS model as well as Spearman's correlation. The diagnostic ability of some parameters for IMT >1.6 mm was evaluated according to receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) results. Cardiometabolic disease risk score was calculated as previously described (Wang et al., 2021a). Briefly, blood biomarker levels were first categorized into quintiles without distinction between pre- and post-treatment by ranking HbA1c, total cholesterol, TG, and high-sensitivity C-reactive protein (hsCRP) from lowest to highest with scores from 1 to 5. For HDL-C, the scoring was reversed. The cardiometabolic disease risk score was then calculated by summing those components, with a higher score indicating a higher risk of cardiometabolic disease. Differences as compared to the baseline regarding cardiometabolic disease risk score and salivary metabolite levels were analyzed with a paired *t* test. An interaction test was performed to determine whether treatment had an effect modification on the association between IMT and salivary metabolites. Spearman's correlation was performed using the GraphPad Prism software package, v.8, and ROC curves and an interaction test were performed with the R package (v. 4.0.3).

Results

After applying the aforementioned exclusion criteria, 25 patients with T2D were available for statistical analysis

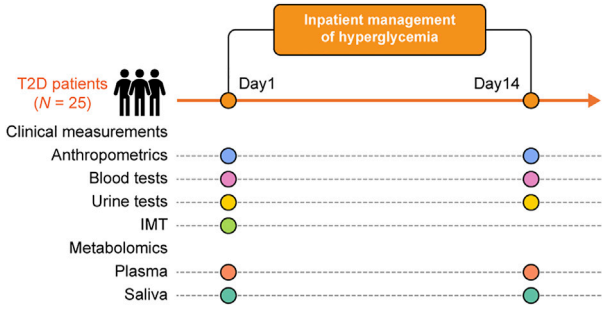


FIGURE 1
Overview of the experimental procedures. This study yielded clinical measures (IMT, anthropometrics, and blood and urine biochemical tests), and metabolome data from plasma and saliva at admission and after treatment for each of the 25 participants.

(Figure 1) and their clinical characteristics are shown in Table 1. Using untargeted GC/MS, 142 salivary and 78 plasma metabolites were identified. Of those, when plasma indoleacetaldehyde and salivary indoleacetic acid were treated as the same due to structural similarity, 62 metabolites were found to be shared between them, seven of which showed a significant positive association at the baseline (Spearman's correlation value > 0.3, $p < 0.05$). Among them, 1,5-anhydroglucitol (1,5-AG) demonstrated the strongest positive association between plasma and saliva ($r = 0.76$, $p = 9.0 \times 10^{-5}$). Overall, rich datasets comprised of three data blocks (44 clinical markers, 142 salivary, and 78 plasma metabolites) obtained at both the baseline and after treatment were generated.

For more accurate characterization of the associations between IMT and clinical and metabolomic parameters at the baseline, and to examine the relative importance of variables in relation to IMT, OPLS was performed with IMT as a Y response variable (Figure 2A). The model showed a moderate predictive ability of 0.362 for Q^2 and reliable performance during the permutation test ($n = 999$ permutations; Figure 2B, inset). Top predictors for IMT from each data block [VIP predictive value > 1.5, p(corr) value <−0.3 or >0.3] included eight clinical markers, and 17 plasma and six salivary metabolites (Figure 3A). Notably, variables that best characterized IMT included the clinical markers HDL-C ($r = -0.578$, $p = 0.002$), glycated albumin (GA) ($r = 0.472$, $p = 0.017$), and TG ($r = 0.581$, $p = 0.002$), the plasma metabolites *N*-acetylglucosamine (GlcNAc) ($r = 0.516$, $p = 0.008$), malate ($r = 0.461$, $p = 0.020$) and inositol ($r = 0.346$, $p = 0.091$), and the salivary metabolites allantoin ($r = 0.496$, $p = 0.012$), 1,5-AG ($r = -0.493$, $p = 0.012$), and malate ($r = -0.437$, $p = 0.029$). Notably, salivary levels of allantoin and 1,5-AG showed AUC values of 0.729 and 0.861, respectively, for diagnosis in the high IMT group (>1.6 mm) (Figure 3B).

Two weeks of treatment during hospitalization resulted in significant improvements in multiple clinical parameters,

TABLE 1 Main clinical characteristics of study participants.

Variables	Diabetes (n = 25)		
	Baseline	Week 2	p-value
Age, years	67.84 (3.75)	—	—
Females, n (%)	14 (56)	—	—
Diabetes duration, years	17.52 (11.76)	—	—
HbA1c, %	8.78 (1.84)	8.16 (1.44)	<0.001
Fasting plasma glucose, mg/dL	135.92 (41.07)	110.32 (19.47)	0.0075
Glycated albumin, %	23.36 (5.82)	19.67 (3.53)	<0.001
Obesity, n (%)	16 (64)	—	—
BMI, kg/m^2	25.99 (3.12)	25.20 (3.01)	<0.001
Waist circumference, cm	97.30 (9.17)	—	—
Hypertension, n (%)	17 (68)	—	—
Systolic BP, mmHg	131.04 (18.10)	124.36 (12.31)	0.10
Diastolic BP, mmHg	74.48 (14.33)	74.92 (11.74)	0.87
Mean BP, mmHg	93.35 (13.90)	91.4 (10.90)	0.46
Dyslipidemia, n (%)	21 (84)	—	—
AST, U/L	27.8 (17.45)	24.04 (14.35)	0.0075
ALT, U/L	26.96 (16.24)	25.12 (19.34)	0.33
γ-GTP, U/L	40.64 (32.52)	32.96 (29.27)	<0.001
Triglycerides, mg/dL	137.76 (83.36)	96.56 (34.01)	0.0034
HDL cholesterol, mg/dL	54.28 (12.44)	52.24 (10.24)	0.11
LDL cholesterol, mg/dL	108.80 (30.02)	82.88 (23.45)	<0.001
Total cholesterol, mg/dL	187.96 (34.86)	154.68 (25.38)	<0.001
hs-CRP, mg/L	1676.32 (3109.72)	630.28 (575.99)	0.10
eGFR, ml/min/1.73 m^2	67.87 (15.24)	63.72 (14.20)	<0.001
Smoking history, n (%)	11 (44)	—	—
Medication use			
Diabetes, n (%) Hypertension, n (%) Dyslipidemia, n (%)	22 (88) 14 (56) 19 (76)	— — —	— — —
Total teeth, n	20.84 (6.14)	—	—
Plaque index	1.02 (0.55)	—	—
Tongue coating index	2.08 (1.49)	—	—
PISA, mm^2	380.41 (232.28)	—	—
IMT, mm	2.05 (0.74)	—	—

Values are presented as the mean (SD), unless otherwise indicated. Paired t tests were used to compare differences between week two and baseline. HbA1c, hemoglobin A1c; BMI, body mass index; BP, blood pressure; AST, aspartate aminotransferase; ALT, alanine aminotransferase; γ-GTP, γ-Glutamyl transpeptidase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; hs-CRP, high-sensitivity C-reactive protein; eGFR, estimated glomerular filtration rate; PISA, periodontal inflamed surface area; IMT, carotid intima-media thickness.

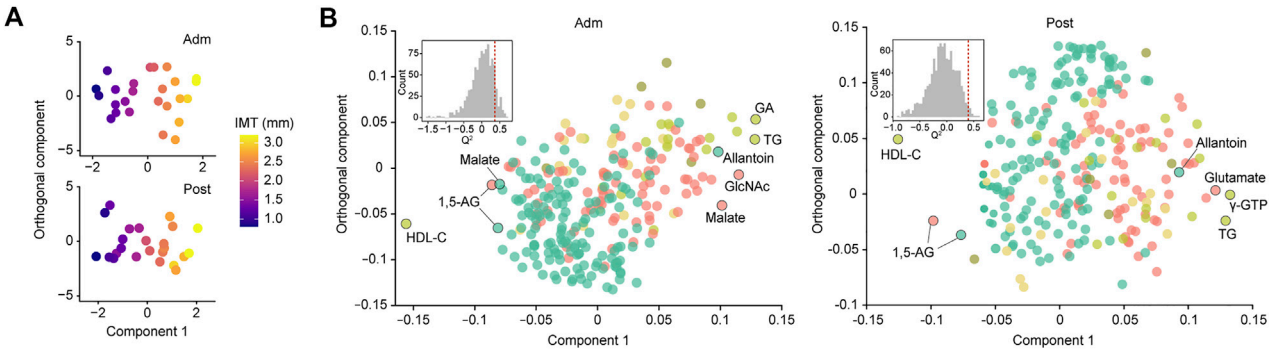


FIGURE 2
Global analysis of correlation of IMT with clinical and metabolomic features at admission (Adm) and after treatment (Post). **(A)** OPLS score plot showing distribution of study participants according to IMT. **(B)** OPLS loading plot showing color-coded distribution of predictors from different data blocks (pear: serum biochemical parameters, olive: urine biochemical parameters, light yellow: other clinical indices, red: plasma metabolome, mint: saliva metabolome), with the right side showing those associated with higher IMT. Inset shows statistical validation using permutation analysis (n = 999 permutations). GA, glycated albumin; TG, triglycerides; γ-GTP, γ-Glutamyl transpeptidase; HDL-C, high-density lipoprotein cholesterol; GlcNAc, *N*-acetylglucosamine; 1,5-AG, 1,5-anhydroglucitol.

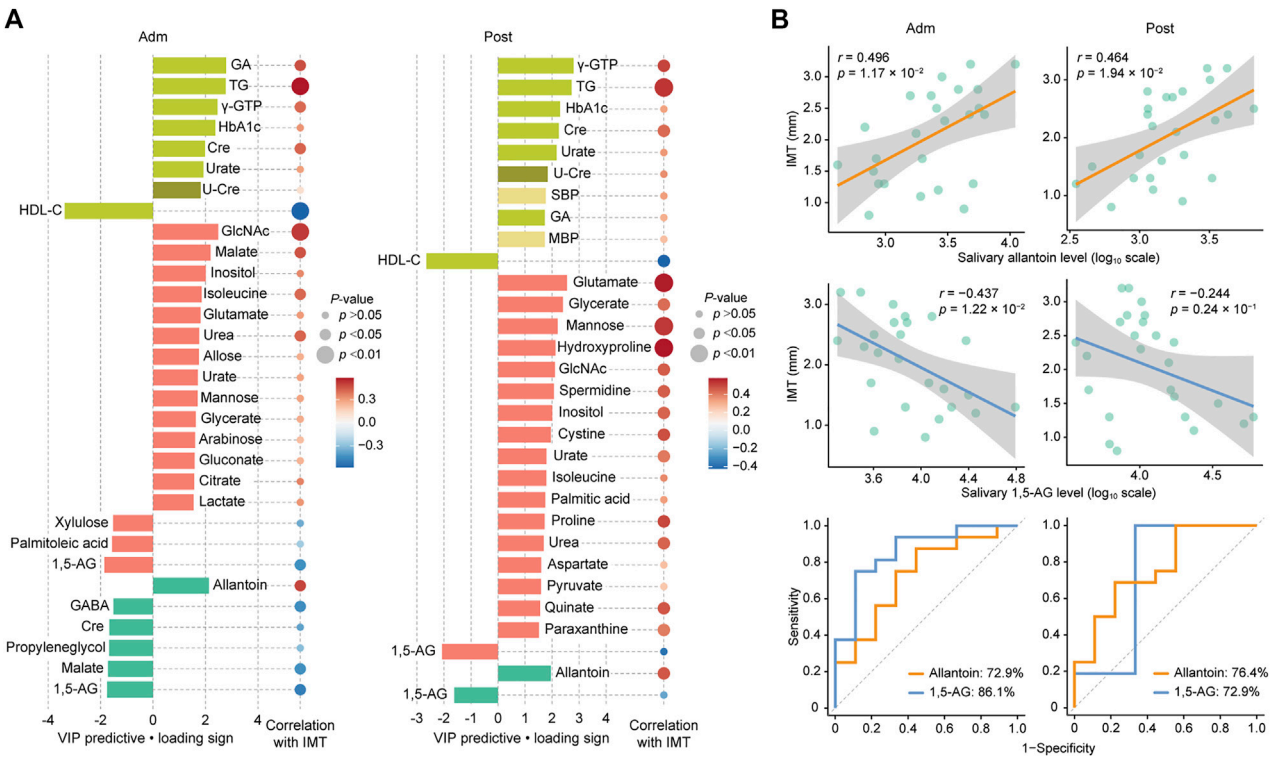


FIGURE 3
Top predictors for IMT from each data block at admission (Adm) and after treatment (Post). **(A)** Bar plots showing the variables with importance for projection of a predictive component (VIP predictive) higher than >1.5 (pear: serum biochemical parameters, olive: urine biochemical parameter, light yellow: other clinical indices, red: plasma metabolome, mint: saliva metabolome). Associations were also assessed using Spearman's correlation. **(B)** Associations of IMT with salivary allantoin and 1,5-AG, and ROC curves for comparing discriminative performance for IMT >1.6 mm using salivary levels of allantoin and 1,5-AG. GA, glycated albumin; TG, triglycerides; γ-GTP, γ-Glutamyl transpeptidase; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; Cre, creatinine; U-Cre, urine creatinine; SBP, systolic blood pressure; MBP, mean blood pressure; GlcNAc, *N*-acetylglucosamine; 1,5-AG, 1,5-anhydroglucitol; GABA, γ-Aminobutyric acid.

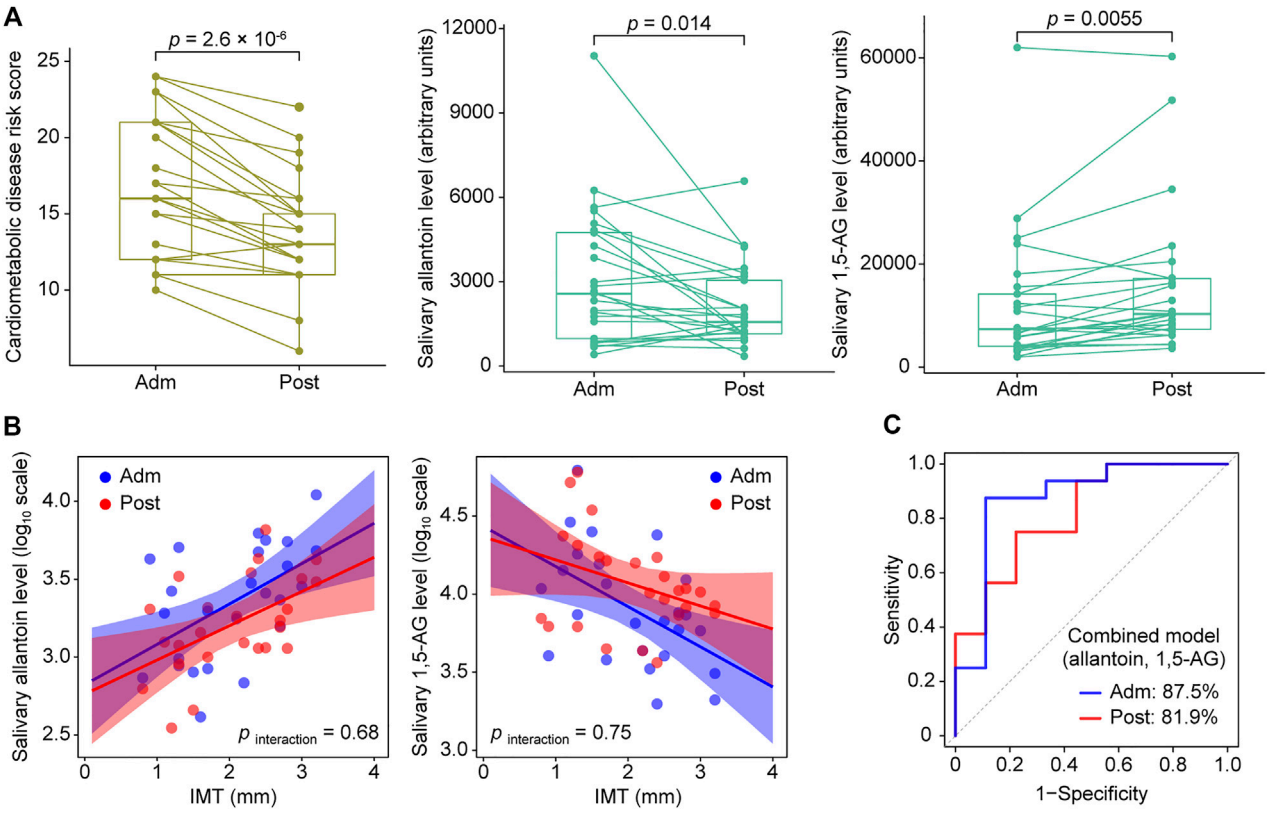


FIGURE 4
Responses of cardiometabolic risk factors and salivary metabolites following 2 weeks of intensive diabetes treatment. **(A)** Effects of intensive diabetes treatment on cardiometabolic disease risk score, and salivary levels of allantoin and 1,5-AG. The score for cardiometabolic disease risk was derived based on a composite score comprised of hemoglobin A1c, total cholesterol, triglycerides, high-density lipoprotein cholesterol, and high-sensitivity C-reactive protein values. *p* values shown are for within-group comparisons and were obtained with a paired *t* test for a comparison of the baseline score (Adm) with response after 2 weeks of treatment (Post). **(B)** Scatterplots of salivary markers with IMT according to hospitalized treatment. **(C)** ROC curves for comparing discriminative performance for IMT >1.6 mm using models combining salivary levels of allantoin and 1,5-AG at admission (Adm) as well as after treatment (Post).

including HbA1c, fasting plasma glucose, GA, and TG (Table 1), as demonstrated in our previous work (Taya et al., 2021). An OPLS model was then constructed using clinical and metabolomic parameters after 2 weeks of treatment for explanation of IMT variations. The model showed a moderate predictive ability of 0.398 for Q^2 and reliable performance during the permutation test (n = 999 permutations; Figure 2B, inset). Specifically, an unfavorable lipid profile remained predominant with the association between GA and IMT less pronounced, while amino acids became the predominant plasma metabolomic predictors for IMT. As for salivary metabolites, only allantoin and 1,5-AG remained relevant for IMT with a reduced predictive ability of 1,5-AG observed. Top predictors for IMT from each data block included the clinical markers γ-GTP (r = 0.503, p = 0.010), TG (r = 0.531, p = 0.0063), and HDL-C (r = −0.424, p = 0.035), the plasma metabolites glutamate (r = 0.571, p = 0.0029), glycerate (r = 0.425, p = 0.034), and mannose (r = 0.530, p = 0.0064), and the salivary metabolites allantoin (r = 0.464, p = 0.0194) and 1,5-AG (r = −0.244, p = 0.24) (Figure 3A). Additionally, salivary levels of allantoin and 1,5-AG showed AUC values of 0.764 and 0.729, respectively, for diagnosis in the high IMT group (Figure 3B).

Based on the OPLS results, we focused on salivary allantoin and 1,5-AG, and investigated the effects of hospitalized treatment on these salivary metabolites. A significant reduction in cardiometabolic disease risk score was noted following 2 weeks of hospitalization based on a composite score that summarized HbA1c, total cholesterol, TG, HDL-C, and hsCRP ($p = 2.6 \times 10^{-6}$) (Figure 4A). The two-week treatment also caused a significant decrease in the level of allantoin (p = 0.014) and an increase in 1,5-AG (p = 0.0055) in saliva (Figure 4A). However, that treatment did not modify the association between IMT and these metabolites ($p_{interaction}$ > 0.05) (Figure 4B), while diagnostic accuracy for the high IMT group was satisfactory using models with combined salivary levels of these metabolites at the baseline (AUC = 0.875) as well as after treatment (AUC =

0.819) (Figure 4C). Collectively, while diabetes treatment in a hospitalization setting induced mitigation of cardiometabolic risk with accompanying alterations in salivary levels of allantoin and 1,5-AG, it did not have an effect modification on the relationship between these salivary metabolites and atherosclerotic burden in T2D patients, with the combination of salivary allantoin and 1,5-AG remaining highly discriminative for those at high risk for CVD regardless of glycemic control status.

Discussion

The present findings demonstrated multivariate patterns of association of carotid wall thickening with metabolomic and clinical factors in patients with T2D. Additionally, they distinguished clinical and metabolomic parameters that change rapidly with improved glycemic control from those affected by relatively prolonged hyperglycemia, helping to find those more appropriate as markers of atherosclerosis, a cumulative disease. In particular, salivary allantoin and 1,5-AG remained the top metabolites reflecting IMT even after 2 weeks of inpatient glycemic control, revealing the potential utility of saliva testing for non-invasive assessment of carotid atherosclerosis severity, which might be useful for cardiovascular risk screening and monitoring of such patients.

Results of OPLS analysis with IMT as the outcome revealed clinical and metabolomic markers that exhibited a covariation with atherosclerotic burden in T2D patients. At the time of admission as well as after treatment, the severity of carotid atherosclerosis was associated with clinical markers in relation to an unfavorable lipid profile, such as lower HDL-C and higher TG, which is in agreement with several previous reports showing that increased TG/HDL-C ratio is an independent predictor of carotid atherosclerosis (Pacifico et al., 2014), as well as increased risk of CVD and all-cause mortality in patients with T2D (Wang et al., 2021b), and also non-diabetic subjects (Prasad et al., 2019; Sultani et al., 2020). The hsCRP is a known risk factor of CVD as well, but our analyses did not find it among its top predictors. The relatively large variance of hsCRP in our dataset may affect the results, however, we confirmed that exclusion of hsCRP did not change the main results in our OPLS models. Therefore, it is unlikely that the large variance of hsCRP undermines the validity of our models. The findings obtained in the present study also demonstrated GA to be more predictive of IMT than HbA1c at baseline. However, 2 weeks of intensive diabetes treatment attenuated the association of GA with IMT, which likely reflected improved glycemic control over a short term (Koga, 2014). It is considered that the associations between IMT and clinical markers shown in the present study corroborate the validity of the present measurements and metabolome analysis findings, as discussed following.

Regarding the relationship of plasma metabolites with IMT, GlcNAc was shown to have a stronger association with severity of carotid atherosclerosis at the baseline as compared to after treatment. Although there are no known clinical reports implicating an effect of plasma GlcNAc in development of atherosclerosis, a higher level of GlcNAc in plasma may well lead to increased protein modification *via* O-linked *β-N*-acetylglucosamine (O-GlcNAcylation), which has been implicated to be involved in development of diabetic cardiovascular complications (Chen et al., 2019; Chatham et al., 2020). Further investigation of the intricate relationship among plasma GlcNAc, O-GlcNAcylation, and atherosclerosis is warranted. The present results also indicated a number of known metabolites related to a higher atherosclerosis burden, such as inositol (Tzoulaki et al., 2019; Omori et al., 2020), isoleucine (Li et al., 2019), glutamate (Lehn-Stefan et al., 2021), urea (Omori et al., 2020), urate (Ishizaka et al., 2007), and hydroxyproline (Milanlouei et al., 2020). Additionally, data from OPLS analysis also indicated that 2 weeks of hospitalized treatment altered the profile of plasma metabolomic predictors for IMT, with amino acids predominant, which is likely associated with improved glycemic control status, though further validation is required.

A key finding in this study is that salivary allantoin was shown to be a potential indicator of IMT and cardiometabolic risk in patients with type 2 diabetes. Allantoin, produced by oxidation of urate, has been proposed as a biomarker for oxidative stress (Martinez-Moral and Kannan, 2019), while other studies have shown that urinary excretion of allantoin is correlated with atherosclerosis extension in mice (Li et al., 2015) and plasma allantoin is correlated with carotid atherosclerosis in humans (Santana et al., 2018). This is the first investigation to find an association between carotid atherosclerosis and salivary allantoin. However, allantoin was not annotated in the present plasma metabolomics results and, as noted in several other studies, no significant association was found between salivary allantoin and plasma urate, precluding mechanistic considerations. Nevertheless, it seems plausible that salivary excretion of allantoin becomes enhanced as a protective response to oxidative stress in association with diabetic angiopathy. Furthermore, since amelioration of hyperglycemia has been shown to reduce oxidative stress (Ohara et al., 2018), the lower level of oxidative stress induced by 2 weeks of intensive diabetes treatment is considered to have contributed to decreased salivary allantoin level, albeit the association with IMT was preserved. Although further research is needed, salivary allantoin might be a useful marker for reflecting the severity of atherosclerosis as part of monitoring CVD risk in patients with T2D.

Additionally, the present results demonstrated a correlation between IMT and 1,5-AG in both plasma and saliva samples, as well as a significant increase in salivary levels of 1,5-AG after 2 weeks of diabetes treatment. Previous studies have demonstrated that 1,5-AG can be used as a marker of short-term glycemic control and established it as a reliable T2D marker in addition to glucose (McGill et al., 2004). A positive association between plasma and salivary 1,5-

AG has been reported by others (Mook-Kanamori et al., 2014; Jian et al., 2020) and was shown in our previous study (Sakanaka et al., 2021), which also presented findings demonstrating the diagnostic utility of salivary 1,5-AG for hyperglycemia when salivary 1,5-AG was used alone as well as in combination with salivary mannose. Furthermore, the results presented here showed that hospitalized treatment attenuated the association between 1,5-AG and IMT. Prior clinical studies did not find a significant association between plasma 1,5-AG and carotid atherosclerosis in examinations of subjects in the general population (Mukai et al., 2015), or of patients with T2D or hypertension (Karrei et al., 2004). Therefore, 1,5-AG is likely a marker of glycemic status rather than atherosclerosis. Nonetheless, it is important to note that the combination of salivary 1,5-AG with salivary allantoin may provide better prediction of cardiovascular risk in type 2 diabetes patients.

The current study has several limitations. Although the aim of this investigation was to comprehensively characterize biochemical and metabolic markers that exhibit covariation with the severity of carotid atherosclerosis by integrating multiple parameters obtained through detailed measurements, the small sample size may not allow for extrapolation of the findings. Future studies that replicate and further develop the present results using a larger sample are required. Additionally, we performed IMT measurement at the site of greatest thickness, including plaque lesions, according to guidelines used in Japan (Terminology and Diagnostic Criteria Committee and Japan Society of Ultrasonics in Medicine, 2009), which are different from European guidelines recommending that IMT measurement should be performed in a region free of plaque, and the distinction between IMT and plaque clearly made (Touboul et al., 2007). Nevertheless, several studies have shown that assessment of carotid plaque is more useful than IMT measured in plaque-free areas for predicting future CVD (Störk et al., 2004; Inaba et al., 2012). Additionally, it has been demonstrated that incorporation of carotid plaque in IMT measurements can better predict cardiovascular events as compared with information derived from plaque alone or IMT without inclusion of plaque (Baldassarre et al., 2012). Hence, the IMT definition employed in the present study seems suitable to achieve the goal of identifying metabolic signatures of atherosclerotic burden in diabetes patients.

Conclusion

The present results show that a combination of salivary metabolites has robust associations with atherosclerotic burden in T2D patients and may be of high value for use in non-invasive identification of those at high risk for CVD in clinical practice. Additionally, they represent a new starting point for further investigations into the role of metabolites for exacerbation of diabetic macroangiopathy as well as their potential use for clinical diagnosis. It is considered that saliva testing will become even more widespread in the future, based on attention it has received due to the coronavirus pandemic, thus analysis of panels of metabolites in saliva may not only become an attractive alternative to blood tests for screening and monitoring of individuals with high risk for CVD, but might also help to reduce the daily burden faced by affected patients who must manage their symptoms over a long period of time. In addition, given the closer relationship developing between dentists and diabetologists, saliva testing during a regular dental visit may enable early warning of increased atherosclerotic burden in patients without subjective symptoms, thus strengthening the cooperation between medicine and dentistry for treatment of diabetes.

Data availability statement

The original contributions presented in the study are publicly available. These data can be found here: [https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST001905] and [https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST001906].

Ethics statement

The studies involving human participants were reviewed and approved by the Osaka University Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AS contributed to study design, data acquisition, analysis, and interpretation, and drafted and critically revised the manuscript. NK contributed to study design, data acquisition, analysis, and interpretation, and critically revised the manuscript. MF contributed to study design, data acquisition, analysis, and interpretation, and critically revised the manuscript. HN, KO, and NT contributed to study design, data acquisition, analysis, and interpretation, and critically revised the manuscript. AI, SM, MI, and ET contributed to data acquisition and critically revised the manuscript. IS, EF, and AA contributed to study design, and data analysis and interpretation, and critically revised the manuscript. MK contributed to study concept and design, data acquisition, analysis, and interpretation, and drafted and critically revised the manuscript. All authors gave final approval for submission of the manuscript and agree to be accountable for all aspects of this work.

Funding

The project was funded by the Japan Agency for Medical Research and Development (AMED-CREST; grant JP18gm0710005), as well as a Japan Society for the Promotion

of Science (JSPS) Grant-in-Aid for Challenging and Pioneering Research (number 20K20393) and JSPS KAKENHI grants (numbers 18H0468, 19H03862, 18K17281, and 20K18827).

Acknowledgments

The authors acknowledge the investigative group of the MEDENGINE project at Osaka University, as well as the study staff and all study participants. The MEDENGINE project is part of a collaboration between medicine, dentistry, and engineering groups at Osaka University, led by EF. The authors also thank Miho Kakiuchi, Harumi Aoki, Hitomi Yadori, and Kyoko Hashimoto for their assistance with data acquisition.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baldassarre, D., Hamsten, A., Veglia, F., de Faire, U., Humphries, S. E., Smit, A. J., et al. (2012). Measurements of carotid intima-media thickness and of interadventitia common carotid diameter improve prediction of cardiovascular events: Results of the IMPROVE (carotid intima media thickness [IMT] and IMT-progression as predictors of Vascular events in a high risk European population) study. *J. Am. Coll. Cardiol.* 60 (16), 1489-1499. doi:10.1016/j.jacc.2012.06.034

Chatham, J. C., Young, M. E., and Zhang, J. (2020). Reprint of: Role of O-linked N-acetylglucosamine (O-GlcNAc) modification of proteins in diabetic cardiovascular complications. *Curr. Opin. Pharmacol.* 54, 209-220. doi:10.1016/j.coph.2020.11.005

Chen, Y., Zhao, X., and Wu, H. (2019). Metabolic stress and cardiovascular disease in diabetes mellitus: The role of protein O-GlcNAc modification. *Arterioscler. Thromb. Vasc. Biol.* 39 (10), 1911-1924. doi:10.1161/ATVBAHA.119.312192

Fan, Y., and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* 19 (1), 55-71. doi:10.1038/s41579-020-0433-9

Haneda, M., Noda, M., Origasa, H., Noto, H., Yabe, D., Fujita, Y., et al. (2018). Japanese clinical practice guideline for diabetes 2016. *J. Diabetes Investig.* 9 (3), 657-697. doi:10.1111/jdi.12810

Inaba, Y., Chen, J. A., and Bergmann, S. R. (2012). Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: A meta-analysis. *Atherosclerosis* 220 (1), 128-133. doi:10.1016/j.atherosclerosis.2011.06.044

Ishizaka, N., Ishizaka, Y., Toda, E., Hashimoto, H., Nagai, R., and Yamakado, M. (2007). Higher serum uric acid is associated with increased arterial stiffness in Japanese individuals. *Atherosclerosis* 192 131, 137. (1). doi:10.1016/j.atherosclerosis.2006.04.016

Jian, C., Zhao, A., Ma, X., Ge, K., Lu, W., Zhu, W., et al. (2020). Diabetes screening: Detection and application of saliva 1, 5-anhydroglucitol by liquid chromatography-mass spectrometry. *J. Clin. Endocrinol. Metab.* 105 dgaa114(6). doi:10.1210/clinem/dgaa114

Karrei, K., Koehler, C., Hanefeld, M., Temelkova-Kurktschiev, T., and Del Prato, S. (2004). Fluctuations in glycaemia in clinical diabetes mellitus type 2 are not associated with carotid intima-media thickening. *Diab. Vasc. Dis. Res.* 1 (1), 51-52. doi:10.3132/dvdr.2004.007

Katakami, N., Omori, K., Taya, N., Arakawa, S., Takahara, M., Matsuoka, T. A., et al. (2020). Plasma metabolites associated with arterial stiffness in patients with type 2 diabetes. *Cardiovasc. Diabetol.* 19 (1), 75. doi:10.1186/s12933-020-01057-w

Koga, M. (2014). Glycated albumin; clinical usefulness. *Clin. Chim. Acta.* 433, 96-104. doi:10.1016/j.cca.2014.03.001

Lehn-Stefan, A., Peter, A., Machann, J., Schick, F., Randrianarisoa, E., Heni, M., et al. (2021). Elevated circulating glutamate is associated with subclinical atherosclerosis independently of established risk markers: A cross-sectional study. *J. Clin. Endocrinol. Metab.* 106 e982, e989. (2). doi:10.1210/clinem/dgaa898

Li, C., He, J., Li, S., Chen, W., Bazzano, L., Sun, X., et al. (2019). Novel metabolites are associated with augmentation index and pulse wave velocity: Findings from the bogalusa heart study. *Am. J. Hypertens.* 32 (6), 547-556. doi:10.1093/ajh/hpz046

Li, D., Zhang, L., Dong, F., Liu, Y., Li, N., Li, H., et al. (2015). Metabonomic changes associated with atherosclerosis progression for LDLR(-/-) mice. *J. Proteome Res.* 14 (5), 2237-2254. doi:10.1021/acs.jproteome.5b00032

Lorenz, M. W., Markus, H. S., Bots, M. L., Rosvall, M., and Sitzer, M. (2007). Prediction of clinical cardiovascular events with carotid intima-media thickness: A systematic review and meta-analysis. *Circulation* 115 (4), 459-467. doi:10.1161/CIRCULATIONAHA.106.628875

Mach, F., Baigent, C., Catapano, A. L., Koskinas, K. C., Casula, M., Badimon, L., et al. (2019). 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur. Heart J.* 41 (1), 111-188. doi:10.1093/eurheartj/ehz455

Martinez-Moral, M-P., and Kannan, K. (2019). Allantoin as a marker of oxidative stress: Inter- and Intraindividual variability in urinary concentrations in Healthy individuals. *Environ. Sci. Technol. Lett.* 6 (5), 283-288. doi:10.1021/acs.estlett.9b00142

McGill, J. B., Cole, T. G., Nowatzke, W., Houghton, S., Ammirati, E. B., Gautille, T., et al. (2004). Circulating 1, 5-anhydroglucitol levels in adult patients with diabetes reflect longitudinal changes of glycemia: A U.S. Trial of the GlycoMark assay. *Diabetes Care* 27 (8), 1859-1865. doi:10.2337/diacare.27.8.1859

Milanlouei, S., Menichetti, G., Li, Y., Loscalzo, J., Willett, W. C., and Barabasi, A. L. (2020). A systematic comprehensive longitudinal evaluation of dietary factors associated with acute myocardial infarction and fatal coronary heart disease. *Nat. Commun.* 11 (1), 6074. doi:10.1038/s41467-020-19888-2

Mook-Kanamori, D. O., Selim, M. M., Takiddin, A. H., Al-Homsi, H., Al-Mahmoud, K. A., Al-Obaidli, A., et al. (2014). 1, 5-Anhydroglucitol in saliva is a noninvasive marker of short-term glycemic control. *J. Clin. Endocrinol. Metab.* 99 E479, E483. (3). doi:10.1210/jc.2013-3596

Morze, J., Wittenbecher, C., Schwingshackl, L., Danielewicz, A., Rynkiewicz, A., Hu, F. B., et al. (2022). Metabolomics and type 2 diabetes risk: An Updated systematic review and meta-analysis of prospective cohort studies. *Diabetes Care* 45 (4), 1013-1024. doi:10.2337/dc21-1705

Mukai, N., Ninomiya, T., Hata, J., Hirakawa, Y., Ikeda, F., Fukuhara, M., et al. (2015). Association of hemoglobin A1c and glycated albumin with carotid atherosclerosis in community-dwelling Japanese subjects: The Hisayama study. *Cardiovasc. Diabetol.* 14, 84. doi:10.1186/s12933-015-0247-7

Ohara, M., Nagaike, H., Goto, S., Fukase, A., Tanabe, Y., Tomoyasu, M., et al. (2018). Improvements of ambient hyperglycemia and glycemic variability are associated with reduction in oxidative stress for patients with type 2 diabetes. *Diabetes Res. Clin. Pract.* 139, 253-261. doi:10.1016/j.diabres.2018.02.017

Olesen, K. K. W., Madsen, M., Egholm, G., Thim, T., Jensen, L. O., Raungaard, B., et al. (2017). Patients with diabetes without significant angiographic coronary artery disease have the same risk of myocardial infarction as patients without diabetes in a real-world population receiving appropriate prophylactic treatment. *Diabetes Care* 40 (8), 1103-1110. doi:10.2337/dc16-2388

Omori, K., Katakami, N., Arakawa, S., Yamamoto, Y., Ninomiya, H., Takahara, M., et al. (2020). Identification of plasma inositol and Indoxyl Sulfate as Novel biomarker Candidates for atherosclerosis in patients with type 2 diabetes. -Findings

from metabolome analysis using GC/MS. *J. Atheroscler. Thromb.* 27 (10), 1053-1067. doi:10.5551/jat.52506

Pacifico, L., Bonci, E., Andreoli, G., Romaggioli, S., Di Miscio, R., Lombardo, C. V., et al. (2014). Association of serum triglyceride-to-HDL cholesterol ratio with carotid artery intima-media thickness, insulin resistance and nonalcoholic fatty liver disease in children and adolescents. *Nutr. Metab. Cardiovasc. Dis.* 24 (7), 737-743. doi:10.1016/j.numecd.2014.01.010

Prasad, M., Sara, J., Widmer, R. J., Lennon, R., Lerman, L. O., and Lerman, A. (2019). Triglyceride and triglyceride/HDL (high density lipoprotein) ratio predict major adverse cardiovascular outcomes in women with non-obstructive coronary artery disease. *J. Am. Heart Assoc.* 8 e009442(9). doi:10.1161/jaha.118.009442

Sakanaka, A., Kuboniwa, M., Katakami, N., Furuno, M., Nishizawa, H., Omori, K., et al. (2021). Saliva and plasma reflect metabolism altered by diabetes and periodontitis. *Front. Mol. Biosci.* 8. 742002, doi:10.3389/fmolb.2021.742002

Santana, M. S., Nascimento, K. P., Lotufo, P. A., Benseaor, I. M., and Meotti, F. C. (2018). Allantoin as an independent marker associated with carotid intima-media thickness in subclinical atherosclerosis. *Braz J. Med. Biol. Res.* 51 e7543(8). doi:10.1590/1414-431x20187543

Sattar, N. (2013). Revisiting the links between glycaemia, diabetes and cardiovascular disease. *Diabetologia* 56 (4), 686-695. doi:10.1007/s00125-012-2817-5

Sharma, K., Karl, B., Mathew, A. V., Gangoiti, J. A., Wassel, C. L., Saito, R., et al. (2013). Metabolomics Reveals signature of Mitochondrial dysfunction in diabetic Kidney disease. *J. Am. Soc. Nephrol.* 24 (11), 1901-1912. doi:10.1681/ASN.2013020126

Störk, S., van den Beld, A. W., von Schacky, C., Angermann, C. E., Lamberts, S. W. J., Grobbee, D. E., et al. (2004). Carotid artery plaque burden, stiffness, and mortality risk in elderly men: A prospective, population-based cohort study. *Circulation* 110 (3), 344-348. doi:10.1161/01.CIR.0000134966.10793.C9

Sultani, R., Tong, D. C., Peverelle, M., Lee, Y. S., Baradi, A., and Wilson, A. M. (2020). Elevated triglycerides to high-density lipoprotein cholesterol (TG/HDL-C) ratio predicts long-term mortality in high-risk patients. *Heart Lung Circ.* 29 (3), 414-421. doi:10.1016/j.hlc.2019.03.019

Taya, N., Katakami, N., Omori, K., Arakawa, S., Hosoe, S., Watanabe, H., et al. (2021). Evaluation of change in metabolome caused by comprehensive diabetes treatment: A prospective observational study of diabetes inpatients with gas chromatography/mass spectrometry-based non-target metabolomic analysis. *J. Diabetes Investig.* 10. doi:10.1111/jdi.13600

Terminology and Diagnostic Criteria Committee, Japan Society of Ultrasonics in Medicine (2009). Standard method for ultrasound evaluation of carotid artery lesions. *J. Med. Ultrason.* 36 (4), 219-226. doi:10.1007/s10396-009-0238-y

Touboul, P. J., Hennerici, M. G., Meairs, S., Adams, H., Amarenco, P., Bornstein, N., et al. (2007). Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the Advisory board of the 3rd and 4th Watching the risk Symposium, 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006. *Cerebrovasc. Dis.* 23 (1), 75-80. doi:10.1159/000097034

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12 (6), 523-526. doi:10.1038/nmeth.3393

Tzoulaki, I., Castagne, R., Boulange, C. L., Karaman, I., Chekmeneva, E., Evangelou, E., et al. (2019). Serum metabolic signatures of coronary and carotid atherosclerosis and subsequent cardiovascular disease. *Eur. Heart J.* 40 (34), 2883-2896. doi:10.1093/eurheartj/ehz235

Wang, D. D., Nguyen, L. H., Li, Y., Yan, Y., Ma, W., Rinott, E., et al. (2021). The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nat. Med.* 27 (2), 333-343. doi:10.1038/s41591-020-01223-3

Wang, L., Cong, H., Zhang, J., Hu, Y., Wei, A., Zhang, Y., et al. (2021). Predictive value of the triglyceride to high-density lipoprotein cholesterol ratio for all-cause mortality and cardiovascular death in diabetic patients with coronary artery disease treated with statins. *Front. Cardiovasc. Med.* 8. 718604, doi:10.3389/fcvm.2021.718604

Wheelock, A. M., and Wheelock, C. E. (2013). Trials and tribulations of 'omics data analysis: Assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* 9 (11), 2589-2596. doi:10.1039/c3mb70194h

Wishart, D. S. (2019). Metabolomics for investigating Physiological and Pathophysiological processes. *Physiol. Rev.* 99 (4), 1819-1875. doi:10.1152/physrev.00035.2018

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fmolb.2023.1154536

OPEN ACCESS

EDITED BY
Andras Szeitz,
University of British Columbia, Canada

REVIEWED BY
Sven Schuchardt,
Fraunhofer Institute for Toxicology and Experimental Medicine (FHG), Germany
Mario Humberto Vargas,
National Institute of Respiratory Diseases-Mexico (INER), Mexico

*CORRESPONDENCE
Alexander Moeller,
alexander.moeller@kispi.uzh.ch

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work and share last authorship

SPECIALTY SECTION
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 30 January 2023
ACCEPTED 16 March 2023
PUBLISHED 31 March 2023

CITATION
Weber R, Streckenbach B, Welti L, Inci D, Kohler M, Perkins N, Zenobi R, Micic S and Moeller A (2023), Online breath analysis with SESI/HRMS for metabolic signatures in children with allergic asthma. *Front. Mol. Biosci.* 10:1154536. doi: 10.3389/fmolb.2023.1154536

Online breath analysis with SESI/HRMS for metabolic signatures in children with allergic asthma

Ronja Weber[1†], Bettina Streckenbach[2†], Lara Welti[1], Demet Inci[1], Malcolm Kohler[3], Nathan Perkins[4], Renato Zenobi[2], Srdjan Micic[1‡] and Alexander Moeller[1*‡]

[1]Department of Respiratory Medicine, University Children's Hospital Zurich, Zurich, Switzerland, [2]Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland, [3]Department of Pulmonology, University Hospital Zurich, Zurich, Switzerland, [4]Division of Clinical Chemistry and Biochemistry, University Children's Hospital Zurich, Zurich, Switzerland

Introduction: There is a need to improve the diagnosis and management of pediatric asthma. Breath analysis aims to address this by non-invasively assessing altered metabolism and disease-associated processes. Our goal was to identify exhaled metabolic signatures that distinguish children with allergic asthma from healthy controls using secondary electrospray ionization high-resolution mass spectrometry (SESI/HRMS) in a cross-sectional observational study.

Methods: Breath analysis was performed with SESI/HRMS. Significant differentially expressed mass-to-charge features in breath were extracted using the empirical Bayes moderated t-statistics test. Corresponding molecules were putatively annotated by tandem mass spectrometry database matching and pathway analysis.

Results: 48 allergic asthmatics and 56 healthy controls were included in the study. Among 375 significant mass-to-charge features, 134 were putatively identified. Many of these could be grouped to metabolites of common pathways or chemical families. We found several pathways that are well-represented by the significant metabolites, for example, lysine degradation elevated and two arginine pathways downregulated in the asthmatic group. Assessing the ability of breath profiles to classify samples as asthmatic or healthy with supervised machine learning in a 10 times repeated 10-fold cross-validation revealed an area under the receiver operating characteristic curve of 0.83.

Discussion: For the first time, a large number of breath-derived metabolites that discriminate children with allergic asthma from healthy controls were identified by online breath analysis. Many are linked to well-described metabolic pathways and chemical families involved in pathophysiological processes of asthma. Furthermore, a subset of these volatile organic compounds showed high potential for clinical diagnostic applications.

KEYWORDS

volatile organic compounds (VOCs), metabolites, allergic asthma, children, breath analysis, SESI/HRMS

1 Introduction

Asthma is the most frequent chronic condition in children in the developed world. The disease is very heterogeneous in its presentation and clinical course. Due to the lack of a well-recognized and easy to apply diagnostic gold-standard (Gaillard et al., 2021), misdiagnosis is relatively common. Reported numbers range from 10% to 62% for underdiagnosis (Kaur et al., 1998; Siersted et al., 1998; van Gent et al., 2007) from 48% to 53% for overdiagnosis (Looijmans-van den Akker et al., 2016; Yang et al., 2017). This has negative impacts on asthma related morbidity, quality-of-life, medication side-effects, prognosis, and health costs. Therefore, the investigation of pediatric asthma and its associated molecular processes including airway inflammation is of high importance for the development of novel, much-needed diagnostic and monitoring applications.

Breath is known to contain several hundreds of metabolites that reflect metabolism as well as disease-specific mechanisms such as airway inflammation (Ferraro et al., 2018). Therefore, there is great interest in discovering endogenous exhaled organic compounds that are linked to diseases and their pathophysiological processes (Neerincx et al., 2017). One of the few clinical tests taking advantage of this is the quantification of exhaled fractional nitric oxide (FeNO), which can be measured in all age groups. FeNO is a biomarker for eosinophilic airway inflammation that is related to allergic asthma (Ferraro et al., 2018). This exemplifies the potential of applying breath analysis to further study allergic asthma and improve the diagnostic power of exhaled biomarkers.

Several breath analysis studies attempted to distinguish children with asthma from healthy controls by different techniques. Dallinga and colleagues compared exhaled breath of children with asthma and a healthy group by gas chromatography mass spectrometry and identified a small set of discriminatory volatile organic compounds (VOCs) that is potentially related to lipid peroxidation, including various hydrocarbons, xylene, benzoic acid, and butanoic acid (Dallinga et al., 2010). A pilot study from van Mastrigt et al. identified VOCs discriminating children with asthma, cystic fibrosis and healthy controls by using a broadband quantum cascade laser spectroscopy technique (van Mastrigt et al., 2016). The distinguishing compound classes included different carboxylic acids, esters, and ethers. Altogether, there is only little overlap between the detected metabolites of different studies and sometimes even conflicting results are reported. Therefore, standardization as well as external validation are challenges that need further research as summarized in recent reviews (Neerincx et al., 2017; Ferraro et al., 2018; Papamichael et al., 2021).

Secondary electrospray ionization high-resolution mass spectrometry (SESI/HRMS) is a technology applied for online breath analysis that links real-time measurements without sample preparation to high mass resolution (Gaugg et al., 2019). The latter strongly improves the confidence in compound identification of the detected mass-to-charge features (*m/z* features). Previous studies confirmed the potential of this technology to identify relevant exhaled organic compounds, including biological metabolites, and reveal altered molecular pathways for different respiratory diseases (Schwarz et al., 2016; Gaugg et al., 2019; Weber et al., 2020). A strength of SESI/HRMS lies in the detection of polar molecules with high molecular masses and low volatility (Gaugg et al., 2016; Bruderer et al., 2019; Chen et al., 2021). Furthermore, its applicability in children was confirmed in our previous study on cystic fibrosis (Weber et al., 2020).

The aim of this study was to identify metabolic signatures in exhaled breath consisting of discriminating organic compounds specific to allergic asthma in children by SESI/HRMS and to assess their biological context.

2 Materials and methods

2.1 Study design, participants, and clinical data

This observational cross-sectional study included children with allergic asthma and healthy controls, aged 5–18 years. Asthmatic patients from the outpatient clinic of the University Children's Hospital Zürich, Switzerland, were recruited for this study. Asthma diagnosis was based on the recent ERS evidence-based practice guidelines (Gaillard et al., 2021) and only children with confirmed asthma were included. Allergic sensitization was defined by either a positive skin prick test or an allergen-specific IgE of >0.35 kU·L–1 by radioallergosorbent test or by ELISA for at least one common aeroallergen. Further, eligible patients were clinically stable enough to temporarily stop the inhalation of long-acting asthma medication at least 1 week before the measurements. Exclusion criteria were the inability to stop medication, and the presence of an acute respiratory infection during the last 2 weeks before the measurement. Clinical data was collected on the same day as breath analysis and is summarized together with anthropometric data in Table 1. Healthy controls without any chronic respiratory symptoms or known lung diseases were recruited from the public. The presence of an acute (respiratory) infection was an exclusion criterion for both groups. The measurement and recruitment period were in parallel and lasted for 15 months. Efforts were put into recruiting participants of both cohorts at a randomized schedule across daytime and throughout the study period. The sample size was based on our previous study with a similar design (Weber et al., 2020). All participants, where appropriate and parents gave their written informed consent in advance. The study was approved by the local ethics committee (KEK-ZH ID 2018–00441) and was conducted in accordance with the Declaration of Helsinki.

2.2 Breath analysis

Online breath analysis was performed using a SESI source (SuperSESI, FIT FossilionTech, Madrid, Spain) connected to a high-resolution time-of-flight mass spectrometer (TripleTOF 5600+, AB Sciex, Concord, ON, Canada). Methodological details and instrumental settings were previously described by our group (Weber et al., 2020). Minor adaptations are specified below. Children were exhaling directly into the instrument in a sitting position. The breathing maneuver consisted of at least three long exhalations at a constant pressure of 5 mbar with short breaks in between. A single-use mouthpiece (product No. 100078, ACE Instruments, Germany) was connected to the ionization source, which was heated to 130°C, by a sterilizable, custom-made polytetrafluoroethylene adapter. Measurements were recorded in

TABLE 1 Participant characteristics.

	Allergic asthma (n = 48)	Healthy controls (n = 56)	*p*-value
Age [y]	12.1 ± 3.1	10.8 ± 4.0	0.07
Male sex [n]	33 (68.8%)	24 (42.9%)	0.01
BMI [kg/m^2]	19.3 ± 4.2	18.3 ± 3.3	0.2
FEV1 [z-score]	−0.6 ± 1.1	−0.1 ± 1.0 †	0.01
FVC [z-score]	0.1 ± 1.0	0.1 ± 0.9 †	0.86
FeNO [ppb]	28.6 IQR 34.4	6.2 IQR 9.6	<0.001
Allergic sensitization [n]	48 (100%)	12 (21.4%)	<0.001

Data are presented as mean ± standard deviation (SD), n (%), or median and interquartile range (IQR). BMI, body mass index, pre-bronchodilator FEV1 = forced expiratory volume in 1 s, pre-bronchodilator FVC, forced vital capacity, FeNO, fractional exhaled nitric oxide. *p*-values were determined by the two sample *t*-test, Fisher's exact test for sex and allergic sensitization distribution, and the Mann-Whitney U test for FeNO, values (no normal distribution). †: 16 spirometries were excluded because of poor quality.

positive (4500 V) and negative (−4500 V) ionization mode between the *m/z* range of 50–500 Da. The accumulation time was set to 0.5 s per scan. The collisionally activated dissociation (CAD) gas was adjusted to 0 to avoid fragmentation. Temperatures of the MS were set to 0, gas 1 was used to pressurize the vial of the electrospray and set to 24, gas 2 was not connected to the SESI source, and the curtain gas was set to 10. The pulser frequency was adjusted to 23.983 kHz and the pulse 1 duration was 2 µs. A net flow of 0.3 L/min was defined by a mass flow controller (Alicat Scientific, Inc., Tucson, AZ, United States) at the exhaust of the ionization source. The nanoelectrospray was generated using silica emitters (50 cm length, 20 µm diameter, New Objective Inc., Woburn, MA, United States) and a 0.1% (v/v) aqueous formic acid solution (Optima LC/MS Grade, Thermo Fisher Scientific, Waltham, MA, United States). All participants were asked not to brush their teeth, consume any food, drinks (except for water), or chewing gums 1 h prior to the measurements (Weber et al., 2020).

2.3 Data preprocessing

All data were recalibrated in PeakView 2.2 (AB Sciex, Concord, ON, Canada) and processed in R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). The conversion and preprocessing of the raw data were done in the same way as described in our previous work (Weber et al., 2020). In brief, the raw mass spectra were resampled by interpolation (Δ*m/z*: 0.0005, *m/z* range: 50–500 Da), peak picking was performed on the average mass spectra associated with exhalation and signal intensities of the *m/z* features were determined by trapezoidal integration. The intensities of the *m/z* features were normalized to the total ion current, $\log_2$-transformed and arranged into a data matrix of breath profiles for further analysis. More details on data preprocessing are given in the Supplementary Material.

2.4 Statistical analysis

To account for confounding influences and reduce the heterogeneity within the groups, batch adjustment was performed by applying surrogate variable analysis (SVA) (Leek and Storey, 2007) on the data matrix of breath profiles. Identification of differentially expressed *m/z* features when comparing cases and controls was assessed by the empirical Bayes moderated t-statistics test (Smyth, 2004). Correction for multiple hypothesis testing was conducted using Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) with significance threshold set to the adjusted *p*-value of 0.05 to determine the statistically significant features. Additionally, the ability of the breath profiles to classify samples as allergic asthmatic or healthy was assessed with the support vector machines algorithm [13] trained and tested in a 10 times repeated stratified 10-fold cross-validation. To avoid using all features for classifier development, Boruta feature selection (Kursa et al., 2010) was applied in each cross-validation iteration to include only the potentially discriminating features between the allergic asthmatic and the healthy control group. In order to prevent bias during cross-validation all preprocessing steps, feature selection and classifier development were strictly conducted on the training data sets, preventing any information leak from the left-out samples (Varma and Simon, 2006). Details on statistical analysis are given in the supplementary material.

2.5 Feature identification

For the 100 most significantly discriminative *m/z* features per study group, compound identification was based on MS^2 spectra that were recorded directly from exhaled breath by SESI/HRMS with the same instrument set up. The settings of the SESI source and TripleTOF MS were identical to the ones described above for the MS1 full scan acquisition, with the following exceptions: the accumulation time was set to 1.0 s per scan and the CAD gas to 6. Precursors were selected with an isolation window of 0.7 Da. Collision energy for precursor fragmentation was set to 20 eV with a ramped energy spread of ± 10 eV. The MS^2 spectra were analyzed by a workflow adapted from a published method (Kaeslin et al., 2021) to detect isotopes, adducts, and losses, and with the SIRIUS software (v4.9.9) (Dührkop et al., 2019) to assign putative molecular formulae and chemical structures. The putatively identified compounds were screened for their biological context and subgrouped into metabolic pathways or chemical families. Additionally, pathway enrichment analysis using the mummichog algorithm (MetaboAnalyst, v5.0) (Pang et al., 2021) was performed for further identification on all

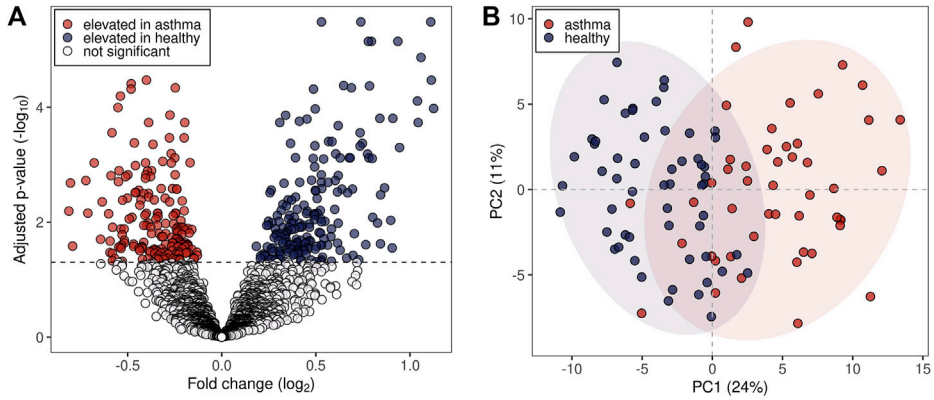


FIGURE 1
Statistical analysis of *m/z* features in breath profiles. **(A)** Volcano plot representing all detected 2,315 *m/z* features. Dashed line: Benjamini-Hochberg adjusted *p*-value of 0.05. **(B)** First two principal components (PCs) score plot of the 134 putatively identified *m/z* features. Blue dots represent healthy probands and red dots asthmatic patients. 95% data ellipses were added per group for visual depiction.

significant features including those without recorded MS^2 spectra, without compound suggestions or with excluded MS^2 spectra (exclusion criteria: see Supplementary Table S1). Lastly, the detected *m/z* features were compared with previously identified compounds from literature. The certainty of identification was indicated by an identification (ID) confidence level ranging from ID 1 to 5, as described by Schymanski and colleagues (Schymanski et al., 2014). More details on the identification approach are included in the supplementary material, including a schematic overview (Supplementary Figure S1).

3 Results

3.1 Participants and clinical data

Exhaled breath samples of 48 allergic asthma patients and 56 healthy control participants, in total 104 children, were included in this study. The age and body mass index values of the two cohorts were comparable, whereas the asthmatic group contained more males than the healthy one. Detailed clinical characteristics of the two individual study cohorts are shown in Table 1. The use of short-acting beta-agonists was allowed until the day before measurements. All children with asthma had a known allergic sensitization to at least one aero-allergen and the asthma severity ranged from mild to moderate. The FeNO values were significantly elevated in the allergic asthma cohort. Additionally, the forced expiratory volume in 1 s (FEV1) of the asthmatics was lower, while the forced vital capacity of the groups was comparable. Twelve children of the healthy control group showed an allergic sensitization according to the skin prick test, but only two reported symptomatic allergies.

3.2 Discriminative breath patterns and their metabolic associations

The pre-processing of the acquired mass spectra of the study subjects revealed 2,315 *m/z* features associated with exhaled breath. 375 *m/z* features were found to be significantly different between the two groups (Benjamini-Hochberg adjusted $p < 0.05$), of which 179 were upregulated and 196 downregulated in the allergic asthma group (Figure 1A). Among those, 134 were assigned to compounds. Inspection of the first two principal components (PCs) of the 134 putatively identified features revealed a moderate separation between the groups along the first PC (24% variance in the data, Figure 1B).

Compound identification revealed several specific metabolic pathways and chemical families with many representatives for both study cohorts (Tables 2, 3). For the allergic asthma group, the chemical families of fatty acid metabolites and monosaccharides as well as the 2-oxocarboxylic acid metabolism and two amino acid pathways, i.e., lysine degradation and tyrosine metabolism, were elevated (Table 2). The relations of metabolites involved in some of these elevated pathways are visualized in Figure 2. For the diminished compounds, arginine pathways were found to be well represented, including both arginine and proline metabolism and arginine biosynthesis. Further, several compounds of the linoleic acid metabolism and of the chemical groups of aldehydes, amides, and fatty acids were identified (Table 3; Figure 3). A full list including more details about the putatively identified compounds can be found in Supplementary Table S2.

The assessment of the classification accuracy in discriminating between the allergic asthmatic and the healthy samples resulted in an area under the curve (AUC) of 0.83, 95% CI: 0.73–0.92, (Figure 4A; Supplementary Table S3; Supplementary Figure S2). When examining feature selection by the Boruta scheme (Kursa et al., 2010) within cross-validation, 57 (±8) *m/z* features were selected on average in each cross-validation iteration, many of which were putatively identified with the compound identification workflow above (Figure 4B). Compounds which were most frequently selected in LOOCV are presented in Figure 4C (for box plots see Supplementary Figure S2) and all the other selected metabolites can be found in Supplementary Table S4.

It is of relevance to note that the adjustment with SVA captures the components of variability within the data and reduces any effect on the intensity levels of *m/z* features arising from other sources than the primary variables of interest (i.e., allergic asthma vs. healthy controls). Hence, any further subgroup analysis or correlation analysis to other clinical parameters could not be performed (Leek and Storey, 2007)

10.3389/fmolb.2023.1154536

TABLE 2 Pathway related metabolites elevated in the allergic asthma cohort.

m/z	Charge	Adj. *p*-value	Log-fold-change	Molecular formula	Ionisation	Δm (ppm)	Compound	Ann.	ID level
Lysine degradation									
131.035	neg	0.0002	−0.39	C5H8O4	[M-H]-	0.1	Glutarate (pentanedioic acid)†	MS^2, Lit.	ID1
117.019	neg	0.0004	−0.37	C4H6O4	[M-H]-	−2.8	Succinate (butanedioic acid)†	MS^2, Lit.	ID1
129.019	neg	0.0008	−0.25	C5H8O5	[M-H2O-H]-	−2.6	2-Hydroxyglutarate (2-hydroxypentanedioc acid)	MS^2	ID3
131.033	pos	0.0031	−0.46	C5H6O4	[M + H]+	−6.8	Glutaconate (2-pentenedioic acid)	MS^2	ID3
161.0435	pos	0.0032	−0.28	C6H8O5	[M + H]+	−5.9	2-Oxoadipate (2-oxohexanedioic acid)†	MS^2	ID3
162.0755	pos	0.0070	−0.25	C6H11NO4	[M + H]+	−3.6	2-Aminoadipate (2-aminohexanedioic acid)	MS^2	ID3
97.029	neg	0.0137	−0.20	C5H8O3	[M-H2O-H]-	−5.2	Glutarate semialdehyde (ω-oxopentanoaic acid)	MS^2	ID3
Tyrosine metabolism									
192.0285	neg	0.0001	−0.28	C9H7NO4	[M-H]-	−9.0	5,6-Dihydroxyindole-2-carboxylate	MS^2	ID3
177.075	pos	0.0017	−0.41	C7H10O4	[M + H2O + H]+	−4.2	Succinylacetone	MS^2	ID3
183.0295	neg	0.0038	−0.30	C8H10O6	[M-H2O-H]-	−2.2	Succinylacetoacetate	MS^2	ID3
215.052	pos	0.0076	−0.41	C9H8O5	[M + H2O + H]+	−14.0	3,4-Dihydroxyphenylpyruvate	MS^2	ID3
181.0505	neg	0.0123	−0.23	C9H10O4	[M-H]-	−0.7	4-Hydroxyphenyllactate	MS^2	ID3
163.039	neg	0.0134	−0.16	C9H8O3	[M-H]-	−6.6	4-Coumarate	MS^2	ID3
149.0245	neg	0.0163	−0.21	C8H8O4	[M-H2O-H]-	0.6	3,4-Dihydroxymandelaldehyde	MS^2	ID3
179.036	neg	0.0214	−0.22	C9H8O4	[M-H]-	5.7	4-Hydroxyphenylpyruvate, 4-Hydroxy-enol-phenylpyruvate	MS^1	ID4
215.052	neg	0.0268	−0.23	C9H10O5	[M-H2O-H]-	5.7	3-Methoxy-4-hydroxymandelate	MS^1	ID4
199.025	neg	0.0225	−0.22	C8H8O6	[M-H]-	0.9	4-Maleylacetoacetate, 4-Fumarylacetoacetate	MS^1	ID4
197.046	neg	0.0268	−0.23	C9H10O5	[M-H]-	2.3	3-Methoxy-4-hydroxymandelate†	MS^1	ID4
167.0345	neg	0.0271	−0.19	C8H8O4	[M-H]-	−2.9	Homogentisate, 3,4-Dihydroxymandelaldehyde, 3,4-Dihydroxyphenylacetate	MS^1	ID4
2-Oxocarboxylic acid metabolism									
169.05	neg	0.0006	−0.27	C8H12O5	[M-H2O-H]-	−3.7	2-Oxosuberate (2-oxooctanedionic acid)	MS^2	ID3
199.058	pos	0.0010	−0.46	C9H12O6	[M-H2O + H]+	−10.5	cis-(Homo)3-aconitate	MS^2	ID3
159.0645	pos	0.0015	−0.40	C7H12O5	[M-H2O + H]+	−4.3	3-Isopropylmalate	MS^2	ID3
161.0435	pos	0.0032	−0.28	C6H8O5	[M + H]+	−5.9	2-Oxoadipate (2-oxohexanedioic acid)†	MS^2	ID3
162.0755	pos	0.0070	−0.25	C6H11NO4	[M + H]+	−3.6	2-Aminoadipate (2-aminohexanedioic acid)	MS^2	ID3
148.06	pos	0.0077	−0.26	C5H9NO4	[M + H]+	−2.9	Glutamate	MS^2	ID3
146.0545	neg	0.0404	−0.31	C5(13C)H10O4	[M(C13)-H]-	7.3	2-Aceto-2-hydroxybutanoate	MS^1	ID4

(Continued on following page)

10.3389/fmolb.2023.1154536

TABLE 2 (*Continued*) Pathway related metabolites elevated in the allergic asthma cohort.

m/z	Charge	Adj. *p*-value	Log-fold-change	Molecular formula	Ionisation	Δm (ppm)	Compound	Ann.	ID level
Fatty acid metabolites									
117.019	neg	0.0004	−0.37	C4H6O4	[M-H]-	−2.8	Butanedioic acid (succinate)†	MS^2, Lit.	ID1
131.035	neg	0.0002	−0.39	C5H8O4	[M-H]-	0.1	Pentanedioic acid (glutarate)†	MS^2, Lit.	ID1
147.0645	pos	0.0070	−0.40	C6H10O4	[M + H]+	−4.7	Hexanedioic acid (adipic acid)	MS^2	ID3
131.033	pos	0.0031	−0.46	C5H6O4	[M + H]+	−6.8	Pentenedioic acid (glutaconate)	MS^2	ID3
143.0345	neg	0.0363	−0.23	C6H8O4	[M-H]-	−3.7	Hexenedioic acid	Lit.	ID4
157.0505	neg	0.0230	−0.26	C7H10O4	[M-H]-	−0.8	Heptenedioic acid†	Lit.	ID4
97.029	neg	0.0137	−0.20	C5H8O3	[M-H2O-H]-	−5.2	ω-Oxopentanoaic acid (glutarate semialdehyde)	MS^2	ID3
125.06	neg	0.0195	−0.23	C7H12O3	[M-H2O-H]-	−6.4	ω-Oxoheptanoic acid	MS^2	ID3
113.024	neg	0.0424	−0.15	C5H8O4	[M-H]-	−3.7	ω-Oxopentenoic acid	Lit.	ID4
155.071	neg	0.0268	−0.20	C8H12O3	[M-H]-	−2.1	ω-Oxooctenoic acid	Lit.	ID4
167.071	neg	0.0264	−0.18	C9H12O3	[M-H]-	−1.9	ω-Oxononadienoic acid	Lit.	ID4
181.086	neg	4.83E-05	−0.48	C10H14O3	[M-H]-	−5.6	ω-Oxodecadienoic acid	Lit.	ID4
87.0445	neg	0.0471	−0.58	C4H8O2	[M-H]-	−7.5	Butanoic acid†	Lit.	ID2
101.0605	neg	0.0214	−0.59	C5H10O2	[M-H]-	−0.2	Pentanoic acid†	Lit.	ID2
197.081	neg	0.0004	−0.46	C10H14O4	[M-H]-	−4.6	2,7-Dimethyl-2,4-octadienedioic acid	MS^2	ID3
129.019	neg	0.0008	−0.25	C5H8O5	[M-H2O-H]-	−2.6	2-Hydroxypentanedioc acid (2-hydroxyglutarate)	MS^2	ID3
178.0355	neg	0.0129	−0.18	C5H9NO6	[M-H]-	−1.2	2-Amino-3,4-dihydroxypentanedioic acid	MS^2	ID3
161.0435	pos	0.0032	−0.28	C6H8O5	[M + H]+	−5.9	2-Oxohexanedioic acid (2-oxoadipate) †	MS^2	ID3
162.0755	pos	0.0070	−0.25	C6H11NO4	[M + H]+	−3.6	2-Aminohexanedioic acid (2-aminoadipate)	MS^2	ID3
133.05	neg	0.0104	−0.39	C5H10O4	[M-H]-	−4.8	2,3-Dihydroxypentanoic acid	MS^2	ID3
Monosaccharides and metabolites									
163.024	neg	0.0002	−0.20	C5H8O6	[M-H]-	−5.0	2-Dehydro-xylonate	MS^2	ID3
151.0585	pos	0.0013	−0.51	C5H10O5	[M + H]+	−10.6	Arabinose	MS^2	ID3
163.0595	pos	0.0158	−0.33	C6H12O6	[M-H2O + H]+	−3.7	Galactose†	MS^2	ID3
193.035	neg	0.0196	−0.24	C6H10O7	[M-H]-	−2.0	Glucuronate	MS^2	ID3
209.03	neg	0.0244	−0.20	C6H10O8	[M-H]-	−1.4	Glucarate	MS^1	ID4
91.04	neg	0.0261	−0.79	C3H8O3	[M-H]-	−0.7	Glycerol	MS^1	ID4
119.0345	neg	0.0319	−0.39	C4H8O4	[M-H]-	−4.1	Erythrulose	MS^1	ID4

Putatively identified compounds elevated in the allergic asthma cohort, grouped by metabolic pathways or chemical families and ordered by their adjusted *p*-value. Exception: fatty acid metabolites are sorted based on their chemical relation. Log-fold-change: negative values indicate higher average expression in the asthmatic group. Log-fold-change was calculated using R-package "limma" (Ritchie et al., 2015); see Supplementary Section S3 for more details. The listed *m/z* values represent the measured values and the mass error (Δm) to the theoretical mass is reported in ppm. Annotation (Ann.) e.g., based on literature (Lit.), references for literature-based identification are included in Supplementary Table S2. †: compounds that were detected several times in different ionisation forms (listed in Supplementary Table S2). MS^1: assignment based on full scan mode by literature match or pathway analysis, MS^2: assignment based on real-time tandem mass spectrometry spectra, ID: identification confidence level ranging from ID1 (high) to ID5 (low).

Nevertheless, we decided to assess whether atopy by itself has an impact on the breath profiles by isolating the group of healthy samples and repeating our analysis pipeline to find differences between the healthy children with sensitization and the ones without. We found no significantly different features between the two groups (see Supplementary Figure S5).

TABLE 3 Pathway related metabolites downregulated in the allergic asthma cohort.

m/z	Charge	Adj. *p*-value	Log-fold-change	Molecular formula	Ionisation	Δm (ppm)	Compound	Ann.	ID level
Arginine and proline metabolism									
104.07	pos	0.0002	0.41	C4H9NO2	[M + H]+	−5.8	4-Aminobutanoate†	MS², Lit.	ID3
60.0805	pos	0.0083	0.80	C4H9NO	[M-CO + H]+	−4.7	4-Aminobutanal	MS¹	ID4
				C4H9NO2	[M-CO2+H]+	−4.7	4-Aminobutanoate	MS¹	ID4
116.07	pos	0.0122	0.42	C5H9NO2	[M + H]+	−5.2	Proline	MS¹, Lit.	ID4
				C5H12N2O2	[M-NH3+H]+	−5.2	Ornithine	MS¹	ID4
193.13	pos	0.0251	0.53	C6H14N4O2	[M + H2O + H]+	2.5	Arginine	MS¹	ID4
118.086	pos	0.0252	0.39	C6H11NO3	[M-CO + H]+	−2.2	4-Acetamidobutanoate	MS¹	ID4
114.0545	pos	0.0264	0.38	C5H7NO2	[M + H]+	−4.0	1-Pyrroline-2-carboxylate†	MS¹	ID4
				C5H9NO3	[M-H2O + H]+	−4.0	Hydroxyproline, Glutamate 5-semialdehyde	MS¹	ID4
112.075	pos	0.0313	0.31	C6H11NO2	[M-H2O + H]+	−6.2	N4-Acetylaminobutanal	MS¹	ID4
102.0545	pos	0.0381	0.31	C5H9NO4	[M-HCOOH + H]+	−4.5	4-Hydroxyglutamate semialdehyde	MS¹	ID4
				C5H7NO3	[M-CO + H]+	−4.5	1-Pyrroline-3-hydroxy-5-carboxylate†	MS¹	ID4
Arginine biosynthesis									
61.039	pos	0.0076	0.36	CH4N2O	[M + H]+	−10.5	Urea	MS²	ID3
96.9925	neg	0.0113	0.25	C4H4O4	[M-H2O-H]-	−6.4	Fumarate	MS²	ID3
				C4H4O4	[M-H2O-H]-	−6.4	Maleate	MS²	ID3
116.07	pos	0.0122	0.42	C5H9NO2	[M + H]+	−5.2	Proline	MS¹, Lit.	ID4
				C5H12N2O2	[M-NH3+H]+	−5.2	Ornithine	MS¹	ID4
193.13	pos	0.0251	0.53	C6H14N4O2	[M + H2O + H]+	2.5	Arginine	MS¹	ID4
102.0545	pos	0.0381	0.31	C5H9NO4	[M-HCOOH + H]+	−4.5	4-Hydroxyglutamate semialdehyde	MS¹	ID4
				C5H7NO3	[M-CO + H]+	−4.5	1-Pyrroline-3-hydroxy-5-carboxylate	MS¹	ID4
Linoleic acid metabolism									
281.2475	pos	7.12E-06	0.79	C18H32O2	[M + H]+	−0.04	Linoleate†	MS²	ID3
295.225	pos	2.09E-04	0.62	C18H32O4	[M-H2O + H]+	−6.0	13(S)-HPODE†	MS¹	ID4
297.242	pos	4.73E-04	0.73	C18H32O3	[M + H]+	−1.4	13(S)-HODE†, 12 (13)-EpOME†, 9 (10)-EpOME†	MS¹	ID4
Aldehydes									
115.075	pos	0.0491	0.49	C6H10O2	[M + H]+	−3.1	4-Hydroxy-2-hexenal†	Lit.	ID2
146.117	pos	0.0210	0.30	C7H12O2	[M + NH4]+	−3.8	4-Hydroxy-2-heptenal	Lit.	ID4
143.106	pos	0.0179	0.41	C8H14O2	[M + H]+	−4.6	4-Hydroxy-2-octenal†	Lit.	ID4
258.243	pos	0.0225	0.34	C15H28O2	[M + NH4]+	1.0	4-Hydroxy-2-pentadecenal	Lit.	ID4
158.1175	pos	0.0082	0.39	C8H12O2	[M + NH4]+	−0.4	4-Hydroxy-2,6-octadienal	Lit.	ID4
172.133	pos	0.0004	0.57	C9H14O2	[M + NH4]+	−1.2	4-Hydroxy-2,6-nonadienal	Lit.	ID2
228.196	pos	0.0238	0.44	C13H22O2	[M + NH4]+	0.9	4-Hydroxy-2,6-tridecadienal	Lit.	ID2
283.191	neg	0.0292	0.46	C15H26O2	[M + HCOO]-	−1.7	4-Hydroxy-2,6-pentadecadienal	Lit.	ID4
253.2155	pos	0.0008	0.49	C16H28O2	[M + H]+	−2.8	4-Hydroxy-2,6-hexadecadienal	Lit.	ID4

(Continued on following page)

TABLE 3 (*Continued*) Pathway related metabolites downregulated in the allergic asthma cohort.

m/z	Charge	Adj. *p*-value	Log-fold-change	Molecular formula	Ionisation	Δm (ppm)	Compound	Ann.	ID level
Fatty amides									
200.201	pos	0.0008	0.63	C12H25NO	[M + H]+	0.5	Dodecanamide	MS^2	ID3
256.263	pos	0.0008	0.80	C16H33NO	[M + H]+	−1.9	Hexadecanamide	MS^2	ID3
302.305	pos	0.0093	0.90	C18H37NO	[M + H2O + H]+	−1.2	Octadecanamide	MS^2	ID3
288.253	pos	0.0003	0.74	C16H33NO3	[M + H]+	−1.1	N,N-bis(2-hydroxyethyl) dodecanamide	MS^2	ID3
316.2845	pos	0.0001	1.04	C18H35NO2	[M + H2O + H]+	−0.4	Palmitoleoylethanolaimde	MS^2	ID3
318.3	pos	0.0001	1.13	C18H37NO2	[M + H2O + H]+	−0.8	Palmitoylethanolamide	MS^2	ID3
Fatty acids									
271.2265	pos	0.0006	0.77	C16H32O4	[M-H2O + H]+	−1.0	10,16-Dihydroxyhexadecanoic acid	MS^2	ID3
220.1905	pos	0.0041	0.37	C11H23NO2	[M + H2O + H]+	−1.0	11-Aminoundecanoic acid	MS^2	ID3
151.096	pos	0.0159	0.29	C6H12O3	[M + H2O + H]+	−3.2	6-Hydroxyhexanoic acid	MS^1	ID4

Putatively identified compounds downregulated on the allergic asthma cohort grouped by metabolic pathways or chemical families and ordered by their adjusted *p*-value. Exception: aldehydes and fatty amides are sorted based on their chemical relation. Log-fold-change: positive values indicate higher average expression in the healthy group. Log-fold-change was calculated using R-packege "limma" (Ritchie et al., 2015); see supplementary material section S3 for more details. The listed *m/z* values represent the measured values and the mass error (Δm) to the theoretical mass is reported in ppm Annotation (Ann.) e.g., based on literature (Lit.), references for literature-based identification are included in Supplementary Table S2. †: compounds that were detected several times in different ionisation forms (listed in Supplementary Table S2). MS^1: assignment based on full scan mode by literature match or pathway analysis, MS^2: assignment based on real-time tandem mass spectrometry spectra, ID: identification confidence level ranging from ID1 (high) to ID5 (low). 12(13)-EpOME: 12,13-Epoxyoctadec-9(Z)-enoic acid; 9 (10)-EpOME: 9,10-Epoxyoctadec-12(Z)-enoic acid; 13(S)-HPODE: 13(S)-Hydroperoxy-9Z, 11E-octadecadienoic acid; 13(S)-HODE: 13(S)-Hydroxy-9Z, 11E-octadeca-dienoic acid.

4 Discussion

We present the first online breath analysis study performed by SESI/HRMS on a pediatric population with allergic asthma. The study revealed group-specific breath patterns with a large number of discriminative *m/z* features, many of which were putatively identified and could be grouped to metabolic pathways or chemical families. Moreover, some of the relevant compounds and pathways were previously published in metabolomic studies in pediatric asthma (Neerincx et al., 2017; Ferraro et al., 2018; Papamichael et al., 2021) or reported in SESI/HRMS studies (see Supplementary Table S2).

As described by Papamichael et al., an altered energy metabolism is expected in children with asthma due to the hypoxic environment, bronchoconstriction, and other associated changes as well as increased efforts for breathing (Papamichael et al., 2021). However, this explanation might not apply to the included asthmatic group of our study, as they did not suffer from acute exacerbations. The lung and gut microbiomes are also potential contributors to the pathophysiology of asthma (Barcik et al., 2020). Several previous breath analysis studies identified compounds and molecular pathways associated with pediatric asthma. However, some of the potential biomarkers were of exogenous origin and only a handful of them were consistently detected in more than one study (Neerincx et al., 2017; Ferraro et al., 2018; Papamichael et al., 2021). The pathways and chemical families identified in our study are biologically relevant and reflect both an altered state of energy metabolism as well as changes in products from the microbiome.

The metabolism of lysine was the most significantly elevated pathway in asthma and all associated compounds were identified based on direct MS^2 spectra. Two different degradation pathways of lysine were found, one is taking place in humans and the other in the gut microbiota (Figure 2). The associated metabolites succinate and glutarate were unambiguously identified (ID1, see Table 2), and have been reported as associated with pediatric asthma in previous metabolomic studies in blood (Chang et al., 2015), urine (Saude et al., 2011), and breath (Carraro et al., 2018). Carraro et al. also reported a decreased level of oxoadipate in early asthma, which is not in line with our findings but could be explained by the different study design focusing on wheezing in preschool children (Carraro et al., 2018). However, a study linked an enzymatic complex involved in the lysine degradation pathway to the formation of reactive oxygen species from 2-oxoadipate (Jordan et al., 2019), which could potentially be a link to asthma pathophysiology.

Tyrosine metabolism was also significantly upregulated in the allergic asthmatic group. As illustrated in Figure 2, some of the metabolites belong to the main human degradation pathway whereas other tyrosine-derived metabolites are of human or microbiotic origin. An increased level of tyrosine in asthmatic children was reported in previous metabolomics studies (Saude et al., 2011; Papamichael et al., 2019; Tao et al., 2019). Additionally, the bacterial tyrosine metabolite 4-hydroxyphenylacetate was reported to be negatively correlated with the FEV1 in urine (Papamichael et al., 2019). It is hypothesized that high levels of tyrosine metabolism might be related to inflammation and oxidative stress in asthma (Papamichael et al., 2021). Also, tyrosine-derived catecholamines are important during conditions of stress and play a role in the regulation of the immune system (Barnes et al., 2015). In contrast to these findings, Carraro et al. reported a lower level of some tyrosine metabolites in children with early asthma compared to transient wheezers (Carraro et al., 2018).

The largest elevated group consisted of 20 fatty acid metabolites, including saturated and unsaturated dicarboxylic acids, ω-oxo-acids,

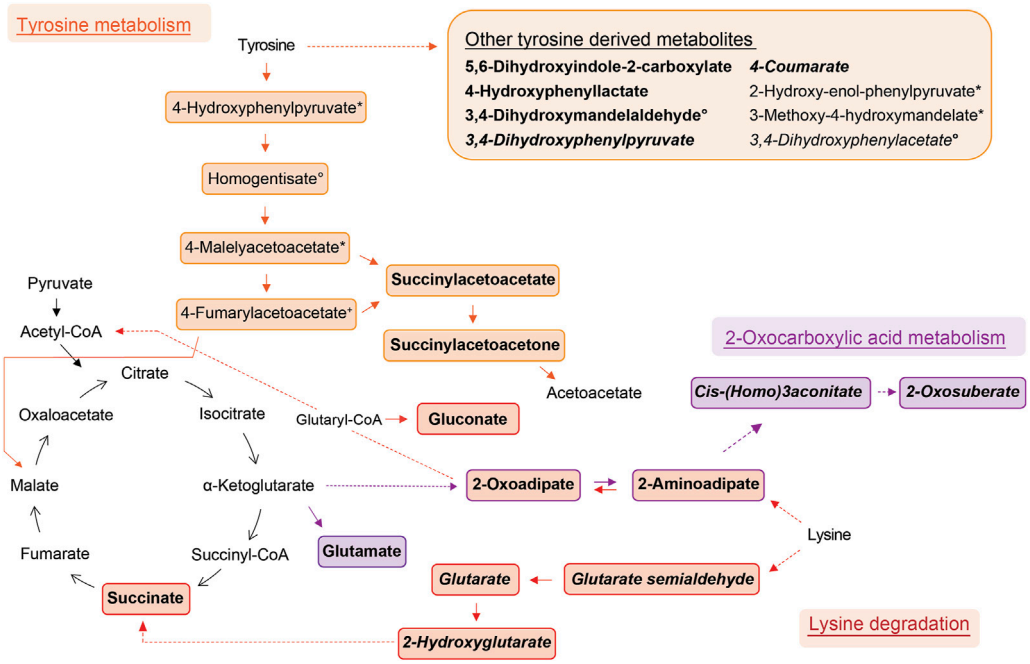


FIGURE 2
Schemes of metabolic pathways well-represented by compounds that were elevated in the allergic asthma group and putatively identified. Tyrosine derived metabolites besides the main degradation pathway in humans are summarised in the box. Two unrelated compounds of the 2-oxocarboxylic acid metabolism are not shown (see Table 2). Solid lines: direct metabolic relations; dashed lines: indirect metabolic relations (metabolites in between were not identified); colored: putatively identified compound; bold: identified by MS^2, regular: identified based on exact mass and pathway mapping, or on literature; italic: metabolites from gut microbiota; *, °, +, #, ^, ‘: several possibilities for 1 *m/z* feature based on exact mass and pathway mapping.

hydroxy-acids, and alkanoic acids. All lysine metabolites are additionally fitting into this chemical family. A large part of these identified fatty acids were previously reported being decreased in chronic obstructive pulmonary disease exacerbations by SESI/HRMS and described as metabolites of the *ω*-oxidation, a minor pathway of the fatty acid oxidation (Gaugg et al., 2017a; 2019). Interestingly, important molecules in asthma pathophysiology including arachidonic acid, leukotrienes, and prostaglandins, although not detected in this study, are also common substrates of the cytochrome P450 *ω*-hydroxylases (Ni and Liu, 2021). Therefore, our findings support the hypothesis that *ω*-oxidation might be upregulated in allergic asthma. Butanoic and pentanoic acid are both identified (ID2, Supplementary Table S2) and were both reported to distinguish asthmatic from healthy children in previous studies (Dallinga et al., 2010; van Mastrigt et al., 2016; van Vliet et al., 2016). Furthermore, the findings from previous metabolomic studies of an altered fatty acid metabolism in asthma is supported by our data (Neerincx et al., 2017; Ferraro et al., 2018; Papamichael et al., 2021).

Further, the 2-oxocarboxylic acid metabolism was also elevated in the allergic asthmatic (Figure 2). 2-Oxoadipate and 2-aminoadipate are overlapping with the lysine degradation pathway and glutamate is a common metabolite involved in multiple metabolic pathways. The metabolism of 2-oxocarboxylic acids is solely happening in archaea, which are also represented in the gut microbiota. A review linked methanogenic archaea as potential important contributors to atopic diseases (Sereme et al., 2019). Additionally, 2-oxoadipate was previously detected as an exhaled metabolite from the gut microbiota in a mice model study (Lan et al., 2022).

Lastly, monosaccharides and derived metabolites were increased in allergic asthma. This difference in carbohydrate metabolism of asthmatic children is expected due to an altered energy demand and metabolism (Papamichael et al., 2021).

A recently published study comparing children with acute asthma exacerbations and healthy controls reported similar results to ours. Despite investigating urine by high-performance liquid chromatography mass spectrometry, they also reported an elevated level of tyrosine metabolism including gentisate and increased glucuronate as well as a downregulated linoleic acid metabolite and palmitic acid in children with acute asthma (Li et al., 2022).

The most prominent group of downregulated metabolites was associated with arginine and proline metabolism as well as arginine biosynthesis. Arginase, an enzyme that converts arginine into ornithine and urea is an important contributor to asthma pathophysiology (Maarsingh et al., 2011). According to Maarsingh et al., increased expression and activity of arginase in asthma mouse models resulted in the promotion of inflammatory processes, decreased arginine and increased ornithine and proline levels (Maarsingh et al., 2011). This is not completely in line with our findings where these downstream pathways are decreased in allergic asthma. However, a study about amino acids in blood serum of asthmatics reported decreased arginine, proline and ornithine levels (Morris et al., 2004).

Further, the linoleic acid metabolism was well-represented amongst the diminished compounds. While conjugated linoleic acid was consistently reported as having anti-inflammatory properties, the effect of linoleic acid especially on asthma is in dispute due to controversial observations in clinical trials

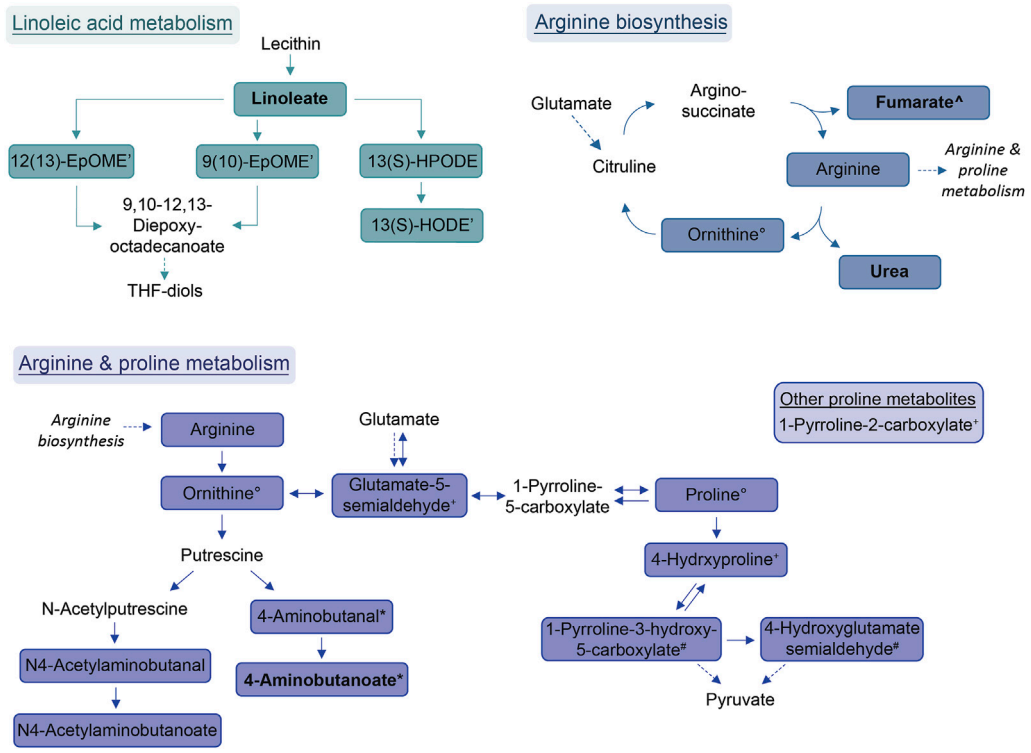


FIGURE 3
Schemes of metabolic pathways well-represented by compounds that were decreased in the allergic asthma group and putatively identified. Proline metabolism: One component is not directly connected to the displayed pathway and is shown in the box. Solid lines: direct metabolic relations; dashed lines: indirect metabolic relations; colored: putatively identified compound; bold: identified by MS^2; regular: identified based on exact mass and pathway mapping, or on literature; *, °, $^+$: several possibilities for 1 *m/z* feature based on exact mass and pathway mapping.

(Wendell et al., 2014). Interestingly, a more recent study found genetically predicted linoleic acid to be associated with a lower risk for asthma, which is in line with our results (Zhao and Schooling, 2019).

Within the group of amides, palmitoylethanolamide (PEA) was found to be decreased in the allergic asthma group. This is in line with the well-studied anti-inflammatory effect of PEA (Clayton et al., 2021). More recently, also an inhibitory effect for the development of allergic airway symptoms was reported for PEA in mice (Roviezzo et al., 2017).

Another group of downregulated compounds was assigned to aldehydes. Aldehydes are indicative of oxidative stress and originate from lipid peroxidation (Jesenak et al., 2017), which is involved in asthma pathophysiology, and are thus expected to be increased in asthmatics. However, the results in literature on breath analysis in asthma are not consistent: Some studies observed an increased level of certain aldehydes in the asthmatic group (Gahleitner et al., 2013; van de Kant et al., 2013; Smolinska et al., 2014), whereas others reported unaltered or even decreased levels (Ibrahim et al., 2011; Sagdic et al., 2011; Caldeira et al., 2012; Riscassi et al., 2022). Beyond this, aldehydes are also used as common additives in cosmetics or food and are known environmental contaminants (Sinharoy et al., 2019), which could influence their exhaled concentrations. Furthermore, the annotation of aldehydes in our study was based solely on exact mass matches with previously published compounds by SESI/HRMS (see Supplementary Table S2).

Altogether, many of the enriched pathways that we reported either elevated or decreased in allergic asthma could be linked to previous findings of metabolomic studies using various methods for blood, urine, or breath analysis. This strengthens the putative compound identification performed in this work and supports the possible biological and diagnostic value of these metabolites.

Assessing the predictability of the disease with supervised machine learning in a 10 times repeated 10-fold cross-validation revealed an AUC of 0.83 (CI: 0.73–0.92), indicating that the metabolic profiles could be applied for potential diagnostic purposes. Some compounds that were allocated to subgroups of metabolic pathways or chemical families were frequently selected during cross-validation (Figure 4C; Supplementary Table S4) suggesting that a smaller group of compounds might not only be pathophysiologically relevant, but also has potential for diagnostic models. The two dicarboxylic acids and lysine metabolites, succinate and glutarate, are promising candidates and were unambiguously identified. Nevertheless, while efforts have been taken to prevent bias by preprocessing data in each cross-validation loop and reducing the dimensionality of the feature set for training the classifier with machine learning, the risk of overfitting cannot be completely ruled out (Vabalas et al., 2019). An independent and increased study cohort would be needed to help in validating the model performance and the selected predictors (Fijten et al., 2017).

Due to a rather large number of significant *m/z* features, a main focus was set on putative compound identification. We aimed at establishing an objective workflow that is based on matching direct MS^2 spectra with database fragment spectra, adapted from previous work (Kaeslin et al., 2021), refined for a more extensive screening of suggested compounds, and expanded by pathway enrichment

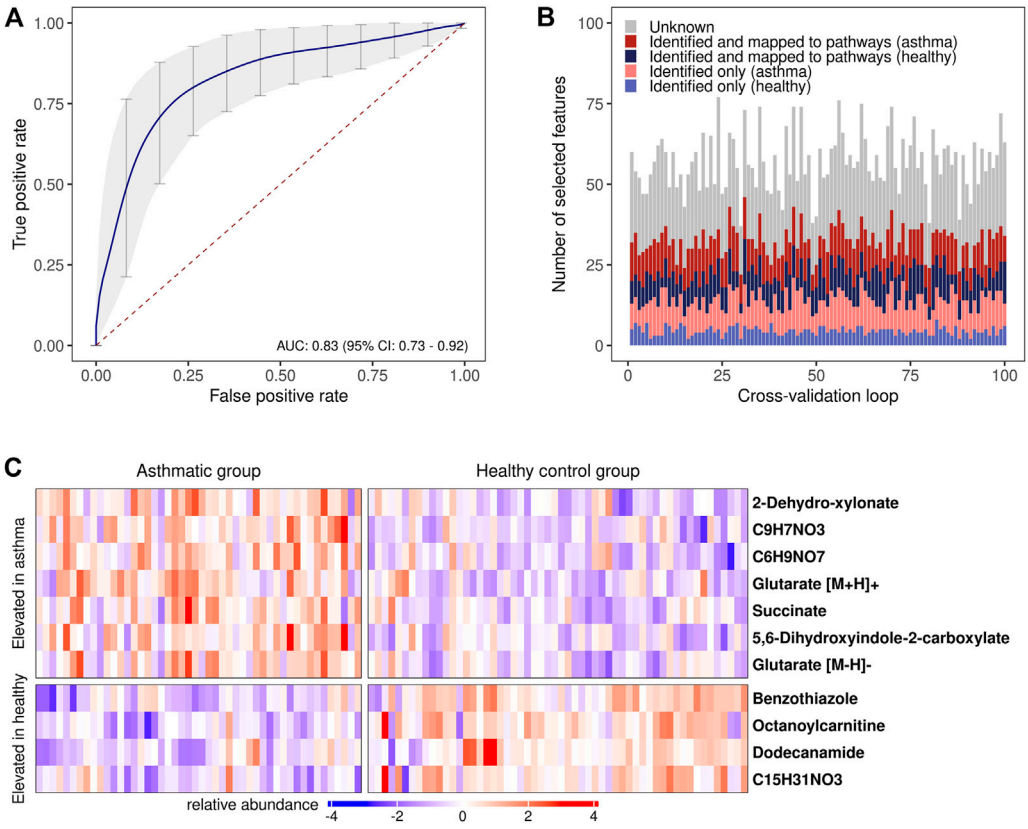


FIGURE 4
Disease prediction based on breath profiles. **(A)** Average receiver operating characteristic curve (ROC) with an average AUC of 0.83 resulting from the 10 times repeated 10-fold cross-validation. ROC curves resulting from predictions on each of the left-out data sets in the cross-validation were used to calculate the average ROC curve (Fawcett, 2006) (vertical averaging). Vertical grey bars: pointwise confidence intervals computed using bootstrapping (10.000 repetitions); red dashed line: line of no discrimination. **(B)** Stacked bar plots of the selected features in each cross-validation iteration. Red/blue color scheme: upregulated features in the allergic asthmatic/healthy group. **(C)** Heat map of the most frequently chosen features (standardized intensities) as predictors in the cross-validation. Columns: study participants; rows: *m/z* features with putatively identified compounds (right), a chemical formula is provided if compound identification was not possible.

analysis to strengthen the feature annotation. A limitation in our identification approach is the lacking chromatographic separation in SESI/HRMS, which hinders the distinction of isomeric compounds. Further, with a minimum isolation window of 0.7 Da, co-fragmentation of several compounds with similar masses can occur, which complicates the annotation of fragment spectra. To address this, we excluded several MS^2 spectra with insufficient quality from further analysis, as specified in Supplementary Table S1. The confidence of identification is only moderate for most compounds, as they are putatively annotated based on fragment spectra analysis with the SIRIUS software that uses computational power to determine chemical structures that potentially have a matching fragmentation pattern (Dührkop et al., 2019) or the exact mass comparison to the literature and/or known metabolites. Therefore, there is a general risk for misclassification, and the confirmation of the unambiguous chemical structures requires further time-consuming experiments, including the measurement of standards. Especially, the molecules that were annotated based on their exact mass or solely detected as adduct or loss species without their primary ion being amongst the significant features need further investigation. However, our aim for this study was to get a broad overview over the potentially involved metabolic pathways rather than accurately identifying a small set of single compounds.

A strength of this study design is that all enrolled patients were taken off long-term therapy at least 1 week prior to the study visit and did not take any short-acting relievers on the day of measurement. While direct breath analysis by SESI bypasses any contamination during sample preparation, this adds up to also diminish confounders and signal interferences from medications in exhaled breath. This is an important aspect, as it was previously shown that the methodology can detect drugs, including the asthma medication Salbutamol, in breath (Gaugg et al., 2017b; Chen et al., 2021; Singh et al., 2021).

While asthma is a heterogeneous disease with different phenotypes, this study focused only on allergic asthma, the most frequent phenotype in children. Therefore, our findings cannot be extrapolated to all forms of pediatric asthma. We included all sensitized healthy controls and all asthmatics with allergic comorbidities such as allergic eczema or hay fever in order to represent the real population for future applications. 21.4% of the healthy cohort had an allergic sensitization to at least one common aeroallergen, which is in line with the estimated prevalence in children (Kölli et al., 2022). However, performing subgroup analysis on the

entire data set was not feasible since surrogate variable analysis is known to adjust for confounding influences and reduce heterogeneity in the data (e.g., demographic variations like age and sex, or disease heterogeneity) (Jaffe et al., 2015). We decided to apply SVA to adjust for unmodeled factors, since the study was conducted over a period of 15 months on a highly sensitive instrument. Therefore, despite a strict adherence to standard operating procedures, we had to assume that apart from demographic variation also unknown environmental or technical confounders might have impacted the *m/z* feature intensities. In order to nevertheless assess the interesting question whether the described markers and pathways might also be related to atopy by itself, we chose to perform an independent subgroup analysis in the healthy cohort. No significant features that could distinguish healthy children with allergic sensitization from the ones without could be found. Therefore, the identified metabolites and pathways represent promising candidate biomarkers for allergic asthma that need to be validated in a larger and independent study cohort.

This study confirms the applicability of SESI/HRMS to a pediatric population and shows its potential to distinguish children with allergic asthma from healthy controls based on their breath signatures. Moreover, well-represented metabolic pathways that are potentially linked to the pathophysiology of allergic asthma in children could be identified. A smaller subset of the differentiating compounds could possibly be used for predictive modelling. These findings might set the path for much-needed, non-invasive clinical applications to improve early diagnosis of asthma.

Data availability statement

TThe MS^2 data generated for this study can be found in the FigShare repository at doi: 10.6084/m9.figshare.21946877. Due to ethical restrictions, other data is available upon request from the authors.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics committe of the canton of Zurich (Kantonale Ethikkommission Kanton Zürich). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

Study design and concept: AM; Data acquisition: BS, RW, LW, and DI; Data processing and analysis: SM, BS, and RW; Data evaluation and interpretation: AM, SM, BS, and RW; Drafting of the manuscript: AM, SM, BS, and RW; Funding acquisition: AM, RZ, and MK; Resources: AM, RZ, MK, and NP; Review and editing of the manuscript: all authors.

Funding

This work was supported by the Swiss National Science Foundation (SNSF) [grant number 326030_177101/1]; the Evi Diethelm-Winteler Foundation; the Childhood Research Center of the University Children's Hospital Zurich; the Heidi Ras Stiftung; the Zurich Foundation; and the Lotte and Adolf Hotz-Sprenger Stiftung.

Acknowledgments

We are very grateful to all the participants for participating in this study. We further like to thank the lung function team at the Children's University Hospital Zurich for their support and Jérôme Kaeslin, ETH Zurich, for the helpful discussions on compound identification. This work is part of Zurich Exhalomics, a flagship project of "Hochschulmedizin Zürich".

Conflict of interest

MK is a founder and board member of Deep Breath Intelligence AG (www.dbi.ch), a company that provides services in the field of breath analysis. RZ is advisor to Deep Breath Intelligence AG.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1154536/full#supplementary-material

References

Barcik, W., Boutin, R. C. T., Sokolowska, M., and Finlay, B. B. (2020). The role of lung and gut microbiota in the pathology of asthma. *Immunity* 52, 241–255. doi:10.1016/j.immuni.2020.01.007

Barnes, M. A., Carson, M. J., and Nair, M. G. (2015). Non-traditional cytokines: How catecholamines and adipokines influence macrophages in immunity, metabolism and the central nervous system. *Cytokine* 72, 210–219. doi:10.1016/j.cyto.2015.01.008

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bruderer, T., Gaisl, T., Gaugg, M. T., Nowak, N., Streckenbach, B., Müller, S., et al. (2019). On-line analysis of exhaled breath *focus review*. *Chem. Rev.* 119, 10803–10828. doi:10.1021/acs.chemrev.9b00005

Caldeira, M., Perestrelo, R., Barros, A. S., Bilelo, M. J., Morête, A., Câmara, J. S., et al. (2012). Allergic asthma exhaled breath metabolome: A challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* 1254, 87–97. doi:10.1016/j.chroma.2012.07.023

Carraro, S., Bozzetto, S., Giordano, G., el Mazloum, D., Stocchero, M., Pirillo, P., et al. (2018). Wheezing preschool children with early-onset asthma reveal a specific metabolomic profile. *Pediatr. Allergy Immunol.* 29, 375–382. doi:10.1111/pai.12879

Chang, C., Guo, Z., He, B., and Yao, W. (2015). Metabolic alterations in the sera of Chinese patients with mild persistent asthma: A GC-MS-based metabolomics analysis. *Acta. Pharmacol. Sin.* 36, 1356–1366. doi:10.1038/aps.2015.102

Chen, X., Zhang, K., Yin, Z., Fang, M., Pu, W., Liu, Z., et al. (2021). Online real-time monitoring of exhaled breath particles reveals unnoticed transport of nonvolatile drugs from blood to breath. *Anal. Chem.* 93, 5005–5008. doi:10.1021/acs.analchem.1c00509

Clayton, P., Hill, M., Bogoda, N., Subah, S., and Venkatesh, R. (2021). Palmitoylethanolamide: A natural compound for health management. *Int. J. Mol. Sci.* 22, 5305. doi:10.3390/ijms22105305

Dallinga, J. W., Robroeks, C. M. H. H. T., van Berkel, J. J. B. N., Moonen, E. J. C., Godschalk, R. W. L., Jöbsis, Q., et al. (2010). Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children. *Clin. Exp. Allergy* 40, 68–76. doi:10.1111/j.1365-2222.2009.03343.x

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A., Meusel, M., et al. (2019). Sirius 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods.* 16, 299–302. doi:10.1038/s41592-019-0344-8

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern. Recognit. Lett.* 27, 861–874. doi:10.1016/j.patrec.2005.10.010

Ferraro, V., Carraro, S., Bozzetto, S., Zanconato, S., and Baraldi, E. (2018). Exhaled biomarkers in childhood asthma: Old and new approaches. *Asthma. Res. Pract.* 4, 9. doi:10.1186/s40733-018-0045-6

Fijten, R. R. R., Smolinska, A., Drent, M., Dallinga, J. W., Mostard, R., Pachen, D. M., et al. (2017). The necessity of external validation in exhaled breath research: A case study of sarcoidosis. *J. Breath. Res.* 12, 016004. doi:10.1088/1752-7163/aa8409

Gahleitner, F., Guallar-Hoyas, C., Beardsmore, C. S., Pandya, H. C., and Thomas, C. L. P. (2013). Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis* 5, 2239–2247. doi:10.4155/bio.13.184

Gaillard, E. A., Kuehni, C. E., Turner, S., Goutaki, M., Holden, K. A., de Jong, C. C. M., et al. (2021). European Respiratory Society clinical practice guidelines for the diagnosis of asthma in children aged 5–16 years. *Eur. Respir. J.* 58, 2004173. doi:10.1183/13993003.04173-2020

Gaugg, M. T., Bruderer, T., Nowak, N., Eiffert, L., Martinez-Lozano Sinues, P., Kohler, M., et al. (2017a). Mass-spectrometric detection of omega-oxidation products of aliphatic fatty acids in exhaled breath. *Anal. Chem.* 89, 10329–10334. doi:10.1021/acs.analchem.7b02092

Gaugg, M. T., Engler, A., Nussbaumer-Ochsner, Y., Bregy, L., Stöberl, A. S., Gaisl, T., et al. (2017b). Metabolic effects of inhaled salbutamol determined by exhaled breath analysis. *J. Breath. Res.* 11, 046004. doi:10.1088/1752-7163/aa7caa

Gaugg, M. T., Gomez, D. G., Barrios-Collado, C., Vidal-de-Miguel, G., Kohler, M., Zenobi, R., et al. (2016). Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry—A pilot study on tobacco smokers. *J. Breath. Res.* 10, 016010. doi:10.1088/1752-7155/10/1/016010

Gaugg, M. T., Nussbaumer-Ochsner, Y., Bregy, L., Engler, A., Stebler, N., Gaisl, T., et al. (2019). Real-time breath analysis reveals specific metabolic signatures of COPD exacerbations. *Chest* 156, 269–276. doi:10.1016/j.chest.2018.12.023

Ibrahim, B., Basanta, M., Cadden, P., Singh, D., Douce, D., Woodcock, A., et al. (2011). Non-invasive phenotyping using exhaled volatile organic compounds in asthma. *Thorax* 66, 804–809. doi:10.1136/thx.2010.156695

Jaffe, A. E., Hyde, T., Kleinman, J., Weinbergern, D. R., Chenoweth, J. G., McKay, R. D., et al. (2015). Practical impacts of genomic data "cleaning" on biological discovery using surrogate variable analysis. *BMC Bioinform* 16, 372. doi:10.1186/s12859-015-0808-5

Jesenak, M., Zelieskova, M., and Babusikova, E. (2017). Oxidative stress and bronchial asthma in children—causes or consequences? *Front. Pediatr.* 5, 162. doi:10.3389/fped.2017.00162

Jordan, F., Nemeria, N., and Gerfen, G. (2019). Human 2-oxoglutarate dehydrogenase and 2-oxoadipate dehydrogenase both generate superoxide/H2O2 in a side reaction and each could contribute to oxidative stress in mitochondria. *Neurochem. Res.* 44, 2325–2335. doi:10.1007/s11064-019-02765-w

Kaeslin, J., Micic, S., Weber, R., Müller, S., Perkins, N., Berger, C., et al. (2021). Differentiation of cystic fibrosis-related pathogens by volatile organic compound analysis with secondary electrospray ionization mass spectrometry. *Metabolites* 11, 773. doi:10.3390/metabo11110773

Kaur, B., Anderson, H. R., Austin, J., Burr, M., Harkins, L. S., Strachan, D. P., et al. (1998). Prevalence of asthma symptoms, diagnosis, and treatment in 12-14 year old children across Great Britain (international study of asthma and allergies in childhood, ISAAC UK). *BMJ* 316, 118–124. doi:10.1136/bmj.316.7125.118

Kölli, F., Breyer, M.-K., Hartl, S., Burghuber, O., Wouters, E. F. M., Sigsgaard, T., et al. (2022). Aero-allergen sensitization in the general population: Longitudinal analyses of the LEAD (lung heart social body) study. *J. Asthma Allergy* 15, 461–473. doi:10.2147/JAA.S349614

Kursa, M. B., Jankowski, A., and Rudnicki, W. R. (2010). Boruta – a system for feature selection. *Fundam. Inf.* 101, 271–285. doi:10.3233/FI-2010-288

Lan, J., Greter, G., Streckenbach, B., Arnoldini, M., Zenobi, R., and Slack, E. (2022). Non-invasive monitoring of microbiota and host metabolism using Secondary electrospray ionization-Mass spectrometry. *bioRxiv* 2022, 493434. doi:10.1101/2022.05.25.493434

Leek, J. T., and Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735. doi:10.1371/journal.pgen.0030161

Li, J., Li, X., Liu, X., Wang, X., Li, J., Lin, K., et al. (2022). Untargeted metabolomic study of acute exacerbation of pediatric asthma via HPLC-Q-Orbitrap-MS. *J. Pharm. Biomed. Anal.* 215, 114737. doi:10.1016/j.jpba.2022.114737

Looijmans-van den Akker, I., van Luijn, K., and Verheij, T. (2016). Overdiagnosis of asthma in children in primary care: A retrospective analysis. *Br. J. Gen. Pract.* 66, e152–e157. doi:10.3399/bjgp16X683965

Maarsingh, H., Dekkers, B. G. J., Zuidhof, A. B., Bos, I. S. T., Menzen, M. H., Klein, T., et al. (2011). Increased arginase activity contributes to airway remodelling in chronic allergic asthma. *Eur. Respir. J.* 38, 318–328. doi:10.1183/09031936.00057710

Morris, C. R., Poljakovic, M., Lavrisha, L., Machado, L., Kuypers, F. A., and Morris, S. M. (2004). Decreased arginine bioavailability and increased serum arginase activity in asthma. *Am. J. Respir. Crit. Care. Med.* 170, 148–153. doi:10.1164/rccm.200309-1304OC

Neerincx, A. H., Vijverberg, S. J. H., Bos, L. D. J., Brinkman, P., van der Schee, M. P., de Vries, R., et al. (2017). Breathomics from exhaled volatile organic compounds in pediatric asthma. *Pediatr. Pulmonol.* 52, 1616–1627. doi:10.1002/ppul.23785

Ni, K. di, and Liu, J. Y. (2021). The functions of cytochrome P450 ω-hydroxylases and the associated eicosanoids in inflammation-related diseases. *Front. Pharmacol.* 12, 716801. doi:10.3389/fphar.2021.716801

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., et al. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396. doi:10.1093/nar/gkab382

Papamichael, M. M., Katsardis, C., Erbas, B., Itsiopoulos, C., and Tsoukalas, D. (2019). Urinary organic acids as biomarkers in the assessment of pulmonary function in children with asthma. *Nutr. Res.* 61, 31–40. doi:10.1016/j.nutres.2018.10.004

Papamichael, M. M., Katsardis, C., Sarandi, E., Georgaki, S., Frima, E.-S., Varvarigou, A., et al. (2021). Application of metabolomics in pediatric asthma: Prediction, diagnosis and personalized treatment. *Metabolites* 11, 251. doi:10.3390/metabo11040251

Riscassi, S., Corradi, M., Andreoli, R., Maccari, C., Mercolini, F., Pescollderungg, L., et al. (2022). Nitric oxide products and aldehydes in exhaled breath condensate in children with asthma. *Clin. Exp. Allergy* 52, 561–564. doi:10.1111/cea.14066

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Roviezzo, F., Rossi, A., Caiazzo, E., Orlando, P., Riemma, M. A., Iacono, V. M., et al. (2017). Palmitoylethanolamide supplementation during sensitization prevents airway allergic symptoms in the mouse. *Front. Pharmacol.* 8, 857. doi:10.3389/fphar.2017.00857

Sagdic, A., Sener, O., Bulucu, F., Karadurmus, N., Özel, H. E., Yamanel, L., et al. (2011). Oxidative stress status and plasma trace elements in patients with asthma or allergic rhinitis. *Allergol. Immunopathol. Madr.* 39, 200–205. doi:10.1016/j.aller.2010.07.006

Saude, E. J., Skappak, C. D., Regush, S., Cook, K., Ben-Zvi, A., Becker, A., et al. (2011). Metabolomic profiling of asthma: Diagnostic utility of urine nuclear magnetic resonance spectroscopy. *J. Allergy Clin. Immunol.* 127, 757–764.e1-6. doi:10.1016/j.jaci.2010.12.1077

Schwarz, E. I., Martinez-Lozano Sinues, P., Bregy, L., Gaisl, T., Garcia Gomez, D., Gaugg, M. T., et al. (2016). Effects of CPAP therapy withdrawal on exhaled breath pattern in obstructive sleep apnoea. *Thorax* 71, 110–117. doi:10.1136/thoraxjnl-2015-207597

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* 48, 2097–2098. doi:10.1021/es5002105

Sereme, Y., Mezouar, S., Grine, G., Mege, J. L., Drancourt, M., Corbeau, P., et al. (2019). Methanogenic archaea: Emerging partners in the field of allergic diseases. *Clin. Rev. Allergy Immunol.* 57, 456–466. doi:10.1007/s12016-019-08766-5

Siersted, H. C., Hansen, H. S., Mostgaard, G., Hyldebrandt, N., Rees, P. J., et al. (1998). Population based study of risk factors for underdiagnosis of asthma in adolescence: Odense schoolchild study. *BMJ* 316, 651–657. doi:10.1136/bmj.316.7132.651

Singh, K. D., Osswald, M., Ziesenitz, V. C., Awchi, M., Usemann, J., Imbach, L. L., et al. (2021). Personalised therapeutic management of epileptic patients guided by pathway-driven breath metabolomics. *Commun. Med.* 1, 21. doi:10.1038/s43856-021-00021-3

Sinharoy, P., McAllister, S. L., Vasu, M., and Gross, E. R. (2019). "Environmental aldehyde sources and the health implications of exposure," in *Aldehyde dehydrogenases:*

From alcohol metabolism to human health and precision medicine. Editors J. Ren, Y. Zhang, and J. Ge (Singapore: Springer Singapore), 35–52. doi:10.1007/978-981-13-6260-6_2

Smolinska, A., Klaassen, E. M. M., Dallinga, J. W., van de Kant, K. D. G., Jobsis, Q., Moonen, E. J. C., et al. (2014). Profiling of volatile organic compounds in exhaled breath as a strategy to find early predictive signatures of asthma in children. *PLoS One* 9, e95668. doi:10.1371/journal.pone.0095668

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1027. doi:10.2202/1544-6115.1027

Tao, J. L., Chen, Y. Z., Dai, Q. G., Tian, M., Wang, S. C., Shan, J.-J., et al. (2019). Urine metabolic profiles in paediatric asthma. *Respirology* 24, 572–581. doi:10.1111/resp.13479

Vabalas, A., Gowen, E., Poliakoff, E., and Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One* 14, e0224365. doi:10.1371/journal.pone.0224365

van de Kant, K. D. G., van Berkel, J. J. B. N., Jöbsis, Q., Lima Passos, V., Klaassen, E. M. M., van der Sande, L., et al. (2013). Exhaled breath profiling in diagnosing wheezy preschool children. *Eur. Respir. J.* 41, 183–188. doi:10.1183/09031936.00122411

van Gent, R., van Essen, L. E. M., Rovers, M. M., Kimpen, J. L. L., van der Ent, C. K., and de Meer, G. (2007). Quality of life in children with undiagnosed and diagnosed asthma. *Eur. J. Pediatr.* 166, 843–848. doi:10.1007/s00431-006-0358-y

van Mastrigt, E., Reyes-Reyes, A., Brand, K., Bhattacharya, N., Urbach, H. P., Stubbs, A. P., et al. (2016). 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. doi:10.1088/1752-7155/10/2/026003

van Vliet, D., Smolinska, A., Jöbsis, Q., Rosias, P. P. R., Muris, J. W. M., Dallinga, J. W., et al. (2016). Association between exhaled inflammatory markers and asthma control in children. *J. Breath. Res.* 10, 016014. doi:10.1088/1752-7155/10/1/016014

Varma, S., and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinform* 7, 91. doi:10.1186/1471-2105-7-91

Weber, R., Haas, N., Baghdasaryan, A., Bruderer, T., Inci, D., Micic, S., et al. (2020). Volatile organic compound breath signatures of children with cystic fibrosis by real-time SESI-HRMS. *ERJ Open Res.* 6, 00171–02019. doi:10.1183/23120541.00171-2019

Wendell, S. G., Baffi, C., and Holguin, F. (2014). Fatty acids, inflammation, and asthma. *J. Allergy Clin. Immunol.* 133, 1255–1264. doi:10.1016/j.jaci.2013.12.1087

Yang, C. L., Simons, E., Foty, R. G., Subbarao, P., To, T., and Dell, S. D. (2017). Misdiagnosis of asthma in schoolchildren. *Pediatr. Pulmonol.* 52, 293–302. doi:10.1002/ppul.23541

Zhao, J. v., and Schooling, C. M. (2019). The role of linoleic acid in asthma and inflammatory markers: A mendelian randomization study. *Am. J. Clin. Nutr.* 110, 685–690. doi:10.1093/ajcn/nqz130

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 04 April 2023
DOI 10.3389/fmolb.2023.1154149

OPEN ACCESS

EDITED BY
Andras Szeitz,
University of British Columbia, Canada

REVIEWED BY
Jasmeen Kaur,
Punjabi University, India
Shane Peter Fitzgerald,
King's College London, United Kingdom

*CORRESPONDENCE
Anas M. Abdel Rahman,
aabdelrahman46@kfshrc.edu.sa

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 30 January 2023
ACCEPTED 03 March 2023
PUBLISHED 04 April 2023

CITATION
Alodaib AN, Nimer RM, Alhumaidy R, Alhenaky A, Abdel Jabar M, AlMalki RH and Abdel Rahman AM (2023), Biomarker discovery in galactosemia: Metabolomics with UPLC/HRMS in dried blood spots. *Front. Mol. Biosci.* 10:1154149. doi: 10.3389/fmolb.2023.1154149

Biomarker discovery in galactosemia: Metabolomics with UPLC/HRMS in dried blood spots

Ahmad N. Alodaib[1,2†], Refat M. Nimer[3†], Rowan Alhumaidy[1], Alaa Alhenaky[1], Mai Abdel Jabar[1], Reem H. AlMalki[1,4] and Anas M. Abdel Rahman[1,2*]

[1]Metabolomics Section, Department of Clinical Genomics, Center for Genomics Medicine, King Faisal Specialist Hospital and Research Centre (KFSHRC), Riyadh, Saudi Arabia, [2]Department of Biochemistry and Molecular Medicine, College of Medicine, Al Faisal University, Riyadh, Saudi Arabia, [3]Department of Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid, Jordan, [4]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia

Introduction: Galactosemia (GAL) is a genetic disorder that results in disturbances in galactose metabolism and can lead to life-threatening complications. However, the underlying pathophysiology of long-term complications in GAL remains poorly understood.

Methods: In this study, a metabolomics approach using ultra-performance liquid chromatography coupled with high-resolution mass spectrometry was used to investigate metabolomic changes in dried blood spots of 15 patients with GAL and 39 healthy individuals.

Results: The study found that 2,819 metabolites underwent significant changes in patients with GAL compared to the control group. 480 human endogenous metabolites were identified, of which 209 and 271 were upregulated and downregulated, respectively. PA (8:0/LTE4) and ganglioside GT1c (d18:0/20:0) metabolites showed the most significant difference between GAL and the healthy group, with an area under the curve of 1 and 0.995, respectively. Additionally, the study identified potential biomarkers for GAL, such as 17-alpha-estradiol-3-glucuronide and 16-alpha-hydroxy DHEA 3-sulfatediphosphate.

Conclusion: This metabolomics study deepened the understanding of the pathophysiology of GAL and presented potential biomarkers that might serve as prognostic biomarkers to monitor the progression or support the clinical diagnosis of GAL.

KEYWORDS

biomarker, dried blood spot, galactosemia, metabolomics, newborn screening (NBS), ultra-performance liquid chromatography

1 Introduction

Galactosemia is an autosomal recessive disorder caused by a defect in the enzyme galactose-1-phosphate uridyltransferase (GALT) (Isselbacher et al., 1956; Berry, 2015). This enzyme is needed to convert galactose (a sugar found in dairy, fruit, and some foods) into the body's primary energy source (Leloir, 1951). If this enzyme is deficient, galactose accumulates to toxic levels in the body, leading to a severe neurological and metabolic

disorder. Symptoms of galactosemia can include vomiting, lethargy, seizures, enlarged liver, jaundice, kidney failure, and changes in brain development.

A defective gene causes the classic form of GAL on chromosome 3. It is characterized by severe deficiency of all three enzymes required for the metabolism of galactose to glucose: galactose-galactokinase (GALK) (GAL type II), galactose-1-phosphate uridylyltransferase (also known as GALT), and UDP-galactose-4-epimerase (GALE) enzymes (GAL type III) (Pasquali et al., 2018). Without these enzymes, galactose is not converted to glucose, accumulates in the body, or is excreted in the urine. The incomplete breakdown of galactose then affects numerous metabolic pathways, which results in several serious symptoms and potential long-term complications such as cataracts, liver damage, and an increased risk of developing neurological disorders. The primary treatment for galactosemia is a strict low-galactose diet. All sources of galactose and lactose, such as cow's milk, need to be eliminated from the diet. An alternative lactose-free, low-galactose breast milk or formula can be used for nutrition. Depending on the individual, the diet may need to be modified periodically. Infants with galactosemia should have regular check-ups with their doctor to ensure proper nutrition, growth, and development.

The Leloir pathway is the metabolic pathway used to convert galactose into glucose. It begins with the conversion of galactose to glucose-1-phosphate, which is catalyzed by the enzyme galactokinase. Glucose-1-phosphate is then converted to glucose-6-phosphate by uridyl transferase and, finally, glucose by the enzyme glucokinase. At the same time, galactose is rapidly metabolized *via* the Leloir pathway once it enters the cell, where initially, GALK catalyzes the phosphorylation of galactose (Walter and Fridovich-Keil, 2019). Then GAL transforms UDP-glucose and galactose 1-phosphate into glucose-1-phosphate and UDP-galactose. Finally, GALE catalyzes the conversion of UDP-galactose to UDP-glucose (Timson, 2016). Additionally, type IV galactosemia is a newly found hereditary metabolic disorder. It is caused by mutations in the galactose mutarotase gene, which results in the diminished activity of the enzyme galactose mutarotase. This enzyme catalyzes the interconversion of the α- and β-anomers of d-galactose and several other monosaccharides (Banford and Timson, 2021). However, the mechanisms of GAL disorder are still poorly understood (van Weeghel et al., 2018). The clinical complications associated with classical galactosemia include cataracts, developmental delays, learning disabilities, speech problems, failure to thrive, intestinal problems, and liver damage. If left untreated, galactosemia can lead to life-threatening conditions such as sepsis, multiple organ failure, and death. It is also associated with an increased risk of ovarian failure in females (Succoio et al., 2022).

The pathophysiology of the long-term complication in GAL needs to be better understood, and predictive biomarkers need to be included (Hermans et al., 2022). Prognostic uncertainty may lead to unnecessary or harmful treatment for patients with GAL, which burden patients and parents (Knerr et al., 2015). Even though neonatal detection and dietary restriction of galactose may change the clinical picture of a newborn, it does not stop long-term problems from happening (Succoio et al., 2022).

Most of the world's newborn screening (NBS) programs include a screen for GAL; despite the level of its false discovery rate, even at early diagnosis, there are often long-term complications (Ohlsson et al., 2011). Fanconi-Bickel disease, liver illness, glycogen storage disease type XI, and even certain drugs may all cause false-positive screening findings in infants (Peduto et al., 2004; Kotb et al., 2019). Most NBS programs depend on measuring GAL activity in DBS to diagnose GAL.

Total galactose (galactose + galactose-1-phosphate) is measured in around 30% of NBS programs as a main screening approach or in conjunction with GAL testing in DBS. However, false negative screening results for GAL may be seen in babies who are given lactose-free formula or who are receiving complete parenteral nutrition if the diagnosis is based only on total galactose (Pasquali et al., 2018). A second analysis of dried blood from the same newborn screening card is undertaken to monitor particular metabolites or metabolic pathways to overcome the issue of non-specific first-line parameters and the resultant high false positive rate (Lehotay et al., 2011). Furthermore, positive results in screening tests, clinical examination, and biochemical and molecular diagnostics are required to confirm patients with GAL. However, based on the biochemical, enzymatic, and genetic information, that is, now available, it is not feasible to provide an accurate prediction of the clinical prognosis at the time of diagnosis.

Because there are no validated biomarkers for the diagnosis and prognosis of GAL, and no specific and reliable treatment regimens, more studies on GAL should be conducted. Metabolomics describes analyzing all metabolites (compounds with low molecular weight, generally 1,500 Da) present in a particular sample acquired from a biological system (Patti et al., 2012; Dahabiyeh et al., 2020; Gu et al., 2020). Metabolomics has emerged as a potentially useful diagnostic and prognostic tool that might explain disease pathogenesis (Masood et al., 2021; Jacob et al., 2022; Jans et al., 2022). There currently needs to be more investigations on the pathophysiology of GAL that concentrate on urine or blood metabolomics profiling (Taylor Fischer et al., 2019; Hermans et al., 2022). Therefore, investigating the GAL metabolomics profile may aid in finding potential biomarkers, shed light on the mechanisms behind the disease's progression, and ultimately aid in its early detection.

In this study, metabolomics employing ultra-performance liquid chromatography coupled with high-resolution mass spectrometry (UPLC/HRMS) was used to detect and quantify differences in metabolite levels between GAL and healthy groups that could potentially serve as biomarkers for the diagnosis or monitoring of GAL and could also provide insights into the underlying biological mechanisms of the disorder.

2 Materials and methods

2.1 Characteristics of the study population

Fifty-four DBS samples were collected from genetically and biochemically confirmed GAL (n = 15) patients at King Faisal Specialist Hospital and Research center (KFSHRC) and healthy controls (n = 39). These healthy controls were age-gender matched with the patient group. 4 out of 19 GAL patients and 7 out of 46 healthy controls were excluded from this study due to 1) inability or unwillingness to provide informed consent or 2) diagnosis with conditions other than GAL. The Research Ethics committee approved this study and Institutional Review Board at

KFSHRC (RAC# 2160027). It was performed following the ethical standards of the Declaration of Helsinki.

2.2 Metabolites extraction

The polar metabolites were extracted from DBS samples using our developed standard protocol (Jacob et al., 2018). Five 3 mm size DBS disks were used for metabolite extraction using methanol, acetonitrile, and water (40:40:20%) for protein precipitation. The mixture was mixed at 25°C and 600 rpm for 2 hours in a thermomixer (Eppendorf, Germany). Pooled QC samples were prepared using aliquots from the study samples. Afterward, the supernatants were transferred to another set of tubes, evaporated in SpeedVacc (Christ, City, Germany), and stored at −80°C until LC-MS analysis.

2.3 UPLC/HRMS

The metabolomics profile for the study samples was collected using our laboratory's applicable standard protocol (Jaber et al., 2022). In detail, the dry extracted samples were resuspended with 50% mobile phase A and B (A: 0.1% formic acid in dH_2O, and B: 0.1% formic acid in 50% MeOH and ACN). The extracted metabolites were chromatographed using an Acquity UPLC using XSelect HSS C18 (100 × 2.1 mm, 2.5 µm) column (Waters Ltd., Elstree, United Kingdom). A gradient mobile phase elution was scheduled in this method as follows: 0–16 min 95%–5% A, 16–19 min 5% A, 19–20 min 5%–95% A, and 20–22 min, 95%–95% A, all at a flow rate of 300 µl/min. The eluted molecules were detected using a Waters Acquity UPLC connected to a Waters Xevo G2-S QTOF high-resolution mass spectrometry system. In separate runs, the molecules were ionized using positive and negative electrospray ionization modes (ESI+, ESI-). In ESI+, the capillary voltage was set to 3.20 kV. The cone voltage was 40 V, the desolvation temperature was 500°C, the nitrogen desolvation gas flow to 800 L/h, and the cone gas flow was 50 L/h. In ESI−, a capillary voltage of −3 kV was used. The collision energies of low and high functions were set at 0 V and 10–50 V, respectively, in MS^E mode. The mass spectrometer was calibrated, as recommended by the vendor, with sodium formate in the range of 100–1,200 Da in both ionization modes. Accurate mass measurements were maintained by continuously infusing leucine-enkephaline lock mass compound (ESI + m/z 556.2771, ESI- m/z 554.2615) and alternating between the sample and the reference every 45 and 60 s for ESI+ and ESI-, respectively. The lock spray was 10 µl/min, 0.5 s scan time, cone voltage 30 V, collision energy 4 eV. The data-independent acquisition was performed in continuum mode with Masslynx™ V4.1 workstation (Waters Corporation, MA, United States).

2.4 Data processing and statistical analysis

Peak picking and alignment of detected ion (m/z, Rt) were processed using Progenesis QI v.3.0 software from Waters (Waters Corporation, MA, United States).

The raw data were deposited in MetaboLight (accession Number MTBLS6996).

Multivariate statistical analysis was performed using MetaboAnalyst v5.0 (McGill University, Montreal, QC, Canada) (Worley and Powers, 2013). Firstly, data were subjected to log transformation, mean centering, and Pareto scaling and then used to generate principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and orthogonal projections to latent structures discriminant analysis (OPLS-DA) models. OPLS-DA models were evaluated using the fitness-of-model (R2Y) and predictive ability (Q2) values.

Univariate analysis was performed using Mass Profiler Professional (MPP) Software (Agilent Technologies, Inc., Santa Clara, CA, United States). The total sample median was used to normalize the signal and ensure normal distribution. Volcano Plot analysis was performed to identify significantly alters between GAL patients and healthy control using Moderated T. Test, false discovery rate (FDR) corrected p-value ≤0.05 and fold change (FC) cut-off of 2. Venn diagrams were developed using MPP Software (Agilent Technologies, Inc., Santa Clara, CA, United States).

Pathway analysis and biomarkers linked with GAL disorder were performed using MetaboAnalyst v5.0 (McGill University, Montreal, QC, Canada)—a pathway view of statistically significant pathways flagged from the metabolome view based on matched metabolites. The pathways are arranged based on the p-value (y-axis), which indicates the pathway enrichment analysis, and pathway impact values (x-axis) representing pathway topology analysis. In addition, Receiver Operating Characteristic (ROC) curves were created using the PLS-DA approach in the MetaboAnalyst v 5.0 for global analysis to identify possible biomarkers. Metabolites were putatively identified based on the exact mass searched against different databases, including Human Metabolome Database and METLIN. The exogenous compounds, such as drugs, food additives, and environmental compounds, were excluded from the final list.

3 Results

3.1 Feature detection and metabolites identification

Using the UPLC/HRMS data, comprehensive untargeted metabolomics analyses were performed on the DBS samples obtained from 15 GAL patients and 39 healthy controls. In total, 25,607 m/z features were detected in positive (n = 12,541) and negative (n = 13,066) ionization modes. After applying the filter of 80% of all samples, 20,775 features remained to statistically evaluate among the patients with GAL and healthy control, as described in the method section.

3.2 Metabolomics profiling for GAL compared to control

Multivariate and univariate analyses were used to determine whether metabolites were significantly different in GAL compared

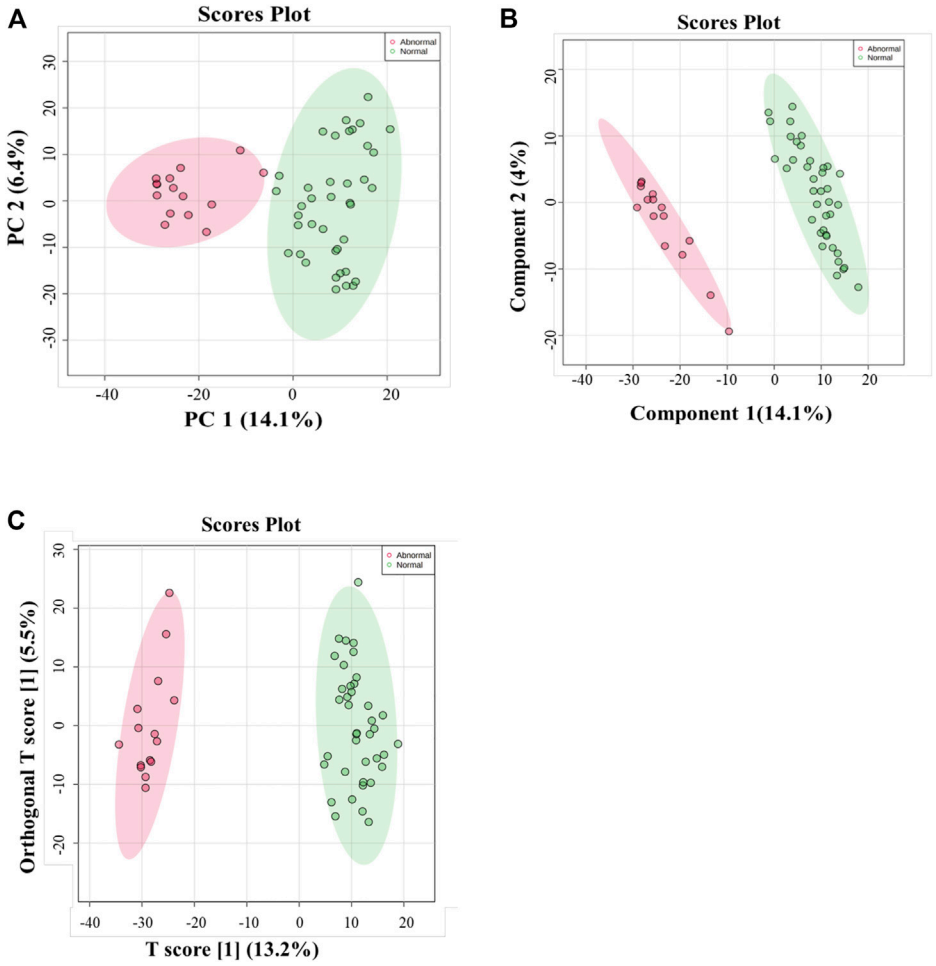


FIGURE 1
(A) Principal component analysis (PCA) model for 54 samples obtained from 15 GAL patients and 39 control that showed a clear separation between the two groups (GAL patients and healthy control). **(B)** PLS-DA Score Plots revealed a clear separation between the groups (GAL patients and healthy control). **(C)** Orthogonal partial least squares-discriminant analysis (OPLS-DA) score plot showed evident separation between two groups (GAL patients and healthy control). The robustness of the created models was evaluated by the fitness of the model (R2Y = 0.972) and predictive ability (Q2 = 0.856) values.

to the control group. As a result of unsupervised, PCA revealed good clustering between GAL patients (red) and the healthy group (green). The total variance of the first two principal components contributed 20.5% in the PCA model for the two study groups (PC 1 = 14.1% and PC 2 = 6.4%) (Figure 1A). Once separation had been assessed, PLS-DA and OPLS-DA were applied to maximize the separation of the groups observed by PCA. The scores plot from the PLS-DA (Figure 1B) and the OPLS-DA (Figure 1C) showed clear group separation, which validated the PCA results. The OPLS-DA model yielded satisfactory fitness of the model (R2Y = 0.972) and predictive ability (Q2 = 0.856) values (Figure 1C). The contributing metabolites in these models' separation between study groups were explored using univariate analysis (Student T. Test and fold change analyses).

Next, a binary comparison between the GAL group and healthy control using volcano plot analysis revealed that 1,300 metabolites were upregulated (red) whereas 1,519 metabolites were downregulated (blue) in GAL patients compared to healthy control (FDR $p \leq 0.05$, FC cut-off of 2), respectively (Figure 2A). Four hundred eighty metabolites were annotated as endogenous human metabolites and are listed in Supplementary Table 1. Further examination using hierarchical clustering analysis (HCA) in Figure 2B depicts differences in the abundance of the top 25 perturbed metabolites between GAL and control groups.

The metabolic pathway analysis revealed that pyrimidine metabolism is the most significantly altered pathway in GAL compared to the control, as displayed in Figure 3.

The ROC analysis was used to identify metabolites that might act as potential biomarkers and to assess their diagnostic accuracy (Figure 4). Multivariate exploratory ROC analysis was created using PLS-DA for classification and feature ranking. Figure 4A shows that the top-ranked metabolites in ROC curves show the area under the curves (AUCs) ranging from 0.992 to 1; confidence Interval (CI): 0.946–1 and 1-1. The selected frequency plots represent the significant features of the expressed metabolites in the patients with GAL and control groups (Figure 4B). The selected

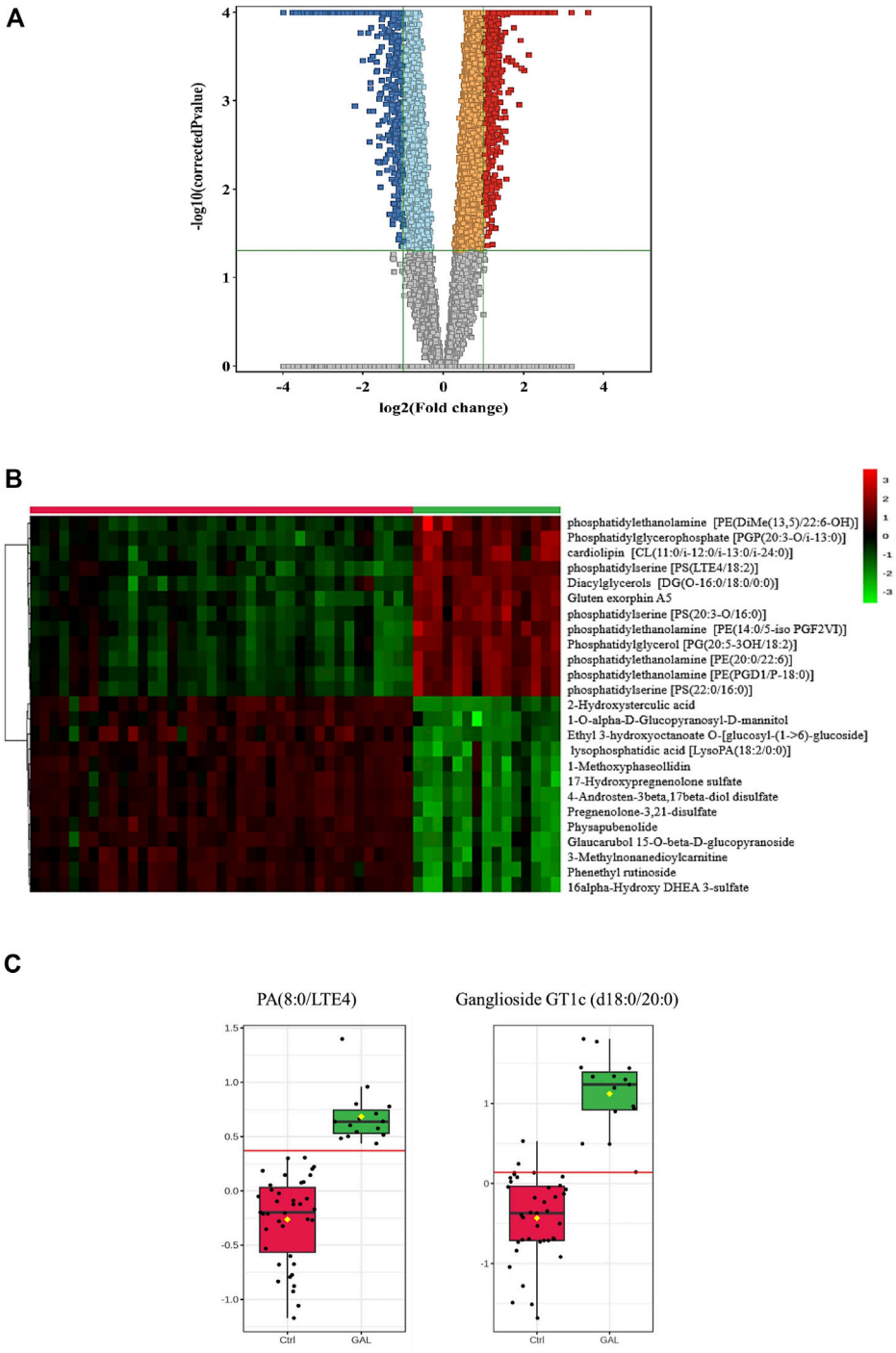


FIGURE 2
(A) Volcano Plot demonstrates the statistically significant altered metabolites filtered between the two groups (GAL patients and healthy control) that 2,819 significantly were dysregulated metabolites (FDR p-value ≤0.05, FC 2), of which 1,300 (red) and 1,519 (blue) metabolites were up-and downregulated in GAL patients compared to healthy control, respectively. Red and blue refer to up-and downregulated metabolites, respectively. Orange and light blue squares refer to metabolites that failed to pass fold change cutoffs and were up- and downregulated, respectively. Gray square metabolites failed to pass both cutoffs. **(B)** Heat map representing the top 25 significantly ($p < 0.05$) altered metabolites between the two study groups; healthy control (red) and GAL patients (green). **(C)** Boxplots for a couple of metabolites [Ganglioside GT1c (d18:0/20:0) and PA (8:0/LTE4)] where green represents GAL patients and red represents Control.

frequency plot shows metabolites, such as 16-alpha-hydroxy DHEA 3-sulfatediphosphate (UDP) and 17-alpha-estradiol-3-glucuronide to be downregulated in patients with GAL in comparison to the control group with AUC 0.997 and 0.961, respectively (Figures 4C, D). In comparison, metabolites such as phosphatidylcholine and diacylglycerols (DG) (20:3n6/0:0/20:4n3) were upregulated.

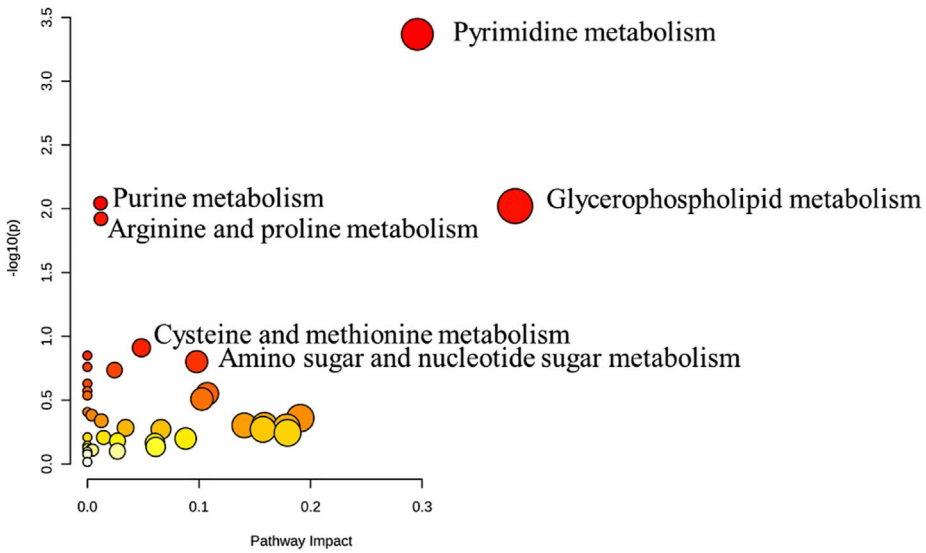


FIGURE 3
Pathway Analysis for the significant metabolites dysregulated between GAL patients and healthy control. Four hundred eighty metabolites were ultimately identified as human endogenous metabolites, where 209 and 271 were up and downregulated, respectively. Colors (varying from yellow to red) mean the metabolites are in the data with different significance levels (*p*-value).

Moreover, phosphatidic acid (PA (8:0/LTE4)) (Figure 4E) and ganglioside GT1c (d18:0/20:0) (Figure 4F) were upregulated in patients with GAL compared to the control group with an AUC 1 and 0.995, respectively shows highest discriminatory power.

4 Discussion

The urgent need for novel biomarkers for early diagnosis and prognosis prediction of GAL disorder has prompted the investigation of potential biomarkers using various experimental approaches. This study conducted metabolomics analyses to identify biomarkers with UPLC/HRMS by following changes in the metabolic profiles of patients with GAL. As an autosomal recessive hereditary genetic disorder, GAL can result in life-threatening health complications unless lactose is eliminated from the diet immediately after birth (Berry, 2021). The clinical outcome of patients with GAL varies widely (Welsink-Karssies et al., 2020). Additionally, pitfalls in diagnosing GAL are present due to false negative and positive newborn screening results (Pasquali et al., 2018). Thus, there is an urgent need to find novel biomarkers for early diagnosis and prognosis prediction of GAL disorder.

Despite the emerging field of metabolomics as the newest Omics platform that focuses on metabolites, small molecules (<1,500 Da) hold promise to shine a light on the molecular mechanisms of several diseases, which may help for diagnostic and therapeutic purposes (Jacob et al., 2019), very few studies have focused on GAL in humans biological fluids (Janeckova et al., 2015; Taylor Fischer et al., 2019; Hermans et al., 2022).

A volcano plot analysis was utilized to identify potential biomarkers of GAL. 2,819 metabolites showed significant differences between the GAL group and the control group. In the heatmap, the top 25 metabolites with the most significant differences in abundance between the groups were visualized and identified as potential biomarkers for GAL. Among these metabolites, phosphatidylethanolamine, which is a category of phospholipids present in biological membranes, was found to be the most significantly upregulated metabolite in the GAL group, confirming the increase of phosphatidylethanolamine in complications of GAL such as neurological impairments and cataracts (Jernigan Jr et al., 2005; López de Frutos et al., 2022).

Pyrimidine metabolism was the most significant pathway that significantly altered between GAL and healthy controls. The pyrimidines are the building blocks of DNA and RNA. They also form active intermediates in carbohydrate metabolism, such as UDP-glucose (Dewulf et al., 2021). Furthermore, UDP-glucose is an organic pyrimidine nucleotide sugar molecule (Ng et al., 2015). In the physiological process, the UDP-glucose is transformed into UDP-galactose in the presence of GAL (Veiga-da-Cunha et al., 2020). While in patients with GAL, the activity of the GAL enzyme is absent or barely detectable (Berry, 2021). Thus, a close link between GAL and alteration in pyrimidine metabolism was shown previously (Taylor Fischer et al., 2019), which matched our result.

We found that ganglioside GT1c and PA (8:0/LTE4) showed the highest discrimination ability between GAL and the control group. Ganglioside GT1c is a glycosphingolipid (ceramide and oligosaccharide) with one or more sialic acids widely distributed throughout the body, especially abundant in the brain and other parts of the central nervous system (Vukelic et al., 2001). Galactosylation of complex molecules and the production of various glycoproteins/glycolipids rely on GALE, which is also responsible for the interconversion of UDP-N-acetylglucosamine (UDP-GlcNAc) and UDP-N-acetylgalactosamine (UDP-GalNAc) (Banford et al., 2021). Therefore, neurological complications in patients with GAL may result from prenatal-neonatal toxicity or

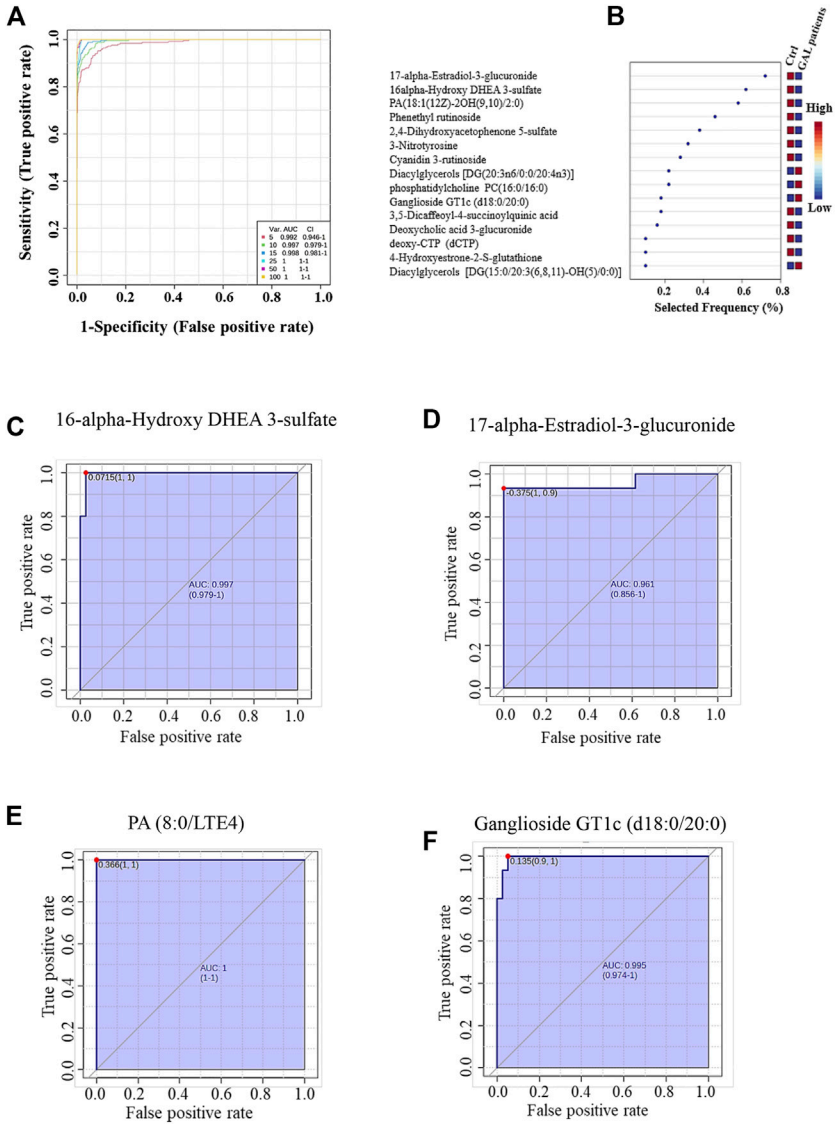


FIGURE 4
(A) The Receiver Operating Characteristics (ROC) curve was generated by the OPLS-DA model, with Area Under the Curve (AUC) values calculated from the combination of 5, 10, 15, 25, 50, and 100 metabolites. **(B)** The frequency plot shows the significantly dysregulated endogenous metabolites between the study groups. Representatives downregulated metabolites with their ROC curves are demonstrated for 16-alpha-hydroxy DHEA 3-sulfatediphosphate (AUC (0.997) **(C)** and 17-alpha-estradiol-3-glucuronide AUC (0.961) **(D)** in GAL patients. Furthermore, PA (8:0/LTE4) AUC (1) **(E)** and ganglioside GT1c (d18:0/20:0) AUC (0.995) **(F)** as examples of upregulated metabolites in GAL patients.

persistent glycoprotein and glycolipid synthesis abnormalities (Coman et al., 2010).

Moreover, oxidized phosphatidic acid is known as PA (8:0/LTE4), which is upregulated in our present study. A phosphate moiety occupies a glycerol substitution site in oxidized phosphatidic acids, which are glycerophospholipids in which at least one of the fatty acyl chains has undergone oxidation (Wishart et al., 2022)—metabolic perturbations in glycerophospholipids found in patients with GAL (Taylor Fischer et al., 2019).

One of the complications of GAL is liver cirrhosis since GAL is a common metabolic liver disorder of childhood (Sahoo et al., 2015). In the liver, glucuronidation takes place, where it is used to assist in the excretion of toxic substances, drugs, or other substances by attaching glucuronic acid *via* a glycosidic bond to the substance. The resulting glucuronide, which has a much higher water solubility than the original substance, is eventually excreted by the kidneys (Wishart et al., 2022). Thus, metabolic liver diseases will affect glucuronidation (Sharma and Nagalli, 2021). Our findings of downregulated 17-alpha-estradiol-3-glucuronide metabolite produced in the liver after glucuronidation of 17-alpha-estradiol by UDP glucosyl transferase are consistent with this hypothesis. Moreover, to our knowledge, 17-alpha-estradiol-3-glucuronide has been identified in the blood of patients with GAL for the first time, which can serve as a potential prognostic biomarker for GAL concerning liver complications.

Furthermore, the metabolite 16 alpha-hydroxy DHEA 3-sulfatediphosphate derived mainly from the fetus and served as a precursor for placental estriol biosynthesis significantly decreased in the GAL group (Schweigmann et al., 2014). 16 alpha-hydroxy DHEA 3-sulfate is a natural human metabolite in pregnant women's placenta and breast milk (Wishart et al., 2022). However, breastfeeding should be avoidable in babies with GAL since breast milk contains lactose (Berry, 2021). Thus, it may explain the decreased level of 16 alpha-hydroxy DHEA 3-sulfatediphosphate in patients with GAL.

This study had some limitations, such as the few patients with GAL and the need for the patient group to be used for external validation. Nevertheless, using the metabolomics approach, our study is one of the few to reveal specific metabolite changes between GAL and healthy controls. Our findings serve as the initial step for further investigations in greater detail.

5 Conclusion

There is currently no biomarker available to predict life-threatening complications in patients with GAL, which are associated with early death among these patients. This study used the HRMS-based metabolomics approach for the first time to gain new insights into the perturbed biochemical pathways in GAL compared to healthy control and to identify potential predictive biomarkers.

A total of 480 endogenous metabolites were identified, and they showed significant dysregulation. These metabolites can provide important insights into the pathophysiological state of GAL disorder.

Two metabolites, ganglioside GT1c and PA (8:0/LTE4), had the highest discrimination between GAL and the healthy group. Moreover, our results showed novel potential biomarkers for GAL, such as 17-alpha-estradiol-3-glucuronide and 16 alpha-hydroxy DHEA 3-sulfatediphosphate.

However, the biomarkers obtained through untargeted metabolomics require additional validation, which may involve the targeted UPLC/HRMS-based method to ensure their accuracy and reliability for clinical use. In addition, further studies are necessary to evaluate these biomarkers' reproducibility, stability, and performance in large separate cohorts to determine their potential clinical value.

Data availability statement

The data were deposited in a MetaboLight accession number MTBLS6996.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board's (IRB) ethics committee at King Faisal Specialist Hospital and Research Centre (KFSHRC) (RAC No. 2160 027). Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

AMA, ANA experimental design, formal analysis, methododolgy, RM, AA; RA, MA, data analysis RM and AMA, supervision AMA and ANA, writing—original draft, ANA and RN; writing—review and editing, AMA, ANA, and RN.

Acknowledgments

The authors would like to express their gratitude to Salah al Baz, Executive Director of the Center of Clinical Genomics, King Faisal Specialist Hospital and Research Centre, to Abdullah N. Alsuwaidan, Chairman of the Center of Clinical Genomics, for the unconditional support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1154149/full#supplementary-material

References

Banford, S., and Timson, D. J. (2021). The structural and molecular biology of type IV galactosemia. *Biochimie* 183, 13–17. doi:10.1016/j.biochi.2020.11.001

Banford, S., McCorvie, T. J., Pey, A. L., and Timson, D. J. (2021). Galactosemia: Towards pharmacological chaperones. *J. Pers. Med.* 11 (2), 106. doi:10.3390/jpm11020106

Berry, G. T. (2015). "Disorders of galactose metabolism," in *Rosenberg's molecular and genetic basis of neurological and psychiatric disease*. Editor R. N. R. a. J. M. Pascual (Elsevier), 615–626.

Berry, G. T. (2021). "Classic galactosemia and clinical variant galactosemia," in *GeneReviews*. Editors D. B. E. Margaret, P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. W. Gripp, et al. (Seattle: University of Washington), 1993–2023.

Coman, D. J., Murray, D. W., Byrne, J. C., Rudd, P. M., Bagaglia, P. M., Doran, P. D., et al. (2010). Galactosemia, a single gene disorder with epigenetic consequences. *Pediatr. Res.* 67 (3), 286–292. doi:10.1203/PDR.0b013e3181cbd542

Dahabiyeh, L. A., Malkawi, A. K., Wang, X., Colak, D., Mujamammi, A. H., Sabi, E. M., et al. (2020). Dexamethasone-induced perturbations in tissue metabolomics revealed by chemical isotope labeling LC-MS analysis. *Metabolites* 10 (2), 42. doi:10.3390/metabo10020042

Dewulf, J. P., Marie, S., and Nassogne, M.-C. (2021). Disorders of purine biosynthesis metabolism. *Mol. Genet. Metab.* 136 (3), 190–198. doi:10.1016/j.ymgme.2021.12.016

Gu, X., Al Dubayee, M., Alshahrani, A., Masood, A., Benabdelkamel, H., Zahra, M., et al. (2020). Distinctive metabolomics patterns associated with insulin resistance and type 2 diabetes mellitus. *Front. Mol. Biosci.* 7, 609806. doi:10.3389/fmolb.2020.609806

Hermans, M. E., van Weeghel, M., Vaz, F. M., Ferdinandusse, S., Hollak, C. E., Huidekoper, H. H., et al. (2022). Multi-omics in classical galactosemia: Evidence for the involvement of multiple metabolic pathways. *J. Inherit. Metab. Dis.* 45 (6), 1094–1105. doi:10.1002/jimd.12548

Isselbacher, K. J., Anderson, E. P., Kurahashi, K., and Kalckar, H. M. (1956). Congenital galactosemia, a single enzymatic block in galactose metabolism. *Science* 123 (3198), 635–636. doi:10.1126/science.123.3198.635

Jaber, M. A., Benabdelkamel, H., Dahabiyeh, L. A., Masood, A., AlMalki, R. H., Musambil, M., et al. (2022). The metabolomics approach revealed a distinctive metabolomics pattern associated with hyperthyroidism treatment. *Front. Endocrinol. (Lausanne)* 13, 1050201. doi:10.3389/fendo.2022.1050201

Jacob, M., Malkawi, A., Albast, N., Al Bougha, S., Lopata, A., Dasouki, M., et al. (2018). A targeted metabolomics approach for clinical diagnosis of inborn errors of metabolism. *Anal. Chim. Acta* 1025, 141–153. doi:10.1016/j.aca.2018.03.058

Jacob, M., Lopata, A. L., Dasouki, M., and Abdel Rahman, A. M. (2019). Metabolomics toward personalized medicine. *Mass Spectrom. Rev.* 38 (3), 221–238. doi:10.1002/mas.21548

Jacob, M., Nimer, R. M., Alabdaljabar, M. S., Sabi, E. M., Al-Ansari, M. M., Housien, M., et al. (2022). Metabolomics profiling of nephrotic syndrome towards biomarker discovery. *Int. J. Mol. Sci.* 23 (20), 12614. doi:10.3390/ijms232012614

Janeckova, H., Kalivodova, A., Najdekr, L., Friedecky, D., Hron, K., Bruheim, P., et al. (2015). Untargeted metabolomic analysis of urine samples in the diagnosis of some inherited metabolic disorders. *Biomed. Pap. Med. Fac. Univ. Palacky. Olomouc Czech. Repub.* 159 (4), 582–585. doi:10.5507/bp.2014.048

Jans, J. J. M., Broeks, M. H., and Verhoeven-Duif, N. M. (2022). Metabolomics in diagnostics of inborn metabolic disorders. *Curr. Opin. Syst. Biol.* 29, 100409. doi:10.1016/j.coisb.2021.100409

Jernigan, H. M., Jr, Blum, P. S., Chakrabarti, I., Su, Y., and Zigler, J. S., Jr (2005). Effects of cataractogenesis on the CDP-choline pathway: Increased phospholipid synthesis in lenses from galactosemic rats and 13/N Guinea pigs. *Ophthalmic Res.* 37 (1), 7–12. doi:10.1159/000082764

Knerr, I., Coss, K. P., Kratzsch, J., Crushell, E., Clark, A., Doran, P., et al. (2015). Effects of temporary low-dose galactose supplements in children aged 5–12 y with classical galactosemia: A pilot study. *Pediatr. Res.* 78 (3), 272–279. doi:10.1038/pr.2015.107

Kotb, M. A., Mansour, L., and Shamma, R. A. (2019). Screening for galactosemia: Is there a place for it? *Int. J. Gen. Med.* 12, 193–205. doi:10.2147/IJGM.S180706

Lehotay, D., Hall, P., Lepage, J., Eichhorst, J., Etter, M., and Greenberg, C. (2011). LC–MS/MS progress in newborn screening. *Clin. Biochem.* 44 (1), 21–31. doi:10.1016/j.clinbiochem.2010.08.007

Leloir, L. F. (1951). The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. *Arch. Biochem. Biophys.* 33 (2), 186–190. doi:10.1016/0003-9861(51)90096-3

López de Frutos, L., Almeida, F., Murillo-Saich, J., Conceição, V. A., Guma, M., Queheberger, O., et al. (2022). Serum phospholipid profile changes in gaucher disease and Parkinson's disease. *Int. J. Mol. Sci.* 23 (18), 10387. doi:10.3390/ijms231810387

Masood, A., Jacob, M., Gu, X., Abdel Jabar, M., Benabdelkamel, H., Nizami, I., et al. (2021). Distinctive metabolic profiles between Cystic Fibrosis mutational subclasses and lung function. *Metabolomics* 17 (1), 4–19. doi:10.1007/s11306-020-01760-5

Ng, B. G., Wolfe, L. A., Ichikawa, M., Markello, T., He, M., Tifft, C. J., et al. (2015). Biallelic mutations in CAD, impair de novo pyrimidine biosynthesis and decrease glycosylation precursors. *Hum. Mol. Genet.* 24 (11), 3050–3057. doi:10.1093/hmg/ddv057

Ohlsson, A., Guthenberg, C., and Döbeln, U. v. (2011). "Galactosemia screening with low false-positive recall rate: The Swedish experience," in *JIMD reports-case and research reports* (Springer), 113–117.

Pasquali, M., Yu, C., and Coffee, B. (2018). Laboratory diagnosis of galactosemia: A technical standard and guideline of the American College of medical genetics and Genomics (ACMG). *Genet. Med.* 20 (1), 3–11. doi:10.1038/gim.2017.172

Patti, G. J., Yanes, O., and Siuzdak, G. (2012). Innovation: Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell. Biol.* 13 (4), 263–269. doi:10.1038/nrm3314

Peduto, A., Spada, M., Alluto, A., Dolcetta, M. L., Ponzone, A., and Santer, R. (2004). A novel mutation in the GLUT2 gene in a patient with Fanconi-Bickel syndrome detected by neonatal screening for galactosaemia. *J. Inherit. Metab. Dis.* 27 (2), 279–280. doi:10.1023/b:boli.0000028841.00833.f4

Sahoo, T., Thukral, A., Agarwal, R., and Sankar, M. J. (2015). Galactosaemia: An unusual cause of chronic bilirubin encephalopathy. *BMJ Case Rep.* 2015, bcr2014206852. doi:10.1136/bcr-2014-206852

Schweigmann, H., Sánchez-Guijo, A., Ugele, B., Hartmann, K., Hartmann, M. F., Bergmann, M., et al. (2014). Transport of the placental estriol precursor 16α-hydroxy-dehydroepiandrosterone sulfate (16α-OH-DHEAS) by stably transfected OAT4-SOAT- and NTCP-HEK293 cells. *J. Steroid Biochem. Mol. Biol.* 143, 259–265. doi:10.1016/j.jsbmb.2014.03.013

Sharma, A., and Nagalli, S. (2021). "Chronic liver disease," in *StatPearls* (Florida: StatPearls Publishing). [Internet].

Succoio, M., Sacchettini, R., Rossi, A., Parenti, G., and Ruoppolo, M. (2022). Galactosemia: Biochemistry, molecular genetics, newborn screening, and treatment. *Biomolecules* 12 (7), 968. doi:10.3390/biom12070968

Taylor Fischer, S., Frederick, A. B., Tran, V., Li, S., Jones, D. P., and Fridovich-Keil, J. L. (2019). Metabolic perturbations in classic galactosemia beyond the Leloir pathway: Insights from an untargeted metabolomic study. *J. Inherit. Metab. Dis.* 42 (2), 254–263. doi:10.1002/jimd.12007

Timson, D. J. (2016). The molecular basis of galactosemia—past, present and future. *Gene* 589 (2), 133–141. doi:10.1016/j.gene.2015.06.077

van Weeghel, M., Welling, L., Treacy, E. P., Wanders, R. J., Ferdinandusse, S., and Bosch, A. M. (2018). Profiling of intracellular metabolites produced from galactose and its potential for galactosemia research. *Orphanet J. Rare Dis.* 13 (1), 146–147. doi:10.1186/s13023-018-0888-1

Veiga-da-Cunha, M., Van Schaftingen, E., and Bommer, G. T. (2020). Inborn errors of metabolite repair. *J. Inherit. Metab. Dis.* 43 (1), 14–24. doi:10.1002/jimd.12187

Vukelic, Z., Metelmann, W., Müthing, J., Kos, M., and Peter-Katalinic, J. (2001). Anencephaly: Structural characterization of gangliosides in defined brain regions. *Biol. Chem.* 382, 259–274. De Gruyter. doi:10.1515/BC.2001.033

Walter, J. H., and Fridovich-Keil, J. L. (2019). "Galactosemia," in *The online metabolic and molecular bases of inherited disease*. Editors D. L. Valle, S. Antonarakis, A. Ballabio, A. L. Beaudet, and G. A. Mitchell (New York, NY: McGraw-Hill Education), 381.

Welsink-Karssies, M. M., Ferdinandusse, S., Geurtsen, G. J., Hollak, C. E. M., Huidekoper, H. H., Janssen, M. C. H., et al. (2020). Deep phenotyping classical galactosemia: Clinical outcomes and biochemical markers. *Brain Commun.* 2 (1), fcaa006. doi:10.1093/braincomms/fcaa006

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., et al. (2022). HMDB 5.0: The human metabolome database for 2022. *Nucleic Acids Res.* 50 (D1), D622–d631. doi:10.1093/nar/gkab1062

Worley, B., and Powers, R. (2013). Multivariate analysis in metabolomics. *Curr. Metab.* 1 (1), 92–107. doi:10.2174/2213235X11301010092

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 13 April 2023
DOI 10.3389/fmolb.2023.1155564

OPEN ACCESS

EDITED BY
Jens Herbig,
Ionicon Analytik, Austria

REVIEWED BY
Marta Laranjo,
University of Evora, Portugal
Philipp Sulzer,
Ionicon Analytik, Austria

*CORRESPONDENCE
Brian Farneti,
brian.farneti@fmach.it

RECEIVED 31 January 2023
ACCEPTED 03 April 2023
PUBLISHED 13 April 2023

CITATION
Farneti B, Khomenko I, Ajelli M, Wells KE, Betta E, Aprea E, Giongo L and Biasioli F (2023), Volatilomics of raspberry fruit germplasm by combining chromatographic and direct-injection mass spectrometric techniques. *Front. Mol. Biosci.* 10:1155564. doi: 10.3389/fmolb.2023.1155564

Volatilomics of raspberry fruit germplasm by combining chromatographic and direct-injection mass spectrometric techniques

Brian Farneti[1]*, Iuliia Khomenko[2], Matteo Ajelli[1], Karen Elizabeth Wells[1], Emanuela Betta[2], Eugenio Aprea[2,3], Lara Giongo[1] and Franco Biasioli[2]

[1]Berries Genetics and Breeding Unit, Research and Innovation Centre of Fondazione Edmund Mach, Trento, Italy, [2]Sensory Quality Unit, Research and Innovation Centre of Fondazione Edmund Mach, Trento, Italy, [3]Center Agriculture Food Environment C3A, University of Trento, Trento, Italy

The application of direct-injection mass spectrometric (DI-MS) techniques, like Proton Transfer Reaction Time of Flight Mass Spectrometry (PTR-ToF-MS) has been suggested as a reliable phenotyping tool for fruit volatilome assessment in both genetic and quality-related studies. In this study the complexity of raspberry aroma was investigated by a comprehensive untargeted VOC analysis, done by combining SPME-GC-MS and PTR-ToF-MS assessments with multi-block discriminant analysis using the DIABLO mixOmics framework. The aim was to acquire an exhaustive characterization of the raspberry volatilome according to different fruit ripening stages (pink, ripe, and overripe) and genetic variances (50 accessions), as well as to investigate the potential of PTR-ToF-MS as a rapid and high throughput VOC phenotyping tool to address issues related to raspberry fruit quality. Results of this study demonstrated the complementarity between SPME-GC-MS and PTR-ToF-MS techniques to evaluate the raspberry aroma composition. PTR-ToF-MS generates reliable raspberry VOC fingerprints mainly due to a reduced compound fragmentation and precise content estimation. In addition, the high collinearity between isomers of monoterpenes and norisoprenoids, discovered by GC analysis, reduces the main analytic limitation of PTR-ToF-MS of not being able to separate isomeric molecules. The high similarity between the VOC matrices obtained by applying PTR-ToF-MS and SPME-GC-MS confirmed the possibility of using PTR-ToF-MS as a reliable high throughput phenotyping tool for raspberry volatiolome assessment. In addition, results provided by the germplasm collection investigation enabled to distinguish the best performing accessions, based on VOCs composition, to be used as superior parental lines for future breeding programs.

KEYWORDS

Rubus idaeus L., PTR-ToF-MS, SPME-GC-MS, germplasm, VOCs, phenotyping, quality

1 Introduction

Volatile organic compound (VOC) directly related to odour composition is one of the main attributes that determines consumers' overall liking of foods compared to other sensory attributes, such as appearance and texture (Klee and Tieman, 2018; Lytou et al., 2019). Therefore, several VOCs are considered key target molecules to improve the flavour perception of fruit and vegetables and, consequently, their marketability. In addition, the end of the "flavour life", mainly due to changes in VOC composition, often precedes the end of shelf life as determined by external modifications (Mazzucotelli et al., 2022). Thus, VOCs should be considered a central trait to determine the optimal cultivation and storage strategies, especially for products for which repeated purchasing behavior and willingness to pay are associated with positive eating experience.

High priority should be given to replace poor flavour cultivars with favorable ones, exploiting the available natural variability. However, the VOC assessment of a very high number of samples, which is necessary to overcome the usually massive biological and genetic variability among fruit samples, may be laborious and time consuming. The so-called "phenotyping bottleneck", caused by the absence of high-throughput and non-invasive methodologies, hampers an effective VOC assessment of broad plant collections (Mazzucotelli et al., 2022)). In fact, the use of conventional gas chromatographic (GC) techniques for VOC analysis presents many analytical limitations. First of all, the process of sample preparation and analysis is time consuming, and second, there are limitations of running exhaustive, complex experimental designs due to the length of analysis. The application of direct-injection mass spectrometric (DI-MS) techniques, like Proton Transfer Reaction Time of Flight Mass Spectrometry (PTR-ToF-MS) or Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), has been demonstrated to be a powerful phenotyping tool for destructive and non-destructive fruit volatilome assessment in both genetic and quality-related studies (Farneti et al., 2017a; Farneti et al., 2020; Di Guardo et al., 2021; Li et al., 2021). Depending on the analytical technique, volatilome assessment can contribute to the discrimination of samples in clusters (Farneti et al., 2017a; Farneti et al., 2020; Di Guardo et al., 2021), the identification of specific biomarkers that are crucial for quality, and the prediction of aspects associated with food quality (Farneti et al., 2013; Bianchi et al., 2020; Li et al., 2021).

VOCs play a key role in the formation of the well-recognized and commonly appreciated aroma of soft fruits (Du and Qian, 2010). In particular, aroma is one of main factors impacting raspberry (*Rubus idaeus* L.) fruit quality and consumer appreciation (Valdés García et al., 2020; Zhang et al., 2021). Raspberry VOC profile of distinct genotypes can be qualitatively and quantitatively very different, and it can be influenced by both agronomic practices and microorganism interactions (Durán-Soria et al., 2021; Sangiorgio et al., 2021; Sangiorgio et al., 2022). Raspberry aroma is a complex blend of almost 300 VOCs with major classes of compounds being terpenes, C13-norisoprenoids, esters, alcohols, aldehydes acids, and ketones (Aprea et al., 2015). However only few of these VOCs have been recognized as major aroma-active compounds in raspberry (Larsen et al., 1991; Aaby et al., 2019; Zhang et al., 2021). Among them, Zhang et al. (2021) identified 14 major aroma-active compounds (volatiles with OAVs ≥1): namely, hexanal, (Z)-3-hexenal, (E)-2-hexenal, eucalyptol, (Z)-3-hexen-1-ol, 1-octen-3-ol, linalool, benzyl alcohol, theaspirane, *ß*-damascenone, dihydro-β-ionone, *a*-ionone, *ß*-ionone, and naphthalene. Additionally, several VOCs, such as benzaldehyde, 1-hexanol, 2-nonanone, (E)-2-hexenal, and (Z)-3-hexenol and monoterpenes, are reported to have inhibitive effects against fungi, especially against *Botrytis cinerea* (Vaughn et al., 1993; Aprea et al., 2010).

Improving fruit flavour, selecting cultivars with premium characters, provides unique challenges in raspberry breeding (Paterson et al., 2013). Selection oriented mostly on productive traits, such as yield, size and shelf-life, have had negative effects on fruit flavour for several fruit species during domestication (Aharoni et al., 2004; Goff and Klee, 2006; Farneti et al., 2017a). Therefore, precise methodological strategies are necessary to support the raspberry breeding activity. This needs a greater comprehension of the genetic control of the pathways involved in VOCs synthesis, environmental factors influencing VOCs production and contributions of impact VOCs to sensory character in raspberry.

In this study the complexity of raspberry aroma was explored by an exhaustive untargeted volatilomic analysis, done by combining SPME/GC-MS and PTR-ToF-MS assessments with multivariate statistic models. The aim of this study was to obtain a thorough characterization of the raspberry volatilome according to different fruit ripening stages and genetic differences, as well as to investigate the potential of PTR-ToF-MS as a rapid and high throughput VOC phenotyping tool to address issues related to raspberry fruit quality.

2 Material and methods

2.1 Plant material and fruit sampling

Raspberry (*R. idaeus* L.) accessions (Supplementary Table S1) were chosen from the experimental field of Fondazione Edmund Mach (FEM) Research and Innovation Center at Pergine (Trento), located in the north of Italy (Trentino Alto Adige region- 46.0744°N, 11,2334°E, 525 m a.s.l.). Plants of raspberry were all grown in 7 L pots, under a hail net and were maintained following standard pruning and agronomical practices (Giongo et al., 2019). Berries were harvested manually at the required stage for each experiment early in the morning and brought to the laboratory within 30 min after picking. Homogeneous fruit, free from external damages or irregularities, were immediately frozen with liquid nitrogen.

For the first experiment, we aimed at acquiring a comprehensive untargeted characterization of the raspberry volatilome according to fruit ripening stages. We employed six raspberry cultivars, namely, "Citria," "Glen Ericht," "Himbotop," "Kweli", "Paris", and "Tulameen". Fruit was collected at three ripening stages based on visual colour evaluation (Pink -P, Ripe-R, and Overripe -OR; Monsalve et al., 2022. "OR" ripening stage corresponds, approximately, to 24 h harvest delay from "R" ripening stage) and analysed by SPME/GC-MS and PTR-ToF-MS.

For the second experiment, aimed at: i) testing the analytical capacity of PTR-ToF-MS as a rapid and high throughput VOC phenotyping tool, and ii) estimating the genetic variability among raspberry cultivars, we employed 50 accessions of raspberry

(34 cultivars and 16 advance selections from the FEM raspberry breeding program; Supplementary Table S1). Fruit was harvested at full ripe stage, which corresponds to the OR commercial maturity stage described above.

2.2 Sample preparation for VOC analysis

Replicates of 0.5 g of raspberry frozen grinded samples, stored at −80°C, were weighed into 20 mL glass vials with screw cap with PTFE/silicone septum (Agilent, Cernusco sul Naviglio, Italy). 0.5 mL of an antioxidant solution (0.5 mL of deionized water, 200 mg of sodium chloride, 2.5 mg of ascorbic acid, and 2.5 mg of citric acid) was added on top according to Farneti et al., 2017b.

2.3 VOC analysis by SPME/GC-MS

Gas chromatographic analysis was performed according to Farneti et al., 2017b. The vials were incubated at 40°C for 10 min constantly stirring before being analysed. Solid-phase microextraction fiber (DVB/CAR/PDMS, 2 cm coating Supelco, Bellefonte, PA, United States) was exposed for 30 min in the vial headspace. Compounds adsorbed by HS-SPME were analysed with a GC interfaced with a mass detector operating in electron ionization (EI) mode (internal ionization source; 70 eV) with a scan range of m/z 33–350 (GC Clarus 500, PerkinElmer, Norwalk CT, United States). HP-INNOWax fused silica capillary column (30 m, 0.32-mm ID, 0.5-μm film thickness; Agilent Technologies, Palo Alto, CA, United States) was used for separation. The initial GC oven temperature was 40°C rising to 220°C at 4°C min^{-1}, the temperature of 220°C was maintained for 1 min, then increased at 10°C min^{-1} until it reached 250°C, which was maintained for 1 min. Helium as a carrier gas was kept at a constant column flow rate of 1.5 mL min^{-1}. Samples were analysed in triplicates. The content of each compound was expressed as μg L^{-1} equivalent of 2-octanol. Compound identification was based on mass spectra matching with the standard NIST/EPA/NIH (NIST 14) and Wiley 7th Mass Spectral Libraries, and linear retention indices (LRI) compared with the literature. LRI were calculated under the same chromatographic conditions after injection of a C7–C30 n-alkane series (Supelco).

2.4 VOC analysis by PTR-ToF–MS

Measurements were performed with a commercial PTR-ToF–MS 8000 apparatus (Ionicon Analytik GmbH, Innsbruck, Austria; Farneti et al., 2017b). The drift tube conditions were as follows: 110°C drift tube temperature, 2.3 mbar drift pressure, 550 V drift voltage with E/N ratio of about 140 Townsend (Td), with E corresponding to the electric field strength and N to the gas number density (1 Td = 10–17 Vcm^2). The acquisition rate of ToF mass spectrometer was 1 spectrum s^{-1} with a mass spectrum ranging up to m/z = 400. The sample headspace was drawn into PTR-MS inlet with 40 sccm flow for 1 min. Flushing of a vial with pure nitrogen during sampling prevented pressure drop inside it. 20 min of sample incubation at 40°C, 1 min of measurement and waiting for 2 min between each measurement were automated by an adapted GC autosampler (MPS Multipurpose Sampler, GERSTEL) coupled to PTR-ToF-MS. The analysis of PTR-ToF-MS spectra proceeded as described in Farneti et al., 2017b.

2.5 Data and statistical analysis

The list of mass peaks detected with PTR-ToF-MS was reduced by applying noise and correlation coefficient thresholds. The first step removed peaks not significantly different from blank samples; the second excluded peaks with over 99% correlation, corresponding for the most part to isotopes of monoisotopic mass peaks (Farneti et al., 2017b).

The internal statistical functions of R.3.4.1. (R Foundation for Statistical Computing, Vienna, Austria) and the external packages "mixOmics", and "ggplot2" were used for the multivariate statistical methods applied in this study.

3 Results and discussion

3.1 Raspberry VOC modification during fruit ripening

3.1.1 SPME-GC-MS headspace analysis

The gas chromatographic analysis by SPME/GC-MS assessed on the headspace of six raspberry cultivars ("Citria," "Glen Ericht," "Himbotop," "Kweli", "Paris", and "Tulameen") harvested at different ripening stages (pink, ripe, over ripe) allowed the detection of 96 VOCs, among which seven were not identified (reported as "Unknown"; Supplementary Table S2). Aldehydes and monoterpenes were the most represented chemical classes, with 16 identified compounds for each chemical class. The other classes of compounds detected in raspberry fruit are alcohols (11 compounds), ketones (11), norisoprenoids (9), esters (8), acids (6), lactones (4), sesquiterpenes (3), hydrocarbons (2), sulfur (1), and furans (1). Based on VOC relative concentration, the ratio between VOC classes differed both regarding fruit ripening and cultivars. Similar to most other fruit species, the production of volatiles in raspberry is integrated with the ripening process with volatile profiles changing dramatically during ripening.

During the complete fruit ripening process, an important VOC profile portion is covered by monoterpenes, norisoprenoids and ketones, while for pink fruit around 70% of the VOC profile is covered by aldehydes and alkenes (Figure 1A). These modifications are strictly cultivar dependent. Indeed, the relative abundance of ketones, monoterpenes, norisoprenoids and sesquiterpenes raised to more than 70% of the overall VOC profile for the cultivars "Citria", "Himbotop", "Kweli" and "Tulameen", while fruit of "Gleen Ericht" and "Paris" did not show a so drastic increment (Figure 1A; Supplementary Figure S1).

Most of the identified VOCs significantly differed both between cultivars and especially between ripening stages (Supplementary Table S2; Supplementary Figure S2). Only few compounds decreased during maturation, like 3-methyl-2-hexene and (E)-2-

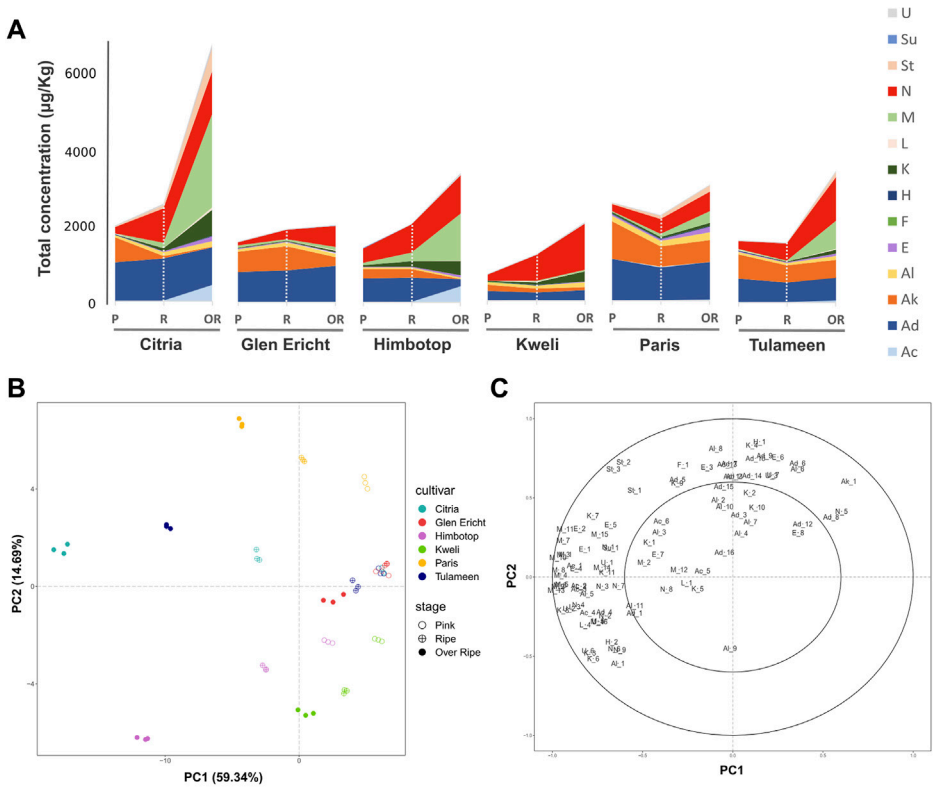


FIGURE 1
Evolution of raspberry VOC profile during fruit ripening assessed by SPME/GC-MS analysis. The stacked area chart reported in the plot **(A)** illustrates the total VOC concentration, expressed as µg/L of 2-octanol, for each raspberry cultivar ("Citria", "Glen Ericht", "Himbotop", "Kweli", "Paris", and "Tulameen") at three ripening stages [pink (P), ripe (R), and overripe (Or)]. Each VOC classes [acids (Ac), aldehydes (Ad), alkenes (Ak), alcohols (Al), esters (E), furans (F), hydrocarbons (H), ketones (K), lactones (L), monoterpenes (M), norisoprenoids (N), sesquiterpenes (St), sulfurs (Su), unknowns (U)] is described with a different color. Plot **(B)** depicts the VOC profile of the raspberry cultivars during fruit ripening over the PCA score plot defined by the first two principal components. Plot **(C)** shows the PCA loading plot of the VOCs identified by SPME/GC-MS analysis.

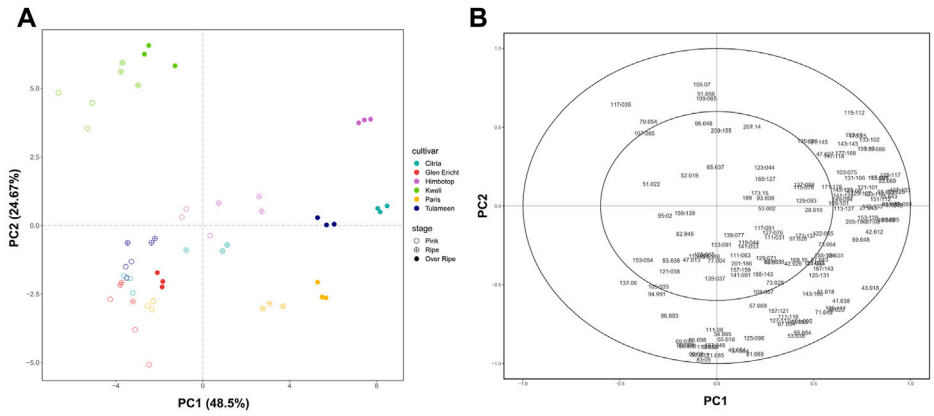


FIGURE 2
Evolution of raspberry VOC profile during fruit ripening assessed by PTR-ToF-MS analysis Plot **(A)** depicts the VOC profile of each raspberry cultivar ("Citria", "Glen Ericht", "Himbotop", "Kweli", "Paris", and "Tulameen") at three ripening stages [pink (P), ripe (R), and overripe (Or)] over the PCA score plotdefined by the first two principal components. Plot **(B)** shows the PCA loading plot of the VOCs identified by PTR-ToF-MS analysis.

hexenal. Compounds that were unaffected by the ripening stage of the fruit are (E)-2-hexenal, octanal, 2-heptanol, (Z)-3-hexen-1-ol, 2-nonanal, 1-heptanol, 2-ethyl-1-hexanol, decanal, benzaldehyde, undecanal, non-anoic acid. In contrast, the concentration of most other compounds increased during ripening especially in fully ripe fruit (OR).

For a rapid and unsupervised data exploration, SPME/GC-MS results were analysed with principal component analysis (PCA). The biplot of the first two PC scores and the loading plot are presented in Figures 1B, C. These VOC profile evolution during fruit ripening is evident in the PCA reported in Figure 2B where nearly 60% of the variability among the fruit VOC profiles (PC1) is strictly associated with the stage of fruit ripening. The variability expressed by the second component (PC2: 15%) is more inherent to differences between accessions. According with these PCA results, the greatest differences among cultivars are detectable among fruits at the most advanced stage of ripening (OR). These differences are mostly related to different concentrations of monoterpenes, norisoprenoids, ketones, sesquiterpenes, and acetate esters (Figure 2C).

The results obtained on these six cultivars support previously published findings on ripening studies performed with different analytical techniques on other raspberry cultivars (Guichard, 1984; Larsen et al., 1991). Guichard (et al., 1984) followed the VOC evolution during ripening in two raspberry cultivars (*R. idaeus* cv. "Lloyd George" and "Rose de Côte d'Or") by analysing the fruit concentrated extracts by gas chromatography. In both cultivars all the terpenes and sesquiterpenes significantly increased during ripening. However, dissimilarly from our results, the ripe and overripe stages were not significantly different. Several acetate esters, like isopentyl-, pentenyl-, (Z)-3-hexenyl- and methyl-acetate, also increased up to 100 fold during ripening. Norisoprenoids had different trends according to the cultivar. Dihydro-β-ionone was at its highest at the ripe stage then decreased. *a*-Ionone moderately increased during fruit ripening in both varieties, while *ß*-ionone reasonably increased only in fruit of the cultivar "Lloyd George" and not at all in "Rose de Côte d'Or". Larsen and co-workers (1991) compared the VOC evolution during the ripening of 10 cultivars (cv. "Camenzind", "Chilcotin", "Glen Prosen", "Glen Moy", "Glen Clova", "Meeker", "Rutrago", "Skeena", "Vaten" and "Zenith"). Concentrated extracts have been analysed using GC with FID detector. They reported relatively great differences in the concentrations of linalool, geraniol, benzyl alcohol, acetoin, acetic acid, and hexanoic acid during ripening. However, only very tiny differences in norisoprenoid concentrations were observed between the cultivars.

Terpenes, terpenoids and norisoprenoid volatile compounds are the major compounds that have been detected for the differentiation of raspberry genotypes (Larsen et al., 1991; Malowicki et al., 2008) as they are highly related to raspberry odour and flavour. The increase of monoterpenes (i.e., *a*-pinene, *a*-phellanderene, *ß*-Phellandrene, o-Cymene, 4-Terpineol) and sesquiterpenes (i.e., trans-Caryophyllene and *a*-Humulene) highlighted by our results can be associated with a different regulation of the isopentenyl diphosphate (IDP) and its isomer dimethylallyl diphosphate (DMADP) synthesis (Paterson et al., 2013; Pazouki and Niinemets, 2016). The formation of both IDP and DMADP is led by the mevalonate (MVA) and the methylerythritol 4-phosphate (MEP) pathways that are active, respectively, in the cytosol and in the plastid of fruit cells. It is widely acknowledged that monoterpenes are synthesized in the plastids whereas sesquiterpenes in the cytosol (Rohmer, 2003). The total content of terpenes and sesquiterpenes among different cultivars, in fact, did not show any significant correlation in our study (Supplementary Figure S3).

Most of the norisoprenoids detected in this study increased during fruit ripening, similarly to monoterpenes. Only damascenone, detected at very low concentration, slightly decreased during fruit ripening. Norisoprenoids, recognized to be important contributors to raspberry fruit aroma, are generated by oxidative cleavage of the carotenoids (Hampel et al., 2007). In particular, the two ionone stereoisomers, *a*-Ionone and *ß*-ionone, are responsible for raspberry-violet-rose fragrance notes (Breitmaier, 2006). These compounds are respectively derived from the degradation of *a* and *ß*-carotene (Paterson et al., 2013). This could be the reason why their content is not correlated in our study (Supplementary Figure S3). Unidentified compounds 4 and 5, based on their fragmentation spectrum (Supplementary Figure S4) and their high correlation with cycloionone (r^2:0.98; Supplementary Figure S2), can be classified as norisoprenoids with high probability. Other norisoprenoids highly correlated with cycloionone are *ß*-ionone (r^2:0.89) and dihydro- *ß*-ionone (r^2:0.85). The high variability in norisoprenoid content between raspberry cultivars is in agreement with Malowicki et al. (2008) and Paterson et al. (2013), both of which reportedlarge variations in *a*-ionone, *ß*-ionone in different raspberry genotypes.

Ketones are other important compounds for characterizing the aromatic profile of raspberries, particularly 2-nonanone, 2-heptanone, 2-undecanone, and 5-nonen-2-one are responsible for a "fruity" and "cheesy" flavour of fruit (http://www.thegoodscentscompany.com/). The concentration of these molecules increased during ripening, depending on the considered cultivar. The cultivars with higher concentrations of ketones are "Citria", "Himbotop" and "Kweli". As also reported previously by Aprea et al. (2009), it was not possible to quantify one of the most characteristic raspberry ketone, raspberry ketone (4-(4-hydroxyphenyl)butan-2-one), since it is not easily detectable without a chemical extraction because of its low volatility.

3.1.2 Direct injection VOC profiling by PTR-ToF-MS

Fruit samples analysed by PTR-ToF-MS were prepared in the same way to the ones used for SPME/GC-MS analysis in order to compare the outcomes of these two methodologies. The whole VOC spectra, assessed in triplicate for each cultivar, were reduced to 148 VOC mass peaks (Supplementary Table S3; Supplementary Figure S5), applying noise and correlation coefficient thresholds. Tentative identification (t.i.) of each mass peak detected by PTR-ToF-MS relied on an in-house library of chemical standards, on the list of compounds detected by SPME/GC-MS analysis, and on compounds reported in the review paper of Aprea et al. (2015). The content of 146 mass peaks was significant different between cultivars, while 112 mass peaks significantly differed between ripening stages (Supplementary Table S3). Among them, only few mass peaks decrease during fruit ripening. These mass peaks, like *m/z* 99.080, *m/z* 81.069, *m/z* 57.033, or *m/z* 43.054, are related to the fragmentation of several c6 aldehydes. The headspace analyses carried out with PTR-ToF-MS allow for the valuation of VOCs that are often omitted from the ordinary gas chromatographic assessments despite their importance for the characterization of fruit quality and freshness, such as methanol (*m/z* 33.033), ethanol

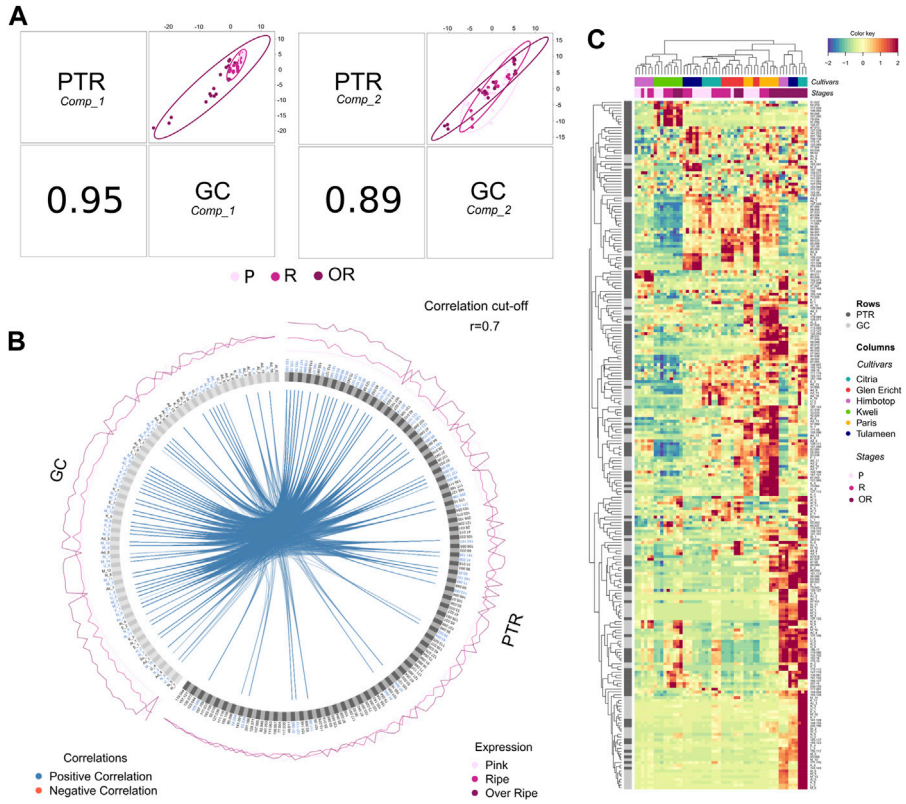


FIGURE 3
Complementarity between SPME/GC-MS and PTR-ToF-MS headspace VOC assessments defined by applying the DIABLO multi-block discriminant analysis. Plot **(A)** depicts a global overview of the correlation structure at the component level. This function allows for the plotting of the components (component 1 and component 2) across the different data sets for a given dimension. Plot **(B)** reports the correlation circular plot between VOCs assessed by SPME-GC-MS and PTR-ToF-MS built based on a similarity matrix. A cut-off level of 0.7 was arbitrarily included to visualize correlation coefficients above this threshold. Plot **(C)** reports the "Clustered Image Map" representing the multi-omics molecular signature expression for each sample. Each cell's colour is based on the values of the similarity matrix performed on the two VOC datasets.

(*m/z* 47.049), acetaldehyde (*m/z* 45.033), or dimethyl sulfide (*m/z* 63.026).

PCA analysis (Figure 2) was carried out to describe differences amongst the raspberry VOC profiles considering both cultivars and ripening stages. 73% of the total variation accounted for the first two principal components (Figure 2A). Likewise in the SPME/GC-MS analysis, differences between fruit sampled at different ripening stages were largely accountable to the first principal component (PC1: 48.5%), while the second component (PC2: 24.7%) predominantly defined differences between cultivars. Based on the correlation circle plot obtained from the PCA analysis, only few mass peaks have a weak association (lower than 0.6) with the first two principal components (Figure 2B). These mass peaks are either present at trace concentration levels or with a non-significant ($p > 0.05$) 'cultivar x stage' interaction (Supplementary Tablee S3). In accordance with SPME/GC-MS analysis, the greatest differences among cultivars are detectable among fruit at the most advanced stage of ripening (OR). These differences between VOC profiles are mainly related to the different concentrations of mass peaks related to monoterpenes (i.e., *m/z* 137.133), norisoprenoids (i.e., *m/z* 191.145, *m/z* 193.16, *m/z* 195.17), sesquiterpenes (i.e., *m/z* 205.196), ketones (i.e., *m/z* 115.112, *m/z* 141.128, *m/z* 143.143), acetate esters (i.e., *m/z* 75.043, *m/z* 89.059), aldehydes (i.e., *m/z* 45.032, *m/z* 73.064, m/z 85.064, *m/z* 87.08, *m/z* 95.085, *m/z* 99.08), and alcohols (i.e., *m/z* 33.033, *m/z* 47.048, *m/z* 83.085). These results agree with the studies of Aprea et al. (2009); Cappellin et al. (2013), in which the VOC profile of some raspberry cultivars was evaluated on intact berries with PTR-MS equipped with a quadrupole mass detector. The main advantages of using PTR-MS equipped with a ToF analyzer are the enhanced mass resolution, allowing the separation of many isobaric compounds and the simultaneous monitoring of multiple peaks at the same nominal mass, and the enhanced speed of analysis since the whole VOC spectrum is acquired in a fraction of a second (Mazzucotelli et al., 2022).

3.2 Complementarity between SPME/GC-MS and PTR-ToF-MS headspace VOC assessments

The results obtained with the two analytical techniques were compared and combined in order to verify both the complementarity of the two methodologies and the possibility of using PTR-ToF-MS as a fast, comprehensive and reliable VOC phenotyping tool for raspberry fruit, despite its known analytical limitations in separating and identifying isomers. In order to identify

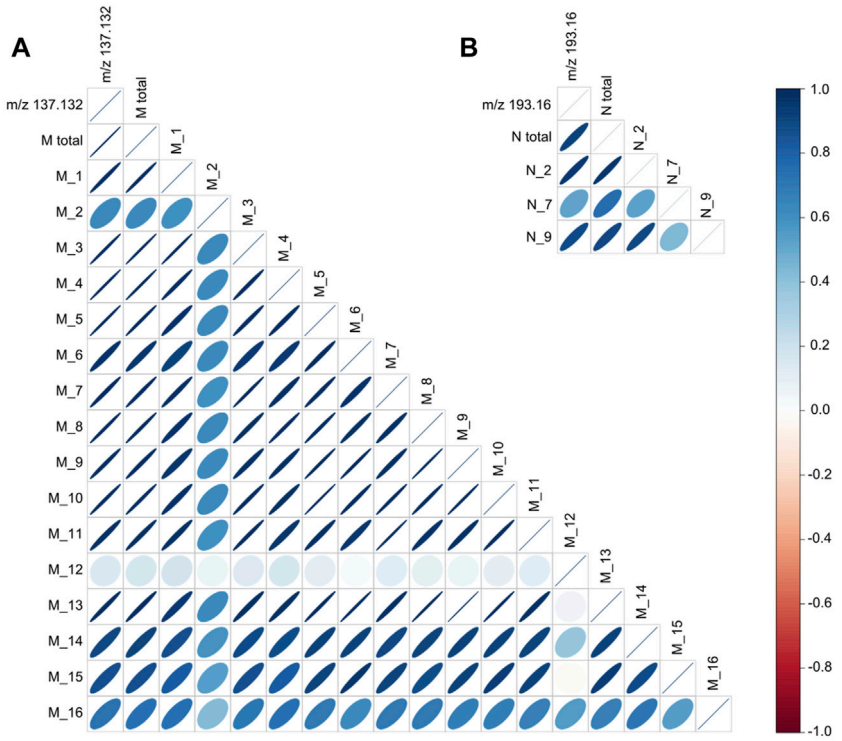


FIGURE 4
Correlation analysis of monoterpenes and norisoprenoids assessed by using SPME/GC-MS and PTR-ToF-MS. Plot **(A)** reports the correlation analysis between the content of the main PTR-ToF-MS mass peak associated with monoterpenes (*m/z* 137.132), the total amount of monoterpenes ("M total"), and the content of each monoterpene compounds (From M_1 to M_16) detected with SPME-GC-MS. Plot **(B)** reports the correlation analysis between the content of the main PTR-ToF-MS mass peak associated with norisoprenoids (*m/z* 193.16), the total amount of norisoprenoids associated with this mass ("N total"), and the content of each norisoprenoid compounds associated with the mass peak m/z 193.16 (N_2, N_7, and N_9) detected with SPME-GC-MS.

a highly correlated VOC signature discriminating known groups of samples (cultivar and/or ripening stages), we applied a multi-block discriminant analysis using the DIABLO mixOmics framework [Data Integration Analysis for Biomarker discovery using a Latent cOmponents (Singh et al., 2019)]. The core DIABLO method extends Generalised Canonical Correlation Analysis, which generalises PLS for multiple matching datasets, and the sparse sGCCA method (Singh et al., 2019). Like PLS, DIABLO generates a pair of components, each associated to each VOC data set. A global overview of the correlation structure at the component level is represented in Figure 3A by using the plotDiablo function. This function allows for the plotting of the components across the different data sets for a given dimension. The results revealed a high correlation between the data obtained by SPME/GC-MS and PTR-ToF-MS, resulting in a correlation of 0.95 and 0.89 for the first and second components, respectively. As reported by previous PCA analyses (Figures 1, 2), values of the first component are more associated with differences in VOC profile due to fruit maturity stages, while in the second component there are more associations with differences between cultivars.

The high similarity between the two matrices is explainable by the great correlation between the content of individual molecules acquired by the two analytical techniques. This result confirms the possibility of using a direct injection mass spectrometry technique (PTR-ToF-MS) for raspberry flavour profile analysis as an alternative to gas chromatographic analysis performed after headspace accumulation (SPME/GC-MS). All the correlations between VOCs assessed by SPME-GC-MS and PTR-ToF-MS are reported in the circos plot (Figure 3B), built based on a similarity matrix. A cut-off level of 0.7 was arbitrarily included to visualize correlation coefficients above this threshold in the multi-omics signature. 60% of the compounds identified and quantified by SPME/GC-MS (58 out of 96) are highly correlated (cut-off of 0.7) with the masses quantified by PTR-ToF-MS (Figure 3B). In contrast, only 32% of the masses considered from the PTR-ToF-MS analyses (48 out of 148) are correlated with the compounds identified by the gas chromatographic analyses. Of these 100 masses without any significant correlation with the gas chromatographic matrix, only a few are present at high concentrations: e.g., *m/z* 33.033 (t.i. methanol), *m/z* 34.995 (t.i. hydrogen sulfide), *m/z* 45.032 (t.i. acetaldehyde), *m/z* 47.048 (t.i. ethanol), *m/z* 107.085 (t.i. ethylbenezene, xylene). These compounds, in fact, cannot be properly identified with the setting of the gas chromatographic methodology applied in this study. Most of the other unrelated masses can be associated with molecules present in raspberry fruit at low concentrations and thus below the LOD of the SPME-GC-MS methodology that we applied. For this reason, the tentative identification of many masses measured by

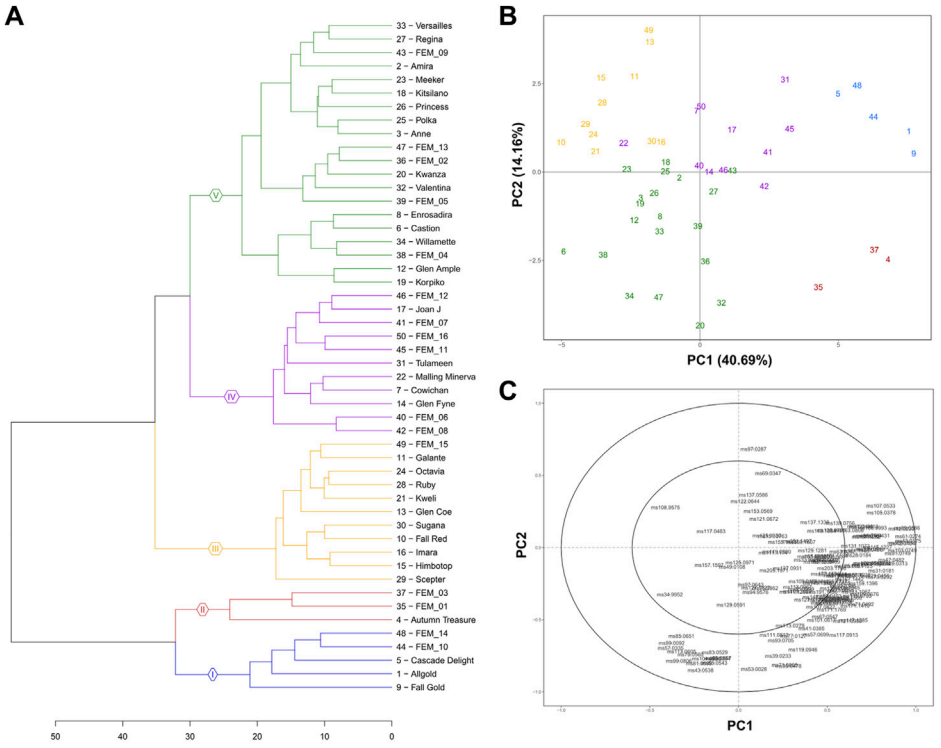


FIGURE 5
Raspberry germplasm volatilome assessed by PTR-ToF-MS. Hierarchical dendrogram **(A)** principal component analysis (PCA) **(B)** and loading plot **(C)** of VOC profiles of 50 raspberry accessions measured by PTR-ToF-MS. Each VOC concentration is the average of three biological replicates. Each colour of the PCA plot depicts one of the five clusters, determined by gap statistics.

PTR-ToF-MS is based on information extrapolated from published studies about raspberry volatilome assessed by applying different analytical techniques (Aprea et al., 2015).

One of the analytical limitations of direct injection mass spectrometry techniques, and thus also of PTR-ToF-MS, is the difficulty in separating and identifying isomers. The VOC profile of raspberry consists of several isomeric compounds important for the aromatic characterization of the fruit, especially in the chemical class of monoterpenes (e.g., *a*-pinene, *a*-phellanderene, *ß*-phellanderene) or norisoprenoids (e.g., *a*-ionone, *ß*-ionone, cycloionone; *ß*-damascenone). As for monoterpenes, the content of *m/z* 137.133 mass is highly correlated with the total monoterpene content (R^2 = 0.99; Figure 4A). However, unexpectedly, the monoterpenes present in raspberry fruit are highly correlated with each other, except for linalool and geraniol, suggesting a very similar biosynthetic pathway regulation of monoterpenes among the raspberry accessions considered in this study. This high collinearity among monoterpenes reduces the limitation of PTR-ToF-MS of not being able to separate isomeric molecules, at least for the raspberry aroma profile. As for the three main norisoprenoids of raspberry fruit, namely, cycloionone, *a*-ionone and *ß*-ionone, their total content is positively correlated with the mass *m/z* 193.16 (R^2 = 0.84, Figure 4B). Unlike what was found with monoterpenes, the correlation is less robust mainly due to the lack of collinearity between the concentrations of *a*-ionone compared to *ß*-ionone and cycloionone. *ß* and *a* ionone are derived from the degradation of *a* and *ß*-carotene (Paterson et al., 2013), respectively. This lack of collinearity between the compounds could result from different carotenoid composition and content among raspberry cultivars.

The collinearity and complementarity of the two analytical techniques used in this study is shown graphically in the Clustered Image Map in Figure 3C. The Clustered Image Map represent the multi-omics molecular signature expression for each sample. Each cell's colour is based on the values of the similarity matrix performed on the two VOC datasets. Dendrograms used along the axes depict how each row (VOCs)/column (cultivars and ripening stages) clusters based on the hierarchical clustering method. Blocks of homogeneous colour depict subsets of features from each dataset which are correlated and suggests a potential causative relationship.

3.3 Phenotyping of raspberry germplasm volatilome

In this trial carried out to both test the applicability of PTR-ToF-MS as a fast phenotyping tool for volatile profiling of raspberry fruit and to verify the variability present within raspberry germplasm, we analysed 50 raspberry accessions (Supplementary Table S1), some of which were collected at different ripening epochs during the season. Of these 50 accessions, 34 were cultivars and 16 were advanced selections from the FEM's breeding program. The whole VOC spectra, assessed in triplicate for each sample, were reduced to

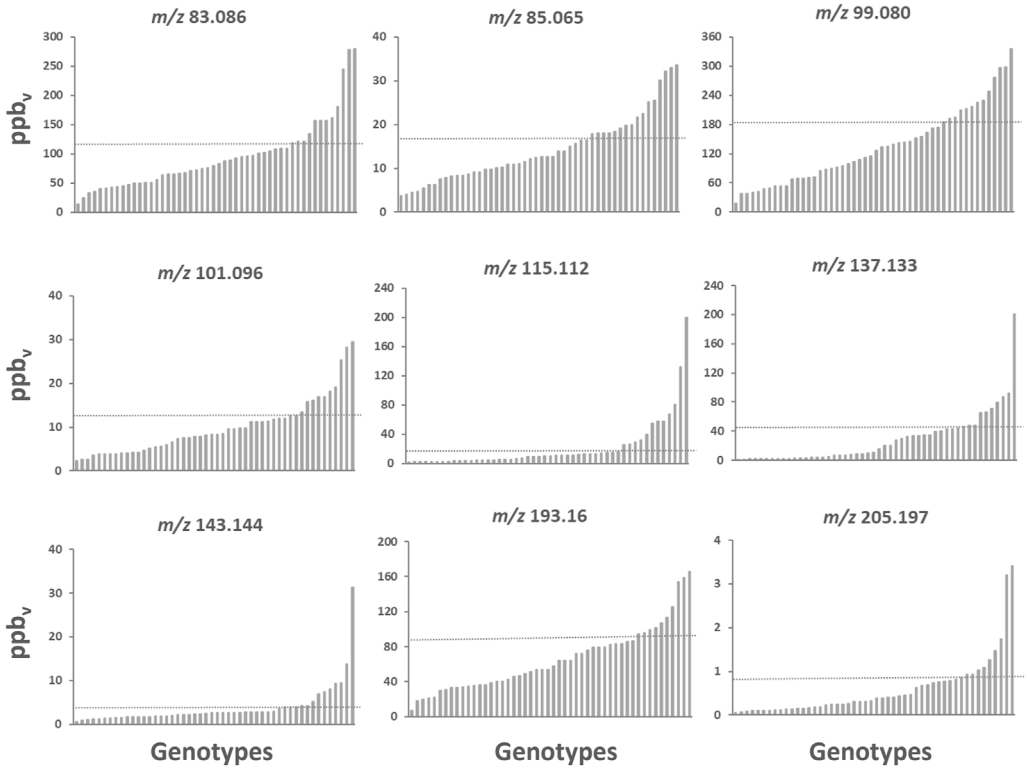


FIGURE 6
Bar chart plots of the main raspberry VOCs assessed by PTR-ToF-MS in the raspberry germplasm collection. Each bar illustrates the average value of three biological replicates. In each graph, accessions were ordered based on the VOC mass peak concentration. The 75th percentile levels are represented by the dashed line. Bar chart plots of all mass peaks with the corresponding complete names of the accessions are reported in the Supplementary Figure S6.

136 VOC mass peaks (Supplementary Table S4), applying noise and correlation coefficient thresholds. The number of mass peaks significantly different from the blank sample is higher than in the previous experiment, most likely because the number of accessions measured is higher and consequently so is the biological variability.

Considering the elevated genetic variability of the germplasm collection employed in this study, we aimed to uncover most of the raspberry natural VOC variability. To avoid any possible statistical bias in the interpretation of results, all data were analysed with unsupervised multivariate statistical methodologies (PCA and hierarchical clustering). Based on the VOCS profile, the 50 accessions were divided into five significantly different clusters (gap statistic) Figure 5A. These five clusters are also distinguishable in the PCA (Figures 5B, C) performed on the raspberry VOC variability defined by the first two PCs, and expressing together 55% of the total variability. According to the correlation loading plot (Figure 5C), the first principal component (PC1), describing 40.7% of the total variability, mainly correlates with VOC concentration magnitude. The second principal component (PC2), instead, resulted mainly related with the VOC chemical composition, describing 14.2% of total aromatic variability. Clusters I and II, in fact, group the 8 accessions with the most intense VOC profile (cluster I: “Fall Gold”, “Allgold”, “Cascade delight”, “FEM_10”, “FEM_14”; cluster II: “Autumn Treasure”, “FEM_01”, “FEM_03”). These two clusters are separated from the remaining 3 clusters (III, IV, V) more by PC1 values. Within cluster V we detected a group of raspberry accessions (“Kwanza”, “Valentina”, “FEM_13”, “Willamette”, “FEM_16”, “FEM_04”, “FEM_05”) that differs from the others by a higher concentration of masses connected with C6 aldehydes (*m/z* 81.069, *m/z* 99.08, *m/z* 101.095) and C6 alcohols (*m/z* 83.085).

Considering the whole VOC profile, although useful for getting a general indication of the possible aromatic fingerprint of a genotype, can still limit the information actually present in the dataset. In particular, we can see that a high statistical weight in the analysis of the PCA of the various accessions is given by molecules present at very high concentrations, two or three orders of magnitude higher than the other compounds, that are, however, not closely related to the characteristic raspberry aromatic profile such as methanol, ethanol or acetaldehyde. However, these molecules are an indicator of the degree of anaerobic fermentation and are often associated with off-flavours, and therefore can be considered as possible biomarkers related to fruit ripeness and shelf life (Beaulieu et al., 1997). Instead, to identify possible biomarkers closely related to the raspberry flavour profile to be considered in qualitative “from farm to fork” studies in our opinion, it is more explanatory to consider the concentration of each mass individually. Figure 6 shows 9 masses (over 136) that can be considered as key elements to describe the raspberry aroma profile according to the loading plots

of the principal component analysis and to the results of previously published articles on raspberry aroma (Aprea et al., 2015; Zhang et al., 2021). High variability was found for each mass within the genotype pool analysed in this study.

Non-etheless, considering each VOC independently (i.e., Supplementary Figure S6) might be useful for the backcross breeding approach, aimed to introduce, or improve, a distinct quality trait to an elite breeding line (Hospital, 2005).

For a more practical application of these results, especially for breeding purposes, the content of each VOC was grouped based on the distribution quantile (low: 0%–25%; middle-low: 25%–50%; middle-high: 50%–75%; high: 75%–100%), calculated for each compound (Supplementary Figure S7). Therefore, all accessions employed in the study can be sorted and clustered according to each VOC mass peak content, which can be arbitrarily chosen, as implemented in the dedicated webpage https://iuliiakhomenko-fmach.shinyapps.io/QualySort_raspberry/.

Considering that no molecular markers are yet available to predict the VOC content of raspberry fruit, the application of reliable phenotyping techniques combined with an array of VOC biomarkers is still essential to support breeding activity.

4 Conclusion

Results of this study confirmed the complementarity between chromatographic and direct-injection spectrometric techniques to study the raspberry aroma composition. The use of PTR-ToF-MS is suitable to generate reliable raspberry VOC fingerprints mainly due to a reduced compound fragmentation and accurate quantification. The high similarity between the VOC matrices obtained by applying PTR-ToF-MS and SPME/GC-MS confirmed the possibility of using a direct injection mass-spectrometry technique (PTR-ToF-MS) as a reliable VOC phenotyping tool in those investigations that require a detailed VOC profile characterization of a large number of raspberry fruit.

A weak aspect of DI-MS methodologies is still represented by compound identification. PTR-ToF-MS can separate many raspberry isobaric compounds; however, many isomers are still not distinguishable without a chromatographic separation. Nevertheless, unexpectedly, most monoterpene and norisoprenoid isomers, important for the raspberry aroma characterization, in this study were highly correlated with each other, suggesting a very similar and solid biosynthetic pathway regulation of both compound classes among the raspberry accessions. This high collinearity, in our opinion, overcomes the main analytic limitation of PTR-ToF-MS of not being able to separate isomeric molecules.

Bearing in mind that the aim of VOC assessment in quality related studies is to obtain an objective estimation of the aroma perceived by the consumer during fruit consumption, we consider unnecessary the application of overly aggressive chemical extraction methodologies (commonly used in several published articles on raspberry VOC profiling). These methodologies are necessary for the quantification of compounds at extremely low concentrations that, in case of raspberry fruit, might be under the perception threshold of the consumer. In addition, results of our study revealed a higher sensitivity of PTR-ToF-MS with respect to SPME/GC-MS, allowing the detection of compounds present in trace amounts.

Pulling together results of the investigations about the role of ripening and of the genetic variability, the array of mass peaks suitable to describe most of raspberry VOC variability can be considerably reduced. This array of VOC biomarkers, in combination with a reliable phenotyping methodology, can be applied for a more targeted VOC assessment for both breeding selection and quality control within the entire production chain. The uncovering of the genetic variability existing within the investigated raspberry germplasm collection allowed us to identify the best performing cultivars, based on VOCs content, to be used as superior parental lines for future breeding programs focused on enhanced fruit quality. In our opinion, a better and more detailed knowledge of the aromatic profile of the fruit is also essential to define the optimal production and storage strategies specific for each genotype. In the case of raspberry, it is evident how anticipated fruit harvesting, which is the common practice to prolong fruit shelf life, may drastically reduce the organoleptic quality of the fruit. These findings should push the forthcoming research activity toward the development of cultivation and storage techniques tailored for enhanced organoleptic quality of the fruit and not just for the external visual quality.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

FB: writing—original draft, investigation, conceptualization, methodology, sample preparation, data analysis. KI: investigation, data analysis, PTR-ToF-MS analysis, writing–review and editing. AM: sample preparation. WE: writing–review and editing. BE: SPME-GC-MS analysis. AE: SPME-GC-MS data analysis, writing–review and editing. GL: conceptualization, funding acquisition, writing–review and editing. BF: supervision, writing–review and editing.

Funding

This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them. This work was financially supported by the AdP of the PAT (Provincia Autonoma di Trento) and by the project AppleBerry (L6/99 of the PAT).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1155564/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Stacked area chart of the relative VOC concentration, expressed as %, for each raspberry cultivar ("Citria," "Glen Ericht," "Himbotop," "Kweli," "Paris," and "Tulameen") at three ripening stages [pink (P), ripe (R), and overripe (Or)]. Each VOC classes [acids (Ac), aldehydes (Ad), alkenes (Ak), alcohols (Al), esters (E), furans (F), hydrocarbons (H), ketones (K), lactones (L), monoterpenes (M), norisoprenoids (N), sesquiterpenes (St), sulfurs (Su), unknowns (U)] is described with a different color.

SUPPLEMENTARY FIGURE S2
Bar graph of the content of each VOC classes, expressed as μg/Kg of 2-octanol, assessed by SPME-GC-MS for each raspberry cultivar ("Citria," "Glen Ericht," "Himbotop," "Kweli," "Paris," and "Tulameen") at three ripening stages [pink (P), ripe (R), and overripe (Or)].

SUPPLEMENTARY FIGURE S3
Correlation plot of volatile organic compounds assessed by SPME/GC-MS.

SUPPLEMENTARY FIGURE S4
Mass spectra of unknown compounds detected by SPME/GC-MS.

SUPPLEMENTARY FIGURE S5
Bar graph of the content of each VOC, expressed as ppbv, assessed by PTR-ToF-MS for each raspberry cultivar ("Citria," "Glen Ericht," "Himbotop," "Kweli," "Paris," and "Tulameen") at three ripening stages [pink (P), ripe (R), and overripe (Or)]. Each value is the average plus standard deviation of three replicates. For each graph lines corresponding to the 25%, 50%, and 75% percentile were added.

SUPPLEMENTARY FIGURE S6
Bar graph of the content of each VOC, expressed as ppbv, assessed by PTR-ToF-MS for each raspberry accession of the raspberry germplasm collection. Each value is the average plus standard deviation of three replicates.

SUPPLEMENTARY FIGURE S7
Print screens of the Shiny web application (Qualysort 1.2; Farneti et al. 2020) with an examples of selected VOC traits (same compounds reported in Figure 6). This Shiny web application (https://iuliiakhomenko-fmach.shinyapps.io/QualySort/) is interactive and gives the opportunity to choose an xlsx file with the dataset of phenotypic traits. It is possible to select the raspberry accessions and the VOCs. Each VOC is classified according to the quartile it belongs to (0–25%, red colour; 25–50%, orange colour; 50–75%, light green colour; 75–100% dark green colour) for each type separately. Button "Calculate" plots the heatmap based on the selected classified data with the hierarchical clustering (Gower's distance and clustering method—average) by samples (rows).

References

Aaby, K., Skaret, J., Røen, D., and Sønsteby, A. (2019). Sensory and instrumental analysis of eight genotypes of red raspberry (rubus idaeus L.) fruits. *J. Berry Res.* 9 (3), 483–498. doi:10.3233/JBR-190387

Aharoni, A., Giri, A. P., Verstappen, F. W., Bertea, C. M., Sevenier, R., Sun, Z., et al. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell. Nov.* 16 (11), 3110–3131. doi:10.1105/tpc.104.023895

Aprea, E., Biasioli, F., Carlin, S., Endrizzi, I., and Gasperi, F. (2009). Investigation of volatile compounds in two raspberry cultivars by two headspace techniques: Solid-phase microextraction/gas Chromatography–Mass spectrometry (SPME/GC–MS) and proton-transfer Reaction–Mass spectrometry (PTR–MS). *J. Agric. Food Chem.* 57 (10), 4011–4018. doi:10.1021/jf803998c

Aprea, E., Biasioli, F., and Gasperi, F. (2015). Volatile compounds of raspberry fruit: From analytical methods to biological role and sensory impact. *Molecules* 20, 2445–2474. doi:10.3390/molecules20022445

Aprea, E., Carlin, S., Giongo, L., Grisenti, M., and Gasperi, F. (2010). Characterization of 14 raspberry cultivars by solid-phase microextraction and relationship with gray mold susceptibility. *J. Agric. Food Chem.* 58 (2), 1100–1105. doi:10.1021/jf902603f

Beaulieu, J. C., Peiser, G., and Saltveit, M. E. (1997). Acetaldehyde is a causal agent responsible for ethanol-induced ripening inhibition in tomato fruit. *Plant Physiol.* 113, 431–439. doi:10.1104/pp.113.2.431

Bianchi, T., Guerrero, L., Weesepoel, Y., Argyris, J., Koot, A., Gratacós-Cubarsí, M., et al. (2020). Linking sensory and proton transfer reaction–mass spectrometry analyses for the assessment of melon fruit (Cucumis melo L.) quality traits. *Eur. Food Res. Technol.* 246, 1439–1457. doi:10.1007/s00217-020-03502-2

Breitmaier, E. (2006). *Terpenes: Flavors, fragrances, pharmaca, pheromones.* Weinheim, Germany: John Wiley and Sons, 48.

Cappellin, L., Aprea, E., Granitto, P., Romano, A., Gasperi, F., and Biasioli, F. (2013). Multiclass methods in the analysis of metabolomic datasets: The example of raspberry cultivar volatile compounds detected by GC–MS and PTR-MS. *Food Res. Int.* 54, 1313–1320. doi:10.1016/j.foodres.2013.02.004

Di Guardo, M., Farneti, B., Khomenko, I., Modica, G., Mosca, A., Distefano, G., et al. (2021). Genetic characterization of an almond germplasm collection and volatilome profiling of raw and roasted kernels. *Hortic. Res.* 8, 27. doi:10.1038/s41438-021-00465-7

Du, X., and Qian, M. (2010). Flavor chemistry of small fruits: Blackberry, raspberry, and blueberry. *January* 1, 27–43. doi:10.1021/bk-2010-1035.ch003

Durán-Soria, S., Pott, D. M., Will, F., Mesa-Marín, J., Lewandowski, M., Celejewska, K., et al. (2021). Exploring genotype-by-environment interactions of chemical composition of raspberry by using a Metabolomics approach. *Metabolites* 11, 490. doi:10.3390/metabo11080490

Farneti, B., Alarcón, A. A., Cristescu, S. M., Costa, G., Harren, F. J. M., Holthuysen, N. T. E., et al. (2013). Aroma volatile release kinetics of tomato genotypes measured by PTR–MS following artificial chewing. *Food Res. Int.* 54 (2), 1579–1588. doi:10.1016/j.foodres.2013.09.015

Farneti, B., Di Guardo, M., Khomenko, I., Cappellin, L., Biasioli, F., Velasco, R., et al. (2017a). Genome-wide association study unravels the genetic control of the apple volatilome and its interplay with fruit texture. *J. Exp. Bot.* 68 (7), 1467–1478. doi:10.1093/jxb/erx018

Farneti, B., Emanuelli, F., Khomenko, I., Ajelli, M., Biasioli, F., and Giongo, L. (2020). Development of a novel phenotypic roadmap to improve blueberry quality and storability. *Front. Plant Sci.* 11, 1140. doi:10.3389/fpls.2020.01140

Farneti, B., Khomenko, I., Grisenti, M., Ajelli, M., Betta, E., Algarra, A. A., et al. (2017b). Exploring blueberry aroma complexity by chromatographic and direct-injection spectrometric techniques. *Front. Plant Sci.* 8, 617. doi:10.3389/fpls.2017.00617

Giongo, L., Ajelli, M., Poncetta, P., Ramos-Garcia, M., Sambo, P., and Farneti, B. (2019). Raspberry texture mechanical profiling during fruit ripening and storage. *Postharvest Biol. Technol.* 149, 177–186. doi:10.1016/j.postharvbio.2018.11.021

Goff, S. A., and Klee, H. J. (2006). Plant volatile compounds: Sensory cues for health and nutritional value? *Science* 311 (5762), 815–819. doi:10.1126/science.1112614

Guichard, E. (1984). formation of volatile components of 2 raspberry cultivars during the ripening. *Sci. Aliments* 4, 459–472.

Hampel, D., Swatski, A., Mosandl, A., and Wüst, M. (2007). Biosynthesis of monoterpenes and norisoprenoids in raspberry fruits (rubus idaeus L.): The role of

cytosolic mevalonate and plastidial methylerythritol phosphate pathway. *J. Agric. Food Chem.* 55, 9296–9304. doi:10.1021/jf071311x

Hospital, F. (2005). Selection in backcross programmes. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 29 (1459), 1503–1511. doi:10.1098/rstb.2005.1670

Klee, H. J., and Tieman, D. M. (2018). The genetics of fruit flavour preferences. *Nat. Rev. Genet.* 19, 347–356. doi:10.1038/s41576-018-0002-5

Larsen, M., Poll, L., Callesen, O., and Lewis, M. (1991). Relations between the content of aroma compounds and the sensory evaluation of 10 raspberry varieties (rubus idaeus L). *Acta Agric. Scand.* 41 (4), 447–454. doi:10.1080/00015129109439927

Li, H., Brouwer, B., Oud, N., Verdonk, J. C., Tikunov, Y., Woltering, E., et al. (2021). Sensory, GC-MS and PTR-ToF-MS profiling of strawberries varying in maturity at harvest with subsequent cold storage. *Postharvest Biol. Technol.* 182, 111719. doi:10.1016/j.postharvbio.2021.111719

Lytou, A. E., Panagou, E. Z., and Nychas, G. J. E. (2019). Volatilomics for food quality and authentication. *Volatilomics food Qual. authentication Curr. Opin. Food Sci.* 28, 88–95. doi:10.1016/j.cofs.2019.10.003

Malowicki, S. M. M., Martin, R., and Qian, M. C. (2008). Volatile composition in raspberry cultivars grown in the pacific northwest determined by stir bar sorptive extraction–gas chromatography–mass spectrometry. *J. Agric. Food Chem.* 56, 4128–4133. doi:10.1021/jf073489p

Mazzucotelli, M., Farneti, B., Khomenko, I., Gonzales-Estanol, K., Pedrotti, M., Fragasso, M., et al. (2022). Proton transfer reaction mass spectrometry: A green alternative for food volatilome profiling. *Green Anal. Chem.* 3, 100041. doi:10.1016/j.greeac.2022.100041

Monsalve, L., Bernales, M., Ayala-Raso, A., Álvarez, F., Valdenegro, M., Alvaro, J. E., et al. (2022). Relationship between endogenous ethylene production and firmness during the ripening and cold storage of raspberry (rubus idaeus 'heritage') fruit. *Horticulturae* 8, 262. doi:10.3390/horticulturae8030262

Paterson, A., Kassim, A., McCallum, S., Woodhead, M., Smith, K., Zait, D., et al. (2013). Environmental and seasonal influences on red raspberry flavour volatiles and identification of quantitative trait loci (QTL) and candidate genes. *Theor. Appl. Genet.* 126, 33–48. doi:10.1007/s00122-012-1957-9

Pazouki, L., and Niinemets, Ü. (2016). Multi-substrate terpene synthases: Their occurrence and physiological significance. *Front. Plant Sci.* 7, 1019. doi:10.3389/fpls.2016.01019

Rohmer, M. (2003) Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution. *Pure Appl. Chem.* 75, 375–388. doi:10.1351/pac200375020375

Sangiorgio, D., Cellini, A., Spinelli, F., Farneti, B., Khomenko, I., Muzzi, E., et al. (2021). Does organic farming increase raspberry quality, aroma and beneficial bacterial biodiversity? *Microorganisms* 9, 1617. doi:10.3390/microorganisms9081617

Sangiorgio, D., Cellini, A., Spinelli, F., Pastore, C., Farneti, B., Savioli, S., et al. (2022). Contribution of fruit microbiome to raspberry volatile organic compounds emission. *Postharvest Biol. Technol.* 183, 111742. doi:10.1016/j.postharvbio.2021.111742

Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., et al. (2019) Diablo: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* 35 (17), 3055–3062. doi:10.1093/bioinformatics/bty1054

Valdés García, A., Maestre Pérez, S. E., Butsko, M., Prats Moya, M. S., and Beltrán Sanahuja, A. (2020). Authentication of "adelita" raspberry cultivar based on physical properties, antioxidant activity and volatile profile. *Antioxidants* 9, 593. doi:10.3390/antiox9070593

Vaughn, S. F., Spencer, G. F., and Shasha, B. S. (1993). Volatile compounds from raspberry and strawberry fruit inhibit postharvest decay fungi. *J. Food Sci.* 58, 793–796. doi:10.1111/j.1365-2621.1993.tb09360.x

Zhang, W., Lao, F., Bi, S., Pan, X., Pang, X., Hu, X., et al. (2021). Insights into the major aroma-active compounds in clear red raspberry juice (Rubus idaeus L. cv. Heritage) by molecular sensory science approaches. *Food Chem.* 336 (17), 127721. doi:10.1016/j.foodchem.2020.127721

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 26 April 2023
DOI 10.3389/fmolb.2023.1160106

OPEN ACCESS

EDITED BY
Andras Szeitz,
University of British Columbia, Canada

REVIEWED BY
Sumit G. Gandhi,
Indian Institute of Integrative Medicine (CSIR), India
Christine Podrini,
The BioArte, Malta
Konstantinos Andreas Kouremenos,
Trajan Scientific and Medical, Australia
Yang Xie,
Brigham and Women"s Hospital and Harvard Medical School, United States
Sabrina Mackinnon,
Newcastle University, United Kingdom

*CORRESPONDENCE
Dominic Fenn,
d.w.fenn@amsterdamumc.nl

SPECIALTY SECTION
This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 06 February 2023
ACCEPTED 30 March 2023
PUBLISHED 26 April 2023

CITATION
Fenn D, Ahmed WM, Lilien TA, Kos R, Tuip de Boer AM, Fowler SJ, Schultz MJ, Maitland-van der Zee AH, Brinkman P and Bos LDJ (2023), Influence of bacterial and alveolar cell co-culture on microbial VOC production using HS-GC/MS. *Front. Mol. Biosci.* 10:1160106. doi: 10.3389/fmolb.2023.1160106

Influence of bacterial and alveolar cell co-culture on microbial VOC production using HS-GC/MS

Dominic Fenn[1,2]*, Waqar M. Ahmed[3], Thijs A. Lilien[2,4], Renate Kos[1], Anita M. Tuip de Boer[2], Stephen J. Fowler[3,5], Marcus J. Schultz[6], Anke H. Maitland-van der Zee[1], Paul Brinkman[1] and Lieuwe D. J. Bos[1,2,6]

[1]Department of Pulmonary medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [2]Laboratory of Experimental Intensive Care and Anaesthesiology, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [3]Division of Immunology, Immunity to Infection and Respiratory Medicine, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom, [4]NIHR-Manchester Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Amsterdam, United Kingdom, [5]Paediatric Intensive Care Unit, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [6]Intensive Care, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

Volatile organic compounds (VOCs) found in exhaled breath continue to garner interest as an alternative diagnostic tool in pulmonary infections yet, their clinical integration remains a challenge with difficulties in translating identified biomarkers. Alterations in bacterial metabolism secondary to host nutritional availability may explain this but is often inadequately modelled *in vitro*. The influence of more clinically relevant nutrients on VOC production for two common respiratory pathogens was investigated. VOCs from *Staphylococcus aureus* (*S.aureus*) and *Pseudomonas aeruginosa* (*P.aeruginosa*) cultured with and without human alveolar A549 epithelial cells were analyzed using headspace extraction coupled with gas chromatography-mass spectrometry. Untargeted and targeted analyses were performed, volatile molecules identified from published data, and the differences in VOC production evaluated. Principal component analysis (PCA) could differentiate alveolar cells from either *S. aureus* or *P. aeruginosa* when cultured in isolation based on PC1 (p = 0.0017 and 0.0498, respectively). However, this separation was lost for *S. aureus* (p = 0.31) but not for *P. aeruginosa* (p = 0.028) when they were cultured with alveolar cells. *S. aureus* cultured with alveolar cells led to higher concentrations of two candidate biomarkers, 3-methyl-1-butanol (p = 0.001) and 3-methylbutanal (p = 0.002) when compared to *S. aureus,* alone. *P. aeruginosa* metabolism resulted in less generation of pathogen-associated VOCs when co-cultured with alveolar cells compared to culturing in isolation. VOC biomarkers previously considered indicative of bacterial presence are influenced by the local nutritional environment and this should be considered when evaluating their biochemical origin.

KEYWORDS

VOC, HS-GC/MS, culture media composition, co-culture model, bacterial VOCs

Introduction

Pulmonary infections remain the leading cause of communicable deaths worldwide (Hubbard, 2006). Continued efforts are being made to modernise clinical microbiology and improve pathogen detection. However, whilst molecular diagnostics are readily becoming recognised as the new gold standard in clinical virology replacing the need for viral cultures

(Hodinka and Kaiser, 2013), the same is not true for clinical microbiology that remains reliant on culture-based methodologies (Didelot et al., 2012; Chanderraj and Dickson, 2018). Consequently, clinical microbiology remains at the mercy of culture dependent analysis and as such, is limited by a) sample availability and b) whether the causative organism can be cultured. Furthermore, conventional cultures are a laborious and inherently biased approach associated with delays in targeted antibiotic therapy and broad-spectrum antibiotic over use (Joo et al., 2014; Hilton et al., 2016). An alternative approach is therefore long overdue.

Exhaled breath metabolomics with its ease of collection and sputum independence represents a technique that could overcome these limitations (Filipiak et al., 2015; Hilton et al., 2016; Kos et al., 2021; Ahmed et al., 2022). In addition, the characterisation and detection of changes in metabolic activity, that may precede disease symptoms, could provide invaluable information regarding bacterial presence, viability and activity (Tounta et al., 2021; Mohd Kamal et al., 2022). However, as with all omics data, the challenge lies in clinically integration (van Karnebeek et al., 2018). Presently, there is a surfeit of discovery exhaled breath studies identifying various volatile organic compounds (VOCs) as candidate biomarkers for specific bacteria (Filipiak et al., 2012; 2015; Fenn et al., 2021). Yet, translational difficulties between the pre-clinically and clinically identified VOCs (Filipiak et al., 2015; Ahmed et al., 2022) prevent their necessary validation and subsequent application.

Successful pathogenic bacterial colonisation is dependent on respiratory tract surface adherence and the procurement of local nutrients for growth (Siegel and Weiser, 2015). Consequently, bacteria are capable of adapting their metabolism in response to an altered nutritional availability at the infection site (Brown et al., 2008; Siegel and Weiser, 2015). The impact of this metabolic alteration is frequently over looked in *vitro* VOC studies and may partly explain the observed translational difficulties. In the present study, two common respiratory pathogens, i.e., *Staphylococcus aureus* (*S.aureus*) and *Pseudomonas aeruginosa* (*P.aeruginosa*) were co-cultured with alveolar epithelial cells, and analyzed using headspace extraction coupled with gas chromatography-mass spectrometry (HS-GC/MS) to explore the impact of the nutritional environment on microbial VOC production. It was hypothesised that bacteria cultured in a growth substrate more representative of *in vivo* conditions would a) yield separate VOCs, and b) the VOCs that have previously been identified *in vivo* would yield higher concentrations in the co-cultured headspace than in regular headspace analysis.

Materials and methods

Cell line and bacterial cultivation

For the cellular component, culture conditions and sample preparation were as described previously (Fenn et al., 2022). In brief, immortalised human alveolar basal epithelial (A549) cells (CCL-185) were cultivated in 75 cm^2 cell culture flasks and incubated at 37 °C in 5% CO2 in Roswell Park Memorial Institute (RPMI) 1,640 medium (Gibco) supplemented with 10% foetal bovine serum (FBS), Penicillin-Streptomycin (5 mL containing 10,000 units per mL Penicillin, 10,000 μg mL^{-1} Streptomycin, Gibco), L-Glutamine, gentamicin and amphotericin B. Cells were passaged every three to 4 days once ≈90% confluent with a similar passage used for each experimental replicate. Prior to headspace collection and bacterial inoculation, they were detached from the culture flask using 0.05% Trypsin-EDTA and resuspended in RPMI-1640 before being seeded (≈1.5 × 10^5) in 1 mL of supplemented RPMI-1640 in 20 mL glass headspace vials (Markes International, Bridgend, Wales) and incubated at 37°C in 5% CO_2 for 22–24 h (Fenn et al., 2022).

The reference strains *S. aureus* (ATCC 29213), *P. aeruginosa* (PAO1) were investigated. For each experimental replicate, strains were sub-cultured from glycerol frozen stocks onto Columbia blood agar plates with 5% sheep blood (43,049, bioMérieux, Marcy-l'Étoile, France) and incubated over night at 37°C to ensure axenic colonies. Isolated colonies were then inoculated and cultivated in 20 mL brain heart infusion (BHI, nr: 116, Amsterdam UMC—location AMC, Netherlands) and grown overnight at 37°C without agitation.

Co-culture infection and treatment experiments

Overnight liquid cultures for both *S. aureus* and *P. aeruginosa* were firstly standardised to a 0.35 OD_{620nm} using RPMI-1640 media without antibiotics. A stepwise dilution was then done to achieve a target of 10^4 colony forming units per ml (CFU/mL), a frequently used diagnostic threshold to define a positive culture *in vivo* (Baselski and Klutts, 2013).

Once the A549 cells had formed a monolayer with a confluence ≈80–90% within the glass headspace vials, the medium was removed. The vials were then washed with phosphate buffered saline to remove any residual traces of RMPI media containing antibiotics. The cells were then replenished with a) 200 ul antibiotic free RPMI-1640 media (control), b) 200 ul standardised *S. aureus* inoculate or c) 200 ul standardised *P. aeruginosa* inoculate prepared in antibiotic free RPMI-1640 media. For the bacteria cultured without cells, the headspace vials were treated in a similar fashion prior to inoculation, using RPMI media alone instead of A549 cells. 200 ul of the standardised inoculate was then added for each bacterium to the glass headspace vials. All vials were sealed with crimp-tops with polytetrafluoroethylene (PTFE)-lined septa (Markes International, Cincinnati, Ohio, United States) and incubated and agitated for 24 h (temperature 37°C, agitation 200 RPM, HiSorb agitator, Markes International, Cincinnati, Ohio, United States). The experiment was conducted in triplicate with three biological replicates per condition per experimental day. Quantitative culturing was performed to determine the CFU/mL of the standardised inoculates for each experiment. Replicates were removed from analysis if target CFU count was not met. Similarly, sterility assessments for the control group were also performed following headspace capture to ensure no evidence of infection.

HS-GC/MS

Headspace sampling was performed using HiSorb high-capacity extraction probes with polydimethylsiloxane sorbent phase (Markes International, Cincinnati, Ohio, United States) as previous described

(Fenn et al., 2022). In brief, following inoculation the 20 mL glass headspace vials were sealed and incubated and agitated for 24 h. Conditioned sorbent probes were then inserted into the headspace for 2 h after 22 h of incubation. The probes were removed and stored in empty stainless steel sorbent tubes (Markes International, Cincinnati, Ohio, United States) ahead of thermal desorption. HS-GC/MS analysis was performed on the same day as VOC capture for all samples as previously described (Fenn et al., 2022). Compounds of interest were identified using the GC/MS Solutions (Shimadzu, Den Bosch, the Netherlands) platform incorporating the National Institute and Technology library as described previously (Bos et al., 2014) and compared to analytical standard where possible. Any compounds that had match score < 80 or could not be accurately detected due to co-elution were excluded from further analysis.

Statistical analysis and data processing

Statistical analysis was performed through the R studio interface using R (version 3.6.1). Raw HS-GC/MS spectra were processed using the R "xcms' package (Scripps Center for Metabolomics, La Jolla, CA, United States) as previously reported (Bos et al., 2014), and underwent denoising, peak detection and alignment to create a three-dimensional data matrix containing sample metadata, retention time and mass-to-charge ratio (m/z), ahead of downstream analysis. Failed analyses and technical errors were excluded by visual inspection of chromatograms after processing using the "xcms" pipeline. Known analytical artefacts, such as siloxanes were also removed. The peak table was then normalised using "limma" package and log scaled to adjust for experimental day differences and to stabilise variance.

An untargeted and targeted approach were used to explore VOC differences in bacteria cultured with and without A549 cells. In the untargeted approach, variation of VOC profiles between bacteria and A549 cells in isolation was assessed using principal component analysis (PCA). Loadings from this PCA were projected on samples from bacteria cultured with cells and differences between group centroids were assessed using pairwise *post hoc* Dunn's analysis correcting for multiple testing. VOC concentrations were then compared between bacteria and bacteria cultured with cells using the Mann-Whitney *U* test, and evaluated alongside a log2 fold change. This was visualised using a volcano plot. A log2 fold change ≥2 and an *p*-value <0.05 was selected to limit false discovery and ensure biologically meaningful differential VOC expression. The analyses were split for each pathogen.

For the targeted approach, species-specific volatile metabolites were identified from previously conducted systematic literature reviews (Kos et al., 2021; Kos et al., unpublished data) and further filtered to include only VOCs that have been reported in one or more *in vivo* studies (Supplementary Table S1). Mann-Whitney U analysis was used to compare VOC concentrations for each pathogen, cultured with and without A549 cells. For both the untargeted and targeted approach, a compound was only considered to result from alterations in bacterial metabolism due to the presence of alveolar cells when it was a) significantly different from bacteria alone and b) significantly different from cells alone.

Results

Following data processing and the removal of known analytical artefacts, a total of 503 features were detected and used in the analysis. The impact of alveolar cell presence on microbial derived volatile metabolites was first evaluated using a PCA model based on all 503 detected compounds comparing bacteria and cells cultured in isolation.

Untargeted analysis

PCA demonstrated separation between A549 cells *versus* *S. aureus* or *P. aeruginosa* in culture based on PC1, which explained 34.4% of the total variation for *S. aureus* and 29.2% for *P. aeruginosa* (adjusted *p*-value: 0.0017 and 0.0498, respectively, Figure 1A, B). However, when *S. aureus* was cultured with alveolar cells, the separation from cells in isolation became less clear (adjusted *p*-value: 0.31, Figure 1A). This was not true for the *P. aeruginosa* co-culture where differentiation was still possible from cells in isolation (adjusted *p*-value: 0.028, Figure 1B). The separation between *S. aureus* or *P. aeruginosa* cultured with and without cells was not possible based on the PC1 (adjusted *p*-value: 0.2 and 1.0, respectively).

Differences in individual VOCs identified in the headspace of these cultures were further explored through the direct comparison of bacteria cultured with and without alveolar cells, as shown in Figures 2A, B. From the initial 503 detected metabolites, ten (1.9%) compounds met the predefined criteria (log2 fold change ≥2 and $p < 0.05$). Of these, six compounds were found in greater concentration in the *S. aureus* co-culture, three compounds were found in greater concentration in the *P. aeruginosa* co-culture, and one compound was found in lower concentrations in both co-cultures (Figures 2A, B). However, of these nine, six metabolites (2-ethyl-1-hexanol, dodecane, nonanal, 2,5-Octadiene,3,4,5,6-tetramethyl-, tridecane, and acetaminde, 2,2-dichloro) showed no significant difference from alveolar cells alone (Figure 3, B-G, $p > 0.05$). An increased concertation was observed in the cellular headspace for one compound (4-ethyl-octane) compared to the *S. aureus* co-culture (Figure 3A, $p = 0.01$). The two remaining compounds (3-methyl-1-butanol and 3-methylbutanal), found both in higher concentration in *S. aureus* cultured with alveolar cells, were the only metabolites that met the criteria to be considered a result of altered bacterial metabolism due to presence of cells (Figures 3H, I). For both bacteria, cyclohexanone, was found in lower concentration in the co-culture headspace (Figures 4A, B).

Targeted analysis

Based on previous systematic reviews (Kos et al., 2021; Kos et al., unpublished data), 38 compounds reported in exhaled breath were considered for targeted analysis, including five compounds shared by both *S. aureus* and *P. aeruginosa*, 18 *S. aureus* specific compounds and 15 *P aeruginosa* specific compounds (Supplementary Table S2).

Three of the *S. aureus* related molecules were identified from headspaces characterised in this work. One molecule, 3-methylbutanal, was pathogen-specific and two were reported in

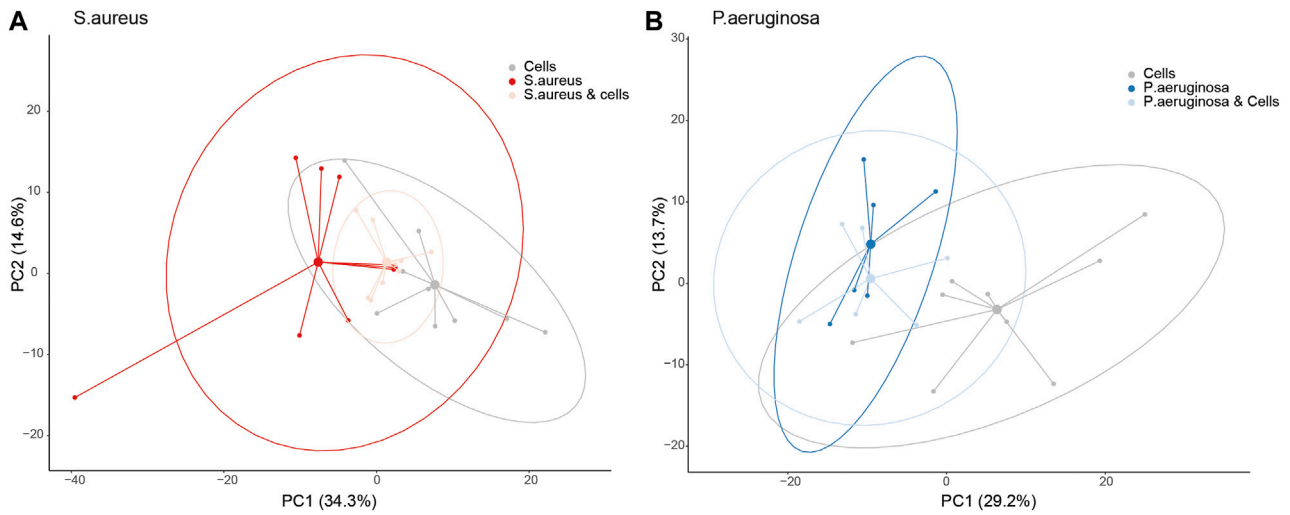


FIGURE 1
A principal component analysis (PCA) plot for bacteria co-cultured with alveolar cells based on the loadings from a PCA between bacteria and cells in isolation, split per pathogen. Panel **(A)** - *S. aureus* and Panel **(B)** - *P. aeruginosa*.

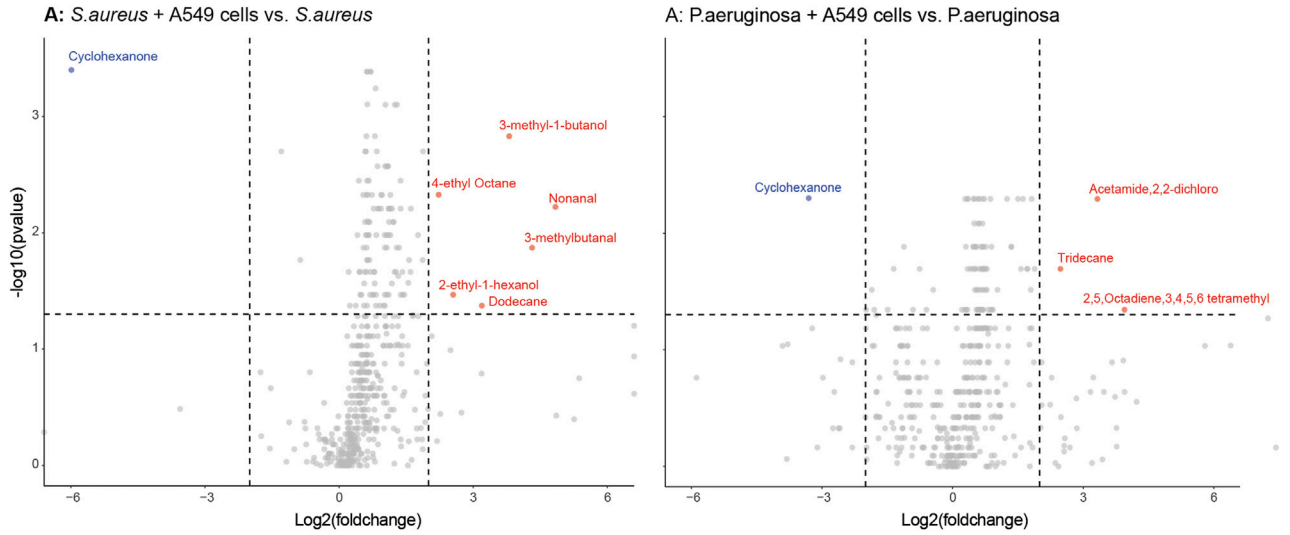


FIGURE 2
Volcano plots of all VOCs significantly released by either *S. aureus* [Panel **(A)**] or *P. aeruginosa* [Panel **(B)**] when cultured with or without alveolar A549 cells. Blue dots represent down-regulated VOCs and red dots represent up-regulated VOCs with a log2 fold change ≥ 2 and $p < 0.05$ depicted by the dashed line.

association with both bacteria (3-methyl-1-butanol and dimethyl disulfide) (Figures 5A–C). Two of three molecules (3-methylbutanal and 3-methyl-1-butanol) showed statistically significant increase in concentration when cultured with alveolar cells compared to alveolar cells in isolation ($p = 0.001$ and $p = 0.001$) or bacteria in isolation ($p = 0.01$ and $p < 0.001$, respectively), see Figures 5A, B.

Ten of the *P. aeruginosa* related molecules were identified from headspaces characterised in this work (Figure 6). However, none of these metabolites showed a significant difference when compared to bacteria in isolation and cells alone. In fact, for the majority of the compounds, no differences in headspace concentration were observed between the three experimental groups (Figures 6A–J). One compound, 1-undecene was detected in greater concentration in the bacterial headspace compared to alveolar cells ($p < 0.001$) but its concentration was not altered by co-culture. Two molecules, 3-methyl-1-butanol and dimethyl disulfide, were both found in *S. aureus* and *P. aeruginosa*. However, unlike for *S. aureus*, 3-methyl-1-butanol was not significantly different between the groups (Figures 6I, J). Dimethyl disulfide was found in higher concentrations in the headspace of *P. aeruginosa* cultured in

10.3389/fmolb.2023.1160106

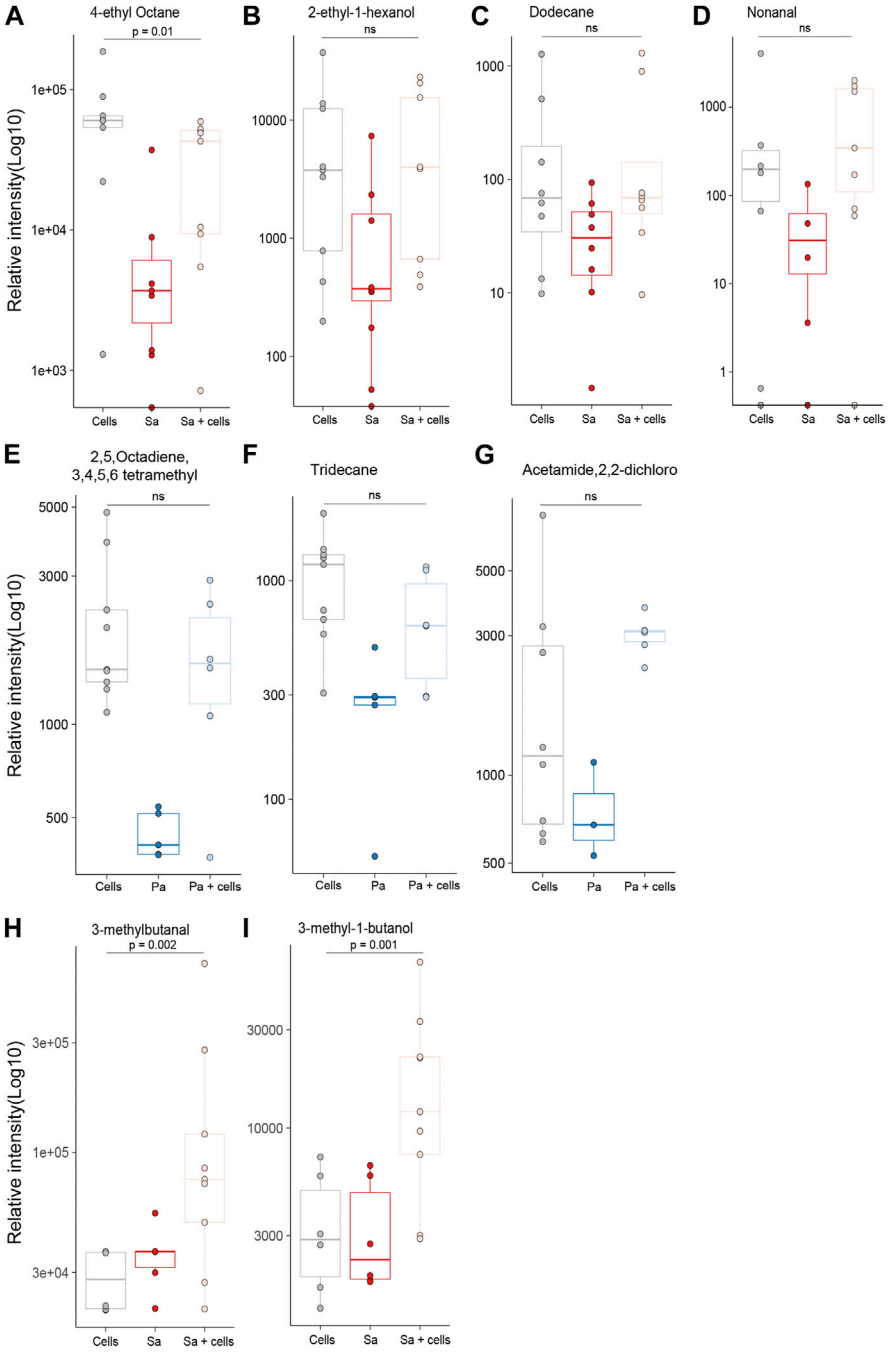


FIGURE 3
All identified VOCs with a log2 foldchange ≥ 2 and *p* value < 0.05 that showed positive association with bacteria cultured with alveolar cells, split per pathogen. Minimum of six repeats were used to generate the boxplots and *p* value calculated using Mann-Whitney U test. (Cells = A549 cells, Sa = *S. aureus*, Pa = *P. aeruginosa*).

isolation, compared to alveolar cells alone and to co-culture (p = 0.01 and p = 0.03, respectively, Figure 6J).

Discussion

HS-GC/MS has been used to provide preliminary data on the influence of clinically relevant nutrients on the production of specific volatile metabolites by *S. aureus* and *P. aeruginosa*, two common respiratory pathogens. Growth of *S. aureus* on alveolar cells rather than culture medium alone resulted in a higher concentrations of 3-methyl-1-butanol and 3-methylbutanal, two important candidate breath biomarkers for *S. aureus* pneumonia.

Growth of *P. aeruginosa* on alveolar cells rather than culture medium alone did not identify statistically significant changes in

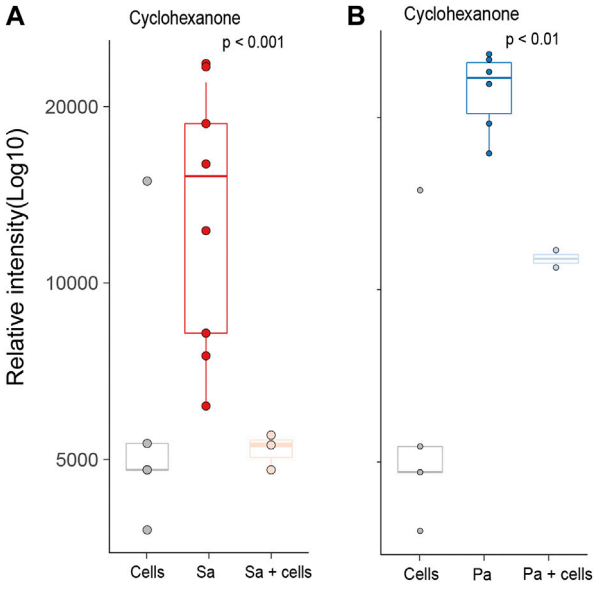


FIGURE 4
All identified VOCs with a log2 foldchange ≥ 2 and *p* value < 0.05 that showed negative association with bacteria cultured with alveolar cells, split per pathogen. Minimum of six repeats were used to generate the boxplots and *p* value calculated using Mann-Whitney U test. (Cells = A549 cells, Sa = *S. aureus*, Pa = *P. aeruginosa*).

previously identified VOCs that are associated with pathogens. Together, our results suggest that production of VOCs considered biomarkers of bacterial presence could be influenced by the nutrient environment of the bacterium.

An untargeted approach identified differences in VOC production between bacteria cultured with and without alveolar cells for *S. aureus*. Most noticeably two metabolites, 3-methyl-1-butanol and 3-methylbutanal. These two compounds were also identified as pathogen associated VOC targets having previously been linked to *S. aureus* in numerous studies, both *in vitro* (Filipiak et al., 2012; Boots et al., 2014; Chen et al., 2017; Lawal and Muhamadali, 2018; Ahmed et al., 2022) and *in vivo* (Filipiak et al., 2015; Ahmed et al., 2022). However, the differences observed in the current study between *S. aureus* co-culture and bacteria alone may provide further insight into their metabolic origin. Bacterial membrane homeostasis is critical to the survival of *S. aureus in vivo*, ensuring optimal compatibility between the host and pathogen, through the upregulation of fatty acid synthesis (Zhang and Rock, 2008; Parsons and Rock, 2013; Frank et al., 2021). Leucine catabolism is a suggested pathway utilised by *S. aureus* to achieve this (Frank et al., 2021) and leads to the formation of both 3-methyl-1-butanol and 3-methylbutanal as found in the current study. Whilst greater elucidation of the metabolic pathway is required, this study presents additional evidence to support their potential use as biomarkers, in particular, 3-methylbutanal that was recently successfully translated for the identification of *S. aureus* in ventilator-associated pneumonia in a large patient cohort (Ahmed et al., 2022) and warrants further investigation.

Changes to *P. aeruginosa* metabolic activity secondary to host interaction are well documented (Palmer et al., 2007; Jurado-Martín et al., 2021). However, despite this, metabolic changes were not observed in the current study for *P. aeruginosa*. One possible explanation may have been the prolonged incubation period utilised in the current study and the rapid metabolic rearrangement attributed to *P. aeruginosa* virulence (Perinbam et al., 2020). As such, the metabolites reflecting these dynamic changes may have been missed.

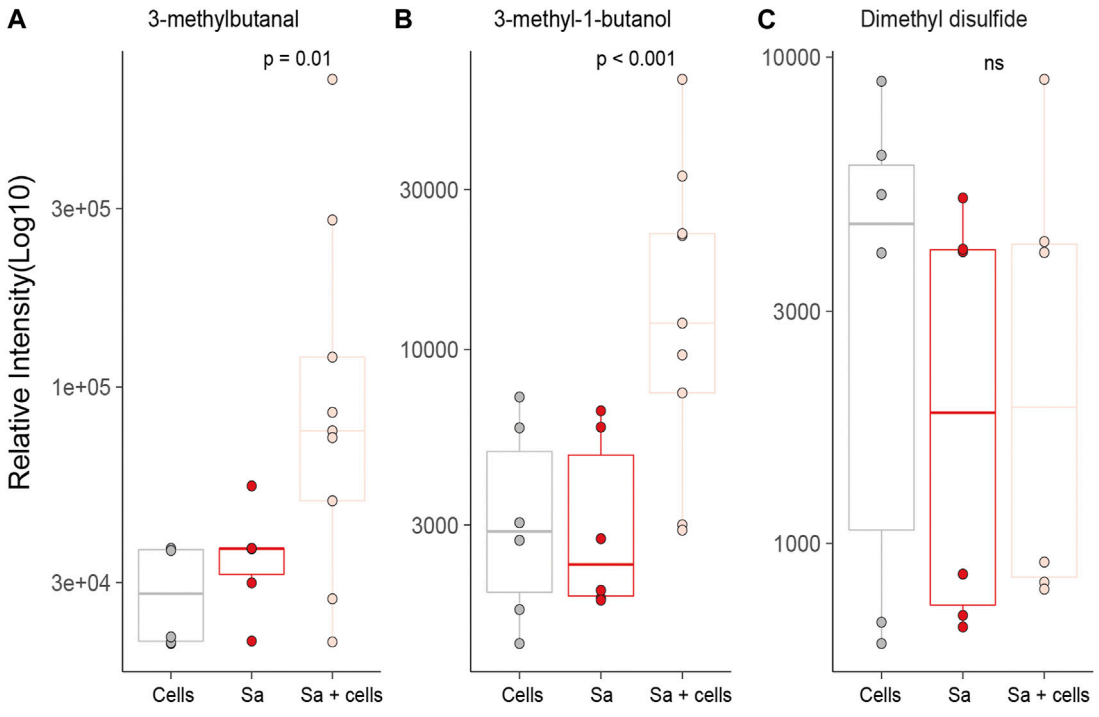


FIGURE 5
All previously identified target VOCs for *S. aureus* observed in *in vitro* bacterial co-culture headspace, split per treatment group with minimum of six repeats used to create boxplots. Differences between experimental groups evaluated using Mann-Whitney U. (Cells = A549 cells, Sa = *S. aureus*.

10.3389/fmolb.2023.1160106

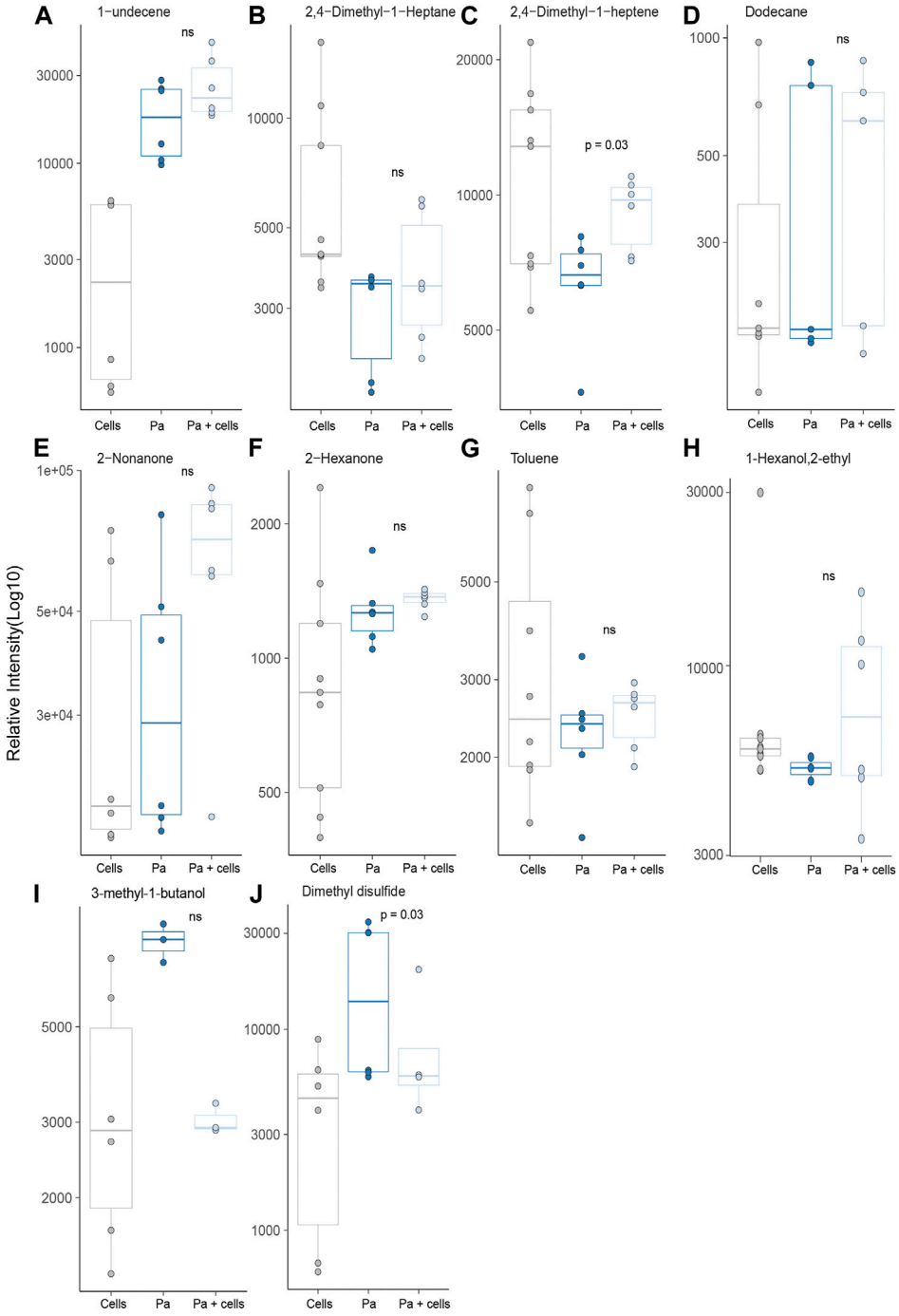


FIGURE 6
All previously identified target VOCs for *P. aeruginosa* observed in *in vitro* bacterial co-culture headspace, split per treatment group with minimum of six repeats used to create boxplots. Differences between experimental groups evaluated using Mann-Whitney U test. (Cells = A549 cells, Pa = *P. aeruginosa*).

However, a previous study that similarly co-cultured *P. aeruginosa* with alveolar cells also showed no differences between bacteria with and without cells despite a shortened incubation window (Lawal and Knobel, 2018). Alternatively, it has been postulated that pseudomonal adaptation and successful colonisation are dependent on the nutritional components found in sputum (Palmer et al., 2007). As such, it is more likely that the co-culture model presented here may have provided inadequate nutritional cues to stimulate meaningful metabolic re-arrangement despite using nutrients more representative of *in vivo* conditions, reinforcing the importance of carefully defining infection site physiology for future studies and considering not only anatomical location but also pathogen prerequisites.

In addition to the untargeted analysis, we also performed targeted analysis in which previously identified microbial VOCS were sought to

evaluate if the use of a more physiologically relevant growth media might lead to an increased headspace concentration. Whilst this was true for *S. aureus* co-cultures for both 3-methyl-1-butanol and 3-methylbutanal, the same cannot be said for *P. aeruginosa,* where differences in VOC release could not be attributed to changes in its metabolism resulting from co-culture with alveolar cells. Furthermore, most of the previously identified *P. aeruginosa* VOC biomarkers showed no significant difference in headspace concentration of *P. aeruginosa* when compared to alveolar cells alone. This differs from the results found in the systematic review conducted by Kos et al., 2021, that identified a number of *P. aeruginosa* specific VOCs. It is important to recognise however, that in the majority of these studies bacteria were cultured in isolation and compared to other bacterial species, not human cells lines. Subsequently, for these studies, VOC release may have been attributed to the bacteria, whereas the current study would suggest their origin is not necessarily only microbially derived and instead could reflect metabolic changes secondary to their pathogenicity. This is particularly evident in the case of dimethyl disulfide, which returned to alveolar cell concentrations when co-cultured with them. As such, the clinical utility of these biomarkers is questionable and caution is warranted for their use *in vivo*. This was not true for all targeted VOCs. 1-Undecene did show a positive association with *P. aeruginosa* greater than alveolar cells in isolation and is in agreement with numerous other studies as evidenced by Kos et al., 2021. However despite this, its translation from pre-clinical to clinical studies is yet to be proven (Kos et al., 2021; Ahmed et al., 2022) and suggests that further work is needed to better determine its *in vivo* biosynthesis before its use as VOC biomarker can be recommended.

A strength of this study lies in the use of a previously developed headspace model that incorporates both glass culture vessels and a sealed system to minimize the impact of plastic contaminants and improve reproducibility (Ahmed et al., 2022; Fenn et al., 2022). Furthermore, the incorporation of bacterial co-culture demonstrated the versatility of said model and could provide an easy and reliable platform to further examine the nutritional effects of more *in vivo* relevant growth media on VOC production. Additionally, the approach taken represents the culmination of comprehensive systematic reviews coupled with a validated *in vitro* model for a reliable VOC assessment utilising both targeted and untargeted analyses.

The key limitation of this approach was the use of a two-dimensional culture model over more physiological relevant three-dimensional models, such as air liquid interface cultures. This model was used however to maximise VOC capture and minimise the influence of VOCs omitted by plastic equipment used in such models. The low number of bacterial species and use of a single cell line is another limitation that warrants caution when extrapolating the current findings to other bacteria or infection sites. Further work therefore should include additional bacterial species and strains co-cultured with different cell types to validate the presented findings. Finally, the respiratory tract is a complex and dynamic community of microbiota (Dickson et al., 2015). As such, the use of bacterial monocultures in the current work is another limitation that should be considered when interpreting our results. The approach taken here likely underestimates *in vivo* bacterial pathogenesis and fails to capture all the interactions between pathogenic bacteria and other respiratory tract flora that could influence VOC production. However, the integration of multiple bacterial interactions represents a challenge when attempting to trace the VOC origin. A challenge faced by *in vivo* exhaled breath analyses and the models that investigate them, such as presented in this study, may be in a better position to address these challenges. Nevertheless, we recognise this as a limitation and advise that caution is taken when interpreting the results in a clinical environment.

Overall, the detection of pathogen associated compounds was poor for both bacteria. Target compounds were either not detected, or showed no significant difference between experimental groups. Similar difficulties have also been observed in other studies attempting to translate volatile metabolites (Filipiak et al., 2015; Kos et al., 2021; Ahmed et al., 2022). However, more often than not, the emphasis is primarily given to the compounds that are found, with the translational discrepancies either not being discussed or attributed to differences in sorbent materials, column affinities or analytical methodologies. Whilst we recognise the validity of such arguments, they represent an innate problem in exhaled breath research that continues to complicate its clinical integration. This burgeoning challenge is further exacerbated by the continued development of new technologies, lack of consensus on which to use and penchants for discovery studies. As such, a plethora of VOC biomarkers continue to be identified that may in fact represent the techniques used rather than their applicability in the pathologic condition studied. Therefore, if exhaled breath is to become clinically applicable, the preference for discovery studies must be disrupted and the emphasis given to a) translational and development studies b) the development of better *in vitro* models c) the use of multiple analytical techniques and d) the cross validation with external research institutions.

Conclusion

HS-GC/MS analyses showed that nutrients more representative of *in vivo* environment had an influence on the production of specific volatile metabolites by *S. aureus* and *P. aeruginosa.* The observed differences for *S. aureus* identified two important candidate biomarkers, 3-methyl-1-butanol and 3-methylbutanal. The fewer discernible differences for *P. aeruginosa* challenges the validity of previously suggested breath biomarkers. Together, our results suggest that VOC biomarkers may indicate the presence of bacterial species, and are influenced by the local nutritional environment and should be considered when evaluating their biochemical origin.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Substantial contributions to the conception or design of the work: DF, WA, SF, RK, AHM, PB, and LB. The acquisition and analysis: DF, WA, TL, RK, AT, PB, and LB interpretation of data for the work: DF, WA, TL, MS, PB, and LB. Drafting the work or revising it critically for important intellectual content: ALL. Final approval of the version to be published: ALL. Agreement to be accountable: ALL.

Funding

LB was supported by the Young Investigator Award and Dirkje Postma Award from the Dutch Lung Foundation (Longfonds) for sample analysis performed in this study, and by the Health Holland and the Amsterdam UMC fellowship, outside of the submitted work. WA and SF are supported by the NIHR-Manchester Biomedical Research Centre.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1160106/full#supplementary-material

References

Ahmed, W. M., Dominic, F., and Iain, R. W. (2022). Microbial volatiles as diagnostic biomarkers of bacterial lung infection in mechanically ventilated patients. *Clin. Infect. Dis.* 76, 1059–1066. doi:10.1093/cid/ciac859

Baselski, V., Klutts, J. S., and Klutts, J. S. (2013). Quantitative cultures of bronchoscopically obtained specimens should be performed for optimal management of ventilator-associated pneumonia. *J. Clin. Microbiol.* 51 (3), 740–744. doi:10.1128/JCM.03383-12

Boots, A. W., Smolinska, A., van Berkel, J. J. B. N., Fijten, R. R. R., Stobberingh, E. E., Boumans, M. L. L., et al. (2014). Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography-mass spectrometry. *J. Breath Res.* 8 (2), 027106. doi:10.1088/1752-7155/8/2/027106

Bos, L. D. J., Weda, H., Wang, Y., Knobel, H. H., Nijsen, T. M. E., Vink, T. J., et al. (2014). Exhaled breath metabolomics as a noninvasive diagnostic tool for acute respiratory distress syndrome. *Eur. Respir. J.* 44 (1), 188–197. doi:10.1183/09031936.00005614

Brown, S. A., Palmer, K. L., and Whiteley, M. (2008). Revisiting the host as a growth medium. *Nat. Rev. Microbiol.* 6 (9), 657–666. doi:10.1038/nrmicro1955

Chanderraj, R., and Dickson, R. P. (2018). Rethinking pneumonia: A paradigm shift with practical utility. *Proc. Natl. Acad. Sci. U. S. A.* 115 (52), 13148–13150. doi:10.1073/pnas.1819024116

Chen, J., Tang, J., Shi, H., Tang, C., and Zhang, R. (2017). Characteristics of volatile organic compounds produced from five pathogenic bacteria by headspace-solid phase micro-extraction/gas chromatography-mass spectrometry. *J. Basic Microbiol.* 57 (3), 228–237. doi:10.1002/jobm.201600505

Dickson, R. P., Erb-Downward, J. R., and Huffnagle, G. B. (2015). Homeostasis and its disruption in the lung microbiome. *Am. J. Physiology - Lung Cell. Mol. Physiology* 309 (10), L1047–L1055. doi:10.1152/ajplung.00279.2015

Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., and Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing', *Nature Reviews Genetics*. *Nat. Publ. Group* 13 (9), 601–612. doi:10.1038/nrg3226

Fenn, D., Mahmoud, A. A., and Paul, B. (2021). Comparison of microbial composition of cough swabs and sputum for pathogen detection in patients with cystic fibrosis. *J. Cyst. Fibros.*, 52–60. Elsevier B.V. doi:10.1016/j.jcf.2021.08.031

Fenn, D., Thijs, A., Laura, A., Marry, R., Nanon, F. L., Anita, M. T., et al. (2022). Validation of volatile metabolites of pulmonary oxidative injury: A bench to bedside study. *ERJ Open Res.* 9, 00427–02022. doi:10.1183/23120541.00427-2022

Filipiak, W., et al. (2015). Breath analysis for *in vivo* detection of pathogens related to ventilator-associated pneumonia in intensive care patients: A prospective pilot study. *J. Breath Res.* 9 (1), 16004. IOP Publishing. doi:10.1088/1752-7155/9/1/016004

Filipiak, W., Sponring, A., Baur, M. M., Filipiak, A., Ager, C., Wiesenhofer, H., et al. (2012). Molecular analysis of volatile metabolites released specifically by staphylococcus aureus and pseudomonas aeruginosa. *BMC Microbiol.* 12, 113. doi:10.1186/1471-2180-12-113

Frank, M. W., Whaley, S. G., and Rock, C. O. (2021). Branched-chain amino acid metabolism controls membrane phospholipid structure in *Staphylococcus aureus*. *J. Biol. Chem.* 297 (5), 101255. Elsevier B.V. doi:10.1016/j.jbc.2021.101255

Hilton, S. K., Castro-Nallar, E., Perez-Losada, M., Toma, I., McCaffrey, T. A., Hoffman, E. P., et al. (2016). Metataxonomic and metagenomic approaches vs. culture-based techniques for clinical pathology. *Front. Microbiol.* 7 (APR), 484–512. doi:10.3389/fmicb.2016.00484

Hodinka, R. L., and Kaiser, L. (2013). Is the era of viral culture over in the clinical microbiology laboratory? *J. Clin. Microbiol.* 51 (1), 2–4. doi:10.1128/JCM.02593-12

Hubbard, R. (2006). The burden of lung disease. *Thorax* 61 (7), 557–558. doi:10.1136/thx.2006.066050

Joo, Y. M., Chae, M. K., Hwang, S. Y., Jin, S. C., Lee, T. R., Cha, W. C., et al. (2014). Impact of timely antibiotic administration on outcomes in patients with severe sepsis and septic shock in the emergency department. *Clin. Exp. Emerg. Med.* 1 (1), 35–40. doi:10.15441/ceem.14.012

Jurado-Martín, I., Sainz-Mejías, M., and McClean, S. (2021). *Pseudomonas aeruginosa*: An audacious pathogen with an adaptable arsenal of virulence factors. *Int. J. Mol. Sci.* 22 (6), 3128–3137. doi:10.3390/ijms22063128

Kos, R., et al. (2021). Targeted exhaled breath analysis for detection of *Pseudomonas aeruginosa* in cystic fibrosis patients. *J. Cyst. Fibros.* Elsevier B.V. doi:10.1016/j.jcf.2021.04.015

Lawal, O., Knobel, H., Weda, H., Bos, L. D., Nijsen, T. M. E., Goodacre, R., et al. (2018). Volatile organic compound signature from co-culture of lung epithelial cell line with: *Pseudomonas aeruginosa*. *Analyst* 143 (13), 3148–3155. doi:10.1039/c8an00759d

Lawal, O., Muhamadali, H., et al. (2018). Headspace volatile organic compounds from bacteria implicated in ventilator-associated pneumonia analysed by TD-GC/MS. *J. Breath Res.* 12 (2). IOP Publishing. doi:10.1088/1752-7163/aa8efc

Mohd Kamal, K., Mahamad Maifiah, M. H., Abdul Rahim, N., Hashim, Y. Z. H. Y., Abdullah Sani, M. S., and Azizan, K. A. (2022) 'Bacterial metabolomics: Sample preparation methods', *Biochem. Res. Int.* 2022, 9186536. doi:10.1155/2022/9186536

Palmer, K. L., Aye, L. M., and Whiteley, M. (2007). Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J. Bacteriol.* 189 (22), 8079–8087. doi:10.1128/JB.01138-07

Parsons, J. B., and Rock, C. O. (2013). Bacterial lipids: Metabolism and membrane homeostasis. *Prog. Lipid Res.* 52 (3), 249–276. Elsevier Ltd. doi:10.1016/j.plipres.2013.02.002

Perinbam, K., et al. (2020). A shift in central metabolism accompanies virulence activation in *Pseudomonas aeruginosa*', *mBio. J. Engel* 11 (2), 1–16. doi:10.1128/mBio.02730-18

Siegel, S. J., and Weiser, J. N. (2015). Mechanisms of bacterial colonization of the respiratory tract. *Annu. Rev. Microbiol.* 69 (1), 425–444. doi:10.1146/annurev-micro-091014-104209

Tounta, V., Liu, Y., Cheyne, A., and Larrouy-Maumus, G. (2021). Metabolomics in infectious diseases and drug discovery', *Molecular Omics. R. Soc. Chem.* 17 (3), 376–393. doi:10.1039/d1mo00017a

van Karnebeek, C. D. M., Wortmann, S. B., Tarailo-Graovac, M., Langeveld, M., Ferreira, C. R., van de Kamp, J. M., et al. (2018). The role of the clinician in the multi-omics era: Are you ready? *J. Inherit. Metabolic Dis.* 41 (3), 571–582. doi:10.1007/s10545-017-0128-1

Zhang, Y. M., and Rock, C. O. (2008). Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* 6 (3), 222–233. doi:10.1038/nrmicro1839

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 11 May 2023
DOI 10.3389/fmolb.2023.1178269

OPEN ACCESS

EDITED BY
Andras Szeitz,
University of British Columbia, Canada

REVIEWED BY
Hiroshi Tsugawa,
Tokyo University of Agriculture and Technology, Japan
Chaevien S. Clendinen,
Pacific Northwest National Laboratory (DOE), United States

*CORRESPONDENCE
Kelly R. Redeker,
kelly.redeker@york.ac.uk

RECEIVED 02 March 2023
ACCEPTED 24 April 2023
PUBLISHED 11 May 2023

CITATION
Issitt T, Reilly M, Sweeney ST, Brackenbury WJ and Redeker KR (2023), GC/MS analysis of hypoxic volatile metabolic markers in the MDA-MB-231 breast cancer cell line. *Front. Mol. Biosci.* 10:1178269. doi: 10.3389/fmolb.2023.1178269

GC/MS analysis of hypoxic volatile metabolic markers in the MDA-MB-231 breast cancer cell line

Theo Issitt[1,2], Matthew Reilly[1], Sean T. Sweeney[1,2], William J. Brackenbury[1,2] and Kelly R. Redeker[1]*

[1]Department of Biology, University of York, York, United Kingdom, [2]York Biomedical Research Institute, University of York, York, United Kingdom

Hypoxia in disease describes persistent low oxygen conditions, observed in a range of pathologies, including cancer. In the discovery of biomarkers in biological models, pathophysiological traits present a source of translatable metabolic products for the diagnosis of disease in humans. Part of the metabolome is represented by its volatile, gaseous fraction; the volatilome. Human volatile profiles, such as those found in breath, are able to diagnose disease, however accurate volatile biomarker discovery is required to target reliable biomarkers to develop new diagnostic tools. Using custom chambers to control oxygen levels and facilitate headspace sampling, the MDA-MB-231 breast cancer cell line was exposed to hypoxia (1% oxygen) for 24 h. The maintenance of hypoxic conditions in the system was successfully validated over this time period. Targeted and untargeted gas chromatography mass spectrometry approaches revealed four significantly altered volatile organic compounds when compared to control cells. Three compounds were actively consumed by cells: methyl chloride, acetone and n-Hexane. Cells under hypoxia also produced significant amounts of styrene. This work presents a novel methodology for identification of volatile metabolisms under controlled gas conditions with novel observations of volatile metabolisms by breast cancer cells.

KEYWORDS

hypoxia, VOC, cancer, breast cancer, volatile flux, hypoxic, GC/MS, metabolism

1 Introduction

The human "volatilome" describes the production and metabolism by the human body of small, carbon-containing compounds called volatile organic compounds (VOCs) which are gaseous at room temperature and pressure (Amann et al., 2014; Drabinska et al., 2021). VOCs can be found in abundance in the breath and are reflective of processes within the body (Drabinska et al., 2021; Issitt et al., 2022a). Although fluctuations of VOCs vary between individuals and throughout the day, disease specific "volatile fluxes," or biomarkers, could provide opportunities to non-invasively diagnose disease, monitor treatment and measure bodily functions (Issitt et al., 2022a; Issitt et al., 2022b).

The clinical potential of VOCs in diagnosis has been shown by a number of published breath studies (Issitt et al., 2022a). Diagnostic accuracy using breath VOC biomarkers has been achieved for a wide range of conditions, including various types of cancer (Issitt et al., 2022a; Jia et al., 2019), liver disease (De Vincentis et al., 2019), diabetes (Das et al., 2016),

transplant rejection (Phillips et al., 2004), infections of the lung (Issitt et al., 2022a; Beccaria et al., 2018), liver function (using labelled VOCs) (Sangnes et al., 2019) and other conditions (Issitt et al., 2022a). Each study may independently achieve high sensitivity of disease detection (i.e., >90%) but the reported compounds often do not translate between studies, slowing clinical application through conflicting and confounding results (Issitt et al., 2022a). However, our recent meta-analysis has shown underlying trends in chemical functional groups from published studies supporting potential clinical application (Issitt et al., 2022a). It is clear that in order to identify effective biomarkers more targeted methodological approaches are required to overcome variability (Issitt et al., 2022a; Hanna et al., 2019).

VOC profiles from cell types associated with pathological conditions have been identified, for example, differences between breast (Issitt et al., 2022b; Lavra et al., 2015), liver (Mochalski et al., 2013) and mesothelioma (Little et al., 2020) cancer cell lines. However, cellular VOC studies tend to be non-stressed cells in high (21%, atmospheric) oxygen conditions, which is not consistent with many disease or normal physiological states. To accelerate biomarker discovery, we propose models of pathophysiological stress. For example; stress from reactive oxygen species (ROS) induces alkane release in breast cancer cells (Liu et al., 2019), VOCs which have been observed in the breath of ROS associated conditions (Issitt et al., 2022a).

Hypoxia is a persistent reduction in oxygen from normal physiological conditions (normoxia). It is characteristic of a range of diseases, including, pulmonary hypertension (Young et al., 2019) and cancer (SamantaandSemenza, 2018). It induces a range of metabolic alterations, including reduction in adenosine triphosphate generation and inhibition of fatty-acid desaturation through hypoxia inducible factor activity (WheatonandChandel, 2011; SamantaandSemenza, 2018; Young et al., 2019), which can produce alterations in a range of associated breath volatiles (Harshman et al., 2015; Mazzatenta et al., 2021). Despite its relevance to pathophysiology, hypoxic volatiles have yet to be investigated *in vitro*. This is partially due to the challenges associated with development of a headspace sampling tool which can maintain an hypoxic environment. While volatile compounds in the available, limited, published studies associated with hypoxia show variation in breath (Harshman et al., 2015; Mazzatenta et al., 2021), translatable studies are required for target biomarker discovery.

Biomarker discovery in appropriate biological models can accelerate clinical delivery by identifying and allowing targeted analytical approaches, separating methodical challenges from pathology, and improving sensitivity. Multi-timepoint sampling and approaches considering local environment will also accelerate clinical application of breath diagnostics and consideration of methodological challenges around clinical application should drive experimental design. We have previously demonstrated a platform and method for both identification of VOC metabolisms in cellular headspace over time and VOC changes in response to cellular stress (Issitt et al., 2022b). However, models of pathological conditions require further investigation to ensure biomarker discovery is translatable from cell to human.

One of the primary sources of variance within the published literature revolves around methodology. Methods of breath VOC analysis can be split into 3 main sections where variability between studies can arise: initial collection, sample transfer and analytical approach. There are many effective breath collection methods for analysis of VOCs, such as simply breathing into a specialised bag or use of specialised technologies (Hanna et al., 2019; Di Gilio et al., 2020). Many studies use single time point collection (Issitt et al., 2022a), considering presence verses absence, which can miss valuable metabolic information, particularly volatile uptake, driven via chemical reactions reflective of cellular state or through cellular metabolism. Furthermore, variability in local environment influences and reduces reported outcome precision (Issitt et al., 2022a; Di Gilio et al., 2020; Doran et al., 2017) and approaches should consider sampling the environment (i.e., ambient air) along with breath (Hanna et al., 2019). A sample, once collected, is then transferred, either directly or indirectly (such as through chemical traps) to an analytical instrument. There are two main analytical approaches for discovery and accurate detection of VOCs: targeted and untargeted. Utargeted approaches, investigating the breath of patients, are capable of identifying relatively concentrated material (ppbv) whereas targeted approaches generally are capable of quantifying lower concentrations (pptv). Untargeted approaches therefore may miss changes in important, low-concentration compounds, while targeted approaches can only look only for a limited number of known compounds of interest, reducing discovery potential.

Here, hypoxic stress is applied to a well-studied breast cancer cell line with the intent of identifying process and disease-linked physiological volatile metabolisms specifically linked to low oxygen conditions, so that more accurate diagnostic tools can be developed and applied in the clinic. Both targeted and untargeted analyses are applied after sampling with a static headspace method that accounts for the ambient air background and allows quantification of cellular uptake of VOCs. It was predicted that upon successful maintenance of a hypoxic environment, cellular VOC profiles from hypoxic versus hyperoxic cellular models would alter significantly.

2 Methods

Methods for culture of MDA-MB-231 cells, headspace sampling from custom chambers and GC/MS analysis have been previously described in detail (Issitt et al., 2022b).

2.1 Cell culture

MDA-MB-231 breast cancer cells (a gift from Professor Mustafa Djamgoz, Imperial College London) were grown in Dulbecco's Modified Eagle Medium (DMEM, Thermo Scientific, Waltham, MA, United States), 25 mM glucose, supplemented with L-glutamine (4 mM) and 5% foetal bovine serum (Thermo Scientific, Waltham, MA, United States). Cell culture medium was supplemented with 0.1 mM NaI and 1 mM NaBr (to model physiological availability of iodine and bromide). All cells were grown at 37°C with 5% CO_2.

Prior to volatile collection, cells were trypsinised, and 500,000 cells were seeded into 8 mL complete media in 10 cm polystyrene cell culture dishes. Cells were then allowed to attach for 3-4 h, washed with warm PBS and 6 mL treatment media was applied. Volatile headspace sampling was performed 24 h later.

2.2 Induction of the hypoxic environment and VOC headspace sampling

Cells were placed in static headspace chambers as previously described (Issitt et al., 2022b) with new, clean silicon gaskets. Low oxygen, hypoxic gas (1% O_2, 5% CO_2, and 94% N_2; purchased from BOC Specialty Gases, Woking, United Kingdom) was flushed through the chambers at a rate of 4 L/min for 10 min (chamber volume = 25 L). Chambers were then closed and placed at 37°C for 2 h to allow residual oxygen in the media to equilibrate with chamber headspace. Chambers were then flushed again at a rate of 4 L/min for 10 min, sealed and returned to 37°C.

After a further 24 h, chambers were flushed again at a rate of 4 L/min for 10 min 15 mL of gas standards (MeCl, 520 ppb (parts per billion); MeBr, 22 ppb; MeI, 26 ppb; DMS, 110 ppb; CFC-11, 400 ppb and $CHCl_3$, 110 ppb; BOC Specialty Gases, Woking, UK) were then injected into the chambers through a butyl seal and time zero sample taken. Injected compounds are either known metabolites for cancer cells, or internal standards (CFC-11) for the analysis and quantification of metabolism. Final chamber concentrations were similar to environmental concentrations, e.g., MeCl, 1.2 ppb and MeBr 0.05 ppb, particularly more polluted urban spaces (Redeker et al., 2007). Injected gases are the same as those used for calibration. Compounds not injected but detected at first time point, due to residual presence from laboratory air, (including isoprene, acetone, 2-MP, 3-MP and n-hexane) were quantified. Two time zero (T0) samples were taken using an evacuated 500 mL electropolished stainless steel canister (LabCommerce, San Jose, United States) through fine mesh Ascarite® traps (Archbold et al., 2005), after which the chamber was resealed and left on a platform rocker on its slowest setting for 120 min, at which point two further air samples (T1) were collected. Duplicate samples were analysed with targeted and untargeted MS approaches.

Cells were removed from the chamber, washed with PBS twice and lysed in 500 µL RIPA buffer (NaCl, 5 M; 5 mL Tris-HCl, 1 M, pH 8.0; 1 mL Nonidet P-40; 5 mL sodium deoxycholate, 10%; 1 mL SDS, 10%) with protease inhibitor (Sigma-Aldrich, Roche; Mannheim, Germany). Protein concentration of lysates were determined using BCA assay (Thermo Scientific, Waltham, MA, United States).

Media alone was taken through exactly the same process as cells. This has been visualised in Supplementary Figure S1. Only acetone was shown to have any significant variability between conditions. These media blank outcome averages were subtracted from respective cellular samples prior to protein normalisation. Comparative controls include lab air blanks and those data available from the dataset and collection method published previously which created and quantified metabolic fluxes of volatile compounds from MDA-MB-231 under hyperoxic (lab air) conditions (Issitt et al., 2022b).

2.3 Sample collection and GC/HID analysis

Ten mL headspace samples were taken from chambers using an airtight syringe (10 mL, SGE, Trajan, Milton Keynes, UK). 1% O_2 (BOC Specialty Gases, Woking, UK) was flushed through sealed chambers containing 6 mL DMEM as described for cell treatments. Samples were taken at 5 and then 10 min post initial flush. In order to replicate cell treatments, the chamber was then closed for 2 h, then flushed for 10 min, after which an air sample was taken. A further 20 min flush with 1% O_2 air was employed and the chamber was closed, placed at 37°C, and left to incubate for 24 h, at which time the final sample was taken.

Air samples were immediately analyzed with a SRI 8610C Gas Chromatograph connected to a SRI 8690-0030 Helium Ionisation Detector (GC/HID (SRI Instruments Europe GmbH, Torrance, CA, United States). Peak separation was achieved using a Restek© PORAPAK Q porous polymer column (1.83 m × 2.1 mm ID × 3.175 mm OD), a solenoid switching valve (for backflushing CO_2) and a Restek© MOLECULAR 5 A sieve column (0.91 m × 2.1 mm ID × 3.175 OD) (Restek©, Bellefonte, PN, United States) connected in series. Helium was used as a carrier gas at 18 psi, and the flow rate and column temperatures (50°C) were maintained during separation. The valve was switched at 1.5 min to backflush the PORAPAK Q column. Measurement of compounds eluted from the MOLECULAR 5 A sieve was achieved by using an SRI 8690-0030 Helium Ionisation Detector. SRI PeakSimple (version 453) software was used to generate a digital chromatograph for each sample and O_2 was quantified by comparing the peak area to known standards.

The standard curve was developed by flushing 120 mL Wheaton vials with butyl stoppers with pure nitrogen (BOC Gases, Woking, UK) for 30 min. Ten mL of nitrogen only was injected to establish a background control. Because atmospheric air at sea level contains 21% O_2, lab air was injected at 1%, 2%, 10%, 20%, and 30% within the N_2-filled vial to generate a standard curve consisting of 0%, 0.21%, 0.42%, 2.1%, 4.2%, and 6.3% and 21% (lab air only). Peak areas were integrated using Graphpad (Prism), and Padé (1, 1). Linear regression demonstrated an R squared value of 0.96.

2.4 GC/MS analysis of VOCs

Collected canister samples were transferred to a liquid nitrogen trap through pressure differential. Pressure change between beginning and end of "injection" was measured, allowing calculation of the moles of canister collected air injected Sample in the trap was then transferred, via heated helium flow, to an Aglient/HP 5972 MSD system (Santa Clara, CA, United States) equipped with a PoraBond Q column (25 m × 0.32 mm × 0.5 µm film thickness) (Restek©, Bellefonte, PN, United States). Targeted samples were analyzed in selected ion monitoring (SIM) mode, and untargeted samples in full scan (SCAN) mode with the mass range of 45–200 amu. The mass spectrometer was operated in electron impact ionization mode with 70 eV ionization energy, and transfer line, ion source, and quadrupole temperatures of 250, 280, and 280, respectively. For details on SIM and significantly altered, identified SCAN compounds, see Table 1. All samples were analysed within 6 days of collection. The oven program for both SIM and SCAN analyses were identical and are as follows: 35°C for 2 min, 10°C/min to 155°C, 1°C/min to 131°C, and 25°C/min to 250 with a 5 min 30 s hold.

Calibration was performed using standard gases (BOC Specialty Gases, Woking, UK). Linear regression of calibration

TABLE 1 Retention times, mass charge ratios and GC/MS modes used to characterise individual VOCs. SIM and SCAN refer to selected ion monitoring and full mass scanning (targeted and untargeted) GC/MS modes.

Compound	Retention time (min)	Mass charge ratio (m/z)
SIM		
Methy l chloride (MeCI)	7.6–7.9	50,52
Methy l bromide (MeBr)	10.3–10.4	94,96
Trichloroflouromethane (CFC-11)	15.0–15.3	101,103
Methy l iodide (Me l)	15.4–15.7	127,142
Dimethyl Sulfide (DMS)	16.2–16.5	62
acetone	18.2–18.4	58
lsoprene	18.4–18.6	Total ion count
Trichloromethane (CHCl3)	25.4–25.7	83,85
2-Methyl pentane (2-MP)	27.6–27.8	43,57
3-Methy lpentane (3-MP)	28.0–28.2	43,57
n-Hexane (n-Hex)	28.5–28.7	43,57
SCAN		
Styrene	33.3–33.5	45–200 amu

curves confirmed strong, positive linear relationships between observed compound peak areas and moles of gas injected for each VOC (r^2 > 0.9 in all cases). For compounds not purchased in gaseous state (BOC Specialty gases, as above), 1-2 mL of compound in liquid phase was injected neat into butyl sealed Wheaton-style glass vials (100 mL) and allowed to equilibrate for 1 h. One mL of headspace air was then removed from neat vial headspace using a gas tight syringe (Trajan, SGE) and injected into the headspace of a second 100 mL butyl sealed Wheaton-style glass vial. This was then repeated, and 1 mL of the 2nd serial dilution vial was injected into the GC/MS system with 29 mL of lab air to give ppb concentrations. This was performed for methanethiol (MeSH, SPEXorganics, St Neots, UK), isoprene (Alfa Aesar, Ward Hill, MA, United States), acetone (Sigma-Aldrich, Burlington, MA, United States), 2- & 3-methyl pentane and n-hexane (Thermo Scientific, Waltham, MA, United States). Reported compounds detected by the GC/MS were confirmed by matching retention times and mass–charge (m/z) ratios with known standards.

$$Equation\ 1: \quad [VOC]\,(ppt) = \frac{CF\ x\ 10^{12}\ x\ Peak\ area\ x\ Calibration\ slope}{n} \qquad (1)$$

Equation 1 outlines the approach to calculating VOC concentrations in parts-per-trillion-by-volume, or pptv. Here *Peak area* refers to the combined peak areas for the mass-charge ratios identified in Table 1. Multiplying *Peak areas* by their associated calibration curves (*Calibration Slope*) generate molar amounts which, when divided by the number of moles of headspace air injected (n), generate a unitless (moles compound/moles of air) ratio. Pptv concentrations are then obtained by multiplying this unitless ratio by 1 × 10^{12}. For clarity, part-per-billion-by-volume values would be obtained by multiplying the unitless ratios by 1 × 10^9, or one billion. Sample VOC concentrations were then normalised to CFC-11 concentrations [240 parts-per-trillion-by-volume (pptv)] through multiplication by a "correction factor," or *CF*, Eq. 1). CFC-11 was used as an internal standard, since atmospheric concentrations of CFC-11 are globally consistent and stable (Redeker et al., 2007). Quantification of Styrene was done as above but normalisation to CFC-11 was not possible under flushed, hypoxic conditions.

To account for differences in rates of cellular proliferation over 24 h, cellular results from GC/MS analyses were normalised to protein content at time of sampling using a BCA assay. When comparing media blanks to cellular assays results are reported in grams compound per Petri dish per hour.

Data has been made publicly available at the National Institute of Health Metabolomics workbench (project PR001638, DOI: http://dx.doi.org/10.21228/M8ZX4D) (Sud et al., 2016).

2.5 Hydrogen peroxide (amplex red) assay

Experiments were performed in phenol red free DMEM. DMEM containing 50 µM Amplex Red reagent (Thermo Scientific, Waltham, MA, United States) and 0.1 U/mL horse radish peroxidase (HRP, Thermo Scientific, Waltham, MA, United States) was added to cells in 12 well dishes (500 µL per well) for 15 min following 24 h in hypoxic or control conditions. Fluorescence at 590 nm was measured with a plate reader (Clariostar, BMG, Ortenberg, Germany) and compared against a H_2O_2 standard curve for quantification.

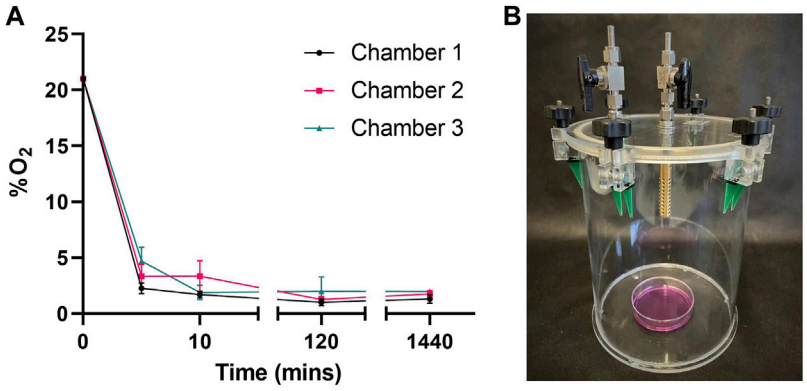


FIGURE 1
Chambers maintain hypoxic conditions over 24 h **(A)** Oxygen (O_2) content in 3 custom made chambers containing 6 mL media was measured following a 10 min flush, 2 h dwell and another 10 min flush (20 mins) with 1% O_2, 5% CO_2 gas mix. O_2°/o was then measured following chambers being sealed for 1440 min (24 h). Mean ± SEM; n = 3 **(B)** Image of collection chamber.

2.6 Statistics

Figures were assembled and statically analysed in Graphpad Prism version 9.3. VOCs were separated based on their flux amount to allow visualisation on the *y*-axis and were analysed this way. Two-way ANOVA with Bonferroni post-hoc analysis was performed for graphs with multiple factors was performed (Figures 2A, B; Supplementary Figures S1A, B). One-way ANOVA with Tukey post-hoc analysis was performed for acetone analysis (Figure 2B; Supplementary Figure S1B). Student's t-test was performed for Styrene analysis against media only as none was detected for control cells, and these were presented on the graph for visual information. Amplex red data was analysed using Student's t-test.

3 Results

3.1 Chambers maintain low oxygen conditions over 24 h

To confirm chambers maintained hypoxic conditions over 24 h we sampled gas from chambers throughout our method, measuring O_2. When flushed with reduced oxygen air (1%) for 5 min, oxygen levels rapidly fell from atmospheric 21% to between 6% and 2% (Figure 1). After 10 min of reduced oxygen flushing, each chamber held less than 5%. Chambers left for 2 h (120 min) to allow media to equilibrate and flushed for 10 min revealed average O_2 levels of 1.15% ± 1.03 (Ch 1), 1.34% ± 0.93 (Ch 2) and 1.98% ± 4.07 (Ch 3) respectively. Sealed chambers maintained low oxygen levels over 24 h with average O_2 levels of 1.31% ± 1.31 (Ch 1), 1.76% ± 1.02 (Ch 2) and 1.96% ± 0.28 (Ch 3) respectively.

3.2 Hypoxia induces differing volatile fluxes in breast cancer cell line MDA-MB-231

Persistent hypoxia over 24 h induced significant changes in flux for 3 targeted compounds (SIM analysis); MeCl, acetone and n-hexane (but not hexane isomers; 2-methyl pentane, or 3-methyl pentane), when compared to control (Figures 2A–C). MeCl was taken up by cells under hypoxia and released by cells under hyperoxic cell culture conditions. n-Hexane was produced by hyperoxic control cells while those under hypoxia consumed hexane.

3.3 Production of styrene under hypoxic conditions

Cells maintained under hypoxic conditions significantly produced styrene as determined by untargeted GG/MS approaches (Figure 3A). Styrene was not found in the headspace of control cells (ND, or not detected) and styrene fluxes in media blanks were not significantly different from zero, while fluxes from hypoxic cells were significantly different from media blanks. Styrene was identified through spectral matching, followed by known standard injections. No other compounds were found to be significantly altered using the untargeted SCAN method.

3.4 Reactive oxygen species are reduced under hypoxia

Changes in volatiles, including alkanes, have been linked to increases in ROS (Calenic et al., 2015). The observed uptake of n-Hexane in hypoxic MDA-MB-231 cells could therefore be correlated with alterations in ROS levels in these cells. Following 24 h exposure to hypoxic conditions, ROS, as determined by Amplex Red assay, showed significant reduction compared to control (Figure 3B).

4 Discussion

Static headspace sampling chamber was demonstrated to be capable of maintaining a low oxygen environment for >24 h, as

10.3389/fmolb.2023.1178269

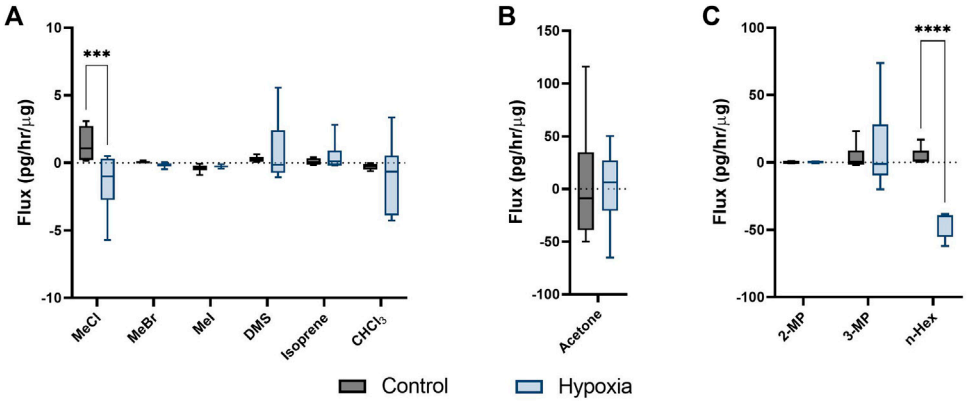


FIGURE 2
Cellular volatile response to hypoxia. Volatile flux (pg/hr/µg) for MDA-MB-231 cells in control conditions or hypoxia (24 h). Media subtracted and protein normalised VOC flux for MDA-MB-231 control cells (n = 6) and cells in hypoxia (n = 6). CHCl3, chloroform; OMS, dimethyl sulfide; MeBr, methyl bromide; MeCl, methyl chloride; MeI, methyl iodide; MeSH, methanoethiol; 2-MP, 2 methyl pentane; 3-MP, 3 methyl pentane; n-Hex, n-hexane. Boxplot whiskers show median ± Tukey distribution, n = 6. Two way ANOVA followed by Bonferroni post-hoc test was performed for **(A,B)**. One way ANOVA with Tukey post-hoc test performed for B; ***$p < 0.001$; ****$p < 0.0001$.

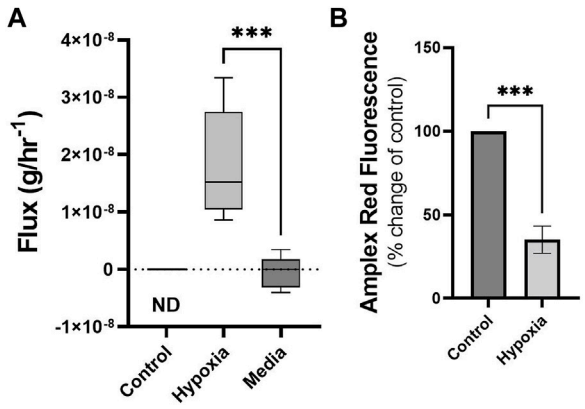


FIGURE 3
Cells under hypoxic conditions produce styrene and exhibit reduced ROS. Volatile flux (g/hr^{-1}) for styrene from MDA-MB-231 cells in control conditions or hypoxia and media only (24 h). Non Detected (ND) for control cells. Amplex Red assay was performed following 24 h incubation as a measure of reactive oxygen species (ROS), H2O2. Shown as percentage change from relative control. Boxplot whiskers show median ± Tukey distribution, A; n = 6. Student's T-test was performed for **(A,B)**, ***$p < 0.001$.

evidenced by chamber concentrations and cellular ROS response. Furthermore, VOCs from cells maintained under low oxygen conditions can be sampled, and that these cells produce a significantly different volatile profile than either media blanks or identical cells exposed to hyperoxic conditions.

Two out of 10 compounds targeted by SIM revealed quantifiable, differential metabolic responses in cells exposed to hypoxic conditions (1% O_2) relative to those maintained in normal laboratory conditions (21% O_2, physiological hyperoxia). Our previous results quantified alterations in MDA-MD-231 cells for these volatiles after treatment with the chemotherapeutic agent Doxorubicin. When placed under cellular stress through Doxorubicin treatment only MeCl showed a similar stress response (enhanced uptake). In contrast, hexane (or hexane isomers) were not consumed or degraded significantly (Issitt et al., 2022b).

Over 24 h of doxorubicin treatment has been shown to increase ROS (Pilco-FerretoandCalaf, 2016) whereas the opposite has been shown in cells maintained in hypoxic conditions (Sgarbi et al., 2018). A significant reduction was demonstrated in ROS in MDA-MB-231 cells following 24 hs of hypoxia (Figure 3B). Cellular stress response mechanics and differences in cellular state could therefore be identified and quantified through volatile metabolic approaches. Alkanes have been positively correlated with ROS previously (Calenic et al., 2015), here a decrease was demonstrated in n-hexane within hypoxic cells (Figure 2C) with diminished ROS content while in cells treated with doxorubicin, non-significant increases were observed (Issitt et al., 2022b). Metabolic consumption n-hexane is through currently unidentified processes, however the demonstration of variable consumption of a compound demonstrates a potential biomarker dynamics missed by studies only focusing on production. Acetone, hexanes and other compounds shown here are commonly found in urban environments (Redeker et al., 2007) and so their expression in the breath is driven through a combination of equilibration in the bloodstream and chemical/biological uptake processes within the body.

The production of styrene by cells under hypoxia could be a defining VOC biomarker for cancer since hypoxia is characteristic of the tumour microenvironment (SamantaandSemenza, 2018). Our recent review showed that, despite substantial variability in reported outcomes, aromatics are powerful descriptors of cancer (Issitt et al., 2022a). Five studies have previously reported styrene in the breath of lung cancer patients using untargeted approaches (Phillips et al., 1999; Chen et al., 2005a; Peng et al., 2009; Rudnicka et al., 2011; Corradi et al., 2015; Koureas et al., 2020). Styrene has also been reported as higher in the breath of lung cancer patients in studies using other approaches (Chen et al., 2005b; Nardi-Agmon et al., 2016;

Wang et al., 2022). However, styrene has been shown to be higher in the breath of smokers (Koureas et al., 2020) and so is often considered, along with other aromatics compounds, to be a confounding contaminant since high percentages of lung cancer patients have a history of smoking. (Issitt et al., 2022a). Styrene has also been reported in the breath of patients with ovarian (Amal et al., 2015), gastric (Amal et al., 2013; Amal et al., 2016) and liver (Qin et al., 2010) cancers.

Styrene utilisation as a breath-based diagnostic biomarker may be challenging since environmental contamination would need to be considered (Hanna et al., 2019). The presented method accounts for environmental VOCs through a flux analysis that incorporates two temporal sampling points, a starting sample following equilibration with the local atmosphere and a second sample at a later time point. This allows us to determine when available environmental volatiles are being added to (metabolically produced) or consumed/degraded by cells. This is important where environmental VOCs may mask effects or differences, such as high traffic, urban environments or perfumed indoor spaces. It is worth stating however, that the observed degradation may be purely non-targeted chemical reactivity with available enzymes or active compounds. However, to some degree whether the process is substrate-specific or nonspecific is unimportant. A different cell response under stress was observed, which points to different cellular states, inclusive of differing enzyme compositions, and points to new and novel potential biomarkers.

Environmental-correction sampling approaches such as this chamber headspace method may present an opportunity to overcome challenges to applications within the clinic, particularly with breath samples taken from ambient air as well as exhalate from the patient. The two time point sampling approach is particularly important since production of compounds with large initial concentrations, or consumption/degradation of compounds are often challenging to detect using single time point sampling methods.

It was observed that cellular consumption of VOCs (MeCl, acetone and n-Hexane) is descriptive of hypoxic stress and that chemotherapeutic stress also induces consumption of VOCs (Issitt et al., 2022b); notably MeCl. To our knowledge this is the first example of a controlled environment experiment performed under low oxygen conditions that both a) quantifies VOC fluxes from a cellular model and b) utilises a VOC injection of gases to monitor ongoing anaerobic metabolism of compounds. We have demonstrated a novel method for induction and maintenance of low oxygen for the study of volatile fluxes. This approach allows new dynamics to be explored for the discovery of cell to patient translational biomarkers. It is perhaps worthy of note that many of the published methods for breath research would not have identified or quantified the methyl chloride or hexane results, due to the small changes (pptv) observed.

It was previously reported that cellular "volatile metabolic flux" can separate cell type and response to chemotherapeutic stress (Issitt et al., 2022b). This chamber-based method has also been successfully used with mice models, quantifying both mouse-breath and faecal volatiles (Issitt et al., 2022b). Here, this chamber-based approach was demonstrated to identify cells under hypoxic stress. A novel method is demonstrated to identify hypoxia-induced VOCs, potential biomarkers of cancer. Importantly these biomarkers are both produced and consumed by cells under hypoxic stress. MeCl, n-hexane and styrene are clinically interesting compounds requiring further investigation. The compounds reported here have been reported as present in human breath (Shahi et al., 2022) and we have shown that these compounds vary in response to cellular stress, from previously published doxorubicin (Issitt et al., 2022b) and here, hypoxic stress. Together this suggests they are able to differentiate cellular response due to pathophysiological differences. These compounds are from diverse functional chemical groups and we have previously demonstrated the ability of functional chemical groups to separate disease groups with greater ability than individually considered compounds (Issitt et al., 2022a). A functionally diverse group of VOCs could give greater power when building a "breath-print" for diagnosis (Issitt et al., 2022a).

5 Conclusion

The work presented here demonstrates a novel methodology investigating volatile metabolisms in a controlled environment for volatile biomarker discovery. Using this method we have shown distinct changes in VOCs, demonstrating the potential for VOCs in defining metabolic alterations to environmental changes.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, TI, WB, and KR; Data curation, TI; Formal analysis, TI; Funding acquisition, SS, WB, and KR; Investigation, TI; Methodology, TI, MR, and KR; Project administration, TI and SS; Resources, MR and WB; Visualization, TI; Writing—original draft, TI; Writing—review and editing, SS, WB, MR, and KR. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the White Rose Mechanistic Biology Doctoral Training Program, supported by the Biotechnology and Biological Science Research Council (BBSRC) BB/M011151/1.

Acknowledgments

The authors would like to acknowledge the support provided by Mark Bentley in the University of York Department of Biology workshop and Professor James Chong.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1178269/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Volatile flux of media controls. Volatile flux in grams per hour (g/hr) for control media (n = 8) or media in hypoxia (n = 6) in 1Dem dishes. CHCl3, chloroform; OMS, dimethyl sulfide; MeBr, methyl bromide; MeCl, methyl chloride; MeI, methyl iodide; MeSH, methanoethiol; 2-MP, 2 methyl pentane; 3-MP, 3 methyl pentane; n-Hex, n-hexane. Boxplot whiskers show median ± Tukey distribution, n = 6. Two way ANOVA followed by Bonferroni post hoc test was performed for **(A,B)**. One way ANOVA with Tukey post-hoc test performed for B; ***$p < 0.001$.

References

Amal, H., Leja, M., Broza, Y. Y., Tisch, U., Funka, K., Liepniece-Karele, I., et al. (2013). Geographical variation in the exhaled volatile organic compounds. *J. Breath. Res.* 7, 047102. doi:10.1088/1752-7155/7/4/047102

Amal, H., Leja, M., Funka, K., Skapars, R., Sivins, A., Ancans, G., et al. (2016). Detection of precancerous gastric lesions and gastric cancer through exhaled breath. *Gut* 65, 400–407. doi:10.1136/gutjnl-2014-308536

Amal, H., Shi, D. Y., Ionescu, R., Zhang, W., Hua, Q. L., Pan, Y. Y., et al. (2015). Assessment of ovarian cancer conditions from exhaled breath. *Int. J. Cancer.* 136, E614–E622. doi:10.1002/ijc.29166

Amann, A., Costello Bde, L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., et al. (2014). The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath. Res.* 8, 034001. doi:10.1088/1752-7155/8/3/034001

Archbold, M. E., Redeker, K. R., Davis, S., Elliot, T., and Kalin, R. M. (2005). A method for carbon stable isotope analysis of methyl halides and chlorofluorocarbons at pptv concentrations. *Rapid Commun. Mass Spectrom.* 19, 337–342. doi:10.1002/rcm.1791

Beccaria, M., Bobak, C., Maitshotlo, B., Mellors, T. R., Purcaro, G., Franchina, F. A., et al. (2018). Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *J. Breath. Res.* 13, 016005. doi:10.1088/1752-7163/aae80e

Calenic, B., Miricescu, D., Greabu, M., Kuznetsov, A. V., Troppmair, J., Ruzsanyi, V., et al. (2015). Oxidative stress and volatile organic compounds: Interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* 13, 1020–1030. doi:10.1515/chem-2015-0105

Chen, X., Cao, M. F., Li, Y., Hu, W. J., Wang, P., Ying, K. J., et al. (2005b). A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method. *Meas. Sci. Technol.* 16, 1535–1546. doi:10.1088/0957-0233/16/8/001

Chen, X., Cao, M., Hao, Y., Li, Y., Wang, P., Ying, K., et al. (2005a). A Non-invasive detection of lung cancer combined virtual gas sensors array with imaging recognition technique. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2005, 5873–5876. doi:10.1109/IEMBS.2005.1615826

Corradi, M., Poli, D., Banda, I., Bonini, S., Mozzoni, P., Pinelli, S., et al. (2015). Exhaled breath analysis in suspected cases of non-small-cell lung cancer: A cross-sectional study. *J. Breath. Res.* 9, 027101. doi:10.1088/1752-7155/9/2/027101

Das, S., Pal, S., and Mitra, M. (2016). Significance of exhaled breath test in clinical diagnosis: A special focus on the detection of diabetes mellitus. *J. Med. Biol. Eng.* 36, 605–624. doi:10.1007/s40846-016-0164-6

De Vincentis, A., Vespasiani-Gentilucci, U., Sabatini, A., Antonelli-Incalzi, R., and Picardi, A. (2019). Exhaled breath analysis in hepatology: State-of-the-art and perspectives. *World J. Gastroenterol.* 25, 4043–4050. doi:10.3748/wjg.v25.i30.4043

Di Gilio, A., Palmisani, J., Ventrella, G., Facchini, L., Catino, A., Varesano, N., et al. (2020). Breath analysis: Comparison among methodological approaches for breath sampling. *Molecules* 25, 5823. doi:10.3390/molecules25245823

Doran, S. L. F., Romano, A., and Hanna, G. B. (2017). Optimisation of sampling parameters for standardised exhaled breath sampling. *J. Breath. Res.* 12, 016007. doi:10.1088/1752-7163/aa8a46

Drabinska, N., Flynn, C., Ratcliffe, N., Belluomo, I., Myridakis, A., Gould, O., et al. (2021). A literature survey of all volatiles from healthy human breath and bodily fluids: The human volatilome. *J. Breath. Res.* 15, 034001. doi:10.1088/1752-7163/abf1d0

Hanna, G. B., Boshier, P. R., Markar, S. R., and Romano, A. (2019). Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis: A systematic review and meta-analysis. *Jama Oncol.* 5, 1070–1070. doi:10.1001/jamaoncol.2018.2815

Harshman, S. W., Geier, B. A., Fan, M., Rinehardt, S., Watts, B. S., Drummond, L. A., et al. (2015). The identification of hypoxia biomarkers from exhaled breath under normobaric conditions. *J. Breath. Res.* 9, 047103. doi:10.1088/1752-7155/9/4/047103

Issitt, T., Sweeney, S. T., Brackenbury, W. J., and Redeker, K. R. (2022b). Sampling and analysis of low-molecular-weight volatile metabolites in cellular headspace and mouse breath. *Metabolites* 12, 599. doi:10.3390/metabo12070599

Issitt, T., Wiggins, L., Veysey, M., Sweeney, S. T., Brackenbury, W. J., and Redeker, K. (2022a). Volatile compounds in human breath: Critical review and meta-analysis. *J. Breath. Res.* 16, 024001. doi:10.1088/1752-7163/ac5230

Jia, Z., Patra, A., Kutty, V. K., and Venkatesan, T. (2019). Critical review of volatile organic compound analysis in breath and *in vitro* cell culture for detection of lung cancer. *Metabolites* 9, 52. doi:10.3390/metabo9030052

Koureas, M., Kirgou, P., Amoutzias, G., Hadjichristodoulou, C., Gourgoulianis, K., and Tsakalof, A. (2020). Target analysis of volatile organic compounds in exhaled breath for lung cancer discrimination from other pulmonary diseases and healthy persons. *Metabolites* 10, 317. doi:10.3390/metabo10080317

Lavra, L., Catini, A., Ulivieri, A., Capuano, R., Baghernajad Salehi, L., Sciacchitano, S., et al. (2015). Investigation of VOCs associated with different characteristics of breast cancer cells. *Sci. Rep.* 5, 13246. doi:10.1038/srep13246

Little, L. D., Carolan, V. A., Allen, K. E., Cole, L. M., and Haywood-Small, S. L. (2020). Headspace analysis of mesothelioma cell lines differentiates biphasic and epithelioid sub-types. *J. Breath. Res.* 14, 046011. doi:10.1088/1752-7163/abaaff

Liu, Y. L., Li, W. W., and Duan, Y. X. (2019). Effect of H2O2 induced oxidative stress (OS) on volatile organic compounds (VOCs) and intracellular metabolism in MCF-7 breast cancer cells. *J. Breath. Res.* 13, 036005. doi:10.1088/1752-7163/ab14a5

Mazzatenta, A., Pokorski, M., and Di Giulio, C. (2021). Volatile organic compounds (VOCs) in exhaled breath as a marker of hypoxia in multiple chemical sensitivity. *Physiol. Rep.* 9, e15034. doi:10.14814/phy2.15034

Mochalski, P., Sponring, A., King, J. L., Unterkofler, K., Troppmair, J., and Amann, A. (2013). Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) *in vitro*. *Cancer Cell Int.* 13, 72. doi:10.1186/1475-2867-13-72

Nardi-Agmon, I., Abud-Hawa, M., Liran, O., Gai-Mor, N., Ilouze, M., Onn, A., et al. (2016). Exhaled breath analysis for monitoring response to treatment in advanced lung cancer. *J. Thorac. Oncol.* 11, 827–837. doi:10.1016/j.jtho.2016.02.017

Peng, G., Tisch, U., Adams, O., Hakim, M., Shehada, N., Broza, Y. Y., et al. (2009). Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat. Nanotechnol.* 4, 669–673. doi:10.1038/nnano.2009.235

Phillips, M., Boehmer, J. P., Cataneo, R. N., Cheema, T., Eisen, H. J., Fallon, J. T., et al. (2004). Heart allograft rejection: Detection with breath alkanes in low levels (the HARDBALL study). *J. Heart Lung Transpl.* 23, 701–708. doi:10.1016/j.healun.2003.07.017

Phillips, M., Gleeson, K., Hughes, J. M. B., Greenberg, J., Cataneo, R. N., Baker, L., et al. (1999). Volatile organic compounds in breath as markers of lung cancer: A cross-sectional study. *Lancet* 353, 1930–1933. doi:10.1016/S0140-6736(98)07552-7

Pilco-Ferreto, N., and Calaf, G. M. (2016). Influence of doxorubicin on apoptosis and oxidative stress in breast cancer cell lines. *Int. J. Oncol.* 49, 753–762. doi:10.3892/ijo.2016.3558

Qin, T., Liu, H., Song, Q., Song, G., Wang, H. Z., Pan, Y. Y., et al. (2010). The screening of volatile markers for hepatocellular carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 19, 2247–2253. doi:10.1158/1055-9965.Epi-10-0302

Redeker, K. R., Davis, S., and Kalin, R. M. (2007). Isotope values of atmospheric halocarbons and hydrocarbons from Irish urban, rural, and marine locations. *J. Geophys. Res. Atmos.* 112, D16307. doi:10.1029/2006JD007784

Rudnicka, J., Kowalkowski, T., Ligor, T., and Buszewski, B. (2011). Determination of volatile organic compounds as biomarkers of lung cancer by SPME-GC-TOF/MS and chemometrics. *J. Chromatogr. B-Analyt. Technol. Biomed. Life Sci.* 879, 3360–3366. doi:10.1016/j.jchromb.2011.09.001

Samanta, D., and Semenza, G. L. (2018). Metabolic adaptation of cancer and immune cells mediated by hypoxia-inducible factors. *Biochim. Biophys. Acta Rev. Cancer.* 1870, 15–22. doi:10.1016/j.bbcan.2018.07.002

Sangnes, D. A., Softeland, E., Teigland, T., and Dimcevski, G. (2019). Comparing radiopaque markers and (13)C-labelled breath test in diabetic gastroparesis diagnostics. *Clin. Exp. Gastroenterol.* 12, 193–201. doi:10.2147/CEG.S200875

Sgarbi, G., Gorini, G., Liuzzi, F., Solaini, G., and Baracca, A. (2018). Hypoxia and IF1 expression promote ROS decrease in cancer cells. *Cells* 7, 64. doi:10.3390/cells7070064

Shahi, F., Forrester, S., Redeker, K., Chong, J., and Barlow, G. (2022). Case Report: The effect of intravenous and oral antibiotics on the gut microbiome and breath volatile organic compounds over one year. *Wellcome Open Res.* 7, 50. doi:10.12688/wellcomeopenres.17450.3

Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. (2016). Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44, D463–D470. doi:10.1093/nar/gkv1042

Wang, P. Y., Huang, Q., Meng, S. S., Mu, T., Liu, Z., He, M. Q., et al. (2022). Identification of lung cancer breath biomarkers based on perioperative breathomics testing: A prospective observational study. *eClinicalmedicine* 47, 101384. doi:10.1016/j.eclinm.2022.101384

Wheaton, W. W., and Chandel, N. S. (2011). Hypoxia. 2. Hypoxia regulates cellular metabolism. *Am. J. Physiol. Cell Physiol.* 300, C385–C393. doi:10.1152/ajpcell.00485.2010

Young, J. M., Williams, D. R., and Thompson, A. A. R. (2019). Thin air, thick vessels: Historical and current perspectives on hypoxic pulmonary hypertension. *Front. Med. (Lausanne).* 6, 93. doi:10.3389/fmed.2019.00093

frontiers | Frontiers in Molecular Biosciences

TYPE Review
PUBLISHED 08 January 2024
DOI 10.3389/fmolb.2023.1295955

OPEN ACCESS

EDITED BY
Andras Szeitz,
University of British Columbia, Canada

REVIEWED BY
Oleg Mayboroda,
Leiden University Medical Center (LUMC), Netherlands
Shangli Cheng,
Karolinska Institutet (KI), Sweden

*CORRESPONDENCE
Antonia García,
antogar@ceu.es

RECEIVED 17 September 2023
ACCEPTED 05 December 2023
PUBLISHED 08 January 2024

CITATION
Bajo-Fernández M, Souza-Silva ÉA, Barbas C, Rey-Stolle MF and García A (2024), GC-MS-based metabolomics of volatile organic compounds in exhaled breath: applications in health and disease. A review.
Front. Mol. Biosci. 10:1295955.
doi: 10.3389/fmolb.2023.1295955

GC-MS-based metabolomics of volatile organic compounds in exhaled breath: applications in health and disease. A review

María Bajo-Fernández[1], Érica A. Souza-Silva[1,2], Coral Barbas[1], Ma Fernanda Rey-Stolle[1] and Antonia García[1]*

[1]Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, Boadilla del Monte, Spain, [2]Departmento de Química, Universidade Federal de São Paulo (UNIFESP), Diadema, Brazil

Exhaled breath analysis, with particular emphasis on volatile organic compounds, represents a growing area of clinical research due to its obvious advantages over other diagnostic tests. Numerous pathologies have been extensively investigated for the identification of specific biomarkers in exhalates through metabolomics. However, the transference of breath tests to clinics remains limited, mainly due to deficiency in methodological standardization. Critical steps include the selection of breath sample types, collection devices, and enrichment techniques. GC-MS is the reference analytical technique for the analysis of volatile organic compounds in exhalates, especially during the biomarker discovery phase in metabolomics. This review comprehensively examines and compares metabolomic studies focusing on cancer, lung diseases, and infectious diseases. In addition to delving into the experimental designs reported, it also provides a critical discussion of the methodological aspects, ranging from the experimental design and sample collection to the identification of potential pathology-specific biomarkers.

KEYWORDS

volatile organic compounds, exhaled breath, breath test, gas chromatography-mass spectrometry, biomarkers

1 Introduction

1.1 Volatile organic compounds

Volatile organic compounds (VOCs) are small molecules (MW <500 Da) with low boiling points and high vapor pressures at ambient temperature. The profile of VOCs released by an organism is called the volatilome, reflecting the metabolic state and playing essential ecological and regulatory roles (Mansurova et al., 2018; Netzker et al., 2020; Sidorova et al., 2021). In humans, VOCs are released through breath, skin, feces, urine, sweat, and saliva, among others (Drabińska et al., 2021), and their origin can be endogenous and exogenous (Pleil et al., 2013). Microorganism-derived VOCs, which include symbionts, commensals, and pathogens, should be considered endogenous since they play significant roles in human health (De Vos et al., 2022).

1.2 Breath test along the history

The origin of the breath test can be traced back to ancient Greece. Hippocrates of Kos (460–370 BC) described specific types of odors associated with physiological imbalance, such as *fetor hepaticus* for liver dysfunction, *fetor oris* for halitosis, the fruity and sweet odor of patients with uncontrolled diabetes, the urine-like smell of kidney failure, and the putrid stench of lung abscess. Paracelsus, in the 16th century, further emphasized the link between "bad" breath and pathology (Fortes et al., 2017).

In the 18th century, Antoine Lavoisier discovered the role of oxygen in combustion and understood the respiratory physiology in animals (Karamanou and Androutsos, 2013). That was the origin of capnography and modern biochemistry. The sensitive detection of VOCs became possible with the introduction of colorimetry in the mid-19th century. Ethanol was isolated from breath by Francis E. Anstie, and acetone was found increased in the breath of diabetes mellitus patients by A. Nebelthau (Phillips, 1992).

Discoveries made in the 20th century are (Amann et al., 2014): mercaptans were detected in the breath of severe liver disease patients by Davidson (1949), connecting them to the *fetor hepaticus* described by Hippocrates of Kos; acetonitrile was detected in the breath of smokers by McKee et al. (1962); methanol was found in human breath (Eriksen and Kulkarni, 1963); volatile fatty acids were reported in patients with cirrhosis (Chen et al., 1970); ammonia was measured spectrometrically by Hunt and Williams (1977); and dimethyl- and trimethylamine were detected in the breath of end-stage renal disease patients (Simenhoff et al., 1977).

The turning point came when Pauling et al. (1971) published a pioneering study using gas–liquid partition chromatography to analyze body fluids and breath to investigate the influence of diet on human microbiota and health. This study detected 250 VOCs in human breath, offering promising prospects for further research in the field.

1.3 Breath test and clinical applications

Breath samples are particularly valuable for VOCs analysis. The gaseous fraction contains over 1,000 VOCs, with acetone and isoprene being the most abundant (Kuo et al., 2020; Drabińska et al., 2021).

Breath tests aim to distinguish between healthy and pathological states by analyzing exhaled breath VOC profiles, identifying pathology-specific compounds and elucidating their biochemical origin. Compared to routine diagnostic methods, they offer several advantages: they are non-invasive, cost-effective, and fast and easy to perform, have an unlimited sample size, and can be safely and repeatedly collected (Sharma et al., 2023). Despite their simplicity, to date, just a few tests are used in clinical practice, such as the fractional exhaled nitric oxide (FeNO) test for asthma diagnosis, the ^{13}C-urea breath test for *Helicobacter pylori* infection, the hydrogen/methane test to detect lactose and/or fructose intolerance, also to detect small intestine bacterial overgrowth, standard capnography based on monitoring CO_2 partial pressure levels during anesthesia and intensive care, and the alcohol breath test used by the police (Simrén and Stotzer, 2006; Buszewski et al., 2007, 2013).

Although many studies propose potential biomarkers for various pathologies, the expected clinical application of breath tests has not progressed as expected (Buszewski et al., 2007; Sharma et al., 2023). Additionally, the link between potential biomarkers and specific pathologies is not clear (Haick et al., 2014; Zou et al., 2022).

1.4 Major sources of endogenous VOCs

Oxidative stress (OS) and cytochrome P450 (CYP) enzymes are the main sources of endogenous VOCs. OS damages cellular components, such as phospholipids, proteins, and DNA, thus being involved in the development of many pathological conditions such as cancer, inflammation, and aging. Lipid peroxidation, especially of polyunsaturated fatty acids (PUFAs), is a significant source of VOCs. The breakdown of lipid peroxides produces a wide range of compounds, such as alkanes, alkenes, alcohols, aldehydes, carboxylic acids, esters, epoxides, and furans (Calenic et al., 2015; Ratcliffe et al., 2020).

CYP enzymes participate in reactive oxygen species (ROS) generation and lipid peroxidation, affecting the oxidation–reduction balance and OS, therefore also contributing to VOC generation. CYP enzymes are found in various tissues, with higher levels in the liver and enterocytes (Murray et al., 2009; Veith and Moorthy, 2018; Behrendorff, 2021).

1.5 VOCs and exhaled breath

As seen in Figure 1, almost 1,000 articles have been published with the aim of finding potential biomarkers and/or therapeutic targets for various pathologies. Lung cancer has been extensively studied (Antoniou et al., 2019; Janssens et al., 2020), although other cancers, pulmonary pathologies [e.g., asthma, chronic obstructive pulmonary disease (COPD), and obstructive sleep apnea (OSA)], gastrointestinal pathologies (e.g., Crohn's and inflammatory bowel pathologies), diabetes, and infectious diseases (e.g., viral infections, tuberculosis, and invasive aspergillosis) have also been investigated (Sethi et al., 2013; Markar et al., 2015; Van Der Schee et al., 2015; Acharige et al., 2018; Saasa et al., 2018; Hanna et al., 2019; Ghosh et al., 2020; Ratiu et al., 2020). The Human Breathomics Database (HBDB) created by Kuo et al. (2020) is a consequence of the relevance of the topic, gathering information on VOCs detected in healthy and pathological subjects.

1.6 Exhaled breath sampling

The average human expiratory volume is 500 mL, comprising three portions: dead space air, air from the airways and alveoli, and alveolar breath. Capnography can monitor the respiratory cycle, as the CO_2 level shows different trends in each portion. Consequently, breath samples can be classified into three types: mixed breath (containing the three portions), late expiratory breath (excluding

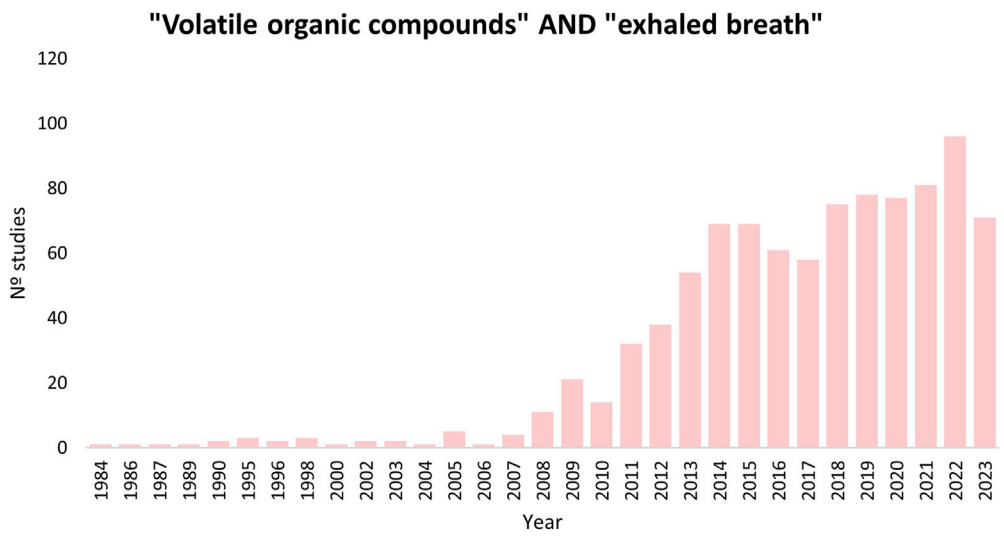


FIGURE 1
Results searching in PubMed using the terms: (volatile organic compounds) AND (exhaled breath).

dead space air), and alveolar breath (containing only the last portion of the expiration) (Beauchamp and Miekisch, 2020).

There are two types of breath analysis: online and offline. Online provides fast results and allows the volatilome to be monitored with minimal sample manipulation. Nevertheless, offline analysis (storing the sample for subsequent analysis) is the most widely used, as it enables sampling at different locations (Sola-Martínez et al., 2022). Sampling, transport, and storage are critical in the offline analysis of gas samples, since the samples may suffer from possible losses, adsorption, and artifact formation (Alonso and Sanchez, 2013). Therefore, the correct choice of sampling methodology is crucial.

Breath sampling methods can be categorized according to the type of sample collected. Devices employed for mixed breath include containers and bags with a valve system to prevent re-breathing, such as Tedlar® and Mylar bags, sorbent tubes, canisters, sampling tubes/bulbs, and the Pneumopipe device (Pennazza et al., 2014; White and Fowler, 2019). Although these devices are simpler to use, they may lead to losses, diffusion, adsorption onto the sampling device material, and potential contamination, especially with reactive VOCs (Miekisch et al., 2012; Tang et al., 2015; Beale et al., 2016). In particular, Tedlar® bags emit contaminants like N,N-dimethylacetamide, phenol, carbonyl sulfide, and carbon disulfide. To preserve sample integrity, storage time should be minimized, and analysis is recommended within 10 h (Beauchamp et al., 2008; Mochalski et al., 2009).

For collecting late expiratory or alveolar breaths, traditional sampling devices present some adaptations, such as a T-shaped mouthpiece, a spirometer system, and CO_2 and pressure sensors (Alonso and Sanchez, 2013; Tang et al., 2015). CO_2 sensors are commonly used for alveolar breath sampling because CO_2 concentrations are highest and constant in the alveolar phase (Lawal et al., 2017). Various devices are available for this purpose, either collecting a final fixed volume, based on the Haldane–Priestly approach, or using CO_2 and pressure sensors: BioVOC®, RTubeVOC, QuinTron AlveoSampler, ReCIVA, the adaptive breath sampler (ABS), breath collection apparatus (BCA), and SOFIA sampler (Phillips, 1997; Basanta et al., 2007; Beale et al., 2016; White and Fowler, 2019).

1.7 Analytical platforms: GC-MS

The concentrations of VOCs in exhalates range from parts-per-million (ppmv) to parts-per-trillion (pptv), requiring highly sensitive analytical techniques to detect these compounds. Analytical platforms used for online and offline analysis include laser spectrometry, selected ion flow tube-mass spectrometry (SIFT-MS), proton transfer reaction-mass spectrometry (PTR-MS), secondary electrospray ionization-mass spectrometry (SESI-MS), ion molecule reaction-mass spectrometry (IMR-MS), and ion mobility spectrometry (IMS). These techniques perform fast analysis and present high sensitivity, although they involve high costs and/or require skilled technicians. An alternative method, also emerging as a point-of-care tool, electronic noses (E-nose) combine selective electronic sensors, offering rapid analysis and affordability. Basically, E-noses are used to detect patterns between the samples, which are further resolved through statistical methods and machine learning. Other online approaches such as optical/laser absorption spectroscopy-based methods detect small molecules with narrow adsorption lines, commonly used for acetone analysis. Additionally, compound identification is limited, and no accepted standards ensure interoperability/normalization of methodologies. A promising approach utilizes nanomaterial-based VOC/gas sensors, which offers a wider dynamic detection range and high selectivity; however, some challenges include receptor immobilization compromising functionality, potentially irreversible reactions between VOCs and the receptor (due to high selectivity), and a reduced likelihood of VOC–receptor interaction due to the small

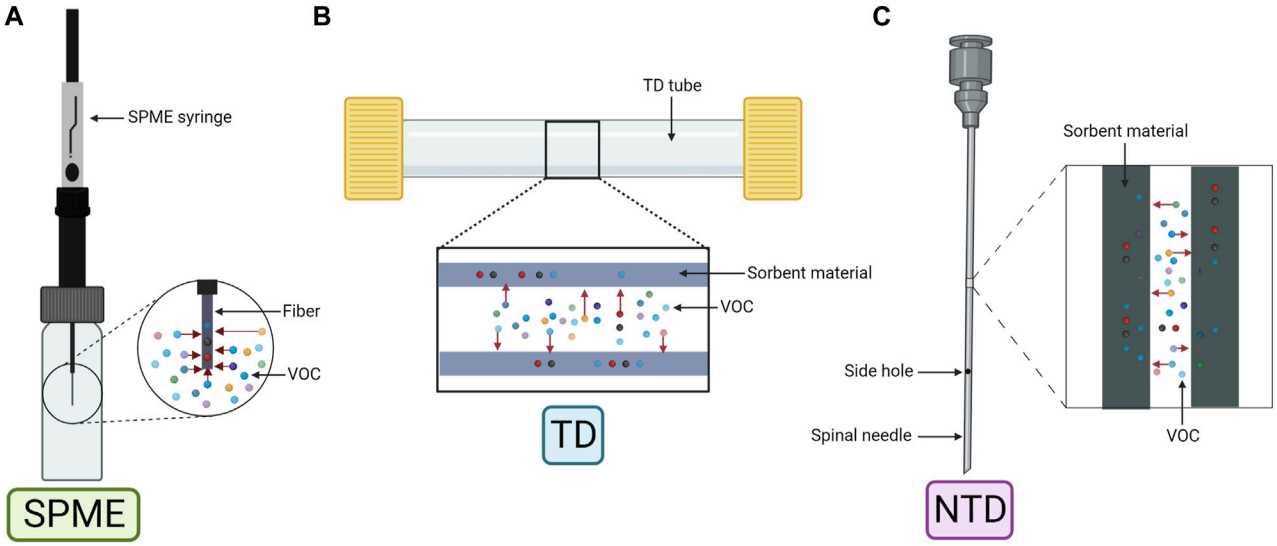


FIGURE 2
Schematic representation of the three main pre-concentration techniques. **(A)** Solid-phase microextraction (SPME). **(B)** Thermal desorption tube (TD). **(C)** Needle-trap device (NTD). Created with Biorender.com.

surface area of nanoscale elements (Buszewski et al., 2013; Bruderer et al., 2019; Wojnowski et al., 2019; Sharma et al., 2023).

GC-MS is a mature technique that is considered the "gold standard" for VOC analysis in exhaled breath (De Lacy Costello et al., 2014; Drabińska et al., 2021). It offers high sensitivity and reproducibility, and the ability to identify and elucidate unknown compounds, especially with high-resolution instruments (Sola-Martínez et al., 2022; Sharma et al., 2023). In addition to requiring an offline approach such as a pre-concentration step, GC-MS applicability may be hampered by its high costs, complex and time-consuming sampling, requirement for standardization and trained personnel, and inapplicability for online analysis (Xu et al., 2016). Nonetheless, its application to the clinical setting is valuable due to its capabilities in biomarker discovery.

1.8 Sample pre-concentration strategies

Exhaled breath samples, especially mixed breath, require enrichment before offline analysis due to low VOC concentration and high water vapor content. Pre-concentration methods usually include two consecutive steps, consisting of trapping VOCs in sorbents followed by their release via thermal desorption. Three main techniques especially suited for GC are used (Figure 2): solid-phase microextraction (SMPE), thermal desorption tubes (TD), and needle-trap devices (NTDs) (Lawal et al., 2017; Sola-Martínez et al., 2022).

SPME (Figure 2A) was first applied to human breath by Grote and Pawliszyn (1997). Equilibrium is established during sampling based on analyte and sorbent physicochemical properties within the fiber (Beauchamp and Miekisch, 2020). The fiber, coated usually with polydimethylsiloxane (PDMS), Carboxen (Car), or divinylbenzene (DVB), can also have a combination of coatings (Car and/or DVB embedded into PDMS) for a wider chemical species extraction (Trujillo-Rodríguez et al., 2020). Moreover, derivatization reactions can be performed by doping the fiber to increase the affinity of the analyte to the coating (Vas and Vékey, 2004).

TD (Figure 2B) allows longer periods of storage and ease of transport without affecting the sample. The device, composed of a stainless steel or a glass tube, contains sorbent materials like organic polymers (e.g., Tenax TA), graphitized carbon (e.g., Carbopack X) or carbon molecular sieves (e.g., Carboxen). TD can have single- or multi-bed sorbents, with the latter covering a wider range of analytes, but compromising reproducibility due to analyte–sorbent interactions (Lawal et al., 2017; Beauchamp and Miekisch, 2020; Sola-Martínez et al., 2022).

NTD (Figure 2C) is less common but shares similarities with SPME and TD. It uses a needle-shaped device filled with sorbent materials to capture compounds by drawing breath through the needle. Similar to SPME, NTD requires a small sample volume, although the sensitivity is volume dependent as in TD (Trefz et al., 2012). Storage and transportation are also similar to that of TD (Lawal et al., 2017).

1.9 Metabolomics

Metabolomics has gained significant attention in clinical research, providing insights into the pathological pathways of various pathologies. These studies can be broadly categorized into two approaches: untargeted and targeted. Untargeted metabolomics is the non-biased approach, which aims to study as many metabolites as possible to discover changes among the groups of samples, while targeted metabolomics focuses on specific metabolites, offering better sensitivity and specificity. Combining both approaches allows for hypothesis

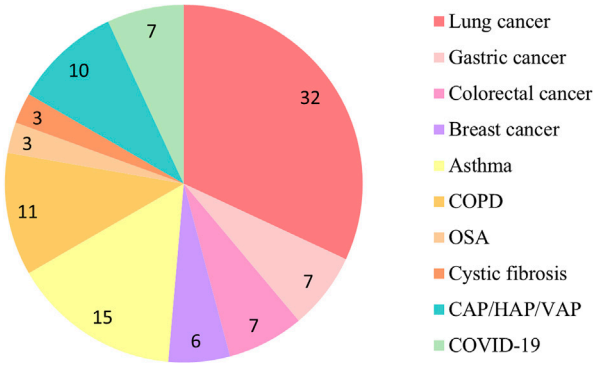


FIGURE 3
Pie plot depicting the proportion of studies for each pathology included in this review. CAP, community-acquired pneumonia; COPD, chronic obstructive pulmonary disease; HAP, hospital-acquired pneumonia; OSA, obstructive sleep apnea; VAP, ventilator-associated pneumonia.

generation (untargeted) and the validation of findings (targeted). Workflows and methodologies for both approaches have subtle differences (Patti et al., 2012).

2 Objectives and literature search

This review aims to identify potential VOC biomarkers that are consistent across different pathologies and to consolidate and discuss the methodologies employed for exhaled breath sampling and analysis. To achieve this, a literature search was conducted, focusing on studies that analyzed human exhaled breath by GC-MS published since 2012. The search strategy utilized specific keywords, such as "volatile organic compounds," "exhaled breath" or "breath test," "gas chromatography," and "mass spectrometry." The databases employed were Scopus and Web of Science. Initially, 377 articles were obtained, which were then narrowed down to 152 after title and abstract evaluation, and the articles were sorted according to the pathology studied. Finally, 70 articles focusing on 10 pathologies of significant interest were included in this review, and categorized in: cancer (such as lung, gastric, colorectal, and breast cancers), other pulmonary pathologies (comprising asthma, COPD, OSA, and cystic fibrosis), and infectious pathologies (encompassing community-acquired pneumonia (CAP)/hospital-acquired pneumonia (HAP)/ventilator-associated pneumonia (VAP) and COVID-19).

3 VOCs in exhaled breath in health and pathology

In the following sections, selected studies for each pathology are discussed, along with the identified candidate VOCs reported as pathology-specific biomarkers. Figure 3 illustrates the distribution of studies, showing that lung cancer has been the most extensively studied, followed by asthma, COPD, and CAP/HAP/VAP.

3.1 VOCs in cancer

3.1.1 Lung cancer

Lung cancer (LC) is the second most diagnosed cancer, and the leading cause of cancer-related deaths (Ferlay et al., 2021). LC comprises two major histological types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) (Rodak et al., 2021). The 5-year relative survival rates for localized NSCLC and SCLC are 65% and 30%, dropping to 9% and 3% when metastasized (2012–2018), respectively (Lung Cancer Survival Rates, 2023). Symptoms may be absent, non-specific, or easily confused with other pulmonary pathologies (Balata et al., 2022).

Low-dose computed tomography (LDCT) is the main screening tool, although it exhibits a high false-positive rate (Nooreldeen and Bach, 2021). Lung tissue biopsy, the gold standard procedure for diagnosis, determines malignancy, histological type, and TNM (tumor, nodule, and metastases) stage. However, this procedure is highly invasive and can lead to complications, such as pneumothorax and pneumonia (Zhang et al., 2020b). Indeed, the development of rapid and non-invasive early diagnostic tests is urgently required, and breath tests offer promising alternatives. Among the pathologies studied in exhaled breath, LC is the most prevalent, as this pathology is directly related to the respiratory tract.

Twenty-three studies focusing on potential biomarkers for LC are summarized in Supplementary Tables S1–S3, referring to metabolomic methodology, group comparisons, and VOC biomarkers, respectively. A Chinese group performed two untargeted studies on the same data, comparing LC patients and healthy controls (HCs). The first study (Zou et al., 2021) developed a prediction model based on the whole breath profile (308 peaks), achieving 85.0% accuracy, 83.0% sensitivity, and 85.0% specificity. Twenty-two discriminative VOCs were annotated, styrene being also found downregulated in LC patients who responded partially to treatment or remained stable (Supplementary Table S3), along with two other VOCs (dodecane, 4-methyl and α-phellandrene) (Nardi-Agmon et al., 2016). The second study (Zou et al., 2022) selected 31 VOCs as biomarkers in the univariate analysis (UVA), which showed 0.787 AUC in the multivariate analysis (MVA) after cross-validation. Additionally, eight VOCs were found to be involved in a total of 18 metabolic pathways, of which 11 were significantly altered.

A Polish group compared LC patients with HCs. Buszewski et al. (2012) divided both groups according to smoking habits, identifying 12 significant VOCs between non-smokers, 7 being upregulated when compared to active smokers. Rudnicka et al. (2014) measured 43 VOCs and developed a model with 88 features, yielding 0.970 AUC, 74.0% sensitivity, and 73.0% specificity, with dimethyl sulfide as the main discriminating VOC. Ligor et al. (2015) applied machine learning algorithms, and the final model formed by eight compounds showed an value (e.i. 0.650) AUC. In a subsequent study (Rudnicka et al., 2019), the model containing seven VOCs selected from the UVA presented an improved performance, showing 86.4% sensitivity and specificity in the test

group. Twelve VOCs were found in common between these four abovementioned studies (Supplementary Table S3).

Schallschmidt et al. (2016) focused on 24 VOCs previously selected as potential LC biomarkers, 20 being also reported in other studies (Supplementary Table S3). In the UVA, 11 and 7 VOCs were significantly altered between LC patients and HCs (non-smokers and active smokers, respectively), 8 VOCs seemingly unrelated to smoking. Moreover, four models were constructed with different subsets of the targeted VOCs, achieving the highest sensitivity (92.0%) with a subset of four VOCs, and the highest specificity (96.0%) with seven VOCs. Ethanol and octane were two target VOCs proposed as potential biomarkers in other studies (Supplementary Table S3).

Sakumura et al. (2017) reported ethanol, along with other four VOCs, in a study classifying LC and HCs using a support vector machine (SVM) algorithm, achieving 89.0% accuracy, a 94.4% true-positive ratio, and a 89.7% true-negative ratio when combining different subsets of five VOCs. Furthermore, the distance to the SVM classification boundary provided information on the cancer stage, with early-stage LC located closer to the boundary than advanced-stage LC.

Two research groups from China and Greece conducted several studies comparing LC patients, pulmonary non-malignant disease (PNMD) patients, and HCs. The Chinese group conducted 4 studies, sharing 27 VOCs (Supplementary Table S3). Wang et al. (2012) found 23 significant VOCs with AUCs >0.6, unrelated to smoking, as potential biomarkers, of which five VOCs were significant between squamous carcinoma and adenocarcinoma LC patients. The discrimination model for LC, PNMD, and HCs could correctly classify 96.5% of LCs. Zou et al. (2014) selected five VOCs as LC-specific biomarkers, achieving AUCs ranging from 0.672 to 1 in a validation cohort, with hexadecanal being the most discriminative. Additionally, Chen et al. (2021) annotated 19 VOCs that could discriminate LC from PNMD, as well as 20 VOCs that differentiated LC from HCs with AUCs of 0.809 and 0.987, respectively. Moreover, LC patients could be distinguished by histology (NSCLC and SCLC) using 20 VOCs value (e.i. 0.939) AUC and stage (early and advanced) with 19 VOCs value (e.i. 0.827) AUC. The Greek group used both targeted and untargeted approaches on the same data set. The targeted study (Koureas et al., 2020) included 19 VOCs, of which 17 VOCs were also found in other studies (Supplementary Table S3). In the UVA, seven VOCs showed significance when comparing LC, PNMD, and HCs, although no single VOC was altered between LC and PNMD. However, LC and HCs were correctly classified by either including 19 VOCs, nine VOCs selected in the UVA (LC vs. HCs), or a subset of VOCs identified by feature selection (FS) (AUCs 0.769–0.970). In the untargeted study (Koureas et al., 2021), 29 features were considered for the analysis, 18 features (12 VOCs annotated) showing significance between LC and HCs, and only 2 (1 VOC annotated) among LC and PNMD. Moreover, LC and HCs were correctly classified using either 29 features or a subset of eight features identified by FS (AUCs 0.940 and 0.960, respectively). In the case of LC and PNMD, three VOCs achieved 75.0% discrimination accuracy value (e.i. 0.820) AUC. Among the features/VOCs from both approaches, three VOCs (one from the targeted and two from the untargeted) achieved an accuracy of 72.0% in discriminating LC and PNMD value (e.i. 0.780) AUC. However, the VOC from the targeted study was detected in extremely low frequencies. Another targeted study focusing on 21 VOCs identified four upregulated VOCs in LC compared to PNMD (Corradi et al., 2015), of which two (hexane and ethylbenzene) were also included in the targeted study by Koureas et al. (2020) (Supplementary Table S3), showing elevated increased levels in adenocarcinoma LC (hexane) and in advanced-stage LC (ethylbenzene) patients.

Furthermore, both untargeted and targeted approaches were performed in the same study comparing LC, COPD, asthmatic patients, and HCs. Monedeiro et al. (2021) built an RF model with the 12 most important VOCs from the untargeted analysis, achieving an overall accuracy of 85.7%. In the following targeted approach, 29 VOCs were preselected, of which 9 were used to build the classification model that provided 91.0% overall accuracy. Additionally, Callol-Sanchez et al. (2017) identified nonanoic acid significantly altered in LC patients compared to both COPD patients and HCs in a targeted study, and Muñoz-Lucas et al. (2020) found elevated levels of propionic acid in LC patients with COPD, mainly detected in advanced-stage LC.

3.1.2 Gastric cancer

Gastric cancer (GaC) is among the five deadliest cancers in 2020, according to the World Health Organization (WHO) (Cancer, 2023). The 5-year relative survival rate is 72% when localized and decreases to 6% when distant at the time of diagnosis (2012–2018) (American Cancer Society, 2017). The main risk factors for GaC, which is predominantly sporadic (90%), include smoking, high meat intake, alcohol consumption, obesity, and *Helicobacter pylori* infection (Conti et al., 2023). Persistent *H. pylori* infection causes chronic inflammation, leading to precursor lesions associated with GaC: atrophy, metaplasia, dysplasia, and carcinoma (Conti et al., 2023).

The gold standard diagnostic technique is upper endoscopy, followed by a biopsy, although it is invasive and requires specialists (Hamashima, 2016). While high-incidence countries have implemented screening programs, low-incidence countries require cost-effective alternatives (Herrera-Pariente et al., 2021). Serum biomarkers, which include carcinoembryonic antigen (CEA), alpha-fetoprotein (AFP), and carbohydrate antigens (CA19-9 or CA72-4), have been used for early diagnosis, but their lack of specificity results in low positive rates and the inability to detect precancerous lesions (Feng et al., 2017).

Five studies focusing on biomarkers for GaC, all employing an untargeted metabolomics approach, are included in Supplementary Tables S1–S3. Two studies, conducted in China and Latvia, compared GaC patients, peptic ulcer disease (PUD) patients, and controls. Xu et al. (2013) identified three upregulated VOCs in GaC and four VOCs upregulated in PUD compared to HCs, with one VOC (furfural) shared among comparisons. Likewise, Amal et al. (2013) found four VOCs upregulated in GaC of which two were also upregulated in PUD. However, no single discriminating VOC between GaC and PUD was identified in any of the studies. Only one VOC was found to be common to different geographical areas, 6-methyl-5-hepten-2-one (Supplementary Table S3).

Tong et al. (2017) reported 11 candidate GaC biomarkers comparing GaC patients with PUD, gastritis patients, and HCs, using UVA and MVA. One VOC, nonanal, was also found by Amal et al. (2013) to be significantly altered between GaC, PUD, and an

additional group stratified based on the operative link on gastric intestinal metaplasia (OLGIM), which classifies patients according to the presence/absence and stage of precancerous lesions. Among the multiple comparisons, eight VOCs showed alterations among groups, seven of which were upregulated in GaC compared to OLGIM, only one VOC being altered between GaC and OLGIM III-IV, and three VOCs in PUD compared to OLGIM.

Lastly, Bhandari et al. (2023) explored the correlation between the fecal microbiome and exhaled breath VOCs. Two VOCs (1-octanol and dioctyl ether) were significantly altered and exclusively present in GaC. Moreover, 14 VOCs from GaC patients were correlated with 33 fecal bacterial taxa, and 7 VOCs from HCs were correlated with 17 bacterial taxa, with no common VOCs between groups.

3.1.3 Colorectal cancer

Colorectal cancer (CRC) ranks among the most common cancers worldwide (Ferlay et al., 2021). The 5-year relative survival drops from 65% to 15.6% when diagnosed at later stages (2013–2019), which represents 23% of cases, as early symptoms are not pathology-specific (Colorectal Cancer—Cancer Stat Facts, 2023). The most applied screening tools are the fecal immunochemical test (FIT) and colonoscopy (Helsingen and Kalager, 2022). The FIT test is based on the measurement of the amount of hemoglobin in feces, and one-third of stage I cancers are missed (Niedermaier et al., 2020). Colonoscopy, while effective, is invasive, time-consuming, and expensive and is performed with conscious sedation, carrying the risk of colonic perforation and major bleeding (Qaseem et al., 2019; Helsingen and Kalager, 2022). Other CRC screening tests such as the guaiac-based fecal occult blood test (gFOBT), sigmoidoscopy, fecal biomarker panel test, and computed tomography (CT) colonography have several limitations, such as false-positive results, invasiveness, and high cost (Qaseem et al., 2019).

Five studies focusing on potential biomarkers for CRC are summarized in Supplementary Tables S1–S3, all employing an untargeted metabolomics approach. The studies that analyzed mixed breath sampled the same cohort of CRC patients. The first study (Altomare et al., 2013) compared CRC and HCs by selecting a pattern of 15 VOCs by UVA to construct the probabilistic neural network (PNN) model, which yielded 76.0% accuracy in the validation cohort. In a subsequent study (Altomare et al., 2015), the data were reprocessed, and 32 of 52 CRC patients were resampled after cancer removal. The PNN model was constructed with 31 VOCs selected by UVA, yielding 97.5% and 97.7% accuracies discriminating pre- and post-surgery CRC patients, and post-surgery CRC and HCs, respectively. Additionally, 11 VOCs shared with the previous study could discriminate pre- and post-surgery CRC patients with 98.8% accuracy. These results demonstrate the metabolic change in exhaled VOC patterns due to cancer cell metabolism and suggest that metabolism does not return to the pre-cancer state after cancer removal.

Another study by the aforementioned research group investigated potential biomarkers of cancer stages (early/I–II or advanced/III–IV) (Altomare et al., 2020). Fifteen VOCs were selected by UVA comparing CRC and HCs, to build a model that included age class (>65 vs. ≤65 year olds). Fourteen identified VOCs could discriminate CRC from HCs, with a 93.0% overall positive predictive value (PPV) after cross-validation, whereas eight and five VOCs could discriminate early-CRC from HCs with an 86.0% PPV and advanced-CRC from HCs with a 91.0% PPV. Three common VOCs between UVA and MVA, namely, ethylbenzene, methylbenzene, and tetradecane, were quantified to establish the threshold concentration values. However, none of these compounds were reported in other studies. Nevertheless, five out of the 15 VOCs were common with previous studies (Altomare et al., 2013, 2015), and three were reported as significantly altered between CRC and HCs: 4-methyloctane and ethanol (research group from Latvia) (Amal et al., 2016) and dodecane (research group from China) (Wang et al., 2014a) (Supplementary Table S3).

Likewise, Amal et al. (2016) found four significantly altered VOCs between CRC and HCs, which were identified by UVA and subsequently quantified: 4-methyloctane and ethanol were downregulated, whereas acetone and ethyl acetate were upregulated. Likewise, Wang et al. (2014a) found nine potential biomarkers (eight upregulated and one downregulated) for CRC patients with adenocarcinoma by MVA.

3.1.4 Breast cancer

Breast cancer (BC) is the most diagnosed type of cancer and the fifth cause of cancer-related mortality (Cancer; Ferlay et al., 2021). While the 5-year relative survival rate stands at 90.8%, it drops dramatically to 31% when diagnosed at a distant stage (2012–2019) (Female Breast Cancer, 2023). The current gold standard screening methods include annual mammography and clinical breast examination for women over the age of 40. Unfortunately, physical breast examinations, even when performed by a physician, fail to reduce mortality (Barba et al., 2021). Regarding mammography, the sensitivity is compromised by breast density (Boyd et al., 2007), and the procedure requires X-ray examination and may lead to overdiagnosis, resulting in unnecessary procedures and treatments (Løberg et al., 2015). Alternative screening approaches, such as digital breast tomosynthesis (DBT), ultrasonography, magnetic resonance imaging (MRI), and positron emission tomography/computed tomography (PET/CT), are hampered by high costs, discomfort, the requirement for trained technicians, and radiation exposure (Barba et al., 2021). Therefore, there is an urgent requirement for innovative screening tools that can overcome these drawbacks, and breath tests show promise as a potential approach.

Four studies focusing on potential BC biomarkers are summarized in Supplementary Tables S1–S3. These studies were conducted in the same geographical area (China). The targeted study by Li et al. (2014) focused on four aldehydes and their potential to discriminate between BC patients, breast non-malignant disease (BNMD) patients, and HCs. All the targeted aldehydes were significantly upregulated in BC, while hexanal was upregulated in BNMD, both compared to HC. Furthermore, nonanal was increased in BC when compared to BNMD. The combination of these VOCs showed 91.7% sensitivity and 95.8% specificity (0.934 AUC) in discriminating early-stage BC from HCs, and the predictive model achieved 80.4% correct classification after leave-one-out cross-validation (LOOCV). Hexanal was also identified as a potential biomarker in a different study (Supplementary Table S3).

Two untargeted studies compared BC with HCs and BNMD. Barash et al. (2015) identified 23 VOCs by UVA, 21 of which showed significant differences between HCs and patients with breast lesions [BC, BNMD, and an additional group of patients with ductal carcinoma *in situ* (DCIS)], and four VOCs were significant between BC and DCIS. The MVA revealed 14 VOCs that could discriminate BC from HC and BNMD, and from DCIS, yielding 72.0% and 81.0% accuracies after LOOCV, respectively. Additionally, two of these 14 VOCs were consistent with findings from other studies (Supplementary Table S3). Wang et al. (2014b) annotated 28 potential biomarkers, of which 21, 6, and 8 VOCs were significantly altered in BC when compared separately to HCs, BNMD (cyclomastopathy and mammary gland fibroma), and DCIS, respectively. Among these, three VOCs, namely, cyclohexanone, 1,4-dimethoxy-2,3-butanediol, and 2,5,6-trimethyloctane, were upregulated in BC compared to both HCs and BNMD. Only cyclohexanone was again reported by Zhang et al. (2020a) (Supplementary Table S3).

Furthermore, Zhang et al. (2020a) subdivided the BC group into DCIS, lymph node metastasis-negative (LNMN), and lymph node metastasis-positive (LNMP), annotating 13, 12, and 17 significant VOCs when compared to HC, respectively. An additional group of GaC patients was included for comparison with BC, yielding 17 significant VOCs. The set of seven overlapping VOCs among all comparisons could discriminate BC and the different subgroups from HCs value (e.i. 0.864–0.943) AUC, sensitivity 80.8%–96.2%, and specificity 71.6%–100%.

3.2 VOCs in other pulmonary pathologies

3.2.1 Asthma

Asthma is a chronic and heterogeneous lung pathology characterized by inflammation and airway obstruction, manifesting with variable symptoms that include cough, wheezing, shortness of breath, and chest tightness (Asthma, 2023; Asthma-Diagnosis, 2023). This pathology places a significant economic burden on healthcare systems, affecting approximately 292 million people worldwide, typically being developed during childhood. Asthma's impact on patients' quality of life and the risk of premature death are major concerns (The Global Asthma Report, 2022).

Diagnosis relies on spirometry, bronchoprovocation tests, peak expiratory flow tests, allergy skin or blood tests, and FeNO tests (Asthma-Diagnosis, 2023). Patients may experience a loss of pathology control and acute exacerbations of symptoms, leading to significant morbidity and a progressive loss of lung function (Castillo et al., 2017). Moreover, the heterogeneity of asthma, concerning severity and response to treatment, is a consequence of the underlying pathophysiological mechanisms. Patients can be classified into different phenotypes based on observable characteristics (steroid response, obesity, allergies, etc.), or endotypes based on the underlying cellular and molecular mechanisms (Kuruvilla et al., 2019). In this context, breath tests offer a non-invasive and easy-to-perform approach for early diagnosis and exacerbation prediction, especially suitable for children, and could be also used to define phenotypes and endotypes by analyzing the profile of endogenous VOCs, which reflects the inflammatory state of the bronchia and underlying molecular mechanisms involved, allowing a significant improvement in treatment effectiveness.

Eleven untargeted studies focused on asthma are included in Tables 1, 2; Supplementary Table S4. Several studies were conducted by the same research group focusing on asthmatic children. Van Vliet et al. (2016, 2017) studied the loss of asthma control and exacerbation episodes over a period of 1 year. In the first study (Van Vliet et al., 2016), a combination of 15 VOCs (10 annotated) showed 86.0% accuracy in classifying persistently controlled and uncontrolled asthma, although no association was found between different exhaled inflammatory markers [FeNO, exhaled breath condensate (EBC), and VOCs] and asthma control. Subsequently (Van Vliet et al., 2017), in a larger cohort of asthmatic children, the combination of seven VOCs used to construct the RF model could predict 88.0% of asthma exacerbation episodes within 14 days. These two studies shared only two VOCs: 1,2-dimethylcyclohexane and 2-methylfuran (Table 2). Additionally, Robroeks et al. (2013) annotated 30 VOCs related to asthma exacerbation, and the models combining six and seven VOCs could correctly classify 96.0% of baseline and exacerbation samples taken from the same patient (100% sensitivity and 93.0% specificity) and 91.0% of patients who would have future exacerbations or not, respectively. These results suggest that the profile of VOCs can identify exacerbations and could be used to predict which patients will suffer these episodes. Additionally, Smolinska et al. (2014) studied a cohort of wheezing children with HCs between the ages of 2 and 4 years until the age of 6 years, to find potential biomarkers for preclinical asthma. A total of 17 VOCs (13 annotated) were selected by comparing asthmatic children with HCs and with transient wheezers, which could correctly classify 80.0% of the wheezing children at inclusion, differentiating those who would develop asthma from those who were transient wheezers. Notably, three VOCs reported in these studies (2-methylfuran, 3-methylfuran, and m-cymene) were also identified by Monedeiro et al. (2021) when comparing LC, COPD asthmatic patients, and HCs. In this study, the model built with 12 VOCs from the untargeted data presented 85.7% overall accuracy, and another with 9 of the 29 targeted VOCs provided 91.0% overall accuracy.

Likewise, two other studies included a cohort of asthmatic children, in this case, compared to HC. Gahleitner et al. (2013) identified a panel of eight candidate VOCs, all of which were upregulated in asthmatic children. Moreover, Caldeira et al. (2012) built a model with the full data set of metabolites (134), yielding a classification rate of 98.0% (96.0% sensitivity and 95.0% specificity). Among these metabolites, six alkanes were related to allergic asthma and four aldehydes and one alkene to HC. The new model that included nine alkanes and aldehydes showed a classification rate of 96.0% (98.0% sensitivity and 93.0% specificity). One VOC from the latter study, decane, was also reported by Sola-Martínez et al. (2021). In this study, a population of women 3 months postpartum was recruited and divided into asthmatics with other coexisting atopic diseases (A-AD) and non-asthmatics, and the latter were further divided into those with and without other atopic diseases (NA-AD and NA-NAD, respectively). Several models were built to compare the different groups, selecting a total of nine VOCs, which could discriminate between asthmatic and non-asthmatic patients, even

TABLE 1 Summary of studies focused on asthma, chronic obstructive pulmonary disease, obstructive sleep apnea, and cystic fibrosis. AB, alveolar breath; ABS, adaptive breath sampler; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; GC, gas chromatography; GC×GC, two-dimensional gas chromatography; LC, lung cancer; MB, mixed breath; MS, mass spectrometry; na, not applicable; nd, not detailed; NIST, National Institute of Standards and Technology; NTD, needle-trap device; OSA, obstructive sleep apnea; SPME, solid-phase microextraction; TD, thermal desorption tube; TOF, time-of-flight; UI, ultra-inert, VOCs, volatile organic compounds.

Reference	Pathology	Methodology	Sample	Sampling	Analysis technique	Sorbent material	Column	IS	Identification	
									Library	Authentic STD
Gahleitner et al. (2013)	Asthma	Untargeted	AB	ABS	TD-GC-MS	Tenax/Carbotrap	nd	No	NIST	Yes
Sola-Martínez et al. (2021)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-MS	Tenax TA	HP-5MS UI (30 m × 0.25 mm × 0.25 μm) (Agilent)	No	NIST	No
Schleich et al. (2019)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS/ TD-GCxGC-TOF-MS	Carbograph 1TD/ Carbopack X and Tenax TA/Carbopack B	RTX-5MS (30 m × 0.25 mm × 1 μm) (Restek) and Rxi-624Sil MS (30 m × 0.25 μm × 1.4 μm) (Restek) 1D and Stabilwax (2 m × 0.25 μm × 0.5 μm) (Restek) 2D	No	NIST	Yes
Brinkman et al. (2017)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-MS	Tenax GR	VF1-MS column (30 m × 0.25 mm × 1 μm) (Varian)	No	NIST	No
Van Vliet et al. (2017)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	nd	No	NIST	No
Van Vliet et al. (2016)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	nd	No	NIST	No
Meyer et al. (2014)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	RTX-5MS (30 m × 0.25 mm × 1 μm) (Restek)	No	nd	No
Smolinska et al. (2014)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	RTX-5MS (30 m × 0.25 mm × 1 μm) (Restek)	No	NIST	No
Robroeks et al. (2013)	Asthma	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Active carbon	RTX-5MS (30 m × 0.25 mm × 1 μm) (Restek)	No	NIST	No
Caldeira et al. (2012)	Asthma	Untargeted	MB	Tedlar® bag	SPME-GC×GC-TOF-MS	DVB/Car/PDMS	HP-5 (30 m × 0.32 mm × 0.25 μm) (Agilent) 1D and DB-FFAP (0.79 m × 0.25 mm × 0.25 μm) (Agilent) 2D	No	In-house library and Wiley and NIST	No

(Continued on following page)

10.3389/fmolb.2023.1295955

TABLE 1 (*Continued*) Summary of studies focused on asthma, chronic obstructive pulmonary disease, obstructive sleep apnea, and cystic fibrosis. AB, alveolar breath; ABS, adaptive breath sampler; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; GC, gas chromatography; GC×GC, two-dimensional gas chromatography; LC, lung cancer; MB, mixed breath; MS, mass spectrometry; na, not applicable; nd, not detailed; NIST, National Institute of Standards and Technology; NTD, needle-trap device; OSA, obstructive sleep apnea; SPME, solid-phase microextraction; TD, thermal desorption tube; TOF, time-of-flight; UI, ultra-inert, VOCs, volatile organic compounds.

Reference	Pathology	Methodology	Sample	Sampling	Analysis technique	Sorbent material	Column	IS	Identification	
									Library	Authentic STD
Monedeiro et al. (2021)	LC/COPD/ Asthma	Untargeted/ targeted	MB	Tedlar® bag	NTD-GC-MS	PDMS/Carbopack/ Carboxen	DB-624 capillary column (60 m × 0.32 mm × 1.8 µm) (Agilent)	No	NIST	Yes
Pizzini et al. (2018)	COPD	Untargeted	AB	Glass syringe	TD-GC-TOF-MS	Carbotrap B 80 mg/ Carbopack X 260 mg	Restek-Q-Bond (30 m × 0.25 mm × 8 µm) (Restek)	No	NIST	Yes
Basanta et al. (2012)	COPD	Untargeted	AB	—	TD-GC-TOF-MS	Tenax TA/Carbotrap	DB5-MS column (30 m × 0.25 mm x 0.25 µm) (Agilent)	D5-Bromobenzene	NIST	No
Phillips et al. (2012)	COPD	Untargeted	AB	Bio-VOC®	TD-GC-MS	Carbograph 1TD/ Carbopack X	HP-5MS (30 m × 0.25 mm × 0.25 µm) (Agilent)	No	NIST	No
van Velzen et al. (2019)	COPD	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Tenax GR	VF1-MS column (30 m × 0.25 mm × 1 µm) (Varian)	No	NIST	No
Gaida et al. (2016)	COPD	Untargeted	MB	Stainless steel tube	TD-GC-MS	Tenax TA	nd	No	NIST	Yes
Cazzola et al. (2015)	COPD	Untargeted	MB	Tedlar® bag	SPME-GC-MS	DVB/Car/PDMS 50/ 30 µg	Equity-5 capillary column (30 m × 0.25 mm × 0.25 µm) (Supelco)	No	NIST	No
Jareño-Esteban et al. (2017)	COPD	Targeted	AB	Bio-VOC®	TD-GC-MS	Tenax TA/graphitized carbon black/carbonized molecular sieve	DB-1 (30 m × 0.25 mm × 1 µm) (Agilent)	Hexamethylcyclotrisiloxane	na	Yes
Bayrakli et al. (2016)	OSA	Targeted	AB	Bio-VOC®	TD-GC-MS	Tenax TA 200 mg	DB-5 (30 m × 0.25 mm) (Agilent)	No	na	Yes
Aoki et al. (2017)	OSA	Targeted	MB	DuPont™ Tedlar® bag	TD/NTD-GC-MS	nd	nd	No	na	Yes
Woollam et al. (2022b)	CF	Untargeted	MB	Tedlar® bag	SPME-GC-MS	DVB/Car/PDMS	HP-5MS (30 m × 0.25 mm × 0.25 µm) (Agilent)	No	nd	No
van Horck et al. (2021)	CF	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	nd	No	NIST	No

10.3389/fmolb.2023.1295955

TABLE 2 VOCs reported in asthma and chronic obstructive pulmonary disease (≥2 studies). ***EO*****, eosinophilic asthma;** ***na*****, not applicable;** ***NEO*****, neutrophilic asthma;** ***ppbv*****, parts per billion by volume; *LOD, Limit of detection.**

Asthma									
No	Compound name	CAS-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
1	1-Propanol	71-23-8	C_3H_8O	Alcohol	Upregulated	9.94	14.59	ppbv	Monedeiro et al. (2021)
					Downregulated (EO)/ Upregulated (NEU)	na	na	na	Schleich et al. (2019)
2	Phenol	108-95-2	C_6H_6O	Alcohol	Downregulated	na	na	na	Meyer et al. (2014)
					Upregulated	<1.43*	<1.43*	ppbv	Monedeiro et al. (2021)
3	Nonanal	124-19-6	$C_9H_{18}O$	Aldehyde	Altered	na	na	na	Van Vliet et al. (2017)
					Altered	na	na	na	Caldeira et al. (2012)
					Upregulated	na	na	na	Schleich et al. (2019)
4	Octanal	124-13-0	$C_8H_{16}O$	Aldehyde	Downregulated	na	na	na	Meyer et al. (2014)
					Altered	na	na	na	Van Vliet et al. (2017)
5	Benzene	71-43-2	C_6H_6	Aromatic hydrocarbon	Upregulated	na	na	na	Meyer et al. (2014)
					Altered	na	na	na	Robroeks et al. (2013)
6	m-Cymene	535-77-3	$C_{10}H_{14}$	Aromatic hydrocarbon	Upregulated	na	na	na	Gahleitner et al. (2013)
					Altered	na	na	na	Van Vliet et al. (2016)
					Upregulated	0.32	0.61	ppbv	Monedeiro et al. (2021)
7	2,4-Dimethylheptane	2213-23-2	C_9H_{20}	Branched hydrocarbon	Downregulated	na	na	na	Meyer et al. (2014)
					Upregulated	na	na	na	Smolinska et al. (2014)
8	2-Methylpentane	107-83-5	C_6H_{14}	Branched hydrocarbon	Upregulated	4.59	1.24	ppbv	Monedeiro et al. (2021)
					Upregulated	na	na	na	Smolinska et al. (2014)
9	3-Methylpentane	96-14-0	C_6H_{14}	Branched hydrocarbon	Upregulated	1.07	0.24	ppbv	Monedeiro et al. (2021)
					Altered	na	na	na	Robroeks et al. (2013)
10	1,2-Dimethylcyclohexane	583-57-3	C_8H_{16}	Cyclic hydrocarbon	Altered	na	na	na	Van Vliet et al. (2016)
					Altered	na	na	na	Van Vliet et al. (2017)

(Continued on following page)

TABLE 2 (*Continued*) VOCs reported in asthma and chronic obstructive pulmonary disease (≥2 studies). *EO*, eosinophilic asthma; *na*, not applicable; *NEO*, neutrophilic asthma; *ppbv*, parts per billion by volume; *LOD, Limit of detection.

Asthma									
No	Compound name	CAS-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
11	2-Methylfuran	534-22-5	C_5H_6O	Ether	Altered	na	na	na	Van Vliet et al. (2016)
					Altered	na	na	na	Van Vliet et al. (2017)
12	Decane	124-18-5	$C_{10}H_{22}$	Hydrocarbon (saturated)	Altered	na	na	na	Caldeira et al. (2012)
					Altered	na	na	na	Sola-Martínez et al. (2021)
13	Dodecane	112-40-3	$C_{12}H_{26}$	Hydrocarbon (saturated)	Downregulated	na	na	na	Meyer et al. (2014)
					Upregulated	6.27	5.18	ppbv	Monedeiro et al. (2021)
					Altered	na	na	na	Caldeira et al. (2012)
14	Tetradecane	629-59-4	$C_{14}H_{30}$	Hydrocarbon (saturated)	Altered	na	na	na	Monedeiro et al. (2021)
					Altered	na	na	na	Caldeira et al. (2012)
15	Undecane	1120-21-4	$C_{11}H_{24}$	Hydrocarbon (saturated)	Upregulated	1.78	0.80	ppbv	Monedeiro et al. (2021)
					Downregulated	na	na	na	Schleich et al. (2019)
16	Acetone	67-64-1	C_3H_6O	Ketone	Altered	na	na	na	Sola-Martínez et al. (2021)
					Downregulated	na	na	na	Smolinska et al. (2014)
17	Acetonitrile	75-05-8	C_2H_3N	Nitrogen-containing	Altered	na	na	na	Brinkman et al. (2017)
					Altered	na	na	na	Monedeiro et al. (2021)
Chronic obstructive pulmonary disease									
No	Compound name	CAS-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
1	Isopropanol	67-63-0	C_3H_8O	Alcohol	Downregulated	na	na	na	Cazzola et al. (2015)
					Upregulated	258.37	10.55	ppbv	Monedeiro et al. (2021)
2	Phenol	108-95-2	C_6H_6O	Alcohol	Altered	na	na	na	Gaida et al. (2016)
					Altered	na	na	na	Phillips et al. (2012)
3	Decanal	112-31-2	$C_{10}H_{20}O$	Aldehyde	Altered	na	na	na	Basanta et al. (2012)
					Altered	na	na	na	Phillips et al. (2012)

(Continued on following page)

TABLE 2 (*Continued*) VOCs reported in asthma and chronic obstructive pulmonary disease (≥2 studies). *EO*, eosinophilic asthma; *na*, not applicable; *NEO*, neutrophilic asthma; *ppbv*, parts per billion by volume; *LOD, Limit of detection.

Chronic obstructive pulmonary disease									
No	Compound name	CAS-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
4	Hexanal	66-25-1	$C_6H_{12}O$	Aldehyde	Upregulated	na	na	na	Jareño-Esteban et al. (2017)
					Altered	na	na	na	Basanta et al. (2012)
					Altered	na	na	na	Phillips et al. (2012)
5	Nonanal	124-19-6	$C_9H_{18}O$	Aldehyde	Upregulated	na	na	na	Jareño-Esteban et al. (2017)
					Altered	na	na	na	Basanta et al. (2012)
6	Benzene	71-43-2	C_6H_6	Aromatic hydrocarbon	Upregulated	na	na	na	Gaida et al. (2016)
					Altered	na	na	na	Phillips et al. (2012)
7	Toluene	108-88-3	C_7H_8	Aromatic hydrocarbon	Upregulated	na	na	na	Gaida et al. (2016)
					Altered	na	na	na	Phillips et al. (2012)
					Altered	na	na	na	van Velzen et al. (2019)
8	Limonene	138-86-3	$C_{10}H_{16}$	Cyclic hydrocarbon	Downregulated	na	na	na	Cazzola et al. (2015)
					Altered	na	na	na	van Velzen et al. (2019)
					Altered	na	na	na	Phillips et al. (2012)
					Upregulated	1.71	1.57	ppbv	Monedeiro et al. (2021)
9	Butane	106-97-8	C_4H_{10}	Hydrocarbon (saturated)	Altered	na	na	na	Phillips et al. (2012)
					Downregulated	na	na	na	Pizzini et al. (2018)
10	Tridecane	629-50-5	$C_{13}H_{28}$	Hydrocarbon (saturated)	Altered	ns	ns	ns	Gaida et al. (2016)
					Upregulated	28.36	3.43	ppbv	Monedeiro et al. (2021)
11	Acetic acid	64-19-7	$C_2H_4O_2$	Organic acid	Altered	na	na	na	Gaida et al. (2016)
					Altered	na	na	na	Phillips et al. (2012)

in the validation cohort (AUCs 0.670–0.900, 71.0%–100% sensitivity, and 60.0%–70.0% specificity), although the accuracy decreased when asthmatic patients were compared to the non-asthmatic groups separately (AUCs 0.680–0.810 for NA-AD and 0.603–0.750 for NA-NAD).

Furthermore, two articles studied asthma phenotypes and endotypes. Schleich et al. (2019) conducted a study on a group of asthmatic patients classified by inflammatory subtypes. From all binary comparisons, 12 VOCs were selected, of which eight were identified as candidate biomarkers. Among them, two VOCs

(hexane and 2-hexanone), along with 1-propanol, were selected from the comparison between eosinophilic and paucigranulocytic value (e.i. 0.680) AUC. Meanwhile, the comparison between neutrophilic and paucigranulocytic yielded two VOCs (3-tetradecene and pentadecene) in the discovery phase and another two (undecane and nonanal) in the replication value (e.i. 0.850 and 0.700, respectively) AUC. Furthermore, when comparing neutrophilic to eosinophilic, three VOCs (3,7-dimethylnonane, 1-propanol, and nonanal) were identified in the discovery phase value (e.i. 0.920) AUC, although only nonanal, along with hexane, showed the best classification performance in the replication phase value (e.i. 0.710) AUC. As a result, two (hexane and 2-hexanone) and three (nonanal, 1-propanol, and hexane) VOCs could discriminate eosinophilic and neutrophilic asthma from other phenotypes value (e.i. 0.720 and 0.730, respectively) AUC. Moreover, Meyer et al. (2014), besides building a model based on 16 VOCs that could discriminate asthmatic patients from HC (100% sensitivity and 91.1% specificity), performed a cluster analysis that included clinical, medication features, and four VOCs that were only present in asthmatic patients, to identify different asthma endotypes. As a result, seven clusters were formed, two with non-allergic asthma and five with allergic asthma. Some clusters presented high clinical similarity but different profiles of VOCs, as well as similar profiles and different clinical symptoms. Although no common VOCs were found between these two studies, eight VOCs were shared with others (Table 2).

3.2.2 COPD

COPD is characterized by chronic respiratory symptoms, such as dyspnea, cough, production of sputum, and/or exacerbations, caused by abnormalities in the airways (bronchitis and bronchiolitis) and/or the alveoli (emphysema), resulting in persistent and progressive airflow obstruction. The causes of the pathology are environmental exposures (tobacco smoking, toxic particles, and gases) and/or genetic risk factors. According to the WHO, 3.23 million people died from COPD in 2019, with 90% of deaths (under the age of 70) occurring in low- and middle-income countries. COPD often coexists with chronic pathologies, such as lung infections and cancer, heart problems, depression, and anxiety (GOLDCOPD, 2023).

COPD diagnosis relies on spirometry, with weak specificity. Additional tests, lung imaging and arterial blood gas tests, can help assess pathology severity. The symptoms develop slowly, and even though COPD is not curable, different treatments can be applied. However, under- or misdiagnosis can lead to lack/incorrect treatment (GOLDCOPD, 2023), and most patients are diagnosed when the lung damage is irreversible (Fazleen and Wilkinson, 2020). Detecting early or pre-COPD cases, where clinical signs are absent or airflow obstruction is not evident in spirometry, can be challenging. Breath tests offer a valuable tool for identifying these cases that diagnostic tests may miss.

Eight studies are indicated in Tables 1, 2; Supplementary Table S4. Jareño-Esteban et al. (2017) targeted five VOCs (hexanal, heptanal, nonanal, propanoic acid, and nonanoic acid) as potential biomarkers. Although hexanal and nonanal were upregulated in COPD patients compared to non-smokers (HC), no significant VOCs were found between COPD patients and active smokers (HC). Both these VOCs were reported in previous studies (Table 2).

Four studies compared COPD patients with HCs, two being performed in the same geographical area (UK), in different research groups. Phillips et al. (2012) applied different machine learning methods, which included a step of FS, to compare the whole group of COPD with HCs, active with former smokers within the COPD group, and COPD with HCs (non-smokers). Of the automatically generated VOCs in the three comparisons (12, 13, and 10, respectively), six overlapped. Likewise, two of these six shared VOCs were reported by Gaida et al. (2016), and another six VOCs were reported in different studies (Table 2). Moreover, Basanta et al. (2012) built a classification model containing 11 VOCs after data reduction (UVA and PCA), with an accuracy of 70.0%. The groups were further divided and compared by smoking status, improving the performance of the model, especially when active smokers were compared (91.0% accuracy). Furthermore, four VOCs were correlated with sputum eosinophils ≥1%, one VOC with sputum eosinophils ≥2%, and four VOCs with exacerbation episodes (≥2/year). The prediction models showed an accuracy of 75.0% and 88.0% for sputum eosinophils ≥1% and sputum eosinophils ≥2%, respectively, and 83.0% for exacerbations, after LOOCV. Of these 11 VOCs, 3 aldehydes (decanal, hexanal, and nonanal) were shared with Jareño-Esteban et al. (2017) and Phillips et al. (2012) (Table 2). Additionally, Gaida et al. (2016) studied two cohorts of COPD patients and HCs from different locations, which were split by smoking habits. Overall, 14 VOCs showed potential as COPD biomarkers, with 4 being reported also by Phillips et al. (2012) (Table 2).

The study by Monedeiro et al. (2021) was previously mentioned in the LC/asthma section, with untargeted and targeted analyses to build classification models that distinguish COPD, LC, asthma patients, and HCs, yielding 85.7% and 91% overall accuracy (untargeted and targeted, respectively). Two of these VOCs (isopropyl alcohol and limonene) were shared with Cazzola et al. (2015), and two additional VOCs were common with other studies (Table 2).

The remaining studies focused on COPD exacerbations. Pizzini et al. (2018) applied UVA and *post hoc* analysis between pairwise combinations, resulting in 12 significant VOCs. Additionally, four VOCs were classified as discriminative for acute exacerbation (A) COPD, two VOCs were classified as discriminative for stable (S) COPD, and two VOCs as associated with COPD. The RF model containing these 12 VOCs could classify COPD patients value (e.i. 0.970, 78.0% sensitivity, and 91.0% specificity) AUC. Meanwhile, van Velzen et al. (2019) sampled the same cohort of COPD patients before (baseline), during, and after (recovery) an exacerbation episode. The UVA between the Clinical COPD Questionnaire (CCQ) symptom scores and VOCs resulted in 10 discriminative compounds. The subsequent MVA discriminated between baseline and exacerbation and between exacerbation and recovery with accuracies of 71.0% and 75.0% , respectively.

3.2.3 OSA

OSA is a respiratory disorder with an incidence of 24% in men and 9% in women (30–60 years of age), affecting nearly 1 billion people worldwide (Lv et al., 2023). OSA is characterized by the repeated collapse of the pharynx, leading to episodes of apnea or hypopnea accompanied by decreased oxygen levels and interruptions in sleep. It is associated with poor sleep quality and daytime sleepiness, as well as an increased risk for several metabolic and cardiovascular pathologies (arterial hypertension, diabetes, etc.), and depression (Lévy et al., 2015; Schwarz et al., 2017; Nowak et al., 2021).

Current OSA diagnosis relies on sleep examination (monitoring sleep stages and cycles), mainly through polysomnography, a costly, time-consuming, and inconvenient test. Although home tests are available, these devices are subject to more measurement errors compared to polysomnography (Kapur et al., 2017). Moreover, several nights should be monitored to obtain a more reliable diagnosis (Stöberl et al., 2017). Therefore, breath tests are presented as a potential tool for both the screening and diagnosis of OSA.

Two independent targeted studies focusing on OSA are included in Table 1 and Supplementary Table S4. Bayrakli et al. (2016) studied the levels of acetone and butanol in patients before and after sleep. Although butanol was upregulated in patients compared to HC (after sleep), this VOC was not significantly increased between patients (before vs. after sleep). Conversely, Aoki et al. (2017) focused on 14 VOCs, which included aromatic, alicyclic, chain hydrocarbons, isoprene, acetone, and ethanol, and classified OSA patients into moderate, severe, and most severe in terms of the apnea–hypopnea index (AHI). The UVA yielded four VOCs upregulated in all OSA patients, four VOCs in severe and most severe OSA patients, and three VOCs exclusively in the most severe OSA patients compared to HCs. Furthermore, four of these VOCs (ethylbenzene, p-xylene, phenylacetic acid, and nonane) showed increased levels according to OSA severity, being correlated with the AHI, arousal index, and duration of percutaneous oxygen saturation (SpO_2) ≤ 90%. Additionally, the levels of acetone and isoprene decreased after continuous positive airway pressure treatment. Nevertheless, no common VOCs were found between these two studies.

3.2.4 Cystic fibrosis

Cystic fibrosis (CF) is an autosomal recessive genetic pathology caused by a mutation in the *cystic fibrosis transmembrane conductance regulator* (CFTR) gene. This mutation disrupts the cells' electrolyte transport system, affecting mainly organs with secretory functions, such as as the lungs, pancreas, and reproductive system (Cystic Fibrosis-Causes, 2023). In the lungs, altered sodium absorption results in thick, hardened secretions, increasing the risk of respiratory infections, inflammation, and oxidative stress (Roesch et al., 2018). Pulmonary exacerbations (PEx) are frequent events in the progression of the pathology, potentially leading to permanent lung function loss, reduced quality of life, and decreased survival. PEx treatment includes antibiotics, but delayed symptom onset worsens outcomes (Goss, 2019). The identification of PEx relies on symptomatology, clinical evaluation, and the measurement of changes in forced expiratory volume in one second (FEV1pp) using spirometry devices (Goss, 2019). The use of breath tests to predict PEx in CF is a promising approach. In two independent untargeted studies, CF PEx in children was studied, as shown in Table 1; Supplementary Table S4. van Horck et al. (2021) performed a 1-year observational pilot study, recruiting patients from three different centers. The RF model with the nine most discriminating VOCs could predict 79.0% of patients with stable or upcoming PEx (within 7 days) (79.0% sensitivity and 78.0% specificity). However, no single VOC was found significantly altered when applying UVA between stable and CF PEx patients. Meanwhile, Woollam et al. (2022b) divided the CF patients into CF baseline (not suffering from PEx) and CF PEx. Four VOCs were found to be correlated with FEV1pp at the time of breath collection, of which two VOCs (4-methyl-octane and 3,7-dimethyldecane) were further correlated with changes in FEV1pp. Moreover, four VOCs were found to be significantly different between CF baseline and CF PEx patients: 3,7-dimethyldecane, durene, and 5-methyltridecane were downregulated, and 2,4,4-trimethyl-1,3-pentanediol 1-isobutyrate was upregulated in PEx patients. Although both studies aimed to identify differential VOCs between CF stable and CF PEx patients, none of the reported were shared.

3.3 VOCs in infectious pathologies

3.3.1 Pneumonia (CAP/HAP/VAP)

CAP, HAP, and VAP are lower respiratory tract infections associated with high morbidity, mortality, and healthcare costs (Ferreira-Coimbra et al., 2020; Munro et al., 2021; Alnimr, 2023). HAP is developed after 48 h of hospitalization, while VAP is the most frequent infection in the intensive care unit (ICU), developed after endotracheal intubation (Modi and Kovacs, 2020). The pathogens involved encompass Gram-positive bacteria (*Staphylococcus aureus* and *Streptococcus pneumoniae*), Gram-negative bacteria (*Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Klebsiella pneumoniae*, and *Acinetobacter baumannii*), and fungi (*Aspergillus* spp. and *Candida* spp.) (Filipiak et al., 2013, 2015).

Current diagnostics rely on clinical, radiological, and microbiological cultures of respiratory samples [endotracheal aspirates, bronchoalveolar lavage (BAL), and protected specimen brush], which present high inter-variability and moderate sensitivity and specificity. The microbiological confirmation can take several days, leading to overtreatment with antibiotics until the specific pathogen is identified (Fernando et al., 2020; Modi and Kovacs, 2020). Therefore, there is an urgent requirement for less invasive and faster diagnostic techniques.

In the case of VAP, van Oort et al. (2017a) presented a protocol for a prospective multicenter study named BreathDx (Molecular Analysis of Exhaled Breath as Diagnostic Test for Ventilator-Associated Pneumonia), aiming to develop a breath test capable of distinguishing suspected VAP patients, with a 99% sensitivity for culture-positive cases. It also aimed to identify unique VOC patterns that could predict specific pathogen infections, holding promise for more efficient VAP diagnosis and treatment.

Seven studies focusing on VAP, and one on CAP/HAP, are summarized in Tables 3–5. To date, two studies have been conducted in relation to BreathDx. Van Oort et al. (2022) performed an untargeted study within a group of intubated and ventilated ICU patients with suspected VAP, further divided into culture-positive (CP) and culture-negative (CN) BAL samples. Moreover, two platforms were used to cover a wider range of compounds: GC-MS-1 for more volatile compounds and GC-MS-2 for heavier and cyclic volatile compounds. The discriminative model that included 20 VOCs previously selected by UVA and MVA showed 0.830–0.870 AUCs, even when applied to a different set of samples. Furthermore, Ahmed et al. (2023) performed a targeted study focusing on microbial VOCs (mVOCs) previously selected from bacterial species associated with VAP (*S. aureus*, *P. aeruginosa*, *K. pneumoniae*, and

Escherichia coli). In the case of CP for *S. aureus*, two VOCs were upregulated compared to the other patients value (e.i. 0.790–0.870) AUC. In the case of CP for *P. aeruginosa*, two VOCs were downregulated compared to CP for other pathogens, and one of these VOCs (identified as 3-methylbutanal) was common with CP for *S. aureus*. Moreover, those VAP patients with CP for bacteria known to metabolize tryptophan (*E. coli*, *Klebsiella oxytoca*, and *H. influenzae*) presented increased levels of indole. Despite the fact that both studies followed the same BreathDx protocol, no shared VOCs were identified. However, two VOCs (dimethyl sulfide and tetrahydrofuran) reported by Van Oort et al. (2022) and another two (3-methylbutanal and acetone) by Ahmed et al. (2023) were also found in other studies (Table 5).

Additionally, several research groups participating in BreathDx had previously conducted studies focusing on CAP/HAP/VAP, one aiming at possible biomarkers for CAP/HAP (Van Oort et al., 2017b). In this study, patients were categorized based on their clinical suspicion, namely, probable CAP/HAP patients (high clinical suspicion), possible CAP/HAP patients (low clinical suspicion), colonized patients (without symptoms of pneumonia), and controls. Additionally, the entire patient cohort was divided into CP and CN. In the UVA, probable CAP/HAP patients and those who were CP presented 11 and 52 downregulated VOCs, respectively, and the classification models could discriminate between groups based on their clinical suspicion, and among CP and CN, even after LOOCV value (e.i. 0.730 and 0.690, respectively) AUC. While this study differed from the others, since they focused on CAP/HAP, several VOCs were shared, such as acetone, which was described by Ahmed et al. (2023), and 2-methylcyclopentanone, as reported by Fowler et al. (2015) (Table 5). The aforementioned study (Fowler et al., 2015) was performed by another research group involved in BreathDx, where ventilated ICU patients were sampled over their stay at five different time points to identify the VOCs that could be used to predict the risk of developing VAP. The model could separate CP and CN patients (sensitivity 98.0% and specificity 97.0%), and eight VOCs were selected as potential predictors (four downregulated and four upregulated). Several of these VOCs were common in different studies, such as ethanol, which was reported in a total of four independent studies (Table 5). In this regard, Schnabel et al. (2015) constructed an RF model based on 12 VOCs, such as ethanol, which correctly classified 74.2% of VAP and non-VAP patients (75.8% sensitivity, 73.0% specificity, and 0.870 AUC). Furthermore, when searching these VOCs in human and VAP-causing bacteria pathways, ethanol was found to be involved in six distinct pathways. Although ethanol seems to be a promising biomarker, its involvement in VAP development should be further studied, as this VOC participates in many physiological and pathological processes, such as OS, and its origin can be attributed to alcohol consumption.

Additionally, two studies focused on mVOCs previously detected *in vitro* from different cultures of pathogens associated with VAP. Filipiak et al. (2015) annotated 13 mVOCs in CP for *S. aureus* and 11 mVOCs in CP for *Candida albicans*. Considering the possible coexistence of VAP-causing pathogens, the study further aimed to explore and assess differential mVOCs that could potentially be associated with the progression of VAP caused by each pathogen. In this regard, 4-heptanone was found to be possibly related to *C. albicans*; propane and butane to *S. aureus*; acetanilide, 2-pentanone, and dimethyl sulfide to *E. coli*; 3-methyl-1-butene to *H. influenzae*; 1-undecene to *P. aeruginosa*; and n-hexane, isobutane, and 2-methyl-1-butene to *S. pneumoniae*. Likewise, Gao et al. (2016) studied the presence of mVOCs in VAP patients, focusing on *A. baumannii*. For this purpose, *A. baumannii* VAP patients, *A. baumannii* colonized patients, and controls were compared, yielding 19 significant VOCs by UVA, 4 being also detected in *in vitro A. baumannii* cultures. Moreover, 8 of these VOCs were considered derived from *A. baumannii*, being able to differentiate *A. baumannii* VAP patients from colonized patients, as well as from controls (0.880 and 0.890 AUCs, respectively). Both studies reported three VOCs in common. Additionally, one VOC was reported by Filipiak et al. (2015). 3-Methylbutanal was also found in the BreathDx study (Ahmed et al., 2023) (Tables 5).

3.3.2 COVID-19

In the past 3 years, COVID-19 has led to approximately 750 million confirmed cases and nearly 7 million deaths worldwide according to the WHO (WHO Coronavirus Dashboard, 2023). Several diagnostic tests were developed to contain the outbreak, such as reverse-transcription polymerase chain reaction (RT-PCR) for SARS-CoV-2 RNA detection in nasopharyngeal or oral swab samples, and antigen tests for spike (S) protein and nucleocapsid (N) protein detection. However, these tests have variable false-negative rates (Kucirka et al., 2020), with antigen tests being less sensitive and specific than RT-PCR (Scohy et al., 2020). Furthermore, these tests require multiple reagents, and in the case of RT-PCR tests, specialized equipment and trained technicians are required.

Despite the vaccination of over 13 million people worldwide (WHO Coronavirus Dashboard, 2023), COVID-19 remains an ongoing public health challenge. The potential emergence of more transmissible variants, changes in clinical symptoms, immune evasion (even in vaccinated individuals), and the possibility of reinfection are significant concerns. Additionally, distinguishing COVID-19 from other upper respiratory infections is crucial for isolation and transmission prevention. Consequently, breath tests, particularly in resource-limited settings, could offer a rapid means of diagnosing COVID-19.

Five studies that focused on COVID-19 are included in Tables 3 and 4. Two studies were conducted within a cohort of hospitalized patients. The targeted study by Berna et al. (2021) was performed in a cohort of pediatric patients. Six of the 84 targeted VOCs were upregulated in COVID-19 patients, which were further validated in an independent cohort. Moreover, the cumulative abundance of these six VOCs was evaluated as a diagnostic strategy (0.920 AUC, 91.0% sensitivity, and 75.0% specificity). Likewise, Ibrahim et al. (2021) identified six VOCs (seven features) that could discriminate COVID-19-positive test patients and COVID-19-negative test patients (0.836 AUC, 68.0% sensitivity, and 85.0% specificity), although the model based on 11 VOCs showed 0.659 AUC, discriminating patients based on clinical suspicion. In both comparisons, only two VOCs, 1-propanol and benzaldehyde, were common, suggesting that the specific metabolic alterations caused by COVID-19 are not necessarily related to symptomatology, especially if the symptoms are shared with other upper respiratory

TABLE 3 Summary of studies focused on pneumonia and COVID-19. AB, alveolar breath; BSG, breath-gas sampler; CAP, community-acquired pneumonia; GC×GC, two-dimensional gas chromatography; HAP, hospital-acquired pneumonia; MB, mixed breath; MS, mass spectrometry; na, not applicable; nd, not detailed; NIST, National Institute of Standards and Technology; QTOF, quadrupole time-of-flight; SPME, solid-phase microextraction; TD, thermal desorption tube; TOF, time-of-flight; VAP, ventilator-associated pneumonia.

Reference	Pathology	Methodology	Sample	Sampling	Analysis technique	Sorbent material	Column	IS	Identification	
									Library	Authentic STD
Van Oort et al. (2022)	VAP	Untargeted	MB	BGS	TD-GC-MS	Carbograph 5TD 300 mg/Tenax GR 90 mg	VF1-MS column (30 m × 0.25 mm × 1 μm) (Varian)	Acetone-D8, hexane-D14, toluene-D8, and xylene-D10	NIST	No
Van Oort et al. (2017a)	CAP/HAP	Untargeted	MB	—	TD-GC-MS	Tenax GR 250 mg	VF1-MS column (30 m × 0.25 mm × 1 μm) (Varian)	No	NIST	No
Gao et al. (2016)	VAP	Untargeted	MB	—	TD-GC-MS	Tenax TA	Rtx-5MS (30 m × 0.25 mm × 0.25 μm)	No	NIST	No
Fowler et al. (2015)	VAP	Untargeted	MB	—	TD-GC-TOF-MS	Tenax TA/Carbotrap	RTX-5 amine column (30 m × 0.25 mm × 0.5 μm) (Restek)	4-Bromofluorobenzene	NIST	No
Schnabel et al. (2015)	VAP	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Carbograph 1TD/ Carbopack X	RTX-5MS (30 m × 0.25 mm × 1 μm)	No	NIST	No
Filipiak et al. (2015)	VAP	Targeted	AB	Glass syringe	TD-GC-MS	Carbotrap B 80 mg/ Carbopack X 260 mg	PoraBOND Q (25 m × 0.32 mm × 5 μm) (Varian)	No	NIST	No
Ahmed et al. (2023)	VAP	Targeted	MB	BGS	TD-GC-MS	Tenax GR 200 mg	DB-5MS (30 m × 0.25 mm × 0.25 μm) (Agilent)	4-Bromofluorobenzene	na	Yes
Cen et al. (2023)	COVID-19	Untargeted	AB	ReCIVA	TD-GC×GC-TOF-MS	Tenax TA/ Carbograph 5TD	DB-624 (30 m × 0.25 mm × 1.4 μm) (Agilent) 1D and DB-WAX column (5 m × 0.25 mm × 0.25 μm) (Agilent) 2D	Bromochloromethane, chlorobenzene-d5, and 1,4-dichlorobenzene-d4	NIST	Yes
Myers et al. (2023)	COVID-19	Untargeted	MB	Tedlar® bag	TD-GC-TOF-MS	Tenax TA/ Carbograph 1TD	nd	Toluene-D8	nd	Yes
Woollam et al. (2022a)	COVID-19	Untargeted	MB	Tedlar® bag	SPME-GC-QTOF-MS	DVB/Car/PDMS	HP-5MS (30 m × 0.25 mm × 0.25 μm) (Agilent)	No	nd	Yes
Ibrahim et al. (2021)	COVID-19	Untargeted	MB	Tedlar® bag	TD-GC-MS	Carbograph 1TD	DB-5MS (60 m × 0.25 mm × 0.25 μm) (Agilent)	Toluene-d8, phenanthrene-d10, and n-octane-d18	In-house library	Yes
Berna et al. (2021)	COVID-19	Targeted	MB	SamplePro FlexFilm Sample Bag	TD-GC×GC-TOF-MS	Tenax/Carbograph/ Carboxen	Stabilwax (30 m × 250 μm × 0.25 μm) (Restek) 1D and Rtx-200MS (5 m × 250 μm × 0.1 μm) (Restek) 2D	Bromochloromethane, 1,4-difluorobenzene, chlorobenzene-D5, and 4-bromofluorobenzene	na	Yes

10.3389/fmolb.2023.1295955

TABLE 4 Summary of group comparisons, statistical approaches, and identified VOCs in the studies focused on pneumonia and COVID-19. CAP, community-acquired pneumonia; CLZ, airway colonized; CN, culture-negative; CP, culture-positive; CTR, controls; FU COVID-19, follow-up samples of COVID-19 patients; HAP, hospital-acquired pneumonia; MVA, multivariate analysis; non-VAC, non-vaccinated; non-VAP, non-ventilator–associated pneumonia; PR CAP/HAP, probable community-acquired pneumonia/hospital-acquired pneumonia; PS CAP/HAP, possible community-acquired pneumonia/hospital-acquired pneumonia; UVA, univariate analysis; VAC, vaccinated; VAP, ventilator-associated pneumonia; VOCs, volatile organic compounds.

Reference	Pathology	Comparison	Statistical approach	Significant VOC	Detail
Van Oort et al. (2022)	VAP	CP (n = 52) vs. CN (n = 56)	UVA/MVA	1-Propenylbenzene (down), 2-bromophenol (down), 2-propenylbenzene (down), 2-methyldecane (up), 2,2-dimethyldecane* (up), 2,2,4,4-tetramethyloctane* (up), 2,6-difluorobenzaldehyde (up), 2,6,7-trimethyldecane (up), 3-methylheptane** (down), 6-methyl-5-hepten-2-one (up), cyclohexane (down), cyclohexanol (up), dimethyl sulfide* (up), enflurane (up), formaldehyde* (up), isopropylbenzene (down), m-di-tert-butylbenzene (down), and tetrahydrofuran (up)	*Significant VOCs in the UVA; **VOC reported in both platforms
Van Oort et al. (2017a)	CAP/HAP	PR CAP/HAP (n = 12) vs. PS CAP/HAP (n = 21) vs. CLZ (n = 13) vs. CTR (n = 47)/CP (n = 25) vs. CN (n = 68)	UVA/MVA	1-Pentanol* (down), 1-propanol** (down), 2-ethoxy-2-methyl-propane** (down), 2-methylcyclopentanone* (down), 5-methyl-2-heptanone* (down), acetone (down), carbon disulfide (down), cyclohexene (down), cyclohexanone* (down), hexafluoroisopropanol (down), methyl isobutyl ketone (down), and sevoflurane (down)	*VOCs colonized vs. non-colonized; **common VOCs PR CAP/HAP vs. CTR and colonized vs. non-colonized
Gao et al. (2016)	VAP	VAP (n = 20) vs. CLZ (n = 20) vs. CTR (n = 20)	UVA/MVA	1,5-Dimethyl-naphthalene (a), 1-undecene** (up), 2,6,10-trimethyl-dodecane* (up), 2-butyl-1-octanol* (up), 2-ethyl-1-hexanol (a), 5-methyl-5-propyl-nonane* (up), benzaldehyde (a), butylated hydroxytoluene (a), cyclohexanone (a), decanal** (up), ethanol (a), isoprene (a), longifolene** (up), n-nonylcyclohexane (a), nonanal* (up), tetradecane** (up), toluene (a), α-cedrene (a), and α-funebrene (a)	*Significant VOCs derived from *Acinetobacter baumannii*; **common VOCs *in vitro* and *in vivo*
Fowler et al. (2015)	VAP	CP (n = 15–26) vs. CN (n = 31–20)	MVA	2,6,11,15-Tetramethyl-hexadecane (up), 2-methyl cyclopentanone (down), 3-carene (up), ethanol (down), heptane (down), n-butyric acid 2-ethylhexyl ester (up), N-cyclohexyl-N′(2-hydroxyethyl)thiourea (down), and nonanal (up)	
Schnabel et al. (2015)	VAP	VAP (n = 32) vs. non-VAP (n = 68)	MVA	Acetone (down), acrolein (down), butane, 2-methyl (up), carane (up), dodecane (down), ethanol (up), ethylbenzene (up), tetrahydrofuran (down), heptane (up), isopropyl alcohol (down), tetradecanal (up), and tetradecane (up)	
Filipiak et al. (2015)	VAP	VAP (n = 22) vs. non-VAP (n = 6)	-	(E)-2-Butene (a), (Z)-2-butene (a), 1,3-butadiene (a), 1-undecene***** (a), 2-methyl-1-butene****** (a), 2-methylpropene (a), 2-pentanone*** (a), 3-methylbutanal (a), 3-methyl-1-butene**** (a), 4-heptanone** (a), acetaldehyde (a), acetic acid (a), acetonitrile*** (a), benzaldehyde (a), butane* (a), dimethyl sulfide*** (a), ethanol (a), ethyl acetate (a), hexanal (a), hexane****** (a), iso-butane****** (a), methacrolein (a), methanol (a), methyl vinyl ketone (a), propanal (a), and propane* (a)	*VOCs related to the course of infection with *Staphylococcus aureus*; **VOCs related to the course of infection with *Candida albicans*; ***VOCs related to the course of infection with *Escherichia coli*; ****VOCs related to the course of infection with *Haemophilus influenzae*; *****VOCs related to the course of infection with *Pseudomonas aeruginosa*; *******VOCs related to the course of infection with *Streptococcus pneumoniae*

(Continued on following page)

TABLE 4 (*Continued*) Summary of group comparisons, statistical approaches, and identified VOCs in the studies focused on pneumonia and COVID-19. CAP, community-acquired pneumonia; CLZ, airway colonized; CN, culture-negative; CP, culture-positive; CTR, controls; FU COVID-19, follow-up samples of COVID-19 patients; HAP, hospital-acquired pneumonia; MVA, multivariate analysis; non-VAC, non-vaccinated; non-VAP, non-ventilator–associated pneumonia; PR CAP/HAP, probable community-acquired pneumonia/hospital-acquired pneumonia; PS CAP/HAP, possible community-acquired pneumonia/hospital-acquired pneumonia; UVA, univariate analysis; VAC, vaccinated; VAP, ventilator-associated pneumonia; VOCs, volatile organic compounds.

Reference	Pathology	Comparison	Statistical approach	Significant VOC	Detail
Ahmed et al. (2023)	VAP	CP (n = 45) vs. CN (n = 59)	UVA	3-Methylbutanoic acid (up), 3-methylbutanal* (down/up*), acetone* (down), and indole** (up)	*Significant VOCs *P. aeruginosa* vs. other pathogen-positive culture; **significant VOC in patients with positive culture for bacteria that can metabolize tryptophan; down/up* different alterations between group comparisons
Cen et al. (2023)	COVID-19	VAC (n = 54) vs. non-VAC (n = 50)	UVA/MVA	2-Methyloctane* (down), 6-methyl-5-hepten-2-one (up), acetonitrile* (down), benzene (down), benzothiazole (up), cyclopentanone (up), hexanal* (down), methanesulfonyl chloride (up), and phenol* (down)	*VOCs in UVA
Myers et al. (2023)	COVID-19	COVID-19 (n = 69) vs. FU COVID-19 (n = 22) vs. CTR (n = 58) vs. HC (n = 21)	UVA/MVA	1-Propene, 1-(methylthio)-, (E)- (down), 2,2,4,6,6-pentamethylheptane**/*** (up), 2,2,4-trimethylpentane*/*** (up), 2-methyldecane*** (down/up*), 2-methylpentane**/*** (up), 2-pentanone*** (up), 3-methylheptane*/*** (up), allyl methyl sulfide*/*** (down/up*), cyclohexanone*** (up), dimethyl disulfide (down), ethyl acetate**/*** (up), heptanal (up), hexane**/*** (up), indole*** (up), methyl acetate**/*** (down), methyl butyrate**/*** (up), sulcatone*/*** (down/up*), α-phellandrene**/*** (down), and γ-terpinene**/*** (down)	*Common VOCs between comparisons: COVID-19 vs. FU COVID-19 and COVID-19 vs. CTR; **significant VOCs in COVID-19 vs. FU COVID-19; ***significant VOCs in the UVA; down/up* different alterations between group comparisons
Woollam et al. (2022a)	COVID-19	COVID-19 (n = 14) vs. HC (n = 12)	UVA/MVA	3,5,5-Trimethylhexanal (up), cedrene (up), and hexyl acetate (up)	
Ibrahim et al. (2021)	COVID-19	COVID-19 (n = 52) vs. CTR (n = 29)	MVA	1-Propanol** (up), 2,2-dimethyl-1-propanol* (a), 3-heptene* (a), 3,6-dimethylundecane (up), 4-ethenyl-1,2-dimethyl-benzene* (a), acetic acid methyl ester* (a), acetoin* (a), benzaldehyde** (a), camphene (up), cyclohexene* (a), iodobenzene (up), octanal* (a), pentadecane* (a), tetrachloroethylene* (a), and β-cubebene (up)	*VOCs identified in clinical suspicion comparison; **VOC identified in both comparisons (COVID-19 test and clinical suspicion)
Berna et al. (2021)	COVID-19	COVID-19 (n = 22) vs. CTR (n = 27)	UVA/MVA	2-Pentyl-furan (up), dodecane (up), heptanal (up), nonanal (up), octanal (up), and tridecane (up)	

infections. These two studies presented one VOC in common, octanal (Table 5).

Conversely, two other studies were conducted on non-hospitalized COVID-19-positive patients. Woollam et al. (2022a) enrolled a cohort undergoing COVID-19 testing due to symptom onset, contact with symptomatic individuals, or mitigation testing. When COVID-19 patients were compared with HC, 41 VOCs were found to be significantly altered, mostly upregulated. Curiously, COVID-19 patients were divided into two subclasses based on their VOC profiles, one of which presented 4 of the 41 VOCs upregulated compared to the other subclass and HCs. Furthermore, the set of 41 VOCs could distinguish among groups with 96.0% accuracy, increasing to 100% when the 16 most significant VOCs were selected. The predictive classification model based on three VOCs (hexyl acetate, cedrene, and 3,5,5-trimethylhexanal) presented 100% sensitivity and 92.0% specificity value (e.i. 0.990) AUC. Lastly, 11 COVID-19 patients were sampled after recovery, and 34 VOCs recovered baseline levels, although five were still upregulated. When including this group in the final model, recovered COVID-19 patients clustered with controls and could be distinguished from COVID-19 patients with 90.0% accuracy. Myers et al. (2023) included patients presenting upper respiratory infections from two different ambulatory

TABLE 5 VOCs reported in pneumonia and COVID-19 (two or more studies). na, not applicable; ppbv, parts per billion by volume. *Downregulated in culture positive for *P. aeruginosa* ventilator-associated pneumonia patients and upregulated in culture positive for *S. aureus* ventilator-associated pneumonia patients; **downregulated in COVID-19 patients compared to follow-up COVID-19 patients and upregulated in COVID-19-positive patients compared to COVID-19-negative patients and healthy controls.

Pneumonia (CAP/HAP/VAP)									
No.	Compound name	Cas-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
1	Ethanol	64-17-5	C_2H_6O	Alcohol	Upregulated	na	na	na	Schnabel et al. (2015)
					Altered	na	na	na	Gao et al. (2016)
					Altered	na	na	na	Filipiak et al. (2015)
					Downregulated	na	na	na	Fowler et al. (2015)
2	3-Methylbutanal	590-86-3	$C_5H_{10}O$	Aldehyde	Altered	na	na	na	Filipiak et al. (2015)
					Downregulated/ upregulated*	na	na	na	Ahmed et al. (2023)
3	Benzaldehyde	100-52-7	C_7H_6O	Aldehyde	Altered	na	na	na	Gao et al. (2016)
					Altered	na	na	na	Filipiak et al. (2015)
4	Nonanal	124-19-6	$C_9H_{18}O$	Aldehyde	Upregulated	na	na	na	Fowler et al. (2015)
					Upregulated	na	na	na	Gao et al. (2016)
5	Tetrahydrofuran	109-99-9	C_4H_8O	Ether	Downregulated	na	na	na	Schnabel et al. (2015)
					Upregulated	na	na	na	Van Oort et al. (2022)
6	Heptane	142-82-5	C_7H_{16}	Hydrocarbon (saturated)	Upregulated	na	na	na	Schnabel et al. (2015)
					Downregulated	na	na	na	Fowler et al. (2015)
7	Tetradecane	629-59-4	$C_{14}H_{30}$	Hydrocarbon (saturated)	Upregulated	na	na	na	Schnabel et al. (2015)
					Upregulated	na	na	na	Gao et al. (2016)
8	1-Undecene	821-95-4	$C_{11}H_{22}$	Hydrocarbon (unsaturated)	Upregulated	na	na	na	Gao et al. (2016)
					Altered	na	na	na	Filipiak et al. (2015)
9	2-Methylcyclopentanone	1120-72-5	$C_6H_{10}O$	Ketone	Downregulated	na	na	na	Fowler et al. (2015)
					Downregulated	na	na	na	Van Oort et al. (2017a)
10	Acetone	67-64-1	C_3H_6O	Ketone	Downregulated	na	na	na	Van Oort et al. (2017a)
					Downregulated	na	na	na	Schnabel et al. (2015)
					Downregulated	na	na	na	Ahmed et al. (2023)

(Continued on following page)

TABLE 5 (*Continued*) VOCs reported in pneumonia and COVID-19 (two or more studies). na, not applicable; ppbv, parts per billion by volume. *Downregulated in culture positive for *P. aeruginosa* ventilator-associated pneumonia patients and upregulated in culture positive for *S. aureus* ventilator-associated pneumonia patients; **downregulated in COVID-19 patients compared to follow-up COVID-19 patients and upregulated in COVID-19-positive patients compared to COVID-19-negative patients and healthy controls.

Pneumonia (CAP/HAP/VAP)									
No.	Compound name	Cas-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	Reference
11	Cyclohexanone	108-94-1	$C_6H_{10}O$	Ketone	Altered	na	na	na	Gao et al. (2016)
					Downregulated	na	na	na	Van Oort et al. (2017a)
12	Dimethyl sulfide	75-18-3	C_2H_6S	Sulfur-containing	Upregulated	na	na	na	Van Oort et al. (2022)
					Altered	0–101.5	na	ppbv	Filipiak et al. (2015)
COVID-19									
No	Compound name	CAS-N	Formula	Chemical class	Sign of alteration	Concentration (patients)	Concentration (controls)	Unit	First author/ year
1	Heptanal	111-71-7	$C_7H_{14}O$	Aldehyde	Upregulated	na	na	na	Berna et al. (2021)
					Upregulated	na	na	na	Myers et al. (2023)
2	Octanal	124-13-0	$C_8H_{16}O$	Aldehyde	Upregulated	na	na	na	Berna et al. (2021)
					Altered	na	na	na	Ibrahim et al. (2021)
3	Methyl acetate	79-20-9	$C_3H_6O_2$	Ester	Altered	na	na	na	Ibrahim et al. (2021)
					Downregulated	na	na	na	Myers et al. (2023)
4	6-Methyl-5-hepten-2-one	110-93-0	$C_8H_{14}O$	Ketone	Upregulated	na	na	na	Cen et al. (2023)
					Downregulated/ upregulated**	na	na	na	Myers et al. (2023)

care settings. Moreover, some COVID-19 patients infected with Alpha, Beta, or Delta variants were resampled after 8–12 weeks (FU COVID-19). In the MVA, 12 VOCs could discriminate between COVID-19 and FU COVID-19 patients value (e.i. 0.825–0.862) AUC. Furthermore, COVID-19 patients and controls (COVID-19-negative test patients presenting symptoms) could be distinguished by 11 VOCs, which were further validated in an independent cohort value (e.i. 0.960, 80.0% sensitivity, and 90.0% specificity) AUC. From both comparisons, four common VOCs (2,2,4-trimethylpentane, sulcatone, allyl methyl sulfide, and isobutyric acid) were identified.

Additionally, Cen et al. (2023) investigated the metabolic reprogramming triggered by the inactivated COVID-19 vaccine, comparing the VOC profiles of COVID-19 vaccinated and unvaccinated subjects. The discriminative model based on nine VOCs (from 21 identified in both UVA and MVA), which included 6-methyl-5-hepten-2-one already found by Myers et al. (2023) (Table 5), exhibited 94.4% overall accuracy, 91.3% sensitivity, and 98.6% specificity value (e.i. 0.995) AUC. Furthermore, the examination of the biomarkers' metabolic pathways demonstrated that the protective metabolic regulation induced by the vaccine influences enzymatic activity and microbial metabolism within the lungs, liver, and gastrointestinal tract.

3.4 Searching for pathology-specific VOCs in human exhaled breath

The search for potential biomarkers in exhaled breath is challenging due to the substantial variability in the concentration of VOCs. This variability is due to metabolic activity but also depends on lifestyle choices (smoking, exercise, diet, etc.) and/or exposure to exogenous factors, such as pollutants and other environmental compounds, among others. Despite this challenge, numerous studies have focused

on identifying specific VOCs associated with a wide range of pathologies. Nevertheless, the results of these studies should be interpreted with caution. In most instances, the origin of these VOCs remains unidentified, which can lead to false discoveries.

The VOCs included in Tables 2, 5; Supplementary Table S3, as classified by Drabińska et al. (2021), are illustrated in Figure 4. As noted, the analysis of exhaled breath covers a wide range of chemical species, although the distribution of these is variable across pathologies. Aldehydes are the most abundant, mainly derived from alcohol metabolism in the liver or the reduction of hydroperoxides during lipid peroxidation (Murray et al., 2009; Hakim et al., 2012), although aldehydes can also come from cigarette smoking or tobacco components' detoxification by cytochrome P450 (Furge and Guengerich, 2006; Papaefstathiou et al., 2020). This chemical group is predominant in CAP/HAP/VAP, COVID-19, and COPD.

Ketones are also strongly represented in CAP/HAP/VAP and COVID-19, mainly resulting from the liver's synthesis of ketone bodies (acetoacetate, acetone, etc.) during conditions like diabetes, fasting, or alcoholism, formed through the metabolism of proteins and/or as secondary products of lipid peroxidation (Vaz and Coon, 1987; Murray et al., 2009). Remarkably, a significant proportion of ethers is observed in COVID-19 and BC, although its origin is commonly attributed to exogenous sources. Other abundant compounds in BC are alcohols, which may come from the gastrointestinal tract or are formed through the hydrocarbon's metabolism or lipid peroxidation (Ortiz De Montellano, 2010; Ratcliffe et al., 2020).

Additionally, hydrocarbons are widely reported in exhalates, primarily saturated, aromatic, branched and cyclic. These compounds are highly represented in LC, asthma, and CRC. Hydrocarbons are mainly produced by lipid peroxidation, in an abnormal metabolic state. Branched-chain hydrocarbons may be of an endogenous origin from bacterial metabolism (Ratcliffe et al., 2020).

The concept of the exposome is gaining popularity, encompassing not only external exposures (chemical agents, radiation, etc.) and associated physiological responses but also internal sources, such as microbiota, and "psychosocial components" (Vineis et al., 2020). Several studies have focused on identifying metabolites related to the exposome, such as the database developed by Neveu et al. (2023), which includes microbial metabolites and is supported by evidence on their origin, and the method developed by González-Domínguez et al. (2020) for exposome research. Furthermore, the effect of the exposome on human health has been widely studied (Morales et al., 2022). Nevertheless, many metabolites associated with the exposome overlap with those produced by human cells/tissues, making it a difficult task to establish what can be considered truly endogenous. This issue is especially challenging for VOCs detected in exhaled breath, since the pulmonary tract is closely associated with environmental exposure.

The full list of reported VOCs was used to identify pathology-specific compounds (Figure 5). The overlapping VOCs may come from exogenous sources (exposome), such as cigarette smoking, environmental pollution, or diet, as well as shared endogenous origins like the ones derived from OS or common VOCs found in breath, such as isoprene and acetone.

Regarding pathology-specific possible biomarkers, several unique VOCs were found (Figure 5), especially in LC, according to the literature reviewed herein. Furthermore, the obtained list of unique pathology-specific VOCs was submitted to searches on KEGG and BioCyc databases with the aim of excluding the VOCs mainly coming from exogenous sources. Upon exclusion of such exogenous VOCs, the final list of pathology-specific possible biomarkers is compiled in Table 6. It is worth mentioning that although these candidate biomarkers might provide useful information, further research is required to establish associations with metabolic alterations in each pathology, as well as to discern between the VOCs that may be related to the exposome and the ones that are truly endogenous.

Furthermore, the correct metabolite identification is a highly important aspect in metabolomics, and different levels can be distinguished based on the reliability of the identification. In this regard, the Metabolomics Standard Initiative (MSI) levels can range from 1 to 4, level 1 being the most rigorous (Sumner et al., 2007).

In the case of LC, four VOCs from Table 6 were reported as potential biomarkers in at least two different studies (Supplementary Table S3). In this regard, 2,3-butanedione and butanal (both MSI level 1) were found to be upregulated. The remaining candidate biomarkers were reported as altered; thus, the trend of their levels should be assessed. Propionic acid (MSI levels 1 and 2) was reported as upregulated and downregulated in two different studies; therefore, it is not an adequate candidate due to the contradictory findings. The remaining metabolites included in Table 6 were reported only once, requiring further study for their use as pathology-specific biomarkers. Additional candidate biomarkers whose endogenous origin has not been established include 2-nonenal (MSI levels 1 and 2), 3-methylhexane (MSI levels 1 and 2), butanal (MSI level 1), pentane (MSI level 1), and propylene (MSI levels 1 and 2) for LC, all of which are reported several times and show a trend toward increased levels; acrylonitrile (MSI level 1) for GaC is reported as upregulated in two independent studies; methacrylic acid (MSI level 2) for BC presents decreased levels; and 1,2-dimethylcyclohexane (MSI level 2) for asthma and 2-methylcyclopentanone (MSI level 2) for CAP/HAP/VAP are reported as downregulated (Tables 2, 5; Supplementary Table S3).

4 Methodologies

The methodologies used for breath sampling and VOCs' pre-concentration and separation in the reviewed studies are presented in Tables 1, 3; Supplementary Table S1 and illustrated in Figure 6.

4.1 Exhaled breath sampling

Breath samples were categorized in mixed or alveolar breath, as late expiratory breath sampling was not specified in any study, being usually confused with alveolar breath. The lack of distinction may be due to the absence of standardized protocols or guidelines for the collection of late expiratory breath. Devices that discard dead space

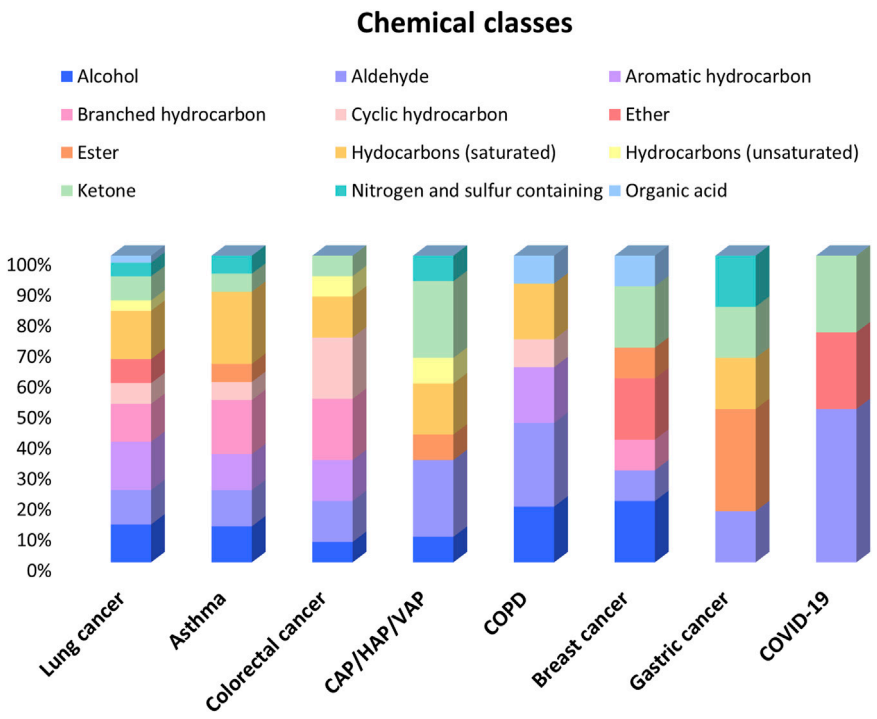


FIGURE 4
Bar chart of the chemical classes of VOCs (referred by two or more studies) reported in each pathology. CAP, community-acquired pneumonia; COPD, chronic obstructive pulmonary disease; HAP, hospital-acquired pneumonia; VAP, ventilator-associated pneumonia.

air may not ensure true alveolar breath sampling; therefore, only those with CO_2 or pressure sensors should be used for this type of sample.

Exhaled breath sampling devices can collect from a few milliliters to 10 L, which depends not only on the device's capacity but also on the fraction of breath sampled, since alveolar breath represents approximately 350 mL of the total expiratory volume. Even though breath samples can be taken from a single expiration or multiple expirations, VOC profiles can vary from breath to breath (Khoubnasabjafari et al., 2022), and the concentrations differ significantly in hypoventilation, hyperventilation, and normal ventilation (Cope et al., 2004).

The selection of sample type depends on the compounds of interest. When studying endogenous VOCs, late expiratory or alveolar breath are preferred, the latter being more convenient due to the higher concentration of VOCs and reduced contamination (Miekisch et al., 2008). However, as depicted in Figure 6A, mixed breath was the most analyzed sample type, probably because the devices used for this type of sample are more affordable and easier to use. However, due to the increasing interest in endogenous VOCs, alveolar breath was analyzed in a significant number of studies, especially in those focused on cancer.

The most common collection device was Tedlar® bags (Figure 6B), consistent with previous reviews (Lawal et al., 2017; Westphal et al., 2023). While these bags are subject to contamination and have limited sample storage time, they are affordable and reusable, with several cleaning protocols available (Westphal et al., 2023). Additionally, devices such as Bio-VOC®, breath-gas sampler, and ReCIVA are gaining popularity, although they are not as widely used.

Sampling methodologies are organized per type of pathology in this section as the severity of the pathology may justify specific approaches.

4.1.1 Cancer

In LC, 14 studies analyzed alveolar breath, while the remaining 9 analyzed mixed breath. The alveolar breath samples were collected using various techniques, such as Bio-VOC®, BCA, Tedlar®/Mylar bags, or other devices (analytical barrier bag and breath reservoir). For mixed breath samples, Tedlar®/Mylar bags were predominantly used, and some studies employed self-developed devices and glass bulbs.

In GaC, alveolar breath was the main type, sampled either with Mylar/Tedlar® bags, GaSampler collection bags (QuinTron), or a custom-built in-house breath sampler. One study used gas-tight syringes for mixed breath sampling. In CRC, two studies sampled mixed breath using Tedlar® bags, and the remaining alveolar breath employed ReCIVA® or other devices [GaSampler Collection Bag (QuinTron) and gas-tight syringes]. In BC, all studies collected alveolar breath, using various devices such as gas-tight syringes, Tedlar® bags, and Bio-VOC®.

4.1.2 Other pulmonary pathologies

In asthma, all the studies collected mixed breath samples using Tedlar® bags, except for one that analyzed alveolar breath and used an ABS. In COPD, alveolar breath and mixed breath were selected

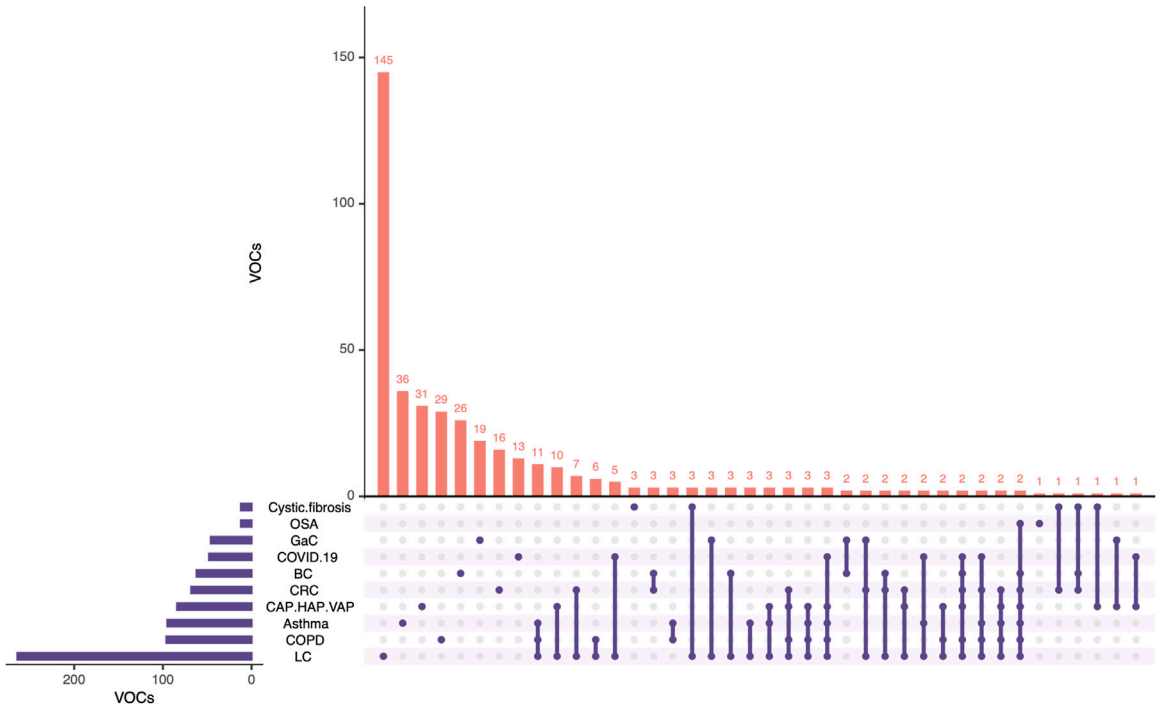


FIGURE 5
Upset plot of the VOCs reported in the reviewed studies. BC, breast cancer; CAP, community-acquired pneumonia; COPD, chronic obstructive pulmonary disease; CRC, colorectal cancer; GaC, gastric cancer; HAP, hospital-acquired pneumonia; LC, lung cancer; OSA, obstructive sleep apnea; VAP, ventilator-associated pneumonia. Created with RStudio (Conway et al., 2017).

TABLE 6 Pathology-specific proposed biomarkers. BC, breast cancer; CAP, community-acquired pneumonia; COPD, chronic obstructive pulmonary disease; CRC, colorectal cancer; GaC, gastric cancer; HAP, hospital-acquired pneumonia; LC, lung cancer; OSA, obstructive sleep apnea; VAP, ventilator-associated pneumonia. *Possible exogenous origin.

Pathology	PubChem ID	Compound*
LC	650	2,3-Butanedione
LC	261	Butanal
LC	1032	Propionic acid
LC	984	Hexadecanal
GaC	225936	2,3-Butanediol
CRC	243	Benzoic acid
CRC	2969	Decanoic acid
BC	10413	4-Hydroxybutanoic acid
Asthma	11005	Myristic acid
Asthma	637540	2-Hydroxycinnamic acid
COPD	2879	p-Cresol
OSA	999	Phenylacetic acid
CAP/HAP/VAP	10430	Isovaleric acid

with equal frequency. The samplers employed were either Tedlar® bags or stainless steel tubes for mixed breath, and Bio-VOC® or glass syringes for alveolar breath. Furthermore, in one study, alveolar breath was also collected directly into the pre-concentration device. In OSA, alveolar and mixed breath samples were sampled with Bio-VOC® and Tedlar® bags, respectively, and in CF, mixed breath was collected using Tedlar® bags.

4.1.3 Infectious pathologies

In CAP/HAP/VAP, one of the targeted studies sampled alveolar breath employing glass syringes, while the remaining studies analyzed mixed breath samples. Since the patients were intubated and mechanically ventilated, most of the sampling was performed directly in sorbent tubes, except for three studies that employed either Tedlar® bags or a breath-gas sampler. In COVID-19, most of the studies analyzed mixed breath, collecting the sample either using Tedlar® bags or SamplePro FlexFilm Sample Bags. Only one study sampled alveolar breath employing ReCIVA.

4.2 Pre-concentration techniques

TD were the most utilized pre-concentration technique, followed by SPME (Figure 6C), consistent with previous reviews (Lawal et al., 2017; Westphal et al., 2023). The widespread use of TD can be attributed to their suitability for long-term sample storage, ease of transport, and the stability of the entrapped compounds. However, SPME requires smaller sample volumes and is less affected by humidity, offering a similar extraction range to TD. Notably, NTD was used in only a few studies, despite presenting aspects of both SPME and TD.

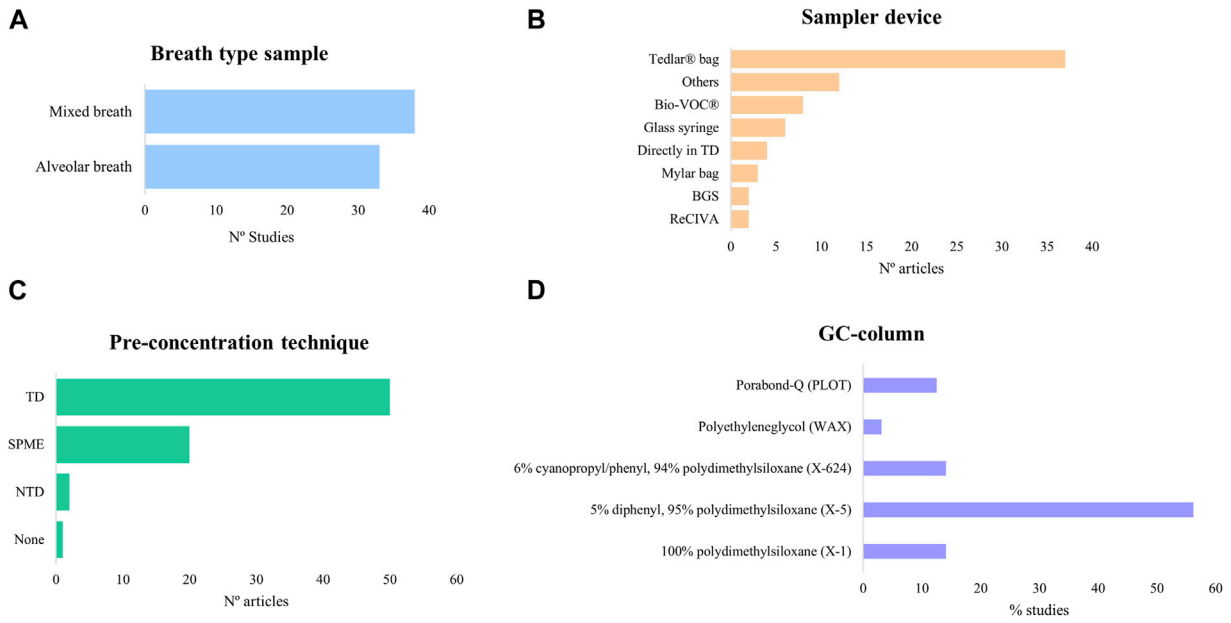


FIGURE 6
Bar charts of the exhaled breath sampling and pre-concentration methods from the reviewed studies. **(A)** Breath-type sample; **(B)** sampler device; **(C)** pre-concentration technique; **(D)** GC-column. BGS, breath-gas sampler; GC, gas chromatography; NTD, needle-trap device; SPME, solid-phase microextraction; TD, thermal desorption tube.

The choice of the sorbent material depends on the chemical nature of the analytes of interest, including polarity and molecular weight (MW). The most used SPME fiber coating material was Car/PDMS (mainly 75 µm thickness), followed by DVB/Car/PDMS, and lastly, PDMS 100 µm and PDMS/DVB (Tables 1, 3; Supplementary Table S1). These fibers are bipolar, except for PDMS (non-polar), with the latter compromising the extraction of polar metabolites (Vas and Vékey, 2004). The wider use of Car/PDMS fiber could be due to the ability of this coating material to better extract low-molecular-weight volatiles (MW 30–225 g/mol) compared to PDMS 100 µm (MW 60–275 g/mol) and DVB/PDMS (MW 80–300 g/mol) (Lawal et al., 2017). However, in a recent study (Schulz et al., 2023), DVB/Car/PDMS turned out to be the most adequate for untargeted studies compared to Car/PDMS and PDMS fibers due to the higher number of extracted VOCs and the stronger overall GC-MS signal.

Regarding TD, Tenax TA is the most used sorbent material (Tables 1, 3; Supplementary Table S1). Although this material captures heavy- and less-volatile compounds, its low affinity to water and the broad sampling range (C6–C30) makes it adequate for untargeted analysis (Wilkinson et al., 2020). Similarly, a few studies have opted for Tenax GR. Other sorbent materials encompass carbon black adsorbents, such as Carbograph 1TD, Carbograph 5TD, Carbopack X, Carbopack B, and Carbotrap. Additionally, carbon-based materials like Carboxen were employed. These alternatives offer a narrower range (C3–C20 and C2–C5), although they facilitate the capture of low-molecular-weight and more volatile compounds (Lawal et al., 2017; Westphal et al., 2023). Many studies have employed multi-bed sorbents, with the most prevalent being Carbograph 1TD/Carbopack X. Furthermore, combinations of the aforementioned materials have been used, such as Tenax TA/Carbograph 1TD and Tenax TA/Carbograph 5TD, both of which have been considered for exhaled breath analysis in previous studies (Wilkinson et al., 2020). Additional combinations, such as Tenax/Carbograph/Carboxen and Carbotrap B/Carbopack X, were also employed.

4.3 GC-MS methods

Considering that the main objective of most studies discussed in the present review was to obtain a snapshot of the VOC content in the breath samples, and also the analysis conditions and the pre-concentration techniques. Conditions of the GC-MS method such as the injector mode, chromatographic column, and type of detector employed for analysis are as important as the preceding pre-concentration technique.

Regarding GC injector parameters, from the 70 works reviewed, an astonishing 61% do not detail the type of injector or injection employed. Such a number is alarming given the fundamental difference between injecting gaseous and liquid samples. In fact, for gaseous injections choosing an injector glass liner of smaller inner diameter would provide a more efficient transfer of analytes onto the GC column, thus yielding more peak capacity efficiency. Nonetheless, of the 39% of works that mentioned the employment of a splitless/split injector, none mentioned the dimensions of the injector glass liner diameter employed. Out of these split/splitless injections, 70% of the injected samples are in the splitless mode, which would indeed be expected for pre-concentration techniques such as direct TD and SPME.

When it comes to GC columns as presented in Figure 6D (compiling the information from Tables 1, 3; Supplementary Table S1), over 50% of the studies herein reviewed employed 5%

diphenyl/95% polydimethylsiloxane phases, here named as X-5 columns (DB-5, RTX-5, SLB-5, HP-5, VF-5, etc.). This type of stationary phase is considered the most versatile owing to the slight polarity imparted by the substitution of 5% of dimethyl groups by diphenyl. This addition also makes this stationary phase suitable for the separation of unsaturated hydrocarbons and aromatic compounds. Conversely, this stationary phase should not be the first choice regarding the analysis of VOCs. In fact, despite its slight polarity, it does not provide sufficient retention and efficiency for the separation of low-molecular-weight polar VOCs, such as alcohols, aldehydes, and organic acids.

The second most used types of GC stationary phase are X-624 and X-1, each being employed in 14% of the studies presented in Tables 1, 3; Supplementary Table S1. The X-624 stationary phase (DB-624, VP-624, SLB-624, etc.) consists of polydimethylsiloxane with 6% of the dimethyl substituted by cyanopropyl and phenyl groups. Therefore, as expected, this is a low-polarity phase, though of higher polarity than its X-5 counterparts. A key characteristic of these columns is the thickness of the stationary phase. While most X-5 columns employ stationary phases of 0.25 µm thickness, X-624 columns are coated with, at least, 1.4 µm of the stationary phase. Therefore, in addition to its chemistry allowing for better selectivity, there is also a considerable gain in retention for low-molecular-weight polar VOCs and a wider range of VOC classes could be successfully analyzed in breath samples. In fact, X-624 stationary phases are the most suitable for VOC analyses, as it is the official stationary phase for a variety of Environmental Protection Agency (EPA) methods dealing with VOCs. X-1 columns (DB-1, RTX-1, SLB-1, HP-1, VF-1, etc.) contain 100% polydimethylsiloxane, the most non-polar stationary phase. Similar to X-5 types of phases, this stationary phase does not provide sufficient selectivity for the separation of small and polar VOCs even when employing thicker phases.

Given the importance of aldehydes and alcohols in the studies herein reviewed, as presented in Figure 4, it may be surprising that only two studies employ polyethylene glycol (WAX) stationary phases, as they are highly selective for polar compounds such as alcohols. A plausible explanation in this type of application might be related to the presence of water in the breath sample. WAX columns are particularly sensitive to moisture in the sample, which may lead to the degradation of this stationary phase.

While all other GC columns mentioned here encompass wall-coated open tubular (WCOT) columns, porous layer open tube (PLOT) columns are highly retentive and, therefore, are primarily employed for the analysis of very low boiling point compounds that are gaseous at room temperature, such as sulfides. Applied in 12% of the studies herein reviewed, Q-PLOT columns are non-polar, as they employ 100% divinylbenzene as adsorbent, therefore imparting great selectivity and retention for low-molecular-weight hydrocarbons.

Most of the studies included in Tables 1, 3; Supplementary Table S1 were performed by one-dimensional (1D) GC-MS. Additionally, five studies applied comprehensive bidimensional gas chromatography (GC×GC). The most common GC×GC setup employs an orthogonal mechanism of separation, using two sequential GC columns with stationary phases of different polarities, with a modulator between them. In short, a narrow band eluting from the first dimension (1D) column is collected and focused on the modulator, and then sent to the second dimension (2D) column (which is much shorter than the 1D). In this way, for example, compounds of similar boiling points coeluting on the 1D could be separated according to their polarity differences on the 2D. This technique offers significant advantages: the extended peak capacity improves peak space separation and allows the detection of coeluting compounds that could be missed by conventional 1D-GC. Moreover, given the acquisition speed required by the narrow bands eluting from the 2D in GC×GC, this technique is often hyphenated with MS detectors with rapid MS analyzers such as TOF, providing also higher sensitivity than that obtained by 1D-GC employing single quadruple MS. As an example of the advanced capacities of GC×GC-TOF-MS, Caldeira et al. (2012) could detect eight-fold more compounds, especially alkanes, alkenes, aldehydes, and ketones, and the concentration range achieved was lower than that of a previous study performed with GC-qMS. Four out of the five studies presented the traditional apolar × polar configuration, employing stationary phases like X-624 and X-5 in the first dimension (1D) and a polar polyethylene glycol-based (WAX) phase in the second dimension (2D) (Caldeira et al., 2012; Pesesse et al., 2019; Schleich et al., 2019; Cen et al., 2023). This combination reduces the interaction of water with the polar stationary phase and provides information on both the volatility (1D) and polarity (2D) of the compounds in the sample (Wilde et al., 2019). Interestingly, Berna et al. (2021) employed a polar × polar setup, using a WAX-based column in the 1D and a trifluoropropylmethyl polysiloxane (RTX-200) column in the 2D. The 2D stationary phase offers a unique selectivity for electron-rich molecules and resolves compounds that could not be resolved by the Wax 1D column. The limited use of this technology can be attributed to the high costs of instrumentation, especially for cryo-based modulators that are the most adequate for applications such as breath analysis due to their ability to successfully trap very volatile compounds. Moreover, method optimization in GC×GC is not as straightforward as in 1D-GC, hence requiring specialized personnel from method development to data process and interpret (Pesesse et al., 2019).

Moreover, 20 studies have employed high-resolution mass spectrometers (TOF-MS), of which more than half were included in studies on other pulmonary diseases (Tables 1, 3; Supplementary Table S1). The high-resolution approach offers notable advantages, especially when performing an untargeted study. In this regard, sensitivity and selectivity are improved compared to the low-resolution approach. The spectral libraries used for compound identification include accurate mass, which further allows for the enhanced structural elucidation of unknown compounds. Nonetheless, the use of this equipment is more complex; data processing requires more time and space and the price is higher (Rey-Stolle et al., 2021).

4.4 Quality control

In breath analysis, as in any metabolomics study, evaluating the quality of the obtained data is crucial. This is essential not only for obtaining reproducible results but also to ensure that the differences observed between groups are attributable to the composition of the samples rather than analytical/instrumentation variations. Such assurance involves the analysis of blanks and the use of internal standards among other strategies detailed below (Dudzik et al., 2018).

The analysis of blanks allows for the identification of contaminants (e.g., Tedlar® bag contaminants) and artifacts (e.g., polydimethylsiloxanes), and its elimination from the data matrix, avoiding false discoveries. The main blanks include collecting device blank, air blank, and vial/tube blank (Westphal et al., 2023). In this regard, several studies have reviewed and conducted analyses of ambient air, yet a considerable number of studies do not include this step, such as that of Gao et al. (2016), Aoki et al. (2017), Van Vliet et al. (2017), and Saidi et al. (2020). Moreover, the inclusion of the remaining blanks is not specified in most of the studies. The injection of a standard mixture over the sequence is recommended, although just a few studies include it in their workflow (Schleich et al., 2019; Koureas et al., 2021).

Despite being of utmost importance to obtain precise results, the addition of an internal standard (IS) to the samples was performed in few studies (Basanta et al., 2012; Corradi et al., 2015; Fowler et al., 2015; Berna et al., 2021; Ibrahim et al., 2021; Van Oort et al., 2022; Ahmed et al., 2023; Cen et al., 2023; Myers et al., 2023) (Tables 1, 3; Supplementary Table S1). Likewise, hexamethylcyclotrisiloxane, a desorption tube bleeding compound, was used as an internal reference compound (Callol-Sanchez et al., 2017; Jareño-Esteban et al., 2017; Muñoz-Lucas et al., 2020). The ISs used include not isotopically labeled compounds such as 1,4-difluorobenzene, 2-methylpentanal, 4-bromofluorobenzene, bromochloromethane, and stable isotopically labeled ones, such as acetone-d8, 1,4-dichlorobenzene-d4, chlorobenzene-d5, bromobenzene-d5, hexane-d14, n-heptane-d16, n-octane-d18, phenanthrene-d10, styrene-d8, toluene-d8, and xylene-d10. Additionally, ISs can be used for data normalization. However, most of the studies reviewed did not add an IS nor did they specify how the data normalization is performed. It is worth mentioning that for a pre-concentration technique based on equilibrium, such as SPME, isotopically labeled ISs present by far the best precision (and accuracy). In fact, inconsistencies during sampling can be normalized as the IS extraction will be influenced to the same extent as the analyte of interest.

Generally, to ensure quality in the analysis of exhaled breath samples by GC-MS, and therefore reliable findings, several key quality assurance and quality control measures are essential. These measures aim to guarantee the accuracy, reproducibility, and reliability of the entire analytical results and are as follows (Li et al., 2019; Becker, 2020; Westphal et al., 2023):

- Calibrants and reference materials: Calibrants and reference materials containing volatile marker compounds are used to establish detector stability, known detection limits, and instrument calibration. These materials should be available consistently during clinical trials and routine applications.
- Training and test samples: For the reproducible identification of specific odor patterns, the availability of appropriate sets of training and test samples is essential. These samples aid in electronic nose-based or canine-based identification methods.
- Standardization and harmonization: Standardization of breath sampling procedures is crucial to minimize inter-observer and intra-observer errors. Researchers involved in breath sampling should undergo certification to ensure uniform and accurate collection processes. Standardization also involves monitoring room air for potential VOC contamination.
- Instrument calibration: Regular instrument calibration using internal standards, such as stable isotope-labeled compounds, enhances QC/QA efforts. It helps in tracking instrument performance and ensuring the accuracy of results.
- Blank analysis: To identify and remove contaminants, blank analyses are essential. These blanks include air blanks, system blanks to identify instrument artifacts, and blanks to account for chemical backgrounds originating from sampling materials like Tedlar® bags.
- Spiking samples with internal standards of known concentrations and different retention times aids in data normalization and enhances data quality.
- Inter-laboratory comparisons: For diagnostic purposes, it is important to compare data obtained from different laboratories and methods. It helps assess the reproducibility and relevance of potential biomarkers.

In summary, ensuring quality in exhalation analysis by GC-MS involves a comprehensive approach that encompasses quality control, standardization, instrument calibration, and data management. All these measures must be implemented without any exception in clinical applications where breath analysis holds potential for disease diagnosis and monitoring.

5 Conclusion

This comprehensive review aims to investigate the potential of GC-MS analysis of VOCs in breath as biomarkers for severe pathologies, such as cancer, pulmonary diseases, and infectious diseases. Critical aspects of the workflow are thoroughly considered and discussed, encompassing the type of exhaled breath, collection devices, pre-concentration techniques, and analysis, as well as the experimental designs, statistical analysis, identification strategies, and proposed potential VOCs biomarkers.

Tedlar® bags and TD are by far the most extended for collection and pre-concentration, respectively. However, the choice of the type of breath sample was more diverse, spanning between mixed and alveolar breath, a critical consideration when aiming to accurately compare and establish levels of endogenous VOCs. Despite the wealth of studies, the conspicuous lack of standardization in the methodological approach and the scarce absolute quantitation of potential biomarkers delay their transference to clinics. Additionally, relatively small cohorts with only a limited model validation in an independent cohort, along with the lack of consensus in altered findings among different studies hindered the identification of a single pathology-specific VOC. A deeper understanding of the endogenous origin of VOCs is imperative to fully grasp the significance of each VOC in discriminating between healthy and pathological states.

Overall, this review underscores the substantial potential of VOCs as biomarkers in health and pathology. Nonetheless, to fully harness this potential, it is crucial to address the lack of standardization in methodological approaches, include larger and well-defined cohorts, and validate models in independent cohorts. As we delve deeper into the complexities of VOCs in exhaled breath, we are poised to advance personalized and non-invasive diagnostic strategies that can revolutionize the detection and management of the pathology, ultimately benefiting public health.

Author contributions

MB-F: investigation, writing–original draft, and writing–review and editing. ES-S: investigation, writing–original draft, and writing–review and editing. CB: funding acquisition and writing–review and editing. MR-S: conceptualization, methodology, supervision, and writing–review and editing. AG: conceptualization, funding acquisition, methodology, project administration, resources, supervision, visualization, and writing–review and editing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Ministry of Science and Innovation of Spain (MCIN) through MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe, under grant number PID 2021-122490NB-I00.

Acknowledgments

MB-F thanks Fundación Universitaria San Pablo CEU for her PhD fellowship. ES-S thanks European Union's Horizon 2020 Research and Innovation Programme for the postdoctoral fellowship received under the Marie Sklodowska-Curie grant agreement no. 101064457.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1295955/full#supplementary-material

References

Acharige, M. J. T., Koshy, S., Ismail, N., Aloum, O., Jazaerly, M., Astudillo, C. L., et al. (2018). Breath-based diagnosis of fungal infections. *J. Breath. Res.* 12, 027108. doi:10.1088/1752-7163/aa98a1

Ahmed, W. M., Fenn, D., White, I. R., Dixon, B., Nijsen, T. M. E., Knobel, H. H., et al. (2023). Microbial volatiles as diagnostic biomarkers of bacterial lung infection in mechanically ventilated patients. *Clin. Infect. Dis.* 76, 1059–1066. doi:10.1093/cid/ciac859

Alnimr, A. (2023). Antimicrobial resistance in ventilator-associated pneumonia: predictive microbiology and evidence-based therapy. *Infect. Dis. Ther.* 12, 1527–1552. doi:10.1007/s40121-023-00820-2

Alonso, M., and Sanchez, J. M. (2013). Analytical challenges in breath analysis and its application to exposure monitoring. *TrAC - Trends Anal. Chem.* 44, 78–89. doi:10.1016/j.trac.2012.11.011

Altomare, D. F., Di Lena, M., Porcelli, F., Travaglio, E., Longobardi, F., Tutino, M., et al. (2015). Effects of curative colorectal cancer surgery on exhaled volatile organic compounds and potential implications in clinical follow-up. *Ann. Surg.* 262, 862–866. doi:10.1097/SLA.0000000000001471

Altomare, D. F., Di Lena, M., Porcelli, F., Trizio, L., Travaglio, E., Tutino, M., et al. (2013). Exhaled volatile organic compounds identify patients with colorectal cancer. *Br. J. Surg.* 100, 144–150. doi:10.1002/bjs.8942

Altomare, D. F., Picciariello, A., Rotelli, M. T., De Fazio, M., Aresta, A., Zambonin, C. G., et al. (2020). Chemical signature of colorectal cancer: case-control study for profiling the breath print. *BJS Open* 4, 1189–1199. doi:10.1002/bjs5.50354

Amal, H., Leja, M., Broza, Y. Y., Tisch, U., Funka, K., Liepniece-Karele, I., et al. (2013). Geographical variation in the exhaled volatile organic compounds. *J. Breath. Res.* 7, 047102. doi:10.1088/1752-7155/7/4/047102

Amal, H., Leja, M., Funka, K., Lasina, I., Skapars, R., Sivins, A., et al. (2016). Breath testing as potential colorectal cancer screening tool. *Int. J. Cancer* 138, 229–236. doi:10.1002/ijc.29701

Amann, A., Costello, B. D. L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., et al. (2014). The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath. Res.* 8, 034001. doi:10.1088/1752-7155/8/3/034001

American Cancer Society (2017). Stomach (gastric) cancer survival rates. American Cancer Society. Available at: https://www.cancer.org/cancer/types/stomach-cancer/detection-diagnosis-staging/survival-rates.html (Accessed July 3, 2023).

Antoniou, S. X., Gaude, E., Ruparel, M., Van Der Schee, M. P., Janes, S. M., Rintoul, R. C., et al. (2019). The potential of breath analysis to improve outcome for patients with lung cancer. *J. Breath. Res.* 13, 034002. doi:10.1088/1752-7163/ab0bee

Aoki, T., Nagaoka, T., Kobayashi, N., Kurahashi, M., Tsuji, C., Takiguchi, H., et al. (2017). Editor's highlight: prospective analyses of volatile organic compounds in obstructive sleep apnea patients. *Toxicol. Sci.* 156, 362–374. doi:10.1093/toxsci/kfw260

Asthma (2023). World health organization. Available at: https://www.who.int/news-room/fact-sheets/detail/asthma (Accessed June 28, 2023).

Asthma-Diagnosis (2023). National heart, lung, and blood Institute. Available at: https://www.nhlbi.nih.gov/health/asthma/diagnosis (Accessed July 25, 2023).

Balata, H., Quaife, S. L., Craig, C., Ryan, D. J., Bradley, P., Crosbie, P. A. J., et al. (2022). Early diagnosis and lung cancer screening. *Clin. Oncol.* 34, 708–715. doi:10.1016/j.clon.2022.08.036

Barash, O., Zhang, W., Halpern, J. M., Hua, Q. L., Pan, Y. Y., Kayal, H., et al. (2015). Differentiation between genetic mutations of breast cancer by breath volatolomics. *Oncotarget* 6, 44864–44876. doi:10.18632/oncotarget.6269

Barba, D., León-Sosa, A., Lugo, P., Suquillo, D., Torres, F., Surre, F., et al. (2021). Breast cancer, screening and diagnostic tools: all you need to know. *Crit. Rev. Oncol. Hematol.* 157, 103174. doi:10.1016/j.critrevonc.2020.103174

Basanta, M., Ibrahim, B., Dockry, R., Douce, D., Morris, M., Singh, D., et al. (2012). Exhaled volatile organic compounds for phenotyping chronic obstructive pulmonary disease: a cross-sectional study. *Respir. Res.* 13, 72. doi:10.1186/1465-9921-13-72

Basanta, M., Koimtzis, T., Singh, D., Wilson, I., and Thomas, C. L. P. (2007). An adaptive breath sampler for use with human subjects with an impaired respiratory function. *Analyst* 132, 153–163. doi:10.1039/b608608J

Bayrakli, I., Öztürk, Ö., and Akman, H. (2016). Investigation of acetone, butanol and carbon dioxide as new breath biomarkers for convenient and noninvasive diagnosis of obstructive sleep apnea syndrome. *Biomed. Chromatogr.* 30, 1890–1899. doi:10.1002/bmc.3757

Beale, D. J., Jones, O. A. H., Karpe, A. V., Dayalan, S., Oh, D. Y., Kouremenos, K. A., et al. (2016). A review of analytical techniques and their application in disease diagnosis in breathomics and salivaomics research. *Int. J. Mol. Sci.* 18, 24. doi:10.3390/ijms18010024

Beauchamp, J., Herbig, J., Gutmann, R., and Hansel, A. (2008). On the use of Tedlar® bags for breath-gas sampling and analysis. *J. Breath. Res.* 2, 046001. doi:10.1088/1752-7155/2/4/046001

Beauchamp, J. D., and Miekisch, W. (2020). "Breath sampling and standardization," in *Breathborne biomarkers and the human volatilome*. Editors J. Beauchamp, C. Davis, and J. Pleil (Amsterdam, Netherlands: Elsevier), 23–41. doi:10.1016/B978-0-12-819967-1.00002-5

Becker, R. (2020). Non-invasive cancer detection using volatile biomarkers: is urine superior to breath? *Med. Hypotheses* 143, 110060. doi:10.1016/j.mehy.2020.110060

Behrendorff, J. B. Y. H. (2021). Reductive cytochrome P450 reactions and their potential role in bioremediation. *Front. Microbiol.* 12, 649273. doi:10.3389/fmicb.2021.649273

Berna, A. Z., Akaho, E. H., Harris, R. M., Congdon, M., Korn, E., Neher, S., et al. (2021). Reproducible breath metabolite changes in children with SARS-CoV-2 infection. *ACS Infect. Dis.* 7, 2596–2603. doi:10.1021/acsinfecdis.1c00248

Bhandari, M. P., Polaka, I., Vangravs, R., Mezmale, L., Veliks, V., Kirshners, A., et al. (2023). Volatile markers for cancer in exhaled breath—could they be the signature of the gut microbiota? *Molecules* 28, 3488. doi:10.3390/molecules28083488

Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., et al. (2007). Mammographic density and the risk and detection of breast cancer. *N. Engl. J. Med.* 356, 227–236. doi:10.1056/NEJMOA062790

Brinkman, P., van de Pol, M. A., Gerritsen, M. G., Bos, L. D., Dekker, T., Smids, B. S., et al. (2017). Exhaled breath profiles in the monitoring of loss of control and clinical recovery in asthma. *Clin. Exp. Allergy* 47, 1159–1169. doi:10.1111/cea.12965

Bruderer, T., Gaisl, T., Gaugg, M. T., Nowak, N., Streckenbach, B., Müller, S., et al. (2019). On-line analysis of exhaled breath focus review. *Chem. Rev.* 119, 10803–10828. doi:10.1021/acs.chemrev.9b00005

Buszewski, B., Grzywinski, D., Ligor, T., Stacewicz, T., Bielecki, Z., and Wojtas, J. (2013). Detection of volatile organic compounds as biomarkers in breath analysis by different analytical techniques. *Bioanalysis* 5, 2287–2306. doi:10.4155/bio.13.183

Buszewski, B., Kesy, M., Ligor, T., and Amann, A. (2007). Human exhaled air analytics: biomarkers of diseases. *Biomed. Chromatogr.* 21, 553–566. doi:10.1002/bmc.835

Buszewski, B., Ligor, T., Jezierski, T., Wenda-Piesik, A., Walczak, M., and Rudnicka, J. (2012). Identification of volatile lung cancer markers by gas chromatography-mass spectrometry: comparison with discrimination by canines. *Anal. Bioanal. Chem.* 404, 141–146. doi:10.1007/s00216-012-6102-8

Caldeira, M., Perestrelo, R., Barros, A. S., Bilelo, M. J., Morête, A., Câmara, J. S., et al. (2012). Allergic asthma exhaled breath metabolome: a challenge for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* 1254, 87–97. doi:10.1016/j.chroma.2012.07.023

Calenic, B., Miricescu, D., Greabu, M., Kuznetsov, A. V., Troppmair, J., Ruzsanyi, V., et al. (2015). Oxidative stress and volatile organic compounds: interplay in pulmonary, cardio-vascular, digestive tract systems and cancer. *Open Chem.* 13, 1020–1030. doi:10.1515/chem-2015-0105

Callol-Sanchez, L., Munoz-Lucas, M. A., Gomez-Martin, O., Maldonado-Sanz, J. A., Civera-Tejuca, C., Gutierrez-Ortega, C., et al. (2017). Observation of nonanoic acid and aldehydes in exhaled breath of patients with lung cancer. *J. Breath. Res.* 11, 026004. doi:10.1088/1752-7163/aa6485

Cancer (2023). World health organization. Available at: https://www.who.int/news-room/fact-sheets/detail/cancer (Accessed May 17, 2023).

Castillo, J. R., Peters, S. P., and Busse, W. W. (2017). Asthma exacerbations: pathogenesis, prevention, and treatment. *J. Allergy Clin. Immunol. Pract.* 5, 918–927. doi:10.1016/j.jaip.2017.05.001

Cazzola, M., Segreti, A., Capuano, R., Bergamini, A., Martinelli, E., Calzetta, L., et al. (2015). Analysis of exhaled breath fingerprints and volatile organic compounds in COPD. *COPD Res. Pract.* 1, 7. doi:10.1186/s40749-015-0010-1

Cen, Z., Lu, B., Ji, Y., Chen, J., Liu, Y., Jiang, J., et al. (2023). Virus-induced breath biomarkers: a new perspective to study the metabolic responses of COVID-19 vaccinees. *Talanta* 260, 124577. doi:10.1016/j.talanta.2023.124577

Chen, S., Mahadevan, V., and Zieve, L. (1970). Volatile fatty acids in the breath of patients with cirrhosis of the liver. *J. Lab. Clin. Med.* 75, 622–627. doi:10.5555/uri:pii:0022214370901605

Chen, X., Muhammad, K. G., Madeeha, C., Fu, W., Xu, L., Hu, Y., et al. (2021). Calculated indices of volatile organic compounds (VOCs) in exhalation for lung cancer screening and early detection. *Lung Cancer* 154, 197–205. doi:10.1016/j.lungcan.2021.02.006

Colorectal Cancer—Cancer Stat Facts (2023). National cancer Institute. Available at: https://seer.cancer.gov/statfacts/html/colorect.html (Accessed May 23, 2023).

Conti, C. B., Agnesi, S., Scaravaglio, M., Masseria, P., Dinelli, M. E., Oldani, M., et al. (2023). Early gastric cancer: update on prevention, diagnosis and treatment. *Int. J. Environ. Res. Public Health* 20, 2149. doi:10.3390/ijerph20032149

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi:10.1093/bioinformatics/btx364

Cope, K. A., Watson, M. T., Foster, W. M., Sehnert, S. S., and Risby, T. H. (2004). Effects of ventilation on the collection of exhaled breath in humans. *J. Appl. Physiol. (1985)* 96, 1371–1379. doi:10.1152/japplphysiol.01034.2003

Corradi, M., Poli, D., Banda, I., Bonini, S., Mozzoni, P., Pinelli, S., et al. (2015). Exhaled breath analysis in suspected cases of non-small-cell lung cancer: a cross-sectional study. *J. Breath. Res.* 9, 027101. doi:10.1088/1752-7155/9/2/027101

Cystic Fibrosis-Causes (2023). National heart, lung, and blood Institute. Available at: https://www.nhlbi.nih.gov/health/cystic-fibrosis/causes (Accessed June 6, 2023).

Davidson, L. S. P. (1949). Mercaptan in the breath of patients with severe liver disease. *Lancet* 2, 197. doi:10.1016/S0140-6736(49)91197-6

De Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., et al. (2014). A review of the volatiles from the healthy human body. *J. Breath. Res.* 8, 014001. doi:10.1088/1752-7155/8/1/014001

De Vos, W. M., Tilg, H., Van Hul, M., and Cani, P. D. (2022). Gut microbiome and health: mechanistic insights. *Gut* 71, 1020–1032. doi:10.1136/gutjnl-2021-326789

Drabińska, N., Flynn, C., Ratcliffe, N., Belluomo, I., Myridakis, A., Gould, O., et al. (2021). A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *J. Breath. Res.* 15, 034001. doi:10.1088/1752-7163/abf1d0

Dudzik, D., Barbas-Bernardos, C., García, A., and Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *J. Pharm. Biomed. Anal.* 147, 149–173. doi:10.1016/J.JPBA.2017.07.044

Eriksen, S. P., and Kulkarni, A. B. (1963). Methanol in normal human breath. *Science* 141, 639–640. doi:10.1126/science.141.3581.639

Fazleen, A., and Wilkinson, T. (2020). Early COPD: current evidence for diagnosis and management. *Ther. Adv. Respir. Dis.* 14, 1753466620942128. doi:10.1177/1753466620942128

Female Breast Cancer (2023). Cancer Stat facts. National cancer Institute. Available at: https://seer.cancer.gov/statfacts/html/breast.html (Accessed May 26, 2023).

Feng, F., Tian, Y., Xu, G., Liu, Z., Liu, S., Zheng, G., et al. (2017). Diagnostic and prognostic value of CEA, CA19-9, AFP and CA125 for early gastric cancer. *BMC Cancer* 17, 737. doi:10.1186/s12885-017-3738-y

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., et al. (2021). Cancer statistics for the year 2020: an overview. *Int. J. Cancer* 149, 778–789. doi:10.1002/ijc.33588

Fernando, S. M., Tran, A., Cheng, W., Klompas, M., Kyeremanteng, K., Mehta, S., et al. (2020). Diagnosis of ventilator-associated pneumonia in critically ill adult patients-A systematic review and meta-analysis. *Intensive Care Med.* 46, 1170–1179. doi:10.1007/s00134-020-06036-z

Ferreira-Coimbra, J., Sarda, C., and Rello, J. (2020). Burden of community-acquired pneumonia and unmet clinical needs. *Adv. Ther.* 37, 1302–1318. doi:10.1007/s12325-020-01248-7

Filipiak, W., Beer, R., Sponring, A., Filipiak, A., Ager, C., Schiefecker, A., et al. (2015). Breath analysis for *in vivo* detection of pathogens related to ventilator-associated pneumonia in intensive care patients: a prospective pilot study. *J. Breath. Res.* 9, 016004. doi:10.1088/1752-7155/9/1/016004

Filipiak, W., Sponring, A., Filipiak, A., Baur, M., Ager, C., Wiesenhofer, H., et al. (2013). "Volatile organic compounds (VOCs) released by pathogenic microorganisms *in vitro*: potential breath biomarkers for early-stage diagnosis of disease," in *Volatile biomarkers*. Editors A. Amann and D. Smith (Amsterdam, Netherlands: Elsevier), 463–512. doi:10.1016/B978-0-44-462613-4.00023-4

Fortes, P. R., Petruci, J. F. S., and Raimundo, I. M., Jr. (2017). *Optical gas sensors for exhaled breath analysis*. SPIE Press. doi:10.1117/3.2284712

Fowler, S. J., Basanta-Sanchez, M., Xu, Y., Goodacre, R., and Dark, P. M. (2015). Surveillance for lower airway pathogens in mechanically ventilated patients by metabolomic analysis of exhaled breath: a case-control study. *Thorax* 70, 320–325. doi:10.1136/thoraxjnl-2014-206273

Furge, L. L., and Guengerich, F. P. (2006). Cytochrome P450 enzymes in drug metabolism and chemical toxicology: an introduction. *Biochem. Mol. Biol. Educ.* 34, 66–74. doi:10.1002/bmb.2006.49403402066

Gahleitner, F., Guallar-Hoyas, C., Beardsmore, C. S., Pandya, H. C., and Thomas, C. P. (2013). Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis* 5, 2239–2247. doi:10.4155/bio.13.184

Gaida, A., Holz, O., Nell, C., Schuchardt, S., Lavae-Mokhtari, B., Kruse, L., et al. (2016). A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD. *J. Breath. Res.* 10, 026006. doi:10.1088/1752-7155/10/2/026006

Gao, J., Zou, Y., Wang, Y., Wang, F., Lang, L., Wang, P., et al. (2016). Breath analysis for noninvasively differentiating Acinetobacter baumannii ventilator-Associated pneumonia from its respiratory tract colonization of ventilated patients. *J. Breath. Res.* 10, 027102. doi:10.1088/1752-7155/10/2/027102

Ghosh, C., Singh, V., Grandy, J., and Pawliszyn, J. (2020). Recent advances in breath analysis to track human health by new enrichment technologies. *J. Sep. Sci.* 43, 226–240. doi:10.1002/jssc.201900769

GOLDCOPD (2023). Global initiative for chronic obstructive lung disease. Available at: www.goldcopd.org.

González-Domínguez, R., Jáuregui, O., Queipo-Ortuño, M. I., and Andrés-Lacueva, C. (2020). Characterization of the human exposome by a comprehensive and quantitative large-scale multianalyte metabolomics platform. *Anal. Chem.* 92, 13767–13775. doi:10.1021/acs.analchem.0c02008

Goss, C. H. (2019). Acute pulmonary exacerbations in cystic fibrosis. *Semin. Respir. Crit. Care Med.* 40, 792–803. doi:10.1055/s-0039-1697975

Grote, C., and Pawliszyn, J. (1997). Solid-phase microextraction for the analysis of human breath. *Anal. Chem.* 69, 587–596. doi:10.1021/ac960749l

Haick, H., Broza, Y. Y., Mochalski, P., Ruzsanyi, V., and Amann, A. (2014). Assessment, origin, and implementation of breath volatile cancer markers. *Chem. Soc. Rev.* 43, 1423–1449. doi:10.1039/c3cs60329f

Hakim, M., Broza, Y. Y., Barash, O., Peled, N., Phillips, M., Amann, A., et al. (2012). Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem. Rev.* 112, 5949–5966. doi:10.1021/cr300174a

Hamashima, C. (2016). Benefits and harms of endoscopic screening for gastric cancer. *World J. Gastroenterol.* 22, 6385–6392. doi:10.3748/wjg.v22.i28.6385

Hanna, G. B., Boshier, P. R., Markar, S. R., and Romano, A. (2019). Accuracy and methodologic challenges of volatile organic compound–based exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. *JAMA Oncol.* 5, e182815. doi:10.1001/jamaoncol.2018.2815

Helsingen, L. M., and Kalager, M. (2022). Colorectal cancer screening — approach, evidence, and future directions. *NEJM Evid.* 1. doi:10.1056/EVIDra2100035

Herrera-Pariente, C., Montori, S., Llach, J., Bofill, A., Albeniz, E., and Moreira, L. (2021). Biomarkers for gastric cancer screening and early diagnosis. *Biomedicines* 9, 1448. doi:10.3390/biomedicines9101448

Hunt, R., and Williams, D. (1977). Spectrometric measurement of ammonia in normal human breath. *Int. Lab.* 11.

Ibrahim, W., Cordell, R. L., Wilde, M. J., Richardson, M., Carr, L., Dasi, A. S. D., et al. (2021). Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry. *ERJ Open Res.* 7, 00139–02021. doi:10.1183/23120541.00139-2021

Janssens, E., van Meerbeeck, J. P., and Lamote, K. (2020). Volatile organic compounds in human matrices as lung cancer biomarkers: a systematic review. *Crit. Rev. Oncol. Hematol.* 153, 103037. doi:10.1016/j.critrevonc.2020.103037

Jareño-Esteban, J. J., Muñoz-Lucas, M. Á., Gómez-Martín, Ó., Utrilla-Trigo, S., Gutiérrez-Ortega, C., Aguilar-Ros, A., et al. (2017). Study of 5 volatile organic compounds in exhaled breath in chronic obstructive pulmonary disease. *Arch. Bronconeumol* 53, 251–256. doi:10.1016/j.arbres.2016.09.003

Kapur, V. K., Auckley, D. H., Chowdhuri, S., Kuhlmann, D. C., Mehra, R., Ramar, K., et al. (2017). Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an american academy of sleep medicine clinical practice guideline. *J. Clin. Sleep. Med.* 13, 479–504. doi:10.5664/jcsm.6506

Karamanou, M., and Androutsos, G. (2013). Antoine-Laurent de Lavoisier (1743-1794) and the birth of respiratory physiology. *Thorax* 68, 978–979. doi:10.1136/thoraxjnl-2013-203840

Khoubnasabjafari, M., Mogaddam, M. R. A., Rahimpour, E., Soleymani, J., Saei, A. A., and Jouyban, A. (2022). Breathomics: review of sample collection and analysis, data modeling and clinical applications. *Crit. Rev. Anal. Chem.* 52, 1461–1487. doi:10.1080/10408347.2021.1889961

Koureas, M., Kalompatsios, D., Amoutzias, G. D., Hadjichristodoulou, C., Gourgoulianis, K., and Tsakalof, A. (2021). 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. doi:10.3390/molecules26092609

Koureas, M., Kirgou, P., Amoutzias, G., Hadjichristodoulou, C., Gourgoulianis, K., and Tsakalof, A. (2020). Target analysis of volatile organic compounds in exhaled breath for lung cancer discrimination from other pulmonary diseases and healthy persons. *Metabolites* 10, 317. doi:10.3390/metabo10080317

Kucirka, L. M., Lauer, S. A., Laeyendecker, O., Boon, D., and Lessler, J. (2020). Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure. *Ann. Intern Med.* 173, 262–267. doi:10.7326/M20-1495

Kuo, T. C., Tan, C. E., Wang, S. Y., Lin, O. A., Su, B. H., Hsu, M. T., et al. (2020). Human breathomics database. *Database (Oxford)* 2020, baz139. doi:10.1093/database/baz139

Kuruvilla, M. E., Lee, F. E. H., and Lee, G. B. (2019). Understanding asthma phenotypes, endotypes, and mechanisms of disease. *Clin. Rev. Allergy Immunol.* 56, 219–233. doi:10.1007/s12016-018-8712-1

Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R., and Fowler, S. J. (2017). Exhaled breath analysis: a review of "breath-taking" methods for off-line analysis. *Metabolomics* 13, 110. doi:10.1007/S11306-017-1241-8

Lévy, P., Kohler, M., McNicholas, W. T., Barbé, F., McEvoy, R. D., Somers, V. K., et al. (2015). Obstructive sleep apnoea syndrome. *Nat. Rev. Dis. Prim.* 1, 15015. doi:10.1038/nrdp.2015.15

Li, J., Peng, Y., Liu, Y., Li, W., Jin, Y., Tang, Z., et al. (2014). Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin. Chim. Acta* 436, 59–67. doi:10.1016/j.cca.2014.04.030

Li, W., Dai, W., Liu, M., Long, Y., Wang, C., Xie, S., et al. (2019). VOC biomarkers identification and predictive model construction for lung cancer based on exhaled breath analysis: research protocol for an exploratory study. *BMJ Open* 9, e028448. doi:10.1136/BMJOPEN-2018-028448

Ligor, T., Pater, Ł., and Buszewski, B. (2015). Application of an artificial neural network model for selection of potential lung cancer biomarkers. *J. Breath. Res.* 9, 027106. doi:10.1088/1752-7155/9/2/027106

Løberg, M., Lousdal, M. L., Bretthauer, M., and Kalager, M. (2015). Benefits and harms of mammography screening. *Breast Cancer Res.* 17, 63. doi:10.1186/S13058-015-0525-Z

Lung Cancer Survival Rates (2023). 5-Year survival rates for lung cancer. American Cancer Society. Available at: https://www.cancer.org/cancer/types/lung-cancer/detection-diagnosis-staging/survival-rates.html (Accessed June 22, 2023).

Lv, R., Liu, X., Zhang, Y., Dong, N., Wang, X., He, Y., et al. (2023). Pathophysiological mechanisms and therapeutic approaches in obstructive sleep apnea syndrome. *Signal Transduct. Target Ther.* 8, 218. doi:10.1038/s41392-023-01496-3

Mansurova, M., Ebert, B. E., Blank, L. M., and Ibáñez, A. J. (2018). A breath of information: the volatilome. *Curr. Genet.* 64, 959–964. doi:10.1007/s00294-017-0800-x

Markar, S. R., Wiggins, T., Kumar, S., and Hanna, G. B. (2015). Exhaled breath analysis for the diagnosis and assessment of endoluminal gastrointestinal diseases. *J. Clin. Gastroenterol.* 49, 1–8. doi:10.1097/MCG.0000000000000247

McKee, H. C., Rhoades, J. W., Campbell, J., and Gross, A. L. (1962). Acetonitrile in body fluids related to smoking. *Public Health Rep.* 77, 553–554. doi:10.2307/4591551

Meyer, N., Dallinga, J. W., Nuss, S. J., Moonen, E. J. C., van Berkel, J. J. B. N., Akdis, C., et al. (2014). Defining adult asthma endotypes by clinical features and patterns of volatile organic compounds in exhaled air. *Respir. Res.* 15, 136. doi:10.1186/S12931-014-0136-8

Miekisch, W., Herbig, J., and Schubert, J. K. (2012). Data interpretation in breath biomarker research: pitfalls and directions. *J. Breath. Res.* 6, 036007. doi:10.1088/1752-7155/6/3/036007

Miekisch, W., Kischkel, S., Sawacki, A., Liebau, T., Mieth, M., and Schubert, J. K. (2008). Impact of sampling procedures on the results of breath analysis. *J. Breath. Res.* 2, 026007. doi:10.1088/1752-7155/2/2/026007

Mochalski, P., Wzorek, B., Śliwka, I., and Amann, A. (2009). Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877, 189–196. doi:10.1016/j.jchromb.2008.12.003

Modi, A. R., and Kovacs, C. S. (2020). Hospital-acquired and ventilator-associated pneumonia: diagnosis, management, and prevention. *Cleve Clin. J. Med.* 87, 633–639. doi:10.3949/ccjm.87a.19117

Monedeiro, F., Monedeiro-Milanowski, M., Ratiu, I. A., Brożek, B., Ligor, T., and Buszewski, B. (2021). Needle trap device-GC-MS for characterization of lung diseases based on breath VOC profiles. *Molecules* 26, 1789. doi:10.3390/molecules26061789

Morales, J. S., Valenzuela, P. L., Castillo-García, A., Butragueño, J., Jiménez-Pavón, D., Carrera-Bastos, P., et al. (2022). The exposome and immune health in times of the covid-19 pandemic. *Nutrients* 14, 24. doi:10.3390/nu14010024

Muñoz-Lucas, M. Á., Jareño-Esteban, J., Gutiérrez-Ortega, C., López-Guijarro, P., Collado-Yurrita, L., Quintana-Díaz, M., et al. (2020). Influence of chronic obstructive pulmonary disease on volatile organic compounds in patients with non-small cell lung cancer. *Arch. Bronconeumol* 56, 801–805. doi:10.1016/j.arbr.2020.10.004

Munro, S. C., Baker, Di., Giuliano, K. K., Sullivan, S. C., Haber, J., Jones, B. E., et al. (2021). Nonventilator hospital-acquired pneumonia: a call to action. *Infect. Control Hosp. Epidemiol.* 42, 991–996. doi:10.1017/ice.2021.239

Murray, R., Rodwell, V., Bender, D., Botham, K. M., Weil, P., Anthony, et al. (2009). *Harper's illustrated biochemistry*. 28th ed New York: Citiseer.

Myers, R., Ruszkiewicz, D. M., Meister, A., Bartolomeu, C., Atkar-Khattra, S., Thomas, C. L. P., et al. (2023). Breath testing for SARS-CoV-2 infection. *EBioMedicine* 92, 104584. doi:10.1016/j.ebiom.2023.104584

Nardi-Agmon, I., Abud-Hawa, M., Liran, O., Gai-Mor, N., Ilouze, M., Onn, A., et al. (2016). Exhaled breath analysis for monitoring response to treatment in advanced lung cancer. *J. Thorac. Oncol.* 11, 827–837. doi:10.1016/j.jtho.2016.02.017

Netzker, T., Shepherdson, E. M. F., Zambri, M. P., and Elliot, M. A. (2020). Bacterial volatile compounds: functions in communication, cooperation, and competition. *Annu. Rev. Microbiol.* 74, 409–430. doi:10.1146/annurev-micro-011320-015542

Neveu, V., Nicolas, G., Amara, A., Salek, R. M., and Scalbert, A. (2023). The human microbial exposome: expanding the Exposome-Explorer database with gut microbial metabolites. *Sci. Rep.* 13, 1946. doi:10.1038/s41598-022-26366-w

Niedermaier, T., Tikk, K., Gies, A., Bieck, S., and Brenner, H. (2020). Sensitivity of fecal immunochemical test for colorectal cancer detection differs according to stage and location. *Clin. Gastroenterol. Hepatol.* 18, 2920–2928. doi:10.1016/j.cgh.2020.01.025

Nooreldeen, R., and Bach, H. (2021). Current and future development in lung cancer diagnosis. *Int. J. Mol. Sci.* 22, 8661. doi:10.3390/ijms22168661

Nowak, N., Engler, A., Thiel, S., Stöberl, A. S., Sinues, P., Zenobi, R., et al. (2021). Validation of breath biomarkers for obstructive sleep apnea. *Sleep. Med.* 85, 75–86. doi:10.1016/j.sleep.2021.06.040

Papaefstathiou, E., Stylianou, M., Andreou, C., and Agapiou, A. (2020). Breath analysis of smokers, non-smokers, and e-cigarette users. *J. Chromatogr. B* 1160, 122349. doi:10.1016/j.jchromb.2020.122349

Patti, G. J., Yanes, O., and Siuzdak, G. (2012). Innovation: metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi:10.1038/NRM3314

Pauling, L., Robinson, A. B., Teranishit, R., and Cary, P. (1971). Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. U. S. A.* 68, 2374–2376. doi:10.1073/pnas.68.10.2374

Pennazza, G., Santonico, M., Incalzi, R. A., Scarlata, S., Chiurco, D., Vernile, C., et al. (2014). Measure chain for exhaled breath collection and analysis: a novel approach suitable for frail respiratory patients. *Sens. Actuators B Chem.* 204, 578–587. doi:10.1016/j.snb.2014.08.007

Pesesse, R., Stefanuto, P. H., Schleich, F., Louis, R., and Focant, J. F. (2019). Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC × GC-TOFMS. *J. Chromatogr. B* 1114–1115, 146–153. doi:10.1016/j.jchromb.2019.01.029

Phillips, C. O., Syed, Y., Parthaláin, N., Zwiggelaar, R. M., Claypole, T. C., and Lewis, K. E. (2012). Machine learning methods on exhaled volatile organic compounds for distinguishing COPD patients from healthy controls. *J. Breath. Res.* 6, 036003. doi:10.1088/1752-7155/6/3/036003

Phillips, M. (1992). Breath tests in medicine. *Sci. Am.* 267, 74–79. doi:10.1038/scientificamerican0792-74

Phillips, M. (1997). Method for the collection and assay of volatile organic compounds in breath. *Anal. Biochem.* 247, 272–278. doi:10.1006/abio.1997.2069

Pizzini, A., Filipiak, W., Wille, J., Ager, C., Wiesenhofer, H., Kubinec, R., et al. (2018). 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. doi:10.1088/1752-7163/aaa4c5

Pleil, J. D., Stiegel, M. A., and Risby, T. H. (2013). Clinical breath analysis: discriminating between human endogenous compounds and exogenous (environmental) chemical confounders. *J. Breath. Res.* 7, 017107. doi:10.1088/1752-7155/7/1/017107

Qaseem, A., Crandall, C. J., Mustafa, R. A., Hicks, L. A., Wilt, T. J., and Clinical Guidelines Committee of the American College of Physicians (2019). Screening for colorectal cancer in asymptomatic average-risk adults: a guidance statement from the American College of Physicians. *Ann. Intern Med.* 171, 643–654. doi:10.7326/M19-0642

Ratcliffe, N., Wieczorek, T., Drabińska, N., Drabińska, N., Gould, O., Osborne, A., et al. (2020). 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. doi:10.1088/1752-7163/ab7f9d

Ratiu, I. A., Ligor, T., Bocos-Bintintan, V., Mayhew, C. A., and Buszewski, B. (2020). Volatile organic compounds in exhaled breath as fingerprints of lung cancer, asthma and COPD. *J. Clin. Med.* 10, 32. doi:10.3390/JCM10010032

Rey-Stolle, F., Dudzik, D., Gonzalez-Riano, C., Fernández-García, M., Alonso-Herranz, V., Rojo, D., et al. (2021). Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: a tutorial. *Anal. Chim. Acta* 1210, 339043. doi:10.1016/j.aca.2021.339043

Robroeks, C. M., Van Berkel, J. J., Jöbsis, Q., Van Schooten, F. J., Dallinga, J. W., Wouters, E. F., et al. (2013). Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study. *Eur. Respir. J.* 42, 98–106. doi:10.1183/09031936.00010712

Rodak, O., Peris-Díaz, M. D., Olbromski, M., Podhorska-Okołów, M., and Dzięgiel, P. (2021). Current landscape of non-small cell lung cancer: epidemiology, histological classification, targeted therapies, and immunotherapy. *Cancers (Basel)* 13, 4705. doi:10.3390/cancers13184705

Roesch, E. A., Nichols, D. P., and Chmiel, J. F. (2018). Inflammation in cystic fibrosis: an update. *Pediatr. Pulmonol.* 53, S30–S50. doi:10.1002/ppul.24129

Rudnicka, J., Kowalkowski, T., and Buszewski, B. (2019). Searching for selected VOCs in human breath samples as potential markers of lung cancer. *Lung Cancer* 135, 123–129. doi:10.1016/j.lungcan.2019.02.012

Rudnicka, J., Walczak, M., Kowalkowski, T., Jezierski, T., and Buszewski, B. (2014). Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography-mass spectrometry versus trained dogs. *Sens. Actuators B Chem.* 202, 615–621. doi:10.1016/j.snb.2014.06.006

Saasa, V., Malwela, T., Beukes, M., Mokgotho, M., Liu, C. P., and Mwakikunga, B. (2018). Sensing technologies for detection of acetone in human breath for diabetes diagnosis and monitoring. *Diagn. (Basel)* 8, 12. doi:10.3390/diagnostics8010012

Saidi, T., Moufid, M., de Jesus Beleño-Saenz, K., Welearegay, T. G., El Bari, N., Lisset Jaimes-Mogollon, A., et al. (2020). Non-invasive prediction of lung cancer histological types through exhaled breath analysis by UV-irradiated electronic nose and GC/QTOF/MS. *Sens. Actuators B Chem.* 311, 127932. doi:10.1016/j.snb.2020.127932

Sakumura, Y., Koyama, Y., Tokutake, H., Hida, T., Sato, K., Itoh, T., et al. (2017). Diagnosis by volatile organic compounds in exhaled breath from lung cancer patients using support vector machine algorithm. *Sensors (Basel)* 17, 287. doi:10.3390/s17020287

Schallschmidt, K., Becker, R., Jung, C., Bremser, W., Walles, T., Neudecker, J., et al. (2016). Comparison of volatile organic compounds from lung cancer patients and healthy controls - challenges and limitations of an observational study. *J. Breath. Res.* 10, 046007. doi:10.1088/1752-7155/10/4/046007

Schleich, F. N., Zanella, D., Stefanuto, P. H., Bessonov, K., Smolinska, A., Dallinga, J. W., et al. (2019). Exhaled volatile organic compounds are able to discriminate between neutrophilic and eosinophilic asthma. *Am. J. Respir. Crit. Care Med.* 200, 444–453. doi:10.1164/rccm.201811-2210OC

Schnabel, R., Fijten, R., Smolinska, A., Dallinga, J., Boumans, M. L., Stobberingh, E., et al. (2015). Analysis of volatile organic compounds in exhaled breath to diagnose ventilator-associated pneumonia. *Sci. Rep.* 5, 17179. doi:10.1038/srep17179

Schulz, E., Woollam, M., Grocki, P., Davis, M. D., and Agarwal, M. (2023). Methods to detect volatile organic compounds for breath biopsy using solid-phase microextraction and gas chromatography–mass spectrometry. *Molecules* 28, 4533. doi:10.3390/molecules28114533

Schwarz, E. I., Engler, A., and Kohler, M. (2017). Exhaled breath analysis in obstructive sleep apnea. *Expert Rev. Respir. Med.* 11, 631–639. doi:10.1080/17476348.2017.1338950

Scohy, A., Anantharajah, A., Bodéus, M., Kabamba-Mukadi, B., Verroken, A., and Rodriguez-Villalobos, H. (2020). Low performance of rapid antigen detection test as frontline testing for COVID-19 diagnosis. *J. Clin. Virol.* 129, 104455. doi:10.1016/j.jcv.2020.104455

Sethi, S., Nanda, R., and Chakraborty, T. (2013). Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin. Microbiol. Rev.* 26, 462–475. doi:10.1128/CMR.00020-13

Sharma, A., Kumar, R., and Varadwaj, P. (2023). Smelling the disease: diagnostic potential of breath analysis. *Mol. Diagn Ther.* 27, 321–347. doi:10.1007/s40291-023-00640-7

Sidorova, D. E., Plyuta, V. A., Padiy, D. A., Kupriyanova, E. V., Roshina, N. V., Koksharova, O. A., et al. (2021). The effect of volatile organic compounds on different organisms: agrobacteria, plants and insects. *Microorganisms* 10, 69. doi:10.3390/microorganisms10010069

Simenhoff, M. L., Burke, J. F., Saukkonen, J. J., Ordinario, A. T., Doty, R., and Dunn, S. (1977). Biochemical profile or uremic breath. *N. Engl. J. Med.* 297, 132–135. doi:10.1056/NEJM197707212970303

Simrén, M., and Stotzer, P. O. (2006). Use and abuse of hydrogen breath tests. *Gut* 55, 297–303. doi:10.1136/gut.2005.075127

Smolinska, A., Klaassen, E. M. M., Dallinga, J. W., Van De Kant, K. D. G., Jobsis, Q., Moonen, E. J. C., et al. (2014). Profiling of volatile organic compounds in exhaled breath as a strategy to find early predictive signatures of asthma in children. *PLoS One* 9, e95668. doi:10.1371/journal.pone.0095668

Sola-Martínez, R. A., Lozano-Terol, G., Gallego-Jara, J., Cánovas Díaz, M., and de Diego Puente, T. (2022). "Offline breath analysis: standardization of breath sampling and analysis using mass spectrometry and innovative algorithms," in *Breath analysis. Bioanalytical reviews.* Editor S. Weigl (Cham: Springer), 19–44. doi:10.1007/11663_2022_21

Sola-Martínez, R. A., Lozano-Terol, G., Gallego-Jara, J., Morales, E., Cantero-Cano, E., Sánchez-Solís de Querol, M., et al. (2021). Exhaled volatilome analysis as a useful tool to discriminate asthma with other coexisting atopic diseases in women of childbearing age. *Sci. Rep.* 11, 13823. doi:10.1038/S41598-021-92933-2

Stöberl, A. S., Schwarz, E. I., Haile, S. R., Turnbull, C. D., Rossi, V. A., Stradling, J. R., et al. (2017). Night-to-night variability of obstructive sleep apnea. *J. Sleep. Res.* 26, 782–788. doi:10.1111/jsr.12558

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* 3, 211–221. doi:10.1007/S11306-007-0082-2

Tang, Z., Liu, Y., and Duan, Y. (2015). Breath analysis: technical developments and challenges in the monitoring of human exposure to volatile organic compounds. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1002, 285–299. doi:10.1016/j.jchromb.2015.08.041

The Global Asthma Report (2022). The global asthma Report 2022. *Int. J. Tuberc. Lung Dis.* 26, S1–S102. doi:10.5588/ijtld.22.1010

Tong, H., Wang, Y., Li, Y., Liu, S., Chi, C., Liu, D., et al. (2017). Volatile organic metabolites identify patients with gastric carcinoma, gastric ulcer, or gastritis and control patients. *Cancer Cell Int.* 17, 108. doi:10.1186/s12935-017-0475-x

Trefz, P., Kischkel, S., Hein, D., James, E. S., Schubert, J. K., and Miekisch, W. (2012). Needle trap micro-extraction for VOC analysis: effects of packing materials and desorption parameters. *J. Chromatogr. A* 1219, 29–38. doi:10.1016/j.chroma.2011.10.077

Trujillo-Rodríguez, M. J., Pacheco-Fernández, I., Taima-Mancera, I., Díaz, J. H. A., and Pino, V. (2020). Evolution and current advances in sorbent-based microextraction configurations. *J. Chromatogr. A* 1634, 461670. doi:10.1016/j.chroma.2020.461670

Van Der Schee, M. P., Paff, T., Brinkman, P., Van Aalderen, W. M. C., Haarman, E. G., and Sterk, P. J. (2015). Breathomics in lung disease. *Chest* 147, 224–231. doi:10.1378/chest.14-0781

van Horck, M., Smolinska, A., Wesseling, G., de Winter-De Groot, K., de Vreede, I., Winkens, B., et al. (2021). Exhaled volatile organic compounds detect pulmonary exacerbations early in children with cystic fibrosis: results of a 1 year observational pilot study. *J. Breath. Res.* 15, 026012. doi:10.1088/1752-7163/abda55

Van Oort, P. M. P., De Bruin, S., Hans, W., Knobel, H. H., Schultz, M. J., Bos, L. D., et al. (2017a). Exhaled breath metabolomics for the diagnosis of pneumonia in intubated and mechanically-ventilated intensive care unit (ICU)-patients. *Int. J. Mol. Sci.* 18, 449. doi:10.3390/ijms18020449

van Oort, P. M. P., Nijsen, T., Weda, H., Knobel, H., Dark, P., Felton, T., et al. (2017b). BreathDx - molecular analysis of exhaled breath as a diagnostic test for ventilator-associated pneumonia: protocol for a European multicentre observational study. *BMC Pulm. Med.* 17, 1. doi:10.1186/s12890-016-0353-7

Van Oort, P. M. P., Nijsen, T. M., White, I. R., Knobel, H. H., Felton, T., Rattray, N., et al. (2022). Untargeted molecular analysis of exhaled breath as a diagnostic test for ventilator-associated lower respiratory tract infections (BreathDx). *Thorax* 77, 79–81. doi:10.1136/thoraxjnl-2021-217362

van Velzen, P., Brinkman, P., Knobel, H. H., van den Berg, J. W. K., Jonkers, R. E., Loijmans, R. J., et al. (2019). Exhaled breath profiles before, during and after exacerbation of COPD: a prospective follow-up study. *COPD* 16, 330–337. doi:10.1080/15412555.2019.1669550

Van Vliet, D., Smolinska, A., Jöbsis, Q., Rosias, P., Muris, J., Dallinga, J., et al. (2017). Can exhaled volatile organic compounds predict asthma exacerbations in children? *J. Breath. Res.* 11, 016016. doi:10.1088/1752-7163/aa5a8b

Van Vliet, D., Smolinska, A., Jöbsis, Q., Rosias, P. P. R., Muris, J. W. M., Dallinga, J. W., et al. (2016). Association between exhaled inflammatory markers and asthma control in children. *J. Breath. Res.* 10, 016014. doi:10.1088/1752-7155/10/1/016014

Vas, G., and Vékey, K. (2004). Solid-phase microextraction: a powerful sample preparation tool prior to mass spectrometric analysis. *J. Mass Spectrom.* 39, 233–254. doi:10.1002/jms.606

Vaz, A. D. N., and Coon, M. J. (1987). Hydrocarbon formation in the reductive cleavage of hydroperoxides by cytochrome P-450. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1172–1176. doi:10.1073/pnas.84.5.1172

Veith, A., and Moorthy, B. (2018). Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Curr. Opin. Toxicol.* 7, 44–51. doi:10.1016/j.cotox.2017.10.003

Vineis, P., Robinson, O., Chadeau-Hyam, M., Dehghan, A., Mudway, I., and Dagnino, S. (2020). What is new in the exposome? *Environ. Int.* 143, 105887. doi:10.1016/j.envint.2020.105887

Wang, C., Ke, C., Wang, X., Chi, C., Guo, L., Luo, S., et al. (2014a). Noninvasive detection of colorectal cancer by analysis of exhaled breath. *Anal. Bioanal. Chem.* 406, 4757–4763. doi:10.1007/s00216-014-7865-x

Wang, C., Sun, B., Guo, L., Wang, X., Ke, C., Liu, S., et al. (2014b). Volatile organic metabolites identify patients with breast cancer, cyclomastopathy, and mammary gland fibroma. *Sci. Rep.* 4, 5383. doi:10.1038/srep05383

Wang, Y., Hu, Y., Wang, D., Yu, K., Wang, L., Zou, Y., et al. (2012). The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomarkers* 11, 129–137. doi:10.3233/CBM-2012-00270

Westphal, K., Dudzik, D., Waszczuk-Jankowska, M., Graff, B., Narkiewicz, K., and Markuszewski, M. J. (2023). Common strategies and factors affecting off-line breath sampling and volatile organic compounds analysis using thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). *Metabolites* 13, 8. doi:10.3390/metabo13010008

White, I. R., and Fowler, S. J. (2019). "Capturing and storing exhaled breath for offline analysis," in *Breath analysis*. Editors G. Pennazza and M. Santonico (Amsterdam, Netherlands: Elsevier), 13–31. doi:10.1016/B978-0-12-814562-3.00002-3

WHO Coronavirus (COVID-19) Dashboard (2023). World health organization. Available at: https://covid19.who.int/ (Accessed June 13, 2023).

Wilde, M. J., Cordell, R. L., Salman, D., Zhao, B., Ibrahim, W., Bryant, L., et al. (2019). Breath analysis by two-dimensional gas chromatography with dual flame ionisation and mass spectrometric detection – method optimisation and integration within a large-scale clinical study. *J. Chromatogr. A* 1594, 160–172. doi:10.1016/j.chroma.2019.02.001

Wilkinson, M., White, I. R., Goodacre, R., Nijsen, T., and Fowler, S. J. (2020). Effects of high relative humidity and dry purging on VOCs obtained during breath sampling on common sorbent tubes. *J. Breath. Res.* 14, 046006. doi:10.1088/1752-7163/AB7E17

Wojnowski, W., Dymerski, T., Gębicki, J., and Namieśnik, J. (2019). Electronic noses in medical diagnostics. *Curr. Med. Chem.* 26, 197–215. doi:10.2174/0929867324666171004164636

Woollam, M., Angarita-Rivera, P., Siegel, A. P., Kalra, V., Kapoor, R., and Agarwal, M. (2022a). Exhaled VOCs can discriminate subjects with COVID-19 from healthy controls. *J. Breath. Res.* 16, 036002. doi:10.1088/1752-7163/ac696a

Woollam, M., Siegel, A. P., Grocki, P., Saunders, J. L., Sanders, D. B., Agarwal, M., et al. (2022b). Preliminary method for profiling volatile organic compounds in breath that correlate with pulmonary function and other clinical traits of subjects diagnosed with cystic fibrosis: a pilot study. *J. Breath. Res.* 16, 027103. doi:10.1088/1752-7163/ac522f

Xu, M., Tang, Z., Duan, Y., and Liu, Y. (2016). GC-based techniques for breath analysis: current status, challenges, and prospects. *Crit. Rev. Anal. Chem.* 46, 291–304. doi:10.1080/10408347.2015.1055550

Xu, Z. Q., Broza, Y. Y., Ionsecu, R., Tisch, U., Ding, L., Liu, H., et al. (2013). A nanomaterial-based breath test for distinguishing gastric cancer from benign gastric conditions. *Br. J. Cancer* 108, 941–950. doi:10.1038/bjc.2013.44

Zhang, Y., Guo, L., Qiu, Z., Lv, Y., Chen, G., and Li, E. (2020a). Early diagnosis of breast cancer from exhaled breath by gas chromatography-mass spectrometry (GC/MS) analysis: a prospective cohort study. *J. Clin. Lab. Anal.* 34, e23526. doi:10.1002/jcla.23526

Zhang, Y., Shi, L., Simoff, M. J., Wagner, O. J., and Lavin, J. (2020b). Biopsy frequency and complications among lung cancer patients in the United States. *Lung Cancer Manag.* 9, LMT40. doi:10.2217/lmt-2020-0022

Zou, Y., Hu, Y., Jiang, Z., Chen, Y., Zhou, Y., Wang, Z., et al. (2022). Exhaled metabolic markers and relevant dysregulated pathways of lung cancer: a pilot study. *Ann. Med.* 54, 790–802. doi:10.1080/07853890.2022.2048064

Zou, Y., Wang, Y., Jiang, Z., Zhou, Y., Chen, Y., Hu, Y., et al. (2021). Breath profile as composite biomarkers for lung cancer diagnosis. *Lung Cancer* 154, 206–213. doi:10.1016/j.lungcan.2021.01.020

Zou, Y., Zhang, X., Chen, X., Hu, Y., Ying, K., and Wang, P. (2014). Optimization of volatile markers of lung cancer to exclude interferences of non-malignant disease. *Cancer Biomarkers* 14, 371–379. doi:10.3233/CBM-140418

frontiers | Frontiers in Molecular Biosciences

TYPE Review
PUBLISHED 30 October 2024
DOI 10.3389/fmolb.2024.1421330

OPEN ACCESS

EDITED BY
Joana Pinto,
University of Porto, Portugal

REVIEWED BY
Paula Guedes De Pinho,
University of Porto, Portugal
Miyako Kusano,
University of Tsukuba, Japan

*CORRESPONDENCE
Steven J. Hallam,
shallam@mail.ubc.ca

RECEIVED 22 April 2024
ACCEPTED 15 October 2024
PUBLISHED 30 October 2024

CITATION
Szeitz A, Sutton AG and Hallam SJ (2024) A matrix-centered view of mass spectrometry platform innovation for volatilome research. *Front. Mol. Biosci.* 11:1421330. doi: 10.3389/fmolb.2024.1421330

A matrix-centered view of mass spectrometry platform innovation for volatilome research

Andras Szeitz[1,2,3], Annika G. Sutton[4] and Steven J. Hallam[1,2,3,5,6,7,8]*

[1]Genome Science and Technology Program, University of British Columbia, Vancouver, BC, Canada, [2]Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada, [3]Life Sciences Institute, University of British Columbia, Vancouver, BC, Canada, [4]School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada, [5]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada, [6]Department of Biochemistry, Chulalongkorn University, Bangkok, Thailand, [7]Bradshaw Research Institute for Minerals and Mining (BRIMM), University of British Columbia, Vancouver, BC, Canada, [8]ECOSCOPE Training Program, University of British Columbia, Vancouver, BC, Canada

Volatile organic compounds (VOCs) are carbon-containing molecules with high vapor pressure and low water solubility that are released from biotic and abiotic matrices. Because they are in the gaseous phase, these compounds tend to remain undetected when using conventional metabolomic profiling methods. Despite this omission, efforts to profile VOCs can provide useful information related to metabolic status and identify potential signaling pathways or toxicological impacts in natural or engineered environments. Over the past several decades mass spectrometry (MS) platform innovation has instigated new opportunities for VOC detection from previously intractable matrices. In parallel, volatilome research linking VOC profiles to other forms of multi-omic information (DNA, RNA, protein, and other metabolites) has gained prominence in resolving genotype/phenotype relationships at different levels of biological organization. This review explores both on-line and off-line methods used in VOC profiling with MS from different matrices. On-line methods involve direct sample injection into the MS platform without any prior compound separation, while off-line methods involve chromatographic separation prior to sample injection and analyte detection. Attention is given to the technical evolution of platforms needed for increasingly resolved VOC profiles, tracing technical progress over time with particular emphasis on emerging microbiome and diagnostic applications.

KEYWORDS

volatile organic compounds, selected ion flow tube-mass spectrometry, ion mobility spectrometry-mass spectrometry, proton transfer reaction-mass spectrometry, secondary electrospray ionization-mass spectrometry, time-of-flight mass spectrometry, comprehensive two-dimensional gas chromatography, high resolution multi-reflecting time-of-flight mass spectrometry

1 Introduction

Volatile organic compounds (VOCs) are gaseous carbon-containing compounds released from biotic and abiotic matrices, manifesting both high vapor pressure and low water solubility (US EPA, 2022). High vapor pressure is correlated with low boiling point

and serves as a measure of compound volatility. In some cases, VOCs are associated with adverse health effects depending on concentration and exposure time. The United States Environmental Protection Agency (US EPA) has established a classification system for VOCs that recognizes three primary categories including very volatile organic compounds (VVOCs) < 0°C to 50–100°C, volatile organic compounds (VOCs) 50–100°C to 240–260°C, and semi-volatile organic compounds (SVOCs) 240–260°C to 380–400°C (US EPA, 2022). Volatilome research arises in part from an awareness that both biotic and abiotic matrices emit VOCs, and that VOC profiles obtained using mass spectrometry (MS) platforms can provide useful information about the metabolic status of biological systems as well as potential signaling pathways or toxicological impacts in natural or engineered environments. For example, researchers have identified VOCs in plant (D'Alessandro and Turlings, 2006; Majchrzak et al., 2020), human (Amann et al., 2014; Drabińska et al., 2021; Bauermeister et al., 2022; Ferrandino et al., 2023; Fu et al., 2023), microbial (Boots et al., 2014; Drabińska et al., 2022), food (Cao et al., 2016; Carraturo et al., 2020; Tiwari et al., 2020), and environmental (Liu and Phillips, 1991; Higashikawa et al., 2013; Ullah et al., 2014) matrices. Moreover, several studies have evaluated the role of VOCs in mediating regulatory and metabolic interactions at the population and community levels of biological organization (Audrain et al., 2015; Weisskopf et al., 2021).

In plants, VOCs including terpenoids, fatty acids, and benzenoids are released under normal growth conditions (Dudareva et al., 2004), or in response to environmental stressors such as increased temperature and salinity or herbivory (Sharkey and Yeh, 2001; Mumm et al., 2003; Schröder et al., 2005). Plant-derived VOCs have gained increasing attention in relation to food security and climate change due to potential applications in promoting crop stress responses, pathogen defense, and enhanced biomass production connected to carbon capture and conversion processes (Materić et al., 2015). For example, gas chromatography-mass spectrometry (GC-MS) investigation of *Xanthomonas c. pv. vesicatoria* 85–10 resolved over 50 VOC compounds including several plant growth promoting ketone and methylketone compounds, and one compound linked to growth inhibition of the necrotrophic fungus *Rhizoctonia solani* (Weise et al., 2012). From an environmental perspective, VOCs associated with 48 *Actinobacteria* species isolated from soil and airborne-dust were profiled using GC-MS, resolving 126 predominantly C1 to C5 compounds, including alcohols, ketones, esters with the potential to mediate metabolic interactions among and between microbes and plants (Choudoir et al., 2019). In humans, VOCs have emerged as biomarkers for diagnostic screening and monitoring disease progression (Janssens et al., 2020; Berenguer et al., 2022), as well as detection of pathogens and antimicrobial resistance (AMR) phenotypes (Dixon et al., 2022). For example, VOC detection has been used for biomarker discovery in pulmonary tuberculosis (Fu et al., 2023), cystic fibrosis (Kaeslin et al., 2021), asthma, chronic obstructive pulmonary disease (Ratiu et al., 2020), lung (Ratiu et al., 2020; Temerdashev et al., 2023), prostate (Berenguer et al., 2022) and other type of cancers (Le and Priefer, 2023), urinary tract (Drabińska et al., 2022) and intestinal infections (John et al., 2021), irritable bowel syndrome (Zhang et al., 2023), as well as neurological disorders (Khan et al., 2021). In addition, several studies using isolated *Escherichia coli* have detected VOCs using solid-phase microextraction-gas chromatography coupled with mass spectrometry (SPME-GC-MS) on liquid cultures (Drabińska et al., 2022), or in strains cultivated on blood agar plates using thermal desorption gas chromatography coupled with time-of-flight MS (TD-GC-TOF-MS) (Boots et al., 2014). Similarly, from an industrial perspective, GC-MS analysis of Chinese milk fan (cheese) containing bacteria affiliated with *Lactococcus, Lactobacillus, Raoultella* and fungi affiliated with *Rhodotorula, Torulaspora,* and *Candida* fungi species identified 60 VOCs, including alcohols, aldehydes, ketones, esters, and aromatic compounds contributing to milk fan aroma (Chen et al., 2021).

The emergence of volatilome research is closely coupled with MS platform innovation instigating new opportunities for VOC detection from previously intractable matrices. MS platforms can use either on-line or off-line VOC detection methods that are closely coupled with increasing throughput and resolving power, respectively. On-line methods utilize direct sample introduction to the MS without upstream sample cleanup and compound separation protocols, while off-line techniques employ various analyte separation methodologies prior to MS detection. Previous reviews have discussed the on-line *versus* off-line methods for VOC detection within specific biological systems (Lindinger et al., 1998; Smith and Španěl, 2005; Materić et al., 2015; Ahmed et al., 2017; Lawal et al., 2017; Gould et al., 2021; Westphal et al., 2023). Here we expand on these accounts by presenting a matrix-centered review of volatilome research in relation to platform innovation over time, providing a practical guide for both practitioners and potential end-users with particular emphasis on emerging microbiome and diagnostic applications.

2 VOC detection platforms

Numerous contemporary reviews on analytical methods for detecting VOCs from different matrices are available (Materić et al., 2015; Ahmed et al., 2017; Lough et al., 2017; Lubes and Goodarzi, 2017; Majchrzak et al., 2018; Gould et al., 2021). Figure 1 provides a graphical overview of MS platforms used for VOC detection over time and Table 1 summarizes key scientific literature in relation to different matrix categories and cognate detection platforms. Emphasis is placed on differentiating between on-line methods in which sample preparation is directly coupled to sample injection and analysis, and off-line methods in which the process of sample preparation and analysis are uncoupled. Table 1 includes an overarching selection of MS platform innovation from the early stage of introducing ion molecule reaction MS in 1993 to the latest technological advancements in GCxGC-TOF-MS, in 2023. Review articles are also included in the table to complement research articles with the aim to offer a comprehensive view on VOC analysis combined with MS techniques, in a wide range of matrices.

2.1 On-line methods

On-line methods associated with real-time detection of analytes require minimal preparation, are compatible with both portable and high-throughput platform integration and provide

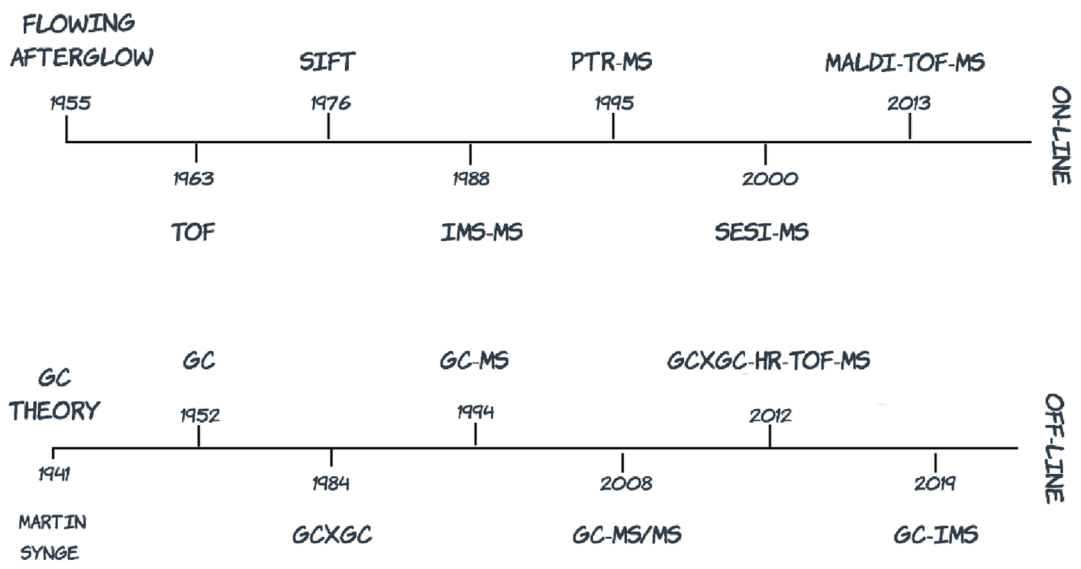


FIGURE 1
Summary of mass spectrometry platform innovation over the past 60 years in relation to volatile organic compound detection (TOF, time-of-flight; SIFT, selected ion flow tube; IMS-MS, ion mobility spectrometry mass spectrometry; PTR-MS, proton transfer reaction mass spectrometry; SESI-MS, secondary electrospray ionization mass spectrometry; MALDI, matrix-assisted laser desorption ionization; GC, gas chromatography; GCxGC, comprehensive two-dimensional gas chromatography; GC-MS, gas chromatography mass spectrometry; GC-MS/MS, gas chromatography triple quadrupole mass spectrometry; HR-TOF-MS, high resolution time-of-flight mass spectrometry).

relatively rapid results with reduced price point per sample. However, complex samples containing multiple structurally related analytes present a particular challenge for on-line detection.

2.1.1 Flowing afterglow and SIFT-MS

The flowing afterglow method was developed over 60 years ago to quantitatively measure ion-molecule reaction rate constants. This method utilized a microwave discharge to ionize a primary gas, with the resulting luminous glow migrating to the reaction area, where an ion-molecule reaction transfers charge to neutral species introduced into the buffer gas (Ferguson et al., 1969). In the 1970s flowing afterglow was implemented in a selected ion flow tube (SIFT) where a positively charged single species low energy primary beam was used to ionize neutral components present in a carrier gas, and the reaction was detected using quadrupole MS (Adams and Smith, 1976). The low energy ionization associated with SIFT-MS resolves fewer analyte fragments with less complex mass spectra formation in gas mixtures and alternative precursor ions can be matched with different matrices to obtain representative product ions. For example, SIFT-MS platforms have been used to quantify trace amounts of gas, even in the parts per billion (ppb) range (Smith and Spanel, 1996) down to parts per trillion volume (pptv) (Bierbaum, 2015; Smith et al., 2022) in both air and human breath using primary ion beams composed of dioxygenyl (O_2^+), hydronium (H_3O^+), or nitrosonium (NO^+) (Španěl et al., 1999; Smith and Španěl, 2005) ions. SIFT-MS has also been used to monitor bacterial growth (Allardyce et al., 2006a; Allardyce et al., 2006b; Scotter et al., 2006) and to explore ion-molecule reaction kinetics in breath collection bags compared to other methods (Malásková et al., 2019).

2.1.2 IMS-MS

First developed in the 1970s, ion mobility spectrometry (IMS) initially known as plasma chromatography (Cohen and Karasek, 1970), typically uses a radioactive source to ionize analyte gasses that are then separated in a drift tube at atmospheric pressure under an electric field and the counterflow of an inert drift gas (Karasek et al., 1976; Karpas et al., 1988). Due to collisions with the drift gas molecules and electric field acceleration, charged species attain an ion mobility proportional to their shape, charge, size, etc., and their different arrival times at the Faraday plate detector enables selective and sensitive measurements (Hill et al., 1990; Fernández Maestre, 2012). IMS has branched into different forms including drift-tube ion mobility spectrometry, traveling-wave ion mobility spectrometry, trapped ion mobility spectrometry, and field-asymmetric waveform ion mobility spectrometry (Dodds et al., 2020). IMS has been used to measure VOCs in environmental samples (Takaya et al., 2022), and for more complex matrices, IMS can be interfaced with an upstream multicapillary column (MCC) containing up to 1,000 parallel capillaries to separate VOCs prior to ionization (Vautz et al., 2006; Handa et al., 2014). More recent integration of IMS with MS detection (Mukhopadhyay, 2008) enables compound detection based on drift time and mass-to-charge (m/z) ratio (Collins and Lee, 2002). Collision cross-section calculations using IMS-MS spectra can also be used to estimate the gas-phase size of ions useful for untargeted analysis (Collins and Lee, 2002; Lapthorn et al., 2013; Dodds and Baker, 2019).

2.1.3 IMR-MS and PTR-MS

In the 1980s ion-molecule reaction mass spectrometry (IMR-MS) emerged as an alternative to SIFT-MS. In contrast to SIFT-MS, IMR-MS employs a two-stage ionization process with

TABLE 1 Summary of analytical techniques and applications used in the analysis of volatile organic compounds with mass spectrometry*.

Matrix	Analytical platform	Objectives	Findings	References
Microbes	HS-GC-TOF-MS	Analysis of VOCs in 200 bacterial headspace samples from various species for bacterial identification is reported	Bacterial identification was possible from VOCs including differentiation between a methicillin-resistant and -sensitive *Staphylococcus aureus*	Boots et al. (2014)
	SPME-GC-MS	Monitoring the VOC profiles after adding cephalosporin antibiotics to *Escherichia coli* strains is discussed	Antibiotic susceptibility was detected in urinary tract infection caused by *Escherichia coli* after 2 h	Drabińska et al. (2022)
	SHS-MCC-GC-IMS	An automated approach to differentiate between *Listeria spp.* from VOCs is presented	SHS-MCC-GC-IMS could differentiate between *Listeria* species using their VOC response	Taylor et al. (2017)
	PTR-TOF-MS, SPME-GC-MS	VOC production by *Porphyromonas gingivalis* after treatment with amoxicillin is investigated	Metabolic effects of amoxicillin were reflected in VOCs produced by *Porphyromonas. gingivalis*	Roslund et al. (2022)
	Review	Recent advancements in microbial VOCs and their roles in microbial ecosystems is reviewed in the context of analytical chemistry techniques	Linking the characteristics of microbial VOCs to analytical chemistry techniques could advance the knowledge on volatile-mediated chemical interactions	Weisskopf et al. (2021)
	Review	An overview on the evolution of VOCs and VOC-labeled enzyme substrates to detect pathogenic bacteria is presented	Colorimetric sensor arrays and MS techniques were useful in diagnostics and decision-making, in healthcare and food industries	Lough et al. (2017)
Plants	DI-MS	A review of the applications of DI-MS in real-time plant volatilomics is presented	DI-MS advancements coupled with omics platforms allowed real-time investigation of plant biogenic VOCs	Majchrzak et al. (2020)
	SESI-HRMS	A platform with a broad metabolome coverage for plant VOCs in real-time is described	SESI-MS detected an excess of stress and light induced VOCs emitted by *Begonia semperflorens*	Barrios-Collado et al. (2016)
	GC-MS	Volatiles released by Scots pine twigs after oviposition is analyzed for compounds attracting egg parasitoids	GC-MS analysis showed increased amounts of (E)-β-farnesene that attract egg parasitoids	Mumm et al. (2003)
	MR-GC-DMS	A novel MR-GC-DMS platform using microcontroller boards is reported	MR-GC-DMS could detect VOCs to distinguish between healthy and infected *Rhododendron* plants	Anishchenko et al. (2018)
	HS-SPME-GCxGC-TOF-MS	An aroma comparison in five pear cultivars is investigated	241 volatiles were identified advancing analytical techniques in aroma evaluation	Wang et al. (2019)
	Review	An overview on the role of herbivore-induced plant volatiles (HIPVs) on analytical chemistry methods is discussed	The importance of selecting the appropriate analytical methods could mitigate the challenges in sampling and analyzing HIPVs	D'Alessandro and Turlings (2006)
	Review	A comprehensive guide of analytical methods for plant foliar VOCs, and sampling approaches for plant sciences is presented	The importance of selecting the appropriate sampling methods and analytical techniques for accurate results was emphasized	Materić et al. (2015)

(Continued on the following page)

TABLE 1 (*Continued*) Summary of analytical techniques and applications used in the analysis of volatile organic compounds with mass spectrometry*.

Matrix	Analytical platform	Objectives	Findings	References
Human Volatilome	Review	An overview of the human volatilome encompassing VOCs in numerous matrices is reported	Comprehensive investigations of VOCs across human matrices contributed to innovative scientific advancements	Amann et al. (2014)
	Review	A review reporting 2,746 VOCs in healthy humans detected mainly with MS techniques is presented	Classified VOCs helped disease diagnosis using the appropriate body matrix for specific research areas	Drabińska et al. (2021)
	Review	An introduction of using untargeted MS techniques to explore human microbiota and its impact on bodily environments	The pivotal role of MS was underlined in mapping microbial functions and physiological impacts	Bauermeister et al. (2022)
Blood	SIFT-MS	SIFT-MS can measure metabolic gases in the headspace of blood culture bottles achieving fast diagnosis in bacteremia or sepsis	Trace gases produced by bacterial cultures were detected after 6 h that was consistent with the gases produced at 24 h	Allardyce et al. (2006b)
	GC-IMS	mVOCs can be analyzed in the headspace of blood culture bottles infected with sepsis-specific pathogens	Using an autosampler with GC-IMS to measure mVOCs could enable point-of-care applications for early detection of sepsis	Drees et al. (2019)
	HS-SPME-GC-MS/MS	A method is developed and validated to quantitate mVOCs as biomarkers in human blood after indoor mold exposure	The method had good linearity, accuracy, precision, limit of detection, and was used to quantify 21 mVOCs in blood	Tabbal et al. (2022b)
	SPME-GC-MS/MS	A method is used to measure BTEX VOCs in human blood at ppt level to conduct PBPK modelling after outdoor and indoor VOC exposure	The method was validated to detect 12 VOCs and applied to analyze samples for a nation-wide biomonitoring study in Canada	Aranda-Rodriguez et al. (2015)
Urine	HS-SPME-GC-MS/MS	A method is developed and validated to quantitate mVOCs as biomarkers in urine after indoor mold exposure	The method had good linearity, accuracy, precision, limit of detection, and was used to quantify 21 mVOCs in urine	Tabbal et al. (2022b)
	Q-MRT-MS	A novel Q-MRT-MS system is introduced featuring an inclined double-orthogonal accelerator and planar gridless ion mirrors with fourth-order energy focusing	Diminishing duty cycles in long flight time instruments could be improved by multiplexing using Encoded Frequent Pushing technology	Cooper-Shepherd et al. (2023)
	PTR-TOF-MS, SIFT-MS	Basic principles and differences between SIFT-MS and PTR-MS are discussed, and real-time breath analysis with the techniques are compared	Real-time analysis of VOCs offered quick results but suffered from interferences. GC separation mitigated interference, but with longer analysis time	Smith et al. (2014)
	PTR-MS PTR-QqQ-MS	PTR-MS can find a marker VOC in the breath of kidney transplant patients that correlates with blood serum creatinine and daily urine production	PTR-TOF-MS and PTR-QqQ-MS confirmed the marker VOC potentially making it possible to monitor kidney functions	Kohl et al. (2013)

(Continued on the following page)

TABLE 1 (*Continued*) Summary of analytical techniques and applications used in the analysis of volatile organic compounds with mass spectrometry*.

Matrix	Analytical platform	Objectives	Findings	References
Breath	SESI-MS, SESI-HRMS	A review of SESI-MS with an emphasis on quality assurance in data analysis of breath samples in the clinic is reported	SESI-MS was suitable to complement molecular diagnostic methods in early-stage biomarker discovery	Blanco and Vidal-de-Miguel (2021)
	SESI-HRMS	First-time use of SESI-HRMS to find metabolic VOCs in children's breath with allergic asthma is reported	The technique was useful to identify children with allergic asthma from breath VOCs	Weber et al. (2023b)
	SESI_TOF, PTR-HRMS	First-time systematic evaluation of SESI-HRMS and PTR-HRMS focusing on their suitability to analyze VOCs in adult breath is presented	The sensitive SESI-HRMS found more features but detected less ions in the low mass region than PTR-HRMS.	Bruderer et al. (2020)
	GC-MS	The breath sample analysis of liver cirrhosis patients with Breath Biopsy OMNI Global VOC service and GC-MS to identify candidate biomarkers is reported	VOCs were found as potential biomarkers for progressive liver disease detection that showed good correlation with biomarkers obtained from serum	Ferrandino et al. (2023)
	GC-MS/MS	A needle trap micro-extraction technique with GC-MS/MS is used to analyze VOCs in the breath of patients with heart failure	The optimized method included hydrocarbons, carbonyls, aromatics, sulfur compounds pptv level	Biagini et al. (2017)
	TD-GC-MS	The Peppermint Initiative: A benchmark to establish standardization protocols for VOC analysis in breath is proposed	The Peppermint Consortium encouraged international, cross-platform and interdisciplinary collaboration	Henderson et al. (2020)
	TD-GC-MS/MS	An optimized method is used to analyze carbonyl compounds in the breath of patients with heart failure	The validated method measured aldehydes and ketones at pptv levels to monitor clinical improvement of the patients	Lomonaco et al. (2018)
	TD-GCxGC-TOF-MS	A method is used to investigate whether malarial infection results in characteristic changes of breath profiles in febrile children	A shift was identified in breath composition of Malawian children with six VOCs enabling classification of infection status	Schaber et al. (2018)
	Review	A review of breath biopsy for the measurement of VOCs to monitor respiratory tract, gastrointestinal disorders and sepsis is presented	Online databases for VOCs were built to support breathomics research in patient cohorts with diverse pathologic states	Belizário et al. (2021)
	Review	An overview of sampling and analytical techniques of breath VOCs to diagnose microbial infections and diseases is presented	Microbial infection could produce distinguishable VOCs whose detection could help understand the complex host-pathogen crosstalk	Ahmed et al. (2017)
Food	GC-MS	Development and validation of a GC-MS technique for the analysis of VOCs in individual food and diet samples are reported	The method showed varying concentrations of VOCs in food composites underlining the importance of diet when assessing VOC exposure	Cao et al. (2016)
	HS-SPME-GC-MS	Use of HS-SPME with GC-MS for real-time detection of bacterial contamination of meat is reported	The technique identified unique VOCs supporting an early-warning system for meat contamination	Carraturo et al. (2020)
	Review	Classification of VOCs from horticultural products and linking them with stress factors for quality management are outlined	VOCs as biomarkers could be used to monitor the quality of horticulture goods during storage	Tiwari et al. (2020)

(Continued on the following page)

TABLE 1 (*Continued*) Summary f analytical techniques and applications used in the analysis of volatile organic compounds with mass spectrometry*.

Matrix	Analytical platform	Objectives	Findings	References
Environment	IMR-MS	An IMR-MS method is used to analyze complex gas mixtures with the advantage of less fragmentation compared to electron-impact ionization	IMR-MS was useful for on-line analysis of car engine exhausts, investigation of catalytic processes and other industrial applications	Lindinger et al. (1993)
	GC-MS/MS	A HS-SPME-GC-MS/MS method is used to measure mVOC biomarkers as occupational health risk agents after indoor mold exposure	The procedure was applied to propose a physiologically based pharmacokinetic model to assess human exposure to indoor mold	Berkane et al. (2023)
	GCxGC-HR-TOF-MS	In diesel fuel samples, advanced data reduction of large datasets obtained with GCxGC-HR-TOF-MS is important for efficient data handling and analysis	Combining Kendrick mass defect and knowledge-based rules, the group type classification reduced complex datasets to a few numerical values	Weggler et al. (2019)
Diagnostics	IMR-MS	The VOC analysis in the headspace of Gram-negative bacterial cultures with IMR-MS as an *in vivo* diagnostic tool is discussed	IMR-MS yielded fast bacterial growth detection and identification offering the possibility of automation	Dolch et al. (2012a)
	SESI-HRMS	Identification of VOC biomarkers in selecting bacterial cultures in cystic fibrosis and assigning molecular structures to features are reported	Several pathogens were distinguished *in vitro* using their VOC profiles and proposed as biomarkers in disease detection in the clinical context	Kaeslin et al. (2021)
	HPPI-TOF-MS	A non-invasive method is implemented to measure VOCs in exhaled breath of TB patients to improve the efficiency of disease diagnosis	The breathomics-based method was accurate, sensitive and specific, offering a simple clinical diagnostic tool for TB screening	Fu et al. (2023)
	Review	MALDI-TOF-MS is described as a method for microbial identification to displace conventional diagnostic techniques	MALDI-TOF-MS would shape and define the clinical microbiology laboratory landscape to accelerate diagnostics and improve patient care	Clark et al. (2013)
	Review	A review of MS techniques with a multiomic approach and integrated science to investigate Enterobacteriaceae phenotypes with carbapenem resistance is reported	Targeted analyses with chromatographic separation and HRMS had potentials to detect molecular signatures and antimicrobial resistance	Dixon et al. (2022)
	Review	VOCs analyzed in headspace of human body fluids with various MS techniques during diagnostic screening and monitoring disease progression in lung cancer are reported	Potential biomarker VOCs in lung cancer showed overlap in human matrices, therefore, standardized trials are needed to validate clinically relevant biosignatures	Janssens et al. (2020)
Multi-omics	GC-MS	Comparative transcriptomics and metabolite profiling are used in tea plants to assess volatile heterosis	Genes and transcription factors with over-dominating expression could serve as candidate genes for breeding high-volatile tea varieties	Zheng et al. (2019)
	GC-MS	Metabolite screening and RNA sequencing of *Magnolia champaca*are employed to discover VOC biosynthesis pathways and floral scent-related genes	GC-MS found 43 VOCs in flowers during sequencing and *de novo*assembly of the transcriptome yielded 47,688 non-redundant unigenes	Dhandapani et al. (2017)

(Continued on the following page)

TABLE 1 (*Continued*) Summary f analytical techniques and applications used in the analysis of volatile organic compounds with mass spectrometry*.

Matrix	Analytical platform	Objectives	Findings	References
Other analyses	Flowing afterglow	Marking the 50th anniversary of the invention of flowing afterglow with a brief overview of its advancements	Contribution of invertors Eldon Ferguson, Fred Fehsenfeld, Art Schmeltekopf was recognized	Bierbaum (2015)
	SIFT-MS	SIFT-MS is introduced to real-time quantification of trace gases in air and breath. Applications in research areas is reviewed	The technique could detect gases at ppb levels in various matrices. Commercial SIFT-MS instruments were reviewed	Smith and Španěl (2005)
	IMS, IMS-MS	IMS platforms are reviewed with their advantages and disadvantages	IMS interfaced with HRMS could separate and identify unique chemical isomers and isobars	Dodds and Baker (2019)
	IMS-MS	A review of IMS-MS applications to small molecules in drug discovery with limitations and potential applications is presented	IMS-MS, in combination with several analytical techniques, offered structural elucidation in milliseconds	Lapthorn et al. (2013)
	SESI-MS	A SESI model with the effects of several technical parameters on ionization is presented	The ionization mechanism was based on gas phase species rather than charged droplets	Vidal-de-Miguel and Herrero (2012)
	PTR-TOF-MS	A workflow is proposed for the measurement of numerous target compounds in humid air and the results are compared to SIFT-MS data	Using PTR-TOF-MS, less fragmentation and similar information-rich data could be obtained as seen with SIFT-MS.	Romano and Hanna (2018)
	GCxGC-HR MR-TOF-MS FFP EFP	The importance of improving the duty cycle to increase sensitivity of HR MR-TOF-MS FFP systems is discussed	Using Encoded Frequent Pushing spectral multiplexing technique improved the duty cycle in HR MR-TOF-MS FFP systems	Willis et al. (2021)
	PTR-TOF-MS, SPME-GC-TOF-MS	Two MS techniques are employed to monitor the VOC production during 3D printing	Quantitative VOC emissions obtained with PTR-TOF-MS were confirmed with qualitative analysis with SPME-GC-TOF-MS.	Wojnowski et al. (2022)
	TD-GCxGC-TOF-MS, PTR-TOF-MS	VOC emission and uptake of components from the breath sampling device ReCIVA are assessed	Thermal pretreatment of ReCIVA components reduced VOC emissions, and uptake differed for various parts of the device	Pham et al. (2023)
	Review	A review on the analyses of aroma VOCs with advanced MS techniques in food products and chemometrics is presented	In volatile metabolomics, lack of metabolite-specific libraries limited the identification and structural elucidation of VOCs	Lubes and Goodarzi (2017)
	Review	An overview of hyphenated, real-time MS techniques with advantages, limitations and examples is presented	Hyphenated techniques had good qualitative data with long run time. Real-time techniques offered good quantitative data in short analysis time, but only tentative qualitative results	Gould et al. (2021)

For abbreviations, please, refer to Supplementary Material.

Krypton (Kr^+) or Xeon (Xe^+) primary beams (Lindinger et al., 1993). During the first stage, primary reagent gas is generated through electron impact ionization, and in the second stage, reagent gas enters a reaction chamber through a lens system where charge is then transferred (Lindinger et al., 1993). Initially, IMR-MS systems were employed to analyze industrial gas mixtures, such as emission from furnaces and motors that required the use of Kr^+ or Xe^+ for efficient ionization (Lindinger et al., 1993). In later manifestations, IMR-MS systems capable of alternating between different primary ion beams, including Kr^+, Xe^+ or Mercury (Hg^+) were developed and used to differentiate VOCs produced by Gram-negative and Gram-positive bacteria in headspace analysis of anaerobic blood samples (Dolch et al., 2012a; Dolch et al., 2012b). IMR-MS was also used to analyze VOCs in exhaled breath (Dolch et al., 2015; Meidert et al., 2021).

Proton transfer reaction MS (PTR-MS) introduced in the 1990s is similar to IMR-MS, but uses an ion beam composed of H_3O^+ with high proton affinity (Lindinger et al., 1993; Hansel et al., 1995). Use of H_3O^+ results in low energy ionization resolving fewer analyte fragments with less complex mass spectra formation in gas mixtures. From a VOC detection standpoint, H_3O^+ is effective in detecting trace gas components in exhaled breath (Lindinger et al., 1993; Hansel et al., 1995; Smith and Spanel, 1996). Additionally, the high proton affinity of H_3O^+ reduces ion-molecule reactions with major air components including nitrogen (N_2), oxygen (O_2), carbon dioxide (CO_2), and water (H_2O) (Lagg et al., 1994), reducing potential matrix effects (Bruderer et al., 2020). The singular use of a H_3O^+ makes PTR-MS less versatile than SIFT-MS with respect to matching precursor ions with different matrices to obtain representative product ions (Smith and Španěl, 2005). Despite this limitation, PTR-MS has diversified into several forms including PTR-single quadrupole MS (PTR-QMS) (Smith et al., 2014), PTR-triple-quadrupole MS (PTR-QqQ-MS) (Kohl et al., 2013), and PTR-TOF-MS (Smith et al., 2014).

PTR-QMS is a scanning MS platform with a relatively slow data acquisition rate and nominal mass resolution. However, it is effective for targeted analyses to quantitate known VOCs with accuracy and precision (Smith et al., 2014). Combined use of PTR-MS and solid phase microextraction gas chromatography time-of-flight mass spectrometry (SPME-GC-TOF-MS) has been used for monitoring applications (King et al., 2010; Majchrzak et al., 2018) including real-time detection of VOCs released from thermoplastics during 3D printing (Wojnowski et al., 2022). In addition, direct infusion (DI) PTR-MS has been used for real-time detection of VOCs released from plants in combination with multi-omic sequencing to establish a monitoring network to refine global emission budgets and observe plant metabolism at different levels of biological organization (Majchrzak et al., 2020). PTR-QqQ-MS is an extension of PTR-QMS providing higher specificity and sensitivity (Kohl et al., 2013). PTR-TOF-MS platforms collect full mass spectra with high mass resolution (Dodonov et al., 2000; Cooper-Shepherd et al., 2023) making them effective in separating isobaric compounds with high sensitivity and potentially accurate mass for untargeted analyses of complex matrices with many analyte fragments (Smith et al., 2014). Biomarkers from exhaled breath of patients with kidney dysfunction have been identified using a combination of PTR-MS, PTR-TOF-MS with structural elucidation using PTR-QqQ-MS (Kohl et al., 2013).

2.1.4 SESI-MS and Orbitrap-MS

First introduced in the late 1980s, electrospray ionization (ESI) made it efficient to ionize liquid phase polar molecules using a sensitive and low energy process (Whitehouse et al., 1986; Fenn et al., 1989; Fenn, 2002). In a variation of ESI called secondary electrospray ionization (SESI) developed in the 2000s, neutral compounds in gas phase were introduced into the nebulized spray of an ESI stream, where ionized droplets transfer charge to the gaseous species followed by MS detection (Wu et al., 2000; Vidal-de-Miguel and Herrero, 2012). SESI-MS has been used to analyze volatile fatty acids (VFAs) in exhaled breath (Martínez-Lozano et al., 2011) and has been used in combination with high-resolution MS (HRMS) time-of-flight (TOF) and Orbitrap instruments. The first commercial Orbitrap platform was introduced in 2005 for high-resolution mass spectrometry. Produced ions enter a curved trap, where they are collisionally cooled and enriched. The concentrated ion packets are injected orthogonally into the orbitrap where they go into an axial oscillation along a central electrode with a frequency that is proportional to their *m/z* ratio. The central electrode has an opposing electrical charge, and ion stability is achieved by high velocity oscillation that prevents the ions from crashing into the electrode. The resulting resonance signal undergoes a mathematical treatment with Fourier transform, where the oscillation signal is converted to a mass spectrum (Makarov, 2000; Zubarev and Makarov, 2013).

Both SESI and Orbitrap are typically combined with off-line chromatography methods to improve compound identification and mass resolution. SESI-HRMS with TOF has been employed to identify biomarkers for cystic fibrosis in bacterial cultures (Kaeslin et al., 2021), to benchmark PTR-HRMS results (Bruderer et al., 2020), and to monitor VOC production in plants over light-dark cycles (Barrios-Collado et al., 2016). More recently SESI has developed into so-called super SESI, an advanced, electrode-free design with less background noise and memory effects (Blanco and Vidal-de-Miguel, 2021). Super SESI-HRMS with TOF and Orbitrap has been used to profile VOCs in exhaled breath (Blanco and Vidal-de-Miguel, 2021; Weber et al., 2023a; Weber et al., 2023b). Orbitrap-HRMS has also been used to detect VOCs associated with cometary ice analogs (Javelle et al., 2021), as well as SVOCs emanating from environmental dust samples (Pourasil et al., 2022), and plant volatilome (Majchrzak et al., 2020).

2.1.5 TOF-MS

Time-of-flight MS (TOF-MS) introduced in the mid-1950s accelerates ions in an electric field within a flight tube. Ions with a smaller *m/z* ratio travel faster, while those with a higher *m/z* travel slower. Ions are simultaneously detected at a microchannel plate detector, where cascades of electrons are converted into photons amplified by a photomultiplier to generate signals for measurement (Cotter, 1989). The difference between the arrival time of ions at the detector decreases as their *m/z* values increase, and ions with the same *m/z* ratio arrive with a time distribution that can overlap with adjacent ions, reducing mass resolution. Linear TOF-MS instruments, particularly matrix-assisted laser desorption ionization TOF MS (MALDI-TOF-MS) systems are now routinely used in analytical labs with applications as varied as proteomics, biomarker discovery, imaging, materials and environmental monitoring (Clark et al., 2013). A more intricate

TOF-MS design is the orthogonal acceleration TOF-MS (oaTOF-MS), where ions are accelerated in the drift region perpendicular to their original direction. Except for the MALDI-TOF-MS, most contemporary TOF-MS platforms use orthogonal acceleration and are commonly referred to as TOF-MS.

High resolution TOF-MS instruments emerged in the 1990s with improved mass resolution using collisional focusing (Douglas and French, 1992) and orthogonal acceleration (Guilhaus et al., 2000), or reflectrons (Mamyrin et al., 1973; Wang et al., 1994; Weggler et al., 2019; Cooper-Shepherd et al., 2023) to increase resolution and the flight path of the ions. A typical drift length in earlier linear oaTOF systems was around 1.5 m (Guilhaus et al., 2000), and this length could be significantly extended by employing different geometric designs. However, increasing the number of passes across mirror grids in reflectrons reduces sensitivity and duty cycle (Cooper-Shepherd et al., 2023). Duty cycle (DuC) describes the proportion of time that ions can enter the TOF for analysis. As mass resolution increases, the duty cycle decreases, with the highest duty cycle attained for the highest *m/z* ion and diminishing for smaller ions (Chernushevich et al., 2017; Willis et al., 2021). The introduction of a linear ion trap/release setup also referred to as "Zeno pulsing" has enabled nearly 100% DuC over a wide *m/z* range using V- (Loboda and Chernushevich, 2009) and W-geometry systems (Chernushevich et al., 2017) in TOF-MS configurations with 20,000 and 90,000 resolving power, respectively. More recently, high resolution, accurate mass, quadrupole-multi-reflecting time-of-flight (Q-MRT) MS instruments featuring open-loop, planar, gridless ion mirrors with fourth-order energy focusing and a 48 m flight path have reached over 200,000 resolving power with sub-ppm mass accuracy (Cooper-Shepherd et al., 2023).

In conventional TOF-MS systems, the push cycle in the accelerator region is followed by a pause period until the largest *m/z* ion arrives at the detector. With longer flight paths, the waiting period may become excessively long, rendering the DuC impractical. Diminishing DuC in instruments with extended flight times can be mitigated by multiplexing using Encoded Frequent Pushing (EFP) technology, where the flight tube is continually filled with ions from subsequent pulses, and the time offset of the pulse frequency is encoded in a sequence that can be accurately decoded to represent individual signals in Q-MRT systems (Willis et al., 2021; Cooper-Shepherd et al., 2023). For example, high resolution multi-reflecting time-of-flight mass spectrometry with folded 20 m flight path (HR-MR-TOF-MS FFP) combined with EFP technology reaches a 50,000 resolving power. Analysis of Egyptian mummy bandage extracts using a GCxGC-HR-MR-TOF-MS FFP with EFP system indicated improvements in signal intensity, dynamic mass range, accurate mass data, and limit of detection confirming the reliable operation of decoding algorithms and hardware with increased transient length in HR-TOF-MS instruments (Willis et al., 2021; LECO Corporation White paper, 2021).

TOF-MS can be combined with numerous off-line methods for VOC detection. For instance, a PTR-TOF-MS method was developed for improved detection of aldehydes, fatty acids, and phenols in humid air (Romano and Hanna, 2018). The same method was used to profile VOCs produced by *Porphyromonas gingivalis*, a common constituent of the human oral microbiome identifying biomarkers in exhaled breath and saliva samples (Roslund et al., 2022). Advances in high-speed electronics, collisional cooling and orthogonal acceleration have increased the resolution of TOF instruments (Douglas and French, 1992; Guilhaus et al., 2000; Chernushevich and Thomson, 2004). In such instruments, ions pass through a multi-pole lens system filled with a low-pressure inert gas and applied radio frequency (RF) resulting in collisional cooling that forms a narrow and dense ion beam. The beam enters an acceleration region, where it is further enriched and the ions are pushed as concentrated packets into the flight tube in an orthogonal direction (Dodonov et al., 1987; Dodonov et al., 1993; Dodonov et al., 2000). High resolution TOF-MS (HR-TOF-MS) has been used to profile VOCs in exhaled breath (Weber et al., 2023a; Weber et al., 2023b), to identify biomarkers for cystic fibrosis in bacterial cultures (Kaeslin et al., 2021), and to benchmark SESI and PTR-HRMS results (Bruderer et al., 2020).

2.2 Off-line methods

Off-line or hyphenated methods integrate chromatographic separation processes upstream of analyte detection. For example, gas chromatography to separate sample components prior to MS can improve resolution, allowing for more precise identification and quantification of the analytes. Gas chromatography (GC), formerly known as gas-liquid chromatography, evolved from liquid chromatography (LC) in 1941, when Martin and Synge proposed during their research on liquid-liquid partition chromatography that the mobile liquid phase could be replaced with a gas phase (Martin and Synge, 1941). The first application with GC was published in 1952 reporting the separation of VFAs from other acidic components and it can be considered the first publication on VOC analysis using GC (James and Martin, 1952). Nowadays, VOC profiling involves various types of GC utilizing open tubular capillary columns with diverse stationary phases connected to a wide range of detectors. These methods require more intensive sample preparation and more time to implement with respect to method development and sample analysis. Although the process of sample preparation can be semi-automated, the throughput of off-line methods tends to be much lower than on-line methods. The primary advantage of using off-line methods is the increased resolution of analytes with less deconvolution needed to interpret mass spectra resulting in improved identification within more complex matrices.

2.2.1 GC-MS

First introduced in the 1950s, GC-MS instruments employed packed columns connected to a magnetic sector mass spectrometer (Grayson, 2016). The first GC with a single quadrupole MS (QMS) became commercially available in 1961 (Finnigan, 2016) and has become one of the most widely used platforms for small molecule detection in liquid or gaseous phase samples across diverse matrices. Currently, most contemporary GC-MS platforms use GCs with capillary columns coupled with a QMS and are commonly referred to as GC-MS. GC-MS instruments are recognized for their robustness, user-friendly configurations, affordability, and high chromatographic resolution with low mass resolution, enabling detection and quantitation of analytes with nominal mass (Rey-Stolle et al., 2022).

Liquid samples can be introduced directly into the injection port, where they are vaporized, focused, and carried to the

column by the carrier gas in either a split or splitless injection mode. Gas-phase samples are injected in a similar way except the analytes are captured from the headspace (HS) of the sample container using static HS (SHS), dynamic HS (DHS), SPME, thermal desorption (TD), purge-trap, needle trap, etc., methods (Sugita and Sato, 2021) followed by sample introduction to the injection port. Thermally labile compounds, instead of being vaporized, are injected via cool on-column injection. Analytes are separated based on their physico-chemical properties in relation to the column's stationary phase, carrier gas, oven temperature programming, and other experimental parameters. Analytes eluted from the column are ionized, typically using electron impact (EI) ionization, where an electron beam emitted by a tungsten filament bombards the molecules, fragmenting them into smaller pieces. This process produces radical cations where the molecular ion is typically not observed. While hard ionization EI is useful for quantitation in targeted analysis, it is less optimal when conducting compound characterization, structural elucidation, or untargeted analysis. Alternatively, analytes can be ionized with chemical ionization, where the electron beam ionizes a reagent gas (e.g., methane, ammonia, etc.) first, then the ionized reagent gas transfers charge to the analyte in a soft ionization process that produces negatively or positively charged ions with the molecular ion and less fragmentation (González et al., 1995; Lubes and Goodarzi, 2017). The ionized species enter a single quadrupole MS, where a constant ratio of direct current (DC)/RF is applied and ramped on the rods of the quadrupole, resulting in ions following either an unstable or stable trajectory. The trajectory of an ion is determined by its *m/z* ratio with unstable ions hitting the rods, discharging, and venting from the system, while stable ions travel through the quadrupole, generating a signal at the detector (Pedder, 2001).

From a VOC detection standpoint, TD-GC-MS has been used to develop standards for exhaled breath (Henderson et al., 2020), and HS-SPME-GC-MS has been used to identify biomarkers of meat spoilage associated with *Salmonella Typhimurium* and *Campylobacter jejuni* reducing the time required for regulatory compliance (Carraturo et al., 2020). Similarly, HS-SPME-GC-MS has been used to profile metabolites associated with the *Candida spp*. volatilome identifying sesquiterpene indicators for *Candida albicans* infection (Fitzgerald et al., 2022). In special cases, GC ionization has been achieved using ESI where the column effluent was introduced into an electrospray plume using a multiple channel ESI MS technique. This method was applied to study the chemical reactions of VOCs with solid catalysts, such as the dehydration of dimethylhydrazine in the presence of mercury oxide (Lee and Shiea, 1998). More recently, GC-IMS has been developed to analyze VOCs in the headspace of both aerobic and anaerobic human blood culture samples, identifying species-specific volatiles that enabled the identification of bacterial strains in bloodstream infections after a 6-h incubations (Drees et al., 2019). Similar to TD-GC-MS, HS-GC-IMS has been used to benchmark VOC detection standards for exhaled breath (Ruszkiewicz et al., 2022) and to profile clinical samples including HS-MCC-GC-IMS system used to detect VOCs associated with *Listeria* spp. infections (Taylor et al., 2017). Anishchenko and colleagues developed a modular and reconfigurable system combined with differential mobility spectrometry, a variant form of IMS, to detect VOCs associated with *Phytophthora ramorum*, a protistan plant pathogen (Anishchenko et al., 2018). Combining GC-MS methods with various collection and detection modules provides a versatile and customizable framework for detecting VOCs from diverse matrices.

2.2.2 GC-MS/MS

The first commercially available tandem GC-MS/MS platform appeared on the market in 2008 incorporating a primary quadrupole (MS1), collision cell, and secondary quadrupole (MS2) with ion optics and focusing lenses. Precursor ions exiting MS1 undergo fragmentation in the collision cell by colliding with an inert, pressurized collision gas (such as helium or nitrogen), and the resulting product ions enter MS2, where they are filtered by applying and ramping a constant DC/RF ratio on the quadrupole rods, as previously explained. The primary advantage of GC-MS/MS lies in its selective and sensitive quantification of targeted analytes in complex matrices, achieved through a multiple reaction monitoring (MRM) mode of operation. In MRM mode, the most abundant precursor and product ions obtained from MS1 and MS2 are paired into a joint MRM signal enabling the targeted detection of the compound with high selectivity and sensitivity. In MRM, unlike ramping voltages, specific DC/RF ratios are applied to MS1 and MS2, ensuring stable trajectories for the precursor and product ions of selected analytes while reducing matrix effects. High-speed electronics enable the rapid interchange of many DC/RF ratios on the quadrupoles in microseconds, allowing simultaneous detection and quantification of numerous analytes in complex matrices (Lubes and Goodarzi, 2017). For example, HS-SPME-GC-MS/MS was used to profile microbial VOCs in human urine and blood as potential biomarkers for indoor mold exposure, resolving 21 analytes as potential occupational health risks (Tabbal et al., 2022b). A modified version of this method was employed to determine urine/air, blood/air, and plasma/air partition coefficients of microbial VOCs in relation to indoor mold exposure (Berkane et al., 2023). In a related study conducted by the Canadian Government, SPME-GC-MS/MS was used to monitor volatile halogenated and BTEX compounds in human blood to determine differences between indoor and outdoor exposure risks (Aranda-Rodriguez et al., 2015). A similar method using HS-GC-MS/MS profiled BTEX and other VOCs in sewage sludge samples from various wastewater treatment plants to better constrain odor control management practices (Byliński et al., 2019). From a biomarker discovery perspective, triple-bed needle trap micro-extraction with GC-MS/MS was used to profile VOCs in exhaled breath of patients with congestive heart failure to identify indicators of disease progression (Biagini et al., 2017; Bellagambi et al., 2021), while the use of a derivatizing agent such as pentafluorobenzyl hydroxylamine pre-loaded on Tenax GR sorbent tubes has been used in conjunction with TD-GC-MS/MS analysis to profile VOCs in exhaled breath from similar patients to monitor clinical improvement over time (Lomonaco et al., 2018).

2.2.3 Multi-dimensional GC (2DGC, GCxGC)

In complex matrices, analytes may persistently coelute even after chromatographic separation using long columns with advanced column chemistry. Efficiently separating multiple coeluting peaks and achieving positive compound identification using deconvolution algorithms can be difficult, especially when working with poorly characterized sample types. Multi-dimensional GC

attempts to address these challenges by separating analytes across different phases (Seeley and Seeley, 2013). For example, in two-dimensional GC (2DGC) effluent from the first-dimension column is diverted onto a second-dimension column with a different stationary phase for added separation. The two columns have separate detectors, generating distinct data files. Another option is comprehensive two-dimensional GC (GCxGC) that employs two ovens and columns featuring different stationary phases, along with a modulator for peak manipulation and a single detector (Giddings, 1984; Liu and Phillips, 1991). Modulation is achieved using rapid flow or thermal separation methods to improve analyte detection (Amaral et al., 2020), e.g., thermal modulation using GCxGC with dual-stage quad-jet thermal modulation and cryogenic cooling. During thermal modulation, coeluting peaks arriving from the first-dimension column undergo added separation by entering the modulator, where the peaks are segmented into slices, focused, injected, and subsequently separated in the second-dimension column, where the sliced sections of the peaks represent analytes resolved from previously coeluting peaks in the first dimension (Sarbach et al., 2013). TD-GCxGC-TOF-MS and PTR-TOF-MS have been used to establish standards for profiling VOCs in exhaled breath using the ReCIVA breath biopsy device including assessment of analyte loss resulting and false positive detection (Pham et al., 2023). Similarly, TD-GCxGC-TOF-MS was used to profile VOCs from the exhaled breath of febrile children infected with the malaria causing parasite *Plasmodium falciparum* resulting in the identification of six potential biomarkers associated with infection status (Schaber et al., 2018). Interestingly, these VOCS were related to terpenes known to attract mosquito vectors involved in malaria transmission. From an industrial perspective, tri-bed SPME-GCxGC-TOF-MS has been used to identify 241 VOCs including esters, alcohols, aldehydes, and alkenes involved in determining pear aroma that could be correlated with genetic differences between cultivars (Wang et al., 2019).

3 Conclusion

This review explores both on-line and off-line methods used in VOC profiling with MS from different matrices. Attention is given to the technical evolution of on-line and off-line methods needed for increasingly resolved VOC profiles, tracing technical progress over time with particular emphasis on emerging microbiome and diagnostic applications (summarized in Table 2). VOC profiling has grown expansively over the past 2 decades across multiple different platforms. It is expected that this trend will continue as more scientists and clinicians turn to increasingly sensitive MS detection platforms including different forms of GCxGC-TOF-MS and PTR-TOF-MS for development of diagnostic and monitoring solutions. At the same time emerging bioinformatics workflows enabling integration of multi-omic data sets (DNA, RNA, proteins, metabolites) promise to invigorate and inform volatilome research across increasingly diverse matrices (Figure 2). For example, Guo and colleagues developed an automated cuvette system in conjunction with on-line PTR-ToF-MS and off-line GC-MS to evaluate fungal VOCs from 43 individual fungal isolates and used the resulting spectra to identify patterns of covariation that informed a machine learning model for biomarker detection within higher level taxonomic groups or functional guilds (Guo et al., 2021).

On-line platforms offer benefits in clinical laboratories due to their quick analysis times, lack of sample preparation, user-friendly software interfaces, and portability to point-of-care (Majchrzak et al., 2018; Gould et al., 2021). Quantitation is possible for targeted analytes with prior calibration of the machine using standards (Španěl et al., 1999; Smith and Španěl, 2005; Fernández Maestre, 2012), while non-target analytes are eliminated from the measurement. SIFT-MS instruments provide the flexibility to choose from various precursor ions, including O_2^+, H_3O^+, and NO^+, which allows users to select the primary ionization agents most suitable for the analyzed matrix. In contrast, PTR-MS employs H_3O^+ as its sole ionizing agent, limiting its versatility but yielding less fragmentation (Smith and Španěl, 2005). On-line techniques are less ideal when conducting untargeted analysis of complex samples where overlapping *m/z* values can limit analyte identification (Kohl et al., 2013; Smith et al., 2014; Lubes and Goodarzi, 2017). Mass resolution can be improved using HRMS systems that reduce the error between accurate mass spectra and predicted mass listed in regulatory-compliant repositories, such as the National Institute of Standards and Technology (NIST) (Ausloos, et al., 1999), Wiley (McLafferty and Sttauffer, 1989), Fiehn (Kind et al., 2009), Golm (Kopka et al., 2004), or other reference libraries. Such repositories utilize information-rich databases and powerful search functions, such as vocBinBase (Skogerson et al., 2011), BinVestigate (Lai et al., 2018) with deconvolution and annotation tools, including MS-DIAL and MS-FINDER (Tsugawa et al., 2015; Tsugawa et al., 2016; Lai et al., 2018), as well as advanced database queries (Kind and Fiehn, 2006). For example, the Wiley Registry/NIST Mass Spectral Library was used to identify various metabolites including terpenoids, saponins, flavonoids, and alkaloids produced by six *Nigella sativa* (black cumin) species (Farag et al., 2014). In a separate integrative study, changes in gene expression and terpenoid production were used to annotate genes and pathways responsible for berry maturation processes (Wang et al., 2017). Despite this potential for multi-omics integration, in the absence of analyte separation even the most accurate mass spectral libraries are confounded by overlapping *m/z* values (Bruderer et al., 2020; Kaeslin et al., 2021).

Off-line platforms, although more labor intensive to operate, offer more detailed chromatographic separations supporting analyte identification, quantification, and structural elucidation (Ahmed et al., 2017; Gould et al., 2021). While development and validation of GC-MS/MS methods requires investment of time and expertise, they remain optimal for profiling VOCs across diverse matrices (Tabbal et al., 2022a; Tabbal et al., 2022b; Berkane et al., 2023). In the analysis of complex samples containing potentially hundreds of analytes, GCxGC-HRMS technology presents several advantages in resolution and analyte identification. The GCxGC module enhances chromatographic peak separation, potentially achieving baseline resolution, and the in-line HRMS system further improves selective spectral identification through high resolution mass measurement of the separated analytes. While low-resolution MS (LRMS) instruments typically measure *m/z* ratios to one decimal place, which suffices for targeted quantification, untargeted analyses require HRMS (Rey-Stolle et al., 2022). HRMS ensures measurement of unknowns with accurate mass where,

TABLE 2 Summary of advantages and disadvantages of on-line and off-line mass spectrometry techniques*.

MS platform	Analytical technique	Advantages	Disadvantages	References
On-line	Flowing afterglow	Pioneering technology for the quantitative measurement of ion reaction constants between charged and neutral species, leading to the development of SIFT-MS.	Charge transfer reactions of positive ions to neutrals can be measured more efficiently in certain cases with more direct methodologies	Ferguson et al. (1969)
	SIFT-MS	Easy to operate and maintain. No sample preparation. Results in minutes. Thermal ionization yields less fragmentation and simple mass spectra. Switching between reagent ions (i.e., O_2^+, H_3O^+, NO^+) can be optimized for best analyte detection. Sensitive and more selective than PRT-MS.	Cannot distinguish between isobars. Quantitation is possible only after calibration with known analytes. Bulky pumping systems are required to maintain separate vacuum regions of the instrument. Requires high purity helium or hydrogen as carrier gas	Prince et al. (2010)
	IMS-MS	Fast, portable, can detect trace amounts of analytes. Separation is based on collisional cross section and separates isomers. High throughput. Suitable for field applications, such as environmental, food, or homeland security samples. Result available in minutes	Sensitive to temperature and humidity changes. Compound resolution efficiency decreases with temperature and humidity that degrade samples. Not suitable for non-volatile analytes. Complex mass spectra, therefore, off-line techniques may be needed to confirm results	Lapthorn et al. (2013)
	IMR-MS	Low-energy chemical ionization. Binary collision process. Yields simple mass spectra with minimal fragmentation and molecular ions observed. Highly sensitive; capable of detecting gaseous compounds at the pptv level, in a few minutes	Selection of primary ion is important to minimize fragmentation when analyzing neutral gases. Complex gas mixtures (e.g., emission from furnaces, motors) require switching between several primary ions, adding analysis time. Differential pumping requires bulky pumping systems	Dolch et al. (2012a), Dolch et al. (2012b)
	PTR-MS	Portable, high throughput with no previous sample preparation or compound separation. User friendly, less expensive than SIFT-MS. Analyte quantitation is possible at pptv levels. Can complement GC-MS methods	Higher collision energies result in fragmentation and inconsistent product ion formation. The availability of H_3O^+ reagent ions only, limits the detection of light hydrocarbons and certain halogenated species	Hansel et al. (1995)
	SESI-MS	Very efficient ionization at atmospheric pressure yields high sensitivity at pptv level. Minimal fragmentation with the molecular ion observed. Easy to interface with commercial ESI MS systems. Suitable for complex biological samples	Vulnerable to matrix-effects causing ion suppression. Not suitable for quantitation. Expensive instrumentation. A nanoflow technique that may result in long sample analysis times	Wüthrich and Giannoukos (2024)
	TOF-MS	A HRMS technique with up to 48 m flight path and 200,000 FWHM mass resolution, at ppb level, across all acquisition speeds. No mass discrimination. Full mass spectra acquired in milliseconds. Can be interfaced with LC, GC.	Duty cycle limited between 1% and 30%. Sensitivity drops with increasing mass resolution. Flight tube sensitive to temperature fluctuation that affects mass accuracy. Not ideal for quantitation	Cooper-Shepherd et al. (2023)

(Continued on the following page)

TABLE 2 (*Continued*) Summary of advantages and disadvantages of on-line and off-line mass spectrometry techniques*.

MS platform	Analytical technique	Advantages	Disadvantages	References
	Orbitrap-MS	Most current HRMS technique with up to 240,000 FWHM mass resolution at *m/z* 400 with sub-ppm accuracy. Internal lock mass calibration with <1 ppm drift over 24 h. High sensitivity. Can be interfaced with off-line techniques	Space charging may limit the dynamic range and sensitivity of low abundance ions. Slow polarity switching in certain models. Increasing scan rates reduce resolution	Zubarev and Makarov (2013)
Off-line	GC-MS	Robust, user-friendly, affordable. High chromatographic resolution with unit mass resolution. Wide linear range. Suitable for quantitation of small molecules in liquid or gaseous samples in diverse matrices. EI detects analytes that are difficult to ionize	Analytes must be easily volatilized. Molecular ion not observed. In convoluted mass spectra, analyte identification may be compromised. Not ideal for isobars and untargeted analysis. Ion source may require frequent cleaning	Lubes and Goodarzi (2017)
	GC-MS/MS	Most sensitive and selective technique for quantification of targeted analytes in complex matrices. Wide linear analytical range; suitable for analytical measurements at sub ppb levels	Upstream sample preparation necessary. Low mass resolution. More expensive than GC-MS. Larger space requirements. Room ventilation may be necessary due to increased heat generation	Tabbal et al. (2022a), Tabbal et al. (2022b)
	GC-TOF-MS	High chromatographic resolution combined with high mass resolution. Simultaneous detection of all ions with accurate mass. Information-rich data acquisition, suitable for complex samples with no loss of data	Mass resolution range is between 30,0000 and 50,000 FWHM with 1–2 ppm mass accuracy. Frequent calibration is necessary. Highly concentrated samples affect accurate mass determinations	Chernushevich et al. (2017)
	GC-Orbitrap MS	Unparalleled mass resolution. Targeted (quantitative) and untargeted (qualitative) analysis, simultaneously. Excellent sensitivity. Compact design. Applications in proteomics, metabolomics, diagnostics, environmental, food, etc., analyses	Limited charge capacity can overfill Orbitrap, reducing mass accuracy. Detection may be affected by ion stability due to conditions in ion injection slit, electrodynamic squeezing, rotational amplitudes, axial oscillation, etc. Expensive instrumentation	Hu et al. (2005), Belarbi et al. (2021)
	GCxGC-HR TOF-MS	Very high peak capacity. Unmatched chromatographic resolution coupled with high mass resolution. FFP design combined with EFP technology improves mass resolution and duty cycle, significantly. Represent the most current GC/MS technology for VOC analysis	Upstream sample preparation is necessary. Very long sample analysis times. Multiple decoding algorithms needed for complex mixtures. Large footprint. Expensive technology	Willis et al. (2021)

*For abbreviations, refer to Supplementary Material.

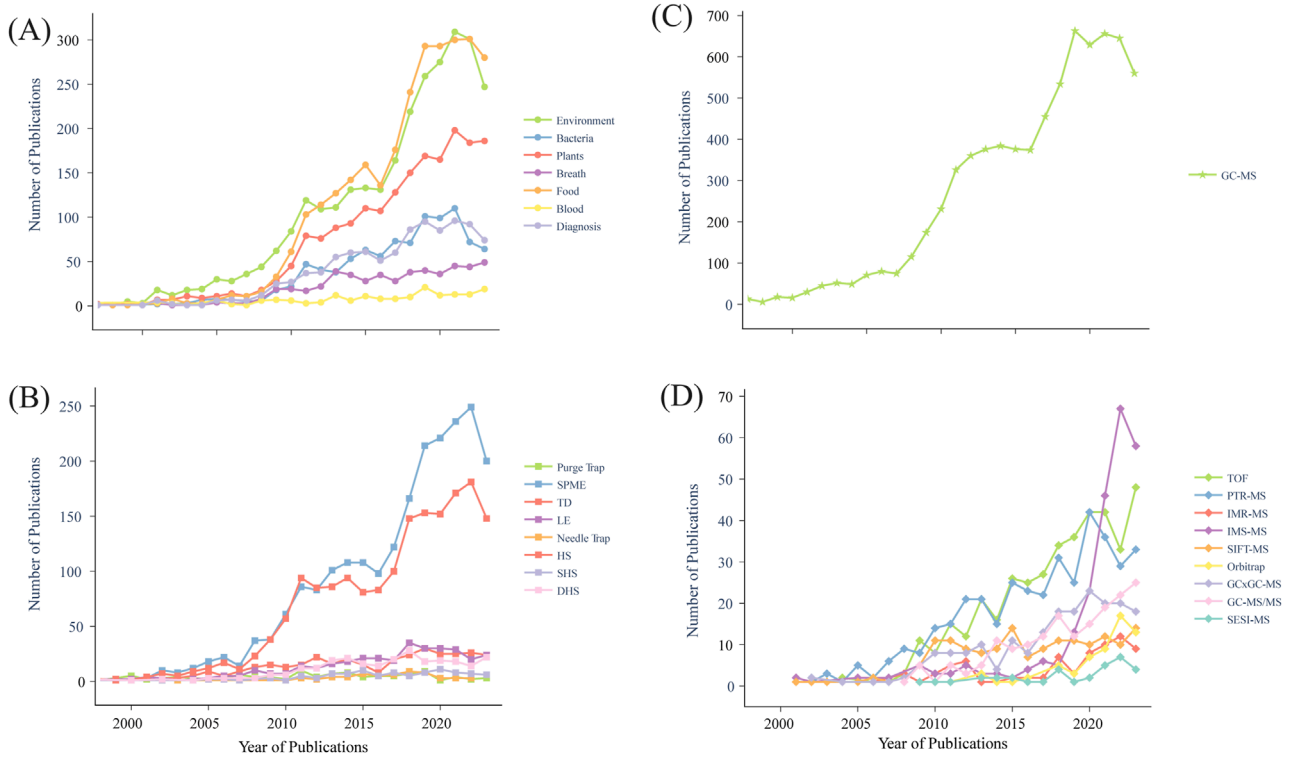


FIGURE 2
Increasing number of publications related to volatile organic compound detection over the past 20 years for different matrices, methods, and platforms based on PubMed searches. **(A)** various matrices, **(B)** extraction methods (SPME, solid phase microextraction; TD, thermal desorption; LE, liquid extraction; HS, headspace; SHS, static headspace; DHS, dynamic headspace), **(C)** GC-MS (gas chromatography mass spectrometry), **(D)** other mass spectrometry (MS) platforms (TOF, time-of-flight; PTR-MS, proton transfer reaction MS; IMR-MS, ion molecule reaction MS; IMS-MS, ion mobility spectrometry MS; SIFT-MS, selected ion flow tube MS; GCxGC-MS, comprehensive two-dimensional gas chromatography MS; GC-MS/MS, gas chromatography triple quadrupole MS; SESI-MS, secondary electrospray ionization MS).

TABLE 3 Determination of compounds, *m/z* 28.0 with accurate mass and the possible chemical formulas.

Element	Atomic mass unit	Chemical formula	Nominal mass	Theoretical exact mass	Measured accurate mass	Mass error
		m/z	*m/z*	*m/z*	*m/z*	ppm
H	1.00783	N_2	28.0	28.00559	28.00562	1.07
C	12.00000	CO	28.0	27.99436	27.99429	−2.54
N	14.00307	CH_2N	28.0	28.01818	28.01827	3.03
O	15.99491	C_2H_4	28.0	28.03077	28.03089	4.28
electron	0.00055					

*$Mass\,error = \frac{mass_{measured} - mass_{theoretical}}{mass_{measured}} * 10^6\,[ppm]$.

for example, compounds with nominal mass *m/z* 28.0 have four possible chemical formulas, as shown in Table 3. LRMS is incapable of distinguishing between these possibilities. HRMS compares the measured accurate mass spectra with predicted mass listed in regulatory-compliant repositories and the analyte with the lowest mass error is selected as the most likely hit; in this case *m/z* 28.00562 (Table 3). The chromatographic separation power and accurate mass measurement of GCxGC-HRMS places these platforms at the forefront of VOC detection with the best analyte separation and accurate mass identification (Weggler et al., 2019; Willis et al., 2021; Cooper-Shepherd et al., 2023; LECO Corporation White paper, 2021).

Living cells and cell systems produce diverse VOC profiles that can be differentially detected using on-line and off-line

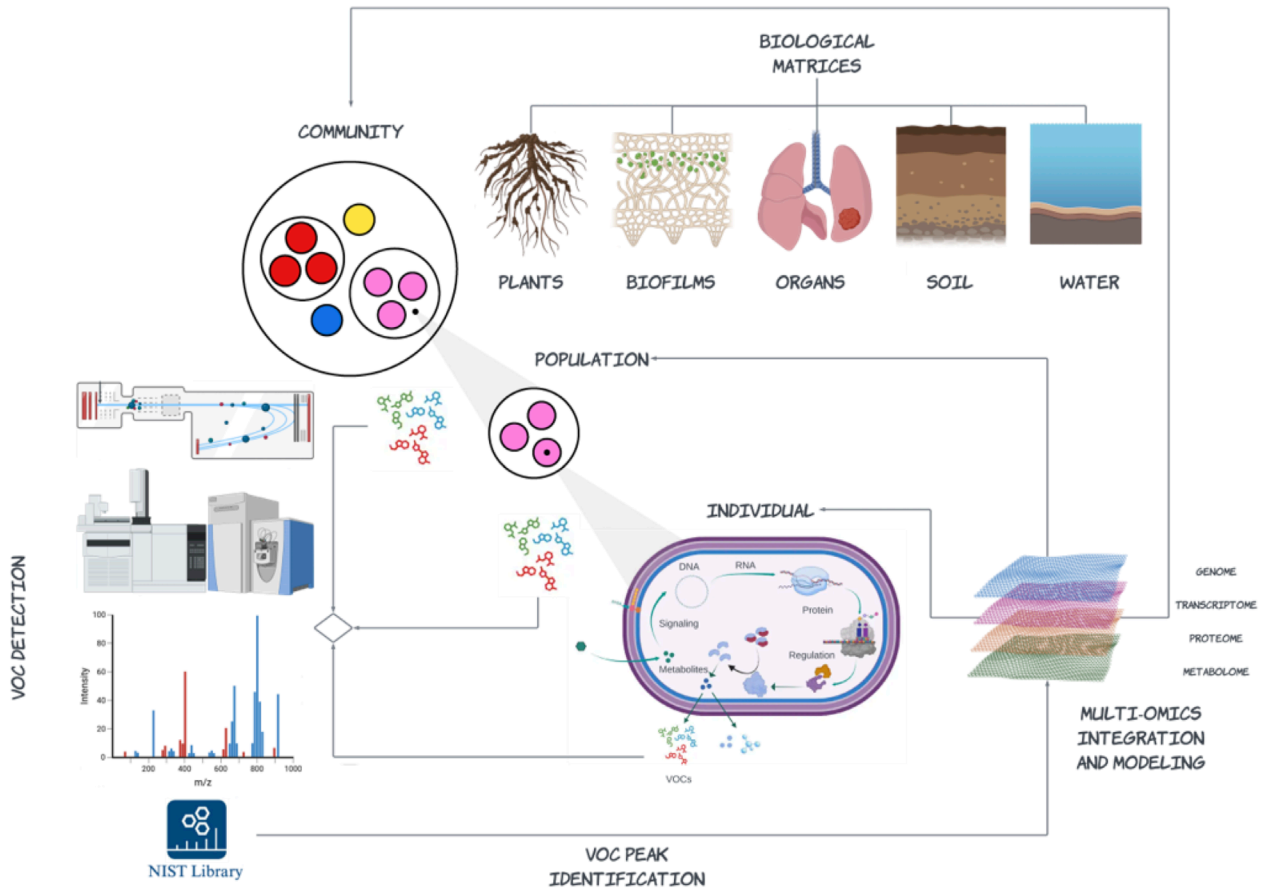


FIGURE 3
An emerging paradigm for volatilome research integrating multi-omic data (DNA, RNA, protein, and metabolites) spanning different levels of biological organization to resolve volatile organic compound (VOC) profiles emerging from biological matrices sourced from natural or engineered environments including our own bodies. This paradigm includes enrichment and isolation methods to more closely link specific VOCs with specific individuals or populations of microorganisms as well as community-level analysis that define emergent patterns of VOC production and signaling among and between different taxonomic lineages. The development of more extensive regulatory-compliant repositories, e.g., Wiley Registry/National Institute of Standards and Technology (NIST) Mass Spectral Library compound reference libraries and integrative methods linking identified VOCs onto the background metabolic network at scale remain ongoing challenges (Created with BioRender.com).

methods. For example, VOCs produced by plants include terpenoids, phenylpropanoids, benzenoids, and fatty acid derivatives (Picazo-Aragonés et al., 2020; Liu et al., 2023); VOCs detected in blood include alcohols, aldehydes, acids, acetone, hydrogen sulfide, methanethiol, dimethyl sulfide, dimethyldisulfide, trimethylamine, indole, aminoacetophenone (Allardyce et al., 2006a; Allardyce et al., 2006b); VOCs detected in breath include hydrocarbons, alcohols, ketones, aldehydes, carboxylic acids, esters, isoprenoids, furan, nitrogen- and sulfur-containing compounds, aromatics, cyclic hydrocarbons (Issitt et al., 2022; Moura et al., 2023) (https://neomeditec.com/VOCdatabase/); VOCs produced by microorganisms include hydrocarbons, alcohols, aldehydes, acids, ketones, esters, aromatics, phenols, nitrogen- and sulfur-containing compounds (Ratiu et al., 2017); VOCs detected in food include alcohols, aldehydes, acids, esters, terpenes, furans, and pyrazines (Starowicz, 2021); and VOCs detected in environmental samples include alkanes, alkenes, alkynes, alcohols, aldehydes, aromatic compounds, ketones, esters, ethers, haloalkanes, nitriles, organic acids, and acrylamide (Lü et al., 2022). While many of these compound classes are shared between sources, more granular analysis reveals a complex array of molecular forms with potential to serve specific signaling or regulatory roles.

In this regard there is increasing interest in VOC profiling to develop new metrics that improve understanding of microbial interactions and trait-based contributions to functions and services in natural and engineered environments including our own bodies (Figure 3). For example, microbial VOCs are increasingly being linked back to antimicrobial properties and plant host interactions including defense mechanisms and root growth (Weisskopf et al., 2021; Razo-Belmán et al., 2023; Schmidt et al., 2023). From a diagnostic or biotechnology innovation perspective this not only applies to the detection of VOCs involved in community-level interactions and host associations, but to development of non-invasive point of care diagnostics in health and disease. For example, lung cancer breath analysis has become a promising method of screening, with ~500 exhaled compounds currently associated with lung cancer status (Schmidt et al., 2023). Identifying

correlations between lung microbiome and VOCs in relation to lung cancer status is an active area of research that requires increased throughput, mass resolution and methods standardization. These requirements also represent common challenges to scaling VOC detection in relation to community level interactions as well as in development of screening paradigms to recover genes or gene cassettes producing VOCs from environmental genomes. Despite these challenges, the trajectory of MS platform innovation and continuous improvement in integrative methods of data analysis and statistical modeling provides an exciting opportunity for researchers to ask fundamental questions with real world implications.

4 Scope statement

Volatile organic compounds (VOCs) are gas-phase small molecules released from biotic and abiotic matrices into the environment. Because they are volatile, VOCs are typically not detected during conventional metabolite analysis and require specific profiling methods and platforms to measure. One of the most popular platforms is gas chromatography-mass spectrometry that is commonly used for detecting VOCs and can be automated with headspace solid-phase microextraction, etc., methods. The study of VOCs in relation to other forms of biological information, e.g., DNA, RNA, protein, and other metabolites encompasses volatilome research. Emerging lines of evidence suggest that the volatilome plays an integral role in signaling and metabolite exchange within natural and engineered environments including our own bodies where the interplay between microorganisms and host cells defines a complex adaptive network. In this review we trace the evolution of mass spectrometry platforms used in the detection of VOCs in relation to different biological matrices and provide contemporary insight into how VOC profiling is becoming increasingly used to develop non-invasive diagnostic tests across a range of application areas in health, industry, and the environment. Since this review examines VOC analysis using various mass spectrometry platforms with a multi-omics approach, it will fit well in this Research Topic.

Author contributions

ASz: Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–original draft, Writing–review and editing. ASu: Data curation, Investigation, Visualization, Writing–original draft, Writing–review and editing. SH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was performed under the auspices of the Natural Sciences and Engineering Research Council (NSERC) of Canada, Genome British Columbia, the Terry Fox New Frontiers Program, and the Canada Foundation for Innovation (CFI).

Acknowledgments

We would like to thank Cody Foley in the Biofactorial automation core facility at UBC, and Jasmeen Kaur for insightful conversations regarding multi-omic approaches and VOC detection, as well as Tony Liu and Juan Camilo Santana Martinez in the Hallam lab for useful discussions related to integrating VOC profiling with multi-omic data sets.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2024.1421330/full#supplementary-material

References

Adams, N. G., and Smith, D. (1976). The selected ion flow tube (SIFT); A technique for studying ion-neutral reactions. *Int. J. Mass Spectrom. Ion. Phys.* 21 (3), 349–359. doi:10.1016/0020-7381(76)80133-7

Ahmed, W. M., Lawal, O., Nijsen, T. M., Goodacre, R., and Fowler, S. J. (2017). Exhaled volatile organic compounds of infection: a systematic review. *ACS Infect. Dis.* 3 (10), 695–710. doi:10.1021/acsinfecdis.7b00088

Allardyce, R. A., Hill, A. L., and Murdoch, D. R. (2006a). The rapid evaluation of bacterial growth and antibiotic susceptibility in blood cultures by selected ion flow tube mass spectrometry. *Diagn. Microbiol. Infect. Dis.* 55 (4), 255–261. doi:10.1016/j.diagmicrobio.2006.01.031

Allardyce, R. A., Langford, V. S., Hill, A. L., and Murdoch, D. R. (2006b). Detection of volatile metabolites produced by bacterial growth in blood culture media by selected ion flow tube mass spectrometry (SIFT-MS). *J. Microbiol. Methods* 65 (2), 361–365. doi:10.1016/j.mimet.2005.09.003

Amann, A., Costello Bde, L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., et al. (2014). The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath. Res.* 8 (3), 034001. doi:10.1088/1752-7155/8/3/034001

Amaral, M. S. S., Nolvachai, Y., and Marriott, P. J. (2020). Comprehensive two-dimensional gas chromatography advances in technology and applications: biennial update. *Anal. Chem.* 92 (1), 85–104. doi:10.1021/acs.analchem.9b05412

Anishchenko, I. M., McCartney, M. M., Fung, A. G., Peirano, D. J., Schirle, M. J., Kenyon, N. J., et al. (2018). Modular and reconfigurable gas chromatography/differential mobility spectrometry (GC/DMS) package for detection of volatile organic compounds (VOCs). *Int. J. Ion. Mobil. Spectrom.* 21 (4), 125–136. doi:10.1007/s12127-018-0240-4

Aranda-Rodriguez, R., Cabecinha, A., Harvie, J., Jin, Z., Marchand, A., Tardif, R., et al. (2015). A method for quantification of volatile organic compounds in blood by SPME-GC–MS/MS with broader application: from non-occupational exposure population to exposure studies. *J. Chromatogr. B* 992, 76–85. doi:10.1016/j.jchromb.2015.04.020

Audrain, B., Farag, M. A., Ryu, C.-M., and Ghigo, J.-M. (2015). Role of bacterial volatile compounds in bacterial biology. *FEMS Microbiol. Rev.* 39 (2), 222–233. doi:10.1093/femsre/fuu013

Ausloos, P., Clifton, C. L., Lias, S. G., Mikaya, A. I., Stein, S. E., Tchekhovskoi, D. V., et al. (1999). The critical evaluation of a comprehensive mass spectral library. *J. Am. Soc. Mass Spectrom.* 10 (4), 287–299. doi:10.1016/S1044-0305(98)00159-7

Barrios-Collado, C., García-Gómez, D., Zenobi, R., Vidal-de-Miguel, G., Ibáñez, A. J., and Martinez-Lozano Sinues, P. (2016). Capturing *in vivo* plant metabolism by real-time analysis of low to high molecular weight volatiles. *Anal. Chem.* 88 (4), 2406–2412. doi:10.1021/acs.analchem.5b04452

Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. V., Jarmusch, A. K., and Dorrestein, P. C. (2022). Mass spectrometry-based metabolomics in microbiome investigations. *Nat. Rev. Microbiol.* 20 (3), 143–160. doi:10.1038/s41579-021-00621-9

Belarbi, S., Vivier, M., Zaghouani, W., Sloovere, A. D., Agasse-Peulon, V., and Cardinael, P. (2021). Comparison of new approach of GC-HRMS (Q-Orbitrap) to GC–MS/MS (triple-quadrupole) in analyzing the pesticide residues and contaminants in complex food matrices. *Food Chem.* 359, 129932. doi:10.1016/j.foodchem.2021.129932

Belizário, J. E., Faintuch, J., and Malpartida, M. G. (2021). Breath biopsy and discovery of exclusive volatile organic compounds for diagnosis of infectious diseases. *Front. Cell. Infect. Microbiol.* 10. doi:10.3389/fcimb.2020.564194

Bellagambi, F. G., Lomonaco, T., Ghimenti, S., Biagini, D., Fuoco, R., and Francesco, F. D. (2021). Determination of peppermint compounds in breath by needle trap micro-extraction coupled with gas chromatography–tandem mass spectrometry. *J. Breath. Res.* 15 (1), 016014. doi:10.1088/1752-7163/abcdec

Berenguer, C. V., Pereira, F., Pereira, J. A. M., and Câmara, J. S. (2022). Volatilomics: an emerging and promising avenue for the detection of potential prostate cancer biomarkers. *Cancers (Basel)* 14 (16), 3982. doi:10.3390/cancers14163982

Berkane, W., El Aroussi, B., Bouchard, M., Marchand, G., and Haddad, S. (2023). Determination of blood:air, urine:air and plasma:air partition coefficients of selected microbial volatile organic compounds. *Chemosphere* 343, 140305. doi:10.1016/j.chemosphere.2023.140305

Biagini, D., Lomonaco, T., Ghimenti, S., Bellagambi, F. G., Onor, M., Scali, M. C., et al. (2017). Determination of volatile organic compounds in exhaled breath of heart failure patients by needle trap micro-extraction coupled with gas chromatography-tandem mass spectrometry. *J. Breath. Res.* 11 (4), 047110. doi:10.1088/1752-7163/aa94e7

Bierbaum, V. M. (2015). Go with the flow: fifty years of innovation and ion chemistry using the flowing afterglow. *Int. J. Mass Spectrom.* 377, 456–466. doi:10.1016/j.ijms.2014.07.021

Blanco, F. G., and Vidal-de-Miguel, G. (2021). Breath analysis by secondary electro-spray ionization - mass spectrometry to interrogate biologically significant metabolites non-invasively. *Crit. Rev. Anal. Chem.* 53, 825–837. doi:10.1080/10408347.2021.1981226

Boots, A. W., Smolinska, A., van Berkel, J. J. B. N., Fijten, R. R. R., Stobberingh, E. E., Boumans, M. L. L., et al. (2014). Identification of microorganisms based on headspace analysis of volatile organic compounds by gas chromatography–mass spectrometry. *J. Breath. Res.* 8 (2), 027106. doi:10.1088/1752-7155/8/2/027106

Bruderer, T., Gaugg, M. T., Cappellin, L., Lopez-Hilfiker, F., Hutterli, M., Perkins, N., et al. (2020). Detection of volatile organic compounds with secondary electrospray ionization and proton transfer reaction high-resolution mass spectrometry: a feature comparison. *J. Am. Soc. Mass Spectrom.* 31 (8), 1632–1640. doi:10.1021/jasms.0c00059

Byliński, H., Aszyk, J., Kubica, P., Szopińska, M., Fudala-Książek, S., and Namieśnik, J. (2019). Differences between selected volatile aromatic compound concentrations in sludge samples in various steps of wastewater treatment plant operations. *J. Environ. Manage.* 249, 109426. doi:10.1016/j.jenvman.2019.109426

Cao, X.-L., Sparling, M., and Dabeka, R. (2016). Occurrence of 13 volatile organic compounds in foods from the Canadian total diet study. *Food Addit. Contam. Part A* 33 (2), 373–382. doi:10.1080/19440049.2015.1129072

Carraturo, F., Libralato, G., Esposito, R., Galdiero, E., Aliberti, F., Amoresano, A., et al. (2020). Metabolomic profiling of food matrices: preliminary identification of potential markers of microbial contamination. *J. Food Sci.* 85 (10), 3467–3477. doi:10.1111/1750-3841.15418

Chen, C., Huang, K., Yu, H., and Tian, H. (2021). The diversity of microbial communities in Chinese milk fan and their effects on volatile organic compound profiles. *J. Dairy Sci.* 104 (3), 2581–2593. doi:10.3168/jds.2020-19053

Chernushevich, I. V., Merenbloom, S. I., Liu, S., and Bloomfield, N. (2017). A W-geometry ortho-TOF MS with high resolution and up to 100% duty cycle for MS/MS. *J. Am. Soc. Mass Spectrom.* 28 (10), 2143–2150. doi:10.1007/s13361-017-1742-8

Chernushevich, I. V., and Thomson, B. A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. *Anal. Chem.* 76 (6), 1754–1760. doi:10.1021/ac035406j

Choudoir, M., Rossabi, S., Gebert, M., Helmig, D., and Fierer, N. (2019). A phylogenetic and functional perspective on volatile organic compound production by Actinobacteria. *mSystems* 4 (2). doi:10.1128/mSystems.00295-18

Clark, A. E., Kaleta, E. J., Arora, A., and Wolk, D. M. (2013). Matrix-assisted laser desorption ionization–time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* 26(3), 547–603. doi:10.1128/cmr.00072-12

Cohen, M. J., and Karasek, F. W. (1970). Plasma chromatography™—a new dimension for gas chromatography and mass spectrometry. *J. Chromatogr. Sci.* 8 (6), 330–337. doi:10.1093/chromsci/8.6.330

Collins, D., and Lee, M. (2002). Developments in ion mobility spectrometry–mass spectrometry. *Anal. Bioanal. Chem.* 372 (1), 66–73. doi:10.1007/s00216-001-1195-5

Cooper-Shepherd, D. A., Wildgoose, J., Kozlov, B., Johnson, W. J., Tyldesley-Worster, R., Palmer, M. E., et al. (2023). Novel hybrid quadrupole-multireflecting time-of-flight mass spectrometry system. *J. Am. Soc. Mass Spectrom.* 34 (2), 264–272. doi:10.1021/jasms.2c00281

Cotter, R. J. (1989). Time-of-flight mass spectrometry: an increasing role in the life sciences. *Biomed. Environ. Mass Spectrom.* 18 (8), 513–532. doi:10.1002/bms.1200180803

D'Alessandro, M., and Turlings, T. C. (2006). Advances and challenges in the identification of volatiles that mediate interactions among plants and arthropods. *Analyst* 131 (1), 24–32. doi:10.1039/b507589k

Dhandapani, S., Jin, J., Sridhar, V., Sarojam, R., Chua, N.-H., and Jang, I.-C. (2017). Integrated metabolome and transcriptome analysis of Magnolia champaca identifies biosynthetic pathways for floral volatile organic compounds. *BMC Genomics* 18 (1), 463. doi:10.1186/s12864-017-3846-8

Dixon, B., Ahmed, W. M., Felton, T., and Fowler, S. J. (2022). Molecular phenotyping approaches for the detection and monitoring of carbapenem-resistant Enterobacteriaceae by mass spectrometry. *J. Mass Spectrom. Adv. Clin. Lab.* 26, 9–19. doi:10.1016/j.jmsacl.2022.09.001

Dodds, J. N., and Baker, E. S. (2019). Ion mobility spectrometry: fundamental concepts, instrumentation, applications, and the road ahead. *J. Am. Soc. Mass Spectrom.* 30 (11), 2185–2195. doi:10.1007/s13361-019-02288-2

Dodds, J. N., Hopkins, Z. R., Knappe, D. R. U., and Baker, E. S. (2020). Rapid characterization of per- and polyfluoroalkyl substances (PFAS) by ion mobility spectrometry–mass spectrometry (IMS-MS). *Anal. Chem.* 92 (6), 4427–4435. doi:10.1021/acs.analchem.9b05364

Dodonov, A. F., Chernushevich, I. V., and Laiko, V. V. (1993). "Electrospray ionization on a reflecting time-of-flight mass spectrometer," in *Time-of-Flight mass spectrometry* (American Chemical Society), 108–123.

Dodonov, A. F., Chernushevich, I. V., Dodonova, T. F., Raznikov, V. V., and Talrose, V. L. (1987). *Mass reflectron*. Moscow, USSR, Soviet Patent no. 1681340A1 (filed February 1987).

Dodonov, A. F., Kozlovski, V. I., Soulimenkov, I. V., Raznikov, V. V., Loboda, A. V., Zhen, Z., et al. (2000). High-resolution electrospray ionization orthogonal-injection time-of-flight mass spectrometer. *Eur. J. Mass Spectrom.* 6 (6), 481–490. doi:10.1255/ejms.378

Dolch, M. E., Choukèr, A., Hornuss, C., Frey, L., Irlbeck, M., Praun, S., et al. (2015). Quantification of propionaldehyde in breath of patients after lung transplantation. *Free Radic. Biol. Med.* 85, 157–164. doi:10.1016/j.freeradbiomed.2015.04.003

Dolch, M. E., Hornuss, C., Klocke, C., Praun, S., Villinger, J., Denzer, W., et al. (2012a). Volatile compound profiling for the identification of Gram-negative bacteria by ion-molecule reaction–mass spectrometry. *J. Appl. Microbiol.* 113 (5), 1097–1105. doi:10.1111/j.1365-2672.2012.05414.x

Dolch, M. E., Hornuss, C., Klocke, C., Praun, S., Villinger, J., Denzer, W., et al. (2012b). Volatile organic compound analysis by ion molecule reaction mass spectrometry for

Gram-positive bacteria differentiation. *Eur. J. Clin. Microbiol. Infect. Dis.* 31 (11), 3007–3013. doi:10.1007/s10096-012-1654-2

Douglas, D. J., and French, J. B. (1992). Collisional focusing effects in radio frequency quadrupoles. *J. Am. Soc. Mass Spectrom.* 3 (4), 398–408. doi:10.1016/1044-0305(92)87067-9

Drabińska, N., Flynn, C., Ratcliffe, N., Belluomo, I., Myridakis, A., Gould, O., et al. (2021). A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *J. Breath. Res.* 15 (3), 034001. doi:10.1088/1752-7163/abf1d0

Drabińska, N., Hewett, K., White, P., Avison, M. B., Persad, R., Ratcliffe, N. M., et al. (2022). Application of a solid-phase microextraction-gas chromatography-mass spectrometry/metal oxide sensor system for detection of antibiotic susceptibility in urinary tract infection-causing *Escherichia coli* – a proof of principle study. *Adv. Med. Sci.* 67 (1), 1–9. doi:10.1016/j.advms.2021.09.001

Drees, C., Vautz, W., Liedtke, S., Rosin, C., Althoff, K., Lippmann, M., et al. (2019). GC-IMS headspace analyses allow early recognition of bacterial growth and rapid pathogen differentiation in standard blood cultures. *Appl. Microbiol. Biotechnol.* 103 (21-22), 9091–9101. doi:10.1007/s00253-019-10181-x

Dudareva, N., Pichersky, E., and Gershenzon, J. (2004). Biochemistry of plant volatiles. *Plant Physiol.* 135 (4), 1893–1902. doi:10.1104/pp.104.049981

Farag, M. A., Gad, H. A., Heiss, A. G., and Wessjohann, L. A. (2014). Metabolomics driven analysis of six Nigella species seeds via UPLC-qTOF-MS and GC–MS coupled to chemometrics. *Food Chem.* 151, 333–342. doi:10.1016/j.foodchem.2013.11.032

Fenn, J. B. (2002). Electrospray ionization mass spectrometry: how it all began. *J. Biomol. Tech.* 13 (3), 101–118.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246 (4926), 64–71. doi:10.1126/science.2675315

Ferguson, E. E., Fehsenfeld, F. C., and Schmeltekopf, A. L. (1969). "Flowing afterglow measurements of ion-neutral reactions," in *Advances in atomic and molecular physics*. Editors D. R. Bates, and I. Estermann (Academic Press), 1–56.

Fernández Maestre, R. (2012). Ion mobility spectrometry: history, characteristics and applications. *Rev. U.D.C.A Act. and Div. Cient.* 15 (2), 467–479. doi:10.31910/rudca.v15.n2.2012.848

Ferrandino, G., De Palo, G., Murgia, A., Birch, O., Tawfike, A., Smith, R., et al. (2023). Breath Biopsy® to identify exhaled volatile organic compounds biomarkers for liver cirrhosis detection. *J. Clin. Transl. Hepatol.* 11 (3), 000–648. doi:10.14218/jcth.2022.00309

Finnigan, R. (2016). "Development of the business of mass spectrometry (1960–75)," in *The encyclopedia of mass spectrometry*. Editors M. L. Gross, and R. M. Caprioli (Boston: Elsevier), 315–317.

Fitzgerald, S., Furlong, C., Holland, L., and Morrin, A. (2022). Multi-strain and -species investigation of volatile metabolites emitted from planktonic and biofilm Candida cultures. *Metabolites* 12 (5), 432. doi:10.3390/metabo12050432

Fu, L., Wang, L., Wang, H., Yang, M., Yang, Q., Lin, Y., et al. (2023). A cross-sectional study: a breathomics based pulmonary tuberculosis detection method. *BMC Infect. Dis.* 23 (1), 148. doi:10.1186/s12879-023-08112-3

Giddings, J. C. (1984). Two-dimensional separations: concept and promise. *Anal. Chem.* 56 (12), 1258A-1260A, 1262A, 1264A passim–1260A. 1262A, 1264A passim. doi:10.1021/ac00276a003

González, M. L., Carnicero, M., de la Torre, R., Ortuño, J., and Segura, J. (1995). Influence of the injection technique on the thermal degradation of cocaine and its metabolites in gas chromatography. *J. Chromatogr. B Biomed. Sci. Appl.* 664 (2), 317–327. doi:10.1016/0378-4347(94)00484-m

Gould, O., Drabińska, N., Ratcliffe, N., and de Lacy Costello, B. (2021). Hyphenated mass spectrometry versus real-time mass spectrometry techniques for the detection of volatile compounds from the human body. *Molecules* 26 (23), 7185. doi:10.3390/molecules26237185

Grayson, M. A. (2016). "A history of gas chromatography mass spectrometry (GC/MS)," in *The encyclopedia of mass spectrometry*. Editors M. L. Gross, and R. M. Caprioli (Boston: Elsevier), 152–158.

Guilhaus, M., Selby, D., and Mlynski, V. (2000). Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom. Rev.* 19 (2), 65–107. doi:10.1002/(SICI)1098-2787(2000)19:2<65::AID-MAS1>3.0.CO;2-E

Guo, Y., Jud, W., Weikl, F., Ghirardo, A., Junker, R. R., Polle, A., et al. (2021). Volatile organic compound patterns predict fungal trophic mode and lifestyle. *Commun. Biol.* 4 (1), 673. doi:10.1038/s42003-021-02198-8

Handa, H., Usuba, A., Maddula, S., Baumbach, J. I., Mineshita, M., and Miyazawa, T. (2014). Exhaled breath analysis for lung cancer detection using ion mobility spectrometry. *PLoS One* 9 (12), e114555. doi:10.1371/journal.pone.0114555

Hansel, A., Jordan, A., Holzinger, R., Prazeller, P., Vogel, W., and Lindinger, W. (1995). Proton transfer reaction mass spectrometry: on-line trace gas analysis at the ppb level. *Int. J. Mass Spectrom. Ion. Process.* 149-150, 609–619. doi:10.1016/0168-1176(95)04294-U

Henderson, B., Ruszkiewicz, D. M., Wilkinson, M., Beauchamp, J. D., Cristescu, S. M., Fowler, S. J., et al. (2020). A benchmarking protocol for breath analysis: the peppermint experiment. *J. Breath. Res.* 14 (4), 046008. doi:10.1088/1752-7163/aba130

Higashikawa, F. S., Cayuela, M. L., Roig, A., Silva, C. A., and Sánchez-Monedero, M. A. (2013). Matrix effect on the performance of headspace solid phase microextraction method for the analysis of target volatile organic compounds (VOCs) in environmental samples. *Chemosphere* 93 (10), 2311–2318. doi:10.1016/j.chemosphere.2013.08.023

Hill, H. H., Jr., Siems, W. F., and St. Louis, R. H. (1990). Ion mobility spectrometry. *Anal. Chem.* 62 (23), 1201A–1209A. doi:10.1021/ac00222a001

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 40 (4), 430–443. doi:10.1002/jms.856

Issitt, T., Wiggins, L., Veysey, M., Sweeney, S. T., Brackenbury, W. J., and Redeker, K. (2022). Volatile compounds in human breath: critical review and meta-analysis. *J. Breath. Res.* 16 (2), 024001. doi:10.1088/1752-7163/ac5230

James, A. T., and Martin, A. J. P. (1952). Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *Biochem. J.* 50 (5), 679–690. doi:10.1042/bj0500679

Janssens, E., van Meerbeeck, J. P., and Lamote, K. (2020). Volatile organic compounds in human matrices as lung cancer biomarkers: a systematic review. *Crit. Rev. Oncol. Hematol.* 153, 103037. doi:10.1016/j.critrevonc.2020.103037

Javelle, T., Righezza, M., and Danger, G. (2021). Identify low mass volatile organic compounds from cometary ice analogs using gas chromatography coupled to an Orbitrap mass spectrometer associated to electron and chemical ionizations. *J. Chromatogr. A* 1652, 462343. doi:10.1016/j.chroma.2021.462343

John, T. M., Shrestha, N. K., Procop, G. W., Grove, D., Leal, S. M., Jr., Jacob, C. N., et al. (2021). Diagnosis of Clostridioides difficile infection by analysis of volatile organic compounds in breath, plasma, and stool: a cross-sectional proof-of-concept study. *PLoS One* 16 (8), e0256259. doi:10.1371/journal.pone.0256259

Kaeslin, J., Micic, S., Weber, R., Müller, S., Perkins, N., Berger, C., et al. (2021). Differentiation of cystic fibrosis-related pathogens by volatile organic compound analysis with secondary electrospray ionization mass spectrometry. *Metabolites* 11 (11), 773. doi:10.3390/metabo11110773

Karasek, F. W., Kim, S. H., and Hill, H. H. (1976). Mass identified mobility spectra of p-nitrophenol and reactant ions in plasma chromatography. *Anal. Chem.* 48 (8), 1133–1137. doi:10.1021/ac50002a017

Karpas, Z., Stimac, R. M., and Rappoport, Z. (1988). Differentiating between large isomers and derivation of structural information by ion mobility spectrometry/mass spectrometry techniques. *Int. J. Mass Spectrom. Ion. Process.* 83(1)**,** 163–175. doi:10.1016/0168-1176(88)80093-4

Khan, A., Kanwal, H., Bibi, S., Mushtaq, S., Khan, A., Khan, Y. H., et al. (2021). "Volatile organic compounds and neurological disorders: from exposure to preventive interventions," in *Environmental contaminants and neurological disorders*. Editors M. S. H. Akash, and K. Rehman (Cham: Springer International Publishing), 201–230.

Kind, T., and Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinforma.* 7 (1), 234. doi:10.1186/1471-2105-7-234

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., et al. (2009). FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.* 81 (24), 10038–10048. doi:10.1021/ac9019522

King, J., Mochalski, P., Kupferthaler, A., Unterkofler, K., Koc, H., Filipiak, W., et al. (2010). Dynamic profiles of volatile organic compounds in exhaled breath as determined by a coupled PTR-MS/GC-MS study. *Physiol. Meas.* 31 (9), 1169–1184. doi:10.1088/0967-3334/31/9/008

Kohl, I., Beauchamp, J., Cakar-Beck, F., Herbig, J., Dunkl, J., Tietje, O., et al. (2013). First observation of a potential non-invasive breath gas biomarker for kidney function. *J. Breath. Res.* 7 (1), 017110. doi:10.1088/1752-7155/7/1/017110

Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., et al. (2004). GMD@CSB.DB: the Golm metabolome database. *Bioinformatics* 21 (8), 1635–1638. doi:10.1093/bioinformatics/bti236

Lagg, A., Taucher, J., Hansel, A., and Lindinger, W. (1994). Applications of proton transfer reactions to gas analysis. *Int. J. Mass Spectrom. Ion. Process.* 134 (1), 55–66. doi:10.1016/0168-1176(94)03965-8

Lai, Z., Tsugawa, H., Wohlgemuth, G., Mehta, S., Mueller, M., Zheng, Y., et al. (2018). Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* 15 (1), 53–56. doi:10.1038/nmeth.4512

Lapthorn, C., Pullen, F., and Chowdhry, B. Z. (2013). Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: separating and assigning structures to ions. *Mass Spectrom. Rev.* 32 (1), 43–71. doi:10.1002/mas.21349

Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R., and Fowler, S. J. (2017). Exhaled breath analysis: a review of "breathtaking" methods for off-line analysis. *Metabolomics* 13, 110. doi:10.1007/s11306-017-1241-8

Le, T., and Priefer, R. (2023). Detection technologies of volatile organic compounds in the breath for cancer diagnoses. *Talanta* 265, 124767. doi:10.1016/j.talanta.2023.124767

LECO Corporation White paper (2021). *High resolution multi-reflecting time-of-Flight mass analyzer with folded flight Path®HR MR-TOFMS FFP®*. St. Joseph, MI, United States: LECO Corporation White paper, 1–19. 209-281-008 - 6/21, 1-19. https://www.leco.com/documents/articles/209-281-008/.

Lee, C. Y., and Shiea, J. (1998). Gas chromatography connected to multiple channel electrospray ionization mass spectrometry for the detection of volatile organic compounds. *Anal. Chem.* 70 (13), 2757–2761. doi:10.1021/ac971325+

Lindinger, W., Hansel, A., and Jordan, A. (1998). On-line monitoring of volatile organic compounds at pptv levels by means of proton-transfer-reaction mass spectrometry (PTR-MS) medical applications, food control and environmental research. *Int. J. Mass Spectrom. Ion. Process.* 173 (3), 191–241. doi:10.1016/S0168-1176(97)00281-4

Lindinger, W., Hirber, J., and Paretzke, H. (1993). An ion/molecule-reaction mass spectrometer used for on-line trace gas analysis. *Int. J. Mass Spectrom. Ion. Process.* 129, 79–88. doi:10.1016/0168-1176(93)87031-M

Liu, Z., and Phillips, J. B. (1991). Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface. *J. Chromatogr. Sci.* 29 (6), 227–231. doi:10.1093/chromsci/29.6.227

Liu, Z., Wang, M., Wu, M., Li, X., Liu, H., Niu, N., et al. (2023). Volatile organic compounds (VOCs) from plants: from release to detection. *Trends Anal. Chem.* 158, 116872. doi:10.1016/j.trac.2022.116872

Loboda, A. V., and Chernushevich, I. V. (2009). A novel ion trap that enables high duty cycle and wide *m/z* range on an orthogonal injection TOF mass spectrometer. *J. Am. Soc. Mass Spectrom.* 20 (7), 1342–1348. doi:10.1016/j.jasms.2009.03.018

Lomonaco, T., Romani, A., Ghimenti, S., Biagini, D., Bellagambi, F. G., Onor, M., et al. (2018). Determination of carbonyl compounds in exhaled breath by on-sorbent derivatization coupled with thermal desorption and gas chromatography-tandem mass spectrometry. *J. Breath. Res.* 12 (4), 046004. doi:10.1088/1752-7163/aad202

Lough, F., Perry, J. D., Stanforth, S. P., and Dean, J. R. (2017). Detection of exogenous VOCs as a novel *in vitro* diagnostic technique for the detection of pathogenic bacteria. *Trends Anal. Chem.* 87, 71–81. doi:10.1016/j.trac.2016.12.004

Lü, F., Chen, W., Duan, H., Zhang, H., Shao, L., and He, P. (2022). Monitor process state of batch anaerobic digestion in reliance on volatile and semi-volatile metabolome. *Bioresour. Technol.* 351, 126953. doi:10.1016/j.biortech.2022.126953

Lubes, G., and Goodarzi, M. (2017). Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chem. Rev.* 117 (9), 6399–6422. doi:10.1021/acs.chemrev.6b00698

Majchrzak, T., Wojnowski, W., Lubinska-Szczygeł, M., Różańska, A., Namieśnik, J., and Dymerski, T. (2018). PTR-MS and GC-MS as complementary techniques for analysis of volatiles: a tutorial review. *Anal. Chim. Acta* 1035, 1–13. doi:10.1016/j.aca.2018.06.056

Majchrzak, T., Wojnowski, W., Rutkowska, M., and Wasik, A. (2020). Real-time volatilomics: a novel approach for analyzing biological samples. *Trends Plant Sci.* 25 (3), 302–312. doi:10.1016/j.tplants.2019.12.005

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* 72 (6), 1156–1162. doi:10.1021/ac991131p

Malásková, M., Henderson, B., Chellayah, P. D., Ruzsanyi, V., Mochalski, P., Cristescu, S. M., et al. (2019). Proton transfer reaction time-of-flight mass spectrometric measurements of volatile compounds contained in peppermint oil capsules of relevance to real-time pharmacokinetic breath studies. *J. Breath. Res.* 13 (4), 046009. doi:10.1088/1752-7163/ab26e2

Mamyrin, B. A., Karataev, V. I., Shmikk, D. V., and Zagulin, V. A. (1973). Mass reflection: a new nonmagnetic time-of-flight high resolution mass-spectrometer. *Sov. Phys. JETP* 37, 45–48.

Martin, A. J., and Synge, R. L. (1941). A new form of chromatogram employing two liquid phases: a theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. *Biochem. J.* 35 (12), 1358–1368. doi:10.1042/bj0351358

Martínez-Lozano, P., Zingaro, L., Finiguerra, A., and Cristoni, S. (2011). Secondary electrospray ionization-mass spectrometry: breath study on a control group. *J. Breath. Res.* 5 (1), 016002. doi:10.1088/1752-7155/5/1/016002

Materić, D., Bruhn, D., Turner, C., Morgan, G., Mason, N., and Gauci, V. (2015). Methods in plant foliar volatile organic compounds research. *Appl. Plant Sci.* 3 (12). doi:10.3732/apps.1500044

McLafferty, F. W., and Sttauffer, D. (1989). Wiley/NBS Registry of mass spectral data (Accessed July 17, 2024).

Meidert, A. S., Choukèr, A., Praun, S., Schelling, G., and Dolch, M. E. (2021). Exhaled breath and oxygenator sweep gas propionaldehyde in acute respiratory distress syndrome. *Molecules* 26 (1), 145. doi:10.3390/molecules26010145

Moura, P. C., Raposo, M., and Vassilenko, V. (2023). Breath volatile organic compounds (VOCs) as biomarkers for the diagnosis of pathological conditions: a review. *Biomed. J.* 46 (4), 100623. doi:10.1016/j.bj.2023.100623

Mukhopadhyay, R. (2008). IMS/MS: its time has come. *Anal. Chem.* 80 (21), 7918–7920. doi:10.1021/ac8018608

Mumm, R., Schrank, K., Wegener, R., Schulz, S., and Hilker, M. (2003). Chemical analysis of volatiles emitted by Pinus svlvestris after induction by insect oviposition. *J. Chem. Ecol.* 29 (5), 1235–1252. doi:10.1023/a:1023841909199

Pedder, R. E. (2001). "Practical quadrupole theory: graphical theory," in *Extrel application note RA_2010A poster presented at the 49th ASMS conference on mass spectrometry and allied topics*.

Pham, Y. L., Holz, O., and Beauchamp, J. (2023). Emissions and uptake of volatiles by sampling components in breath analysis. *J. Breath. Res.* 17 (3), 037102. doi:10.1088/1752-7163/acce34

Picazo-Aragonés, J., Terrab, A., and Balao, F. (2020). Plant volatile organic compounds evolution: transcriptional regulation, epigenetics and polyploidy. *Int. J. Mol. Sci.* 21 (23), 8956. doi:10.3390/ijms21238956

Pourasil, R. S. M., Cristale, J., Lacorte, S., and Tauler, R. (2022). Non-targeted Gas Chromatography Orbitrap Mass Spectrometry qualitative and quantitative analysis of semi-volatile organic compounds in indoor dust using the Regions of Interest Multivariate Curve Resolution chemometrics procedure. *J. Chromatogr. A* 1668, 462907. doi:10.1016/j.chroma.2022.462907

Prince, B. J., Milligan, D. B., and McEwan, M. J. (2010). Application of selected ion flow tube mass spectrometry to real-time atmospheric monitoring. *Rapid Commun. Mass Spectrom.* 24 (12), 1763–1769. doi:10.1002/rcm.4574

Ratiu, I.-A., Ligor, T., Bocos-Bintintan, V., and Buszewski, B. (2017). Mass spectrometric techniques for the analysis of volatile organic compounds emitted from bacteria. *Bioanalysis* 9 (14), 1069–1092. doi:10.4155/bio-2017-0051

Ratiu, I. A., Ligor, T., Bocos-Bintintan, V., Mayhew, C. A., and Buszewski, B. (2020). Volatile organic compounds in exhaled breath as fingerprints of lung cancer, asthma and COPD. *J. Clin. Med.* 10 (1), 32. doi:10.3390/jcm10010032

Razo-Belmán, R., Ángeles-López, Y. I., García-Ortega, L. F., León-Ramírez, C. G., Ortiz-Castellanos, L., Yu, H., et al. (2023). Fungal volatile organic compounds: mechanisms involved in their sensing and dynamic communication with plants. *Front. Plant Sci.* 14. doi:10.3389/fpls.2023.1257098

Rey-Stolle, F., Dudzik, D., Gonzalez-Riano, C., Fernández-García, M., Alonso-Herranz, V., Rojo, D., et al. (2022). Low and high resolution gas chromatography-mass spectrometry for untargeted metabolomics: a tutorial. *Anal. Chim. Acta* 1210, 339043. doi:10.1016/j.aca.2021.339043

Romano, A., and Hanna, G. B. (2018). Identification and quantification of VOCs by proton transfer reaction time of flight mass spectrometry: an experimental workflow for the optimization of specificity, sensitivity, and accuracy. *J. Mass Spectrom.* 53 (4), 287–295. doi:10.1002/jms.4063

Roslund, K., Uosukainen, M., Järvik, K., Hartonen, K., Lehto, M., Pussinen, P., et al. (2022). Antibiotic treatment and supplemental hemin availability affect the volatile organic compounds produced by P. gingivalis *in vitro*. *Sci. Rep.* 12 (1), 22534. doi:10.1038/s41598-022-26497-0

Ruszkiewicz, D. M., Myers, R., Henderson, B., Yusof, H., Meister, A., Moreno, S., et al. (2022). Peppermint protocol: first results for gas chromatography-ion mobility spectrometry. *J. Breath. Res.* 16 (3), 036004. doi:10.1088/1752-7163/ac6ca0

Sarbach, C., Stevens, P., Whiting, J., Puget, P., Humbert, M., Cohen-Kaminsky, S., et al. (2013). Evidence of endogenous volatile organic compounds as biomarkers of diseases in alveolar breath. *Ann. Pharm. Fr.* 71 (4), 203–215. doi:10.1016/j.pharma.2013.05.002

Schaber, C. L., Katta, N., Bollinger, L. B., Mwale, M., Mlotha-Mitole, R., Trehan, I., et al. (2018). Breathprinting reveals malaria-associated biomarkers and mosquito attractants. *J. Infect. Dis.* 217 (10), 1553–1560. doi:10.1093/infdis/jiy072

Schmidt, F., Kohlbrenner, D., Malesevic, S., Huang, A., Klein, S. D., Puhan, M. A., et al. (2023). Mapping the landscape of lung cancer breath analysis: a scoping review (ELCABA). *Lung Cancer* 175, 131–140. doi:10.1016/j.lungcan.2022.12.003

Schröder, R., Forstreuter, M., and Hilker, M. (2005). A plant notices insect egg deposition and changes its rate of photosynthesis. *Plant Physiol.* 138 (1), 470–477. doi:10.1104/pp.105.059915

Scotter, J. M., Allardyce, R. A., Langford, V. S., Hill, A., and Murdoch, D. R. (2006). The rapid evaluation of bacterial growth in blood cultures by selected ion flow tube–mass spectrometry (SIFT-MS) and comparison with the BacT/ALERT automated blood culture system. *J. Microbiol. Methods* 65(3), 628–631. doi:10.1016/j.mimet.2005.09.016

Seeley, J. V., and Seeley, S. K. (2013). Multidimensional gas chromatography: fundamental advances and new applications. *Anal. Chem.* 85 (2), 557–578. doi:10.1021/ac303195u

Sharkey, T. D., and Yeh, S. (2001). Isoprene emission from plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 407–436. doi:10.1146/annurev.arplant.52.1.407

Skogerson, K., Wohlgemuth, G., Barupal, D. K., and Fiehn, O. (2011). The volatile compound BinBase mass spectral database. *BMC Bioinforma.* 12, 321. doi:10.1186/1471-2105-12-321

Smith, D., and Spanel, P. (1996). The novel selected-ion flow tube approach to trace gas analysis of air and breath. *Rapid Commun. Mass Spectrom.* 10 (10), 1183–1198. doi:10.1002/(SICI)1097-0231(19960731)10:10<1183::AID-RCM641>3.0.CO;2-3

Smith, D., and Španěl, P. (2005). Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrom. Rev.* 24 (5), 661–700. doi:10.1002/mas.20033

Smith, D., Španěl, P., Demarais, N., Langford, V. S., and McEwan, M. J. (2022). Recent developments and applications of selected ion flow tube mass spectrometry (SIFT-MS). *Mass Spectrom. Rev. n/a*, e21835. doi:10.1002/mas.21835

Smith, D., Španěl, P., Herbig, J., and Beauchamp, J. (2014). Mass spectrometry for real-time quantitative breath analysis. *J. Breath. Res.* 8 (2), 027101. doi:10.1088/1752-7155/8/2/027101

Španěl, P., Davies, S., and Smith, D. (1999). Quantification of breath isoprene using the selected ion flow tube mass spectrometric analytical method. *Rapid Commun. Mass Spectrom.* 13 (17), 1733–1738. doi:10.1002/(SICI)1097-0231(19990915)13:17<1733::AID-RCM707>3.0.CO;2-S

Starowicz, M. (2021). Analysis of volatiles in food products. *Separations* 8 (9), 157. doi:10.3390/separations8090157

Sugita, K., and Sato, H. (2021). Sample introduction method in gas chromatography. *Anal. Sci.* 37 (1), 159–165. doi:10.2116/analsci.20SAR19

Tabbal, S., El Aroussi, B., Bouchard, M., Marchand, G., and Haddad, S. (2022a). Development and validation of a method for the simultaneous quantification of 21 microbial volatile organic compounds in ambient and exhaled air by thermal desorption and gas chromatography–mass spectrometry. *Mass Spectrom. Atmos.* 13 (9), 1432. doi:10.3390/atmos13091432

Tabbal, S., El Aroussi, B., Bouchard, M., Marchand, G., and Haddad, S. (2022b). A new headspace solid-phase microextraction coupled with gas chromatography-tandem mass spectrometry method for the simultaneous quantification of 21 microbial volatile organic compounds in urine and blood. *Chemosphere* 296, 133901. doi:10.1016/j.chemosphere.2022.133901

Takaya, K., Hagiwara, M., Matoba, S., Takaya, M., and Shibata, N. (2022). A real-time gas monitoring system based on ion mobility spectrometry for workplace environmental measurements. *Ind. Health* 60 (1), 40–46. doi:10.2486/indhealth.2021-0037

Taylor, C., Lough, F., Stanforth, S. P., Schwalbe, E. C., Fowlis, I. A., and Dean, J. R. (2017). Analysis of Listeria using exogenous volatile organic compound metabolites and their detection by static headspace–multi-capillary column–gas chromatography–ion mobility spectrometry (SHS–MCC–GC–IMS). *Anal. Bioanal. Chem.* 409 (17), 4247–4256. doi:10.1007/s00216-017-0375-x

Temerdashev, A. Z., Gashimova, E. M., Porkhanov, V. A., Polyakov, I. S., Perunov, D. V., and Dmitrieva, E. V. (2023). Non-invasive lung cancer diagnostics through metabolites in exhaled breath: influence of the disease variability and comorbidities. *Metabolites* 13 (2), 203. doi:10.3390/metabo13020203

Tiwari, S., Kate, A., Mohapatra, D., Tripathi, M. K., Ray, H., Akuli, A., et al. (2020). Volatile organic compounds (VOCs): biomarkers for quality management of horticultural commodities during storage through e-sensing. *Trends Food Sci. Technol.* 106, 417–433. doi:10.1016/j.tifs.2020.10.039

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* 12 (6), 523–526. doi:10.1038/nmeth.3393

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Anal. Chem.* 88 (16), 7946–7958. doi:10.1021/acs.analchem.6b00770

Ullah, M. A., Kim, K.-H., Szulejko, J. E., and Cho, J. (2014). The gas chromatographic determination of volatile fatty acids in wastewater samples: evaluation of experimental biases in direct injection method against thermal desorption method. *Anal. Chim. Acta* 820, 159–167. doi:10.1016/j.aca.2014.02.012

US EPA (2022). "Technical overview of volatile organic compounds," in *Indoor air quality (IAQ)* (Washington, D.C., United States: United States Environmental Protection Agency). Available at: https://www.epa.gov/indoor-air-quality-iaq/technical-overview-volatile-organic-compounds.

Vautz, W., Zimmermann, D., Hartmann, M., Baumbach, J. I., Nolte, J., and Jung, J. (2006). Ion mobility spectrometry for food quality and safety. *Food Addit. Contam.* 23 (11), 1064–1073. doi:10.1080/02652030600889590

Vidal-de-Miguel, G., and Herrero, A. (2012). Secondary electrospray ionization of complex vapor mixtures. Theoretical and experimental approach. *J. Am. Soc. Mass Spectrom.* 23 (6), 1085–1096. doi:10.1007/s13361-012-0369-z

Wang, C., Zhang, W., Li, H., Mao, J., Guo, C., Ding, R., et al. (2019). Analysis of volatile compounds in pears by HS-SPME-GC×GC-TOFMS. *Molecules* 24 (9), 1795. doi:10.3390/molecules24091795

Wang, T. I., Chu, C. W., Hung, H. M., Kuo, G. S., and Han, C. C. (1994). Design parameters of dual-stage ion reflectrons. *Rev. Sci. Instrum.* 65 (5), 1585–1589. doi:10.1063/1.1144896

Wang, W., Khalil-Ur-Rehman, M., Feng, J., and Tao, J. (2017). RNA-seq based transcriptomic analysis of CPPU treated grape berries and emission of volatile compounds. *J. Plant Physiol.* 218, 155–166. doi:10.1016/j.jplph.2017.08.004

Weber, R., Kaeslin, J., Moeller, S., Perkins, N., Micic, S., and Moeller, A. (2023a). Effects of a volatile organic compound filter on breath profiles measured by secondary electrospray high-resolution mass spectrometry. *Molecules* 28 (1), 45. doi:10.3390/molecules28010045

Weber, R., Streckenbach, B., Welti, L., Inci, D., Kohler, M., Perkins, N., et al. (2023b). Online breath analysis with SESI/HRMS for metabolic signatures in children with allergic asthma. *Front. Mol. Biosci.* 10. doi:10.3389/fmolb.2023.1154536

Weggler, B. A., Gruber, B., and Dorman, F. L. (2019). Rapid screening of complex matrices: utilizing kendrick mass defect to enhance knowledge-based group type evaluation of multidimensional gas chromatography–high-resolution time-of-flight mass spectrometry data. *Anal. Chem.* 91 (17), 10949–10954. doi:10.1021/acs.analchem.9b01750

Weise, T., Kai, M., Gummesson, A., Troeger, A., von Reuß, S., Piepenborn, S., et al. (2012). Volatile organic compounds produced by the phytopathogenic bacterium Xanthomonas campestris pv. vesicatoria 85-10. *Beilstein J. Org. Chem.* 8, 579–596. doi:10.3762/bjoc.8.65

Weisskopf, L., Schulz, S., and Garbeva, P. (2021). Microbial volatile organic compounds in intra-kingdom and inter-kingdom interactions. *Nat. Rev. Microbiol.* 19 (6), 391–404. doi:10.1038/s41579-020-00508-1

Westphal, K., Dudzik, D., Waszczuk-Jankowska, M., Graff, B., Narkiewicz, K., and Markuszewski, M. J. (2023). Common strategies and factors affecting off-line breath sampling and volatile organic compounds analysis using thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS). *Metabolites* 13 (1), 8. doi:10.3390/metabo13010008

Whitehouse, C. M., Meng, C. K., and Fenn, J. B. (1986). *Proceedings of the 34th ASMS conference on mass spectrometry and allied topics*. Denver, CO, 507.

Willis, P., Jaloszynski, J., and Artaev, V. (2021). Improving duty cycle in the Folded Flight Path high-resolution time-of-flight mass spectrometer. *Int. J. Mass Spectrom.* 459, 116467. doi:10.1016/j.ijms.2020.116467

Wojnowski, W., Kalinowska, K., Majchrzak, T., and Zabiegała, B. (2022). Real-time monitoring of the emission of volatile organic compounds from polylactide 3D printing filaments. *Sci. Total Environ.* 805, 150181. doi:10.1016/j.scitotenv.2021.150181

Wu, C., Siems, W. F., and Hill, H. H. (2000). Secondary electrospray ionization ion mobility spectrometry/mass spectrometry of illicit drugs. *Anal. Chem.* 72 (2), 396–403. doi:10.1021/ac9907235

Wüthrich, C., and Giannoukos, S. (2024). Advances in secondary electrospray ionization for breath analysis and volatilomics. *Int. J. Mass Spectrom.* 498, 117213. doi:10.1016/j.ijms.2024.117213

Zhang, V. R., Ramachandran, G. K., Loo, E. X. L., Soh, A. Y. S., Yong, W. P., and Siah, K. T. H. (2023). Volatile organic compounds as potential biomarkers of irritable bowel syndrome: a systematic review. *Neurogastroenterol. Motil.* 35 (7), e14536. doi:10.1111/nmo.14536

Zheng, Y., Wang, P., Chen, X., Sun, Y., Yue, C., and Ye, N. (2019). Transcriptome and metabolite profiling reveal novel insights into volatile heterosis in the tea plant (camellia sinensis). *Molecules* 24 (18), 3380. doi:10.3390/molecules24183380

Zubarev, R. A., and Makarov, A. (2013). Orbitrap mass spectrometry. *Anal. Chem.* 85 (11), 5288–5296. doi:10.1021/ac4001223

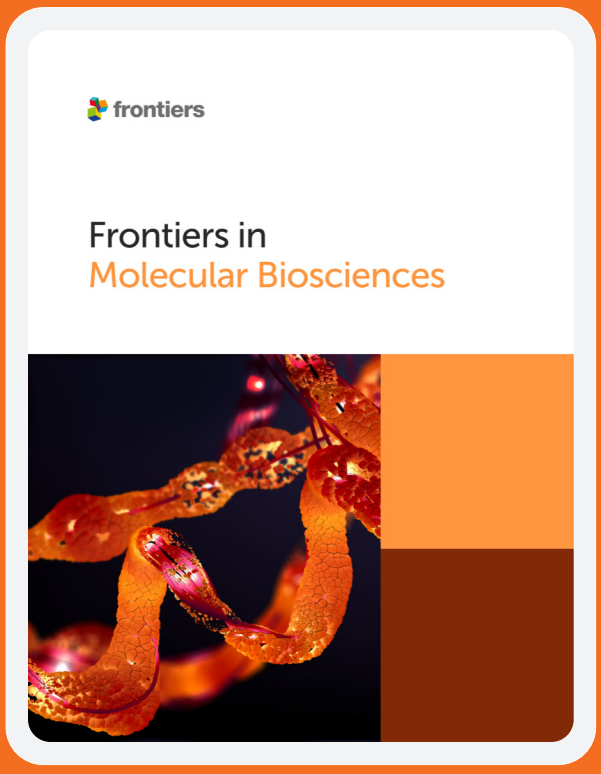
Frontiers in
Molecular Biosciences

Explores biological processes in living organisms on a molecular scale

Focuses on the molecular mechanisms underpinning and regulating biological processes in organisms across all branches of life.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact