



HAL
open science

Volatile Organic Compounds Analysis as a Potential Novel Screening Tool for Breast Cancer: A Systematic Review

Michelle Leemans, Pierre Bauër, Vincent Cuzuel, Etienne Audureau, Isabelle Fromantin

► **To cite this version:**

Michelle Leemans, Pierre Bauër, Vincent Cuzuel, Etienne Audureau, Isabelle Fromantin. Volatile Organic Compounds Analysis as a Potential Novel Screening Tool for Breast Cancer: A Systematic Review. Biomarker Insights, 2022, 17, pp.117727192211007. 10.1177/11772719221100709. hal-03791273

HAL Id: hal-03791273


<https://hal.u-pec.fr/hal-03791273v1>

Submitted on 29 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Volatile Organic Compounds Analysis as a Potential Novel Screening Tool for Breast Cancer: A Systematic Review

Michelle Leemans¹, Pierre Bauër², Vincent Cuzuel³, Etienne Audureau^{1,4} and Isabelle Fromantin^{1,2}

¹Univ Paris Est Créteil, INSERM, IMRB, Créteil, France. ²Institut Curie, Ensemble hospitalier, Unité Plaies et Cicatrisation, Paris, France. ³Institut de Recherche Criminelle de la Gendarmerie Nationale, Caserne Lange, Cergy Pontoise Cedex, France. ⁴Assistance Publique – Hôpitaux de Paris, Hôpital Henri Mondor, Service de Santé Publique, Créteil, France.

Biomarker Insights
Volume 17: 1–19
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11772719221100709



ABSTRACT

INTRODUCTION: An early diagnosis is crucial in reducing mortality among people who have breast cancer (BC). There is a shortfall of characteristic early clinical symptoms in BC patients, highlighting the importance of investigating new methods for its early detection. A promising novel approach is the analysis of volatile organic compounds (VOCs) produced and emitted through the metabolism of cancer cells.

METHODS: The purpose of this systematic review is to outline the published research regarding BC-associated VOCs. For this, headspace analysis of VOCs was explored in patient-derived body fluids, animal model-derived fluids, and BC cell lines to identify BC-specific VOCs. A systematic search in PubMed and Web of Science databases was conducted according to the PRISMA guidelines.

RESULTS: Thirty-two studies met the criteria for inclusion in this review. Results highlight that VOC analysis can be promising as a potential novel screening tool. However, results of *in vivo*, *in vitro* and case-control studies have delivered inconsistent results leading to a lack of inter-matrix consensus between different VOC sampling methods.

DISCUSSION: Discrepant VOC results among BC studies have been obtained, highly due to methodological discrepancies. Therefore, methodological issues leading to disparities have been reviewed and recommendations have been made on the standardisation of VOC collection and analysis methods for BC screening, thereby improving future VOC clinical validation studies.

KEYWORDS: Breast cancer, volatile organic compounds, gas chromatography, biomarker, *in vitro* study, E-nose, mass spectrometry, breath, urine

RECEIVED: December 16, 2021. **ACCEPTED:** April 19, 2022.

TYPE: Systematic review

FUNDING: The author(s) received no financial support for the research, authorship and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

CORRESPONDING AUTHORS: Michelle Leemans, Univ Paris Est Créteil, INSERM, IMRB, 61, Av. du Général de Gaulle, Créteil, France. Email: michelleleemans91@gmail.com

Isabelle Fromantin, Institut Curie, Ensemble hospitalier, Unité Plaies et Cicatrisation, 26, rue d'Ulm, Paris, France. Email: isabelle.fromantin@curie.fr

Introduction

Breast cancer

Breast cancer (BC) is the second most common cancer globally, following lung cancer and is the most prevalent cancer among women. In the United States, about 1 in 8 women are diagnosed with BC during their lifetime.¹ Furthermore, with increasing population age, the incidence rate of BC among older people is augmenting.² The prognoses in patients with BC are mainly dependent on the stage of the disease. In BC, staging is conducted according to the TNM classification system taking into account the size of the primary tumour (T), the spread of locoregional lymph nodes (N), and distant metastasis (M).³ Tumours are separated into 4 different phases, depending on tumour progression. Stage 0 represents non-invasive cancer, meaning that the cancer cells are within the breast lobe and have not yet invaded the surrounding breast tissue.⁴ The stages 1 to 3 are characterised by: (1) invasive BC with no lymph nodes involved, (2) invasive BC with spreading to

nearby lymph nodes and (3) invasive BC with enlarged spreading in the lymph nodes beyond the immediate region of the tumour, respectively, and are considered curable with surgery and/or therapy. In stage 4 BC, the cancer is spread beyond the breast into other organs and is estimated to have relatively low overall survival rate.⁵ In BC, tumour size has been associated with progression-free survival and overall survival.⁶ Therefore, early BC diagnosis can effectively cut down the mortality. Another way to classify BC has been developed based on gene expression analysis, where a unique set of genes is measured within the tumour tissue. Within this classification, the progesterone receptor (PR), the oestrogen receptor (ER) and the HER2 are the most commonly used for BC categorisation.⁷ This classification is termed 'intrinsic subtypes'.

Numerous risk factors have been identified for BC, with lifestyle, genetic- and environmental factors being the most important ones. Other determinants, including lactation, age, socio-economic status, and prior benign breast disease, might play essential roles in the risk of this disease.⁸ Various genes



have been correlated with BC, and abnormal amplification or mutations of specific genes play a vital role in tumour initiation and progression processes. Two famous anti-oncogenes for BC risk are breast cancer-associated genes 1 and 2 (BRCA1 and BRCA2). Mutations in either of these genes increase the lifetime risk of BC up to 60% to 85%.⁹ Yet, only 10% of all BC cases are ascribed to the involvement of hereditary factors.¹⁰

Breast cancer screening

The self-diagnosis of BC is limited, given the disease mainly develops asymptotically in the early phases. Therefore, several countries in Europe and North America successfully organised BC screening programs, which have demonstrated that BC mortality can be reduced by 15% to 20% by periodic mammography screening for women aged 50 to 69 years.^{11,12} In the United States, where 70% of women are subjected to mammography, the latest estimates of diagnostic digital mammography's sensitivity and specificity are between 87.8% and 90.5%, respectively.¹³ Lower sensitivity and higher interval cancer rates were observed in women with a personal history of BC compared with women without or in women with extremely dense breasts.¹⁴ Clinical BC screening is not recommended for women at any age, screening starts mostly between 40 and 44 and continues as a woman is in good health and expected to live at least 10 more years, with a narrow harm-benefit-ratio for younger and older women. A better trade-off seems to be obtained by screening every 2 years for middle-aged women. However, further research is necessary to evaluate the effect of screening intervals.¹⁵ It should be noted that BC at a relatively young age is exhibiting an increasing incidence and that these cases are missed out in the current screening programs.¹⁶

Other limitations of a mammography include anxiety, fear, and worry which are variously described as barriers to widespread BC screening.¹⁷ Furthermore, BC screening every 2 years for 20 years does not guarantee that all BC cases will be discovered due to possible non-visibility of the BC on the mammogram or the BC was not yet developed at the moment of the mammography. Additional supplemental screening with breast magnetic resonance imaging (MRI) may be considered for special high-risk populations, given its high sensitivity (90%) and moderate specificity (72%).¹⁸ For standard breast examination, mammography is preferred over MRI given MRI has various disadvantages, including a significant false-positive rate, longer imaging time, the requirement of an intravenous gadolinium-based contrast agent, its high price tag and problems with claustrophobia. Therefore, the clinical benefits of MRI for BC detection are still a matter of debate.¹⁹

In low- and middle-income countries, the reported sensitivity of diagnostic digital mammography differs from 63% to 95%.²⁰ Furthermore, the accessibility and costs of mammography are important issues for many countries with limited medical resources.²¹ Breast ultrasound, which is employed in high-resource settings to complement mammography in specific

clinical scenarios, gives a potentially applicable alternative for early BC discovery in some resource-limited regions because it is portable, cheaper than mammography, and flexible across a broader range of clinical applications. Sood et al²² conducted a meta-analysis of 26 studies and revealed that portable ultrasound has an overall high sensitivity of 80.1% and specificity of 88.4% for detection of a variety of patient populations. Taken together, new methodologies that can overcome the above limitations are needed to diagnose tumour development at earlier stages of BC.

VOCs

The use of volatiles as non-invasive disease biomarkers originates from a long history of medicine in which Hippocrates described the apparent smell of melaena as early as 400 years BC, and patients with diabetes were characterised as having urine with an odour of decayed apples in ancient Chinese medicine.²³ Nowadays, volatile metabolites have been explored for the diagnosis of a range of other human diseases, including tuberculosis, cystic fibrosis and different cancers.²⁴⁻²⁶ Volatile organic compounds, or VOCs, represent a diverse group of carbon-based molecules, including ketones, alcohols, aldehydes, hydrocarbons, isocyanates, terpenes, sulphides and amines. VOCs produced by the body are liberated into the circulatory system and then infiltrate in biofluids or the air in the lungs. Research on cancer-related VOCs has been conducted in multiple matrices, including breath, blood, urine, saliva, sweat and faeces.²⁷⁻³⁰ VOC emission from these resources can be concentrated in headspace, directly onto thermal adsorbent tubes such as Tenax TA[®], Sorb-Star[®], inert polymer bags or other sorbent materials. Instrumental techniques have shown that hundreds of VOCs are detected from diverse body fluids.³¹ It has been demonstrated that VOCs emitted from the body alter with genetic background, sex, diet and age.³²⁻³⁵ Therefore, body odours can be envisaged as unique 'odour-fingerprints', hence the interest from the forensic field.³⁶ Furthermore, altered metabolic pathways, due to for example infections or cancer, can modify our odour signature by introducing new VOCs or altering the ratio of VOCs usually produced.³⁷ Despite the clinical interest in VOCs and body odours, limited research has been conducted to define these volatiles as diagnostic criteria, either quantitatively or qualitatively.

Next to the endogenously formed and metabolised compounds, VOCs can also consist of exogenous compounds derived from environmental components that are emitted or derived from bacteria on the skin for example. Meaning that volatomics data contain distinct sources of variance, making it a challenge to extract information related to the topic of interest and neglect the irrelevant variance in the data.

In the case of breath collection, additional variance can be obtained given different portions of breath can be collected, which can be divided into end-tidal, late expiratory, and mixed expiratory. Mixed expiratory consists of total exhaled breath including 'dead space air' while other breath categories intend to

minimise contamination from this dead space. Dead space of the respiratory system involves the space in which carbon dioxide (CO₂) and oxygen (O₂) gases are not exchanged across the alveolar membrane in the respiratory tract. End-tidal breath refers to the air that is released at the end of an exhaled breath. Whereas late expiratory breath sampling includes discarding the initial portion of the exhaled breath (estimated dead space) and the consecutive capture of air at the end of the breath cycle.³⁸

VOCs detection techniques

VOCs can be detected using analytic instruments such as gas chromatography coupled with mass spectrometry (GC-MS), proton transfer reaction mass spectrometry (PTR-MS), selected ion flow tube-mass spectrometry (SIFT-MS), infrared spectroscopy or Quadrupole Time-of-Flight GC-MS (GC-MS QTOF).³⁹ GC-MS is the most popular analytical platform for VOC profiling given its reliability and sensitivity in analyte identification. GC-MS QTOF has advantages in detection sensitivity, mass accuracy and resolution compared to the gold-standard GC-MS. Other ionisation techniques such as PTR-MS and SIFT-MS may also be used: the subsequent VOC ions are then detected by a mass analyser. Alternatively, infrared spectroscopy has been reported for the detection of disease-related VOCs and measures the interaction of infrared radiation with the matter by reflection, absorption or emission for compound identification.⁴⁰ Volatile components which may be diagnostically valuable are found in trace concentrations measured in parts per million, parts per billion, or even parts per trillion.⁴¹ Therefore, a pre-concentration step is usually necessary before the analysis. The sample volatiles can be preconcentrated using different methods such as Needle Trap Micro Extraction (NTME) or Solid Phase Micro Extraction (SPME). SPME is one of the most frequently used methodologies for VOC analysis and employs a polymer-coated fused silica fibre exposed to the headspace of the sample containing the VOCs of interest. An interesting development, known as SPME Arrow, offers a higher VOC loading capacity for trace samples, allowing for reduced sampling time and increased sensitivity compared to the original SPME design.⁴² Another possible concentration technique includes stir bar sorptive extraction (SBSE). SBSE potentially improves the concentration capacity and sensitivity of VOCs, compared to SPME, by inducing the amount of adsorbent phase.⁴³ A further extraction technique is liquid-liquid extraction (LLE), a traditional extraction technique in analytical chemistry given its lack of complicated equipment and its simplicity. LLE potentially allows relative high concentrations of a great range of VOCs.⁴⁴

Despite the sensitivity, specificity, and effectiveness of GC-MS technologies, conventional GC-MS remains expensive, time-consuming, requires highly trained and skilled personnel, is laborious to use in daily medical practice, featuring little portability, making its usage possible just in a structured laboratory environment.³⁹ Conversely, electronic noses, or E-noses, have the potential to overcome these disadvantages, given

they provide a cheap, fast, portable, and easy way to analyse gas samples. An E-nose is composed of 2 major components, the sensor array which is in charge of sensing the chemicals and an analysing software component.⁴⁵ When an odour enters the device, it induces a change in current, voltage, resistance parameters or frequency, depending on the type of components in the sensor array.⁴⁶ E-noses can recognise various complex odours by comparing the entering odour with patterns previously learnt.⁴⁷ Therefore, E-noses can potentially be employed for diagnosis, monitoring or phenotyping diseases according to specific volatomics patterns. E-noses can especially be useful once specific VOCs of interest are identified for given research. The amount of research regarding the utilisation of the E-nose has tremendously increased in the field of oncology.⁴⁵ Further advantages include disease detection in bulk host samples and multiple groups of pathogens can potentially be detected using different reference databases with the same e-nose instrument.⁴⁸ Despite this, the device encounters various important disadvantages such as lower sensitivity and/or specificity compared to the analytical approaches. Furthermore, E-noses have potentially poorer reproducibility, given the sensor drift alters over time, lowering the instrument's reliability. Additionally, they can be potentially sensitive to temperature, humidity and light.⁴⁶ Besides, research on specified training of dogs for cancer detection using various odour samples have provided promising results.⁴⁹ A problematic issue that has emerged with canine studies is the large heterogeneity of performance across the dogs in the different studies. Employment of canine cancer detection requires disclosure of what chemical compounds the dogs respond to, as well as the quantity of these compounds. Taken together, it makes studies with GC-MS-based technologies indispensable to accurately identify and quantify volatile biomarkers to expand our basic knowledge and aid the search for potential biomarkers.

Purpose of systematic review

The overall goal of this systematic review is to outline the published research regarding BC-associated VOCs, given no such systematic overview has been previously conducted. Therefore, a state-of-the-art on this scientific subject was needed to demonstrate its limits and potentialities. Specific objectives of this review were to assess the methodological quality of published research; characterise published VOC markers of BC, and examine emerging metabolic signalling pathways. Additionally, methodological issues leading to inconsistencies in research are reviewed and recommendations to identify more accurate biomarkers are given.

Materials and Methods

Literature search

PubMed and Web of Science were independently searched following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses or PRISMA standards, which is the most

frequent method employed for systematic reviews.⁵⁰ Studies about diagnosing BC by VOCs in clinical studies, animal models and *in vitro* studies in the databases from inception to 31 May 2021 were retrieved. The search strategy was adjusted to each database and performed on 31 May 2021. For PubMed, the following string was employed:

((breast) AND (cancer) AND ((smell) OR (odour) OR (Volatile Organic Compounds) OR (volatomics) OR (volatile*) OR ((metabolomic* or volatile*) AND ((GC) or (gas chromatography))) OR ((metabolomic* or volatile*) AND (mass spectrometry)) OR ((exhalation*) OR (breath) or (urine*) AND (sensitivity and specificity))))).

For Web of Science the subsequent sequence was researched: (TI=((breast) AND (cancer) AND ((smell) OR (odor) OR (Volatile Organic Compounds) OR (volatomics) OR (volatile*)OR ((metabolomic* or volatile*) AND ((GC) or (gas chromatography))) OR ((metabolomic* or volatile*) AND (mass spectrometry)) OR ((exhalation*) OR (breath) or (urine*) AND (sensitivity and specificity))))))OR (AB=((breast) AND (cancer) AND ((smell) OR (odor) OR (Volatile Organic Compounds) OR (volatomics) OR (volatile*)OR ((metabolomic* or volatile*) AND ((GC) or (gas chromatography))) OR ((metabolomic* or volatile*) AND (mass spectrometry)) OR ((exhalation*) OR (breath) or (urine*) AND (sensitivity and specificity)))))).

The references of retrieved literature were further reviewed for original articles.

Study selection and eligibility criteria

Data obtained on the search were combined in a worksheet file including, DOI number, title, year of publication, authors, journal and abstracts. In total, 2372 records were identified, 926 from PubMed and 1446 from Web of Science, of which 404 were duplicates (Figure 1). We reviewed the remaining 1706 titles and abstracts to identify studies relevant to the topic. Two researchers (M.L and P.B) independently reviewed the studies, and any disagreements were resolved by discussion. The inclusion criteria were for studies on VOCs related to BC in original articles. Review articles, articles focused on metabolomics, GC-MS studies where no VOCs are identified, and articles not relevant to the topic were excluded. We reviewed 79 full-text papers for inclusion; 44 were excluded after full-text review. A total of 32 papers were included in the systematic review.

Data extraction

The relevant data were extracted from the 32 selected studies. Standardised tables were designed to abstract the studies of interest. Within Table 1, references to each of the studies can be acquired. Supplemental Table 1 includes all VOC sampling details of the included studies. In Supplemental Table 2, all subject details are enclosed. Statistical details on the VOC analysis of all of the included studies are represented in

Supplemental Table 3, whereas a detailed overview of the identified VOCs can be obtained in Supplemental Tables 4 and 5. Details on pathway analysis are represented in Supplemental Table 6. All graphics were generated using Python 3.

Assessment of study quality

Included studies were evaluated using QUADAS-2, a revised tool for the quality assessment of diagnostic accuracy studies.⁸² The QUADAS-2 form is comprised of 4 domains: (1) patient selection, (2) index test, (3) reference standard and (4) flow and timing. For each domain, the risk of bias and applicability were investigated and rated as low, high or unclear risk. Within this review, the reference standard was estimated to be radiologic or histologic confirmation of BC. The QUADAS-2 tool was adapted to have greater relevance to phase 1 biomarker discovery research following Hanna et al.⁸³ Obtained results of the quality assessment were employed to direct the evaluation of the included studies.

Results

Description of included studies

A total of 2372 articles were obtained using the search strategy described above, among which 1967 remained after eliminating duplicated articles. About 1888 studies were excluded from the title and abstract due to irrelevancy or patents. Thus, 79 studies were retrieved for full-text browsing. From these, 47 were further discarded for not meeting the inclusion criteria (Figure 1). Thereby, 32 studies were enclosed within this systematic review, including 16 case-control studies (utilising GC-MS technology), 2 *in vivo*-, 6 *in vitro*-studies, 3 studies focussing on mass spectrometry results solely, 5 E-nose studies, and 5 studies including both E-nose and GC-MS results. References and categorisation of the 32 included studies on BC diagnoses by VOCs are outlined in Table 1. Details on the sampling procedure can be obtained in Supplemental Table 1.

The most commonly used method for the detection of VOCs is GC-MS, which was employed in 24 out of 33 studies. Additionally, 1 study uses GC alone and 1 SESI-MS, 3 studies utilise GC-MS QTOF, and a single study combines research on VOCs employing GC-FID, GC-MS and LC-TOF/MS. Three studies used E-nose solely to characterise the chemical signature of BC, whereas 5 studies identified VOCs both with E-nose and GC-MS (Supplemental Table 1).

A total of 1559 BC patients were included in this systematic review, ranging from 3 to 351 (median = 50) women per study in 14 different countries. Within included studies, BC alone (n = 27) or BC in combination (n = 4) with another cancer was studied (Table 1). Most included studies compared BC patients, of an often mixed histologic subtype with a healthy control population and/or patients with benign conditions (n = 8). Studies tended to include patients with early and advanced tumour stages, although the tumour stage was not reported in

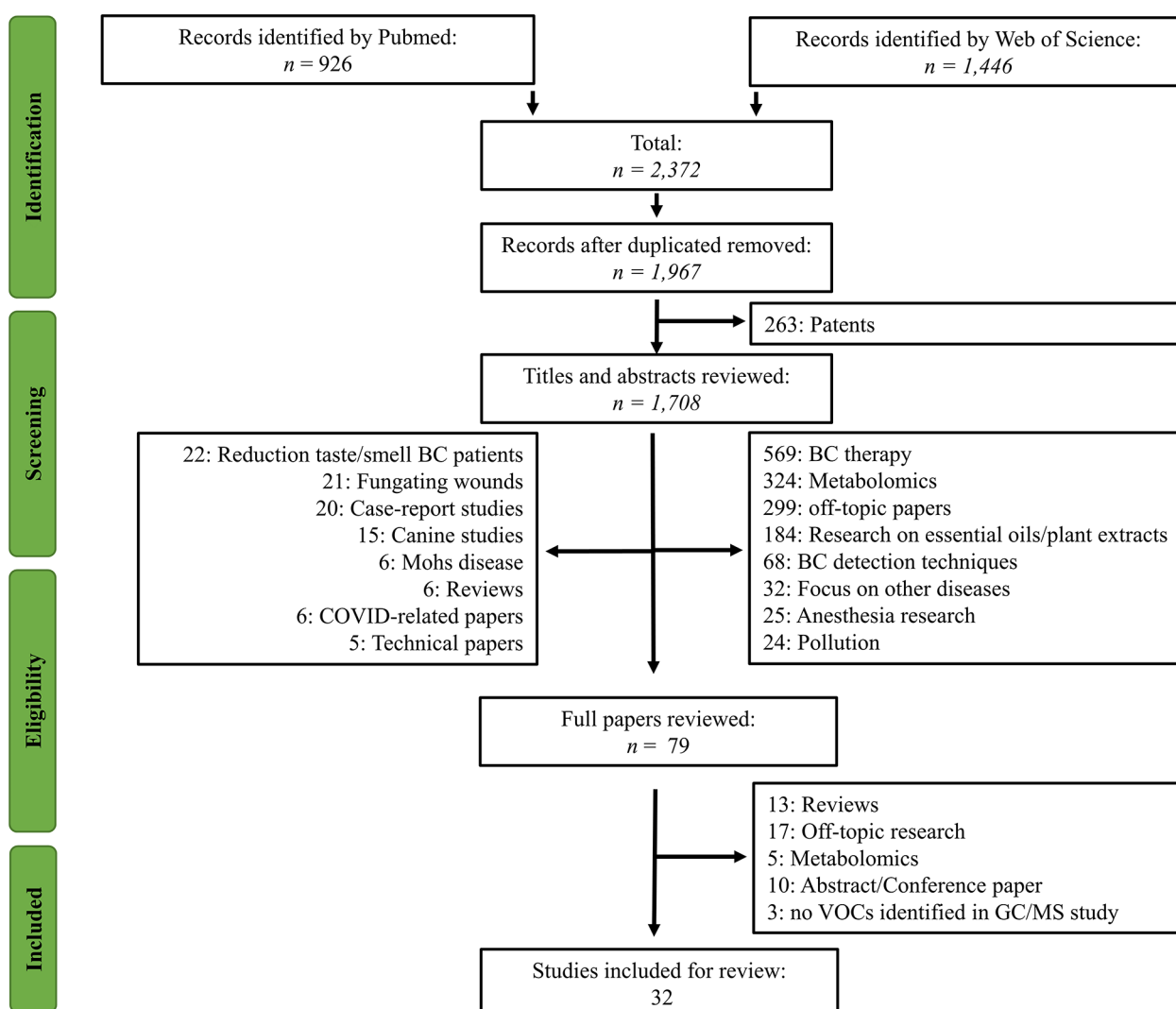


Figure 1. PRISMA flowchart.

12 studies. Interestingly, 8 studies categorised BC and employed categorisation within their data analysis.

A total of 260 VOCs were identified in association with BC within GC-MS studies. Full details of identified BC VOCs are provided in Supplemental Tables 4 and 5. Information on VOC calculation can be obtained in Supplemental Table 3. Results were commonly disposed of accuracy rates for correct VOCs classification of BC samples. The sensitivity of testing for BC diagnosis ranged from 73% to 100%, whereas specificity ranged from 56% to 100%, with a higher specificity for E-nose (85% geometric average) compared to GC-MS studies (80% geometric average). It should be noted that only 14 out of 32 included studies calculated sensitivity/specificity scores. Only 1 out of 32 studies⁶⁴ conducted an external validation of the diagnosing model using a different dataset, while most of the studies conducted internal cross-validation.

Heterogeneity can be observed in all facets of the study procedure, from sample acquisition to data analysis. Given the technical discrepancy and diversity of compounds upon which volatomics models were established, meta-analysis was assumed

unsuitable; alternately, a narrative synthesis of study findings is conferred next.

VOC captivation and conservation. The first step in the process of VOC identification is VOC sampling and collection. Within the included studies different collection containers are employed for the various media, and pronounced differences in storage containers can even be found within a similar matrix. Most studies utilise temporary storage prior to pre-concentration. Various storage containers were employed for breath collection to identify BC associated volatiles. Of the 17 identified breath studies, 5 studies employed Tedlar bags, and 2 studies utilised Mylar bags among other captivation materials. Both urine, saliva and *in vitro* cellular medium were mostly collected into glass vials. After collection, samples are rarely directly processed and are conserved before analysis. Storage conditions were not specified in 12 out of 32 studies. All urine samples were stored at -80°C . Discrepancy is observed for both breath and cellular medium storage, where breath samples were stored at -20°C , 4°C , or

Table 1. Studies on the analysis of patient-derived body fluids, animal-derived fluids, and BC cell lines.

No. of study	Title	Author	Disease
<i>Case-control</i>			
<i>Breath</i>			
1	Early diagnosis of breast cancer from exhaled breath by gas chromatography-mass spectrometry (GC-MS) analysis: a prospective cohort study	Zhang et al ⁵¹	Breast, gastric cancer
2	Differentiation between genetic mutations of breast cancer by breath volatolomics	Barash et al ⁵²	Breast cancer
3	Volatile organic metabolites identify patients with breast cancer, cyclomastopathy, and mammary gland fibroma	Wang et al ⁵³	Breast cancer
4	Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography–mass spectrometry	Li et al ⁵⁴	Breast cancer
5	Volatile organic compounds (VOCs) in exhaled breath of patients with breast cancer in a clinical setting	Mangler et al ⁵⁵	Breast cancer
6	Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors	Peng et al ⁵⁶	Breast, lung, colorectal, prostate cancer
7	Volatile biomarkers in the breath of women with breast cancer	Phillips et al ⁵⁷	Breast cancer
8	Prediction of breast cancer using volatile biomarkers in the breath	Phillips et al ⁵⁸	Breast cancer
9	Volatile markers of breast cancer in the breath	Phillips et al ⁵⁹	Breast cancer
10	Quantitative analysis by gas chromatography of volatile carbonyl compounds in expired air from mice and human	Ebeler et al ⁶⁰	Breast cancer
<i>Urine</i>			
11	Implementing a central composite design for the optimisation of solid phase microextraction to establish the urinary volatome expression: a first approach for breast cancer	Silva et al ⁶²	Breast cancer
12	Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatome biosignature	Porto-Figueira et al ⁶³	Breast, colon cancer
13	A non-invasive approach to explore the discriminatory potential of the urinary volatome of invasive ductal carcinoma of the breast	Taunk et al ⁶⁴	Breast cancer
14	Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—a powerful strategy for breast cancer diagnosis	Silva et al ⁶⁵	Breast cancer
<i>Others</i>			
15	Volatome pattern of breast cancer and cancer-free tissues as a powerful strategy to identify potential biomarkers	Silva et al ⁶¹	Breast cancer
16	Screening of salivary volatiles for putative breast cancer discrimination: an exploratory study involving geographically distant populations	Cavaco et al ³⁰	Breast cancer
<i>In vivo</i>			
<i>Urine</i>			
17	Urinary volatile terpenes analysed by gas chromatography-mass spectrometry to monitor breast cancer treatment efficacy in mice	Woollam et al ⁶⁶	Breast cancer
18	Detection of volatile organic compounds (VOCs) in urine via gas chromatography-mass spectrometry QTOF to differentiate between localised and metastatic models of breast cancer	Woollam et al ⁶⁷	Breast cancer
<i>In vitro</i>			
<i>Culture media</i>			
19	Identification of characteristic compounds of moderate volatility in breast cancer cell lines	Tanaka et al ⁶⁸	Breast cancer

(Continued)

Table 1. (Continued)

20	Extracellular volatilomic alterations induced by hypoxia in breast cancer cells	Taware et al ⁶⁹	Breast cancer
21	Effect of H ₂ O ₂ induced oxidative stress (OS) on volatile organic compounds (VOCs) and intracellular metabolism in MCF-7 breast cancer cells	Liu et al ⁷⁰	Metabolism in MCF-7 cells
22	Volatile metabolomic signature of human breast cancer cell lines	Silva et al ⁷¹	Breast cancer
23	Investigation of biomarkers for discriminating breast cancer cell lines from normal mammary cell lines based on VOCs analysis and metabolomics	Huang et al ⁷²	Breast cancer
24	Investigation of VOCs associated with different characteristics of breast cancer cells	Lavra et al ⁷³	Breast cancer
<i>Mass spectrometry</i>			
25	Breath mass ion biomarkers of breast cancer	Phillips et al ⁷⁴	Breast cancer
26	Secondary electrospray ionisation-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study	Martinez-Lozano Sinues et al ⁷⁵	Breast cancer
27	Fingerprinting breast cancer vs. normal mammary cells by mass spectrometric analysis of volatiles	He et al ⁷⁶	Breast cancer
<i>E-nose</i>			
28	Breath biopsy of breast cancer using sensor array signals and machine learning analysis	Yang et al ⁷⁷	Breast cancer
29	Identification of profiles of volatile organic compounds in exhaled breath by means of an electronic nose as a proposal for a screening method for breast cancer: a case-control study	Díaz de León-Martínez et al ⁷⁸	Breast cancer
30	An in-vitro study for early detection and to distinguish breast and lung malignancies using the PCB technology based nanodosimeter	Venkatraman and Sureka ⁷⁹	Breast, Lung cancer
31	Effect of humidity on nanoparticle-based chemiresistors: a comparison between synthetic and real-world samples	Konvalina and Haick ⁸⁰	Breast cancer
32	Classification of breast cancer precursors through exhaled breath	Shuster et al ⁸¹	Breast cancer
<i>Studies containing both GC-MS analysis and E-nose</i>			
2	Differentiation between genetic mutations of breast cancer by breath volatilomics	Barash et al ⁵²	Breast cancer
6	Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors	Peng et al ⁵⁶	Breast, lung, colorectal, prostate cancer
24	Investigation of VOCs associated with different characteristics of breast cancer cells	Lavra et al ⁷³	Breast cancer
30	An in-vitro study for early detection and to distinguish breast and lung malignancies using the PCB technology based nanodosimeter	Venkatraman and Sureka ⁷⁹	Breast, Lung cancer
31	Effect of humidity on nanoparticle-based chemiresistors: a comparison between synthetic and real-world samples	Konvalina and Haick ⁸⁰	Breast cancer

room temperature and cellular medium at -80°C , -40°C , -20°C or 4°C degrees. Besides storing conditions, also timing between sampling and analysis might differ between studies. Storage timings were only reported in 8 studies with variations ranging between 3 hours post-sampling to 4 months after sample collection.

VOC detection. Twenty-four studies employ the GC-MS separation technique to identify BC related VOCs combined with either SPME or thermal desorption, whereas only 3 studies employ GC-MS QTOF. The number of detected VOCs

ranges between 3 and 646 (median = 109) volatiles with a ratio of detection between 40% and 100% and similarity with the NIST library varying between $>75\%$ and $>90\%$. The purpose of the NIST library is to provide reference data for mass spectrometry-based volatilomics; the higher the similarity with the reference database, the higher the chance of correct volatilomic identification. A better verification of correct compound characterisation can be obtained with compound confirmation by external chemical standards; 8 included studies verified the identity of some of the significantly classified VOCs. The amount of BC significantly related compounds in the 32

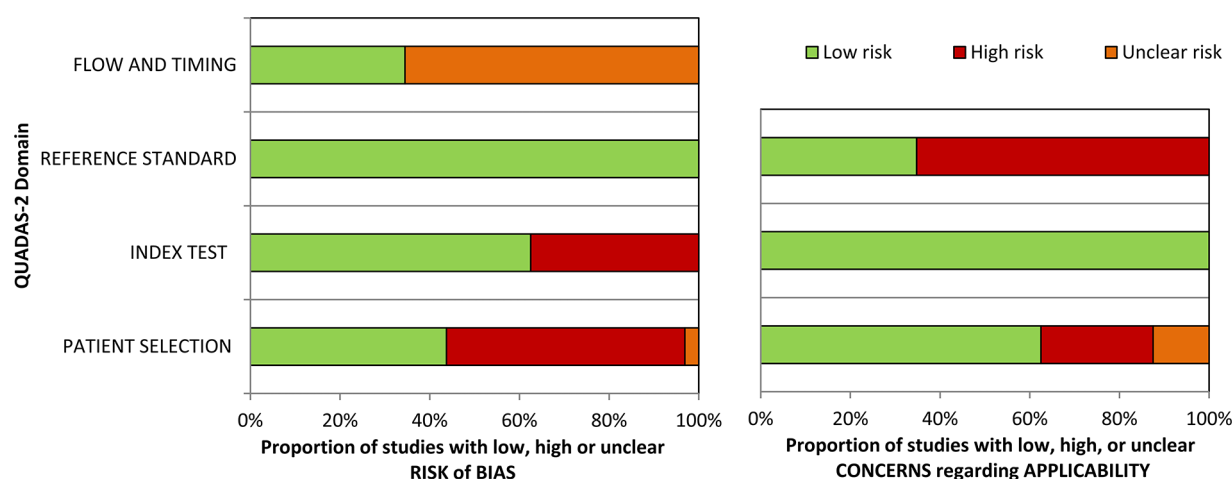


Figure 2. Summary of risk of bias and concerns regarding applicability for included studies: (A) risk of bias and (B) concerns regarding applicability.

included studies ranged between 1 and 75 compounds with a median of 6 compounds. From all included studies, more than half ($n = 17$) of them identified whether the significantly altered compounds were increased or decreased in BC conditions compared to the healthy state. One should bear in mind the fact that chemical identification of VOC biomarkers might be inaccurate, given spectral matching can misidentify the chemical structure. Therefore, mass ions detected by GC-MS have been proposed as intrinsically robust biomarkers, given they are identified without spectral matching. Three out of 32 studies focussed on mass ion detection, wherefrom 2 display the measured mass divided by charge number (m/z value). No commonality between identified mass ions distinguishing BC state from healthy status has been found in these studies.

Seven studies employing an E-nose with a focus on BC have been identified, wherefrom 5 studies utilise additional GC-MS platforms to identify cancer-related VOCs. Remarkably, only 2 E-nose studies employ similar detection apparatus. Seventy-five percent of the E-nose studies conducted VOC analysis on breath samples, remaining studies identified VOC patterns from tissue (1 study) or cell media samples (1 study).

VOC data analysis. For the discovery of unique biomarkers in complex chemical data sets, commonly multivariate methods such as partial least-squares to latent structures (PLS), orthogonal PLS (OPLS) or principal component analysis (PCA) are employed. Another technique for clustering complex data is the usage of hierarchical clustering analysis (HCA). Multiple statistical methods were employed, such as Mann-Whitney U test, Student's t -test and one-way ANOVA. Most studies set significant values if the P -value is below 0.05, rather rarely adjusted P -values were employed. Also, many different algorithms such as Random Forest (RF), Support Vector Machine (SVM), k -nearest neighbour have been employed to conduct Receiver Operator Characteristic (ROC) curve analysis. ROC analysis evaluates the specificity and sensitivity of the relevant

volatiles. For validation of the utilised models, cross-validation and permutation tests were mostly applied.

Quality assessment

Assessment of biases and applicability on outcomes using QUADAS-2 are specified in Figure 2, Supplemental Tables 7 and 8. First, the risk of bias was the highest for patient selection and index test. Elevated risk is obtained for the index test given that multiple studies did not validate their obtained results within the actual dataset nor in an additional, external, dataset. An induced risk of bias for patient selection was observed due to limited exclusion criteria, or no inclusion of 'benign conditions'. Furthermore, unclear risk of bias for 'flow and timing' was obtained given very few articles describe the time between the reference standard and the index test. There was no risk of bias for the reference standard, given that all research employed standardised methods for BC screening.

Second, for risk of applicability, a high risk for the reference standard, an elevated risk for patient selection and a low risk for the index test was determined. The high risk concerning the applicability of the reference standard can be deduced from the fact that limited studies subdivide BC into different subtypes, identifying subtype-specific BC volatiles and differentiating between different BC subtypes. Due to this practice in most selected research articles, BC is treated as a homogenous disease while various subtypes exist with potential different VOC signatures, impacting the applicability of the obtained results to other data sets. Finally, patient selection raises a red flag given that some studies compared VOC patterns from only female BC patients with a mixed-gender control group or significant differences were found between the age of the BC patient and control subjects.

Identified VOCs in different sampling matrices

Sixteen GC-MS case-control studies have been identified from which 10 studies analyse VOCs in breath, 4 in urine and

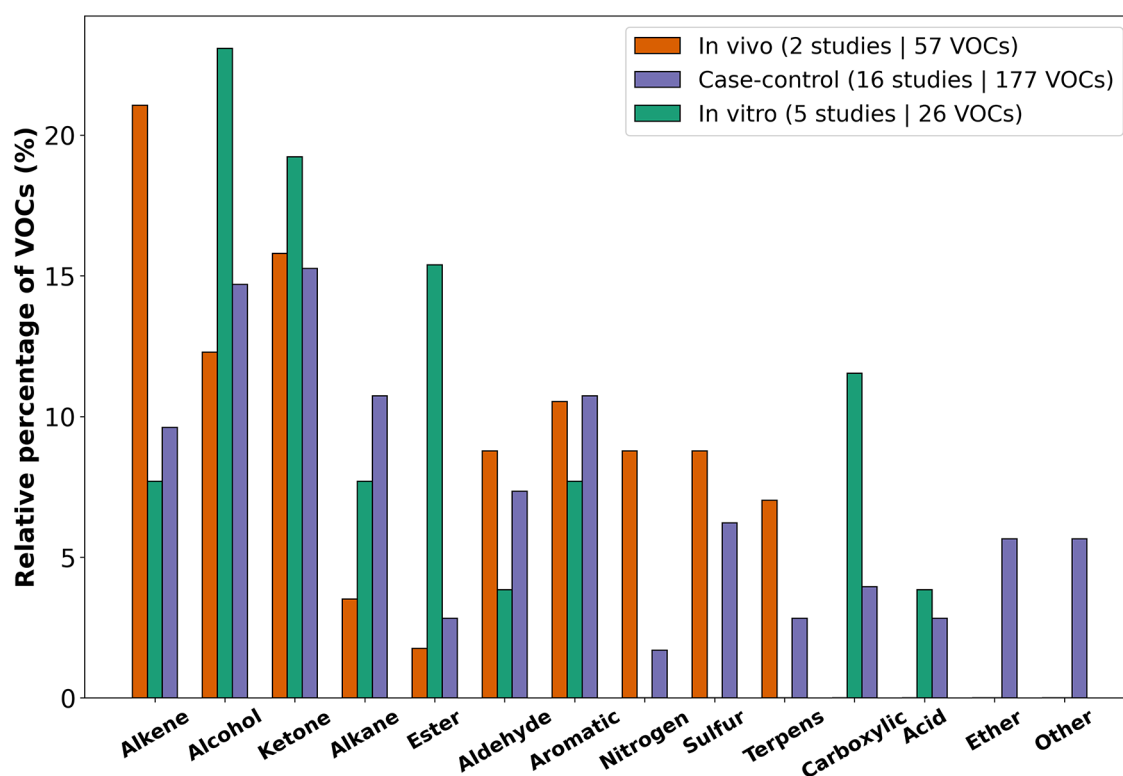


Figure 3. Chemical classification of BC-related VOCs in different experimental settings.

a single study for either tissue and saliva matrix, in which a total of 177 VOCs were associated with BC. About 15% of the identified VOCs were categorised as ketones or alcohols within case-control studies, closely followed by alkanes and aromatic compounds (Figures 3 and 4C).

Breath. Most GC-MS studies characterising BC-related volatometrics within this review were conducted on breath samples. From the 16 case-control studies, 10 studies determined VOCs in the headspace of breath samples from BC subjects using GC-MS (Figure 4D). Two out of the 10 studies were conducted in parallel research on an E-nose. Additionally, 4 out of 5 E-nose studies researched breath sampling and 2 out of 3 MS studies. Within the case-control studies, 3 studies came from the same research group.^{57,58,84} Despite BC analysis in breath matrices containing the largest number of publications,¹⁰ it consists of very little recent research (only 1 study 2020, the oldest study from 1997, median: 2011). Twelve studies from a total of 16 identified breath studies investigating end-tidal breath (=alveolar breath), 4 studies did not specify the breath category, and only 1 study investigated end-tidal expiration. On average, 54 patients were recruited for GC-MS breath studies, low below the average of 159 patients for E-nose breath studies. A maximum of 17 different significantly altered VOCs were identified within 1 study.⁵¹ A total of 62 VOCs have been identified in breath, distributed over 12 different chemical classes. Most VOCs are categorised as ketones, alcohols or alkanes (Figure 4D). Within those 62 VOCs, only 1 ketone

(cyclohexanone) and 1 alkane (n-heptanal) were identified in multiple studies, all other VOCs were uniquely identified.

Urine. Four articles investigated VOCs in the headspace of urine from BC patients and controls, with all 4 studies coming from similar collaborating members. A first study tested the hypothesis of whether BC can be deduced from the urinary metabolomic profile and obtained good discrimination between BC patients and control groups, however, a larger number of patients was required to confirm the results.⁶⁵ In the second study, a larger panel of urine samples from BC patients from a different population was recruited and validated the obtained results in an external cohort.⁶⁴ All studies characterised VOCs in first-morning urine. A total of 105 VOCs were found to be significantly altered within the 4 urine studies (Figure 4E). Up to almost 35% of the VOCs were categorised as ketone, followed by 30% of aromatic compounds and 20% of alcohols. Compared to breath studies, urine studies showed more interconnections within the same matrix (Figure 6). In 3 out of 4 studies, phenol and dimethyl disulphide were associated with BC odour. Furthermore, 5 compounds (p-cresol, acetic acid, 2-amylfuran, furan, guaiacol, 1,2-dihydro-1,1,6-trimethylnaphthalene) were simultaneously found in 2 out of 4 studies.

In vivo. Only 2 studies analysed VOCs in *in vivo* studies. Both studies were conducted by the same research group and identified urinary VOCs in mice after injection with 4T1.2 murine tumour cells in the mammary pad. A total of 57 VOCs were

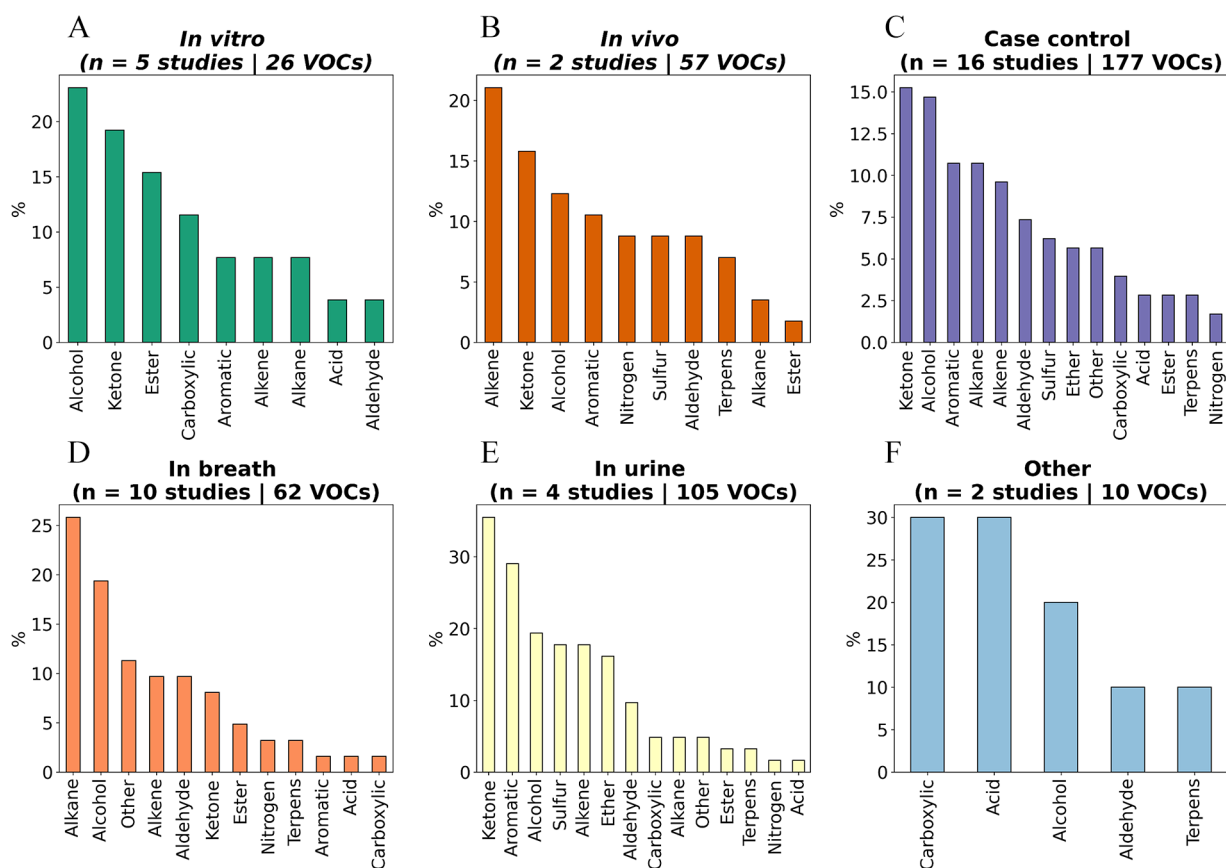


Figure 4. Chemical classification of BC-related GC-MS identified VOCs in different experimental settings and different matrices: (A) *in vitro*, (B) *in vivo*, (C) case-control studies, (D) breath, (E) urine, and (F) other studies representing 1 study in saliva and 1 study in BC tissue.

identified, with major groups being alkenes, ketones and alcohols. From the different experimental settings, the *in vivo* setting identified the most significantly altered terpenes compared to the *in vitro* and case-control settings (Figures 3 and 4B).

In vitro. Six different studies investigating BC smell on *in vitro* studies using GC-MS have been identified, wherefrom one study additionally employs an E-nose. Five out of 6 studies employ MCF-7 cancer cell lines and 4 out of 6 utilise MDA-MB-231 cells. An additional 5 other BC cell lines (SK-BR-3, BT-474, ZR75-1, T-47D and YMB-1) were used within the *in vitro* studies. Remarkably, few of the control cell lines were equal within the multiple studies: KMST-7, HMEC, CCD-1095sK or MCF-10a cell lines were used. Within the 6 different studies, 26 significantly altered VOCs were found with the largest category consisting of alcohols, followed by ketones and esters. Unlike all other experimental settings, no terpenes were identified to be significantly altered within the *in vitro* setting (Figures 3 and 4A).

Other sampling strategies. Other sampling matrices such as saliva and tissue have been employed to decipher BC-associated volatome signature using GC-MS technique. A total of 10 VOCs were identified within these 2 studies (Figure 4F).

Overlaps VOCs in the different matrices. Minimal overlaps in terms of VOCs can be observed between the different studied

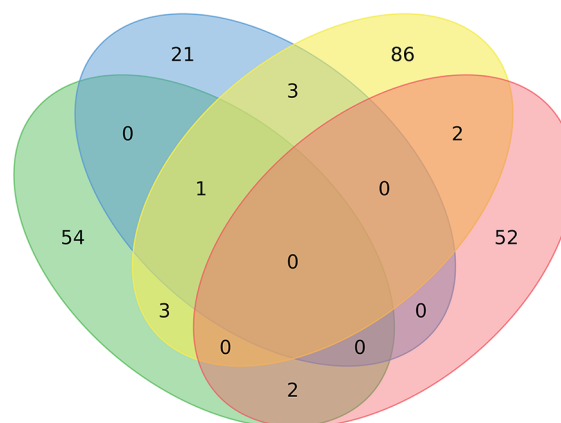
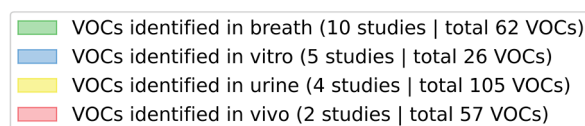


Figure 5. Venn diagram of the altered VOCs in the headspace of different matrices from BC subjects. The presented VOCs alter significantly in BC subjects compared to controls.

matrices characterising BC odour. Only one common alcohol has been identified between breath, urine and *in vitro* media, namely 2-ethyl-1-hexanol (Figure 5, Supplemental Table 5). Most VOC commonality has been found between cell media culture and urine, with 5 common compounds, including

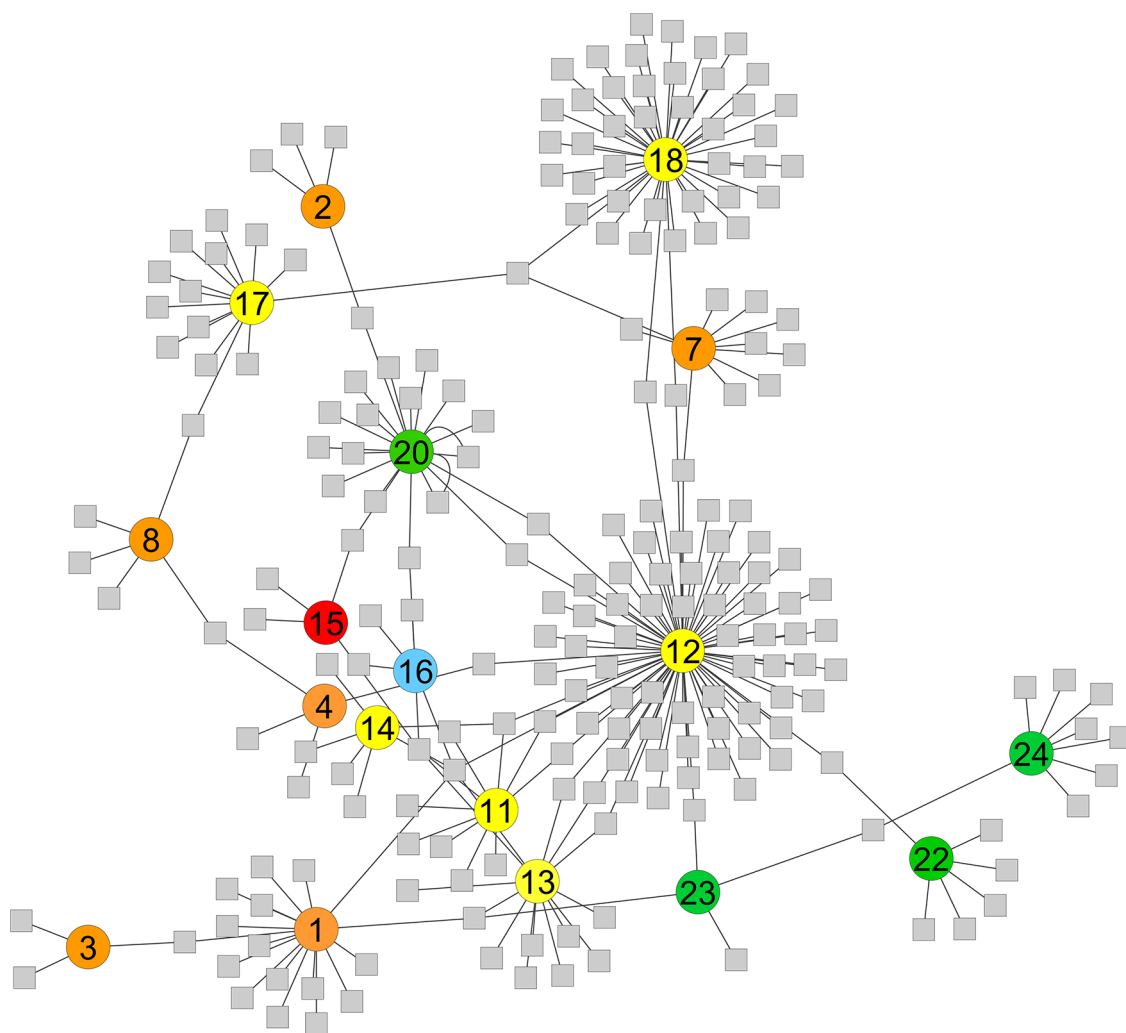


Figure 6. Network analysis of BC studies distribution based on detected volatiles. The number in the circles represents the number of the study, allotted similarly in Table 1; grey squares represent the number of significant altered VOCs. Different colours represent different matrices; yellow: urine, orange: breath, red: tissue, green: cell media, and blue: saliva. Studies that had no VOC connection with any other study were left out of the network.

p-xylene, 2-heptanone, p-tert-butylphenol and octanoic acid. Four inter-mutual VOCs have been characterised between breath and urine wherefrom 2 alcohols (2-ethyl-1-hexanol and phenol), 1 aromatic (durene) and 1 aldehyde (hexaldehyde). Only 2 alcohols have been commonly identified between BC-associated in breath or *in vivo* media (ethanol and 2-ethyl-1-hexanol). When displaying both inter- and intra-matrices VOCs, one can observe that no apparent commonality is obtained among either VOCs obtained within inter- or intra-matrices (Figure 6, Supplemental Table 5). Furthermore, a high divergence can be noted for the amount of significantly altered VOCs among the different studies.

Pathway analysis. Seven out of 32 studies selected studies conducted pathway analysis on significantly altered VOCs to facilitate interpretation of the modified biological processes (Supplemental Table 6). For this, 3 different pathway databases were employed: KEGG, Metaboanalyst and the Metabolome Database. In total, 33 pathways were identified to be affected over the 6 different studies, wherefrom 21 are unique pathways. Interestingly, multiple pathways were identified in several

studies; pyruvate metabolism was determined to be affected in 4 individual studies, whereas sulphur metabolism and fatty acid biosynthesis were both altered in 3 separate studies. Butanoate and tyrosine metabolism and glycolysis or gluconeogenesis are changed in 2 studies; all 16 other pathways are identified to be changed in solely 1 study and can be consulted in Supplemental Table 6.

Discussion

This systematic review aims to estimate the value of VOCs in different matrices for BC screening. Up to now, mammography is mainly employed to screen for BC; however, this screening method has restricted sensitivity and specificity. Compared to the practices described in the introduction, VOC detection in breath or body fluids is non-radiative, non-invasive and potentially economical. Furthermore, the accuracy of classification achieved by volatometric models suggests that VOC-profiling can be potentially used in BC diagnosis and management. Also, this technique could be employed to screen for BC in elderly women, given the little evidence of the risk-benefit ratio of mammography screening in this age group.

In the first part of the discussion, the altered VOCs in BC models/patients in the different matrices will be debated. Although the included studies have all obtained successful results on the use of expelled VOCs for the diagnosis of BC, there is considerable heterogeneity among the diagnostic identified VOCs. Therefore, in the second part, an analysis of possible reasoning behind this discrepancy is displayed.

Breast cancer-related VOCs

BC-related VOCs have been identified in multiple matrices such as breath, urine, saliva, tissue, in *in vivo* and *in vitro* experiments (4.1). The origin of most VOCs in various cancers has not yet been characterised. Comprehension of the metabolic pathways that lead to the generation or elimination of these VOCs will develop a better understanding of the biochemical alteration that arises in cancer (see 4.2). Major findings of VOCs in the matrices are discussed below.

Breath. Human exhaled breath is mainly composed of nitrogen, carbon dioxide, oxygen, water vapour as well as some inert gases.⁸⁵ In the past, more than 3000 different VOCs have been identified in the exhaled breath of healthy subjects.³¹ Many candidate volatiles in exhaled breath of BC subjects have been reported so far, unfortunately, no striking volatiles characterising the onset of BC have yet been elucidated. Of the 62 identified VOCs in breath studies, alkanes and alcohols were the most substantial groups identified (Figure 4D). Induced production of alkanes has been presumed to be due to the enhanced oxidative stress correlated with tumour advancement. The increased metabolism that accompanies tumour cell proliferation drives oxygen free radical production as well as the peroxidation of polyunsaturated fatty acids in membranes and, hence, the emission of alkanes.^{86,87}

Alcohols can be detected due to alcohol ingestion from beverages and food, which are further absorbed through the gastrointestinal tract and then liberated into the bloodstream. Besides, alcohol can be produced from the metabolism of hydrocarbons. Alcohol metabolism might be altered by confounding factors in the body, such as different amounts of fat or water content among different individuals.⁸⁸

Most studies were conducted on end-tidal breath within this BC volatomics review. This type of air is established to contain high levels of endogenous VOCs and minimal contaminants. Lawal et al³⁸ investigated the heterogeneity of breath sampling methods by conducting a review to identify the current state of the art in the breath-omics field. Lawal et al identified late expiratory breath as the most common breath type within a review of 110 breath volatomics articles. Research has demonstrated that the concentration of VOC is different in end-tidal breath and whole breath.⁸⁹ This research underlines the need for standardisation in the breath research community.

Urine. Four studies determined VOCs in the headspace of urine samples from BC patients. 105 VOCs were identified, with the major classes being ketones and aromatic compounds. The source of aromatic compounds is still unexplained and might be possible environmental contaminants.⁹⁰

Other sampling strategies. Other sampling matrices such as saliva and tissue have been employed to decipher BC-associated volatomatic signature.

Cavaco et al³⁰ investigated BC related volatomics in saliva and optimised several parameters that influence the HS-SPME, namely the sample ionic strength, the fibre, volume, pH, agitation, time and temperature of extraction. Furthermore, the same study compared volatomics patterns in saliva samples in 2 distinct geographic regions; in Portugal and India. Interestingly, distinct populations contain different abundant chemical families in their saliva; alkanes, esters and aldehydes were barely detected in Portuguese subjects, whilst they are abundant in the Indian populations. Variations in saliva VOC composition for the distinct groups analysed possibly contribute to multiple factors, with the genetic background and diet probably playing a major role.

Two studies were conducted on biopsy tissue samples. *Ex vivo* analysis of VOCs emitted by tissue samples supplies a unique chance to identify volatomatic patterns correlated with a BC state and can be considered as an additional source of information on volatile biomarkers found in breath, urine or saliva. It should be noted that variation in the form of tissue samples could affect headspace extraction and consequently the repeatability of the measurements. Additionally, the available tissue samples used^{61,79} were rather small (~10 mg or 100 mg respectively), which potentially influence the sensitivity of the analytical method employed and thereby the recognition of volatiles in trace concentrations.

Surprisingly, no case-control studies were found to investigate BC volatomics in either sweat, faeces or blood.

In vivo. Animal studies are a kind of central field between *in vitro* experiments and human clinical studies. The majority of animal studies employ laboratory-bred rats or mice that are mostly genetically identical. Limitations of the included *in vivo* volatomatic studies are that the experiments are carried out in a controlled environment on immunocompromised mice from the same chow infected with the same tumour cells, while greater metabolic heterogeneity will be present in human samples. Interestingly, urinary VOC panels primarily comprising volatile terpenes were able to differentiate healthy control mice from tumour-bearing mice.⁶⁶

In vitro. The analysis of VOCs formed by cancerous cell lines in their microenvironment as the originator of biomarkers should theoretically aid to resolve the question about the VOC origin. Cell lines have a benefit over other matrices in that the environmental variables are more controllable, making results

more easily interpretable and offering a lower cost and better reproducibility. This method can eliminate various confounding factors correlating with the analysis, such as sample collection, storage, manipulation process, and variation in patients' metabolic state, diet, age or gender. Despite this, the experimental setup does not assure that the assembled volatome signature is composed solely of endogenous compounds. There are possibly VOCs generated by the extraction devices, the sampling environment, or the surrounding culture medium. For example, Silva et al⁷¹ investigated the influence of pH on the VOCs identified from culture media and demonstrated that a different pH led to distinct dominant chemical groups in volatome patterns.

Twenty-six VOCs were identified to be related to BC in an *in vitro* setting. Most of those VOCs were categorised as ketones. Ketone production has been shown to drive tumour progression and metastasis.⁹¹ Also, cancer initiation and progression have been correlated with oxidative stress by enhancing DNA mutation or increasing DNA damage.⁹² Therefore, Liu et al⁷⁰ investigated elevated oxidative stress in BC cell lines and identified 15 VOCs that were statistically different between the induced oxidative stress compared to the control group.

A shortcoming of the conducted *in vitro* studies is that the conventional 2D cell culture environment does not resemble the *in vivo* condition. None of the included *in vitro* studies carried out headspace analysis of a 3D model, which better simulates the physiological growth of cancerous cells. Also, cells are handled in an atmospheric hyperoxic environment, while tumours grow in hypoxic conditions, altering cell metabolism and possibly the VOC assembly. Only one study conducted cell culture in hypoxic and normoxic conditions.⁶⁹ Taware et al⁶⁹ identified a total of 27 BC-specific VOCs wherefrom 9 were determined to be uniquely associated with the hypoxic environment, whereas 6 VOCs were specific for the normoxic condition, and 6 VOCs were found to be shared for both conditions. Therefore, *in vitro* research should cultivate cancer cells in hypoxic environments, using 3D models, to come closer to the *in vivo* setting.

Overlaps VOCs in different matrices. Very few overlaps of BC-correlated VOCs have been obtained between the different matrices. 2-Ethyl-1-hexanol was identified as a common VOC between breath, urine and *in vitro* media (Figure 5, Supplemental Table 5). 2-Ethyl-1-hexanol has been demonstrated to play an important role in discrimination between normal lung cell lines and lung cancer cell lines⁹³ and therefore seems rather a general biomarker for cancer and unspecific to BC. Also, the 5 commonly characterised VOCs between cell media and urine have been identified as discriminating VOCs between the healthy and sick states in different cancers. For example, 2-heptanone and p-tert-butylphenol, both characterised in urine and *in vitro* media, have been identified as significantly altered in a

gastric cancer cell line or urine of head and neck carcinoma patients, respectively.^{94,95} Note that many studies obtained a good discrimination on VOC signature between healthy subjects and the evaluated disease, although one of the major limitations is the shortfall of the certainty of discrimination between different illnesses. Interestingly, Rodríguez-Aguilar et al⁹⁶ evaluated the discrimination capacity in chronic obstructive pulmonary disease, lung cancer, BC by chemoresistive sensors and identified a characteristic chemical fingerprint of each disease. Taken together, it should be kept in mind that various of the VOCs for BC have been reported for other diseases and this is a significant hindrance to the translation of VOC analysis to the clinic.

BC-specific VOC pathway analysis

Pathway analysis methods aid researchers identify the biological function of gene sets or volatiles within malignant tissues, thereby stimulating new cancer treatments' design.⁹⁷ Unfortunately, few studies are designed to research the molecular mechanisms and exhalation kinetics underpinning the production of volatile biomarkers in cancer. Such research is necessary to comprehend the components that impact test performance and clarify the diagnostic model. Researchers have hypothesised that VOCs of BC relate to modified oestrogen metabolism mechanisms. This assumption is supported by the fact that oestrogen stimulation can induce the proliferation of both normal and neoplastic mammary epithelial cells.⁹⁸

A defining hallmark of cancer is unrestrained cell proliferation. The proliferation is established once cells have assembled modifications in signalling pathways that control both proliferation and metabolism, wherein the metabolic alterations supply the energetic and anabolic requirements of the induced cell proliferation.⁹⁹ Four individual studies out of 6 studies that conducted pathways analysis demonstrated that pyruvate metabolism was affected in BC-state. Pyruvate has been proved to be one of the key metabolites controlling stem cell function by partly maintaining self-renewal or differentiation. Importantly, medication adjusting pyruvate metabolism was described to modify the balance between differentiation states and self-renewal.¹⁰⁰

Furthermore, sulphur metabolism was shown to be affected in 3 separate VOC studies. Sulphur is derived from the dietary consumption of the proteinogenic amino acids cysteine and methionine, as only lower organisms (plants, fungi, bacteria) can synthesise them *de novo*. The constitution of these sulphur-containing amino acids gives rise to intermediate metabolites that establish the cellular antioxidant system, aid the epigenetic regulation, and intervene in intra- and intercellular signalling, all of which contribute to tumorigenesis.¹⁰¹ Next, fatty acid biosynthesis was found to be affected in 3 separate VOC studies. Fatty acids can serve as fuel sources for energy production in cancerous cells; they serve as vital secondary messengers,

have a principal role as structural components in the membrane matrix and are, therefore, involved in the mechanisms through which cancer cells rewire their metabolism.¹⁰² Because cancer cells have aberrant cell growth and proliferation, they need to accumulate additional metabolic intermediates for cellular building blocks; cancer cells, therefore, exhibit often a shift towards fatty acid synthesis.¹⁰³

Experimental set-up and VOC sampling

In common with various research fields, there is a clash between innovation and standardisation. To date, there is no uniform standard procedure for candidate tumour markers in expelled VOCs from fluids from BC patients. Standardisation of VOC collection and analysis is essential. The little VOC overlaps reveals that the results of VOCs testing depend on various factors related to the patients' physiologic condition, sample collection as well as test environment among others.

Patient selection. Most VOC research highlights the non-invasive nature of VOC sampling which enhances patient acceptability. However, the composition of expelled matrices is influenced by multiple factors, such as hereditary features (gender, nature, diseases, age), as well as behavioural habits (drinking, smoking, . . .). Haze et al³³ demonstrated that 2-nonenal, an unsaturated aldehyde, was detected only in older subjects (40 years or older) and suggested that this change could be due to alterations in skin surface lipids with ageing. Therefore, a source of bias within volatometrics studies is the presence of confounding factors that should be verified within the analysis. Confounders are characteristics that are unequally dispersed between the BC state and healthy controls and can potentially be the underlying reason for the observed differences. Different strategies can be employed to limit confounding factors at the design stage of the experiment. The first strategy for confounder limitation can be obtained by matching control groups to the BC group so that they are equilibrated for potential confounders. Second, experimental conditions should stay constant during all the experiments, such as location and instrument. Third, experiments should be randomised for expected confounders that can be regulated. External validation of the obtained results is the ultimate technique to verify the validity of the predictive model. Clinical confounding factors can be of high importance for VOC generation such as the patient's characteristics, the medical conditions as well as the taken medication. Another disease, such as inflammation or infection in different tissues, may induce a confounding source of oxidative stress and skew the results of the VOC test.⁸⁴ Therefore, it is crucial to get very strict and well-defined inclusion and exclusion criteria where control subjects correspond as much as possible to the subjects of interest. Various articles within this systematic review do not match the age, gender and lifestyle of the healthy controls with the cancer patients and

this might greatly bias the obtained results. Also, the number of recruited subjects is highly variable, with a low sample size potentially highly biasing the obtained results.

Besides, all patients should have conducted reference standard diagnostic testing and both the reference and index test should be performed well before therapeutic intervention. These elements should be accurately stated in the diagnostic study, despite most research included within this systematic review does not report the reference standard test or acknowledge the interval between the reference and index test.

Finally, most research identified within this systematic review includes patients with early as well as advanced BC stages. Even though it is mainly hypothesised that VOC testing might be implemented in the detection of early-stage cancers. In the future, studies should determine the precise role of the tumour stage and other factors, including the association of histologic and molecular subtypes with the diagnostic accuracy of proposed VOC tests.⁵² For this, a threshold for separating patients with cancer of different stages and tumour subtypes is required before embarking on masked validation studies.

Standardisation of sampling. The different sampling methods are prone to different biases. It is important to take into consideration the type and reliability of collection methods as well as the environment. Also, the factors that affect the transfer (time and temperature) of the samples to the laboratory of analysis should be carefully taken into consideration. Research should state clearly the stability of the target VOCs within the collection methods to identify the period before losing VOCs at a certain storage temperature. Also, the timing of sampling during the day could be of importance for certain body fluids such as urine.¹⁰⁴ Many discrepancies have been found in the level of storage conditions, storage time, and storage container (see Supplemental Table 2). Only 1 study described clearly to have researched VOC conservation.⁵²

Not only a good selection of patients among the target population but also a selection of the environment where the VOC test will be performed needs to be carefully picked. A great quantity of VOCs in the diverse matrices are generated from a pure exogenous source, which is neither of bacterial nor human origin. Only compounds produced internally in the body can be examined as biomarkers, which is problematic given the origin of most volatiles is still unknown.¹⁰⁵ Even if internal body fluids are analysed (eg, urine), it does not guarantee the endogenous origin of the analytes, given various inhaled VOCs may dissolve in the blood,¹⁰⁶ be stored in various body compartments and are potentially later excreted. Many exogenous VOCs can be deduced from occupational exposure, environmental pollutants or household chemicals. Therefore, environmental VOCs within the location where the sampling was performed and laboratory air measurements should be taken into consideration. Interestingly, multiple studies constructed a subtraction chromatogram where the abundance of VOCs in

breath was determined minus their abundance in ambient room air.^{54,57}

Another potential strategy for BC diagnosis via VOC is to exploit the volatomics signature by biomarkers found in various sample types. Hanai et al¹⁰⁷ investigated whether unique VOCs could be distinguished in the culture medium of lung cancer cell line in complement to the urine of mice implanted with the same cell line, however, both sample types produced distinct biomarker profiles.

VOC detection techniques

260 VOCs have been identified in the different matrices of BC patients using GC-MS. Nevertheless, in the included studies, only 25 VOCs were detected in more than one study, indicating the volatomics field's low repeatability. The most common identified VOC used to screen BC was phenol, which was detected in 5 different studies, followed by acetic acid, significantly altered in 4 individual studies. Nowadays, internal validation of obtained results is becoming a standard procedure, and as the participant count increases in the more recent studies, the risk of overfitting the diagnostic models will further be diminished. At the same time, the identification of which features of the complex datasets should be enclosed within the analysis has ameliorated. However, compound characterisation continues to be insufficient with minimal studies verifying the putative identification against chemical standards. The chemical identification of the compounds shown should be regarded as tentative given their chemical structures were inferred from the similarity between their obtained mass spectra and mass spectra within the computer-based NIST library, especially when low-match scores are employed. Although broadly utilised as an analytical tool, this approach is susceptible to error. Future research will be necessary to validate the chemical identities of candidate VOC biomarkers of BC by comparing their mass spectra and retention times to those achieved from pure compounds. Better characterisation of chemical compounds will allow the identification of the biological origins of the established distinguishing VOCs. A potential resolution to this problem is to investigate the matrices' mass ions as potential candidate biomarkers of BC. Mass ion may provide a potentially more robust biomarker given it allows diagnosis by pattern recognition without the need for chemical identification of the underlying VOCs.

Furthermore, the observed inter-study variability stated within this review may be linked to instrumental variabilities, such as operator or different location. Comparison of GC-MS data between different laboratories can be challenging given the changeable nature of the technical equipment. Common reasons for reproducibility problems are hypersensitivity to instrument conditions, presence of non-automated procedures and instrumental shifts in time. Potential confounders may be prevented by cautious experimental design. In untargeted

volatomics research where key metabolites are not a priori known it is challenging to keep the sensitivity of the analytical procedure stable. Possible solutions include the utilisation of calibration methods based on both internal as well as external standards to identify correct analytical variation. Therefore, it is suggested to quality control samples to quantitatively measure the variation within and between experiments. Ideally, to assure robustness beyond a particular research laboratory, would be to conduct multicentre studies.

There is dissension regarding the necessary knowledge to accomplish successful VOC diagnostic testing. One can argue that as long as the VOC pattern between diseased and healthy states can be distinguished, there is no requirement to comprehend the origin of the established volatiles. Other experts simultaneously identify environmental VOC contaminants given they might be erroneously appointed as endogenous compounds.⁵⁴ While this may be a point of bias in limited cohort studies, this might minorly affect the results in large patient cohorts. Therefore, the quality of analysis between the comprised research papers is discordant and hampered by the small number of participants in multiple studies. External validation experiments against multicentre studies will be essential to confirm internal and external validity and inform clinical applications.

An alternative procedure to the potential exogenous origin of proposed VOC analysts targets the detection of important, previously demonstrated, chemical groups such as aldehydes solely as makers of cancer.¹⁰⁸ Research has demonstrated that oxidative stress is one of the major causes of cancer progression via the up-regulated production of reactive oxygen species and nitrogen species inducing mutations.¹⁰⁹ Both hydrocarbons and aldehydes are known to be related to oxidative stress, given they are products of lipid peroxidation, but the precise procedure of their appearance in body fluids and breath remains unclear.¹¹⁰

Another possibility for VOC detection can be an electronic sensor device, due to its potential small size and cheaper cost. However, electronic sensors also have several inconveniences depending on the nature of the detector serving as the sensing part of the device. For all sensors, there is a need for controlled airflow, as well as other environmental variables should be regulated including temperature and humidity. Sensor drift, particularly when exposed to multiple gas mixtures, should be taken into account.

Animal BC detection studies have been left out of this review. Despite this, dogs have proven to perceive VOCs related to cancer because of their acute sense of smell. Currently, canine clinical BC-detection studies are ongoing to evaluate the efficacy of such a diagnostic method.¹¹¹ The canine method has multiple advantages as a potential cancer-screening approach, given its ease of testing and interpretation of results. Major limitations of this method are related to the fact that it is still unknown to which chemical compounds or combinations of

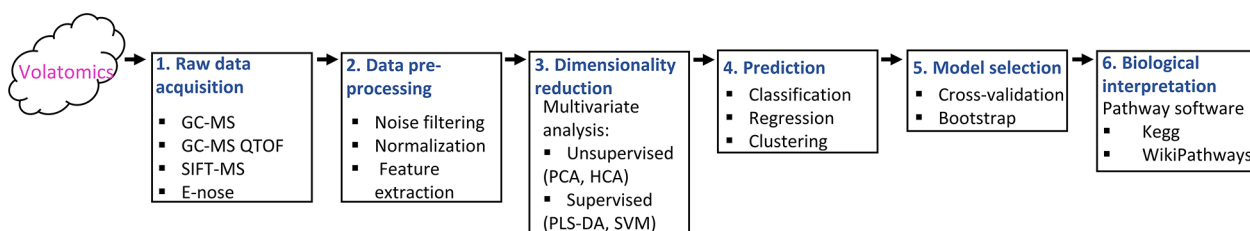


Figure 7. Workflow for volatome analysis. First, measuring devices are needed (eg, GC-MS, GC-MS QTOF, SIFT-MS, E-nose) and the acquired data needs to be pre-processed (eg, normalisation, noise filtering, feature extraction). Next, data pre-processing enhances data quality by discarding variance and bias. Feature selection is employed to characterise the variables that have the predictive capacity for the condition of interest. Further, dimensionality reduction techniques mostly need to be utilised to avoid problems related to the curse of dimensionality. Various statistical techniques are employed to detect patterns in the large data sets with either regard to sample classification (supervised) or inconsiderate of the sample classification (unsupervised). After, a predictive model is built on a certain set of samples (training set) and evaluated on another set (validation set) of samples. Model selection is necessary to pick the optimum model with the best performance. Finally, the biological connotation of the obtained results can be given by the usage of metabolic databases such as KEGG database.

compounds the dogs react. The detection accuracy fluctuates due to failure in the conditioning of the dogs or confounding factors. Canine studies would highly benefit from VOC biomarker identification in the scope to synthesise ‘BC odour’ to produce reproducible training samples for the dogs.¹¹² Taken together, for scientific purposes, it is rather best to employ the different methods (analytical chemistry, E-noses, animal studies) in complement.

VOC data analysis

The major objective of measuring volatiles in different matrices is the characterisation of a specific pool that, at a defined concentration, shows a high level of significance as a putative biomarker for early-stage diagnosis of the disease of interest. The data analysis part is becoming of increasing importance in omics-research with the standardised process depicted in Figure 7. Current detection techniques employed for volatome detection generate enormous possibilities for signal recording, which promotes that the data analysis, as well as the interpretation, might become the bottleneck in the process. In multiple studies, when obtained results cannot be reproduced by different research groups, statistical bias is put forward as a potential explanation for the original over-optimistic results. In many studies included within this systematic review, a *P*-value is employed as a univariate measure of significance for the discrimination of metabolites between BC and healthy conditions. Application of the same statistical test to each of the VOCs in the data would result in an elevated number of false positives (called type I errors). To diminish the chances for type I errors, methods for multiple testing should be employed. A recommended alternative option is to construct predictive classification models to determine the predictive power of the putative biomarkers. For this, the calculated specificity and sensitivity should satisfy the clinical demands of usage.

Unfortunately, in cancer research, most data are derived from small pilot studies which can lead to data overfitting, which consists of modelling noise instead of information.

Overfitting is one of the main components by which some research performs with high performance but does not present a reproducible outcome.¹¹³ In many studies with a low subject number, the data contains much higher dimensionality than the sample count, inducing the overfitting of the model. As a result, the model will fail to adapt to additional data, and this possibly will alter the accuracy of predicting future observations. Therefore, overfitting must be adjusted by the use of validation techniques, adapting the model to the data’s complexity. Partial Least Squares-Discriminant Analysis (PLS-DA) models are prone to overfit¹¹⁴ and verification should be performed. Internal validation of the model mostly consists of partitioning the data into different training and testing sets, whereby the model is built on the training set and its performance is verified on the testing set. Validation of the included predictive models typically relies mostly on cross-validation, Monte Carlo subsampling or bootstrapping. Again, most included research within this systematic review does not include external validation, which is the most adopted method to acquire an unbiased evaluation of the model. External validation includes testing the model both in different instrumental and sampling conditions. Irreproducibility of volatome results is highly correlated with current practices in model validation since external validation is uncommon.

Conclusion and Future Directions

The major aim of volatome studies in clinical research is to enable the detection of a disease of interest with high certainty, noninvasively, potentially cheaper, simpler way and with minimal health risk. Research of the ‘smell of cancer’ is really elegant in the simplicity of the concept; despite there are still restraints to applying this concept clinically. GC-MS is still the gold standard method to analyse VOC biomarkers. However, it is not a simple method to implement in the clinical setting given its non-portability and the expenses that come along. Therefore, it is more probable that, once a disease-associated volatile signature is identified, the forthcoming approach to be used in clinical studies might be the E-nose or animal detection

method due to its ease of testing over GC-MS. For research purposes, it thus seems ideal to use different methods in complement.

Obtained results within this systematic review demonstrate that many VOCs were identified in different body fluids/breath of BC patients and in *in vitro* cell BC-related lines, despite very few of these VOCs could be detected in multiple studies and/or matrices. Therefore, large-scale evaluation studies, preferably with subjects from different countries and ethnicities, are necessary to characterise common VOC profiles in the different matrices. To further foster implementation in practice, there is an urgent need for minimum reporting standards, technical standards, uniform sample processing and standardised data analysis. Additionally, longitudinal multi-centre clinical trials, as well as external validation in independent population samples, will be key for further advancements. Finally, the elucidation of the biochemical pathways involved in VOC production to establish more detailed information on the involved BC mechanisms will be necessary for further research. Armed with these tools, the next generation of volatometrics evidence should show the full potential of expelled VOCs and aid canine cancer detection research.

Author Contributions

ML and IF developed the concept. IF, EA and ML designed the study. ML and PB searched the publications and reviewed the articles. ML, VC and PB contributed to the interpretation of the data and drafts of the manuscript. ML, IF, EA, PB and VC critically reviewed the manuscript and approved the submission.


Author's Information

PB and IF are members of KDOG programme, a research project lead by Institut Curie aiming at evaluating the capacity of trained dogs to detect breast cancer from skin secretion samples. This programme receives funding and support from Royal Canin Company and Seris Security Company, and private donors.

Availability of Data and Materials

All data generated or analysed during this study are included in this published article.

ORCID iD

Michelle Leemans  <https://orcid.org/0000-0002-2028-2610>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- American Cancer Society. *Breast Cancer Facts & Figures 2019–2020*. American Cancer Society, Inc. 2019.
- Hortobagyi GN, de la Garza Salazar J, Pritchard K, et al. The global breast cancer burden: variations in epidemiology and survival. *Clin Breast Cancer*. 2005;6:391-401.
- Cserni G, Chmielik E, Cserni B, Tot T. The new TNM-based staging of breast cancer. *Virchows Arch*. 2018;472:697-703.
- Rosen RD, Sapra A. *TNM Classification*. StatPearls; 2021. Accessed September 13, 2021. <http://europepmc.org/books/NBK553187>
- Hölzel D, Eckel R, Bauerfeind I, et al. Survival of de novo stage IV breast cancer patients over three decades. *J Cancer Res Clin Oncol*. 2017;143:509-519.
- Rezo A, Dahlstrom J, Shadbolt B, Rodins K, Zhang Y, Davis AJ. Tumor size and survival in multicentric and multifocal breast cancer. *Breast*. 2011;20:259-263.
- Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast Cancer subtypes based on ER/PR and her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res*. 2009;7:4-13.
- Momenimovahed Z, Salehiniya H. Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer Target Ther*. 2019;11:151-164.
- Mehrgou A, Akouchekian M. The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. *Med J Islam Repub Iran*. 2016;30:369.
- Martin AM, Weber BL. Genetic and hormonal risk factors in breast cancer. *J Natl Cancer Inst*. 2000;92:1126-1135.
- Göttsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2013;2013:CD001877.
- Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M, et al. The benefits and harms of breast cancer screening: an independent review. A report jointly commissioned by Cancer Research UK and the Department of Health (England) October 2012. *Br J Cancer*. 2013;108:2205.
- Sprague BL, Arao RF, Miglioretti DL, et al. National performance benchmarks for modern diagnostic digital mammography: update from the breast cancer surveillance consortium. *Radiology*. 2017;283:59-69.
- Houssami N, Abraham LA, Miglioretti DL, Sickles EA, Kerlikowske K, Buist DSM, et al. Accuracy and outcomes of screening mammography in women with a personal history of early-stage breast cancer. *JAMA*. 2011;305:790-799.
- WHO. WHO position paper on mammography screening. *Oncol Clin Pract*. 2015;11:16-19.
- Dobi Kelemen G, Kaizer L, Weiczner R, Thurzó L, Kahán Z. Breast cancer under 40 years of age: increasing number and worse prognosis. *Pathol Oncol Res*. 2011;17:425-428.
- Cosedine NS, Magai C, Krivoshekova YS, Ryzewicz L, Neugut AF. Anxiety, worry, and breast cancer screening behavior: a critical review. *Cancer Epidemiol Biomarkers Prev*. 2004;13:501-510.
- Radhakrishna S, Agarwal S, Parikh PM, et al. Role of magnetic resonance imaging in breast cancer management. *South Asian J Cancer*. 2018;7:069-071.
- Menezes GLG, Knuttel FM, Stehouwer BL, Pijnappel RM, Van Den Bosch MAAJ. Magnetic resonance imaging in breast cancer: a literature review and future perspectives. *World J Clin Oncol*. 2014;5:61-70.
- Galukande M, Kiguli-Malwadde E. Rethinking breast cancer screening strategies in resource-limited settings. *Afr Health Sci*. 2010;10:89.
- Rivera-Franco MM, Leon-Rodriguez E. Delays in breast cancer detection and treatment in developing countries. *Breast Cancer Basic Clin Res*. 2018;12:1178223417752677.
- Sood R, Rositch AF, Shakoob D, et al. Ultrasound for breast cancer detection globally: a systematic review and meta-analysis. *J Glob Oncol*. 2019;2019:1-17.
- Porter R. *The greatest benefit to mankind: A medical history of humanity from antiquity to the present*. Harper-Collins. 1999.
- Bobak CA, Kang L, Workman L, et al. Breath can discriminate tuberculosis from other lower respiratory illness in children. *Sci Rep*. 2021;11:1-9.
- Weber R, Haas N, Baghdasaryan A, et al. Volatile organic compound breath signatures of children with cystic fibrosis by real-time SESI-HRMS. *ERJ Open Res*. 2020;6:00171-02019.
- Bond A, Greenwood R, Lewis S, et al. Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer. *Aliment Pharmacol Ther*. 2019;49:1005-1012.
- Sethi S, Nanda R, Chakraborty T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin Microbiol Rev*. 2013;26:462-475.
- Monedeiro F, dos Reis RB, Peria FM, Sares CTG, De Martinis BS. Investigation of sweat VOC profiles in assessment of cancer biomarkers using HS-GC-MS. *J Breath Res*. 2020;14:026009.
- Wang C, Li P, Lian A, et al. Blood volatile compounds as biomarkers for colorectal cancer. *Cancer Biol Ther*. 2014;15:200-206.
- Cavaco C, Pereira JAM, Taunk K, et al. Screening of salivary volatiles for putative breast cancer discrimination: an exploratory study involving geographically distant populations. *Anal Bioanal Chem*. 2018;410:4459-4468.
- Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, Cataneo RN. Variation in volatile organic compounds in the breath of normal humans. *J Chromatogr B Biomed Sci Appl*. 1999;729:75-88.
- Penn DJ, Oberzaucher E, Grammer K, et al. Individual and gender fingerprints in human body odour. *J R Soc Interface*. 2007;4:331-340.
- Haze S, Gozu Y, Nakamura S, et al. 2-Nonenal newly found in human body odor tends to increase with aging. *J Invest Dermatol*. 2001;116:520-524.

34. Mebazaa R, Rega B, Camel V. Analysis of human male armpit sweat after fenu-greek ingestion: characterisation of odour active compounds by gas chromatography coupled to mass spectrometry and olfactometry. *Food Chem.* 2011;128:227-235.
35. Prokop-Prigge KA, Greene K, Varallo L, Wysocki CJ, Preti G. The effect of ethnicity on human axillary odour production. *J Chem Ecol.* 2016;42:33-39.
36. Furton KG, Caraballo NI, Cerreta MM, Holness HK. Advances in the use of odour as forensic evidence through optimizing and standardizing instruments and canines. *Philos Trans R Soc B Biol Sci.* 2015;370:20140262.
37. Janfaza S, Khorsand B, Nikkhab M, Zahiri J. Digging deeper into volatile organic compounds associated with cancer. *Biol Methods Protoc.* 2019;4:1-11.
38. Lawal O, Ahmed WM, Nijssen TME, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis. *Metabolomics.* 2017;13:110.
39. Wallace MAG, Pleil JD. Evolution of clinical and environmental health applications of exhaled breath research: review of methods and instrumentation for gas-phase, condensate, and aerosols. *Anal Chim Acta.* 2018;1024:18-38.
40. Maiti KS, Roy S, Lampe R, Aponolski A. Detection of disease-specific volatile organic compounds using infrared spectroscopy. *Eng Proc.* 2021;8:15.
41. Smith D, Španěl P. The challenge of breath analysis for clinical diagnosis and therapeutic monitoring. *Analyst.* 2007;132(5):390-396.
42. Eckert K, Carter D, Perrault K. Sampling dynamics for volatile organic compounds using headspace solid-phase microextraction arrow for microbiological samples. *Separations.* 2018;5:45.
43. Berrou K, Dunyach-Remy C, Lavigne JP, Roig B, Cadere A. Comparison of stir bar sorptive extraction and solid phase microextraction of volatile and semi-volatile metabolite profile of *Staphylococcus aureus*. *Molecules.* 2019;25:55.
44. Drabińska N, Mlynarz P, de Lacy Costello B, et al. An optimization of liquid-liquid extraction of urinary volatile and semi-volatile compounds and its application for gas chromatography-mass spectrometry and proton nuclear magnetic resonance spectroscopy. *Molecules.* 2020;25:3651.
45. Baldini C, Billeci L, Sansone F, Conte R, Domenici C, Tonacci A. Electronic nose as a novel method for diagnosing cancer: a systematic review. *Biosensors.* 2020;10:84.
46. Karakaya D, Ulucan O, Turkan M. Electronic nose and its applications: a survey. *Int J Autom Comput.* 2020;17:179-209.
47. Swanson B, Fogg L, Julion W, Arrieta MT. Electronic nose analysis of exhaled breath volatiles to identify lung cancer cases: a systematic review. *J Assoc Nurses AIDS Care.* 2020;31:71-79.
48. Wilson AD. Applications of electronic-nose technologies for noninvasive early detection of plant, animal and human diseases. *Chemosensors.* 2018;6:45.
49. Pirrone F, Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. *J Vet Behav.* 2017;19:105-117.
50. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol.* 2009;6:e1000097.
51. Zhang Y, Guo L, Qiu Z, Lv Y, Chen G, Li E. Early diagnosis of breast cancer from exhaled breath by gas chromatography-mass spectrometry (GC/MS) analysis: a prospective cohort study. *J Clin Lab Anal.* 2020;34:1-10.
52. Barash O, Zhang W, Halpern JM, et al. Differentiation between genetic mutations of breast cancer by breath volatolomics. *Oncotarget.* 2015;6:44864-44876.
53. Wang C, Sun B, Guo L, Wang X, Ke C, Liu S, et al. Volatile organic metabolites identify patients with breast cancer, cyclomastopathy, and mammary gland fibroma. *Sci Rep.* 2014;4:5383.
54. Li J, Peng Y, Liu Y, et al. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin Chim Acta.* 2014;436:59-67.
55. Mangler M, Freitag C, Lanowska M, Staack O, Schneider A, Speiser D. Volatile organic compounds (VOCs) in exhaled breath of patients with breast cancer in a clinical setting. *Ginekol Pol.* 2012;83:730-736.
56. Peng G, Hakim M, Broza YY, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br J Cancer.* 2010;103:542-551.
57. Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P, Wai J. Volatile biomarkers in the breath of women with breast cancer. *J Breath Res.* 2010;4:026003.
58. Phillips M, Cataneo RN, Ditkoff BA, et al. Prediction of breast cancer using volatile biomarkers in the breath. *Breast Cancer Res Treat.* 2006;99:19-21.
59. Phillips M, Cataneo RN, Ditkoff BA, Bevers TB. Volatile markers of breast cancer in the breath. *Breast Dis.* 2004;15(1):53.
60. Ebeler SE, Clifford AJ, Shibamoto T. Quantitative analysis by gas chromatography of volatile carbonyl compounds in expired air from mice and human. *J Chromatogr B Biomed Appl.* 1997;702:211-215.
61. Silva C, Perestrelo R, Silva P, Capelinha F, Tomás H, Câmara JS. Volatome pattern of breast cancer and cancer-free tissues as a powerful strategy to identify potential biomarkers. *Analyst.* 2019;144:4153-4161.
62. Silva CL, Perestrelo R, Silva P, Tomás H, Câmara JS. Implementing a central composite design for the optimization of solid phase microextraction to establish the urinary volatome expression: a first approach for breast cancer. *Metabolomics.* 2019;15:1-13.
63. Porto-Figueira P, Pereira JAM, Câmara JS. Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatome biosignature. *Anal Chim Acta.* 2018;1023:53-63.
64. Taunk K, Taware R, More TH, et al. A non-invasive approach to explore the discriminatory potential of the urinary volatome of invasive ductal carcinoma of the breast. *RSC Adv.* 2018;8:25040-25050.
65. Silva CL, Passos M, Câmara JS. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—a powerful strategy for breast cancer diagnosis. *Talanta.* 2012;89:360-368.
66. Woollam M, Teli M, Liu S, et al. Urinary volatile terpenes analyzed by gas chromatography-mass spectrometry to monitor breast cancer treatment efficacy in mice. *J Proteome Res.* 2020;19:1913-1922.
67. Woollam M, Teli M, Angarita-Rivera P, et al. Detection of volatile organic compounds (VOCs) in urine via gas chromatography-mass spectrometry QTOF to differentiate between localized and metastatic models of breast cancer. *Sci Rep.* 2019;9:1-12.
68. Tanaka M, Hsuan C, Oeki M, et al. Identification of characteristic compounds of moderate volatility in breast cancer cell lines. *PLoS One.* 2020;15:e0235442.
69. Taware R, Taunk K, Kumar TV, et al. Extracellular volatome alterations induced by hypoxia in breast cancer cells. *Metabolomics.* 2020;16:21.
70. Liu Y, Li W, Duan Y. Effect of H₂O₂ induced oxidative stress (OS) on volatile organic compounds (VOCs) and intracellular metabolism in MCF-7 breast cancer cells. *J Breath Res.* 2019;13:036005.
71. Silva CL, Perestrelo R, Silva P, Tomás H, Câmara JS. Volatile metabolomic signature of human breast cancer cell lines. *Sci Rep.* 2017;7:43969.
72. Huang Y, Li Y, Luo Z, Duan Y. Investigation of biomarkers for discriminating breast cancer cell lines from normal mammary cell lines based on VOCs analysis and metabolomics. *RSC Adv.* 2016;6:41816-41824.
73. Lavra L, Catini A, Ulivieri A, et al. Investigation of VOCs associated with different characteristics of breast cancer cells. *Sci Rep.* 2015;5:13246.
74. Phillips M, Cataneo RN, Lebauer C, Mundada M, Saunders C. Breath mass ion biomarkers of breast cancer. *J Breath Res.* 2017;11:016004.
75. Martinez-Lozano Sinues P, Landoni E, Miceli R, et al. Secondary electrospray ionization-mass spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study. *J Breath Res.* 2015;9:031001.
76. He J, Sinues PML, Hollmén M, Li X, Detmar M, Zenobi R. Fingerprinting breast cancer vs. normal mammary cells by mass spectrometric analysis of volatiles. *Sci Rep.* 2015;4:1-6.
77. Yang HY, Wang YC, Peng HY, Huang CH. Breath biopsy of breast cancer using sensor array signals and machine learning analysis. *Sci Rep.* 2021;11:1-9.
78. Diaz de León-Martínez L, Rodríguez-Aguilar M, Gorocica-Rosete P, et al. Identification of profiles of volatile organic compounds in exhaled breath by means of an electronic nose as a proposal for a screening method for breast cancer: a case-control study. *J Breath Res.* 2020;14:046009.
79. Venkatraman P, Sureka CS. An in-vitro study for early detection and to distinguish breast and lung malignancies using the PCB technology based nanodosimeter. *Sci Rep.* 2019;9:1-16.
80. Konvalina G, Haick H. Effect of humidity on nanoparticle-based chemiresistors: a comparison between synthetic and real-world samples. *ACS Appl Mater Interfaces.* 2012;4:317-325.
81. Shuster G, Gallimidi Z, Reiss AH, et al. Classification of breast cancer precursors through exhaled breath. *Breast Cancer Res Treat.* 2011;126:791-796.
82. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-536.
83. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis. *JAMA Oncol.* 2019;5:e182815.
84. Phillips M, Cataneo RN, Ditkoff BA, et al. Volatile markers of breast cancer in the breath. *Breast Dis.* 2003;9:184-191.
85. Amann A, Costello BDL, Miekisch W, et al. The human volatome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J Breath Res J Breath Res.* 2014;8:17.
86. Jezierska-Drutel A, Rosenzweig SA, Neumann CA. Role of oxidative stress and the microenvironment in breast cancer development and progression. *Adv Cancer Res.* 2013;119:107-125.
87. Jeejeebhoy KN. In vivo breath alkane as an index of lipid peroxidation. *Free Radic Biol Med.* 1991;10:191-193.
88. Cederbaum AI. Alcohol metabolism. *Clin Liver Dis.* 2012;16:667-685.
89. Doran SLF, Romano A, Hanna GB. Optimisation of sampling parameters for standardised exhaled breath sampling. *J Breath Res.* 2017;12:016007.
90. US EPA. *Introduction to Indoor Air Quality | Indoor Air Quality (IAQ).* US Environmental Protection Agency. 2014. Accessed September 17, 2021. <https://www.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality>

91. Martinez-Outschoorn UE, Lin Z, Whitaker-Menezes D, Howell A, Sotgia F, Lisanti MP. Ketone body utilization drives tumor growth and metastasis. *Cell Cycle*. 2012;11:3964-3971.
92. Mileo AM, Miccadei S. Polyphenols as modulator of oxidative stress in cancer disease: new therapeutic strategies. *Oxid Med Cell Longev*. 2016;2016:1-17.
93. Jia Z, Zhang H, Ong CN, et al. Detection of lung cancer: concomitant volatile organic compounds and metabolomic profiling of six cancer cell lines of different histological origins. *ACS Omega*. 2018;3:5131-5140.
94. Opitz P, Herbarth O. The volatilome – investigation of volatile organic metabolites (VOM) as potential tumor markers in patients with head and neck squamous cell carcinoma (HNSCC). *J Otolaryngol Head Neck Surg*. 2018;47:1-13.
95. Leisher A, Ślefarska D, Leja M, et al. The volatilomic footprints of human HGC-27 and CLS-145 gastric cancer cell lines. *Front Mol Biosci*. 2021;7:471.
96. Rodríguez-Aguilar M, Díaz de León-Martínez L, Gorocica-Rosete P, et al. Application of chemoresistive gas sensors and chemometric analysis to differentiate the fingerprints of global volatile organic compounds from diseases. Preliminary results of COPD, lung cancer and breast cancer. *Clin Chim Acta*. 2021;518:83-92.
97. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol*. 2011;7:517.
98. Subramanian A, Salhab M, Mokbel K. Oestrogen producing enzymes and mammary carcinogenesis: a review. *Breast Cancer Res Treat*. 2008;111:191-202.
99. Fritz V, Fajas L. Metabolism and proliferation share common regulatory pathways in cancer cells. *Oncogene*. 2010;29:4369-4377.
100. Corbet C. Stem Cell Metabolism in Cancer and Healthy Tissues: Pyruvate in the Limelight. *Front Pharmacol*. 2018;8:958. doi:10.3389/fphar.2017.00958
101. Ward NP, DeNicola GM. Sulfur metabolism and its contribution to malignancy. *Int Rev Cell Mol Biol*. 2019;347:39-103.
102. Chen Y, Li P. Fatty acid metabolism and cancer development. *Sci Bull*. 2016;61:1473-1479.
103. Currie E, Schulze A, Zechner R, Walther T, Farese R. Cellular fatty acid metabolism and cancer. *Cell Metab*. 2013;18:153-161.
104. Miller IJ, Peters SR, Overmyer KA, Paulson BR, Westphal MS, Coon JJ. Real-time health monitoring through urine metabolomics. *NPJ Digit Med*. 2019;2:1-9.
105. Kwak J, Preti G. Volatile disease biomarkers in breath: a critique. *Curr Pharm Biotechnol*. 2011;12:1067-1074.
106. Poulin P, Krishnan K. A mechanistic algorithm for predicting blood:air partition coefficients of organic chemicals with the consideration of reversible binding in hemoglobin. *Toxicol Appl Pharmacol*. 1996;136:131-137.
107. Hanai Y, Shimono K, Oka H, Baba Y, Yamazaki K, Beauchamp GK. Analysis of volatile organic compounds released from human lung cancer cells and from the urine of tumor-bearing mice. *Cancer Cell Int*. 2012;12:7-13.
108. Eggink M, Wijtmans M, Kretschmer A, et al. Targeted LC-MS derivatization for aldehydes and carboxylic acids with a new derivatization agent 4-APEBA. *Anal Bioanal Chem*. 2010;397:665-675.
109. Toyokuni S. Molecular mechanisms of oxidative stress-induced carcinogenesis: from epidemiology to oxygenomics. *IUBMB Life*. 2008;60:441-447.
110. Ratcliffe N, Wiczorek T, Drabińska N, Gould O, Osborne A, De Lacy Costello B. 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*. 2020;14:034001.
111. KDOG CLINICAL STUDY. KDOG Cancer Detect Group. Accessed October 25, 2021. <https://kdog.institut-curie.org/page/kdog-clinical-study>
112. KDOG COV. KDOG Cancer Detect Group. Accessed October 25, 2021. <https://kdog.institut-curie.org/page/kdog-cov>
113. Lever J, Krzywinski M, Altman N. Points of significance: model selection and overfitting. *Nat Methods*. 2016;13:703-704.
114. Rodríguez-Pérez R, Fernández L, Marco S. Overoptimism in cross-validation when using partial least squares-discriminant analysis for omics data: a systematic study. *Anal Bioanal Chem*. 2018;410:5981-5992.