Journal of Breath Research

PAPER

Breast cancer detection using volatile compound profiles in exhaled breath via selected ion-flow tube mass spectrometry

RECEIVED
21 July 2022

REVISED
7 November 2022

ACCEPTED FOR PUBLICATION
28 November 2022

PUBLISHED
8 December 2022

Yoshie Nakayama[1,*], Mariko Hanada[2], Hiroshi Koda[2], Masahiro Sugimoto[3], Masahiro Takada[1] and Masakazu Toi[1]

[1] Breast Medical Oncology, Kyoto University Hospital, Shogoin Kawaramachi 54, Sakyoku, Kyoto 606-8507, Japan
[2] Nissha Co., Ltd, Mibu Hanai-cho, Nakagyo-ku, Kyoto 604-8551, Japan
[3] Institute of Medical Science, Tokyo Medical University, 6-1-1 Shinjuku, Shinjuku-ku, Tokyo 160-8402, Japan
[*] Author to whom any correspondence should be addressed.

E-mail: nakayama01077@gmail.com

Keywords: breast cancer, biomarker, exhaled breath, selected ion flow tube–mass spectrometry, volatile compound

Supplementary material for this article is available online

Abstract

This study aimed to evaluate volatile compounds in exhaled breath as a non-invasive screening method to detect breast neoplasms. Exhaled breath samples were collected from patients with breast cancer (BC; $n = 45$) and non-breast cancer (NBC; $n = 51$) controls. Selected ion-flow tube mass spectrometry was used to quantify the volatile compounds. A multiple logistic regression (MLR) model was developed by combining multiple compounds to discriminate between BC and NBC samples. Amongst the 672 quantified peaks, 17 showed significant differences between BC and NBC samples ($P < 0.05$ corrected by false discovery rate). Pathway analysis revealed a significant difference in glycerophospholipid metabolism. The MLR model showed an area under the receiver operating characteristic curve (AUC) of 0.719 (95% confidence interval: 0.615–0.822, $P < 0.0002$). Cross-validation under various conditions resulted in a slight fluctuation in the AUC values, indicating the high generalizability of the MLR model. The model showed a higher BC probability for advanced-stage subjects and higher Ki67 ($\geqslant$30) for BC subjects. This study suggests the potential of volatile compounds in exhaled breath as a noninvasive screening method for BC.

1. Introduction

Breast cancer (BC) is among the most common malignancies in women globally with high mortality, causing millions of deaths annually [1]. Early detection of BC allows for a broader selection of possible treatments, which improves prognosis of the disease. Imaging screening techniques, such as mammography (MMG), ultrasound, and [18F]-fluorodeoxyglucose positron emission tomography, have demonstrated their effectiveness in detecting early-stage disease and decreasing mortality [2–4]. However, the screening systems still need to be improved and optimized, to avoid issues, like misdiagnosis using MMG, which are still common [5].

The existing tumor markers such as cancer antigen 15-3 (CA15-3) and carcinoembryonic antigen (CEA) are used for monitoring treatment and detecting relapse, and not used for screening [6]. Recently, various novel liquid biopsy procedures have been developed using *omics* technologies to comprehensively quantify compounds of interest [7, 8]. For instance, circulating tumor DNA and cells [9], circulating cell-free DNA [10], exosomes [11], and DNA methylation profiles [12] in biofluids are the putative biomarkers that have attracted interest for use in BC diagnosis.

The aberrant metabolism of BC cells, including alterations in major energy-related metabolic pathways, glutaminolysis, the pentose phosphate pathway, and fatty acid biosynthesis, induces genetic and epigenetic processes that mediate oncogenesis [13]. The metabolic profiles of tissues and biofluids have shown potential for diagnosis and stratification of BC, which can assist in planning treatment regimens [14]. Metabolic fingerprinting or metabolic signatures generated from blood-based tests have also

© 2022 IOP Publishing Ltd

been evaluated as an alternative diagnostic tool for BC [15].

Non-invasive diagnostic tools, like urinary biomarkers, are desirable which enable frequent monitoring without causing undue stress or injury to the patient [16]. Volatile organic compounds (VOCs) in breath samples have been used as non-invasive tools as they can be easily collected, stored, and transported [17]. Gas chromatography–mass spectrometry (GC-MS) is the most commonly used technology to quantify VOCs. The Bio-VOC® breath sampler, and GC-MS identified four possible biomarkers for BC, including straight aldehydes [18]. A large cohort study ($n = 203$) identified seven possible biomarkers [19]. The use of machine learning and an electronic nose also showed potential in using VOCs to discriminate BC patients from healthy controls [20, 21]. A systematic review reported 43 previous studies that used VOCs as possible biomarkers for various types of cancer. However, there is a need for sampling standardization, to compare the precision levels between the previously reported biomarkers [22].

Selected ion flow tube–mass spectrometry (SIFT-MS) has recently been used as an alternative to GC-MS for analyzing volatile metabolites in human biofluids, including exhaled breath [23–27]. SIFT-MS has a number of advantages such as detection of permanent gases in the air, at parts-per-trillion levels and also has advantages like instantaneous, quantitative analysis of air and headspace with very high sensitivity and selectivity, direct analysis of high humidity samples, simultaneous analysis of chemically diverse VOCs (e.g. aldehydes, amines, and organosulfur), and simplicity of operation, low maintenance, and long-term stability. Unlike GC-MS, SIFT-MS is easy to maintain and more stable as it does not require chromatography columns and has precise chemical ionization. These features make SIFT-MS a powerful, complementary technique for GC-MS, which, though is a promising compound identification tool, cannot analyze samples rapidly and also requires highly trained operators.

This study evaluated VOCs in exhaled breath to differentiate patients with and without BC using a SIFT-MS. The non-breast cancer (NBC) group included patients with various benign diseases. The discrimination ability of combinations of VOCs in identifying BC was evaluated.

2. Material and methods

2.1. Subjects

This cross-sectional study explored BC-specific salivary metabolite levels and the study was conducted in accordance with the principles of the Declaration of Helsinki. A total of 96 participants were recruited during the period (mention the duration or the dates here). Diagnosis of BC was done by histological examination. None of the patients had received any prior treatment, including hormone therapy, chemotherapy, molecularly targeted therapy, radiotherapy, surgery, or any other alternative therapy. The non-breast cancer (NBC) controls group included healthy subjects with and without benign breast diseases. Written informed consent was obtained from all participants who consented to be enrolled in this study. This study was conducted according to the Declaration of Helsinki principles. The study protocol was approved by the ethics committees of Kyoto University (R0775-3). Written informed consent was obtained from all participants who agreed to be enrolled in this study.

2.2. Exhaled breath collection

On arrival at the hospital, patients were asked to rest on a chair for at least 15 min, and a structured patient history was collected via a questionnaire before sample collection. First, the subjects were asked about their present smoking status, smoking volume, alcohol intake (days per week), comorbidities, height, weight, whether they had undergone fasting for a certain period before arriving at the hospital, whether they had used tooth brush, mouthwash, and/or chewing gum before arriving. Subsequently, breath samples were collected with a Smart Bag PA (GL Science, Tokyo, Japan), a vinyl alcohol-based polymer film bag with low gas adsorption properties and heat resistance, and which prevents the elution of impurities from the material. They were asked to take a deep breath through the nose, followed by a single continuous forced exhalation through the mouth (while keeping their nostrils closed) into a sealed 2 l Smart Bag PA through a straw attached to it for prevention of saliva entry. Harvesting of breath sample in this manner resulted in collection of a mixed alveolar gas sample (mixture of alveolar air and respiratory dead space air). The samples were collected in the same outpatient room and stored at room temperature (room air samples were not collected).

2.3. Measurement of VOCs using SIFT-MS

Upon arrival at the laboratory, samples were stored at room temperature and analyzed within 3 h of collection using SIFT-MS (Voice200 ultra, Syft® Technologies, Christchurch, New Zealand). The SIFT-MS instrument was calibrated using a standard gas mix containing 1,2,3,4-tetrafluoro benzene, benzene, ethylene, isobutane, octafluorotoluene, p-xylene, perfluorobenzene, toluene, and nitrogen (Scotty® specialty gases, Pennsylvania, USA). The corrected intensities for all mass-to-charge ratios were extracted from the data generated by the instruments.

The calibration and measurement parameters of the SIFT-MS were set as previously described [28, 29]. Before analysis, an automated check test of the

flow, temperature, quadrupole performance, and single-point accuracy and precision determinations was performed using a certified standard containing benzene, ethylbenzene, ethylene, hexafluorobenzene, isobutene, octafluorotoluene, perfluoro-2-methyl-2-pentene, perfluorohexane, perfluoroheptane, and octadecafluorooctane. The SIFT-MS sampled the headspace through a septum on the sampling bag, and a stainless-steel needle connected to the sampling line with a flow of 77.3 Pa l/s. The ionized compounds were monitored using a mass spectrometer in full scan mode in the 15–250 m/z range for 70 s (10 s of preparation time and 60 s of sampling time). All samples (including room air) were scanned using the full mass scan mode (15–250 m/z) with three reagent ions (H_3O^+, NO^+, and O_2^+) for 300 ms per m/z measurement. Therefore, the measurement of a sample took 564 s, and the triplicate measurements of a sample took 28.2 min. The three regent ions are sent alternately, not simultaneously, and product ions are measured individually. As a result, three quantitative values are calculated for one compound, the lowest concentration of values displayed. Since there are many chemical species; the product ions are derived from multiple species, which may cause the overlap of multiple peaks. These phenomena would result in a higher concentration than expected. To reduce this problem, we used lower peaks to calculate the concentration. It is a disadvantage of SIFT-MS.

It takes about one hour for the analysis of a breath sample. Breath bags were stored in an air-conditioned analysis room. Before the study started, we had confirmed that the room's gas bag had not changed within 24 h. Two breath samples were collected per patient, one was analyzed immediately, and the other was analyzed after 24 h. It had been confirmed earlier that there is not much difference between the results of both samples. Since the results depend on the type of the sampling bags, we used Smart Bag PA, which has the least gas adsorption and emission according to our previous experience. SIFT-MS had been maintained every six months by the manufacturer to optimize the performance. The gauge needle was placed on the breath arm and into the gas bag. The gauge needle had been heated and kept at a high-temperature state. In addition, the room air was checked for any contamination before sampling. The gauge needles were washed with alcohol once every month.

The part number of the type gauge needle with the included side hole was unknown. The samples were measured in triplicates. No sample treatment was required; thus, the samples were directly injected into the Voice200 ultra SIFT-MS instrument. This MS uses a soft ionization method with reagent ions (H_3O^+, NO^+, and O_2^+). The putative compound names were assigned by matching m/z values with the corresponding data in the 1500 compound database. The data processing was conducted using LabSyft (Syft Technologies, LTD). Similar to the GC-MS, SIFT-MS also requires an understanding of gas analytics. The fewer procedures of sample processing which is helpful for routine usage.

2.4. Statistic analysis

The VOC profiles of each sample were normalized using quantile normalization, and each VOC concentration was transformed into a Z-score. The normalized VOC profiles were visualized using a heatmap and a volcano plot. The Mann–Whitney test was used to compare the BC and NBC groups. Pathway analysis was conducted using the Kyoto Encyclopedia of Genes and Genomes database. Differences in overall profiles were evaluated using principal component analysis (PCA). A multiple logistic regression (MLR) model was developed to differentiate the BC from the NBC group. Before developing the model, stepwise feature selection was performed to identify the minimum number of independent features. The threshold for the removal of a molecule was $P = 0.05$. Subsequently, k-fold cross-validations (k-CVs) were conducted to evaluate the model's generalization ability. The datasets were randomly split into training and validation datasets in a $(k-1)$:1 ratio. The model was developed using training data and evaluated using validation data. This process was repeated k times, and the generalization ability was calculated based on predictions using the validation data. We conducted two-, three-, four, and five-fold CVs 200 times each using different random values. MetaboAnalyst (v5.0) [30], JMP Pro (ver. 14.1.0; SAS Institute Inc., Cary, NC, USA), GraphPad Prism (ver. 7.0.3; GraphPad Software, Inc., La Jolla, CA, USA), and Weka (ver. 3.6.13; University of Waikato, Hamilton, New Zealand) were used for analyses.

3. Results

Table 1 summarizes the characteristics related to the subjects enrolled in this study. Exhaled breath samples were collected from patients with BC ($n = 47$) and NBC ($n = 53$). BC group also included invasive ductal carcinoma of the non-specific type ($n = 2$), invasive lobular carcinoma ($n = 1$), apocrine metaplasia ($n = 1$), and invasive micropapillary carcinoma ($n = 1$).

A heatmap (figure 1) shows the overall VOC profiles, including the normalized data, exhaled breath sample (quantile normalization), and metabolite axes (Z-score). The original data showed that several exhaled breath samples had high intensities of almost all metabolites (dense overall samples; supplementary figure 1). Normalized data did not include any sample-dependent bias. Therefore, subsequent analyses were performed using normalized data.

Table 1. Patient characteristics.

Item		NBC ($n = 51$) n or ave. ± SD	BC ($n = 45$) n or ave. ± SD	P-value
Menopause	Pre/Post	24/26	14/31	0.093[a]
Age		54.1 ± 12.7	61.0 ± 14.6	0.016[b]
Height (cm)		156.3 ± 6.369	156.3 ± 6.912	0.84[b]
Weight (kg)		53.55 ± 9.418	53.57 ± 9.634	0.80[b]
Smoking	No/Yes	27/8	39/5	0.17[a]
Drinking	No/Yes	22/13	31/13	0.48[a]
Anamnesis	No/Yes	20/25	20/25	1.0[a]
Diagnosis	Fiberoadenoma	5		
	Cyst	3		
	Mastopthy	2		
	Sclerosing adenosis	1		
	Cholesterol granuloma	1		
	Diabetic mastopathy	1		
	Subareolar abscess	1		
	Calcification	2		
	Others	35		
Histopathology	Invasive ductal carcinoma (IDC)		40	
	Invasive lobular carcinoma (ILC)		1	
	Invasive micro papilloma		1	
	Metastatic breast Cancer (MBC)		1	
	Apocrine metaplasia		1	
	Accessory breast cancer		1	
Grade	1/2/3		10/20/15	
Stage	1/2/3/4		21/17/4/3	
T	1/2/3/4		25/12/2/6	
N	0/1/2/3/X		32/9/1/2/1	
M	0/1		42/3	
ER	±		33/12	
PR	±		29/16	
HER2	0/1/2/3		17/14/7/7	
Ki67	<30/⩾30		34/11	

a = χ2 test, b = Mann–Whitney test.

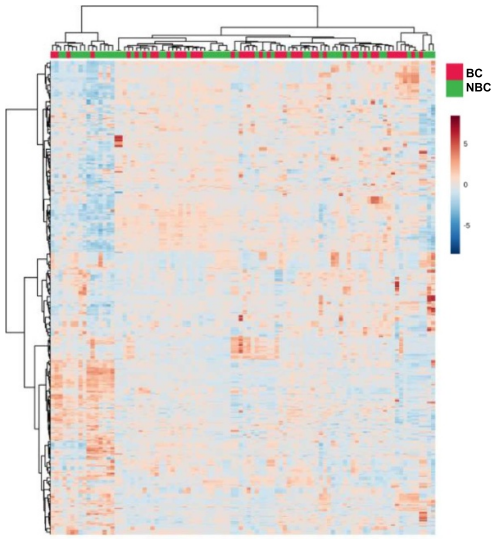

Figure 1. Heatmap showing VOC profiles of normalized data. Clustering is conducted by Euclidian distance with the Ward method. Red and green boxes below the top dendrogram indicate the BC and NBC, respectively. The red and blue show the relative higher and lower intensities for each VOC.

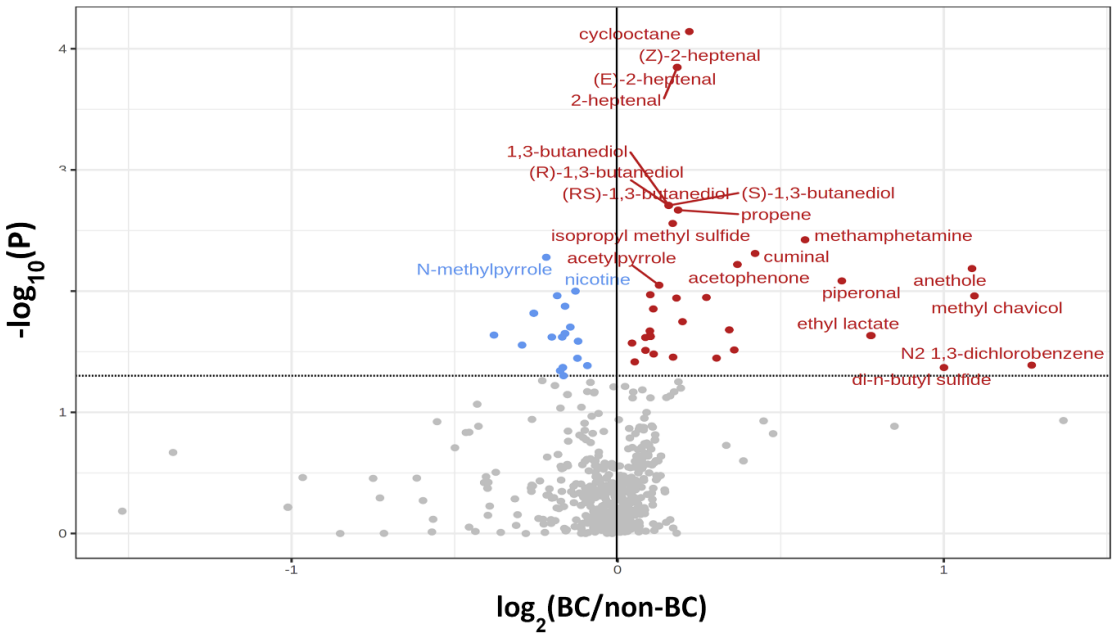

Figure 2. Volcano plots of VOC profiles. *X*-axis indicates log2-transferred fold change (BC/NBC). *Y*-axis indicates log10(*P*-value) between BC and NBC. *P*-values are calculated by Mann–Whitney tests. Elimination of non-significant ($P > 0.1$) peaks resulted in 96 peaks. Subsequently, *P*-values were corrected by false discovery rate, for considering multiple independent tests. *Y*-value >1.30 indicates corrected $P < 0.05$, and the dots above this line are colored in red (higher in BC).

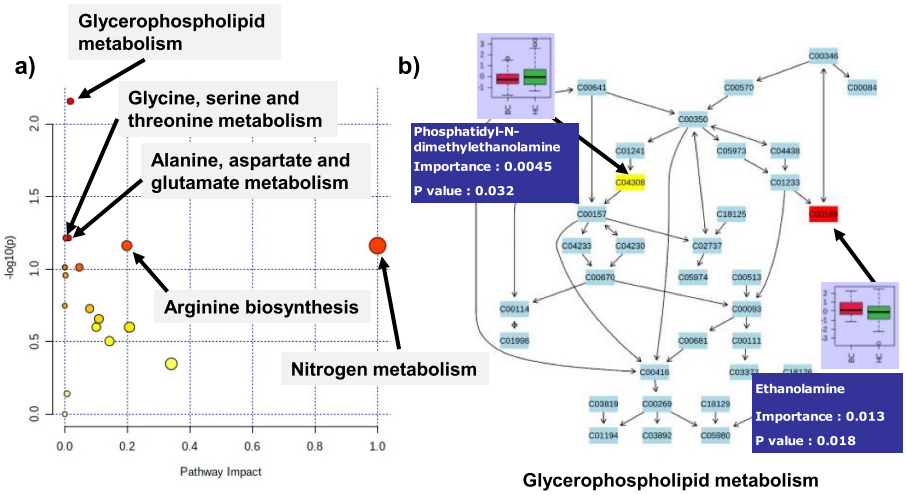

Figure 3. Results of the pathway analysis. (a) Each circle indicates a pathway. The color and size indicate the significance level in the enrichment analysis and pathway impact value from the topology analysis, respectively. (b) Glycerophospholipid metabolism.

The volcano plot (figure 2) shows the significantly different VOCs between BC and NBC. The *y*-axis shows $\log_{10}(P)$ of the *P*-value (Mann–Whitney test) between the BC and NBC groups. A *y*-value >1.30 indicates $P < 0.05$ corrected by FDR. Among the 646 detected components, 17 showed significant differences (plots colored red and blue) between the BC and NBC groups. All molecules showed higher values in the BC group.

Pathway analysis was conducted to assess the pathway-level differences between the BC and NBC groups (figure 3). Only glycerophospholipid metabolism showed significant differences at the pathway level. In this pathway, phosphatidyl-N-dimethylethanolamine levels were lower, whereas ethanolamine levels were higher in BC patients. An MLR model was developed by combining multiple VOCs to discriminate between BC and NBC samples. The model resulted in an area under the receiver operating characteristic curve (AUC) of 0.719 (95% confidence interval CI: 0.615–0.822, $P < 0.0002$) (figure 4(a)). The *k*-CV tests were conducted to

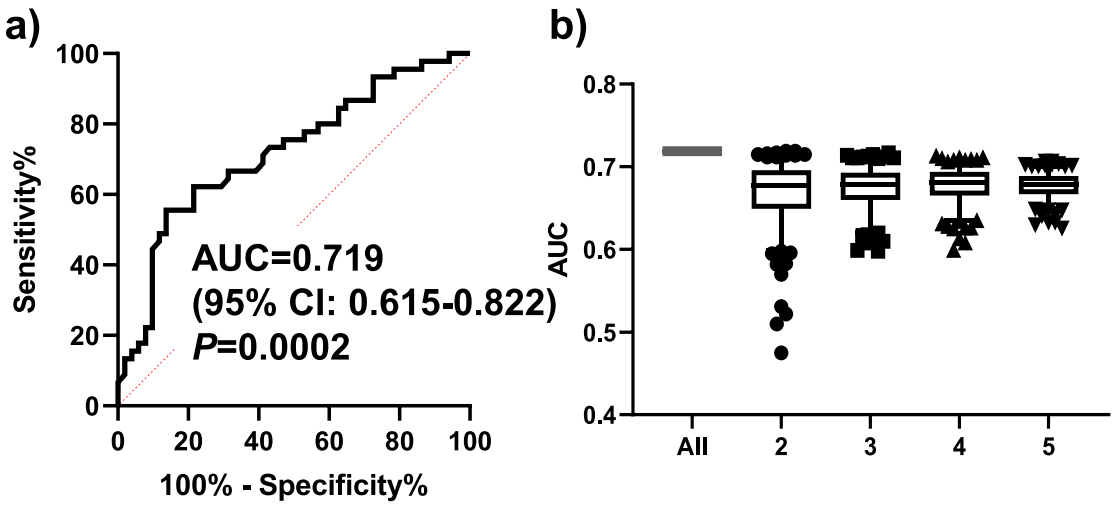

Figure 4. Multiple logistic regression (MLR) model. (a) ROC curve to discriminate BC from NBC. (b) AUC values using all data and *k*-fold cross-validation ($k = 2, 3, 4,$ and 5).

Table 2. MLR model.

Feature	Coefficient	95% CI		Odds ratio	95% CI		*P*-value
3,7-dimethyl-2, 6-octadien-1-ol	−0.629	−1.14	−0.120	0.533	0.32	0.888	0.016
Ethanolamine	0.414	−0.0710	0.900	1.51	0.931	2.46	0.094
Ethyl nonanoate	−0.607	−1.12	−0.0929	0.545	0.326	0.911	0.021
(Intercept)	−0.167	−0.603	0.269				0.45

evaluate the generalizability (figure 4(b)). Although several plots with low AUC values were observed for $k = 2$, the median AUC values for all tests ($k = 2, 3, 4,$ and 5) were almost constant. Therefore, based on the optimal cut-off point, sensitivity, specificity, positive predictive value, and negative predictive value were 86.3%, 55.6%, 68.8%, and 78.1%, respectively.

The model selected three VOCs: 3,7-dimethyl-2,6-octadien-1-ol, ethanolamine, and ethyl nonanoate (table 2). The ethanolamine parameter was negative. A lower intensity of ethanolamine increases the probability of BC. The parameters of the other two features with BC were positive, suggesting that higher intensities increase the likelihood of BC.The relationships between clinical parameters, the predicted probability from MLR, and the three metabolites used in the MLR model were analyzed. Among the NBC samples, the ethanolamine intensity and MLR models showed a significant difference between smokers and non-smokers ($P = 0.0056$ and $P = 0.0323$, Mann–Whitney test; supplementary material-table 1). However, BC samples did not observe this difference (supplementary material table 2). The analysis, including NBC and BC samples (Kruskal–Wallis and Dunn's multiple post-test), showed no significant difference in MLR predictions between smokers and non-smokers in the BC samples.

Among BC samples, a stage-specific difference in the prediction of MLR was observed ($P = 0.0497$, Kruskal–Wallis test) (supplementary material table 2). Ethanolamine showed a grade-specific difference ($P = 0.0302$, Kruskal–Wallis test), whereas there was no significant grade-specific difference in MLR predictions (supplementary material-table 2). In the advanced stage, high pathological grade and proliferation (Ki67 ≥st) (supplementary material table 2). Other parameters, such as node metastasis (N), showed no significant differences. Analyses including NBC and BC samples with stage, grade, Ki67, and N (Kruskal–Wallis and Dunn's multiple post-test) as variables are depicted in figure 5. Figure 5(e) shows that the significant difference (a) between NBC and BC without smoking and (b) between NBC and BC with smoking. The significant difference between NBC without smoking and BC with smoking also observed. However, there was no significant difference in NBC with/without smoking. The comparison of BC with/without smoking also showed no significant differences. Thus, the effect of smoking is less than the difference between NBC and BC. The relationships between these parameters and the normalized VOCs are shown in figures S3 and S4.

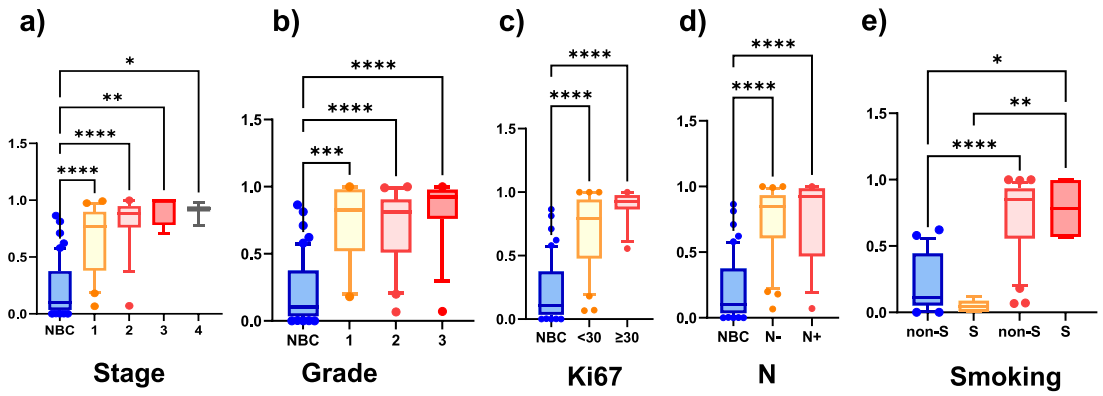

Figure 5. Relationship among clinical parameters and probabilities of prediction probabilities using MLR. (a) Stage, (b) Grade, (c) Ki67, (d) N, and (e) Smoking. Loading plots. Kruskal–Wallis and Dunn's multiple post-test is used to determined *P*-values. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. (e) S and non-S indicate smokers and non-smoker, respectively.

4. Discussion

This study aimed to differentiate BC from NBC participants using VOC profiles in exhaled breath samples. The combination of 3,7-dimethyl-2,6-octadien-1-ol, ethanolamine, and ethyl nonanoate discriminated BC from NBC. 3,7-dimethyl-2,6-octadien-1-ol is emitted from the skin of healthy humans [31]. Ethanolamine is an essential metabolite for cancer cell growth. Uptake and phosphorylation of ethanolamine and choline are crucial in human breast cells [32, 33]. BC adapts to metabolic stress by increasing ethanolamine phospholipid biosynthesis [32]. This metabolite, along with phosphoethanolamine in the cytosol of breast tissue, is correlated with the prognosis of BC patients [34]. BC shows high expression of glucose metabolism-related enzymes related to prognosis [35]. Triple-negative BC, a high-grade BC, mainly depends on glucose metabolism. BCs also harbor various stromal cells, such as cancer-associated fibroblasts and immune cells in the tumor microenvironment that have metabolic interactions with cancer cells. Various glycolytic metabolites, such as serine and glycine, are produced, and other metabolic pathways, such as the pentose phosphate pathway, are also involved. BCs are heterogeneous; consequently, metabolic characteristics are also diverse and dependent on molecular subtype, progression stage, and metastatic site. The concentrations of 3,7-dimethyl-2,6-octadien-1-ol, ethanolamine, and ethyl nonanoate were significantly increased in exhaled breath samples from BC. Changes in glucose metabolism in cells are likely the main reason for the differences in VOC levels. Ethyl nonanoate was detected in the exhaled breath of BC patients by SIFT-MS [25] and microorganism-contaminated foodstuff [36], but there are no reports to the best of our knowledge linking it to cancer.

The expression of Ki67 is strongly associated with tumor cell proliferation and growth, and it is widely used as a proliferation marker in pathological investigations. Nuclear protein Ki67 (pKi67) is an established prognostic indicator for assessing biopsies from cancer patients. Clinically, Ki67 has been shown to correlate with metastasis and the stage of tumors. In addition, it has been demonstrated that Ki67 expression is significantly higher in malignant tissues with poorly differentiated tumor cells than in normal tissue [37]. The Ki67 labeling index is an independent prognostic factor for survival rate, including all stage and grade categories. There was a correlation between the ratio of Ki67 positive malignant cells and patient survival. It has been shown that blocking Ki67 either via antibody microinjection or antisense oligonucleotides leads to an arrest of cell proliferation. Specifically, antisense oligonucleotides and antibodies against pKi67 have been shown to inhibit cell cycle progression. However, further studies are needed, including analyses with more appropriate *in vitro* and *in vivo* models. Ki67 expression induces advanced stage and high-grade BC. The three VOCs described in this study may be useful as an expiratory marker for high proliferation (Ki67 >30) in BC. This result may help establish new screening methods for the increased proliferation seen in BC cells.

Various instruments for analyzing VOCs in exhaled breath from BC patients and discrimination methods have been reported. SIFT-MS has been used to discriminate BC patients from healthy controls [38]. Biomarkers were not identified in this study, and the mass spectrometry pattern was used instead. GC/MS is commonly used for VOC analysis in the exhaled breath of BC patients [17, 39–41]. Partial least squares-discriminant analysis (PLS-DA) can be used as a discrimination model [19, 40], which utilizes PCA as an index to discriminate BC samples from healthy controls, using multiple peaks showing similar changes as PCA. The support vector machine, a machine learning method that is robust against outliers and multi-collinearity of the input features, was also employed to discriminate BC from healthy controls [38]. In contrast, we utilized MLR to utilize

only the minimum features to distinguish BC from NBC samples, which can contribute to developing targeted high-throughput assays.

This study did not evaluate the relationship between the observed VOCs and BC-specific metabolism. Cultured cells, including BC (MCF7), lung cancer (A549 and Calu-3), and non-cancerous lung cells (WI38VA13), showed different patterns of emitted VOCs [42]. The increased oxidative stress of MCF7 caused changes in its secreted VOCs, indicating a link between VOCs and intracellular metabolism [43]. Consistent changes in acetaldehyde and acetone were observed in the exhaled breath of tumor-bearing transgenic mice and women with BC [44]. These studies could help eliminate markers irrelevant to BC.

A limitation of the current research is that it only utilized the fingerprints of the VOCs and did not reveal the underlying biological mechanisms.

This study has several limitations. Firstly, a larger cohort is necessary to validate the developed model. Independent data should also be used to assess generalizability rather than a CV test. Secondly, the sampling protocol requires to be standardized. For example, various factors such as recent diet, sample collection, storage, and preprocessing need to be compared. The minimum requirement to produce reproducible intensities should also be defined before use in a clinical setting. Reporting standardization is also important to allow unbiased decision-making [45]. Thirdly, the specificity of the MLR model for other diseases has to be evaluated. The differences between VOCs in exhaled breath from BC, lung cancer, colorectal cancer, and prostate cancer patients have been previously studied [39, 46]. VOC changes caused by gastrointestinal diseases and colorectal cancer have also been reported [47, 48]. Comparisons with other diseases should be conducted.

The high-throughput feature is one of the advantages of SIFT-MS compared to GC/MS. This study used SIFT-MS to profile hundreds of VOCs and identify diagnostic biomarkers. However, combining only three VOCs was sufficient to discriminate between BC and NBC samples. Therefore, developing more high-throughput and cost-effective assays for the targeted analysis with only a few VOCs is possible and needed.

VOC profiles in exhaled breath were collected from BC ($n = 45$) and NBC ($n = 51$) patients. SIFT-MS successfully profiled 672 peaks, of which 65 peaks showed significant differences between BC and NBC groups. The MLR model combining three compounds of interest discriminated between these two groups and showed high generalizability in CV testing. This diagnostic tool, SIFT-MS for advanced-stage subjects and higher Ki67 ($\geq$30) for BC subjects, is suitable for mass screening because breath collection can be performed in a non-invasive manner.

Data availability statement

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request. Glucose Metabolism and Glucose Transporters in Breast Cancer.

Funding

This work was supported by JSPS KAKENHI Grant No. JS20H0573.

ORCID iDs

Yoshie Nakayama https://orcid.org/0000-0001-5643-4869
Masahiro Sugimoto https://orcid.org/0000-0003-3316-2543
Masahiro Takada https://orcid.org/0000-0002-5954-1296
Masakazu Toi https://orcid.org/0000-0003-3316-2543

References

[1] Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A and Bray F 2021 Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries *CA Cancer J. Clin.* **71** 209–49
[2] Coleman C 2017 Early detection and screening for breast cancer *Semin. Oncol. Nurs.* **33** 141–55
[3] Tohno E, Ueno E and Watanabe H 2009 Ultrasound screening of breast cancer *Breast Cancer* **16** 18–22
[4] Minamimoto R, Senda M, Jinnouchi S, Terauchi T, Yoshida T and Inoue T 2015 Detection of breast cancer in an FDG-PET cancer screening program: results of a nationwide Japanese survey *Clin. Breast Cancer* **15** e139–46
[5] Løberg M, Lousdal M L, Bretthauer M and Kalager M 2015 Benefits and harms of mammography screening *Breast Cancer Res.* **17** 63
[6] Hing J X, Mok C W, Tan P T, Sudhakar S S, Seah C M, Lee W P and Tan S M 2020 Clinical utility of tumour marker velocity of cancer antigen 15-3 (CA 15-3) and carcinoembryonic antigen (CEA) in breast cancer surveillance *Breast* **52** 95–101
[7] Tay T K Y and Tan P H 2021 Liquid biopsy in breast cancer: a focused review *Arch. Pathol. Lab. Med.* **145** 678–86
[8] Li J, Guan X, Fan Z, Ching L M, Li Y, Wang X, Cao W M and Liu D X 2020 Non-invasive biomarkers for early detection of breast cancer *Cancers* **12** 2767
[9] Sant M, Bernat-Peguera A, Felip E and Margelí M 2022 Role of ctDNA in breast cancer *Cancers* **14** 310
[10] Hackshaw A, Clarke C A and Hartman A R 2022 New genomic technologies for multi-cancer early detection: rethinking the scope of cancer screening *Cancer Cell* **40** 109–13
[11] Kumar D N, Chaudhuri A, Aqil F, Dehari D, Munagala R, Singh S, Gupta R C and Agrawal A K 2022 Exosomes as emerging drug delivery and diagnostic modality for breast cancer: recent advances in isolation and application *Cancers* **14** 1435
[12] Vietri M T, D'Elia G, Benincasa G, Ferraro G, Caliendo G, Nicoletti G F and Napoli C 2021 DNA methylation and breast cancer: a way forward (review) *Int. J. Oncol.* **59** 98
[13] Coronel-Hernández J, Pérez-Yépez E A, Delgado-Waldo I, Contreras-Romero C, Jacobo-Herrera N, Cantú-De León D

and Pérez-Plasencia C 2021 Aberrant metabolism as inductor of epigenetic changes in breast cancer: therapeutic opportunities *Front. Oncol.* **11** 676562
[14] Vignoli A, Risi E, McCartney A, Migliaccio I, Moretti E, Malorni L, Luchinat C, Biganzoli L and Tenori L 2021 Precision oncology via NMR-based metabolomics: a review on breast cancer *Int. J. Mol. Sci.* **22** 4687
[15] Zhang A H, Sun H, Qiu S and Wang X J 2013 Metabolomics in noninvasive breast cancer *Clin. Chim. Acta* **424** 3–7
[16] da Costa B R B and De Martinis B S 2020 Analysis of urinary VOCs using mass spectrometric methods to diagnose cancer: a review *Clin. Mass Spectrom.* **18** 27–37
[17] Patterson S G, Bayer C W, Hendry R J, Sellers N, Lee K S, Vidakovic B, Mizaikoff B and Gabram-Mendola S G 2011 Breath analysis by mass spectrometry: a new tool for breast cancer detection? *Am. Surg.* **77** 747–51
[18] Li J, Peng Y, Liu Y, Li W, Jin Y, Tang Z and Duan Y 2014 Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry *Clin. Chim. Acta* **436** 59–67
[19] Zhang Y, Guo L, Qiu Z, Lv Y, Chen G and Li E 2020 Early diagnosis of breast cancer from exhaled breath by gas chromatography-mass spectrometry (GC/MS) analysis: a prospective cohort study *J. Clin. Lab. Anal.* **34** e23526
[20] Yang H Y, Wang Y C, Peng H Y and Huang C H 2021 Breath biopsy of breast cancer using sensor array signals and machine learning analysis *Sci. Rep.* **11** 103
[21] Herman-Saffar O, Boger Z, Libson S, Lieberman D, Gonen R and Zeiri Y 2018 Early non-invasive detection of breast cancer using exhaled breath and urine analysis *Comput. Biol. Med.* **96** 227–32
[22] Oakley-Girvan I and Davis S W 2017 Breath based volatile organic compounds in the detection of breast, lung, and colorectal cancers: a systematic review *Cancer Biomark* **21** 29–39
[23] Hastie C, Thompson A, Perkins M, Langford V S, Eddleston M and Homer N Z 2021 Selected ion flow tube-mass spectrometry (SIFT-MS) as an alternative to gas chromatography/mass spectrometry (GC/MS) for the analysis of cyclohexanone and cyclohexanol in plasma *ACS Omega* **6** 32818–22
[24] La Nasa J, Modugno F, Colombini M P and Degano I 2019 Validation study of selected ion flow tube-mass spectrometry (SIFT-MS) in heritage science: characterization of natural and synthetic paint varnishes by portable mass spectrometry *J. Am. Soc. Mass Spectrom.* **30** 2250–8
[25] Wang M H, Chong K C, Storer M, Pickering J W, Endre Z H, Lau S Y, Kwok C, Lai M, Chung H Y and Ying Zee B C 2016 Use of a least absolute shrinkage and selection operator (LASSO) model to selected ion flow tube mass spectrometry (SIFT-MS) analysis of exhaled breath to predict the efficacy of dialysis: a pilot study *J. Breath Res.* **10** 046004
[26] Smith D and Spanel P 2005 Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis *Mass Spectrom. Rev.* **24** 661–700
[27] Spanël P and Smith D 2011 Progress in SIFT-MS: breath analysis and other applications *Mass Spectrom. Rev.* **30** 236–67
[28] Carrapiso A I, Noseda B, García C, Reina R, Del Pulgar J S and Devlieghere F 2015 SIFT-MS analysis of Iberian hams from pigs reared under different conditions *Meat Sci.* **104** 8–13
[29] Dharmawardana N, Goddard T, Woods C, Watson D I, Ooi E H and Yazbeck R 2020 Development of a non-invasive exhaled breath test for the diagnosis of head and neck cancer *Br. J. Cancer* **123** 1775–81
[30] Pang Z, Chong J, Zhou G, de Lima Morais D A, Chang L, Barrette M, Gauthier C, Jacques P, Li S and Xia J 2021 MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights *Nucleic Acids Res.* **49** W388–w96
[31] de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D and Ratcliffe N M 2014 A review of the volatiles from the healthy human body *J. Breath Res.* **8** 014001
[32] Zhu L and Bakovic M 2012 Breast cancer cells adapt to metabolic stress by increasing ethanolamine phospholipid synthesis and CTP:ethanolaminephosphate cytidylyltransferase-Pcyt2 activity *Biochem. Cell Biol.* **90** 188–99
[33] Kiss Z and Crilly K S 1995 Tamoxifen inhibits uptake and metabolism of ethanolamine and choline in multidrug-resistant, but not in drug-sensitive, MCF-7 human breast carcinoma cells *FEBS Lett.* **360** 165–8
[34] Kano-Sueoka T, Watanabe T, Miya T and Kasai H 1991 Analysis of cytosolic phosphoethanolamine and ethanolamine and their correlation with prognostic factors in breast cancer *Jpn. J. Cancer Res.* **82** 829–34
[35] Shin E and Koo J S 2021 Glucose metabolism and glucose transporters in breast cancer *Front. Cell Dev. Biol.* 2404
[36] Wang Y, Li Y, Yang J, Ruan J and Sun C 2016 Microbial volatile organic compounds and their application in microorganism identification in foodstuff *TrAC Trends Anal. Chem.* **78** 1–16
[37] Li L T, Jiang G, Chen Q and Zheng J N 2015 Ki67 is a promising molecular target in the diagnosis of cancer (review) *Mol. Med. Rep.* **11** 1566–72
[38] Martinez-Lozano Sinues P, Landoni E, Miceli R, Dibari V F, Dugo M, Agresti R, Tagliabue E, Cristoni S and Orlandi R 2015 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.* **9** 031001
[39] Peng G, Hakim M, Broza Y Y, Billan S, Abdah-Bortnyak R, Kuten A, Tisch U and Haick H 2010 Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors *Br. J. Cancer* **103** 542–51
[40] Wang C *et al* 2014 Volatile organic metabolites identify patients with breast cancer, cyclomastopathy, and mammary gland fibroma *Sci. Rep.* **4** 5383
[41] Phillips M, Cataneo R, Lebauer C, Mundada M and Saunders C 2017 Breath mass ion biomarkers of breast cancer *J. Breath Res.* **11** 016004
[42] Thriumani R, Zakaria A, Hashim Y Z H, Jeffree A I, Helmy K M, Kamarudin L M, Omar M I, Shakaff A Y M, Adom A H and Persaud K C 2018 A study on volatile organic compounds emitted by *in-vitro* lung cancer cultured cells using gas sensor array and SPME-GCMS *BMC Cancer* **18** 362
[43] Liu Y, Li W and Duan Y 2019 Effect of H(2)O(2) induced oxidative stress (OS) on volatile organic compounds (VOCs) and intracellular metabolism in MCF-7 breast cancer cells *J. Breath Res.* **13** 036005
[44] Ebeler S E, Clifford A J and Shibamoto T 1997 Quantitative analysis by gas chromatography of volatile carbonyl compounds in expired air from mice and human *J. Chromatogr.* B **702** 211–5
[45] Abderrahman B 2019 Exhaled breath biopsy: a new cancer detection paradigm *Future Oncol.* **15** 1679–82
[46] Rodríguez-Aguilar M *et al* 2021 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* **518** 83–92
[47] Rondanelli M, Perdoni F, Infantino V, Faliva M A, Peroni G, Iannello G, Nichetti M, Alalwan T A, Perna S and Cocuzza C 2019 Volatile organic compounds as biomarkers of gastrointestinal diseases and nutritional status *J. Anal. Methods Chem.* **2019** 7247802
[48] van Keulen K E, Jansen M E, Schrauwen R W M, Kolkman J J and Siersema P D 2020 Volatile organic compounds in breath can serve as a non-invasive diagnostic biomarker for the detection of advanced adenomas and colorectal cancer *Aliment. Pharmacol. Ther.* **51** 334–46