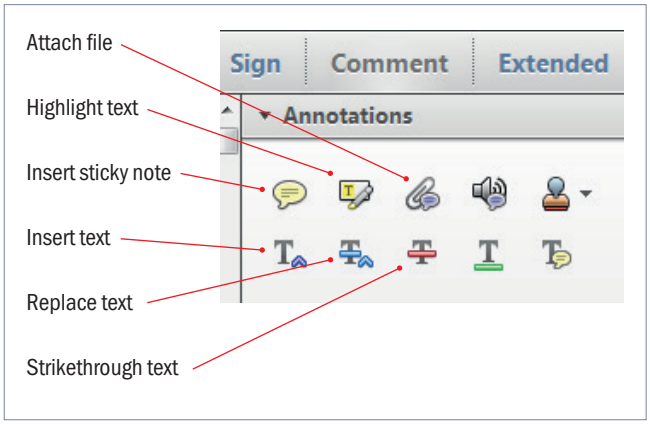
Making corrections to your proof

Please follow these instructions to mark changes or add notes to your proof. Ensure that you have downloaded the most recent version of Acrobat Reader from **https://get.adobe.com** so you have access to the widest range of annotation tools.

The tools you need to use are contained in **Annotations** in the **Comment** toolbar. You can also right-click on the text for several options. The most useful tools have been highlighted here. If you cannot make the desired change with the tools, please insert a sticky note describing the correction.

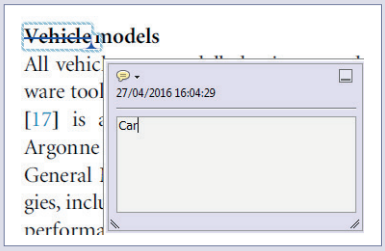
Please ensure all changes are visible via the 'Comments List' in the annotated PDF so that your corrections are not missed.



Do not attempt to directly edit the PDF file as changes will not be visible.

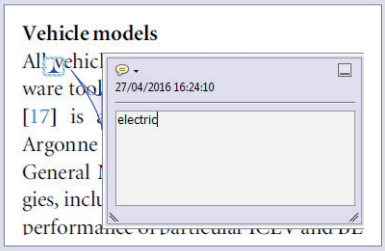
Replacing text

To replace text, highlight what you want to change then press the replace text icon, or right-click and press 'Add Note to Replace Text', then insert your text in the pop up box. Highlight the text and right click to style in bold, italic, superscript or subscript.



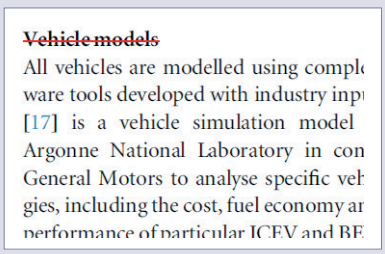
Inserting text

Place your cursor where you want to insert text, then press the insert text icon, or right-click and press 'Insert Text at Cursor', then insert your text in the pop up box. Highlight the text and right click to style in bold, italic, superscript or subscript.



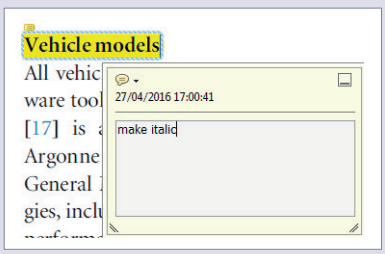
Deleting text

To delete text, highlight what you want to remove then press the strikethrough icon, or right-click and press 'Strikethrough Text'.



Highlighting text

To highlight text, with the cursor highlight the selected text then press the highlight text icon, or right-click and press 'Highlight text'. If you double click on this highlighted text you can add a comment.



QUERY FORM

JOURNAL: Journal of Breath Research

AUTHOR: M Phillips *et al*

TITLE: Breath mass ion biomarkers of breast cancer

ARTICLE ID: jbraa549b

The layout of this article has not yet been finalized. Therefore this proof may contain columns that are not fully balanced/matched or overlapping text in inline equations; these issues will be resolved once the final corrections have been incorporated.

Please check that the names of all authors as displayed in the proof are correct, and that all authors are linked correctly to the appropriate affiliations. Please also confirm that the correct corresponding author has been indicated.

If an explicit acknowledgment of funding is required, please ensure that it is indicated in your article. If you already have an Acknowledgments section, please check that the information there is complete and correct.

Please check that the funding information below is correct for inclusion in the article metadata.
National Breast Cancer Foundation (Australia).

Page 4

Q1
Please check the details for any journal references that do not have a link as they may contain some incorrect information. If any journal references do not have a link, please update with correct details and supply a Crossref DOI if available.

Page 4

Q2
Please provide the volume and page number or article number in reference [10].

IOP Publishing J. Breath Res. 00 (2017) 000000

Journal of Breath Research

RECEIVED
7 July 2016

REVISED
13 December 2016

ACCEPTED FOR PUBLICATION
16 December 2016

PUBLISHED
DD MM 2017

PAPER

Breath mass ion biomarkers of breast cancer

Michael Phillips[1,2,5], **Renee N Cataneo**[1], **Cassie Lebauer**[3], **Mayur Mundada**[1] **and Christobel Saunders**[4]

[1] Breath Research Laboratory, Menssana Research Inc., 211 Warren St, Newark, NJ 07103, USA
[2] Department of Medicine, New York Medical College, Valhalla, NY, USA
[3] Schmitt & Associates, 211 Warren Street, Newark, NJ 07103, USA
[4] School of Surgery, University of Western Australia, Perth, Western Australia
[5] Author to whom any correspondence should be addressed.

E-mail: mphillips@menssanaresearch.com, rcataneo@menssanaresearch.com, cbl8@njit.edu, mayur.mundada16@gmail.com **and** christobel.saunders@uwa.edu.au

Keywords: breath, volatile organic compound, breast cancer, mass ions, cancer detection

Abstract

Breath volatile organic compounds (VOCs) contain biomarkers of breast cancer that are detectable with gas chromatography mass spectrometry (GC MS). However, chemical identification of breath VOC biomarkers may be erroneous because spectral matching can misidentify their structure. Breath mass ions detected with GC MS have been proposed as intrinsically robust biomarkers because they can be identified without spectral matching. We investigated whether breath mass ion biomarkers could identify breast cancer. We re-analyzed data from a previous study of breath VOCs in 54 women with biopsy-proven breast cancer and in 204 healthy controls. Subjects were randomly assigned to a training set (2/3) and a test set (1/3). Chromatograms were processed with metabolomic analysis software (XCMS in R) in order to generate a table listing retention times with their associated ion masses and intensities, and binned into a series of 5 s retention time segments. In the training set, mass ions in each time segment were ranked according to their diagnostic accuracy i.e. the area under curve (AUC) of the receiver operating characteristic (ROC) curve. We employed multiple Monte Carlo simulations to select the biomarker mass ions in each time segment that identified breast cancer with greater than random accuracy and combined those with the highest diagnostic accuracy in a predictive algorithm using multivariate weighted digital analysis (WDA). We then employed this algorithm to predict the diagnosis in the test set. The training set WDA algorithm employing 21 mass ion biomarkers identified breast cancer with ROC curve AUC $=$ 0.79. In the test set, this algorithm predicted breast cancer with ROC curve AUC $=$ 0.77. Breath mass ions biomarkers accurately identified women with breast cancer and could potentially be used in early diagnosis and treatment monitoring.

Normal human breath contains more than 1000 different volatile organic compounds (VOCs), and some have been identified as candidate biomarkers of breast cancer [1–3]. However, breath biomarkers of disease are not yet widely employed in clinical practice, mainly because there is a lack of general agreement about their chemical structures. Probably the best-known example is lung cancer: several clinical studies have found statistically significant evidence of breath VOC biomarkers in patients with lung cancer, but the reported chemical structures of these biomarkers has varied widely between different studies [4–9]. This wide divergence in reported chemical structures may have arisen from several potential causes, including differences in experimental design and methods of sample collection and analysis.

The technical limitations of current analytical tools comprise another potential source of experimental error. The tool most widely employed for breath VOC biomarker discovery is gas chromatography mass spectrometry (GC MS): a sample of concentrated breath VOCs is injected onto a chromatographic column that separates the complex mixture into a series of individual VOCs according to their physicochemical properties such as boiling point and polarity. The separated VOCs then flow into a

detector where they are broken into fragments by a beam of high-energy electrons in a vacuum, and the resulting mass spectrum of fragments comprises a 'fingerprint' that can be used to identify the VOC from a computer-based spectral library.

GC MS is a widely accepted tool, but it can potentially yield erroneous identification of analytes if a mixture as complex and diverse as human breath VOCs overburdens the separation column. If the separation of VOCs is incomplete, then two or more VOCs may enter the MS detector simultaneously, and their combined mass spectra may lead to misidentification of their chemical structures in the spectral library. This may account, in part, for the diversity of breath VOC biomarkers of lung cancer that different investigators have reported.

In response to this problem, we investigated breath mass ions as candidate biomarkers of disease. An individual mass ion may provide an intrinsically more robust biomarker than the mass spectrum of its parent biomarker VOC for two main reasons: first, it is less likely to be affected by coelutions and second, it does not require pattern matching in a spectral library. We recently evaluated this approach in a blinded study of subjects with lung cancer, and found that breath mass ion biomarkers accurately predicted disease [10]. In order to further evaluate the diagnostic potential of breath mass ion biomarkers, we report here a re-analysis of data from a previous study of breath VOC biomarkers of breast cancer [2].

Methods and materials

Human subjects

The clinical study has been previously reported [2]. In summary, we analyzed breath VOCs in 54 women with biopsy-proven breast cancer and 204 cancer-free controls, using gas chromatography/mass spectrometry. Subjects comprised women with biopsy-proven breast cancer and a cancer-free group found either to have no significant abnormalities on routine mammographic screening or who were recalled after screening but subsequently showed not to have breast cancer on further testing. Breath samples were collected from the breast cancer group before they were treated for the disease. All gave their written informed consent to participate in the study, and the Ethics Committees of the University of Western Australia and the Royal Perth Hospital approved the research.

Breath VOC collection and assay

The method has been described [11]. In summary, we employed a portable breath collection apparatus to capture breath VOCs on to sorbent traps that were analyzed by automated thermal desorption, gas chromatography and mass spectrometry. In order to quantify peak areas and to control for drift in instrument performance, an internal standard was run with every chromatographic assay of breath and air (0.25 ml 2 ppm 1-bromo-4-fluorobenzene, Supelco, Bellefonte, PA) (BFB).

Re-analysis of data

All subjects and their respective chromatograms of breath VOCs were randomly assigned to either a training set or a test set in a 2/3: 1/3 split.

Selection of breath mass ion biomarkers of breast cancer

This was performed in the training set; the methods have been described [10]. In summary, we reprocessed chromatograms to generate a table of ion masses with their intensities and retention times normalized to BFB. Aligned data was binned into a series of 5 s retention time segments. We ranked mass ions as candidate biomarkers of breast cancer by comparing their intensity values in subjects with biopsy-proven breast cancer to cancer-free controls. In each 5 s time segment, the diagnostic accuracy of each mass ion was ranked according to the area under curve (AUC) of its receiver operating characteristic (ROC) curve. We then employed multiple Monte Carlo simulations to select the mass ion biomarkers in each time segment that identified breast cancer with greater than random accuracy.

XCMS was only used to convert the original data from the chromatograms into a format that can be stored in a database. The data stored had the fields retention time (in 1/100 of a second), ionic component (M/Z) and measured area.

In a first step all retention times were corrected by the retention time of the BFB marker, reducing the variability from around 12 –3 s.

Further, the area of every entry was divided by the area of the BFP peak yielding a value called abundance. This abundance was used for further analysis.

In order to reduce the data for analysis every 5 s were binned together, i.e. 5 s interval, ionic component and abundance. The value 5 s is a compromise between the typical width of a VOC (7 s) and the reduction factor. From then on every interval, M/Z and abundance was investigated as a potential biomarker.

Development and evaluation of predictive algorithm

We employed multivariate weighted digital analysis (WDA) [12] to construct a predictive algorithm using the mass ions that identified breast cancer with greater than random accuracy in the training set. This algorithm was then employed to predict the presence or absence of breast cancer in the test set.

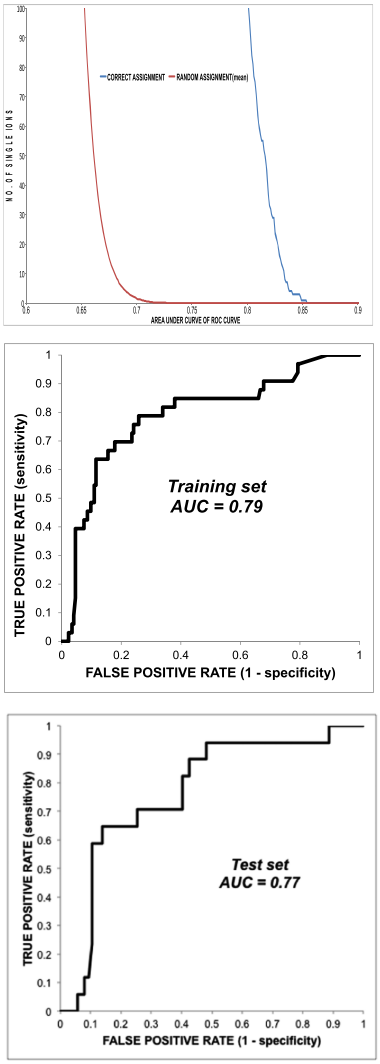


Figure 1. *(Top panel) Monte Carlo statistical analysis of mass ions*. This analysis was performed in a training set that comprised a random selection of two thirds of the study subjects. 21 mass ions exhibited greater than random diagnostic accuracy (C-statistic > 0.6) with correct assignment compared to multiple random assignments. The C-statistic of a mass ion is the area under curve (AUC) of its receiver operating characteristic (ROC). *(Middle panel) receiver operating characteristic (ROC) curve of breath mass ions in training set*. This figure displays the sensitivity and specificity for breast cancer in the training set, employing the multivariate algorithm that was derived from the combination of mass ions exhibiting greater than random diagnostic accuracy. The AUC of a ROC curve (or its C-statistic) indicates the overall accuracy of a test, and may vary from 0.5 (a straight line from bottom left to top right of the graph) to 1.0 (a right angle with its apex at the top left of the graph). A C-statistic of 0.5 indicates that the test performance was no better than random e.g. flipping a coin, while a C-statistic of 1.0 indicates a perfect test with 100% sensitivity and specificity. *(Bottom panel) receiver operating characteristic (ROC) curve in test set*. This figure displays the sensitivity and specificity for breast cancer of the algorithm in the test set that comprised the remaining one third of the study subjects. No data from these subjects was employed in the derivation and construction of the predictive algorithm.

Results

Monte Carlo statistical analysis of mass ions in the training set (figure 1, top panel)

The 'correct assignment' curve displays the number of mass ions as a function of their diagnostic accuracy, defined as the AUC of its associated ROC curve. The 'random assignment' curve similarly displays the number of mass ions as a function of their diagnostic accuracy employing the mean of 50 random assignments of diagnosis ('cancer' or 'cancer-free'). Where the random assignment curve fell to zero at approximately AUC $= 0.6$, 21 mass ions in the correct assignment curve exhibited diagnostic accuracy that was superior to random behavior.

ROC curve in training set (figure 1, middle panel)

The mass ions that exhibited diagnostic accuracy superior to random behavior (identified in top panel) were combined in a multivariate algorithm employing WDA. This algorithm identified breast cancer in the training set with 79% accuracy.

ROC curve in test set: (figure 1, bottom panel)

The multivariate predictive algorithm derived in the test set (top two panels) predicted breast cancer in the test set with 77% accuracy.

Example of mass ion biomarkers in cancer patients and controls

Table 1 displays the aggregated abundance of mass ion biomarkers in a segment of the chromatogram (1350–1375 s) in six randomly selected subjects, three with breast cancer and three cancer-free controls.

Discussion

The main finding of this study was that breath mass ions biomarkers accurately identified women with breast cancer. The training set algorithm employing 21 mass ion biomarkers identified breast cancer with ROC curve AUC $= 0.79$. In the test set, this algorithm predicted breast cancer with ROC curve AUC $= 0.77$. These findings are consistent with previous reports that biomarkers in breath can identify women with breast cancer.

The pathophysiologic mechanisms underlying breath biomarkers of breast cancer are not known. However, increased oxidative stress in breast tissue provides a plausible mechanism that links cancer development with altered biomarkers in breath. In breast cancer, almost 80% of stromal fibroblasts acquire an activated phenotype that produces hydrogen peroxide, which induces tumorigenic alterations in epithelial cells and fuels the growth and survival of cancer cells [13]. Hydrogen peroxide is a powerful oxidant that causes oxidative stress in tissues, which is known to peroxidate polyunsaturated fatty acids in

Table 1. Example of mass ion biomarkers in cancer patients and controls: this table displays the aggregated abundance of mass ion biomarkers (M/Z 56 to 61) in a segment of the breath VOC chromatogram (1350–1375 s) in six randomly selected subjects, three with breast cancer and three cancer-free controls. It is apparent that the abundance of these mass ions was consistently lower in the breast cancer subjects than in the controls.

MZ value	56	57	58	59	60	61
Control #1	2 653 243	3 206 919	3 687 286	487 286	441 990	525 678
Control #2	7 310 221	3 232 056	7 027 963	555 546	238 182	199 806
Control #3	7 607 665	2 467 155	1 887 122	369 912	307 363	135 114
Breast cancer #1	1 600 950	605 261	96 075	63 801	235 209	34 914
Breast cancer #2	1 367 103	365 763	72 178	44 597	605 709	44 103
Breast cancer #3	308 432	284 385	160 276	29 227	99 903	32 764

cell membranes, liberating volatile *n*-alkanes (e.g. ethane and pentane) that are expired in the breath [14]. Future studies of stromal fibroblasts *in vitro* could test the hypothesis that they are a source of the VOCs and mass ions observed in breath as biomarkers of breast cancer.

We previously identified breath mass ion biomarkers of lung cancer employing instruments and methodology similar to those employed in this study [10], and other researchers employing different instrumentation have reported mass ions in breath and in bacteria that also have potential clinical applications for detection of disease [15–17].

We conclude that breath mass ion biomarkers appeared to provide accurate new biomarkers of breast cancer. We employed mass ion biomarkers in order to control for potential coelutions of VOCs that may result from separation with one-dimensional gas chromatography. However, the newer technique of two-dimensional gas chromatography provides improved selectivity of separation of VOCs and could potentially reduce the need for this approach in future studies [1].

The findings of this pilot study are consistent with previous reports of breath biomarkers of breast cancer that employed different detection methods, but this will require validation in future prospective blinded clinical studies.

The ability to detect early breast cancer with a simple breath test provides opportunities to improve patient outcomes. For example, when two tests with different biological mechanisms are employed together to screen for the same disease, the combination may provide greater accuracy than either test when it is used alone [18]. Consequently, it may be possible to increase the sensitivity and the specificity of breast cancer detection by combining a breath test with a screening mammogram. We are currently testing this hypothesis in a prospective clinical study to determine the effects of an ancillary breath test on the sensitivity and specificity of mammography (https://clinicaltrials.gov/ct2/show/NCT02888366).

This includes enhancing the sensitivity and specificity of population mammographic screening, early detection of recurrent cancer to allow change in management, and monitoring response to cancer treatment, enabling early discontinuation of futile, often toxic, therapies, and a switch to potentially more efficacious treatments, thus improving both quality and quantity of life for people with breast cancer.

Acknowledgments

The clinical research was supported by an award from the National Breast Cancer Foundation (Australia). Michael Phillips is President and CEO of Menssana Research, Inc. Schmitt & Associates, Newark, NJ, maintained a database of chromatograms and analyzed the data.

References

Q1

[1] Phillips M *et al* 2013 Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry *PLoS One* **8** e75274
[2] Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P and Wai J. 2010 Volatile biomarkers in the breath of women with breast cancer *J. Breath Res.* **4** 026003
[3] Phillips M *et al* 2014 Rapid point-of-care breath test for biomarkers of breast cancer and abnormal mammograms *PLoS One* **9** e90226
[4] Phillips M *et al* 1999 Volatile organic compounds in breath as markers of lung cancer: a cross- sectional study *Lancet* **353** 1930–3
[5] Amann A, Corradi M, Mazzone P and Mutti A. 2011 Lung cancer biomarkers in exhaled breath *Expert Rev. Mol. Diagn.* **11** 207–17
[6] Deng C, Zhang X and Li N. 2004 Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography-mass spectrometry *J. Chromatogr.* B **808** 269–77
[7] Poli D *et al* 2005 Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study *Respir. Res.* **6** 71
[8] Schumer EM *et al* 2015 High sensitivity for lung cancer detection using analysis of exhaled carbonyl compounds *J. Thorac. Cardiovasc. Surg.* **150** 1517–22
[9] Ma W, Gao P, Fan J, Hashi Y and Chen Z. 2015 Determination of breath gas composition of lung cancer patients using gas chromatography/mass spectrometry with monolithic material sorptive extraction *Biomed. Chromatogr.* **29** 961–5
[10] Phillips M *et al* 2015 Blinded validation of breath biomarkers of lung cancer, a potential ancillary to chest CT screening *PLoS One* (doi:10.1371/journal.pone.0142484)
[11] Phillips M. 1997 Method for the collection and assay of volatile organic compounds in breath *Anal. Biochem.* **247** 272–8

Q2

[12] Phillips M *et al* 2008 Detection of lung cancer using weighted digital analysis of breath biomarkers *Clin. Chim. Acta* **393** 76–84
[13] Jezierska-Drutel A, Rosenzweig SA and Neumann CA 2013 Role of oxidative stress and the microenvironment in breast cancer development and progression *Adv. Cancer Res.* **119** 107–25
[14] Aghdassi E and Allard JP 2000 Breath alkanes as a marker of oxidative stress in different clinical conditions *Free Radic. Biol. Med.* **28** 880–6
[15] Spanel P and Smith D 2007 Selected ion flow tube mass spectrometry for on-line trace gas analysis in biology and medicine *Eur. J. Mass Spectrom.* **13** 77–82
[16] Gu H, Xu N and Chen H. 2012 Direct analysis of biological samples using extractive electrospray ionization mass spectrometry (EESI-MS) *Anal. Bioanal. Chem.* **403** 2145–53
[17] Niyompanich S, Jaresitthikunchai J, Srisanga K, Roytrakul S and Tungpradabkul S. 2014 Source-identifying biomarker ions between environmental and clinical Burkholderia pseudomallei using whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) *PLoS One* **9** e99160
[18] Weinstein S, Obuchowski NA and Lieber ML 2005 Clinical evaluation of diagnostic tests *Am. J. Roentgenol.* **184** 14–9