OPEN

Correspondence and
requests for materials
should be addressed to
E.L. (enyouli@aliyun.
com)

Volatile Organic Metabolites Identify Patients with Breast Cancer, Cyclomastopathy, and Mammary Gland Fibroma

Changsong Wang[1], Bo Sun[1], Lei Guo[1], Xiaoyang Wang[1], Chaofu Ke[2], Shanshan Liu[1], Wei Zhao[1], Suqi Luo[1], Zhigang Guo[1], Yang Zhang[1], Guowang Xu[3] & Enyou Li[1]

[1]Department of Anesthesiology, the First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Biostatistics, School of Public Health, Harbin Medical University, Harbin, China, [3]CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China.

The association between cancer and volatile organic metabolites in exhaled breaths has attracted increasing attention from researchers. The present study reports on a systematic study of gas profiles of metabolites in human exhaled breath by pattern recognition methods. Exhaled breath was collected from 85 patients with histologically confirmed breast disease (including 39 individuals with infiltrating ductal carcinoma, 25 individuals with cyclomastopathy and from 21 individuals with mammary gland fibroma) and 45 healthy volunteers. Principal component analysis and partial least squares discriminant analysis were used to process the final data. The volatile organic metabolites exhibited significant differences between breast cancer and normal controls, breast cancer and cyclomastopathy, and breast cancer and mammary gland fibroma; 21, 6, and 8 characteristic metabolites played decisive roles in sample classification, respectively ($P < 0.05$). Three volatile organic metabolites in the exhaled air, 2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, and cyclohexanone, distinguished breast cancer patients from healthy individuals, mammary gland fibroma patients, and patients with cyclomastopathy ($P < 0.05$). The identified three volatile organic metabolites associated with breast cancer may serve as novel diagnostic biomarkers.

Breast cancer is the most prevalent malignancy among women around the world and a major cause of female deaths. Each year, over 1.3 million women are diagnosed with breast cancer, and approximately 0.5 million women die of this disease[1]. In the United States, nearly 3 million women have a history of invasive breast cancer, and 226,870 new cases of breast cancer were reported in 2012[2]. Breast cancer is a progressive disease in which larger tumor size and the presence of lymph node metastasis are associated with worse prognoses[3]. An early breast cancer diagnosis can effectively reduce patient mortality rates[3–5]. Currently, the most common breast cancer screening method is mammography[4,5]. Many clinical trials have confirmed that mammography screening can significantly decrease mortality rates among breast cancer patients[3,6,7]. However, mammography remains an imperfect test; in particular, one limitation of mammography screening is that it does not detect all breast tumors. In a randomized, controlled mammographic screening trial, mammographic sensitivity to detect breast cancer ranged from 71% to 96%[5]. Patients with dense breast tissue had even lower mammographic sensitivities, from 48% to 70%[5,8].

Ultrasonography is more sensitive than mammography at detecting lesions in women with dense breasts[9,10]. However, ultrasonography does not detect most microcalcifications, which are the typical findings in ductal carcinoma in situ. In fact, 75% of cancers missed by ultrasonography were ductal carcinoma in situ, and 25% were invasive carcinomas[11]. In addition, the results of ultrasonography can vary widely, depending on the expertise of the technician[12,13].

Due to recent advances in analytical chemistry, the association between cancer and volatile organic metabolites in exhaled breaths has attracted increasing attention from researchers[14–16]. Breath analysis is suitable for disease screening because it is non-invasive, rapid, and readily accepted by patients. Preliminary studies have confirmed that analyses of exhaled volatile organic metabolites can differentiate between breast cancer patients and healthy controls[17–19]. Here, we report a systematic study of gas metabolite profiles in human exhaled breath using pattern

recognition methods. Volatile organic compounds from exhaled air in healthy individuals, breast cancer patients, and cyclomastopathy and mammary fibroma patients were used as profile defining variables. Potential biomarkers of breast cancer, cyclomastopathy, and mammary fibroma were analyzed.

Results

Breast Cancer Patients versus Controls. A total of 434 metabolites were consistently detected in 50% of the samples from breast cancer patients and normal controls. While the two-dimensional PCA score plot displayed a good separation trend (Fig. 1A), the OPLSDA score plot demonstrated separation between breast cancer patients and normal controls using one predictive component and three orthogonal components (R2X = 0.51; R2Y = 0.876; Q2 = 0.762; Fig. 1B). Moreover, the R2 and Q2 values calculated from the permutated data were less than the original values in the validation plot, which confirmed the validity of the supervised model (Fig. 1C).

Breast Cancer versus Cyclomastopathy. A total of 406 metabolites were consistently detected in 50% of the breast cancer samples and the patients with mammary gland hyperplasia. While the two-dimensional PCA score plot demonstrated a good separation trend (Fig. 2A), the OPLSDA score plot demonstrated separation between breast cancer patients and patients with mammary gland hyperplasia when using one predictive component and three orthogonal components (R2X = 0.501; R2Y = 0.777; Q2 = 0.565; Fig. 2B). Moreover, all of the R2 and Q2 values calculated from the permutated data were less than the original value in the validation plot, which confirmed the validity of the supervised model (Fig. 2C).

Breast Cancer versus Mammary Gland Fibroma. A total of 408 metabolites were consistently detected in 50% of the breast cancer and mammary gland fibroma samples. While the two-dimensional PCA score plot demonstrated a trend for good separation (Fig. 3A), the OPLSDA score plot demonstrated separation between breast cancer and mammary gland fibroma using one predictive component and one orthogonal component (R2X = 0.225; R2Y = 0.686; Q2 = 0.524; Fig. 3B). Moreover, all of the R2 and Q2 values calculated

from the permutated data were less than the original value in the validation plot, which confirmed the validity of the supervised model (Fig. 3C).

Potential Biomarkers. Among the significant metabolites identified using the VIP values in the OPLSDA model and the FDR values, 29 differential metabolites were annotated using the NIST 11 database with a similarity threshold of 75% (Table 1).

2-acetyl aminopropionic acid, methylacrylic acid, butyl acetate, benzocyclobutene, 4-hydroxybutanoic acid, 1,3,5,7-tetroxane, ethylene carbonate, 2,2-dimethyl decane, 2,3,4-trimethylheptane, 5-methyl-3-hexanol, 5-butylnonane tetradecane, hexadecane, 2,3,6-trimethyloctane, benzenemethanol, alpha,alpha-dimethyl and 2,5-dimethylhexane-2,5-dihydroperoxide exhibited significantly lower levels in the group of individuals with breast cancer than in the group of healthy individuals ($P < 0.05$), and breast cancer patients had significantly higher levels of 2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, cyclohexanone, dimethylacetamide, and trans-2-butene oxide in their exhaled breath ($P < 0.05$).

The exhaled air from the breast cancer group compared with the cyclomastopathy group revealed significant differences in six potential biomarkers. The breast cancer group exhibited significantly lower levels of cyclooctanemethanol and trans-2-dodecen-1-ol ($P < 0.05$), but significantly higher levels of 2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, butyl glycol, and cyclohexanone were found in the breast cancer group ($P < 0.05$).

The comparison of exhaled air from the breast cancer group vs. the mammary gland fibroma group revealed significant differences in eight potential biomarkers. Cyclopentanone exhibited significantly lower levels in the breast cancer group ($P < 0.05$), which had significantly higher levels of 1,2-propanediol, cyclohexanone, butyl glycol,1,4-dimethoxy-2,3-butanediol, 2,5,6-trimethyloctane, 3,4,5,6-tetramethyloctane, and ethylaniline ($P < 0.05$).

2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, and cyclohexanone were increased significantly in breast cancer patients relative to healthy individuals, mammary gland fibroma patients, or patients with cyclomastopathy ($P < 0.05$). Additional information is provided in Table 1.
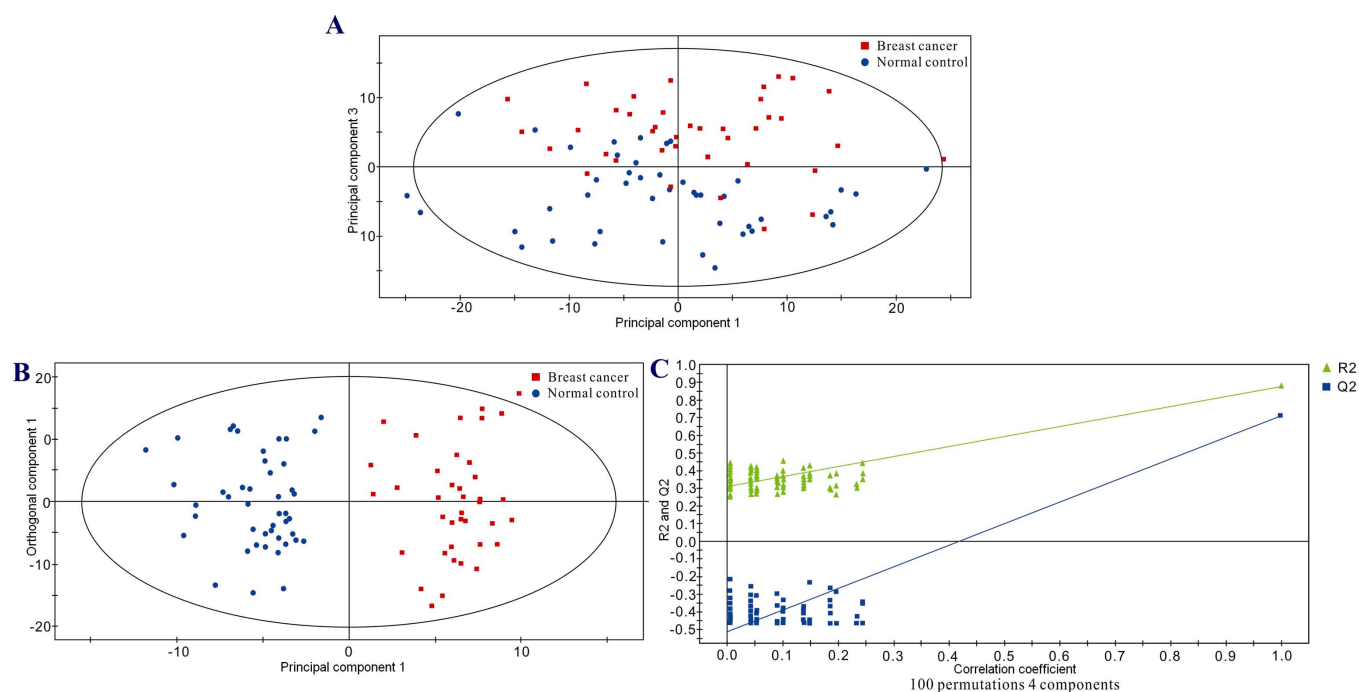
Figure 1 | (A): PCA score plot. (B): OPLSDA score plot (one predictive component and three orthogonal components, R2X = 0.51; R2Y = 0.876; Q2 = 0.762). (C): PLSDA validation plot Intercepts: R2 = (0.0, 0.306); Q2 = (0.0, −0.512).
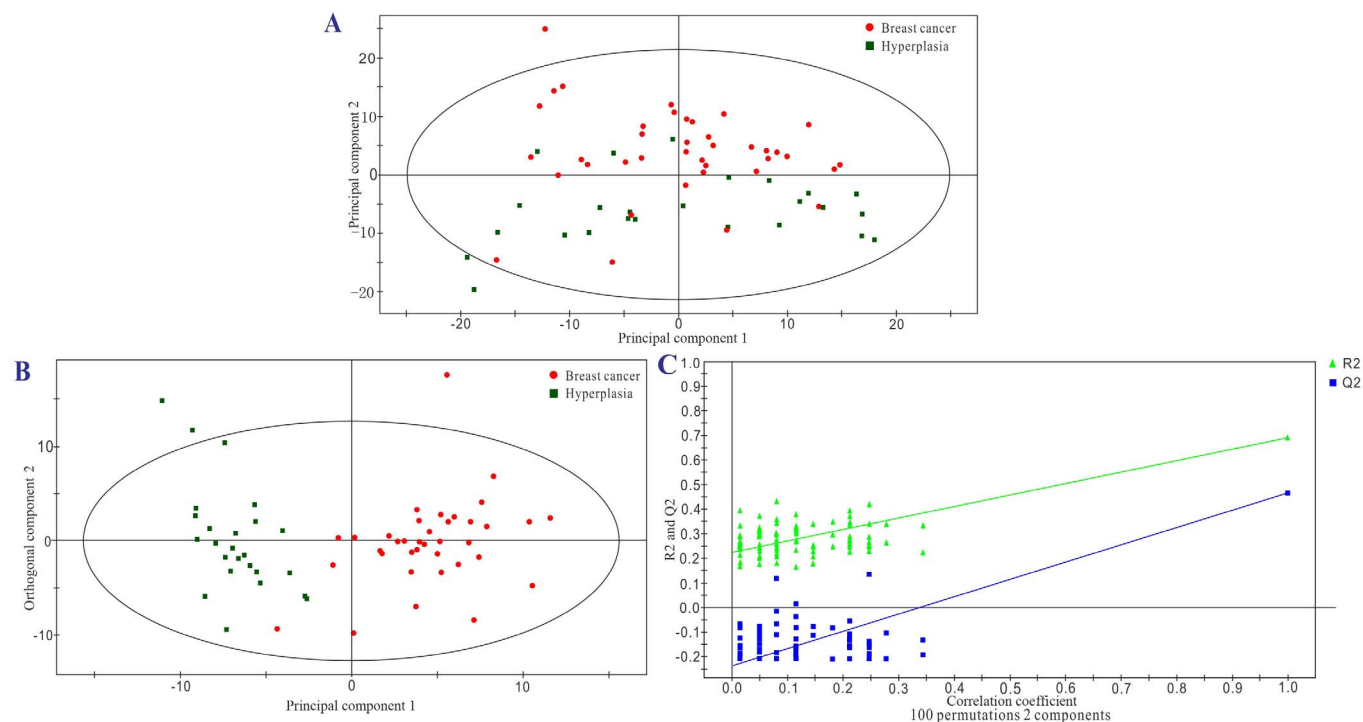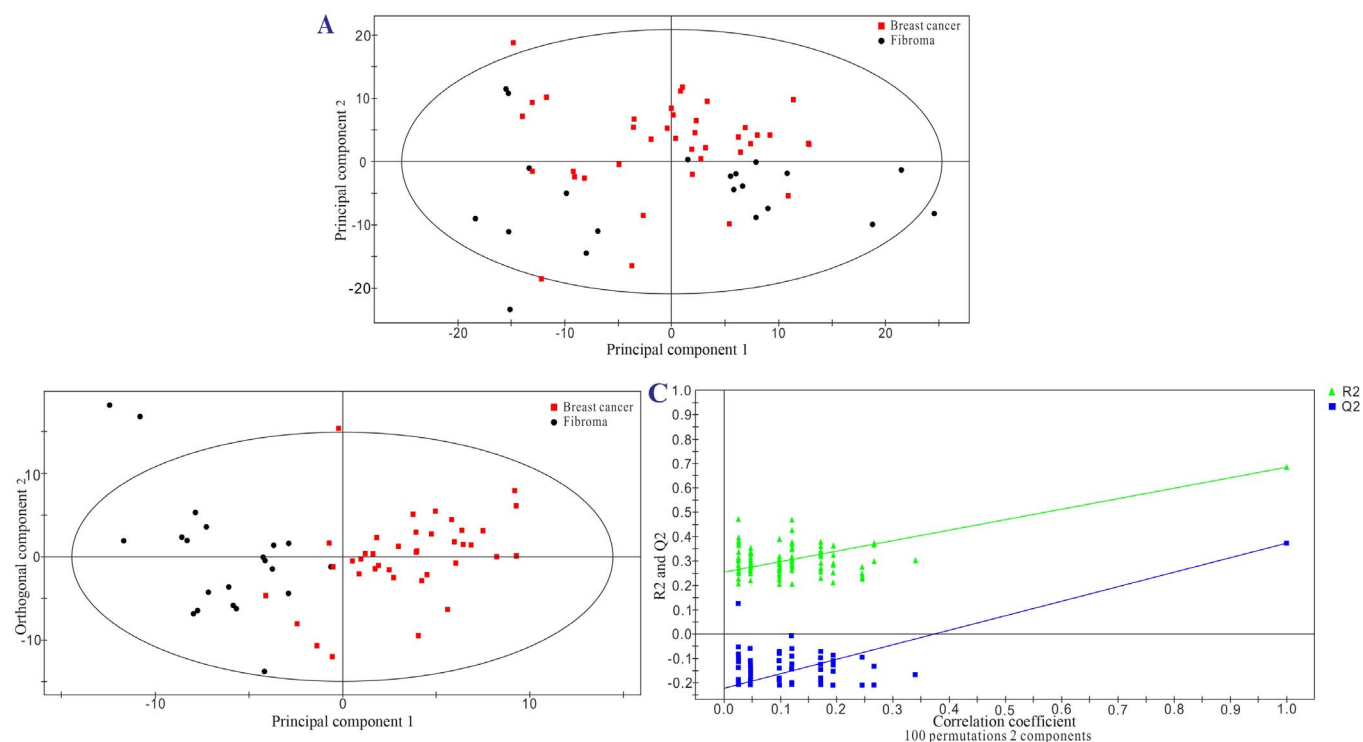
Figure 2 │ (A): PCA score plot. (B): OPLSDA score plot (one predictive component and three orthogonal components, $R2X = 0.501$; $R2Y = 0.777$; $Q2 = 0.565$). (C): PLSDA validation plot intercepts: $R2 = (0.0, 0.226)$; $Q2 = (0.0, -0.239)$.

Discussion

Diseases of the breast are some of the most common types of diseases among women. In particular, breast fibroadenoma, mammary gland hyperplasia, and breast cancer are three major diseases of the breast that are challenging for clinicians to diagnose. Existing screening and diagnostic techniques remain unsatisfactory. For instance, the detection of serum markers not only exhibits poor specificity but also requires wounding patients, which could facilitate the transmission of blood-borne infectious diseases. MRI examinations are expensive and require both state-of-the-art equipment and well-trained physicians with sophisticated technological knowledge and skills.

Figure 3 │ (A): PCA score plot. (B): OPLSDA score plot (one predictive component and one orthogonal component; $R2X = 0.225$; $R2Y = 0.686$; $Q2 = 0.524$). (C): PLSDA validation plot intercepts: $R2 = (0.0, 0.26$; $Q2 = (0.0, -0.222)$.

Table 1 | Potential biomarkers

Potential biomarker	RT	Breast cancer compared with healthy controls			Breast cancer compared with cyclomastopathy			Breast cancer compared with mammary gland fibroma		
		p-value	FC	VIP	p-value	FC	VIP	p-value	FC	VIP
1,2-Propanediol	3.04							3.38E-04	7.59	1.53
2-Acetyl Aminopropionic acid	3.95	2.33E-07	−2.26	1.73						
Cyclopentanone	4.28							2.05E-02	−1.02	1.54
Methylacrylic acid	4.29	3.66E-06	−6.69	1.51						
Butyl acetate	4.29	6.66E-07	−6.23	1.73						
Trans-2-Butene Oxide	5.53	6.28E-07	1.09	1.75						
Dimethylacetamide	5.54	1.04E-09	1.32	2.18						
Benzocyclobutene	6.06	1.06E-14	−8.51	2.02						
Cyclohexanone	***6.14***	***2.22E-10***	***1.81***	***2.16***	***3.30E-08***	***2.58***	***2.1***	***2.36E-07***	***2.76***	***2.37***
Butyl Glycol	6.35				1.03E-05	0.75	1.78	2.17E-07	1.16	2.45
4-Hydroxybutanoic acid	6.46	6.69E-09	−2.12	2.18						
1,3,5,7-Tetroxane	6.59	3.55E-12	−4.68	1.95						
Ethylene carbonate	7.56	9.51E-06	−2.1	1.78						
1,4-Dimethoxy-2,3-butanediol	***7.64***	***5.60E-08***	***1.79***	***2.03***	***1.09E-05***	***3.06***	***1.79***	***1.65E-06***	***6.25***	***2.01***
2,5,6-Trimethyloctane	***7.64***	***2.97E-09***	***1.67***	***2.07***	***9.99E-07***	***1.61***	***1.9***	***2.37E-06***	***2.41***	***1.84***
2,2-Dimethyl Decane	8.36	1.17E-09	−0.95	1.61						
3,4,5,6-Tetramethyloctane	8.37							2.32E-04	0.87	1.51
2,3,4-Trimethylheptane	10.57	4.55E-12	−9.48	1.86						
5-Methyl-3-hexanol	10.57	9.22E-13	−11.42	1.64						
5-Butylnonane	10.58	6.78E-11	−6.72	1.71						
2,3,6-Trimethyloctane	10.58	4.75E-09	−3.53	1.97						
Benzenemethanol, alpha,alpha-dimethyl	10.58	7.86E-09	−2.4	1.64						
Ethylaniline	11.5							7.91E-03	2.45	1.65
Cyclooctanemethanol	11.81				6.27E-04	−0.58	1.5			
trans-2-Dodecen-1-ol	12.5				6.79E-05	−0.62	1.67			
2,5-Dimethylhexane-2,5-dihydroperoxide	16.2	1.94E-14	−13.37	1.92						
Tetradecane	20.57	6.22E-07	−4.14	1.79						
Hexadecane	20.58	1.08E-06	−4.56	1.78						

Abbreviations: RT, retention time; VIP, variable importance in the projection; FC, fold change, defined as: FC=log2(X1/X2), while X1 denoted the arithmetic mean value of certain metabolite in the case group and X2 denoted the arithmetic mean value in the control group. FC with a positive value indicates that the concentration of certain metabolite is up-regulated in breast cancer compared with healthy controls, cyclomastopathy or mammary gland fibroma. FC with a negative value indicates that the concentration of certain metabolite is down-regulated in breast cancer compared with healthy controls, cyclomastopathy, or mammary gland fibroma.

In recent years, several studies have confirmed that certain specific volatile metabolites are present in abnormally high concentrations in the exhalations of breast cancer patients, and the origin of these compounds has been analyzed[15,20,21]. Peng reported that exhaled breaths from breast cancer patients and healthy controls exhibited significantly different levels of five volatile compounds: 3,3-dimethyl pentane; 2-amino-5-isopropyl-8-methyl-1-azulenecarbonitrile; 5-(2-methylpropyl)nonane; 2,3,4-trimethyl decane; and 6-ethyl-3-octyl ester 2-trifluoromethyl benzoic acid[15]. Phillips et al. suggested that alkanes and methylated alkane derivatives could be utilized as specific volatile markers for breast cancer; in particular, the researchers proposed that these compounds were produced because the mitochondrial release of reactive oxygen species (ROS) created oxidative stress that resulted in lipid peroxidation of polyunsaturated fatty acids in the cell membrane[20]. Hietanen et al. observed significantly higher than normal pentane concentrations in exhaled breath samples from breast cancer patients, and they conjectured that this difference originated from the peroxidation of fatty acids in the cell membrane[21].

Previous studies that have addressed the use of exhaled breath analysis in the context of breast disease have had a few shortcomings[15,20,21]. First, these studies have been limited to comparisons between cancer patients and healthy individuals; thus, the results of these investigations can be used to discriminate between these two groups, but are not particularly helpful for the screening and differential diagnosis of breast diseases. Second, various complex pathological types of breast cancer are observed in patients, and these studies do not distinguish among these different pathological types; however, it is known that different pathological types of cancer cells can generate unique volatile metabolites[22,23].

By comparing these three sets of experimental results, we revealed that three potential biomarkers, 2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, and cyclohexanone, were dramatically more concentrated in the exhaled air from breast cancer patients relative to the exhaled air from healthy individuals, mammary gland fibroma patients, or patients with cyclomastopathy. We therefore concluded that these three chemicals were relatively specific for breast cancer. Although butyl glycol levels were similar in the exhaled air from breast cancer patients and healthy individuals, significantly higher butyl glycol levels were detected in the exhaled air of breast cancer patients than in the exhaled air of breast fibroma patients or patients with cyclomastopathy. We speculate that this discrepancy reflects external contamination because we would otherwise have expected to observe a difference between the breast cancer patients and healthy individuals. Butyl glycol is commonly used in paint and ink solvents, metal cleaning agents, and dye dispersants. Each of the remaining examined metabolites exhibited specificity for a particular participant group, without crossover between these groups. Thus, the data from each group could potentially be used to construct specific metabolite models for breast cancer, mammary gland fibroma, and cyclomastopathy that could be used for the clinical screening of these breast diseases.

Most of the volatile markers identified in this study are alkanes, ketones, aldehydes, alcohols, or olefins. The mechanisms through which these metabolites are generated continue to be explored, and no unified consensus has been reached. However, the majority of relevant experimental results support the idea that these compounds result from oxidative stress[24]. Tumor tissues are characterized by vigorous growth and high energy demands. The malignant growth processes of cancer cells can lead to gene mutations and protein

Table 2 | Demographic characteristics of study subjects

	IDC	CMP	MGF	Normal
Subjects (n)	39	25	21	45
Age (mean ± SD)	53.4 ± 10.6	41.8 ± 8.5	36.4 ± 8.0	45.1 ± 10.7
Smokers (n)	2	1	0	0

Abbreviations: IDC, infiltrating ductal cancer; CMP, cyclomastopathy; MGF, mammary gland fibroma.

expression abnormalities. As a result, polyunsaturated fatty acids in cell membranes are subject to excessive oxidation, and an individual may produce excessive ROS[20]. In addition, reductions in other chemicals may result from the consumption of these substances by tumor cells[25]. The specific biological mechanisms that generate volatile metabolites remain under investigation. Phillips has hypothesized that the volatile biomarkers of breast cancer relate to changes in estrogen metabolism mechanisms[17]. This notion is supported by the fact that estrogen stimulation can stimulate the proliferation of both normal and neoplastic mammary epithelial cells[26]. Investigations have also revealed that abnormally high aromatase expression occurs in breast cancer tissues[27,28]. Aromatase (estrogen synthetase) is a component of the cytochrome P450 (CYP) enzyme complex that converts C19 androgen to C18 estrogen, thereby increasing estrogen generation[29]. Other P450 enzymes are also activated in breast cancer tissues, such as CYP1A1, CYP1B1 and CYP3A4[30]. P450 can induce a wide variety of biological response, including promoting the biotransformation of alkanes, alkenes, and aromatic compounds[31]. The metabolism of normal body cells can generate certain volatile metabolites, such as alkanes, that are produced as a result of oxidative stress[32]. Given the multitude of P450 functions, the elevated activity of this enzyme in breast cancer tissues may markedly alter the components of exhaled air. Phillips has suggested that breast diseases are associated with increases in oxidative stress and elevated cytochrome P450 activity[17].

Estrogen is a common precipitating factor of breast diseases, and changes in estrogen metabolism mechanisms play a crucial role in the carcinogenic processes of breast cancers. Spink reported that cytochrome P4501B1 catalyzes the conversion of estrogen to 4-hydroxyestradiol (4-OH-E2), which is a main driver of carcinogenic mechanisms in breast cancer tissues[33]. Liehr et al. demonstrated that the ratio of 4-OH-E2 to 2-hydroxyestradiol (2-OH-E2) is significantly higher in breast cancer tissue than in normal breast tissue[34]. Estrogen metabolites can promote carcinogenesis by damaging cellular macromolecules and promoting the proliferation of injured cells through receptor-mediated processes[34]. Changes in estrogen metabolism can generate certain distinct volatile substances, which may be associated with the chemicals that we observed at elevated levels in the exhaled air from breast cancer patients. However, the detailed mechanisms of these associations must be elucidated through additional clinical research. Furthermore, our observations that the concentrations of certain volatile compounds were at lower than normal levels in the exhaled air from breast cancer patients may relate to cellular consumption and increases in P450 activity. There are certain limitations associated with this study. First, the impact of subject age was not addressed. The individuals in the breast cancer, mammary gland fibroma, and cyclomastopathy groups were approximately 30, 40, and 50 years of age, respectively. Existing studies[35] have demonstrated that increased age is associated with elevated oxidative stress levels, and the resulting oxidation products may damage proteins, DNA, lipids, and other biological macromolecules, thereby generating certain volatile metabolites that may affect the results of this experiment. Second, although the breast cancer patients suffered from the same pathological type of breast cancer, there was no attempt to differentiate among breast cancers of different grades in this experiment. The presence of tumors of different clinical grades may affect the experimental results. In future studies, we will expand our sample size and differentiate among pathological types and stages of tumors in greater detail, allowing for more accurate volatile biomarkers to be identified.

Breast cancer, cyclomastopathy, and breast fibroma exhibit specific metabolic profiles with respect to volatile metabolites. Three volatile organic metabolites(2,5,6-trimethyloctane, 1,4-dimethoxy-2,3-butanediol, and cyclohexanone) associated with breast cancer may serve as novel diagnostic biomarkers.

Methods

Human Subjects. The present experiments were conducted in accordance with the Declaration of Helsinki. The protocol in this study was approved by the Ethics Committee at Harbin Medical University (No.201314), and written informed consent was obtained from patients prior to study enrollment. This study was conducted between May 2011 and April 2012 at the Department of Anesthesiology at the First Affiliated Hospital of Harbin Medical University.

Included in this study were women between 25 and 80 years of age identified as ASA I and II individuals and scheduled for breast surgery. The following exclusion criteria were used: 1) currently breastfeeding, pregnant, or the possibility of becoming pregnant; 2) a diagnosis of a known congenital disease; 3) radiotherapy or chemotherapy treatment prior to testing or another diagnosed malignant cancer at the time of testing; 4) co-existent chronic obstructive lung disease, asthma, or pulmonary tuberculosis and other pulmonary diseases; 5) the presence of a chronic inflammatory disease; and 6) the manifestation of any acute disease symptoms during the preceding two weeks. Moreover, to ensure uniformity of the experimental results by minimizing the impact of diet and environment on the composition of the subjects' exhaled breaths, study participants were asked to strictly fast for 8 hours (h) prior to breath sample collection. In addition to the group of breast cancer patients, this study also examined healthy female volunteers. The inclusion criterion for these individuals was the absence of a history of malignancies or infectious diseases. This study involved a total of 85 patients with histologically confirmed cases of breast disease (including 39 individuals with infiltrating ductal cancer, 21 individuals with mammary gland fibroma, and 25 individuals with cyclomastopathy) and 45 healthy volunteers (who were negative for breast cancer by mammography and ultrasound examination). The demographic characteristics are summarized in Table 2.

Breath Collection. Breath sampling and the parallel collection of ambient air were performed within 24 h after overnight fasting. Alveolar breath sampling was performed as previously described[36,37]. Briefly, 20 ml of exhaled gas were drawn into a gas-tight syringe (50 ml; Agilent Inc., USA). These samples were transferred immediately into evacuated 20 ml glass vials (Supelco Inc., USA). All vials were flushed thoroughly and cleaned with nitrogen gas (purity of 99.999%, Liming gas Inc., China) before being evacuated for breath sampling to remove any residual contaminants[38]. All samples were analyzed within 3 h post-sampling.

Solid-Phase Microextraction (SPME). A manual SPME holder with carboxen/polydimethylsiloxane (CAR/PDMS) fibers of 75 μm thickness was purchased from Supelco (Bellefonte, USA). The SPME fiber was inserted into the vial and exposed to the gaseous sample for 20 min at 40°C. Subsequently, the desorption of volatiles occurred in the hot GC injector at 200°C for 2 min.

Gas Chromatography-Mass Spectrometry (GC/MS) Analysis. Analysis was performed on a GC/MS (Shimadzu GC-MS QP 2010, Shimadzu, Japan) equipped with a DB-5MS (length 30 m × ID 0.250 × film thickness 0.25 μm; Agilent Technologies, USA) plot column. Injections were performed in the splitless mode, and the splitless time was 1 min. The temperature of injector was 200°C. The flow rate of the helium (99.999%) carrier gas was kept constant at 2 ml min^{-1}. The column temperature was held at 40°C for 2 min to concentrate the hydrocarbons at the head of the column and then increased by 70°C min^{-1} to 200°C for 1 min, then ramped 20°C min^{-1} to 230°C for 3 min. The MS analyses were performed in full-scan mode, using a scan range from 35–200 amu. The ion source was maintained at 200°C, and an ionization energy of 70 eV was used for each measurement.

Extraction and Pretreatment of the GC/MS Raw Data. Raw GC/MS data were converted into CDF format (NetCDF) files using Shimadzu GCMS Postrun Analysis software and subsequently processed using the XCMS toolbox (http://metlin.scripps.edu/download/). The XCMS parameters consisted of the default settings with the following exceptions: xcmsSet (fwhm = 8, snthresh = 6, max = 200); retcor (method

= "linear," family = "gaussian," plottype = "mdevden"); and a bandwidth of eight for first grouping command and four for the second grouping command[39,40]. The data set of the aligned mass ions was exported from XCMS and could be further processed using Microsoft Excel to normalize the data prior to multivariate analyses.

Statistical Analyses. The normalized data were exported to SIMCA-p 11.5 for principal component analysis (PCA), partial least-squares discriminant analysis (PLSDA), and orthogonal partial least-squares discriminant analysis (OPLSDA). To prevent overfitting, the default seven-round cross-validation in the SIMCA-p software was applied, and permutation tests using 100 iterations was performed to further validate the supervised model. Additionally, the nonparametric Kruskal-Wallis rank sum test was performed for each metabolite, and the corresponding false discovery rate (FDR) based on p-values was calculated to correct for multiple comparisons. The potential metabolic biomarkers were selected based on variable importance in the projection (VIP) values calculated from the OPLSDA model and FDR values of 1.5 and 0.05.

Abbreviations. VOCs, volatile organic compounds; CAR/PDMS, carboxen/polydimethylsiloxane; SPME, solid-phase microextraction; GC/MS, gas chromatography-mass spectrometry; PCA, principal component analysis; PLSDA, partial least-squares discriminant analysis; FDR, false discovery rate; VIP, variable importance in the projection; CYP, component of the cytochrome P450; 4-OH-E2,4-hydroxyestradiol; 2-OH-E2, 2-hydroxyestradiol.

1. Benson, J. R. & Jatoi, I. The global breast cancer burden. *Future Oncol.* 8, 697–702 (2012).
2. Siege, l. R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA. Cancer J. Clin.* 62, 10–29 (2012).
3. Tabár, L., Vitak, B. & Chen, H. H. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiol Clin. North Am.* 38, 625–651 (2000).
4. Nyström, L., Rutqvist, L. E. & Wall, S. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet.* 341, 973–978 (1993).
5. Humphrey, L. L., Helfand, M. & Chan, B. K. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann. Intern Med.* 137, 347–360 (2002).
6. Tabár, L., Fagerberg, C. J. & Gad, A. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet.* 1, 829–32 (1985).
7. Nystrom, L., Anderson, I. & Bjurstam, N. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lanct.* 359, 909–919 (2002).
8. Pisano, E. D. *et al.* Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N. Engl. J. Med.* 353, 1773–1783 (2005).
9. Zonderland, H. M., Coerkamp, E. G., Hermans, J., Vijver, M. J. & Voorthuisen, A. E. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. *Radiology.* 213,413–422 (1999).
10. Skaane, P. The additional value of US to mammography in the diagnosis of breast cancer. A prospective study. *Acta. Radiol.* 40, 486–490 (1999).
11. Nemec, C. F., Listinsky, J. & Rim, A. How should we screen for breast cancer? Mammography, ultrasonography, MRI. *Cleve. Clin. J. Med.* 12, 897–904 (2007).
12. Akazawa, K. *et al.* Preoperative evaluation of residual tumor extent by three-dimensional magnetic resonance imaging in breast cancer patients treated with neoadjuvant chemotherapy. *Breast J.* 12, 130–137 (2006).
13. Blair, S. *et al.* The efficacy of breast MRI in predicting breast conservation therapy. *J. Surg. Oncol.* 94, 220–225 (2006).
14. Song, G. *et al.* Quantitative breath analysis of volatile organic compounds of lung cancer patients. *Lung Cancer.* 67, 227–231 (2010).
15. Peng, G. *et al.* Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br. J. Cancer.* 103, 542–551 (2010).
16. Poli, D. *et al.* Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS. *J. Chromatogr B. Analyt. Technol. Biomed Life Sci.* 878, 2643–2651 (2010).
17. Phillips, M. *et al.* Volatile biomarkers in the breath of women with breast cancer. *J Breath Res.* 4, 026003 (2010).
18. Patterson, S. G. *et al.* Breath analysis by mass spectrometry: a new tool for breast cancer detection? *Am. Surg.* 77, 747–751 (2011).
19. Shuster, G. *et al.* Classification of breast cancer precursors through exhaled breath. *Breast.* 126, 791–796 (2011).
20. Phillips, M. Volatile markers of breast cancer in the breath. *Breast J.* 9, 184–191 (2003).
21. Kneepkens, C. M., Ferreira, C., Lepage, G. & Roy, C. C. The hydrocarbon breath test in the study of lipid peroxidation: principles and practice. *Clin. Invest. Med.* 15, 163–186 (1992).
22. Barash, O. *et al.* Classification of lung cancer histology by gold nanoparticle sensors. *Nanomedicine.* 8, 580–589 (2012).
23. Chen, X. *et al.* A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis. *Cancer.* 110, 835–844 (2007).
24. Knight, J. A. Free radicals: their history, and current status in aging and disease. *Ann. Clin. Lab. Sci.* 28, 331–346 (1998).
25. Phillips, M. *et al.* Volatile biomarkers in the breath of women with breast cancer. *J. Breath Res.* 4, 026003 (2010).
26. Subramanian, A., Salhab, M. & Mokbel, K. Oestrogen producing enzymes and mammary carcinogenesis: a review. *Breast Cancer Res. Treat.* 111, 191–202 (2008).
27. Richards, J. A., Petrel, T. A. & Brueggemeier, R. W. Signaling pathways regulating aromatase and cyclooxygenases in normal and malignant breast cells. *J. Steroid Biochem. Mol. Biol.* 80, 203–212 (2002).
28. Chen, S. Aromatase and breast cancer. *Front Biosci.* 3, 922–933 (1998).
29. Arun, B. & Goss, P. The role of COX-2 inhibition in breast cancer treatment and prevention. Semin. *Oncol.* 31, 22–29 (2004).
30. Jongen, V. H., Hollema, H., Zee, A. G. & Heineman, M. J. Aromatase in the context of breast and endometrial cancer. A review. *Minerva Endocrinol.* 31, 47–60 (2006).
31. Hannemann, F., Bichet, A., Ewen, K. M. & Bernhardt, R. Cytochrome P450 systems—biological variations of electron transport chains Biochim. *Biophys. Acta.* 1770, 330–344 (2007).
32. Kneepkens, C. M., Lepage, G. & Roy, C. C. The potential of the hydrocarbon breath test as a measure of lipid peroxidation Free Radic. *Biol. Med.* 17, 127–160 (1994).
33. Spink, D. C. *et al.* The effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on estrogen metabolism in MCF-7 breast cancer cells: evidence for induction of a novel 17 beta-estradiol 4-hydroxylase. *J. Steroid Biochem. Mol. Biol.* 51, 251–258 (1994).
34. Liehr, J. G. & Ricci, M. J. 4-Hydroxylation of estrogens as marker of human mammary tumors. *Proc. Natl. Acad Sci. U S A.* 93, 3294–3296 (1996).
35. Phillips, M. *et al.* Gunawardena R. Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress. *J. Lab Clin. Med.* 136, 243–249 (2000).
36. Birken, T., Schubert, J., Miekisch, W. & Noldge-Schomburg, G. A novel visually CO2 controlled alveolar breath sampling technique. *Technol. Health Care.* 14, 499–506 (2006).
37. Miekisch, W. *et al.* Impact of sampling procedures on the results of breath analysis. *J. Breath Res.* 2, 026007 (2008).
38. Bajtarevic, A. *et al.* Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC. Cancer.* 9, 348 (2009).
39. Lin, H. M., Edmunds, S. I., Helsby, N. A., Ferguson, L. R. & Rowan, D. D. Non-targeted urinary metabolite profiling of a mouse model of Crohn's disease. *J. Proteome Res.* 8, 2045–2057 (2009).
40. Gao, X. *et al.* Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization. *Anal. Biochem.* 393, 163–175 (2009).

Acknowledgments

Author contributions
Conception and design: E.L., C.W., G.X. Development of methodology: B.S. Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): X.W., L.G., S.L., W.Z., S.L., Z.G., Y.Z. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.K. Writing, review, and/or revision of the manuscript: E.L., C.W., P.L. Administrative, technical, or material support: G.X.

Additional information
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, C. *et al.* Volatile Organic Metabolites Identify Patients with Breast Cancer, Cyclomastopathy, and Mammary Gland Fibroma. *Sci. Rep.* 4, 5383; DOI:10.1038/srep05383 (2014).

6