A novel non-invasive method for detection of breast and ovarian cancer using volatile organic compounds from urine sample

Mital Nakrani
Gujarat Biotechnology Research Centre, Gandhinagar, Gujarat -382011
Pooja Doshi
Gujarat Biotechnology Research Centre, Gandhinagar, Gujarat -382011
Chaitanya Joshi
Gujarat Biotechnology Research Centre, Gandhinagar, Gujarat -382011
Madhvi Joshi (madhvimicrobio@gmail.com)
Gujarat Biotechnology Research Centre, Gandhinagar, Gujarat -382011

Research Article

Keywords: Breast and ovarian cancer, GC-MS, Non-invasive, Urine Biomarker, Volatile Organic Compound

Posted Date: April 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-397567/v1

A novel non-invasive method for detection of breast and ovarian cancer using volatile organic compounds from urine sample

Mital Nakrani[1], Pooja Doshi[1], Chaitanya Joshi[1], Madhvi Joshi[1]*

1. Gujarat Biotechnology Research Centre, Gandhinagar, Gujarat -382011, India

*corresponding author

Dr. Madhvi Joshi

Joint Director,

Gujarat Biotechnology Research Centre Gandhinagar,

Gujarat (India).

E-mail: madhvimicrobio@gmail.com; jd1-gbrc@guajarat.gov.in

Abstract

The present study describes a novel method for early diagnosis of breast and ovarian cancer from urine samples using a non-invasive method by determining the level of the selected volatile organic compounds (VOCs). In this study, solvent (hexane) was used for the extraction of VOCs and further detected using GC-MS. Total twenty-six compounds were analyzed, out of which, six (Naphthalene; 2-methylOctacosane; Benzaldehyde 2,5dimethyl; 11-Methyldodecanol; Heptane 2,4 dimethyl, and 2,4-Dimethylhept-1-ene) were found to be uniquely present in breast and ovarian cancer patients. These six VOCs could be used as potential biomarkers for early diagnosis of breast and ovarian cancer. Out of these six, two compounds (2-methylOctacosane and 11-Methyldodecanol) were detected for the first time in the aforementioned cancer patients. Quantitation of the remaining twenty VOCs common to both control individuals and subjects revealed significant log2 fold change (P-value<0.05) range from -0.7 to 1.4. The maximum decrease in the concentration level was observed in pentadecane and the maximum increase was observed in trans-1, 2-Diethyl cyclopentane in cancer subjects as compared to the control individuals. The present study may be used to identify these compounds simultaneously from the urine that are potential biomarkers for the diagnosis of breast and ovarian cancer.

Keywords: Breast and ovarian cancer, GC-MS, Non-invasive, Urine Biomarker, Volatile Organic Compound.

1. Introduction

Breast and ovarian cancer (BOC) is the most frequently diagnosed malignancy amongst cancers and the second leading cause of cancer-related deaths among the women worldwide [1]. In India, during last year (2020), total 2,24,062 patients were diagnosed with BOC which is 31% of all the cancers diagnosed in Indian women. Breast cancer is also the most frequently occurring cancer in Indian females with 61% mortality rates [2]. Majority of these mortalities are associated with late diagnosis. There is no precise and non-invasive technique available to screen for the BOC before a confirmatory biopsy is performed.

Early stage detection of BOC will aid to reduce the burden of the disease because treatment could be provided to the patients before metastasis stage is reached [34]. Currently, several invasive methods such as biopsy, ultrasounds, mammography, self-examination and magnetic resonance are widely used. However, these methods are time-consuming, need trained personnel, have low sensitivity and specificity, and discomfort for the patients[56789].

Detection of hyper-methylated DNA in nipple aspirate fluid is another non-invasive breast cancer screening method[5]. The sample collection method for this screening is also a challenge and needs expertise. Thus, there is an urgent need of non-invasive, precise and rapid screening tests for early detection of these cancers.

The study of metabolomics in biotechnology is evolving as a potential tool which helps exploring biological processes occurring in the humans, and also finding applications in disease diagnosis including cancer [69]. It involves the identification of volatile organic compounds (VOCs) that may predict the incidence of metastasis in the patient before its manifestation. Biological samples such as breathe, blood and urine contain thousands of VOCs that could be used as biomarkers for a large number of diseases[8710].Several, VOCs produced by the different cellular metabolic pathways are present in the urine and may serve as a source of biomarkers for breast cancer [111213]. Urine sample analysis for metabolic biomarkers is also relatively economical as compared to other conventional techniques[141516]. Application of such a non-invasive and accurate diagnostic technique for breast cancer will help in early diagnosis. Therefore, this study was conducted with an aim to develop a noninvasive method for the diagnosis of BOC using urine as an input.

2. Results

Representative overlapping chromatograms of the control and cancer subjects highlighting variation in the VOCs profile is shown in fig. 1.

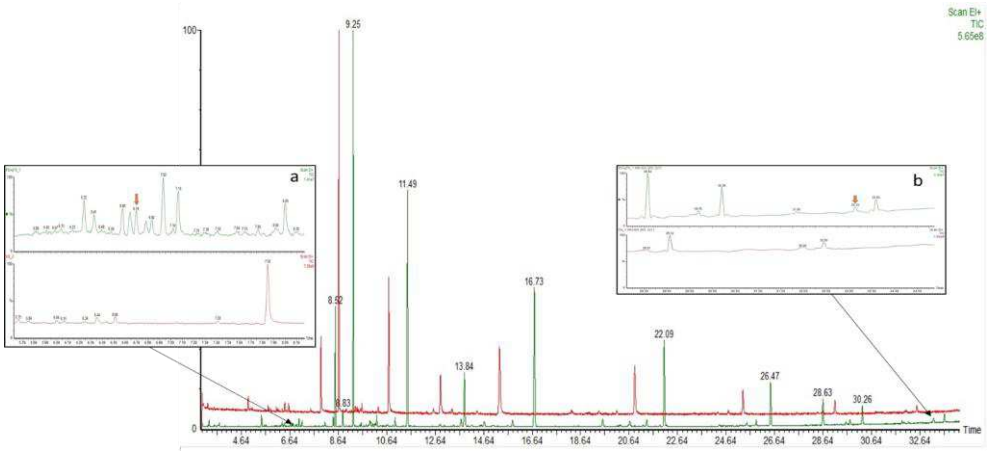

Figure 1: Chromatograms of control and breast and/or ovarian cancer samples

A total of 26 VOCs (Table 1) could be extracted by the developed method and detected through the GC-MS. Metabolites 1-20 in Table 1 are VOCs those were found to be common in both control and cancer subjects.

Table 1: VOC markers (26) in urine sample of breast and ovarian cancer patients and clinically healthy Subjects

SR. NO.	Compound Name	Abbrev.	log2 fold change	T-Test (P Value)	Trait	CAS #
1	Pentadecane	PNTD	-0.70	$1.85E^{-06}$	Down	544-76-3
2	Undecane, 4,7-dimethyl-	UNDM	-0.51	$1.76E^{-23}$	Down	17301-32-5
3	3-Heptyne-2,6-dione,5-methyl-5-(1-methylethyl)-	HDMM	-0.50	$2.86E^{-12}$	Down	63922-44-1
4	Octacosane	OTCS	-0.35	$2.41E^{-05}$	Down	630-02-4
5	Decane	DEC	-0.21	$4.03E^{-11}$	Down	124-18-5
6	Heptadecane, 3-methyl-	HPTDM	-0.19	0.012176	Down	6418-44-6
7	Eicosane, 2-methyl-	ECSH	-0.16	0.014375	Down	1560-84-5
8	2,4-Di-tert-butylphenol	DTBP	-0.12	$5.33E^{-10}$	Down	96-76-4
9	Benzene, 1,3-bis(1,1-dimethylethyl)-	BNDME	-0.03	$3.97E^{-31}$	Down	1014-60-4
10	Dodecane	DDE	0.16	$2.57E^{-08}$	Up	112-40-3
11	Tridecane, 3-methyl-	TDEM	0.28	0.002981	Up	6418-41-3

12	Tetradecane	TTDE	0.42	0.043541	Up	629-59-4
13	Hexadecane	HXDE	0.61	$1.39E^{-11}$	Up	544-76-3
14	Heneicosane	HNCS	0.56	0.001826	Up	629-94-7
15	Undecane, 3-methyl-	UNM	0.57	0.000369	Up	1002-43-3
16	Nonadecane	NNDE	0.75	$1.88E^{-18}$	Up	629-92-5
17	1-Octanol, 2-butyl-	1-OCTA	1.03	0.000421	Up	2/8/3913
18	Heptadecane, 2,6,10,15-tetramethyl-	HTDT	1.09	0.004413	Up	54833-48-6
19	Heptacosane	HPCS	1.37	$2.09E^{-18}$	Up	593-49-7
20	trans-1,2-Diethyl cyclopentane	TDECP	1.44	$1.1E^{-06}$	Up	932-40-1
21	Naphthalene	NPT	–	–	–	91-20-3
22	2-methylOctacosane	MOTDS	–	–	–	900376-72-8
23	Benzaldehyde, 2,5-dimethyl-	BNDM	–	–	–	5779-94-2
24	11-Methyldodecanol	MDD	–	–	–	85763-57-1
25	Heptane, 2,4-dimethyl-	HPDM	–	–	–	2213-23-2
26	2,4-Dimethylheptene	DMHE	–	–	–	19549-87-2

Metabolites 21-26 (Naphthalene, 2-methylOctacosane, Benzaldehyde 2 5-dimethyl-, 11-Methyldodecanol, Heptane 2 4-dimethyl- and 2 4-Dimethylhept-1-ene) are unique to the cancer subjects only. Of these, 2-methylOctacosane and 11-Methyldodecanol are being reported for the first time to be probably associated with breast and ovarian cancer. It was also observed that in VOCs common to both control and cancer subjects, the concentration of Dodecane, Tridecane 3methyl, Tetradecane, Hexadecane, Heneicosane , Undecane 3-methyl , Nonadecane, 1-Octanol 2-butyl, Heptadecane 2,6,10,5-tetramethyl, Heptacosane, trans-1,2-Diethyl cyclopentane was significantly higher in the urine of breast and ovarian cancer patients as compared to control subjects, while the concentration of Pentadecane, Undecane, 4,7-dimethyl-, 3-Heptyne-2,6-dione 5-methyl-5-(1-methylethyl), Octacosane, Decane, Heptadecane 3-methyl, Eicosane 2-methyl, 2,4-Ditertbutylphenol and Benzene 1,3-bis(1,1-dimethylethyl) is lower in cancer patients.

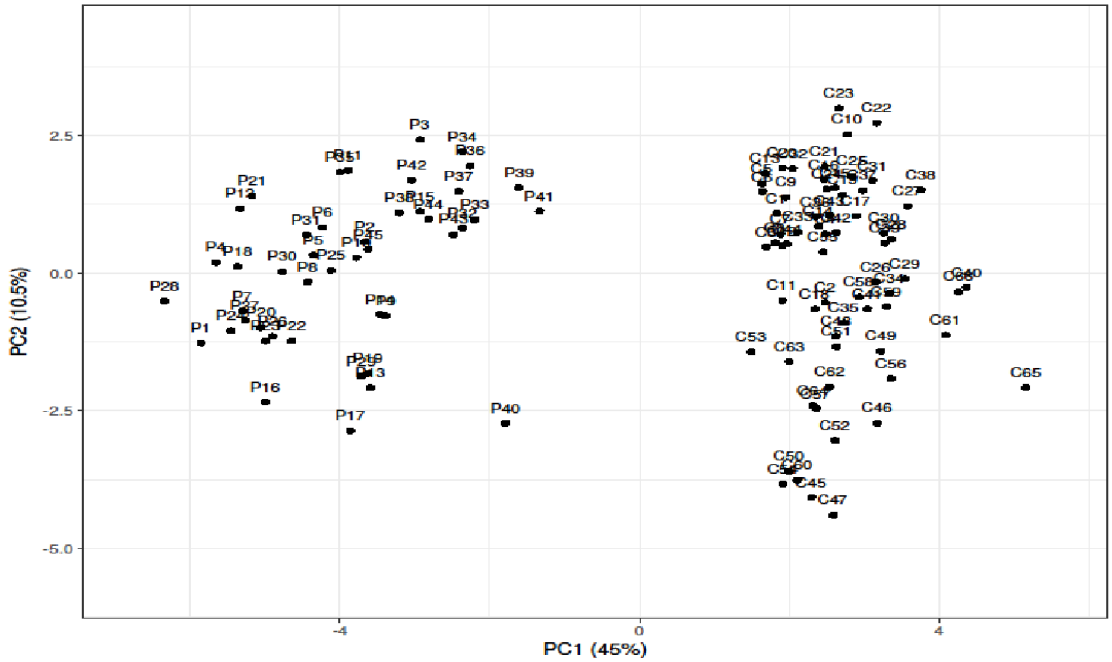

Figure 2a. PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26) observed.

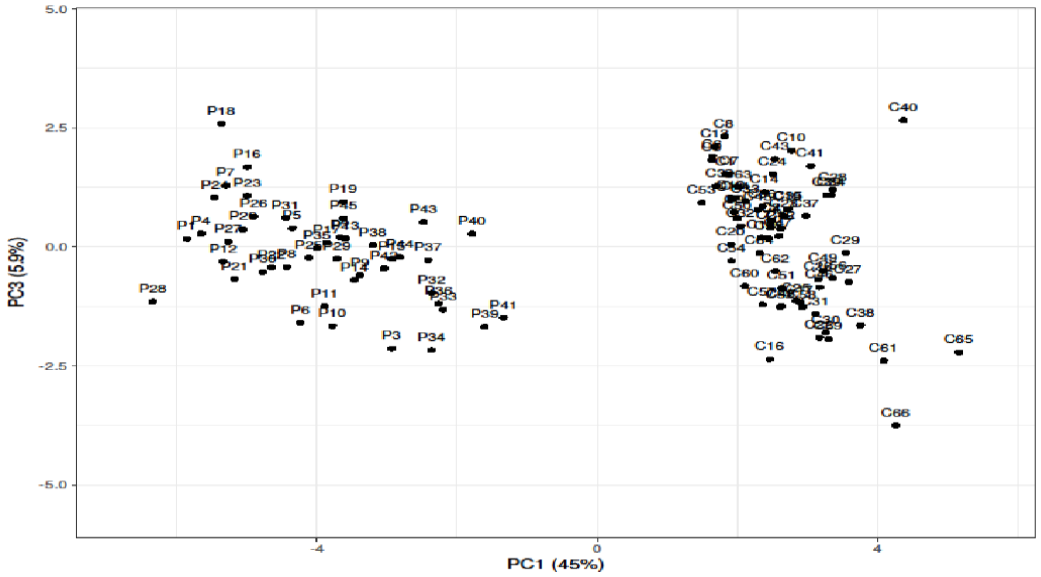

Figure 2b. PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26) observed.

Principal component analysis revealed that VOCs from cancer and control subjects formed two independent clusters indicating significant variation between their concentration levels (Fig. 2). Fig. 2a shows PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26). PC-1 and PC-2 showed 45% and 10.5% of the total variance between control and patient samples. Fig. 2b shows PCA plot of principal component 1 against principal

component 3 utilizing all identified VOCs (26) where 45% and 5.9% of the total variance between control and patient samples was observed.

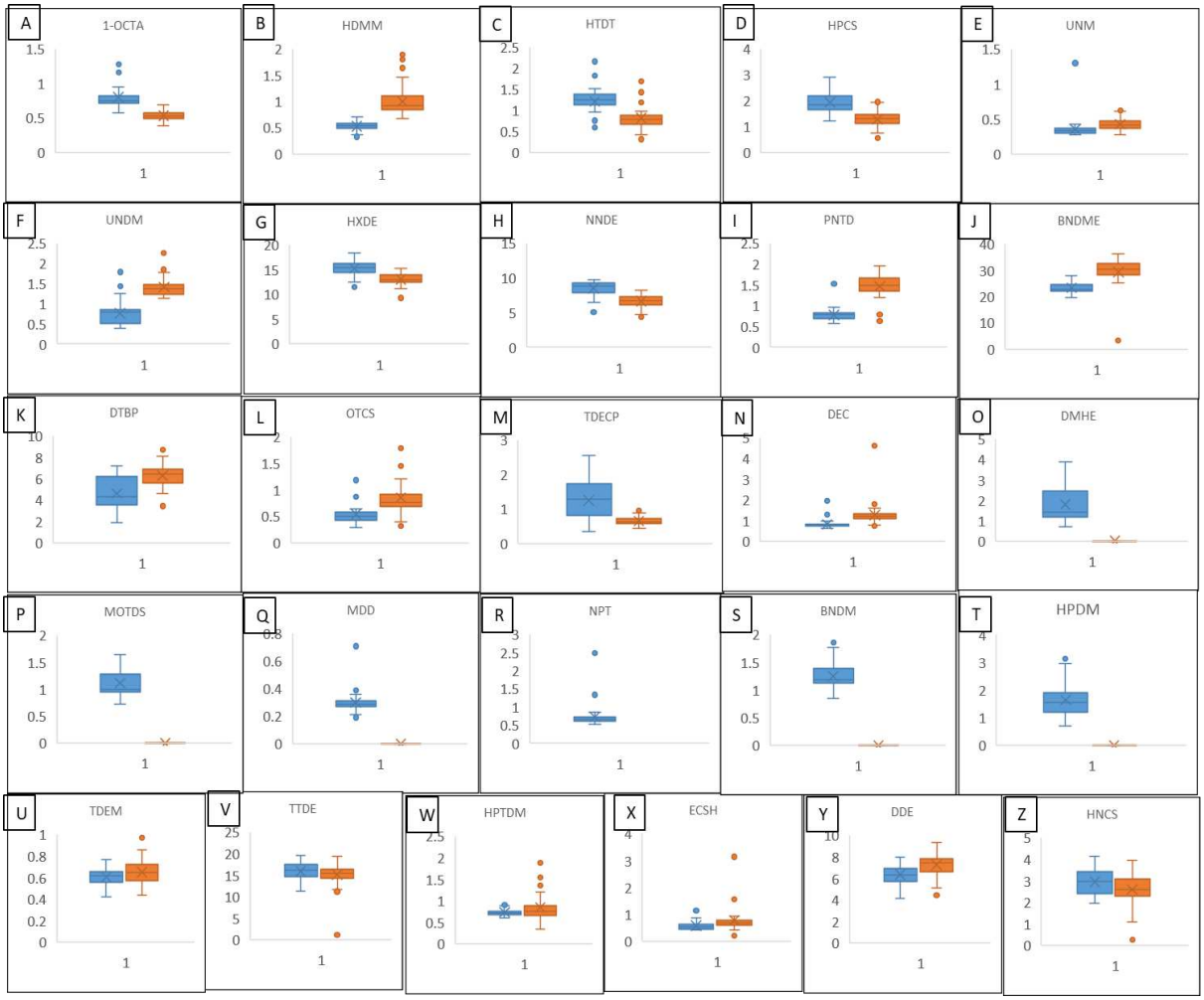

Figure 3: Box-Plot showing average peak area of 26 VOCs

Figure 3 shows Box-Plot of peak areas of 26 VOCs in cancer patients and control subjects. The relative significant difference was observed in the concentration of 20 metabolites (Fig. 3, Plots A to T). However, no significant difference in the concentration was observed in the remaining six compounds (Fig. 3, U to Z). This was also confirmed by ROC analysis where these six compounds were having suboptimal AUC (Table 2). Therefore, 20 compounds could be used as the reliable classifier to distinguish cancer patients from healthy subjects. Further, it was also observed that levels of Dodecane, Tridecane 3-methyl, Tetradecane, Hexadecane, Heneicosane , Undecane 3-methyl, Nonadecane, 1-Octanol 2-butyl, Heptadecane 2,6,10,15-tetramethyl, Heptacosane, trans-1,2-Diethyl cyclopentane in urine of breast and ovarian cancer patients were significantly higher than in clinically healthy subjects, while Pentadecane, Undecane 4,7-dimethyl, 3-Heptyne-2 6-dione 5-methyl-5-(1-methylethyl), Octacosane, Decane, Heptadecane 3-methyl, Eicosane 2-methyl, 2,4 Ditertbutylphenol and Benzene 1,3-bis

(1,1-dimethylethyl) were found to be lower in cancer subjects than that in clinically healthy Subjects. The specificity and sensitivity of different 26 VOC detected in GC-MS in control

Table 2: Specificity and sensitivity of different 26 VOC detected in GC-MS in control subject and cancer patients

Compounds Name	Abbr.	AUC	Standard error (SE at 95% Confidence interval)	Significance level (p value)	Associated criterion	Sensitivity (SE at 95% CI)	Specificity (SE at 95% CI)
Pentadecane	PNTD	0.954	0.896 to 0.985	<0.0001	≤0.963	97.78 (88.2 - 99.9)	92.42 (83.2 - 97.5)
Undecane, 4,7-dimethyl-	UNDM	0.959	0.904 to 0.988	<0.0001	≤0.942	93.33 (81.7 - 98.6)	100 (94.6 - 100.0)
3-Heptyne-2,6-dione,5-methyl-5-(1-methylethyl)-	HDMM	0.999	0.966 to 1.000	<0.0001	≤0.648	97.78 (88.2 - 99.9)	100 (94.6 - 100.0)
Octacosane	OTCS	0.885	0.770 to 0.955	<0.0001	≤0.649	90 (68.3 - 98.8)	88.57 (73.3 - 96.8)
Decane	DEC	0.922	0.855 to 0.964	<0.0001	≤0.934	93.33 (81.7 - 98.6)	87.88 (77.5 - 94.6)
Heptadecane, 3-methyl-	HPTDM	0.608	0.508 to 0.702	0.052	≤0.756	75 (59.7 - 86.8)	55.74 (42.4 - 68.5)
Eicosane, 2-methyl-	ECSH	0.708	0.599 to 0.801	0.0006	≤0.614	72.73 (54.5 - 86.7)	67.31 (52.9 - 79.7)
2,4-Di-tert-butylphenol	DTBP	0.81	0.725 to 0.878	<0.0001	≤4.594	62.22 (46.5 - 76.2)	98.48 (91.8 - 100.0)
Benzene, 1,3-bis(1,1-dimethylethyl)-	BNDME	0.942	0.881 to 0.977	<0.0001	≤25.248	86.67 (73.2 - 94.9)	95.45 (87.3 - 99.1)
Dodecane	DDE	0.788	0.700 to 0.860	<0.0001	≤7.166	84.44 (70.5 - 93.5)	68.18 (55.6 - 79.1)
Tridecane, 3-methyl-	TDEM	0.647	0.551 to 0.735	0.0045	≤0.678	95.56 (84.9 - 99.5)	36.36 (24.9 - 49.1)
Tetradecane	TTDE	0.592	0.495 to 0.685	0.1023	>16.59	44.44 (29.6 - 60.0)	81.82 (70.4 - 90.2)
Hexadecane	HXDE	0.889	0.816 to 0.941	<0.0001	>14.496	77.78 (62.9 - 88.8)	87.88 (77.5 - 94.6)
Heneicosane	HNCS	0.64	0.542 to 0.730	0.011	>3.13375	47.73 (32.5 - 63.3)	79.69 (67.8 - 88.7)
Undecane, 3-methyl-	UNM	0.891	0.806 to 0.947	<0.0001	≤0.386	95.35 (84.2 - 99.4)	68.89 (53.4 - 81.8)
Nonadecane	NNDE	0.922	0.855 to 0.964	<0.0001	>7.726	82.22 (67.9 - 92.0)	90.91 (81.3 - 96.6)
1-Octanol, 2-butyl-	1-OCTA	0.984	0.939 to 0.998	<0.0001	>0.6495	93.33 (81.7 - 98.6)	95.45 (87.3 - 99.1)
Heptadecane, 2,6,10,15-tetramethyl-	HTDT	0.953	0.894 to 0.984	<0.0001	>1.033	100 (92.1 - 100.0)	93.75 (84.8 - 98.3)
Heptacosane	HPCS	0.909	0.833 to 0.958	<0.0001	>1.519	91.11 (78.8 - 97.5)	84.31 (71.4 - 93.0)
trans-1,2-Diethyl cyclopentane	TDECP	0.986	0.942 to 0.999	<0.0001	>0.798	97.62 (87.4 - 99.9)	92.42 (87.4 - 99.9)
Naphthalene	NPT	1	0.967 to 1.000	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)
2-methylOctacosane	MOTDS	1	0.967 to 1.001	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)
Benzaldehyde, 2,5-dimethyl-	BNDM	1	0.967 to 1.001	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)
11-Methyldodecanol	MDD	1	0.967 to 1.001	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)
Heptane, 2,4-dimethyl-	HPDM	1	0.967 to 1.001	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)
2,4-Dimethylheptene	DMHE	1	0.967 to 1.001	<0.0001	>0	100 (92.1 - 100.0)	100 (94.6 - 100.0)

subject and cancer patients as determined by ROC curve are also shown in Table 2. The heatmap of these 26 VOCs is shown in Fig. 4.

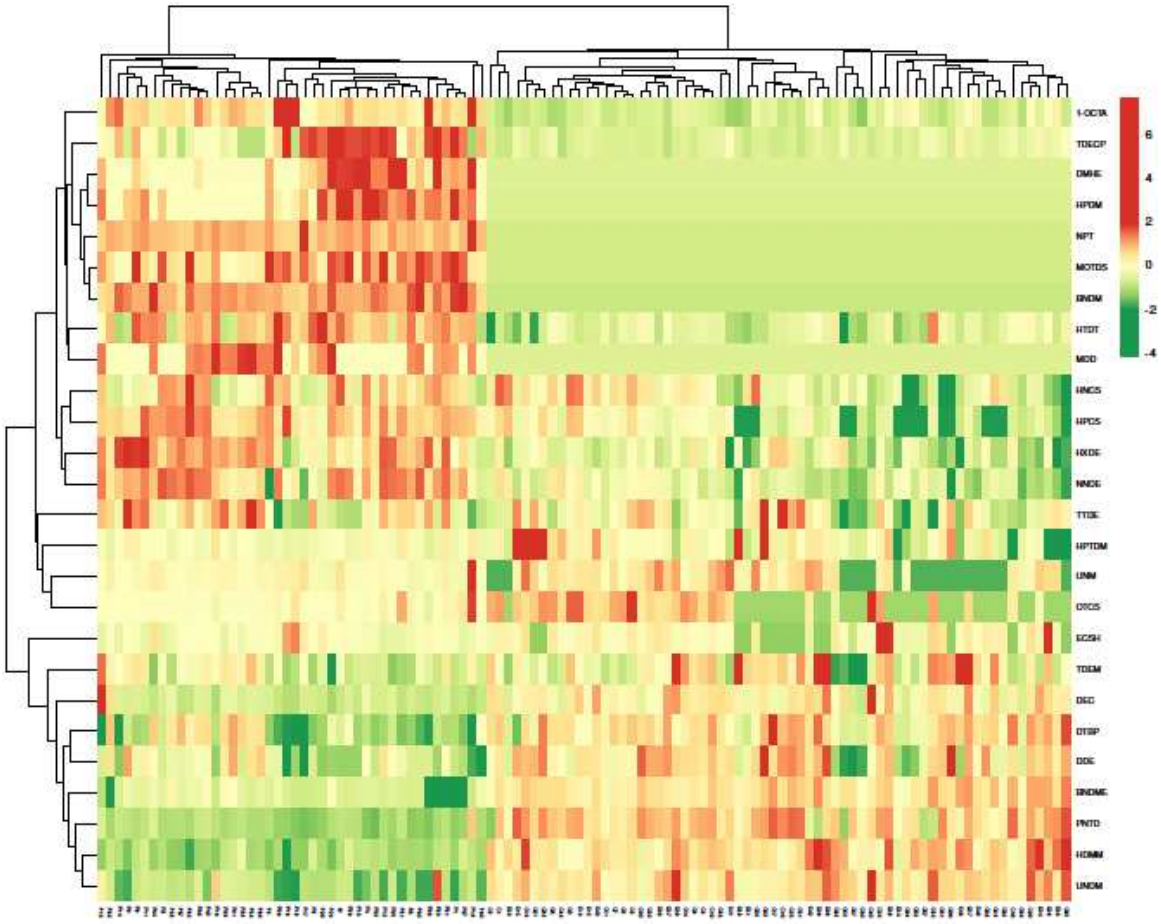

Fig. 4 Heatmap of average peakareas of 26 VOCs showing difference in their concentrations across patient and control subjects

3. Discussion

Volatile organic compound analysis is an efficient alternative approach for the disease diagnosis. Numerous studies have described the analysis of VOCs present in the biological samples using different methods such as GC-MS, Proton Transfer Reaction Mass-Spectrometry (PTR-MS)[17], Electronic nose device [18]. Saliva, breath, and blood are the promising sources for VOCs biomarker identification associated with an array of diseases. Many studies are being performed to explore this area of metabolomics and have reported confirmed, distinct and definite combinations of VOCs from patients with a variety of cancers such as breast, lung and colorectal [19202122]. The metabolites from urine are more useful not only because of the non-invasive method of sampling, but also because of their longer physiological shelf life as compared to those from breath and blood[23]. The meddling caused by the presence of other non-

significant primary and secondary metabolites in breath and blood are avoided in urine and the micromolecular metabolites associated with the breast and ovarian cancer can be easily filtered out and detected into the urine[24]. Additionally, for the extraction of VOCs, most of the methods have used Solid phase micro extraction (SPME)[25]. Hence, the present study considered exploiting the VOCs profiles from the urine of breast and ovarian cancer patients using solvent based extraction. The solvent based method described here provides a less expensive but highly effective method to extract VOCs associated with BOC in urine sample.

An increased oxidative stress and stimulation of polymorphic cytochrome P450 oxidase enzyme[26] are found to be associated with Breast cancer. This in turn is reported to cause peroxidation of lipids in membranes and affect the profusion of VOCs like alkanes and methyl alkanes in breath[27] have also reported a similar hypothesis where altered estrogen metabolism leads to higher level of aromatase (estrogen synthase/cytochrome P450 enzyme complex) expression in cancer tissue that is responsible for alteration in the metabolic profile of the VOCs. Pentadecane a VOC reported to be associated with cancer and also found in this study is also reported to be produced as a result of oxidative stress[28].

Of the 26 VOCs that were found in this study, 18 have already been reported to be associated with various types of cancer. Tridecane[27], Dodecane[27], Tetradecane (20), Undecane, 3-methyl-[29], Pentadecane[28], Benzene 1,3 bis (1,1-dimethyl)[25], 2,4 ditertbutyl phenol[25], 1-Octanol 2-butyl[27] have been found to be reported in breast cancer. Naphthalene[30], Decane[31], Tridecane 3-methyl[29], Hexadecane[32], Heneicosane[32], Heptane 2,4-dimethyl[33], Benzaldehye 2,4 dimethyl[34], Heptane 2,4dimethyl[33], Nonadecane[35], Heptacosane[36], Octacosane[36], Undecane 4,7 dimethyl[37] are also reported in studies carried out on lung cancer. Except for Benzene 1,3 bis (1,1-dimethyl)(22), 2 4 ditertbutyl phenol(22), Benzaldehye 2 4 dimethyl[37] all the reported metabolites have been extracted from breath samples or studies done using cancer cell lines. This also shows that the metabolites that are being reported in this study are also indications of breast and ovarian cancer irrespective of the sample used for the isolation.

To the best of our knowledge 2,4 Dimethylhept-1-ene, 3-Heptyne-2,6-dione, 5-methyl-5-(1-methylethyl), Eicosane 2-methyl, Heptadecane 3-methyl, Heptadecane 2,6,10,15-tetramethyl, trans-1,2-Diethyl cyclopentane, 2-methyl Octacosane, 11-Methyldodecanol are not yet reported to be associated with any types of cancer. Twenty metabolites which are simultaneously analyzed in this study, gives more specific and accuracy to diagnosis of breast and ovarian cancer.

To summarize, VOCs which were detected in higher concentration in cancer subjects as compared to controls as well as the six other VOCs that were found to be present only in cancer

subjects could be considered as promising biomarkers for diagnosing breast and ovarian cancer. For reducing the probability of false positive results, a library of biomarkers of normal and cancer subjects can be created by performing multivariate or chemometric analysis. Besides, GC-MS based analysis, development of point of care device/kit based on the physico-chemical properties of any one or more of these 6 biomarkers will be an advantage for diagnosis of a subject suffering from breast and ovarian cancer by non-invasive urine analysis.

The method described in the present study is a simple, rapid, sensitive, cost effective and non-invasive technique for the diagnosis of BOC using the urine sample. This method may be use for the screening of the breast and ovarian cancer. The study also contributed to the discovery of novel biomarkers of Breast and ovarian Cancer patients which can be further validated for investigations into their related metabolomics pathways and also paved the way for the development of diagnostic tools for Breast and ovarian Cancer patients. Similar studies can also be carried out for different types of cancer which would help to identify VOCs associated with them.

4. Material and Methods

2.1 Sample collection

The present study is approved by the ethics committee of the Gujarat Cancer Research Institute Ahmedabad, Gujarat, India. Urine samples were collected from BOC patients from Gujarat Cancer Research Institute, Ahmedabad, Gujarat. Informed consent for were obtained from participants. A consent form was filled with the details of the patients while collecting the samples. We further conform that all methods were performed in accordance with the relevant guidelines and regulations. Samples were collected from 44 women with the following criteria of individual subjects: breast and ovarian cancer patient (BOC) group diagnosed by clinical procedures, mostly stage I to III, with age between 30-60 years. The controls we selected were from the women participating in the breast cancer screening camp and diagnosed negative by the clinical procedures. Control samples consisted 66 healthy women of the similar age group. The urine samples were collected in sterile 15mL container and were stored at -80°C until analysis.

2.2 Extraction of VOCs and GC-MS

Prior to extraction, samples were defrosted by incubating the container at room temperature. For extraction of VOCs, samples were mixed with n-hexane in 2:1 ratio in sterile tubes and left overnight on tube rotator at room temperature. Then the samples were centrifuged at 3,500 rpm for five minutes and the upper solvent phase was separated carefully. From this, 2µL was used for the injection into the inlet of the GC-MS system at 250°C while the mass transfer line was

also held at 250°C. The injection was made in split mode (50:50). Helium was used as a carrier gas and passed through a helium purification system, oxytrap™ (PerkinElmer, USA). We used 30 meters long Elite-5 MS capillary column with having an inner diameter of 0.25 mm and 0.25 μm film thickness, suitable for the separation of VOCs (PerkinElmer, USA). The optimized GC-MS temperature program of the run was as follows; initial oven temperature hold at 70°C for the first 2 minutes, followed by ramping at a rate of 10°C/min to 150°C and subsequent 5°C/min ramp to 250°C. A 5 minute hold at this temperature was given to obtain a total run time of 40 minutes. The mass spectrometer was run in electron impact (EI) ionization mode and could scan the mass ions in the range of 30–400 at the rate of 0.05scan/sec. A 3 minute solvent delay to exclude solvent peak was used at the start of the run. The data was collected utilizing PerkinElmer Turbomass software. Each urine sample was extracted and analyzed in triplicates.

2.3 Data processing

The GC-MS data were processed using Turbomass software (ver.6.1.0). A matrix was generated by the software that included all the retention times and integrated signals for all extracted VOCs in each sample. The NIST database (version 2.2, 2014) was used to confirm the peak identity of VOC metabolites present in the urine samples (Table 1). A mean value of concentration was determined. The manual alignment was performed for analysis wherein a matrix corresponding to the sample with a maximum number of scans was used as a reference, and all the remaining samples were aligned against it. The compounds thus found were subjected to further statistical analysis.

2.4 Statistical Analysis

Chromatogram obtained in the control and cancer patients were compared for presence of difference in the number of peaks. Percent area under each peak was noted and log2 fold change in percent areas of all compounds in cancer subjects as compared to control was calculated in R statistics using t-test ($p<0.05$). Principal component analysis was performed to determine the pattern of VOCs in control and cancer subjects. In addition, some differential peaks obtained in GC-MS analysis were also considered as the biomarkers for the BOC. Box plots were prepared to represent the distribution of fold change in particular VOC. Receiver Operating Characteristics (ROC) curves were plotted to quantify the sensitivity and specificity of VOC to distinguish between cancer patients and control as well as to determine the cut-off values using MedCalc. The heatmap was generated using a ClustVis by all log2 fold change values of all VOCs detected in each sample.

Acknowledgement:

The authors would like to thank Department of Science and Technology (DST) Government of Gujrat India for funding the project. We also wish to acknowledge Gujrat Cancer Research Institute Ahmedabad for providing urine sample of cancer patients.

Competing interests:

An Indian patent application (202121002808) has been published for the technology/invention disclosed in the present work.

References:

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **0**, 1–41 (2021).
3. Ali, I., A. Al-Othman, Z., Saleem, K., Hussain, A. & Hussain, I. Role of Chromatography for Monitoring of Breast Cancer Biomarkers. *Recent Pat. Biomark.* **1**, 89–97 (2011).
4. Zhang, A. H., Sun, H., Qiu, S. & Wang, X. J. Metabolomics in noninvasive breast cancer. *Clin. Chim. Acta* **424**, 3–7 (2013).
5. Krassenstein, R. *et al.* Detection of Breast Cancer in Nipple Aspirate Fluid by CpG Island Hypermethylation. *Clin. Cancer Res.* **10**, 28–32 (2004).
6. Aboud, O. A. & Weiss, R. H. New opportunities from the cancer metabolome. *Clin. Chem.* **59**, 138–146 (2013).
7. Sethi, S., Nanda, R. & Chakraborty, T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clin. Microbiol. Rev.* **26**, 462–475 (2013).
8. Amann, A. *et al.* The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *J. Breath Res.* **8**, (2014).
9. Lindon, J. C., Holmes, E. & Nicholson, J. K. Metabonomics and its role in drug development and disease diagnosis. *Expert Rev. Mol. Diagn.* **4**, 189–199 (2004).
10. Kelly, R. S., Heiden, M. G. V., Giovannucci, E. & Mucci, L. A. *Metabolomic biomarkers of prostate cancer: Prediction, diagnosis, progression, prognosis, and*

recurrence. *Cancer Epidemiology Biomarkers and Prevention* vol. 25 (2016).

11. Gao, Q. & Lee, W.-Y. Urinary metabolites for urological cancer detection: a review on the application of volatile organic compounds for cancers. *Am. J. Clin. Exp. Urol.* **7**, 232–248 (2019).
12. Monteiro, M. *et al.* GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. *J. Cell. Mol. Med.* **21**, 2092–2105 (2017).
13. Wu, W. & Zhao, S. Metabolic changes in cancer: Beyond the Warburg effect. *Acta Biochim. Biophys. Sin. (Shanghai).* **45**, 18–26 (2013).
14. Rautiainen, S. *et al.* Axillary lymph node biopsy in newly diagnosed invasive breast cancer: Comparative accuracy of fine-needle aspiration biopsy versus core-needle biopsy. *Radiology* **269**, 54–60 (2013).
15. Lavra, L. *et al.* Investigation of VOCs associated with different characteristics of breast cancer cells. *Sci. Rep.* **5**, 1–12 (2015).
16. Ong, M. S. & Mandl, K. D. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $ 4 billion a year. *Health Aff.* **34**, 576–583 (2015).
17. Brunner, C. *et al.* Discrimination of cancerous and non-cancerous cell lines by headspace-analysis with PTR-MS. *Anal. Bioanal. Chem.* **397**, 2315–2324 (2010).
18. De León-Martínez, L. D. *et al.* Identification of profiles of volatile organic compounds in exhaled breath by means of an electronic nose as a proposal for a screening method for breast cancer: A case-control study. *J. Breath Res.* **14**, (2020).
19. Phillips, M. *et al.* Detection of lung cancer with volatile markers in the breath. *Chest* **123**, 2115–2123 (2003).
20. Hietanen, E. *et al.* Diet and oxidative stress in breast, colon and prostate cancer patients: a case-control study. *Eur. J. Clin. Nutr.* **48**, 575—586 (1994).
21. Wang, C. *et al.* Blood volatile compounds as biomarkers for colorectal cancer. *Cancer Biol. Ther.* **15**, 200–206 (2014).
22. Alwis, K. U., Blount, B. C., Britt, A. S., Patel, D. & Ashley, D. L. Simultaneous analysis of 28 urinary VOC metabolites using ultra high performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (UPLC-ESI/MSMS). *Anal. Chim. Acta* **750**, 152–160 (2012).
23. Altomare, D. F. *et al.* Exhaled volatile organic compounds identify patients with colorectal cancer. *Br. J. Surg.* **100**, 144–150 (2013).

24. Wood, D. P. Re: Metabolomics study on the biochemical profiles of odor elements in urine of human with bladder cancer. *J. Urol.* **189**, 1288 (2013).
25. Woollam, M. *et al.* Detection of Volatile Organic Compounds (VOCs) in Urine via Gas Chromatography-Mass Spectrometry QTOF to Differentiate Between Localized and Metastatic Models of Breast Cancer. *Sci. Rep.* **9**, 1–12 (2019).
26. Murray, G. I. The role of cytochrome P450 in tumour development and progression and its potential in therapy. *J. Pathol.* **192**, 419–426 (2000).
27. Phillips, M. *et al.* Volatile biomarkers in the breath of women with breast cancer. *J. Breath Res.* **4**, (2010).
28. Wang, D. *et al.* Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. *Biomed. Reports* **5**, 68–72 (2016).
29. Phillips, M., Cataneo, R. N., Ditkoff, B. A. & Bevers, T. B. Volatile markers of breast cancer in the breath. *Breast Dis.* **15**, 53 (2004).
30. O'Neill, H. J., Gordon, S. M., O'Neill, M. H., Gibbons, R. D. & Szidon, J. P. A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. *Clin. Chem.* **34**, 1613–1618 (1988).
31. Corradi, M. *et al.* Exhaled breath analysis in suspected cases of non-small-cell lung cancer: A cross-sectional study. *J. Breath Res.* **9**, 27101 (2015).
32. Thriumani, R. *et al.* A study on volatile organic compounds emitted by in-vitro lung cancer cultured cells using gas sensor array and SPME-GCMS. *BMC Cancer* **18**, 1–17 (2018).
33. Phillips, M. *et al.* Volatile organic compounds in breath as markers of lung cancer: A cross-sectional study. *Lancet* **353**, 1930–1933 (1999).
34. Stokes, P. Patent et Application. **1**, 10–18 (2018).
35. Wang, Y. *et al.* The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomarkers* **11**, 129–137 (2012).
36. Gang, P. & Data, R. U. S. A. (12) United States Patent. **2**, (2013).
37. Il, T. (12) United States Patent. **2**, (2017).

Figure legends:

Figure 1: Representative chromatograms of VOCs from control and patient urine sample. Enlarged comparative profiles of novel VOCs 11-methyldodecanol (a) and 2-methyloctacosane (b) are shown at their respective retention times.

Figure 2a. PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26) observed.

Figure 2b. PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26) observed.

Figure 3: Box-Plot showing average peak area of 26 VOCs

(Plots A to N show plots of metabolites having significant variation ($p<0.05$) in peak areas of control subjects and patient samples. Plots O to T are plots of metabolites found only in patient samples. Plots U to Z are of metabolites showing no significant difference ($p>0.05$) in the peak areas of control subjects and patient samples.)

Figure 4: Heatmap of average peak areas of 26 VOCs showing difference in their concentrations across patient and control subjects

Figures

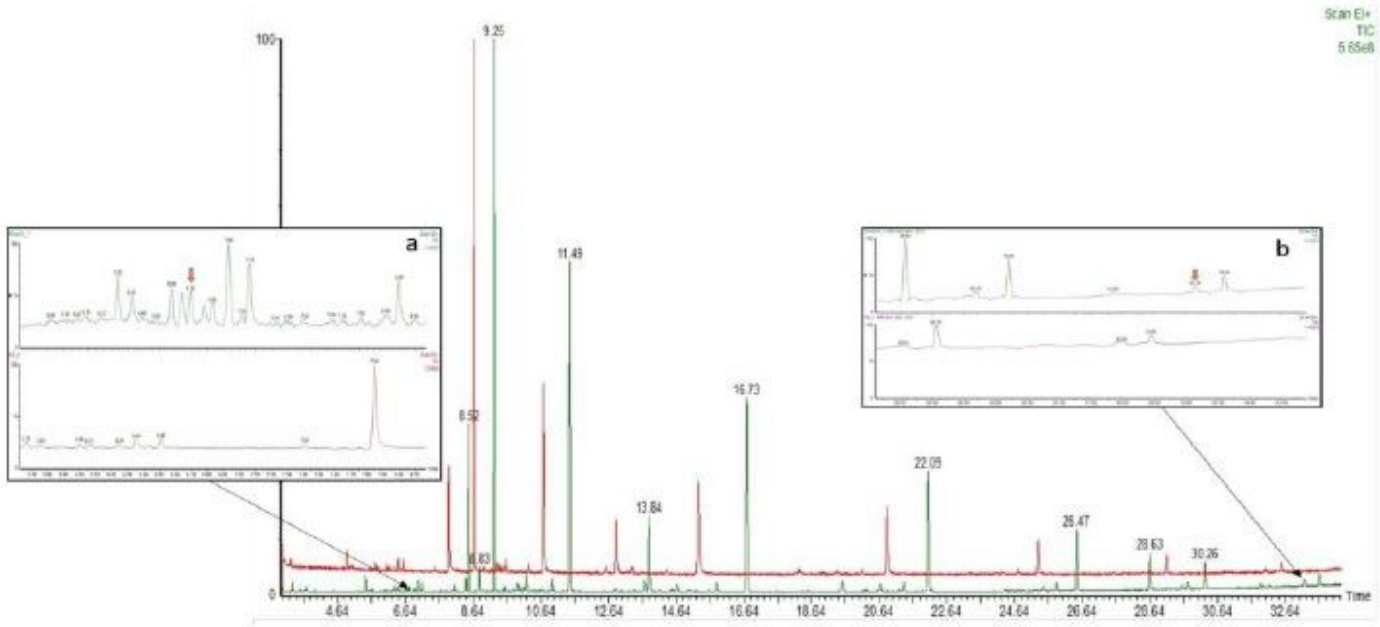

Figure 1

Representative chromatograms of VOCs from control and patient urine sample. Enlarged comparative profiles of novel VOCs 11-methyldodecanol (a) and 2-methyloctacosane (b) are shown at their respective retention times.

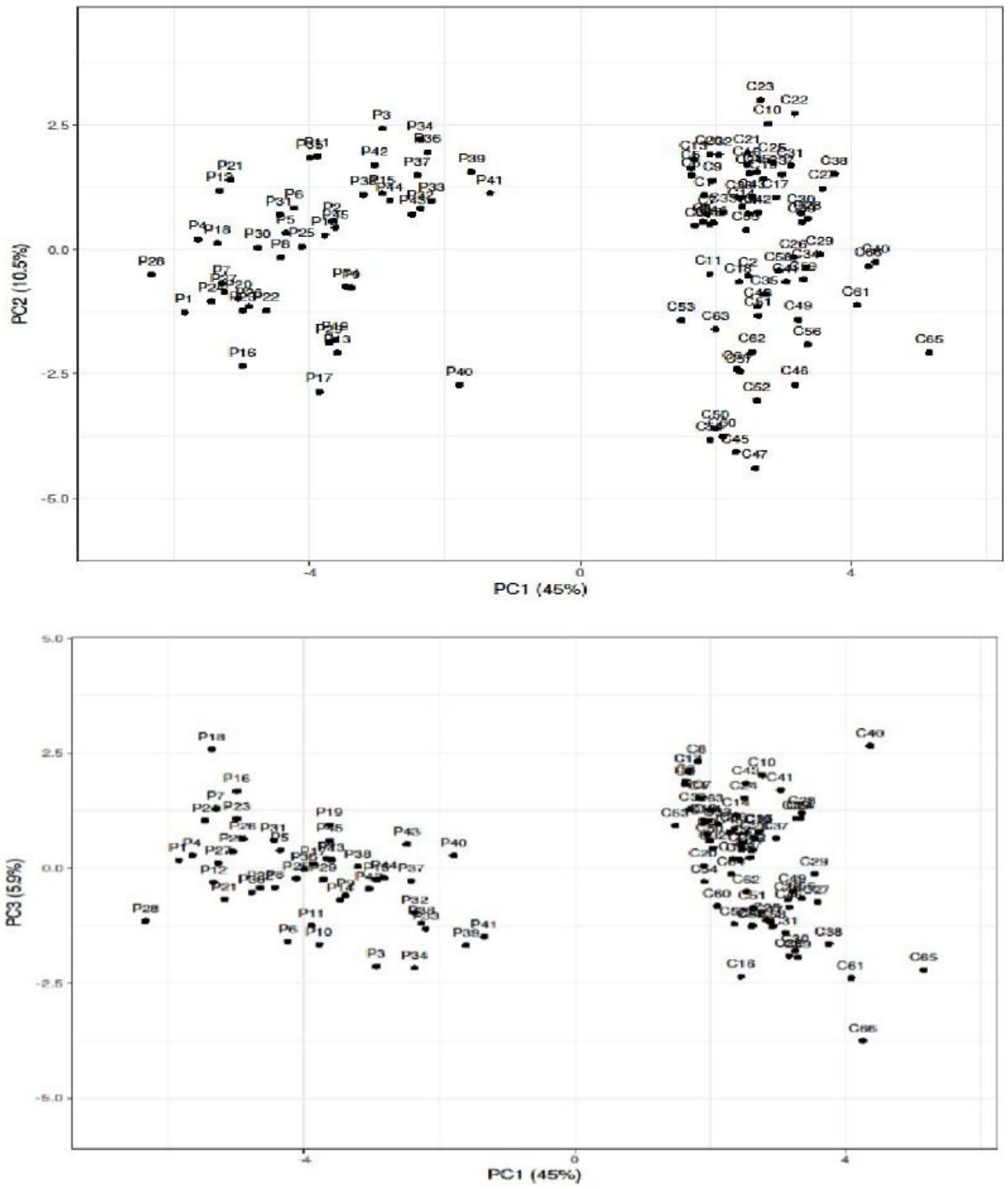

Figure 2

a (top): PCA plot of principal component 1 against principal component 2 utilizing all identified VOCs (26) observed. b (bottom): PCA plot of principal component 1 against principal component 3 utilizing all identified VOCs (26) observed.

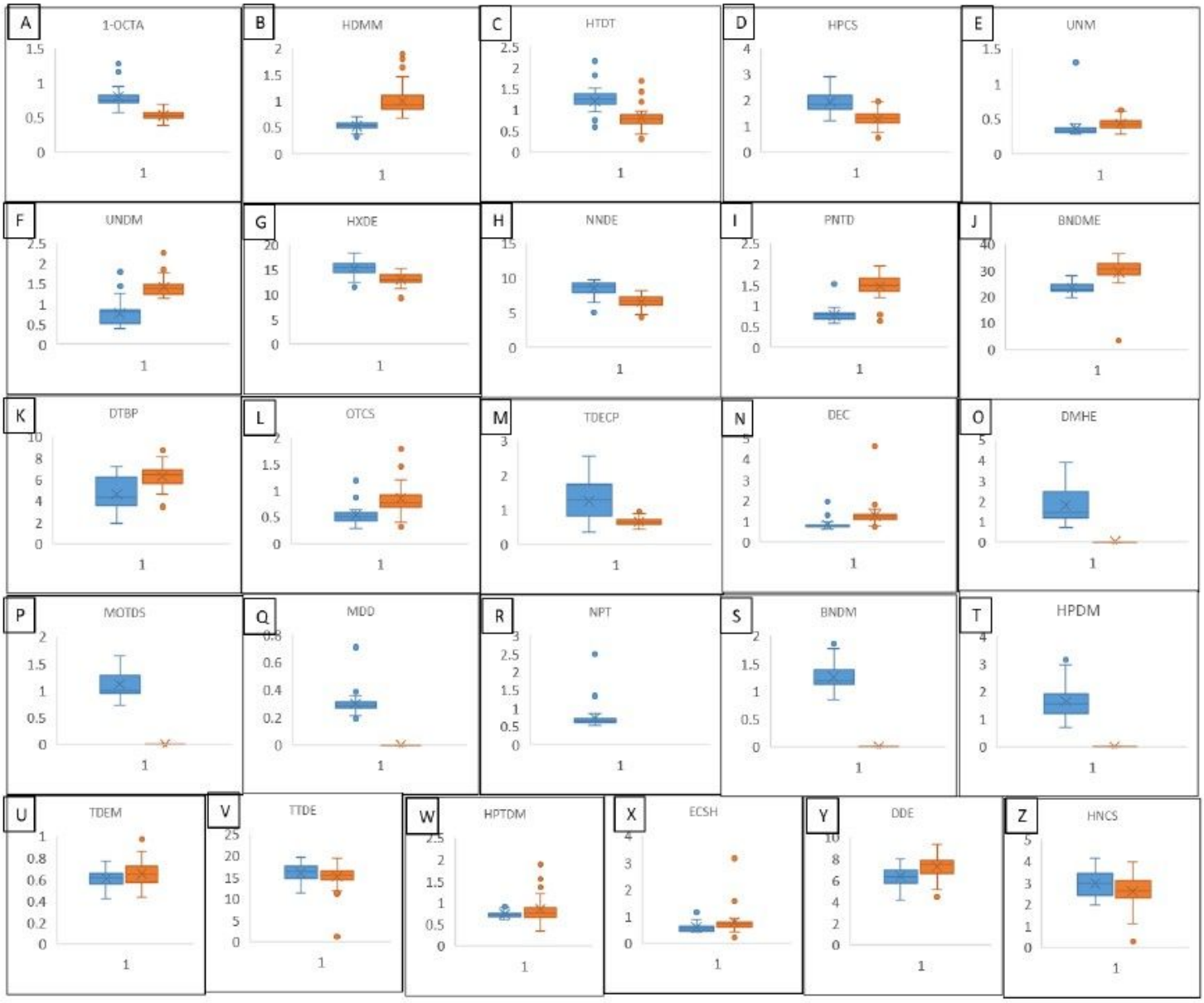

Figure 3

Box-Plot showing average peak area of 26 VOCs. (Plots A to N show plots of metabolites having significant variation ($p<0.05$) in peak areas of control subjects and patient samples. Plots O to T are plots of metabolites found only in patient samples. Plots U to Z are of metabolites showing no significant difference ($p>0.05$) in the peak areas of control subjects and patient samples.)

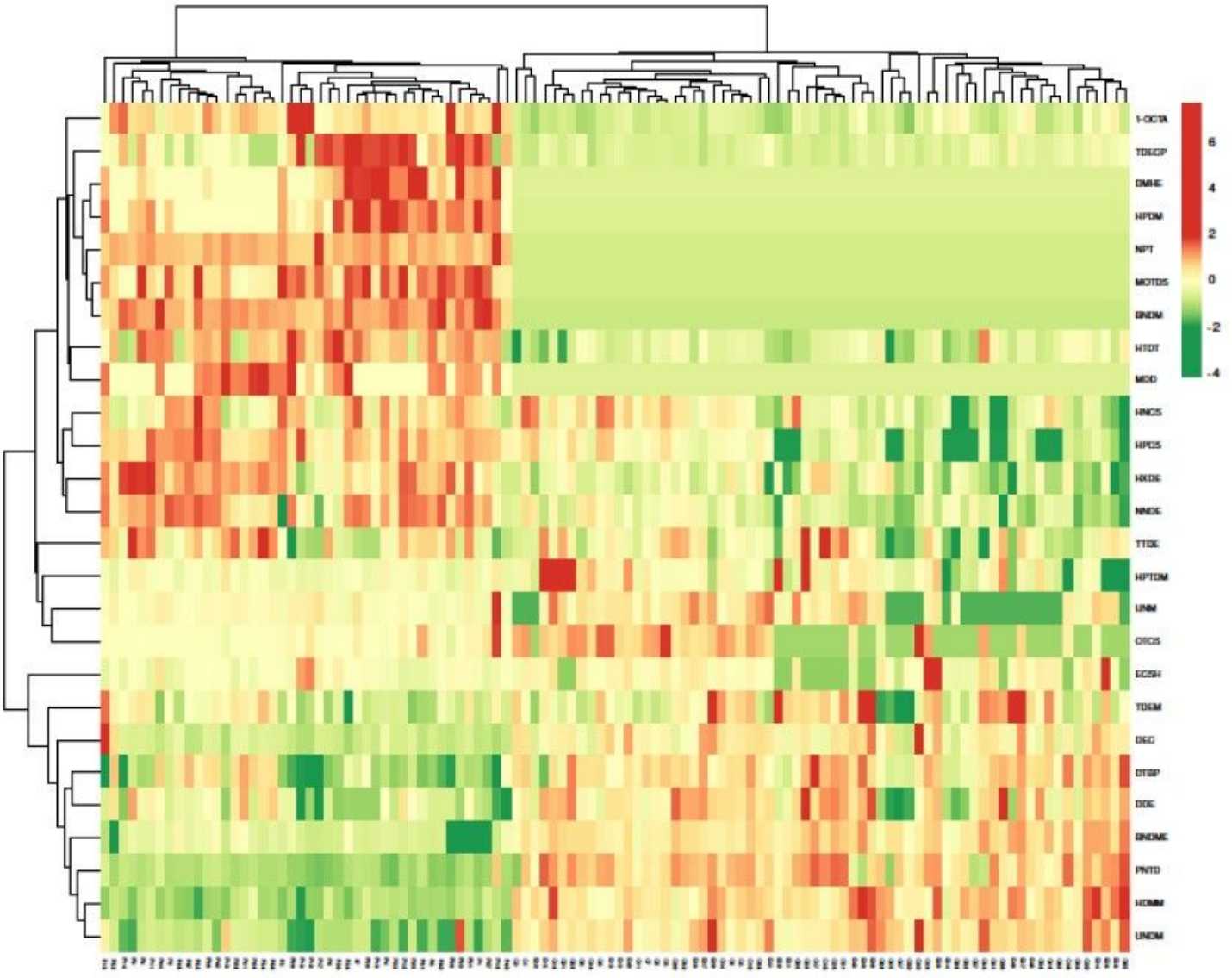

Figure 4

Heatmap of average peakareas of 26 VOCs showing difference in their concentrations across patient and control subjects