

## Biomarkers for the Detection of Pre-Cancerous Stage of Cervical Dysplasia

Paz Elia<sup>1</sup>, Stella Raizelman<sup>1</sup>, Esti Katorza<sup>1,2</sup>, Yosef Matana<sup>1</sup>, Offer Zeiri<sup>3</sup>, Zvi Boger<sup>1,4</sup>, Manuela Cervelli<sup>5</sup>, Paolo Mariotini<sup>5</sup>, Cristina Vallone<sup>6</sup>, Fabrizio Signore<sup>6</sup> and Yehuda Zeiri<sup>1,7\*</sup>

<sup>1</sup>Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>2</sup>Division of Radiotoxicology, Nuclear Research Center - Negev, Beer-Sheva, Israel

<sup>3</sup>Department of Chemistry, Ben-Gurion University of the Negev, Beer-Sheva, Israel

<sup>4</sup>OPTIMAL - Industrial Neural Systems Ltd., Beer-Sheva, Israel

<sup>5</sup>Department of Sciences, University "Roma Tre", Rome, Italy

<sup>6</sup>Department of Gynecology, Breast Surgery and Pathology, San Camillo-Forlanini Hospital, Rome, Italy

<sup>7</sup>Division of Chemistry, Nuclear Research Center - Negev, Beer-Sheva, Israel

### Abstract

**Introduction:** Early diagnosis of cancer can dramatically increase healing probability. However many cancer detection methods are time-consuming, invasive, and require skilled medical staff and/or expensive detection systems. Cervical cancer is the fourth most common malignant disease among women, and the fourth leading cause of cancer death in women worldwide.

**Aim:** This pilot study sought to identify reliable biomarkers indicative of early stages of cervical dysplasia, by analysis of changes in volatile organic compound composition in urine samples.

**Methods:** Urine samples of 17 patients with cervical intraepithelial neoplasia (CIN I) and of 9 healthy female subjects were used. The sample composition was analyzed using Gas-Chromatography-Mass-Spectrometry. The statistical analysis of the data was performed using supervised artificial neural networks.

**Results:** We identified four molecules with potential to serve as biomarkers of cervical dysplasia together with two molecules whose absence in the urine can confirm existence of cervical dysplasia. All indications shows that these six potential biomarkers are produced in the body during various physiological processes enhances in sick women. Hence, these potential biomarkers are not related to environmental or dietary origins.

**Conclusion:** Validation of the statistical method used, indicated that the biomarkers identified are highly reliable for detection of cervical dysplasia.

**Keywords:** Cervical cancer; Cervical dysplasia; Cervical intraepithelial neoplasia; Volatile organic compound; Bio-markers in urine; Supervised artificial neural networks

### Introduction

Early diagnosis of cancer can dramatically increase healing probability. However, many of the methods used for cancer detection are time-consuming, invasive, and require skilled medical staff and/or expensive detection systems such as computer tomography (CT), magnetic resonance imaging (MRI), endoscopy and ultrasonography [1-4]. In recent years much research is focused towards development of safe, reliable, non-invasive and inexpensive early detection schemes. A very promising route is finding specific and reliable biomarkers that are indicative of early stages of cancer. Biomarkers have shown potential for detection of various diseases, including cancer, by their identification in exhaled breath, blood and urine. Urine samples have high potential to contain biomarkers indicative of the physiological condition, indeed, biomarkers were identified for prostate and lung cancers [5,6], tuberculosis [7], exposure to toxic vapors [8] and diabetes mellitus [9]. Identification of biomarkers requires combining skills from different disciplines including analytical chemistry and statistical data analysis [10].

Cervical cancer was the fourth most common diagnosed cancer in women in 2012 and the fourth leading cause of cancer related death in women worldwide [11]. A cervical tumor develops from abnormal cell growth and has been linked to the human Papilloma virus (HPV) [12]. In many cases, infection with certain types of HPV is the first step in the progression from a normal cervix to cervical cancer. It is well established that sexually transmitted HPV induces the growth

of abnormal cells that can become malignant [13,14]. As cancer cells form, cells of abnormal size and shape appear on the surface of the cervix and begin to multiply. Cervical cancer can be detected using a laboratory test that examines cervical cells obtained through a gynecological procedure called a Papanicolaou test (Pap test in short) or by a new technology based on liquid-based cytology [15]. The most effective way of screening for cervical cancer is through routine Pap tests or by testing for human Papilloma virus. Women who undergo routine screening have a better chance of early diagnosis and treatment [16,17].

Cervical dysplasia is the term used to describe the early growth of abnormal cells on the cervix that could progress to cancer. Cervical dysplasia is usually the first stage of cervical cancer, but women with cervical dysplasia do not necessarily develop cancer. Dysplastic cells look like cancer cells, but they are not considered malignant provided that they remain on the surface of the cervix and do not invade

**\*Corresponding author:** Yehuda Zeiri, Biomedical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel, Tel +972-8-6479611; Fax +972-8-6472984; E-mail: [yehuda@bgu.ac.il](mailto:yehuda@bgu.ac.il)

**Received** September 24, 2015; **Accepted** October 28, 2015; **Published** October 30, 2015

**Citation:** Elia P, Raizelman S, Katorza E, Matana Y, Zeiri O, et al. (2015) Biomarkers for the Detection of Pre-Cancerous Stage of Cervical Dysplasia. J Mol Biomark Diagn 6: 255. doi:10.4172/2155-9929.1000255

**Copyright:** © 2015 Elia P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

healthy tissue. Cervical dysplasia is classified by three stages: Cervical Intraepithelial Neoplasia I, II and III (CIN I, CIN II and CIN III, [18], described in more details in the appendix. Women with a pre-cancerous condition will in most cases remain under physician follow-up, since progression to cervical carcinoma in situ (CIS) occurs in approximately 11% of the CIN I cases and progression to invasive cancer occurs in approximately 1% of the cases [19]. Most cases of cervical carcinoma can be prevented through proper screening. In fact, during 2012 in developed countries, with better awareness for early detection and better diagnostic tools, cervical cancer was only the ninth leading cause of cancer death in women, while in undeveloped countries cervical cancer was the third leading cause of cancer related death [11]. In addition, errors in cervical sampling and in sample interpretation are common. Consequently, reports of Pap test sensitivity and specificity vary significantly, hence, current screening is far from being sufficiently accurate [20,21].

Herein, we describe results of a pilot study in which chemical analysis of volatile organic compounds (VOCs) in urine samples are compared between women with cervical dysplasia in stage CIN I and urine samples of healthy women. The small sample size used in this pilot study is related to the limited budget provided for this pilot research. The main goal of the study is to try and identify statistically meaningful potential biomarkers that will allow rapid, non-invasive, reliable and unexpansive test for cervical dysplasia in the CIN I stage.

## Methods

Composition analysis of VOCs in the headspace over urine samples was performed using gas chromatography (GC) combined with mass spectrometry (MS). The volatile organic compounds composition in the headspace of urine samples of 17 patients with CIN I and 9 samples of healthy female subjects were analyzed.

## Subjects

Cervical dysplasia/cancer patients were admitted to the Department of Obstetrics and Gynecology at San Camillo-Forlanini Hospital (Rome, Italy). Colposcopy exams were carried out in all patients before surgery. After LEEP (Loop Electrosurgical Excision Procedure) and pathological analysis, the histological type and grade of the cervical tumors were determined. Only CIN I grade patients were selected for the study and their urine samples were collected. All procedures involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All patients included in the study signed an informed medical consent form, previously approved by the Institutional Ethical Committee and the Medical Board of the San Camillo-Forlanini Hospital. Urine samples were obtained from 17 female CIN I patients (mean age 36.8 years, SD  $\pm$  9.9) and 9 healthy female subjects (mean age 39.6 years, SD  $\pm$  14.2). See details of the subjects in the Appendix Table A2. All healthy women participants did not present any symptoms of any kind of cancer, acute inflammation, flu, pregnancy, or infectious diseases before and during the period of the experiments. Participants were asked to supply 50 ml of urine in sterile screw-top plastic vials. Urine samples were divided into two 25 ml septum topped vials and frozen at  $-18^{\circ}\text{C}$  until use.

## Sample analysis

GC-MS analysis was performed using an Agilent 6890 series GC system (Agilent, USA) connected to Agilent 5973 network mass Selective detector (Agilent). The samples were concentrated prior

to all measurements and introduced into the GC system using static headspace sample extraction, achieved by exposure of a 65  $\mu\text{m}$  polydimethylsiloxane/divinylbenzene (PDMS/DVB) solid-phase microextraction (SPME) fiber (SUPELCO, Bellefonte, PA) to the headspace over the solvent-free urine sample for 20 min at  $80^{\circ}\text{C}$ . A detailed description of the chromatographic analysis procedures used is given in the appendix section.

## Chromatogram pre-processing

The noise level in each chromatogram was evaluated using a home written MATLAB code, according to clinical and laboratory standards institute (CLSI) guideline EP17. The code was used to set the threshold between noise and actual peak according to the Limit of Detection (LOD) definition which is given in the appendix section. The retention times of peaks in the chromatogram were assigned by identification of local maxima that are above the threshold value. The measured chromatograms were represented by a pair of vectors containing peak retention time and their area. The fluctuations in peak positions were measured by analysis of our GC calibration mixture described in the Appendix section. The chromatograms of 40 measurements performed on different days were analyzed to yield a maximum uncertainty at peak position of 1.2 sec. Thus, the entire measurement period was sub-divided into time intervals of 2.4 sec which ensured that identical peaks in two different chromatograms will be located in the same time interval. The area of each peak in the chromatogram was normalized twice consecutively; first by the creatinine level found in the urine sample and then by the largest peak area in the spectrum. Hence, all area values in all spectra were in the range of 0-1.

## Statistical analysis

The statistical analysis used is a non-linear approach, based on Artificial Neural Networks (ANN). The ANN contained two layers of "neurons" appropriately connected by weights. Data inputs were connected to the neurons in the first layer ("hidden" neurons), which were connected in turn to the second layer of the "output" neurons. Adjusting the values of the weights between the neurons during the training of the ANN was carried out using "back-propagation" of the errors between the output neurons and the known data outputs. A constant bias threshold is included as additional input to all hidden and output neurons.

Once the ANN is trained, it is verified by presenting examples not used in the training. ANN modeling has been used in analyzing bio-medical data [22,23], and high dimensional data was successfully modeled and analyzed by the GB ANN algorithm set [24].

The chromatogram obtained for each sample was assigned as belonging to one of two classes: 'Healthy' or 'Sick'. The contribution vector for the separation between the two classes was calculated. The compounds with the largest values in this vector were assigned as possible biomarkers of cervical dysplasia [25-27]. Once a trained ANN is available, it can be analyzed to extract the identification of the more relevant features [28]. The causal index (CI) based algorithm [29] was found to be very useful in revealing the influence of input change on the change in relative magnitude and sign of each output. A detailed description of this approach is given in the appendix section.

## Validation

Two cross-validation methods were used to examine the generalization of the statistical model obtained by the approach described above. Both validation methods used are suitable for

application to cases with small numbers of samples such as those described in this study [30]. The two validation methods used were: Leave one out (LOO) and 7-fold [31,32]. In the 7-fold cross validation the original sample set is randomly partitioned into seven equal size subsamples. One of the subsamples is chosen for validation while the other subsample sets are used as training data. The process is repeated 7 times. The LOO validation approach subdivides the samples space into one sample used for validation and all remaining samples used for training. The process is repeated until all samples have been tested for validation. The classification of each prediction sample for LOO or prediction set for 7-fold was determined by the classification output of each model, as explained in the appendix section.

## Results and Discussion

### GC-MS results

A comparison between typical GC-MS chromatograms of urine samples obtained from healthy and sick women samples is presented in Figure 1. Four compounds that were present in all urine samples examined were identified. These four compounds served as internal references in the data analysis stage. The internal reference compounds peaks are marked by arrows in Figure 1 as Iref. 1 (ammonia), Iref. 2 (hexamethyl-cyclotrisiloxane), Iref. 3 (octamethyl-cyclotetrasiloxane) and Iref. 4 (1,3-dihydro-5-methyl-2H-Benzimidazol-2-one). The origin of these internal reference peaks is related to both the urine sample composition and the experimental system (column and SPME fiber coatings).

Inspection of the chromatograms in Figure 1 shows clear differences between the data of sick and healthy women. There are many peaks that appear in both healthy and sick urine samples but it is clear that each chromatogram exhibits a large number of unique peaks that do not exist in the other chromatogram. This is not surprising, since urine samples of different individuals are expected to exhibit variations due to differences in diet, habits and physiology. Thus, the central goal of the present study is to check whether one can find a group of statistically meaningful peaks that can uniquely identify the urine sample as belonging to a sick woman.

### Data pre-processing

The output obtained by the GC-MS software (enhanced msd ChemStation Vers. E.02.01.1177) is in the form of two column vectors containing the retention time assigned to each peak detected and the area under the peak respectively. All area values in the chromatogram were first normalized according to the concentration of creatinine in

the sample. Next, since ammonia (Iref. 1) exhibited the largest peak area in all chromatograms, its value was used to normalize the area of all other peaks in the chromatogram. The peak retention time vector was mapped onto a vector that contained 625 elements, each of duration of 2.4 sec. This time box size was defined by the statistical error in retention time of peaks corresponding to different known compounds used for the calibration. Thus, each time box could contain only a single peak with retention time equal to the value at the box center. This procedure ensured that identical peaks were assigned identical retention times. The normalized peak area was assigned to the box, and if no peak existed in a given time interval, its value was defined to be zero.

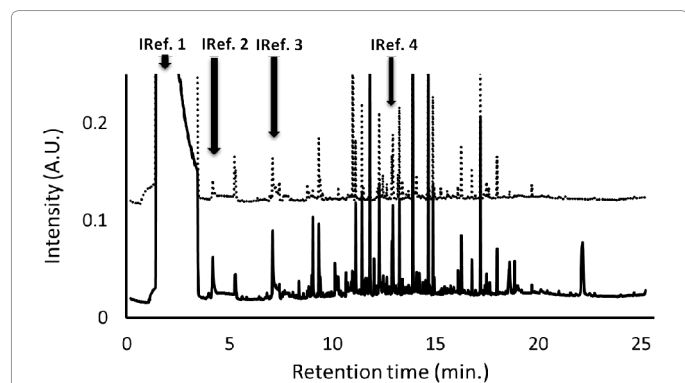
### Statistical analysis and its verification

A supervised ANN was used for the modeling the GC-MS data where all the input features were connected to 5 neurons in the hidden layer and those were connected to a single neuron in the output layer. The back propagation method was used to perform the training of the ANN using 70% (18 out of 26) randomly chosen chromatograms. The convergence of the training set to a model that correctly reproduced all the training samples is shown in the Appendix, Figure A1. The output level contained a single neuron whose value was in the range 0.1-0.9 for all samples. Output values in the range 0.1-0.4 were classified as belonging to sick women; those between 0.6-0.9 as belonging to healthy women, and outputs in the range 0.4-0.6 were defined as undecided. The quality and generalizability of the outcome model was tested on the remaining 30% of the chromatograms in the population that were not used in the training process. The outcome of this generalization was very good as can be seen in Appendix Table A1.

Validation analysis of the ANN based model obtained was carried out using two different procedures, both suitable for small statistical ensembles [33]: the 7-fold variation and the leave one out (LOO) validations. The validation results, using both validation methods, of the ANN based model are presented in Table 1. Inspection of the results clearly suggests that the ANN based model of the GC-MS results are highly accurate. This accuracy is related to the non-linear correlations between the dependent and independent variables.

### Potential biomarkers for CIN I stage of cervical dysplasia

The ANN based model allowed us to identify VOCs in urine samples that may serve as biomarkers for CIN I stage cervical dysplasia. As discussed above, the validation of the model showed that the ANN yields very good results. In the following we shall focus on discussion of the potential biomarkers suggested by the ANN analysis.



**Figure 1:** GC-MS chromatograms results of urine samples from a typical sick woman (bottom-solid line) and a healthy woman (top-round dots).

Data source	Validation type	Group type	Model classification T/S/H/U*	Sensitivity/PPV**	Specificity/NPV***
GC-MS	7-fold	Sick	17/17/0/0	1.00/1.00	0.89/1.00
		Healthy	9/0/8/1		
		Total	26/17/8/1		
	LOO	Sick	17/14/3/0	0.82/0.82	0.89/1.00
		Healthy	9/0/8/1		
		Total	26/14/11/1		
		Healthy	9/2/6/1		
		Total	26/14/8/4		

**Table 1:** Summary of the validation analysis of the ANN models using two different validation methods. T: total members in the group; S: members in sick group; H: members of healthy group; U: Undecided by model; PPV: Positive Predictive Value; NPV: Negative Predictive Value.

The chemical identities of the potential biomarkers obtained by the ANN analysis are listed in Table 2. The peak identification shown in Table 2 was assigned based on the search results of the GC-MS software, ChemStation, using the NIST'08 libraries. Chemical identification was assigned only if the Quality Factor (QF) obtained

by the search code was above 80. In some cases few assignments with QF>80 were obtained, so in these cases all the possibilities are given. For each possible biomarker the QF is given together with molecular structure and percent of repetition in the healthy and sick groups. For each chemically identified potential biomarker we also added possible

RT (min)	Biomarker	Possible origin	QF	% Sick (n=17)	% Healthy (n=9)	Molecule Structure	Known as biomarker for
3.72	3-Hexanone	Dietary [36], Lipid Peroxidation [37]	80	41%	0%		-
3.92	Hexanal	Systemic, dietary [36], Lipid Peroxidation [38]	87	47%	0%		Lung cancer [39], Liver Cancer [40]
6.88	3-Ethylcyclopentanone	Pheromone [41]	83	35.5%	0%		Age and may be related to oxidative stress [42]
7.84	Ethanol, 2-(2-ethoxyethoxy)	Environmental	83	94%	44.5%		-
8.08	1-Propene, 1-(ethylthio)-2-methyl-	Environmental – smoking [43]	87	35.5%	11%		-
8.28	Benzeneacetaldehyde	Environmental	90	70.5%	55.5%		Antioxidant [44]
8.96	Unknown	-	-	76.5%	78%	-	-
9.28	Urea	endogenous metabolites	80	94%	89%		-
9.68	Urea		86	88%	55.5%		-
10.52	Dodecane, 4-methyl-	Probably oxidative stress [45]	89	76.5%	78%		breast cancer [46], Tuberculosis [47] Melanoma [48] and Lung cancer [49]
10.68	3,4-Dimethylbenzaldehyde	Environmental –smoking and pollution [50-52]	93	88%	33.5%		Carcinogen [54]
11.32	Naphthalene, 2-methyl- Or Naphthalene, 1-methyl-	Environmental –smoking [53]	93	47%	44.5%		
11.84	2-Methoxy-4-vinylphenol	Dietary – smoked food [55]	95	82.5%	66.5%		Antioxidant [56]
12.12	Piperitenone	Dietary – mint species [57]	93	47%	11%		Antioxidant [58]
12.28	Toluene-2,4-diisocyanate	Environmental	94	64.5%	33.5%		Carcinogen [59]
12.92	5-Methyl-1,3-dihydro-2H-benzimidazol-2-one	Drug	91	94%	100%		
12.96	Unknown	-	-	94%	100%	-	-
16.96	2-Pentadecanone, 6,10,14-trimethyl-	Dietary Environmental – Smoking [60]	87	59%	22%		Antioxidant
17.56	Lidocaine	Drug	91	70.5%	0%		
19.92	Unknown	-	-	17.5%	0%	-	-
20.12	Bisphenol A	Dietary and Environmental	93	35.5%	22%		Endocrine disruptor [61]
24.00	Unknown	-	-	35.5%	22%	-	-

**Table 2:** Summary of the chemical identity of potential biomarkers obtained using the ANN approach. The chemical identity of a peak is assigned only if the Quality Factor (QF) was greater than 80, otherwise the peak is assigned as unknown. For some peaks in the chromatogram a few possible assignments were possible, in such case all possibilities with QF>80 are shown. There are compounds which occur in both sick and healthy women samples; however their concentration in urine of sick women is 2-3 folds higher.

routes of its generation or entrance to the body and its identification as biomarker in different types of illnesses according to the literature. Inspection of the data in Table 2 shows that the main sources for most of the potential biomarkers are: dietary, peroxidation processes and environmental sources. Bold fonts were used to mark all the names of potential biomarkers that are produced in the body (Table 2). There are four potential biomarkers that are not identified as originating from environmental or dietary sources. The four potential biomarkers that were identified are: 3-Hexanone, Hexanal, Dodecane, 4-methyl- and 3-Ethylcyclopentanone. The physiological production of these potential biomarkers is associated with processes related to oxidative

stress. Three of these compounds were not identified in the urine sample of any member of the healthy women group while the fourth appeared in both groups but their concentration in the urine of sick women was over two fold higher than in samples from healthy women. This group of four compounds have high probability to serve as reliable biomarkers for CIN I stage cervical dysplasia.

A similar set of compounds that characterizes the healthy women group was also identified. The VOCs identified, using the ANN based approach, are presented in Table 3. Two compounds: 2,7-Dimethyloxepin and 1-Butene, 4-isothiocyanato- were detected as


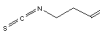
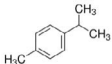


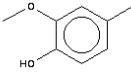
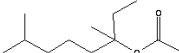
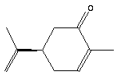
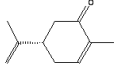
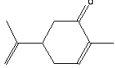





RT (min.)	Name	Molecule Structure	Possible Origin	QF	% in Sick (N=17)	% in Healthy (N=9)
6.36	2,7-Dimethyloxepin		Metabolite of o-Xylene [62] De-aromatization of o-Xylene by P450BM3 (CYP102A1).	96	0	78
7.2	1-Butene, 4-isothiocyanato-		-	87	0	33
7.44	Unknown		-		100	89
7.88	p-Cymene		Food origin [63]	95	18	67
8	D-Limonene		Scent ingredient in perfumes, food and detergents. Cancer protection [64]	96	71	89
8.64	Unknown				17	56
9	β-Linalool		flavor and fragrance compound	96	47	33
10.32	Creosol		Environment (smoking) [65]	96	59	89
10.64	3-Octanol, 3,7-dimethyl-, acetate		-	80	12	22
11	(-)-Carvone		Mint scent	97	82	78
	D-(+)-Carvone		Kummel scent	96		
	Carvone		-	93		
13.28	Unknown		-		88	100
13.36	Butylated Hydroxytoluene		flavor and fragrance compound	91	59	56
13.44	2,5-Cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-		-	90	29	89
15.32	Cyclopentaneacetic acid, 3-oxo-2-pentyl-, methyl ester		Jasmine scent ingredient	91	82	89
22.16	Tetracosane		-	97	18	11
	Octadecane		-	95		

Table 3: List of potential biomarkers that contributed to sample identification of the healthy women according to ANN analysis.

indicative of healthy women but were not identified in any sample of sick women. All other chemicals listed in the two tables appeared in samples of both sick and healthy women with similar probability. This may suggest that the absence of these two compounds in the urine can also serve as potential biomarkers for CIN I stage of cervical cancer if both are absent while the four potential biomarkers described above appear in the urine.

## Conclusions

Potential biomarkers in the urine of stage CIN I of cervical dysplasia were identified. Urine sample analysis was carried out using GC-MS. The data obtained were analyzed using an ANN based statistical approach. The resultant models yield very good separation of the data into two groups: healthy and sick women. The accuracy of the model was examined by two validation methods, both appropriate for examining small statistical groups of data. The validation clearly shows that the ANN-based model is a highly reliable one.

Implementation of the ANN method to the GC-MS data was carefully analyzed. Most of the potential biomarkers in urine samples of sick women that were identified by the ANN method are of environmental or dietary origins. However, the ANN analysis identified four potential biomarkers that are produced by the body. These compounds constitute a sub-set of urine related VOCs, all produced in the body in oxidative stress related processes, have high probability to serve as biomarkers for the CIN I stage of cervical dysplasia. These compounds are (see Table 2 for references):

1. **3-Hexanone**: a product of the lipid peroxidation processes in the body. It has been detected, but not quantified, as a breast cancer biomarker in urine.
2. **Hexanal**: an alkyl aldehyde found in human biofluids including milk samples. It is a mediator of oxidative stress. Hexanal is a volatile compound that has been associated with the development of undesirable flavors. The content of hexanal, which is a major breakdown product of linoleic acid (LA, n - 6 PUFA) oxidation, has been used to follow the course of lipid oxidation. It is a product of the lipid peroxidation process in the body and can be found normally in urine and in cerebrospinal fluid. Abnormal concentrations can be found in urine, blood and exhale breath in lung, liver and breast cancers.
3. **Dodecane, 4-methyl**: This compound has been reported as a biomarker of tuberculosis in a number of publications.
4. **3-Ethylcyclopentanone**: This compound has been reported as a VOC found in the urine whose concentration increases in case of oxidative stress.

In addition, two chemicals that were identified as indicative of urine samples of healthy women but were absent in all samples of the sick women. These compounds are: 2,7-Dimethyloxepin and 1-Butene, 4-isothiocyanato- and their absence in a urine sample of sick women may be considered as a potential biomarker for CIN I cervical dysplasia provided that the four potential biomarkers listed above were identified in the urine. Most of the potential biomarkers identified here for CIN I stage cervical dysplasia have been observed, separately, in previous investigations of different cancers. However, when identified as a group this set of VOCs, together with the two VOCs found only in urine of healthy women but not in samples of sick women, are suggested to be highly suitable to serve as potential biomarkers in the identification of CIN I cervical dysplasia. This set of biomarkers can be used as a simple non-invasive and unexpansive screening procedure to

identify CIN I stage of cervical dysplasia. This constitute an important simple, non-invasive and unexpansive method that compliments the existing cervical dysplasia identification procedure used at present. A last point to be noted is that none of the potential biomarkers identified in the present study were observed in urine samples of smokers or in the urine of individuals whom used to smoke in the past [34,35]. The main drawback of the present study is the small number of urine samples examined for both sick and healthy women. This limitation, as stated in the introduction section, is mainly due to the lack of appropriate financial support for the present investigation. However, the very good results obtained in the validation of the described results suggest that they do have a high potential to become extremely useful. The reliability of these biomarkers has to be further proven in more extensive studies (with larger number of urine samples from sick and healthy women) that are planned to be performed in the future.

## References

1. Akiyama H, Tsurumaru M, Udagawa H, Kajiyama Y (1997) Esophageal cancer. *Curr Probl Surg* 34: 765-834.
2. Lemiere C, Peretti-Viton P, Thomas P, Gaubert JY, di Stefano-Louineau D, et al. (1995) Spreading evaluation in primitive bronchogenic carcinoma: benefit of cerebral MRI compared to CT scan. *Eur J Cancer* 31A: 1715.
3. Bhunchet E, Hatakawa H, Sakai Y, Shibata T (2002) Fluorescein electronic endoscopy: A novel method for detection of early stage gastric cancer not evident to routine endoscopy. *Gastrointest Endosc* 55: 562-571.
4. Kopans DB (1999) Breast-cancer screening with ultrasonography. *Lancet* 354: 2096-2097.
5. Cornu JN, Cancel-Tassin G, Ondet V, Girardet C, Cussenot O (2011) Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *Eur Urol* 59: 197-201.
6. Hanai Y, Shimono K, Matsumura K, Vachani A, Albelda S, et al. (2012) Urinary volatile compounds as biomarkers for lung cancer. *Biosci Biotechnol Biochem* 76: 679-684.
7. Banday KM, Pasikanti KK, Chan EC, Singla R, Rao KV, et al. (2011) Use of urine volatile organic compounds to discriminate tuberculosis patients from healthy subjects. *Anal Chem* 83: 5526-5534.
8. Janasik B, Jakubowski M, Wesołowski W, Kucharska M (2010) Unmetabolized VOCs in urine as biomarkers of low level occupational exposure. *Int J Occup Med Environ Health* 23: 21-26.
9. Inglik N, Rudenko BA, Kakhnovskii IM, Koroleva TV (1989) [Gas chromatographic study of the composition of the volatile components of the urine in normal subjects and patients with diabetes mellitus]. *Lab Delo* 24-7.
10. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, et al. (2010) Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* 49: 5426-5445.
11. (2015) American Cancer Society. *Global Cancer Facts & Figures 3rd Edition*, Atlanta.
12. Lillo FB, Lodini S, Ferrari D, Stayton C, Taccagni G, et al. (2005) Determination of human papillomavirus (HPV) load and type in high-grade cervical lesions surgically resected from HIV-infected women during follow-up of HPV infection. *Clin Infect Dis* 40: 451-457.
13. Bekkers RL, Massuger LF, Bulten J, Melchers WJ (2004) Epidemiological and clinical aspects of human papillomavirus detection in the prevention of cervical cancer. *Rev Med Virol* 14: 95-105.
14. Almog B, Gamzu R, Kuperminc MJ, Levin I, Fainaru O, et al. (2003) Human papilloma virus testing in patient follow-up post cone biopsy due to high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 88: 345-350.
15. Ronco G, Segnan N, Giorgi-Rossi P, Zappa M, Casadei GP, et al. (2006) Human papillomavirus testing and liquid-based cytology: results at recruitment from the new technologies for cervical cancer randomized controlled trial. *J Natl Cancer Inst* 98: 765-74.
16. Parkin DM, Almonte M, Bruni L, Clifford G, Curado M-P, et al. (2008) Burden and trends of type-specific human papillomavirus infections and related diseases in the latin america and Caribbean region. *Vaccine*. 26 Suppl 1: L1-15.

17. Mathew A, George PS (2009) Trends in incidence and mortality rates of squamous cell carcinoma and adenocarcinoma of cervix--worldwide. *Asian Pac J Cancer Prev* 10: 645-650.
18. Richart RM (1969) A theory of cervical carcinogenesis. *Obstet Gynecol Surv* 24: 874-879.
19. Bradshaw KD, Schorge JO, Schaffer J, Halvorson LM, Hoffman BG, eds. (2008) Cervical Intraepithelial Neoplasia, in: Williams' Gynecol., McGraw-Hill Professional.
20. van Ballegooijen M, van den Akker-van Marle ME, Warmerdam PG, Meijer CJ, Walboomers JM, et al. (1997) Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-) effectiveness. *Br J Cancer* 76: 651-657.
21. Goldie SJ, Gaffikin L, Goldhaber-Fiebert JD, Gordillo-Tobar A, Levin C, et al. (2005) Cost-effectiveness of cervical-cancer screening in five developing countries. *N Engl J Med* 353: 2158-2168.
22. Lisboa PJ, Taktak AF (2006) The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw* 19: 408-415.
23. Boger Z (2007) Experience in the applications of artificial neural networks in bio-informatics, In: Meller J, Nowak W (eds.), *Appl. Stat. Mach. Learn. Methods Bioinforma.*, Peter Lang - International Academic Publishers, Pieterlen, Switzerland.
24. Boger Z, Karpas Z (1994) Application of neural networks for interpretation of ion mobility and x-ray fluorescence spectra. *Anal Chim Acta* 292: 243-251.
25. Lancashire LJ, Lemetre C, Ball GR (2009) An introduction to artificial neural networks in bioinformatics--application to complex microarray and mass spectrometry datasets in cancer studies. *Brief Bioinform* 10: 315-329.
26. Lasch P, Beyer W, Nattermann H, Stämmler M, Siegbrecht E, et al. (2009) Identification of Bacillus anthracis by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Appl Environ Microbiol* 75: 7229-42.
27. Xu Y, Selaru FM, Yin J, Zou TT, Shustova V, et al. (2002) Artificial Neural Networks and Gene Filtering Distinguish Between Global Gene Expression Profiles of Barrett's Esophagus and Esophageal Cancer. *Cancer Res* 62: 3493-3497
28. Boger Z (2003) Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Anal Chim Acta* 490: 31-40.
29. Baba K, Enbutu I, Yoda M (1990) Explicit representation of knowledge acquired from plant historical data using neural network, In: 1990 IJCNN Int. Jt. Conf. Neural Networks, IEEE: pp. 155-160 vol.3.
30. Rogers WH, Wagner TJ (1978) A Finite Sample Distribution-Free Performance Bound for Local Discrimination Rules. *Ann. Stat.* 6: 506-514. doi:10.1214/aos/1176344196.
31. Shao J (1993) Linear Model Selection by Cross-validation. *J Am Stat Assoc* 88: 486-494.
32. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection, In: IJCAI, pp. 1137-1145.
33. Molinaro AM, Simon R, Pfeiffer RM (2005) Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 21: 3301-3307.
34. Yuan JM, Butler LM, Stepanov I, Hecht SS (2014) Urinary tobacco smoke-constituent biomarkers for assessing risk of lung cancer. *Cancer Res* 74: 401-411.
35. Carmella SG, Chen M, Han S, Briggs A, Jensen J, et al. (2009) Effects of smoking cessation on eight urinary tobacco carcinogen and toxicant biomarkers. *Chem Res Toxicol* 22: 734-741.
36. Mochalski P, Krapf K, Ager C, Wiesenhofer H, Agapiou A, et al. (2012) Temporal profiling of human urine VOCs and its potential role under the ruins of collapsed buildings. *Toxicol Mech Methods* 22: 502-511.
37. Yancey M, Stuart R, Wiesler D, Novotny M (1986) Quantitative alterations in the metabolism of carbonyl compounds due to diet-induced lipid peroxidation in rats. *J Chromatogr* 382: 47-56.
38. Kappus H, Sies H (1981) Toxic drug effects associated with oxygen metabolism: redox cycling and lipid peroxidation. *Experientia* 37: 1233-1241.
39. Deng C, Zhang X, Li N (2004) Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 808: 269-277.
40. Xue R, Dong L, Zhang S, Deng C, Liu T, et al. (2008) Investigation of volatile biomarkers in liver cancer blood using solid-phase microextraction and gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 22: 1181-1186.
41. Cassidy A, Hanley B, Lamuela-Raventos RM (2000) Isoflavones, lignans and stilbenes - origins, metabolism and potential importance to human health. *J Sci Food Agric* 80: 1044-1062.
42. Miller KV, Jemiolo B, Gasset J, Jelinek I, Wiesler D, Novotny M (1998) Putative Chemical Signals from White-Tailed Deer (*Odocoileus virginianus*): Social and Seasonal Effects on Urinary Volatile Excretion in Males. *J Chem Ecol* 24: 673-683.
43. Kwak J, Grigsby CC, Smith BR, Rizki MM, Preti G (2013) Changes in volatile compounds of human urine as it ages: their interaction with water. *J Chromatogr B Analyt Technol Biomed Life Sci* 941: 50-53.
44. Buszewski B, Ulanowska A, Ligor T, Denderz N, Amann A (2009) Analysis of exhaled breath from smokers, passive smokers and non-smokers by solid-phase microextraction gas chromatography/mass spectrometry. *Biomed Chromatogr* 23: 551-556.
45. Phillips M, Cataneo RN, Greenberg J, Gunawardena R, Naidu A, et al. (2000) 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.
46. Krkosová Z, Kubinec R, Soják L, Amann A (2008) Temperature-programmed gas chromatography linear retention indices of all C4-C30 monomethylalkanes on methylsilicone OV-1 stationary phase. Contribution towards a better understanding of volatile organic compounds in exhaled breath. *J. Chromatogr. A* 1179: 59-68.
47. Phillips M, Cataneo RN, Condos R, Ring Erickson GA, Greenberg J, et al. (2007) Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis (Edinb)* 87: 44-52.
48. Abaffy T, Duncan R, Riemer DD, Tietje O, Elgart G, et al. (2010) Differential volatile signatures from skin, naevi and melanoma: a novel approach to detect a pathological process. *PLoS One* 5: e13813.
49. Phillips M, Altorki N, Austin JH, Cameron RB, Cataneo RN, et al. (2007) Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark* 3: 95-109.
50. Meyer A, Panke S (2008) Genomics in Metabolic Engineering and Biocatalytic Applications of the Pollutant Degradation Machinery, In: Diaz E (ed.), *Microb. Biodegrad. Genomics Mol. Biol.*, Caister Academic Press, Norfolk, UK: pp. 359-388.
51. Thelestam M, Curvall M, Enzell CR (1980) Effect of tobacco smoke compounds on the plasma membrane of cultured human lung fibroblasts. *Toxicology* 15: 203-217.
52. Grosjean D, Grosjean E, Gertler AW (2001) On-road emissions of carbonyls from light-duty and heavy-duty vehicles. *Environ Sci Technol* 35: 45-53.
53. Schmeltz I, Tosk J, Hoffmann D (1976) Formation and determination of naphthalenes in cigarette smoke. *Anal Chem* 48: 645-650.
54. Murata Y, Denda A, Maruyama H, Nakae D, Tsutsumi M, et al. (1997) Chronic toxicity and carcinogenicity studies of 2-methylnaphthalene in B6C3F1 mice. *Fundam Appl Toxicol* 36: 90-93.
55. Gilbert J, Knowles ME (2007) The chemistry of smoked foods: a review. *Int J Food Sci Technol* 10: 245-261.
56. Fukai S, Tanimoto S, Maeda A, Fukuda H, Okada Y, et al. (2009) Pharmacological activity of compounds extracted from persimmon peel (*Diospyros kaki* THUNB.). *J Oleo Sci* 58: 213-219.
57. Zhou S, Koh HL, Gao Y, Gong ZY, Lee EJ (2004) Herbal bioactivation: the good, the bad and the ugly. *Life Sci* 74: 935-968.
58. Nakamura Y, Hasegawa Y, Shiota K, Suetome N, Nakamura T, et al. (2014) Differentiation-inducing effect of piperitenone oxide, a fragrant ingredient of spearmint (*Mentha spicata*), but not carvone and menthol, against human colon cancer cells. *J Funct Foods* 8: 62-67.
59. Dieter MP, Boorman GA, Jameson CW, Matthews HB, Huff JE (1990) The carcinogenic activity of commercial grade toluene diisocyanate in rats and mice in relation to the metabolism of the 2,4- and 2,6-TDI isomers. *Toxicol Ind Health* 6: 599-621.

- 
60. Liu L, Huang Y, Wang J, Tang Z, Lu L, et al. (2013) Study on Discriminating Flue-Cured Tobacco by Volatile Compounds Related to Geographical Origin and Cultivar. *Asian J Chem* 25: 7587-7592.
61. Völkel W, Colnot T, Csanády GA, Filser JG, Dekant W (2002) Metabolism and Kinetics of Bisphenol A in Humans at Low Doses Following Oral Administration. *Chem Res Toxicol* 15: 1281-1287.
62. Whitehouse CJ, Rees NH, Bell SG, Wong LL (2011) Dearomatisation of o-xylene by P450BM3 (CYP102A1). *Chemistry* 17: 6862-6868.
63. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, et al. (2009) Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC Cancer* 9: 348.
64. Steinmetz KA, Potter JD (1996) Vegetables, fruit, and cancer prevention: a review. *J Am Diet Assoc* 96: 1027-1039.
65. Nise G (1992) Urinary excretion of o-cresol and hippuric acid after toluene exposure in rotogravure printing. *Int Arch Occup Environ Health* 63: 377-381.