Volatile Organic Compounds (VOCs) in the exhaled breath as biomarkers for the early detection of lung cancer: application of complementary methodological approaches

Palmisani J.[1], Di Gilio A.[1], Pizzillo V.[1], Tosoni E.[2], Nisi M.R.[1], de Gennaro G.[1], Perbellini L.[3**], Santo A.[2**]*

[1]*Department of Biosciences, Biotechnologies and Environment, University of Bari Aldo Moro, 70125 Bari - Italy*
[2] *Lung Unit, P. Pederzoli Hospital, 37019 Peschiera del Garda, Verona, Italy*
[3] *University of Verona, 37129 Verona, Italy*

**corresponding author: jolanda.palmisani@uniba.it*
***co-last authors*

RATIONALE AND OBJECTIVES OF THE STUDY

Lung cancer (LC) is one of the most aggressive tumors and the leading cause of cancer-related death worldwide. The diagnosis of the disease generally occurs at an advanced stage, making radical surgical treatment possible in only less than 20% of cases. Thus, there is an ever-increasing need to equip National Health Systems with a reliable diagnostic tool alternative to traditional diagnostic exams addressed to large-scale screening programs. Chemical characterization of Volatile Organic Compounds (VOCs) in human breath and the resulting identification of a disease-related biomarkers has been recognized as a promising approach for the early detection and follow-up of oncologic diseases such as LC [1]. To identify a VOCs pattern in human breath as biomarkers of LC exploring potentialities of breath analysis with multiple methodological approaches, a prospective-observational study based on the application of complementary methodological approaches and analytical techniques was carried out.

METHODOLOGY

An overall number of **110** individuals were recruited at the Lung Unit of 'P. Pederzoli' Hospital in Verona (Italy): more specifically, **55** patients affected by LC (n.45 adenocarcinoma, n.8 squamous cell carcinoma, n.2 lung microcytoma) and **55** healthy controls. The enrollment of volunteers in the study fulfilled specific inclusion criteria and the study protocol was approved by the Italian Ethical Committee (Prot. n. 45355). The sampling procedure applied in the study was based on two different methodological approaches: **A**- end-tidal breath collection directly onto two-beds adsorbent cartridges (*Biomonitoring steel tubes, Markes International ltd*) by means of the automated sampler Mistral (*Predict srl, Italy*); **B**- end-tidal breath collection by means of BioVOC® sampler (*Markes International ltd*). Ambient air samples (AA) were simultaneously collected at each sampling session with both the methodological approaches.

End-tidal breath samples as long as AA samples collected with approach **A** were analysed by thermal desorption (*Unity Ultra-xr, Markes International ltd*) and Gas Chromatography/Mass Spectrometry (*GC Agilent 7890/MS Agilent 5975*) at University of Bari (Italy) while end-tidal breath samples and AA samples collected by approach **B** were analysed by Ion Molecular Reaction-Mass Spectrometry (*AirSense Compact analyzer, V&F Analyse- und Messtechnik*) at University of Verona (Italy).

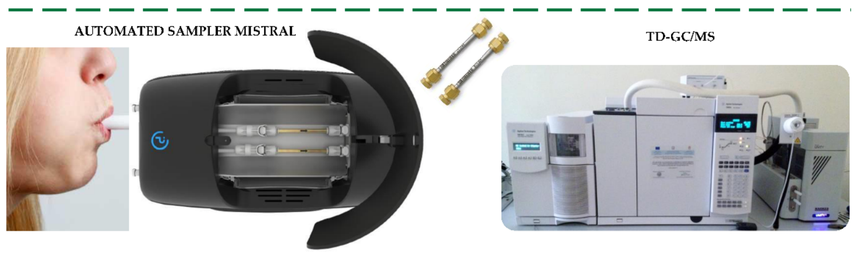
Experimental data were statistically processed by non-parametric Wilcoxon-Kruskal Wallis tests (software R version 3.5.1) in order to identify the most weighting variables in the discrimination between LC and HC breath samples. Linear Discriminant Analysis (LDA) was applied to the dataset to validate breath analysis-based methodology in the discrimination among LC and HC subjects.

Table 1: Study population features.

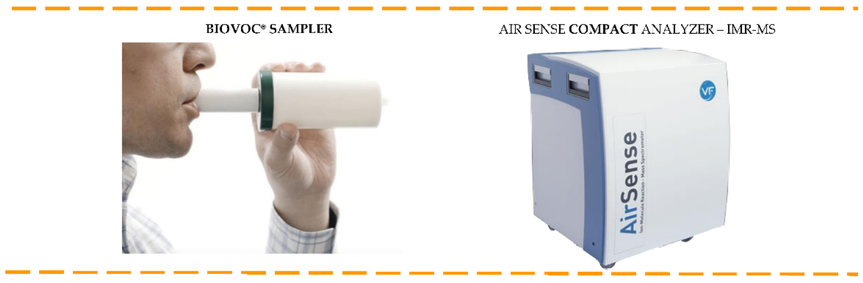
	CTRL (n=55)	**LC** (n=55)
Age (years)	63 (38-83)*	69 (49-85)*
Sex ratio (M:F)	21:34	35:20
BMI	25 (17-42)*	25 (18-32)*
Tobacco smoking	6	8

* Data herein reported are expressed as median and ranges.

METHODOLOGY A



METHODOLOGY B



RESULTS AND DISCUSSION

Non parametric statistical treatment of TD-GC/MS data allowed to identify a selected VOCs pattern (on the overall number of compounds identified or tentatively identified, n.101) showing abundances in end-tidal breath samples of patients with LC diagnosis significantly different from those of CTRLs (variables with significant level of p-values < 0.05, Table 2): pentadecane, isoprene, 2-butoxyethanol, dimethylnonane and a benzene derivative characterized by m/z 118 (not uniquely identified in the present study).

The application of IMR-MS methodology highlighted, in addition, the potential discriminant capability of dimethyl sulfide and dimethyl disulfide between CTRL and LC groups (Table 2). The effect of ambient air contamination (AA samples) was excluded in both cases.

Based on the most weighting variables identified (p-values < 0.05), multivariate statistical analysis e.g., Linear Discriminant Analysis was applied to the dataset providing clustering between CTRL and LC groups (as shown in Figure 1). The two discriminant functions LD1 and LD2 accounted for the 70% and 29% of the total data variance, respectively. LD1 was mainly loaded by *dimethylnonane*, while LD2 by *isoprene*, *2-buthoxyethanol*, *m/z 118* and *pentadecane*.

The predictive model developed on the selected features with a leave-one-out validation approach enabled to discriminate between CTRLs and LC patients with a diagnostic accuracy of 99%, as highlighted by the Receiver Operating Characteristic (ROC) curve (AUC: 0.99, Figure 2).

Table 2: Wilcoxon-Kruskal Wallis tests outcomes: p-values.

	Compound	**p-value**
TD-GC/MS	isoprene	0.04
	2-butoxyethanol	0.02
	m/z 118 (benzene derivative)	$2.0 \cdot 10^{-16}$
	dimethylnonane	$1.5 \cdot 10^{-10}$
	pentadecane	0.009
IMR-MS	dimethyl sulfide	0.02
	dimethyl disulfide	0.02

The discriminant variables identified in this study as biomarkers for LC are in line with scientific literature. A relevant number of studies underlined the occurrence of higher concentration of long chain alkanes (C_{13}-C_{15}), mono- and dimethyl branched alkanes and selected alcohols in the breath of LC patients with respect to CTRLs, due to improved lipid peroxidation in tumor cells [2]. On the contrary, the role of dimethylsulfide and dimethyl disulfide in LC cell metabolism has not been elucidated yet. These compounds have been cited, however, in previous studies on LC and can be considered worthy of future in-depth studies [3,4].

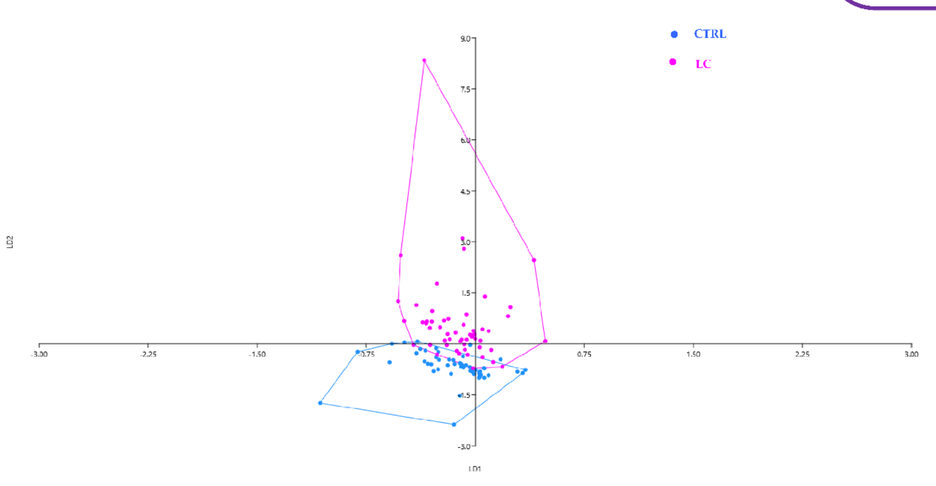


Figure 1: Linear Discriminant Analysis (LDA) plot.

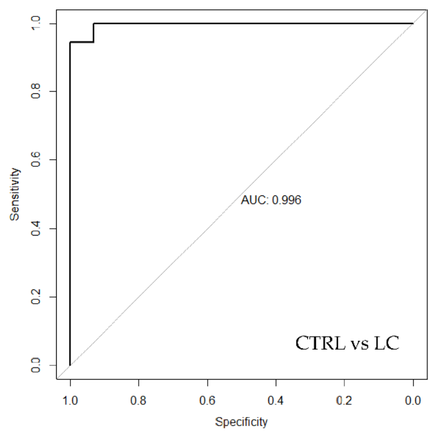


Figure 2: ROC curve.

CONCLUSIONS

The combination of non-parametric tests with a supervised classification method enabled to detect metabolites in human breath with discriminant power between healthy subjects and LC patients. An investigation approach based on the simultaneous application of different analytical techniques may reveal strategic to emphasize the chemical information in complex matrices as human breath, stimulating the debate on 'new compounds' and guiding the future research.

REFERENCES

[1] Phillips et al., *Cancer Biomarkers* 2007; 3(2):95–109.
[2] Poli et al., *Respiratory Research* 2005, 6, 71.
[3] Rudnicka et al., *J. Chromatogr. B* 2011, 879, 3360–3366.
[4] Kischkel et al., Clin. Chim. Acta 2010, 411, 1637-1644.