# On-line breath analysis of volatile organic compounds as a method for colorectal cancer detection

Francisco Zambrana[1], Ana Herrero[2], Guillermo Vidal-de-Miguel[2], Gonzalo Bailador[3], Ernesto Criado[2], Carmen Sánchez-Ávila[3], Isabel Marquina[6], María Sereno[1], César Gómez-Raposo[1], Miriam López-Gómez[1], María Merino[1], Cristina Aguayo[1], Mario Alvarez-Gallego[4], Damián Garcia-Olmo[4], José C. Erdozain[5], Carmen Comas[5], Noemí Manceñidos[5], Jaime Feliu[7], Enrique Casado[1]

[1]Oncology Department, Hospital Universitario Infanta Sofía, Madrid, Spain. [2]European Society of Differential Mobility Analysis (SEADM), Valladolid, Spain. [3]Universidad Politécnica de Madrid, Spain. [4]Surgery Department, Hospital Universitario La Paz, Madrid, Spain. [5]Gastroenterology Department, Hospital Infanta Sofía, Madrid, Spain. [6]Pathology Department, Hospital Infanta Sofía, Madrid, Spain, Oncology Department, [7]Hospital Universitario La Paz, Madrid, Spain.

## Abstract

**Background:** Analysis of exhaled volatile organic compounds (VOCs) in breath is an emerging approach for cancer diagnosis, but little is known about its potential use as a biomarker for colorectal cancer (CRC). We investigated whether a combination of VOCs could distinct CRC patients from healthy volunteers.
**Methods**: In a pilot study, we prospectively analyzed breath exhalations of 38 CRC patient and 43 healthy controls all scheduled for colonoscopy, older than 50 in the average-risk category. The samples were ionized and analyzed using a Secondary ElectroSpray Ionization (SESI) coupled with a Time-of-Flight Mass Spectrometer (SESI-MS). After a minimum of 2 hours fasting, volunteers deeply exhaled into the system. Each test requires three soft exhalations and takes less than ten minutes. No breath condensate or collection are required and VOCs masses are detected in real time, also allowing for a spirometric profile to be analyzed along with the VOCs. A new sampling system precludes ambient air from entering the system, so background contamination is reduced by an overall factor of ten. Potential confounding variables from the patient or the environment that could interfere with results were analyzed.
**Results**: 255 VOCs, with masses ranging from 30 to 431 Dalton have been identified in the exhaled breath. Using a classification technique based on the ROC curve for each VOC, a set of 9 biomarkers discriminating the presence of CRC from healthy volunteers was obtained, showing an average recognition rate of 81.94%, a sensitivity of 87.04% and specificity of 76.85%.
**Conclusions**: A combination of cualitative and cuantitative analysis of VOCs in the exhaled breath could be a powerful diagnostic tool for average-risk CRC population. These results should be taken with precaution, as many endogenous or exogenous contaminants could interfere as confounding variables. On-line analysis with SESI-MS is less time-consuming and doesn't need sample preparation. We are recruiting in a new pilot study including breath cleaning procedures and spirometric analysis incorporated into the postprocessing algorithms, to better control for confounding variables.

## Background

CRC is one of the most common cancers in the Western world, with high incidence and mortality. Slow progression from adenoma to carcinoma and high patient survival in case of early detection makes CRC an ideal candidate for screening. Screening programms for CRC are being developed throughout the Western World with participation rates under 60% in the average-risk population using the recommended tests (gFoBT or endoscopic techniques).

There are somes evidence supporting a different pattern of VOCs in breath in cancer patients, due to the different metabolism of cancer cells producing different VOCs, that may have expression in breath. Actually, dogs trained for scent discrimination were able to discriminate CRC patients from healthy controls (Sonoda et al. 2010).

Mass spectrometry devices have been devoloped to identify the individual VOCs present in breath samples, but there are many methodological differences between the studies without validated results. Results in lung and breast cancer using this approach are promising (Phillips et al, 1999; Patterson et al 2011), but there is few data about its potential role in CRC diagnosis.

Along with this, breath analysis of VOCs could be an easy and very acceptable test to increase the efficiency of a population-based screening program.

## Objectives

To determine whether a combination of VOCs could distinct CRC patients from healthy volunteers. It was our purpose to provide a framework for a new pilot study in this area.

## Methods

**Demographics**: Prospectively enrolled 38 CRC patients (all AJCC stages) and 43 healthy volunteers, diagnosed with colonoscopy.

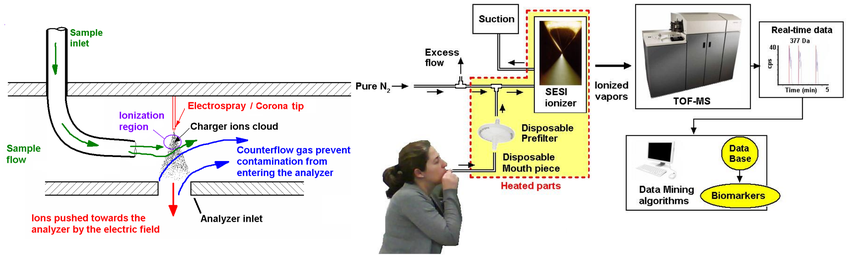
**Confounding variables**: potential interferents were recorded before sampling (comorbidities, smoking, food, medications, cosmetics, enviromental exposure, family history and other CRC risk factors) using a questionnaire.

**Breath sample**: after 2 hours fasting volunteers exhaled into the system directly, without collection of breath in any recipient.

**Mass spectrometry system**: Secondary ElectroSpray Ionization (SESI) coupled with a Time-of-Flight Mass Spectrometer (SESI-MS).

**1.Ion Source: SESI (Secondary ElectroSpray Ionizer)** is a soft ionization technique at atmospheric pressure (API) where a electrospray cloud produces OH-ions that ionize polar molecules.

Introducing humid air and improving fluid configuration reduces required fluid rate and improves ionization efficiency allowing eliminating preconcentration step and on-line analysis (Martínez-Lozano & Fernandez de la Mora, 2007).



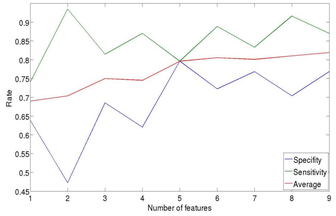

**2.Mass analyzer**: Quadruple time-of-flight mass spectrometer (QqTOF-MS).

**3.Data analysis**: **Features**: Area under the peak from spectrum (AUP). Relative concentrations (Division by the sum of all AUPS). **Preprocessing**: AUP(breath)-AUP(background). **Classifier (for each feature)**: Optimal ROC point. **Fusion several classifiers**: Majority voting. **Feature selection**: Forward selection. **Validation**: Leave-One-Out.

**4.VOCs identification**: The obtained data was compared with the Metlin Database to identify the candidate species using a tolerance of 0.01 Da for the average mass.

## Results

- 255 VOCs, with masses ranging from 30 to 431 Dalton have been identified in the exhaled breath.
- Using a classification technique based on the ROC curve for each VOC, a set of 9 biomarkers discriminating the presence of CRC from healthy volunteers was obtained, showing an average recognition rate of 81.94%, a sensitivity of 87.04% and specificity of 76.85%.




| VOC | Mass interval | | Specificity | Sensitivity | recognition rate | Candidate species |
|---|---|---|---|---|---|---|
| 1 | 149.9775 | 150.0575 | 0.6389 | 0.7407 | 0.6898 | Non identified |
| 2 | 233.0924 | 233.2167 | 0.4722 | 0.9352 | 0.7037 | Valerenic acid |
| 3 | 178.9123 | 179.0030 | 0.6852 | 0.8148 | 0.7500 | Non identified |
| 4 | 99.9526 | 99.9896 | 0.6204 | 0.8704 | 0.7454 | Non identified |
| 5 | 180.9804 | 181.0426 | 0.7963 | 0.7963 | 0.7963 | 2-Oxo-4-phosphonobutanoate |
| 6 | 218.9237 | 218.9916 | 0.7222 | 0.8889 | 0.8056 | DICAMBA 2,4-Dichlorophenoxyacetic acid |
| 7 | 141.0003 | 141.0346 | 0.7685 | 0.8333 | 0.8009 | Kojic acid, Muconic acid |
| 8 | 229.0386 | 229.0971 | 0.7037 | 0.9167 | 0.8102 | Non identified |
| 9 | 182.9143 | 183.0212 | 0.7685 | 0.8704 | 0.8194 | Phosphohydroxypyruvic acid |

- Although these results are quite promising, we have discovered that some of these variables present a temporal evolution which may be produced by contaminants in the environment. Therefore these results should be taken with caution.

## Conclusions

SESI-based analysis shows high sensitivity for the discovery of candidate biomarkers in breath and allows on-line analysis without a pre-concentration step.

We couldn´t find in breath obvious differences in the pattern of VOCs between CRC patients and healthy controls.

Species coming from the inhaled air and those related to the long term distribution seem to be the main confounding variable to control.

Despite these, overall sensitivity and specificity results are encouraging to pursue investigating the potential use of breath analysis for detecting CRC.

We are recruiting in a new pilot study including breath cleaning procedures, random distribution of patients and spirometric analysis incorporated into the postprocessing algorithms, to better control for confounding variables.

## References

1. Phillips M, et al . Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet*. 1999 Jun 5;353(9168):1930-3.
2. Patterson SG, et al. Breath analysis by mass spectrometry: a new tool for breast cancer detection? *Am Surg*. 2011 Jun;77(6):747-51.
3. Sonoda H, et al. Colorectal cancer screening with odour material by canine scent detection. *Gut* 60(6):814-9, 2011.
4. Martínez-Lozano, P. and Fernández de la Mora, J. "Electrospray ionization of volatiles in breath", *Int. J. Mass Spectrometry*, 265 (1): 68-72, 2007.