Identification of volatile signatures for non-invasive cancer detection using Secondary Electrospray Ionization (SESI) – High Resolution Mass Spectrometry and machine learning- based data analysis.

Miceli R*, Locati L*, Segrado F*, Patricola P*, Garrone G*, Cavalleri A*, Mancinelli M*, Granata R*, Agresti R*, Brambilla P°, Milani M#, Licitra L*+ and Orlandi R*

*Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy; °University of Milano-Bicocca, Hospital of Desio, Desio-Milan, Italy; #ASST Cremona, Italy; +University of Milan, Italy

Clinical Studies

Breath analysis offers a prominent potential in cancer diagnosis, due to its non-invasiveness, painlessness, safety, prompt acceptance by patients and the nearly unlimited access to samples.
Prospective studies of biomarker discovery were designed at Istituto Nazionale dei Tumori (INT) to identify cancer-related volatile organic compounds (VOCs) and develop non-invasive diagnostic tools for clinical practice in **head and neck cancer** and **breast cancer.** Exhaled air samples are analysed in **Secondary Electrospray Ionization (SESI) – High Resolution Mass Spectrometry (HR-MS)** with an **untargeted** approach and data are managed following our statistical procedures (P. Martinez-Lozano et al. J Breath Res, 2015). Analysis are performed on patients carrying a **primary tumor** and breath collection always occours **before any clinical treatment**.

Head and neck cancer

Head and neck cancer (HNC) are rare tumors with an incidence of over 500,000 cases annually worldwide. Early detection of **HNC** and identification of **Human Papilloma Virus (HPV) status** are crucial in the patient outcome. HPV is an emerging sexually transmitted infection. Within **HNC**, up to 60% of oropharyngeal carcinomas (OPC) are HPV-related, and the incidence of HPV-OPC is expected to dramatically increase in the next years.

Our objective is the non-invasive detection of HNC with three specific aims:

- Identification of a cancer-related signature for HNC
- Identification of an HPV-related OPC signature. HPV-positive OPC is dramatically diffent from the HPV negative one from both biological and clinical points of view.
- Identification of an HPV-related signature in healthy individuals with oral HPV infection.

The elegible subjects are enrolled in Head and Neck Medical Oncology of Istituto Nazionale dei Tumori (INT) and they come from the whole national territory.

Discovery cohort
- 100 HNC patients (circa 70 HPV positive)
- 70 tumor-free subjects

Test cohort
- 50 HNC patients (circa 35 HPV positive)
- 35 tumor-free subjects

Cohort of 100 healthy subjects
- Gender-related profiles
- Smoking-related profiles
- Food-related profiles

Breast cancer

Breast cancer (BC) is a leading cause of morbidity and mortality among female cancer patients worldwide. Unfortunately, only about 60% of the breast cancers are localized at the time of diagnosis. Non-invasive molecular markers supporting imaging techniques in the specific identification of malignant and aggressive lesions are urgently needed and will provide a real opportunity not only for a personalized BC diagnosis, but also for treatment monitoring and follow-up .

Our objective is the non-invasive detection of BC with two specific aims:

- Development of a volatile cancer signature able to discriminate the breathprints of BC patients and those of subjects with benign disease or tumor-free
- Technical and clinical validation of the final volatile signature in a multicenter cohort of BC patients and controls and in a cohort of high risk women.

The elegible subjects in the discovery cohort are enrolled in Breast Surgery and Breast Imaging of INT, whereas the elegible subjects in the validation cohort are enrolled in 3 hospitals: INT, ASST Monza and ASST Cremona.

Discovery cohort
- 200 BC patients
- 100 benign disease
- 200 healthy controls

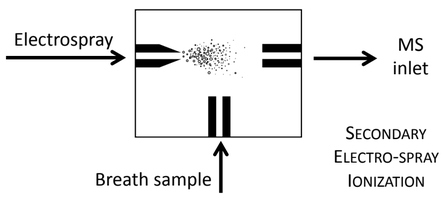
Multicenter validation cohort
- 400 BC including a cohort of BRCA patients
- 150 controls

Breath Collection

Breath samples are collected in the same room for patients and controls using two-ended home made **Nalophan bags** sterilized using Sterrad® method. The online and offline analysis showed a good concordance in terms of Concordance Correlation Coefficient (range is 0.87-0.99). Sample collection, transport to the laboratory and MS analysis strictly adhere to our **Standard Operating Procedures (SOPs)** designed to minimize pre-analytical and analytical variability. Adequate quality controls procedures assess the overall quality of the experimental procedure and its reproducibility.

Mass Spectrometry

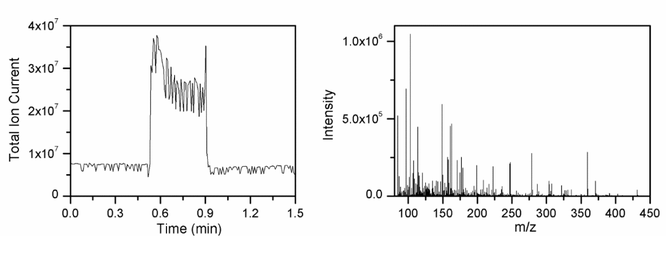
Secondary Electrospray Ionization (SESI) allows real time detection of VOCs in gaseous samples without pretreatment and chromatographic steps. The proton transfer occurs in the ionization chamber where the sample and the ionized electrospray are mixed in vapor phase.



SUPERSESI (FOSSILIONTECH) COUPLED TO THERMO LTQ ORBITRAP ELITE

High resolution mass profiles are obtained with **SESI ion source** coupled with **LTQ Orbitrap Elite**.
According to our analytical SOPs, samples must be analysed within two hours after the collection. Therefore, the overall process involves a strong collaboration between the medical unit and the laboratory team.

The statistical team supervises the quality of data and controls the residual experimental variability by the statistical procedures included in data preprocessing of our pipeline.



Left, Total Ion Current (TIC) vs time. Breath TIC emerges from the background. **Right,** full scan MS profile.

Statistical analysis

DATA EXTRACTION AND CONVERSION
- Data acquisition and extraction
- Peaks alignment
- Generate a [*feature* x *sample*] matrix

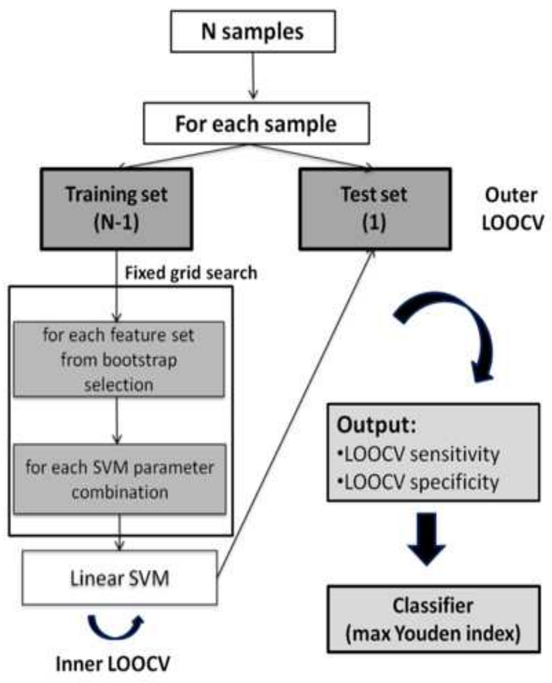
DATA PREPROCESSING

Handling experimental variability
- Quality control in terms of Concordance correlation coefficient (CCC) among breath samples replicates.
- Average of breath samples replicates for each patient
- Data normalization
- Left-censored value imputation
- Batch effect correction

SUPERVISED ANALYSIS

Class Comparison
- Wilcoxon-Mann-Whitney test

Class Prediction (machine learning-based)
- Boostrap selection
- Classifier development



Workflow for class prediction. SVM, support vector machine; LOOCV, leave-one-out cross-validaton.

- The last step of these studies will focus on the characterization of the identified features. m/z values will be linked to molecular structures, possibly with fragmentation experiments. The consistence of the resulting structures will be considered in the context of cancer metabolism.

The HNC project is supported by Bando Ricerca Istituzionale 2016, Direzione Scientifica, Fondazione IRCCS Istituto Nazionale dei Tumori Milano. The BC project is supported by Ricerca Finalizzata 2016, Ministero della Salute.
We thank the Clinic Engeneering Unit for the technical support and Gisella Rivadossi and Nicolino Macchia for the Nalophan bag preparation.

We thank all patients and volunteers