Investigation of various data processing approaches for extracting disease-specific information from GC×GC-(HR)TOFMS breathprint data sets

Romain, Pesesse;[1] Pierre-Hugues, Stefanuto;[1] Florence, Schleich;[2] Renaud, Louis;[2] Jean-François, Focant[1]

[1]Organic and Biological Analytical Chemistry, CART, University of Liège, Belgium
[2]Pneumology and Allergology Unit, University Hospital Center, Liège, Belgium

Abstract

Breath analyses for medical applications (screening, biomarker discovery, biological process description, …) have been gaining more and more attention over the last couple of decades as evidences about links between volatile organic compound (VOC) breath signatures and specific diseases have been accumulating.

Comprehensive two-dimensional gas chromatography GC×GC coupled to time-of-flight mass spectrometry (TOFMS) is a tool of choice to accurately describe complex VOC profiles of exhaled air. The use of high resolution/high accuracy (HR)TOFMS even further enhance proper analyte identification when putative markers of diseases can be isolated from cohort studies. A major point of issue however remains sample integrity as several external factors (e.g. ambient air, time of sampling, mode of sampling, …) can significantly impact the composition of the air patients exhale.

In this study, we collected breath samples using Teddlar™ bags from patients suffering from lung cancer (n=15) as well as from healthy controls (n=15) over a period of 5 months. TD-GC×GC-TOFMS analyses were carried out to highlight prominent analytes in both classes of samples. Several univariate and multivariate data treatment approaches were investigated (e.g. Fisher ratio, dispersion boxes, principal component analysis and clustering) to reduce the weight of external confounding factors to a level that did not impact cohort differentiations anymore. These data mining approaches were further challenged by testing them on additional cohorts. Selected samples were analyzed by GC×GC-HRTOFMS to enhance our confidence in the identification of putative biomarkers.