University of Liège

Faculty of Sciences

Department of Chemistry

Organic and Biological Analytical Chemistry Group

Prof. J.-F. Focant

Contribution of comprehensive two-dimensional gas chromatography to untargeted volatilomics of lung cancer

Dissertation submitted by
Romain PESESSE

To obtain the degree of
Doctor of philosophy in Sciences
Academic year 2018-2019

University of Liège

Faculty of Sciences

Department of Chemistry

Organic and Biological Analytical Chemistry Group

Prof. J.-F. Focant

Contribution of comprehensive two-dimensional gas chromatography to untargeted volatilomics of lung cancer

Dissertation submitted by
Romain PESESSE

To obtain the degree of
Doctor of philosophy in Sciences
Academic year 2018-2019

Thesis prepared at the OBIAchem laboratory, University of Liège, Belgium from October 2013 to September 2017

Thesis submitted at the Faculty of science of the University of Liège, Belgium
April 14 2019

Thesis defended in public at the University of Liège, Belgium
September 10 2019

Jury members:

Supervisor: Prof. J.-F. Focant (ULiège, Belgium)

President: Prof. L. Quinton (ULiège, Belgium)

Secretary: Prof. J.-F. Focant (ULiège, Belgium)

Members: Dr. V. Bertrand (ULiège, Belgium)

Dr. C. Brasseur (Certech, Belgium)

Prof. P. Cardinael (University of Rouen, France)

Dr. G. Purcaro (ULiège, Belgium)

Dr. F. Schleich (ULiège- CHU, Belgium)

“The answer, my friend, is blowin’ in the wind”

Bob Dylan

Acknowledgments

I would like to thank all the people who helped me during my thesis.

First and foremost, I would like to thank my supervisor Prof J.-F. Focant, who gave me the opportunity to complete my thesis in his laboratory. I would especially like to thank him for his advice, explanations, and the valuable access he granted me to the state of the art technological instruments available in his laboratory.

I would also like to thank all my jury members: Prof. P. Cardinael, Prof. L. Quinton, Dr. V. Bertrand, Dr. C. Brasseur, Dr. G. Purcaro and Dr. F. Schleich for their reviewing of this document. Dr. V. Bertrand for her help during my cell lines culture in her laboratory. It was very rewarding to learn about cell culture topics. Dr. F. Schleich for her collaboration with the CHU of Liège and for the access to the samples.

Thanks to the F.N.R.S. for providing me with financial support for my Ph.D. through a F.R.I.A grant.

Thanks to all my old and current colleagues for their help and cheerfulness. Working with you was a real pleasure. I would especially like to thank Nicolas Di Giovanni for all the discussions we had together and also Dr. P-H. Stefanuto for his support, advices and training on the instruments and software.

Thanks to the people from the CART and LSM groups who helped me during my thesis.

Thanks to the personnel at CHU Liège for helping me to collect my breath samples and for the warm welcome I was given in their unit.

Thanks to all the members of the teaching team I worked with to provide great lessons to students.

Thanks to my family and friends who supported me during these last six years. This work would not have been possible without you.

Finally, thanks to Charlotte for her support and encouragement. My thesis contributes to making me a better scientist, but without you, it would have been futile. I am happy to move forward in my career with you by my side.

Abstract

Lung cancer is the most prevalent cancer in term of mortality in developed countries. This is due to the quietness of symptoms at early stage. Thus, the majority of patients are only diagnosed at an advanced stage, resulting in a poor prognosis. As the curability of lung cancer is highly dependent on early diagnosis, there is, therefore, an urgent need to develop earlier diagnostic screening tests allowing detection of lung cancer at a more curable stage.

Human exhaled breath contains several hundreds of volatile organic compounds (VOCs) that can be seen as a fingerprint that could possibly be used to differentiate between individuals exhibiting various health statuses. Breath analysis has been shown to be usable to highlight possible markers of specific diseases in these individuals. Such an approach is particularly adapted to potential early diagnosis of cancer because its low level of invasiveness and relative eases of implementation on a large scale basis. The implementation of an early diagnostic procedure for cancer screening by means of breath analysis could thus contribute to increase the survival rate of diagnosed patients.

Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GC×GC-TOFMS) has been reported to be able to isolate more than a thousand VOCs from one single human breath. Such an approach is, however, still far from clinical use as it still has to go through full clinical validation. Additionally, the routine use of GC×GC-TOFMS in hospitals for cancer screening is probably not the way to go, as other more simple techniques can be implemented to screen for markers of illness (e.g. e-noses, selected ion flow-tube mass spectrometry, …). Nevertheless, GC×GC-(HR)TOFMS is a key step in extracting a list of reliable markers from the complex mixture made by breath VOCs prior clinical use. We are using GC×GC-(HR)TOFMS for the analysis of exhaled air samples taken from lung cancer patients, as well as from the headspace of cancer cell cultures. Solid-phase micro extraction (SPME) and thermal desorption (TD) are used for sampling. Several data mining approaches and statistical tools (Fisher ratio, random forest, principal component analysis, clustering…) have been implemented to digest the large amount of data generated. A short list of potential markers has been extracted among the large number of features detected initially to bring breath analysis closer to clinical use.

Table of contents

Abbreviations

1D: One dimensional technics

2D: Two dimensional technics

CAR: Carboxen

CHU: University Center Hospital

COPD: Chronic Obstructive Pulmonary Disease

CT: Computed Tomography

DMEM: Dulbecco's Modified Eagle Medium

DMSO: Dimethyl Sulfoxide

DVB: Divinylbenzene

EBC: Exhaled Breath Condensate

EI: Electron Ionization

E-nose: Electronic nose

eV: Electron Volt

^{18}F-FDG: Fluorodeoxyglucose

FID: Flame Ionization Detector

FR: Fisher Ratio

GC: Gas Chromatography

GC-MS: Gas Chromatography coupled to mass spectrometry

GC×GC: Comprehensive two-dimensional gas chromatography

GC×GC-TOFMS: GC×GC coupled to time-of-flight mass spectrometer

HCA: Hierarchical Cluster Analysis

HS: HeadSpace

IMS: Ion Mobility Spectrometry

MS: Mass Spectrometry

MPS: MultiPurpose sampler

NIST: National Institute of Standards and Technology

NSCLC: Non-Small Cell Lung Cancer

NTD: Needle-Trap Device

OBiAChem: Organic and Biological Analytical Chemistry

PCA: Principal Component Analysis

PDMS: Polydimethylsiloxane

PET: Positron Emitted Tomography

P_M: Modulation Period

PTR-MS: Proton Transfer Reaction Mass Spectrometry

PUFA: Polyunsaturated Fatty Acid

RF: Random Forest

S/N: Signal to Noise ratio

SPME: Solid Phase MicroExtraction

SIFT-MS: Selected Ion Flow Tube Mass Spectrometry

SCLC: Small Cell Lung Cancer

TD: Thermal Desorption

TOFMS: Time-of-flight Mass Spectrometry

UV: UltraViolet

VOC: Volatile Organic Compounds

Chapter 1:

Introduction

1.1 Cancer

Cells are the basic unit of all living organisms. They constitute the building blocks of life. They multiply themselves and die in a coordinated way known as apoptosis [1]. This process allows the organism to keep its number of cells stable. However, sometimes cells undergo some mutations. At this point, they are no longer involved in the good functioning of the organism and they also stop following the apoptosis process, forming an aggregate of cells called tumors (Figure 1). There are 2 kinds of tumor. When cells stay localized in the tumor and stop to grow, it is called a benign tumor. No major problem is caused by them. But when cells continue to increase without any control, it becomes a malignant tumor. This localized cancer disturbs normal cells around them, leading to an overall malfunctioning of the tissue. Moreover, this tumor can move to other tissue or organs across the organism and create new sources of cancer cells called metastases. The Result of malfunctioning in an organism caused by cancer cells leads to death.

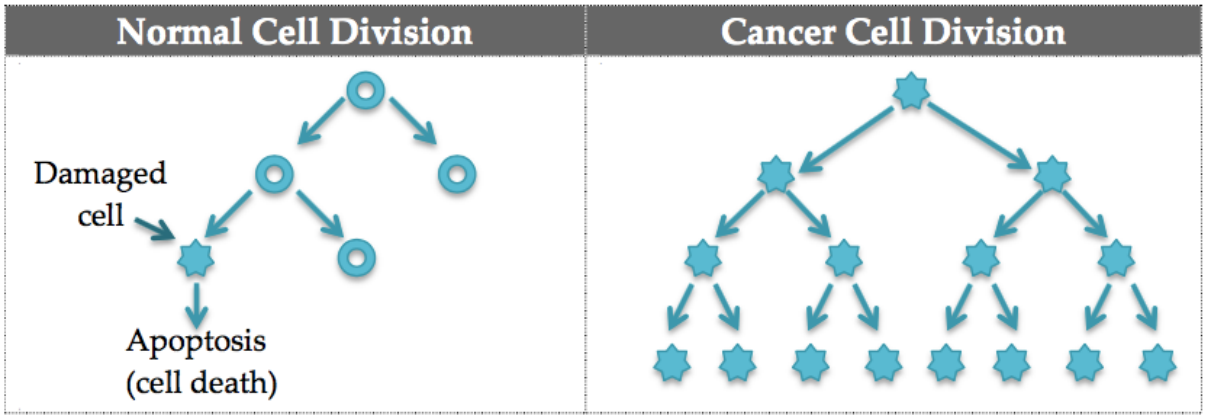

Figure 1: Cell division for normal and cancer cells

Cancer is one of the most deadly diseases that exist today [2] and is present all around the globe (Figure 2). In the United States, cancer is responsible for about one quarter of deaths per year. Figure 3 illustrates the estimated new cases and deaths by gender for US residents in 2018 [2]. Other industrialized countries of the world show similar statistics. Therefore, on average, individuals have almost 50% of chance of developing cancer throughout entire lifetime [2]. The size and colonization degree of cancerous cells is defined in four stages I, II, III and IV depending on the degree of advancement of the pathology. Cancers diagnosed at an early stage (I and II) have a better chance of being cured compared to later stages (III and IV) (Table 1)[3].

Table 1: Five-year survival rate for cancer according to the diagnostic stage

Stage I	Stage II	Stage III	Stage IV
92%	70%	39%	17%

For example, the mortality rate for breast cancer for women is relatively low compared to its incidence. This is because this cancer is easier to detect at an early stage (Figure 3).

Cancer affects all tissues of the organism. Even if the cancer development is the same, each organ reacts differently to the presence of the pathology. As a result, there is no global cancer diagnosis for the first stages. Researchers and clinicians are then forced to establish a specific early diagnosis for each type of cancer.

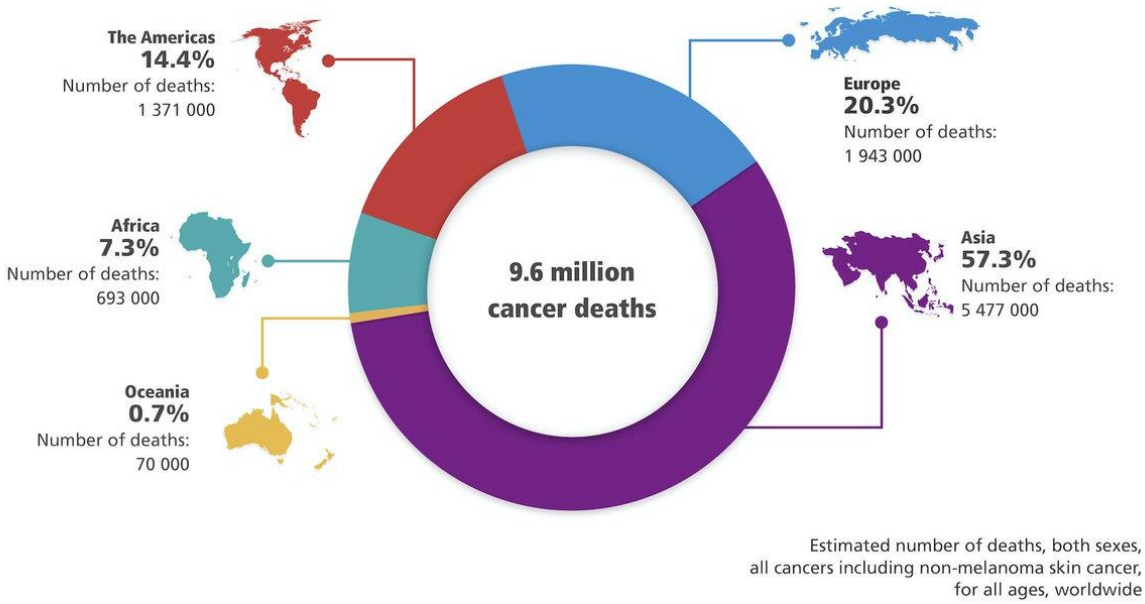

Figure 2: Estimated number of global cancer mortality in 2018 (www.uicc.org)

Estimated New Cases

Males			Females		
Prostate	164,690	19%	Breast	266,120	30%
Lung & bronchus	121,680	14%	Lung & bronchus	112,350	13%
Colon & rectum	75,610	9%	Colon & rectum	64,640	7%
Urinary bladder	62,380	7%	Uterine corpus	63,230	7%
Melanoma of the skin	55,150	6%	Thyroid	40,900	5%
Kidney & renal pelvis	42,680	5%	Melanoma of the skin	36,120	4%
Non-Hodgkin lymphoma	41,730	5%	Non-Hodgkin lymphoma	32,950	4%
Oral cavity & pharynx	37,160	4%	Pancreas	26,240	3%
Leukemia	35,030	4%	Leukemia	25,270	3%
Liver & intrahepatic bile duct	30,610	4%	Kidney & renal pelvis	22,660	3%
All Sites	**856,370**	**100%**	**All Sites**	**878,980**	**100%**

Estimated Deaths

Males			Females		
Lung & bronchus	83,550	26%	Lung & bronchus	70,500	25%
Prostate	29,430	9%	Breast	40,920	14%
Colon & rectum	27,390	8%	Colon & rectum	23,240	8%
Pancreas	23,020	7%	Pancreas	21,310	7%
Liver & intrahepatic bile duct	20,540	6%	Ovary	14,070	5%
Leukemia	14,270	4%	Uterine corpus	11,350	4%
Esophagus	12,850	4%	Leukemia	10,100	4%
Urinary bladder	12,520	4%	Liver & intrahepatic bile duct	9,660	3%
Non-Hodgkin lymphoma	11,510	4%	Non-Hodgkin lymphoma	8,400	3%
Kidney & renal pelvis	10,010	3%	Brain & other nervous system	7,340	3%
All Sites	**323,630**	**100%**	**All Sites**	**286,010**	**100%**

Figure 3: Ten leading cancer types for the estimated new cases and deaths by gender in the United States in 2018 [4]

1.2 Lung cancer

This study only focuses on lung cancer because this pathology is the second most common cancer in term of incidence and it is the number one cancer in terms of mortality for men and women [2, 5]. This is due to its typical silence early on in the course of its development [6, 7]. Therefore, the majority of patients are diagnosed at an advanced stage, resulting in poor prognosis [8]. Statistics reveal a 70% 5-year survival rate for patients diagnosed at an early stage compared to 10-15% at stage IV [9]. In 2018, in the United States only, lung and bronchial cancer deaths are estimated to more than 154,000 [4]. There is, therefore, an urgent need to develop earlier diagnostic screening tests allowing for detection of lung cancer at a more curable stage.

1.2.1 Type of lung cancer

Different types of cancer can affect the lungs. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) represent the most common lung cancers in term of incidence. NSCLC represent around 85% of lung cancer cases and includes adenocarcinoma, squamous cell carcinoma and large cell carcinoma. SCLC represents around 15% of lung cancer cases. Even if it less common than NSCLC, it grows faster, forming metastases more easily. Their treatments are not similar. Usually, SCLC is cured by chemotherapy and radiotherapy whereas NSCLC is cured by surgery [10, 11].

1.2.2 Causes of lung cancer

Lung cancer may be caused by different factors and the most popular is smoking [12-15]. Before 1878, lung cancer represented only 1% of cancer deaths. But since that date, it has increased due to smoking [16]. 75% of diagnosed lung cancers are due to smoking [17, 18]. It is because cigarette smoke contains about 4,000 chemical compounds, including more than 100 mutagens or carcinogenic compounds, such as benzene, arsenic, chrome, formaldehyde and toluene [19, 20]. The relation between the incidence of lung cancer and smoking is clearly defined (Figure 4). Advertising about the dangers of smoking has led to a decrease in the number of smokers these last few years, leading to a clear decrease in estimated new lung cancer cases [21-23]. Inhalation of secondhand smoke from non-smokers individuals also increases the risk of lung cancer. Moreover, other factors described below also lead to this pathology and its incidence remains high.

Exposure to chemical substances is another cause of lung cancer, especially over a long period of time. People living or working in a polluted air environment from industry and vehicles increase their risk to develop lung cancer [24]. Asbestos exposition also increases the risk of lung cancer. This buildings materials has been used extensively in the past but but it is now known that exposure to it may induce lung cancer [25-27]. Since 1987, it was classified as a group 1 carcinogen by the international agency for research on cancer and is currently forbidden in building materials.

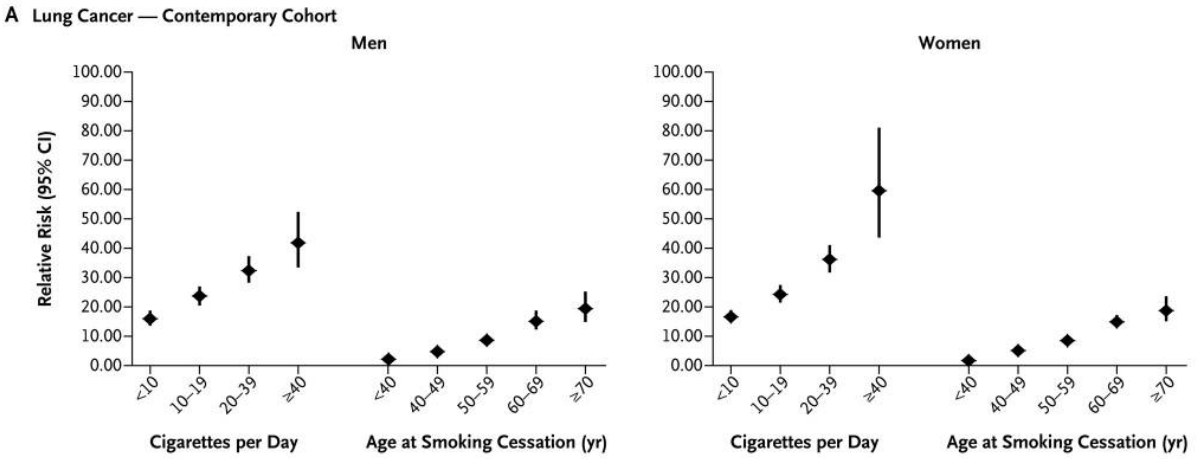

Figure 4: Relative risks of lung cancer among current smokers, according to number of cigarettes smoked per day, and among former smokers, according to age at the time of quitting [28]

Exposition to radon radiation is also often found as a cause of cancer. This radioactive element is a natural gas resulting from the disintegration of uranium in some soils [29]. It is responsible for the largest proportion of human irradiation (about 40% in France)[30]. Radon is related to more than 20,000 cases of lung cancer deaths in the US every year. It mainly affects miners and populations who live on uranium-bearing lands. Usually, the maximum level of radiation tolerated in most countries is 200 Bq / m^3 [31, 32]. However, these doses may be largely carcinogenic [33-35].

Finally, individuals with family history of lung cancer can raise the likelihood of developing lung cancer [22].

1.2.3 Diagnosis and Treatment for lung cancer

There are several methods for diagnosing lung cancer. Traditional methods include medical imagery like pulmonary radiology, computed tomography (CT) and positron emitted tomography (PET) [36]. These methods cannot be used to screen large population of individual because they are expensive, time consuming and require high skill operators. Most of the time, patients undergo radiology or CT for another health problem and lung cancer is detected by 'chance' at the same time. Then, the presence of cancer has to be confirmed by PET scan, blood testing, sputum testing or biopsy. A biopsy shows which cancer type and stage is present and helps to determine which treatment is the most appropriate for the patient.

Pulmonary radiography is a common medical imagery method using x-ray to display lung infections and lung or cardiac pathologies. It could be used to detect lung cancer but it leads to false results and does not detect cancer at the early stage [37]. CT scan is another medical imagery techniques based on x-ray where patient is surrounded by a x-ray beam and a detector (Figure 5)[38]. A computer reconstructs an image of the thoracic cage as a function of the absorbance intensity of the beam. It is possible to detect lesions as small as a millimetre size. However, it also leads to false results, it is expensive and there is a putative risk of radiation from long time exposure [39-41].

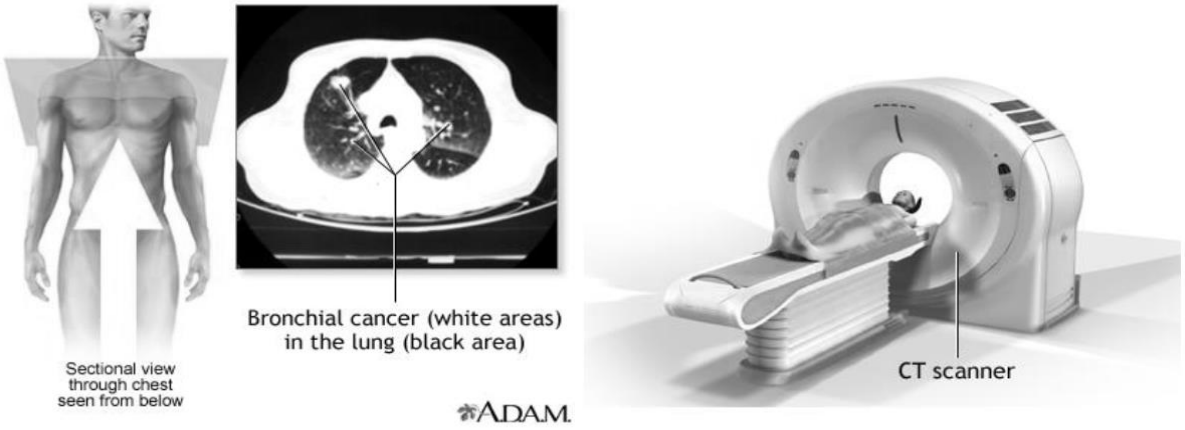

Figure 5: On the left image obtained by tomodensitometry and on the right a representation of the instrument (www.nlm.nih.gov)

Positron emission tomography (PET) is a medical imaging method based on the functioning of the lungs. The first step consists of injecting a radioactive tracer compound into the patient's body. This injection is carried out one hour before the analysis. The typical tracer in the diagnosis of cancer will be fluorodeoxyglucose labelled with radioactive Fluor 18 (^{18}F-FDG). Indeed, this molecule is similar to glucose and will accumulate in cancerous tissues that consume more glucose because of their increased metabolic activity. Once the marker has accumulated in the tissues, the patient is scanned using a PET camera. The positrons emitted by the radioactive element of the marker will annihilate in contact with the electrons present in the medium in order to release two photons of opposite directions and with a well-defined energy (511 keV). A computer reconstructs a functional regions map of the thorax in order to locate the pathological region. It is also possible to couple this equipment to the CT scanner to precisely map the tumor [42, 43]. This method can determine how far the tumor has spread and help the clinician to choose the best solution to cure the patient. However, as PET scan use radioactive element, it is not

recommended for pregnant patients. Synthesis of the radioactive element through a cyclotron makes this method more expensive than other methods.

Usually, next to these analyses, a lung biopsy is performed on the patients to determine the type of cancer cells affecting them. A biopsy is a very invasive technic which consist of sampling a piece of tissue for further analysis. A biopsy can be applied in different ways depending on where the lesion is located. One option consists of a needle inserted through the rib cage to take a pulmonary sample, guided by a CT scan (Figure 6). It is also possible to perform an open lung surgical biopsy under anaesthesia. Finally, another common way is to perform a bronchoscopy of the lungs combined with the biopsy. It consists of observing the inside of the respiratory ducts by means of a bronchoscope apparatus consisting of a camera and a lamp at the end of a flexible hose. The sample is taken directly by the bronchoscope [44, 45].

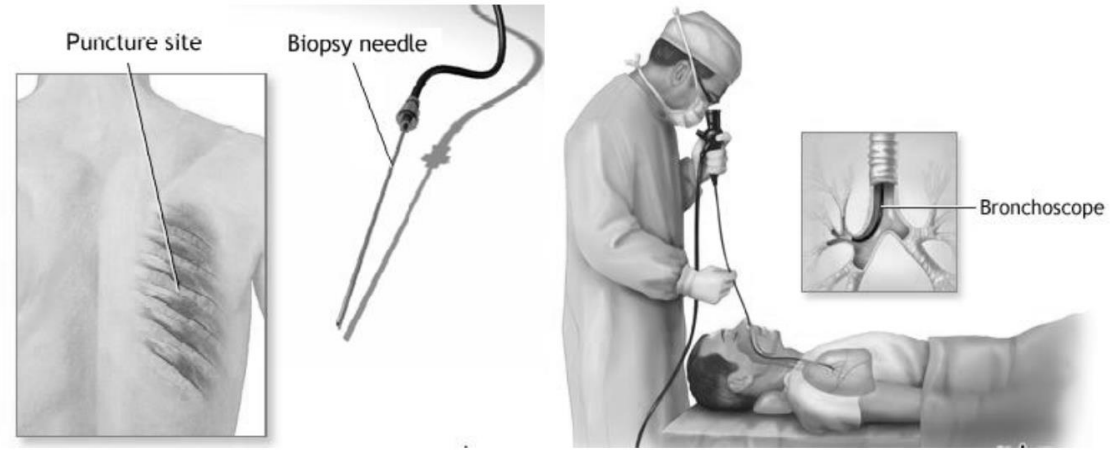

Figure 6: On the left, biopsy needle use for sampling tissue and on the right, its used on patient (www.nlm.nih.gov)

However, these methods are not suitable for early diagnosis of lung cancer because it is not possible to screen large population due to the large amount of resources needed to perform them. These methods are more use to confirm the presence of lung cancer for suspicious patients after a discussion with their regular physician. Therefore, other early alternative methods have to be investigated [46, 47]. The use of gene and protein profiling provides new techniques to detect cancer [48-50]. However, it has some limitations due to tumor heterogeneity, complex interpretation of tumor-host link and redundancy of signal emitted by tumor-cells including epigenetic, genetic, and micro environmental impacts. Moreover, those methods are expensive, require large amounts of biological tissue and are time-consuming.

Medical principles say pathology induces variations all across the body [51]. Hence, it is possible to monitor specifics compounds to detect pathologies. This is already the case with blood analysis [52]. For example, it has been demonstrated that the level of the protein CKAP4 is higher in the blood of lung cancer patients than in healthy individuals. This statement is valid for patients with stage I tumor. Hence, it is possible to monitor the level of CKAP4 in the blood as an early detection of lung cancer [53]. Based on the same idea, it is also possible to monitor compounds present in the exhaled breath to develop early diagnosis of diseases [51, 54]. This relative new way to detect diseases offer several advantages compare to blood analysis and is further describe in the following sections.

1.3 Breath analysis

1.3.1 History of breath analysis

Although Breath analysis has been central to much of the research carried out during the last decade, its origins comes from antiquity, when Hippocrates made the link between *fetor oris* and *fetor hepaticus* and breath odor [55]. The fruity smell emitted in the exhaled breath of diabetic ketoacidosis patients is well known by doctors. Lack of insulin induces the production of ketones because of the inability of the body to prevent the absorption of glucose, and this ketone production is responsible for the specific fruity smell found in exhaled breath [56, 57].

In 1971, Linus Pauling had the great idea to analyse human exhaled breath by GC [58]. He discovered the presence of several compounds, and laid the foundations for modern breath analysis. Since then, several researches programmes have investigated this promising field (Figure 7) and several markers have been already reported [59]. For example, ^{13}C-urea breath test provides a reliable test to detect *Helicobacter pylori* infection [60]. Monitoring nitric oxide for asthma, hydrogen for small intestinal bacterial overgrowth or fructose and lactose malabsorption, CO for neonatal jaundice provide new tools for the diagnosis of such pathologies by breath analysis [61-65].

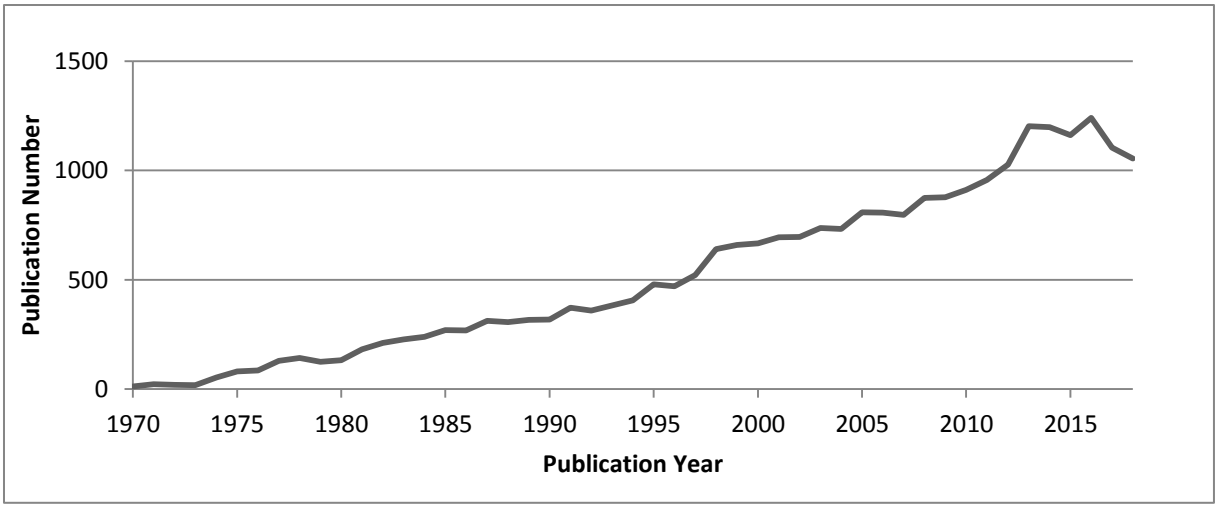

Figure 7: Number of publications per year found in Pub med library with "breath analysis" keys word from 1970 to 2019

Cells, tissues and microbiomes synthetize and release VOCs. They can be a promising source of information about the healthy status of an individual [66, 67]. Hence, scanning such compounds in exhaled breath provides new opportunities for the diagnosis of diseases [68-73]. Moreover, linking VOCs found in exhaled breath and the status of pathologies may lead to a better understanding of how diseases affect the body [74]. Scientific consensus compares healthy volunteers to patient cohorts to seek variation between both groups [75]. It was demonstrated that the pathology affects the VOC profile by comparing VOC profile of cancer patients before and after surgical removal of the tumor [76].

1.3.2 Composition of exhaled breath

Exhaled breath has almost the same composition than the air present in the environment. The biggest difference in term of concentration comes from the transfer of oxygen to carbon dioxide in the lung. Thus, exhaled breath is mainly composed of of N_2 (~75%), O_2 (~15%) and CO_2 (~5%). However, it is also contains VOCs, inorganic species, vapor and other aerosols [77, 78]. VOCs are defined by article 2 of the directive 1999/13/CE from the European legislative body as "any organic compound having at 293,15 K a vapor pressure of 0,01 kPa or more or having a corresponding volatility under the particular conditions of use". They are often analysed in different fields, like in forensics [79, 80], petrochemistry [81], the food industry [82], tobacco combustion [83], medicine [59, 84], environment control [85], etc.

VOCs present in the exhaled breath can come from two different sources. They can be inhaled from the surrounding air and then released during expiration. They do not give any information on the healthy status of individual and are more seen as contamination. Next to them, other possible forms of contamination include VOCs from diet, smoking, lifestyle [86-89]. Those compounds do not point to a healthy status in an individual and they are called exogenous VOCs. For example, acetonitrile had been reported to be present only in the breath of smokers [90]. Sometimes, studies require patients to refrain from eating or smoking for a period of time before taking a sample of the exhaled breath in order to avoid/ reduce the impact of such contaminations. However, it is not possible to complexly remove those exogenous VOCs because some of them can enter the body through the skin and are releases into the breath few hours after their assimilation [7].

The second source of VOCs present in the exhaled breath are call endogenous VOCs. Metabolism occurring in the body releases some sub-products in the blood. When blood comes to the lung to transfer CO_2 and absorb O_2, it releases these small molecules too (Figure 8). Hence, those compounds are present in the exhaled breath. They reflect the healthy status of an individual as pathologies induce trouble in the body that its translated by a variation of his/her VOC profile [91].

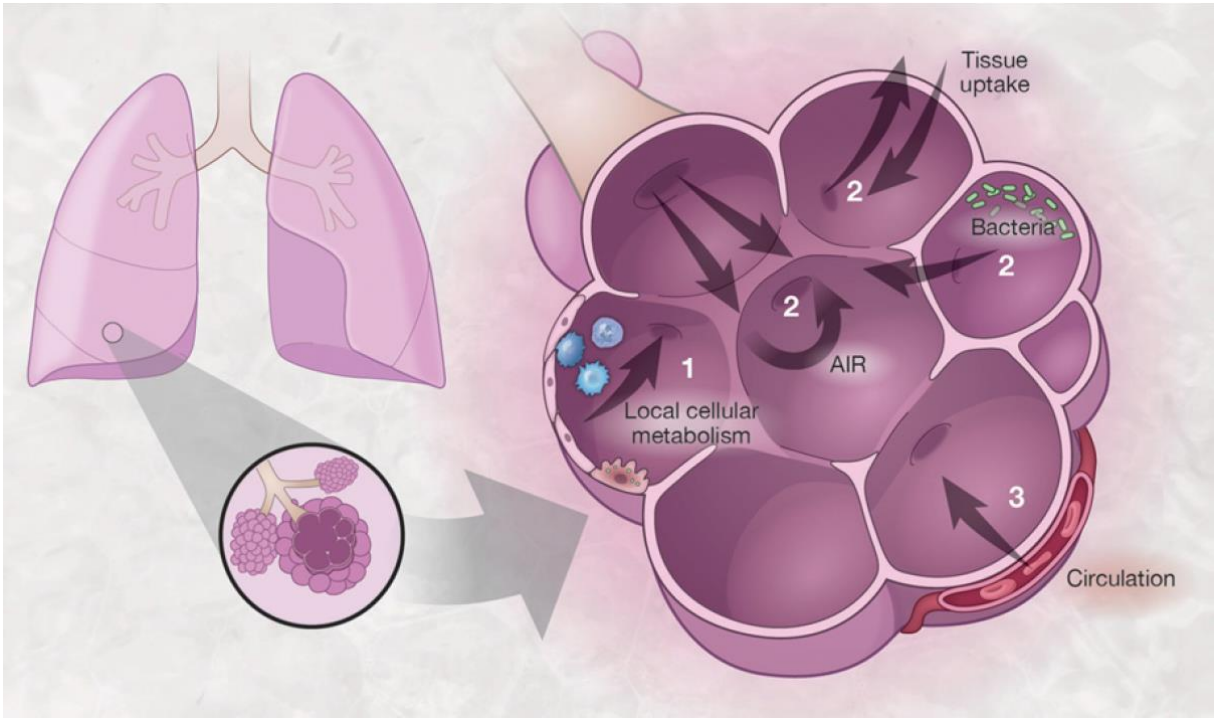

Figure 8: Schematic representation of the exchange of VOCs from the blood to the exhaled breath by using the respiratory system. Bacteria and local cellular metabolism also impact the composition of the VOC profile emitted through the exhaled breath [92]

It is possible to characterize VOCs from all parts of the body as they are transported by the blood to the lungs before being expelled during expiration. Their intensity fluctuates in accordance with the way we breathe. When we exhale, the first part of the gas phase comes from the upper part of the respiratory system (including mouth, nose, pharynx, trachea and bronchi). This is called dead space and represents a volume of around 150 mL. Then, the second part of the expiration coming from the lung is exhaled. The gaseous part of this is called alveolar air. Deeper breath brings more important quantities of alveolar gas, increasing the intensity of endogenous VOCs [91]. Therefore, there are different ways to collect exhaled breath samples: to sample its alveolar air and to sample its total volume. When sampling the total breath of a patient, the ambient air dilutes the alveolar air, and the endogenous VOCs are found in a lower concentration. When sampling the end of the expiration, this alveolar air is found in greater quantity and is therefore more representative of the endogenous VOCs released by the patient. It is possible to sample the alveolar air by measuring the amount of exhaled CO_2 released in the exhalation. Indeed, at the beginning of the exhaled breath, the concentration of CO_2 release increases until it reaches a plateau (Figure 9). And at this point, only alveolar air is present [17, 93, 94].

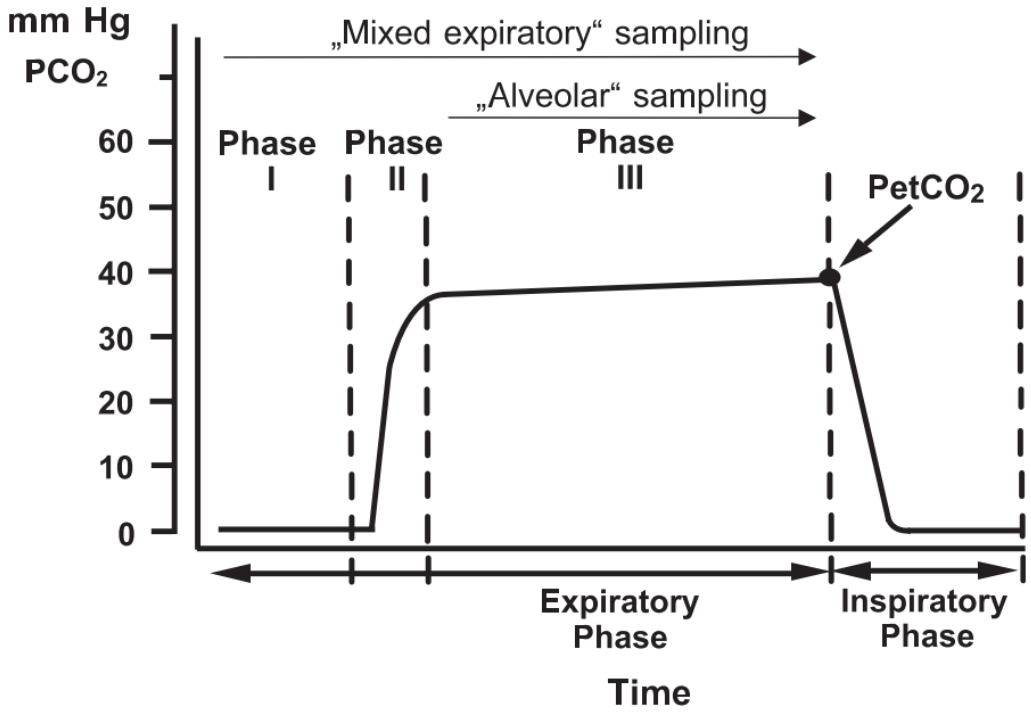

Figure 9: The amount of CO_2 expelled in the exhaled breath in relation to the breathing step [91]. The majority of VOCs are exhaled during phase III

Around one thousand VOCs are detected in a single breath [92] and a majority of studies used multiple breaths to record the VOC signal because their concentrations range from pptv to ppmv [86, 95-98]. However, it is also possible to use one single breath to detect VOCs [99].

Finally, non-volatile compounds emitted in breath can also be detected with exhaled breath condensate (EBC). This method involves collecting the breath, cooling it down to create an accumulation of condensed fluids consisting mainly of soluble and non-volatile compounds. The genomic, transcriptomic, proteomic and metabolomic methods can therefore be applied in order to analyse different molecules present in these condensates to highlight certain pathologies [100-103].

1.3.3 The advantages and disadvantages of breath analysis

Conventional diagnostic procedure uses blood sampling or tissue biopsies. These invasive methods are not always welcomed by patients and providing a non-invasive technique like breath analysis could solve this problem (Figure 10). Several pathologies have an impact on the integrity of exhaled breath. Hence, developing a method to characterize the breath matrix could be used in multiple clinical fields. Moreover, like in the case of blood, it could be possible to screen several factors of diseases through one single breath sample. A study reported the segregation of different cancer patients and also healthy volunteers [104]. This method also has several other advantages. It is pain free, easy to use for the clinician and the patient, inexpensive and easily accessible to all [66, 68]. It can be performed at any time and is unlimited, which it is not the case for the urine matrix or liquid biopsy techniques [105].

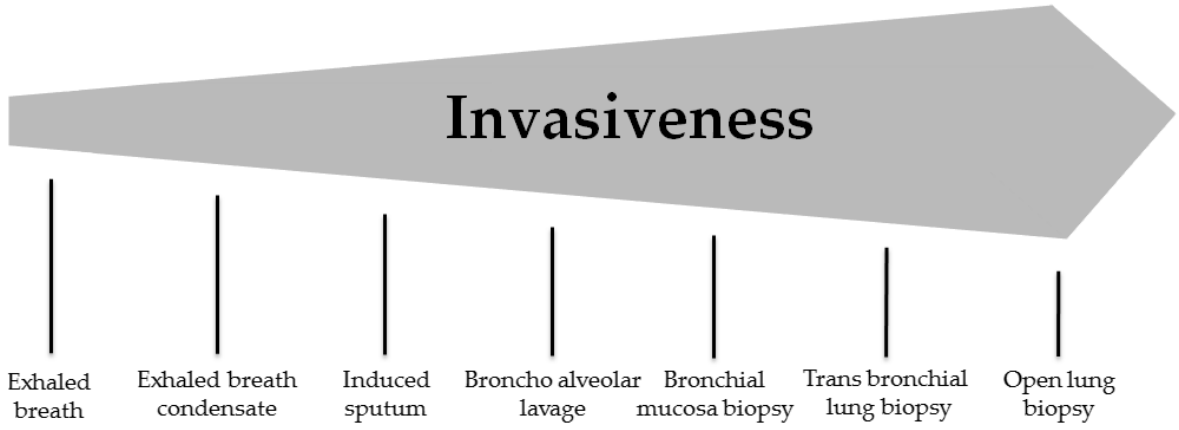

Figure 10: Classification of medical analysis method by their invasiveness

However, these screening methods are not yet able to be used at the clinical level, because the lack of standardized methods or devices for breath collection hampered the robust identification of VOC markers [106-108]. This is partially due the variation in the volume of the exhaled breath collected, the method of sampling (Tedlar® bag, Mylar® bag, home-made device, etc.) and the part of exhaled breath collected (alveolar and total breath). Impact of exogenous VOCs or contamination arising during the storage of breath samples also plays a harmful role in the discovery of putative VOC markers. High variability between samples also leads to difficult data interpretation. Age, gender and smoking, all play an important role in the variability between individuals [109]. The water present in the exhaled breath could potentially condense depending on the sampling method and absorb some VOCs [110, 111]. The high variability from the healthy groups, used for comparison with the patient cohort, induced by smoking factors (non-smokers, ex-smokers or smokers) or higher age variability also lead to difficult data interpretation and make the discovery of putative markers more delicate [86].

1.3.4 Sampling exhaled breath

There are two major ways to analyse exhaled breath samples. With on-line analysis, the sample goes directly from the patient to the instrument to be analysed. Therefore, this kind of analysis does not require sample storage. Usually, proton transfer reaction mass spectrometry (PTR-MS), selected ion flow tube mass spectrometry (SIFT-MS) and e-nose technologies are the most common choice because their time analysis is relatively short and display results in short delay.

On the other hand, in the case of off-line analysis, breath samples are collected and stored before being injected into the instrument. Tedlar® bag is the most common choice to trap breath samples even though other alternative methods could also be used, such as Mylar® bags or homemade devices. These bags consist of an inert membrane of polyvinyl fluoride and are not supposed to change the integrity of the gas sample. They are easy to use for the clinicians because they do not require specific formation or additional equipment to fill them. It is also easy for the patient to blow into them. Different volumes are available (one litre, two litres or five litres for example) and can be selected in accordance with the goal of the study. This kind of equipment is relatively cheap, and it is possible to use a bag several times after cleaning it. However, several studies indicate that they leak and release compounds. Phenol and N-N-dimethyl acetamide are commonly cited in the literature to be the

main compound released by those bags [112-114]. Other studies use alternative bags of alumina foil sheet, known as Mylar® bag, to overcome this issue [8, 115, 116].

There are three main ways to transfer the VOCs collected in the Tedlar® bag to the instrument. The most common choice is using thermal desorption (TD) tubes [117-119]. These tubes are stainless steel or glass tubes filled with one or more adsorbents. These sorbents can be of several types [120]. They may consist of activated carbon, polymer and a carbon molecular sieve which trap the gas inside its structure. All of these sorbents have different properties such as polarity or pore size. This allows the absorption of volatile or semi-volatile organic compounds. VOCs are transferred from the Tedlar® bag to the TD tube by deflating the bag with a pump. Then, the tubes are placed in the thermo desorption unit to allow the heat to release the VOCs trapped inside the tube into the instrument of analysis. The large loading capacity of the sorbent and the absent of solvent are the two main advantage of this technique [121]. The high stability of the trapped VOCs means these can be stored in those tubes in order to inject all samples during a same session into the instrument, to decrease the instrumental variability. These tubes can be conditioned by heating them at high temperature (depending on the nature of the sorbent, usually around 300-350°C) and purged with a nitrogen flow. On the other hand, thermal decomposition of compounds might occur during the heating step necessary to release the compounds from the sorbents to the instrument. And during thermal desorption, polymeric sorbents might depolymerized because of the high temperature. This reaction will release monomeric units and decrease the trapping performance of the tubes for the next samples [122]. For gas analysis, as this technic does not use concentration step, it is necessary to use it with large sample volume to overcome the low concentration of compounds [123].

The second common option is to insert a solid phase micro extraction (SPME) fiber inside the Tedlar® bag through a specific valve to trap some VOCs on the surface of the fiber [90, 124-131]. The SPME technique was invented by Arthur et Pawliszyn in 1989 and consists of a fused silica fiber coated with an sorbent polymer capable of catching and trapping the VOCs present in a gas phase [132]. The coating could be a single sorbent or a mix of different ones. The choice of the coating depends on the kind of compounds analysed [133]. Usually SPME is used to analyse the headspace (HS) of solid and liquid phase or also gaseous phases. It can be applied in the analysis of semi-polar, non-polar, semi-volatile and volatile compounds. It should be noted, however, that the sorption of the SPME fibers is governed by distribution constants and that these fibers are not designed for exhaustive media analyses [134].

Compare to TD tubes which use absorption, trapping of compounds is mainly dictated by adsorption but it might have some competition with absorption depending on the type of fiber used. In addition, the total sorption capacity of the fibers is very limited compare to TD tubes. Coating choice, needle penetration, time of exposition of the fiber above samples and desorption time have to be optimized to reach optimal trapping of compounds. However, this technique is fast, inexpensive, does not consume solvent and could be automated [129, 135-137]. Another advantage is that integration, extraction, concentration and introduction of the sample is carried out in one step, which constitutes a considerable time saving factor in the preparation of the sample (Figure 11). For VOCs analysis, the concentration step allows detection of compounds present at low concentration even with small sample volume [122]. Like with TD tubes, a further advantage is the absence of solvent, deleting the solvent peak in the chromatogram, the dilution of compounds and also the solubility problem induced [138]. However, degradation of thermally unstable compounds inside the injector may occur. This method shows its capacity to characterize VOCs present in human biological specimens [84].

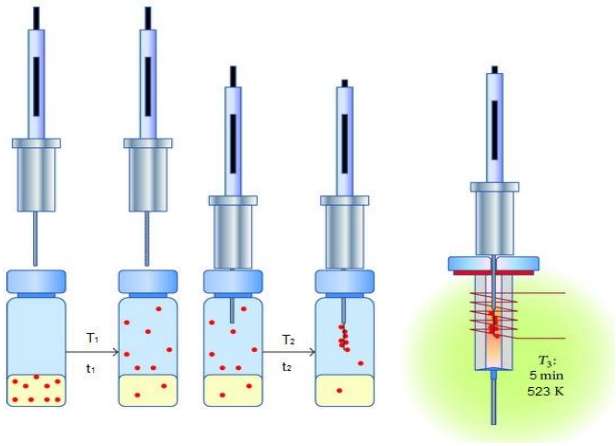

Figure 11: Diagram of the use of SPME fiber to sample the HS of the liquid sample [139]

The last common option to collect VOCs is to use a needle trap device (NTD)[140]. This method consists of using a needle where VOCs can be trapped inside the needle by diffusion of the gas on sorbent. With its small diameter, the needle is introduced in GC inlet to thermally desorb the sample. This method combines the low volume require to perform the sampling like with SPME fiber but it does not have the concentration step. And like TD tubes, it is possible by increasing the sample volume to increase its sensitivity but it is limited by its loading sorbent capacity compare to TD tubes [110]. This is why this method is less popular than SPME and TD and only few publications reported its use [110, 141-143].

Other sampling devices were developed but remain less popular than the technique described above. For example, Markers had developed its own device to trap exhaled air call Bio-VOC™. Breath is collected in a plastic cylinder of 88 mL free from plastic emissions. Then, a TD tube is connected and a piston allows to transferring VOCs from the cylinder to the tube (Figure 12). It is also possible to connect a SPME fiber to collect those VOCs [144]. This kind of device could be used to analyse VOCs with a high concentration but is not suitable for majorities of VOCs found in exhaled breath due to its limited volume [145].

Figure 12: Bio-VOC™ from Markes (www.markes.com)

Owlstone medical in Cambridge (UK) developed its own sampling device, called ReCIVA™, where exhaled breath was directly collected in several TD tubes (Figure 13). Absence of intermediate trapping items reduces the risk of contamination. The presence of the CO_2 pressure sensor makes it possible to select the part of the breath sampled to normalize the amount of endogenous VOCs collected. It contains a single- use masks adapted from children or adults and also includes a bacterial filter to prevent any cross-contamination. This device available since the end of 2017 has been used at over 100 clinical sites around the world and tend to appear to be the standardized method to collect exhaled breath sample.

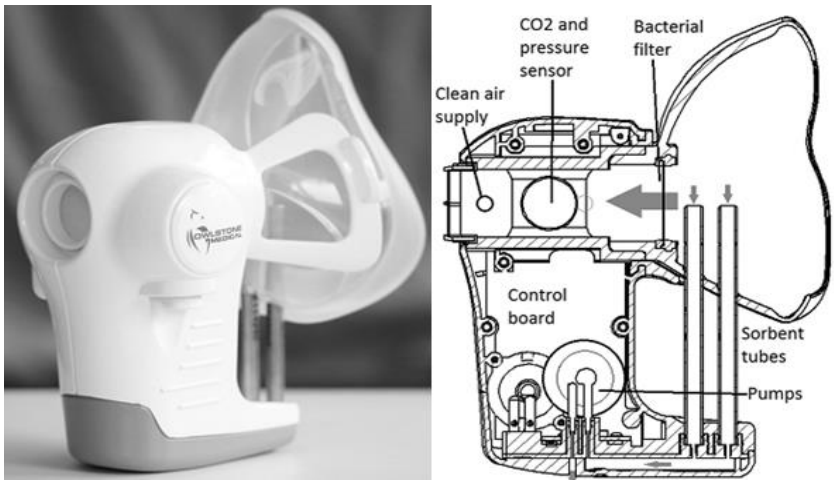

Figure 13: ReCIVA™ sampling device for breath analysis design by Owlstone medical (www.owlstonemedical.com)

Sometimes sampling devices record the amount of CO_2 expelled to normalize the intensity of the VOCs as explained in the previous section. Other teams developed their own device to sample and analyse breath based on the use of e-noses [146-148].

1.3.5 Instruments for exhaled breath analysis

There are several ways to analyse exhaled breath and all instruments have their advantages and drawbacks. The first parameter to take into account is the type of sampling. For on-line analysis, an instrument which displays fast results is needed. Whereas it is not the case for off-line analysis but usually those instruments give a better characterization of the breath matrix. Hence, off-line analysis is most of the time dedicated to untargeted analysis to discover new VOC markers [17]. Those markers could be detected with another instrument by means of on-line analysis directly in the hospital and deliver results to help clinicians in their diagnostic process.

On line analysis mainly includes soft ionization mass spectrometry with selected-ion flow tube (SIFT-MS) [149-151] or proton-transfer reaction mass spectrometry (PTR-MS)[85, 152-157]. On the other hand, GC-MS and IMS are the most common choice for off-line analysis [158]. GC is known as the golden method and represents the major method used for breath analysis in general (Figure 14). All these instruments are briefly described below.

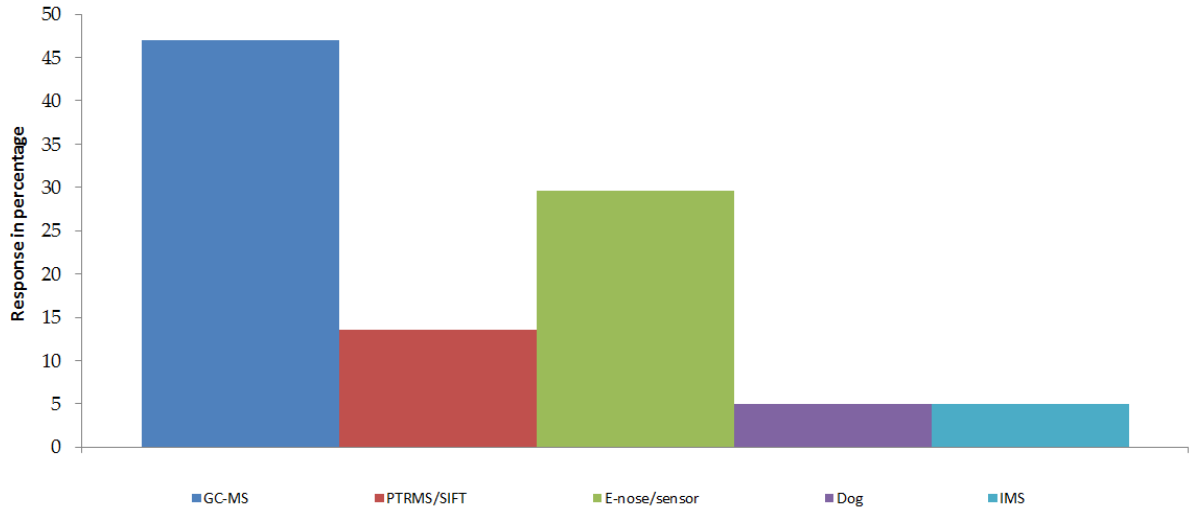

Figure 14: Percentage of the most conventional methods used in publication to characterize breath analysis from lung cancer patients

1.3.5.1 Gas chromatography coupled to mass spectrometry

The golden method to inject and analyse the breath matrix is gas chromatography coupled to mass spectrometry (GC-MS) [70, 84, 159-161]. In this method, VOCs are injected into a chromatography column and transported through it by a carrier gas (He or H_2). The compounds will be separated according to their affinity with the stationary phase of the column and reach the MS, which serves as a detector, at different times. Structural information obtained from mass spectra is used to determine the nature of the VOCs. It is also possible to use the retention time of compounds in the column to determine their structure. GC-MS is characterized by high sensitivity and good robustness and enables qualitative and quantitative analyses. However, the time of analysis for one sample is around one hour and it is necessary to preconcentrate the sample. It is possible to use fast GC runs to reduce the analysis time to few minutes. But the separation power will decrease and, due to the complexity of the sample, it is not possible to use such an alternative for untargeted analysis.

1.3.5.2 Comprehensive two-dimensional gas chromatography

Some studies also used comprehensive two-dimensional gas chromatography (GC×GC) instead of 1D-GC-MS to overcome the issue of the co-elution induced by the complexity of breath samples [111, 141, 162-168]. Phillips demonstrated the utility of such an instrument compared to conventional GC-MS for breath analysis [165]. In his study, he underlined the fact that the exhaled breath sample is very complex by the number of different VOCs constituting this matrix and using GC-MS may miss some crucial information due to the high number of co-elution occurring. Moreover, the sensitivity of GC×GC is higher than with 1D-GC, which is another advantage for samples with low concentration as in the case of exhaled breath analysis. The major drawback of GC×GC-MS is to extract the essential information across the large set of data generated.

1.3.5.3 Proton Transfer Reaction Mass Spectrometry

Proton Transfer Reaction Mass Spectrometry (PTR-MS) is a soft ionization spectrometric method often used for VOC analysis [74, 85, 153, 156, 169]. It consists of using chemical ionization of VOC by using H_3O^+ (Equation 1). This reaction occurs because the affinity of H^+ is higher for VOC than for H_2O or ambient air. Then, protonated VOCs go through the mass spectrometer part to be separated according to their m/z (Figure 15).

$$VOC + H_3O^+ \rightarrow H_2O + HVOC^+ \qquad \text{Equation 1}$$

This method has the advantage of not requiring any method of separation or preconcentration of the sample. In addition, the compounds present in high concentration in air such as N_2, CO_2, O_2 and H_2O do not interfere with the measurements. Finally, this method has a very high sensitivity. However, it is only suitable for targeted analysis as it is necessary to know what is sought beforehand.

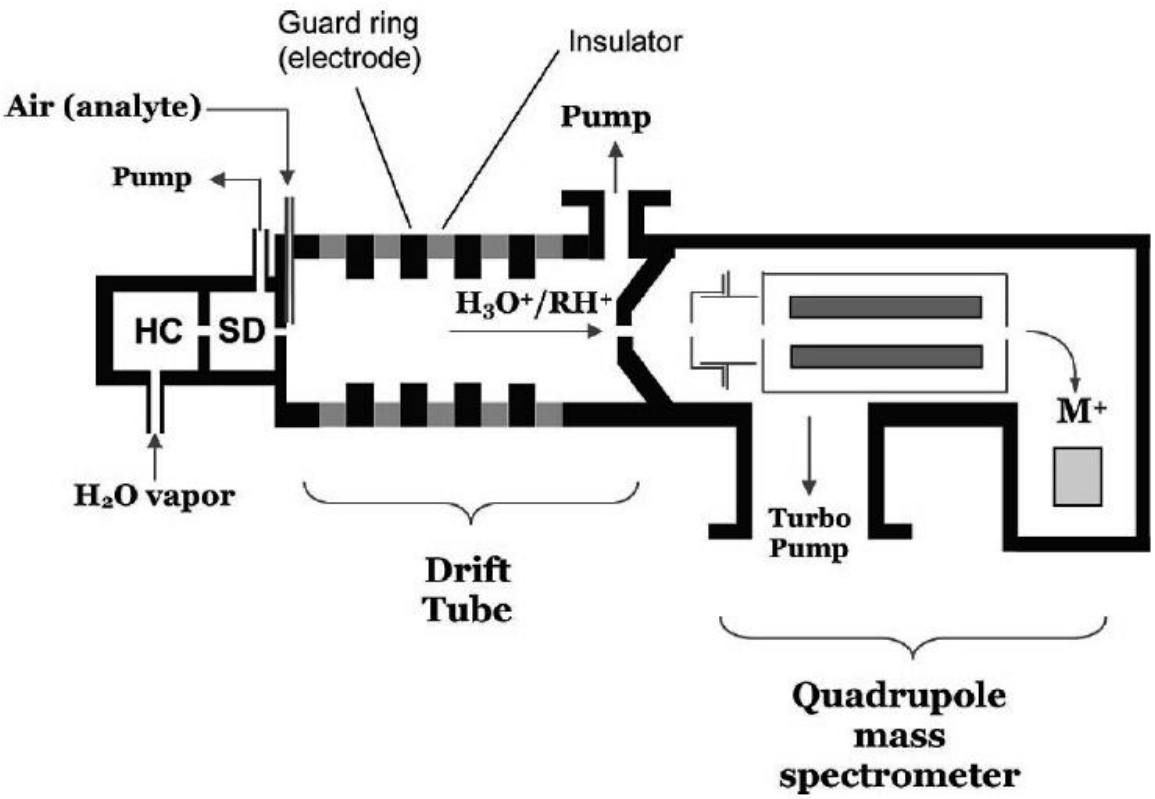

Figure 15: Schematic illustration of a proton-transfer reaction mass spectrometry (PTR-MS) instrument [170]

1.3.5.4 Selected Ion Flow Tube mass spectrometry

The selected ion flow tube mass spectrometry (SIFT-MS) works in much the same way as PTR-MS. With SIFF-MS a mass filter is interconnected to select the ionization agent (Figure 16). In the result, it is possible to use H_3O^+, NO^+ or O_2^+ as precursor ions. Those precursors are then sent to the flow tube to be mixed with VOCs and undergo chemical ionization. Finally, like with PTR-MS, VOC ions go through the spectrometer to be separated according to their m/z [169, 171].

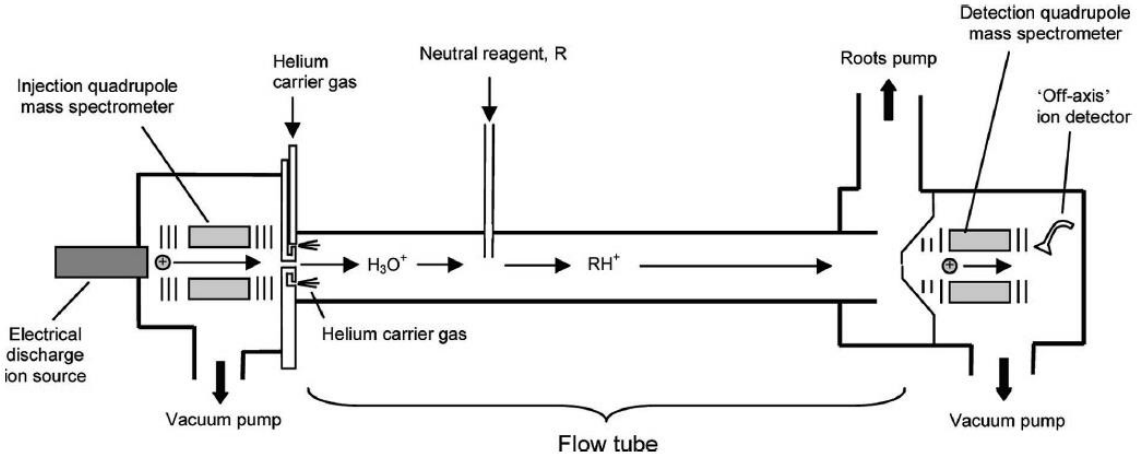

Figure 16: Schematic illustration of a selected ion flow tube (SIFT) instrument [170]

1.3.5.5 Ion mobility spectrometry

Ion mobility spectrometry (IMS) consists in subjecting ionized molecules to an electric field in a gas stream. The ions move according to the electric field at a rate which depends on their interaction with the gas, their mass, their size and their form. Hence, compounds are separated according to their mobility (Figure 17). IMS is a fast and sensitive method with low detection limits (ng to pg / L). It does not requiring preconcentration of the samples. This method is often coupled to one-dimensional chromatography to analysed exhaled breath samples [158, 172-180].

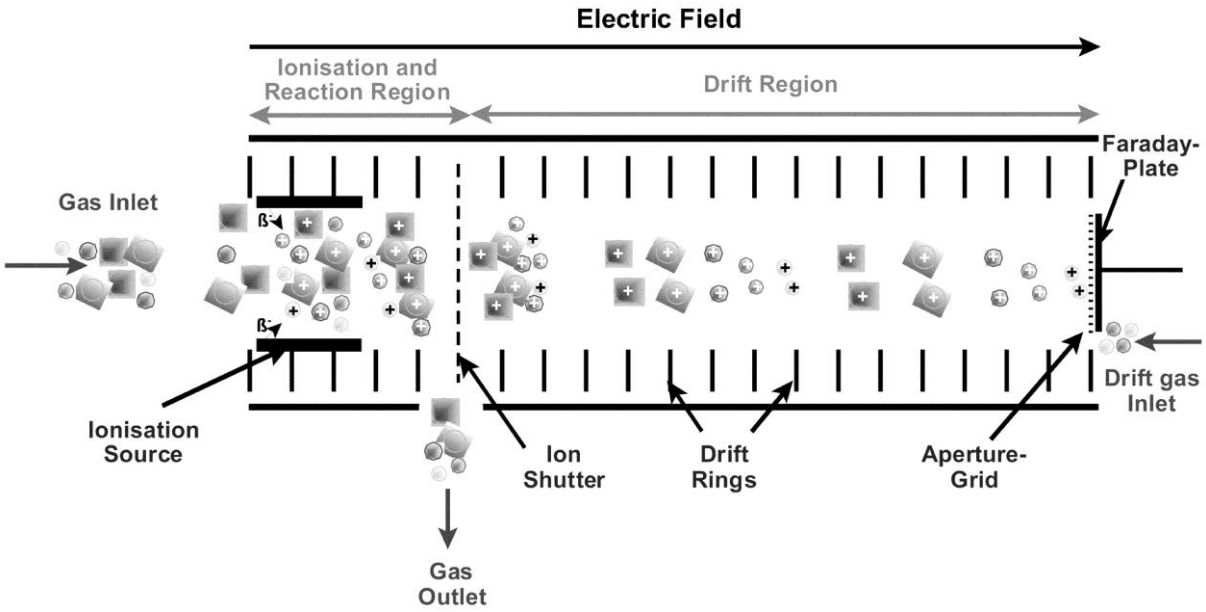

Figure 17: Schematic illustration of ion mobility spectrometry (IMS) instrument [181]

1.3.5.6 Training dogs

There is also a somewhat less conventional method currently being studied in order to analyse breath samples: the canine olfactory sense. Domestic dogs have been trained by humans for thousands of years. They have been selected and raised during all those years to help their masters in daily tasks and this has been made possible by their ability to learn and cooperate with humans.

Smell plays a fundamental role in most animals (finding food, detecting predators, identifying individuals, reproducing, communicating, etc.). Their olfactory sense plays a crucial role in most of the tasks where dogs are used. Indeed, dogs are commonly used today to detect various substances such as explosives, drugs, smuggled food, cigarettes, alcohol, but also search for individuals in the event of landslides, avalanches and even search for human remains. These dogs are also used for more unusual purposes such as the detection of mold or insect pests in buildings.

In 1989, Williams and Pembroke hypothesized that the dog might be able to detect cancer in humans through their sense of smell [182], observing a dog that had shown an interest in a person who turned out to have melanoma. After removal of this melanoma by surgery, the dog no longer showed an interest in the person. This hypothesis has given rise to interest on the part of the scientific community for the use of canine smell in cancer screening. Different laboratories have conducted studies on trained dogs to detect different cancers in various matrices (breath, urine, cancer tissue) but the results remain mixed [183]. However, the development of analytical methods and sensors will probably make the use of these dogs obsolete, but it is very important to note that the complexity of canine smell can inspire us in the development of sensors for the detection of these molecules [17, 184].

1.3.5.7 E-nose

Electronic noses (e-nose) are devices containing chemical or semiconductor sensors that will selectively respond to certain properties of analyses. These devices often include several different sensors that provide combined information. Indeed, an individual sensor is generally not selected for a particular compound. The device combines the information of all the sensors in order to obtain information on the chemical composition of the analyses. Therefore, e-nose used another approach to develop an early diagnosis of lung cancer. Instead of looking for VOC markers of lung cancer as in another method described above, e-nose rather tries to characterize

a pattern of the VOC profile emitted by individual [185].

These devices are generally portable and cheap, have a high sensitivity, perform a rapid analysis and allow trace molecules to be determined. It has a high sensitivity and a rapid return to equilibrium. However, the repeatability of this device is problematic. This device presents risks of contamination and problems of some sensors against water (loss of sensitivity) and too-high concentrations of a compound [186]. In addition, the response of some sensors may change over time and therefore have a short lifetime. The use of these devices can be found in many fields such as cosmetics, food, medicine, forensics and the environment [17, 184]. These devices measure several types of properties like optics, thermal pellistor, electrochemical (chemisistivity, potentiometry), gravimetric, etc. by different devices such as quartz microbalance sensors coated with different molecules, dyes, metal non-particles, polymer-based sensors [187, 188]. Gold nanoparticles e-nose is very common and two instruments, the TECNION and the Cyronose 320 were used in several studies [189].

1.3.6 VOC markers already reported for lung cancer

Currently, various markers of lung cancer have been highlighted in the literature. Since the study by Gordon et al in 1985, which has pointed out the possibility of developing a diagnosis of lung cancer on the basis of breath analysis [190], many laboratories have been working to create lists of lung cancer markers. The comparative table of the 207 markers of lung cancer published is presented in the annex part. To display the variety of putative VOC reported, they were classified according to their chemical function (Figure 18). An important fact to note is that each study highlights several putative markers and not only one. It is because single markers can usually not be accurate enough on their own to detect lung cancer. Instead, a combination of several markers can be used to detect the presence of cancer through exhaled breath [191].

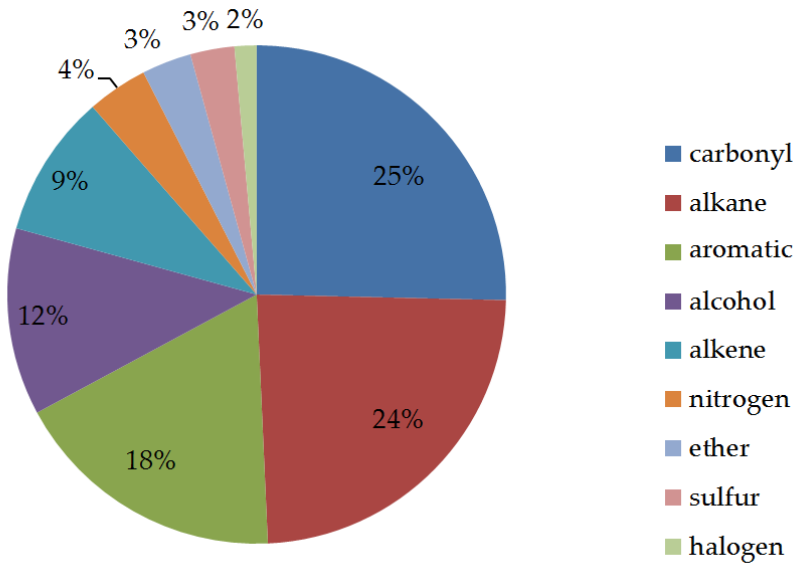

Figure 18: Chemical function abundance for VOC markers of lung cancer reported in the literature [85, 90, 99, 109, 117, 118, 124, 126, 128, 129, 135, 144, 148, 158, 187, 190, 192-209]

Aldehyde, carboxylic acid, ester and ketone are grouped together as carbonyl compounds. One compound could belong to several groups. For example, acetophenone was classified in aromatic and carbonyl compounds. As most VOCs contained an alkyl group, we decided to classify alkane compound as VOCs which only contain alkyl groups, like decane for example. On the other hand, hexanol only belongs to the alcohol group and was not been considered as an alkane compound. To illustrate this classification process, the list of biomarker reported in the literature present in the annex section was classified according to it. With this classification, alkane and carbonyl VOCs represent almost 50% of the chemical function of putative markers of lung cancer detected in exhaled breath. Aromatic compounds are well-represented and are putative compounds to help in the understanding of how the pathology evolves and how the body reacts to it. Indeed, the impact of the pathology will lead to different biochemical syntheses and therefore to a variation of the sub products emitted. Hence, the detection of such compounds in breath could also help clinicians to gain a better understanding of the biochemical process of the pathology. However, those sub-product compounds may undergo some reaction with other species before being released in the exhaled breath and lead to a more complicated interpretation of exhaled breath. The higher stability of aromatic compounds could allow them to avoid these reactions. Finally, halogen, sulfurous and nitrogen compounds represent a tiny part of the chemical function display in putative VOC markers of lung cancer found in exhaled breath.

1.4 Origin of VOCs

The link between VOCs found in exhaled breath and their origins is not clear. Several metabolic processes were proposed to explain the origin of some chemical family such as how the lipid peroxidation of polyunsaturated fatty acids (PUFA) from cell membranes induces the formation of different alkane and aldehyde (Figure 19) [78].

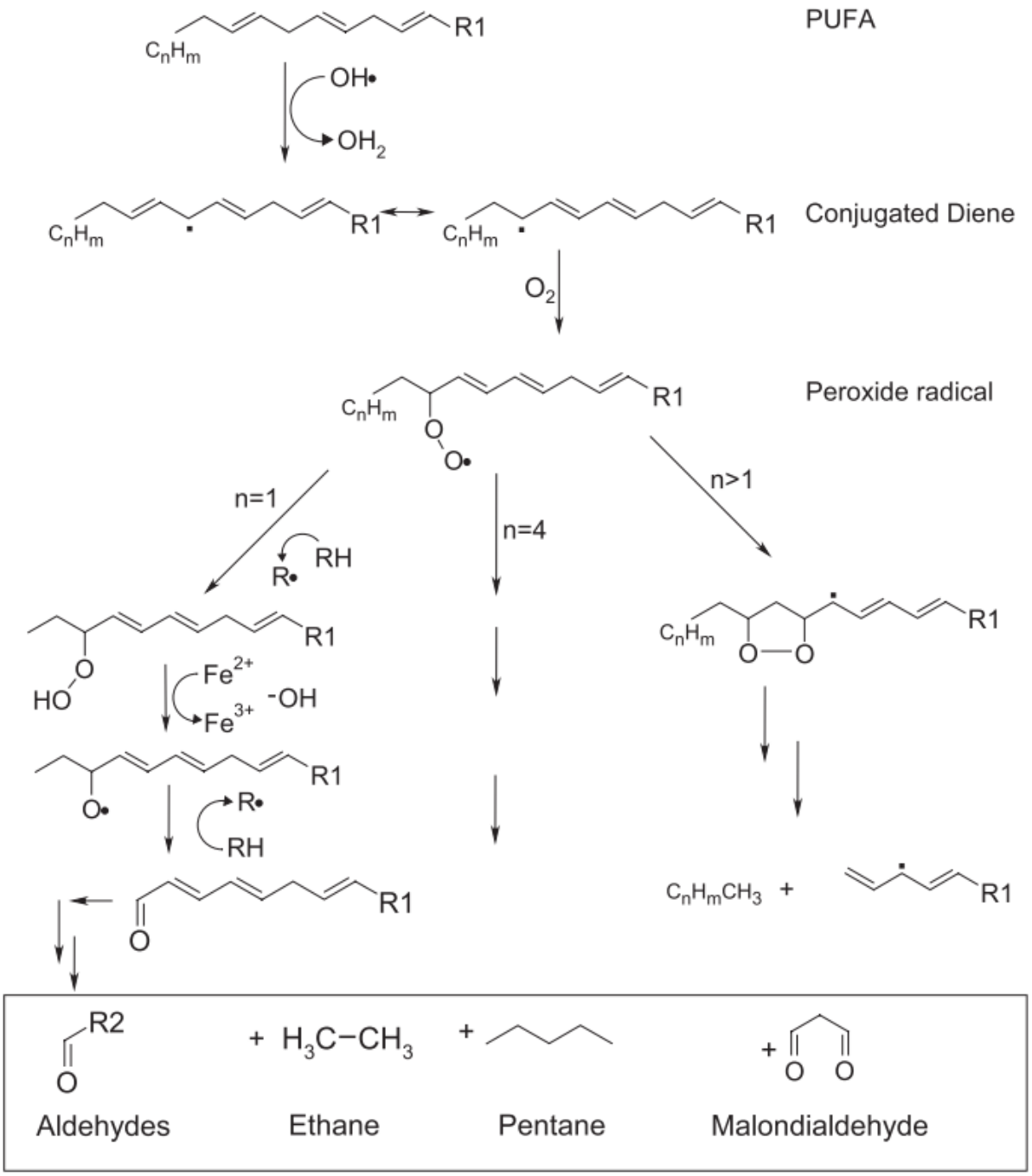

Figure 19: Metabolization of alkanes VOCs with lipid peroxidation of polyunsaturated fatty acids [91]

The production of alcohol compounds could be partially explained by the pyruvate metabolism via the glycolytic fermentation (Figure 20). Alcohol VOCs could also originate from some metabolization occurring in the liver and may react to form carbonyl compounds like aldehydes and ketones by lipid peroxidation.

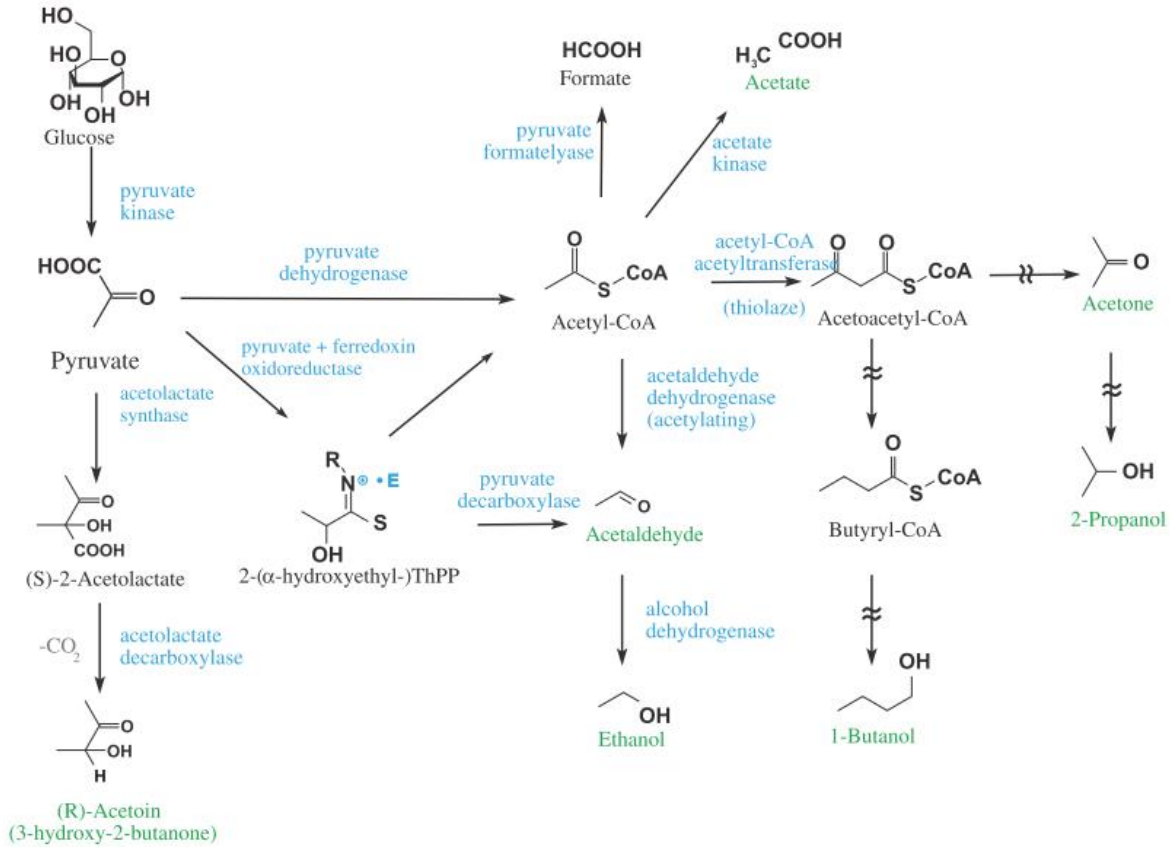

Figure 20: Production of alcohol VOCs during the glycolytic fermentation of pyruvate [210]

Sulphur compounds are produced during the incomplete metabolism of methionine and nitrogen compounds come from the transformation of urea [211, 212]. No biochemical pathways explain the origin of aromatic or halogenated compounds in the exhaled breath.

Nonetheless, there are several biochemical pathways to explain the origin of chemical function families. For example, many studies reported on the presence of isoprene in exhaled breath. It seems to be the case that this VOC could be formed during the synthesis of cholesterol (Figure 21).

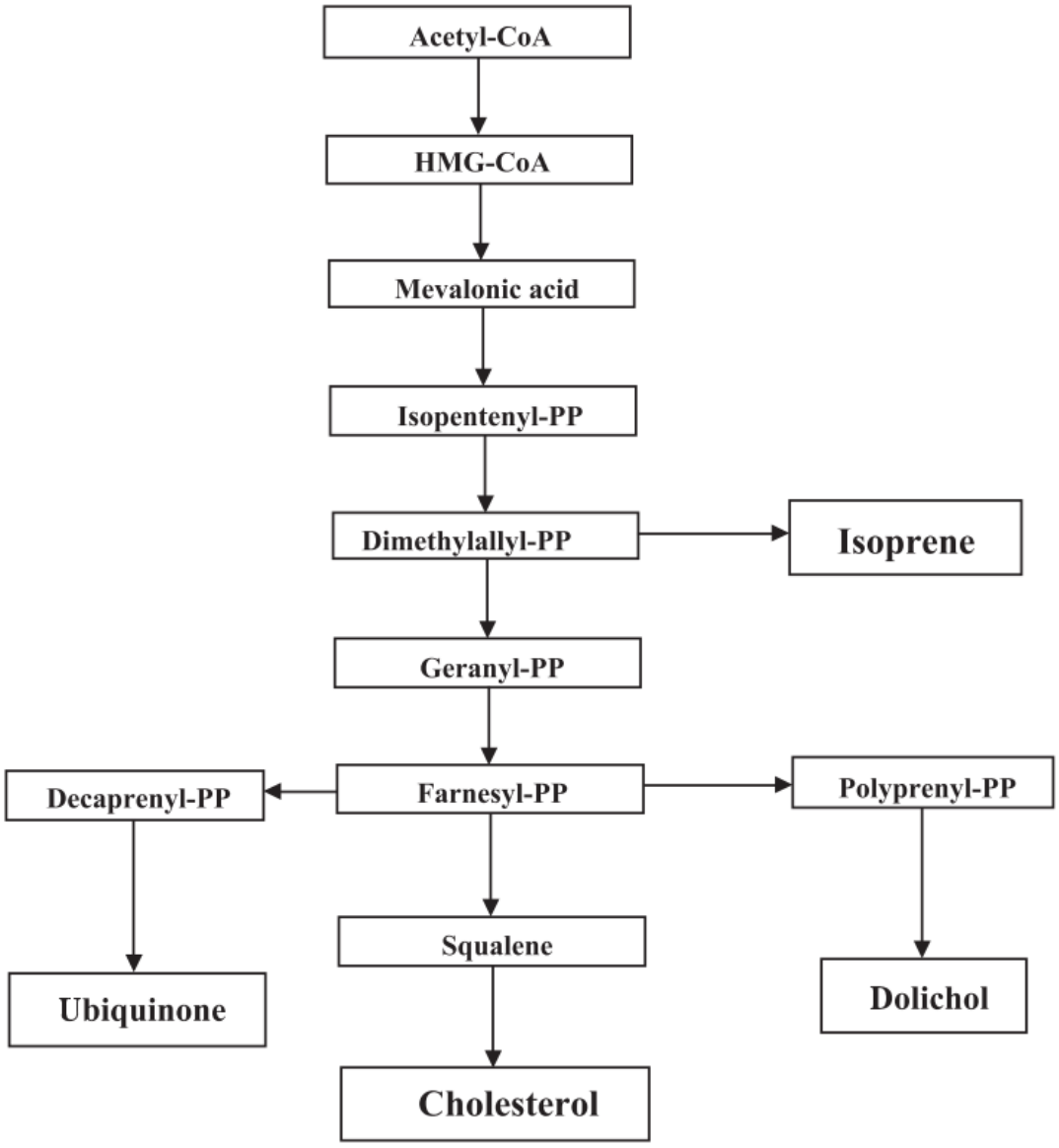

Figure 21: Formation of isoprene during the synthesis of cholesterol [91]

It is also possible to relate the presence of acetone in exhaled breath with the lipid peroxidation of acetoacetate occurring in the liver and the muscles. Indeed, decarboxylation of acetoacetate leads to the formation of acetone (Figure 22).

Although isoprene and acetone formation can be explained, it is not possible to explain the biochemical syntheses of all VOCs detected in exhaled breath. Discussions between physicians, biochemists and chemists could help in the discovery of new understanding of VOC origins. Moreover, cell culture analysis could also help in this direction as cell cultures are known to emit a large number of VOCs that could thus be studied in this context [213, 214].

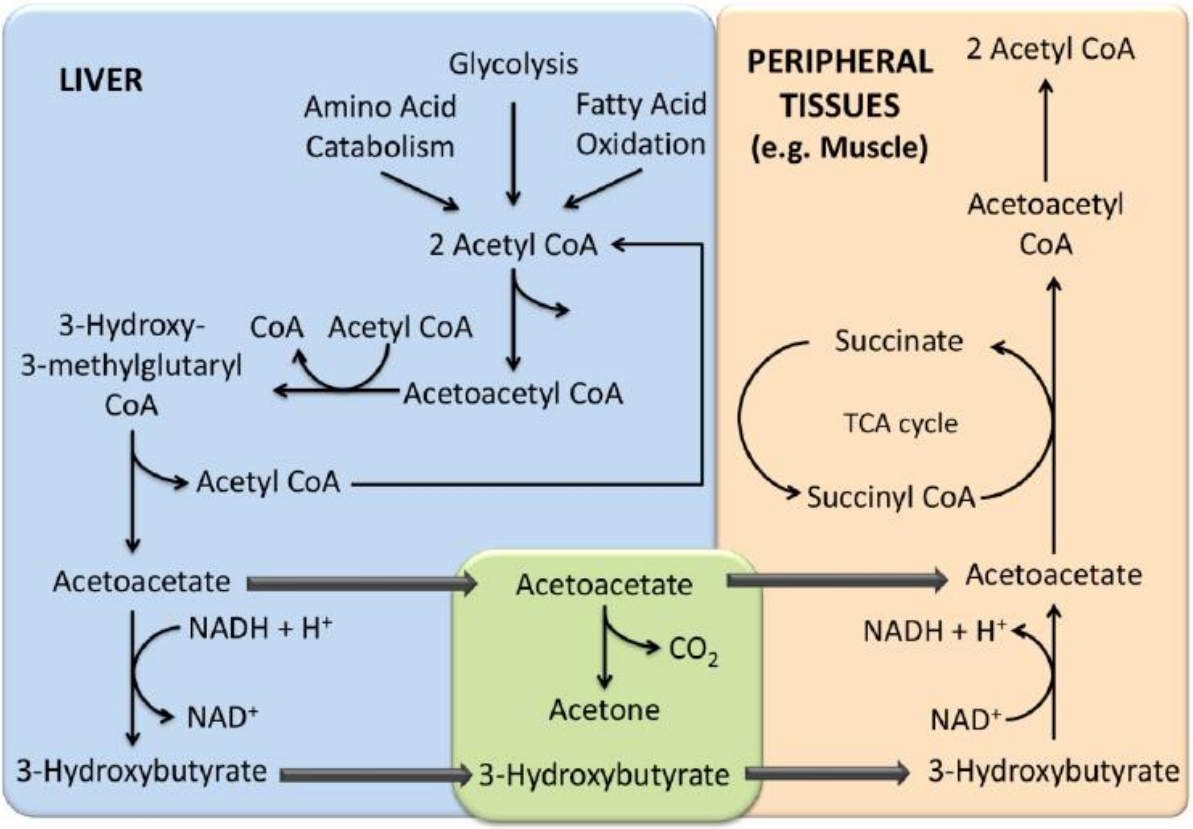

Figure 22: Formation of acetone during the lipid peroxidation of acetoacetate [215]

1.5 Impact of smoking on exhaled breath

The majority of lung cancer patients are or were smokers. Several studies have demonstrated that smoking induces a variation in the VOC profile of individuals. Hence, it is important to pay attention to the smoking factor when seeking for putative markers of lung cancer by being sure that the presence of such markers is due to the pathology and not to the smoking factor [216]. To overcome this problem, it is necessary to include non-smoking patients in a lung cancer cohort and also smokers in a healthy volunteer cohort. Several VOCs have already been reported to be found in smokers or ex-smokers (Table 2).

Table 2: Smoking VOCs markers reported in the literature

#	Name	Reference
1	1,3-Cyclohexadiene	[109]
2	acetaldehyde	[199, 200]
3	acetone	[109]
4	Acetonitrile	[90, 109]
5	Benzene	[199]
6	Furane 2 methyl	[141]
7	Furane 2,5 dimethyl	[109, 199]
8	o-xylene	[109]
9	Toluene	[109]

1.6 Goals and objective of the thesis

The aim of this thesis was to contribute to the development of an early and non-invasive diagnosis of lung cancer by using exhaled breath analysis. Two central lines were investigated. The first pillar consisted in the analysis of exhaled human breath. Investigation of the background level coming from the environment using sampling bags was reported. Then comparison of the exhaled breath of lung cancer patients and healthy volunteers was used to highlight putative markers of such disease. Different statistical tools like Fisher ratio and random forest approaches were applied and compared to help in the data treatment of such datasets. Results related to this pillar are reported in Chapter 2.

The second line focused on the analysis of the headspace of cell cultures. The goal was to compare the VOC profile emitted between different types of cancer cell lines and also between normal and cancer cell lines. Those results were also compared to the VOCs detected in the exhaled breath of individuals. Results related to this pillar are reported in Chapter 3.

Due to the large number of VOCs found in both types of samples, GC×GC-TOFMS techniques were used to overcome the limited peak capacity issues induced by 1D-GC-MS. An important part of the research was devoted to setting up an efficient data treatment approach. This enabled correct processing and interpretation of the complex sets of data collected by the multi-dimensional system from large classes of samples. The following part of this chapter describes in detail the different principles of methods and techniques used during this thesis.

1.7 Comprehensive two-dimensional gas chromatography

1.7.1 Multidimensional methods

It is in human nature to move forward and solve challenges. It is especially true in the field of science. Scientists have been confronted with problems in the separation of complex mixtures such as petrochemicals [217, 218], agro-industry [82, 219, 220] VOC food, perfumery [221, 222], breath [223-225], cigarette smoke [226, 227], forensic [228-230] and pollution [81]. Scientists have therefore developed the so-called multi-dimensional analysis techniques which consist in separating the compounds according to several criteria, and not according to a single one, as is the case with conventional methods. The aim of these techniques is to obtain a better separation power of the analytes and thus allow the separation of compounds which would

normally co-elute together [231]. For this, these methods are called exhaustive or even "comprehensive" and must follow the rules of Giddings [232].

1. All analytes must be separated by at least two dimensions that separate according to different criteria. These dimensions are called orthogonal.
2. All analytes separated in the first dimension must also be separated in the second dimension.
3. The separation obtained according to the previous dimensions must be maintained throughout the process.

The notion of orthogonality must also be introduced. This describes the dependence between the separation criteria used in different dimensions [233]. Thus, a technique that has two dimensions must have different separation criteria according to its dimensions to be orthogonal.

1.7.2 Principle of comprehensive two-dimensional gas chromatography

Comprehensive two-dimensional gas chromatography, also called GC×GC, is a separation method developed by J. Phillips and X. Liu in the early 1990s [234]. This technique is similar to conventional gas chromatography but differs in the fact that analytes are separated not by a single column, but by two. The aim is to obtain a separation based on two criteria to increase the peak capacity.

In practice, GC×GC looks like conventional GC, except that in this case there are two columns of different stationary phases which are coupled together in series through an interface called "the modulator" (Figure 23).

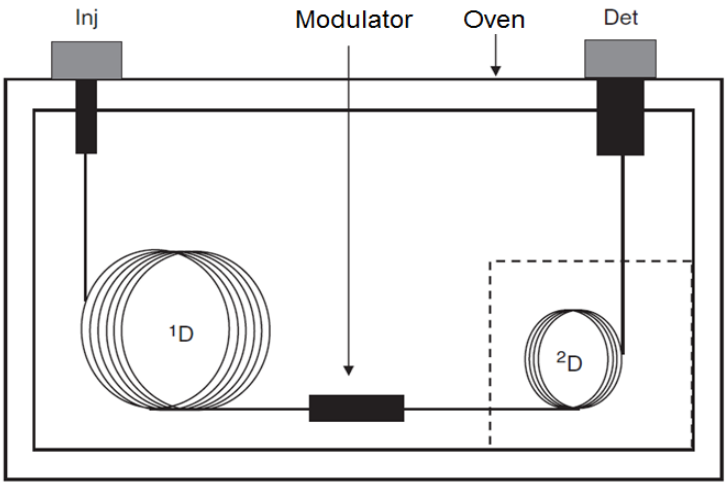

Figure 23: Schematic design of a GC×GC device. Dotted line represent the optional secondary oven for independent temperature control compare to the main oven [235]

The aim of GC×GC is to obtain a better chromatographic resolution by increasing the number of theoretical peaks that can be separated due to the second dimension. When analytes elute at the end of the first column, they are split in different fractions, refocus and reinject, one after the other, in the second column thanks to the modulator. Access to the second dimension leads to analysis of complex mixtures which in conventional 1D-GC would give rise to peak co-elution problems. Three-dimensional chromatograms are obtained, the XY plane of which represents the separations according to the two columns. The Z-axis gives it the intensity of the peaks. It is also common to visualize these data in the form of a two-dimensional chromatogram in which the intensity of the peaks is represented by contour curves, the XY plane always representing the separations according to the two dimensions. Using this multidimensional method, the separation obtained is not on one axis but on a plane (Figure 24). Finally, the modulation step also increases the sensitivity of the system. Before arriving at the modulator, analytes go through the first column, where they have enough time to be spread. Resulting from a certain width of the peaks of the analytes. Refocalisation and reinjection in a shorter column reduces the width of the peak, leading to higher intensities and a better sensitivity. A gain from two to five has been estimated from using GC×GC-TOFMS instead of 1D-GCMS [236].

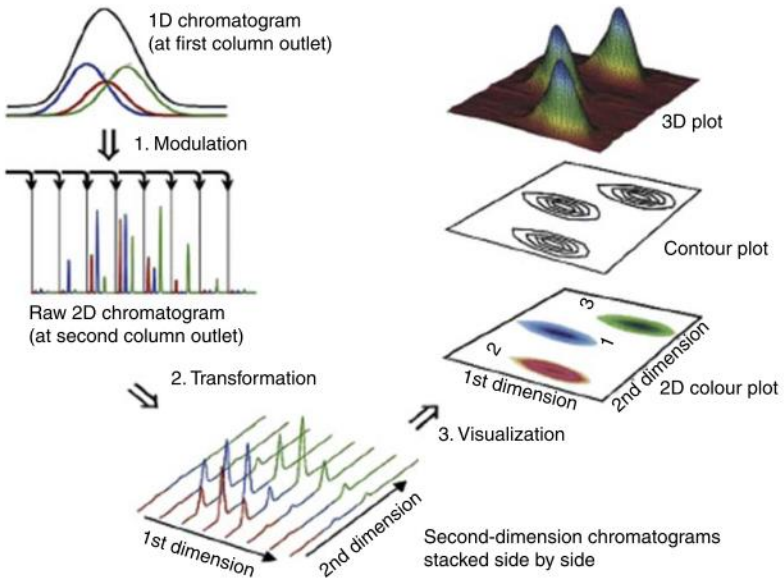

Figure 24**:** Diagram illustrating the modulation principle of GC×GC and peaks reconstruction on 2D chromatogram [237]

However, a high acquisition frequency is required because the analytes leaving the second column have a peak width of the order of 100-200 ms. As it is recognized that it is necessary to acquire at least six to ten data points to properly characterize an elution peak in chromatography [238, 239], it requires the use of a detector having an acquisition speed of at least 25 Hertz. In order to place themselves under more than comfortable analysis conditions, the use of detectors having acquisition speeds of about 100-200 Hertz is a conventional choice.

1.7.3 Modulators

Although the principle remains relatively simple in theory, this is unfortunately not the case in practice. The coupling between the two columns is a key point of the experimental set-up which must allow the analytes leaving the first column to be refocused before being injected into the second dimension while keeping the separation achieved in the first dimension. This is made possible by a device called a "modulator". Different types of modulators exist and all of them have their specific advantages and disadvantages [240-243]. They operate on different mechanisms but have the same role of collecting for a short time the analytes coming out of the first column, in order to refocus them and inject them into the second column. This relocation time is called the "modulation period" (P_M) and is chosen to insurance of at least three to four cuts per peak of the first dimension by the modulator. Usually, P_M is generally around two and eight seconds. If P_M is too long, the separation achieved in the first dimension is lost and rules of Giddings are not applicable anymore. On the other hand, if P_M is too short, new analytes are introduced in the second column when analytes from the previous P_M are still present in the column. This effect is called wraparound and increases the risk of co-elution. To prevent this effect to occur, it is required to use fast GC analysis in the second dimension. Thus, the second column has a smaller internal diameter and length than those of the first column. The purpose of this is to reduce the retention time of the analytes in the column and thus to prevent them from spreading out. In this way, the GC conditions in the second column are similar to fast GC analysis.

The most common modulators are the cryogenic systems. In practice, such modulators rapidly cool, using cryogenic liquid (usually liquid nitrogen), compounds which leave the column in order to immobilize them. Then, they reinject these analytes into the second dimension by means of a hot jet.

In this work, chromatographs are equipped with a cryogenic modulator. The one installed in the Pegasus 4D is a 4-jet thermal modulator (Figure 25). First, a cold jet blocks the compounds at the inlet of the second column. Then the first cold jet becomes a hot pulse to allow these analytes to get back into motion. The next step consists of a second cold jet to stop the analytes again. Finally, the first jet becomes cold to prevent the analytes leaving the first column from returning to mix with the compounds released during the previous modulation period. The latter are re-injected into the second dimension by means of a high-temperature pulse. The cycle is repeated throughout the chromatographic analysis and forms a modulation period.

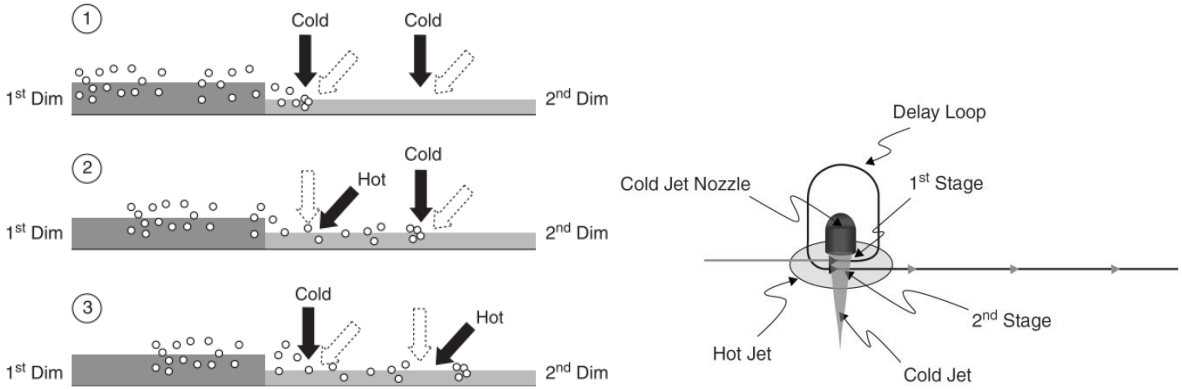

Figure 25: Principle of a four-jet modulator on the right and a loop modulator on the left [123, 244]

The AccuTOF™ GCv 4G from JEOL uses another cryogenic modulator called a "loop modulator". This modulator, designed by Zoex, is a dual-stage thermal modulator where hot and cold jets are used to trap, focus and reinject peaks from the first dimension to the second one. The idea is to use a loop, make with the column call the "delay lop", to use the same cold jet to stop the analyte at two different locations of the column instead of using two different cold jets like with the 4-jet thermal modulator (Figure 25). Peaks coming out of the first dimension are trapped with cold jet and then reinjected into the delay loop by means of a hot jet. A second cold jet, from the same cryogenic jet, trap peaks after moving through the delay loop. P_M ends with a second hot jet to reinject peaks into the second dimension. This jet also allows the next P_M to go through the delay loop. In this study, this modulator used the Peltier effect instead of liquid nitrogen to produce a cold jet. Trapping is less effective for very volatile compounds, starting at C_6, as the reached temperature is higher than with liquid nitrogen. On the other hand, it is cheaper to use and more environmentally friendly.

1.7.4 Injectors

Like with 1D-GC, there are different ways of injecting the sample into the column. Two methods were used during this study. Thermal desorption (TD) technique for breath analysis and headspace with solid phase microextraction (SPME) for cell culture analysis.

1.7.4.1 Thermal desorption tubes

According to the type of compounds analysed, sorbents are selected. In this study, a mix of carbopack® B and Tenax® GR were chosen. This sorbent combination was already reported in the literature for the VOCs analysis [245]. Tenax is a porous polymer of 2.6-diphenylphenol, synthetized with a high level of purity and is stable at high temperatures (up to 350° C). The porous volumes are able to properly trap VOC from $C_{5/6}$ to C_{20}. Thereby 30% of graphite are added to the polymer. Carbopack is an adsorbent based on graphitized black carbon, which can trap compounds from C_3 to C_{20}.

In practice, the sample goes through the tube by using a gas flow, making it possible to trap VOCs onto the sorbent(s) via absorption. Samples can be transported and stored before being thermally desorbed due to the stability of the VOCs inside the tubes. These advantages are needed as samples were collected in the University Hospital of Liège over long periods of time and were sent to the laboratory of Organic and Biological analytical chemistry laboratory to be analysed. Another advantage of TD tubes is their large loading capacity as compared to other techniques allowing more exhaustive analysis.

1.7.4.2 Solid phase microextraction

Three different kinds of fibres were used and investigated in this study to characterize the headspace of cell culture media: the 65 µm polydimethylsiloxane/Divinylbenzene (PDMS/DVB) fused silica fibre, the 100 µm polydimethylsiloxane (PDMS) fused silica fiber and the 50/30 µm Divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) StableFlex fiber. Those fibres were already reported in the literature for the analysis of the headspace of biological matrix [246, 247]. Optimization of their use is described in Chapter 3.

1.7.5 Detectors

All detectors used in conventional GC are not necessarily applicable to GC×GC, because they must have a high sampling rate, since the peaks in the second dimension are very narrow (50 to 600ms). At the first step of GC×GC field, flame ionization detectors (FID) and element-selective detectors were commonly used. Nowadays, mass spectrometer detectors are the most common choice, but new technologies of UV-visible detectors are also now available.

1.7.5.1 Mass spectrometers

Using mass spectrometers as detectors in GC×GC field is called GC×GC-MS and is the most common choice for VOC analysis. There are several types of mass spectrometers but generally time-of-flight (TOF) spectrometers are used as they have a high sampling rate, which is require to proper integrated the short peak elute from the second column [248].

Mass spectrometry used the mass-to-charge ratio of ions to detect them. Thereby, the first step in mass spectrometry is to create ions. There are different ways to ionize. Electron ionization (EI) was the method used during this study. This ionization method consists of a stream of thermionic electrons emitted from a resistively heated filament under vacuum. These electrons are accelerated and interact with gas phase molecules to form molecular ions.

$$M + e^- \rightarrow M^{\cdot +} + 2e^-$$

Electrons get energy of 70 electron volt (eV) and transfer between 10 and 20 eV to the molecule to form the cation. The excess of energy allows fragmentations of the ions and provides structural information.

1.7.5.2 Time-of-flight mass spectrometer

The first time-of-flight mass spectrometer was proposed in 1955 by Wiley-McLaren [249]. TOF analysers measure the time-of-flight of ions in free-field region under vacuum called the flight tube, after applying a difference of potential (Vs) to accelerate them. The potential energy (E_{el}) given by the electro field to the ions depends on their mass (m) and total charge (q=ze). Ions convert this potential energy into kinetic energy (E_k) (Equation 2). As ions have different mass (m), they will have different velocities (v) (Equation 3). Like ions flight the same pathway inside the

flight tube (L), it is possible to transform their speed into time (t) they need to go through the analyser (Equation 4) (Figure 26).

Thus, a relation between the time that ions take to reach the detector and their ratio mass on charge is demonstrated (Equation 5). Calibrations with standard like perfluorokerosene (PFK) or perfluorotributylamine (PFTBA) are performed to tune the system. Thus, it is possible to separate ions with different mass to charge ratio by letting them pass through the flight tube.

$$E_{el} = Ze\,V_s = q\,V_s = \frac{mv^2}{2} = E_K \qquad \text{Equation 2}$$

$$v = (2zeV_s/m)^{1/2} \qquad \text{Equation 3}$$

$$t = L/v \qquad \text{Equation 4}$$

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2eV_s}\right) \qquad \text{Equation 5}$$

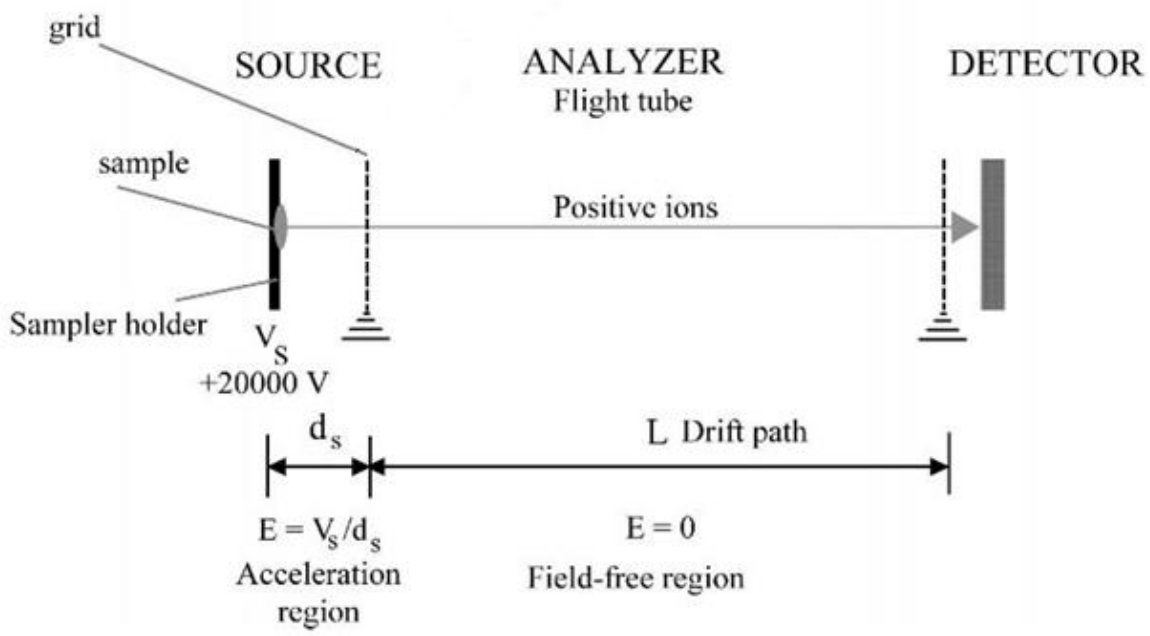

Figure 26: Diagram of Time-of-flight mass spectrometer [250]

TOF has several advantages. Its high transmission efficiency compared to scanning analysers helps to reach high sensitivity, which is important with samples with low concentration analytes as often the case with VOCs. But the main advantage is its high rate acquisition, necessary for GC×GC analysis.

Mass spectrum recorded with this analyser are compared to library to give attribution name to peaks. This attribution depends on the resolution (R) of the instrument. It can be calculated by the relationship between time of flight of ions and their ratio m/z (Equation 6). Derivation of such a relationship according to the mass and the time (Equation 7) leads to a new equation (Equation 8). Thus, the resolution can be calculated (Equation 9), where Δm is the peak widths, Δt is the time scale and Δx is the thickness of an ion packet approaching the detector.

$$\frac{m}{z} = \left({}^{2eV_s}/_{L^2} \right) t^2 \qquad \text{Equation 6}$$

$$\frac{1}{z} dm = 2t \left({}^{2eV_s}/_{L^2} \right) dt \qquad \text{Equation 7}$$

$$\frac{m}{dm} = \frac{t}{2dt} \qquad \text{Equation 8}$$

$$R = \frac{m}{\Delta m} = \frac{t}{2\Delta t} \approx \frac{L}{2\Delta x} \qquad \text{Equation 9}$$

Increasing the drift region length will increase the resolution of the instrument. However, it will also increase the loss of ions. What is reducing the sensitivity of the instrument. Usually, 1- or 2-meter length for the drift region is a good compromise to get good resolution and sensitivity. Use of reflectron and orthogonal acceleration also helped to achieve a better resolution.

A reflectron is an electrostatic reflector, which works like a mirror by creating a retarding field. This field deflects and sends ions back to the flight tube. Thus, ions with higher speed penetrate deeper inside the reflectron, which is increasing their

flight distance compared to ions with lower speed. The goal is to refocus ions with the same m/z which have different kinetic energy, due to energy dispersion, before being accelerated by the difference of potential. Therefore, ions with same m/z will reach the detector all together, increasing the resolution (Figure 27).

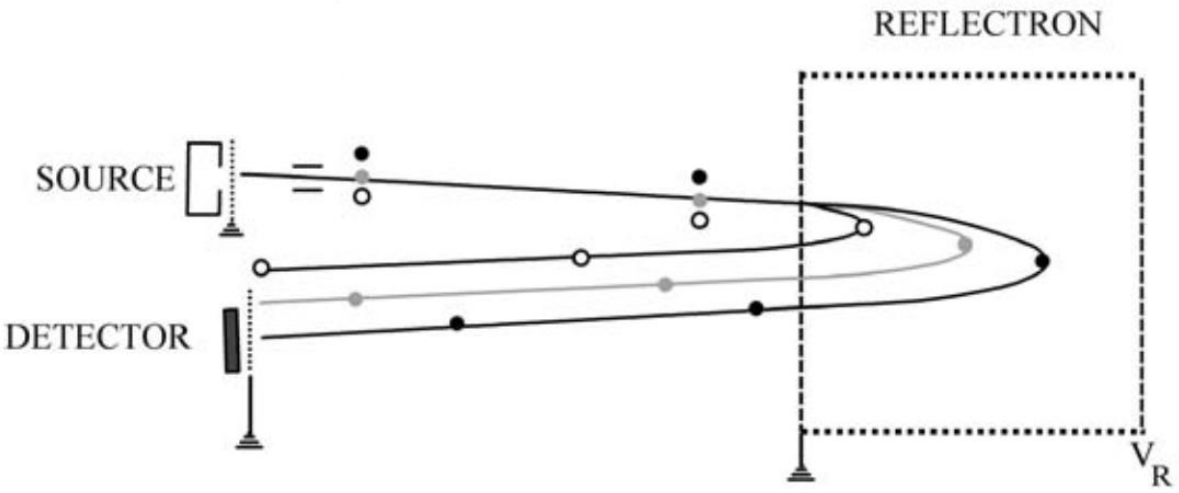

Figure 27: Scheme of time-of-flight mass spectrometer using reflectron [251]

Orthogonal acceleration makes it possible to send ions into the drift tube with a zero kinetic energy in the direction of the flight tube. Ion beams leave the ion source over a narrow slit, which only selects ions with a right direction. Other ions collide with the walls of the slit. Positioning the detector to an orthogonal direction compared to the direction of the beam of ions passing through the slit allows them to have a negligible speed compare to the direction of the flight tube (Figure 28).

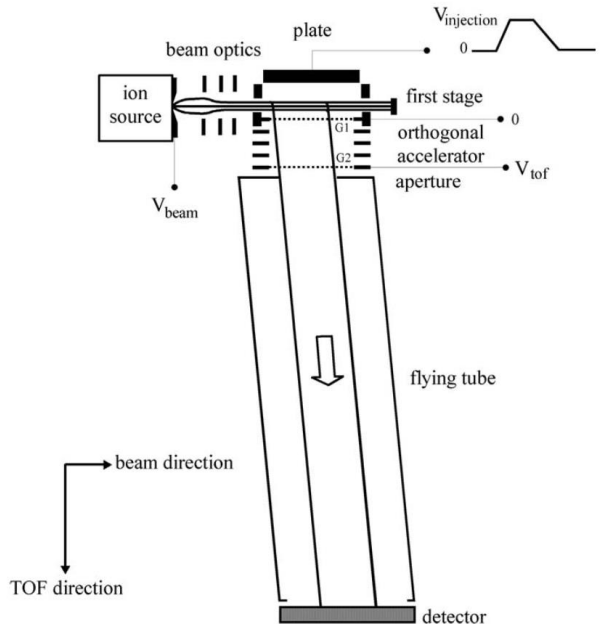

Figure 28: Diagram of time-of-flight mass spectrometer using orthogonal acceleration [251]

Deconvolution is an algorithm which can subtract mass spectrum from multiple mass spectra acquire in the same time. Thus, if several compounds reach the ion source at the same time despite the use of 2 GC columns, it is always possible to separate them according to their mass spectrum. This is particularly useful in non-targeted analysis as in the case of this study. Deconvolution is more easily applicable when proper mass spectra are recorded. Non-scanning instruments like TOF acquire mass spectrum without any shift intensity. Whereas scanning instrument, such as quadrupole-type analyzers, offer potentially skewed mass spectrum. Which gives more complicated access to deconvolution process.

1.8 Sampling breath for analysis

Exhaled breath samples were collected in CHU Liège and analysed in another building in the OBIAChem laboratory. Therefore, off-line analysis was preferred. Exhaled breaths of patients were collected in Tedlar® bags and afterwards transferred to TD tubes with a pump at an approximate flow of 300-400 mL/minute (Figure 29). Tubes were sent to the OBIAChem laboratory and stored at room temperature in the dark waiting to be injected.

Figure 29: Representation of the exhaled breath sampling with on the left the collection of breath in Tedlar® bag and on the right the transfer on TD tube via a pump

1.9 Cell culture

Cell culture corresponds to the maintenance outside the organism of dissociated cells which are not organelle in tissues but capable of dividing in vitro. It has several advantages. It is possible to control the surrounding environment by injecting specific compounds like hormones or controlling the pH. Homogenous populations are obtained, and it is possible to observe cells while they remain alive. Ethical issues are less important than directly working with the human body. On the other hand, cells are genetically unstable and may undergo mutations after several multiplications. *In vitro* environment cannot perfectly mimic the *in vivo* conditions and therefore induce a shift in the development of the cells.

1.9.1 Types of cell culture

First, tissues are removed from animals or plants. It is necessary to extract cells of interest from these tissues with enzymatic and mechanical processes. Then, it is possible to use cell a culture facility to allow these cells to grow. Cells used in this first culture are called primary culture. When cells occupy almost all the space dedicated for their proliferation, they are transferred to a new vessel. This subculture step turns primary culture into cell line nomenclature. Isolating cells from tissues is difficult to do. Thus, many cell line varieties are available on the market. All the cells investigated in this study come from cell lines.

The cell is the building block of living beings. They associate with each other to form tissues, which in turn form organs. Even though cells have a basic common structure, it is possible to classify them according to their principal function and morphology. In this study, only cells from epithelial tissue, call epithelial cells, were investigated to limit the variability between them and to easily highlight the differentiation in the data process.

1.9.2 Cell culture conditions

In practice, cells grow into specific flasks, treated with hydrophilic surface to facilitate the attachment of cells to it. Flasks are filled with culture media which sometimes have to be changed to keep a constant proliferation of the cells. Culture media contains everything cells need to develop. The principal component is water where several amino acids, vitamins, sugar, ions and oligo elements are dissolved. The culture medium used during this study was the Dulbecco's modified Eagle medium (DMEM) with GlutaMAX™ and is described in the annex section.

Serum is added to DMEM to complete the media. Serum usually comes from the soluble coagulated plasma of bovine foetal matter. It contains growing factor, inorganic elements, hormones, minerals and trace elements. Variation from one batch of serum to another is important. It is necessary to use the same serum during all experiments to keep the same culture condition for all cells.

The culture medium used for all cell culture experiments consists of DMEM with GlutaMAX™ and sodium pyruvate from Thermo Scientific (Belgium) and was supplemented with 10% of foetal bovine serum (FBS) and 1% of antibiotic (penicillin–streptomycin 10 000 U/10 000 mg/mL) from Lonza (Belgium).

The culture media also provides a suitable environment for the development of bacteria or other small organisms. Aseptisation methods are used to avoid theses contaminations. All cells manipulations are done under a hood with a laminar flow cleaned with ethanol and each item used must be sterilized.

At the beginning of the culture, cells are introduced into the flask filled with fresh DMEM to reach a surface (confluency) of 30%. Around two or three days are necessary to reach a confluency of 80% depending of the speed of growth of the cells. During this step, cell numbers have to be reduced to avoid problems induced by an overly high concentration of cells. Cells were removed from the flask. Firstly, the medium was removed with a pump while cells stayed attached to the surface. Cells were detached with the addition of trypsin during a short time to avoid auto digestion of the cellular membrane. Then, cells were transferred into fresh media to be separated by centrifuge. Cell pellets were recollected with fresh medium and were able to be used to start new cell culture. Usually at this step, some of the cells were preserved in liquid nitrogen. Addition on DMSO as cryo-conservator at 5% is necessary. This entire procedure is called “the passage” and has to be counted to compare similar cells as they may undergo mutations.

Cell development is observed with a microscope. It also helps to know when cells reach the confluency. As confluency is just based on visual information, one tiny part of the cells obtained in the pellet during the passage were counted by using blue trypan coloration. This compound only colorized dead cells, this method worked in two steps. Firstly, 50µL of solution were mixed with this colorant and analysed by UV method to determine the concentration of dead cells. Then, 50 µL of new solution were also mixed with trypan blue and trypsine to induce cellular death. Results obtain from the UV analysis made it possible to determine the number of cells by subtracting the concentration found with the first step.

1.10 Statistical tools

Breath analysis or cell culture headspace analysis by GC×GC-TOFMS generates a large amount of data. With the objective of highlighting putative markers of lung cancer, essential information had to be extracted from the initial data set. Usually, multivariate analysis can be applied to extract relevant information. Unsupervised methods including principal component analysis (PCA) and hierarchical cluster analysis (HCA) was the first step in data analysis. The goal was to visualize trends grouping and outlier detection. Then other statistical tools can be used to find relevant information.

1.10.1 Principal component analysis

This multivariate statistical tool enables two or three-dimensional visualization of the data using main component axes, which represent a linear combination of the dispersions of the different variables analysed in such way as to express most of the variance within the data set. We went from an "n" dimension system that was not possible to visualize to a three- or two-dimensional system that attempted to best reflect the reality of the initial data. There are two ways to use PCA analysis. When samples are separated in accordance with principal component axes, it is called score plot analysis and makes it possible to visualize sample groups and trends. But it is also possible to display a loading plot where all variables are presented and scattered according to their own influence in the separation of samples. This representation makes it possible to spot which variables play a significant role in the separation of samples and which do not.

1.10.2 Hierarchical cluster analysis

This hierarchical classification displays samples in a dendrogram where they are separated according to a calculated distance between them. Ward's Method consists of classifying each sample in a different cluster and then grouping the two clusters with the lower variability together. Then the process is repeated until we get a single cluster. By deciding the similarity cut-off, which divides the dendrogram into separate clusters, HCA can be used for classification. By comparison with PCA, no overlapping of groups can occur. However, the main disadvantage of HCA is that it does not deliver the information about which variables are responsible for the separation between samples as PCA does with the loading plot.

1.10.3 Date reduction with Fisher ratio approach

Normally, the number of samples (n) must be five times higher than the number of variables (p) when statistical tools are used. As the number of samples collected is low and the number of features detected is high, it is not possible to satisfy this condition. However, it is possible to reduce the initial data set to reduce the data over-fitting issues by using Fisher ratio [252]. This statistical tool calculates the variation in samples belonging to one group (Equation 10). Where $\overline{x}_i$ is the mean of the ith class, $\overline{x}$ is the global mean, n_i is the number of measurements in the ith class and finally k is the number of classes.

$$\sigma_{cl}^2 = \frac{\sum(\overline{x}_i - \overline{x})^2 n_i}{(k-1)} \qquad \text{Equation 10}$$

Then, the variation between the variable across all samples is also calculated (Equation 11), where $\overline{x_{ij}}$ is the ith measurement of the jth class and N is the total number of samples.

$$\sigma_{err}^2 = \frac{\sum\left(\sum\left(\overline{x_{ij}} - \overline{x}\right)^2\right) - \left(\sum(\overline{x}_i - \overline{x})^2 n_i\right)}{(N-k))} \qquad \text{Equation 11}$$

Finally, the ratio between the two terms gives the FR value (Equation 12).

$$FR = \frac{\sigma_{cl}^2}{\sigma_{err}^2} \qquad \text{Equation 12}$$

The higher the Fisher ratio value is, the higher is the specificity of the feature for one class. To determine if a feature is considered as specific to one class, it is necessary to compare its FR value to the critical FR (F_{crit}). This F_{crit} depends on the number of class and samples per class (i.e.: degree of freedom) and also on the level of significance (α). All features with a FR value greater than F_{crit} are declared to as statistically significant for one class at the level of significance selected. An α value of 0.05 is usually used. After the Fisher ratio selection, between 5 to 20 % of features remain. This reduced list can be screened feature by feature to highlight putative markers. Box plots are calculated for each feature and show the distribution of the intensity of samples from a same group. If it is possible to observe significant difference between

groups, a feature is kept and considered as a putative marker. The advantage of box-plot is that it is based on the medium compared to the Fisher ratio. Leading to a reduction in the impact of outliers on the data set. It does not depend on the number of group samples and displays direct results. However, calculated box-plot for each feature is time consuming and explains why it is used on the reduced data set and not immediately on the initial feature table.

1.10.4 Date reduction with random forest approach

Random forest (RF) is a multivariate machine-learning approach build on the decision tree approach. A decision tree consists in describing how to classify a dataset in a homogenous class by using discriminant parameters. At each node of the tree, a different parameter is used to classify the dataset. Based on the quality of the classification, it is possible to judge how effective the role of the parameter is to discriminate the dataset into homogeneous classes. These decision trees are calculated automatically with an algorithm on large datasets, where they can extract the most relevant parameters to classify the population. With RF, multiple decision trees are created and merged together to obtain a more accurate and stable prediction called a forest. After the construction of the classification trees, the variables were ranked according to their importance and their effect on the classification accuracy. Thus, it is possible to extract significant features [253].

Chapter 2:

Analysis of exhaled breath from lung cancer patients and healthy volunteers by TD-GC×GC-TOFMS

This chapter discusses the results obtained by means of exhaled breath analysis. The aim was to highlight markers of lung cancer capable of segregating healthy volunteers and lung cancer patients. This study investigates the impact of the background level coming from the environment using breath sampling bags. The contaminations induced by the bags' membrane and environmental migration during the sampling were investigated. Next, a TD-GC×GC-TOFMS methodology has been used to characterize and compare human breath from 15 lung cancer patients to 14 healthy volunteers. On the processing side, Fisher ratio and random forest approaches were applied and compared with regard to their ability to reduce the data dimensionality and to extract significant information. All experimental parameters and data processing information related to this chapter can be found in the appendix.

2.1 Repeatability of breath sampling

Results linked to the detection of VOC markers of lung cancer by mean of breath analysis are not repeatable from one publication to another, even when same laboratories are involved. The origin of this issue is partly due to the fact that the essential chemical information is hidden under massive amounts of irrelevant signals that make the isolation of putative markers of disease from breath a real analytical challenge. Indeed, such irrelevant signals are made of significant amounts of endogenous VOCs issued from the basic metabolism of the individual and exogenous VOCs related to factors such as food habits, hygiene, tobacco consumption, and ambient air [70, 111, 129] . Access to samples is limited and it is very difficult to get several samples from a same patient in order to replicate analyses to estimate the variability inside such replicates. However, it is easy to get exhaled breath from healthy volunteers. Thus, the first step of this chapter was to demonstrate that exhaled air sampled from the same people over short sampling time, to avoid variation of breath, show low variations. In practice, 5 individuals blew twice into five-litre Tedlar® bags. A pump with an approximate flow of 300 mL/min was used to transfer VOCs from the bag to the TD tube. All samples collected in this study using Tedlar® bags were transferred with the same pump. Then, the samples were injected into the AccuTOF 4G and processed with GC-Image (see annex part for injection parameter). The feature table obtained from the cumulative image of all 2D-chromatograms was used to display a dendrogram (Figure 30). Samples from the same origin showed low dispersion, indicating that variations observed in the literature do not come from the non-repeatable issues

related to the collection of exhaled breath. Thereby, it was possible to use one single sample per patient to discover putative markers from breath analysis.

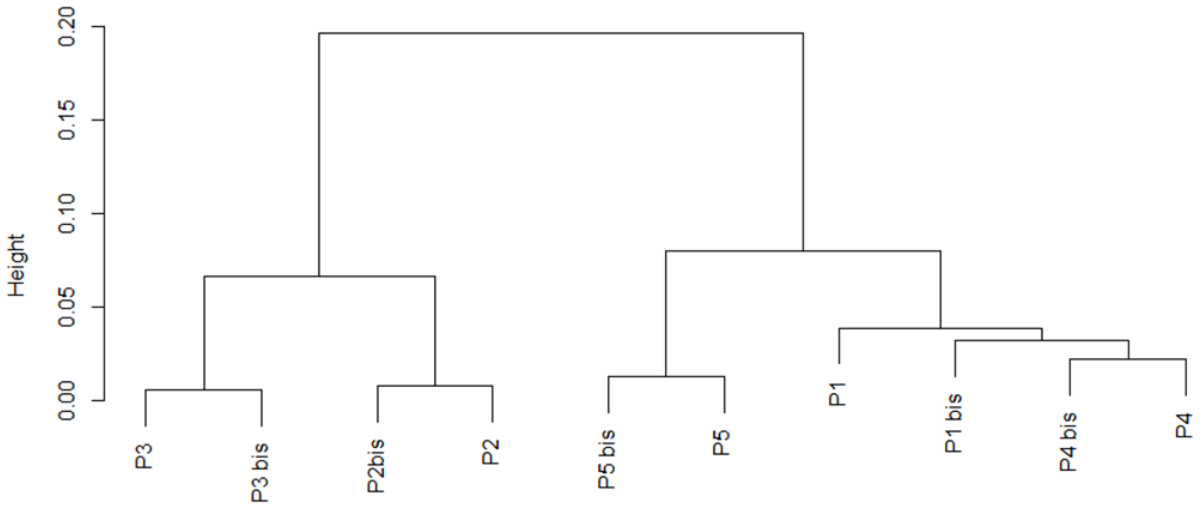

Figure 30: Dendrogram performed with a duplicate breath sample from five patients (P1, P2, P3, P4 and P5) to show the repeatability of samples record by breath analysis

2.2 Tedlar® bag investigation

The second part of this chapter investigated the sampling capacity of Tedlar® bags. Although these bags are easy to use and represent the most common choice for trapping exhaled breath, they are known to possibly be the source of issues in terms of cross contamination, leaching (phenol and acetamide N,N dimethyl) and leaking [114, 254]. A specific exhaustive search of such features and other possible contamination analytes to control and decrease the impact of the use of Tedlar® bags on air sample integrity was conducted. Therefore, the first part of this study was dedicated to the characterization of the background VOC noise of new (unconditioned) Tedlar® bags to know if the membrane was free of VOC contaminants which may alter the integrity of the gas sample. These bags were filled with nitrogen (purity of 99,999%) and kept at room temperature for two hours before using a pump to collect the gas from the bag to the thermal desorption tube.

As observed in Figure 31, phenol and acetamide N,N dimethyl, which are commonly cited in the literature to be leaching from Tedlar® bag, were the 2 major features found in the 2D-chromatogram. However, with GC×GC analysis, it was also possible to detect other features and this lead to a better characterization of the background VOC noise emitted by Tedlar® bags. In the three bags investigated, 2D-

chrommatograms contained an average number of features of 426. It is important to notice that all features detected in these chromatograms do not necessarily come from the Tedlar® bag but could also come from the nitrogen, tubes, the pump and different GC instrumental parts. However, a cleaning step was necessary to reduce the intensity of the background VOC noise emitted by Tedlar® bags.

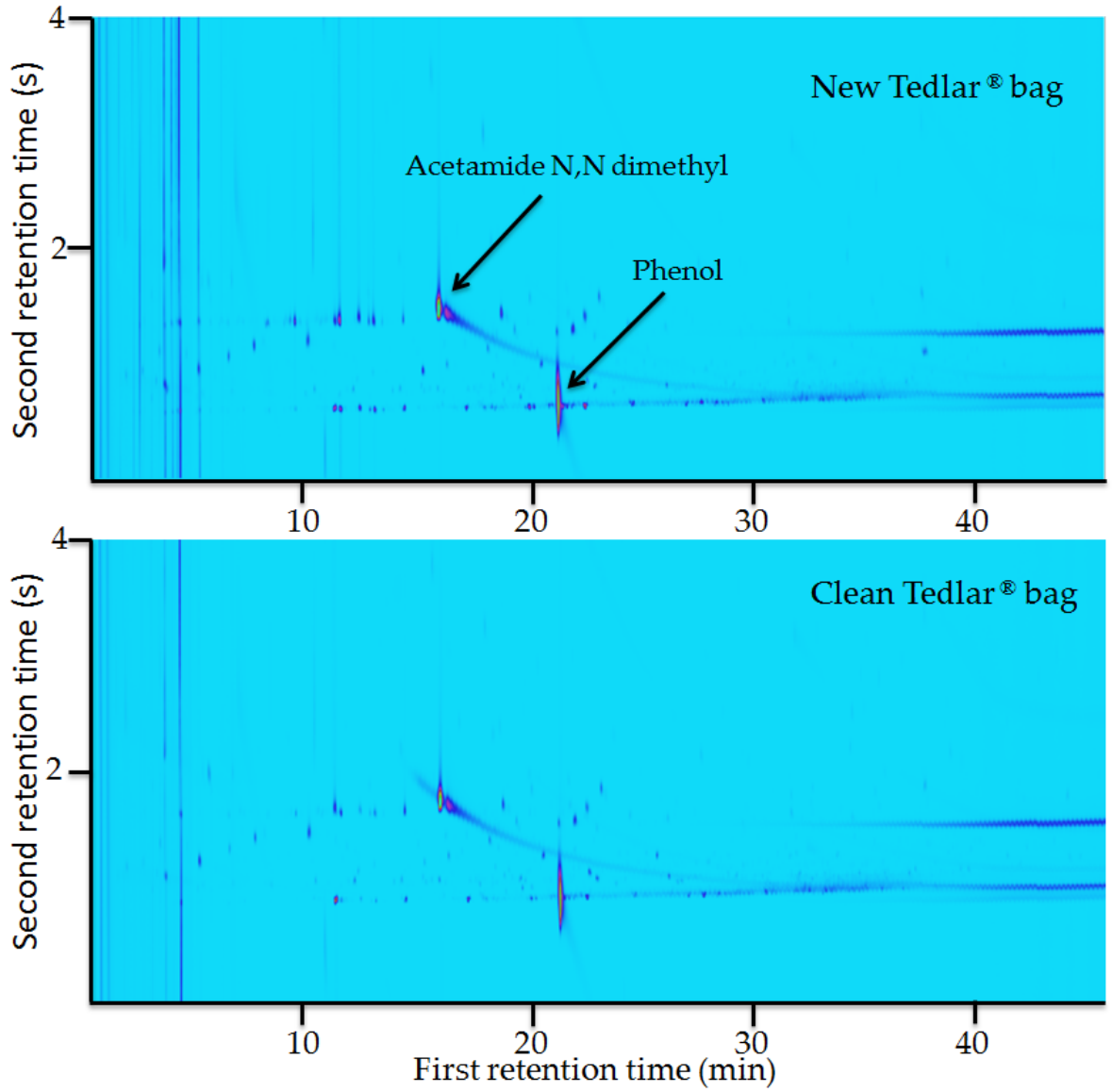

Figure 31: 2D chromatograms of new and clean one-liter Tedlar® bags filled with nitrogen. Clean bags display lower intensity features than new bags

The cleaning step consisted of filling bags with nitrogen, waiting around fifteen minutes, and then deflating them. This procedure was repeated twice to achieve a better cleaning process. To investigate the impact of such cleaning protocol, three other new Tedlar® bags, from the same pack of those used to characterize Tedlar® bag background noise, followed the double-cleaning with nitrogen. They were filled

again with nitrogen and kept at room temperature for two hours. Gas phases were trapped on the TD tube and injected onto the AccuTOF™ GCv 4G as for the previous study. The comparison between uncleaned and double-cleaned bags revealed a decrease in the intensity of contaminant features (Figure 31). Most intense compounds found in bags are reported in Figure 32. This analysis shows that the Tedlar® bag is leaching and releases more compounds than reported in the literature. These compounds need to be taken into account in the data process because they change the integrity of gas samples.

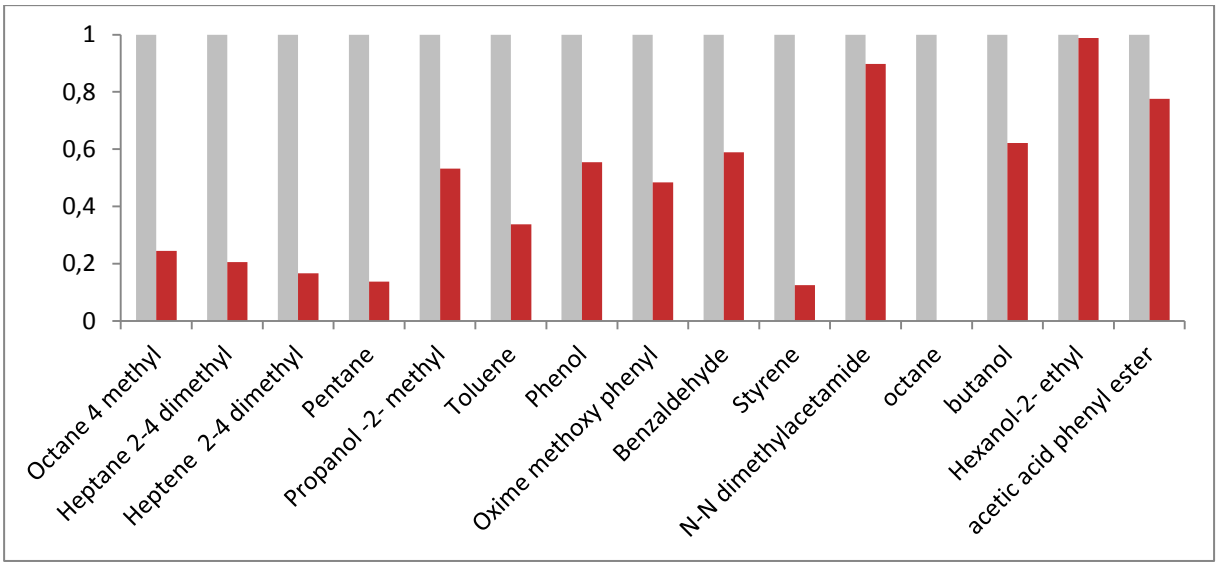

Figure 32: Comparison of the relative intensity of the major features found in new (grey) and twice-cleaned (red) Tedlar® Bags

The second part of this study focused on the permeability of the membrane of Tedlar® bags. It has been reported that the concentration of compounds trapped inside Tedlar® bags decreases over time, underlying limitations of the permeability of the membrane and limited storage potential [114] (Figure 33). However, as far as we know, no studies have yet investigated the resistance of the bag membrane against outside environmental contaminations. To evaluate this effect, 12 Tedlar® bags filled with nitrogen were placed in a box in which the atmosphere was saturated in toluene, methanol, hexane, and dichloromethane. Every two hours, three bags were pulled out of the box and deflated on TD tubes to be analysed. The kinetic study illustrated in Figure 34 shows how relative intensities of solvent peaks increase over time. This time-trend study shows that the relative intensity of each solvent inside the bags increased in accordance with the exposure time. The membrane of Tedlar® bags is thus also prone to permeation of chemicals from the environment to the bag. It can be concluded that the amount of time the sampled breath is kept inside the bags should be kept to a minimum to ensure low impact of the sampling procedure

on sample integrity. Based on the result obtained in this study, no major concentration increase of outside environmental contaminations compounds was detected in the breath sample for a storage time of maximum 8 hours. Furthermore, storage conditions should carefully be described in studies using such bags.

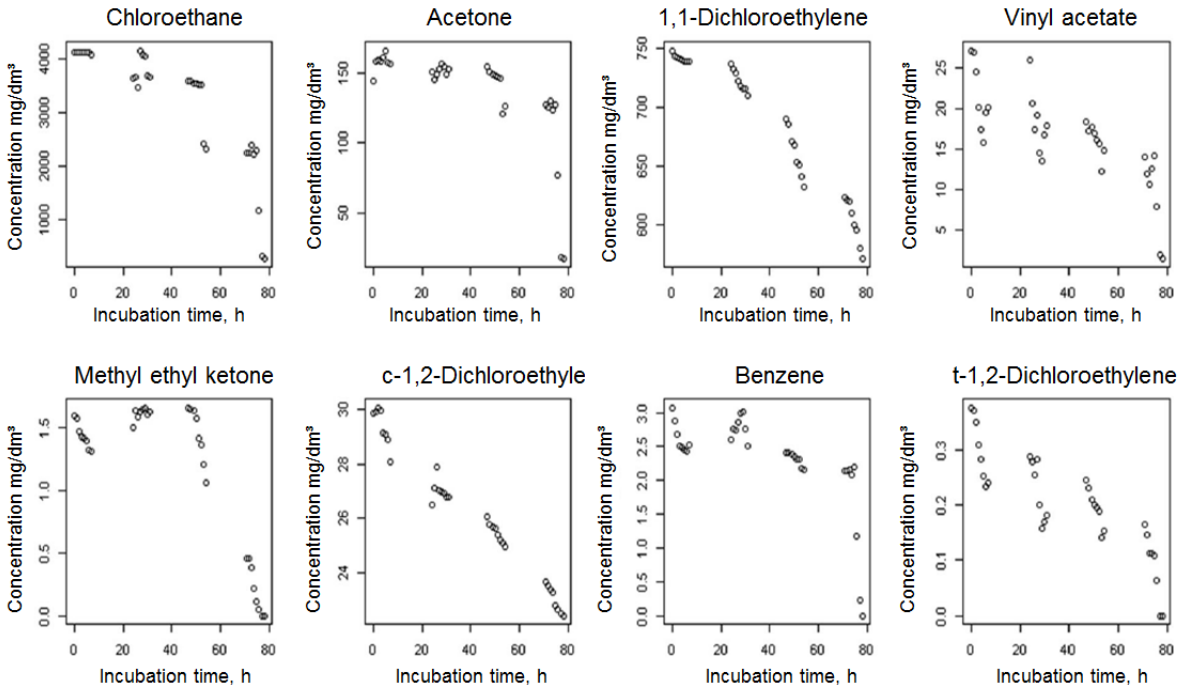

Figure 33: Changes in the concentration of compounds during 78h of incubation in Tedlar® bags [114]

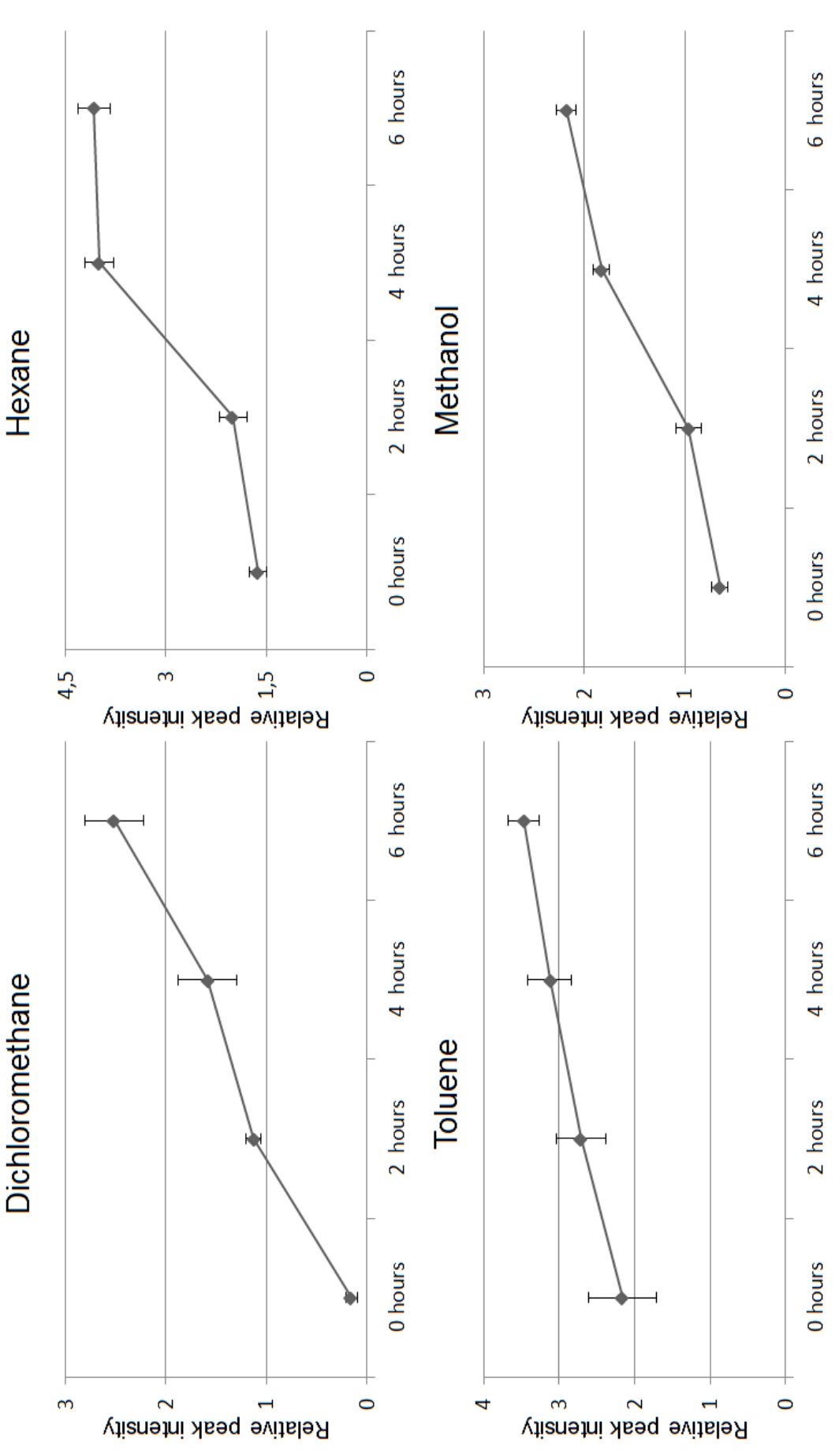

Figure 34: Kinetic study of the permeability of Tedlar® bag membranes for different solvents: Toluene, Hexane, methanol and Dichloromethane

With these results about Tedlar® bags, we demonstrated that it is possible to pursue the goal of trapping exhaled breath in such dedicated bags for VOC measurements. However, they have to be cleaned with nitrogen before being used to reduce background noise and technical staff has to transfer samples onto the TD tube as fast as possible to ensure its integrity. Therefore, breath samples collected in this study were transferred from the bag to the TD tube on the same day of sampling to avoid such sample alterations.

2.3 Lung cancer patient breath analysis

2.3.1 Exhaled breath samples collection

The rest of this chapter was dedicated to the analysis and comparison of exhaled breath from healthy volunteers and lung cancer patient to spot specific putative markers of such disease. The exhaled breath of a total of 29 individuals including 15 lung cancer patients and 14 healthy volunteers was investigated. All subjects were at least 18 years old and they all signed an informed consent to participate in this study after being informed about its goals. The study was approved by the Ethics Committee of the University of Liège (BECT B707201420493) and conducted in conformity with the Declaration of Helsinki. Patients with abnormal chest X-rays who were scheduled for bronchoscopy, in addition to age and gender matching controls, were sampled using Tedlar® bags of 5 L at the pneumology unit of the university hospital of Liège, in Belgium, between January and May 2014 in a series of three sampling campaigns (January, March and May). Each healthy volunteer exhaled breath samples were collected in the same time than the exhaled breath of a lung cancer patient and came from the individual who accompanied the patient, usually his wife and her husband. This procedure helped us to decrease the variability between the patient and the healthy volunteers as we assume that both of them were subject to similar contamination from exogenous VOCs. None of the patients had received any form of anticancer therapy or medication before the sampling. Prior to the sampling, Tedlar® bags were flushed twice with nitrogen (purity >99.99%, Air liquid, BE) to decrease residual contaminants. Characteristics of the study population are reported in Table 3. Exhaled breath was transferred from the bag to a sorbent tube containing Tenax GR and Carbopack B (Markes International Ltd, UK) with a pump at a flow rate of 300 mL/min directly after the sampling to avoid alteration of the samples [80, 245].

Table 3: Information about the set of patients

	Lung cancer patients	Healthy volunteers
Number (%)	15 (52%)	14 (48%)
Age (year), mean (± SD)	62 ± 7	58 ± 11
Gender (M/F)	12/3	9/5
Smoker/ Ex-smoker/ Non-smoker	3/12/0	1/6/7
Sampling period (January/March/May)	8/4/3	8/3/3

2.3.2 Utilization of GC×GC-TOFMS for breath analysis

The eight pairs of breath samples injected in January 2014 were used to corroborate the use of gas chromatography to compare two populations of samples and discover which compounds are responsible for such segregation. After injections into the Pegasus 4D, data was processed in ChromaTOF® software and different signal to noise ratios (S/N) were used to determine which one allows for a better characterization of breath samples. A S/N value of 100 has been selected. This choice explained below in the section 2.3.6.1.

Finally, these 16 samples were also used to show the advantage of using GC×GC-TOFMS instead of 1D-GCMS. 2D-chromatograms reported an average of 1078 features. Although all features are not related to one VOC because of chromatographic artefact and columns bleeding, GC×GC offers a better characterization of human breath than 1D-GC. GC×GC-TOFMS ensures lower co-elution issues, leading to better feature detection (Figure 35). A P_M of 4 seconds is a wise choice as most of the features have a second retention time range from 0 to 3.5 seconds. We could conclude that the experimental aspects used to analyse breath samples and the method of data handling generated was suitable to achieve the aim of this study.

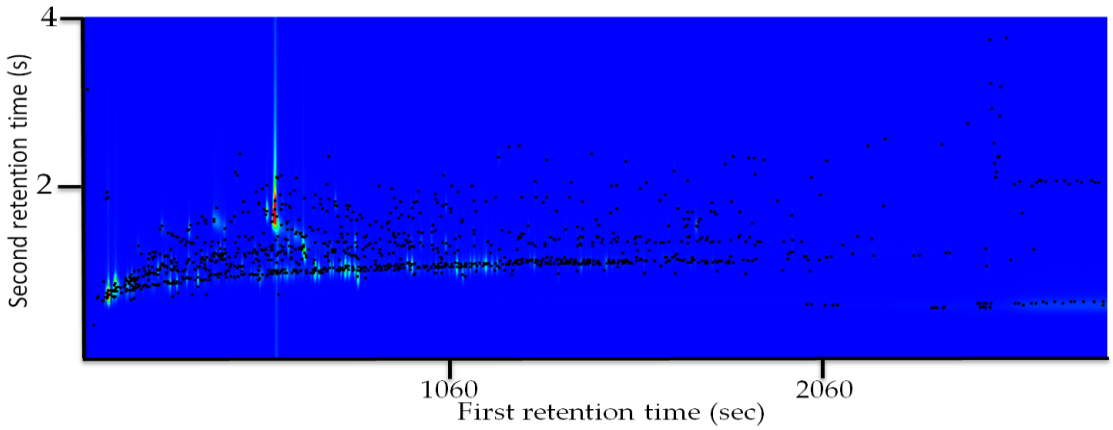

Figure 35: 2D-Chromatogram from the exhaled breath of a lung cancer patient. Each feature is represented by a black square

2.3.3 Analysis of breath samples

Then, other samples of the first set were injected in March and May 2014. A total of 29 samples were used to develop a reliable way to handle data treatment and extract a putative list of markers of lung cancer. The data generated was also imported into ChromaTOF® software for feature detection, mass deconvolution, integration features and library name attribution. As previously mentioned, S/N of 100 was used. The feature table generated by means of the statistical comparison feature of ChromaTOF® contained a total of 1350 features. A non-supervised PCA was performed based on this data set and the resulting plot can be seen in Figure 36. The visualization of such an unsupervised processing showed a clustering of the data according to three apparent batches, each of which appeared to be related to a sampling period (January, March and April) independently of the nature of the samples (patients and controls). This phenomenon demonstrated that the influence of the presence of various levels of background of exogenous VOCs during the sampling was higher than any possible differences related to the health status of the sampled patients, despite the fact that all samples were taken in the same room at the same hospital, with the same method and by the same operating staff.

The hypothesis was that, during the sampling, patients intake VOCs from the environment and then release these VOCs in exhaled breath. As environment VOCs fluctuate daily, it was possible to separate patients in accordance with the sampling period. This hypothesis reinforces the fact that putative biomarkers present in the literature are not coherent and vary a lot from all studies. This might be due to the presence of exogenous VOC which interfere with the detection of putative biomarkers.

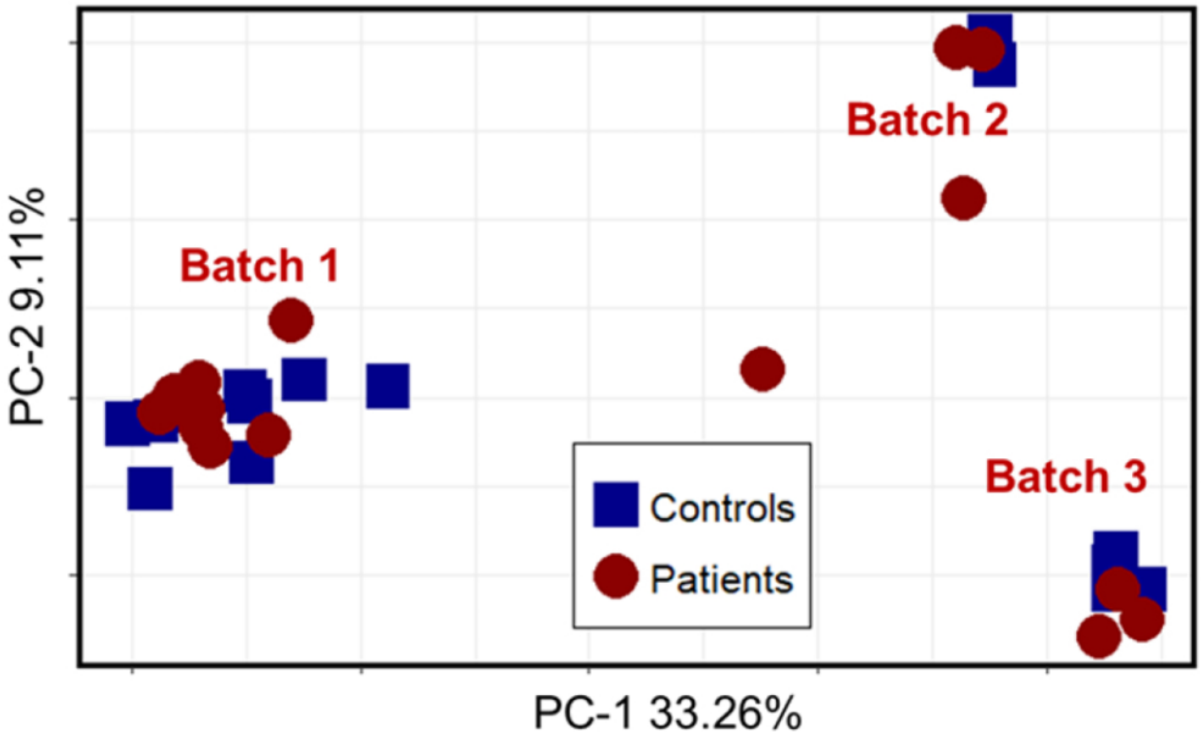

Figure 36: PCA performs with the first set of breath samples (15 lung cancer patients in red and 14 healthy volunteers in blue). Samples were recorded during 3 different periods (batch 1 in January, batch 2 in April and batch 3 in March) and show dispersion according to this criterion

Deeper analysis of this feature table reveals that 671 features were related to sampling periods (Figure 37). Using PCAs with time dependence features displays clear influence of the period of sampling (Figure 38). Moreover, it was also possible to remove these features from the initial feature table to get a new feature table free of sampling period contamination features, and process PCA to show that it is possible to regain the separation between samples according to cancer/non-cancer criteria (Figure 38).

However, although this technique gave good results, it is time consuming. And more problematic as, in this case, there was a clear difference between the three sampling periods because there were one or two months between each of them. But studies recorded during long time periods without the interaction of a sampling process will lead to less clear contamination periods differences and more difficult visualizations of time-dependent features. It was necessary to find other strategies to negate the impact of exogenous features in a goal to detect VOC markers of lung cancer. Three different strategies were tested to keep the essential VOC information.

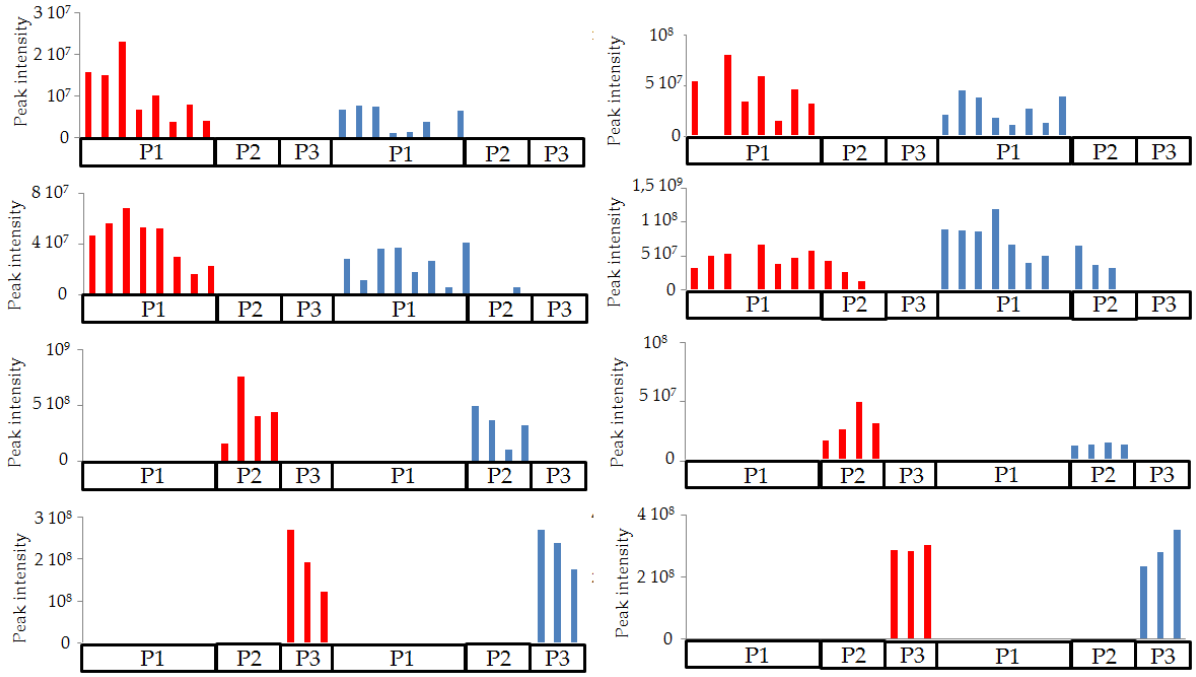

Figure 37: 8 examples of time dependent features with their intensity (from healthy volunteers in blue and lung cancer patients in red) related to the period of sampling (P1 = January, P2= March and P3= May)

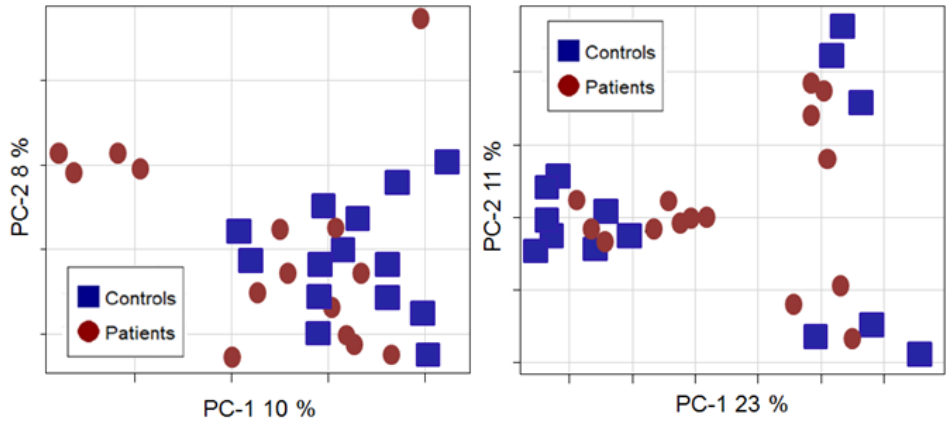

Figure 38: On the left PCA with non-dependent time-sampling features and on the right PCA with dependent time-sampling features for the first set of breath samples

2.3.4 Subtraction blank sample

The first option to delete contamination consisted in collecting blank samples from the room and subtracting this background signal from the VOC profile sample. During the sampling schedule of breath samples, the air of the room, called blanks, were also collected. Results demonstrate that VOC fluctuations profile of the room cannot be fully subtracted using these kinds of methods (Figure 39). For example, with feature 1, variations in this feature intensity found in the room did not reveal the same trend variation in patient-exhaled breath. Feature number 8 shows a relatively stable feature intensity from patients, whereas it shows an important intensity dispersion from room samples. This may be due to the fact that everybody does not react in exactly the same way with the environment. Furthermore, it is highly possible that all exogenous VOCs do not have the same absorption degree and simple blank subtraction cannot fully recover the integrity of exhaled breath samples. Due to the results obtained, this strategy has been abandoned. This information also revealed that exogenous VOC intensity is not stable over time and subsequently, sampling exhaled breath in a same room at different periods will induce a variation that will be difficult to avoid or control.

2.3.5 Medical Air

The second option consisted in using medical air to flush the lungs of patients just before the sampling with the objective of removing all VOCs coming from the environment and having the same background noise for all patients leading to easier seeking of markers. To confirm this original approach, 3 individuals were sampled together outside, in an organic chemistry laboratory corridor and inside an analytical chemistry laboratory where VOC backgrounds were different. Each individual was sampled twice for each area, one after 2 minutes in the specific area to get the equilibrium between the lung and the VOCs present in the environment. And a second breath sample was collected immediately after by breathing medical air during 2 minutes before blowing into a Tedlar® bag. The medical air used came from an air liquid cylinder (Belgium) and was involving the use of a mask and a check valve. Samples recorded were analysed by TD-GC×GC-TOFMS. The feature table obtained from this dataset was used to create a dendrogram (). All samples coming from the first sampling (i.e. without breathing medical air) are not classified according to individuals. It demonstrates the impact of VOC intake from the environment on the VOC profile and may at least partly explain why markers found in the literature are not repeatable from one study to another.

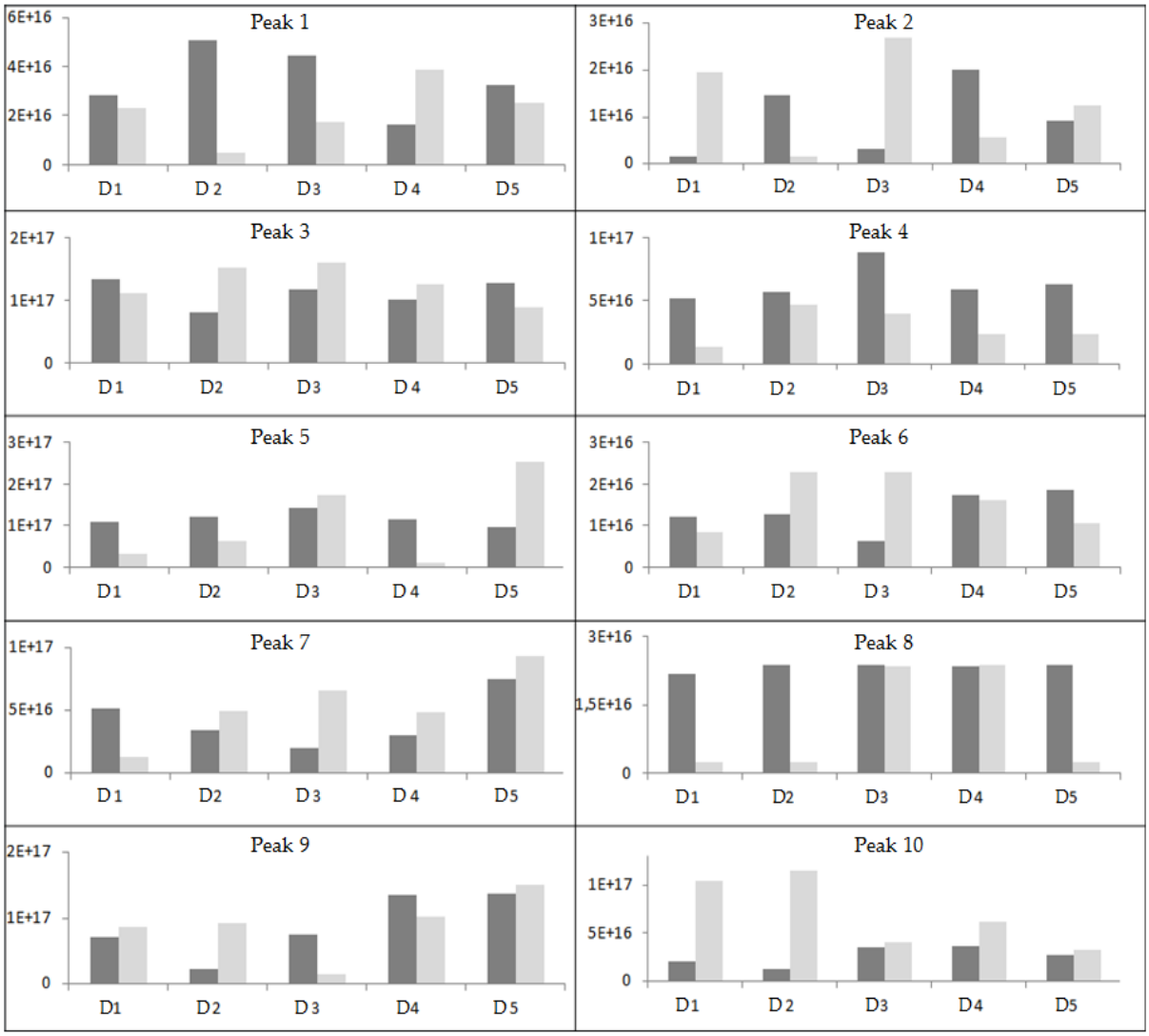

Figure 39: Histograms of 10 features found in exhaled breath (in dark) and room blank (in grey) samples for 5 different days

Moreover, samples are not grouped in accordance with the area, corroborating the fact that each person reacts differently with the environment and using a subtraction of the blank of the room of sampling cannot recover perfectly the impact of exogenous VOCs. Analysis of the samples recorded after breathing medical air shows similar results (Figure 40). Thus, it seems not possible to neglect the impact of exogenous VOC intake from the environment. This means there is no major variation between samples with or without the breathing of medical air before sampling. This experiment demonstrates that breathing medical air for 2 minutes cannot efficiently reduce the noise induced by the surrounding environment. Increasing the time of flushing the lungs may decrease the influence of VOC contaminants. However, in terms of providing an easy diagnosis for lung cancer, this way was not thorough enough.

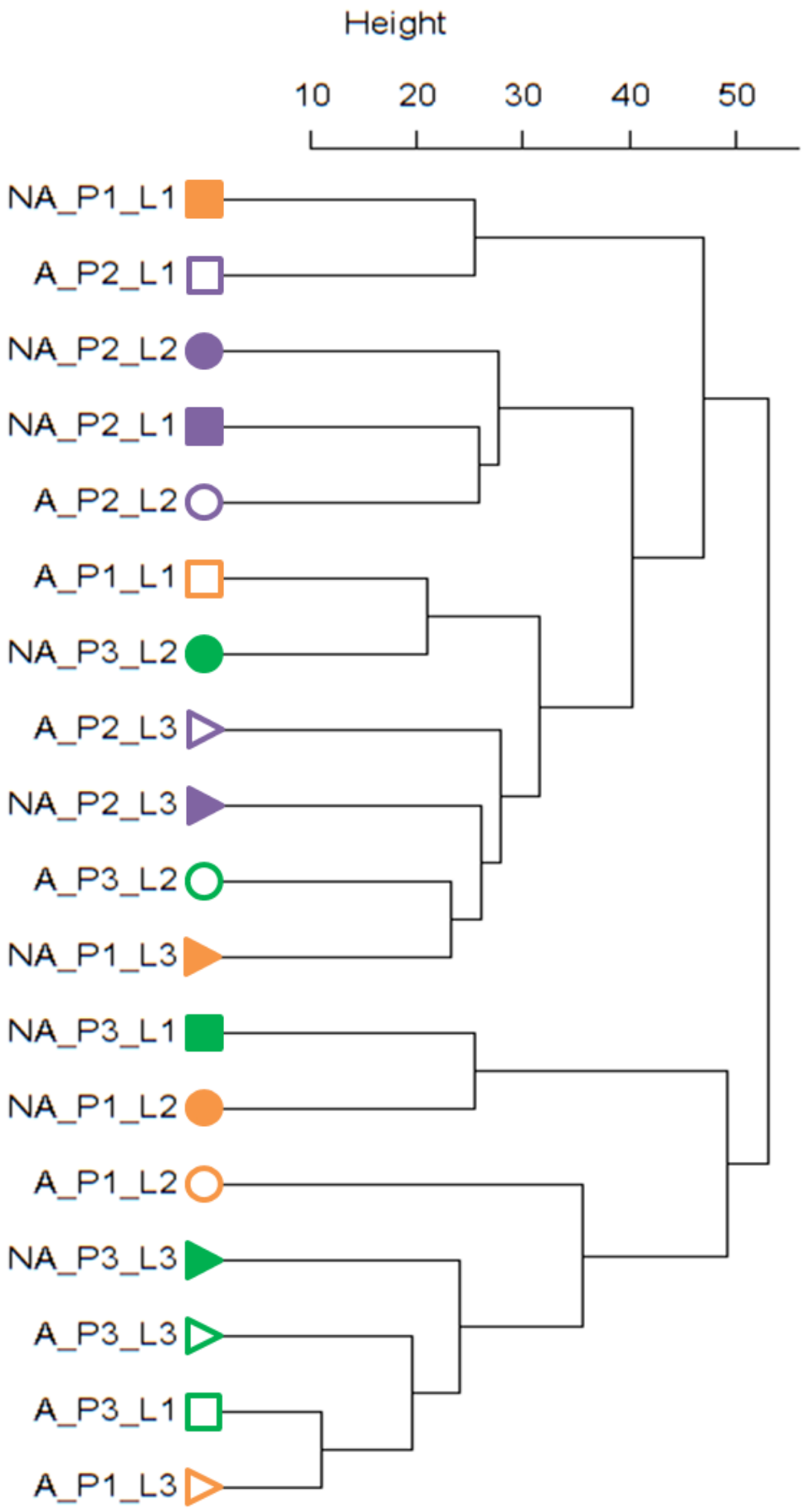

Figure 40: Dendrogram of breath samples recorded from 3 individuals (P1 = orange, P2 = purple and P3 = green) in 3 different locations (L1 =triangle, L2 =square and L3 = dot) without breathing medical air (full) and after breathing medical air for 2 minutes (empty)

2.3.6 Statistical data treatments

The final strategy investigated to reduce VOC contaminant was to use additional data treatments. Data pre-processing and batch effect correction was implemented to reduce the impact of exogenous VOCs. First, all data were normalized using probabilistic quotient normalization (PQN) and log-transformed [255]. The normalization step is a preprocessing method required to scale all samples with the same virtual reference so that they can be compared to each other. And the log-transformed algorithm was used to transform the skewed data to a normal distribution, which is required to have to apply further statistical tools. Then, each batch was individually mean-centred in order to smooth the impact of the sampling dates. This step was possible due to the parallel sampling between patients and controls. This means that the correction was affecting the two classes in the exact same way and it did not generate any overfitting. The corrected data made it possible to more efficiently extract putative biomarkers from the initial raw data. Following this pre-processing step, two different data-processing approaches were investigated: 1) a univariate feature selection tool based on Fisher Ratio calculation; 2) a multivariate approach using random forest algorithms and feature importance ranking.

2.3.6.1 Fisher ratio approach

Due to the large amount of data and features, Fisher ratio calculation was used to decrease the amount of data while maintaining the information about the differences between the two classes. By using a cut-off based on critical Fisher ratio (F_{crit}) with a significant level of 5% [256], it was possible to extract a narrow list of 95 features from the initial list of 1350 features for the first batch of exhaled breath collected (Figure 41). Those features got a retention time in the first dimension ranging from the first minute of the run to the end. Hence, it was not necessary to only apply data treatment on a specific part of the chromatogram. This reduced list could be used to visualized the data reduction by using a PCA (Figure 42). Performing that way resulted in a separation of samples according to cancer/non cancer criteria.

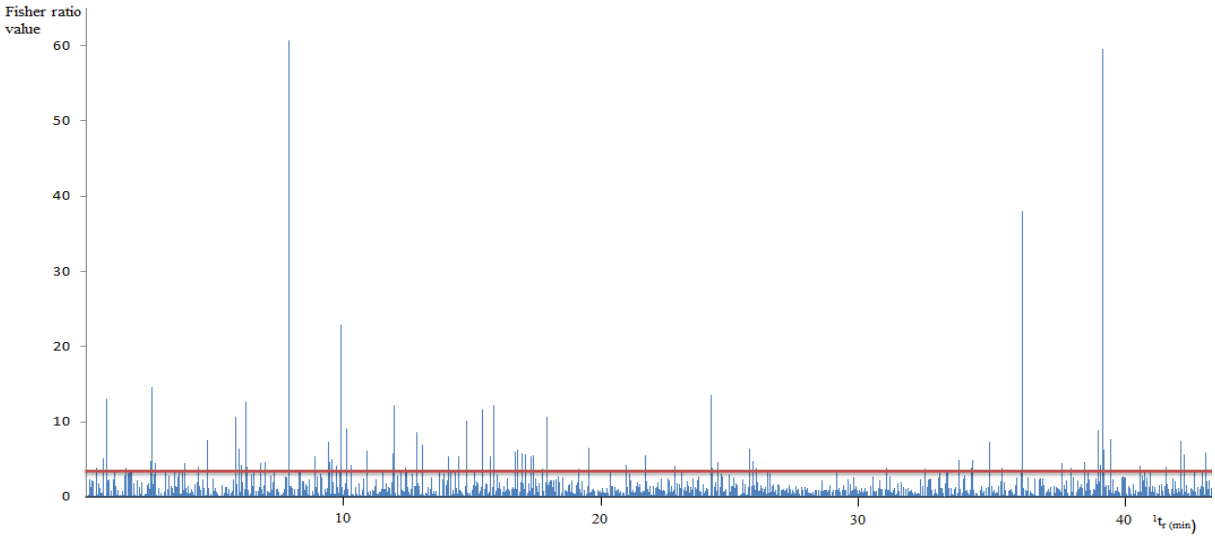

Figure 41: Fisher ratio value for each feature detected in the feature table from the first set of exhaled breath samples according to their retention time in the first dimension. The red line represents the threshold fixed by the critical Fisher value

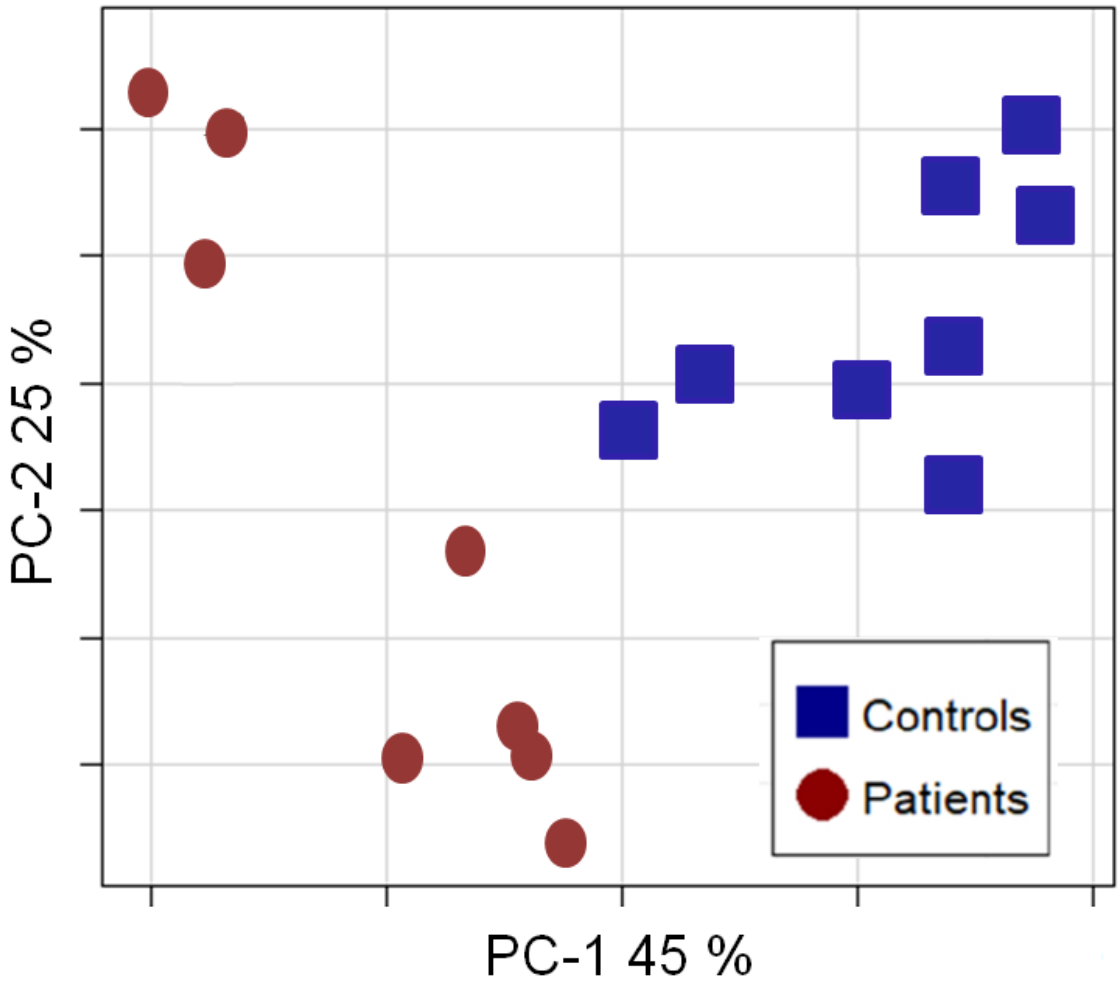

Figure 42: PCA performed with the restricted list of features from the first set of breath samples obtained with Fisher ratio selection

Although Fisher ratio no longer have to demonstrate their ability to reduce the number of datasets while keeping essential information [80, 83, 245, 257-259], the robustness of this statistic tool was test on the data set collected in January with the different S/N investigated. A higher S/N will only detect the most intense feature. Whereas smaller S/N will detect more features but will lead to more difficult data interpretation due to the large amount of features detected. Four S/N were investigated (50, 100, 250 and 1000) and several hundreds of features were detected for all S/N due to the use of GC×GC-TOFMS.

Table 4: Number of features in samples from January 2014 with different S/N investigated

S/N	50	100	250	1000
Number of features	1494	1131	733	380

For all different S/N, a better PC1 and PC2 were obtained for PCA calculated with the restricted list than with all features (Figure 43). It is possible to reduce the amount of data without losing the essential information allowing the separation between the two groups of samples. Moreover, the best separation occurred with S/N of 100, which leads to select S/N during the data processing.

Fisher ratio values were calculated by the statistical comparison feature from ChromaTOF® software and sometimes features got an undefined value instead of a numerical value (42 features out of the 1350 features). Investigation of these undefined features revealed that some of them are very specific to one group and can putatively be markers of lung cancer (Figure 44). This phenomenon occurs for a feature when it is not present in any samples for one class or just in one sample. This is because when a feature is not found in a sample, ChromaTOF® does not give a zero-volume value but simply does not consider the sample for this feature. Therefore, when the software considers no samples or just one, it cannot calculate a standard deviation resulting in an undefined value to this feature. Unfortunately, undefined features also contain features which normally lead to low FR (Figure 44). It is therefore impossible to directly select undefined features as putative markers. In order to remedy to this situation, Fisher ratio were calculated in Excel where a zero-volume value was given to features not present in some samples. Thus, undefined features received a numerical Fisher value and were selected with the F_{crit}. Features represented in all samples were given the same value with the two software calculations. It confirms the absence of mistakes in the formula written in the Excel sheet. But for features, with at least on zero value in one sample, a shift between the two values given by the two software programmes was observed.

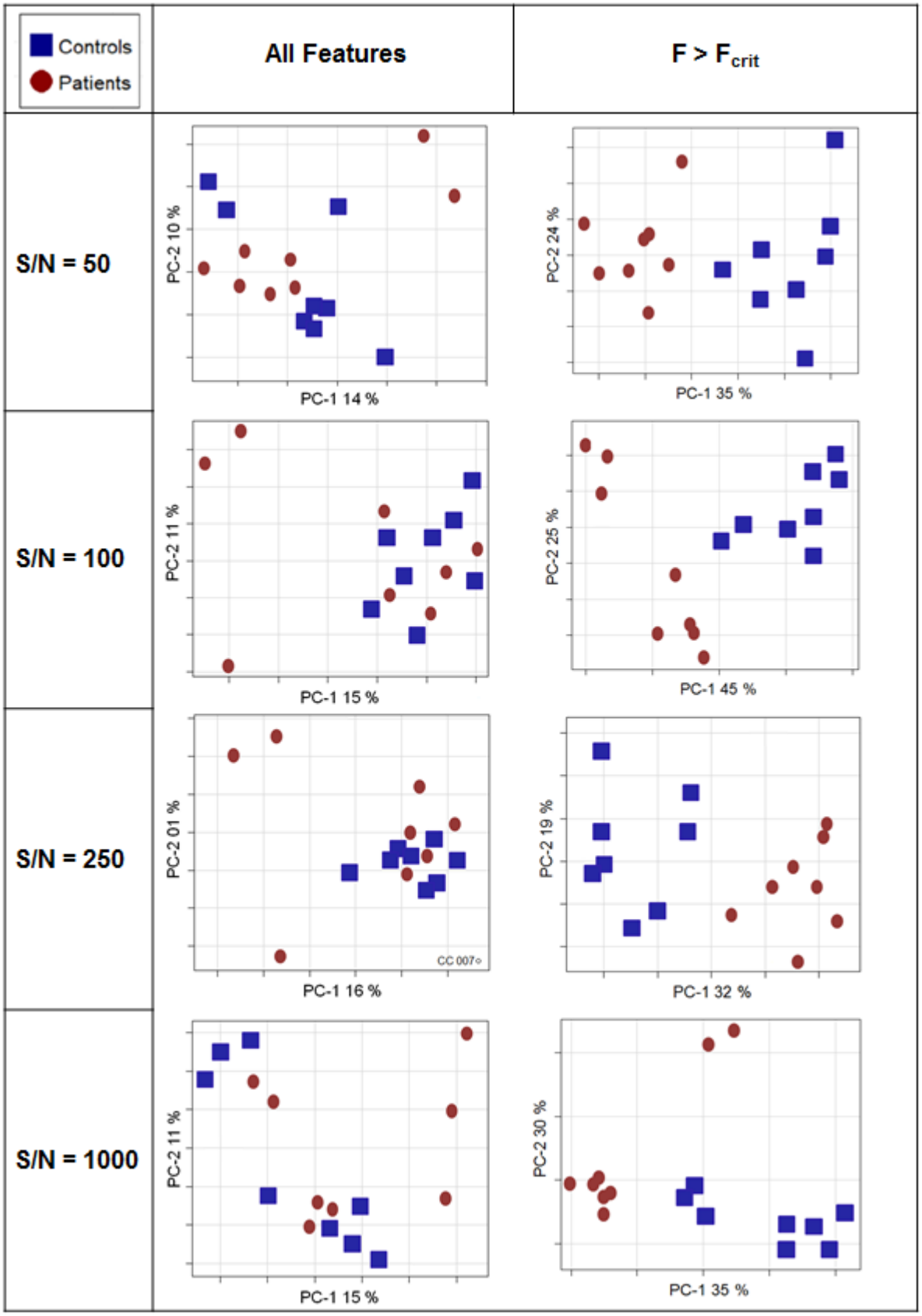

Figure 43: PCAs performed with the four S/N investigated for all feature tables and selected features with Fisher ratio selection

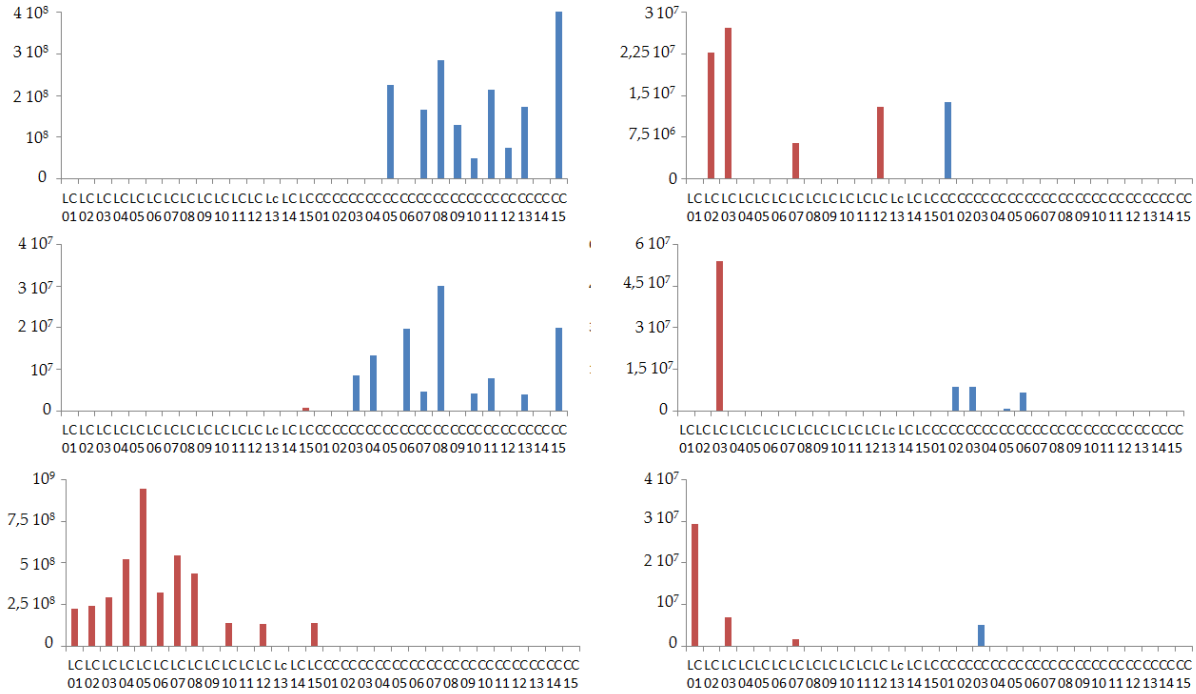

Figure 44: Examples of some undefined features with their intensity for all samples from the first set of breath samples. Specific features related to one population of sample are situated on the left and non-specific features on the right

A theoretical example of such phenomena is explained to illustrate this point. For two classes defined by one variable in three samples. Where at least, one of these variables has a zero value in one of these classes. It is possible to demonstrate the difference of determination for the FR value by means of the two software programmes (Table 5).

Table 5: Theoretical example of two classes defined by one variable in three samples

	Class 1	Class2
Sample 1	13	36
Sample 2	0	33
Sample 3	11	41

To calculate the FR value, Formula from Synovec paper was used (see chapter 1) [252]

	With Excel	With ChromaTOF®
Mean class 1	(13+0+11)/3=8	(13+11)/2= 12
Mean class 2	(36+33+41)/3= 36,66	(36+33+41)/3= 36,66
Global Mean	(13+0+11+36+33+41)/6= 22,33	(13+11+36+33+41)/5= 26,8
Inter variation	$((8-22,33)^2 *3 + (36,66-22,33)^2 *3)/(2-1) = 1232,1$	$((12-26,8)^2 * 2 + (36,66-26,8)^2 *3)/(2-1) = 730$
Inter variation	$((13-22,33)^2 + (0-22,33)^2 + (11-22,33)^2) – ((8-22,33)^2*3) + ((36-22,33)^2 + (33-22,33)^2 + (41-22,33)^2) – ((36,66-22,33)^2*3)) / (6-2) = 94,65$	$((13-26,8)^2 + (11-26,8)^2) – ((12-24)^2*2) + ((36-26,8)^2 + (33-26,8)^2 + (41-26,8)^2) – ((36,66-24)^2*3)) / (5-2) = 361,52$
FR	1232,1/94,65= 13,02	730/ 361,52 = 2,31

The result got by means of ChromaTOF® does not take into account the presence of the variable with a zero value, even if this variable has a value in the other class. Start to this point, it is normal to detect some divergence between the two software results. Hence, Excel was preferred to perform the FR calculation.

After investigations on FR calculation and capability, the feature table obtained from the exhaled breath of 15 lung cancer patients and 14 healthy volunteers was used to extract 73 features. This list was still too large and had to be reduced again. To go deeper into the detection of putative markers, a more selective cut-off was used with 1% of significant level instead of 5%. It was possible to extract a list of 29 features from the 1350 initially present (Figure 45 & table 6). These selected features were screened to avoid chromatographic artefact. No features related to Tedlar® bag leaching were found. It was due to the use of Fisher ratio filters because Tedlar® bag contaminations or even features from other contamination sources were not affecting only samples from one class. Therefore, their Fisher ratio values were too low to be above than the critical Fisher value. From this list, benzene and p-xylene were removed because they are known to be an artefact of smoking. PCA analysis

performed with these 27 features displays a clear separation between lung cancer patients and healthy volunteers (Figure 46). This demonstrates that a proper feature selection approach, such as FR, makes it possible to remove the impact of the background and the extraction of significant information.

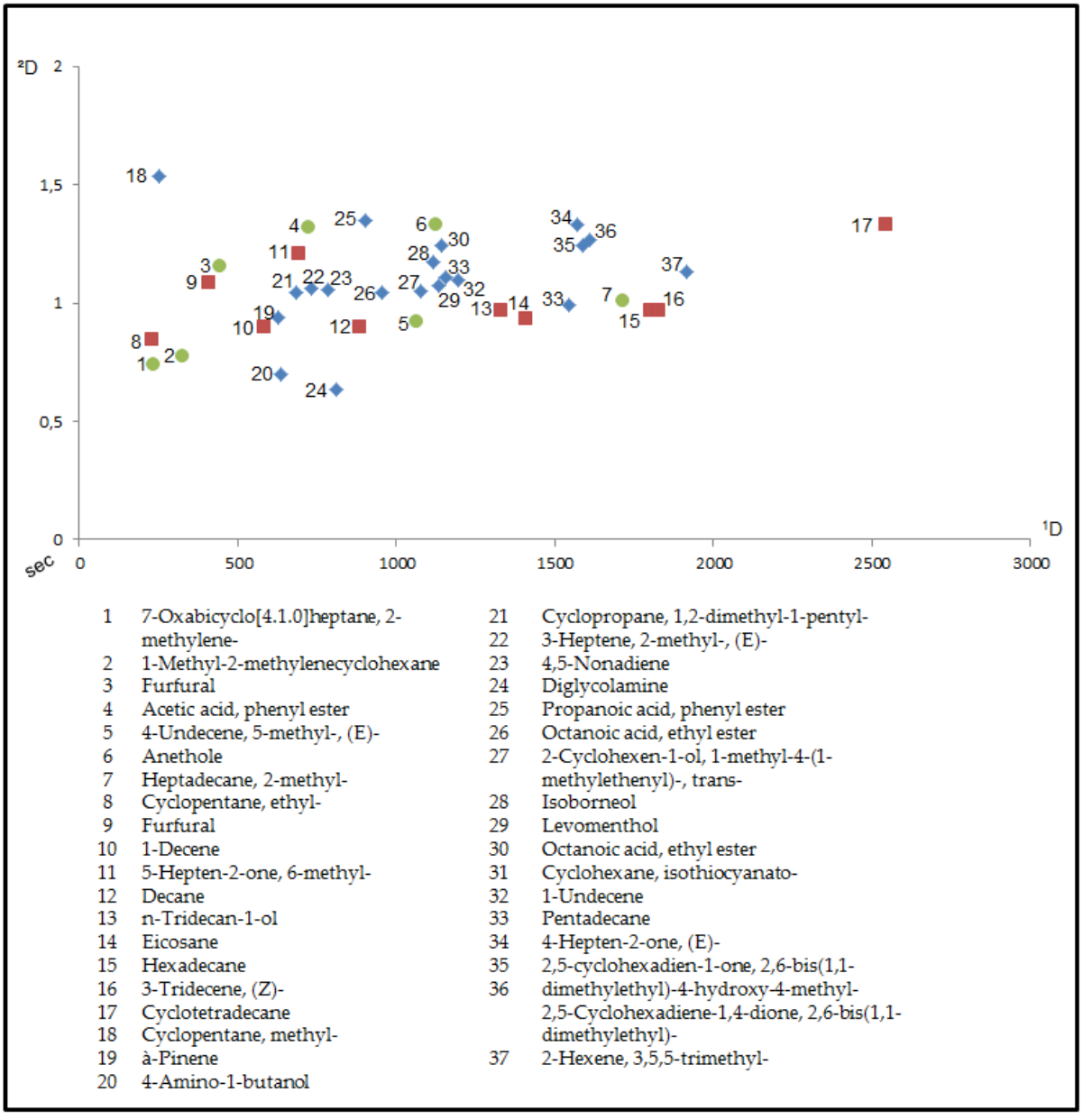

Figure 45: On the top, a reconstructed 2D-chromatogram with the 37 features highlighted, with Fisher ratio in blue diamond, random forest in red square and with both approaches in a green circle. On the bottom, the list of the correspondent feature numbers

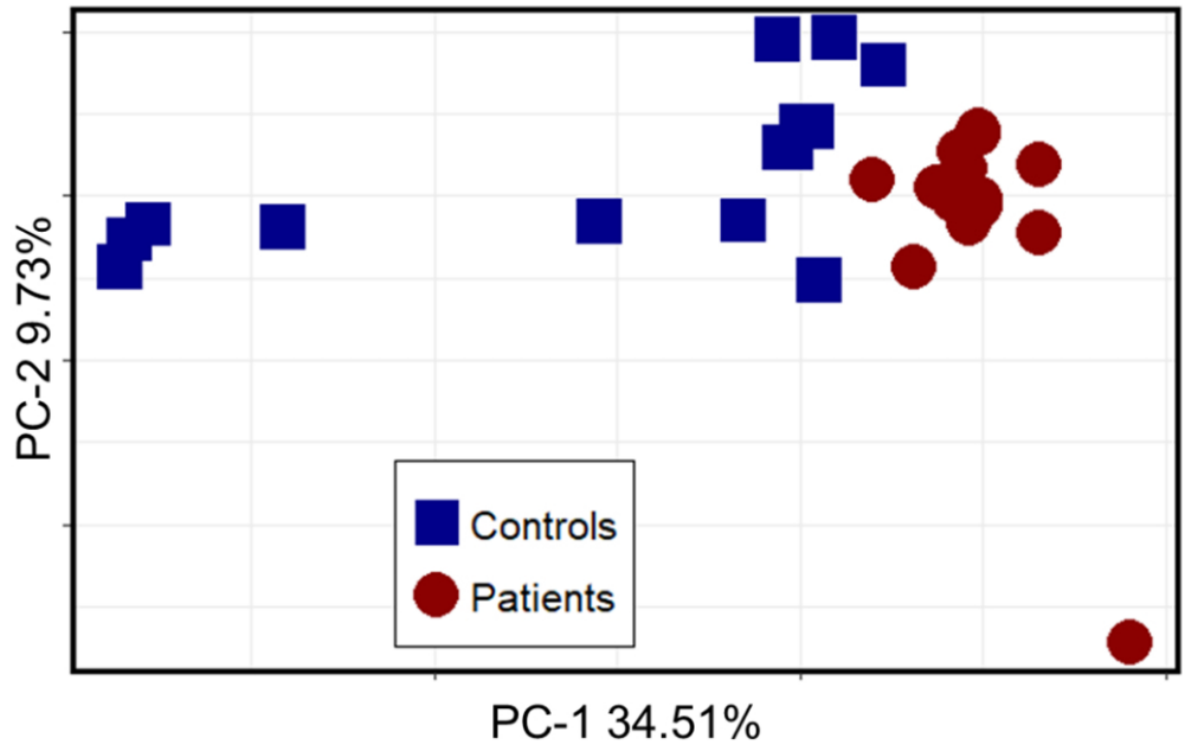

Figure 46 : PCA score plot with the 27 features highlighted with the FR approach in the exhaled breath of healthy volunteers (controls) and lung cancer patients (patients)

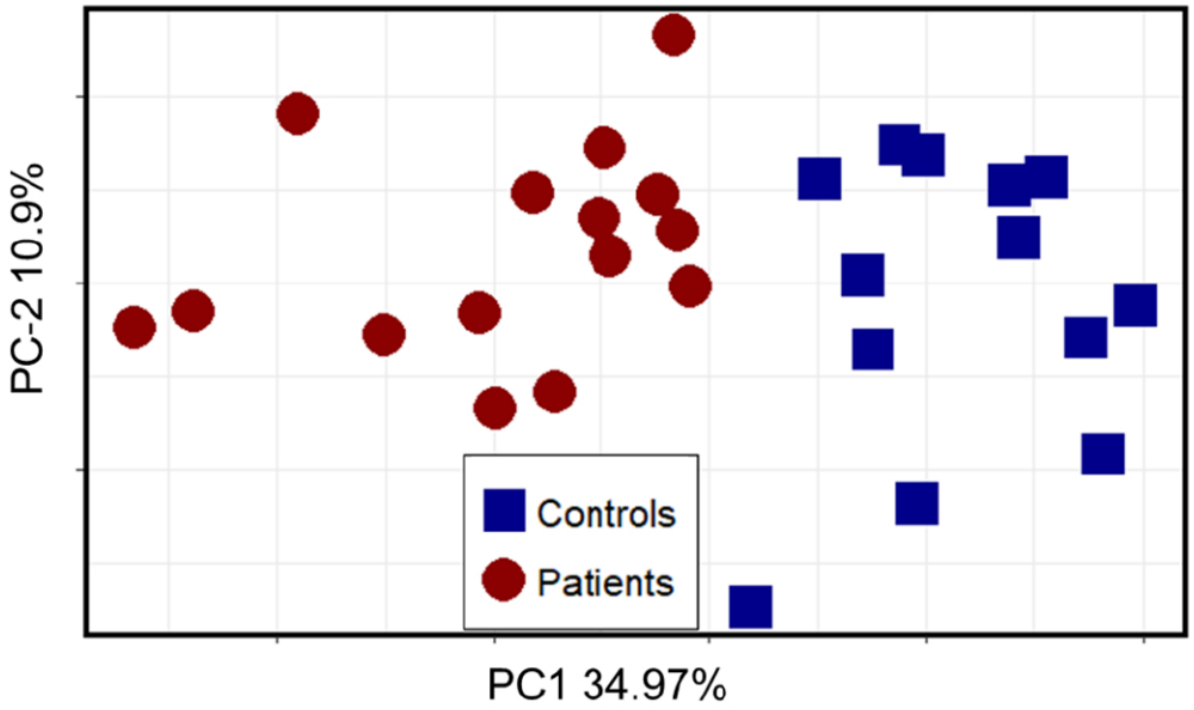

Figure 47: PCA score plot of the healthy volunteers (negative) and lung cancer patients (positive) by using the 17 features highlight with the RF approach

Table 6: Characteristic of the putative list of biomarkers highlighted

Method	^1t_r (s)	^2t_r (s)	Proposed identification	Chemical family	Match factor	Probability
FR	1118	1.17	Isoborneol	Alcohol	862	5091
FR	1134	1.07	Levomenthol	Alcohol	918	2813
FR	1075	1.05	2-Cyclohexen-1-ol, 1-methyl-4-(1-methylethenyl)-, trans-	Alcohol	781	2182
FR	954	1.04	Octanoic acid, ethyl ester (1)	FAME	902	7525
FR	1143	1.24	Octanoic acid, ethyl ester (2)	FAME	867	568
FR	900	1.35	Propanoic acid, phenyl ester	FAME	894	7192
FR	1195	1.10	1-Undecene	HC	830	756
FR	686	1.05	Cyclopropane, 1,2-dimethyl-1-pentyl-	HC	850	1640
FR	1544	0.99	Pentadecane	HC	929	1961
FR	625	0.94	à-Pinene	HC	867	2011
FR	732	1.06	3-Heptene, 2-methyl-, (E)-	HC	937	1471
FR	252	1.54	Cyclopentane, methyl-	HC	846	910
FR	786	1.06	4,5-Nonadiene	HC	857	1683
FR	1157	1.11	Cyclohexane, isothiocyanato-	HC	939	9468
FR	1914	1.13	2-Hexene, 3,5,5-trimethyl-	HC	827	8556
FR	1588	1.25	2,5-cyclohexadien-1-one, 2,6-bis(1,1-dimethylethyl)-4-hydroxy-4-methyl-	Ketone	764	2842
FR	1572	1.33	4-Hepten-2-one, (E)-	Ketone	980	8010
FR	1611	1.27	2,5-Cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-	Ketone	788	5460
FR	812	0.64	Diglycolamine	Miscelaneous	913	5245
FR	635	0.70	4-Amino-1-butanol	Miscelaneous	904	2357
FR-RF	443	1.12	Furfural (1)	Aldehyde	905	7713
FR-RF	722	1.32	Acetic acid, phenyl ester	FAME	940	7818
FR-RF	1714	1.01	Heptadecane, 2-methyl-	HC	940	1852
FR-RF	327	0.78	1-Methyl-2-methylenecyclohexane	HC	878	1233
FR-RF	235	0.74	7-Oxabicyclo[4.1.0]heptane, 2-methylene-	HC	849	2651
FR-RF	1066	0.92	4-Undecene, 5-methyl-, (E)-	HC	856	1301
FR-RF	1127	1.33	Anethole	Miscelaneous	946	5433
RF	1329	0.97	n-Tridecan-1-ol	Alcohol	861	2951
RF	410	1.08	Furfural (2)	Aldehyde	410	9635
RF	231	0.84	Cyclopentane, ethyl-	HC	918	7154
RF	582	0.90	1-Decene	HC	946	3042
RF	1830	0.97	Hexadecane	HC	866	1647
RF	1802	0.97	3-Tridecene, (Z)-	HC	873	346
RF	1410	0.93	Eicosane	HC	927	1570
RF	886	0.90	Decane	HC	898	1410
RF	2546	1.33	Cyclotetradecane	HC	800	458
RF	692	1.21	5-Hepten-2-one, 6-methyl-	Ketone	902	9547

The variability of the healthy volunteer cohort is higher than the variability of lung cancer patients. This may be explained by the fact that healthy volunteer individuals are less homogenous than lung cancer patients. Indeed, the age of both populations shows clearly different dispersions (Figure 48) and it has been reported that age induces variation in the VOC profile of breath [135, 260]. All lung cancer patients were smokers or ex-smokers, but the healthy volunteer class contained non-smokers too. Another factor is the way of life. Lung cancer patients should have quite similar lifestyles as they are all sick, leading to more homogenous characteristics. Whereas, the way of life of the healthy volunteer populations is not restricted by a pathology and influences the composition of the VOC profile exhaled by individual. For example, it has been proven that sport impacts on the VOC profile emitted through the breath [261].

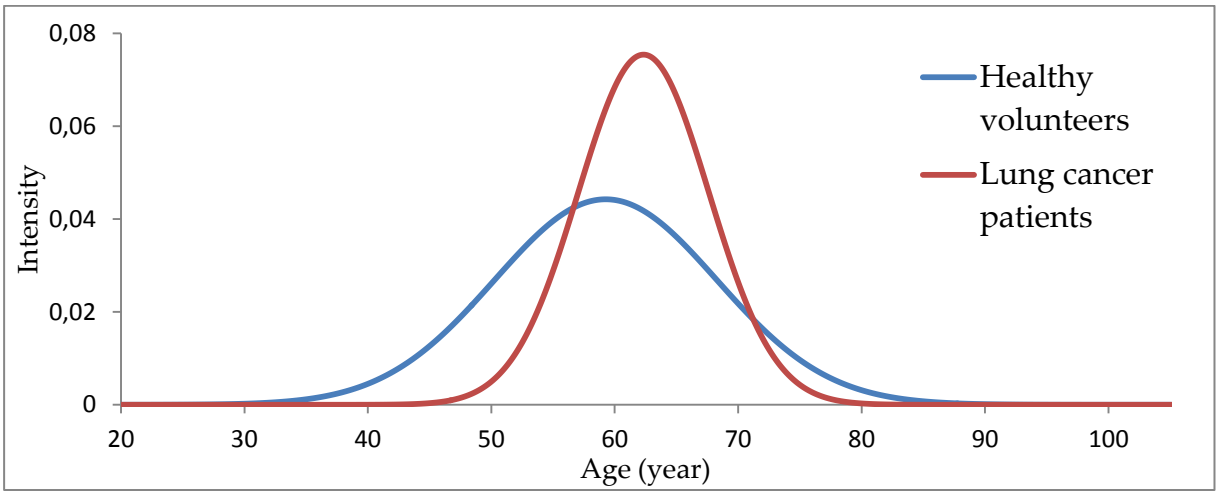

Figure 48: Normal distribution of age from healthy volunteers and lung cancer patients for the set of exhaled breath samples

2.3.6.2 Random forest approach

The second statistical approach for feature selection was based on the use of the random forest algorithm (RF), a multivariate machine-learning approach built on a decision tree approach. Hence, multiple decision trees were created and merged together to obtain a more accurate and stable prediction. After the construction of the classification trees, the variables were ranked according to their importance and effect on the classification accuracy. The significant features were selected based on this ranking and a cut-off value of 0.1 was set in mean decreased accuracy. This resulted in the selection of a set of 17 significant features for the separation of the two populations (Figure 45 & Table 6). As for the univariate FR approach, this RF

demonstrates the potential of this multivariate feature selection approach to properly cluster the two populations in the PCA space based on these 17 features (Figure 47).

2.3.6.3 Comparison of statistical treatment

The general workflow of the data treatment is represented in the Figure 49.

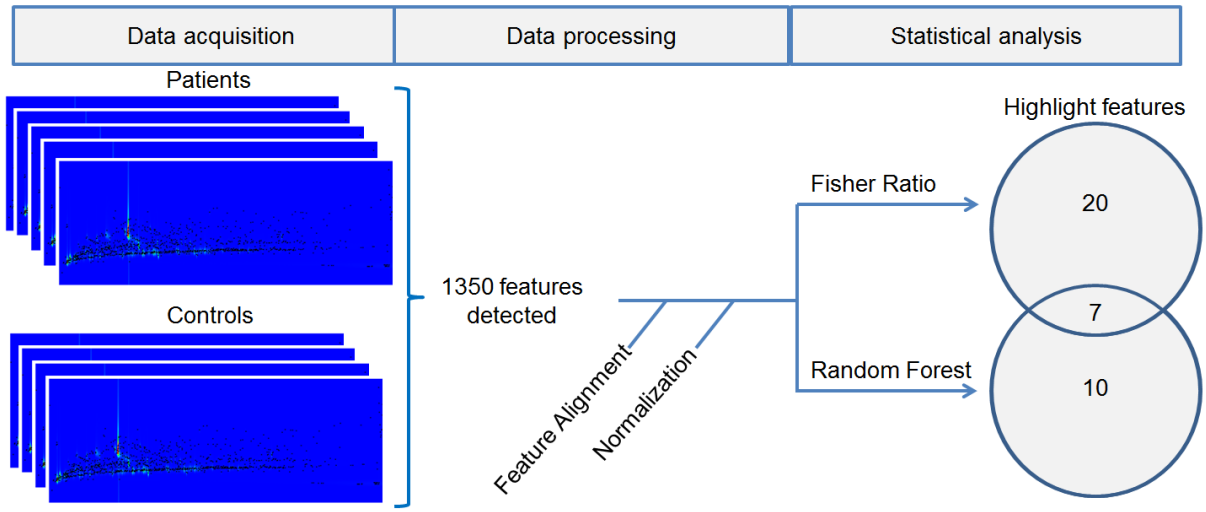

Figure 49 : Diagram of the general workflow used during the data treatment

The main difference between random forest and Fisher Ratio for feature selection is the multivariate dimension. Indeed, the combination of decision trees makes it possible to perform classification based on combined information from different features. Moreover, random forest makes it possible to obtain direct classification performance information in addition to the feature selection possibilities. Both uni- and multivariate methods made it possible to reduce the impact of the presence of variable amounts of exogenous VOCs related to the period of sampling to a level that did not significantly interfere anymore with the extraction of biologically relevant VOC signatures.

The 37 features selected by the two approaches were sorted according to their chemical family (i.e., alcohol, aldehyde, fatty acid methyl ester (FAME), hydrocarbon, ketone, nitrogen containing compounds). From these chemical families, average intensities were calculated (Figure 50) and the ratio Patients-intensity on Controls-intensity were evaluated. For FR features, the highest ratios were obtained for FAMEs and ketones. The overexpression of these compounds in lung patient samples could be explained by inflammation processes inside the lungs. The same family classification process was applied to the 17 features highlighted by the

random forest approach. Interestingly, the two major ratios were also coming from the FAMEs and the ketones. This observation could indicate that the major processes involved in the production of VOCs in the lung of cancer patients could be linked to FAMEs and ketones. These chemical families were already identified in previous studies and they are supposed to come from oxidative stress reactions. This family-based approach in non-targeted screening provides insights regarding the general trends of the samples, which is already informative. Indeed, for non-targeted studies, the full identification of thousands of features is practically impossible, which made the study-to-study comparison highly complicated. However, if a group of compounds are found to be specific in different studies, it could orient the future research towards a predefined group of molecules.

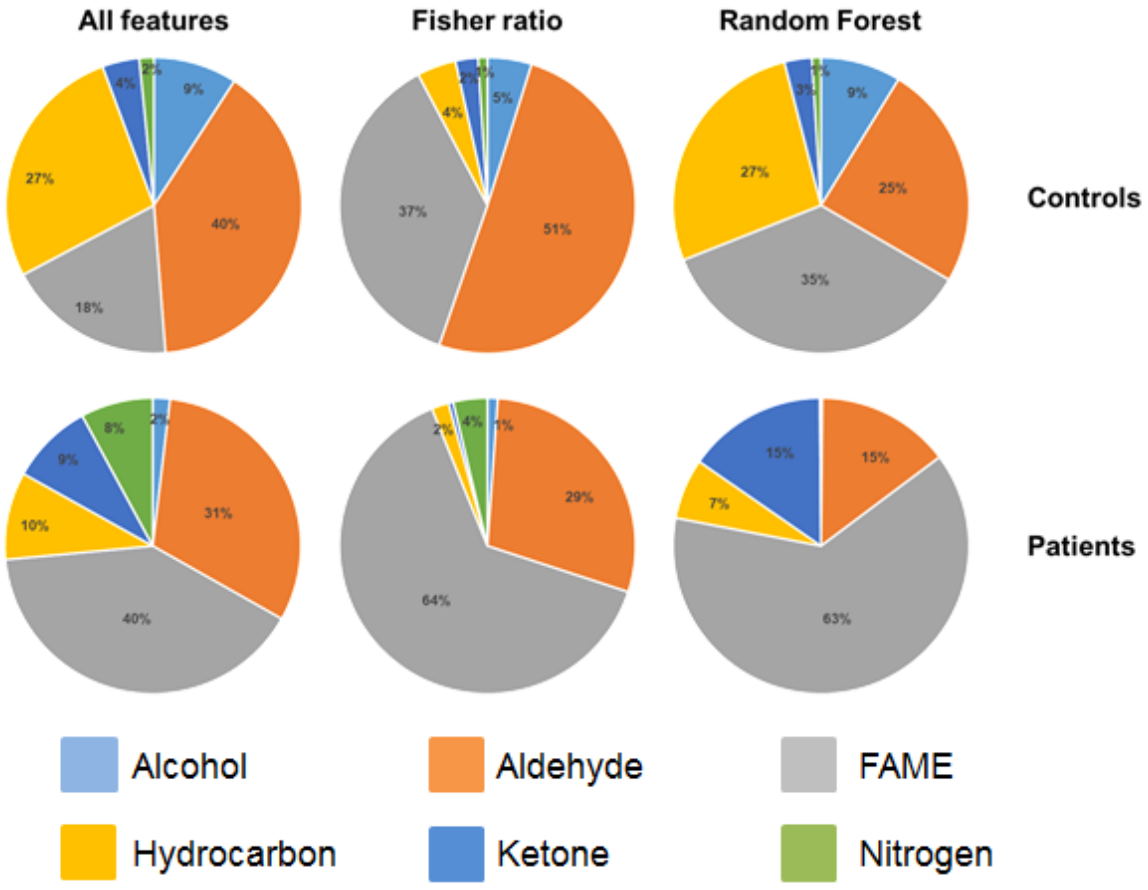

Figure 50: family importance for the selected features in both methods; Fisher Ratio only; random forest only

From the lists of 27 and 17 features, a set of seven features was common to both approaches (Table 6). One of these features, the 7-Oxabicyclo[4.1.0]heptane, 2-methylene- was removed because the impact of this feature does not increase the separation power between the two populations. The six features would be

considered as the most representative markers of differentiation between the two classes. As illustrated in Figure 51, a PCA based on these six markers results in a separation trend between the two classes but the clustering is not as clear as when either FR or RF models are applied separately. It can however be noticed that the percentage of explained variance (67%) is higher than in both separated models. Even if this PCA does not provide extra information, it demonstrates the importance of the technique used for model building and the interest of applying different models in order to validate the data-processing approach.

Only four of the reported features (both models included) are common to compounds previously reported as a lung cancer human breath biomarker in the literature : Cyclopentane, methyl- [118, 196], 2,5-Cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-,[194] Hexadecane [99], and eicosane [135]. However, before making link with biomarkers already reported in the literature, it is necessary to confirm the identification of the features highlight during this work. For example, on the list of 37 features (Table 6), two features get furfural name attribution. It is impossible that both of them a related to furfural. Therefore, there is a miss-identification for at least one of them. Same scenario applies for two other features which get octanoic acid, ethyl ester name attribution. Others wrong identification of compounds list in the Table 6 can be determine based on the retention time of the compounds. For example, it is not possible, with this column set, that eicosane elutes before hexadecane and pentadecane. Thus, utilization of high-resolution detectors and retention index could overcome these miss-identifications and help to confirm the name of the others features.

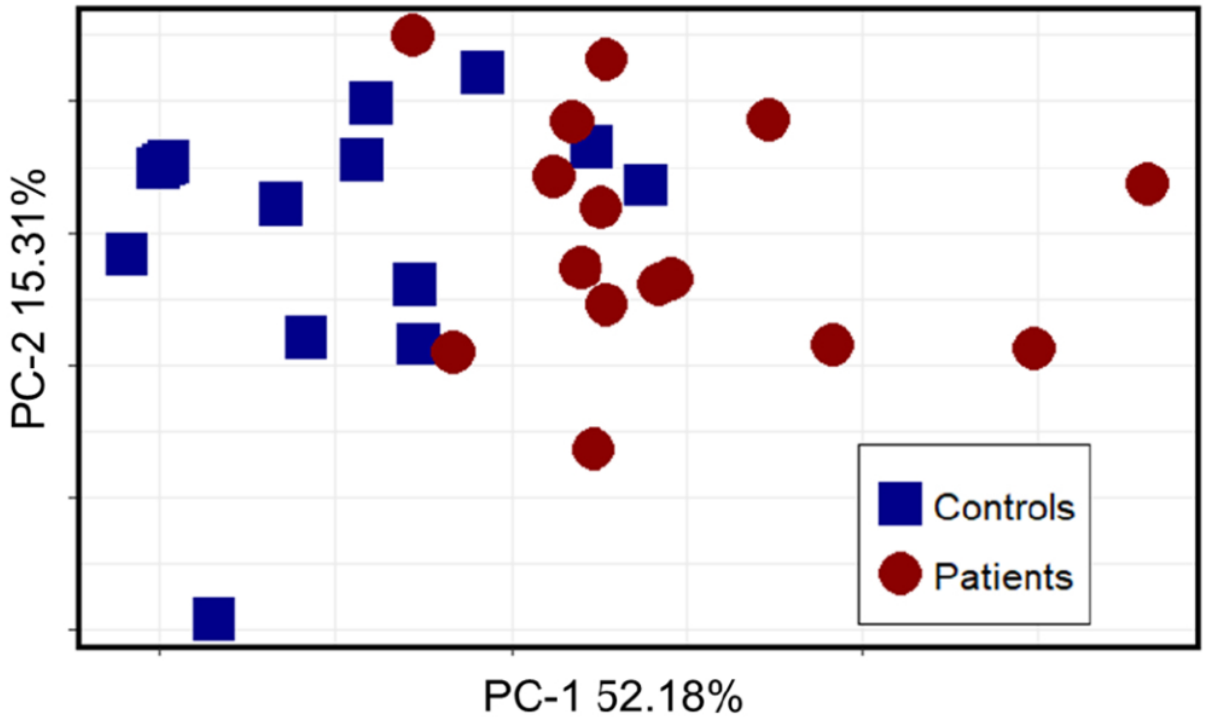

Figure 51: PCA score plot of the healthy volunteers (negative) and lung cancer patients (positive) by using the six features highlight with the RF and FR approaches

Chapter 3:

Headspace analysis of cell cultures media by GC×GC-(HR)TOFMS

This chapter contains results obtained from the analysis of the headspace of cell cultures. The aim was to detect specific VOCs emitted by different cell lines to help in the conception of an early and non-invasive diagnosis of lung cancer. Cell cultures analysis were performed in collaboration with the histology-cytology laboratory of Dr Marie-Claire De Pauw from the University of Liège. The Pegasus 4D equipped with an auto-sampler from Gerstel (MPS) was used for the development method, where cell culture samples were injected manually in the AccuTOF™ GCv 4G. All experimental parameters and data-processing information related to this chapter can be found in the annex.

3.1 Sampling design

As a new field of exploration, the first step was to design a method to collect samples from the headspace of a cell culture and then to optimize it to obtain the best information about the VOCs emitted by cells. The sterile conditions needed for cell culture induces limitations on techniques used for sampling. The hypothesis was that cells interact with media by releasing VOCs into it. Recuperation of the media after passaging cell step into an SPME vial can be used to trap VOCs emitted by the cell through the media matrix. This method was easy to set and dealt with sterilization conditions. SPME vials were heated to 120 ° C for 15 minutes and closed to be free from biological contamination. Then, they were introduced into the laminar hood to be open and filled with DMEM from cell culture. The only drawback of this step was the necessity of cleaning their exterior surface with ethanol before introducing them into the hood. Although alcohol was not present inside the vial, it was almost always possible to find trace residue during the analysis. SPME vials had a capacity of 20 mL making it possible to obtain at least 15 mL of DMEM per cell flask to leave enough space to expose the fiber to the headspace. To produce enough DMEM, flasks with a surface of 75 cm^2 were used as the usual condition is to fill them with 20mL of culture media. Vials were moved from the cellular laboratories to the injection site (same building) to be analysed. There were put into a water bath at 37 °C prior to and during the fiber exposition. Afterward, VOCs were injected in the GC×GC system to be analysed (Figure 52).

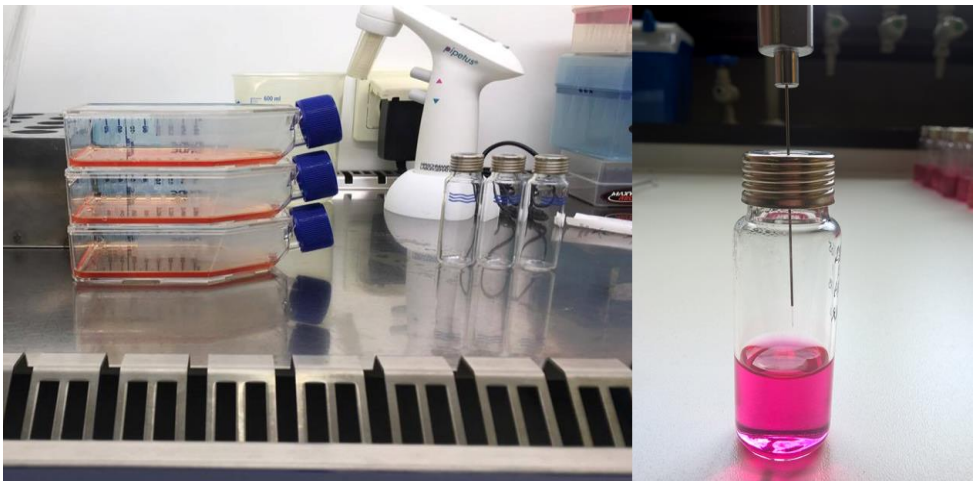

Figure 52: On the left, a flask containing cells and culture media just before the step of passing it inside the laminar hood. SPME vials were ready to be filled with culture media. On the right, the introduction of the SPME fiber to the headspace of culture media

3.2 Headspace optimization

Several parameters in SPME-GC×GC-TOFMS could be optimized. Previous methods of breath analysis by GC×GC-TOFMS was transposed to this study as the VOC matrix was analysed too. Thereby, only SPME parameters were optimized such as the coating fiber choice, the volume of DMEM introduced into the SPME vial, the incubation time in the water bath, exposition or desorption time.

The fiber coating choice was selected on the free DMEM sample injected on the 4G acutof from JEOL. Three SPME fibres were investigated; 65 µm polydimethylsiloxane/Divinylbenzene (PDMS/DVB) fused silica fiber, 100 µm polydimethylsiloxane (PDMS) fused silica fiber, and 50/30 µm Divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) StableFlex fiber. PDMS and DVB/CAR/PDMS fibres were already report in the literature for the analysis of the headspace of cell culture [246, 247]. PDMS/DVB fiber was also chosen as it has been commonly cited for the analysis of the headspace of biological fluids. The 50/30 µm DVB/CAR/PDMS fiber was revealed to be the optimal choice and was chosen for ulterior analysis because more features were detected and the intensity of them was higher than for the 2 other fibres (Figure 53).

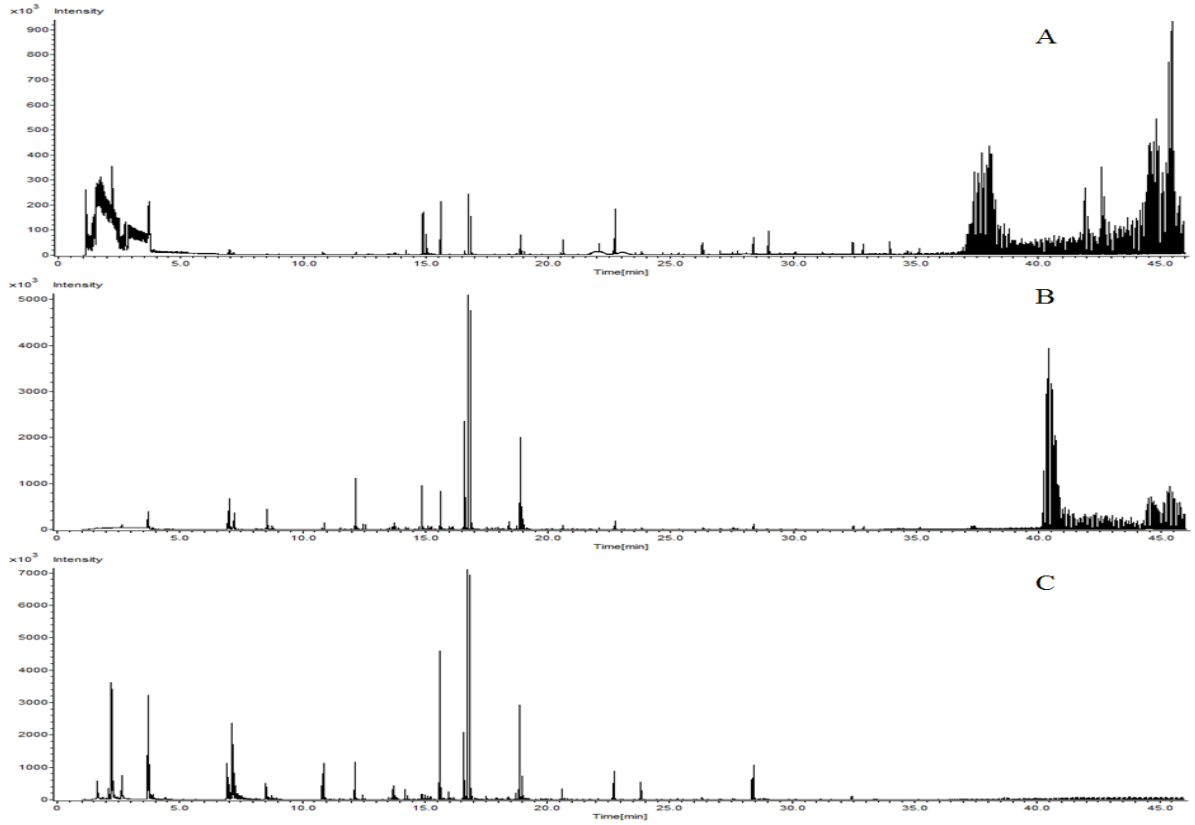

Figure 53: 1D-chromatogram on the headspace of DMEM sample with 3 different SPME fibers (A = PDMS, B= PDMS/DVB and C = DVB/CAR/PDMS)

Optimizations on other SPME parameters were applied on the Pegasus 4D connected with an auto-sampler bench from Gerstel (MPS) with free DMEM samples. Parameters were optimized by using the intensity of three features (A, B and C) present in all chromatograms. These features were selected according to their retention time in the first dimension ($^1t_{r,A}$= 4 minutes, $^1t_{r,B}$= 16 minutes and $^1t_{r,C}$= 32 minutes) to cover as much as possible the separation space and obtain the best optimization.

Incubation time corresponded to the delay of heating the sample at 37 °C in the thermal unit of the MSP2 before exposing the SPME fiber. As volatility of compounds depends on temperature, increasing it made it possible to detect more VOC. However, to be in similar condition to the human body and cell culture growing, a maximum of 37 °C was selected. Two incubation times were selected (10 and 20 minutes) to make it possible to reach the equilibrium between the liquid and the gas phase at 37 ° C. Very close results were obtained (Table 7: lines 1 and 2). It was not necessary to use a longer incubation time than 10 minutes. Lower time was not investigated as we assumed that at least a minimum of 10 minutes are needed to go from room temperature to 37 ° C in the thermal unit.

Longer exposure time made it possible to get higher feature intensity until reaching the equilibrium between the headspace and the coating of the fiber. Table 7 (lines 2-5) display the intensity of the three features for constant incubation time (10 minutes), volume of DMEM (10mL) and storage time of sample (0 day) but several extraction time (20, 50, 90 and 120 minutes). As expected, 20 and 120 minute extraction times gave lower and higher feature intensities respectively. However, 50 minutes showed relatively similar values compared to 90 and 120 minutes extraction time. Choosing it, gave great feature intensity and also allows to fit the sample incubation time with the GC×GC run time.

The next parameter investigated was the quantity of media introduced into the SPME vial. Vials must be filled to maximize the VOC concentration in the gas phase. The partition coefficient, named K, defines the equilibrium distribution of a compound between the liquid and gas phase (Equation 13), where C_L is the concentration of the compound in the liquid phase at the equilibrium and C_g is the concentration of the compound in the gas phase at the equilibrium.

$$K = {C_L}/{C_g} \qquad \text{Equation 13}$$

Compounds with low K values will tend to be more present in the gas phase than compounds with a higher K values. It is possible to decrease the value of K by increasing the temperature. However, to mimic cell growing conditions, the temperature was set at 37 °C. Another option for decreasing K is to add inorganic salt to the liquid phase. High salt concentrations will decrease the solubility of polar compounds in the liquid phase and then increase their concentration in the gas phase. However, due to the initial presence of organic salts found in the culture media, this route was not investigated. The concentration of a compound in the gas phase is also driven by the phase ratio, named β, where Vg is the volume of the gas phase and V_L the volume of the liquid phase (Equation 14).

$$\beta = {V_g}/{V_L} \qquad \text{Equation 14}$$

Changing the volume of culture media introduced in the vial impacts the β value. Adding more samples decreases β value. However, decreasing β value will not

always improve sensitivity by increasing the response of compounds in the gas phase. If the volume of the gas phase is too low, not enough compounds will be trapped into the SPME fiber and therefore affect the sensitivity. Four volumes were investigated and results (Table 7: lines 4, 6-8) displayed an optimum in features intensity for 5 mL.

The delay between the sampling of DMEM in vial and analysis with GC×GC-TOFMS may also have a role to play. Four samples with different volume of fresh media were collected in vials and then stored at 4 °C. This temperature was selected because fresh DMEM was kept in this condition and at low temperature to help keep the integrity of VOC headspace. The result lead to same conclusion about the amount of volume inserted in the vial (Table 7: line 9-12). Compared to direct injection (no storage sample between the collection and the analysis), a diminution of features-intensity was recorded.

The optimal condition for analysing headspace of cell culture was with a 50/30 µm DVB/CAR/PDMS fiber exposed in a SPME vial of 20 mL fill with 5 mL of DMEM during an extraction time of 50 minutes. An incubation time of 10 minutes to reach the temperature sample at 37 °C was also needed. To confirm these results, another injection based on optimum results with another fiber of 50/30 µm DVB/CAR/PDMS was performed and lead to the same features intensity (Table 7: line 13). For the optimization process, design of experiment calculates the optimum for major parameters in a minimum of experiments and showed a correlation between them. This mathematical tool was not used in this thesis as incubation and extraction time were limited to temporal constraints. Indeed, in practice it is also necessary to optimize the number of injections per day. Like GC×GC run lasts around one hour, the time lap dedicated for incubation and extraction time as to be around one hour too. Storage time trends to be as short as possible to keep the integrity of the sample. Thus, only amount volume of liquid phase parameter stay to be optimized. There was no need to use design of experiments to optimize it as it was previously done. This work was necessary to show the correlation between incubation and extraction times of the feature intensity. It also makes it possible to know what was the best fiber coating and what was the optimum amount of DMEM introduced into the SPME vial to obtain the best chromatographic signal.

Table 7: Set of experiences with parameters used to optimize the headspace analysis of cell culture media by SPME based on the response of three features covering the space separation ($1tr_A$= 4 minutes, $1tr_B$= 16 minutes and $1tr_C$= 32minutes)

Number	Incubation time (min)	Extraction time (min)	Volume of DMEM (mL)	Storage time (day)	Intensity feature A	Intensity feature B	Intensity feature C
1	20	120	10	0	$1,8\ 10^5$	$7,0\ 10^5$	$2,2\ 10^5$
2	10	120	10	0	$1,7\ 10^5$	$6,8\ 10^5$	$2,2\ 10^5$
3	10	90	10	0	$1,7\ 10^5$	$6,3\ 10^5$	$2,1\ 10^5$
4	10	50	10	0	$1,4\ 10^5$	$6,2\ 10^5$	$2,0\ 10^5$
5	10	20	10	0	$1,3\ 10^5$	$5,7\ 10^5$	$1,5\ 10^5$
6	10	50	7,5	0	$2,0\ 10^5$	$6,2\ 10^5$	$1,7\ 10^5$
7	10	50	5	0	$2,0\ 10^5$	$6,4\ 10^5$	$2,5\ 10^5$
8	10	50	2,5	0	$1,9\ 10^5$	$5,1\ 10^5$	$1,5\ 10^5$
9	10	50	2,5	1	/	$3,3\ 10^5$	$2,5\ 10^5$
10	10	50	5	1	$2,0\ 10^5$	$5,0\ 10^5$	$1,5\ 10^5$
11	10	50	7,5	1	$1,8\ 10^5$	$6,0\ 10^5$	$2,0\ 10^5$
12	10	50	10	1	$1,2\ 10^5$	$6\ 10^5$	$1,5\ 10^5$
13	10	50	5	0	$2,0\ 10^5$	$6,0\ 10^5$	$1,5\ 10^5$

3.3 Proof of concept

With the optimized conditions, first cell culture could be initiated to demonstrate the feasibility of such kinds of analysis and the utility of using GC×GC-TOFMS instead of GC-MS. The goal was to compare the VOC profiles of 2 cancer cell lines, adenocarcinomic human alveolar basal epithelial A549 (ATCC® CLL-185™) line and caucasian breast adenocarcinoma MCF-7 subtype bos cell line was kindly provided by Dr A.M. Soto (Department of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston), to demonstrate that some VOCs emitted by them were different. Cell-free control culture media headspace was also compared to the VOC signature of these cell lines. These cells lines were cultured in triplicate in Nunc EasYFlasks 75 cm^2 with caps equipped with filter membrane from Thermo Scientific (Belgium) with 20 mL of prepared culture medium. Cell cultures were placed in an incubator at 37 °C and 5% CO^2-enriched atmosphere. On the first day of the culture, the surface occupied by cells was around 20-30% for both cell lines. After 2 days, when the culture was at around 50% of confluency, the culture medium was changed and 20 mL of fresh DMEM was added (Figure 54). When the confluency reached around 80-90%, cell flasks were introduced inside a laminar hood where DMEM was removed and trypsin was used to detach cells from the surface of the flask. Then, they were recollected all together in a flask of 50mL with fresh DMEM to undergo centrifugation at 200 rpm for 5 minutes. DMEM was removed from the flask and the

cellular base was mixed with 3 mL of fresh DMEM. Two cryogenic vials from Sigma-Aldrich (Belgium) were filled each with 950 µL of medium containing cells and 50 µL of DMSO. They were put inside liquid nitrogen for preservation. The 0.9 mL remaining was split into three other new flasks to start a new culture with 25-30% of confluency and completed with 20 mL of fresh DMEM. This procedure was repeated once again for each of the cell lines and provided a total of 12 T-flasks per cell line (six around 50% of confluency and six around 80% of confluency). These vials were transported to the injection site and put on the tray of the MPS2 at room temperature waiting to be injected onto the Pegasus 4D. The SPME-CG×GC-TOFMS methods can be found in annex and were based on previous results.

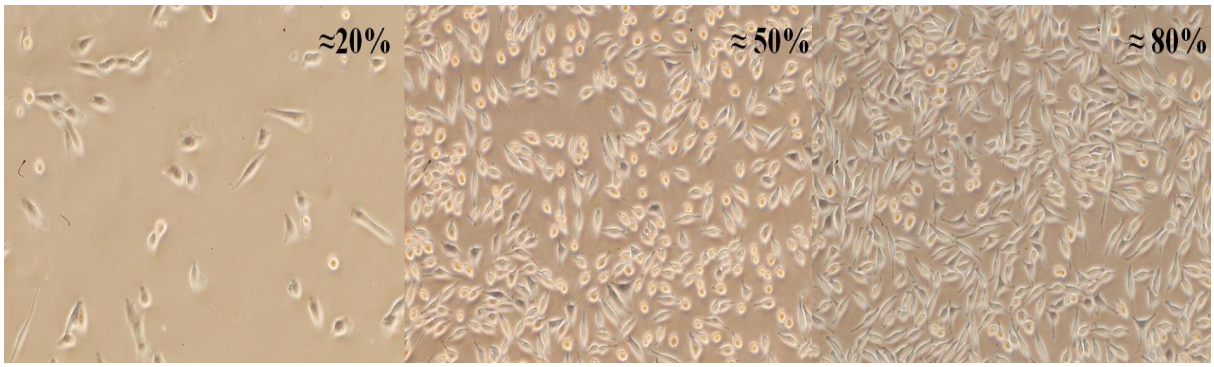

Figure 54: Confluency of MCF7 cell line for the first, the third and fifth day of culture

3.4 Headspace analysis of A549 and MCF-7

Firstly, this study investigated the stability of the sample in the SPME vial. As sextuplicate samples from same origins were collected together and left on the tray of the MPS at room temperature waiting to be injected, there might have been some difference between the first sample injected and the last one. Figure 55 displays the comparison of 2D-chromatograms of the first and the last sample injected of a row of samples with 10 hours gap between these two injections. Only small variations in the intensity of few features change in both chromatograms, meaning there is no alteration of sample one the MPS tray.

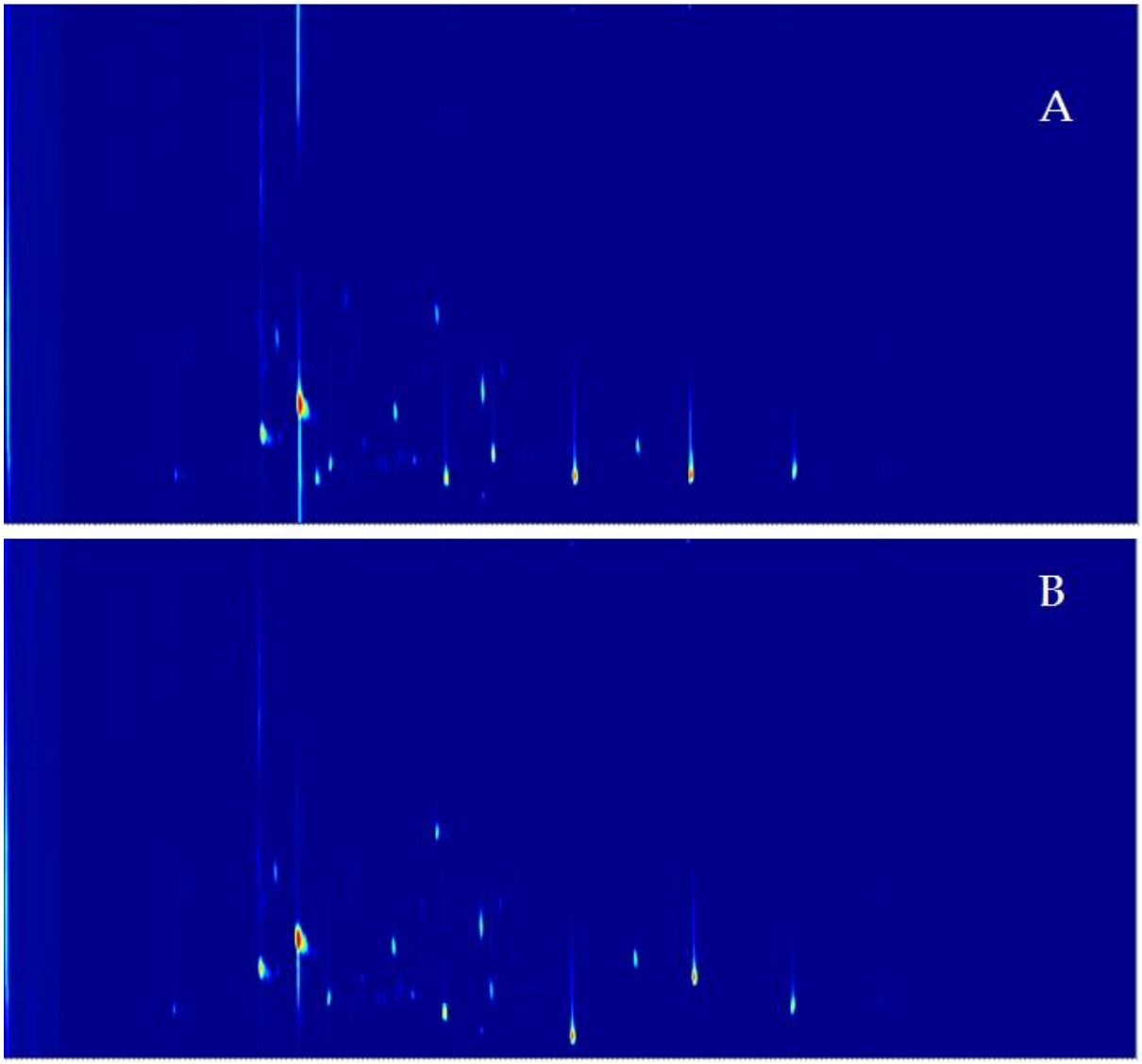

Figure 55: Comparison of the 2-D chromatograms of the first (A) and the last (B) samples of A549 to show the stability of samples on the tray before injection (10hr gap between these 2 injections)

The second part of this study demonstrates that it is possible to use SPME-GC×GC-TOFMS to analyse the headspace of cell culture media to detect some VOCs specific to cell lines. A total of 225 features were detected in the feature table created from all 2-D chromatograms samples with a signal to noise ratio of 100. Without any data treatment, this feature table was used to display a PCA (Figure 56). Clear differences between samples origins indicated that it was possible to characterize specific cells line VOCs from the headspace of the cell medium used during the growth.

As all features from the feature table were not relevant to one kind of sample, Fisher ratio with F_{crit} threshold was used only to select features specific to one population sample like with breath analysis study. This new feature table, including 116 features, was used to calculate box plot as explained in chapter 1. Only 15 features were shown to clearly belong to one group (Figure 57). These box-plots indicate that the hypothesis of observing VOCs belonging to cell thanks to the analysis of the headspace of their growing media is correct. The feature attribution name was not applied because the main goal of this study was to demonstrate that the difference of VOC profile found in the headspace of cell culture by SPME-GC×GC-TOFMS could be spotlight.

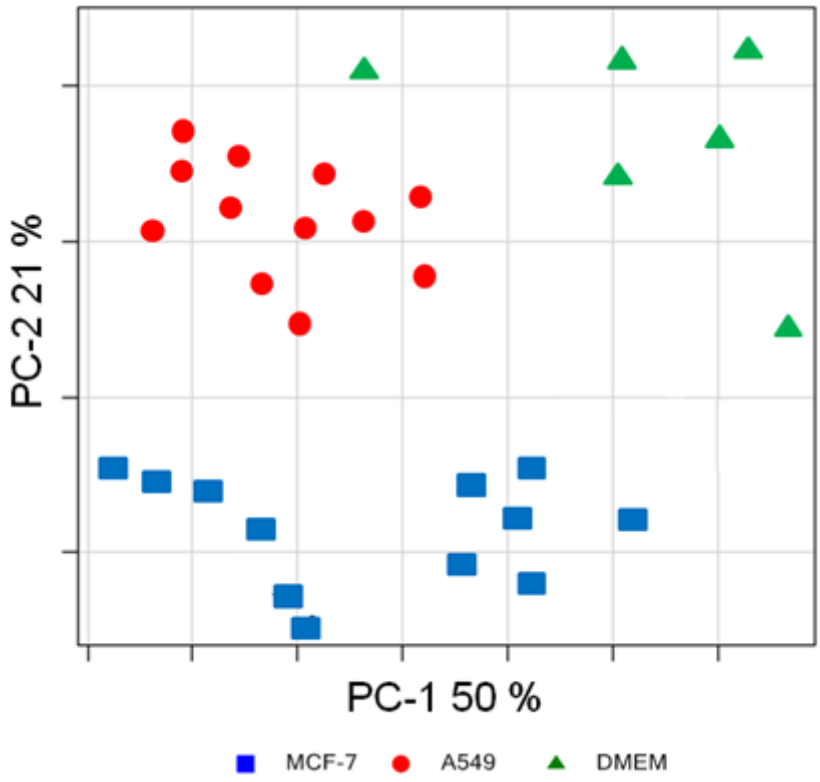

Figure 56: PCA calculated with all features from feature table generated with S/N=100

New PCA and HCA analysis were calculated based on the 15 highlight features (Figure 58). They displayed the same trend where samples were grouped according to their nature. In HCA analysis, one DMEM sample was grouped with MCF-7 samples but it was not the case of the PCA analysis. Principal component one gave a better representation of the variability for the PCA calculated with a short list of features compared to the first PCA, confirming the use of Fisher ratio and box-plot to extract essential information from the large amount of data initially present.

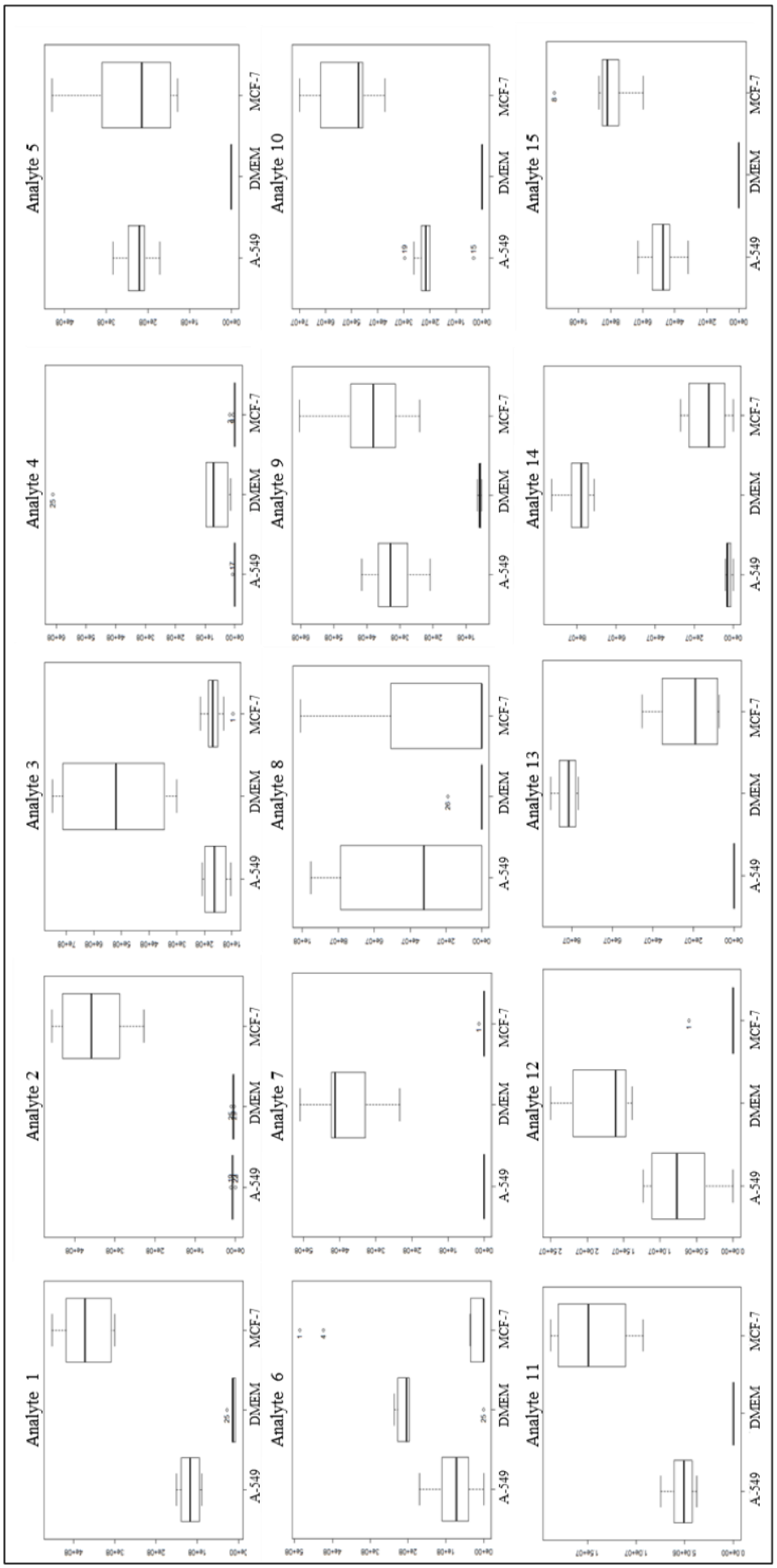

Figure 57: Box-plots of specific features belonging to A549, DMEM or MCF-7

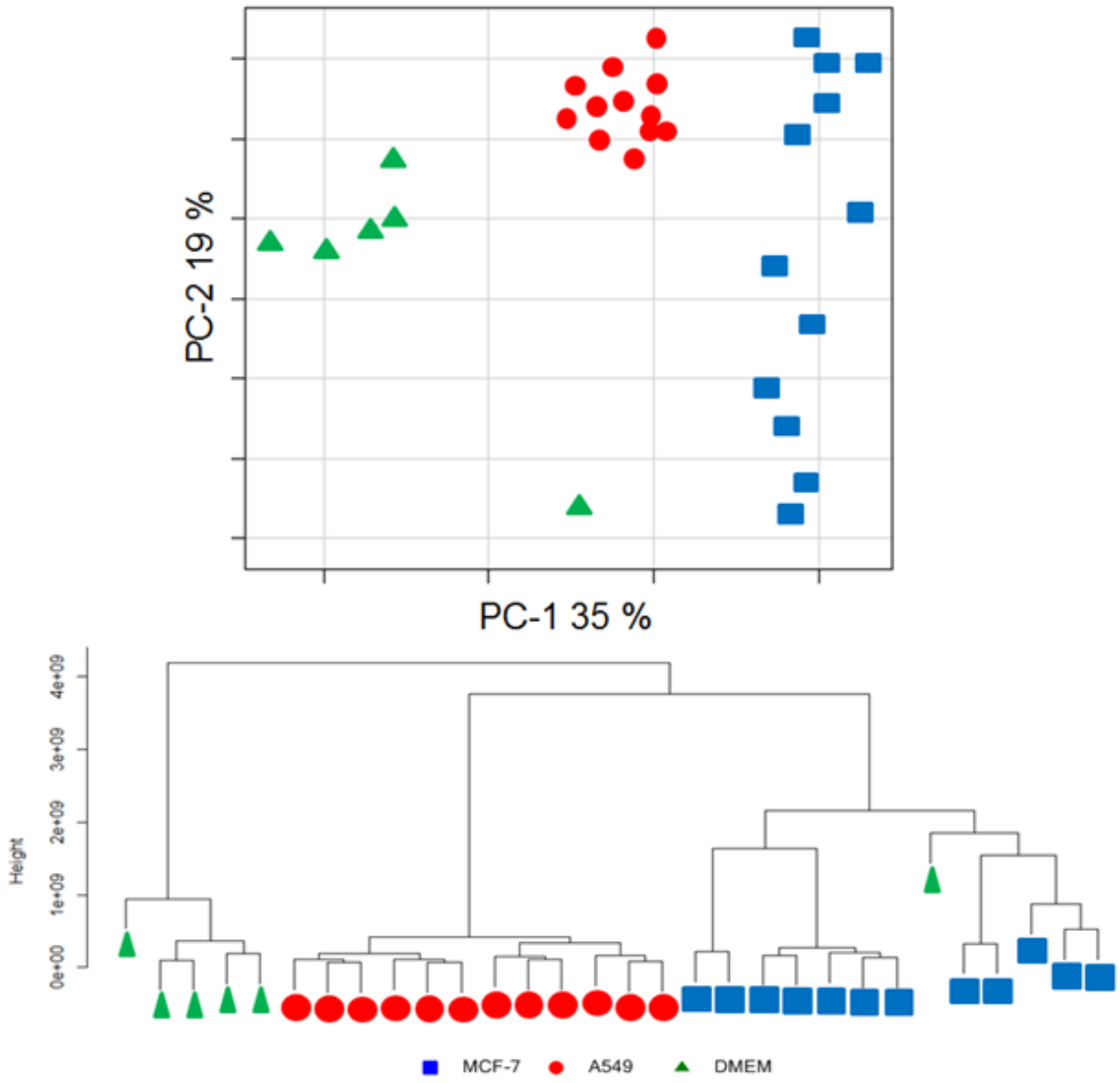

Figure 58: PCA and HCA performed with a restricted list of 15 features from cell analysis

To evaluate the choice of the column set used, reconstructed 2D-chromatogram display to which group features belong (Fisher ratio value higher than the critical Fisher value, Fisher ratio value lower than the critical Fisher value and undefined) and also display the intensity of Fisher ratio value (Figure 59). Thenceforth, visualization of features and features of interest could be obtained. Moreover, the number of undefined features represents almost 40% of the features, whereas for breath analysis study it was around 10%. Lower numbers of samples per class and higher number of classes lead more easily to no value for all samples from the same class for one feature. As explained before in chapter 2, this phenomenon conducts ChromaTOF® to attribute an undefined value in such a case.

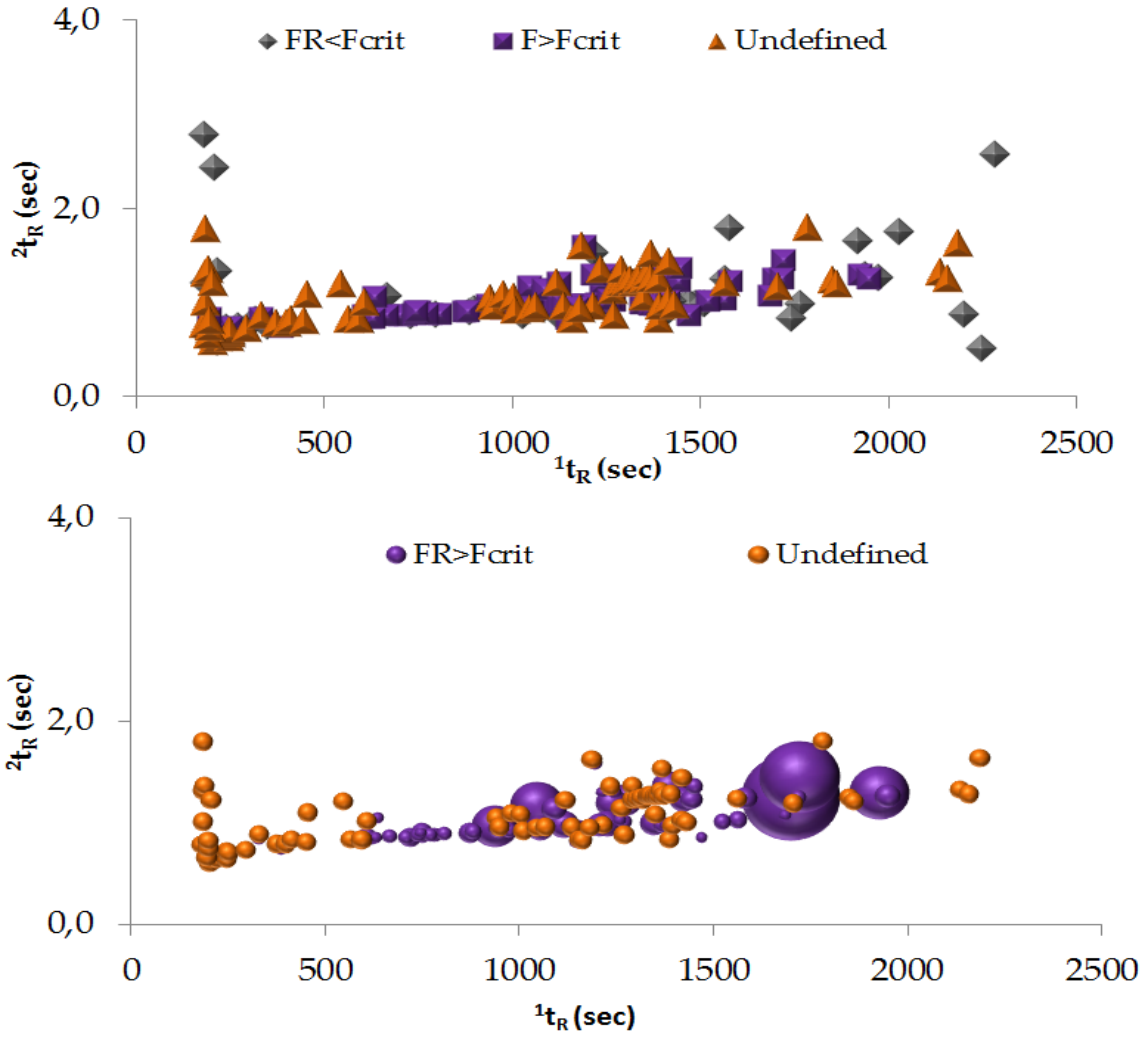

Figure 59 Reconstructed 2D chromatograms of all features detected in the headspace of A549, MCF7 and DMEM samples. 2D-Chromatogram on the top displays the nature of the FR (FR>F_{crit}; FR<F_{crit} and undefined features) and chromatogram on the bottom shows the intensity of the FR with the size of the feature. As the undefined features have no values, they are all given the same random size

To conclude this section, another interesting comparison was to investigate the VOC profile emitted from the same cell line but at a different confluency (50% and 80%). Cells did not interact in the same way in both situations (Figure 60). Reinforcing the hypothesis that cells interact with culture media and this interaction can be observed in the headspace of cells culture. For further analysis, it might be necessary to sample all different cells lines at the same confluency to minimize the intra variability of each type of sample. Counting cells after the passage step to get a "biological "normalization of samples may also lead to a reduction in their intra variability.

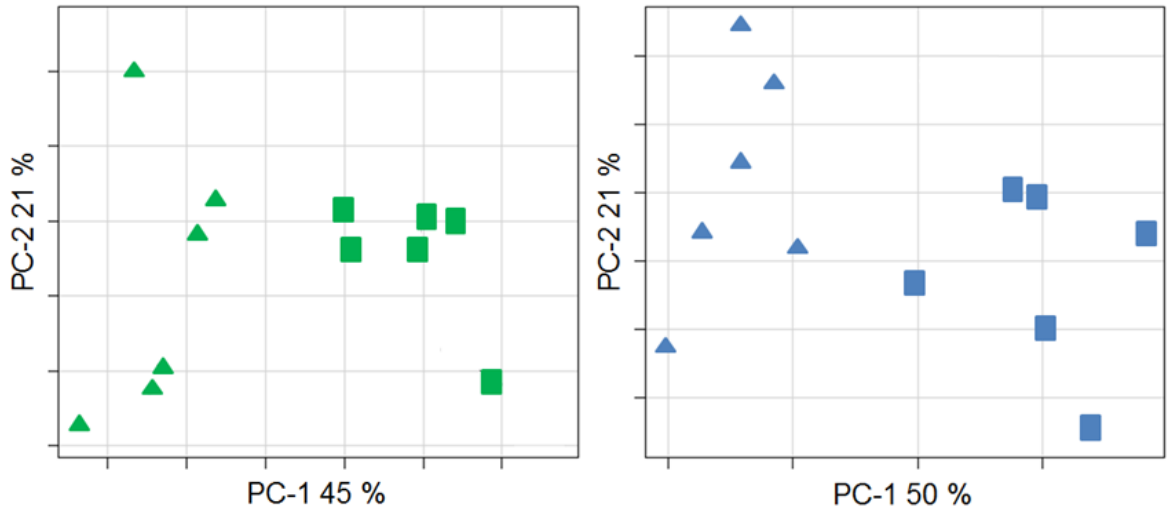

Figure 60: PCAs for same cell culture samples origin (A549 on the left and MCF7 on the right) for 50% (triangle) and for 80% (square) of confluency

This section confirms the hypothesis that it is possible to highlight specific VOCs emitted by cell lines using SPME-GC×GC-TOFMS and allowed us to go deeper in this direction.

3.5 Headspace analysis of six cell lines

To isolate specific VOCs emitted by cell culture, six cell lines were chosen (Table 8) MCF-7 breast cancer cell line and A549 lung cancer cell line were again used. BZR (ATCC® CRL-9483™) and Calu-1 (ATCC® HTB-54™) were also added to get a total of three lung cancer cell lines. BEAS-2B (BEAS-2B ATCC® CRL-9609™) was the only normal cell line investigated. BZR comes initially from BEAS-2B and was transformed into cancer cell line by using a virus. Thus, these two cell lines are only different in terms of the cancer/normal cell criteria and promised interesting comparisons. Finally, HCT-116 colon cell line (HCT 116 ATCC® CCL-247™) was investigated to compare cancer from different kind of tissue. All cell lines came from epithelial tissues to minimize the variation between them.

Table 8: Characteristics of the six cell lines investigated

Cell lines	Tissue	Morphology	Disease
A549	Lung	Epithelial	Carcinoma
BEAS-2B	Lung	Epithelial	Normal
BZR	Lung	Epithelial	Carcinoma
Calu-1	Lung	Epithelial	Carcinoma
HCT 116	Colon	Epithelial	Carcinoma
MCF7	Breast	Epithelial	Adenocarcinoma

3.5.1 Sampling process

Cell lines were cultivated in triplicate with the same cell growing protocol as used before. Sampling was executed at around 80% of confluency just before the passaging step. 5 mL from each flask were transfer inside SPME vials of 20 mL. These vials were previously heated to 120°C for 15 minutes for the sterilization process. After the passaging step, one group of cells was used for cryo-conservation, one other group was put back inside the cell flask for new cultivation and one other group was used to calculate the number of cells to use "biological normalization " during the data processing. Cells were counted by means of the trypan blue coloration method (Table 9). Results revealed a high dispersion of the number of cells collected during the passage step. Hence, using "biological normalization " might help to normalize the data and extract information. New cultures were launched with around 30% of confluency and repeated twice to get a total of 9 samples per cell lines (Figure 61). Nine blank growth media samples were also collected using the same protocol.

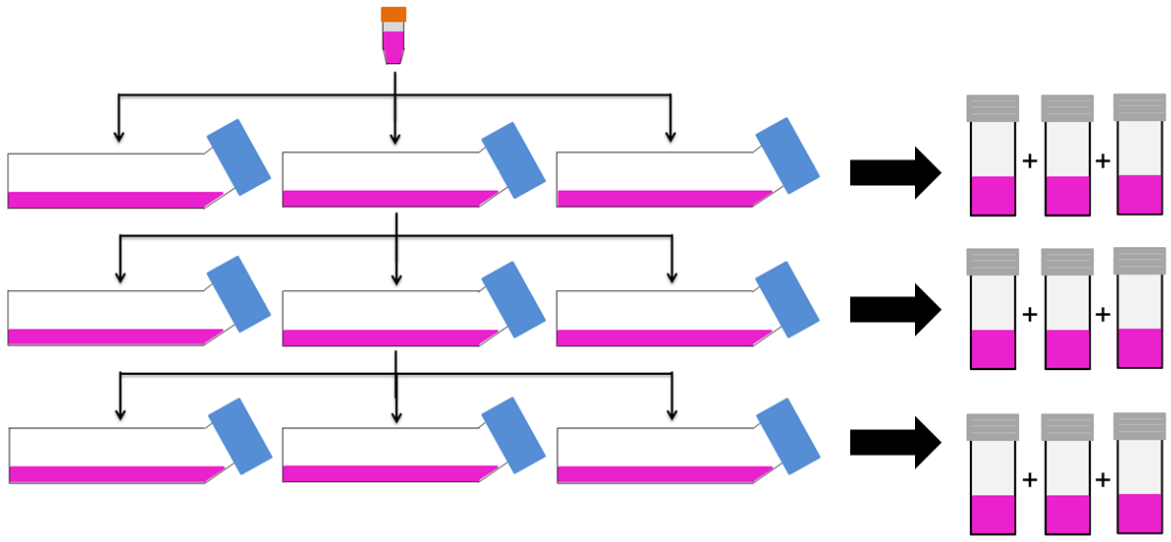

Figure 61: Representation of the design for the culture and sampling for one random cell line

Table 9: Number of cell at 80% of confluency for each triplicate

A549	BEAS-2B	BZR	Calu-1	HCT 116	MCF 7
5 396 000	1 152 550	998 200	966 000	5 965 000	1 590 000
3 574 000	2 274 666	1 940 666	3676 666	6 547 000	7 276 000
5 241 000	4311 000	4 392 000	4 569 000	6 251 000	7 792 000

3.5.2 Cell data Normalization

All vials were transported to the of injection site and put into a bath at 37° C waiting to be injected onto the AccuTOF™ GCv 4G. Before injection, the internal standardization method adapted from Zhao *et al* was used to normalize feature areas during the data treatment [133]. 1mL of 100ppm bromobenzene prepared with methanol (HPLC grade, Sigma-Aldrich) was introduced into a SPME vial of 20 mL at 37° C. The fiber was introduced 30 seconds in the vial before being exposed to the sample during 40 minutes at 37° C. Then the fiber moved to the inlet of the AccuTOF™ GCv 4G. Parameters used for these injections and data processing are detailed in the annex.

The feature table generated was exported to Excel® to follow normalization with the internal standard. The stability of the signal for bromobenzene was investigated and indicated low variation of its intensity (Figure 62). Therefore, the normalization process based on it can be applied.

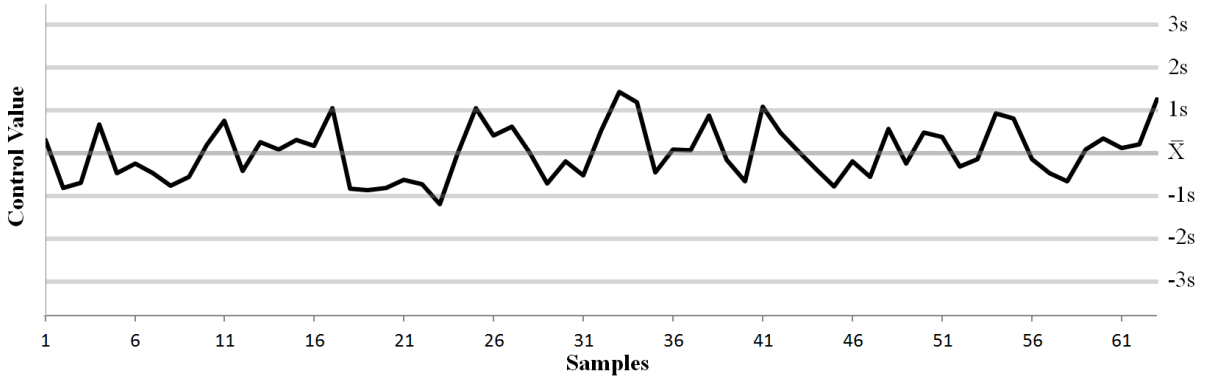

Figure 62: Stability of bromobenzene feature intensity across all samples

Concentration of VOCs from the headspace of culture media depends on the number of cell present in the media. And cells from different cell lines do not grow exactly at the same speed. A second normalization was carried out with the number of cells for each triplicate to delete the impact of the number of cells per sample.

With the AccuTOF™ GCv 4G instrument, it was possible to detect around 370 features per cell line. A total of 544 features were detected in the feature table obtained from the cumulative 2D-chrommatogram after removing column bleed. A 2D-Chromatogram example from BZR cell lines show the complexity of such kinds of matrix and justify the use of two dimensional techniques to characterize them (Figure 63).

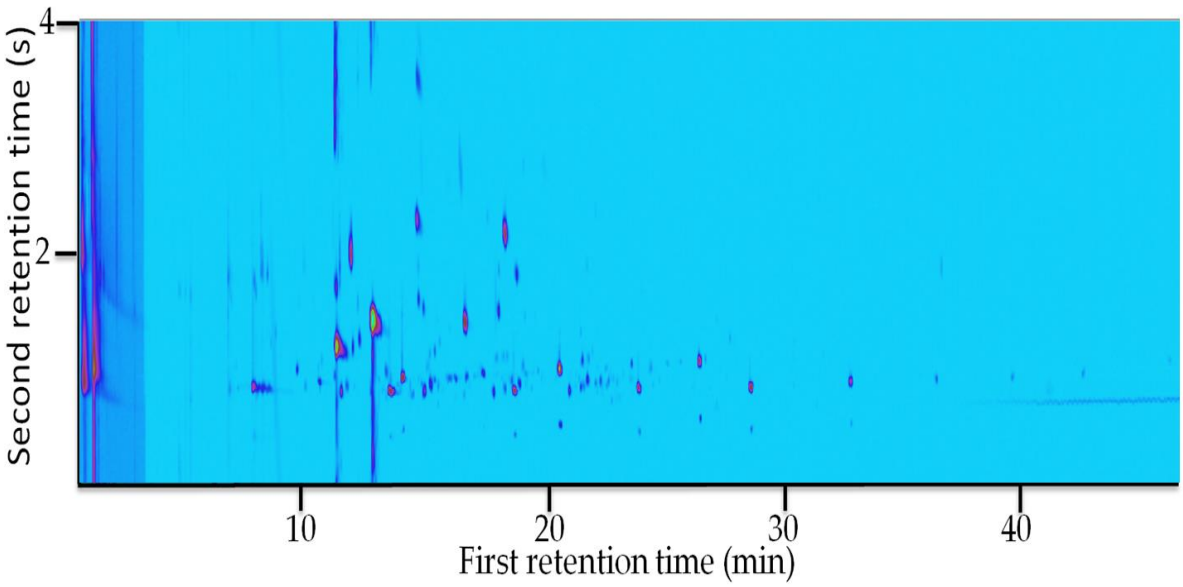

Figure 63: 2-D chromatogram of the BZR cell line headspace

3.5.3 Intra cell variability

Then, as the goal of this study was to compare the VOC profile emitted by the different cell lines investigated, variation inside one group of cell lines was also reported to demonstrate the rightness of such comparison. The cell line selected to verify this aspect was BZR cell line as it showed the most similarities with other cell lines. In a first triplicate of BZR, around 300 features were detected in each 2D-chrommatograms. Low variations between them can be observed (Figure 64), leading to good repeatability of signal record from different samples with the same biological origin. Another verification of the repeatability of samples through the different triplicates of BZR cell lines was also investigated. To lead this second verification, different triplicates were compared with each other with their average feature volume inside their own triplicate. A higher dispersion of feature volume was observed compared to the variation inside triplicate (Figure 64). Cells can undergo small mutations between cell passaging and this may explain the origin of this higher variation. However, variations between all samples for the same cell line are totally acceptable and make it possible to compare different cell lines with each other.

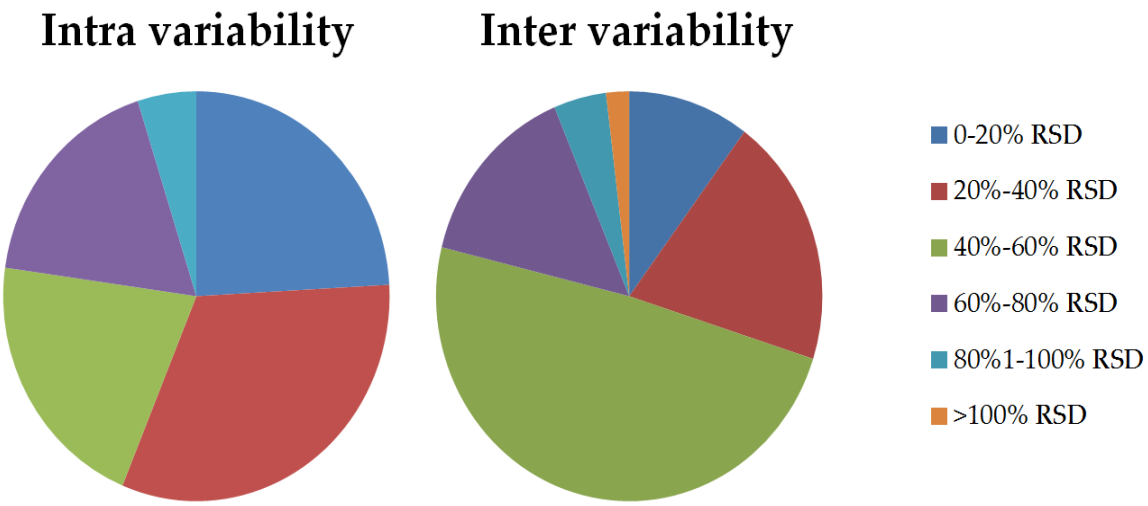

Figure 64: Intra and inter variation of BZR triplicates for each feature detected in their headspace

3.5.4 Comparison of the headspace of the six cell lines investigated

As BZR and BEAS-2B are the same cell line only differentiated by the disease status (respectively carcinoma and normal), special attention was paid to these 2 cell lines. No major variation can be observed between them (Figure 65). Based on the result, the VOC profile emitted by cells are not totally influenced by the type cancer/non-cancer. As explained earlier in this document, some papers discuss the fact that the exhaled breath of cancer patients are different from healthy patients and try to detect which VOCs are responsible for such segregation [197]. According to the results, this difference may come from the interaction of the cancer cells with the organism and not directly from the cancer cells.

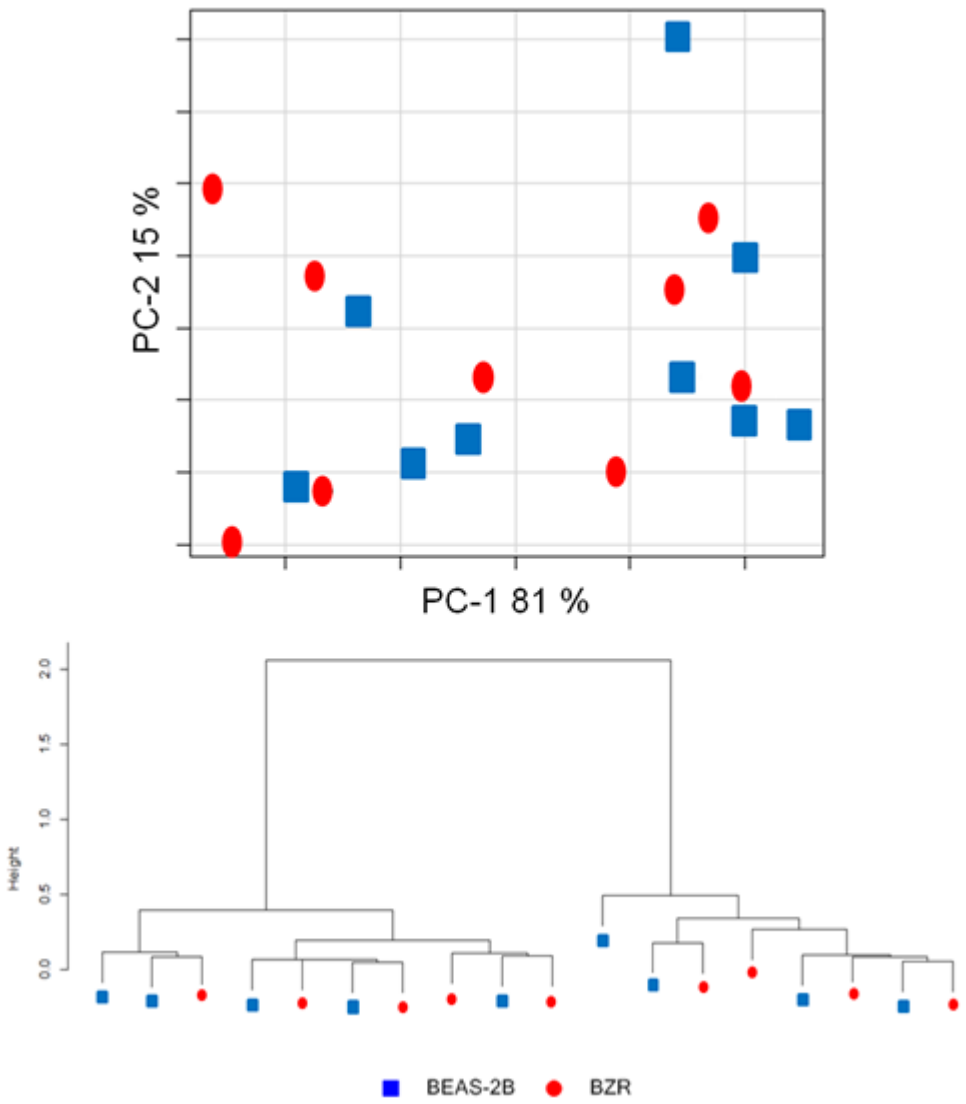

Figure 65: PCA (on the top) and clustering (on the bottom) analysis for BZR and BEAS-2B cell lines

3.5.5 General trend for the six cell lines

PCA and HCA of all headspace samples display grouping of samples from the same origin (Figure 66). Confirming it is possible to characterise cell lines through the headspace of the media needed for their growth and also that the repeatability of the samples from same cell lines is small enough to see a difference between cell lines. It is important to notice relatively similar dispersion for BZR and BEAS-2B. Thus, the kind of cell line is more important than the cancer/normal factor to segregate cell lines according to their headspace analysis. Moreover, based on results gave by such statistical tools, we observed that it is not possible to directly separate cell lines from different tissue by their headspace analysis. Indeed, breast, colon and lung cell lines are not grouped separately.

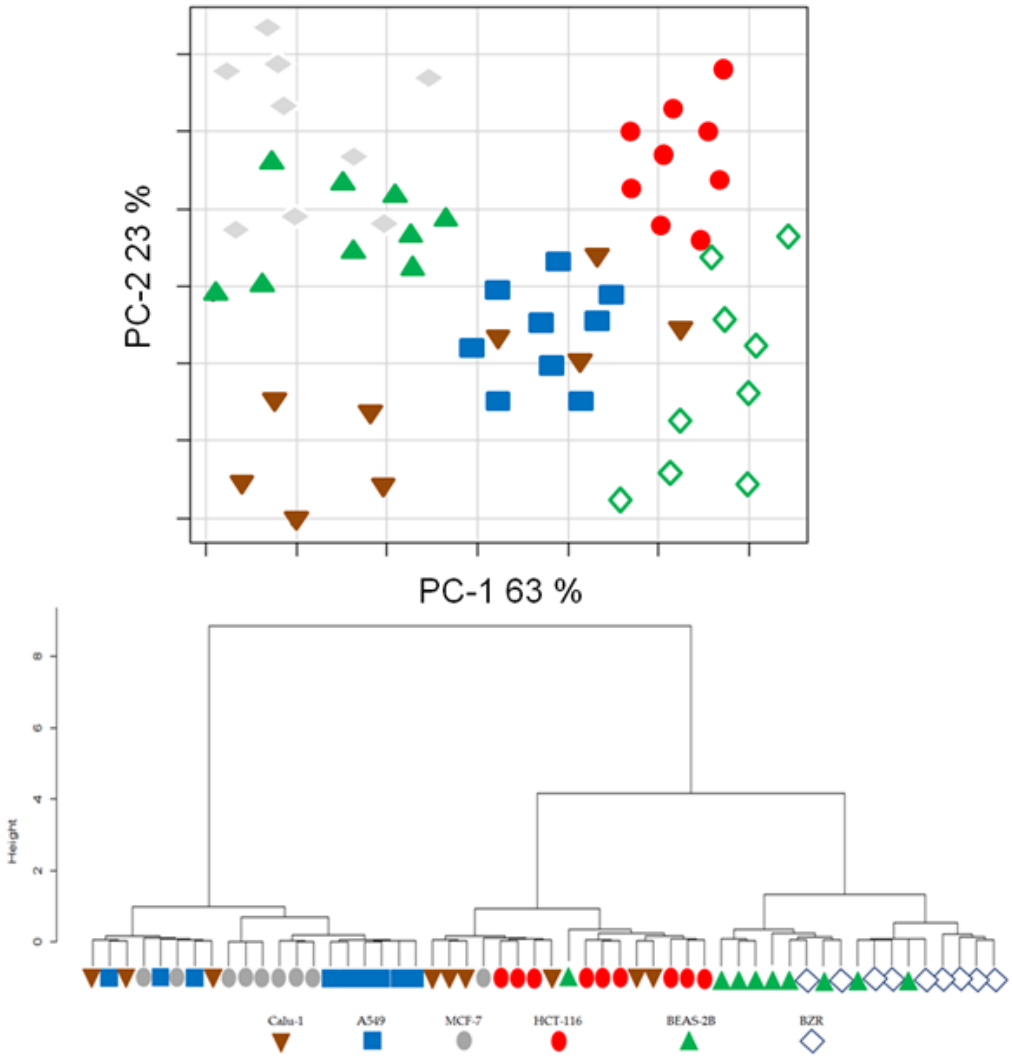

Figure 66: PCA and HCA for the feature table obtained from the 6 cell lines investigated

To find interesting VOCs able to make the difference between cell line categories, statistical tools were used. The first step was to use Fisher ratio or random forest as with breath analysis. However, the large number of classes present in this analysis induces large amounts of the highlight features. Hence, it was not possible to use such statistical tools. The initial feature table of 544 features was screened feature by feature with box-plots.

It revealed that 2,2,5-trimethyl-3,4-hexanedione, 3-buten-2-ol and 3-decanone were only found in all lung cancer cell lines (Figure 67). As 3-Buten-2-ol can follow a tautomeric reaction to form 2-butanone, specific lung cancer VOCs are all ketone. The over activation of the oxidase enzyme call "cytochrome p450" which catalyses the oxidation of organic substances are well known for lung cancer and may be related to the presence of ketone emitted by lung cancer cells.

1,3-dimethoxypropan-2-yl stearate, 2-decanone, 2-ethyl-1-hexanol, 2-octanone, 3-heptanone, 3-nonanone, cyclohexanol, nonanol, octanol and *p*-xylene were only detected in HCT-116. Expected for *p*-xylene, all specific VOCs from colon cell lines own oxygen species. This is not the case for MCF-7 cell line, where only 1-propanol contains an oxygen atom compare to 6-methyloctene, 2,2-dimethylpropanol, , 2,4-dimethylheptene, benzaldehyde, 2-nitro-, diaminomethylidenhydrazone and 2,7-dimethyl-furo[2,3-c]pyridine. However, as only one cell line for breast and colorectal were investigated in this study, it hazardous to try to have a trend of the VOC profile emitted by such cancer tissue. Characterizations with other cell lines of theses tissues have to be accomplished in order to do it.

Acetone and 3-ethyl-3-methylnonadecane were detected in all cancers cell lines investigated in this study. They may be due to a common sub product of metabolization only occurring in cancer cells. Acetone is related to lipolysis, dextrose metabolism and decarboxylation of acetoacetate, which is derived from lipid peroxidation. Lipid peroxidation is also known to generate a straight-chain of hydrocarbons from polyunsaturated fatty acids. Phillips established a list of nine VOCs to differentiate the exhaled breath of lung cancer patients and healthy volunteers mainly based on alkanes compounds [117]. The presence of 3-Ethyl-3-methylnonadecane in cancers cells may be explained and corroborated by them. These two VOCs and the three related to lung cancer are displayed with box-plot to visually indicate their specificity to some cell lines (Figure 67).

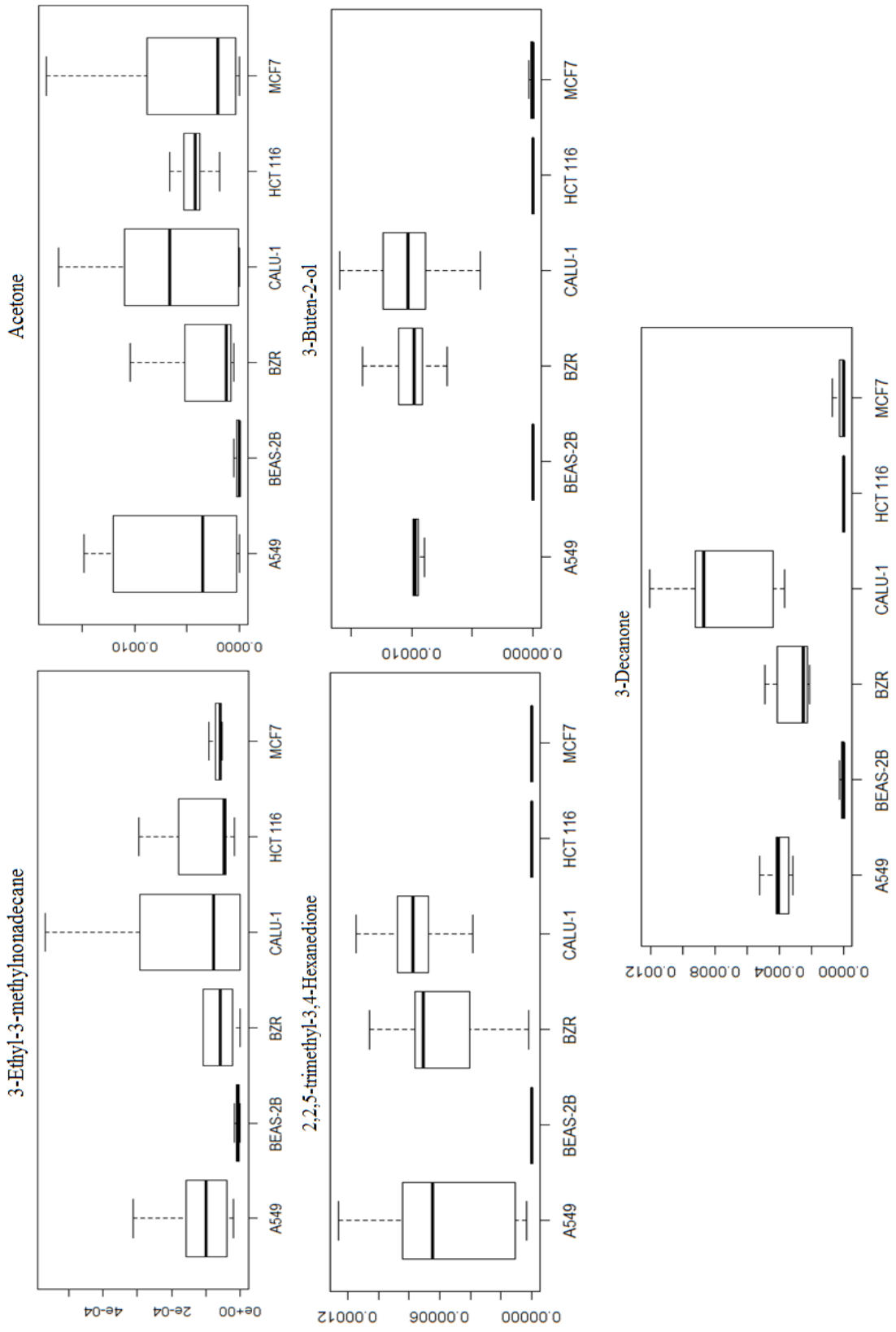

Figure 67: Box plot for the 3 VOCs found in all lung cancer cell lines and also for the 2 VOCs found in all cancer cell lines

3.6 Comparison with Breath Analysis

From the list of putative markers established in chapter 2 to distinguish healthy patients from lung cancer patients. There is no match between them and those detected in this chapter. This could be due to several factors. The sampling method used for breath and headspace cell culture was different. Using TD for both samples may lead to more comparable results. Although *in vitro* analysis makes it easier to study cells, there are some shifts with *in vivo* conditions because the environmental aspect is not taken into account with *in vitro* study. Those shifts may explain why there are no matches between the two central lines of this thesis. Additional normal cell lines should be investigated to also highlight markers emitted by normal cell lines and then try to make a relation between them and those found in the exhaled breath of healthy volunteers. Alkane VOCs like decane, eicosane, hexadecane and 2-methyl-heptadecane have been found in exhaled breath from lung cancer patients and in the headspace of cancer cell lines. This chemical family should be investigated in other studies to try to show the direct relation between cancer cells and VOCs from the exhaled breath.

Chapter 4:

Conclusion and future work

Conclusion

The goal of this thesis was to help in the development of an early and non-invasive diagnosis of lung cancer by using breath analysis. The innovative part was the use of GC×GC-TOFMS instead of conventional 1D-GC-MS to analyse such sample matrices.

Gas samples were collected in Tedlar® bags and transferred into TD tubes to use thermal desorption. The investigation of the membrane permeability of Tedlar® bags revealed that VOCs from the environment could migrate inside the bag, changing the sample integrity. It is therefore recommended to reduce the storage time as much as possible. Then, exhaled breath samples from healthy volunteers and lung cancer patients were compared to determine which VOCs were able to distinguish these two populations. A cohort of 29 samples was used to demonstrate the utility of TD-GC×GC-TOFMS on such samples. The main challenge was to extract the putative VOC markers of lung cancer through the large amount of data generated by 2D-Chrommatograms. The impact of exogenous VOCs on the breath profile was pointed out and three strategies were imagined to negate their impact. Subtraction of blank from the sampling room to the breath sample and breathing medical air before sampling to flush lungs did not give good results. A robust statistical approach however provided good classification performances, overpassing limitations such as environmental contamination and sampling variability. By using a proper pre-processing step, it was possible to neglect the influence of the exogenous VOCs. Two different data analysis strategies (FR and RF) were applied to digest the initial data set and PCA to point out the capacity of such statistical tool. From a feature table of 1350 features, a total of 37 features were detected with both methods and allow distinguishing the two populations of individuals. Among them, six features were detected by both statistical approaches and allowed to separate the two populations. This part of the study underlined the need for analytical controls from sampling to data processing in untargeted biological studies.

The second part of this study focused on the study of VOC profiles emitted by lung cancer cell lines in order to compare them with breath analysis profiles. Firstly, a sampling design using SPME fiber to trap VOC of A549 and MCF-7 from their growing media was conducted, including constraints related to the sterilized conditions required by cell culture to avoid any biological contaminations. SPME parameters were optimized and GC×GC-TOFMS methods from breath analysis were

transposed and produced good chromatography results. A proof of concept revealed with PCA calculation that the headspace from culture media, from one breast adenocarcinoma cell line (MCF7) and from one adenocarcinomic human alveolar cell line (A549) clearly indicated differences of the nature of VOC emitted. With box-plots calculation, it was possible to highlight which VOCs were responsible for such headspace composition differentiation. This proof of concept also underlined the value of GC×GC analysis due to the high amount of VOC collected in the headspace of cell cultures. It was shown that the number of cell present during the collection of the culture media influenced the headspace composition. It was therefore necessary to standardize the data obtained as a function of the number of cells present during sampling. Another standardization with bromobenzene was also used when introducing SPME fiber into the headspace in order to reduce instrumental variability. A study of the intra and inter variability of the BZR cell line between the various samples collected has shown that it would not interfere with the variability present between the different types of cell lines studied.

Finally, this part of this study investigated and compared the headspace of one lung normal cell line, three lung cancer cell lines and two other cancer cell lines (from breast and colon) to point to the presence of ketone VOCs for lung cancer cell lines and alkane VOCs for general cancer cell lines. Comparison of normal (BEAS-2B) and cancer lung (BZR) clone cell lines indicated low variation between them, leading to low dependence of the cancer or normal type on the VOC profile emitted. Unfortunately, it was not possible to link these results with those of breath analysis.

Future work

The next objective is to validate the compromising highlighted VOCs with other breath samples or moreover to use chemical family as an alternative to full identification of large data matrix. Next exhaled breath samples should be collected over long periods of time and in different hospitals to reduce the impact of exogenous compounds. Quantification of such VOCs also needs to be performed but required a more sophisticated sampling process. Indeed, in this study, Tedlar® bags were not calibrated to be filled with the exact same volume, impacting the number of VOCs trapped in the TD tubes. Moreover, as VOCs of interest are located in the lung and the few first seconds of the expiration removes the air from the upper part of the respiratory system, long or short expiration when filling the bag has also an impact on the VOC concentration. It is possible to overcome this issue by calculating the

amount of CO_2 emitted to normalize the volume of exhaled breath collected. With reference to what Wang suggested in 2012, a standard sampling method for breath analysis should be established to allow easier comparison between all studies. ReCIVA® breath sampler from Owlstone medical appear to be the standardize method to collect breath sample since end of 2017. This device combines a reliable, reproducible and universal sampling system, which can be used to reduce the variability between studies.

The utilization of high resolution mass spectrometry technics to obtain accurate mass information, close to ppm level, will also help to identify putative biomarkers with a higher confidence as it measures the exact mass. Then, after biomarkers detection and validation, implementation of specific detectors, such as e-nose, sift-MS, PTR-MS or fast GC-MS, for VOC markers to facilitate their use in a clinical environment also need to be investigated.

In parallel to exhaled breath biomarkers detection and validation, other studies focused on cell cultures using other sampling methods or other cell lines to corroborate results already reported and link the VOCs detected during breath analysis. The idea of transposing the strategy used to detect biomarkers in the exhaled breath by comparing a population of healthy volunteers and lung cancer patients can also be considered for the cell culture. The comparison between BZR and BEAS-2B cell lines has shown that the headspace composition of these two cell lines is very similar. But it is nevertheless possible to observe differences. A promising idea would be to carried out primary cell culture from normal and cancer tissue from the same individual. This would reduce the variability between the two types of cell to focus only on the cancer/non-cancer character. This ambitious approach must, however, deal with the extraction and purification of the cells from the tissues. This strategy can also jointly investigate the composition of the exhaled breath of a particular lung cancer patient and *in vitro* release of compounds from his/her lung cancer cells obtained from tumor resection.

References

1. Elmore, S., *Apoptosis: A Review of Programmed Cell Death.* Toxicol Pathol, 2007. **35**(4): p. 495–516.

2. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2017.* CA: A Cancer Journal for Clinicians, 2017. **67**(1): p. 7-30.

3. Pastorino, U., *Early detection of lung cancer.* Respiration, 2006. **73**(1): p. 5-13.

4. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2018.* CA Cancer J Clin, 2018. **68**(1): p. 7-30.

5. Ma, W., et al., *Determination of breath gas composition of lung cancer patients using gas chromatography/mass spectrometry with monolithic material sorptive extraction.* Biomed Chromatogr, 2015. **29**(6): p. 961-5.

6. Huo, D., et al., *A novel optical chemical sensor based AuNR-MTPP and dyes for lung cancer biomarkers in exhaled breath identification.* Sensors and Actuators B: Chemical, 2014. **199**: p. 446-456.

7. Boedeker, E., G. Friedel, and T. Walles, *Sniffer dogs as part of a bimodal bionic research approach to develop a lung cancer screening.* Interact Cardiovasc Thorac Surg, 2012. **14**(5): p. 511-5.

8. Mazzone, P.J., *Analysis of Volatile Organic Compounds in the Exhaled Breath for the Diagnosis of Lung Cancer.* J Thorac Oncol, 2008. **3**(7): p. 774-780.

9. Filipiak, W., et al., *TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(1): p. 182-95.

10. Collins, L.G., et al., *Lung Cancer: Diagnosis and Management.* American Family Physician, 2007 **75**(1).

11. Yao, S., G. DellaVentura, and C. Petibois, *Analytical characterization of cell-asbestos fiber interactions in lung pathogenesis.* Anal Bioanal Chem, 2010. **397**(6): p. 2079-89.

12. Richqrd Doll, A.B.H., *Smoking and Carcinoma of the Lung.* British medical journal, 1950: p. 739-748.

13. Jha, P., et al., *21st-century hazards of smoking and benefits of cessation in the United States.* N Engl J Med, 2013. **368**(4): p. 341-50.

14. Jemal, A., et al., *Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control.* J Natl Cancer Inst, 2008. **100**(23): p. 1672-94.

15. Thun, M.J., et al., *50-year trends in smoking-related mortality in the United States.* N Engl J Med, 2013. **368**(4): p. 351-64.

16. Witschi, H., *A Short History of Lung Cancer.* TOXICOLOGICAL SCIENCES, 2001. **64**: p. 4-6.

17. Buszewski, B., et al., *Analytical and unconventional methods of cancer detection using odor.* TrAC Trends in Analytical Chemistry, 2012. **38**: p. 1-12.

18. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* Int J Cancer, 2010. **127**(12): p. 2893-917.

19. Borgerding, M. and H. Klus, *Analysis of complex mixtures – Cigarette smoke.* Experimental and Toxicologic Pathology, 2005. **57**: p. 43-73.

20. Talhout, R., et al., *Hazardous compounds in tobacco smoke.* Int J Environ Res Public Health, 2011. **8**(2): p. 613-28.

21. Weiss, W., *Cigarette Smoking and Lung Cancer Trends: A Light at the End of the Tunnel?* Chest, 1997. **111**(5): p. 1414-1416.

22. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.

23. Bunn, P.A., Jr., *Karnofsky Award 2016: A Lung Cancer Journey, 1973 to 2016.* J Clin Oncol, 2017. **35**(2): p. 243-252.

24. Torre, L.A., R.L. Siegel, and A. Jemal, *Lung Cancer Statistics.* Adv Exp Med Biol, 2016. **893**: p. 1-19.

25. Kamp, D.W., *Asbestos-induced lung diseases: an update.* Transl Res, 2009. **153**(4): p. 143-52.

26. Case, B.W., et al., *Applying definitions of "asbestos" to environmental and "low-dose" exposure levels and health effects, particularly malignant mesothelioma.* J Toxicol Environ Health B Crit Rev, 2011. **14**(1-4): p. 3-39.

27. Huang, S.X., et al., *Role of mutagenicity in asbestos fiber-induced carcinogenicity and other diseases.* J Toxicol Environ Health B Crit Rev, 2011. **14**(1-4): p. 179-245.

28. Thun, M.J., et al., *50-year Trends in Smoking-Related Mortality in the United States.* N Engl J Med, 2013. **368**(4): p. 351-364.

29. Samet, J.M., *Radiation and cancer risk: a continuing challenge for epidemiologists.* Environmental Health, 2011. **10**(4).

30. Melloni, B. and F. Bonnaud, *Le radon : un agent carcinogène pulmonaire professionnel et domestique.* Revue des Maladies Respiratoires, 2005. **22**(4): p. 571-575.

31. Barros-Dios, J.M., et al., *Residential radon exposure, histologic types, and lung cancer risk. A case-control study in Galicia, Spain.* Cancer Epidemiol Biomarkers Prev, 2012. **21**(6): p. 951-8.

32. Turner, M.C., et al., *Radon and lung cancer in the American Cancer Society cohort.* Cancer Epidemiol Biomarkers Prev, 2011. **20**(3): p. 438-48.

33. Jiang, J.Q., et al., *Arsenic contaminated groundwater and its treatment options in Bangladesh.* Int J Environ Res Public Health, 2012. **10**(1): p. 18-46.

34. Subramanian, J. and R. Govindan, *Lung cancer in never smokers: a review.* J Clin Oncol, 2007. **25**(5): p. 561-70.

35. Marshall, A.L. and D.C. Christiani, *Genetic susceptibility to lung cancer--light at the end of the tunnel?* Carcinogenesis, 2013. **34**(3): p. 487-502.

36. Silvestri, G.A., et al., *The Noninvasive Staging of Non-small Cell Lung Cancer.* chest, 2003. **123**(1): p. 147-156.

37. Anderson M.D., *CANCER CARE SERIES*, ed. Springer. 2003.

38. Bach, P.B., et al., *Benefits and harms of CT screening for lung cancer: a systematic review.* JAMA, 2012. **307**(22): p. 2418-29.

39. Mazzone, P.J., et al., *Lung cancer screening with computer aided detection chest radiography: design and results of a randomized, controlled trial.* PLoS One, 2013. **8**(3): p. e59650.

40. Wender, R., et al., *American Cancer Society lung cancer screening guidelines.* CA Cancer J Clin, 2013. **63**(2): p. 107-17.

41. Saleh Dammas, E.F.P.J., Philip C. Goodman, *Identification of small lung nodules at autopsy: implications for lung cancer screening and overdiagnosis bias.* Lung Cancer, 2001. **33**: p. 11-16.

42. Chang, C.F., A. Rashtian, and M.K. Gould, *The use and misuse of positron emission tomography in lung cancer evaluation.* Clin Chest Med, 2011. **32**(4): p. 749-62.

43. Yasufuku, K. and S. Keshavjee, *Staging Non–Small Cell Lung Cancer.* Clinical Pulmonary Medicine, 2010. **17**(5): p. 223-231.

44. Mukhopadhyay, S., *Utility of small biopsies for diagnosis of lung nodules: doing more with less.* Mod Pathol, 2012. **25 Suppl 1**: p. S43-57.

45. Thakkar, M.S., F. von Groote-Bidlingmaier, and C.T. Bolliger, *Recent advances in therapeutic bronchoscopy.* Swiss Med Wkly, 2012. **142**: p. w13591.

46. Hasan, N., R. Kumar, and M.S. Kavuru, *Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers.* Lung, 2014. **192**(5): p. 639-48.

47. Marzorati, D., et al., *A review of exhaled breath key role in lung cancer diagnosis.* J Breath Res, 2019.

48. Guo, A., et al., *Signaling networks assembled by oncogenic EGFR and c-Met.* Proceedings of the National Academy of Sciences, 2008. **105**(2): p. 692-697.

49. Yanagisawa, K., et al., *A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer.* J Natl Cancer Inst, 2007. **99**(11): p. 858-67.

50. Maheswaran, S., et al., *Detection of mutations in EGFR in circulating lung-cancer cells.* N Engl J Med, 2008. **359**(4): p. 366-77.

51. Ngoepe, M., et al., *Integration of biosensors and drug delivery technologies for early detection and chronic management of illness.* Sensors (Basel), 2013. **13**(6): p. 7680-713.

52. Kusano, M., E. Mendez, and K.G. Furton, *Comparison of the volatile organic compounds from different biological specimens for profiling potential.* J Forensic Sci, 2013. **58**(1): p. 29-39.

53. Yanagita, K., et al., *Cytoskeleton-Associated Protein 4 Is a Novel Serodiagnostic Marker for Lung Cancer.* The American Journal of Pathology, 2018. **188**(6): p. 1328-1333.

54. Waseem K Jerjes, et al., *The future of medical diagnostics: review paper.* Head & Neck Oncology, 2011. **3**(38).

55. Amann, A., et al., *The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva.* J Breath Res, 2014. **8**(3): p. 034001.

56. Margaret J. Henderson, B.A.K., and Gerald A. Wrenshall,, *Acetone in the Breath: A Study of Acetone Exhalation in Diabetic and Nondiabetic Human Subjects.* Diabetes Res Clin Pract, 1952. **1**(3): p. 188-193.

57. Minh Tdo, C., D.R. Blake, and P.R. Galassetti, *The clinical potential of exhaled breath analysis for diabetes mellitus.* Diabetes Res Clin Pract, 2012. **97**(2): p. 195-205.

58. Linus Pauling, A.B.R., Roy Teranishi, and Paul Cary, *Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography.* Proc. Nat. Acad. Sci, 1971. **68**(10): p. 2374-2376.

59. Sethi, S., R. Nanda, and T. Chakraborty, *Clinical application of volatile organic compound analysis for detecting infectious diseases.* Clin Microbiol Rev, 2013. **26**(3): p. 462-75.

60. Graham, D.Y., et al., *Campylobacter Pyllori Deteected Noninvasively by the 13C-Urea Breath Test.* The Lancet, 1987: p. 1174-1177.

61. Ryter, S.W. and A.M. Choi, *Carbon monoxide in exhaled breath testing and therapeutics.* J Breath Res, 2013. **7**(1): p. 017111.

62. Taylor, D.R., *Advances in the clinical applications of exhaled nitric oxide measurements.* J Breath Res, 2012. **6**(4): p. 047102.

63. Hogman, M., *Extended NO analysis in health and disease.* J Breath Res, 2012. **6**(4): p. 047103.

64. Gelb, A.F., et al., *Review of exhaled nitric oxide in chronic obstructive pulmonary disease.* J Breath Res, 2012. **6**(4): p. 047101.

65. Bucca, C., et al., *Exhaled nitric oxide (FENO) in non-pulmonary diseases.* J Breath Res, 2012. **6**(2): p. 027104.

66. Kim, K.H., S.A. Jahan, and E. Kabir, *A review of breath analysis for diagnosis of human health.* TrAC Trends in Analytical Chemistry, 2012. **33**: p. 1-8.

67. Gasparri, R., et al., *Diagnostic biomarkers for lung cancer prevention.* J Breath Res, 2018. **12**(2): p. 027111.

68. Haick, H., et al., *Assessment, origin, and implementation of breath volatile cancer markers.* Chem Soc Rev, 2014. **43**(5): p. 1423-49.

69. Fens, N., et al., *Breathomics as a diagnostic tool for pulmonary embolism.* J Thromb Haemost, 2010. **8**(12): p. 2831-3.

70. Shirasu, M. and K. Touhara, *The scent of disease: volatile organic compounds of the human body related to disease and disorder.* J Biochem, 2011. **150**(3): p. 257-66.

71. Van Oort, P.M., et al., *The potential role of exhaled breath analysis in the diagnostic process of pneumonia-a systematic review.* J Breath Res, 2018. **12**(2): p. 024001.

72. Davis, M.D., S.J. Fowler, and A.J. Montpetit, *Exhaled breath testing - A tool for the clinician and researcher.* Paediatr Respir Rev, 2018.

73. Afolabi, P.R., et al., *The characterisation of hepatic mitochondrial function in patients with non-alcoholic fatty liver disease (NAFLD) using the (13)C-ketoisocaproate breath test.* J Breath Res, 2018. **12**(4): p. 046002.

74. Zhan, X., J. Duan, and Y. Duan, *Recent developments of proton-transfer reaction mass spectrometry (PTR-MS) and its applications in medical research.* Mass Spectrom Rev, 2013. **32**(2): p. 143-65.

75. Handelman, G., *Breath ethane in dialysis patients and control subjects.* Free Radical Biology and Medicine, 2003. **35**(1): p. 17-23.

76. Broza, Y.Y., et al., *A nanomaterial-based breath test for short-term follow-up after lung tumor resection.* Nanomedicine, 2013. **9**(1): p. 15-21.

77. Beauchamp, J., *Current sampling and analysis techniques in breath research--results of a task force poll.* J Breath Res, 2015. **9**(4): p. 047107.

78. Nardi-Agmon, I. and N. Peled, *Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects.* Lung Cancer (Auckl), 2017. **8**: p. 31-38.

79. Agapiou, A., et al., *Trace detection of endogenous human volatile organic compounds for search, rescue and emergency applications.* TrAC Trends in Analytical Chemistry, 2015. **66**: p. 158-175.

80. Stadler, S., et al., *Characterization of volatile organic compounds from human analogue decomposition using thermal desorption coupled to comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry.* Anal Chem, 2013. **85**(2): p. 998-1005.

81. Tiwari, V., Y. Hanai, and S. Masunaga, *Ambient levels of volatile organic compounds in the vicinity of petrochemical industrial area of Yokohama, Japan.* Air Qual Atmos Health, 2010. **3**: p. 65-75.

82. Camara, M., et al., *Detection and Quantification of Natural Contaminants of Wine by Gas Chromatography−Differential Ion Mobility Spectrometry (GC-DMS).* J. Agric. Food Chem, 2013. **61**: p. 1036-1043.

83. Brokl, M., et al., *Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry.* J Chromatogr A, 2014. **1370**: p. 216-29.

84. Kusano, M., E. Mendez, and K.G. Furton, *Development of headspace SPME method for analysis of volatile organic compounds present in human biological specimens.* Anal Bioanal Chem, 2011. **400**(7): p. 1817-26.

85. Wehinger, A., et al., *Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas.* International Journal of Mass Spectrometry, 2007. **265**(1): p. 49-59.

86. de Lacy Costello, B., et al., *A review of the volatiles from the healthy human body.* J Breath Res, 2014. **8**(1): p. 014001.

87. Filipiak, W., et al., *Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants.* J Breath Res, 2012. **6**(3): p. 036008.

88. Yu, C. and D. Crumpt, *A Review of the Emission of VOCs from Polymeric Materials used in Buildings.* Buiding and environment, 1998. **33**(6): p. 357-374.

89. Krilaviciute, A., et al., *Associations of diet and lifestyle factors with common volatile organic compounds in exhaled breath of average-risk individuals.* J Breath Res, 2019. **13**(2): p. 026006.

90. Bajtarevic, A., et al., *Noninvasive detection of lung cancer by analysis of exhaled breath.* BMC Cancer, 2009. **9**: p. 348.

91. Miekisch, W., J.K. Schubert, and G.F. Noeldge-Schomburg, *Diagnostic potential of breath analysis--focus on volatile organic compounds.* Clin Chim Acta, 2004. **347**(1-2): p. 25-39.

92. van der Schee, M.P., et al., *Breathomics in lung disease.* Chest, 2015. **147**(1): p. 224-231.

93. Maren Mieth, J.K.S., et al., *Automated Needle Trap Heart-Cut GC/MS and Needle Trap Comprehensive Two-Dimensional GC/TOF-MS for Breath Gas Analysis in the Clinical Environment.* Anal. Chem, 2010. **82**: p. 2541-2551.

94. Yu, Y. and J. Pawliszyn, *On-line monitoring of breath by membrane extraction with sorbent interface coupled with CO2 sensor.* Journal of Chromatography A, 2004. **1056**(1-2): p. 35-41.

95. Sacks, J.M.S.a.R.D., *GC Analysis of Human Breath with A Series-Coupled Column Ensemble and A Multibed Sorption Trap.* Anal Chem, 2003. **75**: p. 2231-2236.

96. Hakim, M., et al., *Diagnosis of head-and-neck cancer from exhaled breath.* Br J Cancer, 2011. **104**(10): p. 1649-55.

97. Peng, G., et al., *Diagnosing lung cancer in exhaled breath using gold nanoparticles.* Nature Nanotechnology, 2009. **4**: p. 669-673.

98. Li, M., et al., *A microfabricated preconcentration device for breath analysis.* Sensors and Actuators B: Chemical, 2013. **180**: p. 130-136.

99. Wang, C., et al., *Exhaled volatile organic compounds as lung cancer biomarkers during one-lung ventilation.* Sci Rep, 2014. **4**: p. 7312.

100. Horvath, I., et al., *Exhaled biomarkers in lung cancer.* Eur Respir J, 2009. **34**(1): p. 261-75.

101. Hunt, J., *Exhaled breath condensate: an overview.* Immunol Allergy Clin North Am, 2007. **27**(4): p. 587-96.

102. Emma H. Baker, N.C., Amanda L. Brennan, Donald A. Fisher, Khin M.Gyi, Margaret E. Hodson, Barbara J. Philips, Deborah L. Baines and David M. Wood, *Hyperglycemia and cystic fibrosis alter respiratory fluid glucose concentrations estimated by breath condensate analysis.* J Appl Physiol, 2007. **102**(1969-1975).

103. Kuban, P. and F. Foret, *Exhaled breath condensate: determination of non-volatile compounds and their potential for clinical diagnosis and monitoring. A review.* Anal Chim Acta, 2013. **805**: p. 1-18.

104. E.wood, D., et al., *Lung Cancer Screening, Version 1.2015.* Featured Updates to the NCCN Guidelines, 2015.

105. Buszewski, B., et al., *Human exhaled air analytics: biomarkers of diseases.* Biomed Chromatogr, 2007. **21**(6): p. 553-66.

106. Schmidt, K. and I. Podmore, *Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer.* J Biomark, 2015. **2015**: p. 981458.

107. Alonso, M. and J.M. Sanchez, *Analytical challenges in breath analysis and its application to exposure monitoring.* TrAC Trends in Analytical Chemistry, 2013. **44**: p. 78-89.

108. Lawal, O., et al., *Exhaled breath analysis: a review of 'breath-taking' methods for off-line analysis.* Metabolomics, 2017. **13**(10): p. 110.

109. Kischkel, S., et al., *Breath biomarkers for lung cancer detection and assessment of smoking related effects--confounding variables, influence of normalization and statistical algorithms.* Clin Chim Acta, 2010. **411**(21-22): p. 1637-44.

110. Trefz, P., et al., *Evaluation of needle trap micro-extraction and automatic alveolar sampling for point-of-care breath analysis.* Anal Bioanal Chem, 2013. **405**(10): p. 3105-15.

111. Juan M. Sanchez, a.R.D.S., *Development of a Multibed Sorption Trap, Comprehensive Two-Dimensional Gas Chromatography, and Time-of-Flight Mass Spectrometry System for the Analysis of Volatile Organic Compounds in Human Breath.* Anal Chem, 2006. **78**(9): p. 3046-3054.

112. Bokowa, A.H., *Odour assessment: determining the optimum temperature and time for Tedlar sampling bag pre-conditioning.* Water Sci Technol, 2012. **66**(8): p. 1806-11.

113. Kim, Y.H., et al., *Comparison of storage stability of odorous VOCs in polyester aluminum and polyvinyl fluoride Tedlar(R) bags.* Anal Chim Acta, 2012. **712**: p. 162-7.

114. Szylak-Szydlowski, M., *Odour Samples Degradation During Detention in Tedlar Bags.* Water Air Soil Pollut, 2015. **226**(7): p. 227.

115. Schmekel, B., F. Winquist, and A. Vikstrom, *Analysis of breath samples for lung cancer survival.* Anal Chim Acta, 2014. **840**: p. 82-6.

116. Hertel, M., et al., *Volatile organic compounds in the breath of oral candidiasis patients: a pilot study.* Clin Oral Investig, 2017.

117. Michael Phillips, R.N.C., Andrew R.C. Cummin, Anthony J. Gagliardi, Kevin Gleeson, Joel Greenberg, Roger A. Maxfield, and William N. Rom, *Detection of Lung Cancer With Volatile Markers in the Breath.* CHEST, 2003. **123**: p. 2115-2123.

118. Phillips, M., et al., *Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study.* The Lancet, 1999. **353**(9168): p. 1930-1933.

119. Michael Phillips, N.A., John H.M. Austin, Robert B. Cameron , Renee N. Cataneo, Joel Greenberg, Robert Kloss, Roger A. Maxfield, Muhammad I. Munawar, Harvey I. Pass, Asif Rashid,William N. Rom and Peter Schmitt, *Prediction of lung cancer using volatile biomarkers in breath.* Cancer Biomarkers, 2007. **3**: p. 95-109.

120. Harper, M., *Sorbent trapping of volatile organic compounds from air.* Journal of Chromatography A, 2000. **885**: p. 129-151.

121. Baltussen, E., C.A. Cramers, and P.J. Sandra, *Sorptive sample preparation -- a review.* Anal Bioanal Chem, 2002. **373**(1-2): p. 3-22.

122. Balasubramanian, S. and S. Panigrahi, *Solid-Phase Microextraction (SPME) Techniques for Quality Characterization of Food Products: A Review.* Food and Bioprocess Technology, 2010. **4**(1): p. 1-26.

123. Buszewski, B. and M. Szultka, *Past, Present, and Future of Solid Phase Extraction: A review.* Analytical Chemistry, 2012. **42**(3): p. 198-213.

124. Rudnicka, J., et al., *Determination of volatile organic compounds as biomarkers of lung cancer by SPME-GC-TOF/MS and chemometrics.* J Chromatogr B Analyt Technol Biomed Life Sci, 2011. **879**(30): p. 3360-6.

125. Radomır Hyspler , S.C., Jırı Gasparic , Zdenek Zadak , Marie Cızkova, Vera Balasova, *Determination of isoprene in human expired breath using solid-phase microextraction and gas chromatography–mass spectrometry.* J Chromatogr B, 2000. **739**: p. 183-190.

126. Poli, D., et al., *Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME-GC/MS.* J Chromatogr B Analyt Technol Biomed Life Sci, 2010. **878**(27): p. 2643-51.

127. Huijun Liu · Caixia Li, H.W., Zhongping Huang, Peipei Zhang, Zaifa Pan, Lili Wang, *Characterization of Volatile Organic Metabolites in Lung Cancer Pleural Effusions by SPME–GC/MS Combined with an Untargeted Metabolomic Method.* Chromatographia, 2014.

128. Fuchs, P., et al., *Breath gas aldehydes as biomarkers of lung cancer.* Int J Cancer, 2010. **126**(11): p. 2663-70.

129. Song, G., et al., *Quantitative breath analysis of volatile organic compounds of lung cancer patients.* Lung Cancer, 2010. **67**(2): p. 227-31.

130. Di Francesco, F., et al., *Breath analysis: trends in techniques and clinical applications.* Microchemical Journal, 2005. **79**(1-2): p. 405-410.

131. Pawliszyn, C.G.a.J., *Solid-Phase Microextraction for the Analysis of Human Breath.* Anal Chem, 1997. **69**: p. 587-596.

132. Pawliszyn, C.L.A.a.J., *Solid Phase Microextraction with Thermal Desorption Using Fused Silica Optical Fibers.* Anal Chem, 1990. **62**(19): p. 2145-2148.

133. Zhao, W., G. Ouyang, and J. Pawliszyn, *Preparation and application of in-fibre internal standardization solid-phase microextraction.* Analyst, 2007. **132**(3): p. 256-61.

134. Wang, Y., et al., *Equilibrium in-fibre standardisation technique for solid-phase microextraction.* Journal of Chromatography A, 2005. **1072**(1): p. 13-17.

135. Wang, Y., et al., *The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath,tissues and cell lines.* Cancer Biomarkers, 2012. **11**: p. 129-137.

136. Lord, H. and J. Pawliszyn, *Evolution of solid-phase microextraction technology.* Journal of Chromatography A, 2000. **885**: p. 153-193.

137. Jelen, H.H., M. Majcher, and M. Dziadas, *Microextraction techniques in the analysis of food flavor compounds: A review.* Anal Chim Acta, 2012. **738**: p. 13-26.

138. Ramirez, N., et al., *Comparative study of solvent extraction and thermal desorption methods for determining a wide range of volatile organic compounds in ambient air.* Talanta, 2010. **82**(2): p. 719-27.

139. Koziel, J.A. and J. Pawliszyn, *Air Sampling and Analysis of Volatile Organic Compounds with Solid Phase Microextraction.* Journal of the Air & Waste Management Association, 2011. **51**(2): p. 173-184.

140. Lord, H.L., W. Zhan, and J. Pawliszyn, *Fundamentals and applications of needle trap devices: a critical review.* Anal Chim Acta, 2010. **677**(1): p. 3-18.

141. Maren Mieth, J.K.S., Thomas Groger, Bastian Sabel, Sabine Kischkel, and D.H. Patricia Fuchs, Ralf Zimmermann, and Wolfram Miekisch, *Automated Needle Trap Heart-Cut GC/MS and Needle Trap Comprehensive Two-Dimensional GC/TOF-MS for Breath Gas Analysis in the Clinical Environment.* Anal Chem, 2010. **82**: p. 2541–2551.

142. Filipiak, W., et al., *Optimization of sampling parameters for collection and preconcentration of alveolar air by needle traps.* J Breath Res, 2012. **6**(2): p. 027107.

143. Mochalski, P., et al., *Blood and breath levels of selected volatile organic compounds in healthy volunteers.* Analyst, 2013. **138**(7): p. 2134-45.

144. Poli, D., et al., *Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study.* Respir Res, 2005. **6**: p. 71.

145. Kwak, J., et al., *Evaluation of Bio-VOC Sampler for Analysis of Volatile Organic Compounds in Exhaled Breath.* Metabolites, 2014. **4**(4): p. 879-88.

146. Tirzite, M., et al., *Detection of lung cancer in exhaled breath with an electronic nose using support vector machine analysis.* J Breath Res, 2017. **11**(3): p. 036009.

147. Nakhleh, M.K., et al., *Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules.* ACS Nano, 2017. **11**(1): p. 112-125.

148. Oguma, T., et al., *Clinical contributions of exhaled volatile organic compounds in the diagnosis of lung cancer.* PLoS One, 2017. **12**(4): p. e0174802.

149. Smith, D. and P. Španěl, *Status of selected ion flow tube MS: accomplishments and challenges in breath analysis and other areas.* Bioanalysis, 2016. **8**(11): p. 1183-1201.

150. Spanel, P., K. Dryahina, and D. Smith, *A quantitative study of the influence of inhaled compounds on their concentrations in exhaled breath.* J Breath Res, 2013. **7**(1): p. 017106.

151. Turner, C., et al., *An exploratory comparative study of volatile compounds in exhaled breath and emitted by skin using selected ion flow tube mass spectrometry.* Rapid Commun Mass Spectrom, 2008. **22**(4): p. 526-32.

152. King, J., et al., *Dynamic profiles of volatile organic compounds in exhaled breath as determined by a coupled PTR-MS/GC-MS study.* Physiol Meas, 2010. **31**(9): p. 1169-84.

153. Amann, A., et al., *Applications of breath gas analysis in medicine.* International Journal of Mass Spectrometry, 2004. **239**(2-3): p. 227-233.

154. Herbig, J., et al., *On-line breath analysis with PTR-TOF.* J Breath Res, 2009. **3**(2): p. 027004.

155. Beauchamp, J., F. Kirsch, and A. Buettner, *Real-time breath gas analysis for pharmacokinetics: monitoring exhaled breath by on-line proton-transfer-reaction mass spectrometry after ingestion of eucalyptol-containing capsules.* J Breath Res, 2010. **4**(2): p. 026006.

156. Schwoebel, H., et al., *Phase-resolved real-time breath analysis during exercise by means of smart processing of PTR-MS data.* Anal Bioanal Chem, 2011. **401**(7): p. 2079-91.

157. Winkler, K., J. Herbig, and I. Kohl, *Real-time metabolic monitoring with proton transfer reaction mass spectrometry.* J Breath Res, 2013. **7**(3): p. 036006.

158. Handa, H., et al., *Exhaled breath analysis for lung cancer detection using ion mobility spectrometry.* PLoS One, 2014. **9**(12): p. e114555.

159. Raed A Dweik, a.A.A., *Exhaled breath analysis: the new frontier in medical testing.* J. Breath Res, 2008. **2**(3).

160. Bos, L.D., et al., *A simple breath sampling method in intubated and mechanically ventilated critically ill patients.* Respir Physiol Neurobiol, 2014. **191**: p. 67-74.

161. Phillips, M., et al., *Prediction of breast cancer risk with volatile biomarkers in breath.* Breast Cancer Res Treat, 2018. **170**(2): p. 343-350.

162. Libardoni, M., et al., *Analysis of human breath samples with a multi-bed sorption trap and comprehensive two-dimensional gas chromatography (GCxGC).* J Chromatogr B Analyt Technol Biomed Life Sci, 2006. **842**(1): p. 13-21.

163. Koek, M.M., et al., *Semi-automated non-target processing in GC x GC-MS metabolomics analysis: applicability for biomedical studies.* Metabolomics, 2011. **7**(1): p. 1-14.

164. Das, M.K., et al., *Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS.* Anal Chem, 2014. **86**(2): p. 1229-37.

165. Michael Phillips, M., Urvish Patel, Renee N. Cataneo, Xiang Zhang, *Detection of an Extended Human Volatome with Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry.* Plos one, 2013. **9**(8): p. e75274.

166. Hongying Ma, X.L., Jianmin Chen, Huijie Wang, Tiantao Chena, Kai Chen, Shifen Xu, *Analysis of human breath samples of lung cancer patients and healthy controls with solid-phase microextraction (SPME) and flow-modulated comprehensive two-dimensional gas chromatography (GC × GC).* Anal. Methods, 2014. **6**: p. 6841-6849.

167. Beccaria, M., et al., *Preliminary investigation of human exhaled breath for tuberculosis diagnosis by multidimensional gas chromatography - Time of flight mass spectrometry and machine learning.* J Chromatogr B Analyt Technol Biomed Life Sci, 2018. **1074-1075**: p. 46-50.

168. Wilde, M.J., et al., *Breath analysis by two-dimensional gas chromatography with dual flame ionisation and mass spectrometric detection - Method optimisation and integration within a large-scale clinical study.* J Chromatogr A, 2019.

169. Boots, A.W., et al., *The versatile use of exhaled volatile organic compounds in human health and disease.* J Breath Res, 2012. **6**(2): p. 027108.

170. Blake, R.S., P.S. Monks, and A.M. Ellis, *Proton-Transfer Reaction Mass Spectrometry.* Chem. Rev, 2009. **109**: p. 861-896.

171. Smith, D. and P. Spanel, *Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis.* Mass Spectrom Rev, 2005. **24**(5): p. 661-700.

172. Westhoff, M., et al., *Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study.* Thorax, 2009. **64**(9): p. 744-8.

173. Maurer, F., et al., *Calibration and validation of a MCC/IMS prototype for exhaled propofol online measurement.* J Pharm Biomed Anal, 2017. **145**: p. 293-297.

174. Lamote, K., et al., *Detection of malignant pleural mesothelioma in exhaled breath by multicapillary column/ion mobility spectrometry (MCC/IMS).* J Breath Res, 2016. **10**(4): p. 046001.

175. Allers, M., et al., *Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using closed gas loop GC-IMS and GC-APCI-MS.* J Breath Res, 2016. **10**(2): p. 026004.

176. Peng, L., et al., *Online Measurement of Exhaled NO Concentration and Its Production Sites by Fast Non-equilibrium Dilution Ion Mobility Spectrometry.* Sci Rep, 2016. **6**: p. 23095.

177. Allafchian, A.R., et al., *A novel method for the determination of three volatile organic compounds in exhaled breath by solid-phase microextraction-ion mobility spectrometry.* Anal Bioanal Chem, 2016. **408**(3): p. 839-47.

178. Ulanowska, A., et al., *Determination of Volatile Organic Compounds in Exhaled Breath by Ion Mobility Spectrometry.* Chem. Anal, 2008 **53**: p. 953-965.

179. Romero, K.I. and R. Fernandez-Maestre, *Ion mobility spectrometry: the diagnostic tool of third millennium medicine.* Rev Assoc Med Bras (1992), 2018. **64**(9): p. 861-868.

180. Nissinen, S.I., et al., *Detection of Pancreatic Cancer by Urine Volatile Organic Compound Analysis.* Anticancer Res, 2019. **39**(1): p. 73-79.

181. Westhoff, M., P. Litterst, and J.I. Baumbach, *Ion mobility spectrometry in the diagnostic of sarcoidosis: Results of a fesability study.* Journal of physiology and pharmocology, 2017. **58**(5): p. 739-751.

182. H, W. and P. A., *Sniffer Dogs in the Melanoma Clinic?* The lancet, 1989. **8640**: p. 734-7.

183. Amundsen, T., et al., *Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer.* Acta Oncol, 2014. **53**(3): p. 307-15.

184. Luque de Castro, M.D. and M.A. Fernández-Peralbo, *Analytical methods based on exhaled breath for early detection of lung cancer.* TrAC Trends in Analytical Chemistry, 2012. **38**: p. 13-20.

185. Nasiri, N. and C. Clarke, *Nanostructured Chemiresistive Gas Sensors for Medical Applications.* Sensors (Basel), 2019. **19**(3).

186. Di Natale, C., et al., *Solid-state gas sensors for breath analysis: a review.* Anal Chim Acta, 2014. **824**: p. 1-17.

187. Machado, R.F., et al., *Detection of lung cancer by sensor array analyses of exhaled breath.* Am J Respir Crit Care Med, 2005. **171**(11): p. 1286-91.

188. Di Natale, C., et al., *Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors.* Biosensors and Bioelectronics, 2003. **18**(10): p. 1209-1218.

189. Fernandes, M.P., S. Venkatesh, and B.G. Sudarshan, *Early Detection of Lung Cancer Using Nano-Nose - A Review.* The Open Biomedical Engineering Journal, 2015. **9**: p. 228-233.

190. Gordon, S.M., Szidon, J.P., Krotoszynski, B.K., Gibbons, R.D., O'Neill, H.J, *Volatile organic compounds in exhaled air from patients with lung cancer.* Clin Chem Lab Med, 1985. **31**: p. 1278-1282.

191. Capuano, R., et al., *Sensors for Lung Cancer Diagnosis.* J Clin Med, 2019. **8**(2).

192. Amann, A., et al., *Analysis of exhaled breath for disease detection.* Annu Rev Anal Chem (Palo Alto Calif), 2014. **7**: p. 455-82.

193. Peng, G., et al., *Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors.* Br J Cancer, 2010. **103**(4): p. 542-51.

194. Phillips M, A.N., Austin JH, Cameron RB, Cataneo RN, Greenberg J, Kloss R, Maxfield RA, Munawar MI, Pass HI, Rashid A, Rom WN, Schmitt P., *Prediction of lung cancer using volatile biomarkers in breath.* Cancer biomark, 2007. **3**: p. 95-109.

195. Phillips, M., et al., *Detection of lung cancer using weighted digital analysis of breath biomarkers.* Clin Chim Acta, 2008. **393**(2): p. 76-84.

196. Chen, X., et al., *A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis.* Cancer, 2007. **110**(4): p. 835-44.

197. Amann, A., et al., *Analysis of exhaled breath for screening of lung cancer patients.* memo - Magazine of European Medical Oncology, 2010. **3**(3): p. 106-112.

198. Rudnicka, J., et al., *Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs.* Sensors and Actuators B: Chemical, 2014. **202**: p. 615-621.

199. Ligor, M., et al., *Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry.* Clin Chem Lab Med, 2009. **47**(5): p. 550-60.

200. Bousamra, M., 2nd, et al., *Quantitative analysis of exhaled carbonyl compounds distinguishes benign from malignant pulmonary disease.* J Thorac Cardiovasc Surg, 2014. **148**(3): p. 1074-80.

201. Schallschmidt, K., et al., *Comparison of volatile organic compounds from lung cancer patients and healthy controls-challenges and limitations of an observational study.* J Breath Res, 2016. **10**(4): p. 046007.

202. Schumer, E.M., et al., *High sensitivity for lung cancer detection using analysis of exhaled carbonyl compounds.* J Thorac Cardiovasc Surg, 2015. **150**(6): p. 1517-22.

203. Li, M., et al., *Breath carbonyl compounds as biomarkers of lung cancer.* Lung Cancer, 2015. **90**(1): p. 92-7.

204. Shah R, S.S., Richardson J, Means AJ, Goulden C, *Results of surgical treatment of stage I and II lung cancer.* J Cardiovasc Surg, 1996. **37**: p. 169-172.

205. O'Neill, H.J., Gordon, S.M., O'Neill, M.H., Gibbons, R.D., Szidon, J.P., *A Computerized Classification Technique for Screening for the Presence of Breath Biomarkers in Lung Cancer.* 34, 1988. **8**(1613-1618).

206. Ma, H., et al., *Analysis of human breath samples of lung cancer patients and healthy controls with solid-phase microextraction (SPME) and flow-modulated comprehensive two-dimensional gas chromatography (GC × GC).* Analytical Methods, 2014. **6**(17): p. 6841.

207. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD.* Lung Cancer, 2009. **64**(2): p. 166-70.

208. Jouyban, A., et al., *Co-liquefaction with acetone and GC analysis of volatile compounds in exhaled breath as lung cancer biomarkers.* Bioimpacts, 2017. **7**(2): p. 99-108.

209. Saalberg, Y. and M. Wolff, *VOC breath biomarkers in lung cancer.* Clin Chim Acta, 2016. **459**: p. 5-9.

210. Filipiak, W., et al., *Volatile Organic Compounds (VOCs) Released by Pathogenic Microorganisms in vitro: Potential Breath Biomarkers for Early-Stage Diagnosis of Disease.* 2013: p. 463-512.

211. Davies, S., P. Spanel, and D. Smith, *Quantitative analysis of ammonia on the breath of patients in end-stage renal failure.* Kidney International,, 1997. **52**: p. 223-228.

212. Scislowski, P.W.D. and K. Pickard, *The regulation of transaminative flux of methionine in rat liver mitochrondria.* Archives of biochemistry and biophysics, 1994. **314**(2): p. 412-416.

213. Yamaguchi, M.S., et al., *Headspace sorptive extraction-gas chromatography-mass spectrometry method to measure volatile emissions from human airway cell cultures.* J Chromatogr B Analyt Technol Biomed Life Sci, 2018. **1090**: p. 36-42.

214. Bischoff, A.C., et al., *Smell of cells: Volatile profiling of stem- and non-stem cell proliferation.* J Breath Res, 2018. **12**(2): p. 026014.

215. Pereira, J., et al., *Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview.* Metabolites, 2015. **5**(1): p. 3-55.

216. Krilaviciute, A., et al., *Detection of cancer through exhaled breath: a systematic review.* Oncotarget, 2015. **6**(36): p. 38643-38657.

217. Giri, A., et al., *Molecular Characterization of Volatiles and Petrochemical Base Oils by Photo-Ionization GCxGC-TOF-MS.* Anal Chem, 2017. **89**(10): p. 5395-5403.

218. Eisele, A.P., et al., *Volatile organic compounds at two oil and natural gas production well pads in Colorado and Texas using passive samplers.* J Air Waste Manag Assoc, 2016. **66**(4): p. 412-9.

219. Perez-Hurtado, P., et al., *Direct Analysis of Volatile Organic Compounds in Foods by Headspace Extraction Atmospheric Pressure Chemical Ionisation Mass Spectrometry.* Rapid Commun Mass Spectrom, 2017.

220. Kfoury, N., et al., *Direct Contact Sorptive Extraction: A Robust Method for Sampling Plant Volatiles in the Field.* J Agric Food Chem, 2017.

221. Xiao, Z., et al., *Characterization of the key odorants of fennel essential oils of different regions using GC–MS and GC–O combined with partial least squares regression.* Journal of Chromatography B, 2017. **1063**: p. 226-234.

222. Niu, Y., et al., *Characterization of the key aroma compounds in different light aroma type Chinese liquors by GC-olfactometry, GC-FPD, quantitative measurements, and aroma recombination.* Food Chem, 2017. **233**: p. 204-215.

223. Altomare, D.F., et al., *Exhaled volatile organic compounds identify patients with colorectal cancer.* Br J Surg, 2013. **100**(1): p. 144-50.

224. Ferreira, S.L., et al., *Statistical designs and response surface techniques for the optimization of chromatographic systems.* J Chromatogr A, 2007. **1158**(1-2): p. 2-14.

225. Demarche, S.F., et al., *Asthma Control and Sputum Eosinophils: A Longitudinal Study in Daily Practice.* J Allergy Clin Immunol Pract, 2017. **5**(5): p. 1335-1343 e5.

226. Brokl, M., et al., *Analysis of mainstream tobacco smoke particulate phase using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry.* J. Sep. Sci, 2013. **36**: p. 1037-1044.

227. Giri, A., Z. Zelinkova, and T. Wenzl, *Experimental design-based isotope-dilution SPME-GC/MS method development for the analysis of smoke flavoring products.* Food Addit Contam Part A Chem Anal Control Expo Risk Assess, 2017.

228. Dekeirsschieter, J., et al., *Enhanced characterization of the smell of death by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GCxGC-TOFMS).* PLoS One, 2012. **7**(6): p. e39005.

229. Forbes, S.L., et al., *Comparison of the decomposition VOC profile during winter and summer in a moist, mid-latitude (Cfb) climate.* PLoS One, 2014. **9**(11): p. e113681.

230. Perrault, K.A., et al., *Detection of decomposition volatile organic compounds in soil following removal of remains from a surface deposition site.* Forensic Sci Med Pathol, 2015. **11**(3): p. 376-87.

231. Bertoncini, F., et al., *Apport de la chromatographie en phase gazeuse bidimensionnelle pour la caractérisation de matrices.* SPECTRA ANALYSE, 2005. **247**: p. 26-31.

232. Giddings, J.C., *Concepts and Comparisons in Multidimensional Separation.* J. High Resol. Chromatogr, 1987. **10**: p. 319-323.

233. Omais, B., et al., *Considerations on orthogonality duality in comprehensive two-dimensional gas chromatography.* Anal Chem, 2011. **83**(19): p. 7550-4.

234. Phillips, J.B. and J. Xu, *Comprehensive multi-dimensional gas chromatography.* Journal of Chromatography A, 1995. **703**: p. 327-334.

235. Barcelo, D., et al., *Comprehensive Two Dimensional Gas Chromatography*. First edition ed. COMPREHENSIVE ANALYTICAL CHEMISTRY. Vol. 55. 2009: Elsevier.

236. Korytár, P., et al., *Analysis of polychlorinated biphenyls by comprehensive two-dimensional gas chromatography.* Journal of Chromatography A, 2002. **958**: p. 203-18.

237. Adahchour, M., et al., *Recent developments in comprehensive two-dimensional gas chromatography (GC×GC).* TrAC Trends in Analytical Chemistry, 2006. **25**(8): p. 821-840.

238. Adahchour, M., et al., *Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: principles and applications.* Journal of Chromatography A, 2005. **1067**(1-2): p. 245-254.

239. Liu, H.C., Q.W. Li, and L.B. Tang, *Capillary gas chromatographic determination of dimethachlon residues in fresh tobacco leaves and cut-tobacco.* J Zhejiang Univ Sci B, 2007. **8**(4): p. 272-6.

240. Tranchida, P.Q., *Comprehensive two-dimensional gas chromatography: A perspective on processes of modulation.* Journal of Chromatography A, 2018. **1536**: p. 2-5.

241. Cordero, C., et al., *Method translation and full metadata transfer from thermal to differential flow modulated comprehensive two dimensional gas chromatography: Profiling of suspected fragrance allergens.* J Chromatogr A, 2017. **1480**: p. 70-82.

242. Bahaghighat, H.D., C.E. Freye, and R.E. Synovec, *Recent advances in modulator technology for comprehensive two dimensional gas chromatography.* TrAC Trends in Analytical Chemistry, 2018.

243. Magagna, F., et al., *Advanced fingerprinting of high-quality cocoa: Challenges in transferring methods from thermal to differential-flow modulated comprehensive two dimensional gas chromatography.* J Chromatogr A, 2018. **1536**: p. 122-136.

244. Focant, J., et al., *Measurement of Selected Halogenated Contaminants in Human Serum and Milk using GCxGC-IDTOFMS.* Organohalogen Compounds, 2004. **66**: p. 804-811.

245. Stefanuto, P.H., et al., *GC x GC-TOFMS and supervised multivariate approaches to study human cadaveric decomposition olfactive signatures.* Anal Bioanal Chem, 2015. **407**(16): p. 4767-78.

246. Hoffman, E.M., et al., *Characterization of the volatile organic compounds present in the headspace of decomposing human remains.* Forensic Sci Int, 2009. **186**(1-3): p. 6-13.

247. Lorenzo, N., et al., *Laboratory and field experiments used to identify Canis lupus var. familiaris active odor signature chemicals from drugs, explosives, and humans.* Anal Bioanal Chem, 2003. **376**(8): p. 1212-24.

248. Leclercq, P.A. and C.A. Cramers, *Review HIGH-SPEED GC-MS.* Mass Spectrometry Reviews, 1998. **17**: p. 37-49.

249. Price, D., *Time-of-Flight Mass Spectrometry: The Early Years as Chronicled by the European Time-of-Flight Symposia.* ACS Symposium Series, 1994. **549**(Chapter 1): p. 1-15.

250. Hoffman, E.d. and V. Stroobant, *Mass Spectrometry Principles and Applications*. Third Edition ed, ed. Wiley. 2007.

251. Hoffmann E. Stroobant V. *Mass Spectrometry Principles and Applications, third edition* , ed. Wiley 2007.

252. Karisa M. Pierce, J.C.H., Janiece L. Hope, Petrie M. Rainey, Andrew N. Hoofnagle, Rhona M. Jack, Bob W. Wright, and Robert E. Synovec, *Fisher Ratio Method Applied to*

Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts. Anal Chem, 2006. **78**: p. 5068-5075.

253. Purcaro, G., et al., *SPME-GCxGC-TOF Ms fingerprint of virally-infected cell culture: Sample preparation optimization and data processing evaluation.* Anal Chem, 2018. **1027**: p. 158-167.
254. Mochalski, P., et al., *Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags.* Analyst, 2013. **138**(5): p. 1405-18.

255. Dieterle, F., et al., *Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics.* Anal Chem, 2006. **78**: p. 4281-4290.

256. P-H, S., et al., *Advanced method optimization for volatile aroma profiling of beer using two-dimensioanal gas chromatography time-of-flight mass spectrometry.* J Chromatogr A, 2017. **1507**: p. 45-52.

257. P.-H. Stefanuto, K.A.P., R. M. Lloyd,. Stuart, Rai, S. L. Forbes and J.-F. Focant, *Exploring new dimensions in cadaveric decomposition odour analysis.* Anal. Methods, 2015. **7**: p. 2287-2294.

258. Perrault, K., et al., *A New Approach for the Characterization of Organic Residues from Stone Tools Using GC×GC-TOFMS.* Separations, 2016. **3**(2): p. 16.

259. Stefanuto, P.-H., et al., *Fast Chromatographic Method for Explosive Profiling.* Chromatography, 2015. **2**(2): p. 213-224.

260. Phillips, M., et al., *Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress.* Journal of Laboratory and Clinical Medicine, 2000. **136**(3): p. 243-249.

261. Paschke, K.M., A. Mashir, and R.A. Dweik, *Clinical applications of breath testing.* F1000 Med Rep, 2010. **2**: p. 56.

262. Joanna Rudnicka, M.W., Tomasz Kowalkowski, Tadeusz Jezierski, Bogusław Buszewski, *Determination of volatile organic compounds as potential markers of lung cancer by gas chromatography–mass spectrometry versus trained dogs.* Sensors and Actuators B, 2014. **202**: p. 615–621.

263. Michael Phillips, R.N.C., Andrew R.C. Cummin, Anthony J. Gagliardi, Kevin Gleeson, Joel Greenberg, Roger A. Maxfield, and William N. Rom, *Detection of Lung Cancer With Volatile Markers in the Breath.* CHEST, 2003. **123**(6): p. 2115-2123.

264. Ma, H., et al., *Analysis of human breath samples of lung cancer patients and healthy controls with solid-phase microextraction (SPME) and flow-modulated comprehensive two-dimensional gas chromatography (GC × GC).* Anal. Methods, 2014. **6**: p. 6841-6849.

Annex

List of VOC markers reported in the literature

#	Name	Chemical classification	References
1	1,10-(1-butenylidene)bisbenzene	Alkene and aromatic	[193]
2	10,11-dihydro-5H-dibenz-(B,F)-azepine	Aromatic and nitrogen	[194]
3	1,1-(1,2cyclobutanediyl)bis-benzene	Aromatic	[195]
4	1,1-[1-(ethylthio)proprylidene] bis-benzene	Alkene, aromatic and sulphur	[195]
5	1,1,2-triclhoro-1,2,2-trifluoro ethane	Halogen	[195]
6	1,1-ethylidenebis-4-ethyl-benzene	Aromatic	[195]
7	1,1-oxybis-benzene	Aromatic and ether	[194]
8	1,2,3,4-tetrahydro-9-propyl-anthracene	Aromatic	[195]
9	1,2,3,4-tetrahydro-isoquinoline	Aromatic and nitrogen	[90]
10	1,2,4,5-tetramethyl-benzene	Aromatic	[99]
11	1,2,4-trimethyl-benzene	Aromatic	[118, 196]
12	1,2-benzenedicarboxylic acid, diethyl ester	Aromatic and carbonyl	[194]
13	1,2-dichlorobenzene	Aromatic and halogen	[109]
14	1,2-dimethyl-benzene	Aromatic	[109]
15	1,3-cyclohexadiene	Alkene	[197]
16	1,3-dioldiisobutyrate-2,2,4-trimethyl-pentane	Carbonyl	[194]
17	1,4-dimethyl-benzene	Aromatic	[118]
18	1,4-pentadiene	Alkene	[262]
19	1,5,9-trimethyl-1,5,9-cyclododecatriene	Alkene	[194]
20	1,6-dioxacyclododecane-7,12-dione	Carbonyl	[99]

#	Name	Chemical classification	References
21	1,7,7-trimethyl-bicyclo-2,2,1-heptan-2-one	Carbonyl	[195]
22	1-butanol	Alcohol	[129]
23	1-cyclopentene	Alkene	[90]
24	1-heptene	Alkene	[118]
25	1-hexene	Alkene	[118, 196]
26	1-methanol-4-trimethyl- 3-cyclohexene	Alcohol and alkene	[195]
27	1-methyl-1,3-cyclopentadiene	Alkene	[90]
28	1-methyl-2-pentyl-cyclopropane	Alkane	[118]
29	1-methyl-ethenyl-benzene	Aromatic	[118]
30	1 methylthio-propene	Alkene and sulphur	[195]
31	2-[(2-ethoxy-3,4-dimethyl,2-cyclohexen-1-ylidene)methyl]-furan	Alkene, aromatic and ether	[195]
32	2,2,4,6,6-pentamethyl-heptane	Alkane	[99, 118]
33	2,2,4-trimethyl-1,3-pentadiol diisobutyrate	Carbonyl	[195]
34	2,2,4-trimethyl-3-carboxyisopropyl-isobutyl ester pentanoic acid	Carbonyl	[195]
35	2,2,7,7-tetramethyl-tricyclo-6,2,1,0(1,6)undec4-en-3-one	Alkene and carbonyl	[195]
36	2,3,4-trimethyl-decane	Alkane	[99]
37	2,3,4-trimethyl-hexane	Alkane	[193]
38	2,2-diethyl 1,1-biphenyl	Aromatic	[194]

#	Name	Chemical classification	References
39	2,2-dimethyl-butane	Alkane	[109]
40	2,2-dimethyl-decane	Alkane	[99]
41	2,2-dimethyl-pentane	Alkane	[109]
42	2,3-bifuran-octahydro	Ether	[99]
43	2,3-dihydro-3-phenyl-1,1,3-trimethyl-1H-Indene	Aromatic	[194]
44	2,3-butadione	Carbonyl	[90]
45	2,3-dimethyl-2-Butanol	Alcohol	[90]
46	2,3-dimethyl-butane	Alkane	[109]
47	2,3-hexadione	Carbonyl	[195]
48	2,4-dimethyl-3-pentanone	Carbonyl	[194]
49	2,4-dimethyl-heptane	Alkane	[109, 198]
50	2,4-dimethyl-pentane	Alkane	[109, 262]
51	2,5-cyclohexadien-1-one, 2,6 bis (1,1-dimethylethyl)-4-ethyldiene	Alkene and carbonyl	[195]
52	2,5-cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-	Alkene and carbonyl	[194]
53	2,5-dimethyl-2,4-hexadiene	Alkene	[194]
54	2,5-dimethyl-furan	Aromatic and ether	[109, 194]
55	2,6,10,14-tetramethyl-pentadecane	Alkane	[135]
56	2,6,10-trimethyltetradecane	Alkane	[135]
57	2,6,11-trimethyl-dodecane	Alkane	[135]

#	Name	Chemical classification	References
58	2,6-dimethyl-naphthalene	Aromatic	[135]
59	2,6- di-tert-butyl-4-methyl-phenol	Alcohol and aromatic	[135]
60	2-butanone	Carbonyl	[90, 99, 109, 199-203]
61	2-butenal	Carbonyl	[109]
62	2-ethyl-1-hexanol	Alcohol	[99]
63	2-ethyl-9,10-anthracenediol	Alcohol and aromatic	[195]
64	2-ethyl-hexyl-nonyl ester sulfurous acid	Carbonyl and sulphur	[99]
65	2-hydroxy-acetaldehyde	Alcohol and carbonyl	[200, 202]
66	2-hydroxy-iso-butyrophenone	Alcohol, aromatic and carbonyl	[99]
67	2-methyl-1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester propanoic acid	Carbonyl	[194] [109] [195]
68	2-methyl-2,2-dimethyl-1-(2-hydroxy-1-methylethyl)-propyl ester propanoic acid	Alcohol and carbonyl	[99]
69	2-methoxy-2-methyl-propane	Ether	[195]
70	2-methyl-3-hydroxy-2,4,4-trimethylpentyl ester propanoic acid	Alcohol and carbonyl	[99]
71	2-methyl-1,3-butadiene	Alkene	[109]
72	2-methyl-2-Butene	Alkene	[90]
73	2-methyl-3-hexanone	Carbonyl	[195]
74	2-methyl-butane	Alkane	[90, 109]

#	Name	Chemical classification	References
75	2-methyl-hendecanal	Carbonyl	[135]
76	2-methyl-heptane	Alkane	[118]
77	2-methyl-hexane	Alkane	[109, 144, 263]
78	2-methyl-naphthalene	Aromatic	[135]
79	2-methyl-pentane	Alkane	[109, 199]
80	2-methyl-butyl-acetat or 2-Hexanol	Carbonyl or alcohol	[158]
81	2-methyl-propanal	Carbonyl	[109]
82	2-penta-decanone	Carbonyl	[135]
83	2-pentanone	Carbonyl	[203]
84	2-propanol	Alcohol	[262]
85	2-Propenal	Carbonyl	[109]
86	3,3-dimethyl-pentane	Alkane	[109, 193]
87	3,3,6,6-tetraphenyl-1,2,4,5-tetroxane	Aromatic and ether	[195]
88	3,4,5,6-tetramethyl-octane	Alkane	[99]
89	3,7-dimethyl-decane	Alkane	[135]
90	3,7-dimethyl-pentadecane	Alkane	[135]
91	3,7-dimethyl-undecane	Alkane	[90]
92	3,8-dimethyl-hendecane	Alkane	[135]
93	3-Butyn-2-ol	Alcohol and alkyne	[90, 199]
94	3-hydroxy-2-butanone	Alcohol and carbonyl	[90, 200, 202-204]

#	Name	Chemical classification	References
95	3-methyl-4-heptanone	Carbonyl	[99]
96	3-methyl-5-propyl-nonane	Alkane	[99]
97	3-methyl-butanol	Alcohol	[158]
98	3-methyl-hexane	Alkane	[263]
99	3-methyl-nonane	Alkane	[118]
100	3-methyl-octane	Alkane	[118]
101	3-methyl-pentane	Alkane	[109, 199]
102	3-methyl-tridecane	Alkane	[263]
103	3-phenyl-2-oxazolidinone	Aromatic, carbonyl and nitrogen	[99]
104	4-ethoxy-ethyl ester benzoic acid	Aromatic, carbonyl and ether	[194, 195]
105	4-hydroxy-hexenal	Alcohol, alkene and carbonyl	[200, 202, 203]
106	4-methyl-decane	Alkane	[194]
107	4-methyl-tetradecane	Alkane	[135]
108	4-methyl-toluene	Aromatic	[99]
109	4-penten-2-ol	Alcohol and alkene	[195]
110	5-(2-methyl-)propyl-nonane	Alkane	[135]
111	5,5-dimethyl-1-3-hexadiene	Alkene	[195]
112	5-butyl-nonane	Alkane	[135]
113	5-isopentyl-2-methyl-7-oxabicyclo(4,1,0)-heptan-2-ol	Alcohol, alkene and ether	[195]
114	5-methyl-decane	Alkane	[263]

#	Name	Chemical classification	References
115	5-propyl-tridecane	Alkane	[135]
116	7-methyl-hexadecane	Alkane	[135]
117	8-hexyl-pentadecane	Alkane	[135]
118	7-methyl-tridecane	Alkane	[263]
119	8-methyl-heptadecane	Alkane	[135]
120	Acetaldehyde	Carbonyl	[109, 128, 200, 203]
121	Acetone	Carbonyl	[85, 109, 187, 200, 205, 207, 264]
122	Acetonitrile	Nitrogen	[109, 197]
123	Acetophenone	Aromatic and carbonyl	[90]
124	Acrolein	Alkene and carbonyl	[187, 203]
125	Alpha isomethyl ionone	Alkene and carbonyl	[195]
126	Benzaldehyde	Aromatic and carbonyl	[90, 199]
127	Benzene	Aromatic	[109, 118, 144, 187, 196, 207]
128	Benzophenone	Aromatic and carbonyl	[195]
129	Benzothiazole	Aromatic and sulphur	[99]
130	Butanal	Carbonyl	[109, 126, 128, 201]

#	Name	Chemical classification	references
131	Butane	Alkane	[109, 205, 207, 263]
132	Butanol	Alcohol	[109, 201]
133	Butyl acetate	Carbonyl	[90]
134	Butyl glycol	Alcohol and ether	[99]
135	Camphor	Carbonyl	[195]
136	Caprolactam	Carbonyl and nitrogen	[99]
137	Carbon disulfide	Sulphur	[109, 187]
138	Chlorobenzene	Aromatic and halogen	[109]
139	Cyclobutyl-benzene	Aromatic	[90]
140	Cyclohexane	Alkane	[109, 118, 148]
141	Cyclohexanone	Carbonyl	[99, 158]
142	Cyclooctane-methanol	Alcohol	[99]
143	Cyclopentanone	Carbonyl	[99]
144	Decanal	Carbonyl	[118, 201]
145	Decane	Alkane	[118, 144, 196, 207]
146	Dimethyl-formamide	Carbonyl and nitrogen	[109]
147	Dimethyl-phenyl-carbinol	Alcohol and aromatic	[99]
148	Dimethyl sulfide	Sulphur	[109, 187, 262]
149	Dodecane	Alkane	[193]
150	Eicosane	Alkane	[135]

#	Name	Chemical classification	References
151	Ethanol	Alcohol	[109, 187]
152	Ethyl-acetate	Carbonyl	[262]
153	Ethylaniline	Aromatic and nitrogen	[99]
154	Ethyl-benzene	Aromatic	[144, 201]
155	Ethyl-benzol	Aromatic	[158]
156	Ethylene carbonate	Carbonyl	[99]
157	Ethyl-hexanol	Alcohol	[99]
158	Ethylenimine	Nitrogen	[90]
159	Formaldehyde	Carbonyl	[85, 128]
160	Heptanal	Carbonyl	[109, 118, 126, 158, 190, 196]
161	Heptane	Alkane	[109, 263]
162	Hexadecanal	Carbonyl	[135]
163	Hexadecane	Alkane	[99]
164	Hexanal	Carbonyl	[109, 118, 126, 128, 158, 190]
165	Hexane	Alkane	[109, 199, 205]
166	Hexanol	Alcohol	[196]
167	Hydroxy-2-nonenal	Alcohol, alkene and carbonyl	[203, 263]
168	Hydroxyacetaldehyde	Alcohol and carbonyl	[203]
169	Isobutane	Alkane	[187, 262]
170	Isolongifolene-5-ol	Alcohol and alkene	[99]

#	Name	Chemical classification	References
171	Isoprene	Alkene	[85, 118, 144, 187, 196, 205, 207, 264]
172	Isopropanol	Alcohol	[109, 187, 195]
173	Isopropyl-amine	Nitrogen	[158]
174	Isopropyl-benzene	Aromatic	[99]
175	Malondialdehyde	Carbonyl	[90, 203]
176	Methanol	Alcohol	[187, 264]
177	Methyl-cyclopentane	Alkane	[118, 196]
178	Methyl-propyl-sulfide	Sulphur	[90]
179	Naphthalene	Aromatic	[99]
180	n-Dodecane	Alkane	[158]
181	n-Nonal or Cyclohexanon	Carbonyl	[158]
182	Nonadecanol	Alcohol	[135]
183	Nonanal	Carbonyl	[126, 128, 190]
184	Nonanoic acid	Carbonyl	[158, 190]
185	Octanal	Carbonyl	[109, 126, 128, 190, 208]
186	Octane	Alkane	[144]
187	o-toluidine	Aromatic and nitrogen	[85]
188	Pentanal	Carbonyl	[90, 109, 126, 128, 203]

#	Name	Chemical classification	References
189	Pentamethyl-heptane	Alkane	[144]
190	Pentane	Alkane	[109, 144, 187, 199, 207, 263, 264]
191	p-menth-1-en-8-ol	Alcohol and alkene	[195]
192	Propanal	Carbonyl	[109, 126, 201, 205]
193	Propane	Alkane	[109]
194	Propanoic acid	Carbonyl	[190, 195]
195	Propanol	Alcohol	[85, 90, 109, 194, 199, 264]
196	Propyl-benzene	Aromatic	[118, 196]
197	Styrene	Aromatic	[118, 196]
198	Tetradecane	Alkane	[99]
199	Tetramethyl-urea	Carbonyl and nitrogen	[90]
200	Toluene	Aromatic	[109, 144, 187, 193]
201	trans-Caryophyllene	Alkene	[194]
202	Tridecanone	Carbonyl	[135]
203	Trichlorofluoro-methane	Halogen	[118]
204	Tridecane	Alkane	[135]
205	Undecane	Alkane	[90, 118, 196]
207	Xylene	Aromatic	[90, 144, 148]

Composition of DMEM (concentration in mg/L)

Amino Acids (mg/L)	Glycine	30	L-Alanyl-L-Glutamine	862	L-Arginine hydrochloride	84
	L-Cystine 2HCl	63	L-Histidine hydrochloride	42	L-Isoleucine	105
	L-Leucine	105	L-Lysine hydrochloride	146	L-Methionine	30
	L-Phenylalanine	66	L-Serine	42	L-Threonine	95
	L-Tryptophan	16	L-Tyrosine	72	L-Valine	94
Vitamins (mg/L)	Choline chloride	4	D-Calcium pantothenate	4	Folic Acid	4
	Niacinamide	4	Pyridoxine hydrochloride	4	Riboflavin	0.4
	Thiamine hydrochloride	4	i-Inositol	7.2		
Inorganic Salts (mg/L)	Calcium Chloride	264	Ferric Nitrate	0.1	Magnesium Sulfate	200
	Potassium Chloride	400	Sodium Bicarbonate	3700	Sodium Chloride	6400
	Sodium Phosphate monobasic	141				
Other Components	D-Glucose (Dextrose)	25	Phenol Red	15	Sodium Pyruvate	1

Method parameter for samples injected on the Pegasus 4D

This GC×GC-TOFMS instrument was provide by LECO (Corp., St. Joseph, MI) and it was used for the analysis of the exhaled breath samples and for the optimization of the analysis for culture cell line headspace.

Parameters for analysis of the cohort of individual breath samples

Pegasus 4D was connected to Unity 2 series thermal desorber (Markes International Ltd.). The modulator was mounted in an Agilent 7890 gas chromatograph equipped with a secondary oven and a quad-jet dual stage modulator working with liquid nitrogen to create the cold jets. Details regarding the system have been reported elsewhere [261]. The GC column set used was a combination of a 30 m × 0.25 mm i.d. Rxi-5Sil, 5% phenyl 95% dimethylpolysiloxane (Restek Corp., Bellefonte, PA) with a film thickness of 0.25 μm as 1D and a more polar 1.2 m × 0.10 mm i.d. mid polar BPX-50, 50% phenyl polysilphenylene-siloxane (SGE, Austin, TX) with a film thickness of 0.10 μm as 2D. This column set was already successfully used in previous VOC mixtures analysis [78,222,262].

In the thermal desorber Unity 2, samples were first purge with dry nitrogen during 1 min to remove the amount of water found in the sample. Then, tubes were heating at 300 °C during 5 min and VOCs sample were recollected on the cold trap of the thermal desorber at -10 °C. Samples were injected in the system with a fast heating of the cold trap at 300 °C during 3 min. In the GC system, helium was used as carrier gas with a constant flow of 1mL/min. The main oven held an initial temperature of 35 °C during 5 min and then increased until 240°C at a rate of 5 °C/min. The temperature offset for the secondary oven was 5 °C and operated in the iso-ramping mode. The modulation period was 4 s with a hot pulse duration set at 700 ms and a cooling time between stages of 1300 ms. Modulation was carried out on the very beginning of the 2D column with an offset of 10 °C compared to the temperature of the primary GC oven. The end of the second column went through the transfer line heating at 250 °C to reach the MS source at the same temperature. Analytes ionization was applied by an electron ionization mode with a filament voltage of −70 eV. The data acquisition rate was set at a frequency of 100 spectra/s for a mass range from 29 to 450 m/z. The detector voltage was set on 1500 V. Tuning and mass calibration were performed every days with perfluorotributylamine (PFTBA).

Data from Pegasus 4D were acquired and processed with the ChromaTOF® 4.5 software (LECO Corp.). Feature finding, mass deconvolution, integration feature and library searching were carrying out by this software. Mass spectral identification used Wiley (2011) and NIST (2011) databases with a match factor threshold >600. Statistical compare option of ChromaTOF® 4.5 software was used to align 2D chromatograms and built several feature tables which contains every feature found in each samples with a signal to noise ratio of 50, 100, 250 and 1000. Fisher ratio calculation was executed at this step with the statistical compare feature. Feature tables created were extracting in Excel® and normalize with the total amount volume. Then, non-specific compounds, such as siloxane compounds coming from the bleeding of the columns were manually removed. Critical Fisher value threshold was applied in Excel and reduce feature list was import in R 3.4.3 using the Rstudio interface (Free Software Foundation's GNU project) to perform random forest. The packages used were: MetabolAnalyze, FactoMineR, ggfortify, random forest, ggplot2, pca3d, RColorBrewer. This software was also used to carry out PCA and clustering analysis with euclidian distance and Ward method.

Parameters for the analysis of the cell culture optimization

For this study, headspace analysis was used instead of thermal desorption. The Pegasus 4D was connected to a multipurpose sampler (MPS) from Gerstel to perform automatic SPME injection. Vials were let on the tray at room temperature, and heated in the thermal unit at 37°C during different time laps (from to minutes). Fiber was then introduced to the headspace with a needle penetration of 38 mm and exposed during different time laps (from to minutes). The fiber moved to the inlet of the - Pegasus 4D (to be desorbed at 250°C during 120 s with a penetration of 54 mm. GC×GC and mass spectrometry parameters were exactly the same than for the analysis of breath samples. ChromaTOF® 4.5 was used to acquire the data obtained with a S/N of 100. Same data treatment than for breath samples was used with Fisher ratio, PCA and HCA calculation.

Method parameter for samples injected on the AccuTOF™ GCv 4G

This GC×GC-TOFMS instrument was provide by Jeol and it was used for the analysis of Tedlar® bags, for the investigation against endogenous VOCs found in breath samples, and also for the analysis of the headspace of culture cell line.

Parameters for the analysis of the headspace of cell line cultures

After incubation time, the fiber moved to the injector of the AccuTOF™ GCv 4G to be desorbed at 250°C during 5 minutes. The column set used is an Rxi-624 SilMS (30m x 0.25mm inside diameter x 1.40 µm film thickness; Restek Corporation., Bellefonte, PA, USA) as first dimensions and a stabilwax (2m x0.25mm x0.5 µm; Restek corporation) as the second dimension. This set was installed in an Agilent 7890 GC oven (Palo Alto, Ca, USA) where a ZX2 dual-stage thermal loop modulator (Zoex Corporation, Houston, TX, USA) was installed. The modulation period was 4 s with a hot pulse duration of 700ms and a cooling time between stages of 1300 ms. In the GC system, helium was used as carrier gas with a constant flow of 1 mL/min. The oven temperature program follows the same parameter than with the Pegasus 4D (35°C initially during 5 min and then increase until 204°C with a rate of 5°/min). The end of the second column went through the transfer line heating at 250°C to reach the MS source at the same temperature. Analytes ionization was performed by an electron ionization mode with a filament voltage of −70 eV and ionizing current of 300 µA. The data acquisition rate was set at a frequency of 50 spectra/s for a mass range from 30 to 400 *m*/*z*. The detector voltage was set on 2000 V. An acquisition delay of 1 minute was used. Tuning and mass calibrations were performer every 2 days with perfluorokerosene (PFK) (Tokyo Chemical Industry Co. Ltd., Tokyo, Japan) with a mass resolution of 8054 at m/z 293. Data were acquired using MassCenter (version 2.6.2b) (JEOL Ltd) and process in GC image 2.5 HR (Zoex Corporation). GC Project and Image Investigator features of GC-image were used to built a template 2D-chrommatogram from a cumulative image from all samples. Then, this template was applied to the centroided samples files with a minimum blob volume detection of 200 000. Mass spectral identification used Wiley (2011) and NIST (2011) databases with a match factor threshold >600. Feature table created was extract in Excel® sheet and R software to perform box-plot calculation.

Scientific communications

Scientific communications

Results related to this thesis were published in scientific papers and presented trough international symposiums via oral presentations and posters**.** All scientific communications are list below.

Publications:

R. Pesesse, P.-H. Stefanuto, F. Schleich, et al., Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC×GC-TOFMS, Journal of Chromatography B, https://doi.org/10.1016/j.jchromb.2019.01.029

Oral presentations:

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

Pesesse R, Stefanuto P-H, Schleich F, Louis R, Focant J-F
41th International Symposium on Capillary Chromatography & 14th GC×GC Symposium, Fort Worth, Texas, may 2017

GC×GC-(HR)TOFMS in Cancer Research

Pesesse R, Stefanuto P-H, Bertrand V, De Pauw-Gillet M-C, Schleich F, Meuwis M-A, Louis E, Louis R, Focant J-F
40th International Symposium on Capillary Chromatography & 13th GC×GC Symposium, Riva del Garda, Italia, May 2016

Characterization of human exhaled breath for non-invasive detection of diseases by GC×GC-TOFMS

Pesesse R, Stefanuto P-H, Schleich F, Meuwis M-A, Louis E, Louis R, Focant J-F
14th International Symposium o hyphenated techniques in chromatography and separation technology, Ghent, January 2016

GC×GC-TOFMS for reliable detection of candidate biomarkers in breath analyses

Pesesse R, Stefanuto P-H, Schleich F, Louis R, Focant J-F
International analysis of breath research summit 2015, Vienna, September 2015

In vitro and in vivo volatile fingerprinting of lung cancer

Stefanuto P-H, Pesesse R, Schleich F, De Pauw-Gillet M-C, Louis R, Focant J-F
39th International Symposium on Capillary Chromatography & 12th GC×GC Symposium, Fort Worth, Texas, may 2017

Investigating the Volatile Profile of Lung Cancer Cell Cultures

Stefanuto P-H, Pesesse R, Bertrand V, De Pauw-Gillet M-C, Focant J-F
38th International Symposium on Capillary Chromatography & 11th GC×GC Symposium, Riva del Garda, Italia, May 2014

Posters:

Detection of VOC biomarkers of lung cancer in exhaled breath by GC×GC-(HR)TOFMS

Pesesse R, Stefanuto P-H, Schleich F, Louis R, Focant J-F
Department day, University of Liège, Belgium, July 2017

Investigation of the possible effect of sampling on human exhaled breath sample integrity

Pesesse R, Stefanuto P-H, Schleich F, Meuwis M-A, Louis E, Louis R, Focant J-F
40th International Symposium on Capillary Chromatography & 13th GC×GC Symposium, Riva del Garda, Italia, May 2016

Screening for Putative Biomarkers of Lung Cancer in Exhaled Air

Pesesse R, Stefanuto P-H, Schleich F, Louis R, Focant J-F
Biomedica life sciences summit, Genk, Belgium, June 2015

Exhaled Air Analysis for Early Detection of Lung Cancer

Pesesse R, Stefanuto P-H, Schleich F, Louis R, Focant J-F
38th International Symposium on Capillary Chromatography & 11th GC×GC Symposium, Riva del Garda, Italia, May 2014