Cancer biomarker discovery using

Selected Ion Flow Tube Mass Spectrometry

A Thesis presented to

The Faculty of Graduate Studies

of

Lakehead University

by

Randa Babgi

In partial fulfillment of requirements

for the degree of

Master of Science in Biology

Thunder Bay, Ontario, Canada

Dec9th, 2016

Acknowledgments

Firstly, I would like to thank my supervisor, Dr. Brian Ross, for his support of my research project. He was always helpful and without his guidance and assistance during my scientific endeavours, my research work would not have been a success. Due to my interaction with Dr. Ross, I have improved my knowledge and research skills during the past two years. I would also wish to thank Dr. Suntres for his support and for providing me with tissue samples. I am also grateful to Hanan Alsarmi & Donna Newhouse for collecting blood samples at the Northern Ontario School of Medicine. Many thanks are given also to my friends Sultanah Albalawi, Catherine Chambers for their support.

To my family members, I would like to thank you for your support particularly my dear husband, my Mom and Dad who have been very supportive along my academic journey to complete my thesis. Without their support and words of encouragement, I would not have achieved what I have achieved today. I cannot forget to thank my sisters and brothers especially Ali Babgi who is spending many times looking after my child when I was at schoolat school. I am most grateful to him for the support and encouragement he gave me. During the difficult times, I always found solace and encouragement from him which motivated me to move forward.

Finally, I am grateful to the Saudi Bureau in Ottawa for the support and funding that facilitated my studies at Lakehead University.

Table of contents

List of Figures

List of Tables:

Table 15: Correlational analysis of normalised product ion count rates in the reaction with H_3O^+ and NO^+ precursor ions from the SIFT-MS analysis of breath and blood headspace.

List of Abbreviations

SIFT-MS	Selected Ion Flow Tube Mass Spectrometry
FS	Full Scan
MIM	Multiple ion monitoring
m/z	mass to charge ratio
GC-MS	Gas chromatography–mass spectrometry
VC	Volatile compound

Abstract

Lung cancer is one of the most common types of neoplasm and is a leading cause of death in Canada. Accurate early diagnosis is key to the effective treatment of the disease, however the current means to do so, such as Computed Tomography diagnostic imaging, are costly, invasive, and not practical for routine use. The detection of volatile cancer biomarkers in breath represents an attractive non-invasive means to diagnose the disease. It is unclear, however, which, if any, breath chemicals have diagnostic utility. In this thesis I used Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), a trace gas analytic method to (a) identify potential volatile cancer biomarkers in blood, and (b) investigate whether these markers are present in breath. Potential biomarkers were identified by comparing products ions formed in the reaction between hydronium (H_3O^+) and nitronium (NO^+) precursor ions and trace gases present in the headspace of plasma obtained from patients with breast cancer, colorectal or lung cancer, and healthy controls. Using this approach product ions of interest were identified which derive from a wide range of chemical classes including aldehydes, acids, alcohols and sulphides, including some which have been identified previously by other investigators. Many of these ions could be quantified in the breath of healthy controls and therefore be suitable for quantification by breath analysis. The production rate of most of these ions was, however, poorly correlated between those formed in the reactions between nasal breath and those formed in reactions with blood headspace, even when using in samples collected and analysed simultaneously from the same participants. The lack of correlation suggests that the breath trace gases from which these product ions are formed are not dependent on the blood concentration of the same gas, but likely derive mainly from the airways. As such while my data suggest that cancer

biomarkers may be found in the bloodstream, breath analysis is not a suitable means to non-invasively detect these cancer markers, in particular cancers of tissues other than those found in the airway. On the other hand, my data suggests that the detection of airway disease, including that of lung cancer, may be suitable candidates for the diagnosis and/or screening using breath analysis.

Chapter 1: Introduction

1. Background

Cancer is a class of disease characterized by uncontrolled cell division and the invasive spread of cancer cells[1] which if left untreated can lead to ill-heath and death. In 2011 almost 29.9 % of deaths in Canada were due to cancer. In addition, in 2015 around 196,900 new cancer cases occurred in Canada while the disease led to 78,000 deaths.[2] The risk of developing cancer is increased by extrinsic factors, for example, tobacco use and certain dietary habits[1], and intrinsic factors such as inherited susceptibility genes. Indeed, in the USA in 2015, approximately 171,000 of the evaluated 589,430 cancer deaths were attributed to tobacco smoking. Similarly, the World Cancer Research Fund has estimated that up to 33% of cases in the USA are associated with obesity, poor diet, and a sedentary life style.[1] These components may act either together or in succession to bring about disease in a complex etiological process that hampers screening and diagnosis. On the other hand, ten or more years frequently go between the initiating event and clinical presentation allowing for effective treatment to occur if the tumour is detected, although it should be noted that some aggressive tumors progress much more rapidly.[1]

1.1. Lung cancer

Out of all types of cancer lung cancer has the highest worldwide incidence. According to a recent report, lung cancer caused 158,040 deaths in the USA during 2015, comprising 27% of all cancer deaths. The high prevalence of lung cancer is primarily due to the environmental exposure to substances such as asbestos, arsenic and polycyclic hydrocarbons,[3] and those related to lifestyle, primarily to smoking.[4,5] Due to these causative

factors being so common lung cancer is likely to remain a significant global mortality risk for the foreseeable future.[6]

The 5 year mortality rate of lung cancer from the time of its presentation is very high, estimated to be about 85-90% even though surgical and chemotherapeutic treatment is available and commonly used.[7] Encouragingly, many reports suggest that if the lung cancer is detected at its early stage it can be easily treated.[8,9] However, there are limited techniques available such as Computerized Tomography (CT) Scan or, Magnetic Resonance imaging (MRI) which made difficult to detect which are generally too costly to be used as a general screening tool. However, early detection is made difficult due to the limited techniques available such as Computerized Tomography (CT) Scan or, Magnetic Resonance imaging (MRI) that which are generally too costly to be used as a general screening tool as described below.[10]

1.2 The challenges of early and timely diagnosis of lung cancer.

Patients with lung cancer are frequently symptomatic for a long time before they seek medical attention.[11] They generally present with numerous symptoms including coughing, blood or hemoptysis, shortness of breath or altered breathing, wheezing, chest pain, weight loss, and fatigue.[11–13] The fact that patients frequently ignore the initial symptoms obviously delays diagnosis but also worsens the prognosis, whereas enhancing patient awareness of lung cancer symptoms results in earlier stage diagnosis.[11,14] Indeed, early diagnosis generally increases the effectiveness of treatment, reducing mortality and morbidity, since the tumour can be treated at an earlier stage of the disease when it has caused minimal tissue damage and before it has metastasised. In particular, a timely and

accurate diagnosis aids in the determination of a course of action, in terms of selecting a treatment, and increases the treatment's effectiveness[14]. Ideally a diagnostic test should be accurate, cause no discomfort or risk to health, and be non-invasive.[15]

Since the early 20th century, there are various techniques have been used for detecting the presence of lung cancer, including chest radiography, histological assessment, and sputum cytology. These tools are not suitable for population-based screening due to the risk associated with these invasive procedures such as radiation exposure, and/or the involvement of technically difficult and expensive techniques such as gas chromatography mass spectrometry. As a result, they are not widely used for the early detection of cancer.[16] Morphological abnormalities such as lesions have also been used to assess the disease, but this is usually done when a person has passed the early stage (stage-1) of the cancer when the disease is most amenable to treatment. Indeed, the biopsy of tissues, such as for suspected basal cell hyperplasia and squamous metaplasia, types that are increased in smokers, are usually only used in the late stages of lung cancer and have, at best, limited usefulness.[17]

Most recently, CT scans have proven to be more useful for the early diagnosis of lung cancer compared to traditional radiography. In CT X-rays are used to form 3 dimensional images of the body, which assists in the detection of small early stage tumours. Recently, CT has being augmented by Positron Emission Tomography, which increases the diagnostic accuracy. An advanced form of CT imaging called 'spiral' or 'helical' CT scan, which provides more accurate images of internal organs, have allowed for the detection of tumours as small as 1-5mm.[8] Although this is an effective technique to detect lung cancer,

it is expensive, invasive and associated with some risk due to radiation exposure which makes it unsuitable for the sort of regular 'health check' which would revolutionise the early diagnosis of lung cancer and allow treatment to occur in its early stages. There is therefore a need for new diagnostic techniques to be developed, such as those utilising so-called cancer 'biomarkers', particularly when used as pre-imaging screening test to select those who should undergo further testing.[8]

1.3 Biomarker based cancer tests.

A biomarker can usefully be defined as a molecule that is associated with a particular physiological state, including pathological disease states; for example, plasma glucose concentration is a biomarker of diabetes. Ideally disease biomarkers would specifically and sensitively reflect a pathological state which could be utilized for diagnosis, estimating prognosis, treatment selection, and/or for monitoring the efficacy of treatment.[18,19] While biomarker development for clinical use is not without difficulties they have great appeal given they are relatively simple and inexpensive to use. Volatile compounds found in the breath are a particular type of biomarker that are particularly attractive since they can be quantified using an entirely non-invasive process. Specifically, 'breath analysis' involves analyzing the chemical composition of trace gas volatile inorganic and organic compounds in the exhaled breath as the end products of metabolic processes which may be able to tell us something about physiological and pathological states. The latter could well from the basis of a diagnostic test for cancer and other diseases and is the subject of this thesis.[8]

1.4 Volatile Organic and Inorganic Compounds as a type of biomarker.

Volatile Compounds (VC) have a high vapour pressure at room temperature under normal pressure conditions and therefore exist, to varying degrees, in the gas phase. They can be organic aliphatic or aromatic compounds, or inorganic such as nitric oxide. Consisting of nitrogen, oxygen, carbon dioxide, inert gases and water, breath also consists of approximately 1000 trace VC. The concentrations of VCs in the breath range from parts per million (PPM) to parts per trillion (PPT) with some of the most abundant being isoprene, acetone, ammonia, and propanol.[20] It has been proposed that VC can be used as disease markers which have the potential to form the basis of diagnostic tests, particularly exhaled breath, with a growing body of evidence supporting that claim (see Section 1.7).[20] Their interpretation is complicated, however, by the fact that the compounds one breathes out have a number of sources of origin all of which can be present simultaneously. [20]

1.5 Sources of VC in breath.

Many of the VC in the breath are exogenous in origin, that is, what is breathed out derives from what is breathed in. Indeed, atmospheric air has been identified as the main source of breath VC originating from both natural and human-made sources.[21] For example, chemicals including trichloroethene, toluene and tetrachloroethylene are commonly found in the bloodstream but are thought to be exclusively exogenous in origin.[21] Aside from occupational chemical exposure applications such exogenous compounds are not of great interest as biomarkers. Endogenous VC, on the other hand, derived from metabolic processes taking place in the body including the airways, bloodstream (cells and plasma), and other tissues. VC as disease markers should, ideally,

not be present at all in ambient air. This is not true, however, for most common breath VC which do occur in ambient air, making the interpretation of breath concentrations difficult. For example, lipid peroxidation of unsaturated fatty acids produces volatile alkanes (ethane and pentane) and aldehydes (propanal and hexanal)[22], which are breathed out but also have exogenous sources such as the burning of hydrocarbons or are released by plants.[23] Another abundant breath gas, acetone is created by hepatocytes via decarboxylation of excess acetyl-CoA.[22] but is also a very commonly used solvent. Even isoprene, which is formed during the metabolism of mevalonate during cholesterol biosynthesis, is found exogenously to varying extents.[23] Adding to this complexity endogenous compounds are not always formed in the patients' own tissues. For example, ethanol and methanol in the breath is derived from intestinal bacterial flora.[23] Sulphur containing compounds in the breath can emanate in the liver and lungs, but predominantly derived from the gastrointestinal tract and oral cavity.[24] Similarly while ammonia in the breath can indicate kidney failure, the gas mostly originates from microorganisms in the mouth.[25] Moreover, the relationship between the VC in each body "compartment" (such as the circulation and various body tissues) is unclear even though much of the diagnostic potential of breath testing relies on there being a direct relationship between VC in the diseased tissue and VC in breath. It is therefore difficult to determine the actual source of breath VC and, hence, what any changes in their concentration may mean. In this study I minimised the confounding effect of ambient air by collecting breath samples in the same location throughout. I also collected breath in a manner which minimised the contribution of VC emanating from mouth microorganisms. Therefore, the impact of the problems of different VC sources cannot be underestimated

and will have to be taken into account when breath analysis is used in 'real world' clinical applications.

1.6 Use of VC in diagnostic tests.

In order to use VC as a clinical tool, the VC profiles (also known as volatomes) of healthy and unwell individuals, need to be well defined. Additionally, the source of each VC needs to be studied. As mentioned above, exhaled air is a blend of various sources of air. Specifically these include alveolar air, derived from lung and non-lung sources, dead space air in the airways of the lung, and ambient air. Dead space air includes the mouth, bronchiole, nose, and pharynx all of which can possess VC.[26,27] As such the manner in which breath is collected is of importance. Breath samples are collected from participants in different ways such as capnography, direct on-line measurements, or in Tedlar bags.[26–28] These collection techniques vary in terms of the contribution of dead space air to the volatome. When one is interested in VC which do not originate within the deadspace air of the airways, nasal air is preferable to collect, compared to that emanating from the mouth, since it minimizes the contribution of oral cavity microorganisms. Indeed, nasal and mouth air can contain markedly different VC concentrations. For example, Schmidt et al. reported that the concentration of ammonia in the mouth was approximately 20-fold higher than nasal air since most ammonia breathed out of the mouth is generated on the dorsal surface of the tongue.[29] Even using nasal air we cannot, of course, be certain of the actual tissue source of VC, although it does minimize the contribution of mouth microorganisms.

1.7 Use of VC as disease biomarkers.

To date the only 'breath test' in common clinical practice is used to diagnose the presence of *Helicobacter pylori* in the stomach. In that test ingestion of isotopically labelled urea is catabolized by urease present in the bacterium leading to the release of labeled carbon dioxide, which can be detected in the breath.[23] Another less common, but commercialized application, measures nitric oxide as a measure of airway inflammation[20], while other applications are still in development such as the detection of hydrogen cyanide as a marker of lung *Pseudomonas aeruginosaaeroginosa* infection.[30] The catabolism of isotopically labelled erythromycin to carbon dioxide has been used to estimate the clearance rate of the chemotherapeutic drug docetaxel as a means to detect hypo-metabolisers who will experience severe toxic reactions.[31] Finally, the catabolism of glucose to hydrogen has been assessed using a breath test to determine bacterial growth rates in the gastrointestinal tract as may occur in several bowel disorders.[32]

1.8 Use of VC as diagnostic or screening test for cancer.

Encouraged by such applications many researchers have attempted to identify biomarkers of lung cancer as summarized in Table 1. As can be seen there have been a large number of potential biomarkers identified although none have been developed into a routinely used clinical test. The most developed, from a commercial perspective at least, emanate from the research group lead by Philips who have made use of Gas Chromatography Mass Spectrometry (GC/MS) to identify lung cancer markers (mostly alkanes) and have reported that these markers can detect lung cancer with an approximate sensitivity and specificity of 80%.[33–36]

Table 1. Various volatile organic compounds of biomarkers for lung cancer in literature.

Biomarkers	Disease	n	Age (Years)	Detection Method	Sensitivity %	Specificity %	Ref.
S, Hpp, Hm, D, Pb, U, C, Cmp, Mt, B, Bt, Ibm, O, He, No, Hp, Bd, Hpd, Hx, Cy, Bm, Hl.	LC	60	67	GC/MS	100	81.3	[33]
An, OT, C.	LC	28	60	Gas Sensor array	85	100	[37]
Pa, But,Pt, Hx, Hl, Oc, N	NSCLC	40	68	SPME-GC/MS	NC	NC	[38]
S, D, I, B, U, He, Hx, Pb, Tb, Hl, Mc.	LC	29	>50	SPME-GC/MS	86	69	[16]
Bu, Mo, E, Ac, Pn, I, Pr, Ds, Cd, B, T	LC	14	64	Sensor array	71	92	[39]
Bu, T3m, T7m, O4m, H3m, Hp, H2m, Pn, D5m	LC	178	64	GC/MS	90	82	[34]
Cdd, Pdt, Ba, Pam, Dd, Cdb, Bo, Fu, Bdl, Pd, Tc, Hid, Pr, Dm, Bade, Hd.	LC	193	66	GC/MS	85	80	[35]
I, Mp, Pn, Eb, X, Tb, T, B, Hp, D, S, O, Pen	NSCLC	36	67	GC/MS	72	94	[40]
F, Pr, Ac, I, Ot	LC	17	62	PTR/MS	54	99	[41]
Pr, Po,Ett, Pmm, Pmo, Hxd, Dh, Hxm, Hid, Cp, Bht, Cmt, Pm, Imoh, Ii, Tu, Tpd, Ba, Bnda, Patc, Pam, Tet, Ben, Cyd, Fu, Bc, Bcp, Atp, Ae, Bee.	LC	193	NS	GC/MS	85	81	[36]

Abbreviations: Where LC: Lung Cancer, NSCLC: Non-Small Cell Lung Cancer, n: number of patients, Age: Average Age, NC: Sensitivity and Specificity are not calculated. NS: Not stated.

Biomarker abbreviations: ***Ac*****:** Acetone (3). ***Ae:*** 2-ethyl-9,10-anthracenediol, (1). ***An****:* Aniline (1). ***Atp:*** 1,2,3,4-terahydro-9-propyl-anthracene,. (1). ***B*****:** benzene (4). ***Ba*****:** Ethyl-4-ethoxy benzoate(2). ***Bc:*** 1,1-(1,2-cyclobutanediyl)bis-,cis-benzene, (1). ***Bd:*** 1,4-dimethyl-benzene, (1). ***Bt:*** 1,2,4-trimethyl-benzene, (1). ***Bo*****:** 1-oxybis-benzene (1). ***Bu*****:** Butane (2). ***Bm:*** 1-methylethenyl-benzene, (1). ***Ben*****:** Benzophenone (1). ***Bade:*** Diethyl benzene-1,2-dicarboxylate (1). ***Bdl:*** 2,2-diethyl-1,1-Biphenyl, (1). ***Bee:*** 1,1-ethylidene-bis[4-ethyl]-benzene, (1). ***Bep:*** 1,1-[1-(ethylthio)propylidene]bis-Benzene, (1). ***Bht:*** 7,7-trimethyl-(1S)-bicyclo[2.2.1]heptan-2-one, , (1). ***But:*** Butanal, (1). ***Bnda*****:** 5-(Ethoxycarbonyl)bicyclo[3.2.2]nonane-1-carboxylic acid (1). ***C:*** methyl-cyclopentane (2). ***Cd*****:** Carbon disulfide (1). ***Cp*****:** Camphor (1). ***Cy***: cyclohexane (1). ***Cdd:*** 1,5,9-trimethyl-1,5,9-Cyclododecatriene, (1). ***Cdb*****:** 2,5- 2,6-bis (1,1-dimethylethyl)-cyclohexadiene-1,4-dione, (1). ***Cmp*****:** 1-methyl-2-pentyl-Cyclopropane, (1). ***Cmt*****:** α,α-4-trimethyl-3-cyclohexene-1-methanol (1).***Cyd:*** 2,6-bis(1,1-dimethylethyl)-4-ethylidene-2,5-cyclohexadien-1-one, (1). ***D*****:** decane (3). ***Dd*****:** 10,11- dihydro-5H-dibenzo-(B,F)-azepin (1). ***Dh:*** 5,5-dimethyl-1,3-hexadiene (1).***Dm*****:** 4-methyl-decane, (1). ***DS*****:** dimethyl sulfide (1). ***E*****:** ethanol (1). ***D5m*****:** 5-methyl-decane (1). ***Eb*****:** Ethylbenzene (1). ***Ett*****:** 1,1,2-trichloro-1,2,2-trifluoro-ethane (1). ***F*****:** formaldehyde (1). ***Fu*****:** 2,5-dimethyl-furan, (2). ***Hd*****:** 2,5-dimethyl- 2,4-hexanedione, (1). ***He*****:** 1-hexene (2). ***Hl:*** Heptanal (3). ***Hm***; ,2-methyl heptane (1). ***Hp;*** 1-heptene (3). ***Hx*****:** Hexanal (3). ***Hpd:*** 2,4-dimethyl-heptane, (1). ***Hpp:*** 2,2,4,6,6-pentamethyl-heptane, (1). ***Hid:*** 2,3-dihydro-1,1,3-trimethyl-3-phenyl-1-H-indene, (2). ***Hxd*****:** 2,3-hexadiene (1). ***Hxm****:* 2-methyl-3-hexanone, (1). ***H3m:*** 3-methyl-hexane, (1). ***H2m*****:** 2-methyl hexane, (1). ***I***: isoprene (4). ***Ibm:*** 1,3-butadiene, 2-methyl-isoprene (1) ***Imoh:*** 5-isopropenyl-2-methyl-7-oxabicyclo[4.1.0]heptan-2-ol (1). ***Ii:*** a isomethyl ionone (1). ***Mc:*** Methyl cyclopropane (1), ***Mt*****:** trichlorofluoro-methane, (1). ***Mp*****:** 2-methyl-pentane (1). ***Mo:*** methanol (1). ***N:*** nonanal (1). ***No*****:** 3-methyl-nonane, (1). ***O*****:** 3-methyl-octane,)(2). ***Oc:*** Octanal (1). ***OT:*** o-toluidine (2). ***O4m*****:** 4-methyl-octane, (1). ***Pa:*** propanal (1). ***Pd*****:** 2,4-dimethyl-3-pentanone (1). ***Pb*****:** propyl benzene (2). ***Pm*****:** *p*-menth-1-en-8-ol (1). ***Pn*****:** pentane (2). ***Po*****:** 4-penten-2-ol (1). ***Pr*****:** propanol (4). ***Pt***: pentanal (1). ***Pdt*****:** 2,2,4-trimethyl-pentan-1,3-dioldiisobutyrate, (1). ***Pam*****:** propanoic acid, 2-methyl-, 1-(1,1-diamethylethyl)-2-methyl-1,3-propanediyl ester (2). ***Patc:*** propanoic acid, 2,2,4-trimethyl-3-carboxyisopropyl, isobutyl ester (1). ***Pen:*** pentamethylheptane (1). ***Pmm*****:** 2-methoxy-2-methyl-propane , (1). ***Pmo:*** 1-(methylthio)-(E)- (1)1-propene,. ***S*****:** styrene (3). ***T*****:** toluene (2). ***Tb:*** 1,2,4-trimethyl benzene (2). ***Tc*****:** trans-caryophyllene (1). ***Tu:*** 2,2,7,7-tetramethyltricyclo[6.2.1.0(1,6)]undec-4-en-3-one (1). ***T3m:*** 3-methyl tridecane, (1). ***T7m*****:** 7-methyl-tridecane, (1). ***Tet*****:** 1,2,4,5- 3,3,6,6-tetraphenyl-tetroxane, (1). ***Tpd*****:** 2,2,4- Trimethyl-1,3-pentanediol diisobutyrate.(1). ***U***: Undecane (2). ***X:*** Xylene (1).

Many of the cancer biomarker studies has a low sample size, which reduces their statistical power and hence the robustness of their findings, possibly explaining the huge variety of potential biomarkers explored. Some putative markers have been reported in multiple studies however with these being listed in Table 2. Compounds in diverse classes such as ketones, aldehydes, aromatic compounds, alcohol, alkenes, and particularly alkanes have been identified in multiple studies from multiple research groups adding credence to their veracity as actual markers. On the other hand, for many of these compounds, such as toluene and methyl-cyclopentane, their biological source remains obscure thereby decreasing their plausibility as biomarkers, while others, such as propanol and acetone, are well characterised metabolically[16,33–41] and hence are somewhat more credible biologically as cancer markers.

A second, arguably more important problem with the current literature, is that all previous studies, including those conducted by the Phillips group looked at symptomatic patients only. This raises the question of whether the marker is actually present in the pre-symptomatic period when they would be most useful, or whether they are mere 'epiphenomena' occurring subsequent to the primary pathophysiological process (such as the cachexia which occurs frequently in later stage cancer patients). This would require researchers to conduct so-called prospective investigations in either general or 'at risk' populations to determine if any of the potential biomarkers identified can detect lung cancer in the pre-symptomatic stage of the disease, a requirement for any useful diagnostic or screening test. Prospective studies are very expensive to perform given they require large sample sizes and long follow-up times to achieve the statistical power necessary to determine the utility of any marker, a fact that likely explains the lack of such data in the

literature. A less costly alternative aimed at generating evidence that would justify the cost of a prospective analysis, is to use a cross-sectional design that includes multiple cancer types. Specifically, a marker, which occurred in a single type of cancer is more likely to be a primary marker, at the very least being one worthy of the time and expense of undertaking a large prospective study. In my study I will therefore use a larger sample size than many studies and include several different types of cancer (lung, colorectal, and breast).

1.9 Selected Ion Flow Tube Mass Spectrometry (SIFT/MS).

The gold standard for VC analysis is GC/MS which has been used in many cancer biomarker identification studies.[34,35,37,40] GC/MS is however technically demanding and costly. Other techniques have been developed including electronic sensors, ions mobility spectrometry, and various mass spectrometric techniques all of which avoid chromatographic separation and are generally technically easier, cheaper, and faster to perform. One such technique, Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), will be used in this study. This method quantifies trace gases using chemical ionization by reacting charged precursor ions (usually H_3O^+, NO^+ or O_2^+)[42] with the gases in a breath sample. The product ions produced in this reaction are characteristic of the VC present, while their rate of formation allows absolute quantification to be performed, that is, without the need for calibration standards.

The workings of the SIFT-MS are shown in Figure 1. Precursor ions are produced from low pressure water vapour using a microwave. The desired precursor ions are then selected by a quadrupole mass filter and carried in a stream of helium gas into the flow tube where they mix with the gas sample introduced into the flow tube using negative pressure. Most trace gases present react with the precursor ions to form distinctive product ions,

which are then quantified using a quadrupole mass spectrometer and photomultiplier combination. Once the rate of reaction between the precursor ion and trace gases is known, the absolute concentration of the gas can be calculated without the need for calibration standards. Importantly, the availability of three precursor ions can help differentiate between isomeric and isobaric compounds which generally do not produce the same product ions with all 3 precursors given that the H_3O^+ precursor usually reacts by proton transfer, while the NO^+ precursor reacts to form charge transfer products, including the molecular ion, or NO^+ -adducts depending on the ionisation energy of the trace gas.[42] The more energetic $O_2^{+\cdot}$ reacts similarly to NO^+ for many compounds, but produces more fragmentation product ions.[42] The instrument can be operated in Full Scan (FS) mode over a range of *m/z* values (generally 10 – 200 *m/z*) which is useful for determining which VC are present in a sample but provides low measurement precision due to the short length of time that each *m/z* value is quantified. This lack of precision can be overcome by using Multi-Ion Monitor (MIM) mode which quantifies a much more limited range of ions for a longer period of time thereby increasing the measurement precision of trace gases selected *a priori*.[42] Typical data produced using both FS and MIM modes are illustrated in Figures 2 and 3.

Table 2: Volatile compounds with altered abundance in human with lung cancer reported in more than one study.

Name of compounds	Times reported	Expected ions (*m/z*)		Reference
		H_3O^+	NO^+	
Acetone	2	59,77	88	[39,41]
Benzene	4	79	78	[16,33,39,40]
Butane	2	No reaction	No reaction	[34,39]
Decane	3	161	141	[16,33,40]
Ethyl-4-ethoxy benzoate	2	Unknown	Unknown	[35,36]
2,5-dimethyl-furan	2	Unknown	Unknown	[35,36]
Hexene	2	85	86	[16,33]
Heptanal	3	97, 115	113	[16,33,38]
Heptene	3	Unknown	Unknown	[33,34,40]
Hexanal	3	83, 101, 119	99	[16,33,38]
2,3-dihydro-1,1,3-trimethyl-3-phenyl-1-H-indene	2	Unknown	Unknown	[35,36]
Isoprene	4	No reaction	66, 68	[16,36,39,40]
3-methyl-octane	2	Unknown	Unknown	[33,40]
Propanoic acid, 2-methyl-1-(1,1-diamethylethyl)-2-methyl-1,3-propanediyl ester	2	Unknown	Unknown	[31,36]
Pentane	3	No reaction	No reaction	[34, 39,40]
Propanol	4	43, 61, 79, 97	59	[35, 36, 39, 41]
Styrene	3	105	104	[16,33,40]
o-toluidine	2	Unknown	Unknown	[37,41]
Toluene	2	93	92	[39,40]
Methyl-cyclopentane	2	Unknown	Unknown	[33,37]
1,2,4-trimethyl benzene	3	120	121	[16,33,40]
Undecane	2	161	141	[16,33]

The compounds listed in Table 1 which appear in more than one report in the literature are described. The reported product ions in reactions with H_3O^+ and NO^+ as detected by SIFT-MS are listed when these are known.[43–48]

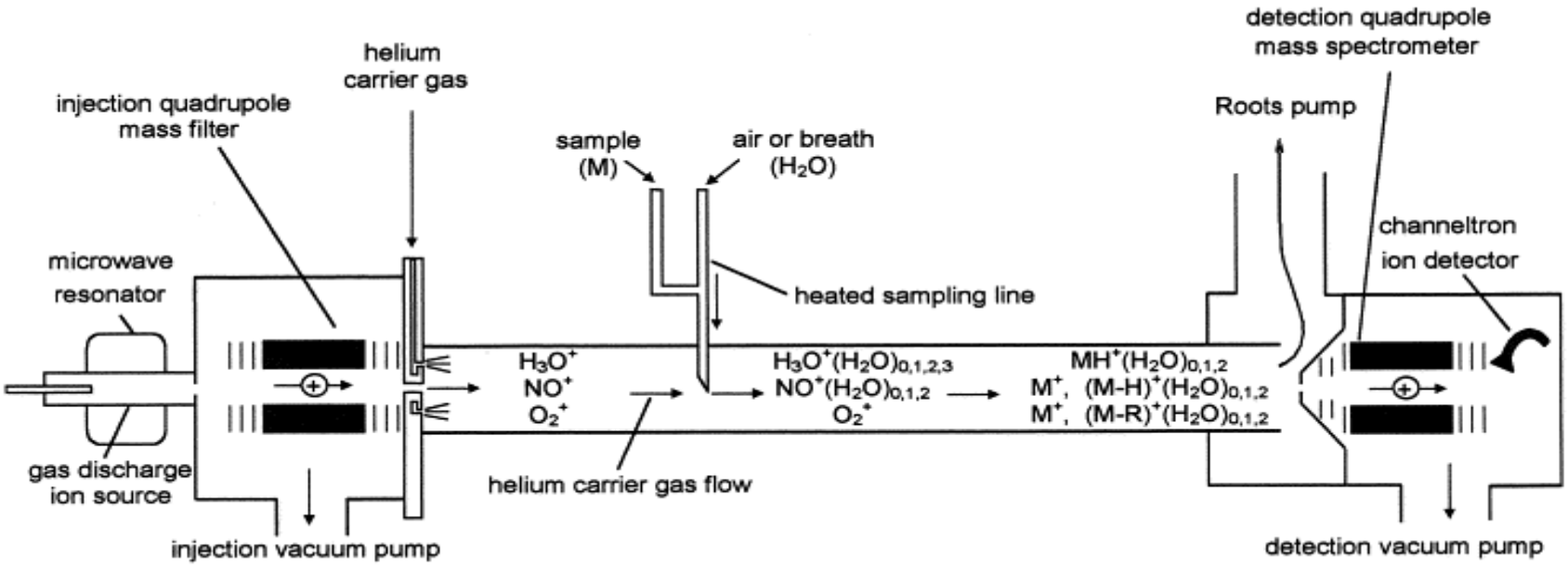

Figure 1: Schematic diagram of the Selected Ion Flow Tube Mass Spectrometer (SIFT-MS) used in this study. Some typical reactions of the H_3O^+, NO^+ and $O_2^{+\cdot}$ precursor ions are shown.[49]

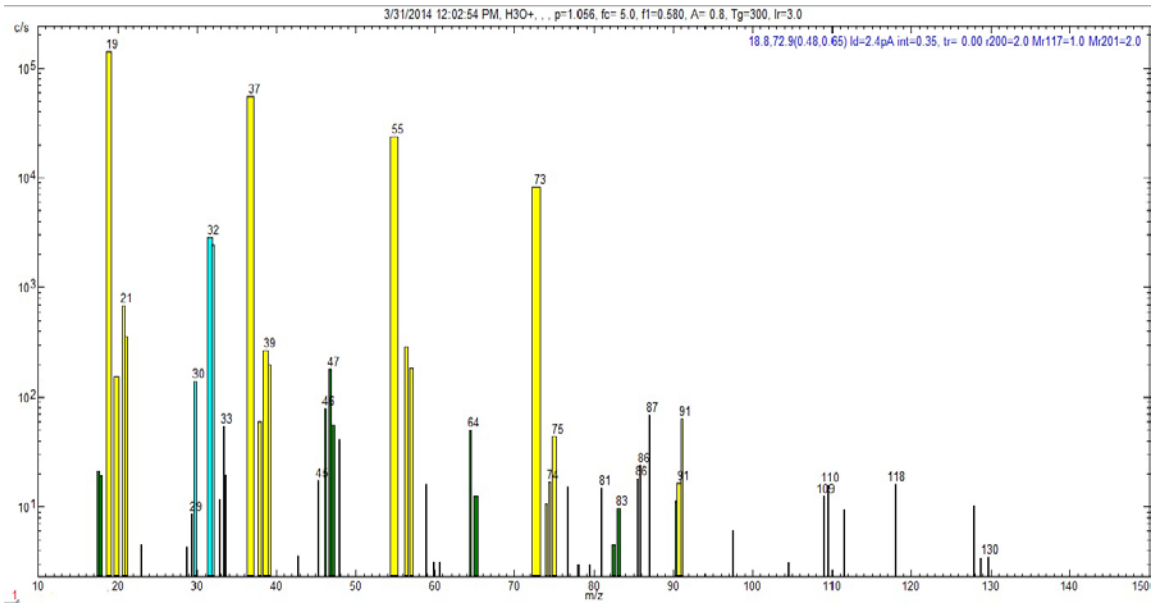

Figure 2. Sample of a mass spectrum produced using the FS mode of the SIFT-MS using the H_3O^+ precursor. Ion count rates, c/s (vertical axis),, plotted against that mass-to-charge ratio, *m/z*. The numbers over the peaks give the *m/z* value, the most abundant being the H_3O^+ ion (*m/z* 19) and its mono (*m/z* 37), di (*m/z* 55) and tri (*m/z* 73) hydrates. Clearly visible is the *m/z* 47 ion deriving from ethanol.

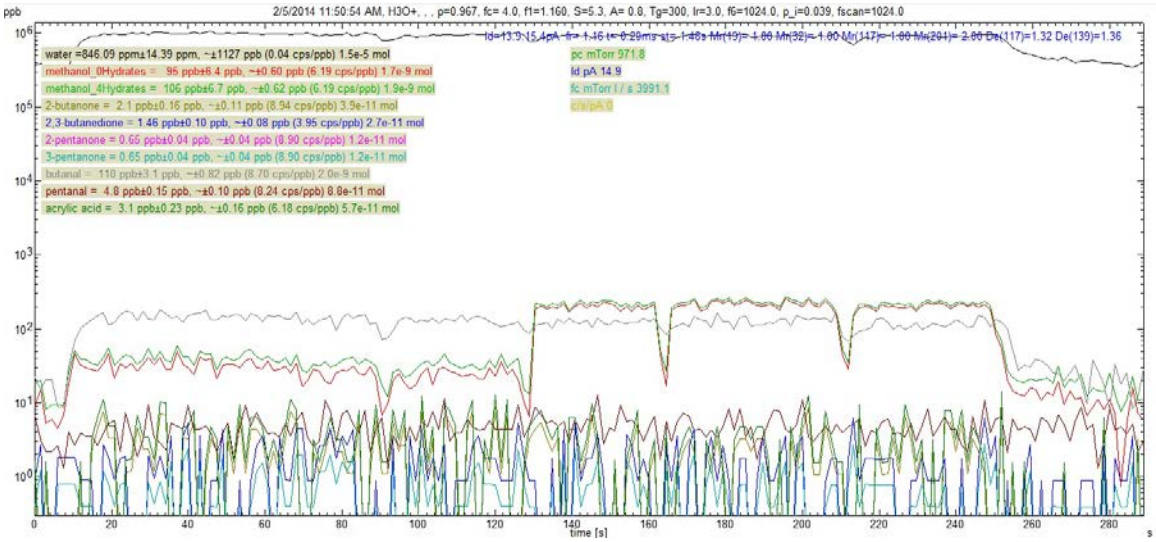

Figure 3. Illustration of the MIM mode of the SIFT-MS. Product ion count rates have been converted to concentrations using reaction rate constants to calculate absolute amounts of the gas in a unit time, with knowledge of the sampling flow rate being used to convert those to concentrations. The concentration of each compound is averaged over a selected region of data and displayed by the software. The most abundant compound shown is water vapour which varies as different samples are introduced into the instrument.

2. Objectives

The purpose of this study was two-fold. Firstly, the blood headspace volatome from patients with lung, breast and colon cancer and healthy controls were compared in order to identify potential cancer biomarkers. Secondly, the abundance of these putative biomarkers was measured in breath, and the relationship between blood and breath concentration investigated.

3. Hypothesis

I expect that (i) the volatome of blood headspace will differ between cancer and control groups, that is that the concentration of gases present in blood headspace will differ between participant groups, (ii) that one or more of the VC changes in abundance will be limited to a single type of cancer, and (iii) that these VC will be present in breath as well as blood headspace.

4. Materials and Methods

4.1 Participants.

The test populations for biomarker screening (study #1) included 390 participants - 287 cancers patients (112 of patients with breast cancer, 78 of patients with colorectal cancer, 97 of patients with lung cancer) and 103 healthy controls. Plasma samples had previously been collected and were made available to me under the existing consent which allowed for the analysis of biomarkers related to cancer (TBRHSC REB Approval 2004266). Due to plasma sample volume limitations the FS and MIM studies analyzed had to be conducted using different participants. A summary of the characteristics of the patients is shown in Table 3. The groups were not age matched (the control group was significantly ($P < 0.05$; post-ANOVA Tukey test) younger than each of the patient groups), nor sex-matched (the breast cancer group was all female which was not the case for the other groups). Further, the two analyses, termed FS and MIM in the Table (see below for further details) were not conducted using the same participants due to sample limitations.

Volunteers for the study comparing breath and blood (study #2) were recruited by advertisement under a protocol approved by the LU REB (REB Approval # 074 15-16). After giving informed consent demographic information, tobacco use, and health information were collected. No participants had a current or previous diagnosis of cancer and were otherwise healthy too. The participant characteristics are shown in Table 4.

Table 3: Participant characteristics for study #1.

Variable	Control n=103		Breast n=112		Colorectal n=78		Lung n=97	
	FS	MIM	FS	MIM	FS	MIM	FS	MIM
Sample size	50	53	50	62	35	43	42	55
Age (mean ± SD)	44 ± 14[1]	46 ± 13[1]	65 ± 12	66 ± 13	73 ± 12	71 ± 11	70 ± 10	70 ± 9
Sex (M, F)	11/39	9/44	0/50	0/62	15/20	23/20	19/23	22/33

Abbreviations: FS = full scan, MIM = multiple ion monitoring (see Section 1.10). M = male, and F = female. Participants were in stage II and III of their disease (cancer spread to nearest (II) or further (III) lymph nodes). [1] one-way ANOVA analysis showed a significant difference between groups for FS ($F_{3,173}$= 53.0; $P < 0.001$) and MIM ($F_{3,173}$= 52.0; $P < 0.001$) study groups with post-hoc Tukey tests indicating that the Control group differed significantly ($P < 0.001$) with all other groups.

Table 4: Participant demographics for study # 2.

Participants	Healthy controls
Sample size	30
Age (mean ± SD)	26.5 ± 3.77
Sex (M, F)	15 / 15

Abbreviations: M = male, and F = female.

4.2 Breath collection (study #2)

Breath was collected in bags made from poly vinyl fluoride (Tedlar) film (SKC Inc, Eighty Four, PA, USA) inflated and deflated with N_2 gas (99.99% purity) three times before use to remove residual volatiles emanating from the bag material. All analyses of breath took place within 2 hours of collection. Participants were asked to sit and relax for 3 minutes and breathe through their nose. Then they were asked to inhale fully through their nose, hold their breath for 5 seconds, and then exhale completely in the bag via disposable poly tetra fluoro ethylene (PTFE) tube. The first 5 seconds of exhalation were vented since this represents mainly bronchiolar tidal air, air more heavily contaminated with ambient trace gases, with the remainder being alveolar air, air which is closer to equilibrium with bloodstream VC. If the collected volume was not sufficient, participants were asked to repeat the procedure until the desired volume was attained. After that the bag was sealed using the PTFE valve integrated into the bag and taken to the laboratory for analysis. A bag of ambient air was collected at the same time using a sampling syringe. In the laboratory the bag was heated to human body temperature (37°C) for 20 minutes in an incubator to volatilize any condensed chemicals

4.3 Blood collection (study #2)

Blood was collected by a licenced phlebotomist. 6 ml of blood was withdrawn from the antecubital vein from a seated subject using a needle and vacutainer combination containing the anticoagulant ethylenediaminetetraacetic acid (EDTA). Blood samples were placed on ice and transferred to the lab. A 500µl aliquot of blood from the 6ml into a vacutainer transferred into a 250 ml flask, the flask incubated at 37°C for 10 minutes, and headspace gases analysed. All analyses took place within 2 hours of collection.

4.4 SIFT-MS

The SIFT-MS analysis was conducted using a profile 3 SIFT-MS Instrument (Instrument Sciences, UK) utilizing the H_3O^+, NO^+ and O^{2+} precursor ions as required. The flow tube pressure used was approximately 1 Torr and the temperature approximately 27 °C in all experiments, reaction conditions which are typical of other studies in the literature. The sample gas inlet temperature used was at 100 °C chosen to prevent volatile chemical condensing in the sampling line (particularly that of water vapour), and the sampling rate used was 0.2ml/s. The instrument was allowed to stabilise for between 30 minutes and 1 hour before tuning of the mass filter and source to optimise pre-cursor ion count rates and ion purity by a combination of changing the ion source pressure, altering the injected *m/z* setting, or altering the ion energy setting of the mass filter. Moreover, the mass resolution and mass precision for the downstream mass spectrometer was routinely checked. The selected precursor ions made up over 99% of total precursor ions, and precursor ions count rates were approximately 800,000 for H_3O^+ (including hydrates), and 600,000 for NO^+. Product ion count rates were corrected for diffusion loss. Analysis breath and blood samples proceeded in two ways: (i) full scan (FS) mode generates mass spectrum with each *m/z* quantified for a short period of time with resulting low measurement precision particularly of low abundance ions, and (ii) multiple ion monitoring (MIM) in which a limited subset of ions related to particular volatile compounds are quantified in any given time period hence increasing sensitivity and measurement precision and, as a consequence, sensitivity. FS mode generates a value for each *m/z* in counts per second, whereas MIM converts product ion count rates into actual concentrations using predetermined reaction rate constants.[50] For FS analysis the mass range used was from *m/z* 10 to 200, with each scan taking 185 seconds using the H_3O^+ and NO^+ precursor ions. Mass spectra were generated using both blood headspace and ambient air. The MIM analysis mode

of the SIFT-MS was conducted over a period of 180s per sample using precursor and product ion combinations and reaction rate coefficients obtained from the literature (Table 5).

4.5 Statistical Analysis

Product ion count rates are partly dependent on precursor ion count rates. This is taken into account when calculating volatile chemical concentrations using MIM mode but 'raw' FS mode mass spectra are not directly comparable between analyses because precursor count rates vary somewhat between analyses runs. To allow mass spectra to be compared they are 'normalised' by dividing the count rate of each *m/z* by the precursor ion count rate (including hydrate) for each spectrum. (Data was analysed using the Statistica statistical analysis software (Dell Statsoft, Tulsa, OK, USA) as described in the Results section. A significance level of one in twenty (0.05) was routinely used.

Table 5: Summary of H_3O^+ and NO^+ product ions derived from specific compounds measured in this study.

Compound	Precursor ion	Reaction rate coefficient (x 10^{-9} cm^3 s^{-1})	Product ions (*m/z*)	Ref.
Acetone	H_3O^+ (19, 37, 55, 73)	3.9	59 (1), 77 (1)	47
Formaldehyde	H_3O^+ (19, 32)	3.4	31 (1), 33 (-100.4), 83 (-100.9)	51
Acetaldehyde	H_3O^+ (19, 37, 55, 73)	3.7	45 (1.576), 81 (1.831)	51
Hydrogen Sulphide	H_3O^+ (19, 37, 55, 73)	1.9	35 (1), 53 (1)	52
Propanol	NO^+ (30, 48)	2.3	59 (1)	46
Propanal	NO^+ (30, 48)	3.0	55 (-0.006), 57 (1.0)	47
Hexanal	NO^+ (30, 48)	2.5	99 (1)	47
Malondialdehyde (MDA)	NO^+(30, 48)	4.8	89 (2.5), 102 (2.0)	53
Acetic Acid	NO^+ (30, 48)	0.9	90 (1)	54
Isoprene	NO^+ (30, 48)	1.7	66 (-0.006), 68 (1)	55

The table describes the precursor and product ions (with the multiplier value for each ion) and the reaction rate constant used to calculate gas concentrations using SIFT-MS as described in the referenced papers to which the reader is referred for further details.

5. Results

5.1 Analysis of traces gases present in headspace of blood obtained from cancer patients.

Volatile chemicals in the headspace of blood plasma obtained from patients with lung, breast, or colorectal cancers, and from healthy controls, was analysed using SIFT-MS. The MIM and FS modes were used to quantify the headspace concentration of particular gases (MIM mode) expected to be found in plasma headspace and/or which had previously been shown to be altered in cancer as described in Table 6. The headspace concentration of acetaldehyde differed between subject groups considering, in the first instance, both sexes combined (C in the Table) (P<0.05; one way ANOVA) with post-hoc analysis indicating reduced acetaldehyde concentration in all three patient groups compared to healthy controls (Tukey test; P<0.05). Propanal concentration also differed significantly between groups (P<0.05) with post-hoc testing indicating significantly increased concentrations in the headspace of plasma from patients with breast cancer, but not colorectal or lung cancers, compared to healthy controls (P<0.05). Applying a Bonferroni correction resulted in no significant differences being detected (Table 6). The analyses of normalised mass spectra obtained using the FS mode (Tables 7 and 8) showed statistically significant differences in the rate of formation of a number of product ions formed in the reaction with H_3O^+ with blood headspace. For many of these significant differences (post-hoc Tukey; $P < 0.05$) were found in all 3 patients groups compared to controls (*m/z* 63, 85, 89, 103, 104, 119, 121, and 124), some in only 2 groups (*m/z* 42, 51, 69, 86, 102, 107,and 125), and some in only one these being *m/z* 60, 70, and 179 in lung cancer only, *m/z* 45, 81, 99, and 108 in breast cancer only, and *m/z* 96, 169, and 170 in colorectal cancer only. For the reaction with NO^+ precursor ion many, though fewer than with H_3O^+, differed between groups (Table 8). For many of these significant

differences (post-hoc Tukey; $P < 0.05$) were found in all 3 patients groups compared to controls (*m/z* 114 only), none in only 2 groups, and some in only one these being *m/z* 90 and 195 in lung cancer only, *m/z* 29, 80, 102, and 116 in breast cancer only, and none in colorectal cancer only.

Given the large number of simultaneous comparisons a Bonferroni correction was applied based on 190 comparisons per precursor ion giving a threshold p-value of 0.0002.[56] This resulted in only *m/z* 99 from the H_3O^+ reaction differing between groups for combined male and female participants and female only, and *m/z* 124 with both sexes combined, with no significant differences being detected post-Bonferroni for the NO^+ reaction.

Given that gender and age were not matched across groups the effect of each was investigated by both including only one sex in the analysis or by directly comparing values in males and females within each participant group (Tables 6-11). For the MIM experiments, acetaldehyde concentration was significantly ($P < 0.05$) lower than healthy controls in both male and females, with the exception of females with lung cancer, although the difference was more pronounced in men compared with women (Table 6). Indeed, comparison of concentrations between men and women showed that acetaldehyde concentration was higher in males compared to female healthy controls (t-test; $P < 0.05$). For propanal no significant differences were found between groups when only male or female subjects were included in the analysis ($P > 0.05$). No significant correlations (Table 10) were found between the measured volatile chemical concentration except between acetone and age in the healthy control group ($P < 0.05$). Concluding that there was no effect of sex in all compounds except in acetaldehyde.

For the FS data, segregating by sex retained many statistically significant differences with the normalised counts differing between groups more often in females than males (Tables 7 and 8), although for many significant differences were observed in both sexes. In no case was any product ion that was significantly different between groups in the combined data set found

to not differ significantly between group in both the male and female sub-groups. Comparison of normalised count rates between male and female participants in each subject group did reveal differences (Tables 7 and 8). ($P < 0.05$; one way ANOVA).

The effect of age was investigated using linear regressional analysis (Tables 9-11). For the chemical concentration data collected using MIM acetone levels were weakly ($r^2 < 0.20$) correlated to age in the control group only, and malondialdehyde in the breast cancer group only; no other significant correlations were observed (Table 9). For the FS analysis (Tables 10 and 11) a number of significant correlations were observed although these were mostly weak ($r^2 < 0.20$). No correlations for either the MIM or FS data set were observed following application of the Bonferroni correction.

The product ions which were found to differ between participant groups are summarised in Table 12, listed by whether they differed from the control groups in all cancer groups, in two out of three, or just in one. The possible identity of the chemical from which each ion may derive is also shown.

Table 6: Headspace concentrations of selected volatile chemicals measured using MIM mode SIFT-MS in male and female in patients with cancer and healthy controls.

Compounds	Sex	Control (n=53)	Breast (n=62)	Colorectal (n= 43)	Lung (n= 55)	ANOVA F-statistic
Acetone	C	252.9 ±186.2	202.7 ± 141.0	238.8 ± 210.8	346.6 ± 530.8	$F_{3, 209}$=1.82
	F	256.1 ± 201.2	202.7 ±141.0	236.6± 121.5	381.6 ± 675.3	$F_{2, 155}$=1.70
	M	235.8 ± 72.3	NA	240.9 ± 273.8	295.1 ± 182.4	$F_{2, 51}$= 0.35
Formaldehyde	C	7.6 ± 6.7	7.9 ± 5.6	9. 7 ± 5.1	7.5 ± 5.2	$F_{3,209}$= 1.06
	F	8.3 ± 6.9	7.9 ± 5.6	11.4 ± 5.7	7.4 ± 5.8	$F_{2, 255}$= 1.66
	M	4.4 ± 4.3	NA	8.1± 6.0	7.7 ± 4.4	$F_{2, 51}$=1.55
Acetaldehyde	C	13.8 ± 18.9	7.5 ± 4.1[1]	8.2 ± 3.6[1]	9.0 ± 4.9[1]	$F_{3,209}$= 3.48[1, no]
	F	10.5 ± 5.3[a]	7.5 ± 4.1[1]	7.8 ± 3.6[1]	9.3 ± 5.6	$F_{2, 255}$=3.37[1, no]
	M	31.4 ± 43.8[b]	NA	8. 6 ± 3.6[2]	8.6 ± 3.7[2]	$F_{2, 51}$=4.89[2, no]
Hydrogen Sulphide	C	4.0 ± 2.5	3.9 ± 2.2	4.8 ± 3	4.2 ± 2.2	$F_{3,209}$= 1.09
	F	3.8 ± 2.1	3.9 ± 2.2	4.6 ± 2.8	4.2 ± 2.4	$F_{2, 255}$=0.58
	M	4.9± 4.0	NA	5.0 ± 3.3	4.2 ± 2	$F_{2, 51}$=0.32
Propanol	C	21.7 ± 25.7	23.2 ± 18.6	26.7 ± 23.3	27.1 ± 29.2	$F_{3,209}$=0.52
	F	21.7 ± 27.3	23.2 ± 18.6	29.2± 27.2	27.1 ± 28.6	$F_{2, 255}$=0.61
	M	21. 6 ± 27.3	NA	24.3 ± 19.4	25.5 ± 31.3	$F_{2, 51}$=0.07
Propanal	C	20.4 ± 18.9	31.2 ± 24.5[1]	20.9 ± 19.3	21.4 ± 16.3	$F_{3,209}$= 3.13[1, no]
	F	21.4 ± 20.2	31.2 ± 24.5	26.1 ± 22.7	21.7 ± 16.5	$F_{2, 255}$=1.97
	M	15.5 ± 8.1	NA	15.1 ± 13.6	21.0 ± 16.6	$F_{2, 51}$=0.81
Hexanal	C	439.1 ± 1189.8	153.9 ± 228.6	150.9 ± 130	173.1 ± 203.4	$F_{3,209}$= 2.19
	F	389.5 ± 1118.8	153.9 ± 228.6	170.2 ± 159.3	182.3 ± 199	$F_{2, 255}$=1.18
	M	699.5 ± 1576.4	NA	132.7 ± 95.9	156.5 ± 217.1	$F_{2, 51}$=2.08
Malondialdehyde	C	38.2 ± 21.8	30.6 ± 16.4	33.0± 17.0	40.9 ± 34.7	$F_{3,209}$= 1.81
	F	39.7 ± 23.1	30.6 ± 16.4	30.0 ± 18.3	43.8 ± 40.2	$F_{2, 255}$=2.28
	M	30.3 ± 10.8	NA	35.9± 15.7	35.8 ± 21.9	$F_{2, 51}$=0.32
Acetic Acid	C	34.1 ± 17.7	30 ± 14.0	27 ± 10.2	32.4 ± 14.8	$F_{3,209}$= 1.79
	F	34.3 ± 18.4	30 ± 14.0	26.2 ± 9.6	33.2 ± 15.2	$F_{2, 255}$=1.40
	M	33.2 ± 14.8	NA	27.7 ± 10.9	30.9 ±14.5	$F_{2, 51}$=0.54
Isoprene	C	10.4 ± 8.3	17.1 ± 16.7	13.5 ± 17.9	17.1 ± 17.6	$F_{3,209}$= 2.10
	F	10.7 ± 8.3	17.1 ± 16.7	14.6 ± 20.6	16.3 ± 14.1	$F_{2, 255}$=1.60
	M	9.2 ± 8.7	NA	12.5 ± 15.4	18.6 ± 23	$F_{2, 51}$= 0.86

The concentrations of the indicated gases were quantified in plasma headspace of samples obtained from patients with breast, colorectal, and lung cancers, and from healthy controls. Data are shown as mean PPB ± SD. C= combined male and female, F= female only, M=male only. Concentrations were compared between groups using a one-way ANOVA and post-hoc Tukey test conducted as appropriate. 1: P<0.05, 2: P<0.01. Whether the ANOVA was significant after Bonferroni correction (for 30 comparisons) is indicated (Yes or No). Concentrations in males and females were compared by t-test, a: P<0.05, b: P<0.01.

Table 7: Normalised product ion count rates of the reaction between plasma headspace obtained from patients with cancer and healthy controls and H_3O^+ ions.

Ions (*m/z*)	Sex	Control (n= 50)	Breast (n= 50)	Colorectal (n=35)	Lung (n= 42)	ANOVA F-statistic
41	C	20 ± 30	10 ± 20	10 ± 30	10 ± 20	$F_{3,173}$= 2.03
	F	20 ± 30	10 ± 20[1]	20 ± 40	4 ± 10[1]	$F_{2,,128}$=2.95[1, no]
	M	10 ± 20	NA	10 ± 20	10 ± 20	$F_{2,42}$=0.10
42	C	100 ± 100	30 ± 100[2]	40 ± 100[2]	100 ± 100	$F_{3,173}$= 5.84[2, no]
	F	100 ± 100	30 ± 100[2]	30 ± 100[2]	100 ± 100	$F_{2,128}$=7.01[2, no]
	M	100 ± 100	NA	100 ± 100	100 ± 100	$F_{2,42}$=0.60
43	C	400 ± 400	200 ± 200	300 ± 300	300 ± 200	$F_{3,173}$= 2.56
	F	400 ± 400	200 ± 200[1]	300 ± 100[1]	400 ± 200	$F_{2,128}$=3.02[1, no]
	M	400± 200	NA	400 ± 400	300 ± 200	$F_{2,42}$=0.86
42	C	100 ± 200	40 ± 40[2]	100 ± 100	100 ± 40	$F_{3,173}$= 4.21[2, no]
	F	100 ± 100	40 ± 40	40 ± 40	50 ± 50	$F_{2,128}$=2.11
	M	400 ± 400[c]	NA	100 ± 200[1]	100 ± 100[1]	$F_{2,42}$=6.38[1, no]
51	C	1600 ± 1100	900 ± 500[1]	1400 ± 2200	1100 ± 500[1]	$F_{3,173}$= 3.27[1,no]
	F	1500 ± 1000	900 ± 500[2]	1300 ± 1200	1100 ± 600[2]	$F_{2,128}$=4.11[2, no]
	M	1900 ± 1500	NA	1700 ± 3200	1000 ± 400	$F_{2,42}$=0.81
52	C	30 ± 40	10 ± 20	20 ± 40	10 ± 20	$F_{3,173}$= 5.04
	F	30 ± 40	10 ± 20[2]	10 ± 30[2]	10 ± 20[2]	$F_{2,128}$=4.80[2,no]
	M	40 ± 40	NA	30 ± 100	10 ± 20	$F_{2,42}$=1.10
54	C	2300 ± 1400	2000 ± 1600	3200 ± 2400	4400 ± 8900	$F_{3,173}$= 2.48
	F	2000 ± 1200[b]	2000 ± 1600	2800 ± 1700	5500 ± 11900[1]	$F_{2,128}$=2.95[1, no]
	M	3300 ± 1900	NA	3700 ± 3200	3000 ± 1800	$F_{2,42}$=0.39
56	C	4700 ± 600	4500 ± 900	4500 ± 700	4700 ± 600	$F_{3,173}$=1.86
	F	4800 ± 600	4500 ± 900	4700 ± 600	5100 ± 1000[1]	$F_{2,128}$=2.87[1, no]
	M	4600 ± 400	NA	4300 ± 800	4500 ± 700	$F_{2,42}$=0.90
60	C	200 ± 200	100 ± 100	200 ± 200	300 ± 300[2]	$F_{3,173}$=4.10[2, no]
	F	200 ± 200	100 ± 100[2]	200 ± 200	300 ± 200[2]	$F_{2,128}$=4.66[2, no]
	M	200 ± 100	NA	200 ± 300	300 ± 300	$F_{2,42}$=0.51
63	C	300 ± 500	100 ± 100[3]	100 ± 200[3]	100 ± 100[3]	$F_{3,173}$= 6.43[3, no]
	F	200 ± 300[c]	100 ± 100[1]	100 ± 100[1]	100 ± 200[1]	$F_{2,128}$=2.84[1, no]
	M	800 ± 800[c]	NA	200 ± 300[3]	200 ± 100	$F_{2,42}$=10.28[3, no]
69	C	6800 ± 4500	4400 ± 2300[1]	6000 ± 6800	4700 ± 4500[1]	$F_{3,173}$= 3.41[1, no]
	F	6100 ± 3000[a]	4400 ± 2300[1]	4900 ± 3300[1]	4900 ± 2400[1]	$F_{2,128}$=2.95[1, no]
	M	9700 ± 7700	NA	7400 ± 9700	4600 ± 1700	$F_{2,42}$=1.97
70	C	100 ± 100	100 ± 100	100 ± 100	100 ± 100[2]	$F_{3,173}$=4.46[2, no]
	F	100 ± 100	100 ± 100	100 ± 100	100 ± 100	$F_{2,128}$=2.60
	M	200 ± 100	NA	100 ± 100	100 ± 100	$F_{2,42}$=2.86

Cont.

Ions (*m/z*)	Sex	Control (n=50)	Breast (n=50)	Colorectal (n=35)	Lung (n=42)	ANOVA F-statistic
81	C	400 ± 500	100 ± 100[2]	200 ± 200	200 ± 100	$F_{3,173}$= 5.56[2, no]
	F	200 ± 300	100 ± 100	100 ± 100	200 ± 100	$F_{2,128}$=2.31
	M	800 ± 900[b]	NA	200 ± 200[3]	200 ± 100[3]	$F_{2,42}$=8.33[3, no]
85	C	2500 ± 4000	300 ± 200[3]	1100 ± 4200[3]	400 ± 300[3]	$F_{3,173}$= 6.33[3, no]
	F	1900 ± 3700[a]	300 ± 200[3]	300 ± 200[3]	500 ± 300[3]	$F_{2,128}$=5.73[3, no]
	M	4700 ± 4300[a]	NA	2100 ± 6400	400 ± 300[1]	$F_{2,42}$=3.46[1, no]
86	C	100 ± 200	30 ± 40[2]	100 ± 100	30 ± 40[3]	$F_{3,173}$=5.51[2, no]
	F	100 ± 200	30 ± 40[2]	20 ± 40[2]	20 ± 40[2]	$F_{2,128}$=4.43[2, no]
	M	100 ± 200	NA	100 ± 200	30 ± 40[1]	$F_{2,42}$=3.29[1, no]
89	C	300 ± 400	100 ± 100[3]	100 ± 100[3]	200 ± 400[3]	$F_{3,173}$= 5.58[3, no]
	F	300 ± 400	100 ± 100[3]	100 ± 100	200 ± 200	$F_{2,128}$=6.52[3, no]
	M	300 ± 400	NA	100 ± 200	300 ± 600	$F_{2,42}$=0.56
92	C	500 ± 200	500 ± 200	400 ± 300	500 ± 200	$F_{3,173}$= 1.46
	F	500 ± 200	500 ± 200	300 ± 200[1]	400± 200	$F_{2,128}$=3.00[1, no]
	M	500 ± 200	NA	500 ± 300	500 ± 200	$F_{2,42}$=0.02
96	C	400 ± 400	300 ± 300	200 ± 300[2]	400 ± 300	$F_{3,173}$= 3.91[2, no]
	F	400 ± 400	300 ± 300	200 ± 300[1]	400 ± 300	$F_{2,128}$=3.07[1, no]
	M	400 ± 300	NA	300 ±300	400 ± 400	$F_{2,42}$=1.15
99	C	500 ± 500	200 ± 100[3]	300 ± 200	300 ± 200	$F_{3,173}$= 8.26[3, yes]
	F	400 ± 400[c]	200 ± 100[1]	200 ± 100[2a]	300 ± 200	$F_{2,128}$=3.79[1, no]
	M	1100 ± 700[c]	NA	300 ± 200[3a]	300 ± 200[3]	$F_{2,42}$=14.22[3, yes]
100	C	100 ± 200	30 ± 100	40 ± 100	40 ± 100	$F_{3,173}$= 2.23
	F	100 ± 100	30 ± 100	20 ± 30	50 ± 100	$F_{2,128}$=1.35
	M	300 ± 400[b]	NA	100 ± 100[1]	40 ± 100[1]	$F_{2,42}$=3.28[1, no]
102	C	400 ± 900	100 ± 100[1]	200 ± 1100	100 ± 100[1]	$F_{3,173}$= 3.08[1, no]
	F	300 ± 1000[b]	100 ± 100	30 ±40[1]	100 ± 50[a]	$F_{2,128}$=2.98[1, no]
	M	1100 ± 1400	NA	500 ± 1600[b]	100 ± 100	$F_{2,42}$=2.51
103	C	1100 ± 3300	100 ± 100[2]	100 ± 200[2]	100 ± 100[2]	$F_{3,173}$=4.53[2, no]
	F	1400 ± 3700	60 ± 100[1]	40 ± 100[1]	100 ± 100[1]	$F_{2,128}$=3.24[1, no]
	M	600 ± 3400	NA	100 ± 300	50 ± 50	$F_{2,42}$=1.90
104	C	1000 ± 3000	40 ± 100[2]	100 ± 100[2]	100 ± 100[2]	$F_{3,173}$= 4.56[2, no]
	F	1100 ± 3200	40 ± 100[1]	100 ± 100	100 ± 100	$F_{2,128}$=3.30[1, no]
	M	900 ± 2400	NA	100 ± 100	100 ± 100	$F_{2,42}$=1.95
107	C	5100 ± 8100	600 ± 700[3]	1000 ± 1700[3]	2900 ± 5500	$F_{3,173}$= 7.65[3, no]
	F	5000 ± 8500	600 ± 700[3]	800 ± 1700[3]	2500 ± 3300	$F_{2,128}$=6.56[3, no]
	M	5600 ± 6500	NA	1300 ± 1600	3400 ± 7400	$F_{2,42}$=1.65
108	C	200 ± 300	30 ± 100[3]	100 ± 100	100 ± 200	$F_{3,173}$= 9.11[3, no]
	F	200 ± 300	30 ± 100[3]	100 ± 100	100 ± 100	$F_{2,128}$=9.37[3, no]
	M	200 ± 200	NA	100 ± 100	200 ± 300	$F_{2,42}$=1.17

Cont.

Ions (*m/z*)	Sex	Control (n= 50)	Breast (n= 50)	Colorectal (n=35)	Lung (n= 42)	ANOVA F-statistic
117	C	100 ± 200	100 ± 100	100 ± 100	100 ± 100	$F_{3,173}$= 5.04
	F	100 ± 200	100 ± 100[2]	100 ± 100[2]	100 ± 100	$F_{2,128}$=4.23[2, no]
	M	100 ± 100	NA	100 ± 100	100 ± 100	$F_{2,42}$=1.04
119	C	300 ± 300	200 ± 100[2]	200 ± 200[2]	200 ± 100[2]	$F_{3,173}$= 4.12[2, no]
	F	300 ± 300	200 ± 100	100 ± 100[1]	200 ± 100	$F_{2,128}$=3.23[1, no]
	M	300 ± 300	NA	200 ± 200	200 ± 100	$F_{2,42}$=2.17
121	C	3900 ± 11600	40 ± 100[2]	100 ± 100[2]	50 ± 100[2]	$F_{3,173}$= 4.52[2, no]
	F	3900 ± 12000	40 ± 100[1]	40 ± 40[1]	40 ± 100[1]	$F_{2,128}$=3.27[1, no]
	M	3500± 10900	NA	100 ± 200	100 ± 100	$F_{2,42}$=1.80 [no]
122	C	300 ± 700	20 ± 40[2]	20 ± 40[2]	10 ± 30[2]	$F_{3,173}$= 4.83[2, no]
	F	200 ± 700	20 ± 40[1]	20 ± 40[1]	10 ± 20[1]	$F_{2,,128}$=3.40[1, no]
	M	300 ± 800	NA	20 ± 30	10 ± 30	$F_{2,42}$=2.06
123	C	200 ± 300	100 ± 100	200 ± 800	70 ± 80	$F_{3,173}$= 1.85
	F	200 ± 300[a]	100 ± 100[1]	100 ± 100[1]	100 ± 100[1]	$F_{2,128}$=3.12[1, no]
	M	400 ± 400	NA	400 ± 1300	100 ± 100	$F_{2,42}$=1.01
124	C	50 ± 100	4 ± 10[3]	10 ± 40	10 ± 20	$F_{3,173}$= 8.37[3, yes]
	F	40 ± 100[a]	4 ± 10[2]	10 ± 20	10 ± 30	$F_{2,128}$=5.63[2, no]
	M	100 ± 100[a]	NA	20 ± 50[2]	10 ± 30	$F_{2,42}$=6.19[2, no]
125	C	500 ± 600	100 ± 100[3]	300 ± 700	200 ± 200[3]	$F_{3,173}$= 6.53[3, no]
	F	500 ± 600	100 ± 100[3]	200 ± 100[3]	200 ± 200[3]	$F_{2,128}$=9.42[3, no]
	M	600± 600	NA	400 ± 1100	300 ± 300	$F_{2,42}$=0.80
126	C	40 ± 100	10 ± 20[2]	20 ± 100	10 ± 30	$F_{3,173}$= 3.99[2, no]
	F	40 ± 100	10 ± 20[2]	10 ± 20	10 ± 20[2]	$F_{2,128}$=3.97[2, no]
	M	40 ± 100	NA	30 ±100	10 ± 40	$F_{2,42}$=0.79
139	C	1100 ± 2300	100 ± 100	5800 ± 34200	100 ± 100	$F_{3,173}$= 1.24
	F	600 ± 1300[c]	100 ± 100[2]	100 ± 100[2]	100 ± 200[2]	$F_{2,,128}$=5.08[2, no]
	M	3100 ± 3900	NA	13600 ± 52300	100 ± 100	$F_{2,42}$=0.85
140	C	100 ± 100	3 ± 10	400 ± 2000	5 ± 20	$F_{3,173}$= 1.29
	F	50 ± 100[b]	3 ± 10[2]	10 ± 40	1 ± 20	$F_{2,128}$=3.75[2, no]
	M	200 ± 200	NA	800 ± 3100	2 ± 10	$F_{2,42}$=0.88
169	C	200 ± 500	30 ± 100	30 ± 50[1]	20 ± 30	$F_{3,173}$= 3.41[1,no]
	F	200 ± 500	30 ± 100	30 ± 50	20 ± 30	$F_{2,128}$=2.22
	M	200 ± 300	NA	30 ± 50	20 ± 30	$F_{2,42}$=3.10
170	C	20 ± 100	3 ± 10	3 ± 10[1]	10 ± 30	$F_{3,173}$=2.87[1,, no]
	F	20 ± 100	3 ± 10	0 ± 0	10 ± 30	$F_{2,,128}$=2.29
	M	20 ± 40	NA	10 ± 20	10 ± 20	$F_{2,42}$=0.82
179	C	10 ± 30	10 ± 20	10 ± 30	30 ± 100[1]	$F_{3,173}$=2.81[1, no]
	F	10 ± 20	10 ± 20	5 ± 20	50 ± 100[1]	$F_{2,128}$=3.59[1, no]
	M	10 ± 30	NA	10 ± 40	20 ± 30	$F_{2,42}$=0.17

Values shown are mean ± SD of normalized count rates multiplied by 10,000. C= combined male and female, F= female only, M=male only. Only the *m/z* values which have at least one ANOVA indicating a significant difference between groups in the combined, male or female participants are shown; all other ions showed no significant group differences. One-way ANOVA and post-hoc Tukey test were conducted as appropriate. 1: $P < 0.05$, 2: $P < 0.01$. Whether the ANOVA was significant after Bonferroni correction (for 190 comparisons) is indicated (Yes or No). The concentrations in males and females were compared by t-test, a: $P < 0.05$, b: $P < 0.01$

Table 8: Normalised product ion count rates of the reaction between plasma headspace obtained from patients with cancer and healthy controls and NO^+ ions.

Ions (*m/z*)	Sex	Control (n= 50)	Breast (n= 50)	Colorectal (n=35)	Lung (n= 42)	ANOVA F-statistic
26	C	0 ± 0	0 ± 0	0 ± 0	10 ± 30	$F_{3,173}$= 2.20
	F	0 ± 0	0 ± 0	0 ± 0	10 ± 30[2]	$F_{2,128}$=3.42[2, no]
	M	0 ± 0	NA	0 ± 0	0 ± 0	Cannot be calculated
29	C	5000 ± 2000	3900 ± 2300[2]	5700 ± 2000	4900 ± 2300	$F_{3,173}$= 5.22[2, no]
	F	4800 ± 1900	3900 ± 2300	5700 ± 1900[2]	5000 ± 2200	$F_{2,128}$=4.22[2, no]
	M	5900 ± 0.22	NA	5600 ± 2200	4800 ± 2500	$F_{2,42}$=0.82
80	C	500 ± 500	300 ± 300[1]	400 ± 500	300 ± 300	$F_{3,173}$= 2.92[1, no]
	F	500 ± 400[a]	300 ± 300	200 ± 200b	300 ± 300[b]	$F_{2,128}$=2.48
	M	900 ± 700[a]	NA	700 ± 700b	300 ± 300	$F_{2,42}$=2.72
90	C	200 ± 300	200 ± 200	100 ± 100	300 ± 400[1]	$F_{3,173}$= 2.71[1, no]
	F	200 ± 300	200 ± 200	200 ± 300	200 ± 300	$F_{2,128}$=0.84
	M	300 ± 100	NA	100 ± 100	400 ± 500	$F_{2,42}$=1.83
102	C	300 ± 400	200 ± 200[1]	200 ± 300	200 ± 300	$F_{3,173}$= 2.87[1, no]
	F	300 ± 400	200 ± 200	100 ± 200[1]	200 ± 200	$F_{2,128}$=3.44[1, no]
	M	300 ± 200	NA	200 ± 400	200 ± 400	$F_{2,42}$=0.16
114	C	1000 ± 3100	100 ± 400[2]	100 ± 200[2]	40 ± 100[2]	$F_{3,173}$=4.58[2, no]
	F	1100 ± 3700	100 ± 400	100 ± 300	40 ± 100[1]	$F_{2,128}$=3.21[1, no]
	M	800 ± 1600	NA	100 ± 100	40 ± 100	$F_{2,42}$=2.05
116	C	300 ± 400	100 ± 200[1]	200 ± 300	100 ± 200	$F_{3,173}$= 3.53[1, no]
	F	300 ± 400	100 ± 200[1]	200 ± 300	100 ± 200[1]	$F_{2,128}$=2.92[1, no]
	M	300 ± 600	NA	200 ± 200	100 ± 200	$F_{2,42}$=0.90
117	C	500 ± 800	100 ± 100	40700 ± 239200	200 ± 300	$F_{3,173}$= 1.36
	F	500 ± 800	100 ± 100	70900 ± 316500	200 ± 300	$F_{2,,128}$=1.94
	M	800 ± 800	NA	400 ± 800[1]	200 ± 200[1]	$F_{2,42}$=3.37[1, no]
119	C	300 ± 400	100 ± 100	41400 ± 246400	200 ± 300	$F_{3,173}$=1.36
	F	200 ± 400	100 ± 100	72400 ± 323300	200 ± 300	$F_{2,128}$=1.94
	M	400 ± 400	NA	100 ± 100[2]	200 ± 200[2]	$F_{2,42}$=5.16[2, no]
128	C	100 ± 200	20 ± 40	100 ± 200	200 ± 600	$F_{3,173}$= 1.69
	F	100 ± 200	20 ± 40[2]	30 ± 100[2]	100 ± 100	$F_{2,128}$=4.77
	M	100 ± 100	NA	100 ± 200	300 ± 900[3]	$F_{2,42}$=0.60[3, no]

Cont.

Ions (*m/z*)	Sex	Control (n=50)	Breast (n=50)	Colorectal (n=35)	Lung (n=42)	ANOVA F-statistic
142	C	20 ± 100	30 ± 100	100 ± 300	100 ± 500	$F_{3,173}$= 0.82
	F	20 ± 100	30 ± 100	200 ± 500[1]	20 ± 100	$F_{2,128}$=3.12[1, no]
	M	20 ± 40	NA	20 ± 50	200 ± 700	$F_{2,42}$=0.52
152	C	100 ± 300	20 ± 100	100 ± 100	100 ± 50	$F_{3,173}$= 0.95
	F	100 ± 300	20 ± 100	100 ± 100	20 ± 60	$F_{2,128}$=0.55
	M	20 ± 50	NA	100 ± 100[1]	10 ± 20	$F_{2,42}$=3.70[1, no]
157	C	30 ± 100	30 ± 100	10 ± 30	10 ± 100	$F_{3,173}$= 1.33
	F	20 ± 50	30 ± 100	100 ± 300	20 ± 80	$F_{2,128}$=0.55
	M	100 ± 100	NA	0 ± 0	5 ± 20[2]	$F_{2,42}$=5.25[2, no]
195	C	3 ± 20	0 ± 0	3 ± 20	40 ± 100[2]	$F_{3,173}$= 4.00[2, no]
	F	4 ± 20	0 ± 0	10 ± 30[2]	50 ± 100	$F_{2,128}$=3.93[1, no]
	M	0 ± 0	NA	0 ± 0	20 ± 70	$F_{2,42}$=1.08
200	C	3 ± 20	0 ± 0	10 ± 40	0 ± 0	$F_{3,173}$=1.79
	F	0 ± 0[a]	0 ± 0	20 ± 100[2]	0 ± 0	$F_{2,,128}$=3.98[2, no]
	M	20 ± 60	NA	0 ± 0	0 ± 0	$F_{2,42}$=1.98

Values shown are mean ± SD of normalized count rates multiplied by 10,000. C= combined male and female, F= female only, M=male only. Only the *m/z* values which have at least one ANOVA indicating a significant difference between groups in the combined, male or female participants are shown; all other ions showed no significant group differences. One-way ANOVA and post-hoc Tukey test were conducted as appropriate. 1: $P < 0.05$, 2: $P < 0.01$. Whether the ANOVA was significant after Bonferroni correction (for 190 comparisons) is indicated (Yes or No). The concentrations in males and females were compared by t-test, a: $P < 0.05$, b: $P < 0.01$

Table 9: Comparison between plasma headspace volatile chemical concentration and subject age in patients with cancer, and in healthy controls.

Compounds	Regressional analysis	Control (n= 53)	Breast (n= 62)	Colorectal (n= 43)	Lung (n= 55)
Acetone	r^2	**0.10**	0.00	0.00	0.04
	P	**0.03**	0.66	0.96	0.18
Formaldehyde	r^2	0.01	0.03	0.00	0.01
	P	0.55	0.23	0.74	0.55
Acetaldehyde	r^2	0.01	0.00	0.00	0.01
	P	0.42	0.74	0.69	0.46
Hydrogen Sulphide	r^2	0.00	0.02	0.05	0.02
	P	0.71	0.37	0.19	0.35
Propanol	r^2	0.01	0.03	0.02	0.03
	P	0.51	0.22	0.37	0.31
Propanal	r^2	0.00	0.05	0.01	0.00
	P	0.94	0.10	0.63	0.68
Hexanal	r^2	0.00	0.01	0.01	0.01
	P	0.77	0.61	0.52	0.55
Malondialdehyde	r^2	0.05	0.08	0.04	0.00
	P	0.12	0.14	0.26	0.76
Acetic Acid	r^2	0.02	0.03	0.00	0.01
	P	0.34	0.21	0.87	0.55
Isoprene	r^2	0.02	0.01	0.00	0.02
	P	0.36	0.50	0.74	0.39

Headspace concentration of selected compounds was compared with subject age using regressional analysis. The squared Pearson correlation coefficient (r^2) and the statistical significance (*P*) is shown. Correlations which are statistically significant are bolded.

Table 10: Comparison between the rate of product ions formation produced in the reaction between plasma headspace and H_3O^+ ions and subject age in patients with cancer, and in healthy controls.

Ions *(m/z)*	Regression analysis	Control (n= 50)	Breast (n= 50)	Colorectal (n= 35)	Lung (n= 42)
41	r^2	0.06	0.01	0.00	0.01
	P	0·11	0.41	0.92	0.48
42	r^2	0.02	0.01	**0.19**	0.01
	P	0.37	0.50	**0.01**	0.48
43	r^2	0.01	0.01	**0.19**	0.00
	P	0.48	0.42	**0.01**	0.99
45	r^2	0.04	0.02	0.10	0.00
	P	0.15	0.29	0.06	0.80
47	r^2	0.05	0.03	**0.12**	0.04
	P	0.11	0.24	**0.04**	0.19
48	r^2	0.04	0.01	**0.12**	0.0
	P	0.18	0.53	**0.04**	0.95
49	r^2	**0.12**	0.0	0.11	0.17
	P	**0.02**	0.81	0.05	0.01^2
50	r^2	0.00	0.01	0.02	0.02
	P	0.94	0.53	0.40	0.41
51	r^2	0.01	0.02	0.10	0.07
	P	0.40	0.36	0.10	0.13
52	r^2	0.03	**0.09**	**0.13**	**0.11**
	P	0.24	**0.04**	**0.03**	**0.03**
54	r^2	0.15	0.0	0.05	0.03
	P	0.01^2	0.66	0.19	0.31
56	r^2	0.05	0.00	0.03	0.03
	P	0.13	0.87	0.34	0.24
60	r^2	0.01	0.00	**0.19**	0.00
	P	0.47	0.90	**0.01**	0.86
63	r^2	0.05	0.07	0.09	0.00
	P	0.11	0.07	0.08	0.82
69	r^2	0.00	0.04	0.11	0.01
	P	0.66	0.16	0.05	0.65
70	r^2	0.00	0.04	0.08	0.00
	P	0.90	0.16	0.10	0.93
81	r^2	0.01	0.05	0.02	0.00
	P	0.62	0.12	0.41	0.82
85	r^2	0.00	0.03	0.11	0.03
	P	0.88	0.22	0.05	0.29
86	r^2	0.04	0.01	0.10	0.04
	P	0.18	0.53	0.06	0.20
89	r^2	0.00	0.02	0.01	0.00
	P	0.76	0.33	0.55	0.81
91	r^2	0.02	0.00	0.03	0.03
	P	0.35	0.87	0.33	0.25
92	r^2	0.01	0.00	0.03	0.01
	P	0.45	0.82	0.36	0.45

Ions *(m/z)*	Regression analysis	Control (n= 50)	Breast (n= 50)	Colorectal (n= 35)	Lung (n= 42)
96	r^2	0.05	0.00	**0.24**	0.00
	P	0.13	0.76	**0.00**	0.69
99	r^2	0.01	0.01	0.01	0.00
	P	0.51	0.59	0.60	0.72
100	r^2	0.00	0.00	0.09	0.02
	P	0.86	0.76	0.07	0.32
102	r^2	0.00	0.01	**0.13**	0.00
	P	0.68	0.41	**0.03**	0.89
103	r^2	0.02	0.01	0.10	0.04
	P	0.30	0.58	0.10	0.22
104	r^2	0.02	0.01	0.01	0.01
	P	0.32	0.51	0.68	0.65
107	r^2	0.06	**0.12**	0.05	0.00
	P	0.08	**0.01**	0.20	0.90
108	r^2	0.02	0.01	0.03	0.00
	P	0.28	0.44	0.34	0.94
109	r^2	0.00	0.01	0.04	0.05
	P	0.88	0.42	0.26	0.14
117	r^2	0.04	0.00	0.00	0.02
	P	0.18	0.67	0.72	0.36
119	r^2	0.02	0.01	**0.14**	0.01
	P	0.31	0.61	**0.03**	0.54
121	r^2	0.03	0.01	0.03	0.00
	P	0.27	0.44	0.36	0.86
123	r^2	0.00	0.00	**0.12**	0.05
	P	0.79	0.67	**0.04**	0.17
124	r^2	0.00	0.04	0.02	0.01
	P	0.83	0.15	0.45	0.57
125	r^2	0.01	**0.22**	0.10	0.00
	P	0.49	**0.00**	0.07	0.66
126	r^2	0.01	0.06	**0.11**	0.00
	P	0.61	0.09	**0.05**	0.68
139	r^2	0.01	0.00	**0.12**	0.02
	P	0.44	0.74	**0.04**	0.38
140	r^2	0.00	0.01	**0.12**	0.03
	P	0.73	0.59	**0.04**	0.28
169	r^2	0.04	0.00	0.00	0.05
	P	0.18	0.89	0.91	0.15
170	r^2	0.01	0.00	0.01	0.03
	P	0.59	0.81	0.53	0.29
179	r^2	0.06	0.01	0.00	0.01
	P	0.10	0.42	0.85	0.59

The rate of product ion formation was compared with subject age using regressional analysis. The squared Pearson correlation coefficient (r^2) and the statistical significance (*P*) is shown. Only those ions differing in normalized count rates between groups in Table 7 are shown. Statistically significant correlations are shown bolded.

Table 11: Comparison between the rate of product ion formation produced in the reaction between plasma headspace and NO^+ ions and subject age in patients with cancer, and in healthy controls.

Ions *(m/z)*	Regression analysis	Control (n= 50)	Breast (n= 50)	Colorectal (n= 35)	Lung (n= 42)
26	r^2	0.00	0.00	0.00	0.03
	P	0.00	0.00	0.00	0.25
29	r^2	0.00	0.04	0.00	0.00
	P	0.08	0.19	0.83	0.83
80	r^2	0.01	0.06	0.03	0.01
	P	0.55	0.09	0.32	0.44
90	r^2	0.02	0.03	0.21	0.00
	P	0.34	0.26	0.01	0.99
102	r^2	0.01	0.00	0.10	0.00
	P	0.57	0.71	0.06	0.69
114	r^2	0.03	0.03	0.03	0.00
	P	0.27	0.22	0.35	0.95
116	r^2	0.01	0.07	0.00	0.02
	P	0.59	0.07	0.69	0.43
117	r^2	0.00	0.01	0.02	0.00
	P	0.78	0.52	0.47	0.93
119	r^2	0.01	0.01	0.02	0.02
	P	0.42	0.45	0.47	0.35
128	r^2	0.06	0.02	0.00	0.01
	P	0.11	0.32	0.81	0.54
145	r^2	0.01	0.01	0.03	0.04
	P	0.52	0.51	0.34	0.22
152	r^2	0.01	0.01	0.00	0.09
	P	0.48	0.58	0.96	0.06
157	r^2	0.01	0.05	0.00	0.02
	P	0.44	0.13	0.81	0.39
195	r^2	0.00	0.00	0.04	0.01
	P	0.00	0.00	0.27	0.61
200	r^2	0.00	0.00	0.00	0.00
	P	0.90	0.00	0.83	0.00

The rate of product ion formation was compared with subject age using regressional analysis.
The squared Pearson correlation coefficient (r^2) and the statistical significance (*P*) is shown. Only those ions differing in normalized count rates between groups in Table 8 are shown.

Table 12: Summary of H_3O^+ and NO^+ product ions and compounds derived from study #1.

Groups	Product ions (m/z)	Precursor ion	Possible chemical identity from literature	Ref.
In all groups	63	H_3O^+	dimethyl sulphide	[52]
	85	H_3O^+	pentanoic acid, methanthiol	[52,54]
	89	H_3O^+	methyl propionate, ethyl acetate, butyric acid, malondialdehyde, putrescine, pentanol	[47,53,54,57,58]
	103	H_3O^+	pentanoic acid, trimethylacetic acid, ethyl propionate, methyl butyrate, cadaverine	[54,57,58]
	104	H_3O^+	methyl acetate	[57]
	119	H_3O^+	hexanal, hexanone	[47]
	121	H_3O^+	propyl benzene, phenylethanone, pentanoic acid	[43,54,59]
	122	H_3O^+	phenylethanol	[59]
	114	NO^+		
In two groups	42	H_3O^+		
	51	H_3O^+	methanol	[46]
	69	H_3O^+	pentanal, methanol	[46]
	86	H_3O^+	cadaverine	[58]
	102	H_3O^+		
	107	H_3O^+	pentanol, xylene, butyric acid, ethyl acetate, putrescine, benzaldehyde	[47,54,57,58]
	125	H_3O^+	pentanol, butyric acid	[47,54]
In Lung disease group	60	H_3O^+		
	70	H_3O^+	2-methyl-1-butanol	[60]
	179	H_3O^+		
	90	NO^+	acetic acid	[54,57]
	195	NO^+		
In Breast disease group	45, 81	H_3O^+	acetaldehyde	[47]
	99	H_3O^+	hexenal	[47]
	108	H_3O^+		
	124	H_3O^+	spermidine	
	126	H_3O^+		[58]
	57	NO^+	Propanal, 2-methy-2-propanol, octanol, propionic acid, methyl propionate, ethyl propionate	[46,47,54,57]
	80	NO^+		
	102	NO^+	butanone, malondialdehyde, acrylic acid	[47,53,54]
	116	NO^+	pentanone, hydroxybutyric acid	[47,61]
In Colorectal disease group	96	H_3O^+		
	169	H_3O^+		
	170	H_3O^+		

The ions listed are those found to differ significantly between participant groups (combined sexes) as described in Table 7 and 8, including the number of groups the differed in compared to healthy controls. The possible chemical identity of these ions is shown by comparing the *m/z* values to the SIFT-MS literature.

5.2 Investigation of putative cancer markers in healthy controls.

In the previous experiment a series of putative cancer markers, in the form of SIFT-MS product ions, were identified by means of analysing blood headspace obtained from patients with cancer. The first step in determining whether these have utility in cancer diagnosis using breath analysis is to determine whether they are present in breath. Furthermore, even if they are present in breath the precise relationship between blood and breath concentrations needs to be determined. As such breath and blood headspace was compared in a group of healthy controls. Participants provided a sample of blood and nasal breath, the normalised count rates of the putative markers ions quantified in blood headspace (and a water headspace control) and nasal air (and an ambient air control), and the count rates compared (Tables 13 and 14).

The mean values for each product ion were compared between the two groups to get the correlation between blood and breath and as illustrated in Table 15. Comparing the controls (water headspace or ambient air) to their respective biological samples showed that only some ions showed values significantly higher than these control samples. Specifically, for H_3O^+ reaction with blood headspace products of *m/z* 45, 60, 81, 99 and 169 differed significantly (t-test; $P < 0.05$) from controls but *m/z* 70, 96, 108, 124, 126, 170, and 179 did not; for nasal breath *m/z* 60, 70, 81, 99, 108, 126 and 179 differed from controls ($P < 0.05$) but *m/z* 45, 96, 124, 169, and 170 did not. A similar situation was evident for the NO^+ reaction products with only *m/z* 90 differing significantly from controls in blood headspace ($P < 0.05$) while *m/z* 29, 57, 80, 102, 116 and 195 did not. For nasal breath *m/z* 57, 90, 102 and 116 differed from ambient air ($P < 0.05$), while *m/z* 29, 80, and 195 did not. Comparing blood headspace and nasal breath showed that *m/z* 99,

108, 126, and 179 formed in the reaction with H_3O^+ differed significantly (paired t-test; $P < 0.001$) between blood headspace and breath, while *m/z* 90, 102 and 116 from the reaction with NO^+ ions differed between blood headspace and breath (paired t-test; $P < 0.05$). Finally, the ratios between the mean normalised ions count rates deriving from blood headspace and nasal breath were calculated and observed to vary widely between ions.

Normalised ion count rates for nasal breath and blood headspace were also compared using correlational analysis (Table 15). For H_3O^+ reaction products normalised ion count rates correlated between breath and blood headspace for product ion *m/z* 170, and for the NO^+ reaction *m/z* 29 and 116.

Table 13: Quantification of putative cancer marker products formed in the reaction with H_3O^+ ions in the blood headspace and nasal breath obtained from healthy controls using SIFT-MS.

Product ions	Water headspace	Blood headspace	Ambient air	Nasal breath	Ratio of breath to blood
45	121 ± 57	530 ± 234[3]	256 ± 148	413 ± 575	0.78
60	178 ± 50	682 ± 327[3]	266 ± 158	750 ± 410[3]	1.10
70	35 ± 53	49 ± 107	43 ± 31	220 ± 86[3]	4.49
81	121 ± 75	462 ± 204[3]	101 ± 54	281 ± 138[1]	0.61
96	147 ± 76	2365 ± 6718	288 ± 207	410 ± 280	0.17
99	219 ± 98	502 ± 363[2]	158 ± 54	250 ± 71[1,c]	0.50
108	36 ± 29	46 ± 56	193 ± 125	445 ± 293 [3,c]	9.67
124	26 ± 28	19 ± 53	25 ± 30	58 ± 125	3.05
126	22 ± 24	30 ± 47	36 ± 26	255 ± 227[3,c]	8.50
169	70 ± 57	125 ± 94[2]	48 ± 43	103 ± 305	0.82
170	59 ± 34	45 ± 24	74 ± 45	86 ± 26[c]	2.15
179	6 ± 18	5 ± 16	77 ± 55	179 ± 178[2,c]	35.8

Blood and nasal breath were obtained from 30 healthy controls. The formation rate of the ions of interest (Table 12) were quantified using SIFT-MS in MIM mode in blood headspace, water headspace (for comparison with blood), and breath and ambient air (for comparison with breath). Values shown are mean normalized count rates multiplied by 10000 ± SD. Water and blood headspace, or ambient air and breath, were compared using a t-test with the statistical significance shown (1: $P < .05$, 2: $P < 0.01$, 3: $P < 0.001$). Breath and blood headspace were also compared using a paired t-test (a: $P < 0.05$, b: $P < 0.01$, c: $P < 0.001$). The ratio of mean blood headspace ion abundance to the mean breath abundance is also shown.

Table 14: Quantification of putative cancer marker products formed in the reaction with NO^+ ions in the blood headspace and nasal breath obtained from healthy controls using SIFT-MS.

Product Ions	Water headspace	Blood headspace	Ambient air	Nasal breath	Ratio of breath to blood
29	23588 ± 2847	22847 ± 4509	24888 ± 7307	22230 ± 7592	0.97
57	81 ± 27	83 ± 30	59 ± 18	126 ± 44[1]	1.52
80	144 ± 121	138 ± 127	10 ± 39	22 ± 370	0.17
90	143 ± 81	240 ± 228[1]	504 ± 264	1017 ± 476[3,c]	4.24
102	50 ± 18	78 ± 64	99 ± 83	155 ± 82[1,c]	1.99
116	102 ± 45	118 ± 36	157 ± 134	2176± 1532[3,c]	18.44
195	6 ± 21	14 ± 37	4 ± 20	12 ± 42	0.86

Blood and nasal breath were obtained from 30 healthy controls. The formation rate of the ions of interest (Table 12) were quantified using SIFT-MS in MIM mode in blood headspace, water headspace (for comparison with blood), and breath and ambient air (for comparison with breath). Values shown are mean normalized count rates multiplied by 10000 ± SD. Water and blood headspace, or ambient air and breath, were compared using a t-test with the statistical significance shown (1: $P < .05$, 2: $P < 0.01$; 3: $P < 0.001$). Breath and blood headspace were also compared using a paired t-test (a: $P < 0.05$. b: $P < 0.01$, c: $P < 0.001$). The ratio of mean blood headspace ion abundance to the mean breath abundance is also shown.

Table 15: Correlational analysis of normalised product ion count rates in the reaction with H_3O^+ and NO^+ precursor ions from the SIFT-MS analysis of breath and blood headspace.

Precursor ion	Product Ion	Pearson correlation coefficient	P-value
H_3O^+	45	0.04	0.83
	60	0.28	0.14
	70	0.02	0.92
	81	0.24	0.20
	96	-0.12	0.52
	99	0.19	0.31
	108	0.08	0.67
	124	0.03	0.87
	126	0.04	0.84
	169	-0.12	0.52
	170	**0.53**	**0.00**
	179	0.03	0.87
NO^+	29	**0.66**	**0.00**
	57	0.20	0.30
	80	0.11	0.55
	90	0.20	0.29
	102	-0.03	0.87
	116	**0.38**	**0.04**
	195	0.29	0.11

The normalised count rates of various SIFT-MS product ions generated in the reaction between H_3O^+ or NO^+ ions and blood headspace or nasal air obtained from 30 healthy controls were compared using linear regression. The Pearson correlation coefficients and p-values are shown. Significant (P <0.05) correlations are **bolded.**

6. Discussion

The major findings of my thesis are that SIFT-MS can be used to identity putative biomarkers of cancer including those that may be markers of only one type of cancer, thereby support my first and second hypothesis which are the volatome of blood headspace differ between cancer and control groups, and one or more of the VC changes in abundance are limited to a single type of cancer. Moreover, some of these markers are present in the human breath of healthy controls which suggest the potential of breath testing for cancer diagnostics, in support of my third hypothesis that these VC present in breath as well as blood headspace.

I did, however, find that the correlation between blood headspace and breath was generally poor in terms of product ion formation rates suggesting that VCs in the bloodstream may not exert a large influence over the concentration of the same VC in the airways. This raises the question of whether such breath markers are useful for cancer in general, or should just be aimed at the diagnosis of lung cancer and perhaps other pulmonary diseases.

6.1 SIFT-MS product ions which may be cancer biomarkers.

In this thesis I explored the question of whether the volatome of blood headspace is influenced by the person having cancer when compared to that of healthy controls, and whether one or more of the VC changes in abundance will be limited to a single type of cancer rather than a general marker of the disease. Considering some common breath compounds and a full spectral analysis of product ions formed in the reaction of trace gases and H_3O^+ or NO^+ ions, I found that many ions varied in their rate of formation between the

four participant groups (see Table 12). Although the groups were not sex-matched I found little evidence of an effect of sex upon product ion formation rates and hence collected the results I obtained using combined male and female participants. Although I had age matched my participants I also investigated the effect of age on my results but found little evidence of there being any effect

The large number of ions identified by my analysis may at least partly relate to the number of simultaneous comparisons being high (over both precursor ions there were 380 product ions considered). That is, the number of ions quantified using this type of spectral analysis is not based on a specific hypothesis regarding a particular ion, but rather includes all possible product ions over the range *m/z* 10 – 200. Indeed, applying a 'Bonferroni' correction, which divides the generally accepted statistical significance level of 0.05 amongst all the comparisons, resulted in only *m/z* 99 from the H_3O^+ reaction being considered significantly different between groups (see Tables 8 and 9) (*m/z* 124 from the H_3O^+ reaction differed when only males were included). The Bonferroni correction has been criticised as being overly conservative, and frequently inappropriate,[62] when considering a hypothesis such as that in this thesis, that is one which is not directed at a particular ion, but rather that any ion is a biomarker of cancer, and hence does not rule out that some of the significantly altered ion count rates do indeed differentiate cancer from controls. However the method I employed using full spectrum analysis can be viewed as lacking statistical power and would require much larger sample sizes to result in definitive conclusions.

However, it is worth considering what the potential identity of these ions are, compare them to what others have found, and consider their biological plausibility, that is whether it is likely that a human cell, normal or cancer, would generate such a biomarker. A wide variety of compound classes emerge from the SIFT-MS analysis including alcohols, organic acids, polyamines, aldehydes, sulphides, and esters, identified by comparing the product ions I have identified as possible cancer biomarkers, with the SIFT-MS spectra for known compounds described in the literature (see Table 12). As can be seen from the Table many ion products can originate from multiple compounds due to isobaric or isomeric compounds frequently possessing overlapping mass spectra. Comparing the possible chemicals identified in Table 12 to the expected ions from biomarkers identified in other studies described in the Introduction (Tables 1 and 2) shows some commonality, with hexanal (*m/z* 119 with H_3O^+) having been associated with the presence of cancer in two or more other studies[16,33,35], while dimethyl sulphide (*m/z* 63 with H_3O^+)[16], propyl benzene (*m/z* 121 with H_3O^+)[33], methanol (*m/z* 51 and 69 with H_3O^+)[36] and xylene (*m/z* 107 with H_3O^+)[39] have been identified in at least one other study. It seems unlikely that esters, such as ethylacetate, are the true identity of these ions given that metabolic processes capable of producing are unknown. The significance of alcohol changes is unclear given that longer-chain alcohols like pentanol most likely derive from gut microflora,[63,64] although this may indicate disordered gut function in patients, as would changes in the concentration of propionic acid - another gut derived compound.[65,66] Compounds such as cadaverine and putrescine generally occur in putrefaction but have been reported to possess altered abundance in patients with cancer, possibly relating to necrosis occurring in the tumour or surrounding tissue.[67,68] The alkanes and aldehyde are a common chemical class reported to be a lung cancer biomarker

(Table 12) and are of note since both classes of compounds can derive from lipid peroxidation, which may change subsequent to altered oxidative stress in cancer.[23] Notably other authors have reported altered abundance of this compound class in cancer which lends support to the possibility that these ions are true markers of, at least, symptomatic cancer.[35, 40,69] It should be pointed out, however, that increased oxidative stress is expected in cancer while the data in Tables 7-9 indicate that although propanal concentrations were increased in patients with breast cancer, acetaldehyde, and the production rates of ions which may derive from pentanal and malondialdehyde, were decreased in one or more cancer groups contrary to the hypothesised pathophysiological mechanism. It is presently unclear as to the reason for apparently reduced aldehyde concentrations, although further study of what occurs to the abundance of these chemicals during the progression of cancer may shed some light on this discrepancy. Indeed, this study, and others like it, are only conducted using patients whose disease are advanced enough to have received a clinical diagnosis of cancer raises the problem, discussed in the Introduction, that the altered volatome may be due to physiological effects occurring secondary to the primary pathophysiological mechanisms and therefore be of no use for the diagnosis of cancer in its pre-symptomatic stage. Without a prospective study being carried out it is not possible to answer this question in any definitive way. However, it is less likely that ions, the abundance of which were changed in only one type of cancer, are due to general changes in illness associated with cancer such as cachexia, although this does not rule out secondary effects associated with just one type of tumour. Although not conclusive of being associated with a primary pathophysiological process I decided to pursue further experiments only with those ions which were identified as putative markers of a singly type of cancer.

6.2 Putative cancer markers in breath and blood

Although blood based biomarkers are of interest, mass screening using them would require each persons to donate a sample of blood. Even if using 'pin prick' types of collection this is unlikely to be used by a significant number of people in good health due to the costs, pain and infection risk associated with blood collection. Breath analysis overcomes these difficulties and, as such, I investigated whether the ions I identified in my first study with blood headspace could also be produced in during reactions between breath and H_3O^+ or NO^+.

For my initial investigation I utilised healthy controls, and therefore my study was unable to detect the breath production of trace-gas derived ions, which *only* occur in cancer. However, since my first study frequently observed decreased ion production rates in cancer, this possibility is less likely for many of the product ions under investigation. My first question was whether the formation rate of product ions deriving from breath could be differentiated statistically from ambient air samples. For many of the putative marker ions this was indeed the case which indicates that these ions are likely to be at least partly derived from endogenously produced compounds. Of course a failure to differentiate breath from ambient does not mean there are no endogenous sources of these ions, merely that ambient air also contains higher quantities of these chemicals making them unsuitable for use as breath based biomarkers. For comparison I also tried to differentiate blood headspace from water headspace and found that not all ion count rates differed significantly from this control sample. This is surprising given that these ions were identified using plasma samples in study 1. A comparison to a water headspace was not carried out for the first study, however, and so it may be that there was no above background formation of these ions in study 1

either. However this seems unlikely given that differences were observed between groups of participants, the very reason these ions were chosen for further study. Study 1 and study 2 differed slightly in that whole blood was used in the latter and plasma in the former, although it is not immediately obvious why this would explain any differences between each. Alternatively, samples in study 2 were freshly collected and analysed immediately, while study 1 plasma samples were frozen and stored at -80°C for over a year before analysis. Conceivably a storage effect could also play a role given that others have observed degradation of samples stored at anything above liquid nitrogen temperatures.[71,72,73] Nevertheless, it remains notable that many of the putative markers identified using blood headspace are also present in human breath and could therefore be used as breath-based cancer markers.

The use of breath analysis to detect cancers occurring out within the pulmonary system also necessarily requires that breath concentrations of the marker are directly correlated with the bloodstream given that the circulatory system is the means for the VC marker to get into the breath. A significant correlation was observed for the NO^+ reaction product 116, potentially being pentanone, and the H_3O^+ reaction product *m/z* 170 (which matches no known ion product in the literature), indicating that concentrations in blood and breath are likely related. The other product that was found to be correlated (NO^+ product ion *m/z* 29) could be due to miscounting of the precursor ion *m/z* 30 as 29. That is that the correlation is due to an artefact related of the mass resolution of the mass spectrometer given that *m/z* 30 is being counted as both *m/z* 30 and 29. This conclusion is supported by the fact that *m/z* 29 was not formed at greater rates that that of the control sample in neither blood headspace or nasal air. Aside from *m/z* 116, however, the formation rate of the other

putative cancer marker product ions were not significantly correlated between blood and breath. This is a surprising finding considering that these samples were collected and analysed at the same time from the same participants (see Table 15). No such correlation would be expected, of course, for ions, which are not present above control levels (water headspace or ambient air) for either breath and/or blood (e.g. the H_3O^+ reaction product *m/z* 108 or the NO^+ product *m/z* 102). In addition reaction products found in very low concentrations (e.g. the H_3O^+ reaction product *m/z* 57) may not be quantified with sufficient precision to result in a correlation being observed with the sample size used (n=30). That being said, the lack of correlation for some higher abundance ions, such as the H_3O^+ reaction product *m/z* 60 or the NO^+ reaction product *m/z* 90, is unlikely to be due to these measurement effects. It is possible that blood concentrations actually do not predict breath concentrations implying that most of these product ions are derived from gases originating somewhere in the airways and not from the circulation. Such a possibility is consistent with the large range in relative abundance between breath and blood fractions for each ion (see Table 14 and 15). These ratios ranged from below unity (blood > breath) to well above (breath > blood) again suggesting that no strong relationship exists between blood and breath fractions in terms of trace gases. In other words, that breath trace gases are much more dependent upon gases originating in the airways than those elsewhere in the body. While for some compounds, such as acetone or isoprene, a bloodstream source may be likely although certainly not proven, my data is reminiscent of the finding that many compounds found in mouth-collected breath mainly originate in the oral cavity.[22,23] Given that I used nasal rather than mouth breath, a fraction which minimises the contribution of the oral cavity, my data are unlikely to have the same explanation however this does rule out the

trace gases in nasal air mainly originate elsewhere in the airway. While measurement sensitivity and precision problems described above certainly leave such a conclusion open to challenge, my data do suggest that (a) identifying putative breath-based non-airway cancer markers using blood samples may not be a useful strategy, and (b) that using breath analysis to diagnose cancers other than those within the airway will not be possible. On the other hand breath analysis is suitable for diagnosis of airway conditions as has already been shown for airway inflammation (ethane and NO^+) and infection (hydrogen cyanide).[20,30]

6.3 Conclusion and Perspectives

In conclusion my data suggest that the use of SIFT-MS to identify putative disease biomarkers present in the blood stream has some value although the full mass spectrum approach I used is likely to generate many false positives and would require the use of larger sample size. My data also suggest that screening blood samples to identify markers that will ultimately be applied to breath diagnostic testing is of limited value given that the relationship between bloodstream volatile chemicals and those in the breath is, at best, weak. Future work using breath analysis to identify disease biomarkers and apply those to the screening and diagnosis of illness, including those for lung cancer, should therefore focus on airway disease and be confined to the utilization of actual breath samples.

References

1. Cancer Facts & Figures 2015 - acspc-044552.pdf.
2. Cancer statistics at a glance - Canadian Cancer Society. *www.cancer.ca* Available at: http://www.cancer.ca/en/cancer-information/cancer-101/cancer-statistics-at-a-glance/?region=on. (Accessed: 20th September 2016)
3. Toh, S. *et al.* Selective serotonin reuptake inhibitor use and risk of gestational hypertension. *Am. J. Psychiatry* **166,** 320–328 (2009).
4. Manton, K. G., Akushevich, I. & Kravchenko, J. in *Cancer Mortality and Morbidity Patterns in the U.S. Population* 247–311 (Springer New York, 2009). doi:10.1007/978-0-387-78193-8_7
5. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61,** 69–90 (2011).
6. Minna, J. D., Roth, J. A. & Gazdar, A. F. Focus on lung cancer. *Cancer Cell* **1,** 49–52 (2002).
7. Beckles, M. A., Spiro, S. G., Colice, G. L. & Rudd, R. M. Initial evaluation of the patient with lung cancer*: Symptoms, signs, laboratory tests, and paraneoplastic syndromes. *Chest* **123,** 97S–104S (2003).
8. Hirsch, F. R., Franklin, W. A., Gazdar, A. F. & Bunn, P. A. Early Detection of Lung Cancer: Clinical Perspectives of Recent Advances in Biology and Radiology. *Clin. Cancer Res.* **7,** 5–22 (2001).
9. Safety, feasibility and clinical benefit of localized chemotherapy using microencapsulated cells for inoperable pancreatic carc. Available at: http://www.cancer-therapy.org/CT/v6/B/HTML/51._Kalogerakos_et_al,_463-476.html. (Accessed: 1st February 2016)

10. Cutler, D. M. Are We Finally Winning the War on Cancer? *J. Econ. Perspect.* **22,** 3–26 (2008).
11. Corner, J., Hopkinson, J., Fitzsimmons, D., Barclay, S. & Muers, M. Is late diagnosis of lung cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis. *Thorax* **60,** 314–319 (2005).
12. Spiro, S. G., Gould, M. K. & Colice, G. L. Initial evaluation of the patient with lung cancer: Symptoms, signs, laboratory tests, and paraneoplastic syndromes*: accp evidenced-based clinical practice guidelines (2nd edition). *Chest* **132,** 149S–160S (2007).
13. Hamilton, W., Peters, T. J., Round, A. & Sharp, D. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* **60,** 1059–1065 (2005).
14. Athey, V. L., Suckling, R. J., Tod, A. M., Walters, S. J. & Rogers, T. K. Early diagnosis of lung cancer: evaluation of a community-based social marketing intervention. *Thorax* **67,** 412–417 (2012).
15. Peeling, R. W., Smith, P. G. & Bossuyt, P. M. M. A guide for diagnostic evaluations. *Nat. Rev. Microbiol.* **8,** S2–S6 (2008).
16. Chen, X. *et al.* A study of the volatile organic compounds exhaled by lung cancer cells in vitro for breath diagnosis. *Cancer* **110,** 835–844 (2007).
17. Auerbach, O. *et al.* Changes in the Bronchial Epithelium in Relation to Smoking and Cancer of the Lung. *N. Engl. J. Med.* **256,** 97–104 (1957).
18. Fung, E. T., Wright, G. L. & Dalmasso, E. A. Proteomic strategies for biomarker identification: progress and challenges. *Curr. Opin. Mol. Ther.* **2,** 643–650 (2000).
19. ASEE-GSW_2015_submission_75.pdf.

20. Cao, W. & Duan, Y. Breath Analysis: Potential for Clinical Diagnosis and Exposure Assessment. *Clin. Chem.* **52,** 800–811 (2006).
21. Gerber, D. E. An Emerging Lung Cancer Biomarker. *Sci. Transl. Med.* **6,** 230ec58-230ec58 (2014).
22. Horvath, G. *et al.* Mice deficient in pituitary adenylate cyclase activating polypeptide display increased sensitivity to renal oxidative stress in vitro. *Neurosci. Lett.* **469,** 70–74 (2010).
23. Miekisch, W., Schubert, J. K. & Noeldge-Schomburg, G. F. E. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347,** 25–39 (2004).
24. Hakim, M. *et al.* Volatile Organic Compounds of Lung Cancer and Possible Biochemical Pathways. *Chem. Rev.* **112,** 5949–5966 (2012).
25. Narasimhan, L. R., Goodman, W. & Patel, C. K. N. Correlation of breath ammonia with blood urea nitrogen and creatinine during hemodialysis. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4617–4621 (2001).
26. Teshima, N., Li, J., Toda, K. & Dasgupta, P. K. Determination of acetone in breath. *Anal. Chim. Acta* **535,** 189–199 (2005).
27. Miekisch, W. & Schubert, J. From highly sophisticated analytical techniques to life-saving diagnostics: Technical developments in breath analysis. *TrAC Trends Anal. Chem.* **25,** 665–673 (2006).
28. Buszewski, B., Kesy, M., Ligor, T. & Amann, A. Human exhaled air analytics: biomarkers of diseases. *Biomed. Chromatogr.* **21,** 553–566 (2007).

29. Schmidt, F. M. *et al.* Ammonia in breath and emitted from skin. *J. Breath Res.* **7,** 017109 (2013).

30. Enderby, B., Smith, D., Carroll, W., & Lenney, W. Hydrogen cyanide as a biomarker for Pseudomonas aeruginosa in the breath of children with cystic fibrosis. Pediatric pulmonology, **44**, 142-147 (2009).

31. Rivory, L. P., Slaviero, K. A., Hoskins, J. M. & Clarke, S. J. The Erythromycin Breath Test For the Prediction of Drug Clearance. *Clin. Pharmacokinet.* **40,** 151–158 (2012).

32. Rhodes, J. M., Middleton, P. & Jewell, D. P. The lactulose hydrogen breath test as a diagnostic test for small-bowel bacterial overgrowth. *Scand. J. Gastroenterol.* **14,** 333–336 (1979).

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

34. Phillips, M. *et al.* DEtection of lung cancer with volatile markers in the breath*. *Chest* **123,** 2115–2123 (2003).

35. Phillips, M. *et al.* Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark. Sect. Dis. Markers* **3,** 95–109 (2007).

36. Phillips, M. *et al.* Detection of lung cancer using weighted digital analysis of breath biomarkers. *Clin. Chim. Acta* **393,** 76–84 (2008).

37. D'Amico, A. *et al.* An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer* **68,** 170–176 (2010).

38. 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* **878,** 2643–2651 (2010).

39. Machado, R. F. *et al.* Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath. *Am. J. Respir. Crit. Care Med.* **171,** 1286–1291 (2005).
40. 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.* **6,** 71 (2005).
41. Wehinger, A. *et al.* Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas. *Int. J. Mass Spectrom.* **265,** 49–59 (2007).
42. Spaněl, P. & Smith, D. Progress in SIFT-MS: breath analysis and other applications. *Mass Spectrom. Rev.* **30,** 236–267 (2011).
43. Španěl, P. & Smith, D. *Selected ion flow tube studies of the reactions of H3O+, NO+, and O+2 with several aromatic and aliphatic hydrocarbons*. **181,** (1998).
44. Francis, G. J., Wilson, P. F., Milligan, D. B., Langford, V. S. & McEwan, M. J. GeoVOC: A SIFT-MS method for the analysis of small linear hydrocarbons of relevance to oil exploration. *Int. J. Mass Spectrom.* **1,** 38–46 (2007).
45. Diskin, A. M., Wang, T., Smith, D. & Španěl, P. A selected ion flow tube (SIFT), study of the reactions of H3O+, NO+ and O2+ ions with a series of alkenes; in support of SIFT-MS. *Int. J. Mass Spectrom.* **218,** 87–101 (2002).
46. Spanel, P. & Smith, D. SIFT studies of the reactions of H3O+, NO+ and O2+ with a series of alcohols. *Int. J. Mass Spectrom. Ion Process.* **167,** 375–388 (1997).
47. Španěl, P., Ji, Y. & Smith, D. SIFT studies of the reactions of H3O+, NO+ and O2+ with a series of aldehydes and ketones. *Int. J. Mass Spectrom. Ion Process.* **165,** 25–37 (1997).
48. Jordan, A. *et al.* An online ultra-high sensitivity Proton-transfer-reaction mass-spectrometer combined with switchable reagent ion capability (PTR + SRI − MS). *Int. J. Mass Spectrom.* **286,** 32–38 (2009).

49. Selected Ion Flow Tube Mass Spectrometry. Available at: http://www.sift-ms.com/. (Accessed: 22nd April 2016)
50. Spaněl, P., Rolfe, P., Rajan, B. & Smith, D. The selected ion flow tube (SIFT)--a novel technique for biological monitoring. *Ann. Occup. Hyg.* **40,** 615–626 (1996).
51. Španěl, P. & Smith, D. Quantification of trace levels of the potential cancer biomarkers formaldehyde, acetaldehyde and propanol in breath by SIFT-MS. *J. Breath Res.* **2,** 046003 (2008).
52. Španěl, P. & Smith, D. *Selected ion flow tube studies of the reactions of H3O+, NO+ and O2+ with some organosulphur molecules*. **176,** (1998).
53. Shestivska, V. *et al.* Direct detection and quantification of malondialdehyde vapour in humid air using selected ion flow tube mass spectrometry supported by gas chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom. RCM* **29,** 1069–1079 (2015).
54. Ŝpaněl, P. & Smith, D. SIFT studies of the reactions of H3O+, NO+ and O+2 with a series of volatile carboxylic acids and esters. *Int. J. Mass Spectrom. Ion Process.* **172,** 137–147 (1998).
55. Smith, D. *et al.* Isoprene levels in the exhaled breath of 200 healthy pupils within the age range 7-18 years studied using SIFT-MS. *J. Breath Res.* **4,** 017101 (2010).
56. Bauer, C. Exploiting Proteomics Data. (Freie Universität Berlin, Freie Universität Berlin, Germany, 2014).
57. Ross, B. M., Babay, S. & Ladouceur, C. The use of selected ion flow tube mass spectrometry to detect and quantify polyamines in headspace gas and oral air. *Rapid Commun. Mass Spectrom. RCM* **23,** 3973–3982 (2009).

58. Iachetta, L., Malek, L. & Ross, B. M. The reactions of H3O+, NO+ and O 2+ with several flavourant esters studied using selected ion flow tube mass spectrometry. *Rapid Commun. Mass Spectrom.* **24,** 815–822 (2010).
59. Wang, T., Spanel, P. & Smith, D. A selected ion flow tube study of the reactions of H3O+, NO+ and O2+- with some phenols, phenyl alcohols and cyclic carbonyl compounds in support of SIFT-MS and PTR-MS. *Int. J. Mass Spectrom.* **239,** 139–146 (2004).
60. Wang, T., Smith, D. & Španěl, P. Selected ion flow tube, SIFT, studies of the reactions of H3O+, NO+ and O2+ with compounds released by Pseudomonas and related bacteria. *Int. J. Mass Spectrom.* **233,** 245–251 (2004).
61. Wang, T., Španěl, P. & Smith, D. Selected ion flow tube mass spectrometry of 3-hydroxybutyric acid, acetone and other ketones in the headspace of aqueous solution and urine. *Int. J. Mass Spectrom.* **272,** 78–85 (2008).
62. Perneger, T. V. What's wrong with Bonferroni adjustments. *BMJ* **316,** 1236–1238 (1998).
63. Francavilla, R. *et al.* Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatr. Allergy Immunol.* **23,** 420–427 (2012).
64. Taneyo Saa, D. *et al.* Impact of Kamut® Khorasan on gut microbiota and metabolome in healthy volunteers. *Food Res. Int.* **63, Part B,** 227–232 (2014).
65. Braga, M. Gut Function and Immune and Inflammatory Responses in Patients Perioperatively Fed With Supplemented Enteral Formulas. *Arch. Surg.* **131,** 1257 (1996).

66. Al-Lahham, S. H., Peppelenbosch, M. P., Roelofsen, H., Vonk, R. J. & Venema, K. Biological effects of propionic acid in humans; metabolism, potential applications and underlying mechanisms. *Biochim. Biophys. Acta* **1801,** 1175–1183 (2010).
67. Löser, C., Fölsch, U. R., Paprotny, C. & Creutzfeldt, W. Polyamines in colorectal cancer. Evaluation of polyamine concentrations in the colon tissue, serum, and urine of 50 patients with colorectal cancer. *Cancer* **65,** 958–966 (1990).
68. Russell, D. H., Levy, C. C., Schimpff, S. C. & Hawk, I. A. Urinary Polyamines in Cancer Patients. *Cancer Res.* **31,** 1555–1558 (1971).
69. Fuchs, P., Loeseken, C., Schubert, J. K. & Miekisch, W. Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer* **126,** 2663–2670 (2010).
70. Queralto, N. *et al.* Detecting cancer by breath volatile organic compound analysis: a review of array-based sensors. *J. Breath Res.* **8,** 027112 (2014).
71. Sheng, Z. Z. & Hermann, A. M. Superconductivity in the rare-earth-free Tl Ba Cu O system above liquid-nitrogen temperature. *Nature* **332,** 55–58 (1988).
72. Malinin, T. I., Pegg, D. E., Perry, V. P. & Brodine, C. E. Long-term storage of bone marrow cells at liquid nitrogen and dry ice temperatures. *Cryobiology* **7,** 65–69 (1970).
73. Fox, H. *et al.* Degradation of specific polyunsaturated fatty acids in red blood cells stored at −20°C proceeds faster in patients with schizophrenia when compared with healthy controls. *Prostaglandins Leukot. Essent. Fatty Acids* **69,** 291–297 (2003).