



Swansea University
Prifysgol Abertawe



Swansea University E-Theses

Biological tools in the diagnosis of lung cancer.

Ghosal, Robin

How to cite:

Ghosal, Robin (2010) *Biological tools in the diagnosis of lung cancer..* thesis, Swansea University.
<http://cronfa.swan.ac.uk/Record/cronfa43007>

Use policy:

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>



Swansea University
Prifysgol Abertawe

**Biological Tools in the Diagnosis of
Lung Cancer**

Robin Ghosal

MBBCH (Wales), MRCP (UK)

**Submitted to the University of Wales in fulfilment of
the requirements for the Degree of Doctor of Medicine**

2010

ProQuest Number: 10821397

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10821397

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346



Summary

Lung cancer is the most common cancer worldwide with 1.3 million new cases diagnosed each year. The 5-year survival rate is much lower than other common cancers such as breast and prostate cancer. Several large-scale screening programmes using existing technologies over the past 40 years have not yet reduced mortality rates from lung cancer. We have studied new technologies on sputum and exhaled breath to assess their potential for diagnosis. Reliable, non-invasive and cheap diagnostic tests are the cornerstone for any future screening programme. The first study tests the sputa of patients with suspected lung cancer and healthy controls with Fourier Transform Infra-Red (FTIR) spectroscopy. We developed a predictive model based on two wavenumbers, to differentiate those with proven lung cancer versus healthy controls with a sensitivity of 93% and specificity of 91%. When we included the sputa of patients having tests for lung cancer initially but with no evidence of cancer after one year ("high-risk" group), this only partially reduced the model's predictive ability. The second study assessed the sputa from the same cohorts with a panel of gene antibodies (p16, p53, p63, EGFR and cyclin D1). Results were not discriminatory with low sensitivity (8-42%), suggesting immunohistochemistry on sputa cells will not be a useful diagnostic tool. Our final study assessed exhaled volatile organic compounds (VOCs) in the breath of newly diagnosed lung cancer patients and in healthy controls using gas chromatography – mass spectrometry. 29 cancer-exclusive VOCs were identified and 25 further VOCs were universally higher in the cancer cohort, allowing correct classification of 89% of cancer patients. We conclude that two of the three novel techniques (sputum FTIR and exhaled VOCs) could successfully distinguish cancer from healthy control subjects and show potential as screening modalities in further larger scale studies.

Declaration and statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ... (candidate)

Date 21/6/10

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed (candidate)

Date 21/6/10

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date 21/6/10

Table of Contents	Page
Acknowledgements	6
Index of figures	11
Index of tables	14
Abbreviations	16
Publications and presentations	18
Part 1: Introduction	
Chapter	
1. An overview of lung cancer	22
2. Screening in lung cancer	49
3. The effect of smoking in lung cancer	71
4. Fourier Transform Infra-Red (FTIR) spectroscopy and medicine	79
5. Volatile Organic Compounds (VOCs) in lung disease	88
Part 2: Methods	
Chapter	
6. Biological screening for lung cancer with sputum	108
7. Biological screening for lung cancer with exhaled breath	125
Part 3: Results	
Chapter	
8. Fourier Transform Infra-Red and Immunohistochemical analysis of sputum in lung cancer	143

9.	Analysis of Volatile Organic Compounds in exhaled breath in lung cancer	186
----	--	-----

Part 4: Discussion

Chapter

10.	FTIR and Immunohistochemistry assessment of sputum – discussion	205
11.	Exhaled breath in lung cancer – discussion	231

Appendix		253
-----------------	--	-----

Bibliography		278
---------------------	--	-----

Acknowledgements:

Dr Keir Lewis, Senior Lecturer, School of Medicine, University of Swansea and Consultant Respiratory Physician, Hywel Dda Health Board, Prince Philip Hospital, Llanelli, who is my main supervisor. He guided me throughout the research, reviewing each chapter of this thesis, helping me understand statistical analysis both in terms of statistical tests and software packages, providing guidance and support whilst writing Abstracts and Papers as well as contributing to the design and methodology of the research. He was a co-applicant on Grant Applications and Local Ethics Committee Applications. He also contributed from his research funds.

Professor Tim Claypole, Welsh Centre for Printing and Coating, Swansea University, who is my second supervisor. He has provided support and guidance especially on the volatile organic compounds component of the research studies.

Dr Paul Lewis, Lecturer in Bioinformatics, Institute of Life Sciences, Swansea University, who had the original idea to look at biological techniques such as panels of biomarkers and Fourier Transform Infra-Red (FTIR) in sputum for the early detection of lung cancer. Based on this original concept, the study design then evolved and continues to do so. He helped analyse and interpret the results of the immunohistochemistry and even more so with FTIR where he helped me understand and guide me through the principals of multivariate analysis and the R statistical package. He reviewed Abstracts and Papers, and made many helpful comments on Chapters 6, 8 and 10.

Dr Chris Philips, Senior Research Assistant, Welsh Centre for Printing and Coating, Swansea University. Chris has worked side by side with me throughout the exhaled breath component of the research. Whilst I recruited the patients and collected the breath samples, he processed them and generated the data that the statistical model was derived from. He also helped with the interpretation of the VOC results.

Mrs Christine Davies, Senior Technician in Pathology, Singleton Hospital, Swansea, helped with the immunohistochemistry that was performed on sputum. She helped prepare the slides and stain them. She also taught me about immunohistochemistry scoring and was the co-scorer on the sputum slides.

Dr Jonathan Mullins, Lecturer in Bioinformatics, Swansea University and Dr Roisin Mullins, Lecturer in Computer Science and Informatics at Lampeter University who used complex statistical analysis based on the data generated from the analysis of the volatile organic compounds to develop methods of discriminating cases as described in Chapters 7, 9 and 11 and helped guide me through this.

Dr Philip Kloer, Consultant Respiratory Physician, Hywel Dda Health Board, and lung cancer lead for Carmarthenshire, who helped recruit patients from his clinics and bronchoscopy lists, and verified important data such as lung cancer staging and final diagnosis. He also provided guidance during the writing of relevant Abstracts and Papers and supported my research by releasing me from many clinical duties.

Dr Masood Yousef, Lecturer in Biological Sciences, originally approached our research group with the idea of assessing exhaled breath in lung cancer. It was from this idea that we developed the protocol and techniques used.

Mrs Lynda Hopkins, Deputy Service Manger, Immunohistochemistry, Singleton Hospital, Swansea, who helped with teaching me techniques in immunohistochemistry and also helped perform the immunohistochemical staining.

Dr Hemal Shukla, FP1 in Medicine, Prince Philip Hospital, Llanelli, who helped with the collection of exhaled breath samples and carbon monoxide levels from patients who were recruited into the study.

Sion Bayliss, PhD student, University of Swansea, who helped in recruiting the healthy control sputa that were analysed in the FTIR study.

Miss Claire Guy and Mrs Lois Davies, Secretaries at Prince Philip Hospital, Hywel Dda Health Board for their continued help with secretarial components and also in collection of various patients details often at very short notice. Without them things would have been significantly more difficult.

Professor Michael Phillips, Clinical Professor of Clinical Medicine, New York Medical College, who helped advise about some of the methodology he used in his studies and gave guidance on how to proceed from this point in time.

I would also like to acknowledge Dr Diane Parry, Dr Rohan Mehta, Princess of Wales Hospital and their secretary Lynne Roberts; Dr Kim Harrison and Dr Emrys Evans, Morriston Hospital, Abertawe Bro Morgannwg University Health Board; Dr Carol Llewellyn-Jones and Dr Gareth Collier, Hywel Dda Health Board for their collaboration and help with recruitment from the various hospitals through the studies.

Finally and most importantly I would like to thank my wife, Angharad who helped, supported and advised me throughout the research period and my daughter, Gwenan who was born in the last 5 months of it, both of whom I wish to dedicate this work to.

My contribution to this Thesis started in 2006 when Dr Paul Lewis had the idea of using novel approaches to try and detect panels of early biomarkers in patients who were at greatest risk of developing lung cancer. In collaboration with Dr Keir Lewis and myself we gradually designed the methodology of the study and defined what the study population would be and made successful funding applications to the Welsh Office of Research and Development (WORD) for £111,000 (Application no. H07-3-31) and Trust Research and Development Approvals and Local Research and Ethics Committee (LREC) applications and amendments.

Recruitment started in late 2006. I attended lung cancer clinics and obtained informed consent before recruiting all study participants undergoing bronchoscopy for potential lung cancer. I obtained nearly all research specimens and Dr Kloer and Dr K Lewis took any in my absence. The final diagnosis and staging of lung cancer was made at the Trust lung cancer multi-disciplinary team meetings attended by Consultant Radiologists, Pathologists, Chest Physicians and Oncologists which I attended whenever possible.

Processing and slide preparation for the sputum samples in preparation for immunohistochemistry and FTIR, as described in Chapter 6, was performed at the University of Swansea, initially by members of Dr P Lewis's research team and then by myself. The immunohistochemistry preparation was performed by myself alongside and under the supervision of Christine Davies and members of her pathology team. Interpretation of the immunohistochemistry was by myself under the supervision of Christine Davies who independently scored some of the slides.

In 2007, Dr Masood Yousef approached our research group with a developing technology, using exhaled breath to assess volatile organic compounds (VOCs) within the breath in the early detection of lung cancer. From this meeting we incorporated the technique into our study, also looking at disease progression, and the effect of continued smoking on lung cancer survival.

Follow-up of the VOC study cohort with collection of the appropriate samples was carried out by a junior doctor (Dr Shukla). She performed home visits under my supervision to obtain VOCs and exhaled CO levels on 15 patients. All other follow ups were carried out by myself.

The VOCs were processed and analysed by Dr Yousef initially and then by Dr Chris Phillips as the VOCs study branched off and the methodology elaborated upon in the Welsh Centre of Printing and Coating at Swansea University. They taught me further analytical techniques during 2008 and 2009.

Statistical techniques using more common tests such as Mann-Whitney U Test and non-paired Student t-test were taught to me by Dr Keir Lewis and the analysis was then performed by myself. More complex statistical analysis for sputum biomarkers requiring multivariate analysis was supervised by Dr Paul Lewis who coached me with regards the techniques used and the R statistical package for which he developed codes. Using this we developed the models described in Chapter 8.

Statistical analysis for VOCs studies were performed by Dr Jonathan Mullins, Lecturer in Bioinformatics, School of Medicine, Swansea University with support from Dr Roisin Mullins, Lecturer in Computer Science and Informatics at Lampeter University.

All Literature reviews, Conclusions and Discussions are my own personal opinions with the support of my supervisors and co-investigators.

Index of Figures	Page
1.1 Relative survival for lung cancer, England and Wales, 1971 – 2001	24
1.2 The 20 most commonly diagnosed cancers in the UK 2005, excluding non-melanoma skin cancer	26
1.3 TNM staging in NSCLC	30
1.4 Left upper lobe rounded opacity consistent with lung cancer	33
1.5 CT Thorax of patient in Figure 1.4. The distance and optimal direction for a planned CT-guided biopsy is marked	33
1.6 Erosion of the right main bronchus by a tumour seen by Bronchoscopy	35
1.7 ECOG/WHO Performance Status	40
1.8 Proposed mechanism of how EGFR may cause or support the inappropriate cell growth that leads to proliferation, migration, and survival of cancer cells	45
1.9 A typical diagnostic pathway for the diagnosis and management of lung cancer	48
3.1 The different sites of cancer caused by cigarette smoking	72
3.2 Overall survival according to smoking status. The differences between current smokers versus non-smokers, former smokers and recent quitters	74
4.1 An FTIR machine	80
4.2 The FTIR process	82
5.1 Lipid peroxidation	93
5.2 Schematic of potential mechanism of VOCs in lung cancer	96
5.3 Potential sources of VOCs in alveolar air	98
5.4 Potential sources of VOCs in alveolar air of lung cancer patients with possible mechanisms of production and how the illness may affect them	99
5.5 ‘Smellprint’ derived from eNOSE from exhaled VOCs from a lung cancer patient	101

6.1	Patient flow chart for biological screening for lung cancer with sputum	110
6.2	Screening questionnaire for the healthy control cohort for biological screening for lung cancer with sputum	114
6.3	IHC scoring system for both nuclear and cell membrane antigens for biological screening for lung cancer with sputum	119
7.1	Patient flow chart for biological screening for lung cancer with exhaled breath	127
7.2	Screening questions to establish 'healthy' status of controls for biological screening for lung cancer with sputum exhaled breath	129
7.3	Bio-VOC [®] breath sampler	131
7.4	Pushing the trapped alveolar air with the plunger into the 2-bedded thermal desorption tube	133
7.5	The 2-bedded thermal desorption tubes	133
7.6	Screwing the brass caps onto the sorbent tubes	133
7.7	Schematic of components of GC-MS	136
7.8	Actual GC-MS used for sample processing	136
7.9	Schematic of simple mass spectrometer	138
8.1	PCA of example set of triplicate FTIR spectra. Triplicates of 50 samples are shown.	154
8.2	FTIR spectra of cancer and healthy controls within the wavenumber range 599.75 and 3997.7 cm ⁻¹ .	156
8.3	Median absorbance values for the wavenumber for each of the 3 groups	157
8.4	The 2 nd derivative spectra for the cancer, high-risk and healthy controls	158
8.5	PCA of cancer, healthy control and high-risk spectra. The image shows a 2D scatterplot of the first 2 principals	160
8.6	The PCA scatterplot represented in Figure 8.5 but the cases are labelled with their study numbers for identification	161
8.7	1 st and 3 rd principal components showing separation of controls from the cancer and high-risk	162

8.8	HCA of the 3 cohorts, cancer, high-risk, controls. The controls are seen to group together, as do many of the cancer cases with a few of the high-risk also mixed in	164
8.9	Representation of the 126 significant wavenumbers in order of significance	167
8.10	Discrimination line that maximally discriminates cancer and healthy controls	169
8.11	The leave-one-out cross validation of sensitivity	172
8.12	Control sample of EGFR (cell-membrane) staining	178
8.13	Actual patient staining with EGFR	178
8.14	Control sample immunostained with p63 (nuclear staining)	179
8.15	Actual patient sample stained with p63	179
8.16	Hierarchical cluster analysis of the sputa based on IHC scores	184
9.1	Polynomial equations applied to derive two curves which best discriminates lung cancer and control cases	193
9.2	29 VOCs which are exclusive to the lung cancer cohort and were able to correctly classify 31 of lung cancer patients	195
9.3	25 VOCs which were universally higher in lung cancer patients and allowed 8 further subjects to be correctly identified	197
9.4	12 VOCs which are exclusive to healthy control subjects and allow correct identification of 21 as controls	200
9.5	15 VOCs which were universally higher in the healthy controls compare to the lung cancer subjects and allowed 7 subjects to be correctly classified as controls	202
10.1	Comparison of FTIR of normal and malignant exfoliated cervical cells. Malignant cells have much less absorption in the glycogen-rich area, 1025 cm^{-1} and 1047 cm^{-1} , compared to normal cervical cells	210
11.1	An example of the breath collection apparatus used by Phillips <i>et al</i>	237
11.2	Post-test probability of lung cancer by breath VOC assay	238
11.3	Breath collection and VOC extraction	241

Index of tables	Page
1.1 Subtypes of non-small cell lung cancer in smokers and never smokers	28
8.1 Demographic of the cancer, high-risk and controls cohorts	145
8.2 Breakdown of histological subtypes within the cancer cohort	147
8.3 Details of the cancer cohort	148
8.4 Details of the high-risk cohort	149
8.5 Details of the healthy control cohort	150
8.6 Mean counts for bronchial epithelial cells in groups: cancer, high-risk patients and healthy control	152
8.7 Statistical comparison of groups for bronchial epithelial counts	152
8.8 126 significant wavenumbers after Holm's sequential Bonferroni <i>p</i> value correction	166
8.9 Calculating sensitivity / specificity for cancer and healthy control cases for discriminatory model	171
8.10 Calculating sensitivity / specificity for cancer and non-cancer (high-risk and healthy control) cases for discriminatory model	174
8.11 Demographics of cancer and high-risk groups that underwent sputum IHC analysis	176
8.12 Cohens kappa data test statistically assessing the concordance in IHC scoring between the two independent scorers	180
8.13 Differences between the cancer and high-risk cohort for each antibody using Mann-Whitney U Test	182
8.14 Sensitivity and specificity of antibodies for in sputum of cancer and high-risk cohort (with the uninterpretable sputum slides removed from analysis)	183
9.1 Comparison of baseline demographic data of the cancer and healthy control cohorts	187
9.2 Breakdown of histological subtypes within the cancer cohort	189
9.3 Summary of the VOC profile cancer cohort	190
9.4 Summary of healthy control group	191
9.5 List of 29 VOCs exclusive to lung cancer cohort and the patients in which they occurred	196

9.6	25 VOCs which were universally higher in the lung cancer cohort compared to healthy controls	198
9.7	12 VOCs that were exclusive to the healthy control cohort and the controls in which they were detected	201
9.8	15 VOCs that were universally higher in healthy controls compared to lung cancer subjects	203

Abbreviations

ABC	Avidin Biotin Complex
AMU	Atomic mass units
ANOVA	Analysis of variance
AUC	Area under the curve
CF	Cystic Fibrosis
CO	Carbon monoxide
CO ₂	Carbon dioxide
COPD	Chronic Obstructive Pulmonary Disease
COX	Cyclooxygenase
CT	Computed Tomography
CXR	Chest X-ray
CYP	Cytochrome P450
DDT	Dithiothreitol
EBC	Exhaled breath condensate
EBUS	Endobronchial ultrasound
eCO	Exhaled carbon monoxide
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
ELF	Epithelial lining fluid
eNOSE	Electronic nose
FEV ₁	Forced expiratory volume in 1 second
FISH	Fluorescent in situ hybridization
FNA	Fine needle aspiration
FTIR	Fourier Transform Infra-Red
FVC	Forced vital capacity
GC-MS	Gas chromatography-mass spectrometry
GP	General Practitioner
HCA	Hierarchical Cluster Analysis
Hx	History
IHC	Immunohistochemistry
IMS	Ion Mobility Spectrometry
LDA	Linear discriminant analysis

LDCT	Low-dose spiral CT
LREC	Local Research and Ethics Committee
MDT	Multi-disciplinary team
MGMT	O ⁶ -methylguanine-DNA methyltransferase
MI	Myocardial infarction
NCI	National Cancer Institute
NHS	National Health Service
NICE	National Institute of Clinical Excellence
NIST	National Institute of Standards and Technology
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NNN	N'-nitrosonornicotine
NSCLC	Non-small cell lung cancer
PAH	Polycyclic aromatic hydrocarbons
PBM	Probability based matching
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PIS	Patient Information Sheet
PSA	Prostate specific antigen
QOL	Quality of life
RCT	Randomised controlled trials
ROS	Reactive oxygen species
RT	Retention time
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SVCO	Superior vena cava obstruction
TB	Tuberculosis
TBNA	Transbronchial needle aspiration
TNM	Tumour Node Metastasis
UKCRN	UK Clinical Research Network
VEGF	Vascular endothelial growth factor
VOCs	Volatile organic compounds
WHO	World Health Organisation
WORD	Welsh Office of Research and Development

Publications (to date):

Papers:

Ghosal R; Kloer P; Lewis KE. **A review of novel biological tools used in screening for the early detection of lung cancer.** *Postgrad Med J.* 2009.

85(1005): p. 358–63

Winner of Welsh Thoracic Society publication prize 2010

Lewis PD, Lewis KE, Ghosal R, Bayliss S, Wills J, Godfrey AR, Kloer PJ, Mur L.

Evaluation of FTIR Spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* (in submission)

Book Chapter:

Ghosal R. **Smoking and cancer.** *Smoking Cessation.* Oxford University Press.

(in print)

Abstracts:

Ghosal R; Lewis KE; Kloer PJ; Davies C; Lewis P. **Assessing a panel of five biomarkers in the sputum of patients with and without lung cancer.**

Thorax, Dec 2009; 64 (Suppl IV):A80

Lewis KE, Philips CO, Mullins R, Shukla H, Ghosal R, Kloer PJ, Mullins JG,

Lewis PD, Claypole T. **Differences in profiles of volatile organic compounds in the breath of lung cancer patients versus controls.** *Thorax*,

Dec 2009; 64 (Suppl IV):S119

Ghosal R; Kloer P; Mur L; Lewis KE; Lewis P. **Fourier Transform InfraRed spectroscopy and metabolic profiling in lung cancer.** *Thorax.* Dec 2008;

63(Suppl VII):A60 (spoken presentation)

Ghosal R; Kloer P; Thomas C; Lewis PD; Lewis KEL. **Smoking habits in patients undergoing bronchoscopy for possible lung cancer.** *American Journal of Respiratory and Critical Care Medicine*. Apr 2008;177(Abstract Issue):A905 (poster presentation)

Ghosal R; Kloer P; Thomas C; Lewis PD; Lewis KEL. **Assessment of smoking habits in patients undergoing initial investigation for possible lung cancer.** *Lung Cancer*. Apr 2008;60(S1):S17 (poster presentation)

Ghosal R; Lewis KE; Kloer P; Mehta R; Parry DS; Llewellyn-Jones C; Mur L; Prior S; Blaser J; Lewis PD. **Using Fourier Transform Infra-Red (FTIR) spectroscopy to evaluate metabolic markers in sputum in patients with and without lung cancer.** *Lung Cancer*. Apr 2008;60(Suppl I):S11 (poster presentation)

Ghosal R; Lewis KE; Kloer P; Mehta R; Parry DS; Llewellyn-Jones C; Mur L; Prior S; Lewis PD. **Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer.** *Thorax*. Dec 2007; 62(Suppl III):A10 (spoken presentation)

Non-abstracted Presentations:

Biological tools in diagnosing lung cancer: results of the 2 year pilot studies. Welsh Thoracic Society. March 2010

Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer. British Association of Lung Research, Manchester. July 2009 (Poster presentation)

Volatile Organic Compounds and Lung Cancer. Welsh Thoracic Society. March 2009

Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer. Hywel Dda Research Study Day. Dec 2008.

Winner of best research project 2008

Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer. Welsh Society of Cytology. October 2008.

Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer. Swansea University Research Study Day. June 2008

Fourier Transform Infra-Red (FTIR) measuring metabolic markers in sputum in patients with and without Lung Cancer. Welsh Thoracic Society. September 2007.

Winner of best research presentation 2007

PART 1: INTRODUCTION

Chapter 1. An overview of lung cancer

Introduction

Lung cancer is a disease which occurs when there is uncontrolled cell growth within the lung fields or the bronchial airways. [1] The cancerous growth may either invade major structures such as the heart and mediastinum locally or metastasise to organs outside the thoracic cavity such as liver, adrenal glands, brain or bones. Lung cancer is the most common cause of cancer related death in men and second in women worldwide. [2, 3] The most common cause of lung cancer is tobacco smoking, which contributes to 85-90% of all lung cancers. [4-7] The potential mechanisms of how smoking may cause cancer are discussed in Chapter 3. However, other factors have also been implicated in lung carcinogenesis including genetic factors, radon gas, asbestos exposure and air pollution such as metal air pollutants and passive (second hand) smoking. [8-16] It is thought that the majority of lung cancers result from many complex genetic and environmental factors over time. [17, 18]

Lung cancer is divided histologically into two main groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). It is an important differentiation to make as the treatment modalities will be dependent on the histological subtype and the prognosis also varies depending on the tissue typing. Despite advances in radiological techniques, chemotherapeutic drugs and radiotherapy, the overall 5-year survival rates remain low (10.9% in Europe) [19] and have not changed significantly over the last 30 years. [20] This can be seen in Figure 1.1

One of the main reasons for such poor survival rates is the late presentation of patients to medical practitioners. This is predominantly due to symptoms only becoming apparent in a later, more advanced stage of disease. There is a further delay as smokers who do have symptoms often do not seek help immediately, are misdiagnosed or are partially treated with antibiotics, inhalers and painkillers. Therefore, various methods have been studied to attempt to detect lung cancer at an earlier stage thus allowing potentially curative / radical

therapy. Although the technology is advancing there has been as yet no definitive screening tool that has reduced mortality rates significantly. These studies will be discussed further in Chapter 2.

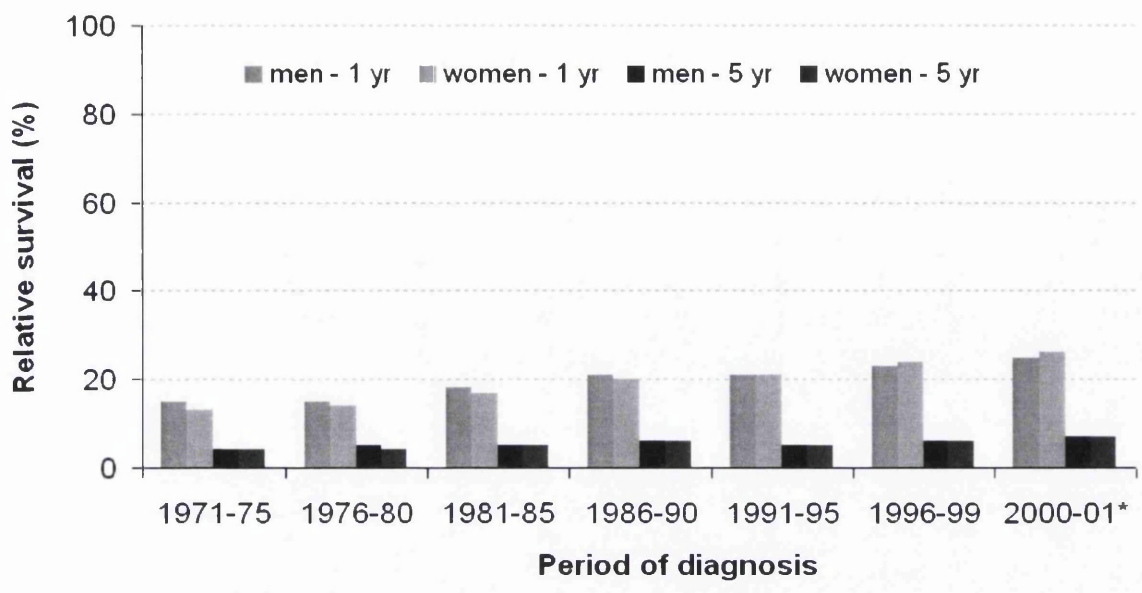


Figure 1.1: Relative survival for lung cancer, England and Wales, 1971 – 2001 [20]

Epidemiology of lung cancer

Worldwide

Lung cancer is the most common cancer worldwide in terms of both incidence and mortality with 1.3 million new cases being diagnosed each year and 1.3 million deaths, with Europe and North America having the highest rates. [21-23] Famous people who have died from lung cancer include Walt Disney, Nat “King” Cole and Yul Brenner. With increasing smoking in developing countries, the incidence is expected to increase in the next few years, especially in China and India. [24, 25]

United Kingdom

In the UK, lung cancer is the second most commonly diagnosed cancer overall with only breast cancer having more cases. There were 33,000 new cases diagnosed in England and Wales alone in 2006. [26] 2146 new cases were diagnosed in Wales in 2005, which is higher than the UK average. [27] In men there are less lung cancer cases diagnosed per year than prostate cancer, however there are more lung cancer deaths. In women, less lung cancer cases are diagnosed per year than breast cancer, but again there are more lung cancer deaths. The incidences of the most common cancers are shown in Figure 1.2. [28-31]

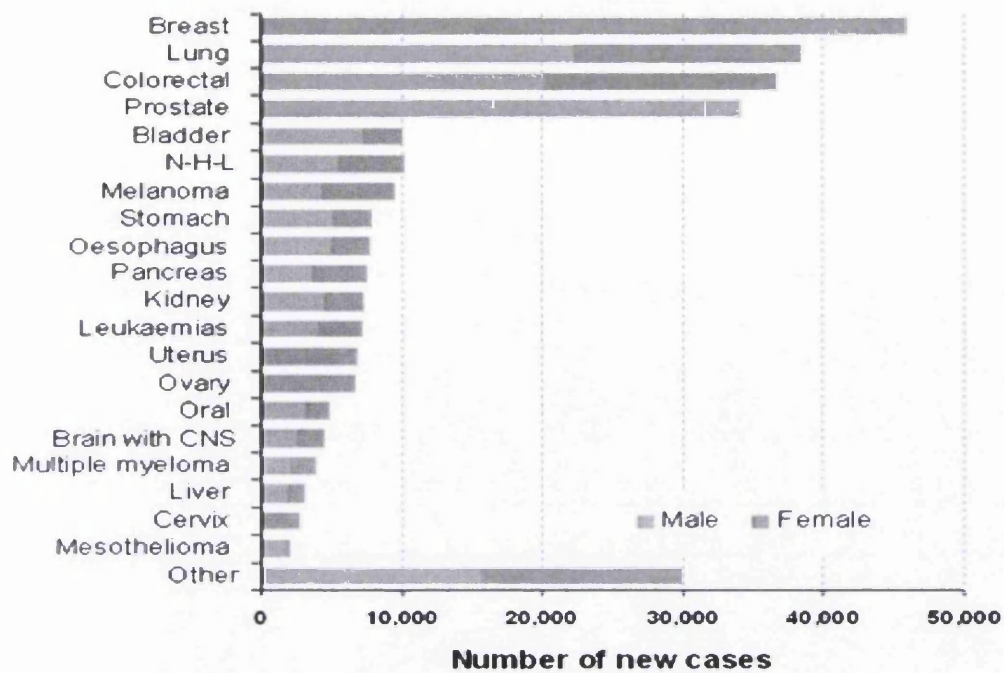


Figure 1.2: The 20 most commonly diagnosed cancers in the UK 2005, excluding non-melanoma skin cancer [28-31]

Histological subtype

In terms of histological subtype, NSCLC (80.4%) occurs more frequently than SCLC (16.8%). The remaining 2.8% is made up of carcinoid, sarcoma and unspecified lung cancer. [32-34] The prevalence of NSCLC subtypes are further delineated in Table 1.1, which is based on a retrospective cohort study of 730 NSCLC patients diagnosed in USA between 1999 and 2005.

Histological sub-type		Frequency of non-small cell lung cancers (%)	
		Smokers	Never-smokers
Squamous cell lung carcinoma (SCC)		42	33
Adenocarcinoma	Adenocarcinoma	39	35
	Bronchioalveolar carcinoma	4	10
Carcinoid		7	16
Other		8	6

Table 1.1: Subtypes of non-small cell lung cancer in smokers and never smokers [35]

Survival in lung cancer

Lung cancer is a devastating disease with a low 5-year survival rate, especially if one compares it with other common and usually more publicised cancers such as prostate or breast cancer. Recent data shows that the 5-year survival rates for lung cancer in Europe are 10.9%, compared with a 5-year survival of 79% for breast cancer and 77.5% for prostate cancer. [19] The overall 5-year survival rate in America for lung cancer is slightly higher than in Europe at 16% although this figure is possibly less accurate of true survival rates and has been criticised because it is calculated from only a quarter of the population and also excludes those without histological confirmation of diagnosis. [36] These latter patients are likely to be more elderly and frail so unable to tolerate radical treatments and are therefore likely to have a shorter survival times.

One of the major reasons for the poor survival data is the identification of lung cancer at an already advanced stage of disease. Compared with breast and prostate cancer there is as yet no suitable screening program. In breast oncology, mammograms have been used for screening since 1976 and for prostate cancer the availability of a simple blood test such as Prostate Specific Antigen (PSA) can help identify those with the disease, importantly often whilst it is still localised. To illustrate the importance of early detection of lung cancer one can reflect on the survival data according to radiological staging. Figure 1.3 shows the Tumour Node Metastasis (TNM) staging for NSCLC. The TNM system for staging cancer was first proposed by Denoix in 1946 and developed for lung cancer by The Union Internationale Contre le Cancer (UICC) and the American Joint Committee for Cancer Staging (AJCC) in 1972 and modified now to the international classification system applied by all teams managing lung cancer. [37-40]

In late 2009, a new TNM staging system for lung cancer was introduced for all those newly diagnosed. [41] We used the TNM staging described in Figure 1.3 as our patients were all recruited before the establishment of the 7th edition of the TNM staging.

TNM STAGING OF LUNG CANCER

Supraclavicular Scalene(ipsi-/contralateral)	Mediastinal		Subcarinal	Hilar		Peribronchial (ipsilateral)	LYMPH NODE (N)	Stage IV M1 (any T, any N)				
	(contralateral)	(ipsilateral)		(contralateral)	(ipsilateral)			Stage III B				
+	/	+	/	+			N3	Stage III A				
-	-	-	+ &/ +	-			N2	Stage II A Stage II B				
-	-	-	-	-	+ &/ +		N1	Stage I A Stage I B Stage II B				
-	-	-	-	-	-	-	N0	Stage I A Stage I B Stage II B				

Stage 0 (Tis, N0, M0)					
T1	T2	T3	T4	PRIMARY TUMOR (T)	
a&b&c	any of a,b,c,d	(a&c)/b/d	(a&c)/d	Criteria	
≤ 3 cm	> 3 cm	any	any	a. Size	
No invasion proximal to the lobar bronchus	Main bronchus (≥ 2 cm distal to the carina)	Main bronchus (< 2 cm distal to the carina)	-	b. Endo-bronchial location	
surrounded by lung or visceral pleura	Visceral pleura	Chest wall **/ diaphragm/ mediastinal pleura/ parietal pericardium	Mediastinum/ trachea/heart/ great vessels/ esophagus/ vertebral body/ carina	c. Local Invasion	
-	Atelectasis/ obstructive pneumonitis that extends to the hilar region but doesn't involve the entire lung	Atelectasis/ obstructive pneumonitis of the entire lung	Malignant pleural/peri-cardial effusion or satellite tumor nodule(s) within the ipsilateral primary-tumor lobe of the lung	d. Other	

METASTASES (M)
M0 : Abscent
M1 : Present
 Separate metastatic tumor nodule(s) in the ipsilateral nonprimary-tumor lobe(s) of the lung also are classified M1

Tis : Carcinoma in situ

Staging is not relevant for Occult Carcinoma (Tx, N0, M0)

* Including direct extension to intrapulmonary nodes
 ** Including superior sulcus tumor

(& : and) (/ : or) (&/ : and/or)

Figure 1.3: TNM staging in NSCLC [37]

Following complete surgical resection for stage IA disease the 5-year survival is 67% and with stage IB disease, 57%. [38] The 5-year survival rate for stage IV NSCLC, which is the most advanced stage (with distant metastases), is only 1%. [4] Small cell lung cancer in particular has usually spread widely (hence the staging term 'extensive') by the time the patient presents with symptoms. The 5-year survival rate for small cell lung cancer is 5%. [42]

Signs and symptoms of lung cancer

The most common symptoms that a patient with lung cancer may present with include new onset of breathlessness (dyspnoea), coughing of blood (haemoptysis), chronic cough or change to regular cough pattern, chest pain, weight loss (cachexia), hoarse voice (dysphonia) and less common presentations such as difficulty swallowing (dysphagia), finger clubbing and facial swelling which suggests superior vena cava obstruction (SVCO). [43] Long-term smoking, even in the absence of lung cancer can result in many of these symptoms, such as breathlessness, cough and occasional haemoptysis and contributes to patients and medical practitioners dismissing it as related to smoking rather than to potential cancer. This is a contributory factor to the late diagnosis of lung cancer.

The type of symptoms the patient presents with may give a clue to the location of the primary tumour. Thus, if the symptoms include wheeze particularly inspiratory or monophonic in nature, or the patient has stridor, then this suggests a central endobronchial (airway) tumour or a tumour pressing on the proximal airway. Similarly, bright red / fresh haemoptysis may signify endobronchial bleeding from a proximal endobronchial tumour. If the predominant symptom is of chest pain then this may reflect a peripheral tumour abutting the chest wall and causing damage to nearby sensory nerves. However, to confirm the presence and location of the tumour, further imaging is required (first line is a chest X-radiograph (CXR)). Some symptoms / signs are not directly due to the primary lesion itself, but from its spread to the lymph nodes or metastases to other parts of the body. Examples of this include SVCO or dysphagia caused by enlargement of the mediastinal lymph nodes, jaundice

caused by metastatic spread to the liver or musculoskeletal pain from bone metastases. The most likely sites of metastasis from lung cancer are contralateral (opposite side) lung, liver, adrenal glands, bone and brain and symptoms such as back pain, headaches or fits can often be the first manifestation of lung cancer and difficult symptoms to control as the cancer progresses.

Only about 10% of people with lung cancer do not have symptoms at diagnosis with the cancer being found incidentally, for example on a CXR arranged for alternative reasons. [42]

Diagnosis of lung cancer

Patients are usually referred to specialist chest clinics when they have developed symptoms and often will have a CXR that identifies an abnormality consistent with an underlying tumour. Figure 1.4 shows the typical finding on a CXR. Once an abnormality such as this has been identified then the patient usually undergoes an urgent Computed Tomography (CT) of their thorax (Figure 1.5) and abdomen to obtain greater detail about the size, location and nature of the primary lesion, whether there is lymph node involvement and whether there is metastatic spread above or below the diaphragm.

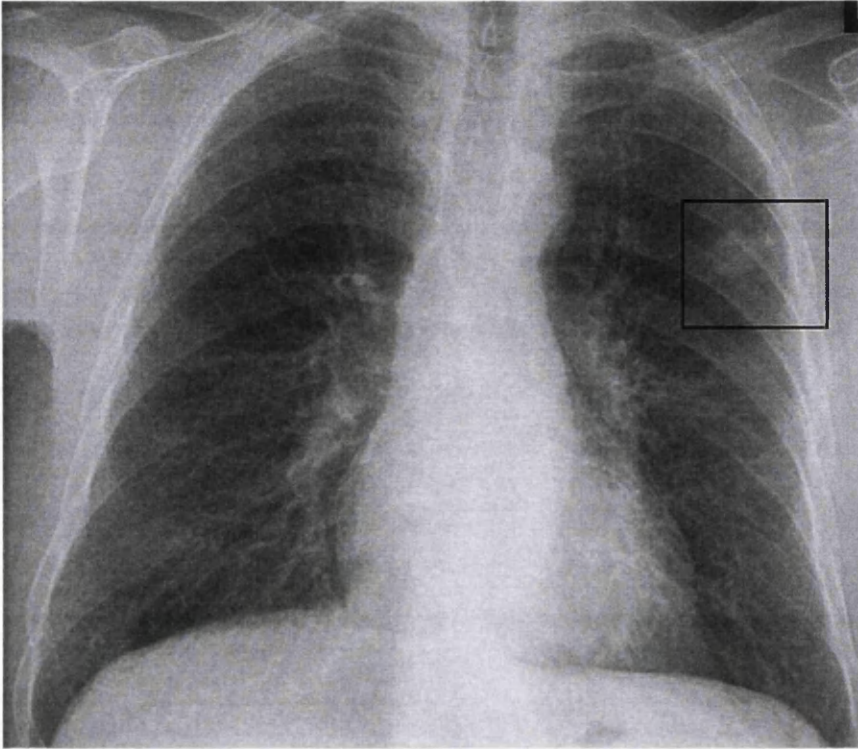


Figure 1.4: Left upper lobe rounded opacity consistent with lung cancer

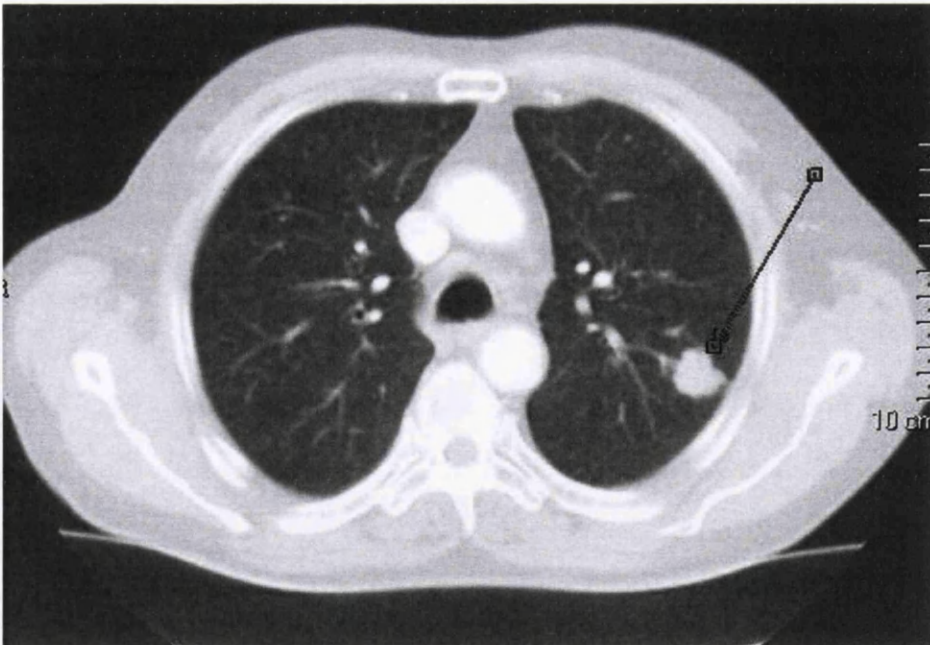


Figure 1.5: CT Thorax of patient in Figure 1.4. The distance and optimal direction for a planned CT-guided biopsy is marked (images taken from www.wikipedia.org)

Sputum cytology was first used in the mid-1930s but studies have shown that conventional cytology has a low sensitivity, especially with peripheral lesions. A systematic review included 16 studies and pooled 28,477 patients using sputum cytology and the overall sensitivity and specificity was 66% and 99% respectively. The indication for the sputum cytology was mixed in the studies which may have led to a degree of heterogeneity of the results. Thus 8 studies of 2,455 patients having sputum evaluated pre-bronchoscopy (for suspected lung cancer) resulted in sensitivity of only 22%. [44] An additional study not in the review included 60 consecutive patients suspected of lung cancer. Again the sensitivity in this group was very low at only 33%. [45] Thus the National Institute of Clinical Excellence (NICE) guidelines suggest that the use of sputum cytology may only be useful in those with central masses who are unable or unwilling to undergo bronchoscopy or any other invasive procedure. [46]

The majority of the patients will undergo a fibre-optic bronchoscopy to visualise the trachea and early bronchial divisions. During this procedure, if abnormal mucosa or tumour is seen (Figure 1.6 shows an example of this) then brushings and biopsies of the lesion are taken to provide a cytological/histological diagnosis.

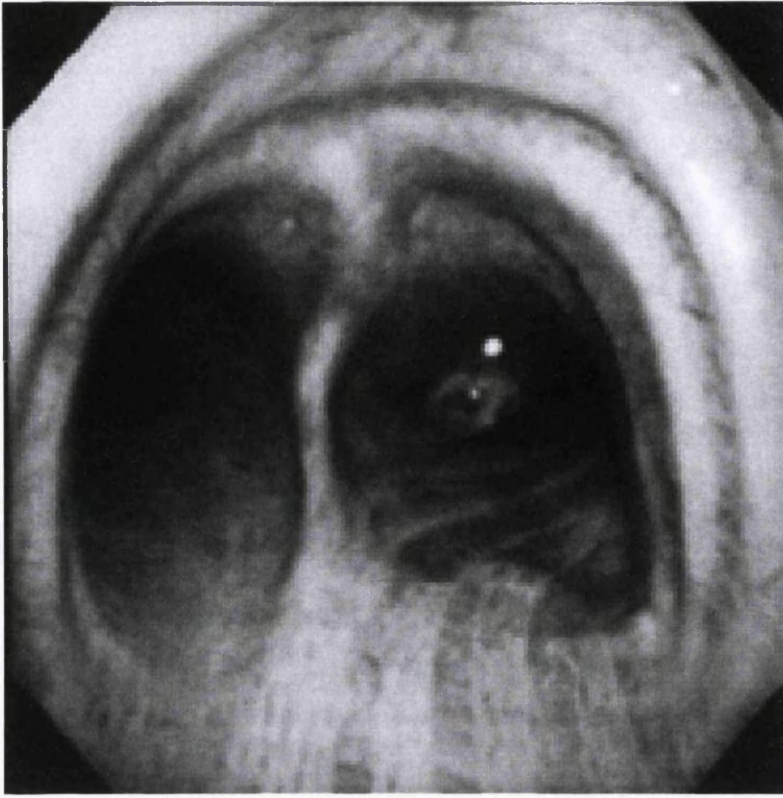


Figure 1.6: Erosion of the right main bronchus by a tumour seen by bronchoscopy [47]

If an abnormality is not identified, then distal bronchial washings (obtained by washing a particular lobe with 10 – 20ml of normal saline) are often performed in the lobe of the lung where the abnormality is seen on CT. Alternatively, if the mass is very peripheral and close to the chest wall, then a diagnosis may be obtained by a biopsy or Fine Needle Aspiration (FNA) performed under CT guidance (see Figure 1.5). On the rare occasions when a diagnosis is not possible using these means then a biopsy may be performed on a metastatic lesion (e.g. adrenal or bone metastasis). Other ways of obtaining a diagnosis are by sampling abnormal looking mediastinal or hilar lymph nodes (based on the CT) by techniques such as Transbronchial Needle Aspiration (TBNA) which uses the CT scan to 'map out' the lymph nodes which can be sampled, Endobronchial Ultrasound (EBUS) where the lymph nodes are directly visualised by ultrasound then a needle used to sample them, or even a mediastinoscopy which is a surgical procedure where the lymph nodes within the mediastinum are sampled. If the patient is too frail to have any invasive investigations or it is proving difficult to obtain suitable material for a histological diagnosis, then a diagnosis of lung cancer can be made radiologically and clinically.

Treatment of lung cancer

The basis of any treatment regime planned for a patient depends on the type of lung cancer they have, their physical and mental well-being (co-morbidity), the staging of the disease and their personal wishes. Surgery is an option in patients with early stage lung cancer who have no serious co-morbidities such as severe coronary heart disease, and have reasonable respiratory reserve as assessed by pulmonary function tests. Since the vast majority of lung cancers are smoking related, a large proportion of potentially operable cancers will not proceed to surgery due to the presence of significant chronic obstructive pulmonary disease (COPD). A recent study of over 4,000 patients showed that surgery itself has an operative death rate of 4.4%, depending on the patient's age, lung function and other risk factors. [48] If the tumour is operable but the patient is either not fit enough for surgery or decides against it, then radical radiotherapy can be given with curative intent. Should the cancer be deemed

inoperable then radiotherapy and chemotherapy can be given on a palliative basis. Patients who are too unwell for any palliative chemotherapy will usually be treated very conservatively with the management aim being symptom control.

Treatment options should be put to the patients and their relatives and they should be allowed time and questions before any decision is made on treatment. Specialist nurses often see patients repeatedly before and after they have had tests and treatment to provide some continuity in what can be a confusing and frightening process. The various treatment modalities used in lung cancer will now be discussed in more detail.

Surgery

Surgery as a management option for lung cancer is usually performed with curative intent. In 1933, Graham and Singer performed the first pneumonectomy (removal of a whole lung) for lung cancer in a one stage operation for a carcinoma which originated in the left upper bronchus. [49] This was thought to be the standard surgical operation of choice for lung cancers for several years. In the 1950s and 60s, the role of pneumonectomy and lobectomy (removal of a lobe of the lung) with mediastinal lymph node dissection was standardised and are similar to the operations carried out today. [50, 51]

Surgery is usually an option in NSCLC limited to one lung up to stage IIIA. [38] The earlier the cancer is detected the better the staging is likely to be and if surgically resected the better the chance of long-term survival. The 5-year survival rate for fully resected stage IA disease is 67%, IB is 57%, falling to 23% for IIIA disease. [38] Should the patient have stage II or III disease then adjuvant chemotherapy (chemotherapy after surgery) has been shown to increase the 5-year survival by up to 15%. [52, 53]

Once lung cancer has been diagnosed and CT-staging favours surgical resection, then a Positron Emission Tomography (PET) scan should be

performed to confirm the staging of the disease. The PET scan uses the increased glucose metabolism seen by cancer cells to detect where the cancerous cells are in the body. If the PET scan corroborates the staging, then extensive preoperative investigations must be carried out to assess the suitability of the individual for surgery.

Age:

Although age is not a contra-indication to surgery, increasing age has been associated with increased peri-operative morbidity. Elderly patients undergoing lung resection are more likely to require intensive peri-operative support. [54]. There is a higher mortality risk if over 80, and if pneumonectomy rather than lobectomy (14% mortality versus 7% respectively). [55-57]

Lung function (Spirometry):

Since smoking is the major contributing factor for lung cancer, impaired spirometry is likely to stop a patient from undergoing surgery. Three large studies in the 1970s had data from over 2,000 patients and showed that mortality of under 5% should be expected if the pre-operative Forced Expiratory Volume in 1 second (FEV₁) is > 1.5 litres for a lobectomy and > 2 litres for a pneumonectomy. [58-60]

Cardiovascular disease:

Pre-existing coronary artery disease increases the risk of non-fatal myocardial infarction or death within 30 days of non-cardiac surgery. [61] Patients who have had a myocardial infarction (MI) within 6 weeks should not be operated on. If they have had an MI in the last 6 months and are planning to have thoracic surgery then a cardiology opinion is required. [54]

Cerebrovascular disease:

All patients who have a history of previous stroke, transient ischaemic attacks, or carotid bruits should be assessed by carotid doppler studies (ultrasound of the carotid arteries). Patients with evidence of significant stenosis (for example, over 70%) should be assessed by a vascular surgeon or consultant in stroke

medicine pre-operatively and their management discussed with the thoracic surgeon. [54]

Nutrition and performance status:

A history of weight loss, poor nutritional status and poor performance status (World Health Organisation (WHO) or Eastern Cooperative Oncology Group (ECOG) scales, see Figure 1.7) have all been independently associated in lung cancer with advanced disease and a poor overall prognosis. [62-64]

0 – Fully active, no restrictions on activities

1 – Unable to do strenuous activities, but able to carry out light housework and sedentary activities

2 – Able to walk and manage self-care, but unable to work. Out of bed more than 50% of waking hours

3 – Confined to bed or a chair more than 50% of waking hours. Capable of limited self-cares

4 – Completely disabled. Totally confined to a bed or chair. Unable to do any self-care

Figure 1.7: ECOG/WHO Performance Status [65]

Types of surgery

There are several approaches to surgery, which can be taken depending on the location of the tumour. Localized tumours for patients with a good FEV₁ can be treated by lobectomy or bi-lobectomy, or a pneumonectomy if the tumour involves more than one or two lobes or has ipsilateral (same side) hilar node involvement. Pneumonectomy and lobectomy have mortality rates of 6-8% and 2-4% respectively. [66-70]

Sublobar resection was common in the treatment of tuberculosis and it was only a matter of time before it was used in the management of lung cancer. [71] It involves resecting the tumour with very little of the lung parenchyma removed. This is used usually when the tumour is peripheral with clear regional lymph nodes and when the pre-operative FEV₁ is markedly reduced. The mortality rates from this higher risk group of patients range from 1.4% to 3.5%. [72-75] However, the local recurrence rate is higher, up to 23%. [76] Long-term survival is 5 to 10%, which is worse than with a lobectomy. [74, 76]

Sleeve resections involve a lobectomy and removal of the part of the bronchus involving the tumour, forming an anastomosis (joining) between the bronchus proximal and distal to it. This particular surgical procedure may avoid the need for a pneumonectomy.

Post-operative complications

Such major surgery always entails the risk of post-operative complications. These include the risk of broncho-pleural fistula (connection between the bronchus and pleura), respiratory failure, infection, phrenic nerve damage causing diaphragm paralysis, recurrent laryngeal nerve damage leading to a hoarse voice and prolonged chest wall pain.

Chemotherapy

Small cell lung cancer is the most aggressive all the lung cancer subtypes. At presentation the disease has usually already spread beyond the point that surgery will be curative. Hence, the usual treatment modalities involve chemotherapy with radiotherapy. In small cell lung cancer, cisplatin and etoposide are most commonly used chemotherapeutic agents. [77]

There has been increasing evidence for the role of chemotherapy in advanced NSCLC. Benefits have been shown in palliation, survival, quality of life, symptom control and cost. There have been 13 studies that have assessed the effect of combination chemotherapy versus supportive care only. Two thirds of the trials showed a survival benefit, which was statistically significant and so unlikely to have occurred through chance. The median survival differences from the studies ranged from 9 – 22 weeks which is likely to be clinically important for patients. [78-90] The American College of Chest Physicians has guidelines which summarise how symptom control has been demonstrated showing an improvement for cough, haemoptysis, pain, dyspnoea, weight loss, anorexia and malaise. [91] Eight trials assessed quality of life, with a significant improvement in the chemotherapy arm in all but one of the trials. [78, 81, 88, 89, 92-95] Interestingly, chemotherapy appears cheaper than best supportive care alone. [96] Platinum based regimes are the most commonly used first line treatments combined with older drugs such as vindesine, mitomycin C and / or ifosfamide, or new agents such as gemcitabine, docetaxel or vinorelbine. Polychemotherapy is associated with better results than single agent treatment. [97-99] Two drug regimes are superior to one drug in terms of tumour response and improved survival. Adding a third drug has a weaker effect on tumour response and no effect on survival. [100]

When chemotherapy is used in addition to (after) surgery this is termed adjuvant chemotherapy, and it is used to improve outcomes. During surgery, lymph nodes are also sampled. If these contain cancer cells then depending on which lymph nodes are involved, the patient has stage II or III disease. Adjuvant chemotherapy in these situations has been shown to improve survival at 5 years by up to 15%. [52, 53]

Chemotherapy does, however, carry the risk of side effects and complications. Patients can develop anorexia, nausea, vomiting, fatigue and neuropathy. The more severe complications include toxicity and neutropenic sepsis (secondary to profound bone marrow suppression).

Radiotherapy

Radiotherapy can be given alone with curative intent in patients with early stage NSCLC who are not eligible for surgery. This is termed radical radiotherapy and involves high intensity radiotherapy. It is also given in combination with chemotherapy such as in limited disease (confined to one hemithorax) small cell lung cancer, which is potentially curable. [101] Patients with limited small cell lung cancer may also receive prophylactic cranial irradiation (PCI) to reduce the risk of brain metastases. [102] There is now increasing evidence that PCI has a role in extensive small cell lung cancer as it has been shown to reduce the cumulative risk of brain metastases within one year from 40.4% to 14.6%. The same study also shows a 1-year survival in the irradiated group of 27.1% compared with 13.3% in the control group. [103]

The role of adjuvant radiotherapy in patients whom have undergone curative surgery is limited. If there is a role then, its benefit has only been shown in patients who had N2 disease. No benefit was identified in those with N0 or N1 disease. [104]

Radiotherapy has a specific role in helping NSCLC and small cell lung cancer patients with significant symptoms such as bone pain or haemoptysis. Palliative radiotherapy involves smaller doses of radiation to the chest wall and it is aimed to provide symptom control.

Brachytherapy (localised radiotherapy) is a more recent introduction. It can be given directly into the airway via bronchoscopy when the cancer affects a short section of bronchus. It can also be used as a palliative modality when there is blockage of a large airway by an inoperable lung cancer. [105, 106]

Targeted Therapy

Receptor tyrosine kinases (RTK) are transmembrane receptor-linked kinases that have a pivotal role in diverse cellular activities including growth, differentiation, metabolism, adhesion, motility, and death. [107] Many RTKs are involved in oncogenesis either by gene mutation or chromosome translocation. [108] In recent years tyrosine kinase inhibitors have been developed for the treatment of advanced lung cancer which targets the tyrosine kinase domain of the epidermal growth factor receptor (EGFR) expressed in many cases of NSCLC. Erlotinib (Tarceva) has been shown to increase survival in lung cancer patients and has been approved as a second line treatment of NSCLC. [109]

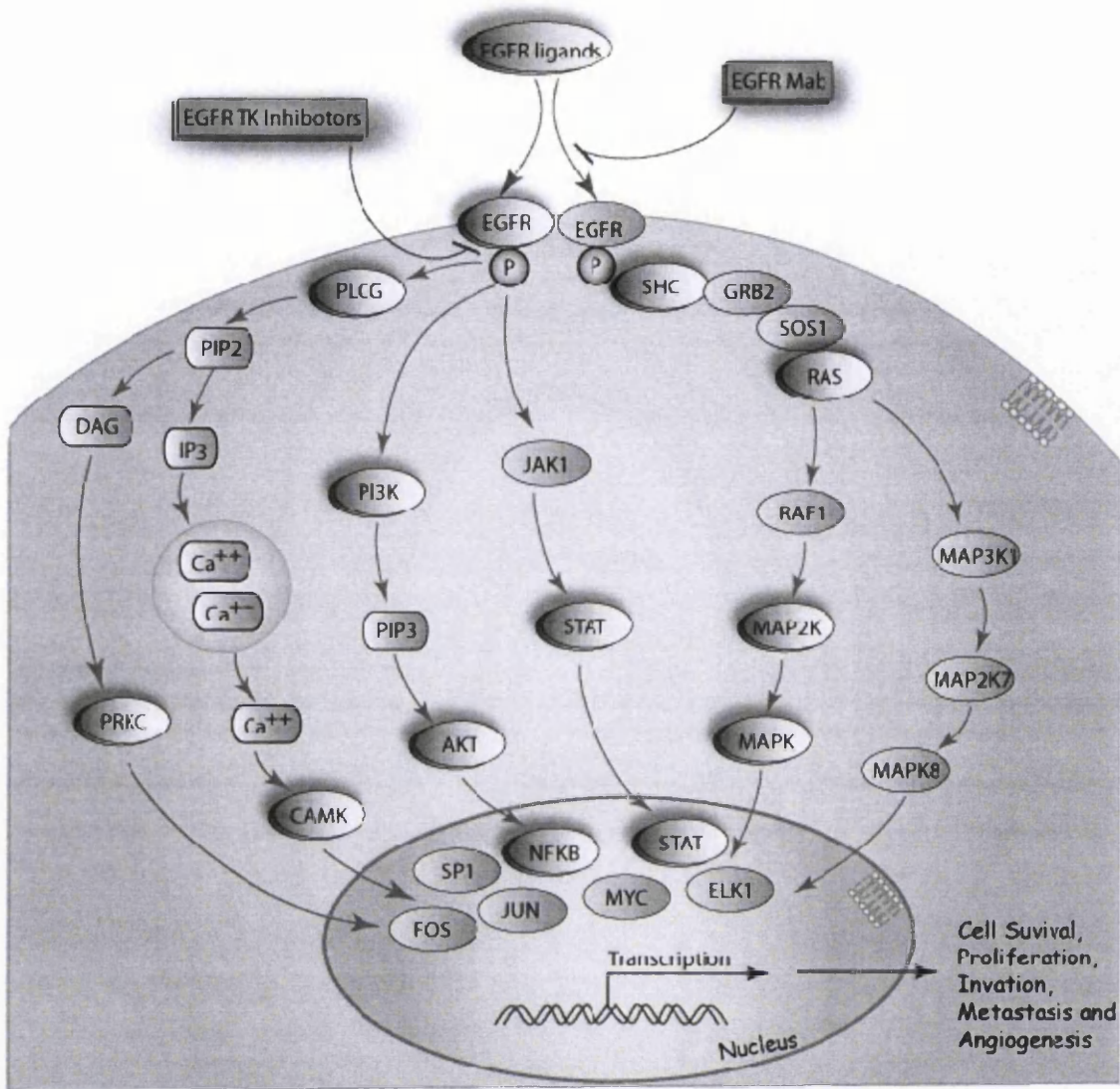


Figure 1.8: Proposed mechanism of how EGFR may cause or support the inappropriate cell growth that leads to proliferation, migration, and survival of cancer cells (image taken from www.gene.com)

Angiogenesis, the growth of new vessels from pre-existing vessels, is a fundamental step in tumour growth and progression. Vascular endothelial growth factor (VEGF) is a key angiogenic factor implicated in tumour blood vessel formation and permeability. Over-expression of VEGF has been observed in a variety of cancers and has been associated with a worse relapse-free and overall survival. The angiogenesis inhibitor bevacizumab (a monoclonal antibody directed against VEGF) in combination with paclitaxel and carboplatin improves survival of patients with advanced NSCLC (stage IIIb and IV) but does increase the risk of lung bleeding, particularly in patients with squamous cell carcinoma. [110, 111]

Supportive treatment only (Palliation)

Many of the patients diagnosed with advanced lung cancer will have significant co-morbidity and poor performance status and the decision has to be made whether to give any treatment such as palliative chemotherapy or radiotherapy or just to give best supportive care. The patient themselves are central in the decision making along with the multi-disciplinary team (MDT) which usually consists of a radiologist, chest physician, oncologist, thoracic surgeon and palliative care specialist. Once best supportive care is decided as the best management course then it is important to make sure the patient is comfortable in their remaining days. Adequate analgesia, diet and even oxygen may be required as well as palliative radiotherapy if deemed appropriate.

Figure 1.9 shows a typical diagnostic pathway for the diagnosis and management of lung cancer.

Conclusion: An overview of lung cancer

Lung cancer is usually a devastating disease and is often silent until the advanced stages. There are various ways a patient may present and then certain pathways in which to diagnose the cancer. Only a small number will be diagnosed early enough and will be fit enough, for potential surgery and thus curative treatment. The majority, however, will have inoperable lung cancer and will therefore require a combination of chemotherapy or radiotherapy alone or in combination or management by best supportive care. Whatever the treatment modality overall 5-year survival rates for lung cancer are worse than other common cancers.

Two potential ways of tackling lung cancer are:

- 1) to detect the cancer early enough so that it can be treated with curative intent;
- 2) to prevent it from occurring.

One of the key ways of detecting the disease early is by screening and this will be discussed in Chapter 2. Smoking causes 85 - 90% of lung cancers, thus by reducing the number of people who smoke and the number of cigarettes that are smoked, this would hopefully reduce the number of lung cancer cases in the future. It is important to develop smoking cessation programmes worldwide in order to help individuals to stop smoking and I have discussed the role smoking has in cancer in Chapter 3.

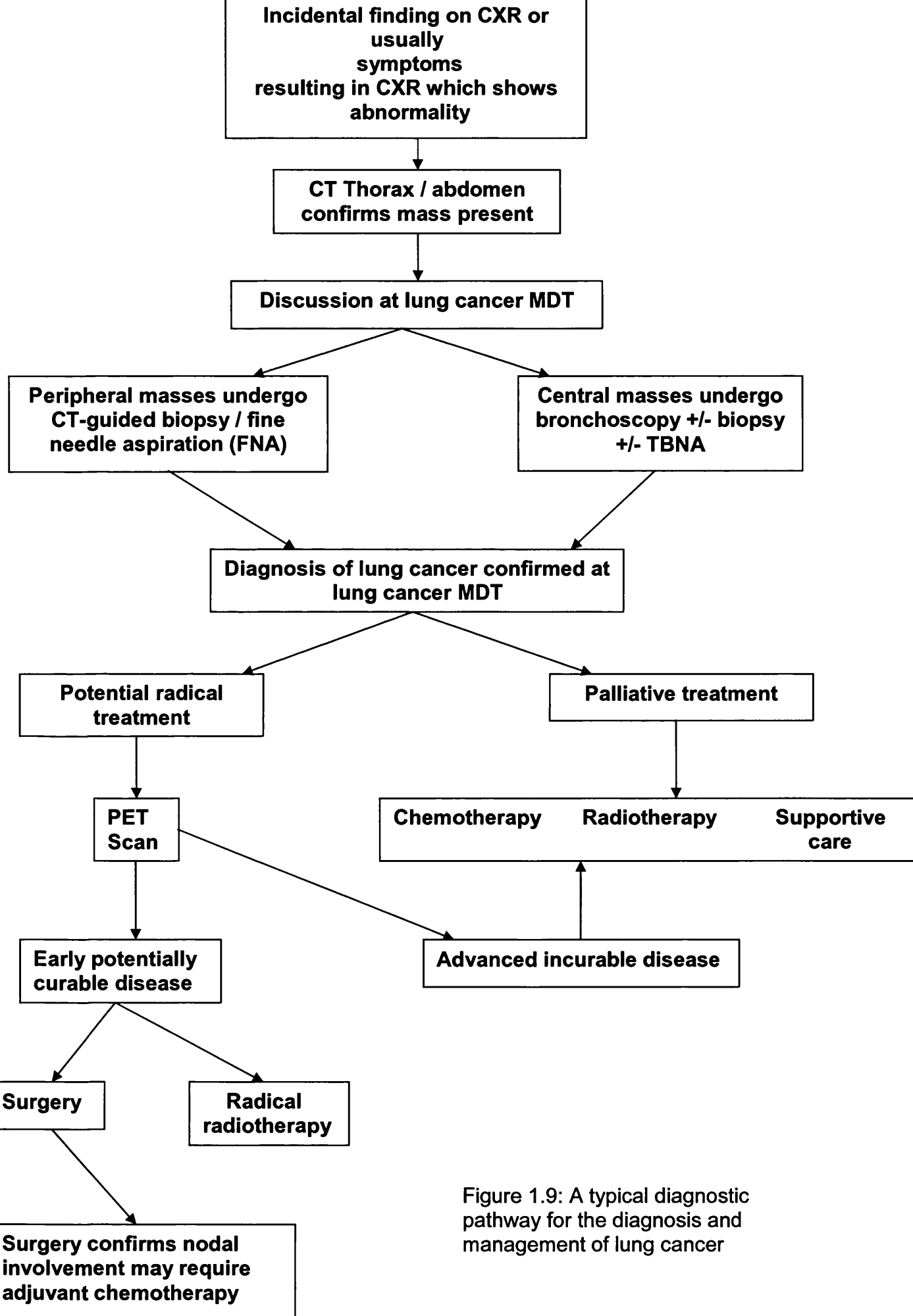


Figure 1.9: A typical diagnostic pathway for the diagnosis and management of lung cancer

Chapter 2. Screening in lung cancer

Principles of screening

Screening is an important process in medicine, which is used to identify a disease process within a defined population early thus enabling earlier intervention and management in the hope to reduce mortality and suffering from a disease. [112] In 1968 the World Health Organisation published guidelines on the principles of screening. [113] For a screening tool to be effective, certain principles must be adhered to:

a) The condition should be an important health problem.

Lung cancer is the most common cancer worldwide with 1.3 million new cases diagnosed each year. [21] There were an estimated 33,000 new cases diagnosed in the England and Wales alone in 2006. [26]

b) The disease has significant mortality and morbidity.

In Europe, lung cancer has an overall 5-year survival of only 10.9% (compared to a 5-year survival of 79% for breast cancer and 77.5% for prostate cancer). [19]

c) There should be a latent phase of the disease.

Lung cancer usually presents when a patient develops symptoms. In the majority of cases, the cancer is already in an advanced stage at the time symptoms develop. The aim of screening is to detect these cancers in the earlier asymptomatic / latent phase.

d) Intervention earlier in the disease process can improve outcomes.

The 5-year survival for NSCLC is much higher in less advanced disease. Patients with Stage IA NSCLC who have undergone complete surgical resection of the tumour have a 5-year survival rate of 67%. This compares with a 5-year survival of 1% for Stage IV NSCLC. [38]

e) The screening test itself should have certain characteristics.

It must be sensitive (to avoid false negatives and inappropriate reassurance) and specific (to avoid too many false positives leading to unnecessary worry and further expensive / invasive testing). The test should also cause little harm (or much less than the disease), and should be acceptable to the population to be screened.

f) The cost of finding a case using the screening technique should be balanced in relation to medical expenditure as a whole.

During this chapter I will discuss various screening modalities which have been studied over the last 30 years, firstly reviewing the role of radiology, then more simple sputum cytological techniques used in conjunction with simple chest radiographs before discussing in more detail newer biological techniques which have been studied and are still currently being evaluated.

Low-dose spiral CT scans and screening for lung cancer

This is probably the most widely tested screening modality for lung cancer to date. There have been many observational studies assessing the feasibility of using low-dose spiral CT (LDCT) of the thorax in detecting lung cancer at an earlier stage and thus allowing curative treatment. The advantage of using such a technology is that it allows detection of much smaller nodules (2-3mm), which would almost certainly not be visible on a CXR. One of the earliest studies was carried out in Japan by Kaneko *et al*, where 3,457 people underwent LDCT scan and CXR. The results confirmed detection of peripheral lung cancers in 15 of the 3,457 cases (0.3%). 11 of the 15 cases had negative CXRs (73%). Thus the detection rates were 0.43% (15 from 3457) with CT scans and 0.12% (4 from 3,457) with CXRs. 14 of the 15 new cases (93%) of cancer were stage I. [114] This early study strongly suggested that low-dose spiral CT was a superior radiological modality compared with CXR in the early detection of lung cancer.

The Early Lung Cancer Action Project (ELCAP) was a major study published in 1999 by Henschke *et al*, where a study population of 1,000 participants underwent yearly LDCT scans and CXR in a single-arm study. They compared the number of lung nodules detected by CXR and CT. They went on to analyse how many of these pulmonary nodules were malignant and the radiological staging of the cancer cases. 233 (23%) participants were found to have non-calcified nodules by CT scan at baseline compared with 68 (7%) by CXR. 27 nodules (2.7%) were found to be malignant, with 20 of these not being found on CXR. 23 out of 27 (85%) of these malignant nodules were classified as Stage I with 26 (97%) being resectable. In the second year of screening, a further 7 cancers were detected with 6 of these in Stage I. The conclusions of the study were that LDCT scan was four-times as sensitive as CXR in detecting lung cancers, and six times more sensitive at detecting stage I disease. [115]

Subsequent studies continue to show a similar trend. Soube *et al* screened 1,611 participants every 6 months with LDCT scan, CXR and sputum cytology. 186 (11.5%) of the 1,611 patients were found to have a "positive" CT scan

compared with 55 (3.4%) with a "positive" CXR. 14 (0.87%) of the 1,611 cases were found to have lung cancer with 77% of these in stage I. On repeated screening, 7,891 examinations were carried out and 22 (0.28%) new lung cancers were diagnosed, 82% of these stage I. The 5-year survival rate for all incidence cancers discovered via screening was 65%. [116] Swensen *et al* discussed their Mayo Clinic results in 2003. The study involved 1,520 participants aged 50 years or older with a greater than 20 pack year smoking history. Participants had 3 annual LDCT scans of thorax / abdomen and sputum cytology. 1,049 (69%) had non-calcified pulmonary nodules (2,832 nodules were identified in total) with 26 baseline cancer cases diagnosed. 10 further cases were diagnosed during screening on CT alone. 2 were picked up on sputum cytology alone and 2 further cases presented as interval cancers (i.e. developed symptoms in between scans). Thus 40 primary lung cancer cases were diagnosed (2.6%) which equates to 1.4% of the 2,832 nodules. [117]

Diederich *et al* studied 817 individuals. These were deemed a high-risk population with age ≥ 40 years and smoking history ≥ 20 pack years. Each of the participants underwent LDCT scan of thorax. 350 (43%) out of 817 had non-calcified pulmonary nodules. If the nodules were greater than 10mm, they were considered potentially malignant and either biopsy was attempted, or follow-up CT scans were recommended to assess for any growth. For all other lesions, follow-up CT scan was recommended. 32 nodules in 29 participants were > 10 mm. Biopsy of 15 lesions revealed 12 lung cancers in 11 cases (1.3%) and of these 7 (64%) were stage I and 2 (18%) stage II. The 17 other nodules > 10 mm were followed up for 24 months with serial CT scans and did not show any growth. [118] Diederich *et al* published further work on the same cohort of participants assessing the annual follow up with CT scan in asymptomatic patients. Follow up of non-calcified nodules present at baseline CT scan showed growth in 11 of 792 subjects. 8 of the 11 growing nodules were biopsied, identifying 7 lung cancers. A further 174 new nodules were identified, 3 representing lung cancer. 7 of the 10 newly diagnosed cancers were stage I. [119] Further studies continue to show this trend of CT scans being much more sensitive for detecting lung cancers than plain CXRs. [120-122]

All the above trials are very promising but questions along the same themes were raised throughout:

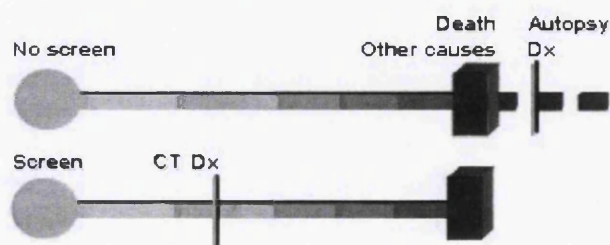
a) False positives:

There were an extremely large number of non-calcified nodules detected within the screening population. The majority of these nodules were ultimately benign. However, after detecting such nodules at baseline, they required following up with interval scans for at least 2 years, and some participants would have undergone biopsies on the lesions before they could be concluded as being benign. The presence of an abnormality on a CT scan such as a nodule is likely to cause a high degree of anxiety and this will be protracted over a 2 year period within which follow-up scans are performed. Bearing in mind the large number of the study population with nodules found on their CT scan, this could be a concern. Many of these healthy individuals will elect to undergo invasive procedures such as biopsy or even surgical resection, with the potential morbidity associated with it.

b) Overdiagnosis, lead-time bias and length-time bias.

Even nodules that turn out to be lung cancer may be indolent, very slow growing and ultimately would not have spread or caused death in the individual due to their co-existing morbidity. Thus in normal clinical practice these tumours are unlikely to be diagnosed prior to death and if they were found on screening and treated would not have altered the life expectancy of the individual.

Overdiagnosis Bias (Pseudodisease)

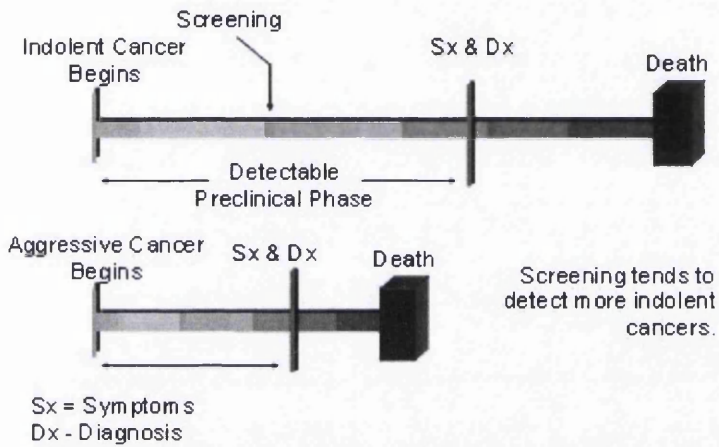


Screening detects cancer (pseudodisease) that would remain subclinical before death from other causes.

Dx = Diagnosis

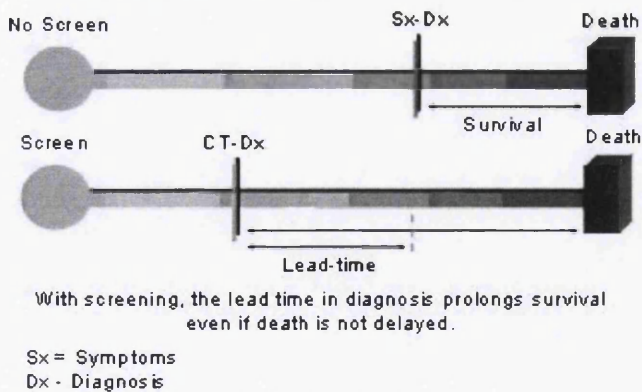
This is termed overdiagnosis and when a disproportionately large percentage of these slow growing tumours are detected by screening it results in a higher 5-year survival rate. This is called length-time bias. Again this carries the burden of anxiety to the individual diagnosed with lung cancer in this situation within the screened population.

Length Bias



Lead-time bias is when the lung cancer has been detected earlier by the screening modality but the time of death has remained the same. This again gives the impression of improved survival rates; however this is purely due to earlier detection rather than later death.

Lead-time Bias



Thus, it is due to overdiagnosis, lead-time bias and length-time bias, that an increase in 5-year survival rates can be quoted in the studies, however the lung cancer mortality rates are similar to the unscreened population. To be beneficial screening tests should detect disease before signs or symptoms occur and *must* lead to decreased mortality.

c) Cost-effectiveness:

LDCT screening on a population level carries a significant cost implication. The initial screening with the requirement of follow-up scan for the large number with nodules, along with the potential for invasive procedures and specialist consultations will all amount to a high cost. This cost of course may be acceptable, but this will only be if the screening reduces mortality rates.

d) Direct harm:

There is also the controversial issue of risk of cancer associated with diagnostic CT scanning. Berrington de Gonzalez and Darby [123] estimated that the attributable risk percentage of cancers from diagnostic x-rays ranged from 0.6% to 1.8% of all cancers in most developed countries. However this is likely to be debated for many years to come.

LDCT scanning as a screening tool in early lung cancer is promising. There are issues associated with detection of large numbers of benign lesions, and overdiagnosis of malignant lesions that will not have an impact on life expectancy. Both these issues could adversely affect patients. CT screening has detected more stage I cancers, and if used with a good diagnostic algorithm, may still prove effective in improving survival and mortality rates. In terms of meeting the criteria of what makes a beneficial screening test, this continues to be debated. At present the studies have not shown a significant reduction in mortality rates, and some individuals have required follow-up tests, some of which have been invasive and most of the studies have been observational, proof-of-principle studies. What is urgently required is a large, randomised, controlled trial to see if CT scanning will have an impact on lung cancer mortality.

One such study has just been published. The DANTE (Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays) trial was a randomised control trial looking at LDCT in one arm and a non-investigated control arm. All subjects (males aged 60 – 75 years, > 20 pack year smoking history) had a baseline CXR and 3 day pooled sputum sample. Those excluded were those with less than 5 year life expectancy, those with cancer within the last 10 years, and those unable to comply with follow-up. The LDCT arm (n=1,276) had a scan performed at baseline and then once a year for four further years (with a yearly medical interview) and the control arm (n=1,196) had a yearly medical interview and provided there was no clinical indication, no further investigations were required. The results showed that at a median follow up of 33 months, lung cancer was detected in 60 (4.7%) of the LDCT arm and 34 (2.8%) of the control arm ($p=0.016$) with the resectability rate similar in both groups. More patients in the LDCT arm were detected in stage I (54% v 34%; $p=0.06$) and the number of advanced cases were similar for both arms. 20 (1.6%) of patients in the LDCT died from lung cancer as did 20 (1.7%) from the control group, again showing that additional scanning of the study arm had no impact on the overall mortality. [124]

Two further studies are ongoing: The Prostate, Lung, Colorectal and Ovarian Cancer screening trial (PLCO study) has been instituted by the National Cancer Institute (NCI) and includes over 150,000 men and women aged 55 to 74 who are split into two trial arms. One arm undergoes intervention (screening) whilst those in the other control arm continue their normal health care routine. [125] The National Lung Screening Trial (NLST), launched in 2002, is a randomised control study, looking at LDCT and CXRs as screening techniques for lung cancer. Nearly 50,000 subjects have been enrolled into the study consisting of current and former smokers. They have recruited a large enough cohort in order to determine if there is a 20% reduction in mortality using LDCT compared to CXR. [126] The long-term survival outcomes from these studies are due to be reported in the next 5 to 10 years.

Sputum cytology and screening for lung cancer

Sputum is a good biological specimen to conduct screening because it is usually readily accessible (i.e. non-invasive). Sputum cytology in combination with CXRs has been attempted as a screening strategy for detecting (early) lung cancer.

The Memorial Sloan-Kettering Study

The Memorial Sloan-Kettering Study was sponsored by the NCI; it commenced in 1974. The study cohort consisted of 10,040 male volunteers, from New York, over 45 years old who smoked. They had a baseline postero-anterior and lateral CXR, then 4,968 were randomised to have annual CXRs together with sputum cytological examinations every 4 months for 5-8 years (dual – screened arm) and the other 5,072 men received an annual CXR for 5-8 years. At baseline, the dual screening group had 30 (0.6%) lung cancers detected (29 of which were NSCLC), with sputum cytology alone detecting 9, CXR alone detecting 14 and 7 cancers detected by both. There were 23 lung cancers (all NSCLC) detected at baseline in the CXR only arm. Complete resection was possible in 18 of the 30 baseline cancers in the dual group and 11 of the 23 in the CXR only group. On follow-up screening, 114 incident cases were identified in the dual screening arm (14% by sputum cytology alone) and 121 were detected in the CXR alone arm. Of the 288 cancers diagnosed throughout the study a high proportion (40%) were stage I (whose 5-year survival was 76%). The overall 5-year survival in both arms was identical at 35% and was higher than the generally reported 5-year survival of 10% at the time. However, there were no differences in mortality rate of 2.7 per 1,000 person-years. The potential reasons for the higher survival rates are discussed later. [127-129]

The John Hopkins Study

The John Hopkins Study applied a similar protocol to a target population from the Baltimore metropolitan area. 10,387 male, smoking (and ex-smoking) volunteers had baseline CXRs. The 5,226 receiving annual CXRs combined

with 4 monthly sputum cytology examinations were found to have 39 (0.75%) lung cancers (36 NSCLC) on initial screening (11 by cytology alone, 20 by CXR alone, 8 by both modalities). The 5,161 men receiving CXRs only had a similar number of 40 (0.78%) lung cancers (32 NSCLC) detected at baseline and the numbers of incident cancers were similar in both groups over the next 8 years (194 in the dual group and 202 in the annual CXR only group). The overall 5-year survival rate for those diagnosed with lung cancer in both groups was 20% and most importantly, the overall group mortality rates were similar not only to each other but to mortality rates of the unscreened population. [127, 130]

The Mayo Lung Project

The Mayo Lung Project was the third NCI sponsored randomised controlled trial (RCT) investigating lung cancer screening. Outpatients undergoing medical examinations in the Mayo Clinics were invited for screening if they were men aged 45 years or older and smoked at least 20 cigarettes/day in the preceding year. 10,933 men, had baseline CXRs and 3 days pooled sputum cytology examinations. 79 cases of NSCLC and 12 cases of SCLC were diagnosed (initial CXR alone n=59, sputum cytology alone n=17, both n=15). The remainder either received a CXR and 3 day pooled sputum examination every 4 months or every 12 months (at that time deemed 'usual care'), for 6 years. Of the original 91 (0.83%) lung cancer cases diagnosed, the 5-year survival was good at 40%, increasing to 70%, for those with stage I or II disease.

During the follow-up period, more cancers (n=206) were diagnosed and more of these were resectable (48%) in the 'screened' group than in the 'usual care' group (n=160 of which 32% were resectable). The 5-year survival in the screened group was also higher at 33% compared with 15%. Initially this was felt to be very promising but the improved survival in the screened group could have been explained by overdiagnosis, as there were more lung cancers in this group overall, than the control group despite randomisation. This is supported by the lack of statistically significant advantage in the overall mortality rates (3.2 versus 3.0 per 1000 person-years, respectively) at 6 years or even after 20 years. [131-134]

The Czechoslovakia Study

The final major study investigating the role of CXR combined with sputum cytology screened 6,364 males aged 40 – 64 years, whom had smoked approximately 150,000 cigarettes throughout their lifetime and were smoking at the time of enrolment. 18 cancers (0.3%), were diagnosed on initial testing, 6 (33%) of which underwent curative resection. The remaining 6,346 either received CXRs and sputum cytology performed every 6 months for 3 years, or CXR and sputum cytology performed 3 years after the initial examination. Subjects in both arms were then screened annually for 3 further years. There was statistically no difference in the survival or mortality rates between the two arms and indeed the majority of cancers in the control arm were found from interim tests as new symptoms developed. [135] There was no difference in mortality rates after 15 years. [136]

These studies suggest that screening with CXRs does not influence mortality and adding regular sputum cytology screening to CXRs also offers no further long-term survival advantage for the screened population. Some of the improved early survival rates were once again likely to be due to a combination of overdiagnosis, lead-time bias and length-time bias.

An interesting and consistent finding from the studies is the high frequency of detection of squamous cell carcinoma (SCC) in the sputum cytology groups, which is theoretically the most amenable to radical treatment.

As CT screening and CXRs with or without sputum cytology have made little impact on lung cancer mortality, other researchers are looking at other biomarkers for diagnosing early lung cancer.

The principle of biomarkers and lung cancer- underlying histological and cellular changes

A biomarker can be defined as a substance used as an indicator of a biologic state. [137] A good biomarker should require none or a minimally invasive technique of sampling. [138] As the understanding of the biology of lung cancer develops, easily accessible body tissues such sputum or serum will be sampled for other non-cytological biomarkers (or even a panel of biomarkers) that could help in the early detection of early or even premalignant lesions. Biomarkers are also being evaluated to monitor and even predict treatment response. An example of this would be the serum biomarker, prostate-specific antigen (PSA) that is used to monitor response after radical surgery or chemotherapy in prostate cancer.

Both squamous dysplasia and carcinoma-in-situ are deemed pre-neoplastic changes, which often but not always lead to squamous cell carcinoma of the lung. [139] Pre-neoplastic changes in lung adenocarcinoma and carcinoid include atypical adenomatous hyperplasia and diffuse idiopathic pulmonary neuroendocrine cell hyperplasia. [139] When developing biomarkers, researchers have monitored these cellular alterations, in particular the gene expression and chromosome structure, which take place within these pre-neoplastic lesions. Some of these changes include mutations in the p53 and *ras* genes that are associated with hyper-proliferation and loss of cell cycle control, aberrant gene promoter methylation, increased vascular growth and altered protein expression. [139-141]

Sputum biomarkers and screening for lung cancer

Gene Promoter Methylation

This is a very promising modality for lung cancer screening. Early in the development of lung cancer, methylation of promoter sequences in multiple different tumour suppressor genes occurs associated with the silencing of transcription and inactivation of these tumour suppressor genes. This has a

crucial role in triggering malignant transformation and progression. [141]

Detection of methylation is performed by specific polymerase chain reaction analysis. Hypermethylation was investigated in a small 'proof-of-concept' study looking at methylation of p16 and O⁶-methylguanine-DNA methyltransferase (MGMT) gene promoters in a high-risk population. The sputum samples of 21 patients with squamous cell carcinoma were investigated. The presence of one or both gene promoters was found in all 21 sputum samples irrespective of whether they were taken up to 3 years (n=10) before or at the time of diagnosis (n=11). This compares to 15% (methylation of p16) and 25% (methylation of MGMT) in 123 controls, deemed high risk through a smoking history and/or radon exposure. Most interestingly, 48% of the squamous cell cancer patients had methylation of both genes compared to only 4% of controls ($p < 0.001$). On follow-up, 3 lung cancers were diagnosed in the controls between 1 and 3 years after sputum collection with the MGMT gene being methylated in two of the subjects. [142]

Other studies report the potential of gene-specific methylation as a biomarker for lung cancer. Belinsky *et al* postulated that a panel of methylated genes would yield greater sensitivity and specificity. They recruited subjects older than 25 years, with ≥ 30 pack-year smoking history and a forced expiratory volume in 1 s (FEV₁) $\leq 75\%$ and FEV₁/FVC (forced vital capacity) ≤ 0.75 . They excluded anyone with any cancer diagnosed in the previous 5 years. 3,259 subjects provided pooled sputum for 3 days in 1 container and the next 3 days in a second container. The sputum from the second container was used as samples for the study. 182 subjects were diagnosed with (incident) lung cancer with 1,353 cohort deaths. Once sufficient quality of DNA was analysed the study cohort consisted of 98 subjects (cases) and the non-cancer cohort was used to form a group of 92 controls matched for age, gender and month of enrolment. There were 26 current smokers in the control group compared with 42 in the cases; however the total number of pack years was similar. Fourteen genes were analysed for promoter methylation, including p16 and MGMT. They found that 6 of the 14 genes (p16; MGMT; DAPK; RASSF1A; PAX5 β and GATA5) were individually associated with a $> 50\%$ increased risk of lung cancer. Sputum collected within 18 months of a diagnosis of lung cancer had

more methylated gene promoters than sputum from the same subject collected more than 18 months before the diagnosis. The concomitant methylation of 3 or more genes was associated with a 6.5-fold increased risk of lung cancer with receiver operating characteristics of a specificity of 64% and sensitivity of 64%. [143] This is the first study to show how a panel of genes can be used as potential biomarkers in screening a high-risk population for lung cancer. They did conclude that the level of specificity obtained is not yet high enough for prospective screening studies but called for more evaluation candidate gene panels – especially as technology improves.

Serum biomarkers and screening for lung cancer

Serum biomarkers have been used to screen for and assess treatment response in different cancers. These include prostate-specific antigen (PSA) for prostate cancer, carcino-embryonic antigen (CEA) and carbohydrate antigen (CA) 19-9 in colonic cancer and CA125 in ovarian cancer. Serum has been used as a screening biofluid for lung cancer.

Circulating DNA and genetic changes

Elevated levels of circulating cell-free DNA in the blood of cancer patients were first reported in 1987. Although the precise mechanism of DNA release into the blood remains obscure, it appears that much of this circulating DNA is derived from apoptotic and necrotic tumour cells. [144] Patients with lung cancer have genetic and epigenetic changes consisting of chromosome loss, oncogene activation and tumour suppressor gene methylation [145, 146] and if this abnormal DNA can be detected in the serum, it may allow development of specific markers. Levels of circulating cell-free DNA are generally higher in the blood of patients with lung cancer compared with healthy controls [147, 148], but their specificity / sensitivity and receiver operator curves specific to differentiate lung cancer have still not been clarified.

K-*ras* and p53 are well characterised common mutations in human lung cancer and pre-neoplasia, thus are the most commonly studied genetic mutations.

Overall, *k-ras* mutations have been found in serum of 20-30% of lung cancer patients with only one study detecting the mutation within the control group, although the characteristics of the controls were not specified. [149-152] P53 mutations have been found in the serum of 10-30% of lung cancer patients compared to minimal detection in the healthy control groups. [153] One study found p53 mutations in 41% of lung tumours with the identical mutation identified in the plasma of 73% of them. [154]

Vascular Endothelial Growth Factor (VEGF)

The role of VEGF in carcinogenesis has been discussed in Chapter 1. Recent preliminary data found significantly increased VEGF concentration (measured using enzyme immunoanalysis) in the serum of 7 patients with NSCLC compared to 5 healthy smokers and 7 healthy non-smoking controls ($p=0.033$). [155]

Proteomics

Proteomics is the study of proteins and two approaches have been applied to lung cancer. Protein profiling is where patterns of protein expression are used and the other is to identify individual proteins. A recent study analysed serum from 158 lung cancer patients and 50 controls using a mass spectrometry technique. 74 lung cancer serum samples and 20 healthy controls were used to develop a training set. From this, a specific pattern consisting of five protein peaks in the serum was chosen as serum biomarkers that could diagnose NSCLC and this pattern was used (in a blinded fashion) to try and discriminate the remaining serum samples of 84 lung cancer patients and 30 healthy controls. The pattern of proteins was used to differentiate the samples with a sensitivity of 86.9% and specificity of 80%. [156] These proteomic techniques are also being evaluated to look for patterns in serum that are associated with premalignant changes in the lungs of high-risk individuals. [157]

Unfortunately, the lack of adequate sensitivity, specificity and reproducibility has meant that no single specific biomarker associated with specific cell

changes has yet been identified. By understanding the biology of lung carcinogenesis in more detail, various groups are developing high throughput techniques, using panels of biomarkers associated with these cellular changes for early detection of lung cancer.

Breath biomarkers and screening for lung cancer

Volatile Organic Compounds in Breath (VOCs)

Exhaled breath samples are even more accessible than sputum or serum and more than 200 different measurable chemicals are exhaled in the human breath. [158] In 1985, Gordon *et al* studied VOCs in exhaled air from 12 patients with histologically confirmed NSCLC and 9 healthy controls (2 of which were heavy smokers). Gas chromatography combined with mass spectrometry was used to identify peaks representing individual and combination of VOCs from the breath samples. The results showed 4 separate peaks occurred in more than half of the lung cancer patients, which were completely absent in the control group; the authors suggested that unique VOCs are exhaled in the breath of lung cancer patients, which could have potential diagnostic usefulness. However, they recommended that this technique be tested in a larger population to assess its true effectiveness. [159]

Phillips *et al*, from the USA, collected exhaled breath in 108 fasting patients about to undergo a bronchoscopy for an abnormal CXR. Lung cancer was confirmed histologically in 60 patients (50 NSCLC, 10 SCLC) and excluded in 48. Many VOCs were common to both sets of breath samples but a group of 22 of these VOCs were able to distinguish between lung cancer and control cases according to risk weighting attached to the test. For example, the VOCs had 100% sensitivity (i.e. no false negatives) and 81% specificity for stage I lung cancer if the receiver operator accepted a post-test probability of 0.46. A post-test probability of 0.9 yielded lower 66.7% sensitivity but 100% specificity (no false positives). The abnormal VOCs consisted of mainly alkanes and benzene derivatives. Smoking status alone did not account for the benzene derivatives since these were also present in the breath of non-smokers and ex-smokers, so

it was likely to be a genuine marker for lung cancer. Surprisingly, there were no significant differences in sensitivity and specificity of VOCs between early and advanced lung cancer. These researchers also concluded that patterns of VOCs could act as a 'finger-print' for lung cancer but would need validation in a larger general population before they could recommend the test as a diagnostic screening tool. [160]

The same researchers compared exhaled VOCs in patients with biopsy proven primary lung cancer (n=67); non-lung cancer metastasising to the lungs (n=15); abnormal CXRs but no histological evidence of lung cancer (n=5) and healthy volunteers from the general population (n=41). The breath test identified over 80 different alkanes and mono-methylated alkanes that were then used via discriminant functional analysis to generate a predictive model using a panel of 9 VOCs. These 9 VOCs, in combination yielded a sensitivity of 89.6% and a specificity of 82.9% (using a post-test probability of 0.5) of identifying primary or secondary lung cancer. Again, there were only minor differences when subjects were stratified according to history of tobacco smoking, histological cancer type, or lung cancer staging. [161] The authors concluded that their test was sensitive, cost effective and easy to perform and could be used to complement other modalities, but again called for further evaluation.

Poli *et al*, from Italy, looked at VOCs in a smaller number of lung cancer cases but a wider variety of control groups. They measured 13 VOCs (7 aliphatic and 6 aromatic compounds) in 36 early NSCLC cases, 25 controls with clinically stable mild-moderate chronic obstructive pulmonary disease (COPD), 35 asymptomatic smokers with normal spirometry and 50 healthy non-smokers with normal spirometry. The NSCLC, COPD and smoking controls had generally higher levels of exhaled VOCs than the non-smoking controls suggesting, unlike the American group, that smoking status was very important. Although no single VOC could distinguish NSCLC from the other groups, their panel of 13 VOCs could be used to correctly classify 72.2% of subjects as having NSCLC. On the basis of the results the overall sensitivity was 72.2% and specificity 93.6%. All 36 NSCLC cases underwent surgical resection and 26 of the NSCLC patients agreed to have post-operative breath collection. In

this subgroup, post-operatively there was a significant reduction in exhaled isoprene and decane levels. However, the low sensitivity and specificity of the VOCs for such an important disease led Poli *et al* to conclude that 'using VOCs alone in the early detection of lung cancer cannot be recommended at this stage'. [162]

The most recent study, published early this year, looked at exhaled breath from 43 NSCLC patients and 41 healthy controls who were non-smokers all from China. Of the cancer arm, 11 patients received 1 cycle of chemotherapy and the breath sample was obtained 4 weeks after the treatment, and the rest of the cancer cohort did not receive any type of treatment. The patients were fasted overnight and 15 ambient air samples were taken from the room where the patients gave their samples. All the samples were analysed using GC-MS. They identified 2 particular VOCs which were significantly higher in the cancer group compared to the controls ($p < 0.001$). [163]

The studies discussed thus far involve mass spectrometry for the measurement of VOCs. This equipment is expensive (over £300,000) and available only in research facilities. Moreover the technique is often laborious and time consuming needing highly specialised technicians. New techniques to detect VOCs have recently been developed. One such system consists of a colorimetric sensor array which has 36 spots made up of different chemically sensitive compounds. The colours of the spots change depending on the chemical it comes into contact with. A study using this technology was performed by Mazzone *et al*. 143 subjects were included in the study, 49 had NSCLC, 73 had various chronic lung diseases such as COPD, idiopathic pulmonary fibrosis, pulmonary arterial hypertension and sarcoidosis. There were 21 healthy controls. A prediction model was developed using 70% of the subjects, then tested. It was able to predict the presence of lung cancer in the remaining 30% of subjects with a sensitivity of 73.3% and specificity of 72.4%. Although this suggests only moderate accuracy of diagnosis, the study does show the potential of this measuring technique, which is a lot cheaper and quicker to employ than gas chromatography. [164]

Bronchoscopy and screening for lung cancer

The role of bronchoscopy in the diagnosis and particularly treatment of lung cancer is still developing. White light bronchoscopy is used conventionally as a diagnostic tool for staging and obtaining endobronchial samples mainly in symptomatic patients or those with imaging suggesting a central cancer. There have also been studies examining the role of bronchoscopy in the early detection of (central) lung cancer, i.e. before it has become advanced and still in the premalignant cyto-pathological stages.

Autofluorescent Bronchoscopy

The principle behind autofluorescent bronchoscopy is that when blue light (wavelength 380 – 460nm) is shone onto abnormal mucosa there is reduced fluorescence, thus giving potential for abnormal mucosa and thus premalignant lesions to be identified. A review suggests that autofluorescent bronchoscopy (in experienced hands) can diagnose carcinoma *in situ* in 1.6% and moderate and severe dysplasia in 19% of current heavy or former smokers with sputum atypia. Moreover, the pre-invasive lesions found were small with 55% being \leq 1.5mm in greatest diameter. [165] Over 1,000 cases comparing white light and autofluorescent bronchoscopy has shown that 40% of pre-invasive lesions were detected by white light bronchoscopy alone but an average of 80% were detected with the addition of autofluorescent bronchoscopy. [166]

A recent study combining autofluorescent bronchoscopy, sputum cytology and spiral CT surveillance in a very high-risk group was published in 2007. In order to be eligible for screening, participants required two or more of: \geq 20 pack year history of tobacco use; asbestos-related lung disease on CXR; $FEV_1 < 70\%$ of predicted or treated prior aero-digestive cancer, with no evidence of disease for > 2 years. 186 patients were enrolled and 169 completed the baseline tests. 66% of the patients had squamous metaplasia or worse with 13 /169 (7%) diagnosed with lung cancers. Sputum cytology missed 100% of the dysplasia and 68% of the metaplasia detected by autofluorescent bronchoscopy, and failed to detect any cases of carcinoma or carcinoma in situ. Patients who had

pulmonary nodules on spiral CT scan were 3.2 times more likely to exhibit premalignant changes on autofluorescent bronchoscopy and this was statistically significant ($p < 0.001$). The study concluded that autofluorescent bronchoscopy should be carried out on high-risk individuals regardless of the results of conventional sputum cytology. [167]

Other studies have found no increased detection rates of carcinoma *in situ* or dysplasia using autofluorescent bronchoscopy compared to white light bronchoscopy in current and former smokers. [168] This suggests that autofluorescent bronchoscopy is best used in highly selected cohorts of patients, such as those with sputum atypia or if with multiple risk factors for lung cancer and is very operator dependant.

Genetic testing of epithelial cells from endobronchial brushings obtained at bronchoscopy

Spira *et al* investigated whether gene expression profiling of large airway epithelial cells could be used as biomarkers in histologically normal bronchial mucosa obtained by bronchial brushings. 129 current or former smokers who underwent fibre-optic bronchoscopy for a clinical suspicion of lung cancer were included in the study. Each subject was followed up until a final diagnosis was made. 60 were diagnosed with lung cancer and 69 without lung cancer. Lung cancer was diagnosed if bronchoscopy or subsequent lung biopsy yielded lung tumour cells. Subjects were classified as non-cancer if their investigations yielded a non-cancer pathology or if the radiological abnormality resolved on follow-up. From a training set ($n=77$), which represented a spectrum of clinical risk for lung cancer, an 80-gene biomarker panel was identified and was able to distinguish smokers with and without lung cancer. The panel of genes was tested prospectively on the remaining 52 cases, obtaining an accuracy of 83% (80% sensitivity, 84% specificity). A comparison was made with traditional cytopathology of cells obtained by biopsy, washings and brushings of the affected area on bronchoscopy. 32 of the same 60 (53%) lung cancer cases and 5 of 69 (7%) non-cancer pathologies were diagnosed by traditional bronchoscopy techniques. Among the non-diagnostic bronchoscopies ($n=92$), the panel of

gene biomarker's accuracy was 85% (89% sensitive, 83% specific). By combining the tests such that a diagnosis of lung cancer could be made from either traditional cyto-pathology or the biomarker panel, the diagnostic sensitivity improved to an impressive 95% with a 95% negative predictive value. They concluded that gene-expression in cytologically normal bronchial epithelial cells can serve as a lung cancer biomarker. [169]

Conclusions: screening in lung cancer

Lung cancer continues to be a major cause of mortality and morbidity. Currently, only a small proportion of patients presenting to physicians for help, survive more than 5 years, often because the disease is far advanced at presentation. Screening studies using conventional techniques such as CXR and sputum cytology analysis have not shown group mortality benefit. Low-dose spiral CT scans are still being evaluated but advancing technology is allowing a range of biological markers to be tested in the early detection of lung cancer. Better understanding of the biology of lung carcinogenesis and improved techniques in DNA analysis are being combined to develop cost-effective, high throughput technologies that are being applied to readily accessible bio-fluids such as blood, sputum and even exhaled breath. It is likely that combinations of biomarkers, perhaps using combinations of bio-fluids will lead to non-invasive, and hopefully cost-effective screening tools to improve outcomes in this devastating and common disease.

Chapter 3. The effect of smoking in cancer

Introduction

Years of research have shown clear links between tobacco smoking and cancer. Smoking is the single biggest cause of cancer worldwide and accounts for 25% of UK cancer deaths. It has been a causal or contributory agent in many cancers including lung, oropharynx, oesophagus, liver, pancreas, kidney, stomach, bladder, cervix and myeloid leukaemia. Figure 3.1 shows the organs affected.

Why does smoking cause cancer?

Each cigarette contains nearly 60 known carcinogens i.e. tumour initiators, or tumour promoters. Particularly 'strong' carcinogens are the tobacco-specific nitrosamines, such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone also known as NNK and N'-nitrosonornicotine (NNN); other strong carcinogens include polycyclic aromatic hydrocarbons, such as benzo[a]pyrene and aromatic amines. 'Weak' carcinogens like acetaldehyde are even more abundant. The total amount of carcinogen adds up to 1-3mg per cigarette resulting in a cumulative dose effect over a life-time of smoking. [170]

These different agents act on different biological pathways to have their carcinogenic effect. P53 is a tumour suppressor gene that detects any errors when cell division takes place and triggers DNA repair or cell death (apoptosis) if the damage persists. Benzo[a]pyrene, a potent mutagen in cigarette smoke, directly damages p53 rendering it inactive. Identical damage to p53 has been demonstrated in lung cancer patients. [171] Nicotine itself activates protein kinase C which reverses the growth inhibition induced by cellular opioids in lung cancer cells, thus reducing apoptosis and acting as a tumour promoter [172]. In-vitro studies show that nicotine also inhibits apoptosis through multiple intracellular signal transduction pathways such as activation of the protein kinase signalling pathway seen in lung cancer [173].

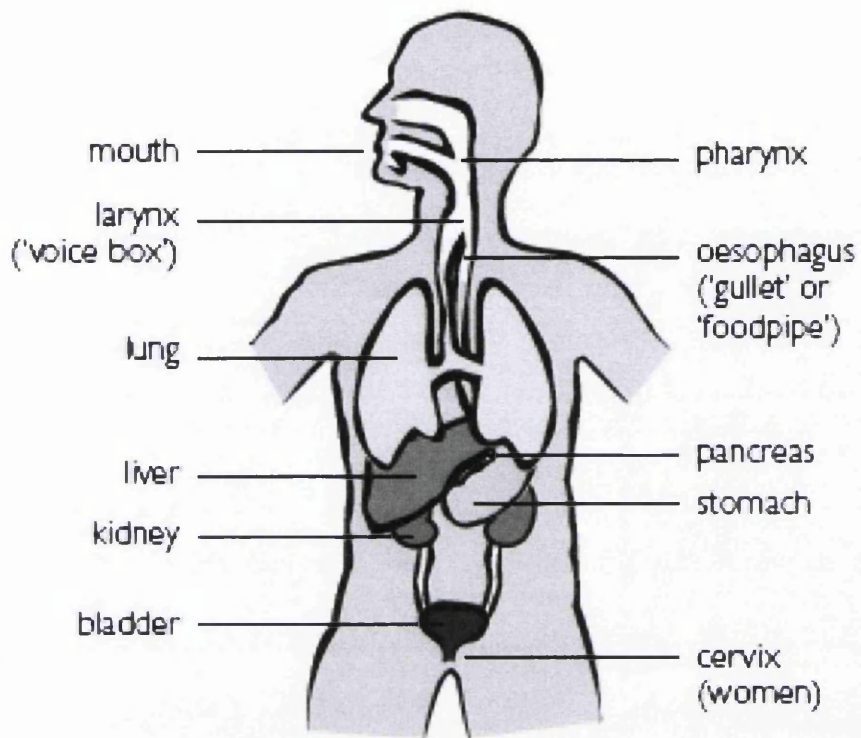


Figure 3.1: The different sites of cancer caused by cigarette smoking (image taken from www.cancerresearchuk.org)

Tobacco-specific nitrosamines such as NNK have induced lung tumours when applied directly to the skin of mice, or injected into the bladders or swabbed in the oral cavities of rats.

Smoking and lung cancer

The most widely studied cancer with respect to smoking is lung cancer. As well as the main cause of lung cancer, smoking status appears to be important after diagnosis. There is increasing evidence to suggest that smoking cessation even at the time of lung cancer diagnosis has a significant beneficial effect.

Smoking and survival in lung cancer

An observational study followed 112 patients with small cell lung cancer. 20 patients had stopped smoking permanently before diagnosis, 35 stopped at diagnosis and 57 continued to smoke. The patients had a treatment regime consisting of chemotherapy +/- radiotherapy. Those who had already stopped smoking had the best survival followed by those who stopped at the time of diagnosis. The group who continued smoking had the worst survival rates with no-one surviving more than 2 years compared with 9-15% in the other groups surviving (disease free) for at least 2-4 years. These results were statistically significant. [174].

A more recent paper in 2005, prospectively recorded smoking status in 311 patients operated for non-small cell lung cancer (NSCLC) by a single surgeon. At surgery, 54% were current smokers, 8% non-smokers, 26% former smokers and 12% were recent quitters. Even in this best lung cancer prognostic cohort, current smoking was an independent predictor of reduced survival ($p=0.001$) (along with older age and presence of lymph node metastases). Compared to current smokers, non-smokers ($p=0.042$), former smokers ($p=0.006$) and even *recent quitters* ($p=0.004$) all had a significantly better prognosis and disease-free survival. See Figure 3.2. These results suggest that it is beneficial to stop smoking prior to surgery for NSCLC and that continued smoking is associated with a poorer prognosis. [175]

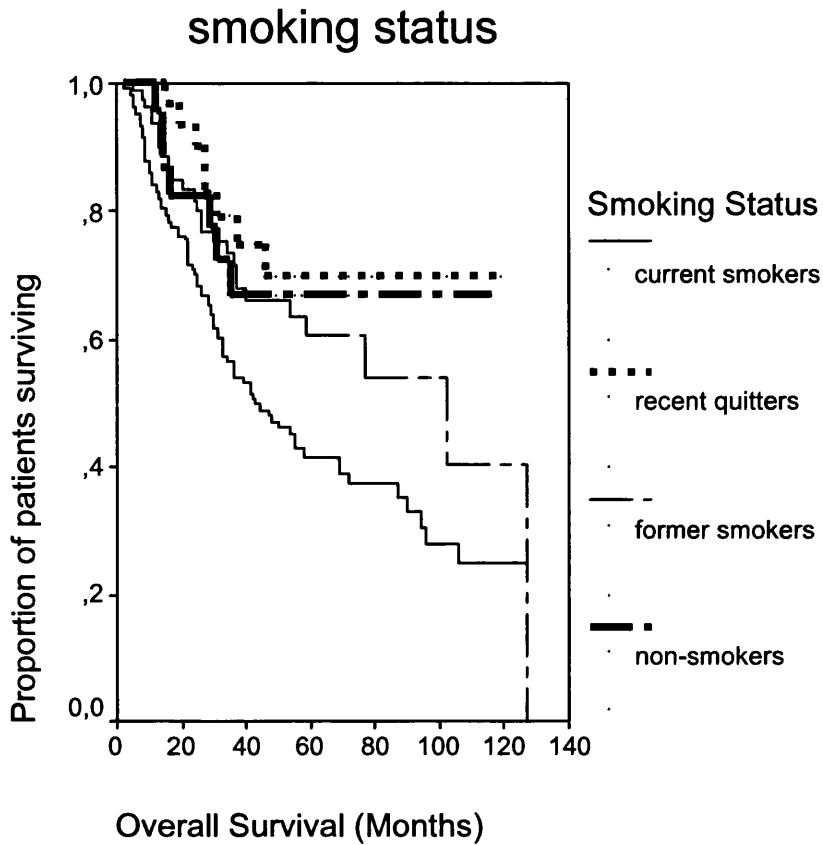


Figure 3.2: Overall survival according to smoking status. The differences between current smokers versus non-smokers ($p=0.0263$), former smokers ($p=0.0109$) and recent quitters ($p=0.0051$) were significant. [175]

A retrospective study looking at a 10 year period followed 237 patients who received radiotherapy alone or chemoradiotherapy for NSCLC and had smoking status recorded. Among those with less advanced disease radiologically (stage I/II), current smokers had a 2-year survival of 41% (median survival 13.7 months) compared to non-smokers 2-year survival of 56% (median survival 27.9 months). Groups were similar at baseline and again this was statistically significant ($p=0.01$) as well as being clinically important. There was no statistical differences among those with more advanced (stage III) disease. [176]

Smoking complicating treatment for lung cancer

Does smoking status influence complication rates associated with chemotherapy and radiotherapy? Monson *et al* investigated the side effects of treatment in 83 patients receiving curative radiotherapy for lung cancer between 1989 and 1993. 20% developed clinically detectable radiation pneumonitis with an incidence which was significantly increased in those with low performance status, low pulmonary function test, co-morbid lung disease, no surgical resection and those with a significant smoking history. [177]

A retrospective review of 186 patients receiving six-cycles of chemotherapy with concurrent radiotherapy during cycle 2 or 3 for limited-stage small cell lung cancer (between 1989 and 1999) found that continuing smokers (42%) did not have a greater incidence of toxicity-related treatment breaks, but those who did continue to smoke and did need a treatment break had the poorest survival outcome. Median survival (18 versus 13.6 months) and 5-year survival (8.9% versus 4%) was greater for former smokers than for continued smokers. [178]

Smoking and quality of life in lung cancer

A large study investigated the quality of life (QOL) experienced by lung cancer survivors, using a lung cancer symptom scale. 1,028 patients were included in the study and the group differences adjusted for age, gender, stage and time of assessment. At time of diagnosis 18% were never smokers, 58% were former smokers and 24% were current smokers. Appetite, fatigue, cough, shortness of

breath, illness affecting normal activities and overall QOL were all statistically and clinically worse in the current smokers group compared with the never smokers. It was concluded that persistent smoking after lung cancer diagnosis negatively impacts QOL scores. [179]

Smoking and other cancers

Continued smoking can impact on other cancers once diagnosed, although this has been less well studied. 610 men with localised prostate cancer were assessed between 1994 and 1997. High-risk cancers (those with PSA greater than 20 nmol/L, Gleason grading greater than 7 or stages T3-4) were more common in the 15% who currently smoked compared to former or non-smokers ($p=0.017$). Outcomes and overall mortality following curative external beam radiation therapy were poorer in the smoking group, although there was no difference in prostate cancer specific mortality. [180]

A small number of studies have looked at the rate of occurrence of second primary tumours in patients initially diagnosed with breast cancer. The frequency of second malignancies consistently appear to be greater in the those who continue to smoke after initial diagnosis with the predominant malignancy being lung cancer. [181, 182] Radiation therapy appears to increase these risks further. [182]

In patients undergoing breast reconstruction for breast cancer, there is a significantly higher risk of mastectomy skin flap necrosis, abdominal flap necrosis and hernia from the donor site in current smokers compared to former or non-smokers, especially in those with more than a 10-pack year smoking history. [183] A case –control study looked at a cohort of patients treated with radiotherapy for Hodgkin's lymphoma. They found that there was a six-fold increased risk of developing lung cancer if the patients had a greater than 10 pack year smoking history compared to those with less than a one pack year history. For those that continued to smoke after diagnosis, the increase in risk of lung cancer was greater as the radiation dose increased compared to those who abstained from smoking. [184]

What potential mechanisms does continued smoking have on cancer cells?

There are biological reasons why continued smoking leads to worse drug toxicity and side effects, including direct damage to lung tissue itself, further impairment of immune function and increased incidence of infection. Other potential hypotheses for why continued smoking is associated with reduced survival and treatment efficacy is that tobacco carcinogens are 'fuelling' the cancer, hence causing more rapid progression and death. Certainly various tobacco chemicals can interfere with biological pathways in which chemotherapy would usually work, hence inhibiting its action. There is potential for nicotine, for example, to decrease the efficacy of various drugs via induction of hepatic enzymes, and thus may have some detrimental role in chemotherapy, although this needs to be investigated further. [185, 186] Many retrospective studies not showing any significant differences in outcome between smokers and non-smokers have been criticised for not validating or even recording smoking status and amount during treatment. [187] Better clinically designed studies, hopefully well designed intervention trials and further biochemical research identifying these cellular pathways are being planned.

Conclusions: smoking and cancer

All health care professionals try to prevent disease wherever possible, and it is very clear that they should be advising all patients of the role smoking has in causing cancer and in encouraging and offering cessation advice to reduce the cancer burden to society. Probably less known are the benefits of cessation in cancer patients. Traditionally, healthcare professionals may not have been active in helping patients with cancer quit smoking perhaps believing that the benefits in quality and length of life are not worthwhile. [188, 189] There is considerable and growing observational evidence to show this attitude is probably wrong. Meanwhile, any research into early diagnosis, screening or monitoring strategies in lung cancer must account for smoking history and validate current smoking status at every assessment for the above reasons.

Chapter 4. Fourier Transform Infra-Red (FTIR) Spectroscopy in Medicine

Introduction

Our first study applies Fourier-transform Infrared Spectroscopy (FTIR) to sputum as a way of diagnosing lung cancer. Jean Baptiste Joseph Fourier (March 21st, 1768 – May 16th, 1830) was a French mathematician and physicist. He has been credited with the discovery of the Greenhouse effect, where gases in the atmosphere are thought to increase the surface temperature on earth. Fourier Transform is also named in his honour. [190] In mathematics, Fourier Transform is an operation that transforms one complex-valued function of a real variable into another. [191]

How FTIR works

FTIR spectroscopy involves shining an infra-red light through a biological sample. Some of the infra-red radiation is absorbed by the sample (specifically the molecular bonds) and some of it passes through (transmitted). The resulting spectrum represents the molecular absorption and transmission, creating a molecular fingerprint of the sample. As with fingerprints, no two unique molecular structures produce the same spectra. The advantage of such a technique is that it can be used to identify any type of compound such as carbohydrates, lipids, proteins and DNA. It can determine the quality or consistency of a sample and also the amount of components in a mixture. Figure 4.1 is an example of an FTIR machine.

The other advantage of FTIR is it allows high-throughput, cost effective analysis of the samples. The analysis of one sample is in the region of a few pence. It is also more sensitive due to the detectors which are employed, and the optical throughput which is higher, resulting in much lower statistical noise levels. The moving mirror in the interferometer is the only continuously moving part in the instrument, thus giving very little chance of the machine breaking down.



Figure 4.1: An FTIR machine

The sample analysis process

The normal instrumental process is as follows.

The source: the infra-red light is emitted from a glowing black-body source and passes through an aperture which controls the amount of energy presented to the sample.

The interferometer: the beam enters the interferometer and undergoes “spectral encoding”. This results in an interferogram signal which then exits the interferometer.

The sample: the beam enters the sample compartment where it hits the molecular bonds of the sample causing them to rotate and vibrate, giving an absorption peak. The rest of the light passes through (transmission).

The detector: the beam finally passes through to the detector for final measurement.

The computer: the measured signal is then digitalised and sent to the computer where it undergoes Fourier transformation (a specific mathematical equation) with the output being an infra-red spectrum for further analysis.

Figure 4.2 represents a schematic of the sample processing.

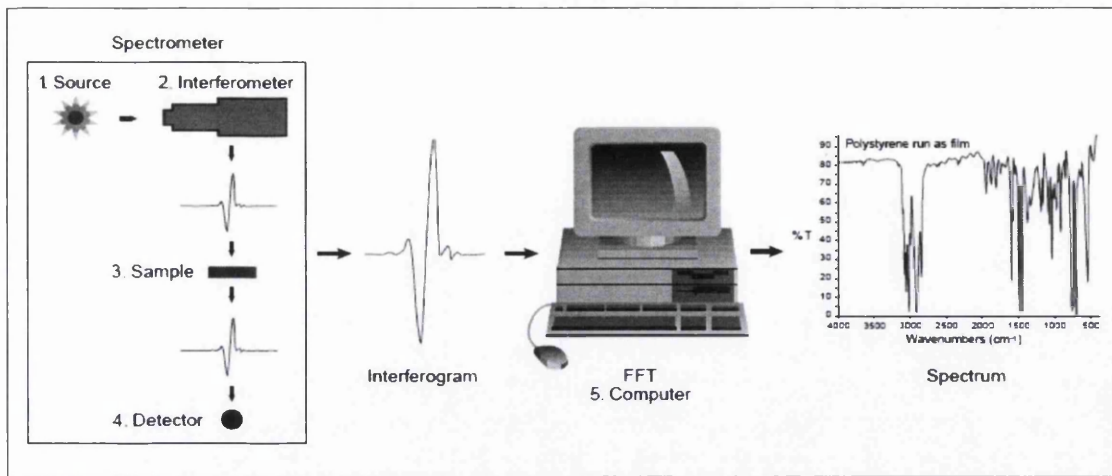


Figure 4.2: The FTIR process (image taken from www.thermonicolet.com)

As there needs to be a relative scale for the absorption intensity, a background spectrum must be measured. This is normally a measurement with no sample in the beam. This can then be compared to the measurement with the sample in the beam to determine the “percent transmittance”. This allows for all the instrument characteristics to be removed so one is left with purely the characteristics of the sample.

Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical method routinely used to analyse interrelationships among large numbers of variables revealing common underlying factors or components and is used in FTIR analysis. The objective of PCA is to reduce the dimensionality (number of variables) of the dataset but retain most of the original variability in the data. PCA is referred to as a data reduction method. PCA examines the correlations between the original variables (in our study the spectra represent the variables) and condenses the information contained within these variables into a smaller group of components with minimal loss of information. Thus, PCA could reduce a large group of individual cases into smaller groups (components) of related cases. The association of a case with each component depends on the correlation values (loadings) calculated by PCA and the variance shared by the component and variable is equal to the square of the correlation. The higher the loading value, the better the description of the variable by the component. The first component represents the best linear combination of variables, similar to a line of best fit, so, the first component may be described as the single best summary of linear relationships within the data. The second component is derived from the proportion of the variance remaining after the first component has been extracted and is the second best linear combination of variables, orthogonal (at right angles) to the first component and so on. There is no inter-correlation between components. PCA is extremely useful for visualizing the relationships between cases as the loadings of cases for two or three components can be visualized using two dimensional (2D) and three dimensional (3D) scatter plots.

FTIR and cancer

The role of FTIR in medicine has been developing over recent years. Its role in lung cancer has been studied on lung cancer tissue and normal lung tissue specimens. Using a sample size of 26 patients, Yano *et al* found that at certain wavenumbers there was significantly higher absorbance found in lung cancer specimens relative to non-cancer control tissue from the same patients. They also found that absorbances for these wavenumbers were higher for adenocarcinoma than for SCC. [192, 193] Wang *et al* reported that there were large differences in peak absorbance at certain wavenumbers between lung tumour and normal lung cells in pleural fluid of the same patients although this was assessed on only 8 subjects. [194]

FTIR has been studied in relation to other cancers such as cervical cancer. FTIR and PCA was applied to the exfoliated cervical cells of 272 patients. Six spectra were performed for each patient and they were assessed and split into two groups. Type 1 exhibited a profile characteristic of normal cervical epithelium. Type 2 exhibited features suggestive of dysplasia or malignant change. Of the 272 patients, 68.6% had type 1 spectra for all six performed, 29.4% had at least one type 2 spectra and 2% were inconclusive. Of the 68.6%, 86% were diagnosed normal by cervical smear with no formal biopsy required. 7% were diagnosed abnormal by biopsy and 5% normal by biopsy and 2% remained inconclusive. Of the remaining 29.4% of which at least one type 2 spectra was obtained, 71% had shown abnormal smear results and went on to have biopsies which confirmed 87% of them as abnormal. Thus the study showed the early potential of FTIR. [195]

A study in 2004 looked at the potential detection of biomarkers using FTIR in cervical cancer and melanoma. Studies were conducted with formalin-fixed biopsies of melanoma and cervical cancer by FTIR to detect common biomarkers, which occurred in both types of cancer distinguishing them from the respective non-malignant tissues. The spectra were analysed for changes in levels of biomolecules such as RNA, DNA, phosphates and carbohydrates (such as glycogen). The results showed promise for cervical cancer but not so

for melanoma. [196] More recently, a group from King's College, London further looked at FTIR in cervical cancer. Using a cohort of 53 patients, they compared the cervical smear test and FTIR technique against the findings of the colposcopists (gynaecologist). The smear test achieved an overall classification of 43%, whereas FTIR achieved a rate of 72%. [197] Further work on cervical smears was published in 2008. 800 cervical scrapings were analysed using FTIR and by cytology which is the gold standard. FTIR was successful at distinguishing between cancer and normal tissue with a sensitivity of 85% and specificity of 91%. [198]

In 2002, Lasch *et al* analysed FTIR spectra performed on 26 patients tissue samples which were pathologically diagnosed as colorectal adenocarcinoma. Using multivariate statistical techniques such as hierarchical cluster analysis (HCA) they were able to correctly characterise 95% of the cases correctly. [199]

FTIR on cancerous breast tissue has shown distinctive spectra different to that of normal breast tissue. Using the spectra, close to 100% diagnostic accuracy could be obtained and it was also able to discriminate between subtypes of breast cancer, such as apocrine, tubular, intraductal and mucinous carcinomas and invasive infiltrating ductal carcinomal tissues. [200]

FTIR has also been evaluated in prostate cancer. It has been used on prostate biopsies and has shown that it is possible to distinguish between prostate confined disease and extra-prostatic cancer. This is not possible with existing criteria and is of importance especially when considering treatment regimes. FTIR with PCA was also able to distinguish between low and higher grade cancers with a sensitivities and specificities as high as 83.6% and 86% respectively. [201, 202]

The technique has also been used on oesophageal, gastric and oral tissue with varying results. [203-206]

FTIR and lung disease

In order to use FTIR effectively, suitable biofluids are required for sampling. In lung disease, sputum is readily available and this has been assessed in chronic obstructive pulmonary disease (COPD) albeit with small numbers of study patients. 15 patients with COPD were included in the study (FEV₁ of 40 – 65% predicted, 10 – 40 pack year smoking history, 2 of which were current smokers). The control group consisted of 15 healthy volunteers who were never smokers and had no evidence of respiratory disease. The results showed no significant difference in spectra between current smokers and ex-smokers in the COPD group. Significant differences were noted between the spectra of COPD and healthy patients in three particular regions, amide A, amide II and glycogen rich region. [207] This was the first study using FTIR on sputum. A different study looked at pleural fluid, identifying spectral differences between the cells for pleural fluid of patients who were normal, had lung cancer and those with tuberculosis. Eight patients were included in the study and the FTIR spectra on the cell pellets were noted to be different between the 3 groups especially the lung cancer group at certain wavenumbers. [194]

In patients with Cystic Fibrosis (CF), sputum production is usually a significant symptom due to the presence of bronchiectasis (dilatation and damage to bronchial airways, which usually results in chronic sputum production). CF patients often get different respiratory infections as their disease progresses and identification is important as aggressive, early treatment prevents further damage. FTIR has been studied on the sputum of 150 CF patients to identify if there are spectral differences that can help identify certain infective organisms. The clinical isolates were identified using standard guidelines and then the samples underwent FTIR. There were spectral differences and this allowed the correct identification of organisms such as *Pseudomonas aeruginosa*, *Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans*, *Acinetobacter* spp., *Ralstonia pickettii*, and *Burkholderia cepacia* complex bacteria with up to 98% accuracy. [208]

Conclusion : FTIR in medicine

There are a small but growing number of studies, which have shown the potential of FTIR in medicine. However, the majority of these have involved mainly biopsy tissue obtained in an invasive manner. The main aim of the technology is to facilitate diagnosis and in terms of cancer the technology could be used to detect changes at an earlier more treatable stage. This is particularly important in lung cancer for reasons already discussed. There have been only a few attempts at studying FTIR in lung cancer. One such study has used FTIR analysis on lung cancer tissue and one on pleural fluid of lung cancer patients; Two studies have looked at sputum in COPD and CF patients, respectively, but no-one, has yet assessed FTIR in the sputum of lung cancer patients.

Chapter 5. Volatile organic compounds (VOCs) in lung disease

Introduction

Our third study measures volatile organic compounds (VOCs) in subjects with lung cancer and controls. VOCs are organic chemicals that easily vaporise at room temperature. [209] As they contain the element carbon within their molecular structure they are called organic. They have no colour or taste and include a wide range of individual substances, such as hydrocarbons (e.g. benzene and toluene), halocarbons and oxygenates. Examples of VOCs include methane, which is an important hydrocarbon and has an environmental impact principally with its contribution to global warming and the ozone. Some VOCs are harmful, including benzene, polycyclic aromatic hydrocarbons (PAH) and 1,3 butadiene. Benzene, toluene and xylene may increase susceptibility to leukaemia if there is extensive prolonged exposure. PAHs in particular, can cause cancer as mentioned in Chapter 3. There is also an apparent correlation between butadiene exposure and higher risk of cancer. Sources of 1,3 butadiene include manufacturing of synthetic rubbers, petrol driven vehicles and cigarette smoke.

VOCs and lung disease

The current studies looking at the role of VOCs in diagnosing lung cancer have been summarised in Chapter 2. As well as the potential for certain patterns of VOCs developing as a result of certain disease processes, VOCs within the ambient air in certain environments may have a causative effect in respiratory disease. Inhalation of significant amounts of VOCs may lead to respiratory problems due to irritation of the airways and can cause symptoms such as one gets with asthma.

VOCs and airway disease

Various studies have shown a correlation with increased airway hyper-responsiveness in workers exposed to VOCs especially if they are smokers and have a background of atopy. [210-213] A study looking specifically at the role of VOCs on physician diagnosed asthma showed that there was a significantly higher odds of diagnosis following exposure to aromatic compounds. In those who were not officially given a diagnosis of asthma, exposure to aromatic compounds and chlorinated hydrocarbons significantly increased the odds of one or two wheezing attacks. [214]

VOCs and Infection

Phillips *et al* has studied panels of VOCs for the early detection of lung cancer (see Chapter 2). He recently expanded on the potential of the technique to look for alterations in VOCs for patients with suspected pulmonary tuberculosis (TB). Statistical analysis consisting of linear discriminant analysis and principal component analysis built various receiver operator curves that eventually allowed a 100% sensitivity and specificity when distinguishing between those suspicious of TB and those who were healthy controls. Pattern recognition analysis was also used to distinguish those who were sputum culture positive (i.e. infective) TB cases with 82.6% sensitivity and 100% specificity. The study conclusions suggested that VOCs could be used in pulmonary TB to distinguish between well vs. sick (i.e. healthy controls vs. hospitalised due to suspicion of TB) and non-infective vs. infective (i.e. those with negative vs. positive sputum cultures for TB). [215]

As the understanding of VOCs becomes clearer then one can consider and investigate the relationship between VOCs in interstitial lung disease, chronic obstructive pulmonary disease and bronchiectasis, especially in exacerbations of such diseases.

VOCs and other cancers

The early mutagenic and potential carcinogenic role of reactive oxygen species and oxidative stress is discussed below. In some cancers such as breast cancer, it is thought that as well as increased oxidative stress being involved in carcinogenesis, the presence of breast cancer itself causes a further increase in oxidative stress. [216] Phillips *et al* used this theory and assessed exhaled VOCs in 51 women with breast cancer and 42 healthy controls and achieved a sensitivity of 94% and 74% specificity using a panel of 8 VOCs. When assessing 50 patients with abnormal mammogram but no cancer on biopsy the sensitivity dropped to 63% and specificity improved to 84%. [217] Further re-evaluation with a reduced panel of 5 VOCs, showed that when comparing the cancer group with the healthy control, specificity was improved to 84% whilst maintaining the high sensitivity (94%). [218]

Potential mechanisms of VOCs production in breath

Only recently has the pathophysiological processes of VOC production emerged due to better understanding of the mechanisms behind VOC synthesis and clearance. Largely, VOC production revolves around production of reactive oxygen species (ROS).

Reactive oxygen species

ROS are ions or very small molecules that include oxygen ions, free radicals, and peroxides, both inorganic and organic. They are highly reactive due to the presence of unpaired electrons on their outermost electron shell. ROS are formed as a natural byproduct of normal oxygen metabolism and have important roles in cell signaling. ROS levels increase dramatically due to environmental stress such as ultraviolet light / heat exposure and cigarette smoke which can result in significant cell structure damage. This is known as oxidative stress. ROS can also be formed by ionizing radiation, which can penetrate cells and create ions in the cell contents which in turn can cause permanent alterations in DNA. [219]

The general harmful effects of ROS on the cells are:

- Damage of DNA and RNA
- Oxidations of polyunsaturated fatty acids in lipids
- Oxidations of amino acids in protein
- Oxidatively inactivating specific enzymes by oxidation of co-factors

ROS are constantly being produced in the mitochondria and leak into the cytoplasm where they cause damage to proteins, polyunsaturated fatty acids and DNA by peroxidation. Studies support the experimentally based notion of oxidative DNA damage as an important mutagenic and apparently carcinogenic factor, however the proof of a direct causal relationship is still lacking. VOCs such as alkanes and methylated alkanes in breath are apparent markers of oxidative stress, being produced as a result of lipid peroxidation from ROS.

Lipid Peroxidation

Lipid peroxidation refers to the oxidative degradation of lipids. It usually affects polyunsaturated fatty acids as they have multiple double bonds in between which lie methylene -CH₂- groups that possess especially reactive hydrogen. During the process, free radicals 'steal' electrons from the lipids cell membranes, resulting in cell damage. A free radical chain reaction then causes the process to proceed. The process takes place in three steps:

- 1) Initiation – this is where a fatty acid radical is produced with ROS being the most notable of initiators. The ROS, typically OH[•] combines with the hydrogen atom to make water and fatty acid radical.
- 2) Propagation – the fatty acid radical is unstable, thus reacts readily with oxygen, producing peroxy-fatty acid radical. This new product is also unstable and reacts with another free fatty acid producing a free fatty acid radical and a lipid peroxide or cyclic peroxide if it had reacted with itself. Since the new free fatty acid radical reacts in the same way, the cycle continues.

3) Termination – the process is called a free radical chain reaction as when a free radical reacts it always produces another free radical. The only time this will stop is if two radicals react and produce a non-radical molecule and this can occur when there is a high concentration of radical species thus giving a high probability of two radicals colliding. [220]

If the chain reaction is not terminated early enough, then damage to the cell membrane ensues which consists mainly of lipids. The end products of lipid peroxidation may be carcinogenic. [221] Figure 5.1 is a schematic of lipid peroxidation.

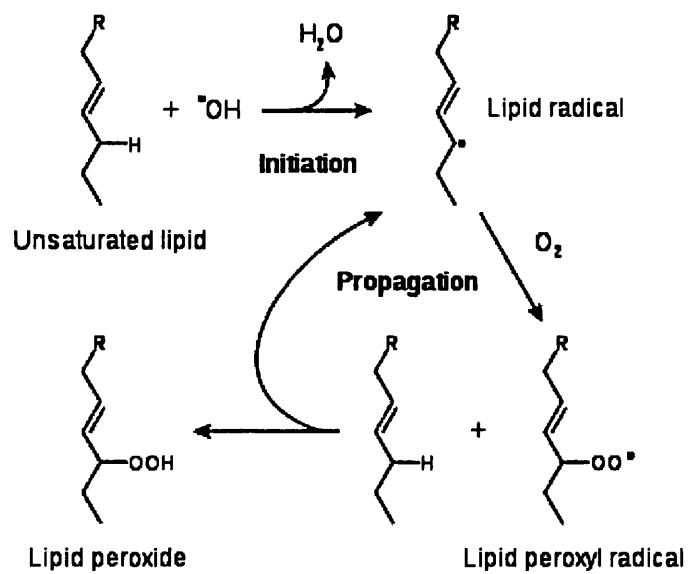


Figure 5.1: Lipid peroxidation (image taken from www.wikipedia.org)

Oxidative stress in cancer patients

It is thought that not only does oxidative stress have a role in the aetiology of cancer, but in some cancers, such as breast, the disease itself causes increased oxidative stress, and is thus self-perpetuating. [216] Emerging research suggests a similar effect in lung cancer. [222-224] Potential mechanisms for increased oxidative stress in patients who already have cancer have been suggested. The first includes altered energy metabolism which may result in symptoms such as cachexia / anorexia, nausea and vomiting, which leads to reduced nutritional intake of glucose, proteins and vitamins, and this leads to accumulation of ROS. [225] A further mechanism is one of chronic activation of the immune system and thus excessive production of pro-inflammatory cytokines, which may in turn lead to increased ROS production. [226] Interestingly cachexia has also been shown to be associated with high levels of pro-inflammatory cytokines. [226-228] 60% of lung cancer patients present with cachexia, which is characterised by progressive weight loss with predominantly depletion of skeletal muscle and to a lesser extent, reduced adipose tissue, reduced dietary intake and poor performance status. [229] This increases to greater than 80% as the disease advances. [230] There have been several links between increased oxidative stress and cachexia including the role of TNF- α inducing oxidative stress, a potential role of IL-6 and high levels of glycolytic activity and lactate production. [225, 231-236]

As increased oxidative stress appears to be increased in some cancers, and VOCs such as alkanes are markers of oxidative stress then one can postulate these levels will be higher in the breath of cancer patients. [237-241]

VOCs in exhaled breath

Alkanes, such as ethane and pentane are generated by lipid peroxidation of polyunsaturated fatty acids in cell membranes and these are excreted in breath. Breath methylated alkanes may be the result of the same process. [239, 240, 242] The alkanes produced by lipid peroxidation are then metabolised by

cytochrome P450 (CYP)-mixed oxidase enzymes and result in alkyl alcohols. [243]

A number of studies have shown certain CYP enzyme genotypes, such as CYP1A1, to be activated in lung cancer. [244-247] There are genetic polymorphisms in the inducibility of CYP1A1, with some evidence that high inducibility is more common in patients with lung cancer. [248] For example, two enzymes of cytochrome P450, CYP1A1 and CYP1A2, are inducible by tobacco carcinogens, and animal studies evidenced a genetic polymorphism of CYP1A1 associated with tumour occurrence after administration of a polycyclic aromatic hydrocarbon. [244] Induction of CYP activity results in accelerated metabolism of a number of drugs and the activation of some procarcinogens. CYP1A1 has been the most studied genotype with relation to lung cancer and has been associated with higher disease recurrence rates and poorer survival rates. [249] Thus activation of CYP enzymes in individuals with lung cancer may accelerate the degradation of VOCs that are produced as a result of oxidative stress. This provides the basis behind exhaled breath testing and is represented by the following schematic, Figure 5.2.

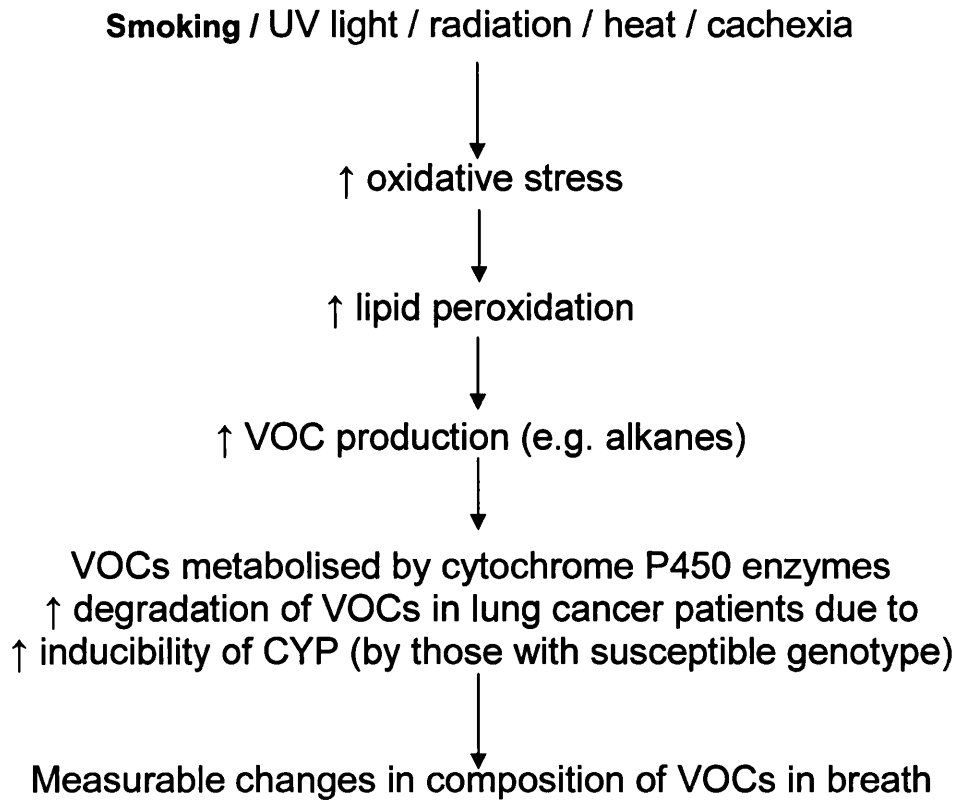


Figure 5.2: Schematic of potential mechanism of VOCs in lung cancer

Phillips *et al* suggested a schematic of potential sources of endogenous / exogenous VOC production as shown in Figure 5.3. [250] In Figure 5.4, I have incorporated the potential mechanisms of VOCs production in lung cancer patients discussed in this Chapter into the schematic.

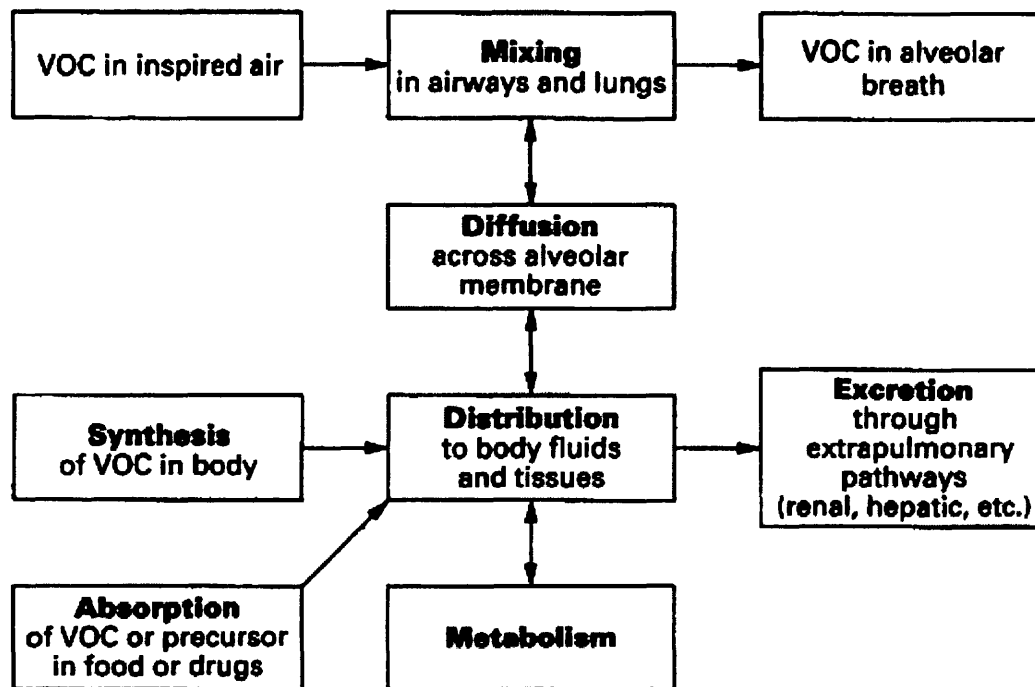


Figure 5.3: Potential sources of VOCs in alveolar air. This figure shows that VOCs in the breath may be derived from sources either inside the body (as products of metabolism) or outside the body (VOCs in the inspired air, food, and drugs). [250]

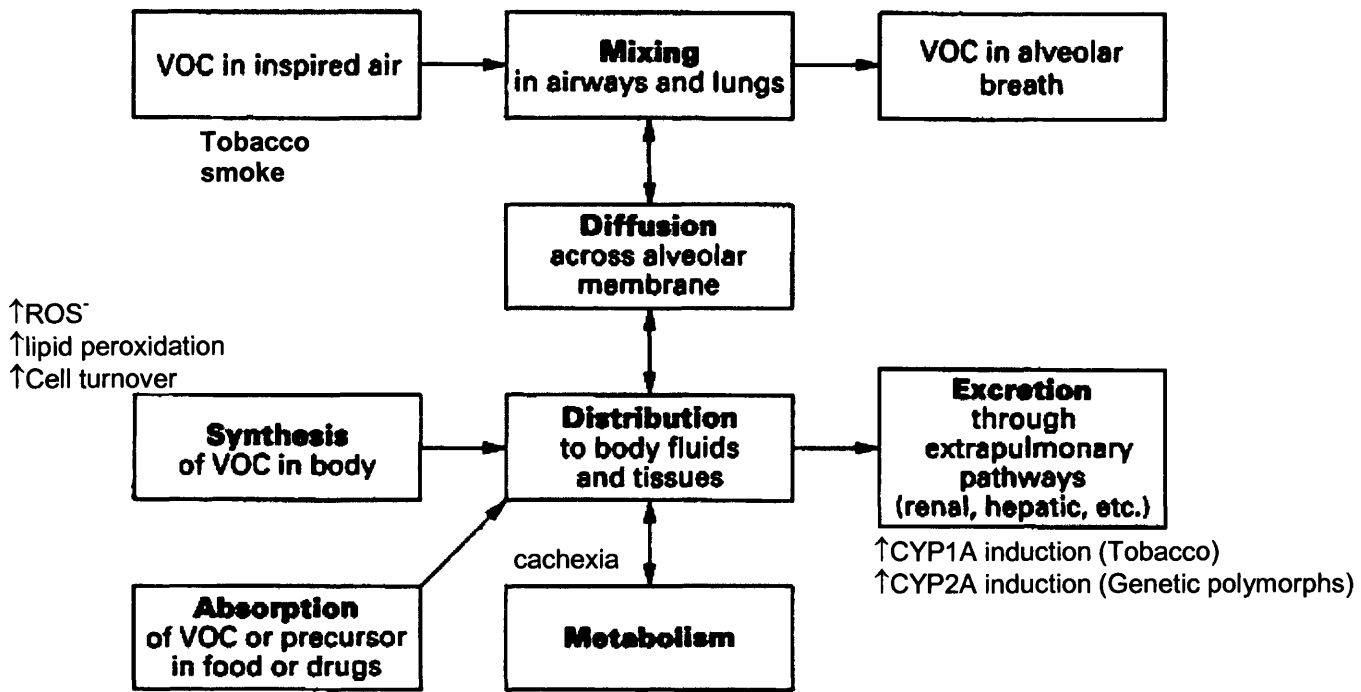


Figure 5.4: Potential sources of VOCs in alveolar air of lung cancer patients with possible mechanisms of production and how the illness may affect them

Different techniques used to trap then measure Volatile Organic Compounds

Mass Spectrometry

Many of the studies using VOCs discussed in Chapter 2 involved the collection of exhaled breath and then analysing the breath using gas chromatography – mass spectrometry (GC-MS). The results of these studies are promising but the accuracies are not yet high enough to be clinically useful. The advantage of the GC-MS system is that they are sensitive and can detect specific VOCs and measure their concentrations at very small levels. The GC-MS systems themselves are currently expensive and analysis can be laborious and time-consuming. Also the breath has to be collected and then transported to the GC-MS machines which can lead to sample storage / leak and degradation problems en route also negating the advantage of patient nearside testing favoured by patients and clinicians.

Gaseous chemical sensing devices / electronic nose (eNOSE)

Gaseous chemical sensing and identification devices such as eNOSEs are able to detect a single (or pattern of) odorant molecule(s) such as VOCs, mimicking human olfaction. The premise of most of these systems is that absorption of gases onto the sensor system causes a change in the conductivity, mass, vibration, or colour of the system, thus altering its' output. This is then translated into a digital value (a 'smellprint'). Figure 5.5 is a representative 'smellprint' of VOC from a lung cancer patient.

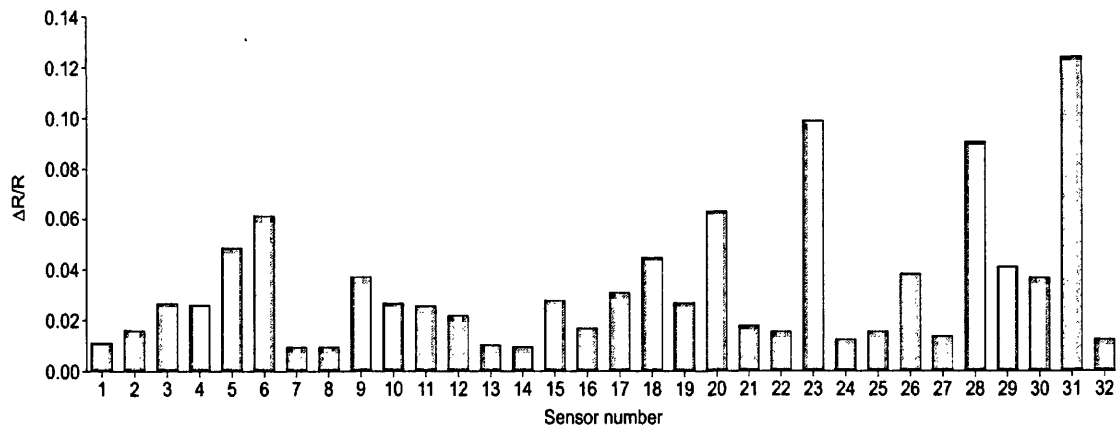


Figure 5.5: 'Smellprint' derived from eNOSE from exhaled VOCs from a lung cancer patient [251]



The system can be set up to consist of an array of sensors, which can be tuned depending on the task. This method has been used in studies looking at lung cancer detection from exhaled breath with varying success. The first studies in 2003 and 2005 looked at variation in oscillating sensor frequencies and the reversible change in resistance across the sensor, respectively. [252, 253] The first such study identified a 'smellprint' which characterised lung cancer and control cohorts with 90% accuracy. The subsequent model correctly classified 100% of cancer patients and 94% of controls. [253] Machado *et al* used a 32 sensor eNOSE and found differences in the 'smellprint' of lung cancer subjects compared to controls, but no signal differences between patients with differing disease severity and no confounding effect of cigarette smoking or co-existent airways disease (asthma or COPD). [252] Then in 2007, Mazzone used the colorimetric sensor described in Chapter 2. [164]

More recently, a study compared 10 patients with lung cancer, 10 COPD patients and 10 healthy controls, and using an eNOSE was able to discriminate each of the cohorts. [254] This has been developed further looking at non-lung cancer cohorts, concentrating on asthma, COPD, smoking controls and non-smoking controls using eNOSE. The asthma and COPD cohorts could be distinguished with 96% accuracy, as could the asthma and smoking and non-smoking controls. Interestingly the COPD group could not be distinguished from the non-smoking controls. [255]

The advantage of this technique of detection over GC-MS is that it is relatively inexpensive, and easy to use as a point-of-care test. However and importantly, they may not be sensitive enough to detect all the potentially important VOCs in breath and it is not possible to identify specific chemical compounds using this technique.

Ion Mobility Spectrometry (IMS)

IMS is based on ionisation of gaseous molecules that are separated by short impulses in drift tubes of a few centimetres. The small electrical current, which is then, generated at a Faraday plate (detection plate) forms the spectrum of

the running time of the ions. This technique allows a ten-fold higher detection rate of VOCs. It offers an immediate two-fold separation of VOCs with visualisation in a three-dimensional chromatogram (separating VOCs by retention time, drift (mobility) and concentration). This method has been evaluated in a study of 32 patients with histologically proven lung cancer (24 men, 8 women), with a mean age of 65.1 (+/- 9.6) years and 54 healthy controls. Of the lung cancer cohort, 17 patients were former smokers, 6 had never smoked and 7 were current smokers. Using this method, a combination of 23 VOCs from exhaled breath were able to correctly classify lung cancer patients or healthy controls with 100% accuracy. The method is quick and reliable based on early data; however it does not give any clues as to the identity of the molecules which the VOCs actually represent. [256]

Other breath tests for lung cancer:

Exhaled breath condensate (EBC)

EBC is a biological fluid that mainly consists of water, but also contains small droplets of airway lining fluid (epithelial lining fluid (ELF)). The advantage of EBC is collection of the samples is also completely non-invasive. The EBC can then be used to look for any changes in DNA. An example of this technique involved patients breathing tidally through a mouthpiece for 20 minutes with the condensate collection taking place on ice at -20°C. The patients were informed to swallow any saliva produced. The samples were then stored in 1.5ml tubes at a temperature of -70°C. [257]. This particular study involved 30 patients with NSCLC who were compared with a healthy control group of 20 subjects. The results showed 89% of the study group having genetic alterations in the DNA from the EBC detected against 35% of the healthy controls. These genetic alterations consisted of loss of DNA heterogeneity and microsatellite instability. It is thought these microsatellite alterations (MA) take place on chromosome 3p where tumour suppressor genes are located. The same research group found that not only are there significantly more MAs in the DNA of NSCLC patients compared to healthy controls, but also that the MA profile from the EBC

corresponded to that of the actual tumour tissue of each patient with NSCLC.

[258]

Two further potential biomarkers in EBC have recently been studied. Cigarette smoke has been thought to be the main factor causing the inflammation of the airways described in lung cancer patients. Two key components thought to be involved in this process are cyclooxygenase (COX) and survivin, a COX-2 dependent factor of apoptosis. In a study comparing not only lung cancer versus healthy controls but also looking at the smoking history and status, higher levels of survivin and COX-2 were found in the EBC of NSCLC compared to healthy smokers and non-smokers. There were also higher VOC levels noted in the smokers (containing both healthy and lung cancer patients) and ex-smokers, compared to the non-smokers, and a positive correlation between EBC survivin and COX-2 levels and the number of cigarettes smoked (pack/yr). There was also a correlation identified between levels of the two markers and progression of disease. Thus using EBC, survivin and COX-2 are thought to have potential at detecting early lung cancer in the smoking population. [259]

p53 mutations have also been detected in EBC. In a small study, p53 mutations at exons 5-8 were identified in 4 out of 11 (36.3%) of NSCLC patients and none of the healthy volunteers. [260]

Thus the use of EBC has identified various potential markers, however all of the studies so far have involved a small number of patients and further analysis is required. The technique is non-invasive thus amenable to patients but does require some time in terms of sample collection. EBCs contain both volatile and non-volatile substances. Volatile substances that are breathed out in a gaseous state can be dissolved in condensed water during EBC collection depending on their physico-chemical properties. [261] Non-volatile substances, such as salts and proteins, are mainly expired in small droplets, and further diluted with exhaled water vapours. It is thought that the droplets are formed as a result of random convective processes, and may not be directly related to water vapour production. This variability in droplet dilution as well as questions over the

source of the droplets (i.e. not necessarily from the epithelial lining of the airways) raises questions over EBC interpretation. Reported increases in EBC concentrations may reflect proportionate increases in the total volume rather than the concentration of ELF droplets in the collected samples. [261, 262]

Conclusion: breath sampling for VOCs and other biomarkers in lung disease

When considering a test to screen or diagnose a condition, a very important consideration is what will be the most comfortable and acceptable test for patients to endure. Breath sampling is a non-invasive procedure that is comfortable for nearly all patients. There are firm biological reasons why one expects differences in the exhaled breath of patients with lung cancer when compared to their healthy counterparts. More work needs to be done in refining and testing the new technology so eliminate confounders and improve accuracy. Studies comparing ways of analysing exhaled breath continue to try and answer many remaining questions and hopefully to find a modality which will successfully show differences between lung cancer and control or high-risk patients to a level of ease, cost and accuracy that would satisfy our screening criteria.

PART 2: METHODS

Chapter 6. Biological screening for lung cancer with sputum

Background

We obtained approval from the Local Research and Ethics Committee (LREC), (05/MMW01/75) and Hywel Dda Trust Research and Development Unit. Site-specific approvals (Bridgend and Swansea) were obtained as the study progressed. Recruitment took place between 2007 and 2008 across 3 hospitals, Prince Philip Hospital (Llanelli), West Wales Hospital (Carmarthen) and Princess of Wales Hospital (Bridgend). All sample processing and analysis was performed at Swansea and Aberystwyth Universities.

During the recruitment period, the study was awarded a grant from the Welsh Office of Research and Development (WORD) (Application no. H07-3-31) and has also been registered with the UK Clinical Research Network (UKCRN) portfolio database (UKCRN ID 4682), the Wales Cancer Trials Network (WCTN) portfolio and was listed with the National Cancer Institute (ID: NCT00899262) in the United States (<http://clinicaltrials.gov>).

MATERIALS AND METHODS

a) PATIENTS

Patients were recruited predominantly from Respiratory Out-Patient clinics although a small number were recruited after they were admitted to hospital with an acute illness. We approached any patient in whom the clinical diagnosis of lung cancer was being considered. These included those with a CXR that was abnormal and those with particular symptoms (e.g. haemoptysis and normal CXR) for which a bronchoscopy was requested to help diagnose or exclude lung cancer. CXR abnormalities included an opacity or mass within the lung fields, enlarged hilar, unresolving consolidation, unexplained lobar or complete lung collapse or cavitation. During the first consultation the patient was made aware that they could have a tumour although other diagnoses were possible and thus further investigation was needed. At the end of this first clinical consultation, the outline of the study was discussed and patients were given a Patient Information Sheet (PIS) with further details of the study and what it would entail as a participant within it. The PIS included information with regards to the sputum samples if they agreed to be part of the study (See Appendix 1). We excluded any patient who refused or did not have capacity to provide written consent (See Appendix 2), anyone under the age of 18, and any patient in whom a bronchoscopy was not planned as part of the diagnostic process. See Figure 6.1.

Patient Flow Chart

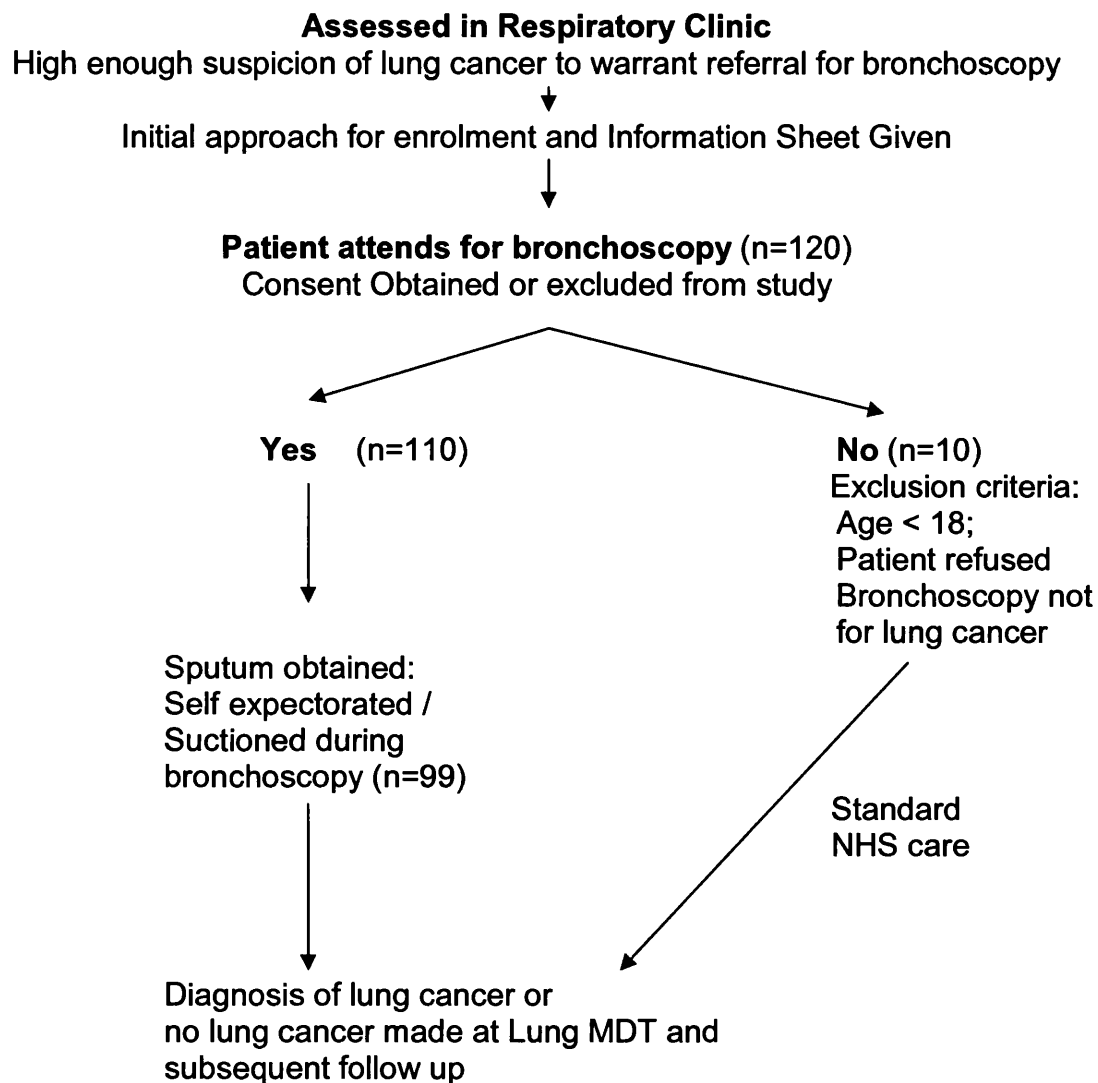


Figure 6.1: Patient flow chart for biological screening for lung cancer with sputum

All participants had at least 24 hours to read their PIS before attending for their bronchoscopy and before signing and dating the Consent Form; a copy of this Consent Form was given to the patient, one placed in the medical notes and a copy kept in the master research file. Their General Practitioner (GP) was then also informed of their patient's participation (See Appendix 3).

Data and sample collection

Sociodemographic and medical data was collected at initial interview and cross-checked from hospital notes. This included: past medical history (including previous cancers), current prescribed medication, smoking history, current smoking status, family history of cancer, and any recent respiratory tract infection. Specific symptoms were noted: weight loss, hoarse voice, haemoptysis, cough, shortness of breath or chest pain. We also recorded CXR findings, FEV₁, WHO performance status and current or previous occupation.

Smoking status was validated using exhaled carbon monoxide (CO) levels. The patients were asked to take a deep breath in and hold at peak inspiration for 15 seconds, then to exhale slowly into the CO monitor (Bedfont-Micro Smokerlyzer ®). The peak number registered on the CO monitor was recorded. Those reporting smoking or anyone with a CO level ≥ 10 ppm (particles per million) were considered current smokers. [263] Ex-smokers were those reporting smoking no cigarettes within the last 12 months and having exhaled CO (eCO) levels <10 ppm. Never smokers were those reporting less than 100 cigarettes in their lifetime, none within 12 months and having eCO < 10 ppm.

See Appendix 4 for data collection sheet.

A universal container was provided for the patient to cough sputum prior to bronchoscopy. All patients were fasted as per protocol for bronchoscopy (6 hours). The self-expectorated sputum sample was frozen immediately on dry ice.

The patient then underwent bronchoscopy. They had an intravenous cannula inserted through which they had sedation (midazolam). 2% Lignocaine was sprayed within the nasal area and oropharynx and then they had further lignocaine injected through the cricoid membrane in order to anaesthetise the proximal trachea and vocal cords. In a small number of patients, endobronchial anaesthesia was performed rather than trans-cricothyroid, due to patient preference, recent haemoptysis, thick neck, enlarged thyroid or complex anatomy. When the patient was sedated adequately they were usually laid flat (supine) and bronchoscopy commenced via nasal or oral approach according to clinical discretion. During the procedure, airway secretions were suctioned by nursing staff using a standard suction yankeur (Pennine Healthcare, London, UK) and placed into a universal container. We therefore obtained endobronchial secretions for those patients who were unable to self-expectorate any sputum. These samples were frozen immediately on dry ice. It was noted whether the sputum samples were self-expectorated (n=42) or taken through the bronchoscope.

The bronchial tree was inspected and diagnostic samples taken for clinical purposes. Endobronchial appearances were coded in the research database as:

1 = normal

2 = external compression of bronchus/ trachea from the tumour

3 = abnormal mucosa

4 = definite tumour endobronchially

Obtaining diagnosis

All the patients had a contrast-enhanced CT scan of thorax and abdomen +/- brain. Clinical details, radiology and cyto-histopathology results were discussed at each hospital's Lung Cancer Multi-Disciplinary Team (MDT) meeting. Those patients with confirmed lung cancer were split according to histological subtype (small cell or NSCLC) and extent of disease (TNM staging). The clinical management plan was then formulated which took the course of potential

surgery, radical radiotherapy, palliative radiotherapy, chemotherapy or best supportive care according to current guidelines. [46]

Patients classed as not having lung cancer had to have a normal bronchoscopy and radiological abnormalities that resolved, with no evidence of cancer at 1 year follow-up. Because of their symptoms and usually smoking history they were deemed as 'high-risk' compared to the background population.

Healthy control cohort

In order to identify what a "normal" sputum FTIR spectra would represent, we recruited healthy individuals, that were a mixture of current, former and never smokers, who did not have any known respiratory disease nor symptoms suggestive of respiratory disease or intercurrent chest or nose infections. They were also excluded as controls if they had previous or any type of current cancer. Figure 6.2 illustrates the screening questions proforma we used for healthy controls.

If they fulfilled the inclusion criteria for a healthy control they were given a Healthy Control Information Sheet (Appendix 5) and given at least 24 hours to consider participation. Once written informed consent was obtained, they were asked to self-expectorate sputum into a universal container, which was then immediately stored at -80°C.

Do you smoke?

Do you currently or have you ever smoked?

Have you smoked for a period longer than 10 years?

Have you lost weight recently (unintentionally)?

Are you troubled by any of the following symptoms at the moment?

-cold or flu

-chest pain

-wheeze (including at night)

-cough

-increased sputum or spit production

-breathlessness

Do you take any inhalers or other medications for your chest

Do you have any respiratory disease or history of this?

Have you ever had cancer of any kind?

Figure 6.2: Screening questionnaire for the healthy control cohort for biological screening for lung cancer with sputum

b) PROCEDURE

Sputum processing for FTIR

Preparation of mucolytic agent

Whilst stored, the unprocessed sputum samples were kept in a freezer (New Brunswick Scientific, New Jersey, USA) at -80°C. The samples were frozen for between 1-60 days before being processed. They were defrosted at room temperature for 12-24 hours. The sputum / bronchial cells were isolated by breaking down the mucus with a solution consisting of 2.5g Dithiothreitol (DTT) in 31ml Cytolyt (Fluka Biochemika Sigma-Aldrich Chemie GmbH Switzerland) mixture which digested the mucus in the specimen. The lifespan of the mixture is only a few days and it had to be kept refrigerated.

Preparation of sputum cells for FTIR

15ml Cytolyt was added to the universal container containing the sputum sample with a pipette and to this 0.5ml of DTT working solution was added. The sample was then placed on the Vortex (IKA Vibrax VXR Basic, IKA ® Werke GmbH & Co, Staufen, Germany) for approximately 15 minutes in order to obtain a sample of fluid consistency. The sample then underwent centrifugation (Hettich Rotina 38, GmbH & Co, Tuttlingen, Germany) at 3000 rpm for 10 minutes to concentrate the cells. The supernatant was poured off leaving a pellet of cells. A proportion of cells were pipetted into a 1.5ml Eppendorf tube. These pellets were freeze-dried over night, diluted in 200 μ L of sterile distilled water, agitated for 5 minutes and split into 20 μ L aliquots, which were frozen immediately in liquid nitrogen and stored at -80 °C in preparation for FTIR. The remaining portion of the cell pellet was prepared for cellblocks to be used in cytology and immunohistochemistry.

Estimating the proportion of bronchial epithelial cells per sputum sample

For each sample, 1000 cells were counted per slide. The percentage of bronchial epithelial cells for that sample was then estimated from this total

count. All samples (i.e. cancer, high risk and healthy control) were assessed for the presence of bronchial epithelial cells but for analysis purposes, each group were represented by randomly selected cases. There were 40 selected cases from the cancer group and healthy control group and 36 high-risk cases. These were further divided into induced sputum and non-induced cases. The scores were recorded and comparisons were made between groups for mean counts using the Mann-Whitney U test.

Sputum analysis with FTIR

Samples were spotted on to 96 'well' re-usable silicon sample carrier plates (LNC Technology Ltd., Ystrad Mynoch, Hengoed, UK). The silicon plates were cleaned in warm 0.5% SDS, rinsed with distilled water, soaked overnight in 5 M nitric acid, rinsed again with distilled water and air-dried. The sputum samples (after being thawed for 12–24 hrs) were plated randomly across a plate, permitting possible variations within or between plates to be taken into account during analysis. The loaded sample plates were oven dried at 50°C for 30 minutes (Sanyo Gallenkamp plc., Loughborough, UK) in order to remove extraneous moisture prior to FTIR analysis. Prepared plates were allowed to cool and then inserted onto the motorised stage of the diffuse reflectance absorbance-scanning accessory, connected to the FTIR spectrometer. We used the VERTEX 70 spectrometer (Bruker Optics Ltd, Banner Lane, Coventry, UK), equipped with a mercury-cadmium-telluride (MCT) detector cooled with liquid nitrogen where the spectra were obtained in reflectance mode. In order to minimise carbon dioxide (CO₂) and water vapour interference peaks in the mid-infrared region of collected spectra, the sampling compartments and microscope stage were purged with dry CO₂ free air produced from a Peak Scientific compressor (Peak Scientific Ltd. Paisley, UK). The spectrum for each sample contained 1763 data points, ranging from 4000 to 600 cm⁻¹ wavenumbers. Each spectrum represented the average of 256 scans which improved the signal to noise ratio. The FTIR process took approximately 2 hours.

Sputum processing for immunohistochemistry

The preparation of cell pellets was as described above. After a portion of the individual cell pellets was used for FTIR, the remaining sample was prepared for immunohistochemistry.

Preparation of sputum cells for immunohistochemistry

Whilst preparing the sputum cells for cell block, they had to be formed into a clot using plasma and thrombin. Plasma was made up by adding 1ml of distilled water to the plasma powder (Dade Ci-trol 1, Dade Behring, Marburg, Germany) and allowing the solution to tilt for 15 minutes. The thrombin (Dade Behring, Marburg, Germany) was made up by adding 5ml of distilled water to the thrombin powder, gently swirling the solution until it had completely dissolved. To the centrifuged cell pellet 0.25ml plasma was added and to this a further 0.5ml of thrombin was added. A clot was formed within seconds and this was then used to make a paraffin cell block.

Cell pellets reserved for immunohistochemical analysis were fixed in 10% buffered formal saline for 1 to 2 hours, embedded in paraffin and 3-4 μm thick sections prepared on positively charged slides before immunohistochemical studies. The sections were put directly on Benchmark XT which used EZ prep to de-wax the slides. The tissue sections were incubated with p63 antibody (1:50 dilution, Menarini, Florence, Italy); EGFR antibody kit (Ventana Biotek Solutions, Tucson, Arizona, USA); p16 (pre-diluted, mtm labs, Heidelberg, Germany); p53 (1:50 dilution, Novocastra Laboratories Ltd, Newcastle upon Tyne, UK) or cyclin D1 (1:100 dilution, Labvision, Newmarket, Suffolk, UK) at a concentration of 4 μg /slide. For antigen retrieval, the slides were incubated in CC1 buffer (Ventana Biotek Solutions, Tucson, Arizona, USA) for an hour on heated plates at 100°C on a Benchmark XT processor. Primary antibody incubation ranged from 32 to 40 minutes (depending on the antibody) at dilution 1:100 at 37°C. Positive immunostaining was detected through interaction of Avidin Biotin Complex (ABC) with biotin conjugated secondary antibody using a Ventana I View DAB detection kit (Ventana Biotek Solutions, Tucson, Arizona,

USA). The ABC method is a widely used technique for immunohistochemical staining. Avidin, a large glycoprotein, can be labelled with peroxidase or fluorescein and has a very high affinity for biotin. Biotin, a low molecular weight vitamin, can be conjugated to a variety of biological molecules such as antibodies. [264] The slides were subsequently counter stained with haematoxylin, dehydrated, cleared and mounted in DPX mountant to be examined under light microscopy. Each antibody required a control: p16 – cervix; EGFR – skin; p53 – breast; p63 – skin; cyclin D1 – mantle cell lymphoma. A negative (no antibody) control section of sputum cells was also used.

Sputum analysis with immunohistochemistry

We used a standard immunohistochemical scoring system (IHC). [265, 266] Immunohistochemical sections were analysed by myself and Mrs Christine Davies (an experienced cytological analyst) using a multi-headed microscope (Olympus BX51 at x 20 magnification and Olympus WHN 10 x 22 eyepiece). Both assessors were blinded to the patient's diagnosis, demographics, and any other clinical metadata at the time of assessing the slides, using only a unique study identifier number. After processing the 99 samples for FTIR, there were only 83 samples with enough of a cell pellet left for IHC. The first 66 sputa were assessed together with a consensus to allow myself to be trained to an adequate level before we assessed the remaining 17 samples individually. We applied scoring systems based on techniques established by others. [265, 266] The scoring system used varied depending on the antibody. EGFR is a cell membrane antigen, whereas p16, p53, p63 and cyclin D1 are nuclear antigens. Figure 6.3 shows the IHC scoring systems looking at cell positivity.

Nuclear Antigens (p16, p53, p63, cyclin D1)

0 = less than 5% of tumour cells staining;

1 = 5–50% of tumour cells staining;

2 = more than 50% of tumour cells staining;

5 = uninterpretable.

Cell Membrane Antigens (EGFR)

0 score = No staining observed, or membrane staining in < 10% neoplastic cells. Negative.

1 score = Weak complete and/or incomplete membrane staining in > 10% neoplastic cells. Positive.

2 score = Moderate complete and/or incomplete membrane staining in > 10% neoplastic cells. Positive.

3 score = Strong complete and/or incomplete membrane staining in >10% neoplastic cells. Positive

Figure 6.3: IHC scoring system for both nuclear and cell membrane antigens for biological screening for lung cancer with sputum [265, 266]

c) STATISTICS

FTIR analysis of sputum

Analysis of the demographic data was performed using the Statistical Package for Social Sciences (SPSS) version 13.0 (© 2000, The Apache Software Foundation, Chicago, Illinois). [267] Data was assessed for normality using the Shapiro Wilk and Kolmogorov-Smirnov tests. If there was a discrepancy between the two tests then if the sample size was < 50, the result of the Shapiro Wilk test was used, and if > 50 samples, Kolmogorov-Smirnov was used. Numerical data (e.g. age of subject) that were normally distributed were compared using the non-paired Student's t-test, or if there were more than two variables, the analysis of variance (ANOVA) was performed. If the variables were not normally distributed then they were compared using the non-parametric Mann-Whitney U test. Analysis of categorical data, such as smoking status, was performed using Chi-squared.

All FTIR statistical analysis was performed using the R (CRAN) version 8.1 statistical analysis and programming environment (The R Foundation for Statistical computing). [268] FTIR data distributions were assessed for normality using the Shapiro Wilk Test. The FTIR spectra were then compared against each other using PCA as described in Chapter 4.

FTIR spectra analysis

The raw FTIR spectra initially underwent a 3 stage process:

i) Normalisation.

The various sputa contained a different number of bronchial cells in comparison to each other, which may or may not be proven to be statistically significant. As a result of this, spectra produced may have been similar with regards to peaks, but variable with regard to absorbance. To standardise this, the greatest peak on each spectra is given the value 1. The height of the 2nd highest peak is a proportion of 1. For example, if the 2nd highest peak was at half the absorbance

of the highest peak then it has the value 0.5. The y-axis then runs from 0 to 1. All the spectra thus were normalised to a similar scale.

ii) Baseline correction.

Many of the spectra may have a sloping baseline. The most important reason for the sloping baseline is the formation of a film over the samples, which results in scattering of IR light. A two-point baseline correction was applied which is a standard procedure. This involves two points of the spectral baseline being selected and a straight line drawn between the two points. The spectra was then shifted to nullify the slope. We took the first point at wavenumber 1800 cm^{-1} and the second point was the final wavenumber, as this was within the fingerprint region of interest.

iii) 2nd derivative

The curves produced then underwent a two-step process to develop a 2nd derivative spectra. The usually jagged peaks were smoothed out in the first step, and then the 2nd derivative spectra were produced where not only the height of the peaks is relevant but also the gradient leading up to and away from the peak too. This ensured that “peaks” which were not obviously visible by looking at 1st derivative spectra were not omitted from analysis. These peaks may or may not have been important. Savitzky and Golay described this technique in 1964. [269]

Hierarchical Cluster Analysis (HCA)

HCA is a multivariate technique where a set of variables (in this study it will be the spectra from each of the cases) is grouped together in clusters depending on how similar the cases are. It is developed in a tree-building technique starting at the bottom where all cases are independent. As we take the first step up the tree (dendrogram), the cases with the highest similarity are linked together by a single-linkage. The next cases of highest similarity are then linked and so on. Once the tree is fully built all the cases will be linked, however, the links at the top of the tree represent the least similar cases (variables) of the whole patient sample. HCA therefore gives an immediate idea of how groups of

cases cluster together according to similarities. Thus, with FTIR spectra, those with similar metabolic profiles should cluster together.

Selection of significant wavenumbers

Differences between median absorbances of pairs of wavenumbers were then evaluated using the Mann-Whitney U Test. Initial analysis was between the cancer and healthy control groups. When applying multiple statistical tests simultaneously there is an increasing risk of generating a higher rate of false positive results by falsely rejecting the null hypothesis (multiplicity problem). We therefore had to address this issue by applying a correction to procedure to adjust significance levels (i.e. p values) to minimize the risk of generating false positives. A simple procedure was to apply a Bonferroni correction where the p value (in this case 0.05) was divided by the number of tests (n) being applied (in this case 1763). [270] The problem with the Bonferroni correction is that it is too conservative in trying to prevent false positive results, increasing the risk of false negatives, so a more robust procedure was called for. A popular variant of the Bonferroni correction is the Holm's sequential Bonferroni correction. [271] This procedure ranks the significant p values from the most significant to least significant. It then divides the level of significance p value (0.05) by n and then sequentially works through the ranked list dividing the p value (0.05) by $n -$ the rank number, to give a new p value. If the old p value $<$ new p value then it kept its significance. If the old p value $>$ new p value then it was rejected.

To determine the most optimal wavenumber pair that would discriminate between cancer and healthy controls, evaluation was performed using logistical regression of each wavenumber pair. A pair of wavenumbers that generated the highest accuracy were those that, after being incorporated into a predictive model, yielded the lowest rate of false positives (specificity) and false negatives (sensitivity). To determine the most optimal wavenumber pair, all pairwise combinations of wavenumbers were sequentially evaluated using the R statistical package. For each wavenumber pair, a linear model was generated that separated cancer and healthy control cases, according to their absorbance values. The discriminating line was positioned so that it separates cancer and

healthy control cases as best it could whilst minimizing both false positives and negatives.

Cross-validation with leave-one-out method

Once the model was generated using a combination of wavenumbers, it was cross-validated using the leave-one-out method. Using this method, a case was taken out of the model and the sensitivity was retested. Subsequently, that case was reinstated and the next case was withdrawn and the sensitivity retested. This was continued until each case had been “left out”. The results of the sensitivities were then plotted on a histogram and in order for the model to be robust, the sensitivities need to have been consistently maintained with very little spread.

Predictive model equation

Once the model had been developed, the third cohort (high-risk group) was then applied to this model using the equation:

$z = (\text{coef1} \times \text{value for wavenumber 1}) + (\text{coef2} \times \text{value for wavenumber 2}) + \text{intercept}.$

A further equation was then applied (logit function):

$$f(z) = 1/(1+e^{-z})$$

where $f(z)$ is the probability of a particular outcome. The output is between 0 and 1. Those with an output > 0.5 were predicted as cancer. Those with an output < 0.5 were predicted as not cancer.

IHC analysis of sputum

All statistical analysis was performed using SPSS for Windows and the R statistical analysis and programming environment. Sociodemographic data was assessed for normality using Shapiro Wilk / Kolmogorov-Smirnov tests and then for normally distributed data, the non-paired Student's t-test was performed, and for non-parametric data, Mann-Whitney U Test. Categorical data was assessed by Chi-squared. This allowed a comparison of the cohorts baseline characteristics. As the immunostained slides were scored independently by two assessors, the Cohen kappa data test was performed to analyse concordance between the two sets of scores. [272]

Mann-Whitney U Test was used to look for differences in number of positive cells between the individual antibodies between the cancer and high-risk group. The sensitivity and specificity for each antibody was then calculated. HCA was performed to see if any clusters emerged suggesting similarities in cases based on the IHC panel of 5 antibodies.

Chapter 7. Biological screening for lung cancer with exhaled breath

Background

We obtained approval from the Local Research and Ethics Committee (LREC), (08/WMW01/21), and Hywel Dda Trust Research and Development Unit in March 2008. Site-specific approval in Swansea was obtained as the study progressed. Recruitment took place between 2008 and 2009 across 2 hospitals, Prince Philip Hospital (Llanelli) and Morriston Hospital (Swansea). All sample processing took place at The Welsh Centre of Printing and Coating, Swansea University and statistical analysis at The Institute of Life Sciences, Swansea University and Lampeter University.

MATERIALS AND METHODS

a) PATIENTS

We approached those patients who were newly clinically diagnosed with lung cancer. They were recruited within the hospital after they presented to either the lung cancer or other respiratory clinics or from the acute medical intake. The majority (n=35) had histologically confirmed evidence of lung cancer, and a small number were diagnosed based on their clinical presentation and radiology results. The exclusion criteria were those whom were unable to give consent, those under the age of 18 years old, and those unable to carry out the breath tests correctly (see Figure 7.1). All were given a PIS and at least 24 hours to consider whether to take part in the study (See Appendix 6). Following written informed consent, physiological and sociodemographic data including lung cancer subtype, cancer staging, age, sex, FEV₁, past medical history (including previous cancer), smoking status, smoking history and date of diagnosis were recorded on a study specific form (See Appendix 7). Three consent forms were signed, one given to the patient, one for the medical notes

and one for the research master file (Appendix 8). Their GP was also informed (See Appendix 9).

Patient flow Chart

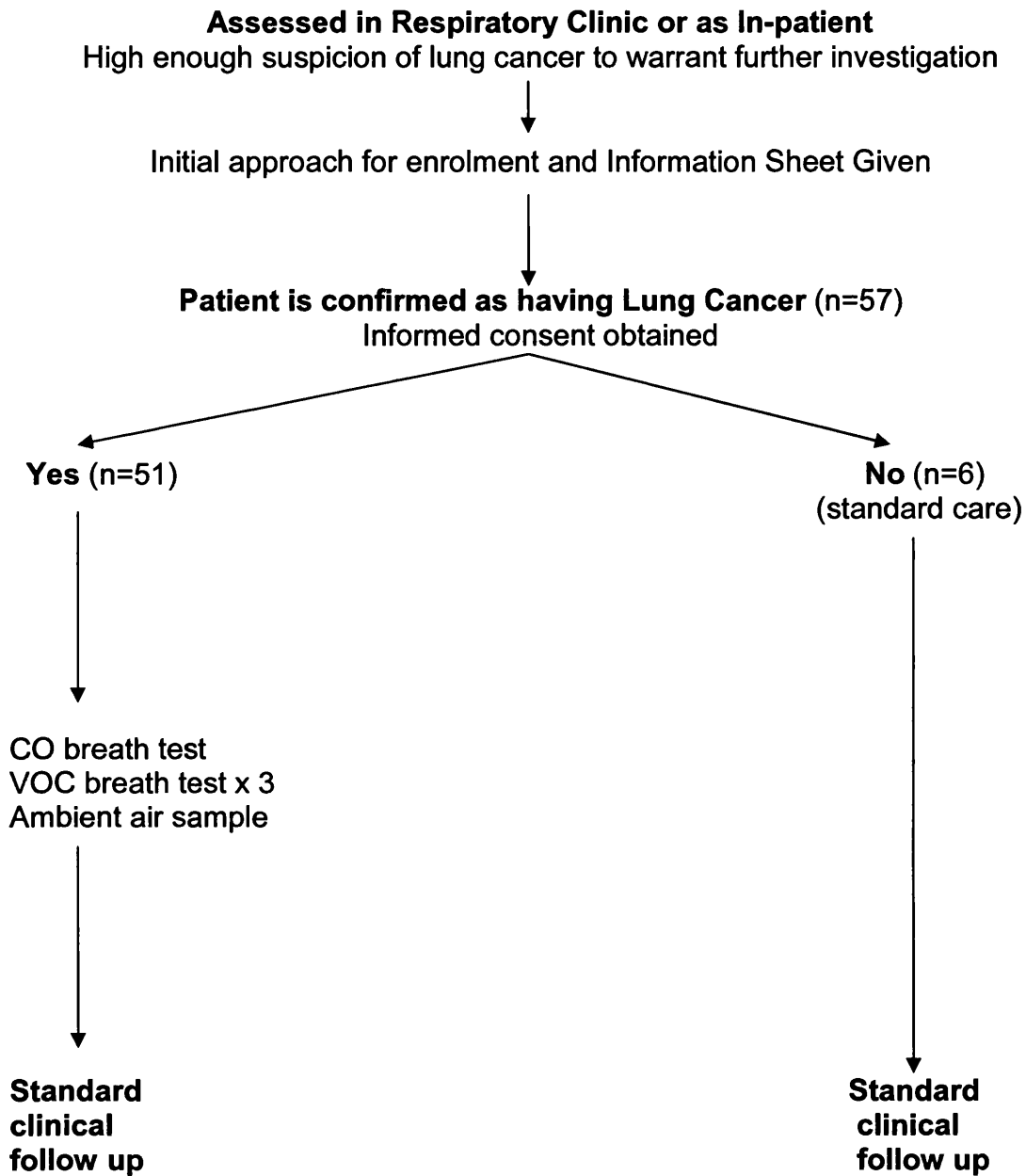


Figure 7.1: Patient flow chart for biological screening for lung cancer with exhaled breath

Our control group consisted of 'healthy' subjects who had no pre-existing lung disease according to our simple screening questionnaire as detailed in Figure 7.2.

Do you smoke?

How many years have you smoked for and how many per day?

Have you lost weight recently (unintentionally)?

Are you troubled by any of the following symptoms at the moment?

-cold or flu

-chest pain

-wheeze (including at night)

-cough

-increased sputum or spit production

-breathlessness

-do you take any inhalers or other medications for your chest

Figure 7.2: Screening questions to establish 'healthy' status of Controls for biological screening for lung cancer with sputum exhaled breath

Controls were excluded if they had a known diagnosis or symptoms of underlying lung disease or current chest infection. Smokers and non-smokers were both included. The potential controls were given a Healthy Controls Information Sheet (See Appendix 10) and given at least 24 hours to consider their participation. If they agreed to be involved then formal consent was obtained and a record of age, sex and smoking status/history was recorded.

b) PROCEDURE

Sample Collection

When collecting the breath samples, all participants were fasted for at least 4 hours. Although there was no clear guidance in the literature over fasting, it was felt that eating certain food or drinking certain beverages might alter the VOC components of the breath. Previous studies measuring VOCs fasted the patients overnight and performed the breath tests before bronchoscopy. [160, 161] We felt that 4 hours was a reasonable length of time to be fasted for this group of patients. The majority of our patients were sampled immediately prior to bronchoscopy but the controls and those who provided breath samples from wards or their houses were also fasted.

All subjects were rested in a room with windows and doors shut for 20 minutes pre breath test, as there is evidence that acute aerobic and anaerobic exercise can increase the oxidative stress within the body. [273, 274] We felt that 20 minutes provided enough time for the individual to fully recover from any exertion and for them to return back to their resting heart and respiratory rate.

We used a Bio-VOC[®] breath sampler (Markes International Limited, Llantrisant, Wales, UK). (see Figure 7.3)



Figure 7.3: Bio-VOC[®] breath sampler

The Bio-VOC[®] harnesses the fact that the concentration of VOCs in the blood is in equilibrium with the concentration in the air in the alveolar portion of the lungs. This assumes that free exchange of chemicals can occur through the thin alveolar wall and capillary blood vessel walls and subject's alveolar (end-tidal) VOC concentrations are proportional to those in their blood.

Before using the Bio-VOC[®] sampler the container was cleansed by repeatedly pumping air in and out with the plunger, 10 times, to ensure any previous patients VOCs had been removed. When using the Bio-VOC[®] breath sampler, the subject was asked to take a deep breath in and then exhale slowly into the sampler via a disposable mouth piece until full expiration was completed (a slow vital capacity breath). An individual exhaling fully can usually expel over 4 litres (L) of air, although this may be reduced in patients with lung disease. 2000ml of this air comes from the alveolar component of the breath and it is this later air that is the most important to trap. The sampler holds 129ml of air, so typically 1L of alveolar air needs to pass through the sampler to ensure the presence of undiluted alveolar air.

Once the air was trapped, the plunger replaced the mouthpiece again and the air pushed into a 2-bedded thermal desorption tube (Markes International Limited, Llantrisant, Wales, UK) as demonstrated in Figures 7.4 and 7.5.

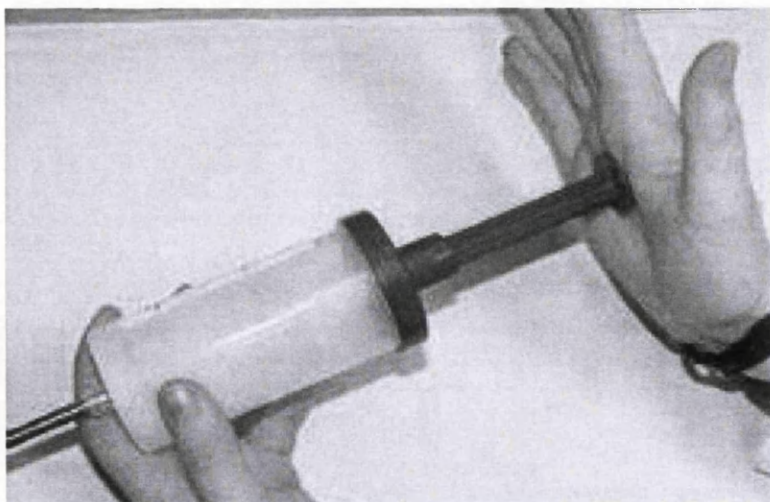


Figure 7.4: Pushing the trapped alveolar air with the plunger into the 2-bedded thermal desorption tube



Figure 7.5: The 2-bedded thermal desorption tubes

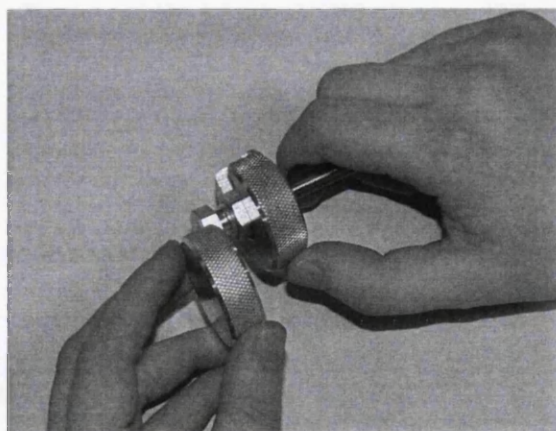


Figure 7.6: Screwing the brass caps onto the sorbent tubes

Once the trapped alveolar air was pushed into the desorption tube, an airtight 2-piece, brass, ¼ inch screw cap was fitted to each end to prevent gases entering or leaving the tube, and these caps were tightened with a spanner (Figure 7.6). These types of seals have been evaluated in several studies and have been demonstrated to ensure minimal ingress of external artefacts and minimal loss of the sample over a 27 month period. [275-277] The thermal desorption tube contained two graphitized carbon black sorbents (Carbograph 1TD and Carbopack X) with a total sorbent mass approximately 470mg which absorbed the volatile compounds with the more volatile being trapped first. Each tube was preconditioned by flowing helium through at 75 ml/min at 320°C to ensure there were no residual VOCs within them. Each tube had a unique number that was recorded on the participant's data sheet. The procedure was repeated two further times so that we had 3 alveolar breath samples per patient. (Following external review on early pilot data, we amended our protocol and collected the patient samples in triplicate after the first 16 patients). Each tube number was recorded on the patient data sheet in the order in which they were taken. The tubes were stored in lined steel cans in between desorption stages to reduce contamination.

An ambient air sample was taken each time a new set of patient samples was taken. The eventual readings of ambient air sample's VOCs would be subtracted from the subject's sample to give the true proportion of VOCs from the subject's body to allow for the changing levels of VOCs from day to day within the same environment. For example, within a bronchoscopy suite, which is in close proximity to operating theatres and an intensive care unit, there is likely to be an increased background concentration of anaesthetic gases. Within Outpatient departments, smokers may be walking through the area even if the subject is a non-smoker. Thus background changes in VOCs as a confounder is reduced as much as possible.

After the VOCs had been sampled, the patient's current smoking status and number of cigarettes currently smoked per day was recorded. Smoking status was validated using a CO monitor (Bedfont-Micro Smokerlyzer ®) using the technique described in Chapter 6.

Processing the samples

The samples were analysed using GC-MS. In this method, as the sample gets heated up, the more volatile gases are detected first. Analysis of the gas is then performed by mass spectrometry. The schematic flow chart is illustrated in Figure 7.7, with the actual GC-MS machine used illustrated in Figure 7.8.

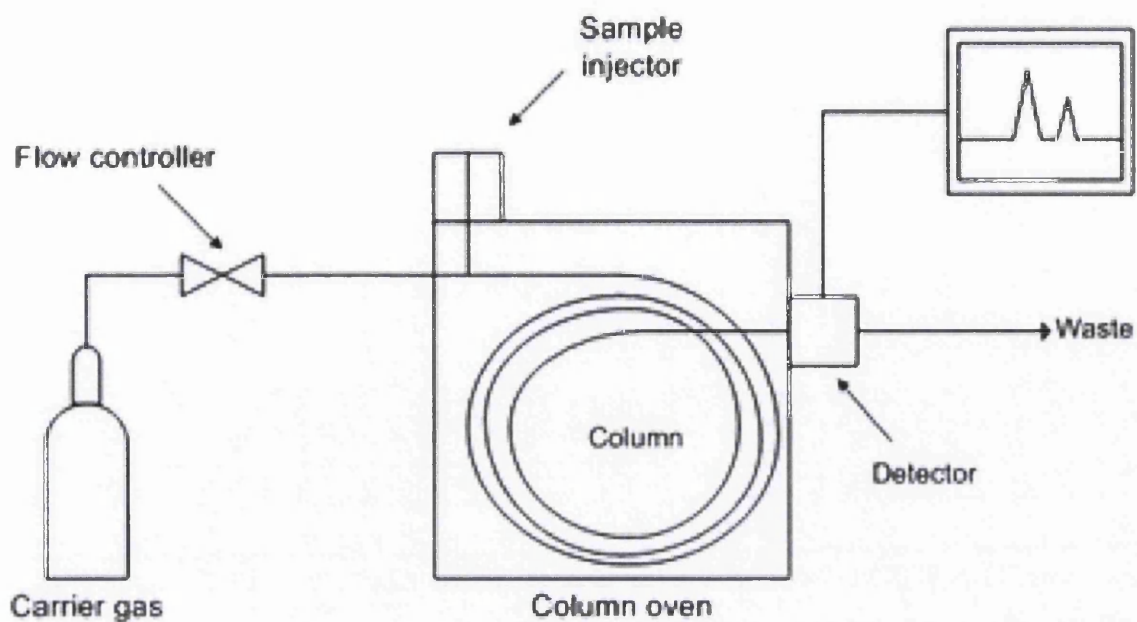


Figure 7.7: Schematic of components of GC-MS (image taken from www.wikibooks.org)

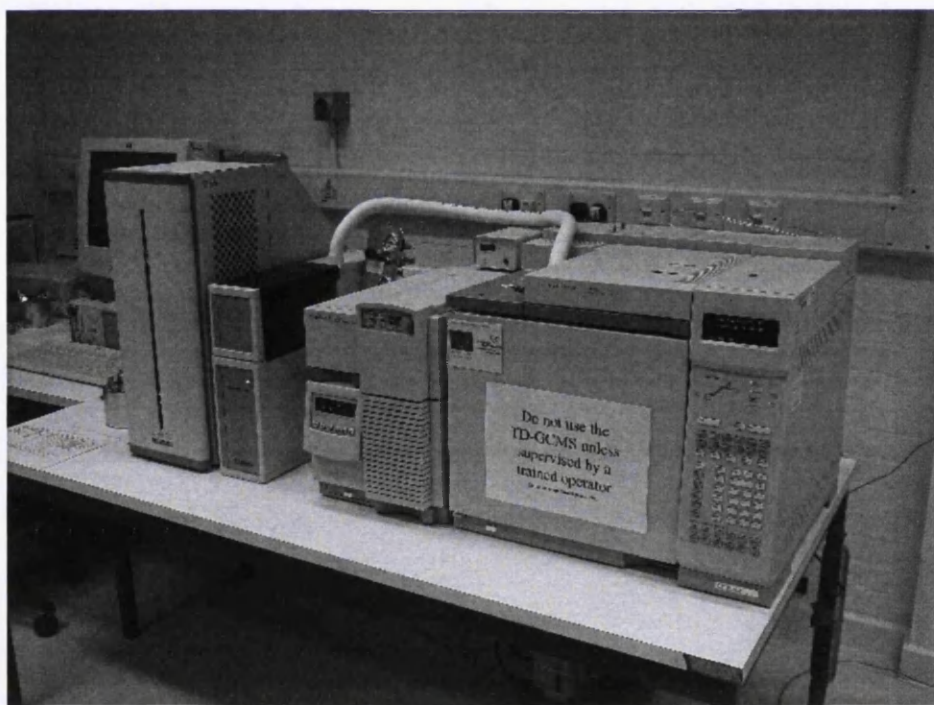
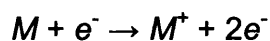


Figure 7.8: Actual GC-MS used for sample processing in Swansea University (Markes International, Llantrisant, Wales)

Batches of tubes were loaded into an Ultra unit (Markes International, Llantrisant, Wales) for automated processing via a Unity thermal desorption unit (Markes International, Llantrisant, Wales) and fed with inert helium at 10 psi which was used to desorb the tubes in a 30 ml/min stream of inert helium at 300°C for 5 minutes. The flow was driven onto a cold trap (U-T11GPC, general purpose graphitized carbon C4/5-C30/32) set at -10°C. The trap was then desorbed at 300°C for 3 minutes. To allow sufficient flow through the trap, 5ml of the sample was vented and the remainder injected onto an Agilent Technologies 6890N gas chromatograph. The VOCs were then separated using a capillary column 30m x 0.25mm id coated with a film thickness of 0.25µm. The column temperature was initially set at 40°C and then increased to 200°C at a rate of 5°C/min.

As the volatile compounds are separated from the sample (the more volatile being separated first) they undergo electronic ionization which leaves the compound positively charged. When a gas undergoes electronic ionization, it is bombarded with electrons resulting in the gas molecule becoming a positive ion. [278] The following describes the electron ionization process. [279] The VOC is represented by M and electrons by e^- .



Passing through a magnet, the compound is then deflected, with the degree of deflection being influenced by the mass-to-charge ratio of the ionized compound. [280] Figure 7.9 is a schematic of a simple mass spectrometer.

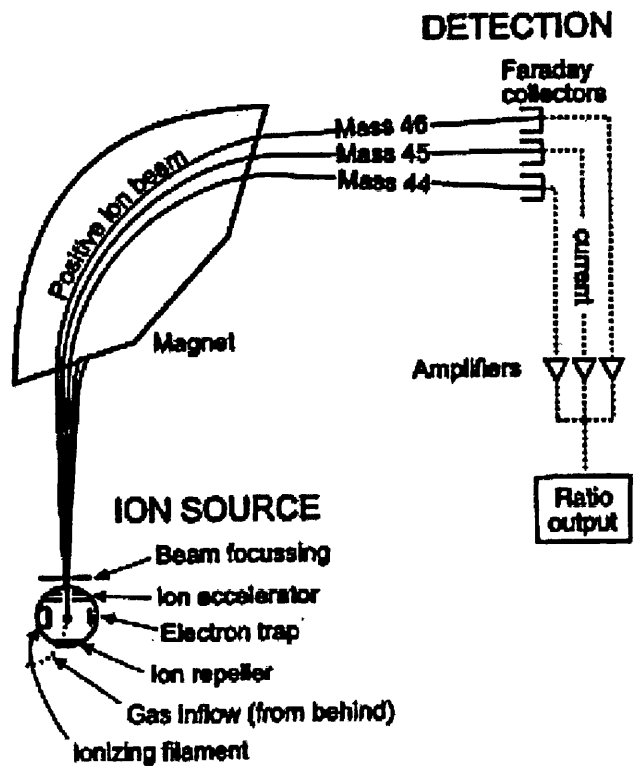


Figure 7.9: Schematic of simple mass spectrometer. (Image taken from www.wikipedia.org)

Mass spectral data were obtained in the SCAN mode range between 40 - 550 amu (atomic mass units). The 40 amu cut-off was used to eliminate water, air etc from the spectra. Data is outputted as a chromatogram, a plot of abundance against retention time. Automated peak detection was used to calculate peak area (area under the curve) and retention time for each compound, with the same peak integration parameters used for all samples using an automated library search report by Chemstation (© Agilent Technologies, 1999). The area under the curve (AUC) in the retention time/abundance chromatogram is analogous to the amount of a VOC for a particular sample and the retention time (RT) is the time at which the VOC exits the column and is detected. This also aids its identity as the same compound occurring in different samples should have the same retention time if settings are unchanged. The National Institute of Standards and Technology (NIST) 98 mass spectral library (Version 1.7a, build 07 / 18 / 2000 Distributed by the Standard Reference Data Program of the National Institute of Standards and Technology © 1997-1998) was used to identify the compounds using the spectra obtained at the apex of each peak and utilising probability based matching (PBM). [281] Each match was also assigned a quality rating based (from 1 to 100) on how well it matched the library. Extensive manual peak identification was also performed, within the NIST program, to check the validity of the automated matches. Automated matches are tentative and should ideally be confirmed by comparing retention times and mass spectra of pure chemicals injected directly into a sorbent tube.

Each tube was run twice with the first run used for analysis and the second analysed to ensure that there was no carry-over of volatiles onto subsequent tubes and to check for VOCs inherent in the system such as siloxanes from the column. After each set of tubes was desorbed an additional tube conditioning stage was used with 75ml/min helium at 320°C to ensure complete removal of any residual VOCs.

In order to monitor system performance over the duration of the experimental program, a solution of toluene in methanol (10µL per 100ml of methanol) was directly injected (in varying quantities) into sorbent tube with a stream of helium passed over the tube for 2 minutes. The tube was then desorbed and analysed

using the same methodology as the sample testing. There was no drift in detected levels over time.

c) STATISTICS

Sociodemographic data was analysed using the Statistical Package for Social Sciences (SPSS) version 13.0 (© 2000, The Apache Software Foundation, Chicago, Illinois). Mann-Whitney U Test, non-paired Student's t-test and Chi-squared were all used depending on the data distribution. Graphs of AUC against RT, polynomial equations and compound-by-compound analysis of VOCs were performed using Microsoft Excel for Windows, version 2003 (© 1985-2003 Microsoft corporation). [282]

Assessing clusters of VOCs

All the cancer and control samples were analysed for VOCs and all VOCs that occurred greater than 5 times overall were deemed valid. This was an arbitrary figure chosen as it was felt by our VOC lead technician (personal communication with Dr Chris Phillips, 2010), that those VOCs occurring less than 5 times were generally low quality matches and were unlikely to be relevant. We plotted graphs of AUC against RT, in order to identify visually any clusters of VOCs. As the patterns obtained clearly did not follow a linear or exponential relationship, we applied polynomial equations to develop mathematical curves that best separated lung cancer cases from controls. These equations represent the closest mathematical relationship that could be established between AUC and RT for lung cancer cases, and separately for controls. The entire spectrum for each individual (lung cancer and controls) was tested against these equations, using the relative deviation from each of the curves to check a final discrimination method and we used multiregression analysis to develop the R squared value. The R squared value is an indicator of the correlation of all the points relating to the compounds with that polynomial expression. If the R squared value is close to 1 then the correlation (percentage variability) is statistically significant suggesting that the majority of points lie

close to or on the polynomial trendline. The equation lines are also very useful to identify particular zones for discriminating cancer patients.

Identification of discriminating VOCs by compound-by-compound analysis

All of the cancer samples were initially included (i.e. VOCs from all 3 samples when the majority of patients were taken in triplicate and all of the single (very early) samples). However, it was felt appropriate to use only the first tube of the triplicates of the control samples (or the single sample if only one was collected) in this statistical analysis as there was no guidance in the literature with regards to which tubes to use and there was little variability across the 3 control samples of VOCs. We could find no published guidelines or indeed any studies where samples were taken in triplicate. A comparison was then made between the total VOCs detected in the cancer group and those in the control group. Any VOCs that were present in any one of the samples for the individual cancer patients (for example one of the 3 triplicates, or in the single sample for the early subjects) and in none of the control subjects samples were deemed cancer-exclusive VOCs. These cancer-exclusive VOCs were then combined to see how many of the cancer patients could be correctly identified by the panel. The same approach was taken for those VOCs present in the control patients' tube but not present in any of the cancer patients sets of samples, and these were control-exclusive VOCs. Again we assessed how many controls could be correctly classified using these VOCs.

We performed similar analysis of VOCs that were universally higher in (but not exclusive to) cancer patients and then for VOCs that were universally higher in (but not exclusive to) controls.

Analysing the ambient air samples

All the discriminatory VOCs were individually compared with their concurrent ambient air sample. If the VOCs consistently appeared in the ambient air samples then a subtraction process would take place, subtracting the ambient air VOCs from the alveolar air VOCs (alveolar-air gradient). [283]

PART 3: RESULTS

Chapter 8. Fourier Transform Infra-Red and Immunohistochemical analysis of sputum in lung cancer

EVALUATION OF SPUTUM WITH FTIR

Patient Demographics

120 patients referred for bronchoscopy with a possible diagnosis of lung cancer were approached. 8 refused consent and 2 were unable to give informed consent. Of the remaining 110 patients whom were recruited, we were only able to obtain sputum (either self-expectorated or suctioned) in 99.

63 of the 99 patients were given a formal diagnosis of lung cancer by the lung MDT.

The remaining 36 patients did not have any evidence of lung cancer on 1-year clinical and radiological follow up (diagnosis made: pneumonia, n=13; pulmonary fibrosis, n=5; old TB, n=3; Mycobacterium Avium Intracellulare (MAI), n=1; haemoptysis on background of COPD bronchiectasis / asthma, n=12 ; B-cell lymphoma, n=1; empyema, n=1).

93 people were approached to be healthy controls. 4 refused and 11 were not eligible after completing the screening questionnaire (symptoms suggestive of an underlying respiratory disease, n=3; pre-existing respiratory disease, n=5; symptoms consistent with current respiratory tract infection, n=3). 8 were unable to provide a sputum sample.

After analysis of the spectra, 9 of the cancer group, 16 of the healthy control and 12 of the high-risk spectra were excluded. This was because the initial part of the spectra was completely different (in fact inverted) from all other spectra suggesting a technical problem. The fingerprint region of the spectra we were to be interested in (wavenumbers $1800 - 900 \text{ cm}^{-1}$) appeared consistent with all other tracings, however due to the inconsistent early component of the spectra

they were excluded from further analysis to ensure the least confounders for the model. On further investigation it was identified that a film was on the processing tray where the cell pellets lay and probably affected the spectra detection.

Thus the 3 cohorts consisted of cancer subjects (n=54); high-risk subjects (n=24) and healthy controls (n=54). Table 8.1 compares the baseline sociodemographic characteristics of the 3 cohorts.

	Cancer group n = 54	High-risk group n = 24	Control group n = 54	p-value
Mean Age (yrs)	66.6 +/- 8.7	65.1 +/- 13.6	51.1 +/- 15.3	<0.001 [*]
% male	52	75	52	0.12 ^{**}
Smoking Status				
Current	19	9	24	<0.001 ^{**}
Former	28	15	1	
Never	7	0	29	
Median no. of pack years (IQR)	30 (20 – 45)	40 (22 – 50)	n/a	0.18 ^{***}
Mean predicted FEV ₁ (%) ^{****}	69.5 +/- 17	65.8 +/- 20	n/a	0.48 (CI -6.8 to 14.1)

Table 8.1: Demographic of the cancer, high-risk and controls cohorts

*post-hoc tests show no significant difference between ages of cancer and high-risk group, but controls were significantly younger than both.

**calculated using chi-squared

***data not normally distributed, thus Mann-Whitney U test applied

****n=38 in cancer group and n=18 in high-risk group, 22 were not recorded across the two cohorts

n/a = not available as we did not have ethics to do spirometry and pack-year history on healthy controls

The diagnosis of cancer was confirmed histologically in 46 cases and clinico-radiologically in 8. Of the 54 patients with a diagnosis of primary lung cancer (cancer group), 39 were NSCLC, 9 were small cell lung cancer, 5 were clinical diagnosis and 1 had carcinoid.

Table 8.2 has the histological breakdown of the cancers and Table 8.3 has the details of the cancer cohort. Tables 8.4 and 8.5 have the details of the high-risk and healthy control cohorts respectively.

Lung Cancer Type	Histologic Sub-Type	Number
Non-small cell (NSCLC)	Squamous Cell Carcinoma (SCC)	17
	Adenocarcinoma	10
	Large cell	2
	Unknown subtype *	9
	Broncho-alveolar Cell	1
Small Cell (SCLC)		9
Lung – clinical diagnosis		5
Carcinoid		1
	Total	54

Table 8.2: Breakdown of histological subtypes within the cancer cohort

* This group were noted to have cells consistent with NSCLC but the pathologists was unable to identify the histological subtype

Case ID	Age	Sex	Smoking Status	Smoking history pack years	Previous Cancer?	Final clinical diagnosis	Final histology
LC04	67	M	ex	30	No	NSCLC	Bronchoalveolar
LC06	77	M	ex	15	No	Lung – clinical diagnosis	Not known (NK)
LC07	61	M	ex	30	No	NSCLC	SCC
LC08	60	M	ex	20	No	NSCLC	SCC
LC10	58	F	current	48	No	NSCLC	Adenocarcinoma
LC11	81	M	current	120	No	NSCLC	SCC
LC14	83	F	ex	25	No	Lung – clinical diagnosis	NK
LC15	56	M	current	30	No	NSCLC	NK
LC16	55	F	current	20	No	NSCLC	Adenocarcinoma
LC30	69	M	current	90	No	NSCLC	SCC
LC33	68	F	ex	20	breast	NSCLC	SCC
LC36	70	F	ex	12	No	Lung – clinical diagnosis	NK
LC37	67	M	ex	10	No	NSCLC	Adenocarcinoma
LC42	69	F	never	0	No	carcinoma in lung	Carcinoid
LC43	64	M	ex	25	No	NSCLC	NK
LC44	58	M	current	45	No	Small Cell	small cell
LC49	66	F	ex	45	breast	NSCLC	SCC
LC50	68	M	ex	40	No	Small Cell	small cell
LC54	77	F	ex	60	No	NSCLC	SCC
LC56	67	M	current	77	Dukes A colon	Small Cell	small cell
LC58	87	F	never	0	colon/ovarian	NSCLC	Adenocarcinoma
LC61	52	M	current	40	No	Small Cell	small cell
LC62	50	F	ex	36	No	NSCLC	Adenocarcinoma
LC64	71	F	ex	30	skin	Lung – clinical diagnosis	NK
LC68	72	F	current	60	colon	NSCLC	Adenocarcinoma
LC69	48	F	never	0	No	NSCLC	Large cell
LC70	62	M	ex	100	No	NSCLC	NK
LC71	77	F	ex	7	No	NSCLC	Adenocarcinoma
LC72	69	M	current	25	No	NSCLC	SCC
LC75	81	F	current	30	No	Lung – clinical diagnosis	NK
LC76	77	M	ex	25	No	NSCLC	SCC
LC77	69	M	current	40	No	Small Cell	small cell
LC78	81	M	ex	60	larynx; bladder	NSCLC	SCC
LC79	62	M	current	47	No	NSCLC	SCC
LC80	66	M	ex	20	No	Small Cell	small cell
LC81	66	F	never	0	No	NSCLC	Large cell
LC82	71	M	ex	50	No	NSCLC	SCC
LC85	54	M	never	0	No	NSCLC	Adenocarcinoma
LC86	60	M	ex	30	No	Small Cell	small cell
LC87	68	F	ex	30	No	NSCLC	NK
LC89	60	F	ex	40	No	NSCLC	Adenocarcinoma
LC91	62	F	current	40	breast	NSCLC	NK
LC92	81	F	current	30	No	NSCLC	SCC
LC93	63	F	ex	40	No	NSCLC	NK
LC94	65	F	never	0	lung	NSCLC	SCC
LC95	74	M	ex	50	No	NSCLC	SCC
LC97	62	F	current	45	No	NSCLC	NK
LC101	54	M	ex	40	No	NSCLC	NK
LC103	69	M	current	40	No	NSCLC	NK
LC105	57	F	never	0	No	Small Cell	small cell
LC107	68	M	current	50	No	Small Cell	small cell
LC108	62	F	current	30	No	NSCLC	SCC
LC109	64	F	ex	20	No	NSCLC	Adenocarcinoma
LC110	69	M	ex	20	No	NSCLC	SCC

Table 8.3: Details of the cancer cohort

Case ID	Age	Sex	Smoking Status	Smoking History pack years	Previous Cancer?
LC03	78	M	ex	10	No
LC19	65	M	ex	100	No
LC34	84	M	ex	20	No
LC40	59	M	ex	40	No
LC48	59	M	current	20	No
LC52	42	F	current	20	No
LC53	73	M	ex	110	prostate
LC55	69	M	current	53	No
LC57	73	F	current	25	No
LC59	48	M	current	30	No
LC60	69	F	ex	50	No
LC65	60	M	current	40	No
LC67	78	M	ex	45	No
LC73	48	M	ex	20	No
LC74	58	M	ex	50	No
LC83	88	M	current	21	No
LC84	61	M	ex	35	No
LC88	82	M	ex	45	No
LC90	44	M	ex	33	No
LC96	70	F	current	25	No
LC98	84	M	ex	60	No
LC100	43	F	current	40	No
LC102	64	F	ex	40	No
LC106	64	M	ex	120	No

Table 8.4: Details of the high-risk cohort

Control ID	Age	Sex	Smoker?	Smoked period longer 10 yrs
SC01	58	F	N	N
SC04	24	M	N	N
SC05	53	M	Y	Y
SC06	79	M	Y	Y
SC07	39	M	N	N
SC08	61	M	Y	N
SC09	59	M	Y	N
SC10	67	F	N	N
SC12	35	F	Y	Y
SC13	55	M	N	N
SC14	47	M	N	N
SC15	45	F	Y	N
SC16	76	F	N	N
SC17	55	M	Y	N
SC18	44	M	N	N
SC19	68	M	Y	N
SC20	30	M	N	N
SC21	58	M	N	N
SC22	43	M	Y	Y
SC23	55	M	Y	Y
SC24	49	M	N	Y
SC25	26	F	Y	N
SC26	66	F	N	N
SC27	43	F	Y	Y
SC28	32	M	N	N
SC29	50	M	Y	Y
SC30	38	M	N	N
SC32	83	F	Y	Y
SC34	79	F	N	N
SC35	75	F	Y	Y
SC36	66	F	N	N
SC37	56	F	Y	Y
SC39	30	F	Y	N
SC41	50	F	N	N
SC42	38	M	N	N
SC43	61	F	Y	Y
SC44	39	F	N	N
SC45	41	M	Y	Y
SC46	27	M	N	N
SC47	59	M	Y	Y
SC48	41	F	N	N
SC49	72	F	Y	Y
SC50	43	F	Y	Y
SC51	35	F	N	N
SC52	26	F	Y	N
SC53	71	F	N	N
SC54	69	M	Y	Y
SC55	39	F	N	N
SC56	64	M	N	N
SC57	62	M	Y	Y
SC58	55	M	N	N
SC59	39	F	N	N
SC60	51	M	Y	Y
SC61	43	M	N	N
SC62	54	F	N	N

Table 8.5: Details of the healthy control cohort

Estimating the proportion of bronchial epithelial cells per sputum sample

All samples, regardless of group, contained bronchial epithelial cells within the section. Although difficult to predict (given variation in cell pellet size), we estimated that the likely concentration of bronchial epithelial cells per sample would lie between 10,000 and 100,000.

As a quality control check, we counted 1000 total cells from sections created from randomly selected sputum sample pellets (40 lung cancer, 36 high risk and 40 healthy controls from the original 169 study participants). The proportion of bronchial epithelial cells for that sample was then estimated from this total count. The scores were recorded (Table 8.6) and comparisons were made between groups for mean counts (Table 8.7) using the Mann-Whitney U Test.

Samples	N	Induced	Mean %	SE	Range (%)
Cancer	20	Yes	3.8	1.0	2.1-8.4
	20	No	3.6	0.9	3.0-7.6
High-risk	13	Yes	4.0	1.0	1.4-10.2
	23	No	3.8	1.0	2.9-10.6
Control	40	No	3.3	1.1	0.8-8.2

Table 8.6: Mean counts for bronchial epithelial cells in groups: cancer, high-risk patients and healthy control

Group comparison	<i>p</i> -value
cancer v. high-risk	0.5057
cancer v. healthy control	0.3547
high-risk v. healthy control	0.6395
cancer induced v. cancer non-induced	0.7704
non-cancer induced v. non-cancer non-induced	0.9352

Table 8.7: Statistical comparison of groups for bronchial epithelial counts

Table 8.7 shows that when all groups were compared there was no statistically significant difference in cell counts between any groups for mean cell count. Importantly, Table 8.7 also shows there was no significant difference within the cancer (or high-risk) groups when the sputum had been obtained by induction / endobronchial suction or just coughed up naturally (self-expectorated).

FTIR ANALYSIS OF SPUTUM

FTIR spectra were generated for all 63 cancer (54 histological and 9 clinical diagnosis), 36 high-risk and 70 healthy control cases. FTIR processing and analysis was repeated three times for each case as a measure of internal consistency. Spectra contained data for 1763 data points within the mid infrared wavenumber range 599.75 and 3997.7 cm^{-1} range. All spectra were standardised (sometimes referred to as normalised) by zero centering using the 2264 cm^{-1} wavenumber as this always approximates 0 in absorbance in studies on human tissue. Certain spectra were excluded as described earlier, leaving 54 cancer, 24 high-risk and 54 healthy control spectra.

Assessing the reproducibility of FTIR on sputum samples

All samples were processed in triplicate and spectra were assessed for similarity in pattern of wavenumber absorbance using principal component analysis (PCA). PCA allowed a visual inspection of the similarity of replicates on a case by case basis using a scatterplot of scores for each spectrum on the first two principal components. An example of such an analysis on 50 randomly selected triplicate spectra is shown in Figure 8.1.

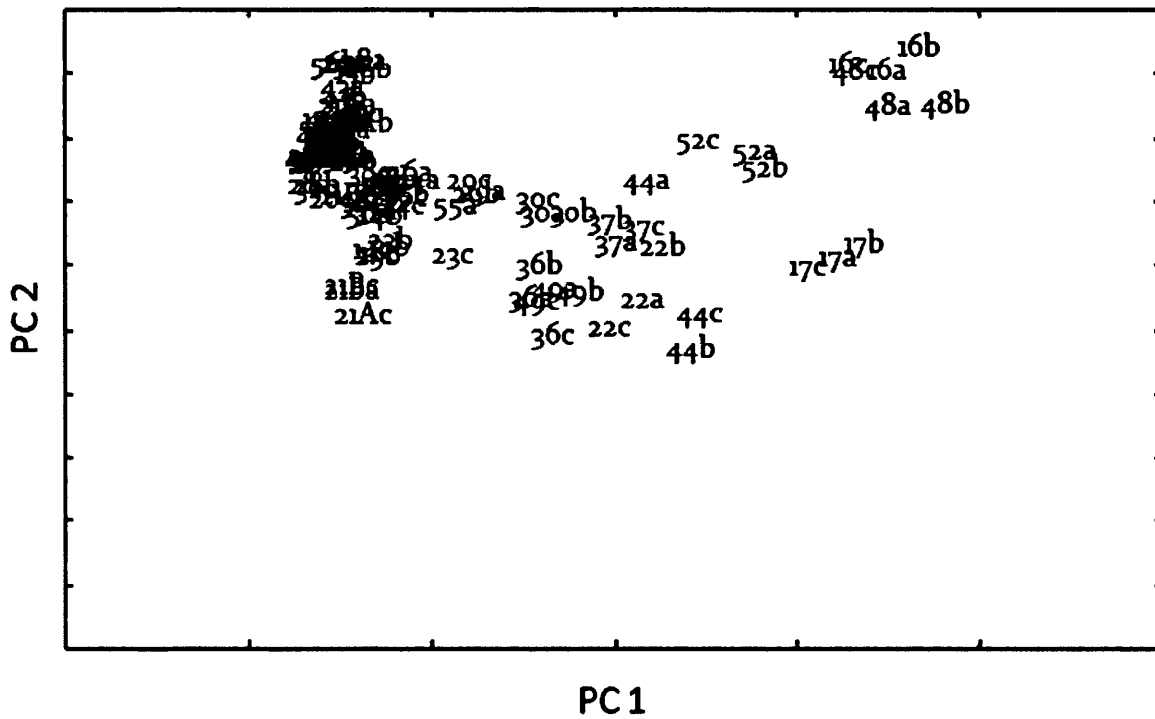


Figure 8.1: PCA of example set of triplicate FTIR spectra. Triplicates of 50 samples are shown. For example, samples 17a, b and c from the same sputum, aggregate together

Using this type of approach we observed that over all 507 FTIR runs, only 16 samples showed a replicate that deviated in position from other replicates in a scatterplot. Even then, these samples showed close similarity by position for the remaining two replicates and deviations were not consistent over further runs. An example of such a deviation is shown in Figure 8.1 by the replicate 44a which is observed in a different position to the other replicates 44b and 44c. Using this approach we were able to establish that the overall reproducibility for FTIR in producing similar patterns for the same samples within the same run and between runs was at least 99%.

Analysis of FTIR spectra from cancer versus healthy control sputa

Representative raw FTIR spectra were selected for the 54 cancer and 54 healthy sputum samples (Figure 8.2). The median absorbances for each wavenumber across the 3 cohorts based on the raw spectra shows potential wavenumbers with potential differences between the groups. (Figure 8.3)

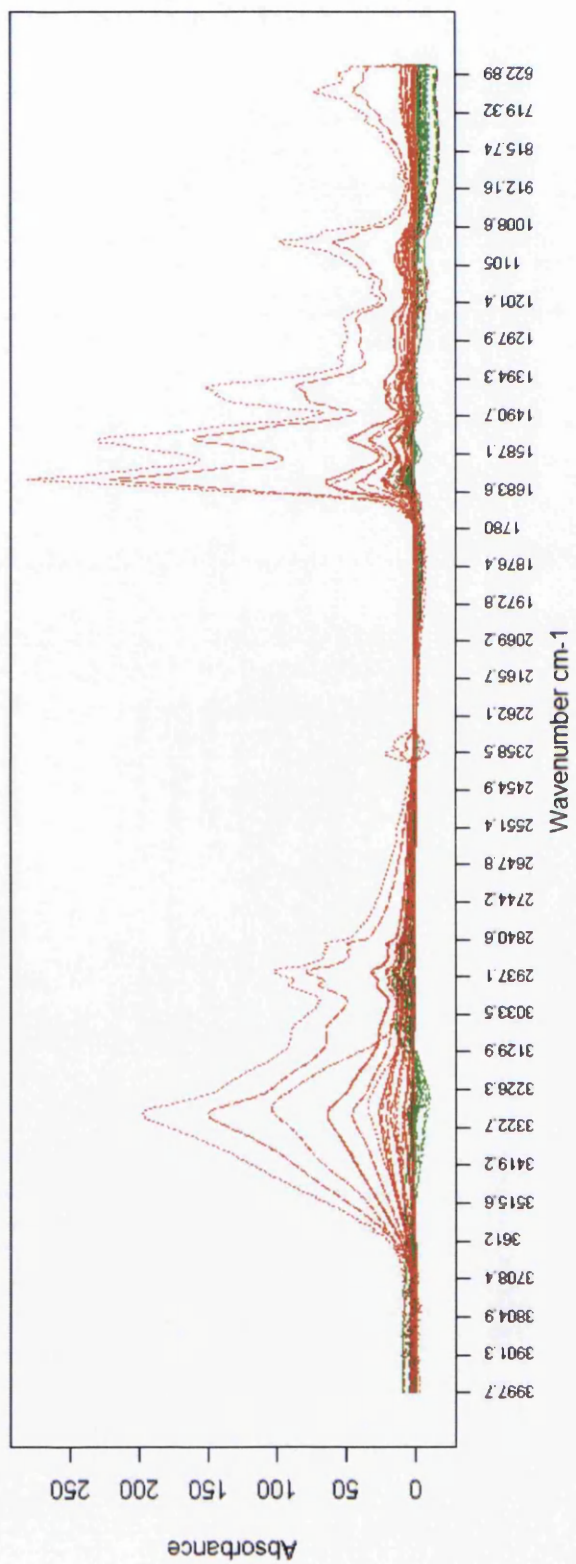


Figure 8.2: FTIR spectra of cancer and healthy controls within the wavenumber range 599.75 and 3997.7 cm^{-1} . Red lines = cancer; green lines = healthy controls

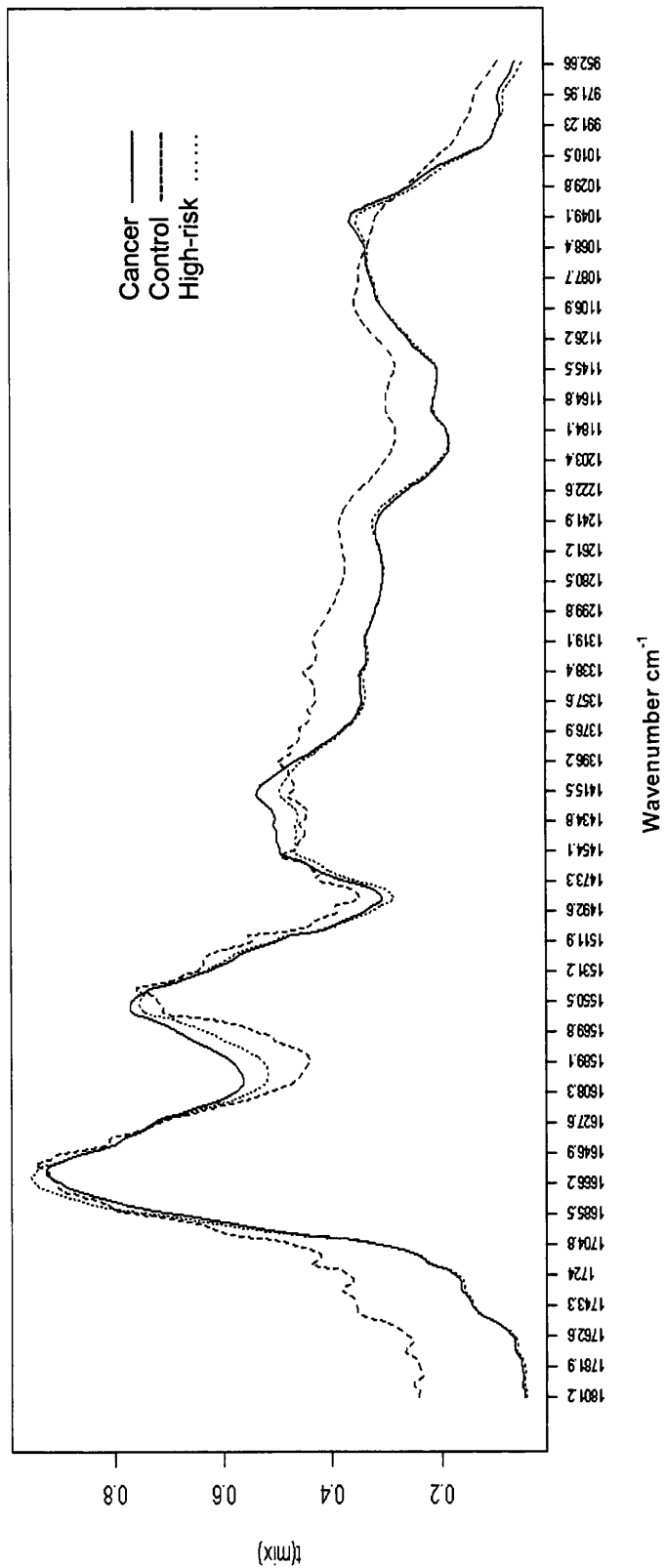


Figure 8.3: Median absorbance values for the wavenumber for each of the 3 groups. Cancer = solid line; healthy controls = dashed line; high-risk = dotted line.

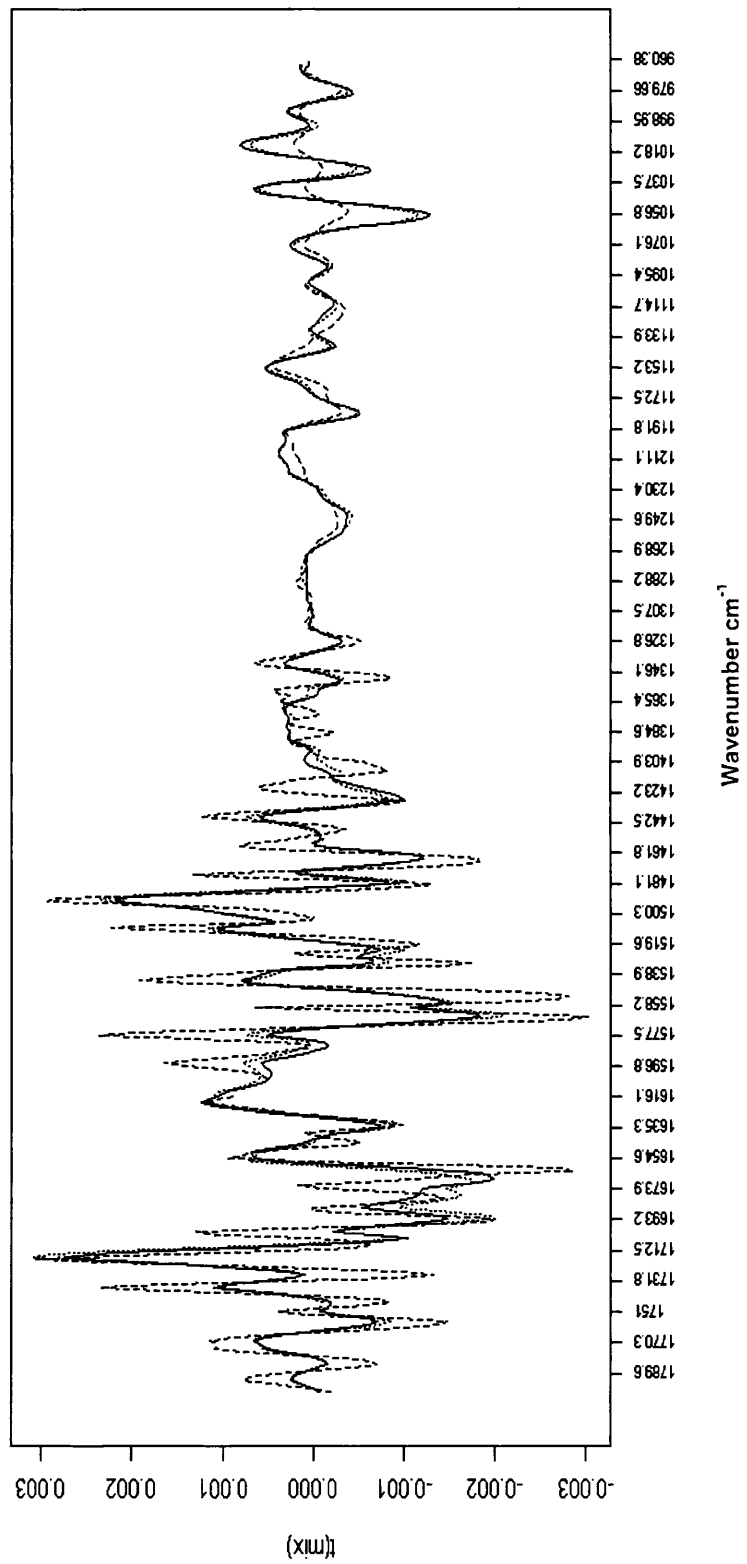


Figure 8.4: The 2nd derivative spectra for the cancer (solid line), high-risk (dotted line) and healthy controls (dashed line).

Figure 8.3 which represents the median absorbance values across the three groups show in particular how the healthy controls and cancer cases differ particularly in the regions: $1400 - 1060 \text{ cm}^{-1}$; $1060 - 1025 \text{ cm}^{-1}$ and $1400 - 1670 \text{ cm}^{-1}$. Our spectral images show that many cancer spectra differed from healthy control spectra at a number of wavelength ranges. More specific spectral differences for shorter ranges of wavenumbers can be identified using the 2nd derivative spectra represented by Figure 8.4. The wavenumbers below 900 cm^{-1} usually display a degree of variability so are difficult to interpret.

Principal component analysis to determine how the 3 cohorts FTIR spectra cluster

Prior to determining which wavenumbers differed by absorbance between groups, an overall assessment of the similarity between all spectra, regardless of group, using PCA, was made. This gave an idea of how many groupings of similar spectra existed and whether there was a separation between cancer, healthy controls and high-risk for these groupings. In this analysis, the first three principal components explained more than 90% of the variation in the data and were thus retained for interpretation. Loading scores of all spectra in the 3 cohorts are shown on the first 2 principal components in Figure 8.5 and 8.6. The third principal component is shown in Figure 8.7.

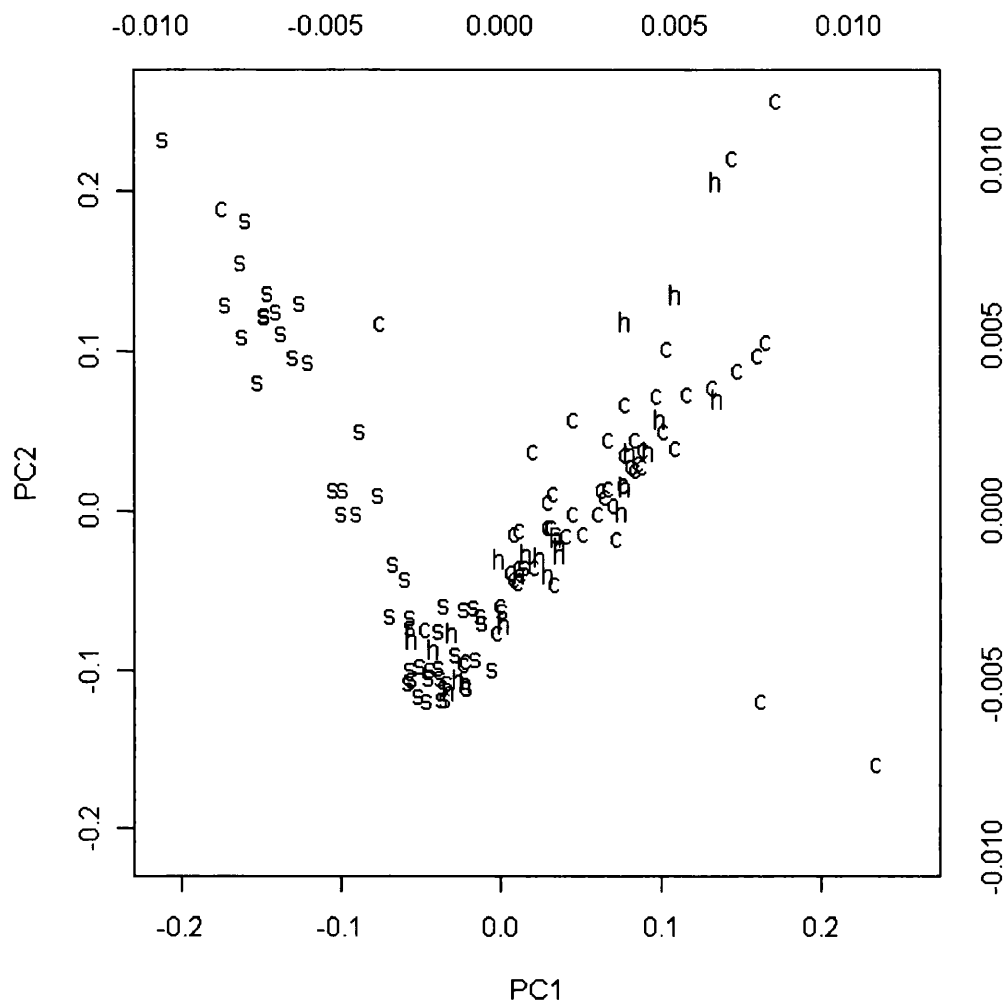


Figure 8.5: PCA of cancer (c), healthy control (s) and high-risk (h) spectra. The image shows a 2D scatterplot of the first 2 principals. Two arms representing the controls and cancers can be identified

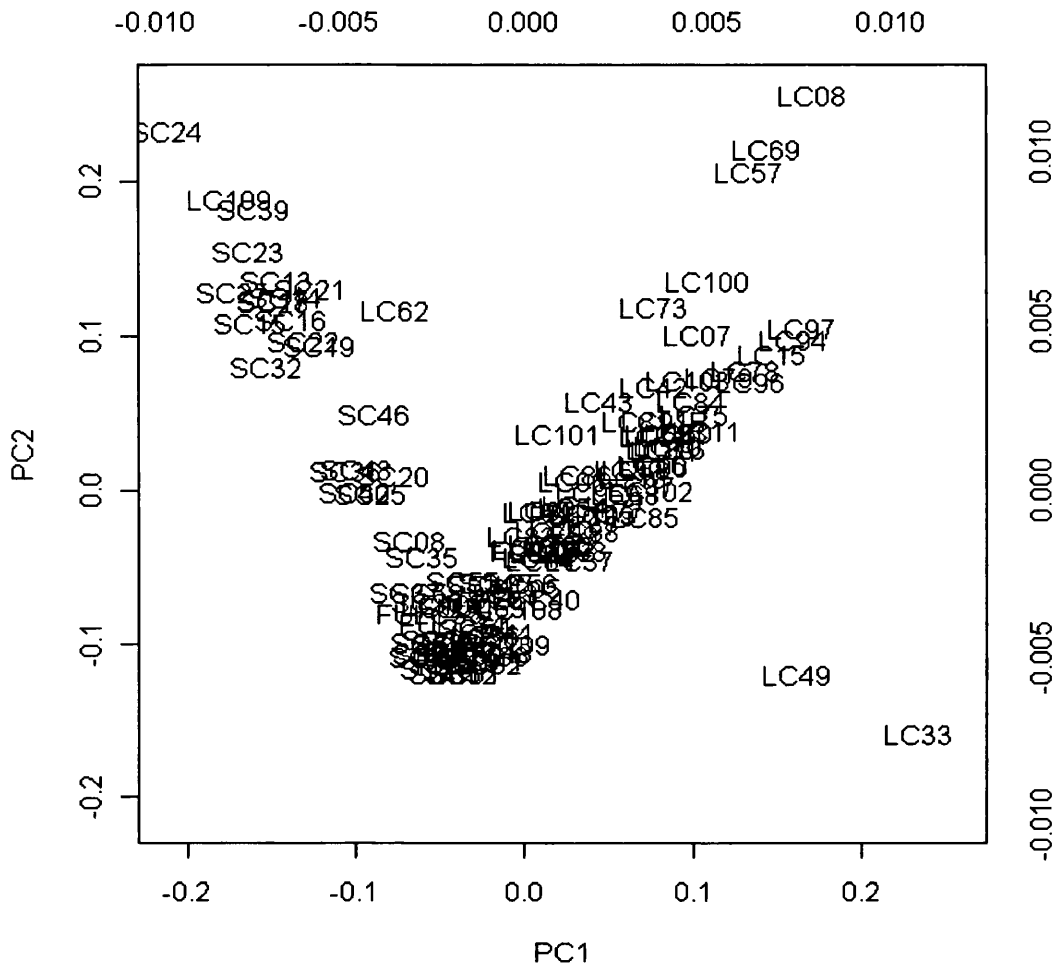


Figure 8.6: The PCA scatterplot represented in figure 8.5 but the cases are labelled with their study numbers for identification.

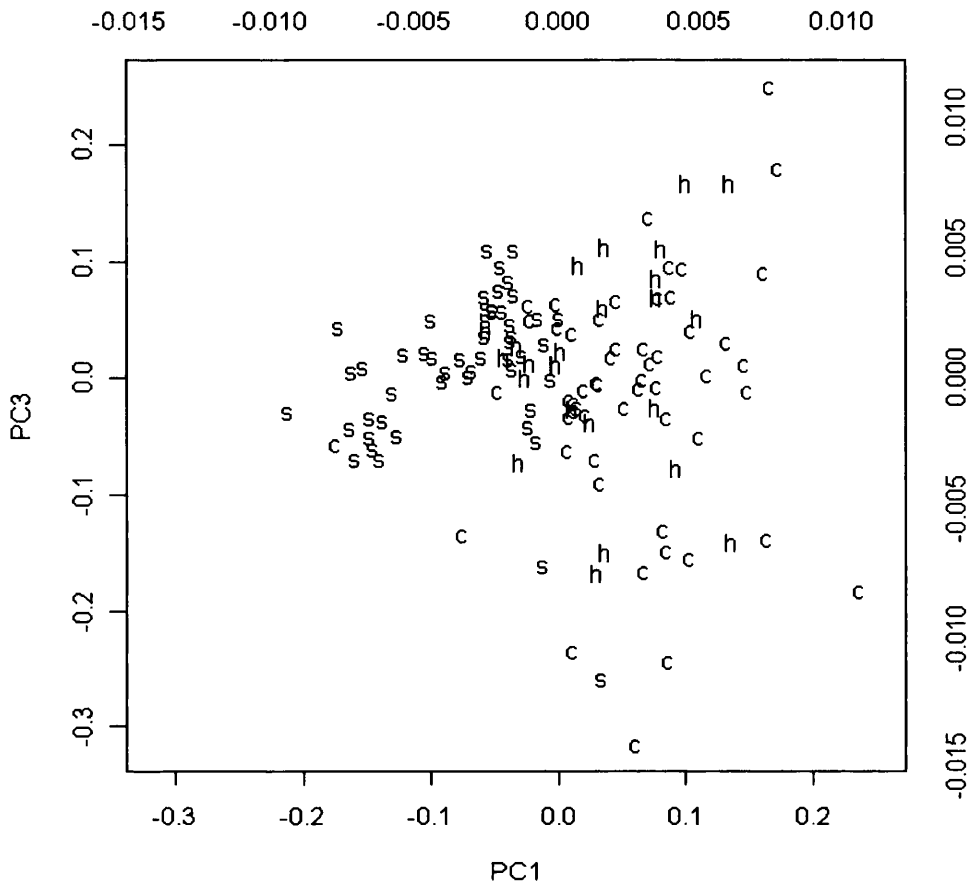


Figure 8.7: 1st and 3rd principal components showing separation of controls (s) from the cancer (c) and high-risk (h). The controls are seen to bunch closer together than the more widely spread cancer cohort.

A pattern emerged between the cancer and healthy control cases within the scatterplot. The first and second components yield a pattern whereby the cancer and healthy control cases begin to separate in different directions away from a closely clustered group of spectra from both groups. There appear to be two 'branches' of spectra separating in different directions from a general group where one branch represents cancer spectra and the other healthy controls. Two cancer cases sit away from all other cases, LC33 and LC49. It emerges that these subjects both have a history of breast cancer.

HCA also shows 3 distinct clusters (see Figure 8.8). The 1st cluster seen at the top consists predominantly of cancer cases although there are also some high-risk cases included. The middle group consists almost entirely of controls, with occasional cancer cases in the cluster. The 3rd cluster (see at the bottom) is a heterogeneous group consisting of all the cohorts. The two cancer cases with a history of breast cancer group separately from the rest and are deemed very different as they are the last to be linked with the rest of the study cohort.

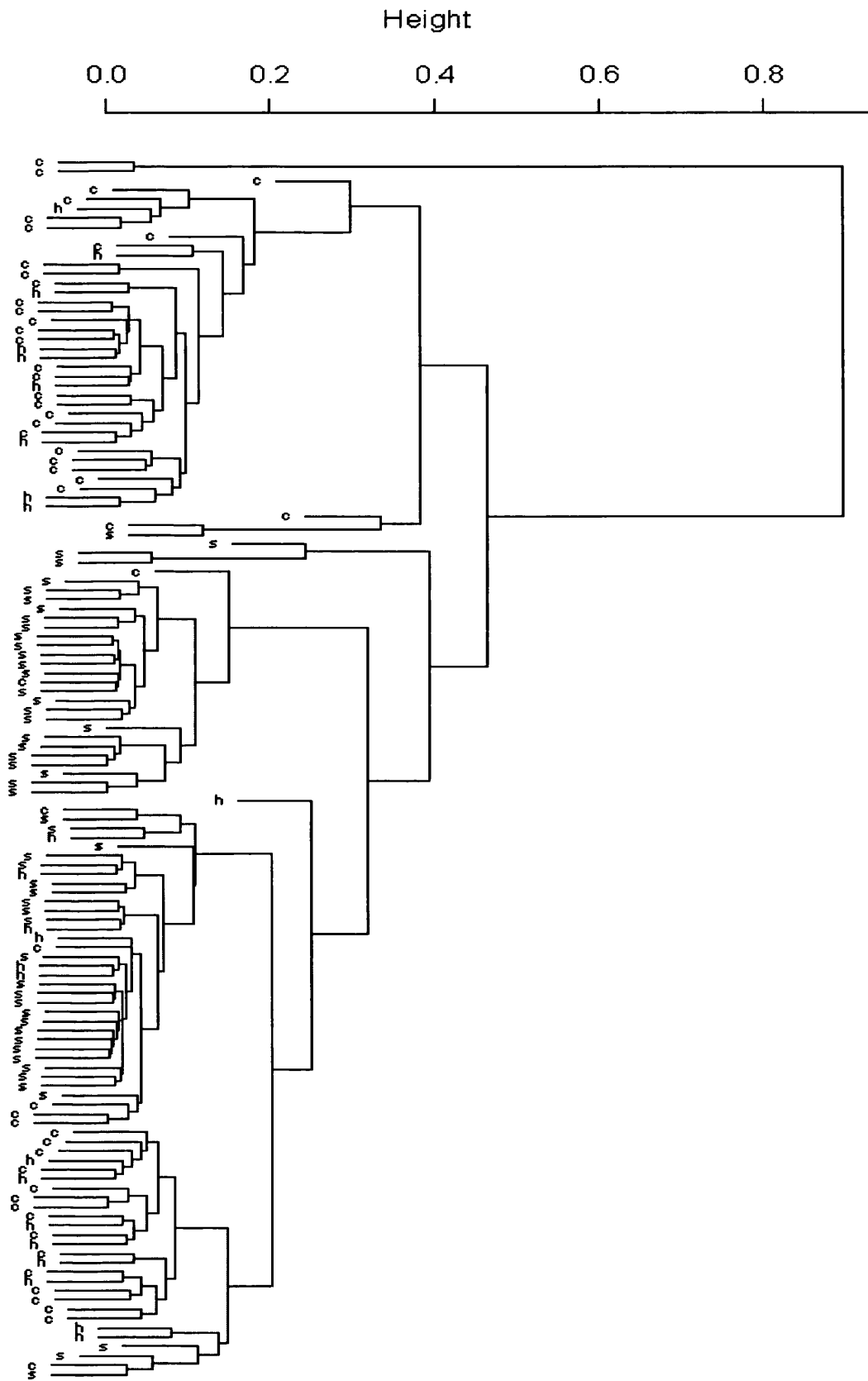


Figure 8.8: HCA of the 3 cohorts, cancer (c), high-risk (h), controls (s). The controls are seen to group together, as do many of the cancer cases with a few of the high-risk also mixed in

Although the PCA and HCA results gave an encouraging sign as to separation especially between cancer and healthy control spectra the findings were based on all wavenumbers and the next step was to determine which wavenumbers were significantly different between the cancer and control groups. In the initial analysis and modelling we used only the cancer and healthy control subjects as the high-risk may have underlying metabolic changes which could introduce confounders. The cancer and control groups represented the “cleanest” data.

Determination of discriminating wavenumbers between cancer and healthy controls

The initial step in analysis of individual wavenumbers was to determine the corresponding data distributions. Using the Shapiro Wilk test for normality, the distribution of the median absorbance values for every wavenumber was found to be non-normally distributed. Thus, differences within each wavenumber between cancer and healthy control groups were assessed using the non-parametric Mann-Whitney U Test using the median absorbance value for each wavenumber for the cancer and healthy control groups. There were 624 significant wavenumbers ($p=0.05$) and after applying Holm's sequential Bonferroni correction to the list, we generated a reduced list of 126 wavenumbers. These wavenumbers and adjusted p values are shown in Table 8.8.

Wave Number	P-Value	Wave Number	P-Value	Wave Number	P-Value	Wave Number	P-Value
640.25	8.01E-05	875.52	8.7E-05	1569.8	9.51E-05	1068.4	9.9E-05
1031.7	8.03E-05	863.95	8.71E-05	1201.4	9.52E-05	960.38	9.92E-05
655.68	8.04E-05	1589.1	8.73E-05	1639.2	9.54E-05	860.1	9.94E-05
1392.4	8.05E-05	1151.3	8.74E-05	1307.5	9.56E-05	740.53	9.96E-05
1409.7	8.06E-05	877.45	8.76E-05	869.74	9.58E-05	910.24	9.98E-05
1413.6	8.08E-05	1332.6	8.77E-05	663.39	9.6E-05	842.74	0.0001
1411.6	8.09E-05	1575.6	8.79E-05	1400.1	9.62E-05	790.67	0.0001
1394.3	8.1E-05	1637.3	8.8E-05	692.32	9.63E-05		
1049.1	8.12E-05	1635.3	8.82E-05	1147.4	9.65E-05		
1045.2	8.13E-05	1386.6	8.83E-05	1064.5	9.67E-05		
661.46	8.14E-05	879.38	8.85E-05	1571.7	9.69E-05		
1390.4	8.16E-05	997.02	8.87E-05	1382.7	9.71E-05		
1047.2	8.17E-05	1035.6	8.88E-05	1344.1	9.73E-05		
1033.7	8.18E-05	644.11	8.9E-05	1012.4	9.75E-05		
1407.8	8.2E-05	1027.9	8.91E-05	1556.3	9.77E-05		
1051	8.21E-05	1429	8.93E-05	1419.4	9.78E-05		
1018.2	8.22E-05	1022.1	8.94E-05	1417.4	9.8E-05		
1396.2	8.24E-05	1421.3	8.96E-05	1199.5	9.82E-05		
1020.2	8.25E-05	1205.3	8.98E-05	651.82	9.84E-05		
1016.3	8.26E-05	1041.4	8.99E-05	1313.3	9.86E-05		
1043.3	8.28E-05	912.16	9.01E-05	755.96	9.88E-05		
642.18	8.29E-05	1153.2	9.03E-05	1068.4	6.734E-05		
1004.7	8.31E-05	1633.4	9.04E-05	960.38	6.752E-05		
659.54	8.32E-05	1311.4	9.06E-05	860.1	4.901E-05		
1388.5	8.33E-05	865.88	9.07E-05	740.53	6.717E-05		
1002.8	8.35E-05	1384.6	9.09E-05	910.24	6.787E-05		
788.74	8.36E-05	1149.4	9.11E-05	842.74	4.697E-05		
657.61	8.38E-05	784.89	9.12E-05	790.67	7.283E-05		
1581.3	8.39E-05	862.02	9.14E-05	1569.8	3.201E-05		
1398.1	8.4E-05	796.46	9.16E-05	1201.4	6.682E-05		
1052.9	8.42E-05	1554.3	9.17E-05	1639.2	6.7E-05		
786.81	8.43E-05	1203.4	9.19E-05	1307.5	4.911E-05		
1000.9	8.45E-05	1209.1	9.21E-05	869.74	7.717E-05		
1587.1	8.46E-05	881.31	9.23E-05	663.39	8.052E-05		
1579.4	8.47E-05	653.75	9.24E-05	1400.1	7.242E-05		
1577.5	8.49E-05	867.81	9.26E-05	692.32	6.648E-05		
1006.7	8.5E-05	1309.4	9.28E-05	1147.4	6.665E-05		
1585.2	8.52E-05	1591	9.29E-05	1064.5	1.547E-05		
1583.3	8.53E-05	1573.6	9.31E-05	1571.7	6.805E-05		
1207.2	8.55E-05	1334.5	9.33E-05	1382.7	7.263E-05		
1014.4	8.56E-05	1631.5	9.35E-05	1344.1	1.548E-05		
998.95	8.58E-05	995.09	9.36E-05	1012.4	1.549E-05		
1405.9	8.59E-05	1155.2	9.38E-05	1556.3	8.077E-05		
638.32	8.61E-05	1066.4	9.4E-05	1419.4	1.575E-05		
1425.1	8.62E-05	1054.9	9.42E-05	1417.4	1.573E-05		
1415.5	8.64E-05	1025.9	9.43E-05	1199.5	3.333E-05		
1423.2	8.65E-05	914.09	9.45E-05	651.82	1.576E-05		
1427.1	8.67E-05	798.38	9.47E-05	1313.3	1.546E-05		
1029.8	8.68E-05	1346.1	9.49E-05	755.96	6.823E-05		

Table 8.8: 126 significant wavenumbers after Holm's sequential Bonferroni p value correction ($p < 0.0001$)

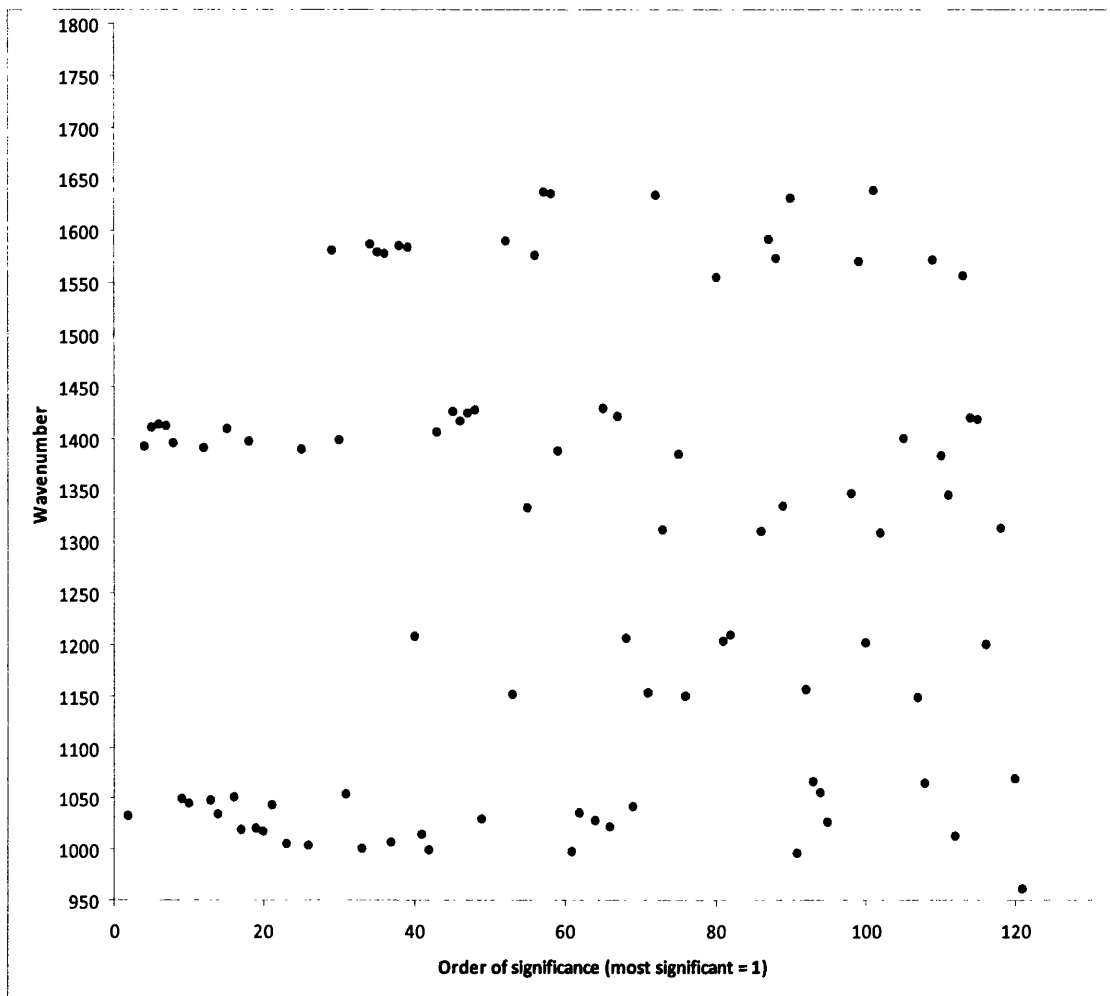


Figure 8.9: Representation of the 126 significant wavenumbers in order of significance (i.e. 1 = most significant wavenumber)

Figure 8.9 shows the 3 main regions which are represented by significant wavenumbers. These involve 2 areas within $1400 - 1700 \text{ cm}^{-1}$ (which represents proteins including amide I and amide II regions). The other region is from $1000 - 1080 \text{ cm}^{-1}$ which includes the glycogen-rich and DNA / RNA regions (for example around 1080 cm^{-1} is the phosphodiester groups of nucleic acid). Lipids have typical absorbance values greater than 3000 cm^{-1} . [194, 207, 284, 285]

Determination of optimal pair of wavenumbers that discriminate between cancer and healthy controls.

The list of 126 wavenumbers was used to find a pair of wavenumbers that could discriminate between cancer and healthy controls with the highest accuracy.

Using logistical regression described in the statistical methods in Chapter 6, we determined that the most optimal discriminating pair of wavenumbers between cancer and healthy controls was 1031.7 cm^{-1} and 1409.7 cm^{-1} . The 1031.7 cm^{-1} wavenumber falls within a region assigned to C-O stretching coupled with bending of C-OH groups of glycogen. The 1409.7 cm^{-1} wavenumber signifies a carboxylate (COO^-) group typically found on amino acids (and thus protein).

The logistical regression model using 1031.7 cm^{-1} and 1409.7 cm^{-1} was plotted as in Figure 8.10. The cancer cases (circles) can be seen to be more widely scattered above the discriminating line and this may well represent biological differences. The controls (triangles) appear to be less scattered and more grouped together. The false positive (controls which appear on the cancer side) lie in close proximity to the discriminating line.

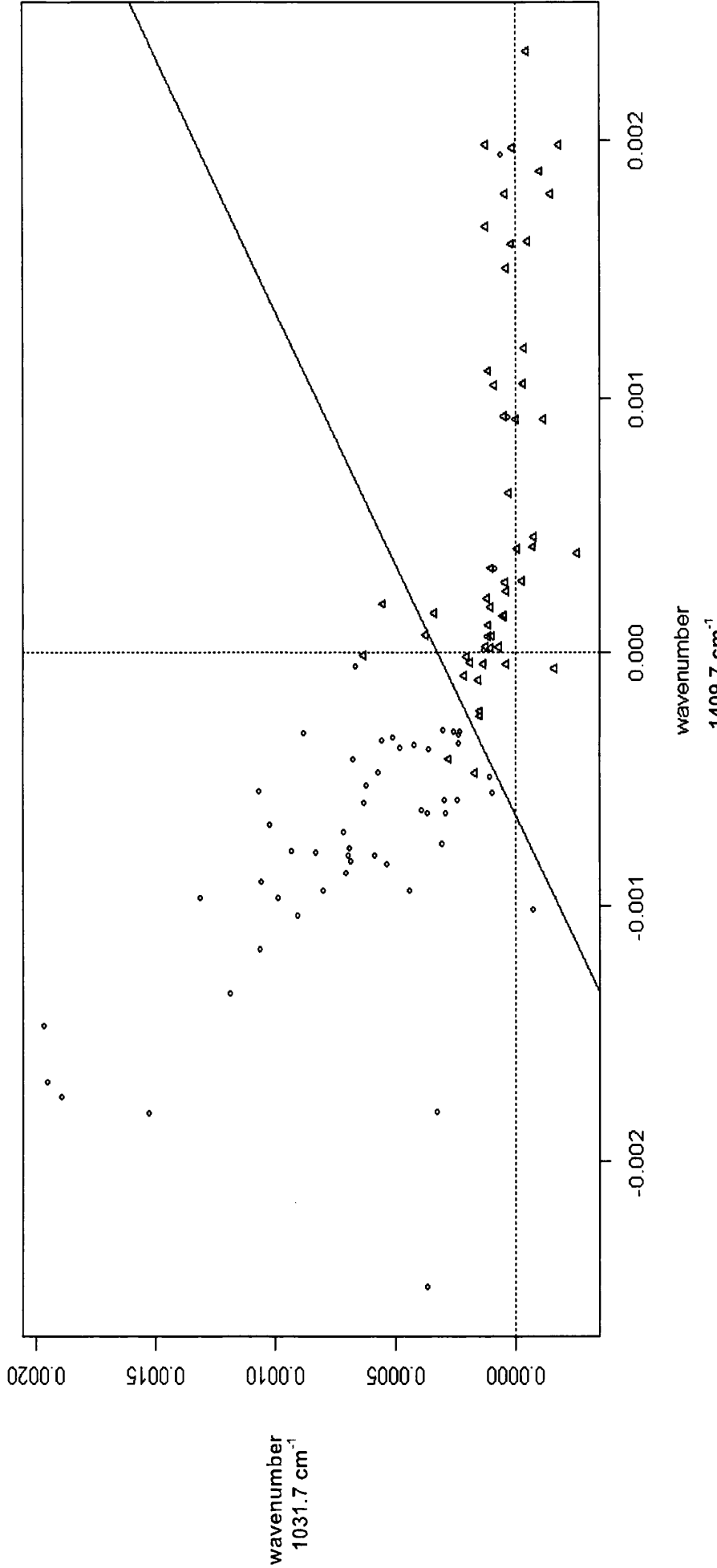


Figure 8.10: Discrimination line that maximally discriminates cancer (circles) and healthy controls (triangles)

Overall sensitivity and specificity was measured for the model based on the data used to generate the model (i.e. the training data set). Numbers of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are shown in Table 8.9. The sensitivity $[TP/(TP+FN)]$ of the model was 92.6% and the specificity $[TN/(TN+FP)]$ was 90.7% when comparing lung cancer against *healthy controls*.

The strength of the model was assessed by cross-validation. The results of the leave-one-out cross validation are in Figure 8.11. The leave-one-out technique shows that the using this particular pair of wavenumbers continues to give a consistently high level of sensitivity, at close to 90% with little variation.

		Predicted	
		Cancer	Non-cancer
Actual	Cancer	50 (TP)	4 (FN)
	Non-cancer	5 (FP)	49 (TN)

Table 8.9: Calculating sensitivity / specificity for cancer and healthy control cases for our discriminatory model

Histogram of correct

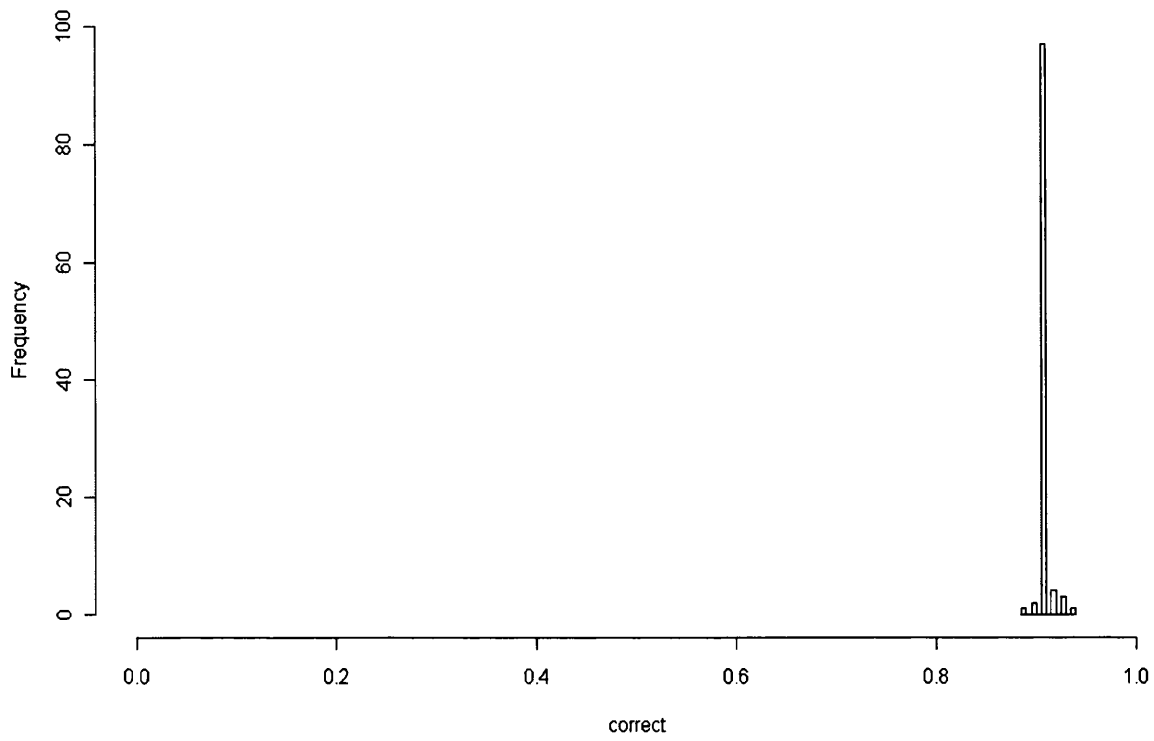


Figure 8.11: The leave-one-out cross validation of sensitivity

Analysis of FTIR spectra of high-risk cohort using predictive model

Once the model had been developed using the wavenumbers 1031.7 cm^{-1} and 1409.7 cm^{-1} , we added the *high-risk* cases to it to assess how the model would predict their outcome based of their FTIR spectra using the equations:

$$z = (-2337.0649 \times \text{value for } 1031.7 \text{ cm}^{-1}) + (4636.0538 \times \text{value for } 1409.7 \text{ cm}^{-1}) - 1.51444$$

then, probability of cancer = $1/(1+e^{-z})$

Values > 0.5 predict cancer;

Values < 0.5 predict no cancer.

There were 24 high-risk spectra included in the analysis and of those 17 (71%) would be predicted from the model to have 'cancer'.

Of these 17 cases, 3 had pneumonia, 1 had pulmonary fibrosis, 1 had B-cell lymphoma, 1 had MAI, 2 had old TB and 3 of the others had a history of asthma. The sensitivity and specificity when comparing the high-risk and cancer groups with the developed model are 92.6% and 29.3% respectively (i.e. 71% predicted to have cancer).

Table 8.10 represents the data for *all* study groups based on the discriminating model developed. The overall sensitivity and specificity of the model developed is 92.6% and 69% respectively.

		Predicted	
		Cancer	Non-cancer
Actual	Cancer	50	4
	Non-cancer	22	56

Table 8.10 Calculating sensitivity / specificity for cancer versus non-cancer (combining high-risk and healthy controls as a single group) for discriminatory model

EVALUATION OF SPUTUM WITH IMMUNOHISTOCHEMISTRY

Patient demographics

The patients consist of the same high-risk and cancer cohort described earlier in this chapter. Only 83 of the original 99 patients had enough of a sputum cell pellet to be used for IHC after the FTIR processing. Analysis was therefore limited to 51 in the lung cancer cohort and 32 in the high-risk cohort.

Of the 51 patients with lung cancer, 36 had NSCLC (16 squamous cell carcinoma, 11 adenocarcinoma, 3 large cell, 6 were of unknown subtype); 9 had small cell lung cancer; 1 eventually was proved to have carcinoid and 5 had a clinico-radiological diagnosis of lung cancer. The full demographic data of the groups are shown in Table 8.11.

	Cancer group n = 51	High-risk group n = 32	p-value
Mean Age (yrs)	66.8 +/- 8.6	62.0 +/- 14.2	0.59 CI -0.19 to 9.8
% Male	51.0	71.9	0.06*
Smoking Status			
Current	18	12	0.60*
Former	25	18	
Never	7	2	
Unknown	1	0	

Table 8.11: Demographics of cancer and high-risk groups that underwent sputum IHC analysis

* Chi-squared

Analysis of IHC scoring of the biomarkers

The 5 antibodies that were used to stain the antigens were:

Cell membrane antigen – EGFR

Nuclear antigens – cyclin D1; p16; p53; p63

Figures 8.12 and 8.13 represent the actual control and patient sample for cell membrane staining and Figures 8.14 and 8.15 represent nuclear staining with p63.

Mrs Christine Davies and I scored the first 66 slides (cases) together and then the last 17 slides independently. Statistically there was no difference in our scores using the Cohen kappa data test (Table 8.12).

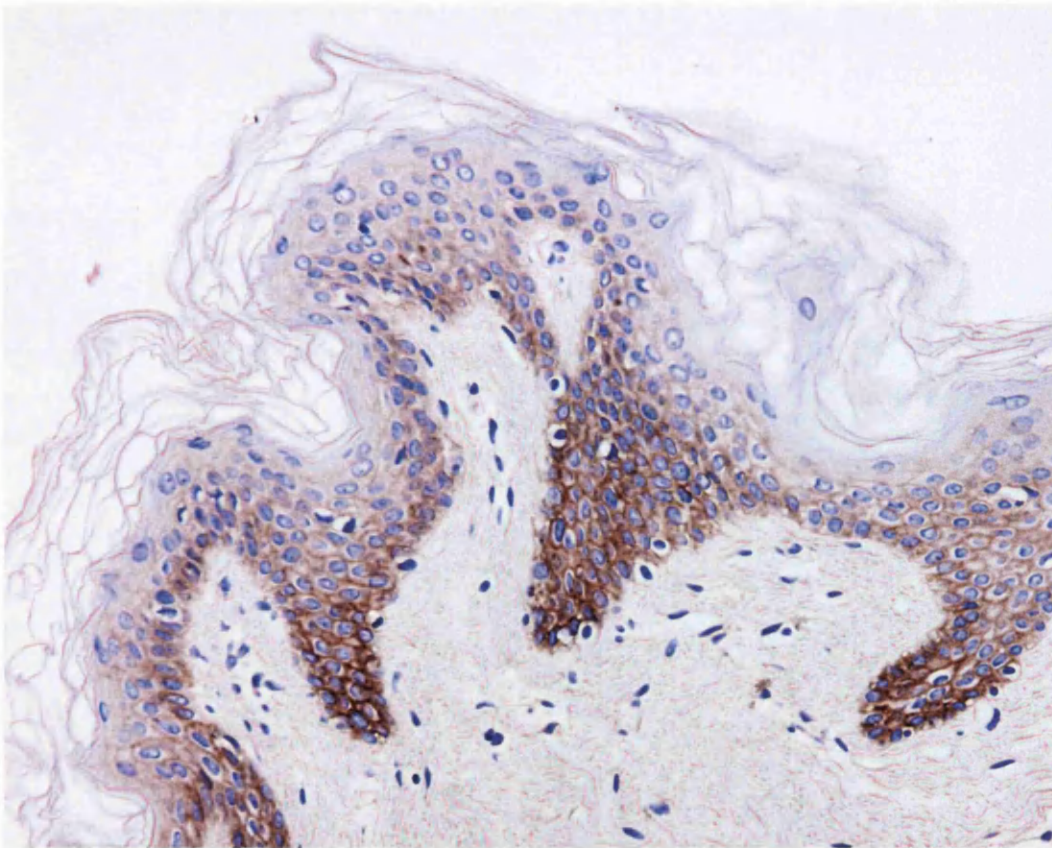


Figure 8.12: Control sample (skin) of EGFR (cell-membrane) staining

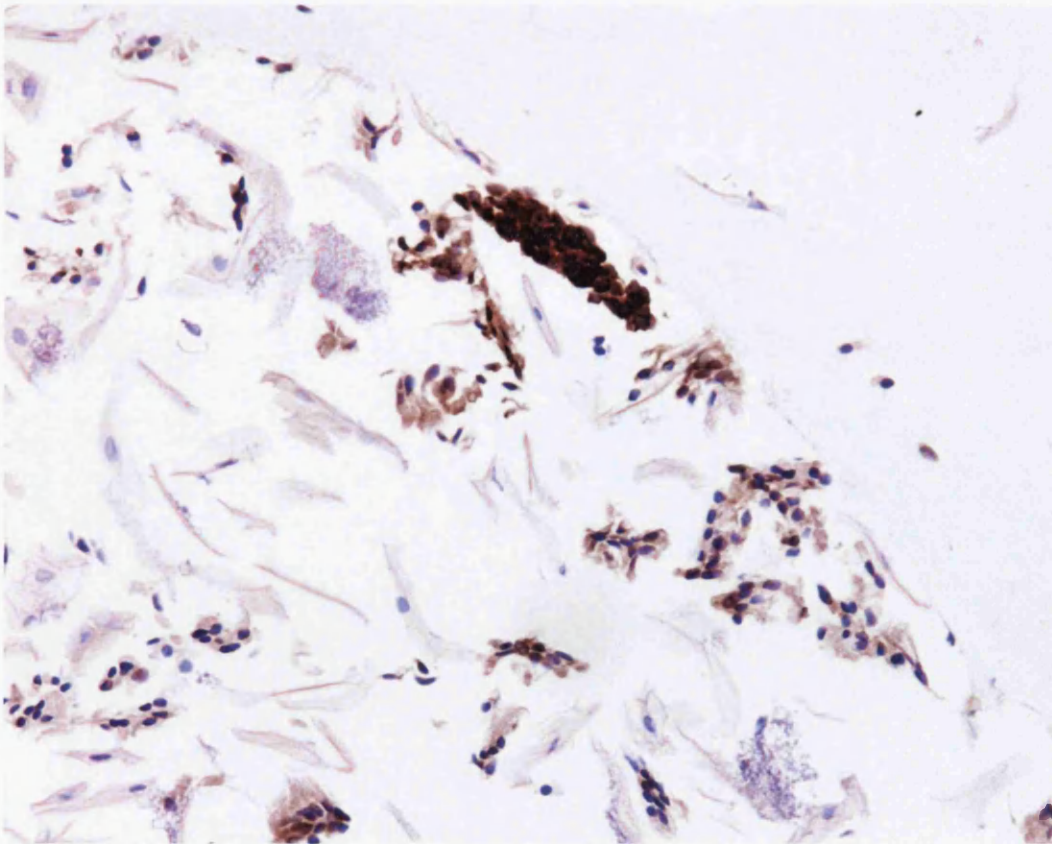


Figure 8.13: Sputum from one of our lung cancer patients staining with EGFR

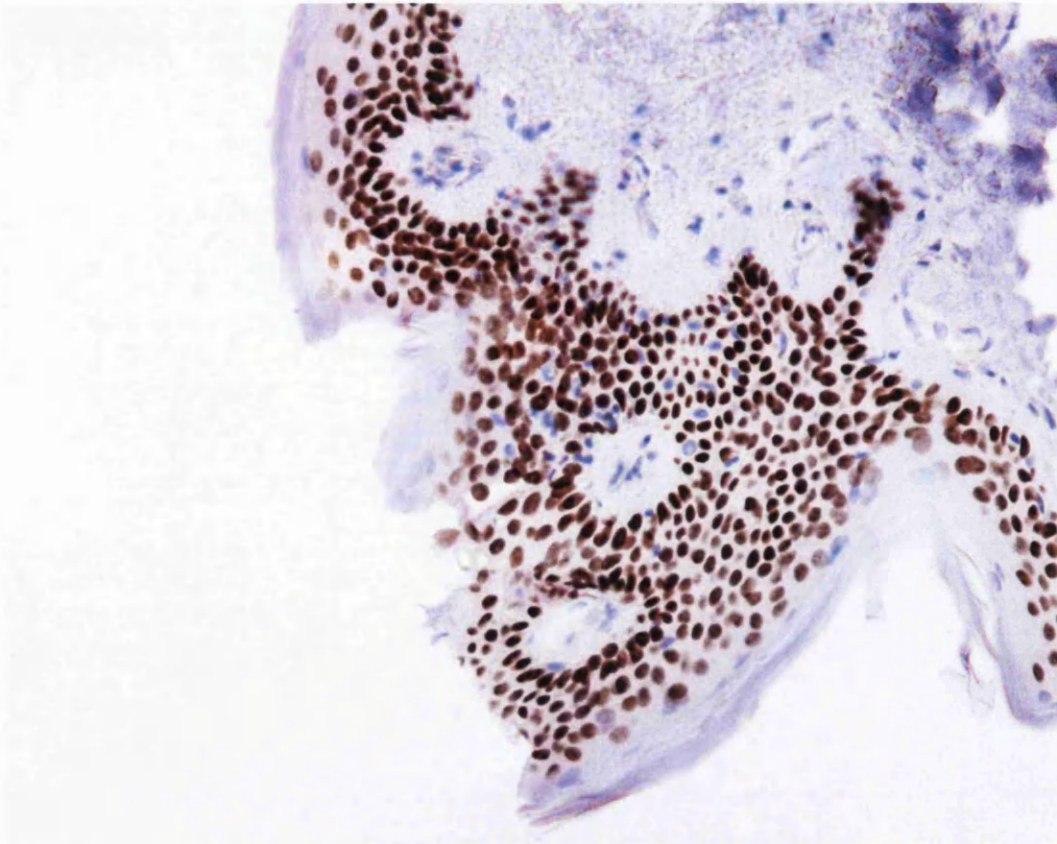


Figure 8.14: Control sample (skin) immunostained with p63 (nuclear staining)

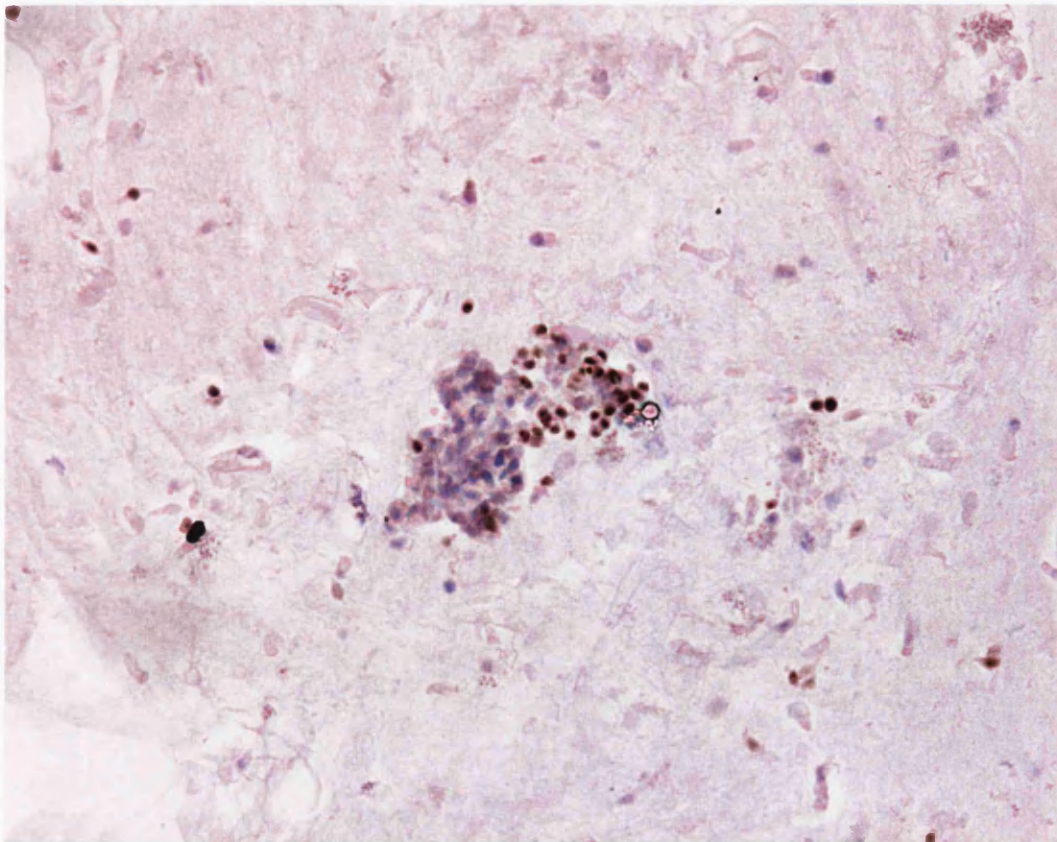


Figure 8.15: Sputum from lung cancer patient stained with p63

Antibody	Concordance between independent scorers assessed by Cohens kappa data test: <i>p</i> value
Cyclin D1	0.61
p53	0.54
EGFR	0.54
p16	0.63
p63	0.51

Table 8.12: Cohens kappa data test comparing the concordance in IHC scoring between the two independent scorers

Table 8.13 shows there was no statistical difference between the two cohorts for any of the stains on sputum cells and the sensitivities of the antibodies was very low (i.e. only a small number of the cancer sputa stained positively).

Some of the sputum slides were deemed uninterpretable (n=16) due to a paucity of bronchial cells or the sample not being of good enough quality.

Importantly, the relatively low number of slides that had positive staining gives rise to the higher specificity (i.e. very few non-cancer cases stained positively) (Table 8.14).

Antibody	Statistical difference in number of positive cells of antibody: <i>p</i>-value	Sensitivity (%)	Specificity (%)
Cyclin D1	0.58	7.8	93
P53	0.71	5.9	96.9
EGFR	0.06	35.3	87.5
P16	0.10	25.5	81.3
P63	0.93	17.6	78.1

Table 8.13: Differences between the cancer and high-risk cohort sputum staining for each antibody using Mann-Whitney U Test

Antibody	Sensitivity (%)	Specificity (%)
Cyclin D1	9.8	93
P53	7.7	96
EGFR	41.9	85.7
P16	31.7	79.3
P63	23.1	72.0

Table 8.14: Sensitivity and specificity of antibodies in sputum of cancer and high-risk cohort (with the uninterpretable sputum slides removed from analysis)

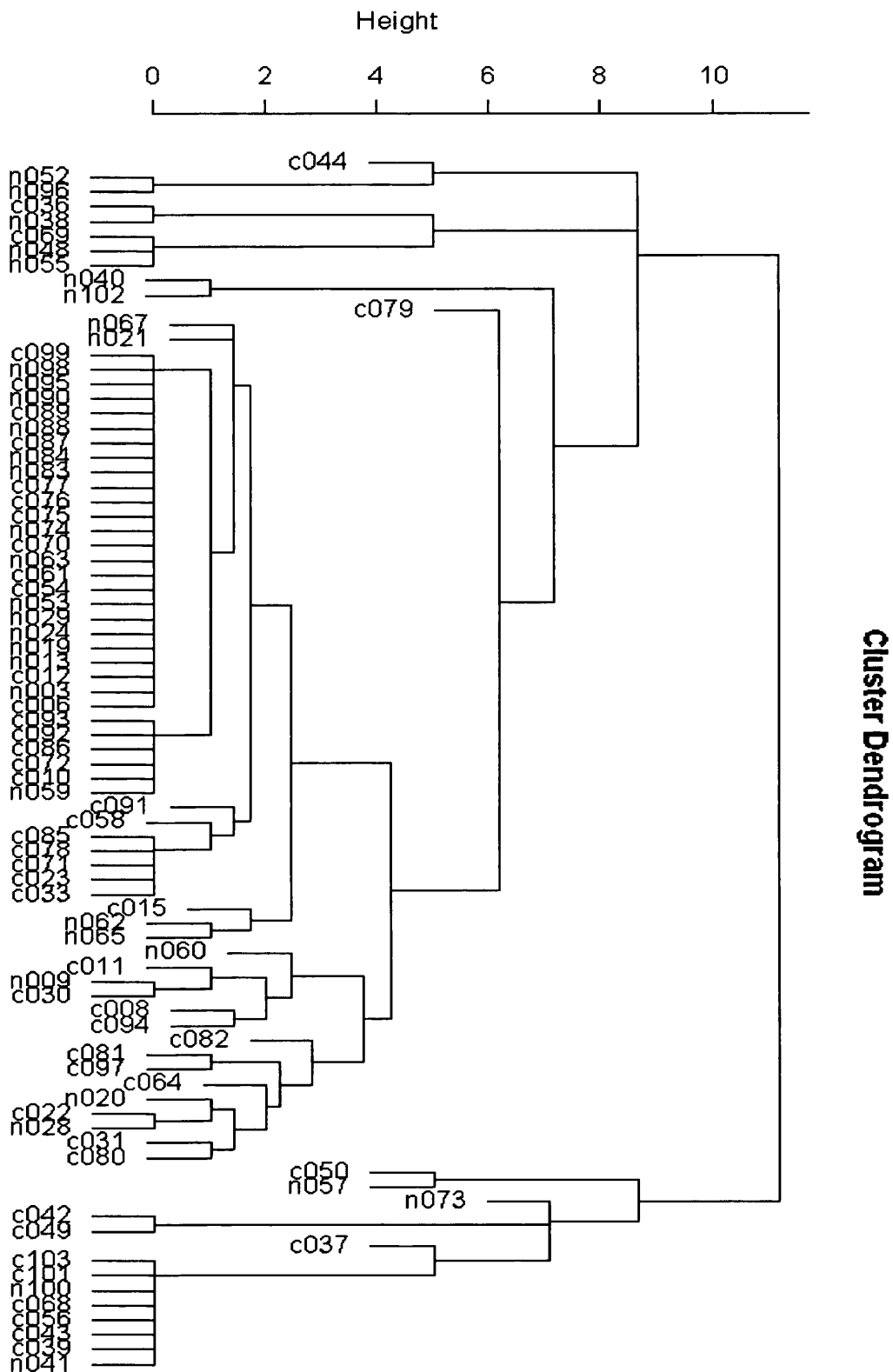


Figure 8.16: Hierarchical cluster analysis of the sputa based on IHC scores.

n = non-cancer and c = cancer

The hierarchal cluster analysis (Figure 8.16) shows similar cases that group together with the largest cluster in the middle. These very 'similar' cases are clustered since there were a large number of slides with zero positively stained cells from both the cancer and high-risk cohorts. The other clusters are a mixture of both cohorts in the majority of cases, again showing that the antibodies have not been able to show much differentiation between the cancer and non-cancer groups.

Chapter 9. Analysis of Volatile Organic Compounds in exhaled breath in lung cancer

Patient Demographics

57 patients who had a provisional diagnosis of lung cancer were approached. 5 patients refused participation, none (approached) were unable to give informed consent and one patient was unable to perform the test with adequate technique so was excluded. Of the 51 patients recruited into the study, 3 more were excluded; 1 patient was thought to have endobronchial disease (squamous dysplasia noted on biopsy) but subsequently had two normal bronchoscopies; 2 other patients had their histology (lymph node biopsies) re-examined and it was felt that they had metastatic breast cancer rather than a primary lung cancer. 4 further patients were excluded from analysis as their final VOC profiles had less than 20 peaks, which was significantly less than the other subjects. The reasons for this are discussed below.

50 subjects were approached to be healthy controls. All agreed to be included in the study but 3 were excluded from analysis as their final VOC profiles again had less than 20 peaks.

A review of the 7 patients with less than 20 peaks on their VOC profile revealed their samples were all processed at the same time. On this occasion, the GC-MS had developed a pressure loss in the helium supply during the cold trap secondary desorption phase leading to a leak within the system.

The sociodemographic data comparison of the cancer cohort (n=44) and healthy controls (n=47) is shown in Table 9.1.

	Cancer n = 44	Healthy controls n = 47	p-value
Mean Age (yrs)	70.5 +/- 8.1	52.1 +/- 11.6	< 0.001**
% Male	45.5	34	0.27*
Smoking Status			
Current	19	9	< 0.001*
Former	22	14	
Never	3	24	
% smoked within 24 hours of sampling (eCO ≥ 10 ppm)	15	12	0.83*

Table 9.1: Comparison of baseline demographic data of the cancer and healthy control cohorts.

** Data normally distributed. 95% CI for difference 14.2 to 22.5

* chi-squared

Diagnosis of cancer was confirmed histo-cytologically in 35 cases and clinico-radiologically in 9. Of the 44 patients with a diagnosis of primary lung cancer (cancer group), 31 were NSCLC, 4 were small cell lung cancer and 9 were clinical diagnosis. Table 9.2 has the histological breakdown of the cancers. Table 9.3 has a summary of the characteristics of the cancer cohort, and Table 9.4 of the healthy control group.

Lung Cancer Type	Histologic Sub-Type	Number
Non-small cell (NSCLC)	Squamous Cell Carcinoma (SCC)	10
	Adenocarcinoma	6
	Large cell	9
	Unknown subtype	6
Small Cell (SCLC)		4
Lung – clinical diagnosis		9
	Total	44

Table 9.2: Breakdown of histological subtypes within the cancer cohort.

Study No.	Sex	Age	Hx of Cancer [*]	Smoking Hx (pack yr)	smoking status	Diagnosis	Cancer subtype
S01	m	68	N	50	current	SCLC	small
S02	f	62	N	30	current	NSCLC	squamous
S03	f	64	N	20	ex	NSCLC	adenocarcinoma
S04	f	65	N	90	ex	Ca (radiological)	
S05	m	64	N	100	current	Ca (radiological)	
S06	m	78	Y - skin	120	ex	NSCLC	squamous
S07	f	71	Y - breast	40	ex	NSCLC	unknown
S08	f	57	N	0	never	SCLC	small
S09	f	67	N	25	ex	NSCLC	large cell
S10	f	69	N	40	current	NSCLC	large cell
S12	f	73	N	50	current	Ca (radiological)	
S14	f	61	N	45	current	NSCLC	squamous
S15	m	64	N	100	ex	NSCLC	large cell
S16	f	82	Y - breast	64	current	NSCLC	unknown
S18	f	76	N	8	ex	NSCLC	squamous
S19	f	73	N	90	ex	NSCLC	large cell
S20	m	67	N	40	ex	NSCLC	squamous
S21	f	63	N	20	ex	NSCLC	adenocarcinoma
S23	m	78	N	25	ex	Ca (radiological)	
S24	f	74	N	60	current	Ca (radiological)	
S25	f	71	N	8	current	NSCLC	squamous
S26	m	78	Y - bladder	65	current	NSCLC	squamous
S27	m	65	N	100	ex	NSCLC	large cell
S28	m	69	N	75	ex	NSCLC	unknown
S29	f	67	N	40	current	NSCLC	adenocarcinoma
S30	m	64	N	100	current	NSCLC	adenocarcinoma
S31	f	62	Y - skin	20	ex	NSCLC	adenocarcinoma
S32	m	59	Y - leukaemia	25	current	NSCLC	squamous
S33	f	68	N	40	current	NSCLC	squamous
S34	m	81	N	60	ex	Ca (radiological)	
S35	f	75	N	0	never	Ca (radiological)	
S36	f	78	N	64	current	Ca (radiological)	
S37	m	77	N	15	current	NSCLC	large cell
S38	f	73	N	20	ex	NSCLC	large cell
S39	m	75	N	30	ex	SCLC	small cell
S40	m	79	N	70	ex	NSCLC	unknown
S41	f	61	N	15	current	NSCLC	large cell
S42	m	77	Y - leukaemia	30	current	NSCLC	squamous
S43	m	46	N	30	current	SCLC	small cell
S44	m	82	Y - prostate	45	ex	NSCLC	adenocarcinoma
S45	m	73	N	100	ex	NSCLC	unknown
S46	m	85	Y - breast	0	never	Ca (radiological)	
S49	f	81	N	50	ex	NSCLC	large cell
S51	f	78	N	30	ex	NSCLC	unknown

Table 9.3: Summary of the VOC profile cancer cohort

* Only co-existent cancers at recruitment were patients with leukaemia

Study No.	Sex	Age	Smoking status
C01	f	52	current
C02	m	37	never
C03	f	74	never
C04	m	32	never
C05	f	56	never
C06	m	38	never
C07	m	38	never
C08	f	61	never
C09	f	58	ex
C10	f	33	ex
C11	f	76	ex
C12	m	77	ex
C14	m	62	ex
C15	f	37	never
C16	f	61	ex
C17	m	68	ex
C18	f	62	never
C19	m	76	never
C20	f	61	ex
C21	f	53	ex
C22	f	50	never
C23	f	48	current
C24	f	55	ex
C25	m	56	never
C26	f	44	current
C27	m	36	current
C28	m	44	current
C29	f	51	never
C30	f	56	never
C31	f	41	never
C32	f	43	ex
C33	f	47	current
C34	m	59	ex
C35	f	51	current
C36	m	45	never
C37	m	38	never
C38	f	44	current
C39	f	51	never
C40	f	45	never
C41	m	57	never
C42	f	53	never
C43	f	47	ex
C44	f	52	never
C45	f	53	never
C46	f	55	ex
C47	m	50	never
C50	f	65	current

Table 9.4: Summary of healthy control group

Scatterplot of VOCs and polynomial equations

A scatterplot of area under the curve (AUC) on the y-axis against retention time (RT) on the x-axis was created. (Figure 9.1) Two distinct polynomial equations were then generated representing different overall patterns between lung cancer cases and controls.

$$\text{Lung cancer: } y = -14.259x^5 + 1217.3x^4 - 37821x^3 + 517134x^2 - 3E+06x + 6E+06$$
$$R^2 = 0.0512$$

$$\text{Controls: } y = -8.5934x^5 + 731.48x^4 - 22939x^3 + 324106x^2 - 2E+06x + 4E+06$$
$$R^2 = 0.1064$$

The R squared (R^2) value for the lung cancer polynomial was close to 0 ($R^2=0.051$), as was the R^2 value for the control polynomial ($R^2=0.1$), thus not statistically significant, which suggests that the scatter around the lines for each cohort may have occurred by chance. There is no strong correlation presented here and the line would not be a good predictor of cancer patients or controls. The lack of statistical significance is not surprising as the variance for the majority of compounds is so high. However, the lines were useful in identifying particular zones for discriminating cancer patients.

Polynomial equations and curves for lung cancer cases and controls identified a period of 8-17 seconds retention time in the GC-MS, representing certain VOCs, which corresponded to the greatest deviation between the two curves for lung cancer and control data sets.

P1 and P2 represent two zones that are exclusive to VOCs from lung cancer patients, accounting for approximately 27% of the lung cancer cohort and 0% of controls. This data is represented in Figure 9.1.

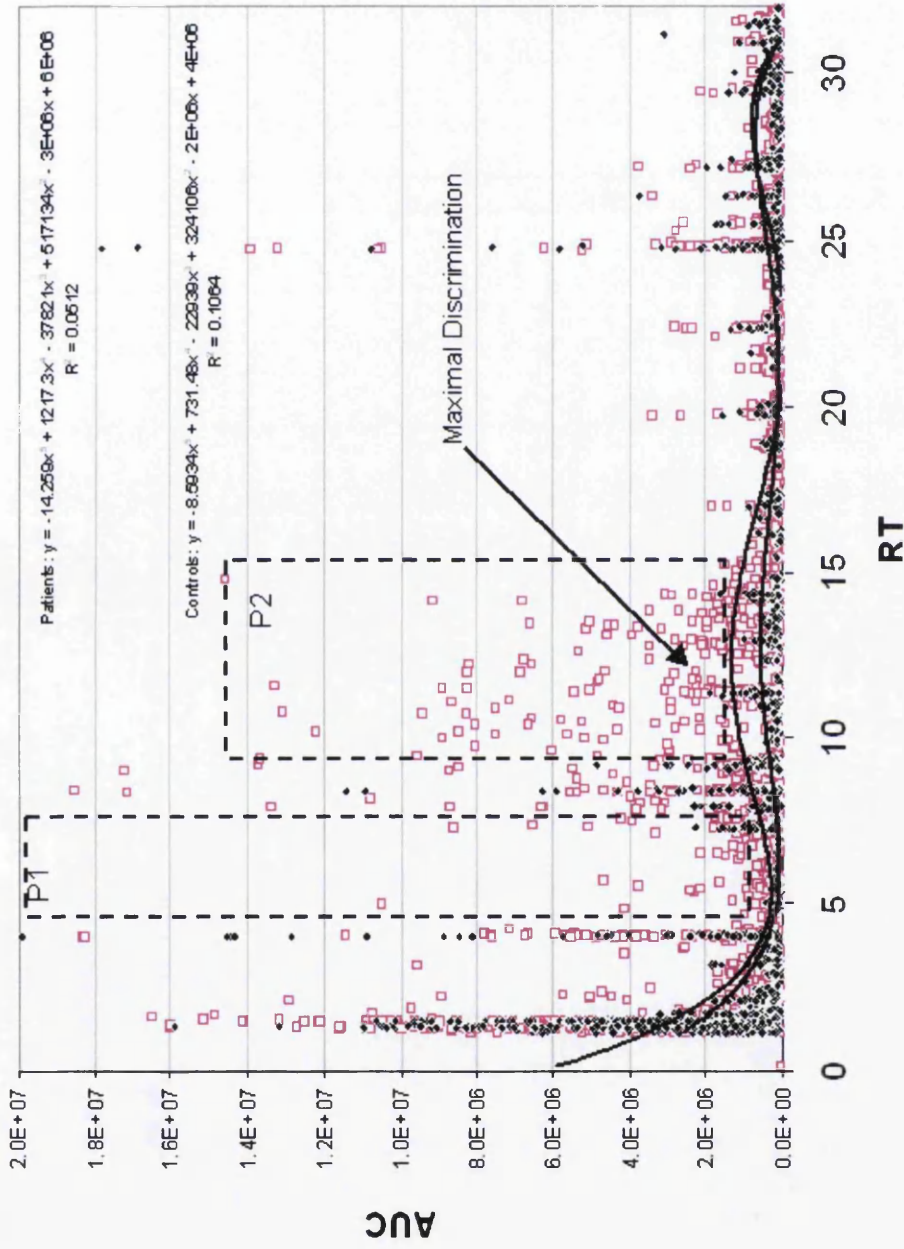


Figure 9.1: Polynomial equations applied to derive two curves, which best discriminates lung cancer and control cases. P1 and P2 represent zones where VOCs are exclusive to cancer patients. Pink = lung cancer VOCs; Black = healthy control VOCs. RT (retention time); AUC (area under curve).

Compound-by-compound analysis for identification of specific discriminating VOCs

246 VOCs were identified in total from all samples (all subjects).

29 compounds were found to be exclusively in lung cancer patients (see Figure 9.2). Using these 29 VOCs, 70% (31 of 44) of lung cancer patients could be correctly classified. Thus using these 29 cancer-exclusive VOCs on the study group gives a sensitivity of 70% and specificity of 100%. Interestingly, the points in zones P1 and P2 correspond to a proportion of the cancer-exclusive compounds.

However, no cancer subject had more than 5 of the 29 exclusive VOCs in their breath suggesting the heterogeneous nature of lung cancer pathways.

A further 25 VOCs were universally higher in lung cancer patients and although present in 23 of the 47 controls the levels of VOCs were low. Thus using the 25 VOCs we were able to correctly identify a further 18% (8 of 44) of lung cancer patients. These compounds are represented in Figure 9.3.

Thus using these 54 VOCs allowed the correct classification of 39 out of 44 (88.6%) of lung cancer patients. There were 5 cancer patients who were not identified with the panel, i.e. false negatives S06, S19, S25, S34, S42. Looking at the individual cases that were false negatives, 3 were SCC and 1 large cell with the other a clinical diagnosis. There were no other similarities identified.

The overall sensitivity and specificity using the 54 VOCs was 88.6% and 100% respectively.

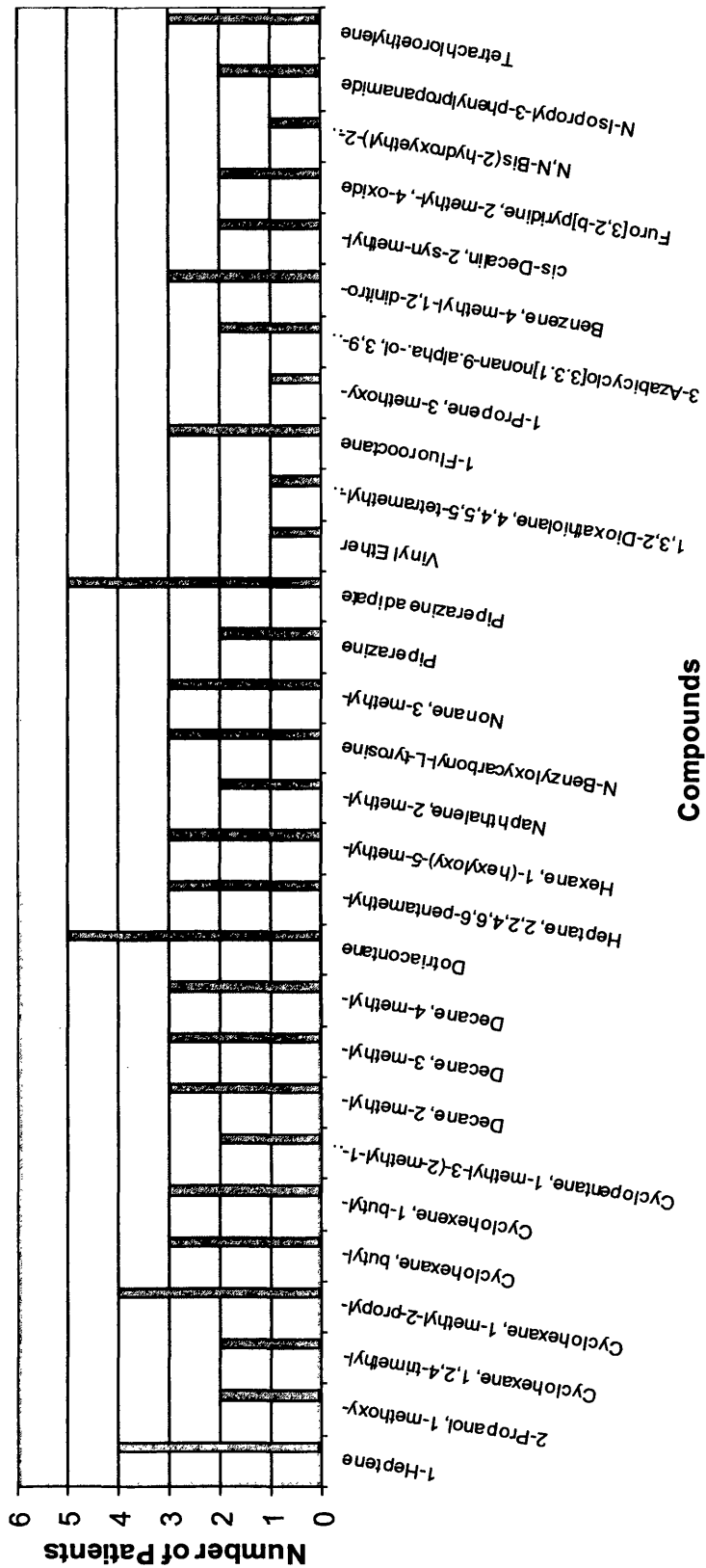


Figure 9.2: 29 VOCs which are exclusive to the lung cancer cohort and were able to correctly classify 31 (70%) of lung cancer patients

Compounds	Cases with exclusive VOCs				
1-Heptene	S16	S20	S24	S45	
2-Propanol, 1-methoxy-	S49	S36			
Cyclohexane, 1,2,4-trimethyl-	S08	S15			
Cyclohexane, 1-methyl-2-propyl-	S07	S08	S10	S15	
Cyclohexane, butyl-	S08	S10	S15		
Cyclohexene, 1-butyl-	S09	S18	S26		
Cyclopentane, 1-methyl-3-(2-methyl-1-propenyl)-	S08	S43			
Decane, 2-methyl-	S08	S10	S15		
Decane, 3-methyl-	S08	S10	S15		
Decane, 4-methyl-	S08	S10	S15		
Dotriacontane	S02	S12	S26	S32	S51
Heptane, 2,2,4,6,6-pentamethyl-	S08	S10	S33		
Hexane, 1-(hexyloxy)-5-methyl-	S10	S27	S38		
Naphthalene, 2-methyl-	S38	S43			
N-Benzoyloxycarbonyl-L-tyrosine	S18	S23	S33		
Nonane, 3-methyl-	S08	S10	S15		
Piperazine	S39	S40			
Piperazine adipate	S09	S23	S28	S39	S40
Vinyl Ether	S46				

Table 9.5: List of 29 VOCs exclusive to lung cancer cohort and the patients in which they occurred

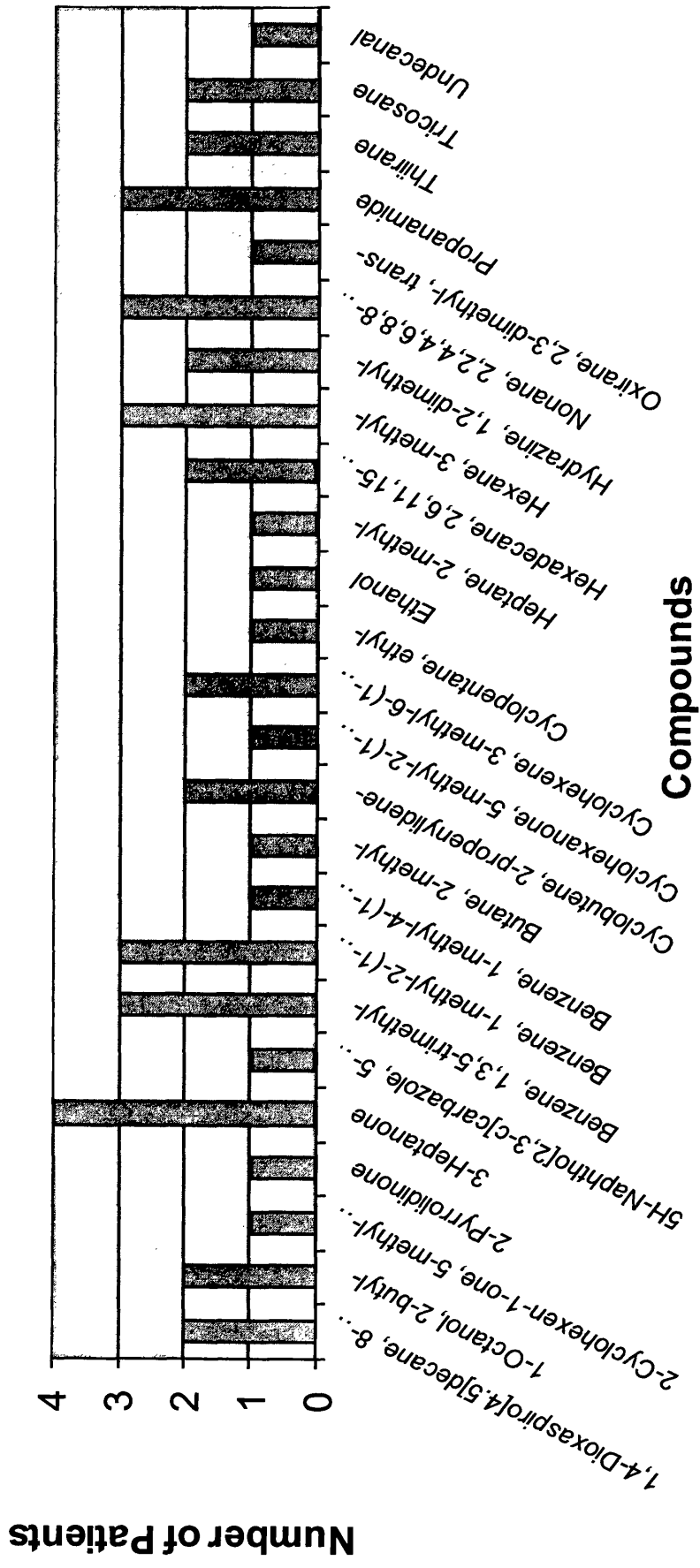


Figure 9.3: 25 VOCs which were universally higher in lung cancer patients and allowed 8 (18%) further subjects to be correctly identified

Compounds	Cases with higher levels of specific VOCS			
1,4-Dioxaspiro[4.5]decane, 8-				
(methylthio)-	S03	S05		
1-Octanol, 2-butyl-	S15	S44		
2-Cyclohexen-1-one, 5-methyl-2-(1-methylethyl)-	S43			
2-Pyrrolidinone	S24			
3-Heptanone	S02	S08	S09	S14
5H-Naphtho[2,3-c]carbazole, 5-methyl-	S04			
Benzene, 1,3,5-trimethyl-	S15	S26	S35	
Benzene, 1-methyl-2-(1-methylethyl)-	S08	S23	S35	
Benzene, 1-methyl-4-(1-methylethenyl)-	S43			
Butane, 2-methyl-	S30			
Cyclobutene, 2-propenylidene-	S45	S50		
Cyclohexanone, 5-methyl-2-(1-methylethyl)-, trans-	S43			
Cyclohexene, 3-methyl-6-(1-methylethyl)-	S08	S43		
Cyclopentane, ethyl-	S29			
Ethanol	S16			
Heptane, 2-methyl-	S29			
Hexadecane, 2,6,11,15-tetramethyl-	S26	S35		
Hexane, 3-methyl-	S08	S10	S29	
Hydrazine, 1,2-dimethyl-	S09	S39		
Nonane, 2,2,4,4,6,8,8-heptamethyl-	S04	S08	S37	
Oxirane, 2,3-dimethyl-, trans-	S37			
Propanamide	S01	S35	S39	
Thiirane	S28	S45		
Tricosane	S02	S41		
Undecanal	S04			

Table 9.6: 25 VOCs which were universally higher in the lung cancer cohort compared to healthy controls

12 VOCs were found to be exclusive to control subjects and were able to correctly classify 21 of 47 (45%) as controls (Figure 9.4).

A further 15 VOCs were universally higher in the control group compared to the cancer cohort and when applied were able to correctly classify 7 of 47 (15%) as healthy controls (Figure 9.5).

Thus 27 VOCs allowed 60% of subjects to be correctly classified into the healthy control group.

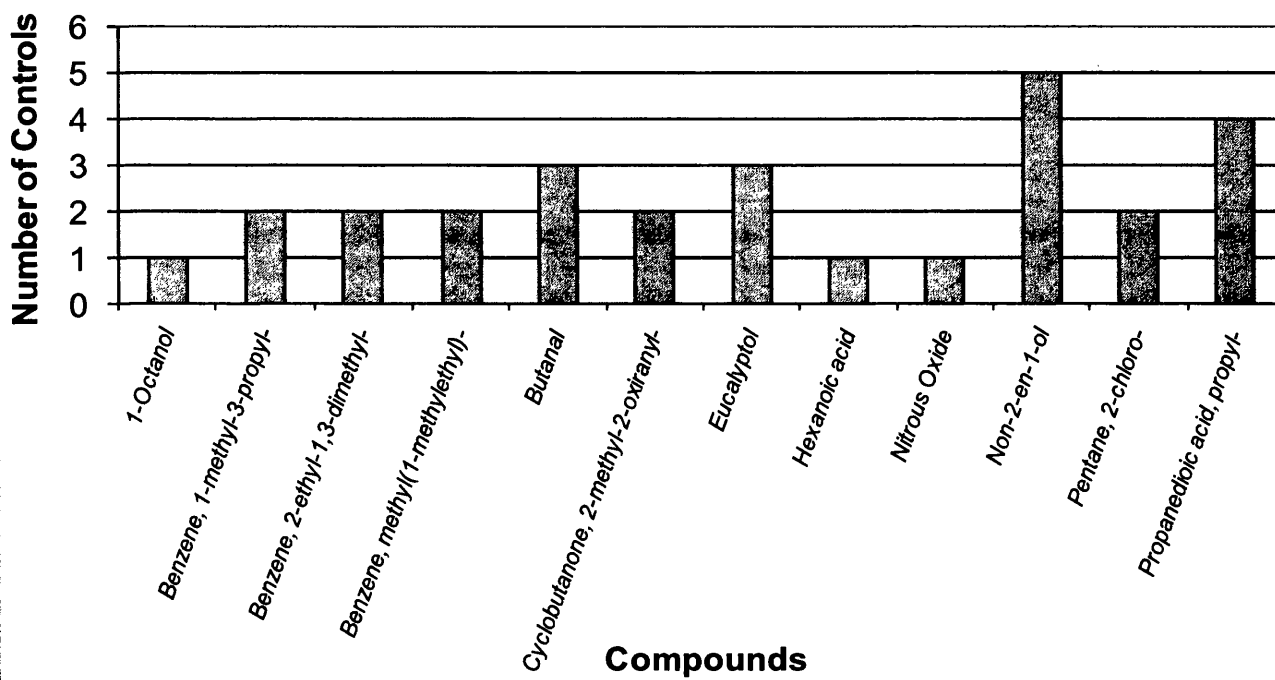


Figure 9.4: 12 VOCs which are exclusive to healthy control subjects and allow correct identification of 21 (45%) as controls.

Compounds	Cases with exclusive VOCs				
1-Octanol	C33				
Benzene, 1-methyl-3-propyl-	C03	C50			
Benzene, 2-ethyl-1,3-dimethyl-	C30	C50			
Benzene, methyl(1-methylethyl)-	C35	C37			
Butanal	C05	C11	C17		
Cyclobutanone, 2-methyl-2-oxiranyl-	C05	C50			
Eucalyptol	C24	C28	C35		
Hexanoic acid	C10				
Nitrous Oxide	C44				
Non-2-en-1-ol	C10	C19	C23	C40	C42
Pentane, 2-chloro-	C04	C25			
Propanedioic acid, propyl-	C10	C12	C14	C23	

Table 9.7: 12 VOCs that were exclusive to the healthy control cohort and the controls in which they were detected.

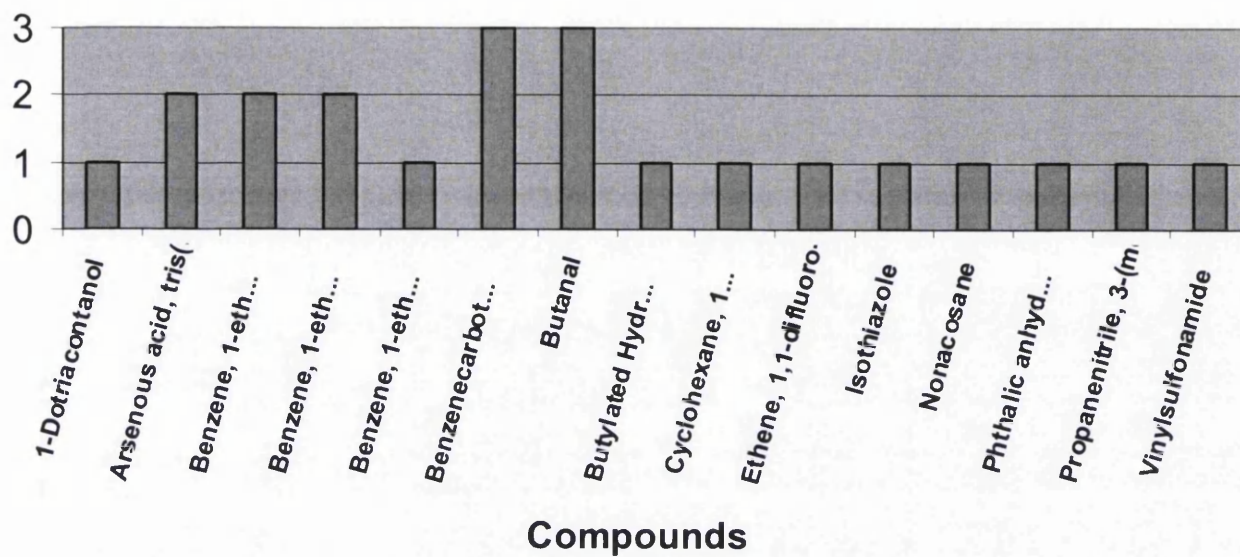


Figure 9.5: 15 VOCs which were universally higher in the healthy controls compare to the lung cancer subjects and allowed 7 (15%) subjects to be correctly classified as controls.

Compounds			
1-Dotriacontanol	C11		
Arsenous acid, tris(trimethylsilyl) ester	C18	C37	
Benzene, 1-ethyl-2,3-dimethyl-	C03	C12	
Benzene, 1-ethyl-2,4-dimethyl-	C03	C04	
Benzene, 1-ethyl-4-methyl-	C15		
Benzenecarbothioic acid	C33	C38	C43
Butanal	C05	C11	C17
Butylated Hydroxytoluene	C16		
Cyclohexane, 1,4-dimethyl-	C17		
Ethene, 1,1-difluoro-	C12		
Isothiazole	C16		
Nonacosane	C08		
Phthalic anhydride	C23		
Propanenitrile, 3-(methylamino)-	C06		
Vinylsulfonamide	C40		

Table 9.8: 15 VOCs that were universally higher in healthy controls compared to lung cancer subjects

PART 4: DISCUSSION

Chapter 10 FTIR and Immunohistochemistry

Assessment of sputum – Discussion

Summary of key findings

The main objective of the study on sputum was to assess the role of FTIR in distinguishing between cancer patients and those who were cancer-free.

Sputum can be obtained in a non-invasive manner and FTIR provides a cost-effective high-throughput method of analysis, thus addressing two criteria when considering screening a population or part of it. We generated FTIR spectra on all sputum samples from three cohorts, cancer, those deemed high-risk enough of having lung cancer to warrant a diagnostic bronchoscopy and healthy controls, with very little smoking history and no respiratory or constitutional symptoms. The spectra were obtained within the mid-infrared range between wavenumbers at 600 cm^{-1} and 4000 cm^{-1} . We focused on this range as it represents the wavenumbers attributed to most biological differences observed in cancerous human tissues. It was evident that cancer spectra yielded higher absorbances than healthy control spectra at different regions of wavenumbers. The areas of clear differences were in the regions of wavenumbers 1400 to 1060 cm^{-1} ; 1060 to 1025 cm^{-1} and 1400 to 1670 cm^{-1} . These wavelengths reflect chemical signature bonds representing protein areas (especially amides I and II), glycogen-rich and RNA / DNA regions. The fact that most of the sputum samples gave good quality FTIR spectra strongly suggests that this biofluid is a useful non-invasive tissue source for FTIR analysis in the detection of lung cancer. We encountered and were able to identify the technical flaws that led to the grossly abnormal spectra and have since modified our laboratory training. This compares with similar difficulties encountered with the immunohistochemistry staining.

Multivariate analysis (PCA) was able to demonstrate the reproducibility of FTIR on sputum samples at different times as well as consistency in processing replicates of the same sample. We did not observe groupings of cancer cases

according to sub-types such as squamous, adenocarcinoma or small cell carcinomas but are looking at sub-group patterns in more samples.

A preliminary analysis of entire FTIR spectra for cancer and healthy control cases provided initial insight into whether FTIR could discriminate these two groups. By retaining the first three principal components we observed, using scatterplots, a pattern of cancer and healthy control spectra 'branches' separating in different directions. This approach was useful as it allowed us to quantitatively assess the variation in absorbance values across spectra across cancer and control cases. The first three principal components alone explained greater than 90% of the overall variation across the spectra meaning that the separating pattern we were observing really was attributable to metabolic differences between cancer and healthy control cases.

The next step involved identification of a small number of infrared wavenumbers that could optimally discriminate between cancer and healthy control cases. The number of wavenumbers needed to evaluate in combination was unknown at the time and the analysis process started with just two wavenumbers. We envisaged that if two wavenumbers in combination could not discriminate between cancer and healthy control cases with a high degree of accuracy then we would incrementally increase the wavenumbers accordingly. Using the R statistical software, we were able to arrive at a combination of just two wavenumbers, 1031.7 cm^{-1} and 1409.7 cm^{-1} , which discriminated cancer and healthy controls with a high degree of sensitivity and specificity. The discriminatory model developed using wavenumbers 1031.7 cm^{-1} and 1409.7 cm^{-1} had a sensitivity of 92.6%, a specificity of 90.7%. This compares very well with the immunohistochemistry analysis of sputum which had a sensitivity ranging from 10 to 42% and specificity of 72 to 96%.

When the patients who underwent bronchoscopy and lung cancer was written as potential differential diagnosis - but ultimately turned out not to have lung cancer after 1 year clinical / bacteriological and radiological follow up (high-risk cohort) were added to the discriminatory model, 71% of these were predicted to have cancer. Although this figure appears high (i.e. a lot of false positives), this

group were already deemed of high enough risk to warrant an invasive investigation such as bronchoscopy. All had radiographic changes or symptoms compatible with lung cancer together with risk factors such as a significant smoking history or a family history of cancer. Although at 1 year follow up there was no evidence of lung cancer, they may still have early biological changes of cancer that are not detectable by current means. We continue to monitor the clinical progress of this very important sub-group.

How do the results compare with previous studies of FTIR in cancer?

The ability of FTIR to detect cancer has been tested in a small number of studies. In our study, the 1409.7 cm^{-1} wavenumber signified a difference in concentration of amino acids and consequently protein(s). 1409.7 cm^{-1} is associated with carbon (C) oxygen (O), COO- stretching and C-(hydrogen)H bending. Our other important number, the 1031.7 cm^{-1} wavenumber falls within a region assigned to C-O stretching coupled with bending of C-OH groups of glycogen. [206]

This result is quite intriguing as glycogen has previously been investigated in lung cancer tissue using FTIR by Yano *et al.* [192, 193] Using a sample size of 26 patients, they found that absorbances at 1045 cm^{-1} wavenumber, representing glycogen, were significantly higher in lung cancer biopsy specimens relative to non-cancer control tissue from the same patients. They also found that absorbances for this wavenumber were higher for adenocarcinoma than for SCC although we did not observe this for wavenumber 1031.7 cm^{-1} . Wang *et al* reported that there were large differences in peak absorbance at wavenumber 1030 cm^{-1} (which represents glycogen rich region) between lung tumour and normal lung cells in pleural fluid of the same patients although this was assessed on only 8 subjects. [194] They also identified differences at 1080 cm^{-1} which represents the phosphodiesteres. Light and electron microscopic studies have shown that glycogen is mostly free within cells in different types of malignant lung tumours but in benign tumours it is mostly membrane-bound. [286, 287] The results of this study as well as the previous studies suggest that glycogen levels are raised in lung tumours

relative to normal lung tissue and that glycogen is not confined to membranes within the cell. Interestingly, the higher absorbance in phosphodiester (1080 cm^{-1}) is also a notable difference when comparing oesophageal cancerous cells with its normal counterpart. [203]

Greater levels of glycogen detected by FTIR (wavenumber 1030 cm^{-1}) were also identified when comparing nine gastric cancerous tissues with normal gastric tissue from the same patients with samples obtained immediately post gastrectomy. Phosphodiester (1081 cm^{-1}) were again greater in the malignant samples. [204]

Cervical cancer appears to have a reduced absorption within the wavenumbers representing glycogen. A study in Australia assessed FTIR on exfoliated cervical cells of 227 patients. Six spectra were recorded for each patient and the spectra were described as either type 1 or type 2. Spectra designated type 1 exhibited a profile characteristic of normal epithelial cells, with intense glycogen bands at 1022 cm^{-1} and 1150 cm^{-1} , and a pronounced symmetric phosphate stretch at 1078 cm^{-1} . Spectra designated type 2 exhibited features suggestive of dysplastic or malignant transformation, with pronounced symmetric and asymmetric phosphate modes and a reduction in glycogen-band intensity. Of the 272 patients, 68.6% of samples exhibited only type 1 profiles for all six recorded spectra, 29.4% of samples yielded had at least one type 2 spectrum and 2% of samples were inconclusive. Of the 68.6% with normal type 1 spectra, 86% were diagnosed normal by Papanicolaou smear with no follow up biopsy ordered, 7% were diagnosed abnormal by biopsy, 5% normal by biopsy and 2% were still inconclusive. For the remaining 29.4% of abnormal type 2 spectra on FTIR testing, most (71%) had shown an abnormal smear result. These 71% were subsequently biopsied, and 87% were confirmed abnormal or cancer by traditional histopathological methods. [195]

The lowering of cellular glycogen content during carcinogenesis of the cervix has been well documented. [288, 289] In 1991, a study looked at exfoliated cervical cells collected from 156 women by brushing. The cell pellets were split so that one could undergo FTIR, the other microscopic examination by two

pathologists. 136 of the samples was normal, 12 were cancerous and 8 dysplastic. The normal samples had FTIR spectral peaks particularly at 1025 cm^{-1} and 1047 cm^{-1} , thus corresponding to the glycogen rich area.

The infrared spectrum of glycogen from mammalian liver lies within the frequency region $975 - 1060\text{ cm}^{-1}$. Glycogen, therefore, made a large contribution to the intensity of these two IR bands of the normal cervical cells.

When the 12 cancer samples underwent FTIR, there were significantly reduced absorptions at the 1025 cm^{-1} and 1047 cm^{-1} wavenumbers, represented in Figure 10.1. [289]

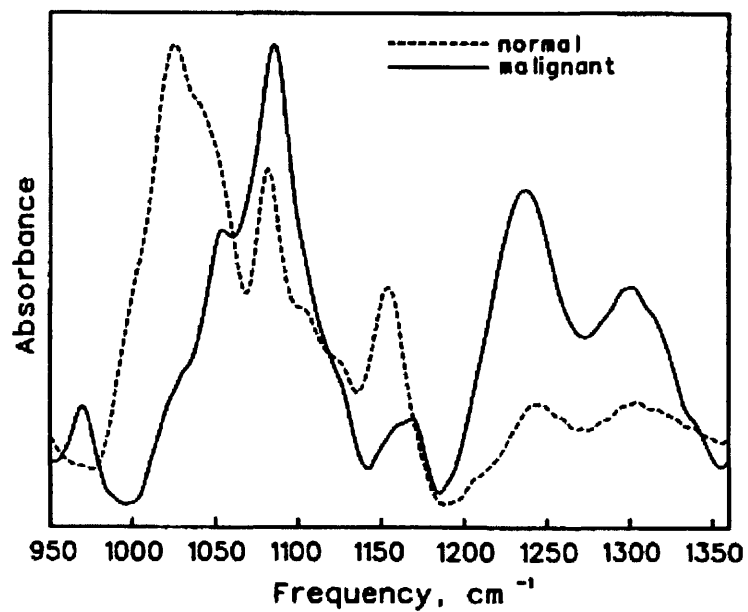


Figure 10.1: Comparison of FTIR of normal and malignant exfoliated cervical cells. Malignant cells have much less absorption in the glycogen-rich area, 1025 cm⁻¹ and 1047 cm⁻¹, compared to normal cervical cells. [289]

All the studies of FTIR in cancer described above follow similar techniques with regard to FTIR analysis. The difference between this study and the others is that they all involve analysis on biopsy tissue that offers no comfort advantage to the patient undergoing a surgical biopsy. In such histological specimens, FTIR could still look for signs of early dysplasia, or as an adjunct to improve receiver operator curves for histological staining alone.

The main advantage of our study is that the sputum is obtained in a non-invasive manner. Moreover, most people at risk of lung cancer should be able to produce sputum – even if they are too frail to undergo endobronchial / CT-guided or even open-lung biopsies.

FTIR applied to sputum in other diseases

A recent paper looked at the role of FTIR in sputum of COPD patients to help monitor the disease status. 15 consecutive patients with moderate to severe COPD were enrolled along with 15 healthy controls who had never smoked. Sputum was collected by induction with nebulised hypertonic (3%) saline. Similar to our data, as their data was not normally distributed, they performed a Mann-Whitney U Test to detect any statistical differences between the COPD and control sputum. Within the COPD group, no statistical difference was identified between current or former smokers. Slightly different to the other FTIR studies, this group looked at the shift in the peak of absorbance as well as the intensity of absorbance. They noted, in the glycogen-rich area, a shift in peak from normal of around 1077 cm^{-1} towards around 1073 cm^{-1} in 66% of patients with COPD (10/15). They also found a three-fold increase in absorbance of the same 10 COPD patients at wavenumber 1073 cm^{-1} and a shift in peak from 1559 cm^{-1} to 1561 cm^{-1} , compared to controls - which is in the amide II region. [207] The authors also concluded that sputum is a suitable biofluid for FTIR.

This study as well as our own shows an increase absorbance in the glycogen-rich region as previously noted in the lung cancer studies, although the peak absorbances were in a different position. In our study the key wavenumber from

the glycogen-rich region is 1031.7 cm^{-1} with the key wavenumbers from the other studies being close to but not identical to this. Although the FTIR technique is similar, our sputum collection and processing was different. We relied on self-expectorated sputum and for study purposes suctioned sputum from bronchoscopy. For large screening studies in the community, the ideal would be self-expectorated sputum only. Inducing sputum with nebulised hypertonic saline adds to cost, and time (14 minutes based on this particular study protocol). The study numbers are small and they haven't taken into account the subjects that smoke but do not have COPD. Although there was no statistical difference between the spectra of former and current smokers within the COPD cohort, smoking status was not validated, and simply taken on asking the individuals.

Strengths and weaknesses of FTIR and sputum study

We tried a novel approach. We could find no published data on the use of FTIR in the sputum of lung cancer patients. Sputum is a good biofluid to study as it can be obtained in a non-invasive manner for the subject and poses less infectious risk to researchers than blood. When one considers a screening modality these are important issues to consider.

Strengths of the study:

Participant selection:

The patients were recruited in a prospective manner from respiratory clinics in district general and teaching hospitals and are typical of those undergoing bronchoscopy in an everyday clinical setting. Patients were recruited prior to definitive diagnosis. In targeting this group of patients we have assessed the molecular changes at the earliest point possible using untested technology. All the patients in the cancer and high-risk group underwent bronchoscopy, thus the Respiratory Physician assessing the patient must have felt there was a potential for lung cancer to be present for the patient to undergo an invasive procedure. Usually this group of patients will have smoked for many years or

have a family history of cancers. It is likely that should any screening modality be developed for lung cancer then this population of patients could be targeted.

The healthy controls consisted of those with no respiratory disease, cancer or worrying symptoms that may indicate underlying cancer. The cohort consisted of a mixture of current, former and never smokers.

Sample size:

As the study was experimental it was difficult to power the study initially. However by having 169 patients in the study (132 of the 169 spectra were used in analysis), it provided a large enough sample to see if the technology worked in principle. The model was developed after analysis of 108 spectra (54 cancer and 54 healthy control). This compared favourably with Yano (n=26), Wang (n=8), Wang (n=27), Wong (n=12) and Whiteman (n=30) all of whom had smaller sample sizes. [192, 194, 203, 207, 289]

Demographics:

There were no significant differences between the 3 cohorts with regards gender. Although the high-risk group appeared to have a much larger proportion of males compared to the other 2 groups, this was not significant. The high-risk and cancer groups were well matched for number of smoking pack years and FEV₁.

Sputum:

Sputum is an ideal specimen to assess in this group of patients. Often with many years of smoking, sputum expectoration is increased (chronic bronchitis) thus making it a potentially accessible biofluid to test. The samples can be stored in normal multi-specimen containers which are readily available and cheap so in practice can be collected by the bedside or in the GP surgery.

FTIR process:

The advantage of FTIR is that it is cost-effective and high-throughput. Although the FTIR equipment is expensive, the processing only costs 5-10 pence per sample each. Moreover the process is automated and each specimen tray has

96 'wells' that can be used for up to 96 individual samples (or still 32 patients, if taken in triplicate). The process itself takes approximately 2 hours for all 96 samples compared with 5-10 minutes per sample for cytopathology review.

Statistics:

We applied standard statistics and tested for normality throughout. PCA has been used previously in the analysis of FTIR data. PCA allowed us to get an idea about the pattern of data, particularly how compounds of similar molecular structures group together. Logistical regression allows the formation of a linear model based on the differences between two objects (in this case cancer and healthy control cohorts). In multivariate statistical analysis, these methods are well used and recognised. We specifically used the 'leave-one-out cross validation' method to test how robust our model using the optimal discriminating pair of wavenumbers was. By taking cases out in sequence and still confirming that the sensitivity maintained a high level of consistency, we were satisfied that the two chosen wavenumbers were a good model.

Weaknesses of the study

Participants:

The cancer and high risk cohorts were recruited from secondary care and all had symptoms and/or abnormal radiology that was recognised both by them and their General Practitioners / hospital specialists. They do not represent the majority of smokers e.g. over 50 years in the community, where a large screening trial would be based. Our healthy control age group were significantly younger than both the cancer and the high-risk group and this represents the Swansea University students and bias to a working-age population of hospital staff. We know of no published work that sputum or any FTIR spectra are influenced by age per se but ideally the groups should all have been matched for age.

Advanced stage of disease within cancer group:

As one would expect, most of the patients presented when they had already developed symptoms and in 41 (76%) the cancer was in an advanced stage of

the disease (Radiological TNM stage > III). Although the aim of the study was to see how FTIR could differentiate cancer from non-cancer cases, the ideal in the future is to produce a screening test that is applicable to all stages of lung cancer, particularly the early forms amenable to curative treatment. The changes in wavenumbers maybe more pronounced in a more advanced disease state as greater tumour load should result in more metabolic changes. It is hoped that these changes will still be present to a high enough degree in the very early, curative, stages of the disease, as only then will this have an impact on reducing mortality rates. In further studies, or as the number of recruited patients increase, analysis of the early disease states can be assessed. This may well continue to prove difficult for the reasons explained throughout the thesis with regards to these patients presenting late. We do not have large enough numbers yet to make meaningful comparisons according to tumour histology or TNM staging.

Smoking status:

There were statistically significant differences in the smoking status between the 3 groups. Although the cancer and healthy controls had a similar number of current smokers (19 vs. 24), there were very few former smokers in the control group (n=1) compared to the cancer group (n=28). Although exhaled CO levels in the high-risk and cancer cohort validated smoking status, it was not performed in the healthy control group. Future studies should always biologically validate smoking status, perhaps even long-term abstinence with (much more expensive) markers such as salivary or urinary cotinine, as eCO only reflects smoking within 24 hours.

Size of high-risk cohort:

Although there were similar number of patients in the cancer and healthy control cohorts (n=54), the high-risk group had fewer than half the number of patients (n=24). We could not predict at the time of recruitment and bronchoscopy which of the 99 patients would have lung cancer. Our eventual cancer rate of 2/3 from all those undergoing bronchoscopy reflects national practice. [290]

Location of tumour

The patients were selected as those who were fit enough and had appropriate radiology to undergo bronchoscopy. Thus they were selected as tending to be fitter and to have more central lesions. We initially wanted to recruit these patients as

- a) We thought they would have the highest histopathological diagnostic rate
- b) We had ethics to obtain biopsy material (a future comparison is biopsy FTIR spectra versus sputa FTIR spectra)
- c) It was a preset time and focal point for me to attend and recruit from (given my other clinical commitments, I could not always attend CT sessions, clinics etc).

This meant that patients with more peripheral lesions, e.g. who underwent CT-guided biopsy or those refusing or often deemed unfit for bronchoscopy, were not approached and this should be acknowledged. The sputa from more central lesions are more likely to contain dysplastic / malignant cells, and thus potentially show greater biological changes in their sputum. Even after selection, there were some patients who had a diagnosis of lung cancer but had no visible endobronchial lesion. These were too few in number to allow meaningful analysis at this point. It can be difficult to diagnose early lung cancer in subjects who have a small peripheral nodule in the middle of their lung which is too peripheral to see endobronchially but often too small to safely perform a CT-guided biopsy / fine needle aspiration. These patients are now being included in a modified study to see if they also have changes in their sputum FTIR spectra as those with more central lesions, particularly as many of these will have early stage disease.

Sputum in those unable to self-expectorate:

When collecting the sputum, not all the samples were self-expectorated. Some of the samples had to be collected at the time of bronchoscopy via suction and this may have meant that some of the samples contained normal saline or e.g. cleaning chemicals from the bronchoscope collection channel. Importantly, these samples did not show a statistically significant difference in the number of

bronchial cells present when a random selection of sputa were assessed. Ideally, all the sputum samples should be examined for the presence of adequate numbers of bronchial cells, however, this would have taken several days to be performed and would still require an experienced pathologist to do it optimally. This was not feasible within the constraints of this study. There were some delays in freezing some of the sputum samples which may have led to deterioration of the samples. In an ideal situation, all samples should be immediately frozen.

Histological diagnosis missing:

Excluding those patients who were too unwell to undergo invasive investigations and were therefore given a clinico-radiological diagnosis of lung cancer, there remained 9 patients out of 54 who had no definite histological subtype identified. The lung cancer MDT, diagnosed these as having NSCLC, according to typical clinical findings and radiological appearances. We wanted to maintain numbers and use a final (clinical) diagnosis of lung cancer reflecting a clinical service.

Sputum cell preparation for FTIR:

Preparation of the samples for FTIR took between one and two hours (for 20 samples) requiring manual laboratory trained technicians. This again may change as technology advances and more automated facilities are developed but limits its general applicability.

Spectral anomalies

A small number of spectra were fundamentally different from most samples and also from the recognised patterns from medical literature. By cross checking records and sample coordinates, we realised that these spectra were analysed in one particular batch and all the abnormal spectra came from samples that were grouped in a corner of a single tray upon which there was a bio film (probably contamination bacteria) present which altered the absorptions thus resulting in very different spectra. These spectra were not included in the final analysis, as they were deemed uninterpretable.

Transforming the raw spectra:

The raw spectra underwent various transformations to create the 2nd derivative spectra. This included a baseline correction, which altered some of the sloping and negative values derived. Although recognised techniques were used in the process, whilst transforming data there is always the potential for 'losing' data. Although this is a potential weakness, the overall process is standard technique and important in order to standardise the data and thus allow sensible interpretation. Although we validated our model internally, a better test of our model would be to apply it to a new set of prospectively gathered data (preferably where we are blinded to the clinical diagnosis).

The way forward: FTIR and sputum

We have shown that there are significant spectral differences between cancer, high-risk and healthy controls when applying FTIR to sputum. Our most significant wavenumbers are similar to those reported in other studies looking at cancers of other organs and using a variety of tissue samples. We have generated a predictive model with good sensitivity and specificity and the future plans are to further test the accuracy of the model of all stages of lung cancer and more smokers over the age of 50 years that are more likely to represent the screened 'at risk' population.

Recruitment:

We are continuing to recruit with two English NHS Trusts and a further Welsh Health Board gaining site-specific approval. By studying a larger number of patients within each cohort (especially those at high risk), we will be able to see how successful the model is at predicting cancer. It is plausible that the wavenumbers used in the model may change slightly.

Methodology:

Rather than recruit solely the patients who are undergoing bronchoscopy, we have amended our ethics application to just collect sputum. We are now collecting samples from people with peripheral lesions and those not deemed fit for bronchoscopy. The location of the lesion on CT and findings (proximal lesion, distal lesion, no lesion) at any bronchoscopy are being noted.

Internal quality control checks and accurate sample labelling detected some technical errors, encountered in the sputum processing. We hope to learn from these errors.

Further assessment of spectral breakdown according to cancer subtype and staging should provide additional information. Staging is particularly important to identify if the predictive model maintains the sensitivity and specificity in early, potentially curative, stages of disease. Identification of these early cases

will be the most likely way to impact the mortality rates from lung cancer as there are few new treatment regimes being developed.

Further analysis of the actual compounds in the sputum by mass spectrometry and DNA analysis is being undertaken in Swansea University. We hope to more accurately identify the actual important molecular differences between the three clinical cohorts and confirm if the FTIR signals representing chemical bonds do identify specific compounds that relate to carcinogenesis. This could lead to a better understanding of lung cancer molecular pathways and even provide targets for future therapies.

These improvements in study design are essential to further validate our early results. It is important that this work evolves, as a cost-effective, high-throughput screening technique for lung cancer, using a non-invasive biofluid, remains elusive.

IMMUNOHISTOCHEMISTRY AND SPUTUM

Summary of key findings

An additional arm of the main study was assessing established immunohistochemistry techniques but in two original ways. We tested a panel of antibodies, which have not been evaluated in combination, and we looked at sputum cells. We assessed antibodies to 5 established lung cancer markers; p16, p53, p63, cyclin D1 and EGFR. In comparing the cancer and high-risk cohorts (we did not collect sputum for this test of healthy controls), there were no significant differences in the staining for the individual antibodies. Although the specificity was high, they had very low sensitivity. Thus, although very few *without* cancer would be wrongly diagnosed as having lung cancer there were an unacceptably high number who actually did have lung cancer but had negative IHC (low sensitivity). This continued to be the case even when all the samples that were deemed uninterpretable were removed from the dataset. This was further proven when hierarchical cluster analysis continued to show heterogeneity of groups. As a statistical modality it can be used as a visual inspection of how similar results group together. No clear-cut group of all lung cancer cases were seen to branch together, confirming that this particular group of antibodies do not differentiate well between cancer and non-cancer cases based on sputum.

How do the results compare with other studies?

p16

P16 is thought to be a tumour suppressor gene. Its alteration or inactivation has been implicated in various malignant tumours, including the lung. [291, 292] p16 expression in 135 lung cancer specimens was studied to evaluate the presence of genetic alterations. P16 alteration was found to be a frequent event and independent of histological subtype of lung cancer. The reduction or loss of p16 expression was also associated with a worse prognosis in lung cancer. [293] A study performed on 171 patients with NSCLC who underwent surgery, had p16 immunoanalysis of their tumour specimens. 62 (36.3%) were classed

as p16-negative. There were significantly more p16-negative squamous cell carcinoma (SCC) compared to adenocarcinomas ($p=0.039$) and these p16-negative SCC had a significantly lower survival compared to those with p16-positive SCC ($p=0.001$). [294] One particular way of inactivating the p16 gene is by promoter methylation. A study in Chile found 59 out of 74 (79.7%) patients with NSCLC had their p16 inactivated by promoter methylation and this was significantly higher in those with SCC (91%) compared with adenocarcinoma (70%, $p=0.029$). [295]

P16 evaluation in sputum has been discussed in Chapter 2.

Methylation of p16 and O⁶-methylguanine-DNA methyltransferase (MGMT) in a high-risk population was studied. The sputum samples of 21 patients with squamous carcinoma were investigated. The presence of one or both gene promoters was found in all 21 sputum samples irrespective of whether they were taken up to 3 years ($n=10$) before or at the time of diagnosis ($n=11$). This compares to 15% (methylation of p16) and 25% (methylation of MGMT) in 123 controls, deemed high risk through a smoking history and/or radon exposure. 48% of the squamous cell cancer patients had methylation of both genes compared to only 4% of controls ($p<0.001$). [142]

Belinsky *et al* studied 3,259 subjects who provided pooled sputum for 3 days in 1 container and the next 3 days in a second container. The sputum from the second container was used as samples for the study. 182 subjects were diagnosed with (incident) lung cancer with 1,353 cohort deaths. Once sufficient quality of DNA was analysed the study cohort consisted of 98 subjects (cases) and the non-cancer cohort was used to form a group of 92 controls matched for age, gender and month of enrolment. P16 was one of the fourteen genes analysed for promoter methylation. They found that 6 (including p16) of the 14 genes were individually associated with a > 50% increased risk of lung cancer. Sputum collected within 18 months of a diagnosis of lung cancer had more methylated gene promoters than sputum from the same subject collected more than 18 months before the diagnosis. The concomitant methylation of 3 or more genes was associated with a 6.5-fold increased risk of lung cancer with receiver operating characteristics of a specificity of 64% and sensitivity of 64%. [143]

A more recent study looked at a heavy smoking but cancer-free cohort of subjects. Sputum samples were collected from 820 Caucasian subjects by self-expectoration on 3 consecutive days. They were all current or former smokers having built up at least a 20-pack year smoking history and were cytologically and radiologically free of cancer. The analysis of the sputum with various antibodies including p53 and p16 identified 56 individuals (6.9%) with at least one molecular alteration. P16 methylation occurred in 5.1%, which was the most prevalent of all the alterations. P53 mutation was the next most common with 1.9% and one of the patients with the p53 mutations developed early lung cancer on follow-up within 3 years. [296]

p53

p53 is a tumour suppressor gene. It has an important role, especially in multicellular organisms, where it regulates the cell cycle and thus acts as a tumour suppressor. [297] The role of p53 has been studied previously, and like p16, has also been shown to have an independent effect on lung cancer prognosis.

A study of 156 resected NSCLC cases was carried out using immunostaining with p53. 103 specimens (66%) expressed p53 with greater expression in SCC compared to adenocarcinomas. The cases were divided into three groups, as follows: p53-negative ($\leq 0.1\%$ stained, $n=53$), low p53 (0.1% to 50%, $n=54$), and high p53 ($> 50\%$, $n=49$). Overall, patients in the high-p53 group survived longer than those in the low or negative groups, with respective median survival durations of more than 65, 26, and 33 months ($p=0.002$). The survival difference among the three groups was statistically significant for non-squamous cell ($p=0.008$), but not for SCC ($p=0.17$). The authors concluded that high expression of the p53 oncoprotein is a favourable prognostic factor in a subset of patients with NSCLC. [298] These findings were further evaluated on 179 surgically resected NSCLC specimens where a significant relationship between strong p53 expression and patient survival was identified. In a multivariate analysis, strong expression ($> 50\%$) of the p53 oncoprotein is an independently favourable prognostic factor. [299]

It has been postulated that p53 mutation / alteration has a role in the aetiology of lung cancer. Persistent smoking dramatically increases the risk of death from lung cancer. [300] Of all of the carcinogenic effects of tobacco smoke, its mutagenic action is certainly believed to be the major cause of human lung malignancy. [301, 302] This has been discussed in detail in Chapter 3. The tobacco carcinogen Benzo[a]pyrene has been shown to have direct effects on mutating p53, with these mutations being greater in lung cancer patients. Most specifically these patterns of p53 mutations are greater in the smoking cancer group compared to non-smokers. The studies suggest that p53 mutations in lung cancers can be attributed to direct DNA damage from cigarette smoke carcinogens rather than to selection of pre-existing endogenous mutations. [171, 303, 304] Other studies have continued to confirm the presence of p53 alterations and its increased levels in NSCLC surgical specimens at a greater rate than normal tissue and most commonly in SCC. [305-307]

The role of p53 alterations in sputum has also been studied. As a follow-up of the John Hopkins Lung Project [130], 15 patients were identified who later developed adenocarcinoma or large cell cancer of the lung. The primary lung carcinomas from 10 of these 15 patients contained either a *ras* or a p53 gene mutation. Using a polymerase chain reaction-based assay (PCR), stored sputum samples obtained prior to clinical diagnosis were examined for the presence of these same oncogene mutations. In 8 of 10 patients, the identical mutation identified in the primary tumour was also detected in at least one sputum sample. The earliest that these were detected were in a sputum sample obtained 1 year prior to clinical diagnosis. [308] A study analysed sputum from 15 Chinese women with lung cancer. Of the 15 patients there were mutations in 7 (46.7%) patients, including 5 patients with p53 mutations, 1 patient with a *K-ras* mutation, and 1 patient with *K-ras* and p53 mutations. [309]

EGFR

Epidermal Growth Factor Receptor (EGFR) is a cell-surface receptor and is activated by ligands such as epidermal growth factor and transforming growth factor α . Once activated, a signalling pathway is commenced which ultimately leads to DNA synthesis and cell proliferation. Thus EGFR signalling pathway is

one of the most important pathways that regulate growth, survival, proliferation, and differentiation in mammalian cells. [310] Over-expression of EGFR (upregulation) has been associated with lung cancer.

EGFR is most frequently expressed in SCC but also frequently expressed in adenocarcinoma and large cell cancer. This was identified when 183 NSCLC tumour specimens were assessed for EGFR expression by IHC and gene copy numbers were identified by fluorescent in situ hybridization (FISH). FISH is a cytogenetic technique used to detect and localize the presence or absence of specific DNA sequences on chromosomes. EGFR protein over-expression was observed in 62% of the NSCLC, more frequently in SCC than non-SCC (82% v 44%; $p < 0.001$), and in 80% of the bronchoalveolar carcinomas. The prevalent FISH patterns were balanced disomy (40%) for EGFR gene which meant that 60% had an increased number of gene copies per cell, most of which were balanced trisomy or polysomy. Gene amplification was seen in 9% of the patients. Gene copy number correlated with protein expression ($r = 0.4$; $p < 0.001$) but EGFR over-expression or high gene copy numbers had no significant influence on prognosis. [311]

Further studies showed the over-expression of EGFR in lung cancer. 57 consecutive patients who underwent surgical resection of NSCLC had paired samples of cancerous tissue and uninvolved lung. Over-expression was identified by IHC. 88 (44 pairs) samples were of good enough quality to be examined for EGFR over-expression of which 82 exhibited EGFR expression. 20 of the 44 tumour samples (45%) had EGFR over-expression. [312] A comparison study looking at EGFR expression by IHC in SCC tumour samples compared to normal lung tissue, again showed a much higher rate of expression in the SCC group. In fact, a statistically significant stepwise increase in expression from uninvolved bronchial epithelium to precancerous lesions ultimately to SCC was observed. [313]

A meta-analysis of 18 studies, involving close to 3,000 patients analysed the role of EGFR over-expression and survival prognosis. EGFR over-expression differed between histological types: 39% in adenocarcinomas, 58% in

squamous cell carcinomas, 38% in large cell carcinomas, and 32% in cancers in a miscellaneous category ($p < 0.0001$). The combined hazard ratio (HR) was 1.14 (95% CI 0.97 to 1.34; $p = 0.103$), indicating that EGFR over-expression has no significant impact on survival. When only the 15 IHC-based studies were considered, the combined HR was 1.08 (95% CI 0.92 to 1.28; $p = 0.356$), again suggesting that EGFR over-expression has no impact on survival. [314] Studies involving sputum and EGFR expression are limited. In one of the few studies looking to identify cells that carry chromosomal alterations indicative of malignancy, the overall sensitivity of the FISH assay on induced-sputum from 52 lung cancer patients was 71% and the specificity was 100%. The most frequently detected gains were at 7p12 (EGFR) in 17 of 24 completely resectable early-stage (II+IIIA) non-small cell lung cancers (NSCLC). There was a statistically significant increase in the proportion of cases with gains of EGFR in SCC, compared with adenocarcinomas ($p = 0.017$), and a higher average EGFR gene copy number in the SCCs than in the adenocarcinomas ($p = 0.013$). [315]

Cyclin D1

Cyclin D1 has a key role in the cell cycle being of particular importance in the G1 to S-phase transition. Mutations, amplification and overexpression of this gene, which alters cell cycle progression, are observed frequently in a variety of tumors and may contribute to tumorigenesis.

Cyclin D1 was studied in 60 surgically resected NSCLC specimens. There was expression in all the samples and over-expression in 50% but there were no statistical differences within the different histological subtypes or with different stages of disease. [316] When examined on 135 surgically resected lung cancer specimens, cyclin D1 expression was found to be a negative prognostic marker ($p < 0.00005$) whose expression correlated with a shorter patient survival time. [317] The relationship of cyclin D1 over-expression and shorter survival in lung cancer was further evidenced by a study of 69 surgically resected NSCLC specimens examined by IHC, ranging from stage I and IIIa. 24 samples stained positive for cyclin D1 with over-expression being significantly higher in those with lymph node metastases (50.0% v 14.4%, $p = 0.002$). The patients with

cyclin D1 expression had a significantly shorter survival time (29.7 \pm 6.1 months vs. 74.6 \pm 8.6 months, $p=0.0066$). [318]

We could find no publications relating to cyclin D1 expression in sputum samples from lung cancer patients.

p63

p63 is part of the same family of genes as p53. Since they are structurally very similar, p63 is thought to have a similar role to p53 with regards to cell cycle arrest and apoptosis. In 89 subjects who underwent potentially curative surgery for lung cancer, the resected specimens were analysed for p63 expression using IHC. p63 expression was found in 47 and in 37/40 of the squamous cell carcinomas (SCC), 7/39 of the adenocarcinomas, 1/5 of the small cell carcinomas, and 0/5 of the large cell carcinomas. The authors concluded that p63 may have a role in the pathogenesis of SCC of the lung. [319] This was further evaluated and concluded on a study of 221 tumour specimens from patients with stage I NSCLC. p63 expression was seen in 109/118 (92%) squamous cell carcinomas, 2/2 (100%) adenosquamous carcinomas and 4/6 (67%) large cell carcinomas but only 15/95 (16%) adenocarcinomas. Furthermore, the prevalence of p63-immunoreactive cells increased progressively from pre-neoplastic and pre-invasive lesions to invasive squamous cell carcinomas. [320]

A study by Pierre Massion *et al* suggested that p63 over-expression was associated with prolonged survival. They also found that p63 expression was amplified in 88% of SCC, 42% of large cell carcinomas, and in 11% of adenocarcinomas of the lung. Both, p63 genomic amplification and protein staining intensity was associated with better survival. They also found a significant increase in the copy numbers, as assessed by FISH, in preinvasive lesions that were graded as being severe dysplasia or higher. [321, 322]

We could find no publications on p63 staining in sputum of lung cancer patients.

Strengths and weaknesses of IHC and sputum study

The strengths and weaknesses regarding patient selection are similar to those discussed earlier in the FTIR and sputum section of this Chapter.

The two cohorts had similar age, number of smokers with the high-risk cohort tending towards having more males ($p=0.06$). It is important to match for smoking status as it is biologically plausible that smoking independently affects gene methylation and pre-neoplastic differences in marker expression, before progressing to cancer.

Strengths of the study

Unique panel of gene markers:

We chose antibodies that have all been studied in cancer and more specifically lung cancer. The theoretical action of each gene makes it plausible that if altered in any way this can have a biological role in carcinogenesis. It is also more likely that a combination of alterations can lead to cancer (the 'dual hit' hypothesis) so a combination of markers may at least improve diagnostic sensitivity. Most previous studies have looked at IHC staining on tissue biopsies and surgically resected specimens. Very few studies have looked at sputum IHC in lung cancer and none have used this panel of biomarkers on sputum.

Validation of scoring:

I performed the initial scoring alongside Christine Davies (who has over 20 years experience in cytology and IHC) in order learn the scoring system and cell recognition. We then performed the second half ($n=17$) independently and these later scores were assessed for agreement. Scoring independently is important as any subjective process can be influenced by the comments or immediate presence of others. In normal clinical practice, scoring is generally performed on two separate occasions to aim for a consensus.

Weaknesses of the study

No healthy control cohort:

We did not use a healthy control group. Although a potential weakness, the need for healthy controls to retest an established method is debatable. The high-risk group acted as 'controls' and to a better degree represent the background target population for any screening validation.

Time consuming:

The process of immunostaining and interpretation is not only time consuming but the antibodies are also very expensive. It takes time to process the cells and then subsequently cut and stain the samples, and again more time to interpret and score the individual slides. Each sample required staining by 5 individual antibodies (5 x 83) giving 415 slides that required scoring.

Preparation of the slides from start to finish can take 2 hours and then it takes approximately 4 hours to process an IHC slide, although 30 can be done in a single batch.

Poor quality slides:

It became increasingly apparent during the scoring process that some of the samples were of poor quality. This was either through the technical process of staining and slide preparation or through the sputum cells themselves being of low viability for staining. (The good quality cell pellets were allocated to the FTIR study first).

The way forward: IHC and sputum

This study suggests that our techniques applying this panel of antibodies for IHC analysis of sputum, is not accurate in differentiating lung cancer from non-cancer cases. Although each of the antibodies has a potential role in different aspects of the cell cycle and possibly carcinogenesis, they had very poor sensitivity and were negative in many cancer cases. Moreover, IHC is currently an expensive and labour intensive technique. It has an important role in defining certain types of cancer and will remain a necessary tool in medicine, but as a screening modality, its role is far less certain. Following this study, we do not feel that continued assessment on sputum with this panel of antibodies is worthwhile.

Chapter 11 Exhaled Breath in Lung Cancer – discussion

Summary of key findings

The main objective of this study was to assess if volatile organic compounds detected in (resting) exhaled breath can be used to distinguish subjects with newly diagnosed lung cancer from healthy controls. As breath is easily obtained non-invasively, it is again more favourable to patients when considering a screening / diagnostic modality. It also has less infectious potential than sputum or serum. Importantly, the technology allows storage of the breath samples at room temperature in non-expensive (and reusable) facilities for long periods of time. The apparatus used for sampling was also portable which is also desirable in clinical practice for example in General Practice / nearside testing.

We prospectively collected 51 breath samples from patients at the time they were diagnosed with lung cancer (i.e. before any treatment had been commenced). We also collected 50 controls that had no pre-existing history or symptoms of respiratory disease or cancer. After exclusions, the spectra from the breaths of 47 controls and 44 lung cancers were measured.

Polynomial equations were used to develop mathematical curves that best separated lung cancer cases from controls. These equations represented the closest mathematical relationship that could be established between retention time and area under the curve for lung cancer cases, and separately, controls. On visualising the curves, two distinct zones, P1 and P2, are identified representing chemical compounds with a retention time from 8 to 17 seconds in the GC-MS. These zones contained VOCs only from lung cancer patients (27%) and no controls. We could find no other papers using this statistical approach to VOCs in lung cancer.

When all the individual VOCs were identified and a discriminating panel considered, 29 VOCs were noted to be exclusive to lung cancer patients (i.e. present in patients with lung cancer but not at all in the control cohort). Using a panel of 54 VOCs gave a sensitivity of 88.6% and specificity of 100% of diagnosing lung cancer from healthy controls.

The cellular pathway of many endogenous VOCs is unknown and needs further research, however their endogenous production can be potentially explained by whether the VOC is a saturated hydrocarbon, unsaturated hydrocarbon, oxygen-containing VOC or nitrogen-containing VOC. [323] These chemical groups will now be discussed:

- a. The production of VOCs which are saturated hydrocarbons (the cyclohexanes; cyclopentanes; decanes; dotriacontane; heptanes; nonane,3-methyl) arise from lipid peroxidation of fatty acid components of cell membranes, triggered by reactive oxygen species (ROS). This pathway is discussed in Chapter 5. Smaller quantities may result from protein oxidation and colonic bacterial metabolism. Due to their low solubility in the blood they are excreted within minutes of formation. [324]
- b. Unsaturated hydrocarbons (1-heptene;naphthalene,2-methyl-) may be produced by the mevalonic pathway of cholesterol synthesis. An example of this pathway is in the use of isoprene (an unsaturated hydrocarbon) which is normally found in human breath. The presence of isoprene in breath was noted to decrease once subjects had commenced cholesterol-limited diets and cholesterol lowering drugs. [325] There is also experimental evidence that isoprene exhalation may be produced by damage to the fluid lining of the lung and the body. [326, 327]
- c. Oxygen-containing compounds (2-propanol, 1-methoxy-; vinyl ether; N-benzyloxycarbonyl-L-tyrosine; hexane, 1-(hexyloxy)-5-methyl-) such as acetone are found in the breath and are produced by decarboxylation of acetoacetate which is derived from lipolysis or lipid peroxidation. [323] 2-propanol is postulated to be a product of an enzyme-mediated reduction

of acetone. [328] Endogenous ethanol is probably produced by intestinal bacterial flora. [329]

- d. Nitrogen-containing compounds (piperazine; piperazine adipate) can be elevated in the breath of subjects with liver failure or uraemia (high urea levels). The odour of uraemic breath is due to elevated levels of dimethylamine and trimethylamine. [330]

Although we still do not fully understand how VOCs are made and metabolised, or how VOCs are involved in certain molecular pathways, tentative links can be postulated. A study looking at exhaled aldehydes in 12 lung cancer, 12 healthy controls and 12 healthy smokers found that the cancer patients had higher levels of pentanal, nonanal, octanal and hexanal compared to the other 2 cohorts. They noted that hexanal is known to be generated through oxidative cleavage of unsaturated fatty acids, for example, arachidonic acid. [331, 332] In recent years, there has been increasing evidence suggesting that the arachidonic acid pathway plays a role in lung cancer proliferation. [333-335]

In interpreting the presence of VOCs, one can also look at the absence of certain compounds, since carcinogenesis may be causing suppression of particular VOCs. 12 VOCs were found only in the breath of our controls and absent from *all* the cancer subjects. Using these 12 VOCs, 45% of controls were correctly identified. A further 15 VOCs were universally higher in the control group and these VOCs allowed another 15% to be identified as controls. Therefore this panel of 27 VOCs unique or universally higher in controls were able to correctly identify 60% of controls. These VOCs were either metabolised more quickly by those with lung cancer, were absorbed into tissues or simply had their production suppressed / diverted into other pathways.

Comparison with other studies of exhaled breath

There are several techniques that have been used in exhaled VOC analysis in lung cancer. Many of these techniques have been discussed in Chapter 5. Our study involved GC-MS that is still the most widely used method of measuring VOCs.

In the mid-1970s a method of sampling exhaled breath was developed using a new sorbent called Tenax (a polymer based on 2,6-diphenyl-*p*-phenylene oxide). [336, 337] This method was capable of measuring sub-parts per billion levels. The subject inhales pure humidified and charcoal-scrubbed air from a 20L Tedlar bag that has previously been filled from a pure air cylinder and then exhales into a second 20L Tedlar bag. This bag is then emptied by pumping through a glass cartridge containing 1.5g Tenax. [338] An early study was carried out based on this by Gordon *et al*, looking at the exhaled breath of 12 lung cancer patients and 17 healthy controls. The patients had to inhale purified air for 5 minutes and then exhaled 40L of breath into a sampling bag on two occasions. On the third occasion this sample was drawn into the cartridge. The Tenax cartridges were analysed for volatile organic compounds by a thermal desorption GC-MS procedure. The frequency of peak generation was noted and the peaks were then matched against a historical library program. [339] 49 peaks were identified which had statistically significant difference in peak occurrence. 4 peaks were present in more than half the lung cancer cohort but were absent from the healthy controls. [159]

Phillips *et al* developed the technique further. They recruited patients who had a suspicious CXR and were undergoing bronchoscopy for this. Exhaled breath samples were collected within 24 hours pre-bronchoscopy after the subjects being fasted overnight, using a portable electrical device. [340] Patients wore a nose clip while breathing in and out of the device for 5 minutes. A 1.0L sample of breath was pumped through a sorbent trap that contained activated carbon and captured the VOCs for analysis. Ambient air samples were also collected after each patient sample. The VOCs were then separated by gas chromatography and analysed by mass spectrometry. 108 patients were

recruited to the study, lung cancer being confirmed histologically in 60 patients (50 NSCLC, 10 SCLC) and excluded in 48. Many VOCs were common to both sets of breath samples but a group of 22 of these VOCs (identified by discriminant analysis) were able to distinguish between lung cancer and control cases according to risk weighting attached to the test. For example, the VOCs had 100% sensitivity (i.e. no false negatives) and 81% specificity for stage I lung cancer if the receiver operator accepted a post-test probability of 0.46 (see Figure 11.2). A post-test probability of 0.9 yielded lower 66.7% sensitivity but 100% specificity (no false positives). The abnormal VOCs consisted of mainly alkanes and benzene derivatives. Smoking status alone did not account for the benzene derivatives since these were also present in the breath of non-smokers and ex-smokers, so it was likely to be a genuine marker for lung cancer. Of the 22 VOCs they used as their panel, 3 were exact compounds in our 29 cancer-exclusive panel (heptane, 2,2,4,6,6-pentamethyl; nonane,3-methyl; 1-heptene) and 3 were similar VOCs (decane; cyclohexane; cyclopentane,methyl-) which accounted for a further 7 of our VOC panel. A further 3 VOCs from their panel of 22 appear in our VOCs which were "higher in cancer" group, either identically (heptane,3-methyl-) or similar (benzene; benzene,1,2,3-trimethyl-). [160]

Using similar methodology, the same researchers compared exhaled VOCs in patients with biopsy proven primary lung cancer (n=67); non-lung cancer metastasising to the lungs (n=15), abnormal CXRs but no histological evidence of lung cancer (n=5) and healthy volunteers from the general population (n=41). The breath test identified over 80 different alkanes and mono-methylated alkanes that were then used via discriminant functional analysis to generate a predictive model using a panel of 9 VOCs. These 9 VOCs, in combination yielded a sensitivity of 89.6% and a specificity of 82.9% (using a post-test probability of 0.5) of identifying primary or secondary lung cancer. Again, there were only minor differences when subjects were stratified according to history of tobacco smoking, histological cancer type, or lung cancer staging. Interestingly, none of our cancer-exclusive VOCs was in their panel of 9, although they did have decane and hexane compounds, which were present in our 29. Butane was their best single discriminator, and although not present in

the 29 exclusive VOCs, we had butane, 2-methyl- present in the higher in cancer group as well as hexane,3-methyl-. [161]

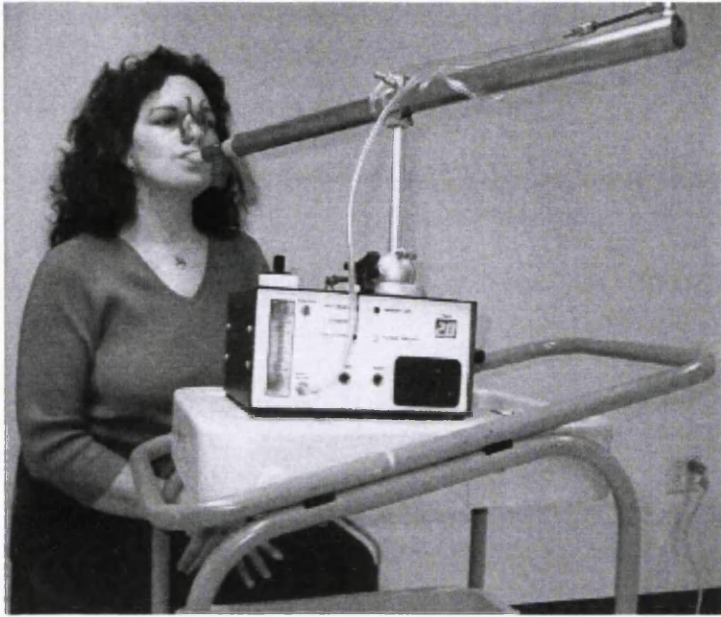


Figure 11.1: An example of the breath collection apparatus used by Phillips *et al.* [161]

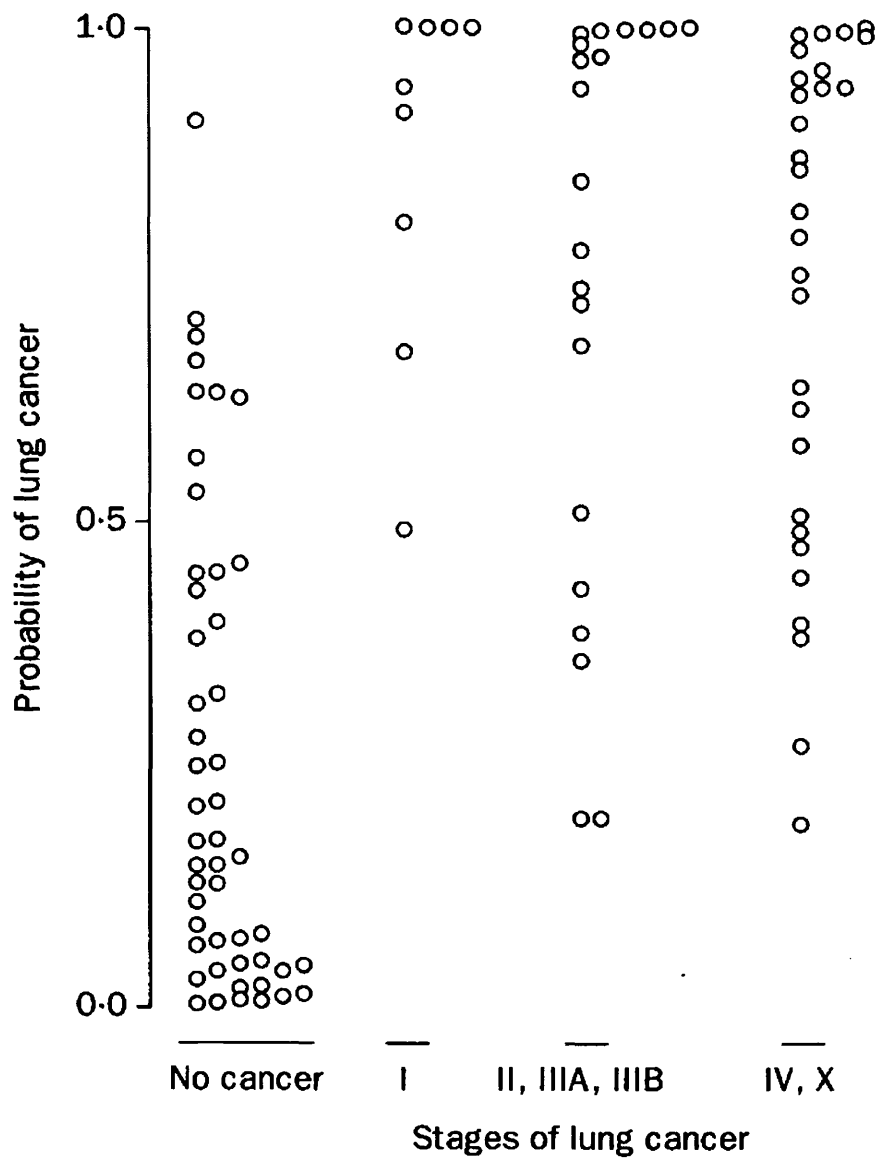


Figure 11.2: Post-test probability of lung cancer by breath VOC assay [160]

An Italian research group had previously completed a study using the most similar apparatus to ours on 36 patients with surgically resected NSCLC, 25 COPD controls and 50 healthy controls. The cancer breath samples were all collected pre-operatively. They used the same breath collection apparatus i.e. the BioVOC sampler (Markes International Ltd, Rhondda Cynon Taff, UK) as us. After 60 minutes rest, the subjects were asked to perform a single slow vital capacity breath into a one-way valve connected to a Teflon®-bulb, which traps the last portion of exhaled air (150 ml) (see Figure 11.3). It was not reported whether the patients were fasted. Twenty environmental samples were taken from the rooms in which the subjects performed the test, in order to compare breath and ambient air VOC levels - but the timing of these samples was not published. The VOCs were extracted by gas chromatography and analysed by mass spectrometry. The NSCLC, COPD and smoking controls had generally higher levels of exhaled VOCs than the non-smoking controls suggesting, unlike Phillips' group, that smoking status was very important. Although no single VOC could distinguish NSCLC from the other groups, a panel of 13 VOCs could be used to correctly classify 72.2% of subjects as having NSCLC. On the basis of the results the overall sensitivity was 72.2% and specificity 93.6%. Of their panel of VOCs they included pentamethylheptane and decane which were similar to our panel of cancer exclusive VOCs. They also recorded trimethyl-benzene, which in our study was one of the VOCs universally higher in cancer cases; they also had octane, heptane and benzene and we found similar if not exact compounds in the "universally higher in cancer" panel. [162]

Early this year, 2010, Song *et al* published a study looking at exhaled breath from 43 NSCLC patients and 41 healthy controls who were non-smokers, all from China. Of the cancer group, 11 patients received 1 cycle of chemotherapy and the breath sample was obtained 4 weeks after the treatment, and the rest of the cancer cohort did not receive any type of treatment. The patients were fasted overnight and 15 ambient air samples were taken from the room where the patients gave their samples. The subjects were then asked to breathe deeply onto 4L Tedlar® bags. Mixed expiratory samples were collected with no restriction on the particular part of breath. All the samples were analysed using standard GC-MS. Statistical analysis between the two groups was performed

using Wilcoxon rank sum test (Mann-Whitney U test). They identified 2 particular VOCs, 1-butanol and 3-hydroxy-2-butanone which were significantly higher in the cancer group compared to the controls ($p < 0.001$). There was no significant difference between the level of the 2 VOCs and the different stages of lung cancer. They noted that the VOCs they identified were not seen in previous panels of VOCs (and are not present in ours), and suggested the differences maybe due to the ethnicity of the study population or the fact that the study population was small with some of the patients having late stage lung cancer. They also commented that their differences may be due to different extraction methods, but interestingly did not comment on the fact that 11 of the 43 cancer patients had undergone chemotherapy, which in itself may have altered the endogenous VOC profile. [163]

The main finding from our study was the presence of 29 cancer-exclusive VOCs. Many were similar to VOCs reported others but our panel also included previously unreported VOCs.

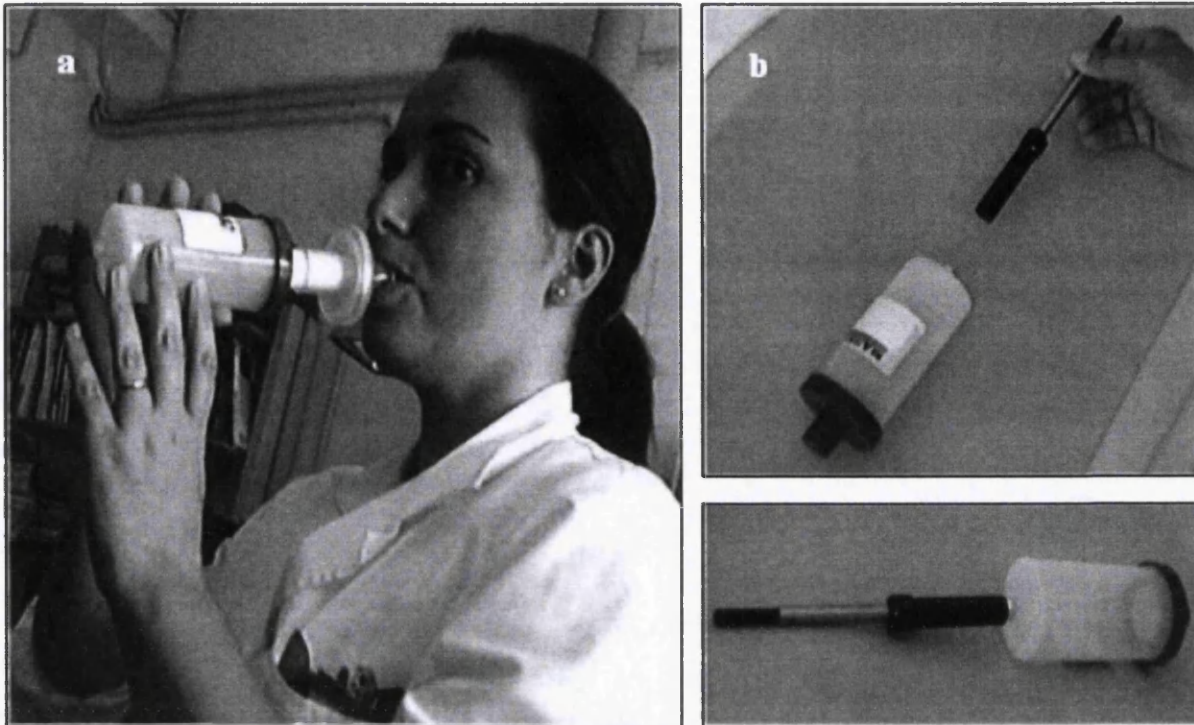


Figure 11.3: Breath collection and VOC extraction. The subjects performed a single slow vital capacity into a Teflon® bulb (Bio-VOC® breath sampler) (a) which traps the last portion of exhaled air (150 mL); the VOCs were extracted by directly inserting a 75 mm Carboxen/PDMS SPME fiber (30 min) into the bulb (b). [162]

Strength and weaknesses of exhaled VOCs study

Strengths of the study

Sample size:

Although we recruited 101 patients initially, for various reasons only 91 patients were included in the data analysis. This sample size compares well with others. Gordon *et al* had a study population of 29 (12 cancers vs. 17 controls); Poli *et al* had 111 (36 cancers vs. 25 COPD vs. 50 healthy controls) and Phillips *et al* initially had 108 (60 cancers vs. 48 controls) and then subsequently changed the arms of his study to have 128. [159-162]

Timing of recruitment:

The patients were recruited as early as possible i.e. at the time of diagnosis and before treatment that could change levels of VOCs.

Control group:

The healthy control group consisted of subjects who had no symptoms or history of respiratory disease including infection or cancer which may affect VOCs. The groups had similar gender distribution helping to eliminate this as a potential confounder. Although there was a significant difference in self-reported smoking status (especially never smokers) between the two groups ($p < 0.001$), similar proportions had smoked within 24 hours of their breath sampling as judged by eCO levels > 10 ppm ($p = 0.83$). This is probably of more direct relevance, as smoking within 24 hours should have greater effect on the exhaled VOCs. Some previous studies have shown no significant effect of smoking status on VOCs, whilst others have shown increased levels of a large number of VOCs when comparing smoking controls with their non-smoking counterparts. [161, 162] It is possible that although some subjects said they were smokers at time of diagnosis / recruitment, the fear of having cancer may have shocked them into stopping smoking prior to further investigations.

Validation of smoking status:

All previous studies on lung cancer and VOCs did not validate smoking status but relied on self-reporting. We are the first to validate smoking status using a standard definition of exhaled CO level above 10 ppm to confirm the consumption of a cigarette in the preceding 24 hours. 15% of our cancer subjects and 12% of controls had smoked within 24 hours prior to sampling. As each cigarette contains over 4000 noxious chemicals and burns at over 400° centigrade, it seems plausible they will acutely affect VOCs but the timescale taken for VOCs to return to baseline after each cigarette is unknown.

Fasting participants pre-test:

We fasted all our patients for 4 hours, which we felt was both feasible and realistic bearing in mind our study population, many of whom had co-morbidities. Although there was no clear guidance in the literature over fasting, it was felt that eating certain foods or drinking certain beverages might alter the VOC components of the breath. Phillips *et al*, found evidence in 12 healthy subjects that many VOCs had higher alveolar concentrations than ambient air concentrations suggesting these VOCs were either synthesised in the body, or absorbed from another site, possibly as a drug or in the food. Figure 5.3 is a schematic of the potential mechanisms of endogenous / exogenous VOC production postulated by Phillips *et al*. [250] There are other studies which have also reported the presence of VOCs in food products. [341, 342] Unfortunately the dietary intake of the subjects was not recorded.

Resting participants pre-test:

It was important that all subjects were rested for 20 minutes pre breath test for three reasons.

- 1) strenuous exertion can cause increased oxidative stress and thus changes in VOC concentration [273, 274]
- 2) by remaining in a single room for 20 minutes the subject was able to inhale the surrounding environmental air to allow better equilibrium for alveolar air. Although there is no clear indication from the literature as to how long a subject should be inside a room, pharmacokinetic models

assumes that inhaled air resides inside the alveoli for sufficient time to allow VOCs to reach equilibrium with arterial blood. [338]

- 3) blood flow rate through the alveoli is not linear with breathing rate, thus different levels of exertion could potentially have an effect on equilibrium being reached.

We felt that 20 minutes provided enough time for the individual to fully recover from any exertion and for them to return back to their resting heart and respiratory rate whilst also keeping the methodology simple and time efficient.

Ambient air sampling:

We took ambient air samples from the same room immediately before each set of breath samples were taken. This allowed us to analyse how many of the discriminating VOCs were present in the air and if these were affecting the results. A study looking at the variation of VOCs within normal human breath took 50 fasted subjects and analysed their exhaled breath using GC-MS. There was a mean of 204 VOCs (range 157 – 241) per individual breath, and a total of 3,481 VOCs were observed at least once (the majority were only identified once). Only 27 VOCs were present in all subjects. Ambient air samples were taken for each individual and any VOCs were subtracted from their corresponding breath sample to create “alveolar gradients”. A positive alveolar gradient suggests there was more in the breath than in the air and vice versa for the negative alveolar gradient. Of the 3,481 VOCs detected, 1,753 had positive alveolar gradients and 1,728 had negative ones, although the majority of the negatives were found in only one subject. It has been postulated that when there is more of a certain VOC in the ambient air sample compared to the exhaled breath (i.e. negative alveolar gradient) then this corresponds with increased clearance of the VOCs via hepatic and/or renal pathways. [283]

Phillips *et al*, has considered how researchers have approached the situation of when VOCs are present in both the ambient air sample and the subject's exhaled breath, raising the question of whether the VOC originated from within the body or whether it is just contamination from inhaling the surrounding air. They considered three approaches [283]:

- 1) Ignore the problem, either by not collecting a corresponding ambient air sample like some [159, 164], or by collecting them but not fully incorporating the air samples in the overall analysis. [162, 256]
- 2) Provide the subjects with VOC-free air to breathe or to breathe through apparatus to filter out VOCs, prior to sample collection. [252, 343].
Although this seems reasonable, Phillips felt that it is virtually impossible to achieve VOC-free air in practice.
- 3) Correct for VOCs in the ambient air by, subtracting it from the exhaled breath VOCs [160, 161, 283], or by filtering out those VOCs which are higher in the air sample or only marginally less than the exhaled breath sample. [344]

In the studies discussed earlier in this Chapter, Poli *et al* took twenty ambient air samples from the rooms in which the subject samples were taken but it was not commented whether this was done at the same time as the subject sample or randomly or even on different days. [162] Song *et al*, took 15 ambient air samples from the rooms that the patients samples were collected, however, it was not clear at which point in time these air samples were taken. [163] Phillips *et al* did take ambient air samples at the same time as the subject samples. [160, 161]

In our study, when considering the 29 cancer-exclusive VOCs, some of these were only present in the corresponding air samples of 2 patients. In fact S15 accounted for the majority of the cancer-exclusive VOCs found in their air samples, with the air and patient samples being taken at the subject's home. We cannot recall anything unusual about his home and it was not next to a main road or near a factory and the background (ambient air sample) showed no unusual spectra. S15 was a male who was previously fit and well. He had an extensive smoking history of greater than 100-pack year history. Thus in the 31 patients correctly classified as having lung cancer based on the panel of 29 VOCs, 29 had none of the discriminating VOCs in their ambient air samples, so subtraction or calculation of ambient gradients is not needed. Of the 12 VOCs, found exclusively in controls, only in 2 patients were some VOCs found in their

corresponding air sample meaning that at least 19/21 of the controls had these exclusive VOCs from an endogenous source rather than from background air. Reassuringly, none of the VOCs that were universally higher in controls were present in their corresponding air samples.

We believe that only by simultaneously assessing the ambient air samples for every single breath test can we really be sure that compounds are endogenous rather than contaminants.

Triplicate samples:

Following external review on early pilot data, we amended our protocol and collected the patient samples in triplicate after the first 16 patients. This was to evaluate the consistency of VOC production / exhalation and measurement within the individual subjects in the same 5 minute session. This does not appear to have been done before in lung cancer using GC-MS. The other advantage of performing multiple samples is to ensure that important VOCs are not missed due to poor subject technique (for example not full exhalation) or individual tube leaks.

After taking the triplicate samples, we noticed some inconsistencies in VOCs measured for each subject. Although the actual (qualitative measurements) VOCs themselves appeared similar across the three samples, after the first sample some of the subsequent samples appeared to have lower levels of the same VOCs. There were other occasions when the 2nd or 3rd patient sample contained the higher level of VOCs. Possible explanations could be some of the patients providing good slow vital capacity initially but then becoming fatigued with subsequent attempts to fully exhale not being as good quality.

Alternatively, where the VOCs concentrations were higher in the later tubes participants could be improving their technique with practice. There may also be variations in the sealing of the tubes and connecting tubes to the GC-MS.

Breath sample collection apparatus:

The apparatus used for sample collection was small, lightweight so very portable. (Indeed we have sent some equipment to Canada recently). It was

also straightforward to use for the investigator and patient. The subjects were merely required to perform a slow vital capacity breath and then the sample could be stored at room temperature (in a tin) for up to several months and not require freezing or specialist storage facilities. Such a technique would be feasible to use in clinical practice (e.g. as a bedside test).

Weaknesses of the study

Sample size:

I discussed our sample size with Michael Phillips, Clinical Professor of Clinical Medicine, New York Medical College; he stated:

“A good rule of thumb in multivariate analysis is that you need at least 6 subjects in the disease group for each variable in your predictive algorithm”.

Taking this into account, using each of the 29 cancer-exclusive VOCs as a variable, we would ideally require at least 174 (i.e. 29 x 6) lung cancer patients. For multivariate analysis, our study is likely to be underpowered.

Participant demographics:

Our control group was significantly younger than the lung cancer group ($p < 0.001$). Phillips *et al* studied the effect of aging on VOCs, where they investigated 102 subjects ranging from the ages of 9 to 89. The abundance of VOCs increased significantly with advancing age ($p < 0.001$), which could represent increased oxidative stress as one gets older although there may be contribution from the decline in clearance of the cytochrome P450 system. [242] Bearing this in mind, it would be ideal to have a more age-matched group (e.g. spouses rather than medical staff) to eliminate this confounder.

Sampling from multiple rooms:

Ideally the subject samples should all have been taken in a single room to minimise the effect of different ambient air mixtures on the VOCs. Not only can these directly ‘contaminate’ the collection apparatus (e.g. tubes and plunger) but they are inhaled and then exhaled by subjects. However, even in the same room, there are continuous changes in VOC concentrations, so we relied on

taking ambient air samples before each set of breath samples. We did not take ambient air samples in triplicate because of resource implications (number of tubes and GC-MS technician time). Phillips *et al* took ambient air samples in the bronchoscopy suite after the patient samples. Poli *et al* took 20 ambient air samples from the room but did not state at which point these were taken and neither did Song *et al* (See previous section comparing other studies).

Equipment failure:

During late April 2009, a set of 7(x3) patient samples and their ambient air samples (28 tubes in total) were processed but were identified to have a very diminished level of total VOCs (<20 VOC peaks) and low levels of each VOC detected. Laboratory procedure records were reviewed and it was noted that there was a pressure loss in the helium supply during the secondary desorption phase on this sample run. Also, the heavy use of the equipment at this time required the cold trap to be reconditioned. Replacing the helium cylinder and reconditioning of the cold trap later corrected this. These 28 samples were not included in the analysis.

Improving trapping of VOCs:

The Teflon bulb used only collected the last 129ml of exhaled breath. If we had a larger collection device or a method of collecting multiple breaths on to one tube, this would mean a greater volume of breath and a larger number of VOCs. There is also the problem of breath condensation within the tubes. The human breath contains much more vapour than VOCs and although the sorbents within the tubes that trap the VOCs are hydrophobic and should repel the vapour, it is thought that the water can compete with the VOCs in reaching the binding sites of the sorbent thus interfering with the trapping of the VOCs. One solution to this potential issue would be having equipment which removes the water vapour from the exhaled breath prior to VOC sampling. The breath vapour itself might also contain VOCs.

One further technical issue which may have affected the trapping of VOCs was the varying speeds at which the plunger was pushed to transfer the exhaled

breath into the sorbent tube. Excessive speed will prevent trapping of VOCs. An automated system would remove operator influence.

Separation column:

We used non-polar columns in the GC-MS. If we had used a combination of non-polar and polar columns in series, then polar VOCs such as alcohols would be more readily detected and better separated. Also, the starting temperature in the column was 40°C which was then heated to 200°C. By having a GC oven which started at a lower temperature, for example 0°C, more highly volatile compounds might be detected. This requires an additional cooling unit which was not present on the equipment.

Mass spectrometer:

Some mass spectrometers are more sensitive and thus pick up lower concentration VOCs and also have less background noise. This might offer improvements given the low abundance of VOCs in breath.

VOC identification and validation:

Automated peak detection was used to calculate peak area (area under the curve) and retention time for each compound. Each match was also assigned a quality rating based on how well it matched the chemical library. There is some debate on what is deemed an acceptable quality rating and in this preliminary analysis we wanted to include all VOCs before excluding a potentially key compound. [281] Further analysis is being planned where we will include only those compounds with a quality rating above 50% (most of our VOCs) to see if this alters any key findings. Extensive manual peak identification was also performed to check the validity of the automated matches; ideally however, the identity of the compounds should be verified by comparing retention times of pure chemicals injected into the sorbent tubes (and tested as with the breath samples) allowing precision and accuracy testing in our environment.

Unfortunately we did not have the resources for this, as each standard VOC sample is expensive and calibration is time consuming.

Number of VOCs detected:

Having discarded VOCs which occurred on less than 5 occasions overall, the total number of VOCs identified was 246. In comparison, using Phillips' technique, we have already discussed how in normal healthy human breath he was able to detect 3,481 VOCs in total and a mean of 204 VOCs per patient from the 50 patients sampled. [283] The relative lack of total number of VOCs may well represent trapping difficulties when the patients are giving exhaled breath, or GC-MS detection difficulties especially of VOCs low in concentration. The fact that we discarded those VOCs which occurred less than 5 times will have some impact on the total number of VOCs but this impact is likely to be small, with these compounds tending to be poor quality matches.

Validation of cancer-exclusive compounds and statistical techniques:

Although 29 cancer-exclusive VOCs were identified some occurred in just 1 lung cancer patient and not in the other 43 and it may have occurred in only 1 out of that patient's triplicate samples. The validity of these VOCs is difficult to assess at this stage but we are considering further collaborations with the Department of Engineering to apply further discriminatory statistics based on categorical clusters on larger sample numbers. Here for example, we would first concentrate on VOCs found in all 3 samples of lung cancer subjects but in 0/3 samples of all controls. A secondary analysis would be looking at a VOC found in 2/3 samples of lung cancer patients and 0/3 of controls etc. Another approach which is being undertaken is to take the median values of each VOC across the triplicates and compare them between the cancer and control cohorts to see if any are exclusive or universally higher.

With sufficient data available for these 29-54 VOCs, a further way to validate these tentative findings is to use a method such as logistic regression to generate a linear model using a training dataset and test it prospectively on many more subjects. The resulting linear model equation would allow prediction of cases as being either lung cancer or non-lung cancer.

The way forward: exhaled VOCs and lung cancer

This study has identified certain volatile organic compounds, which are present exclusively in lung cancer patients and are able to classify a high percentage of them correctly. Even on 50-80 samples, two distinct polynomial equations representing lung cancer and controls could be developed. Further testing on at least 180 treatment naive patients with lung cancer, together with refinements in technique (single room for collection, comparing against age / smoking-matched controls, continuous monitoring for helium leaks) should test the specificity and sensitivity of our test. Recruitment continues locally with other centres (e.g. Canada) joining the project.

The various studies assessing VOCs in lung cancer have developed varying panels and this will continue to be the case until technology is refined and clinical sampling is standardised. We too have applied unique techniques (triplicate samples, simultaneous ambient air and novel statistical modelling) and added to the body of knowledge in VOC analysis in lung cancer to show some promising results. Our group wants to pursue and expand on the early pilot work with the ultimate goal of developing non-invasive diagnostic and ultimately screening tests for lung cancer. However, the concepts remain experimental at this stage, and the optimal methodology is still being worked out. Better understanding of VOCs may even give some ideas of the biological pathways for lung cancer opening up new targets for future therapies.

Closing statement

Lung cancer is a common and devastating disease for which most people will have no cure. During the last 4 years I have developed a better understanding of research and how to approach it. I have learnt laboratory skills such as slide preparation and scoring for immunohistochemistry as well as techniques such as Fourier Transform Infra-Red and gas chromatography – mass spectrometry. It has also given me an opportunity to learn complex multivariate statistical analysis as well as more standard statistical methods and to identify weaknesses of the methodology in order to strengthen the studies. The studies are to be pursued with the ultimate goal of developing a screening test that can identify lung cancer patients early enough to have a positive impact in reducing mortality rates.

APPENDIX

Appendix 1

HEADED NOTE PAPER

PATIENT INFORMATION SHEET -version 2, 8th February 2006

Title: Assessing early detection biomarkers in lung cancer

You are being invited to take part in a research study. Before you decide it is important to explain why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Consumers for Ethics in Research (CERES) publish a leaflet entitled "Medical Research and You". This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained without charge from CERES, PO BOX 1365, LONDON N16 0BW.

1. What is the purpose of the study?

We believe that chemicals and tiny cell abnormalities may be important in detecting lung cancer. These chemicals or biomarkers may be detectable even before patients develop symptoms or abnormal chest x-rays. We want to test if these chemicals can be detected in the spit and biopsy specimens of lung cancer and also if they are detectable in patients who do not yet have visible lung cancers.

2. Why have I been chosen?

Your doctor is concerned that you may be suffering from lung cancer and we would like you to be one of our trial subjects whilst you attend our clinic and bronchoscopy.

3. Do I have to take part?

It's up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. If you withdraw after a long period of time, you must be aware that we may have already measured your samples. Withdrawal will not affect the standard of care you receive.

4. What will happen to me if I take part?

Following standard assessment and care in our clinic, we would like to send 2 extra samples of your spit for chemical analysis, when you attend for bronchoscopy. During the bronchoscopy, we will also take 2-4 small samples in addition to the standard tests. These samples will be taken to Swansea University for special analysis and disposed of after 10 years.

After the bronchoscopy, you will receive our standard treatment for your condition. Normally, if all our tests are clear and you get better, we would discharge you after 1 year. For this research however, we would like to see you in clinic for a clinical review once a year for 5 years, even after all other tests and treatment, or simply for a check-up. This research does not involve any extra or different treatment to normal but we will take extra samples in the beginning and follow you over a longer time.

5. What do I have to do?

There are no lifestyle restrictions and we want you to continue with standard treatment at all times.

6. What are the side-effects of taking part?

We are not trying out any new treatments but this would involve the inconvenience of providing an extra spit test and extra samples at bronchoscopy.

You will not feel the extra samples being taken at bronchoscopy.

Each sample is about 1-2 mm in size. Normally there is a small (1 in 100) chance of bleeding but the extra research biopsies are 1 in 50 chance of bleeding. The procedure will be performed by an experienced doctor and won't be performed until standard samples are taken and only if you are still well. It should not mean any extra time in hospital. You will be asked to attend clinic 1 year later and once a year for five years, even if you are well. This is to see if any of the biomarkers predict disease developing later on. These annual visits may entail extra inconvenience.

7. What are the possible benefits of taking part?

The extra annual appointments for five years may detect future cancer early, before it causes symptoms or spreads. The results of the research using your samples, and those from others – may benefit lung cancer patients in the future. We cannot pay you for your participation in the study.

8. What happens when the research study stops?

After five years, we will review you in clinic only if needed as part of standard NHS care.

9. What happens if something goes wrong?

If you have reason to complain about any aspect of the way you have been approached or treated during the course of the study, the normal National Health Service complaints mechanisms are available to you.

10. Will my taking part in this study be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information that leaves the hospital will be coded so that you cannot be recognised from it. Your GP will be notified that you have helped us with this study but won't know the results of the research. Information will not be made available to any life insurance or private medical insurance companies.

11. What will happen to the results of the research study?

The results may be published in scientific journals and presented at conferences. You are welcome to contact the researchers for a report, when this study ends (2014). Individuals will not be identified in any report/publication. The researchers in the hospital or those giving treatment will not know the results of the biomarker tests because these biomarkers are not currently part of standard treatment.

12. Who is organising and funding the research?

This is jointly organised through our local hospitals and University of Wales Swansea.

13. Who has reviewed the study?

This has undergone peer review by academic colleagues within Wales and has been presented to the Research Ethics Committees in Swansea and Dyfed Powys.

Contact for further information.

If you have any further queries please contact:

**Dr Keir Lewis
Consultant Physician in Respiratory Medicine,
Carmarthenshire NHS Trust and
Senior Lecturer,
School of Medicine,
Swansea University.**

☎ 01554 783133

Fax : 01792 – 513054

Email k.e.lewis@swansea.ac.uk

Or

**Dr Philip Kloer
Consultant Physician in Respiratory Medicine,
Carmarthenshire NHS Trust**

☎ 01554 783569

Or

**Dr Carol Llewellyn Jones
Consultant Physician in Respiratory Medicine,
Carmarthenshire NHS Trust**

☎ 01267 227616

Appendix 2
HEADED PAPER

CONSENT FORM-Version 1, 8th November 2005

Study Number ...05/WMW01/75

Patient Identification number for this trial

N.B. Three copies will be made For
(1) patient
(2) researcher
(3) hospital notes

Title of Project: Assessing early detection biomarkers in lung cancer

Name of Researchers: **Dr Keir Lewis, Dr Paul Lewis, Dr Phil Kloer, Dr Carol Llewellyn –Jones, Dr Robin Ghosal.**

Contact Telephone Number: **01554 783133**

Please Initial

Box

- | | | |
|----|---|---|
| 1. | I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions. | <input style="width: 50px; height: 20px;" type="text"/> |
| 2. | I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. | <input style="width: 50px; height: 20px;" type="text"/> |
| 3. | I understand that sections of any of my medical notes may be looked at by responsible individuals from Dr Lewis' research team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. | <input style="width: 50px; height: 20px;" type="text"/> |
| 4. | I agree to take part in the above study. | |

Name of Patient	Date	Signature
Name of Person taking consent	Date	Signature
Researcher	Date	Signature

Appendix 3

HEADED NOTEPAPER

GP INFORMATION SHEET (version 2: 8th February 2006)

Title of study:

Assessing early detection biomarkers in lung cancer
(LREC number 05/MMW01/75)

Your patient *has agreed to participate in the above study that is being organised through Carmarthenshire NHS Trust and Swansea University.*

This involves standard assessment in our lung cancer clinic but an extra sputum test prior to bronchoscopy and 2-4 extra biopsies taken during standard bronchoscopy. They will then also be asked to attend chest clinic, annually for five years, to see if certain biomarkers predict later development of lung cancer.

This is an observational study and all standard NHS treatment will continue throughout the study. Management will not be compromised in any way.

All information will be kept in the strictest confidence and no individual patients can be identified by anyone other than the lead researcher.

Subjects have provided written consent and can withdraw at any time from the study.

Dr Keir E Lewis
Senior Lecturer Swansea School of Medicine and Consultant in
Respiratory Medicine
Carmarthenshire NHS Trust

Dr Paul Lewis
Lecturer South West Wales Cancer Institute

Appendix 4

Clinical Record Form (version 2: 15th Aug 2007)

ID Number..... Surname..... Forename.....
Hospital No..... Consent Date.....

Consent Taker..... Visit No..... Hospital... POW/PPH/WWGH

Medical Hx:
.....
.....

Past Hx of Cancer...Y/ N Type of Cancer.....

Drug Hx:
.....
.....

Smoking Hx Current/ Ex/ Never No. of Pack Yrs.....

Occupation..... Hx of Asbestos Exposure...Y/ N

FHx of Cancer.....Y/ N Type/Relative.....

FEV₁ ..n/a or% of Pred

Performance Status ...0/ 1/ 2/ 3/ 4 Suspected Infection....Y/ N

CO level.....ppm O₂ Sats pre-bronch..... %

CXR

Symptoms

Wt loss....Y/ N Haemoptysis....Y/ N Unexplained SOB....Y/N
Cough....Y/ N Chest Pain.... Y/ N Hoarse Voice....Y/ N

Examination

Clubbing...Y/ N Lymphadenopathy....Y/ N SVCO... Y/N
Organomegaly... Y/ N Cachexia... Y/N Monophonic Wheeze. Y/N
Pleural Effusion...Y/ N

Bronchoscopic View.....

Diagnosis

Clinical / Radiological diagnosis of Lung Cancer.... Y/ N

Sputum Cytology....Positive/ Negative/ Not applicable
(report)
.....

Bronchoscopy Histology
.....

Final Diagnosis (histology) ...NSCLC / SCLC/ Other
Other details.....

Lung Cancer Staging.... (NSCLC) T N M
 Limited / Extensive
 (SCLC)

Sample Collection

Sputum Time:
Biopsy Time:

Samples Collected by..... Date.....

Appendix 5

HEADED NOTEPAPER

HEALTHY CONTROLS INFORMATION SHEET- version 2, 30th September 2008

Title: Assessing early detection biomarkers in lung cancer

You are being invited to take part in a research study. Before you decide it is important to explain why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

1. What is the purpose of the study?

We believe that chemicals and tiny cell abnormalities may be important in detecting lung cancer. These chemicals or biomarkers may be detectable even before patients develop symptoms or abnormal chest x-rays. We want to test if these chemicals can be detected in breath, spit and biopsy specimens of patients suspected of having lung cancer compared to healthy controls and patients with chronic lung disease.

2. Why have I been chosen?

We would like you to be one of our healthy control subjects.

3. Do I have to take part?

It's up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

4. What will happen to me if I take part?

Following screening questions, we will ask you for your age and then will ask you to breathe deeply into a container that measures the amount of carbon monoxide in your breath. This number will be recorded on a data sheet and is a marker of pollution and cigarette smoke exposure that can

alter chemicals in your breath. We will then ask you to spit into a collection container and these sputum samples will be taken to Swansea University for special analysis looking at different types 'biomarkers' before being stored in freezers. The samples will be disposed of after 10 years.

5. What do I have to do?

There are no lifestyle restrictions and we want you to continue with your normal activities unchanged.

6. What are the side-effects of taking part?

There is the inconvenience of taking 5 minutes of your time to answer the 6 simple screening questions.

7. What are the possible benefits of taking part?

The results of the research using your samples, and those from others – may benefit lung cancer patients in the future. We cannot pay you for your participation in the study.

8. What happens when the research study stops?

Your sputum will be stored in the University but there is no other clinical contact needed.

9. What happens if something goes wrong?

If you have reason to complain about any aspect of the way you have been approached or treated during the course of the study, the normal National Health Service complaints mechanisms are available to you.

10. Will my taking part in this study be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information that leaves the hospital will be coded so that you cannot be recognised from it. Information will not be made available to any life insurance or private medical insurance companies.

11. What will happen to the results of the research study?

The results may be published in scientific journals and presented at conferences. You are welcome to contact the researchers for a report, when this study ends (2009/10). Individuals will not be identified in any report/publication. The researchers will not know the results of the biomarker tests because these biomarkers are not currently part of standard treatment.

12. Who is organising and funding the research?

This is jointly organised through our local hospitals, Swansea University and funded by the Welsh Assembly Government.

13. Who has reviewed the study?

This has undergone peer review by academic colleagues within Wales and has been presented to the Research Ethics Committees in Swansea and Dyfed Powys.

Contact for further information.

If you have any further queries please contact:

**Dr Keir Lewis
Consultant Physician in Respiratory Medicine,
Carmarthenshire NHS Trust and
Senior Lecturer,
School of Medicine,
Swansea University.**

☎ 01554 783133

Fax : 01792 – 513054

Email k.e.lewis@swansea.ac.uk

Or

Dr Philip Kloer

**Consultant Physician in Respiratory Medicine,
Carmarthenshire NHS Trust**

☎ 01554 783569

Or

Dr Paul Lewis,

Lecturer in Biomedical Science,

Institute of Life Science,

Swansea School of Medicine,

Swansea University,

SA2 8PP

☎ 01792 295222

Appendix 6

HEADED NOTE PAPER

PATIENT INFORMATION SHEET - version 2, 3rd June 2008

Title: Smoking status, exhaled volatile organic compounds and lung cancer

You are being invited to take part in a research study. Before you decide it is important to explain why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

1. What is the purpose of the study?

We believe that smoking status at the time of diagnosis of lung cancer and after may have an impact on the course of the condition, and we want to test to see if this is the case. Smoking is one of the main causes of lung cancer and we believe that continued smoking even after diagnosis may “fuel” the cancer as well as interfere with treatment compared with not smoking. We also want to see if chemicals present in your breath (but not detected by routine tests) will have an impact on your condition and whether these chemicals change over time.

2. Why have I been chosen?

Your doctor is concerned that you may be suffering from lung cancer and we would like you to be one of our trial subjects if this diagnosis is confirmed.

3. Do I have to take part?

It's up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. If you withdraw after a long period of time, you must be aware that we may have already measured your samples. Withdrawal will not affect the standard of care you receive.

4. What will happen to me if I take part?

If, following standard assessment and care in our clinic, we believe you do have lung cancer, we will obtain special breath test samples. The first of these will involve you taking a breath in and then holding it for 15 seconds before slowing breathing out into a monitor, which has a disposable mouthpiece. The second breath test will involve you simply taking a couple of deep breaths in and out, then blowing slowly into a piece of measuring apparatus. Together, these procedures should take under 3 minutes.

We will see you once a month in the hospital (this will coincide with your normal clinic appointments wherever possible). On each visit, the two breath tests are repeated and you will have a brief examination and any problems will be addressed. Should you become unwell or are unable to attend hospital, two of the doctors can come to your home if you are happy for the breath tests to be performed there and this visit could be at the same time as any other planned home visits by e.g. specialist nurses, to avoid further inconvenience. All extra travelling and car parking costs as a result of the study will be reimbursed by the research team.

5. What do I have to do?

There are no lifestyle restrictions and we want you to continue with standard treatment at all times.

6. What are the side-effects of taking part?

We are not trying out any new treatments but participation in the study will involve the inconvenience of attending the hospital once a month for the next 12 months. These consultations will only last 10 minutes, and you will not have to wait around for too long prior to being seen. The breath tests do not hurt.

7. What are the possible benefits of taking part?

The extra monthly appointments during the year will allow us to assess your general health and address any problems. It will give you more contacts than usual with the doctors treating you. The results of the research may benefit lung cancer patients in the future.

8. What happens when the research study stops?

After one year, we will review you in clinic only if needed as part of standard NHS care.

9. What happens if something goes wrong?

If you have reason to complain about any aspect of the way you have been approached or treated during the course of the study, the normal National Health Service complaints mechanisms are available to you.

10. Will my taking part in this study be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information that leaves the hospital will be coded so that you cannot be recognised from it. If you agree, your GP will be notified that you have helped us with this study and they will be allowed access to any published results, although individual patients results will not be available to them. Information will not be made available to any life insurance or private medical insurance companies.

11. What will happen to the results of the research study?

The results may be published in scientific journals and presented at conferences. You are welcome to contact the researchers for a report, when this study ends (2009). Individuals will not be identified in any report/publication.

12. Who is organising and funding the research?

This is jointly organised through Carmarthenshire NHS Trust and Swansea University.

13. Who has reviewed the study?

This has undergone peer review by academic colleagues within Wales and has been presented to the Research Ethics Committees in Dyfed Powys.

Contact for further information.

If you have any further queries please contact:

**Dr Robin Ghosal,
Clinical Research Fellow,
Hywel Dda NHS Trust,
☎ 01554 756 567 pager 873**

Or

**Dr Keir Lewis
Consultant Physician in Respiratory Medicine,
Hywel Dda NHS Trust and
Senior Lecturer,
School of Medicine,
Swansea University.**

☎ 01554 783133

Fax : 01554 783597

Email k.e.lewis@swansea.ac.uk

Or

**Dr Philip Kloer
Consultant Physician in Respiratory Medicine,
Hywel Dda NHS Trust**

☎ 01554 783569

Appendix 7

Lung Cancer Clinical Data Proforma

Pt Name..... DOB.....

Patient Number..... Telephone number

Age.....

Date of consultation..... Consultant.....

CXR.....

Symptoms: Cough / Wt loss / Haemoptysis / Hoarse voice / Chest pain / SOB

Co-morbidities COPD IHD DM CVA Obesity ↑BP Ca
Other.....

Tumour Position : Central / Peripheral **FHx** Y / N

Occupation **Asbestos exposure** Y / N

FEV₁.....% of Pred **Performance Status** 0 1 2 3 4

CO level at clinicppm **Date of 1st CO**

Smoking Hx Current / Ex / Never / Passive **No. of pack years**.....

Current no. cigarettes per day **Referred for smoking cessation**
Y N Declined

Final Diagnosis No Ca / Small Cell / NSCLC / Other
.....

Date of diagnosis
.....

Subtype Squamous / Adenocarcinoma / Large cell / Undifferentiated

CT scan staging T N M **Stage**..... Limited / Extensive

Cancer treatment Surgery / Radical Radiotherapy / Chemotherapy /
Radiotherapy / Palliative

Smoking Status Validation:

Month	CO level (PPM)	Current no. of Cigarettes/ day	Date	Month	CO level (PPM)	Current no. of Cigarettes/ day	Date
1				5			
2				6			
3				7			
4				8			

D o D..... **Signed**

Version 1. Ghosal R / Kloer PJ / Guy C. Feb 2008

Appendix 8

HEADED PAPER

CONSENT FORM-Version 1, 23rd March 2008

Study Number ...08/WMW01/21

Patient Identification number for this trial

N.B. Three copies will be made for
(1) patient
(2) researcher
(3) hospital notes

Title of Project: Smoking status, exhaled volatile organic compounds and lung cancer

Name of Researchers: Dr Robin Ghosal, Dr Keir Lewis, Dr Phil Kloer, Dr Elinor Young, Dr Jacqui Orme

Contact Telephone Number: 01554 783133

Please Initial

Box

1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.
3. I understand that sections of any of my medical notes may be looked at by responsible individuals from Dr Ghosal's research team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.
4. I agree to my GP being informed of my participation in the study.
5. I agree to take part in the study.

Name of Patient	Date	Signature
Name of Person taking consent	Date	Signature
Researcher	Date	Signature

Appendix 9

HEADED NOTEPAPER

GP INFORMATION SHEET (version 1: 23rd March 2008)

Title of study:

Smoking status, exhaled volatile compounds and lung cancer
(REC number: 08/MMW01/21)

Your patient *has agreed to participate in the above study that is being organised through Carmarthenshire NHS Trust and Swansea University.*

This involves standard assessment in our lung cancer clinic plus two additional breath tests, to measure exhaled carbon monoxide and volatile compounds. Patients are then also asked to attend for assessment and follow up breath tests on monthly basis for the next 12 months, to see if these volatile compounds in their breath and validated smoking status have any ability to predict clinical outcomes (survival or response to treatment). If patients cannot come to hospital for these extra tests, we will seek their permission to visit them in their home.

This is an observational study only and all standard NHS treatment will continue throughout. Management will not be compromised in any way.

All information will be kept in the strictest confidence and no individual patients can be identified by anyone other than the lead researcher.

Your patient has provided written consent and has agreed that you be informed of their participation. They are fully aware that they can withdraw at any time from the study without giving a reason.

Dr Robin Ghosal
Clinical Research Fellow
Carmarthenshire NHS Trust

Dr Keir E Lewis
Senior Lecturer, Swansea School of Medicine and Consultant in Respiratory Medicine,
Carmarthenshire NHS Trust

Dr Philip Kloer
Consultant in Respiratory Medicine and Clinical Lead in Lung Cancer,
Carmarthenshire NHS Trust

Appendix 10 HEADED NOTEPAPER

HEALTHY CONTROL INFORMATION SHEET- version 1, 10TH October 2008

Title: Smoking status, exhaled volatile organic compounds and lung cancer

You are being invited to take part in a research study. Before you decide it is important to explain why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

1. What is the purpose of the study?

We want to see if smoking status at the time of diagnosis of lung cancer and afterwards has an impact on the course of the illness and response to treatment. We also want to see if chemicals present in patient's breath (but not detected by routine tests) will help determine diagnosis and longer term outcome and whether these chemicals change over time.

2. Why have I been chosen?

We would like you to be one of our 50 or so healthy control subjects to compare levels of these chemicals in your breath with 50 patients diagnosed with lung cancer .

3. Do I have to take part?

It's up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

4. What will happen to me if I take part?

Following screening questions (see attached), to check if you are eligible, we will ask you for your age and then will ask you to breathe deeply into a container that measures the amount of carbon monoxide in your breath. This number will be recorded on a data

sheet and is a marker of pollution and cigarette smoke exposure that can determine chemicals in your breath. We will then ask you to take a deep breath in and exhale fully into another container where we will trap your breath and these samples will be analysed at Swansea University for levels of these chemicals in their mass spectrometry department. We will only be required to do these tests on one occasion and your active participation will end at this point.

5. What do I have to do?

In order to take part you must be fasting (only had water) for the previous 4 hours. You must also not exert yourself beyond light walking for 20 minutes just before the breath tests.

6. What are the side-effects of taking part?

There is the inconvenience of taking 5 minutes of your time to answer the 6 simple screening questions and do two breath tests.

7. What are the possible benefits of taking part?

The results of the research using your samples, and those from others - may help develop less invasive screening and monitoring tools to benefit lung cancer patients in the future. We cannot pay you for your participation in the study.

8. What happens when the research study stops?

Your breath samples will have been analysed, but no further clinical follow up is required. The samples will be discarded immediately after analysis but the (anonymous) results will be kept on file for ten years.

9. What happens if something goes wrong?

If you have reason to complain about any aspect of the way you have been approached or treated during the course of the study, the normal National Health Service complaints mechanisms are available to you.

10. Will my taking part in this study be kept confidential?

All information collected about you during the course of the research will be kept strictly confidential. Any information that leaves the hospital will be coded so that you cannot be recognised from it. Information will not be

made available to any life insurance or private medical insurance companies.

11. What will happen to the results of the research study?

The results may be published in scientific journals and presented at conferences. You are welcome to contact the researchers for a report summary, when this study ends (2009). Individuals will not be identified in any report/publication.

12. Who is organising and funding the research?

This is jointly organised through Hywel Dda NHS Trust and Swansea University.

13. Who has reviewed the study?

This has undergone peer review by academic colleagues within Wales and has been presented to the Research Ethics Committees in Dyfed Powys.

Contact for further information.

If you have any further queries please contact:

Dr Robin Ghosal,
Clinical Research Fellow,
Hywel DDa NHS Trust,
☎ 01554 756 567 pager 873

Or

Dr Keir Lewis
Consultant Physician in Respiratory Medicine,
Hywel Dda NHS Trust and
Senior Lecturer,
School of Medicine,
Swansea University.

☎ 01554 783133

Fax: 01554 783597

Email k.e.lewis@swansea.ac.uk

Or

Dr Philip Kloer

**Consultant Physician in Respiratory Medicine,
Hywel Dda NHS Trust**

☎ 01554 783569

BIBLIOGRAPHY

1. *Lung Cancer*, in www.wikipedia.org.
2. *Lung Cancer Facts (women)*, in *National Lung Cancer Partnership*. 2006.
3. WHO, *Deaths by cause, sex and mortality stratum*, in *The World Health Report*. 2004. p. 120-125.
4. Hong, W.K. and A.S. Tsao, *Lung Carcinoma: Tumors of the Lungs*, in *Merck Manual Professional Medical Library Online Edition*. 2008.
5. Peto, R., et al., *Mortality from smoking worldwide*. *Br Med Bull*, 1996. **52**(1): p. 12-21.
6. Peto, R., et al., *Mortality from tobacco in developed countries: indirect estimation from national vital statistics*. *Lancet*, 1992. **339**(8804): p. 1268-78.
7. Biesalski, H.K., et al., *European Consensus Statement on Lung Cancer: risk factors and prevention*. *Lung Cancer Panel*. *CA Cancer J Clin*, 1998. **48**(3): p. 167-76; discussion 164-6.
8. *IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans.*, in *Tobacco smoke and involuntary smoking*. 2004, WHO International Agency for Research on Cancer.
9. Carmona, R.H., *The health consequences of involuntary exposure to tobacco smoke: A report of the Surgeon General*. 2006, U.S Department of Health and Human Services.
10. Catelinois, O., et al., *Lung cancer attributable to indoor radon exposure in france: impact of the risk models and uncertainty analysis*. *Environ Health Perspect*, 2006. **114**(9): p. 1361-6.
11. Chiu, H.F., et al., *Outdoor air pollution and female lung cancer in Taiwan*. *Inhal Toxicol*, 2006. **18**(13): p. 1025-31.
12. Coyle, Y.M., et al., *An ecological study of the association of metal air pollutants with lung cancer incidence in Texas*. *J Thorac Oncol*, 2006. **1**(7): p. 654-61.
13. Hackshaw, A.K., M.R. Law, and N.J. Wald, *The accumulated evidence on lung cancer and environmental tobacco smoke*. *Bmj*, 1997. **315**(7114): p. 980-8.

14. Kabir, Z., K. Bennett, and L. Clancy, *Lung cancer and urban air-pollution in Dublin: a temporal association?* Ir Med J, 2007. **100**(2): p. 367-9.
15. O'Reilly, K.M., et al., *Asbestos-related lung disease.* Am Fam Physician, 2007. **75**(5): p. 683-8.
16. Tredaniel, J., et al., [*The consequences of passive smoking in adults*]. Rev Mal Respir, 2006. **23 Suppl 2**: p. 4S67-4S73.
17. Kiyohara, C., et al., *Genetic polymorphisms and environmental risk of lung cancer: a review.* Rev Environ Health, 2004. **19**(1): p. 15-38.
18. Kiyohara, C., et al., *Genetic polymorphisms and lung cancer susceptibility: a review.* Lung Cancer, 2002. **37**(3): p. 241-56.
19. Verdecchia, A., et al., *Recent cancer survival in Europe: a 2000-02 period analysis of EURO CARE-4 data.* Lancet Oncol, 2007. **8**(9): p. 784-96.
20. *Lung Cancer survival statistics*, in www.cancerresearchuk.org.
21. Ferlay, J., et al., *Estimates of the cancer incidence and mortality in Europe in 2006.* Ann Oncol, 2007. **18**(3): p. 581-592.
22. *Commonly diagnosed cancers worldwide.* 2005, Cancer Research UK.
23. *Cancer*, in www.who.int.
24. Behera, D. and T. Balamugesh, *Lung cancer in India.* Indian J Chest Dis Allied Sci, 2004. **46**(4): p. 269-81.
25. Liu, B.Q., et al., *Emerging tobacco hazards in China: 1. Retrospective proportional mortality study of one million deaths.* Bmj, 1998. **317**(7170): p. 1411-22.
26. *Key findings about the quality of care for people with Lung Cancer in England and Wales - Report for the audit period 2006*, in *National Lung Cancer Audit.* 2006. p. 26.
27. *UK Lung Cancer incidence statistics*, in www.cancerresearchuk.org.
28. *Cancer Incidence and Mortality.* 2008, Northern Ireland Cancer Registry.
29. *Cancer Incidence in Wales.* 2008, Welsh Cancer Intelligence and Surveillance Unit.
30. *Cancer incidence, Mortality and Survival data.* 2008, ISD Online.
31. *Cancer Statistics registrations: registrations of cancer diagnosed in 2005, England.* Series MB1 no.36. 2005, Office for National Statistics.

32. Etienne-Mastroianni, B., et al., *Primary sarcomas of the lung: a clinicopathologic study of 12 cases*. Lung Cancer, 2002. **38**(3): p. 283-9.
33. Morandi, U., C. Casali, and G. Rossi, *Bronchial typical carcinoid tumors*. Semin Thorac Cardiovasc Surg, 2006. **18**(3): p. 191-8.
34. Travis, W.D., L.B. Travis, and S.S. Devesa, *Lung cancer*. Cancer, 1995. **75**(1 Suppl): p. 191-202.
35. Bryant, A. and R.J. Cerfolio, *Differences in epidemiology, histology, and survival between cigarette smokers and never-smokers who develop non-small cell lung cancer*. Chest, 2007. **132**(1): p. 185-92.
36. Butler, C.A., et al., *Variation in lung cancer survival rates between countries: do differences in data reporting contribute?* Respir Med, 2006. **100**(9): p. 1642-6.
37. Lababede, O., M.A. Meziane, and T.W. Rice, *TNM staging of lung cancer: a quick reference chart*. Chest, 1999. **115**(1): p. 233-5.
38. Mountain, C.F., *Revisions in the International System for Staging Lung Cancer*. Chest, 1997. **111**(6): p. 1710-7.
39. Denoix, P.F., *Enquete permanent dans les centres antercancereux*. Bull Inst Nat Hyg, 1946. **1**: p. 70.
40. Hassan, I., *Lung Cancer, Staging*, in www.emedicine.medscape.com. 2009.
41. Goldstraw, P., et al., *The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours*. J Thorac Oncol, 2007. **2**(8): p. 706-14.
42. Minna, J.D., *Harrison's Principles of Internal Medicine*. 2004: McGraw-Hill. 506-516.
43. Hamilton, W., et al., *What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study*. Thorax, 2005. **60**(12): p. 1059-65.
44. Schreiber, G. and D.C. McCrory, *Performance characteristics of different modalities for diagnosis of suspected lung cancer: summary of published evidence*. Chest, 2003. **123**(1 Suppl): p. 115S-128S.
45. Agusti, C., et al., *Induced sputum in the diagnosis of peripheral lung cancer not visible endoscopically*. Respir Med, 2001. **95**(10): p. 822-8.

46. *The diagnosis and treatment of lung cancer*, in www.nice.org.uk. 2005.
47. McLennan, G., et al., *Bronchoscopy Atlas*, in www.int-med.uiowa.edu. 2006.
48. Strand, T.E., et al., *Risk factors for 30-day mortality after resection of lung cancer and prediction of their magnitude*. *Thorax*, 2007. **62**(11): p. 991-7.
49. Graham, E.A. and J.J. Singer, *Landmark article Oct 28, 1933. Successful removal of an entire lung for carcinoma of the bronchus. By Evarts A. Graham and J. J. Singer*. *Jama*, 1984. **251**(2): p. 257-60.
50. Cahan, W.G., *Radical lobectomy*. *J Thorac Cardiovasc Surg*, 1960. **39**: p. 555-72.
51. Cahan, W.G., W.L. Watson, and J.L. Pool, *Radical pneumonectomy*. *J Thorac Surg*, 1951. **22**(5): p. 449-73.
52. Douillard, J.Y., et al., *Adjuvant vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIa non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial*. *Lancet Oncol*, 2006. **7**(9): p. 719-27.
53. Winton, T., et al., *Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer*. *N Engl J Med*, 2005. **352**(25): p. 2589-97.
54. *BTS guidelines: guidelines on the selection of patients with lung cancer for surgery*. *Thorax*, 2001. **56**(2): p. 89-108.
55. Ginsberg, R.J., et al., *Modern thirty-day operative mortality for surgical resections in lung cancer*. *J Thorac Cardiovasc Surg*, 1983. **86**(5): p. 654-8.
56. Massard, G., et al., *Bronchogenic cancer in the elderly: operative risk and long-term prognosis*. *Thorac Cardiovasc Surg*, 1996. **44**(1): p. 40-5.
57. Yellin, A., L.R. Hill, and Y. Lieberman, *Pulmonary resections in patients over 70 years of age*. *Isr J Med Sci*, 1985. **21**(10): p. 833-40.
58. Miller, J.I., Jr., *Physiologic evaluation of pulmonary function in the candidate for lung resection*. *J Thorac Cardiovasc Surg*, 1993. **105**(2): p. 347-51; discussion 351-2.

59. Boushy, S.F., et al., *Clinical course related to preoperative and postoperative pulmonary function in patients with bronchogenic carcinoma*. Chest, 1971. **59**(4): p. 383-91.
60. Wernly, J.A., et al., *Clinical value of quantitative ventilation-perfusion lung scans in the surgical management of bronchogenic carcinoma*. J Thorac Cardiovasc Surg, 1980. **80**(4): p. 535-43.
61. Eagle, K.A., et al., *Cardiac risk of noncardiac surgery: influence of coronary disease and type of surgery in 3368 operations*. CASS Investigators and University of Michigan Heart Care Program. Coronary Artery Surgery Study. Circulation, 1997. **96**(6): p. 1882-7.
62. Thatcher, N., M. Ranson, and H. Anderson, *Carcinoma of the lung*. European Respiratory Monograph, 1995: p. 270.
63. Feld, R., R. Arriagada, and D. Ball, *Prognostic factors in non-small cell lung cancer: a consensus report*. Lung Cancer, 1991. **7**: p. 3-5.
64. Radzikowska, E., P. Glaz, and K. Roszkowski, *Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases*. Ann Oncol, 2002. **13**(7): p. 1087-93.
65. Oken, M.M., et al., *Toxicity and response criteria of the Eastern Cooperative Oncology Group*. Am J Clin Oncol, 1982. **5**(6): p. 649-55.
66. Pagni, S., J.A. Federico, and R.B. Ponn, *Pulmonary resection for lung cancer in octogenarians*. Ann Thorac Surg, 1997. **63**(3): p. 785-9.
67. Pierce, R.J., et al., *Preoperative risk evaluation for lung cancer resection: predicted postoperative product as a predictor of surgical mortality*. Am J Respir Crit Care Med, 1994. **150**(4): p. 947-55.
68. Au, J., R. el-Oakley, and E.W. Cameron, *Pneumonectomy for bronchogenic carcinoma in the elderly*. Eur J Cardiothorac Surg, 1994. **8**(5): p. 247-50.
69. Deslauriers, J., et al., *Prospective assessment of 30-day operative morbidity for surgical resections in lung cancer*. Chest, 1994. **106**(6 Suppl): p. 329S-330S.
70. Naunheim, K.S., et al., *Lung cancer surgery in the octogenarian*. Eur J Cardiothorac Surg, 1994. **8**(9): p. 453-6.

71. Jacobson, M.J., et al., *A comparison of wedge and segmental resection of the lung*. Thorax, 1976. **31**(4): p. 365-8.
72. Weissberg, D., C.J. Straehley, and N.M. Scully, *Less than lobar resections for bronchogenic carcinoma*. Scand J Thorac Cardiovasc Surg, 1993. **27**: p. 121-6.
73. Temeck, B.K., P.W. Schafer, and N. Saini, *Wedge resection for bronchogenic carcinoma in high-risk patients*. South Med J, 1992. **85**(11): p. 1081-3.
74. Jensik, R.J., L.P. Faber, and C.F. Kittle, *Segmental resection for bronchogenic carcinoma*. Ann Thorac Surg, 1979. **28**(5): p. 475-83.
75. Errett, L.E., J. Wilson, and R.C. Chiu, *Wedge resection as an alternative procedure for peripheral bronchogenic carcinoma in poor-risk patients*. J Thorac Cardiovasc Surg, 1985. **90**: p. 656-61.
76. Warren, W.H. and L.P. Faber, *Segmentectomy versus lobectomy in patients with stage I pulmonary carcinoma. Five-year survival and patterns of intrathoracic recurrence*. J Thorac Cardiovasc Surg, 1994. **107**(4): p. 1087-93; discussion 1093-4.
77. Murray, N. and A.T. Turrisi, 3rd, *A review of first-line treatment for small-cell lung cancer*. J Thorac Oncol, 2006. **1**(3): p. 270-8.
78. Spiro, S.G., et al., *Chemotherapy versus supportive care in advanced non-small cell lung cancer: improved survival without detriment to quality of life*. Thorax, 2004. **59**(10): p. 828-36.
79. Cellerino, R., et al., *A randomized trial of alternating chemotherapy versus best supportive care in advanced non-small-cell lung cancer*. J Clin Oncol, 1991. **9**(8): p. 1453-61.
80. Cartei, G., et al., *Cisplatin-cyclophosphamide-mitomycin combination chemotherapy with supportive care versus supportive care alone for treatment of metastatic non-small-cell lung cancer*. J Natl Cancer Inst, 1993. **85**(10): p. 794-800.
81. Cullen, M.H., et al., *Mitomycin, ifosfamide, and cisplatin in unresectable non-small-cell lung cancer: effects on survival and quality of life*. J Clin Oncol, 1999. **17**(10): p. 3188-94.
82. Cormier, Y., et al., *Benefits of polychemotherapy in advanced non-small-cell bronchogenic carcinoma*. Cancer, 1982. **50**(5): p. 845-9.

83. Ganz, P.A., et al., *Supportive care versus supportive care and combination chemotherapy in metastatic non-small cell lung cancer. Does chemotherapy make a difference?* Cancer, 1989. **63**(7): p. 1271-8.
84. Helsing, M., et al., *Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. Joint Lung Cancer Study Group.* Eur J Cancer, 1998. **34**(7): p. 1036-44.
85. Kaasa, S., et al., *Symptomatic treatment versus combination chemotherapy for patients with extensive non-small cell lung cancer.* Cancer, 1991. **67**(10): p. 2443-7.
86. Leung, W.T., et al., *Combined chemotherapy and radiotherapy versus best supportive care in the treatment of inoperable non-small-cell lung cancer.* Oncology, 1992. **49**(5): p. 321-6.
87. Quoix, E., A. Dietemann, and J. Charbonneau, *La chimiothérapie comportant du cisplatine est-elle utile dans le cancer bronchique non microcellulaire au stade IV? Resultats d'une etude randomisee. [Is chemotherapy with cisplatin useful in bronchial nonsmall cell at stage IV? The results of a randomised study.* Bull Cancer, 1991. **78**: p. 341-346.
88. Rapp, E., et al., *Chemotherapy can prolong survival in patients with advanced non-small-cell lung cancer--report of a Canadian multicenter randomized trial.* J Clin Oncol, 1988. **6**(4): p. 633-41.
89. Thongprasert, S., et al., *Relationship between quality of life and clinical outcomes in advanced non-small cell lung cancer: best supportive care (BSC) versus BSC plus chemotherapy.* Lung Cancer, 1999. **24**: p. 17-24.
90. Woods, R.L., et al., *A randomised trial of cisplatin and vindesine versus supportive care only in advanced non-small cell lung cancer.* Br J Cancer, 1990. **61**(4): p. 608-11.
91. Socinski, M.A., et al., *Chemotherapeutic management of stage IV non-small cell lung cancer.* Chest, 2003. **123**(1 Suppl): p. 226S-243S.
92. *Effects of vinorelbine on quality of life and survival of elderly patients with advanced non-small-cell lung cancer. The Elderly Lung Cancer*

- Vinorelbine Italian Study Group. J Natl Cancer Inst, 1999. 91(1): p. 66-72.*
93. Anderson, H., et al., *Gemcitabine plus best supportive care (BSC) vs BSC in inoperable non-small cell lung cancer--a randomized trial with quality of life as the primary outcome. UK NSCLC Gemcitabine Group. Non-Small Cell Lung Cancer. Br J Cancer, 2000. 83(4): p. 447-53.*
94. Ranson, M., et al., *Randomized trial of paclitaxel plus supportive care versus supportive care for patients with advanced non-small-cell lung cancer. J Natl Cancer Inst, 2000. 92(13): p. 1074-80.*
95. Roszkowski, K., et al., *A multicenter, randomized, phase III study of docetaxel plus best supportive care versus best supportive care in chemotherapy-naive patients with metastatic or non-resectable localized non-small cell lung cancer (NSCLC). Lung Cancer, 2000. 27(3): p. 145-57.*
96. Jaakkimainen, L., et al., *Counting the costs of chemotherapy in a National Cancer Institute of Canada randomized trial in nonsmall-cell lung cancer. J Clin Oncol, 1990. 8(8): p. 1301-9.*
97. Marino, P., et al., *Single-agent chemotherapy versus combination chemotherapy in advanced non-small cell lung cancer: a quality and meta-analysis study. Lung Cancer, 1995. 13(1): p. 1-12.*
98. Lilenbaum, R.C., P. Langenberg, and K. Dickersin, *Single agent versus combination chemotherapy in patients with advanced nonsmall cell lung carcinoma: a meta-analysis of response, toxicity, and survival. Cancer, 1998. 82(1): p. 116-26.*
99. Lilenbaum, R.C., et al., *Single-agent versus combination chemotherapy in advanced non-small-cell lung cancer: the cancer and leukemia group B (study 9730). J Clin Oncol, 2005. 23(1): p. 190-6.*
100. Delbaldo, C., et al., *Benefits of adding a drug to a single-agent or a 2-agent chemotherapy regimen in advanced non-small-cell lung cancer: a meta-analysis. Jama, 2004. 292(4): p. 470-84.*
101. Wagner, H., Jr., *Radiation therapy in the management of limited small cell lung cancer: when, where, and how much? Chest, 1998. 113(1 Suppl): p. 92S-100S.*

102. Ng, M., et al., *Tolerability of accelerated chest irradiation and impact on survival of prophylactic cranial irradiation in patients with limited-stage small cell lung cancer: review of a single institution's experience*. J Thorac Oncol, 2007. **2**(6): p. 506-13.
103. Slotman, B., et al., *Prophylactic cranial irradiation in extensive small-cell lung cancer*. N Engl J Med, 2007. **357**(7): p. 664-72.
104. Lally, B.E., et al., *Postoperative radiotherapy for stage II or III non-small-cell lung cancer using the surveillance, epidemiology, and end results database*. J Clin Oncol, 2006. **24**(19): p. 2998-3006.
105. Celebioglu, B., et al., *High dose rate endobronchial brachytherapy effectively palliates symptoms due to inoperable lung cancer*. Jpn J Clin Oncol, 2002. **32**(11): p. 443-8.
106. Raben, A. and B. Mychalczak, *Brachytherapy for non-small cell lung cancer and selected neoplasms of the chest*. Chest, 1997. **112**(4 Suppl): p. 276S-286S.
107. Bhise, S.B., A.D. Nalawade, and H. Wadhawa, *Role of protein tyrosine kinase inhibitors in cancer therapeutics*. Indian Journal of Biochemistry & Biophysics, 2004. **41**: p. 273-280.
108. Gunby, R.H., et al., *Oncogenic Fusion Tyrosine Kinases As Molecular Targets for Anti-Cancer Therapy*. Anti-Cancer Agents in Medicinal Chemistry, 2007. **7**(6): p. 594-611.
109. Bencardino, K., et al., *Epidermal growth factor receptor tyrosine kinase inhibitors for the treatment of non-small-cell lung cancer: results and open issues*. Intern Emerg Med, 2007. **2**(1): p. 3-12.
110. Sandler, A., *Bevacizumab in non small cell lung cancer*. Clin Cancer Res, 2007. **13**(15 Pt 2): p. s4613-6.
111. Sandler, A., et al., *Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer*. N Engl J Med, 2006. **355**(24): p. 2542-50.
112. *Screening (medicine)*, in www.wikipedia.org.
113. Wilson, J.M.G. and G. Jungner, *Principles and Practice of Screening for Disease*. WHO Chronicle, 1968. **22**(11): p. 473.
114. Kaneko, M., et al., *Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography*. Radiology, 1996. **201**(3): p. 798-802.

115. Henschke, C.I., et al., *Early Lung Cancer Action Project: overall design and findings from baseline screening*. Lancet, 1999. **354**(9173): p. 99-105.
116. Sobue, T., et al., *Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project*. J Clin Oncol, 2002. **20**(4): p. 911-20.
117. Swensen, S.J., et al., *Lung cancer screening with CT: Mayo Clinic experience*. Radiology, 2003. **226**(3): p. 756-61.
118. Diederich, S., et al., *Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers*. Radiology, 2002. **222**(3): p. 773-81.
119. Diederich, S., et al., *Screening for early lung cancer with low-dose spiral computed tomography: results of annual follow-up examinations in asymptomatic smokers*. Eur Radiol, 2004. **14**(4): p. 691-702.
120. Nawa, T., et al., *Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies*. Chest, 2002. **122**(1): p. 15-20.
121. Sone, S., et al., *Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner*. Br J Cancer, 2001. **84**(1): p. 25-32.
122. Sone, S., et al., *Mass screening for lung cancer with mobile spiral computed tomography scanner*. Lancet, 1998. **351**(9111): p. 1242-5.
123. Berrington de Gonzalez, A. and S. Darby, *Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries*. Lancet, 2004. **363**(9406): p. 345-51.
124. Infante, M., et al., *A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial*. Am J Respir Crit Care Med, 2009. **180**(5): p. 445-53.
125. Oken, M.M., et al., *Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial*. J Natl Cancer Inst, 2005. **97**(24): p. 1832-9.
126. *National Lung Screening Trial*, www.cancer.gov/nlst.
127. Berlin, N.I., *Overview of the NCI Cooperative Early Lung Cancer Detection Program*. Cancer, 2000. **89**(11 Suppl): p. 2349-51.

128. Flehinger, B.J., et al., *Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Memorial Sloan-Kettering study*. *Am Rev Respir Dis*, 1984. **130**(4): p. 555-60.
129. Melamed, M.R., et al., *Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York*. *Chest*, 1984. **86**(1): p. 44-53.
130. Frost, J.K., et al., *Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins study*. *Am Rev Respir Dis*, 1984. **130**(4): p. 549-54.
131. Marcus, P.M., et al., *Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up*. *J Natl Cancer Inst*, 2000. **92**(16): p. 1308-16.
132. Fontana, R.S., et al., *Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study*. *Am Rev Respir Dis*, 1984. **130**(4): p. 561-5.
133. Fontana, R.S., et al., *Lung cancer screening: the Mayo program*. *J Occup Med*, 1986. **28**(8): p. 746-50.
134. Fontana, R.S., et al., *Screening for lung cancer. A critique of the Mayo Lung Project*. *Cancer*, 1991. **67**(4 Suppl): p. 1155-64.
135. Kubik, A. and J. Polak, *Lung cancer detection. Results of a randomized prospective study in Czechoslovakia*. *Cancer*, 1986. **57**(12): p. 2427-37.
136. Kubik, A.K., D.M. Parkin, and P. Zatloukal, *Czech Study on Lung Cancer Screening: post-trial follow-up of lung cancer deaths up to year 15 since enrollment*. *Cancer*, 2000. **89**(11 Suppl): p. 2363-8.
137. *Biomarker*, in www.wikipedia.org.
138. Field, J.K., et al., *Consensus statements from the Second International Lung Cancer Molecular Biomarkers Workshop: a European strategy for developing lung cancer molecular diagnostics in high risk populations*. *Int J Oncol*, 2002. **21**(2): p. 369-73.
139. Greenberg, A.K., H. Yee, and W.N. Rom, *Preneoplastic lesions of the lung*. *Respir Res*, 2002. **3**: p. 20.
140. Brabender, J., et al., *Comprehensive expression analysis of retinoic acid receptors and retinoid X receptors in non-small cell lung cancer:*

- implications for tumor development and prognosis. Carcinogenesis, 2005. 26(3): p. 525-30.*
141. Belinsky, S.A., *Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer, 2004. 4(9): p. 707-17.*
 142. Palmisano, W.A., et al., *Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res, 2000. 60(21): p. 5954-8.*
 143. Belinsky, S.A., et al., *Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. Cancer Res, 2006. 66(6): p. 3338-44.*
 144. Jahr, S., et al., *DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res, 2001. 61(4): p. 1659-65.*
 145. Sozzi, G., et al., *Detection of microsatellite alterations in plasma DNA of non-small cell lung cancer patients: a prospect for early diagnosis. Clin Cancer Res, 1999. 5(10): p. 2689-92.*
 146. Chen, X.Q., et al., *Microsatellite alterations in plasma DNA of small cell lung cancer patients. Nat Med, 1996. 2(9): p. 1033-5.*
 147. Xue, X., Y.M. Zhu, and P.J. Woll, *Circulating DNA and lung cancer. Ann N Y Acad Sci, 2006. 1075: p. 154-64.*
 148. Ludovini, V., et al., *Plasma DNA, microsatellite alterations, and p53 tumor mutations are associated with disease-free survival in radically resected non-small cell lung cancer patients: a study of the perugia multidisciplinary team for thoracic oncology. J Thorac Oncol, 2008. 3(4): p. 365-73.*
 149. Betta, P.-G., et al., *Codon 12 K-ras mutations in plasma DNA are not an indicator of disease in non-small cell lung cancer (NSCLC) patients. J Clin Oncol, 2005. 23: p. 7204.*
 150. Camps, C., et al., *Is there a prognostic role of K-ras point mutations in the serum of patients with advanced non-small cell lung cancer? Lung Cancer, 2005. 50(3): p. 339-46.*
 151. Aviel-Ronen, S., et al., *K-ras mutations in non-small-cell lung carcinoma: a review. Clin Lung Cancer, 2006. 8(1): p. 30-8.*

152. Pathak, A.K., et al., *Circulating cell-free DNA in plasma/serum of lung cancer patients as a potential screening and prognostic tool*. Clin Chem, 2006. **52**(10): p. 1833-42.
153. Bremnes, R.M., R. Sirera, and C. Camps, *Circulating tumour-derived DNA and RNA markers in blood: a tool for early detection, diagnostics, and follow-up?* Lung Cancer, 2005. **49**(1): p. 1-12.
154. Andriani, F., et al., *Detecting lung cancer in plasma with the use of multiple genetic markers*. Int J Cancer, 2004. **108**(1): p. 91-6.
155. Montero-Martinez, C., et al., *Vascular Endothelial Growth Factor (VEGF): A potential biomarker for early detection of Non-small Cell Lung Cancer (NSCLC) in risk subjects*. American Journal of Respiratory and Critical Care Medicine, 2008. **177**(Abstracts issue): p. A310.
156. Yang, S.Y., et al., *Application of serum SELDI proteomic patterns in diagnosis of lung cancer*. BMC Cancer, 2005. **5**: p. 83.
157. Field, J.K., et al., *The Liverpool Lung Project research protocol*. Int J Oncol, 2005. **27**(6): p. 1633-45.
158. Pauling, L., et al., *Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography*. Proc Natl Acad Sci U S A, 1971. **68**(10): p. 2374-6.
159. Gordon, S.M., et al., *Volatile organic compounds in exhaled air from patients with lung cancer*. Clin Chem, 1985. **31**(8): p. 1278-82.
160. Phillips, M., et al., *Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study*. Lancet, 1999. **353**(9168): p. 1930-3.
161. Phillips, M., et al., *Detection of lung cancer with volatile markers in the breath*. Chest, 2003. **123**(6): p. 2115-23.
162. Poli, D., et al., *Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study*. Respir Res, 2005. **6**: p. 71.
163. Song, G., et al., *Quantitative breath analysis of volatile organic compounds of lung cancer patients*. Lung Cancer, 2010. **67**(2): p. 227-31.

164. Mazzone, P.J., et al., *Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array*. Thorax, 2007. **62**(7): p. 565-8.
165. Lam, S., C. MacAulay, and B. Palcic, *Detection and localization of early lung cancer by imaging techniques*. Chest, 1993. **103**(1 Suppl): p. 12S-14S.
166. Lam, S., et al., *Detection and localization of early lung cancer by fluorescence bronchoscopy*. Cancer, 2000. **89**(11 Suppl): p. 2468-73.
167. Loewen, G., et al., *Autofluorescence bronchoscopy for lung cancer surveillance based on risk assessment*. Thorax, 2007. **62**(4): p. 335-40.
168. Kurie, J.M., et al., *Autofluorescence bronchoscopy in the detection of squamous metaplasia and dysplasia in current and former smokers*. J Natl Cancer Inst, 1998. **90**(13): p. 991-5.
169. Spira, A., et al., *Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer*. Nat Med, 2007. **13**(3): p. 361-6.
170. Hecht, S.S., *Tobacco carcinogens, their biomarkers and tobacco-induced cancer*. Nat Rev Cancer, 2003. **3**(10): p. 733-44.
171. Pfeifer, G.P., et al., *Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers*. Oncogene, 2002. **21**(48): p. 7435-51.
172. Maneckjee, R. and J.D. Minna, *Opioids induce while nicotine suppresses apoptosis in human lung cancer cells*. Cell Growth Differ, 1994. **5**(10): p. 1033-40.
173. Heusch, W.L. and R. Maneckjee, *Signalling pathways involved in nicotine regulation of apoptosis of human lung cancer cells*. Carcinogenesis, 1998. **19**(4): p. 551-6.
174. Johnston-Early, A., et al., *Smoking abstinence and small cell lung cancer survival. An association*. Jama, 1980. **244**(19): p. 2175-9.
175. Sardari Nia, P., et al., *Prognostic value of smoking status in operated non-small cell lung cancer*. Lung Cancer, 2005. **47**(3): p. 351-9.
176. Fox, J.L., K.E. Rosenzweig, and J.S. Ostroff, *The effect of smoking status on survival following radiation therapy for non-small cell lung cancer*. Lung Cancer, 2004. **44**(3): p. 287-93.

177. Monson, J.M., et al., *Clinical radiation pneumonitis and radiographic changes after thoracic radiation therapy for lung carcinoma*. *Cancer*, 1998. **82**(5): p. 842-50.
178. Videtic, G.M., et al., *Continued cigarette smoking by patients receiving concurrent chemoradiotherapy for limited-stage small-cell lung cancer is associated with decreased survival*. *J Clin Oncol*, 2003. **21**(8): p. 1544-9.
179. Garces, Y.I., et al., *The relationship between cigarette smoking and quality of life after lung cancer diagnosis*. *Chest*, 2004. **126**(6): p. 1733-41.
180. Pickles, T., et al., *The effect of smoking on outcome following external radiation for localized prostate cancer*. *J Urol*, 2004. **171**(4): p. 1543-6.
181. Prochazka, M., et al., *Lung cancer risks in women with previous breast cancer*. *Eur J Cancer*, 2002. **38**(11): p. 1520-5.
182. Neugut, A.I., et al., *Increased risk of lung cancer after breast cancer radiation therapy in cigarette smokers*. *Cancer*, 1994. **73**(6): p. 1615-20.
183. Chang, D.W., et al., *Effect of smoking on complications in patients undergoing free TRAM flap breast reconstruction*. *Plast Reconstr Surg*, 2000. **105**(7): p. 2374-80.
184. van Leeuwen, F.E., et al., *Roles of radiotherapy and smoking in lung cancer following Hodgkin's disease*. *J Natl Cancer Inst*, 1995. **87**(20): p. 1530-7.
185. Dresler, C.M., *Is it more important to quit smoking than which chemotherapy is used?* *Lung Cancer*, 2003. **39**(2): p. 119-24.
186. Dresler, C.M. and E.R. Gritz, *Smoking, smoking cessation and the oncologist*. *Lung Cancer*, 2001. **34**(3): p. 315-23.
187. Gritz, E.R., C. Dresler, and L. Sarna, *Smoking, the missing drug interaction in clinical trials: ignoring the obvious*. *Cancer Epidemiol Biomarkers Prev*, 2005. **14**(10): p. 2287-93.
188. Martinez, E., et al., *Issues related to implementing a smoking cessation clinical trial for cancer patients*. *Cancer Causes Control*, 2009. **20**(1): p. 97-104.
189. Knudsen, N., et al., *Insights on how to quit smoking: a survey of patients with lung cancer*. *Cancer Nursing*, 1985. **8**(3): p. 145-150.
190. *Joseph Fourier*, in www.wikipedia.org.

191. *Fourier Transform*, in www.wikipedia.org.
192. Yano, K., et al., *Evaluation of glycogen level in human lung carcinoma tissues by an infrared spectroscopic method*. *Cancer Lett*, 1996. **110**(1-2): p. 29-34.
193. Yano, K., et al., *Direct measurement of human lung cancerous and noncancerous tissues by fourier transform infrared microscopy: can an infrared microscope be used as a clinical tool?* *Anal Biochem*, 2000. **287**(2): p. 218-25.
194. Wang, H.P., H.C. Wang, and Y.J. Huang, *Microscopic FTIR studies of lung cancer cells in pleural fluid*. *Sci Total Environ*, 1997. **204**(3): p. 283-7.
195. Wood, B.R., et al., *An investigation into FTIR spectroscopy as a biodiagnostic tool for cervical cancer*. *Biospectroscopy*, 1998. **2**(3): p. 143-145.
196. Mordechai, S., et al., *Possible common biomarkers from FTIR microspectroscopy of cervical cancer and melanoma*. *J Microsc*, 2004. **215**(Pt 1): p. 86-91.
197. Njoroge, E., et al., *Classification of cervical cancer cells using FTIR data*. *Conf Proc IEEE Eng Med Biol Soc*, 2006. **1**: p. 5338-41.
198. El-Tawil, S., et al., *Comparative study between Pap smear cytology and FTIR spectroscopy: a new tool for screening for cervical cancer*. *Pathology*, 2008. **40**(6): p. 600-603.
199. Lacsh, P., et al., *Characterization of colorectal adenocarcinoma sections by spatially resolved FT-IR microspectroscopy*. *Applied Spectroscopy*, 2002. **56**(1): p. 1-9.
200. Gao, T., J. Feng, and Y. Ci, *Human breast carcinomal tissues display distinctive FTIR spectra: implication for the histological characterization of carcinomas*. *Anal Cell Pathol*, 1999. **18**(2): p. 87-93.
201. Baker, M.J., et al., *Investigating FTIR based histopathology for the diagnosis of prostate cancer*. *J Biophotonics*, 2009. **2**(1-2): p. 104-13.
202. Baker, M.J., et al., *FTIR-based spectroscopic analysis in the identification of clinically aggressive prostate cancer*. *Br J Cancer*, 2008. **99**(11): p. 1859-66.

203. Wang, J.S., et al., *FT-IR spectroscopic analysis of normal and cancerous tissues of esophagus*. World J Gastroenterol, 2003. **9**(9): p. 1897-9.
204. Fujioka, N., et al., *Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy*. Cancer Detect Prev, 2004. **28**(1): p. 32-6.
205. Fukuyama, Y., et al., *A study on the differences between oral squamous cell carcinomas and normal oral mucosas measured by Fourier transform infrared spectroscopy*. Biospectroscopy, 1999. **5**(2): p. 117-126.
206. Maziak, D.E., et al., *Fourier-transform infrared spectroscopic study of characteristic molecular structure in cancer cells of esophagus: an exploratory study*. Cancer Detect Prev, 2007. **31**(3): p. 244-53.
207. Whiteman, S.C., et al., *FTIR spectroscopic analysis of sputum: preliminary findings on a potential novel diagnostic marker for COPD*. Ther Adv Respir Dis, 2008. **2**(1): p. 23-31.
208. Bosch, A., et al., *Fourier transform infrared spectroscopy for rapid identification of nonfermenting gram-negative bacteria isolated from sputum samples from cystic fibrosis patients*. J Clin Microbiol, 2008. **46**(8): p. 2535-46.
209. *Volatile Organic Compounds*, in www.wikipedia.org.
210. Jang, A.S., et al., *Volatile organic compounds contribute to airway hyperresponsiveness*. Korean J Intern Med, 2007. **22**(1): p. 8-12.
211. Jones, K.R. and N. Brautbar, *Reactive airway disease in patients with prolonged exposure to industrial solvents*. Toxicol Ind Health, 1997. **13**(6): p. 743-50.
212. Pappas, G.P., et al., *The respiratory effects of volatile organic compounds*. Int J Occup Environ Health, 2000. **6**(1): p. 1-8.
213. Norback, D., et al., *Asthmatic symptoms and volatile organic compounds, formaldehyde, and carbon dioxide in dwellings*. Occup Environ Med, 1995. **52**(6): p. 388-95.
214. Arif, A.A. and S.M. Shah, *Association between personal exposure to volatile organic compounds and asthma among US adult population*. Int Arch Occup Environ Health, 2007. **80**(8): p. 711-9.

215. Phillips, M., et al., *Volatile biomarkers of pulmonary tuberculosis in the breath*. Tuberculosis (Edinb), 2007. **87**(1): p. 44-52.
216. Brown, N.S. and R. Bicknell, *Hypoxia and oxidative stress in breast cancer. Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast cancer*. Breast Cancer Res, 2001. **3**(5): p. 323-7.
217. Phillips, M., et al., *Volatile markers of breast cancer in the breath*. Breast J, 2003. **9**(3): p. 184-91.
218. Phillips, M., et al., *Prediction of breast cancer using volatile biomarkers in the breath*. Breast Cancer Res Treat, 2006. **99**(1): p. 19-21.
219. *Reactive oxygen species*, in www.wikipedia.org.
220. *Lipid peroxidation*, in www.wikipedia.org.
221. Marnett, L.J., *Lipid peroxidation-DNA damage by malondialdehyde*. Mutat Res, 1999. **424**(1-2): p. 83-95.
222. Esme, H., et al., *High levels of oxidative stress in patients with advanced lung cancer*. Respirology, 2008. **13**(1): p. 112-6.
223. Gupta, A., et al., *Smoking intensity, oxidative stress and chemotherapy in nonsmall cell lung cancer: A correlated prognostic study*. Biosci Trends, 2009. **3**(5): p. 191-9.
224. Gupta, A., et al., *Oxidative stress in non-small cell lung cancer patients after chemotherapy: Association with treatment response*. Respirology, 2010. **15**(2): p. 349-356.
225. Mantovani, G., et al., *Cancer-related anorexia/cachexia syndrome and oxidative stress: an innovative approach beyond current treatment*. Cancer Epidemiol Biomarkers Prev, 2004. **13**(10): p. 1651-9.
226. Mantovani, G., et al., *Cytokine activity in cancer-related anorexia/cachexia: role of megestrol acetate and medroxyprogesterone acetate*. Semin Oncol, 1998. **25**(2 Suppl 6): p. 45-52.
227. Mantovani, G., et al., *Serum levels of leptin and proinflammatory cytokines in patients with advanced-stage cancer at different sites*. J Mol Med, 2000. **78**(10): p. 554-61.
228. Mantovani, G., et al., *Serum values of proinflammatory cytokines are inversely correlated with serum leptin levels in patients with advanced stage cancer at different sites*. J Mol Med, 2001. **79**(7): p. 406-14.

229. Dewys, W.D., et al., *Prognostic effect of weight loss prior to chemotherapy in cancer patients. Eastern Cooperative Oncology Group.* Am J Med, 1980. **69**(4): p. 491-7.
230. Bruera, E., *ABC of palliative care. Anorexia, cachexia, and nutrition.* Bmj, 1997. **315**(7117): p. 1219-22.
231. Buck, M. and M. Chojkier, *Muscle wasting and dedifferentiation induced by oxidative stress in a murine model of cachexia is prevented by inhibitors of nitric oxide synthesis and antioxidants.* EMBO J, 1996. **15**: p. 1753-65.
232. Tisdale, M.J., *Loss of skeletal muscle in cancer: biochemical mechanism.* Front Biosci, 2001. **6**: p. 164-74.
233. Shaw, J.H. and R.R. Wolfe, *Glucose and urea kinetics in patients with early and advanced gastrointestinal cancer: the response to glucose infusion, parenteral feeding, and surgical resection.* Surgery, 1987. **101**(2): p. 181-91.
234. Tayek, J.A., *A review of cancer cachexia and abnormal glucose metabolism in humans with cancer.* J Am Coll Nutr, 1992. **11**(4): p. 445-56.
235. Wilmore, D.W. and L.H. Aulick, *Metabolic changes in burned patients.* Surg Clin North Am, 1978. **58**(6): p. 1173-87.
236. Roth, E., et al., *Metabolic disorders in severe abdominal sepsis: glutamine deficiency in skeletal muscle.* Clin Nutr, 1982. **1**(1): p. 25-41.
237. Ray, G., et al., *Lipid peroxidation, free radical production and antioxidant status in breast cancer.* Breast Cancer Res Treat, 2000. **59**(2): p. 163-70.
238. Loft, S. and H.E. Poulsen, *Cancer risk and oxidative DNA damage in man.* J Mol Med, 1996. **74**(6): p. 297-312.
239. Kneepkens, C.M., et al., *The hydrocarbon breath test in the study of lipid peroxidation: principles and practice.* Clin Invest Med, 1992. **15**(2): p. 163-86.
240. Kneepkens, C.M., G. Lepage, and C.C. Roy, *The potential of the hydrocarbon breath test as a measure of lipid peroxidation.* Free Radic Biol Med, 1994. **17**(2): p. 127-60.

241. Hietanen, E., et al., *Diet and oxidative stress in breast, colon and prostate cancer patients: a case-control study*. Eur J Clin Nutr, 1994. **48**(8): p. 575-86.
242. Phillips, M., et al., *Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress*. J Lab Clin Med, 2000. **136**(3): p. 243-9.
243. Remmer, H., et al., *Cytochrome P-450 oxidation of alkanes originating as scission products during lipid peroxidation*. Xenobiotica, 1984. **14**(1-2): p. 207-19.
244. Benhamou, S. and C. Bonaiti-Pellie, [*Susceptibility to bronchial cancer: an example of genetic-environmental interaction*]. Ann Biol Clin (Paris), 1995. **53**(9): p. 507-13.
245. Anderson, L.M., et al., *Promotion of mouse lung tumors by bioaccumulated polychlorinated aromatic hydrocarbons*. Exp Lung Res, 1991. **17**(2): p. 455-71.
246. Hecht, S.S., *Approaches to cancer prevention based on an understanding of N-nitrosamine carcinogenesis*. Proc Soc Exp Biol Med, 1997. **216**(2): p. 181-91.
247. Watanabe, M., *Polymorphic CYP genes and disease predisposition-- what have the studies shown so far?* Toxicol Lett, 1998. **102-103**: p. 167-71.
248. Zevin, S. and N.L. Benowitz, *Drug interactions with tobacco smoking. An update*. Clin Pharmacokinet, 1999. **36**(6): p. 425-38.
249. Goto, I., et al., *Prognostic significance of germ line polymorphisms of the CYP1A1 and glutathione S-transferase genes in patients with non-small cell lung cancer*. Cancer Res, 1996. **56**(16): p. 3725-30.
250. Phillips, M., J. Greenberg, and J. Awad, *Metabolic and environmental origins of volatile organic compounds in breath*. J Clin Pathol, 1994. **47**(11): p. 1052-3.
251. Horvath, I., et al., *Exhaled biomarkers in lung cancer*. Eur Respir J, 2009. **34**(1): p. 261-75.
252. Machado, R.F., et al., *Detection of lung cancer by sensor array analyses of exhaled breath*. Am J Respir Crit Care Med, 2005. **171**(11): p. 1286-91.

253. Di Natale, C., et al., *Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors*. *Biosens Bioelectron*, 2003. **18**(10): p. 1209-18.
254. Dragonieri, S., et al., *An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD*. *Lung Cancer*, 2009. **64**(2): p. 166-70.
255. Fens, N., et al., *Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma*. *Am J Respir Crit Care Med*, 2009. **180**(11): p. 1076-82.
256. Westhoff, M., et al., *Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study*. *Thorax*, 2009. **64**(9): p. 744-8.
257. Carpagnano, G.E., et al., *3p microsatellite alterations in exhaled breath condensate from patients with non-small cell lung cancer*. *Am J Respir Crit Care Med*, 2005. **172**(6): p. 738-44.
258. Carpagnano, G.E., et al., *3p microsatellite signature in exhaled breath condensate and tumor tissue of patients with lung cancer*. *Am J Respir Crit Care Med*, 2008. **177**(3): p. 337-41.
259. Carpagnano, G.E., et al., *Cigarette smoke and increased COX-2 and survivin levels in exhaled breath condensate of lung cancer patients: How hot is the link?* *Lung Cancer*, 2009.
260. Gessner, C., et al., *Detection of p53 gene mutations in exhaled breath condensate of non-small cell lung cancer patients*. *Lung Cancer*, 2004. **43**(2): p. 215-22.
261. Dwyer, T.M., *Sampling airway surface liquid: non-volatiles in the exhaled breath condensate*. *Lung*, 2004. **182**(4): p. 241-50.
262. Effros, R.M., et al., *The promise and perils of exhaled breath condensates*. *Am J Physiol Lung Cell Mol Physiol*, 2004. **287**(6): p. L1073-80.
263. West, R., et al., *Outcome criteria in smoking cessation trials: proposal for a common standard*. *Addiction*, 2005. **100**(3): p. 299-303.
264. *Introduction to immunohistochemistry*, in <http://www.ihcworld.com>.
265. Gamboa-Dominguez, A., et al., *Epidermal growth factor receptor expression correlates with poor survival in gastric adenocarcinoma from*

- Mexican patients: a multivariate analysis using a standardized immunohistochemical detection system.* Mod Pathol, 2004. 17(5): p. 579-87.
266. Au, N.H., et al., *Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers.* J Pathol, 2004. 204(1): p. 101-9.
267. SPSS, www.spss.com, Editor.
268. *A language and environment for statistical computing.* R Foundation for Statistical Computing, 2007.
269. Savitzky, A. and M.J.E. Golay, *Smoothing and Differentiation of Data by Simplified Least Squares Procedures.* Anal Chem, 1964. 36(8): p. 1627-1639.
270. *Bonferroni Correction*, in www.wikipedia.org.
271. Holm, S., *A simple sequential rejective multiple test procedure.* Scandinavian Journal of Statistics, 1979. 6: p. 65-70.
272. Cohen, J., *A coefficient of agreement for nominal scales.* Educ Psych Meas, 1960. 20: p. 37-46.
273. Fisher-Wellman, K. and R.J. Bloomer, *Acute exercise and oxidative stress: a 30 year history.* Dyn Med, 2009. 8: p. 1.
274. Pleil, J.D. and A.B. Lindstrom, *Measurement of volatile organic compounds in exhaled breath as collected in evacuated electropolished canisters.* Journal of Chromatography, 1995. 665: p. 271-279.
275. Wright, M.D., et al. *Storage stability study of TO-14 compounds on single and multi-bed thermal desorption tubes.* in *Measurement of Toxic and Related Air Pollutants, A&WMA Conference Proceedings.* 1998.
276. Linqvist, F. and H. Bakkeren, *Stability of chlorinated hydrocarbons on Tenax.* TNO Division of Technology for Society, 1990.
277. Vandendriessche, S. and B. Griepink, *The certification of benzene, toluene and m-xylene sorbed on tenax in tubes.* Community Bureau of Reference, 1989.
278. *Electron Ionization*, in *Compendium of Chemical Terminology*, International Union of Pure and Applied Chemistry.

279. Davis, R. and M. Frearson, *Mass Spectrometry - Analytical Chemistry by Open Learning*. 1987, London: John Wiley & Sons.
280. Sparkman, D.O., *Mass spectrometry desk reference*. Global View Pub, 2000.
281. Stauffer, D.B., et al., *Probability-Based-Matching algorithm with forward searching capabilities for matching unknown mass spectra of mixtures*. *Anal Chem*, 1985. **57**: p. 1056-1060.
282. *Microsoft Excel*, in www.office.microsoft.com.
283. Phillips, M., et al., *Variation in volatile organic compounds in the breath of normal humans*. *J Chromatogr B Biomed Sci Appl*, 1999. **729**(1-2): p. 75-88.
284. *FT-IR Spectroscopy - An introduction in measurement techniques and interpretation*, in www.pharmtech.uni-erlangen.de.
285. McCrae, K.C., et al., *Analysis of neoplastic changes in mouse lung using Fourier-transform infrared microspectroscopy*. *Vibrational Spectroscopy*, 2002: p. 189-197.
286. Sidhu, G.S. and E.M. Forrester, *Glycogen-rich clara cell-type bronchio-alveolar carcinoma: light and electron microscopic study*. *Cancer*, 1977. **40**: p. 2209-2215.
287. Isao, M., et al., *A case of primary lung cancer in which almost all of the tumor contained glycogen-rich clear cytoplasm*. *Journal of the Japanese Association for Chest Surgery*, 2003. **17**(7): p. 782-7.
288. Cohenford, M.A. and B. Rigas, *Cytologically normal cells from neoplastic cervical samples display extensive structural abnormalities on IR spectroscopy: Implications for tumor biology*. *Proc Natl Acad Sci U S A*, 1998. **95**: p. 15327 - 15332.
289. Wong, P.T., et al., *Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis*. *Proc Natl Acad Sci U S A*, 1991. **88**(24): p. 10988-92.
290. Mak, V.H., et al., *Value of washings and brushings at fiberoptic bronchoscopy in the diagnosis of lung cancer*. *Thorax*, 1990. **45**(5): p. 373-6.

291. Reed, A.L., et al., *High frequency of p16 (CDKN2/MTS-1/INK4A) inactivation in head and neck squamous cell carcinoma*. *Cancer Res*, 1996. **56**(16): p. 3630-3.
292. Nobori, T., et al., *Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers*. *Nature*, 1994. **368**(6473): p. 753-6.
293. Groeger, A.M., et al., *Independent prognostic role of p16 expression in lung cancer*. *J Thorac Cardiovasc Surg*, 1999. **118**(3): p. 529-35.
294. Huang, C.-I., et al., *p16 protein expression is associated with a poor prognosis in squamous cell carcinoma of the lung*. *Br J Cancer*, 2000. **82**: p. 372-380.
295. Guzman, L.M., et al., *High frequency of p16 promoter methylation in non-small cell lung carcinomas from Chile*. *Biol Res*, 2007. **40**(3): p. 365-72.
296. Baryshnikova, E., et al., *Molecular Alterations in Spontaneous Sputum of Cancer-Free Heavy Smokers: Results from a Large Screening Program*. *Clin Cancer Res*, 2008. **14**(6): p. 1913 - 1919.
297. *p53*, in www.wikipedia.org.
298. Lee, J.S., et al., *Expression of p53 oncoprotein in non-small-cell lung cancer: a favorable prognostic factor*. *J Clin Oncol*, 1995. **13**(8): p. 1893-903.
299. Tan, D.F., et al., *Prognostic significance of expression of p53 oncoprotein in primary (stage I-IIIa) non-small cell lung cancer*. *Anticancer Res*, 2003. **23**(2C): p. 1665-72.
300. Doll, R. and A.B. Hill, *Smoking and carcinoma of the lung. Preliminary report. 1950*. *Bull World Health Organ*, 1999. **77**(1): p. 84-93.
301. Greenblatt, M.S., et al., *Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis*. *Cancer Res*, 1994. **54**(18): p. 4855-78.
302. Hernandez-Boussard, T.M. and P. Hainaut, *A specific spectrum of p53 mutations in lung cancer from smokers: review of mutations compiled in the IARC p53 database*. *Environ Health Perspect*, 1998. **106**(7): p. 385-91.

303. Denissenko, M.F., et al., *Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53*. Science, 1996. **274**(5286): p. 430-2.
304. Rodin, S.N. and A.S. Rodin, *Human lung cancer and p53: the interplay between mutagenesis and selection*. Proc Natl Acad Sci U S A, 2000. **97**(22): p. 12244-9.
305. Top, B., et al., *Comparative analysis of p53 gene mutations and protein accumulation in human non-small-cell lung cancer*. Int J Cancer, 1995. **64**(2): p. 83-91.
306. Chiba, I., et al., *Mutations in the p53 gene are frequent in primary, resected non-small cell lung cancer*. Lung Cancer Study Group. Oncogene, 1990. **5**(10): p. 1603-10.
307. Kishimoto, Y., et al., *Aberrations of the p53 tumor suppressor gene in human non-small cell carcinomas of the lung*. Cancer Res, 1992. **52**(17): p. 4799-804.
308. Mao, L., et al., *Detection of oncogene mutations in sputum precedes diagnosis of lung cancer*. Cancer Res, 1994. **54**(7): p. 1634-7.
309. Keohavong, P., et al., *Detection of K-ras and p53 mutations in sputum samples of lung cancer patients using laser capture microdissection microscope and mutation analysis*. Anal Biochem, 2004. **324**(1): p. 92-9.
310. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling*. Mol Syst Biol, 2005. **1**: p. 2005 0010.
311. Hirsch, F.R., et al., *Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis*. J Clin Oncol, 2003. **21**(20): p. 3798-807.
312. Rusch, V., et al., *Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung*. Cancer Res, 1993. **53**(10 Suppl): p. 2379-85.
313. Piyathilake, C.J., et al., *Differential expression of growth factors in squamous cell carcinoma and precancerous lesions of the lung*. Clin Cancer Res, 2002. **8**(3): p. 734-44.

314. Nakamura, H., et al., *Survival impact of epidermal growth factor receptor overexpression in patients with non-small cell lung cancer: a meta-analysis*. Thorax, 2006. **61**(2): p. 140-5.
315. Kang, J.U., et al., *Gain of the EGFR gene located on 7p12 is a frequent and early event in squamous cell carcinoma of the lung*. Cancer Genet Cytogenet, 2008. **184**(1): p. 31-7.
316. Caputi, M., et al., *Prognostic role of cyclin D1 in non small cell lung cancer: an immunohistochemical analysis*. Eur J Histochem, 1997. **41**(2): p. 133-8.
317. Caputi, M., et al., *Prognostic role of cyclin D1 in lung cancer. Relationship to proliferating cell nuclear antigen*. Am J Respir Cell Mol Biol, 1999. **20**(4): p. 746-50.
318. Keum, J.S., et al., *Cyclin D1 overexpression is an indicator of poor prognosis in resectable non-small cell lung cancer*. Br J Cancer, 1999. **81**: p. 127-132.
319. Lim, S., J.Y. Kim, and K. Park, *Expression of p63 in Lung Cancer*. The Korean Journal of Pathology, 2003. **37**: p. 100-107.
320. Pelosi, G., et al., *p63 immunoreactivity in lung cancer: yet another player in the development of squamous cell carcinomas?* J Pathol, 2002. **198**(1): p. 100-9.
321. Massion, P.P., et al., *Significance of p63 amplification and overexpression in lung cancer development and prognosis*. Cancer Res, 2003. **63**(21): p. 7113-21.
322. Massion, P.P., et al., *Role of p63 amplification and overexpression in lung cancer development*. Chest, 2004. **125**(5 Suppl): p. 102S.
323. Miekisch, W., J.K. Schubert, and G.F. Noeldge-Schomburg, *Diagnostic potential of breath analysis--focus on volatile organic compounds*. Clin Chim Acta, 2004. **347**(1-2): p. 25-39.
324. Mazzone, P.J., *Analysis of volatile organic compounds in the exhaled breath for the diagnosis of lung cancer*. J Thorac Oncol, 2008. **3**(7): p. 774-80.
325. Stone, B.G., et al., *Effect of regulating cholesterol biosynthesis on breath isoprene excretion in men*. Lipids, 1993. **28**(8): p. 705-8.

326. Mendis, S., P.A. Sobotka, and D.E. Euler, *Expired hydrocarbons in patients with acute myocardial infarction*. Free Radic Res, 1995. **23**(2): p. 117-22.
327. Foster, W.M., et al., *Breath isoprene: temporal changes in respiratory output after exposure to ozone*. J Appl Physiol, 1996. **80**(2): p. 706-10.
328. Davis, P.L., L.A. Dal Cortivo, and J. Maturo, *Endogenous isopropanol: forensic and biochemical implications*. J Anal Toxicol, 1984. **8**(5): p. 209-12.
329. Cope, K., T. Risby, and A.M. Diehl, *Increased gastrointestinal ethanol production in obese mice: implications for fatty liver disease pathogenesis*. Gastroenterology, 2000. **119**(5): p. 1340-7.
330. Simenhoff, M.L., et al., *Biochemical profile of uremic breath*. N Engl J Med, 1977. **297**(3): p. 132-5.
331. Fuchs, P., et al., *Breath gas aldehydes as biomarkers of lung cancer*. Int J Cancer, 2009.
332. Shahidi, F., *Headspace volatile aldehydes as indicators of lipid oxidation in foods*. Adv Exp Med Biol, 2001. **488**: p. 113-23.
333. Hanaka, H., et al., *Microsomal prostaglandin E synthase 1 determines tumor growth in vivo of prostate and lung cancer cells*. Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18757-62.
334. Avis, I., et al., *Inhibitors of the arachidonic acid pathway and peroxisome proliferator-activated receptor ligands have superadditive effects on lung cancer growth inhibition*. Cancer Res, 2005. **65**(10): p. 4181-90.
335. Moreno, J.J., *New aspects of the role of hydroxyeicosatetraenoic acids in cell growth and cancer development*. Biochem Pharmacol, 2009. **77**(1): p. 1-10.
336. Krotoszynski, B.K., G.M. Bruneau, and H.J. O'Neill, *Measurement of chemical inhalation exposure in urban population in the presence of endogenous effluents*. J Analyt Toxicol, 1979. **3**: p. 225-234.
337. Krotoszynski, B.K., G. Gabriel, and H.J. O'Neill, *Characterization of human expired air: a promising investigation and diagnostic technique*. J Chromatogr Sci, 1977. **15**: p. 239-244.

338. Wallace, L., et al., *Breath Measurements as Volatile Organic Compound Biomarkers*. Environ Health Perspect, 1996. **104**(Supplement 5): p. 861-869.
339. Smith, D.H., M. Achenbach, and W.J. Yeager, *Quantitative comparison of combined gas chromatographic/mass spectrometric profiles of complex mixtures*. Anal Chem, 1977. **49**: p. 1623-1632.
340. Phillips, M., *Method for the collection and assay of volatile organic compounds in breath*. Anal Biochem, 1997. **247**: p. 272-278.
341. Uhler, A.D. and G.W. Diachenko, *Volatile halocarbon compounds in process water and processed foods*. Bull Environ Contam Toxicol, 1987. **39**(4): p. 601-7.
342. Entz, R.C. and G.W. Diachenko, *Residues of volatile halocarbons in margarines*. Food Addit Contam, 1988. **5**(3): p. 267-76.
343. O'Neill, H.J., et al., *A computerized classification technique for screening for the presence of breath biomarkers in lung cancer*. Clin Chem, 1988. **34**(8): p. 1613-8.
344. Wehinger, A., et al., *Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas*. Mass Spectrometry, 2007. **265**: p. 49-59.