UNIVERSITY OF LIVERPOOL

Institute of Translational Medicine

Department of Molecular and Clinical Cancer Medicine

Georgios Nikolaidis

DNA Methylation-assisted diagnosis of Lung Cancer in bronchial washings

Thesis submitted in accordance with the requirements of

The University of Liverpool

for the degree of Doctor in Philosophy

December 2011

Supervisors: Dr. Triantafillos Liloglou and Prof. John K Field

PAGE NUMBERING AS ORIGINAL

Acknowledgements

I would like to acknowledge my supervisors, Dr Triantafillos Liloglou and Prof. John K Field for their continuous guidance and support from the very first to the last step of this work. Their valuable assistance, enabled me to develop my perceptive on the subject and complete this thesis.

I am also grateful to my colleagues, Dr. Olaide Raji, Dr. Alexandros Daskalos, Dr. Julie Bryan, Mrs Jennie Goggin, Mrs Staphanie Unsworth and all the members of the Liverpool Lung Project for the valuable support throughout these years.

Moreover, I should like to acknowledge the contribution of Prof. J.R. Gosney and Mr. Andrew Dodson regarding the immunohistochemisty experiments.

In addition, I would like to express my thankfulness to my loved ones, my wife Lydia and my parents Christos and Caterina for all the support that all these years passionately provided to help me completing this study.

Finally, I would like to express the deepest appreciation to Roy Castle Lung Cancer foundation for funding this piece of research.

Contents

ABSTRACT

Lung cancer is the most lethal malignancy worldwide and late diagnosis is a significant factor contributing this. The Liverpool Lung Project (LLP) aims to reduce lung cancer mortality through the development of a molecular-epidemiological risk assessment model which will facilitate early detection of lung cancer and thus early intervention. LLP encompasses retrospective and prospective sub-studies. DNA methylation is an epigenetic modification with key role in gene transcriptional control, embryonic development, imprinting and cancer. A large number of studies in lung cancer have revealed abnormal DNA methylation patterns involving a variety of genes.

The aim of this study was to construct and evaluate a panel of DNA-methylation biomarkers which can be applied to bronchial washings (BWs) material and assist in diagnosis of lung cancer. The discovery and validation process followed the guidelines set by the NCI- Early Detection Research Network (EDRN) and the CR-UK diagnostic biomarker roadmap. Specific objectives included (a) the discovery of promoter targets with high frequency of hypermethylation in primary lung cancer, (b) the development of a highly sensitive and specific DNA methylation assay fitting to clinical standards and (c) the validation of these targets in BWs utilising a longitudinal retrospective case-control design.

Targets from previous high throughput approaches were validated in an independent set of 48 non-small cell lung cancer samples and paired normal tissues using Pyrosequencing methylation analysis (PMA). PMA confirmed significant hypermethylation in the primary NSCLC tissue for the following promoters: RASSF1, p16, WT1, CYGB, RARβ, CDH13, DAPK, p73, TMEFF2, and TERT. In addition, immunohistochemical staining for p16 and WT1 was performed in a 20 non-small cell lung cancer samples. Quantitative methylation-specific PCR

(qMSP) assays were subsequently developed and tested for reliability and robustness for these ten candidates. These assays were used to screen 655 BWs from the Liverpool Lung Project (LLP) subjects divided into a training (194 cases and 214 Controls) and validation (139 cases and 109 controls) sets. Multifactor Dimensionality Reduction (MDR) was used to select the best subset of the markers with good discrimination. Analysis in the training BWs set demonstrated significant differences in the detected hypermethylation frequency in cases over controls for RASSF1, p16, WT1, CYGB, RARβ and TERT. The diagnostic efficiency of this panel was evaluated in the independent validation set. A logit method was used to obtain the sensitivity and specificity of the six markers. LogicF analysis demonstrated that the top five predictors were WT1, cytology, RASSF1, TERT and p16. The overall performance the latter panel demonstrated no diagnostic bias in different groups of gender, age or smoking status. While cytology alone provides a 49.5% sensitivity, 99.5% specificity, the addition of the four methylation markers provided 76.2% sensitivity, 92.3% specificity (AUC=0.89).

Although the diagnostic efficiency of this panel must be improved by incorporating additional promoters, our findings clearly demonstrate the impact of DNA methylation-based assays in the diagnosis of cytologically occult lung neoplasms.

Abbreviation Table

5-mC	5-Methylcytosine
AUC	Area Under the Curve
ASP	Asparigine
BW	Bronchial Washings
ChIP	Chromatin Immunoprecipitation
CpG	Cytosine – Guanine dinucleotide
CIMP	CpG Island Methylator Phenotype
COPD	Chronic Obstructive Pulmonary Disease
CE	Conformité Européenne (European Conformity)
CRUK	Cancer Research United Kingdom
CT	X-Ray Computed Tomography
Ct	Cycle threshold
CYGB	Cytoglobin
DAB	Diaminobenzidine
DNMT	DNA Methyltransferases
EDRN	Early Detection Research Network
GWAS	Genome-Wide Association Study
IASLC	International Association For The Study Of Lung Cancer
IHC	Immunohistochemistry
JPL	Jet Propulsion Laboratory
LLP	Liverpool Lung Project
MEP	Methylation Enrichment Pyrosequencing
miRNA	Micro - RNA
MRI	Magnetic Resonance Imaging
MtI	Methylation Index
MSP	Methylation Specific PCR
MSRE	Methylation-Specific Restriction Enzyme

NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute
NSCLC	Non Small Cell Lung Carcinoma
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
PET-CT	Positron Emission Tomography - Computed Tomography
PMA	Pyrosequencing-based Methylation Analysis
PWWP	Pro-Trp-Trp-Pro motif
REMARK	REporting recommendations for tumour MARKer prognostic studies
SAM	S-Adenosylmethionine
SCLC	Small Cell Lung Carcinoma
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedure
STARD	STAndards for the Reporting of Diagnostic accuracy studies
SSCP	Single Strand Conformation Polymorphism
Taq	Thermus Aquaticus
TKI	Tyrosine Kinase Inhibitor
TNM	Tumour-Nodel-Metastasis
tRNA	Transfer Ribonucleic Acid
TSG	Tumour Suppressor Gene
UADT	Upper Aerodigestive Tract
UKLS	United Kingdom Lung cancer Screening
WGA	Whole Genome Amplification
WHO	World Health Organisation

Chapter 1. Lung Cancer

Lung Cancer is the most lethal type of human neoplasia (Franklin, 2000a). It is a typical environmental cancer where carcinogen exposure leads to accumulation of genetic and epigenetic damage (Figure 1.1) eventually leading to increased proliferation and malignant transformation (Miller, 2005). This is in agreement with the fact that lung cancer is an age-related disease and tumours are clinically detectable with higher frequency in older individuals.

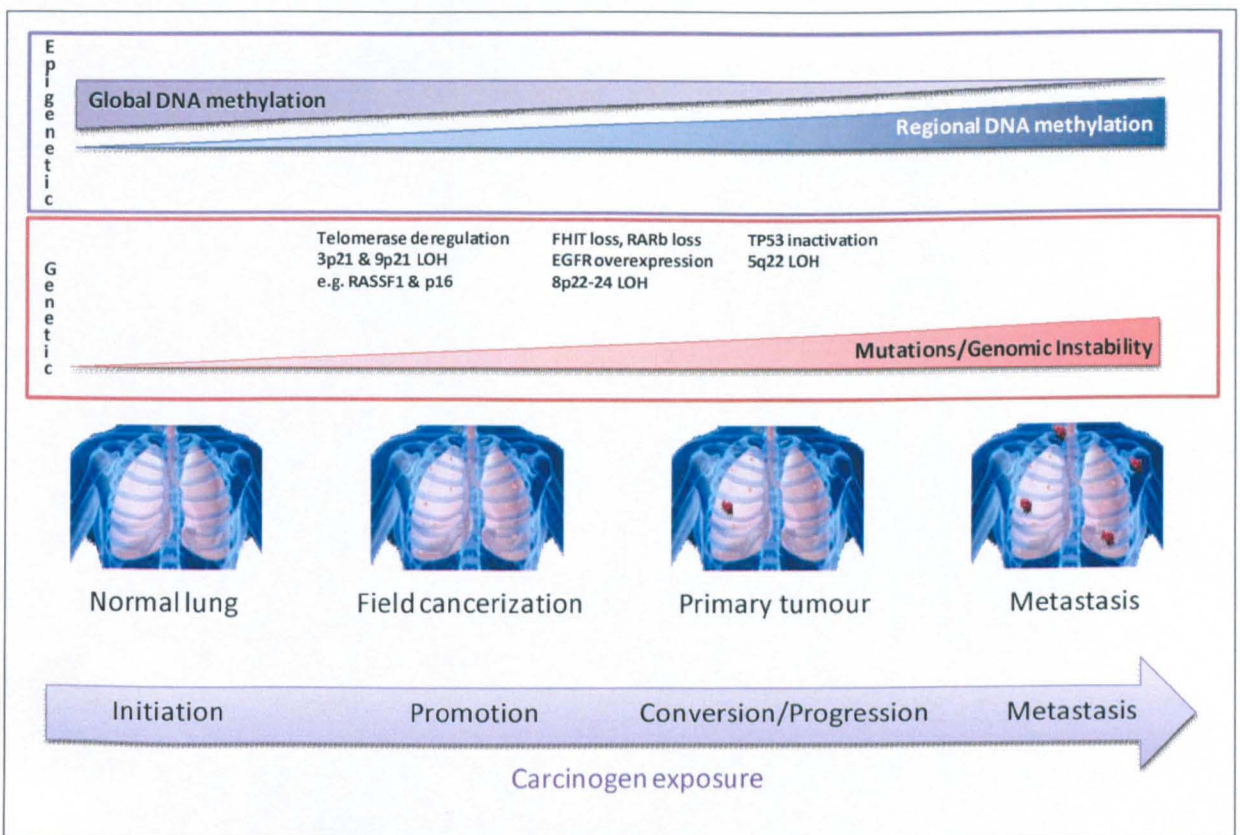


Figure 1.1. Steps in lung cancer development. Environmental carcinogens induce the formation of DNA adducts which subsequently lead to mutations and epimutations in the DNA. Initiation events may result in benign clonal outgrowths that, by acquiring further genetic and epigenetic changes, convert into localised malignant tumours. The latter have the ability to invade neighbouring tissue and metastasise via the blood and lymphatic systems spreading throughout the body.

1.1. Incidence and mortality of lung cancer

Lung cancer is one of the most commonly occurring malignancies worldwide (Broker & Giaccone, 2002) and the second most common malignant neoplasm in the UK following breast cancer. According to the CRUK CancerStats (http://info.cancerresearchuk.org/cancerstats), lung cancer is the most common cancer in the world with 1.3 million new cases diagnosed every year (2008 data). In UK, it was the most frequently occurring cancer till the late 1990s when breast cancer took the leading position. It is also well established that lung malignancies are rarely met in people younger than 40 and rise in the over 60 years old individuals with greater frequency (Figure 1.1.1.).

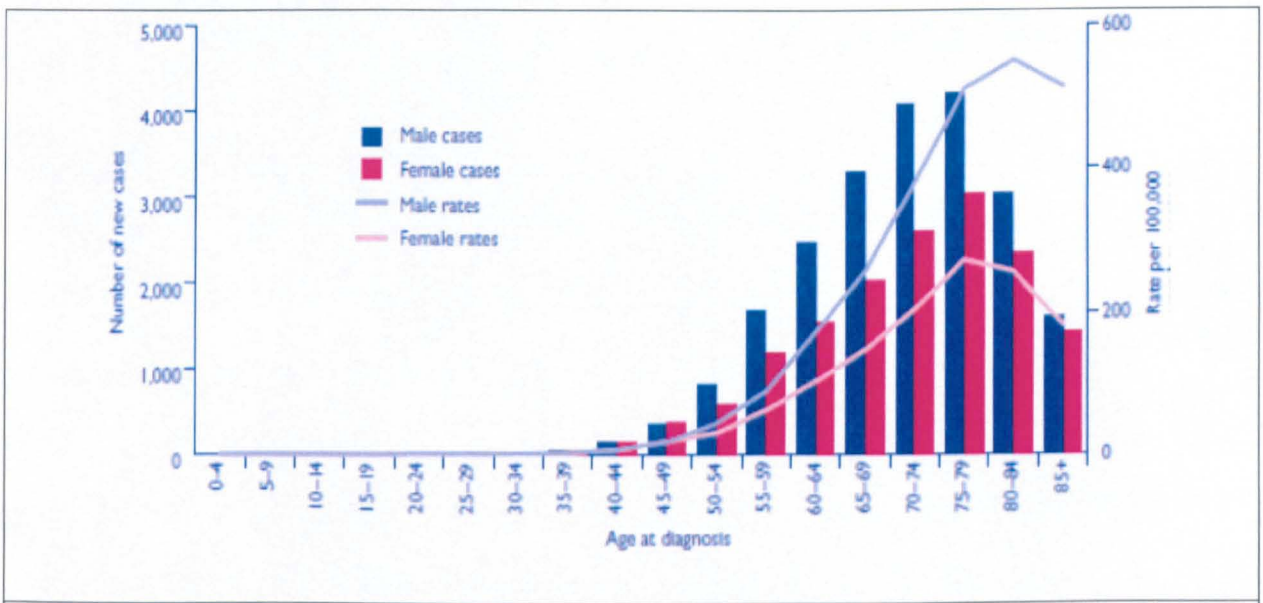
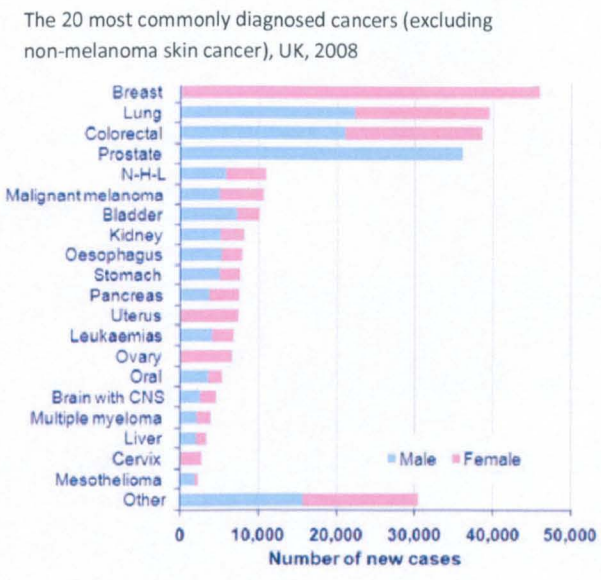


Figure 1.1.1. Illustrative representation of incidence rates for lung cancer by age and sex.

Between 1978 and 2007, incidence rates for cancer in Great Britain increased by 25%, with a 14% increase in men and a 32% increase in women. However, in the last decade incidence rates have remained fairly constant.

There are more than 200 different types of cancer, but four of them - breast, lung, large bowel (colorectal) and prostate - account for over half (54%) of all new cases. Breast cancer is the most common cancer in the UK even though it is rare in men [CancerStats].



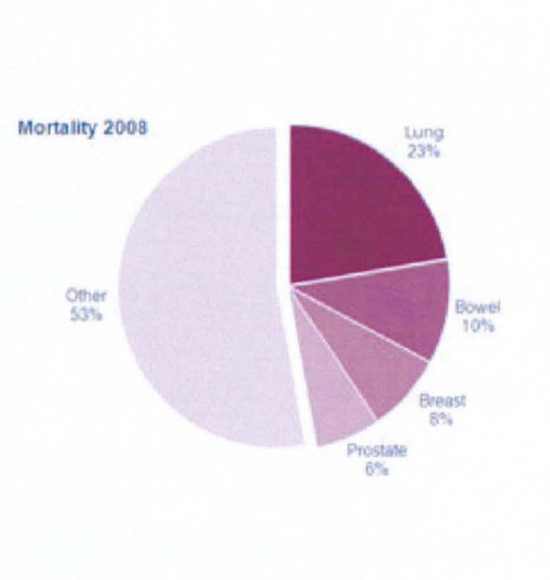


Figure 1.1.2. Incidence (left) and mortality (right) of lung cancer in the UK in comparison to other cancers. Lung cancer is the second most frequent and by far the most lethal neoplasia in the UK. Diagrams adapted from Cancer Research UK, Cancer Statistics.

Lung cancer is causing more deaths than any other neoplasia both in the USA (Jemal et al, 2010) and the UK (Jack et al, 2011) and late detection is a major contributor to this high mortality rates (Mulshine & van Klaveren, 2011). This is why, although accounting for 14% of all cancer diagnoses, it is responsible for 23% of cancer deaths (Figure 1.1.2.). These figures indicate the very poor prognosis of this tumour type and the lack of progress in treatment. Therapy is hampered by the tendency for lung cancer to be diagnosed at a late stage; hence the need to develop biomarkers for early detection is absolutely critical.

Furthermore, within countries, differences in incidence according to the ethnicity, geographical location and low and middle income are often observed mainly because there is a great variability on tobacco consumption and life-style (Brennan et al, 2011).

1.2. Histological Classification

The most widely accepted lung tumour classification schema is that of the World Health Organisation (WHO). This classification system is the result of years of collaborative effort by a panel of pathologists with expertise in lung cancer (Franklin, 2000b; Brambilla et al, 2001).

Lung cancer is pathologically divided into small cell lung cancer (SCLC), which accounts for approximately 20% of all cases, and non-small cell lung cancer (NSCLC) that represents approximately 80%. The most frequent types of NSCLC are squamous cell carcinoma, adenocarcinoma and large cell carcinoma (Figure 1.2.1), although the latter is recently considered to be closer to SCLC because of its neuroendocrine features (Giaccone, 2002).

Squamous cell carcinomas are considered to arise through a series of morphological changes beginning with basal cell hyperplasia, leading to squamous cell metaplasia – characterised by increasing severity of cellular atypia (dysplasia) – to carcinoma in situ and finally squamous cell carcinoma. On the other hand, adenocarcinoma appears to arise from atypical adenomatous hyperplastic cells within the pulmonary parenchyma (Belinsky et al, 2004). Squamous cell lung carcinoma usually originates near a central bronchus where adenocarcinoma is usually originates in peripheral lung tissue. Small cell lung cancer develops from a neuroendocrine cell within the lungs. It tends to arise in the larger airways (primary and secondary bronchi) and grows rapidly into a considerably large tumour mass (Collins et al, 2007).

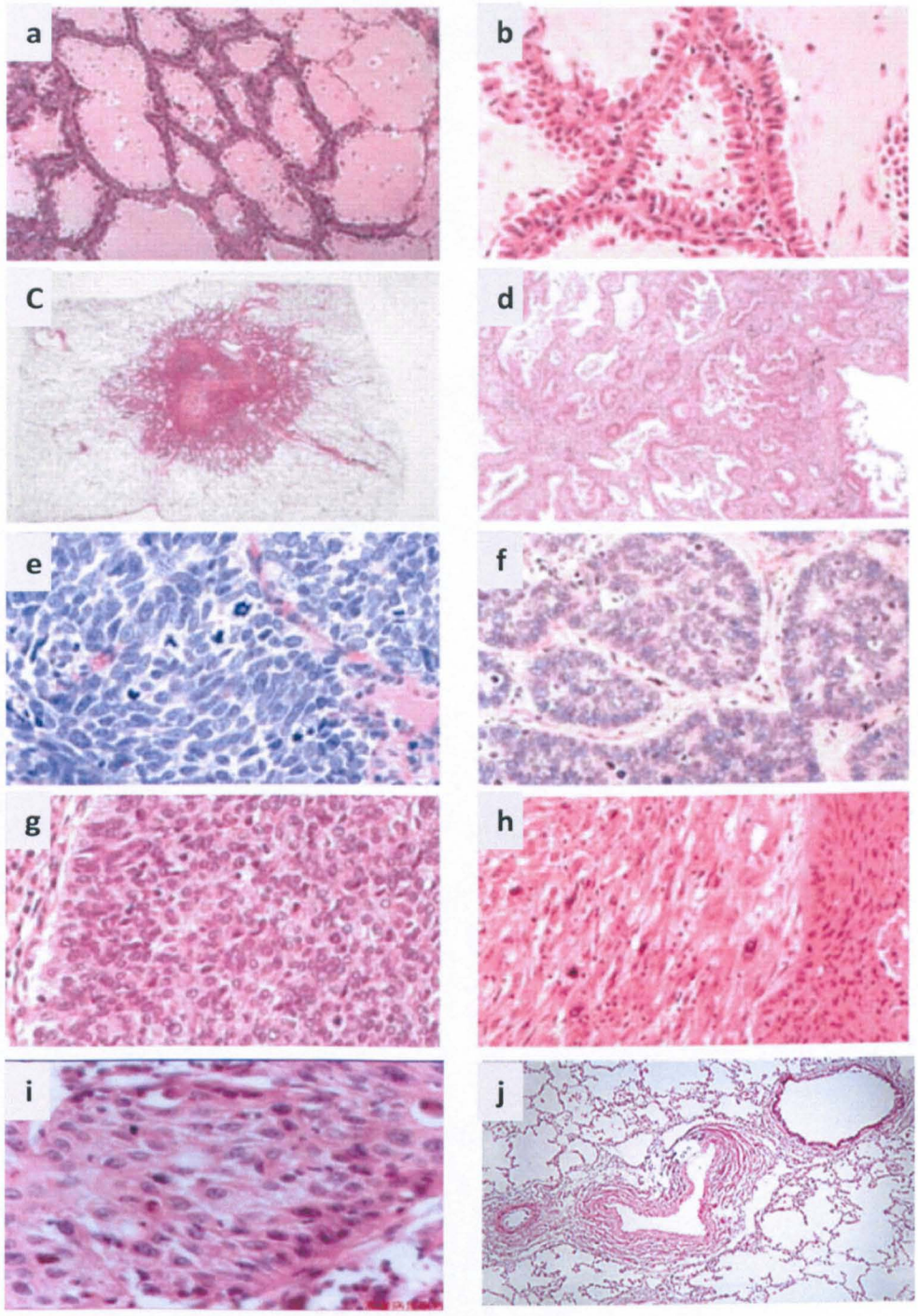


Figure 1.2.1. Microscopic images of normal lung and the major types of lung cancer. a) Bronchioloalveolar carcinoma, nonmucinous b) Bronchioloalveolar carcinoma c) Adenocarcinoma with mixed subtypes in low magnification d) Same adenocarcinoma at higher magnification e) Small cell carcinoma f) Large cell neuroendocrine carcinoma g) Basaloid carcinoma h) Pleomorphic carcinoma i) squamous cell carcinoma j) normal lung. Adapted from (Brambilla et al, 2001).

Lung cancer may arise from the major bronchi (central tumours) or small bronchi, bronchioles, or alveoli (peripheral tumours) of the distant airway of the lung. Squamous cell carcinomas usually arise centrally, whereas adenocarcinomas and large cell carcinomas usually arise peripherally.

Surgery remains the only curative treatment, however, only 15-20% of tumours can be radically resected. Factors that turn tumours to operable or non-operable are: the individual's age, the presence of other systemic diseases, and the local extension of the tumour and lung function. Depending on the size and spread of the tumour, surgery can involve a wedge resection, segmental resection, lobectomy (surgical excision of a lobe) or a pneumonectomy (surgical removal of a lung) (Figure 1.2.2) (Gridelli et al, 2003). Pneumonectomy was considered the only appropriate treatment of primary lung cancer. However, because of the unacceptably high mortality rate of around 40% associated with pneumonectomy, lobectomy evolved as the treatment of choice for resectable peripheral cancers (Narsule et al, 2011). Currently, lobectomy is the preferred treatment for early stage, non-small cell lung cancer primarily because of the increased local recurrence rate that has been reported with sublobar resection (Kates et al, 2011). Tumour size, indolent behaviour, pulmonary function and brachytherapy are the factors that need to be carefully assessed prior to lobar or sublobar resection (Narsule et al, 2011).

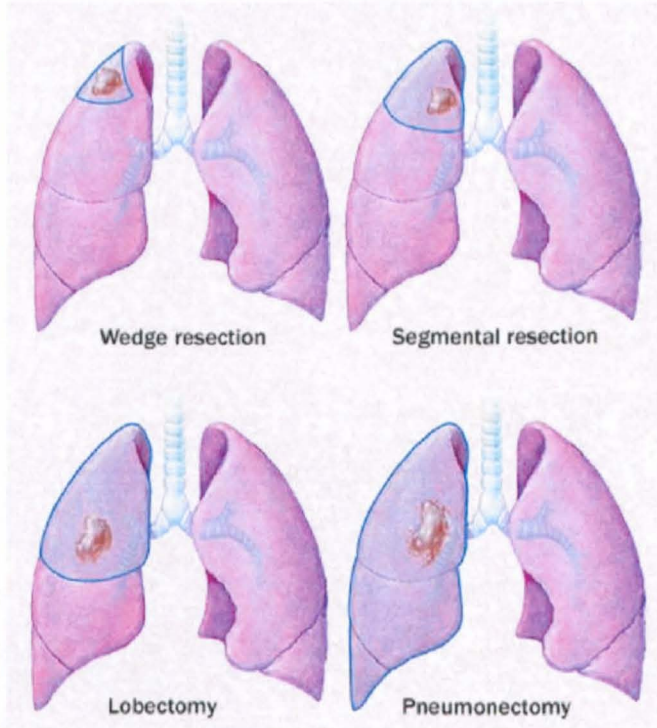


Figure 1.2.2. Schematic representation of the possible surgical operations that can be employed in lung cancer condition.

Image adapted from http://www.mayoclinic.com/health/medical/IM04119

Moreover, effective chemotherapy agents have been developed that possess favourable toxicity profiles and that are directed towards important biological pathways such as those necessary for cell proliferation, differentiation and apoptosis (Hirsch et al, 2002) Early disease treatments facilitate chemoprevention which may arrest tumour development. Current knowledge of lung cancer staging and progression has led to possible modes of intervention (Mulshine & Hirsch, 2003). Combined treatment involving radiation and chemotherapy improves the outcome for patients with locally advanced tumours and is associated with a curative potential. In the case of advanced disease, chemotherapy prolongs and improves the quality of patient life (Ramalingam et al, 2004). Radiotherapy is often used in conjunction with chemotherapy and in some cases surgery in the treatment of

locally advanced disease is necessary. It is also used as a palliative treatment in patients with advanced disease to control symptoms and in some patients, to prolong life (Gridelli et al, 2003; Macbeth & Stephens, 2004). Finally, symptomatic treatment includes the use of pain relief medication, analgesics and sedatives. There is strong evidence for differential efficacy by histology for various cytotoxic agents like cisplatin and carboplatin (Ardizzoni et al, 2007) as well as by subtyping of NSCLC when considering the use of molecular targeted agents like tyrosine kinase inhibitors (TKIs) (Gazdar, 2009).

Table 1.2.1 Molecular therapies and their potential biomarkers. Adapted from (Kerr, 2011).

Agents	Target	Potential biomarkers
Erlotinib, Gefitinib	EGFR internal TK	EGFR mutation EGFR gene copy number KRAS mutation MET amplification Not EGFR IHC
Cetuximab	EGFR extracelluar domain	EGFR IHC
Bevacizumab	VEGFR	None known
Crizotinib	Chimeric ALK TK	ALK1 protein EML4–ALK gene translocation (FISH)
IGFR1 inhibitors (CP751,871 etc.)	Insulin growth factor receptor TK	IGFR1 protein
Multi-targeted TKIs	EGFR, VEGFR, HER2 etc.	EGFR mutation

Table 1.2.2. Cytotoxic therapies and their potential biomarkers. Adapted from (Kerr, 2011).

Drug agent/class	Action	Potential biomarker
Platinum-based	Disrupt DNA synthesis Induce apoptosis	ERCC1 Serpin B3, p27, rap80, Abraxas BRCA1
Taxanes	Stabilize microtubules Disrupt cell division Induce apoptosis	BRCA1 Beta tubulin III
Vinca alkaloids	Prevent microtubule assembly Disrupt cell division Induce apoptosis	BRCA1 Beta tubulin III
Pemetrexed	Inhibits thymidylate synthase Disrupts DNA synthesis	Thymidylate synthase (TS)
Gemcitabine	Pyrimidine antimetabolite Disrupts DNA synthesis	Ribonucleotide reductase subunit M1 (RRM1)

For small cell lung cancer, non small cell lung cancer and bronchopulmonary carcinoids an international staging system is used, the TNM staging (Kerr, 2011). This is carefully conducted by the International Association for the Study of Lung Cancer (IASLC) Prospective Lung Cancer Staging Project (Kerr, 2011). The node and metastasis classification of malignant tumours is periodically revised. Its seventh edition includes the updated classification for lung cancer, based on the analyses of the International Association for the Study of Lung Cancer international database. It is the largest validation ever carried out to date: 100,869 patients registered in 46 databases from 20 countries (Rami-Porta et al, 2009).

T descriminator : It is determined by the size of the primary tumour as measured in the long-axis diameter; extend of invasion of the primary tumour and presence or absence of satellite nodules. Tumours up to 3cm are T1, tumours up to 5cm are T2, up to 7 cm are T3

and the larger ones are T4. Contrast- enhanced CT or magnetic resonance imaging (MRI) are used to describe this factor (Travis et al, 2008).

N descriminator : it describes the presence or absence of metastatic involvement of the lymph nodes throughout the thorax. Patients without nodal metastatic disease are designated as N0. Patients with N1 disease are defined as having metastatic involvement of lymph nodes in the ipsilateral peripheral or hilar zones (Vallieres et al, 2009). The N2 designation signifies metastatic extension to lymph nodes in the ipsilateral mediastinal (upper, aorticopulmonary, lower) or subcarinal lymph node zones. The N3 nodal designation includes metastatic involvement of any nodes in the supraclavicular lymph node zone or nodes in contralateral mediastinal, hilar–interlobar, or peripheral zones Again, contrast-enhanced CT is used; however positron emission tomography (PET) or PET-CT is more accurate and specific (Travis et al, 2008; Carvalho et al, 2009).

M descriptor is the presence or absence of metastases within or outside the thorax. M0 represents the absence of an extranodal metastatic disease. M1a are metastatic nodules to the opposite lung and M1b are metastatic nodules outside the thorax. Suitable imaging procedures for detecting distant metastases are contrast-enhanced cranial CT or MRI, bone scintigraphy, ultrasonography, CT or MRI of the liver and adrenals and PET, PET-CT (Travis et al, 2008).

Table 1.2.3. Descriptors proposed T, N and M categories and Stage Groupings as described by Goldstraw et al, 2007.

Sixth Edition T/M Descriptor	7th Edition T/M	N0	N1	N2	N3
T1 (less than or equal to 2 cm)	T1a	IA	IIA	IIIA	IIIB
T1 (>2-3 cm)	T1b	IA	IIA	IIIA	IIIB
T2 (less than or equal to 5 cm)	T2a	IB	IIA	IIIA	IIIB
T2 (>5-7 cm)	T2b	IIA	IIB	IIIA	IIIB
T2 (>7 cm)	T3	IIB	IIIA	IIIA	IIIB
T3 invasion		IIB	IIIA	IIIA	IIIB
T4 (same lobe nodules)		IIB	IIIA	IIIA	IIIB
T4 (extension)	T4	IIIA	IIIA	IIIB	IIIB
M1 (ipsilateral lung)		IIIA	IIIA	IIIB	IIIB
T4 (pleural effusion)	M1a	IV	IV	IV	IV
M1 (contralateral lung)		IV	IV	IV	IV
M1 (distant)	M1b	IV	IV	IV	IV

Cells in bold indicate a change from the sixth edition for a particular TNM category

1.3. Aetiology and Epidemiology

It has been proposed that cancer is predominantly an environmental disease, as only a small proportion of cancers follow a Mendelian pattern of inheritance (Trichopoulos et al, 1996). Furthermore, the incidence of cancer changes when a population is exposed to different cultural and lifestyle conditions.

Increased evidence suggests that lung cancer, like other solid tumours, is the result of a multistep process, rather than sudden transformation of previously normal epithelium (Franklin & McCubrey, 2008). Field carcinogenesis is supported by two contradicting scenarios. On one hand the biological explanation based on high exposure of respiratory epithelium to multiple carcinogens that lead to different mutational progression at dispersed sites in the airways. On the other hand, the hypothesis of a single, mutant, progenitor epithelial clone that may expand over time to populate widespread areas of the respiratory tract stands for a different scientific approach (Sozzi et al, 1995; Sidransky & Messing, 1992).

Humans are constantly exposed to chemical carcinogens in their everyday lives but only a proportion develops lung cancer. Increased risk of environmentally induced cancer is associated with exposures and host factors (including carcinogen metabolism). Since many carcinogenic compounds require metabolic activation to enable them to react with cellular macromolecules, individual features of carcinogen metabolism may play an essential role in the development of environmental cancer (Hietanen, 1997). The idea that environmental factors could cause cancer goes back to Percival Pott in 1775. He was the first to discover environmental carcinogens and occupational cancer. While exposure to tobacco smoke is widely accepted as the major etiologic factor in lung cancer, differences in individual susceptibility have been inferred from the observation that only a minority of cigarette smokers has diagnoses of lung cancer. Variation in the ability to metabolize xenobiotics has been considered as a possible explanation for this phenomenon.

The past four decades of epidemiological research have yielded valuable information on the risks of populations to environmental exposures such as tobacco, asbestos, and dietary

components (Feigelson et al, 1996). For such effort different practical approaches to molecular epidemiology of human cancer took place (Sugimura et al, 1991). Franz Hermann Muller first reported in 1939 a smoking link with lung cancer. Later on, scientists have focused their studies in epidemiological studies and by 1950 three epidemiological studies have been reported to give power to the statement (Levin et al, 1950; Wynder & Graham, 1950; Doll & Hill, 1950). A prospective study of 40,000 male doctors that compared cigarette consumption with lung cancer deaths carried out by Doll and Hill was reported in 1964, resulting in a strong dose response relationship. Further studies gave strength to the statement that *all major histological types of lung cancer are associated with smoking* (Sun et al, 2007). Stronger association of smoking with SCLC and SCC than adenocarcinoma is a reality. Beyond any doubt, tobacco smoke exposure is by far the greatest risk for conducting the disease as tobacco itself contains more than 60 carcinogens (Hecht, 2003). However, according to statistical reviews, only a small fraction (10-20%) of the life-time smokers develop lung cancer suggesting that genetic factors as well as other environmental factors may affect an individual's susceptibility to lung cancer (Hall et al, 1997).

The hazardous nature of asbestos has been recognised for over a hundred years yet (Budgen, 2004). The association of asbestos exposure and cancer has been studied by many investigators and a strong correlation has been established (Kazan-Allen, 2005). Moreover, radon and its daughter decay products are thought to be the cause of 5% of lung cancer in the UK (Bowie & Bowie, 1991). However, results from case – controls studies investigating the association between radon exposure and lung cancer risk remain controversial (Darby et al, 2005; Krewski et al, 2006). In addition, lung cancer incidence in female non-smokers has led to studies to evaluate possible role of polycyclic aromatic hydrocarbons. Cooking fume

exposure during frying have been evaluated in various studies and showed an significant association with an increased risk of lung cancer (Yu et al, 2006).

1.4. The Liverpool Lung Project

The Liverpool Lung Project (LLP) was initiated in 1998 and it is funded by the Roy Castle Lung Cancer Foundation. Its aim is to produce a molecular-epidemiological risk assessment model in order to facilitate early detection of lung cancer (Field et al, 2005b). The program encompasses two sub-studies; a prospective cohort study of 7,500 individuals and a retrospective Case/Control with 1500 cases and 3000 controls. Detailed epidemiological and molecular profiling is utilised to identify risk factors and early biomarkers. Genetic predisposition is a developing integral component of the risk assessment model and studies on specific single nucleotide polymorphisms (SNPs) have been undertaken (Liloglou et al, 2002; Gorlov et al, 2007). The latter has been already constructed based on the epidemiological profiling of cases and controls (Cassidy et al, 2008). The group has participated in a multi centre European study based on a genome-wide approach with a total of 6,158 cases and 9,732 controls (Hung et al, 2008). In the overall analysis, data from 14 studies for 18 sequence variants in 12 DNA repair genes were incorporated. Four of the variants were found to be weakly associated with lung cancer risk. Aiming on marker development another study based on a genome wide approach took place with the group's participation. 3,259 cases and 4,159 in the training set and 2,899 cases and 5,573 controls in the validation set provided two uncorrelated disease markers. The susceptibility regions found to have possible role in lung cancer aetiology, contain TERT and CLPTM1L (McKay et al, 2008). Furthermore, based on previous findings in the correlation of the genetic variants

in 15q25 (Hung et al, 2008) a genotyping study were conducted to confirm the strong association between the 15q variants and lung cancer. An independent association with smoking quantity as well as an association with upper aerodigestive tract (UADT) cancers (Lips et al, 2010). Moreover, as genome-wide studies proved to be successful in identifiying genetic variation involved in susceptibility to etiologically complex disease. GWAS was carried out, based on the Illumina HumanHap300 beadchips. 19 top-ranked variants were further analysed to present evidence for significant association (McKay et al, 2011) Simultaneously, there is a continuous effort for developing biomarkers for early diagnosis utilising sputum and bronchial lavage samples (Field et al, 1999; Liloglou et al, 2000; Liloglou et al, 2001; Field et al, 2005a; Ehrich et al, 2005; Ehrich et al, 2006).

Chapter 2. Molecular alterations in NSCLC

In the last three decades, scientists have been investigating the molecular origins of lung cancer and the impact of genetic and epigenetic abnormalities. Both phenomena appear to play key role in the development of this disease.

2.1. Genetic Alterations

One of the key molecular elements of lung tumours is genetic instability (Field et al, 1996); this is the loss of the ability of cells to pass their genetic footprint into the next generation in the classical hemi-conservative manner. Mounting evidence suggests that aneuploid cells are genetically unstable and, ultimately, lead to cancer development (Erenpreisa & Cragg, 2007). Lung cancer development is associated with DNA replication stress from the early premalignant stages. It has been shown that allelic imbalance occurs preferentially at loci that are prone to DNA double-strand break formation when DNA replication is compromised. The DNA replication stress progressively leads to DNA double-strand breaks, genomic instability and in the long term, selective pressure for p53 mutations (Gorgoulis et al, 2005).

Cancer cells represent one class of cells that often has altered ploidy. For epithelial tumours, a near-triploid DNA content is common and a near-tetraploid DNA content also occurs (Nigg, 2002). However, cancer cells are usually not polyploid but rather aneuploid, which means that, in addition to alterations in the total chromosome number, they contain a variety of other gross chromosomal rearrangements: amplifications, deletions and nonreciprocal

translocations (Matzke et al, 2003). It has been shown that the most frequent chromosomal alterations in NSCLC occur at 3p, 9p5qand 17p (Liloglou et al, 1997; Liloglou et al, 2000; Xinarianos et al, 2000; Mariatos et al, 2000; Sikking et al, 2003). These loci host a large number of genes with high importance in cancer development such as MLH1, FHIT (3p), p16, p15 (9p), p53 (17p) and APC (5q). In addition, gene-specific alterations have been reported in NSCLC such as mutations in many genes including p53 (Liloglou et al, 1997) K-ras (Li et al, 2003; Clayton et al, 2000), EGFR (Gazdar & Minna, 2008; Mounawar et al, 2007) and LKB1 (Koivunen et al, 2008). In squamous cell carcinomas, the sequential molecular abnormalities appear to follow a certain model. Mutations arise progressively at multiple 3p chromosome sites and then 9p21 (p16) as the earliest detectable changes. Later changes occur at 8p21-23, 13q14 (RB), and 17p13 (TP53) (Wistuba, 2005). In contrast, in lung adenocarcinomas, the most important early finding is the presence of KRAS mutations in as many as 39% of hyperplasias (Westra, 2000). Moreover, EGFR mutations have been reported and significantly associated with adenocarcinoma histology, non- or light- smokers status, females, and east Asian ethnic groups (Shigematsu et al, 2005).

The molecular aberrations that are frequent in NSCLC are detectable in sputum, bronchial lavage and plasma and thus serve as potential early detection markers (Liloglou et al, 2001; Baryshnikova et al, 2008; Li et al, 2007; Wang et al, 2006; Keohavong et al, 2005). Moreover, there is increasing evidence for use in patient stratification for therapeutic purposes such as prediction to response in particular regimes (Kondo et al, 2005; van Zandwijk et al, 2007; Tsao et al, 2006).

2.2. Epigenetic Alterations in NSCLC

Epigenetic modifications are heritable changes of the genome that alter gene expression without involving alterations in the primary structure (nucleotide sequence) of DNA (Bird, 1996). Such changes are DNA methylation, histone modifications, chromatin remodelling and RNA interference (Esteller et al, 2001; Tong, 2006). Epigenetic abnormalities are among the most frequent changes in human cancer (Baylin & Ohm, 2006; Esteller, 2007b; Jones & Baylin, 2007). In fact, it is considered that all human cancers contain multiple epigenetic abnormalities and this emphasizes the potential value of cancer epigenetic traits for clinical exploitation.

2.2.1. DNA Methylation

DNA methylation is currently the most intensively studied epigenetic modification and competes for a position amongst the most promising of the DNA mechanisms to provide biomarkers for cancer management, including risk modeling, early detection, prediction of relapse, therapeutic stratification and monitoring.

DNA methylation in eukaryotes involves the addition of a methyl group to the carbon 5 position of the cytosine ring (Singal & Ginder, 1999). 5-methylcytosine (5-mC) was first identified in 1948. This covalent modification is occurs at cytosines followed by guanine residues (CG or CpG dinucleotides) in human DNA. CpG dinucleotides are generally under-represented in mammalian genome and are often found as discrete clusters referred to as *CpG islands* (Magewu & Jones, 1994). DNA methylation tend to maintain patterns established by the coordinated action of the DNA methyltransferases (DNMTs) and

associated factors, such as polycomb proteins, in the presence of S-adenosylmethionine (SAM) that serves as methyl donor (Robertson, 2005). A family of DNA methyltsansferases (DNMTs) catalyse the addition of a methyl group from S-adenosyl-methionine to cytosine residues (Strathdee & Brown, 2002). There are currently five known DNMTs, namely DNMT1, DNMT2, DNMT3a, DNMT3b and DNMT3L (Rountree, 2001). Each one appears to play a distinct role in the cell (Bestor, 2000). Only DNMT1, DNMT3a and DNMT3b interplay to produce the global cytosine methylation pattern. These independently encoded proteins are classified as maintenance enzyme (DNMT1) and *de novo* enzymes (DNMT3a and DNMT3b) (Kulis & Esteller). The *de novo* DNMTs can effectively methylate C to 5-mC in unmethylated DNA, whereas maintenance DNMT preferentially attaches a methyl group to hemimethylated DNA during replication (Margot et al, 2003). Both isoforms have similar domain arrangements containing a variable region at N terminus, followed by the PWWP domain that may involved in nonspecific DNA binding (Qiu et al, 2002). Moreover, as much research took place for its role in embryogenesis and germ cell development, major findings have been proposed in relation with the *de novo* DNMTs. Inactivation of each of these genes lead in severe phenotypes, at the time at DNMT3a seems to be responsible for the methylation of sequences critical for late embryonic development whereas DNMT3b is more important for early development (Okano et al, 1999). The last member of the DNMT3 family known as DNMT3L is not thought to function as cytosine methyltransferase, however was shown to stimulate *de novo* methylation by DNMT3a and DNMT3b and to mediate transcriptional repression through interaction with histone deacetylase 1 (Chedin et al, 2002; Deplus et al, 2002). DNMT1 is connected with maintenance of parental DNA methylation patterns in newly synthesized DNA daughter strands (Turek-Plewa & Jagodzinski, 2005). Also, DNMT1 exhibits a preference for hemimethylated DNA and it

possesses a domain targeting to replication foci (Hemann et al, 2004). Finally, DNMT2 has very weak methylation activity (Margot et al, 2003). The function of DNMT2 is highly conserved, and human DNMT2 protein restored methylation in vitro to tRNA(Asp) (Goll et al, 2006). As the cancer methylome is highly disrupted, this epigenetic effect serves as an excellent target for anti-cancer therapies. Several small synthetic and natural molecules, like 5'-azacitidine and 5'-azadeoxycitidine, are able to reverse the DNA hypermethylation through inhibition of DNA methyltransferase (DNMT) (Ren et al, 2011).

Genomic DNA methylation patterns are not randomly distributed. Regions with repetitive or parasitic DNA are found to be hypermethylated in normal mammalian DNA, while CpG islands, often associated with regulatory sites of genes, are hypomethylated (Yoder et al, 1997). DNA methylation is of particular importance in key aspects of mammalian life such as embryonic development, X-chromosome inactivation and imprinting. During embryonic development, genome-wide demethylation after fertilisation is followed by *de novo* methylation upon embryo implantation (Reik et al, 2001). X-chromosome inactivation in females is strongly associated with widespread methylation of CpG islands (Heard et al, 1997). Genomic imprinting, a process involving inactivation of the paternally or maternally inherited allele, also heavily relies on DNA methylation (Bartolomei & Tilghman, 1997).

In the last decade, particular attention has been given to the critical role of abnormal DNA methylation patterns in human cancer. The original observation that inactivation of genes in cancer cell lines are frequently associated with methylation of CpG islands (Antequera et al, 1990) was followed by a massive expansion in this field. To date, the terms hypo- and hyper-methylation are well established (Ehrlich, 2002) to describe increase or loss of the 5-methyl cytosine content of particular DNA stretches. Tumours are characterised by a genome-wide

hypomethylation while, at the same time, certain promoters gain specific patterns of hypermethylation (Esteller, 2007b). Between these two phenomena, hypermethylation at CpG islands has been far more studied and has a much clearer role in tumour development. Such studies have widely been facilitated with the introduction of methylation-specific PCR (MSP) (Herman et al, 1996). The impact of hypermethylation of CpG islands during tumour development have been reported for a large number of genes, (e.g p16 and p73 that involve in the loss of cell cycle control, MLH1 and MGMT that contribute in an increased mutation rate and RASSF1 that leads in aberrant signal transduction) (Kerr et al, 2007; Shames et al, 2007a; Strathdee & Brown, 2002) using a wide variety of detection methods (Fraga & Esteller, 2002). The logistics of DNA methylation as well as the physiological function have been under intense investigation. Two mechanisms have been proposed to provide explanation for the association of gene expression inactivation and DNA methylation. The first proposed that the presence of 5-mC directly downregulates expression by inhibiting the binding of transcription factors (Baylin et al, 1998; Jones & Baylin, 2002; Figure 2.2.1.1). The second suggests that DNA methylation triggers the recruitment of a protein complex resulting in histone modification and subsequent chromatin reorganisation (Bird & Wolffe, 1999) (Figure 2.2.1.2.).

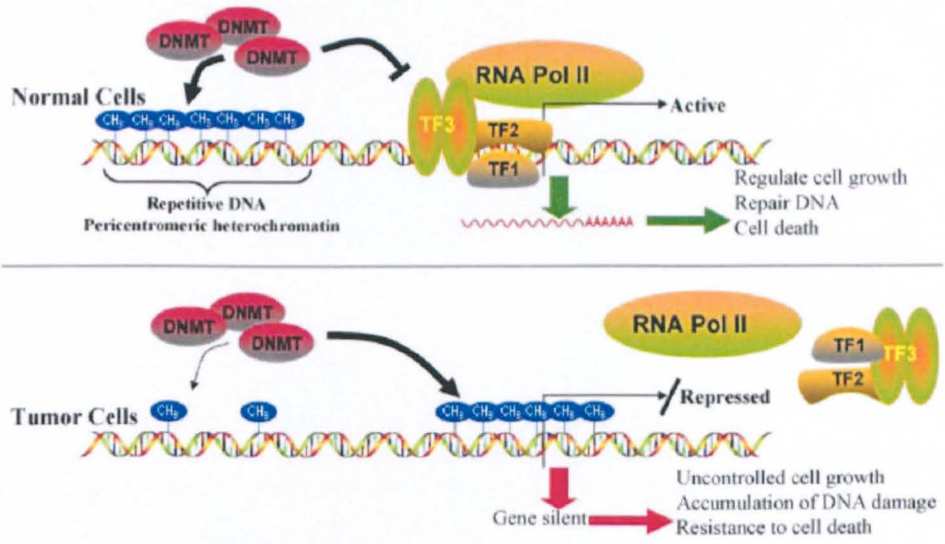


Figure 2.2.1.1. Model suggesting a direct inhibition of the transcription by DNA methylation. The presence of 5-mC creates a stereochemical hurdle for the transcriptional machinery to physically "dock" on the promoter and initiate transcription. (figure adapted from http://www.med.ufl.edu/biochem/keithr/research.html)

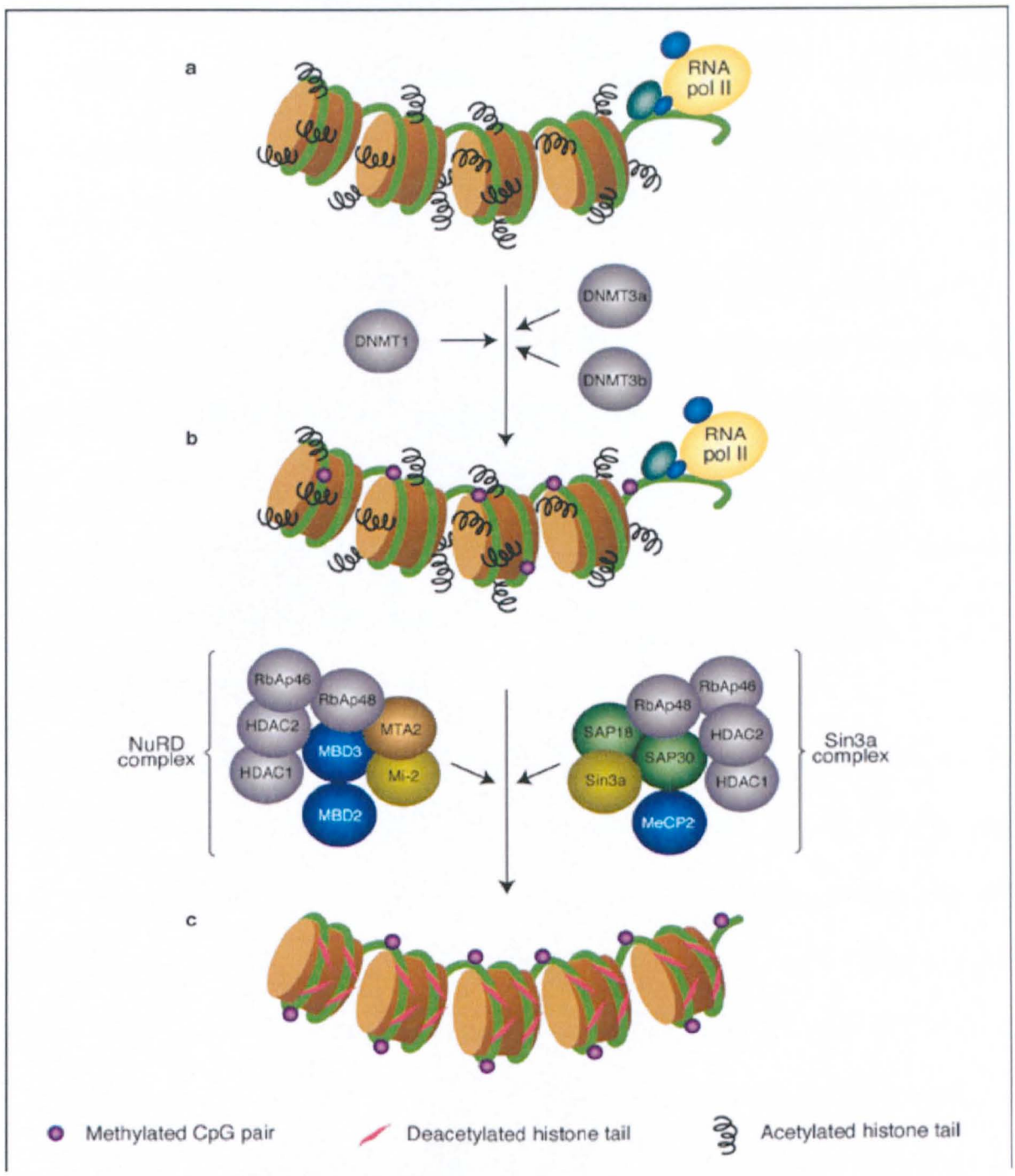


Figure2.2.1.2. Proposed model, explaining the transcriptional repression induced by DNA methylation. The model suggests that DNA methylation is a target for methylcytosine-binding proteins such as MBD2 and MeCP2. These proteins are associated with protein complexes that include histone deacetylases HDAC1 and 2. The latter deacetylate the histone tails resulting in reorganisation of chromatin in a tighter form which makes it inaccessible to the transcription machinery (Strathdee & Brown 2002).

Studies of DNA methylation in lung cancer to date strongly suggests that the analysis of DNA methylation profiles will be of great utility both for understanding the molecular basis of cancer development as well as for developing early diagnosis tools (Kerr et al, 2007). To date, many studies on methylation status of various genes have been performed on resected NSCLCs in order to stratify genes as potential biomarkers. Multiple lines of evidence, employing gene-specific or genome-wide methodologies demonstrate a high frequency of hypermethylation in particular genes in NSCLC (Zochbauer-Muller et al, 2001; Zochbauer-Muller et al, 2002; Dunn et al, 2004; Toyooka et al, 2006; Xinarianos et al, 2006; Rauch et al, 2006; Rauch et al, 2007; Rauch et al, 2008). Particular attention has been given to evaluating the potential clinical significance of methylation of TSG in patients with NSCLC (Minna et al, 2002; Belinsky et al, 2004).

The epigenetic inactivation of tumour suppressor genes by DNA methylation is tumour type-specific (Esteller et al, 2001) and affects all cellular pathways. Examples of genes undergoing this aberrant DNA methylation include genes involved in the cell cycle (p15 and p16), DNA repair (MGMT, BRCA1, hMLH1), carcinogen metabolism (GSTP1), cell adherence (CDH1, CDH13), and apoptosis (DAPK, TMS1) (Baylin et al, 2000; Herman, 1999). The cyclin-dependent kinase inhibitor 2A (p16) blocks the action of cyclin dependent kinases CDK4 and CDK6 leading to G1 arrest and senescence (Hara et al, 1996). p16 is one of the most cited genes regarding its methylation status in human tumour of the respiratory tract (Belinsky et al, 2007; Shivapurkar et al, 2007; Shaw et al, 2006). Although promoter hypermethylation of p16 was not significantly related to conventional clinicopathological characteristics of patients or tumours by some studies (Ulivi et al, 2006; Wang et al, 2008) results from the European lung cancer consortium (EUELC) suggest that p16 methylation correlates with tumour relapse within a year from resection (Liloglou et al, In preparation). This is in

agreement to findings from Kim *et al* (2006) and Nakata *et al* (2006) supporting a prognostic role for p16 hypermethylation in lung tumours.

MGMT encodes a DNA repair protein responsible for removing alkylation adducts from the O(6)-position of guanine in DNA (Jacinto & Esteller, 2007). The transcriptional silencing of MGMT by promoter hypermethylation instigates an important mutator pathway in human cancer, because the O^6-methylguanine adducts, resulting from alkylating agents are not removed, thereby producing G:C to A:T transitions (Horsfall et al, 1990).

Following the example of the "mutator" phenotype in cancer (Loeb et al, 2008), researchers have defined a "CpG island methylator" phenotype (CIMP) in various types of cancers, including NSCLC (Liu et al, 2008; Marsit et al, 2006; Suzuki et al, 2006). The term is used to describe tumours with irregular function of the DNA methylation machinery leading to abnormal methylation in multiple genes (Issa, 2004).

2.2.2. Histone Modifications

The epigenetic reprogramming of human cancer involves a close cooperation of DNA methylation and a number of histone modifications (Esteller et al, 2001). This cooperation has been shown to contribute to the transcriptional silencing of tumour suppressor genes in human lung cancer (Risch & Plass, 2008). The N-terminus domain of all core histones is subject to chemical modifications including acetylation, methylation, phosphorylation, ubiqitination, sumoylation and other covalent modifications at certain residues (Jenuwein & Allis, 2001; Clapier & Cairns, 2009; Kouzarides, 2007). These distinct modifications that play repressive and/or activating roles in histones are considered to be part of the regulatory mechanisms that determine cell fate, by altering the normal gene expression (Sharma et al,

2010). Changes in the positions of the nucleosomes, which is the basic repeating unit of chromatin structure, control transcription by blocking/allowing the binding of transcription factors and basal transcription machinery to promoter regulatory sequence (Cairns, 2001). Histone tail acetylation is a key element influencing the condensation state of chromatin and thus transcriptional activity (Strahl & Allis, 2000). Furthermore, phosphorylation of Ser-10 at histone H3 has also been considered to affect transcriptional activation as well as chromosome condensation during mitosis (Cheung et al, 2000). Additional histone modifications play an important role in the chromatin reorganisation. Histone H3 K9-methylation and K4-methylation oppose one another to create gene-silencing heterochromatic chromatin versus gene activating-chromatin (Noma et al, 2001). Histone modification in normal cells and tumour development is an exploding field providing new insights on the human epigenome and its potential exploitation in cancer diagnosis and therapy (Esteller, 2007a).

2.2.3. Chromatin remodelling

Chromatin is known as the complex in which DNA is packaged together with proteins. The role of chromatin remodelling and nucleosome positioning have been characterized as epigenetic effects that regulate gene expression. Initial findings on the involvement in histone modification have been reported as one of the fundamental epigenetic events on cancer (Jones & Baylin, 2002). Nucleosomes are barrier to transcription that blocks access of activators and transcription factors to their sites on DNA. At the same time, they serve as inhibitors for the transcript elongation by engaged polymerases (Ho & Crabtree, 2010). Nucleosome displacements of as few as 30 bp at transcription start site have been

implicated in changing the activity of RNA polymerase II (Schones et al, 2008). Nucleosome remodelling and the incorporation of histone variants are largely accomplished through the action of ATP-dependent chromatin remodelling complexes (Berger et al, 2009). These complexes are a diverse family grouped into SWI/SNF, ISWI, CHD, or INO80 sub-families, based upon sequence homology of the associated ATPase (Clapier & Cairns, 2009). All families of chromatin remodelers have been tied to cancer, although the molecular mechanisms are not serving a clear function and remain under investigation (Portela & Esteller, 2010). The fundamental mechanism mediated by these complexes is thought to be the movement or exchange of nucleosomes to regulate transcription. (Hargreaves & Crabtree, 2011). NURF, for instance, an ATP-dependant chromatin complex specifically targeted to chromatin through interactions with sequence specific transcription factors and modified histones (Alkhatib & Landry, 2011).

2.2.4. RNA inhibition (RNAi)

MicroRNAs (miRNAs) represent a class of naturally occurring small noncoding RNA molecules, distinct from, but similar to siRNAs (Fabbri et al, 2007). Many microRNAs (miRNAs) target mRNAs involved in processes aberrant in tumourigenesis, such as proliferation, survival, and differentiation (Kumar et al, 2008). Components of miRNA-machinery have been implicated in tumourigenesis. Expression of Dicer, which is a protein with double-strand RNA-cleavage activity and initiates RNA interference (Macrae et al, 2006), has been shown to be downregulated in lung cancer (Karube et al, 2005). Moreover, a certain fraction of miRNAs proved to be regulatory elements of cancer-related processes such as cell growth and tissue differentiation and therefore might themselves function as

oncomirs (Esquela-Kerscher & Slack, 2006). miRNA deregulation is associated with carcinoma development at various sites, including those of the lung, breast, pancreas, ovary and the lower digestive tract (Galasso et al, 2012; Kumar et al, 2007; Volinia et al, 2006). One of the best studied miRNA families is the let-7 family which has been shown to regulate the expression of Ras oncogenes (Kumar et al, 2008).

Chapter 3. Molecular tools for early diagnosis of NSCLC.

It is a fact that more than half of lung cancer cases are diagnosed at an advanced stage which is usually beyond effective treatment. Consequently, lung cancer is a prime candidate disease for early screening. This is clearly demonstrated by the fact that the overall 5-years survival for non-small cell lung cancer remains at only 15%, and at the same time a diagnosis at stage IA increases the 5-year survival rate to 80% (Mulshine & Sullivan, 2005). Thus, there is a clear need for improved clinical stratification methods that can identify patients with early-stage disease or even high-risk individuals (Sidransky, 2002).

Biomarkers are defined as "*a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*" (NIH definition) (Atkinson et al, 2001) Biomarkers are measurable in biological media, such as in tissues, cells, or fluids. Lung cancer, as most human cancers, is characterised by a plethora of genetic and epigenetic alterations, as discussed previously. Gene mutations, allelic imbalance and irregular gene expression products and abnormal DNA methylation are affluent in malignancies and have been employed as potential tools for early diagnosis (Dunn et al, 2003; Slebos et al, 1990; Horio et al, 1993; Liloglou et al, 2001; Field et al, 1999).

DNA methylation abnormalities comprise a common phenomenon in human lung cancer and thus it is suggested that DNA methylation based biomarkers may assist in diagnosis, prognosis and patient stratification towards individualised therapeutic regimens. It is already stated that screening for promoter hypermethylation of tumour suppression genes, in biological fluids such as sputum and bronchial lavage can be positively detected several

years before the clinical diagnosis (Belinsky et al, 2006). Although a lot of data has already been produced, the spectrum of DNA methylation markers with the high potential in clinical use is not yet clear. Large retrospective and prospective studies with high statistical power are still required to provide adequate sensitivity and specificity.

Since the high frequency of epigenetic abnormalities in human primary tumours was made apparent, in the early 2000s, many groups have started investigating the feasibility of detecting such epigenetic aberrations in body fluids. To date the vast majority of such reports have examined DNA methylation, while there are fewer on detecting histone modifications (Deligezer et al, 2008; Paul et al, 2010). This clearly reflects the relevant literature on surgical tumour specimens but also underlines the possibly higher utility of nucleic acid changes in clinical specimens because of their stability. The detection of microRNAs in body fluids has also been reported (Gilad et al, 2008; Keller et al, 2009) and is expected to significantly contribute to cancer diagnostics. The major challenge regarding the detection of cancer specific DNA methylation in body fluids is, as in the case of all other DNA abnormalities, that cancer cells and/or cancer circulating DNA are present in small quantities amongst the normal contaminating DNA.

3.1. DNA methylation detection techniques and methodologies.

The need for accurate detection of minute amounts of abnormally methylated DNA in body fluids inevitably leads to discussing methodology issues and the appropriateness of different approaches. As 5' methyl cytosine pairs to guanine as effectively as unmethylated cytosine, common sequence detection techniques are inadequate for this purpose. A large number of DNA methylation detection methods and modifications exist to date, each having certain

advantages and disadvantages. There are no golden, totally bulletproof methods. There are good methods which, provided the necessary quality control measures are undertaken, can successfully address the research questions in particular projects.

There are four major approaches regarding the means of discrimination between 5' methyl cytosine and cytosine:

1. Chromatin immunoprecipitation (ChIP), which utilize anti-5' methyl cytosine (Webber et al, 2005) or methyl binding domain proteins (MBD) (Ballestar et al, 2003).
2. Methylation-Specific Restriction Enzyme (MSRE) based methods, which take advantage of the differential recognition of methylated DNA sequences over their unmethylated counterparts (Schumacher et al, 2008).
3. Bisulphite-based methods, which exploit the conversion of cytosine - but not 5' methyl cytosine to uracil, converting this epigenetic modification into a sequence difference (Herman et al, 1996).
4. Mixed methods utilizing combinations of these principles (Xiong & Laird, 1997).

Among the applications downstream of the above mentioned processes are microarrays and PCR-based methods. The latter can provide a direct result, such as methylation specific PCR (MSP) or prepare templates for subsequent detection with single strand conformation polymorphism (SSCP) analysis, high resolution melting (HRM), sequencing and Pyrosequencing.

One easily recognizes the particularities in detecting abnormal DNA methylation in body fluids. These are considered to carry a tiny load of abnormal nucleic acids in the presence of high amounts of "contaminating" normal DNA. In addition, the overall amount of DNA yielded may be relatively low, frequently due to the fact that only a small part of this clinical

specimen becomes available for research purposes. Exception to this is the detection of DNA methylation in the peripheral leukocytes in leukaemias or for monitoring therapeutic schemes with epigenetic drugs (Stewart et al, 2009) where the target DNA is ample.

The reliable detection of minimal residual disease associated nucleic acids in patient with solid tumours in samples such as plasma/serum, urine, sputum, bronchial washings (BWs), saliva etc. is demanding and devious. The most widely used approach has been bisulphite conversion of DNA followed by methylation specific PCR (MSP), in either its endpoint or real time (qMSP, Methylight) version. The concept here is that methylation-specific primers bearing cytosines (forward) and guanines (reverse) preferably at the 3' end of both primers will generate amplicons from only the methylated DNA copies of the target sequence under optimal conditions. The long experience of the research community in PCR amplification techniques allows, in combination with the availability of new engineered hot start *Taq* (Thermus Aquaticus) polymerases and reliable thermocycling hardware, to reach high levels of fidelity. Endpoint qMSP is a method which revolutionized epigenetic research (Herman et al, 1996). Its main disadvantage is the lack of quantitation efficiency as well as problems in sensitivity and specificity. The latter can be overcome by using methylation enrichment pyrosequencing (MEP), which employs Pyrosequencing to confirm the status of an MSP amplicon (Shaw et al, 2006). The real time version of MSP (qMSP or Methylight) overcomes many of the endpoint assay problems. It is highly sensitive (especially when using fluorescent probes) in visualizing minute amounts that would never been seen on gels. The use of fluorescent probes adds one more level of sequence specificity and allows for multi-color detection of internal controls to normalize for DNA input. What is currently missing from qMSP/Methylight is an internal control of bisulphite conversion.

As mentioned above, an important problem in body fluids is the low availability of DNA. To make the problem more challenging, bisulphite conversion diminishes DNA quality and the subsequent cleanup reduces significantly recovery. It is widely accepted today in the epigenetic biomarker research community that it is unlikely to find the perfect single marker. Most studies point to the discovery of panels of biomarkers, thus multiple assays. Therefore, the tiny amount of DNA recovered has to be split in different reactions. One way to overcome this problem is to multiplex four targets per reaction using probes with different fluorescent dyes (Fackler et al, 2009). Of course the level of optimization this requires to prove equal amplification factors over 5 concentration logs between the different amplicons is significant. The abundance of each target is different, thus in the absence of such calibration the high abundance target (10^{-1}-10^{-2} per genome) will probably consume resources impairing amplification of the rare ($\leq 10^{-3}$) copies. A different approach is the post-bisulphite whole genome (bisulfitome) amplification (Vaissiere et al, 2009) which demonstrates promising findings, but further research is required to prove the potential extent of its use.

A last important point, frequently bypassed, regarding methodology is the amount of DNA added in a qMSP reaction. qMSP users claim detection of targets in dilution with normal DNA as low as 10^{-4}. In this case one needs to ensure 10^4 genome equivalents (21 ng) are added in the reaction to have one abnormal copy keeping in mind that MSP targets only one bisulphite converted strand). This is also an important issue for consideration in whole bisulphite amplification.

Overall, huge progress has been demonstrated in the last decade in methodological approaches for DNA methylation detection in body fluids. Many academic groups have

involved themselves in this research and the contribution of the biotechnical industry must also be acknowledged. Although there are still issues regarding the quality control of the used techniques, there should be no doubt that these will be overcome and a clinically useful method will be very soon available.

3.2 DNA methylation detection in sputum and bronchial washings.

Lung cancer is the first cause of cancer-related deaths in the western world and a clear example of the lack of early diagnostics and screening programmes (Field & Duffy, 2008; Smith et al, 2008a). The cytological examination of BWs (also refered to as bronchial or bronchoalveolar lavages) and sputum (induced or spontaneous) is the standard practice in assisting in lung cancer diagnosis, following suspicious symptoms and/or imaging outcome (Smith et al, 2008b). However, cytology alone has generally low efficiency in diagnosing lung tumours (Dobler & Crawford, 2009). Studies on DNA methylation aberration in lung tumours have produced a long list of genes that may serve as candidate biomarkers for lung cancer diagnosis (Divine et al, 2005; Ehrich et al, 2006; Field et al, 2005a; Tessema et al, 2009). Many of these genes have been tested in sputum and BWs providing diverse results.

The feasibility of detecting p16 promoter hypermethylation in bronchial lavage was shown in an early study by Ahrendt et al (Ahrendt et al, 1999) while subsequent studies investigated panels of promoters (Kim et al, 2004; Topaloglu et al, 2004). DNA methylation analysis of APC, p16, RASSF1 and RARβ have been showed, while the same group reported good specificity for RASSF1 (Grote et al, 2006), as well as for p16 and RARβ the same group reported good specificity (Grote et al, 2005). Analysis of DNA methylation in bronchial lavage appears to achieve higher sensitivity and specificity in detecting central rather than

peripheral tumours (de Fraipont et al, 2005) which is consistent with the nature of the bronchoscopic examination.

DNA methylation in sputum of multiple genes (among p16, MGMT, DAPK, RASSF1A, and GATA5) was shown to be a promising predictor for lung cancer (Palmisano et al, 2000; Belinsky et al, 2005; Belinsky et al, 2006). Increased methylation in sputum from lung cancer patients has been shown for TCF21 (Shivapurkar et al, 2008) and RASSF1 (van der Drift et al, 2008). ACS/TMS1 methylation was demonstrated in the sputum of 41% of patients with stage III NSCLC, 15% of patients with stage I NSCLC and 2% of cancer-free smokers (Machida et al, 2006). A panel of four genes (p16, DAPK, PAX5beta and GATA5) assayed in sputum was reported to reflect DNA methylation in biopsies, with the highest specificity shown for p16 (Belinsky et al, 2007). Specificity and the positive and negative predictive values of DNA methylation biomarkers are still unclear. Although the prevalence of DNA methylation of specific genes is higher in the sputum/lavage of lung cancer patients, it is also reported at diverse frequencies in the sputum/lavage of cancer-free controls (Baryshnikova et al, 2008; Cirincione et al, 2006; Hsu et al, 2007; van der Drift et al, 2008). This probably is indicative of a field cancerization effect and a manifestation of early preneoplastic foci in smokers (Zochbauer-Muller et al, 2003; Verri et al, 2009; Russo et al, 2005). This smoking–related methylation "noise" poses certain challenges in biomarker implementation as it may lead to overdiagnosis. Thus the landscape of methylation marker specificity is far from being clear; keeping in mind that additional to tobacco carcinogens may also cause similar noise effects. As will be discussed further below, different studies utilize different techniques, promoter regions, primers and enzymes. Also, the quality control implemented in academic studies is most usually inferior to that required for clinical diagnostics. A significant issue in most studies mentioned is that the cancer-free smokers utilized as controls are not followed up to

5 years, which is required to establish a real clinical gold standard for estimating lung cancer risk (Baryshnikova et al, 2008). Thus, the frequency of real positives in the sputum and BWs of genuinely confirmed cancer-free individuals is not yet clear. Molecular data on sputum and BW available in research papers are not always followed by cytological reports, thus the relationship between adequacy (presence of lung macrophages) and positive DNA methylation results is not totally clear. It has been reported that DNA methylation of particular genes is found in samples independent of their adequacy report (Belinsky et al, 2005; Belinsky et al, 2006). This can be interpreted two-fold. It may suggest the presence of circulating cancer DNA in the specimen and/or it may be indicative of a field cancerization effect. In the latter case, one is not detecting an abnormality from the tumour cells themselves but abnormalities from histologically normal bronchial or oral cells. Although this is considered as background or "noise", it can significantly assist diagnosis; instead of a direct indicator of a cancer lesion, this could be considered as a surrogate marker of increased risk prompting patient follow up over a definitive time.

There are a number of issues one should consider when designing biomarker studies utilizing BWs and sputum. Assuming that molecular assays used for the methylation detection are "bulletproof" (this will be discussed below), the basic limitations come from the very nature of these specimens. The bronchial washing is a representative specimen of a particular bronchus. Thus, the molecular detection of a tumour in BW depends on the proximity of the bronchus examined to the tumour. BWs are usually rich in bronchial cells and frequently contain traces of blood but lack contaminating oral squamous epithelial cells. The latter are frequently present in sputum, which theoretically represents a wider area of the lung; however, its cell content strongly depends on the training given to the patient of how to produce a good sputum specimen. Sputum can be induced or spontaneous; for

example three early morning sputum collections in order to increase the number of cells in the specimen. It is common understanding for researchers using BW and sputum that, as both specimens originate from the main bronchi, they will provide higher sensitivity in the detection of central (usually squamous and small cell carcinomas) than peripheral (frequently adenocarcinomas) tumours (Field et al, 1999; Liloglou et al, 2012). It is also known that there is low consistency of the cell profile in the two specimen types among the different patients in the same clinical setting. This variability increases between different clinical settings. Sputum samples contain a large number of oral squamous cells while one cannot exclude the presence of cells exfoliating from the esophagus. These issues pose a limitation in the molecular diagnosis of lung cancer and are frequently overlooked by researchers, who often concentrate more on the efficiency of the chemistry of their PCR assays, commonly tested on standard DNA dilutions.

3.3 DNA methylation detection in blood, plasma and serum

Plasma and serum are frequently used in clinical research as potential sources of low invasiveness specimen sampling for DNA methylation-based diagnostics. The relative suitability between these two sample types is not totally clear, as very few studies have used both types of specimens in parallel; however they appear to be comparable (Wong et al, 2003). Plasma seems to be used in preference to serum. That could be supported probably due to the concern that cancer circulating DNA might be trapped in the coagulating clot. The great advantage of plasma/serum is that it is probably present with the highest uniformity of specimen collection and preparation in comparison to any other clinical sample. It is less subject to site- or operator- dependent variability and can

nevertheless be easily standardized among hospitals. In contrary, the main disadvantage of plasma /serum, at least for as long as site specific markers do not exist, is that the detected methylation abnormalities may have originated from anywhere in the body. Thus, although the feasibility of detecting tumour-specific epigenetic abnormalities has been demonstrated, the sources of potential contaminating signal are increased, skewing the results and escalating false positives. It is currently difficult to envisage how a plasma/serum positive screening assay would point the clinician towards the site of malignancy.

As in the case of the plasma and serum, different genes have been tested providing diverse and sometimes conflicting results. DNA methylation abnormalities have been reported in the plasma and serum of patients with a variety of cancers. A summary of these studies is depicted in the Table 3.3.1.

Table 3.3.1. Reported genes found hypermethylated in particular cancers in the literature.

Cancer	Genes	Reference	
Lung	p16, MGMT, RASSF1	Belinsky Klinge 2005	PLASMA
Breast	RARβ, TWIST, Cyclin D2,	Bae Shim 2005	
	GSTP1, RARβ, RASSF1A, APC	Hoque Feng 2006	
Prostate	GSTP1, CEA, PSMA, PBMC	Papadopoulou Davilas 2004	
	GSTP1, RASSF2, HIST1H4K, TFAP2E	Payne Serth 2009	
Gastric	p15, MGMT, hMLH1	Kolesnikova Tamkovich 2008	
Pancreatic	CCND2, SOCS1, THBS1, PLAU, VHL	Melnikov, Scholtens 2009	
	BRCA1, CCND2, hMLH1, CDKN1C, PGR, SYK, VHL	Ligget Melnikov 2010	
Colorectal	SEPT9	Grutzmann Molnar 2008	
	TMEFF2, NGFR, SEPT9	Lofton-Day, Model 2008	
	SEPT9, ALX4	Tanzer Balluff 2010	
Hepatocellular	APC, WIF-1, RUNX-3, DLC-1, SFRP-1, DKK, E-CAD	Liu Zhang 2011	
Ovarian	BRCA1, EP300, NR3C1, MLH1, CDKN1C, PGR, THBS1	Melnikov, Scholtens 2009	
Glioma	p16, MGMT, p73, RARβ	Weaver Grossman 2006	
Lung	p16, MGMT, RASSF1, DAPK, H-CAD, PAX5α, PAX5β	Belinsky Klinge 2005	SERUM
	p16, DAPK, PAX5β, GATA5	Belinsky Grimmes 2006	
	DAPK, MGMT, GSTP1	Hoffmann Kaifi 2009	
Breast	APC, RASSF1A, ESR1	Van der Auwera Elst 2009	
Gastric	p16	Abbaszadegan Moaven 2008	
	RUNX3	Sakakura Hamada 2009	
Colorectal	MGMT, P16, RARβ2, RASSF1A, APC	Taback Saha 2006	
	p16, MGMT, hMLH1	Wang Sasco 2008	
Hepatocellular	p16, p15, RASSF1	Zhang Wu 2007	
Glioma	p16	Wakabayashi Natsume 2009	
Neuroblastoma	RASSF1	Misawa Tanaka 2009	

The studies mentioned above mainly focus on early diagnosis of disease; however, these sample types have also demonstrated a benefit in other aspects of cancer management, such as establishing associations with risk factor exposure (Brait et al, 2009), prediction of relapse (Hoffmann et al, 2009) and therapeutic monitoring (Aparicio et al, 2009; Sonpavde et al, 2009).

While plasma and serum samples are used to target cell-free circulating DNA from solid tumours, white blood cells are the apparent target sample for leukemia-related studies (Bullinger et al, 2010; Dunwell et al, 2010; Figueroa et al, 2010; Milani et al, 2010). In addition, methylation profiling of WBCs has been shown to provide important information

on risk prediction for breast (Widschwendter et al, 2008), bladder (Wilhelm et al, 2010) and gastric (Hou et al, 2010) cancer, as chronic exposures to carcinogens most probably leave their epigenetic imprint in blood cells.

DNA methylation is known for its critical developmental role in cancer biology (Geiman & Muegge, 2010). Furthermore, there is also an increasing line of evidence implicating DNA methylation deregulation in many human non-malignant diseases including neurodegenerative conditions (Kronenberg et al, 2009; Wang et al, 2011), cardiovascular disease (Gluckman et al, 2009) and diabetes (Ling & Groop, 2009). Thus the use of DNA methylation biomarkers in blood/plasma/serum for the clinical management of other conditions grows continuously. There are reports demonstrating the use of DNA methylation biomarkers in blood products to diagnose fragile X syndrome (Godler et al, 2010), multiple sclerosis (Liggett et al, 2010), insulin resistance (Gemma et al, 2010), underweight state in anorexia nervosa (Ehrlich et al, 2010) coronary artery disease (Sharma et al, 2008) and hypertension (Smolarek et al, 2010). In addition, hypermethylation of RASSF1 (Tsui et al) and Maspin (Chim et al, 2008) in the maternal plasma has been associated with pre-eclampsia although there are contradicting reports (Bellido et al, 2010). Hypermethylated RASSF1 is a fetal DNA marker that can be readily detectable in maternal plasma (Chan et al, 2006) while detection of placental DNA methylation patterns of chromosome 21 in maternal plasma have been suggested as a noninvasive means of prenatal diagnosis for trisomy 21 (Chim et al, 2008). DNA methylation in both cord blood and plasma have also been suggested as potential biomarkers for skewed X chromosome inactivation in IVF-conceived infants (King et al, 2010) and as a biological measure of maternal folic acid intake (Fryer et al, 2009).

3.4. Biomarker validation

The development, evaluation, and validation of biomarkers for earlier lung cancer detection and risk assessment is still at basic levels, with hundreds of published papers on small preclinical studies but lacking clinical validation similar to drug clinical trials. The Early Detection Research Network (EDRN) was created within the National Cancer Institute (NCI) in the US to address and deal with this problem (www.edrn.nci.nih.gov). The network's mission is to translate newly emerging molecular knowledge into practical clinical tests to detect cancer and cancer risk. EDRN has suggested a 5-phase procedure for biomarker validation for clinical use (Pepe et al, 2001) (Figure 3.4.1).

Currently, the Network consists of experts in basic molecular science, laboratory technology, clinical studies, biometry, and epidemiology. The expertise in laboratory science includes conducting research on the biology of incipient neoplasia encompassing the development, characterization and testing of biomarkers of early cancer and risk, development of relevant technologies for biomarker detection, and analytical tools for the evaluation of biomarkers for detection and risk assessment. The expertise in laboratory validation includes knowledge and practice of Standard Operating Procedures (SOPs), and experience in the statistical evaluation of accuracy, precision, reproducibility, and performance characteristics of tests in multi-centre settings. Expertise in patient accrual and associated clinical issues for studies will be needed to apply basic science discoveries to clinical settings. Computational and informatics needs of the Network are provided by a Data Management and Coordinating Centre and the JPL.

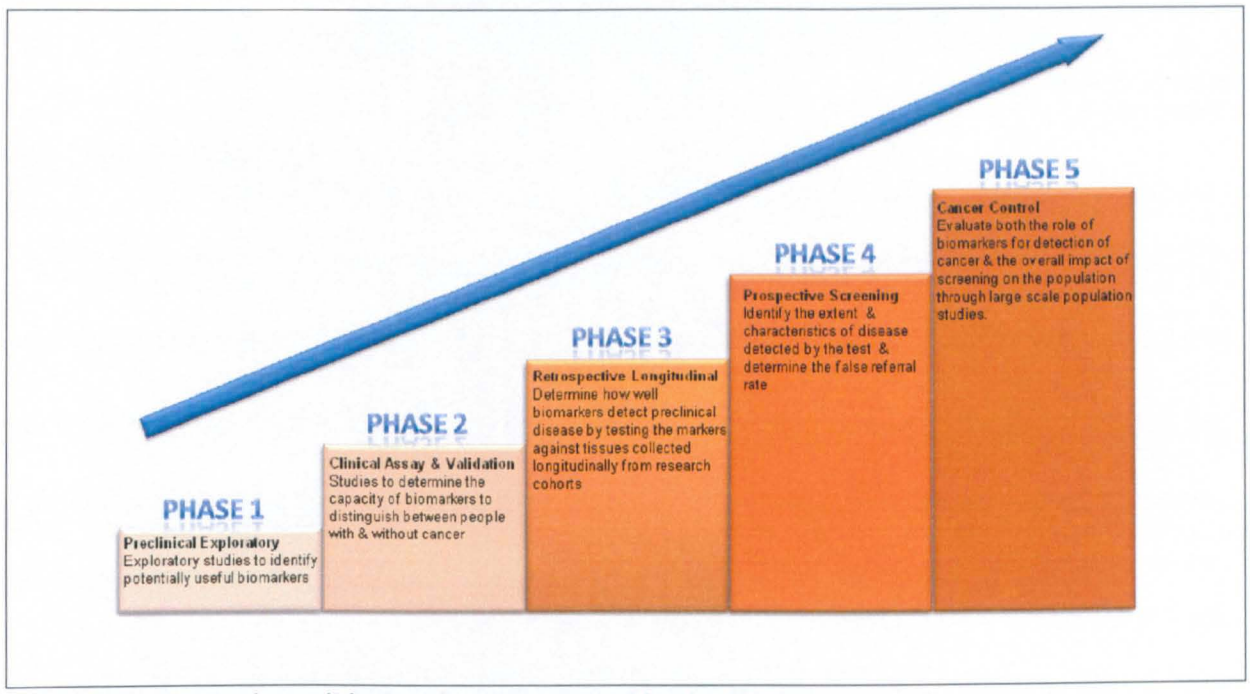


Figure 3.4.1. Biomarker validation phases suggested by the EDRN.

3.5 Aims and Objectives

The aim of this study was to construct and evaluate a panel of DNA-methylation biomarkers to assist in diagnosis of lung cancer in BWs material. Specific objectives included:

(a) The discovery of promoter targets with high frequency of hypermethylation in primary lung cancer.

(b) The development of a highly sensitive and specific DNA methylation assay fitting to clinical standards.

(c) The validation of these targets in BWs utilising a longitudinal retrospective case-control design.

Chapter 4. DNA methylation Biomarker Selection

4.1. Introduction

This chapter describes the selection of target promoters from the subsequent methylation profiling of the BWs. For this phase, it was very important to have quantitative information from a longer stretch of the promoters, thus Pyrosequencing (sequencing-by-synthesis) has been employed. Analysis of DNA methylation patterns by pyrosequencing combines an easy-to-use protocol that conferring high reproducibility and accuracy.

4.1.1. Early detection program of the Liverpool Lung Project (LLP)

As previously mentioned, the aim of the LLP is to deliver a molecular-epidemiological risk assessment model, which will facilitate early detection of lung cancer and thus reduction of the overall mortality by the disease (Cassidy et al, 2008; Field et al, 2005b). Detailed epidemiological and molecular profiling is used to identify risk factors and early biomarkers. Genetic predisposition studies are performed in the Case/Control context. In-depth interviews are carried out using structured and semi-structured questionnaires. Sputum and blood are collected from virtually all recruits while other specimens such as bronchoalveolar lavage, mouth swabs and surgical specimens are collected from smaller subsets within the study. These specimens are utilised for the development and validation of specific molecular markers (*e.g.* genetic instability, mutation and expression profiling, and methylation status) (Liloglou et al, 2001; Field & Youngson, 2002; Heighway et al, 2002; Ehrich et al, 2005; Ehrich et al, 2006).

In the last ten years the LLP biomarkers group has developed a particular interest in DNA methylation abnormalities of lung tumours and their potential utilisation for diagnostic and

therapeutic stratification purposes. The LLP Biomarker Discovery strategy is depicted in Figure 4.4.1. The team has collaborated with major biotech partners such as Sequenom (Palo Alto, CA) and Epigenomics (Berlin) providing a number of potential diagnostic targets (Field et al, 2005a; Ehrich et al, 2005; Ehrich et al, 2006). Using MALDI-TOF mass spectrometry technology, a wide variety of gene promoters has been screened, covering a total of 1426 CpG positions. Taking into account the outcome of this novel experimental approach for methylation detection – despite the fact it has a detection limit of 5% methylation DNA (Ehrich et al, 2005) – this study was the first to suggest that DNA methylation analysis can be used in combination with gene expression profiling to discover a clinically meaningful molecular marker set (Ehrich et al, 2006). Moreover, using microarray-based assay approach the methylation patterns of 245 CpGs in 59 candidate genes were examined in lung adenocarcinomas and squamous cell carcinomas of the lung, as well as in matched normal controls. ARH1, MGMT, GP1bβ, RARβ and TMEFF2 have been proved to be the best discriminators between squamous cell carcinomas and normal tissues. In contrast, when comparing adenocarcinomas and normal adjacent tissues, only TMEFF2 and MGMT are significant discriminators (Field et al, 2005a).

Simultaneously, targets reported in the published literature or selected through sub-studies within LLP are evaluated. One such example of a strong candidate biomarker is cytoglobin (CYGB). This gene was shown to present with significant of genetic and epigenetic changes in non-small cell lung carcinomas (Xinarianos et al, 2006), oesophageal cancer (McRonald et al, 2006) and head and neck cancer (Shaw et al, 2006).

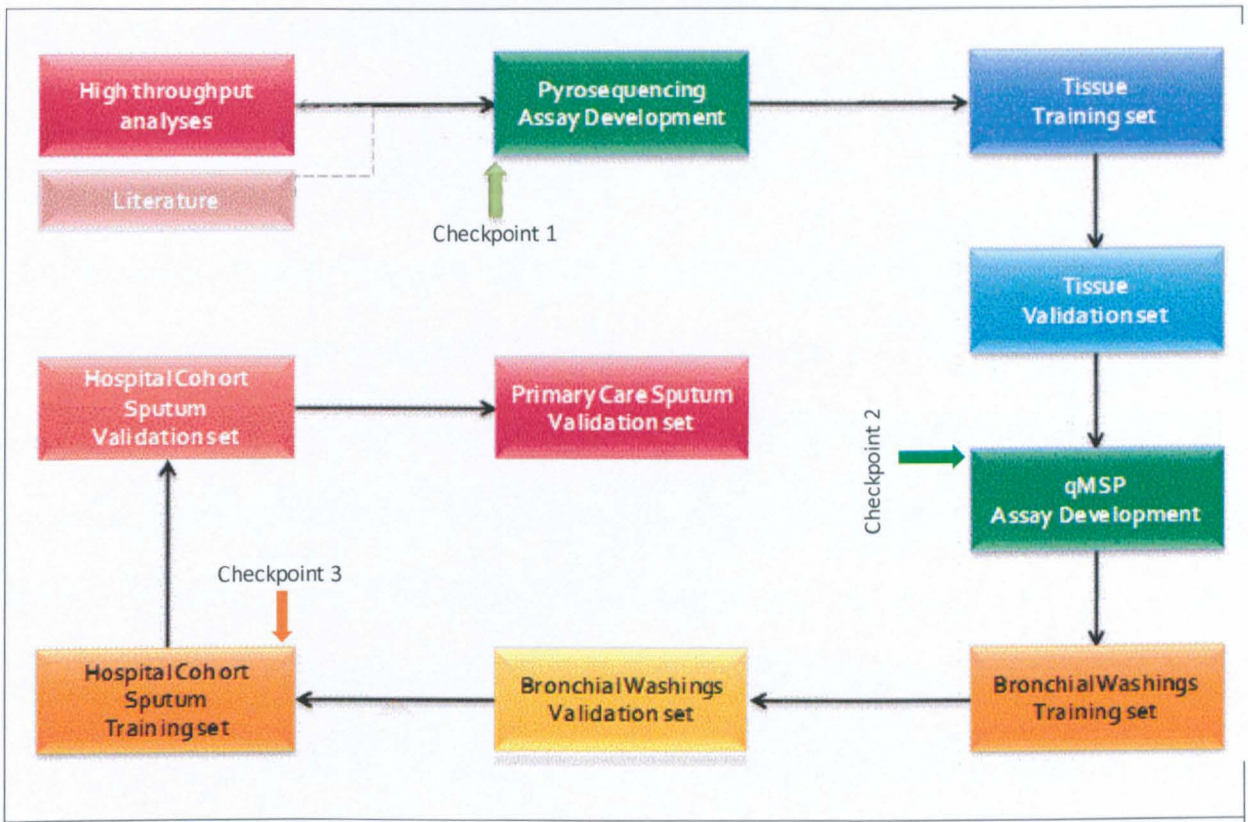


Figure 4.1.1. Epigenetic biomarker discovery strategy within LLP. At checkpoint 1, the successful candidates from previous microarray studies and/or the literature are selected for pyrosequencing assay design. Following technical validation to verify the primary observation, the targets are tested for frequency of abnormality in two independent sets of surgical material (lung tumour and normal adjacent tissue). The selected targets from this step (Checkpoint 2) are subject to qMSP assay development and optimisation. qMSP assays are then utilised to screen BWs from retrospective hospital case/control sets in order to construct diagnostic panels and configure the relevant algorithms. At checkpoint 3, once the diagnostic efficiency is considered adequate, the diagnostic accuracy of the biomarker panel(s) is tested in prospective populations.

4.2 Materials and Methods

4.2.1. Samples and Patients

The LLP is an ongoing case-control molecular-epidemiological study of lung cancer in Liverpool, United Kingdom (Field et al, 2005b). LLP has received Ethical approval from the Liverpool Research Ethics Committee and all the patients have provided informed consent. Inclusion criteria for the case arm of this study part were availability of frozen surgical tissue, histological confirmation and informed consent.

The tissue set used for the technical validation by pyrosequencing consisted of 48 NSCLCs (24 adenocarcinomas and 24 squamous carcinomas) and paired adjacent normal tissues. Twenty six (54%) patients were male and twenty two female (46%). Age of the patients ranged between 46 and 80 years (mean=64). All specimens were of advanced stage (43 T2, 4 T3 and 1 T4). It has to be mentioned at this point that there is a recognised bias in our tissue collection favouring a high frequency of T2s, as T1s are frequently too small (thus the pathologist correctly refuses to provide a research sample) and T3s/T4s are normally inoperable. Most T3/T4s availiable in the tissue bank are due to pre-surgical under-staging.

In addition, 10 adenocarcinomas and 10 squamous carcinomas were examined for p16 and WT1 protein expression by immunohistochemistry (IHC) (refer to page 75). Concerning pT stage, the set consisted of five T1, eleven T2, three T3 and one T4). Age of the patients ranged between 52 to 76 years (mean=63).

4.2.2. DNA preparation

DNA extraction from primary tissues was undertaken utilising 20×40 µm sections which were cut from frozen tissue. The first and last sections underwent pathological review from Professor J. Gosney, to ensure at least 80% tumour cell content. DNA extraction of the samples was performed using the DNeasy 96 Blood and Tissue kit (Qiagen) following the manufacturer's protocol. Briefly, tissue was lysed with 360 µl of ATL reagent and 40 µl Proteinase K solution (Qiagen) and incubated at 56°C overnight in an orbital shaking incubator at 200 rpm. 820 µl of premixed AL buffer with ethanol were added and after mixing, lysates were transferred in two "twin" 96-well plates with silica based membranes. The samples were then centrifuged at 6,000 rpm for 10 min and washed once with 500 µl buffer AW1. After centrifugation at 6,000 rpm the samples were washed again with 500 µl buffer AW2. After centrifugation at 6,000 rpm 55µl of AE buffer pre-warmed at 60°C was added to each sample and DNA was recovered by centrifugation at 6,000 rpm for 5 min.

For the DNA extraction from cell lines the DNeasy Blood and Tissue kit (Qiagen) was used. A maximum of $5x10^6$ cells were pelleted at 300 x g for 5 min and the pellet was re-suspended in 200 µl PBS. 20 µl proteinase K and 4 µl - 100 mg/ml - of RNase A (Qiagen) were added, the lysate was then mixed by vortexing and incubated at room temperature for 10 minutes. Subsequently 200 µl Buffer AL were added and the lysate was mixed thoroughly by vortexing and incubated at 56°C for 10 min. 200 µl of ethanol (96–100%) were added to the sample which was mixed thoroughly by vortexing. The mixture was transferred into the DNeasy Mini spin column (which carries a silica based membrane) placed in a 2 ml collection tube, and was centrifuged at 6000 x g (8000 rpm) for 1 min. 500 µl of Buffer AW1 were added, and the sample was centrifuged for 1 min at 6000 x g (8000 rpm). 500 µl of Buffer

AW2 were then added and the sample was centrifuged for 3 min at 20,000 x g (14,000 rpm) to dry the DNeasy membrane. The DNeasy mini spin column was then placed in a 1.5 ml microcentrifuge tube and the DNA was recovered into 200 µl Buffer AE with centrifuging at 6000 x g for 1 min. DNA quality and quantity was assessed by spectrophotometry at 260/280 nm wavelength.

4.2.3. *In vitro* methylation of genomic DNA

In vitro DNA methylation was performed in order to prepare our positive standard dilutions. Ten µg DNA were mixed with 4 µl 10x NEBuffer 2, 1 µl S-adenosyl methionine (SAM) and 1µl (4U/µl) SssI methylase in a total volume of 40 µl. Following incubation of at one hour, an additional µl of SAM was added and the reaction was further incubated for 6-16 hours. Inactivation of the enzyme was achieved at 65°C for 20 min.

4.2.4. Bisulphite convertion using ZymoResearch kit

Preparation of CT conversion reagent and M-Wash buffer were carried out following the supplier's protocol giving particular attention to the accuracy of pipetting and reagent mixing order as in preparation as the reaction is highly sensitive to pH fluctuations.

130 µl of the CT Conversion Reagent were added to 20 µl (1 µg) of the DNA sample in a PCR tube. If the volume of the DNA sample is less than 20 µl, water was added. The reactions were mixed by pipetting and centrifuged to collect all liquid to the bottom of the tube. The thermal profile was 98°C for 10 min, 64°C for 2.5 hours, 4°C storage up to 20 hours. After this, 600 µl of M-Binding Buffer were added to a Zymo-Spin™ IC Column. The column was

then mounted into a provided collection tube. The samples were loaded into the Zymo-Spin™ IC Column containing the MBinding Buffer and they were mixed by inverting the tubes several times. Centrifugation at full speed (>10,000 x *g*) for 30 sec was followed and the flow-through was discarded. 200 µl of M-Wash Buffer were added to the column and the samples were centrifuged at full speed for 30 sec. Addition of 200 µl of M-Desulphonation Buffer to the column was followed and the samples stand at room temperature (20°C – 30°C) for 15 - 20 min. After the incubation, they were centrifuged at full speed for 30 sec. Two washes using 200 µl of M-Wash Buffer and centrifuging at full speed for 30 sec and 5 min respectively needed to follow. Finally, the Zymo-Spin columns were placed into 1.5 ml eppendorf tubes and 30 µl of M-Elution Buffer were applied directly to the column matrix. A 30 seconds centrifugation required at full speed to elute the DNA. According to the protocol, an 80% yield is expected.

4.2.5. Agarose gel electrophoresis

In order to check and identify the PCR product an electrophoresis in an agarose gel was performed. In a 100 ml conical flask, 30 ml of 0.5x TBE and 0.6 g agarose added. The mixture placed in microwave until it boiled resulting in a crystal clear solution. The agarose was eyed-checked to maintain no cloudy solution. Then, 5 µl (3 mg/ml) SafeView (NBS Biologicals) added and mixed well mounted on an shaker for 4 minutes to facilitate temperature fall to around 60°C. The gel mix was transferred to the casting tray and left to set. The comb was removed from the set gel and the gel was placed into an electrophoresis tank. 0.5x TBE was added to cover the gel. 3 µl PCR product mixed with 2 µl loading buffer

and loaded into the wells. The gel was run at 80 V for approximately 30 min. After the run was completed, the bands were visualised on a UV transilluminator at 365nm.

4.2.6. Pyrosequencing

The first level validation in independent sets of NSCLC surgical samples was undertaken by Pyrosequencing (Qiagen). Pyrosequencing is a "Sequencing-by-synthesis" method based on sequential addition and incorporation of nucleotides in a primer-directed polymerase extension (Tost & Gut, 2007).

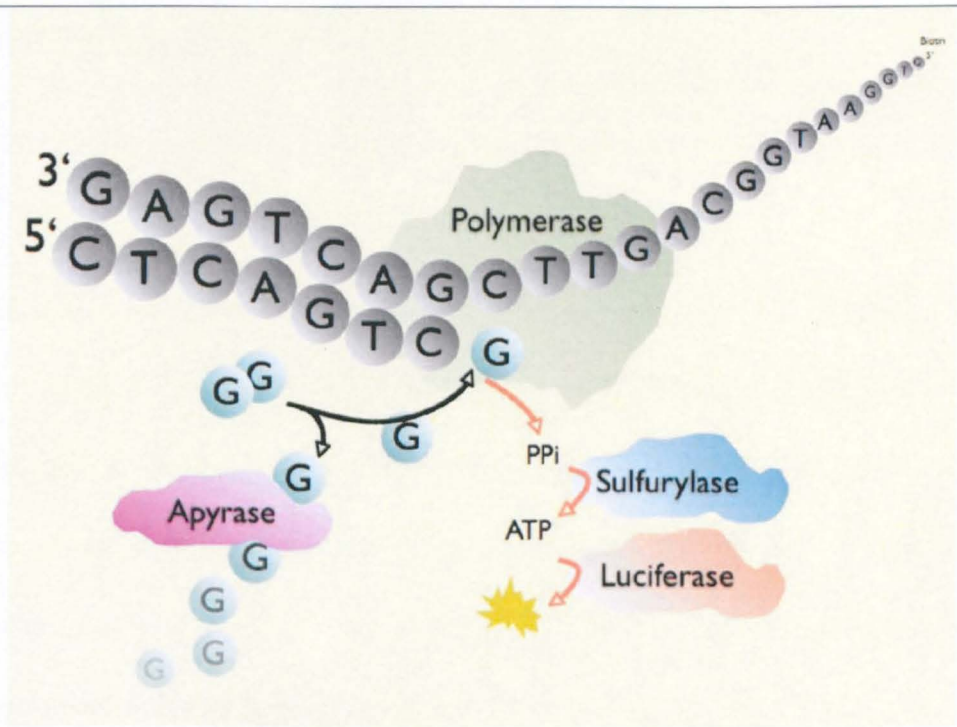


Figure 4.2.6.1. Principle of Pyrosequencing. Nucleotides are added stepwise to the template-primer hybrid. The PPi released by the DNA polymerase-catalyzed nucleotide incorporation is detected by a coupled reaction involving ATP sulfurylase and luciferase The remaining non-incorporated nucleotides are degraded by apyrase (Adapted from www.varionostic.de/technology/pyrosequencing.html).

The main advantages of pyrosequencing for DNA methylation analysis include the ability of direct quantitative sequencing, reproducibility, speed and ease-of-use (Tost & Gut, 2007). Moreover, the ability to use the same PCR product using different sequencing primers is another advantage as it reduces the cost and saves time as well as saves DNA samples which are valuable in most cases and it is now called "serial pyrosequencing" (Tost et al, 2006). Pyrosequencing-based Methylation Analysis (PMA) investigates quantitatively the degree of methylation at CpG positions in close proximity after bisulphite treatment of genomic DNA (Colella et al, 2003). PMA has now become the gold standard for quantitative methylation studies (Shaw et al, 2006; Xinarianos et al, 2006; Chen et al, 2006; Kang et al, 2006; Issa et al, 2005; Daskalos et al, 2008; Gao et al, 2008).

4.2.7. Primer design

Analysis of DNA methylation by Pyrosequencing is a process that depends upon PCR amplification yield. A strong and specific product is critical for precise quantitative analysis. For efficient primer design, Qiagen provides specialized software under the commercial name Pyromark Assay Design 2.0. After defining the targets region of interest, usually within the promoter or the 5' UTR of a gene, the optimal PCR primers and pyro-sequencing primer(s) are predicted. The basic guidelines to be followed for primer design are:

1. Target >4 CpGs, preferably avoiding T runs prior to the CpGs.
2. PCR primers should not contain CpGs but must include at least 3 non-CpG cytosines to provide specificity over partial bisulphite conversion.
3. Avoid mispriming of the sequencing primer, duplex formation with the biotinylated PCR primer and 3' loops of the template.

4. Keep template size below 250 bp (ideally 80-120 bp) to reduce probability of secondary structure formation.

The sequence of interest was downloaded from Genebank (NCBI) and an *"in silico"* bisulphite convertion is facilitated by the software. An important issue is the inclusion of (multiple if possible) bisulphite control sites within the sequenced region. These are preferably ACA and GCA (converted to ATA and GTA accordingly) trinucleotides. Incorporation of C at these sites during pyrosequencing indicate poor bisulphite conversion.

Oligonucleotides were ordered by MWG (Germany), dissolved in 10mM Tris-HCl ph 8.0, 1 mM EDTA, 50% Glycerol and stored at -20°C.

Figure 4.5.1.1. Example of PCR primers and pyrosequencing primer selection for RASSF1 methylation analysis. The target region is highlighted in blue. The screen shows relevant information such as amplicon size, primer Tms and notifications.

4.2.8. Pyrosequencing methylation analysis.

The primers used for pyrosequencing analysis of our targets, following the above guidelines, are presented in (Table 4.5.2.1.). Thermal profiles for each PCR reaction varied depending to the melting temperature (Tm) of each primer pair. The HotStarTaq Plus Master Mix (Qiagen) was employed for all PCR reactions following the protocol provided by supplier (Table 4.5.2.2.). The thermal profile was 95 °C for 5 min followed by 40 cycles of 95 °C for 30 seconds, annealing temperature (Table 4.5.2.3.) for 30 sec and extension of 30 sec. A final extension of 10 min was used to allow *Taq* finishing partial products of previous cycles. In total, twenty one genes were examined:, p16, CYGB, TMEFF2, p73, WT1, CDH13, DAPK1, TERT, RASSF1, RARb, FHIT, SERPINB5, ARHI, TIMP3, ATM, STAT1, p14, p15 and MLH1. These were selected on the basis of high frequency of hypermethylation on primary lung carcinomas either on our own microarray studies or frequently found on the literature (see biomarker strategy in page 52).

Table 4.2.8.1. Primer sequences used for Pyrosequencing Methylation Analysis (PMA)

Primer Name	Sequence 5'→3'	modification
p16meth_F	AGGGGTTGGTTGGTTATTAG	
p16meth_R	TACCTACTCTCCCCCTCTC	biotinylated
p16meth_Pseq	GGTTGGTTATTAGAGGGT	
RASSF1meth_F	AGTATAGTAAAGTTGGTTTTTAGAAA	
RASSF1meth_R	CCCTTCCTTCCCTCCTT	biotinylated
RASSF1_Pseq	AAGTTGGTTTTTAGAAATA	
TMEFF2meth_F	AGGGTGGAGGGAGAGTTAA	
TMEFF2meth_R	ACTAAAAACCTACTACTTCCCAAA	biotinylated
TMEFF2meth_Pseq	GTGGAGGGAGAGTTAAG	
TERTmeth_F	GAGGGGTTGGGAGGGTT	biotinylated
TERTmeth_R	TCCTACCCCTTCACCTTCCA	
TERTmeth_Pseq	CCCTTCACCTTCCAACT	
CYGBmeth_F	GGGAATTGATTTAAAGTTTA	biotinylated
CYGBmeth_R	AAAAAACCCAACTAAATCCAC	
CYGBmeth_Pseq	ACCCAACTAAATCCAC	
RARbmeth_F	GTTAAAGGGGGGATTAGAAT	biotinylated
RARbmeth_R	CTCCTTCCAAATAAATACTTACAA	
RARbmeth_Pseq	ACCCAAACAAACCCT	
DAPK1meth_F	GGAGTTGGGAGGAGTAG	
DAPK1meth_R	ACCAATAAAAACCCTACAA	biotinylated
DAPK1meth_Pseq	GGAGTTGGGAGGAGTA	
p73meth_F	GGTTATATTTTTTGTTTTTTGGA	
p73meth_R	GGTGTTAGGAAAGATGGGT	biotinylated
p73meth_Pseq	GTTTTTTGGATTTTAAG	
WT1meth_F	TTAGTAGTTGGGGTGAGG	
WT1meth_R	ACCAAACTCCCACACTAA	biotinylated
WT1meth_Pseq	TTAGTAGTTGGGGTGA	
CDH13meth_F	GTTGATGATTAGGATTAATGG	
CDH13meth_R	AACAAATAAAATACCACCTCC	biotinylated
CDH13meth_Pseq	GATTAATGGTTTTATAAGA	
MGMTmeth_F	GGGATATGTTGGGATAGTT	
MGMTmeth_R	CCCAAACACTCACCAAAT	biotinylated
MGMTmeth_Pseq	GGATATGTTGGGATAGT	
EDRNBmeth_F	AGTATAGTAAAGTTGGTTTTTAGAAA	
EDRNBmeth_R	CCCTTCCTTCCCTCCTT	biotinylated
EDRNB_Pseq	AAGTTGGTTTTTAGAAATA	
SERPINB5meth_F	AGTTGTTAAGAGGTTTGAGTAG	
SERPINB5meth_R	CTACTACCCCACCTTACTT	biotinylated
SERPINB5_Pseq	TTGAGTAGGAGAGGAGTGT	

Table 4.2.8.1. (cont.)

Primer Name	Sequence 5'→3'	modification
ARHImeth_F	TTTGGGTAGGGTTTATTAGTAGG	
ARHImeth_R	TCTAAAACCCCAAACTTCCA	biotinylated
ARHI_Pseq	TTATTAGTAGGGTTAGATGAG	
TIMPmeth_F	TAGTTGGAGTTTGGGGGATTG	
TIMPmeth_R	AAAACATCTTCCCCTCTCAACTAT	biotinylated
TIMP_Pseq	AGTTTGGGGGATTGG	
ATMmeth_F	AAGAGGGTGGGTGAGAGT	biotinylated
ATMmeth_R	CCATATCCACCAATAACCAA	
ATM_Pseq	CCATATCCACCAATAACC	
STAT1meth_F	AGGTTAGTTGTTAAAGGGAGTT	
STAT1meth_R	ACTAAATAAACTACAACCCAATC	biotinylated
STAT1_Pseq	AAGGGAGTTTTTAGAATGA	
P14meth_F	GGTTGTTTTTGGTAGGGT	biotinylated
P14meth_R	CCACCACCATCTTCCCA	
P14_Pseq	CCACCATCTTCCCACC	
P15meth_F	GTTGGTTTTTTATTTTGTTAGAG	
P15meth_R	TAAACTCAACTTCATTACCCTC	biotinylated
P15_Pseq	GGTTTTTGAGAGTTAGGAA	
MLH1meth_F	TTTTAGGAGTGAAGGAGG	biotinylated
MLH1meth_R	TAAAACCCTATACCTAATCTAT	
MLH1_Pseq	AACCCTATACCTAATCTATC	
FHITmeth_F	TAAGTGGAATATTGTTTTTGG	biotinylated
FHITmeth_R	TCCAAACAAAAACCCACC	
FHIT_Pseq	CCACCCCACTAAACTCC	

Table 4.2.8.2. PCR reaction components for Pyrosequencing template preparation. .

Reagent	Volume (μ l)	Final Conc
HotStar *Taq* Mix	12.5	1X
Primer	1	200 nM
H2O	8.5	
DNA	3	2ng/μl
Total Vol	25	

Table 4.2.8.3. PCR thermal profiles.

		Temperature (°C)	Time (sec)
Initial denaturation		95	300
Denaturation		94	15
Annealing	p16	57	30
	RASSF1	55	
	TMEFF2	49	
	TERT	59	
	CYGB	49	
	RARb	56	
	DAPK1	58	
	p73	52	
	WT1	57	
	CDH13	59	
	EDRNB	57	
	FHIT	54	
	SERPINB5	56	
	ARHI	58	
	GPIb	52	
	TIMP	55	
	ATM	54	
	STAT1	59	
	p14	51	
	p15	53	
	MLH1	54	
	MGMT	52	
Extension		72	30
Final extension		72	600

For pyrosequencing analysis, PCR products were mixed with 50 µl Binding Buffer (Qiagen), 3 µl Streptavidin coated sepharose beads (GE Healthcare) and 22 µl ddH_2O.

After incubation for 10 min at room temperature beads were captured onto a vacuum-operated 96-pin tool (Figure 4.5.1) and washed sequentially in 70% EtOH, 0.2M NaOH and 10 mM Tris-HCl pH 7.5 for 10 sec each. Beads were subsequently dispensed into the pyrosequencing plate containing the sequencing primer in annealing buffer (Qiagen) and after a 2 min denaturation at 80°C the plate was put into a Pyromark 96 ID machine (Qiagen) for the pyrosequencing reaction to take place (Figure 4.5.2.2.).

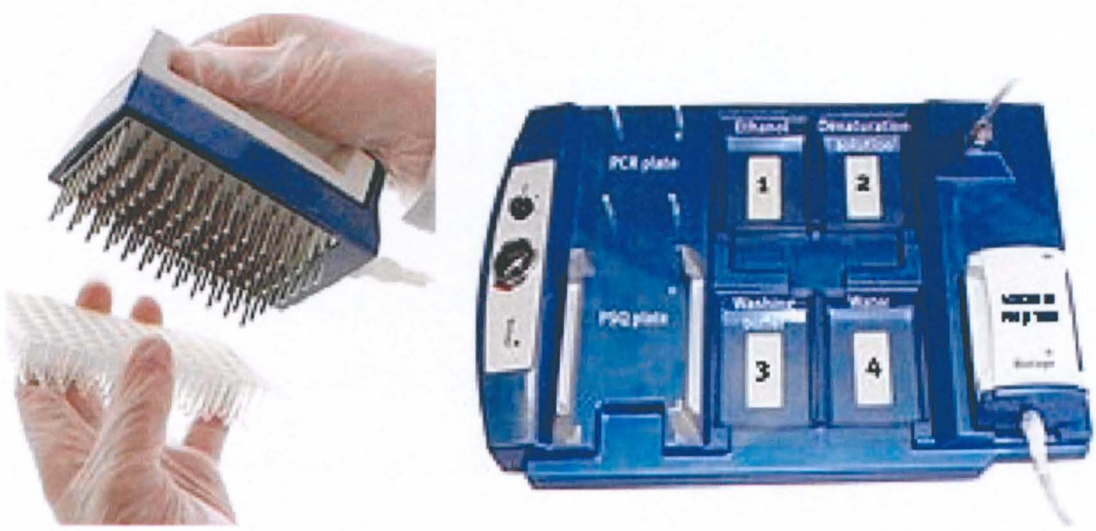


Figure 4.2.8.1. PyroMark Vacuum Prep Tool and PyroMark Vacuum Prep Worktable (Qiagen) for preparation of the pyrosequencing template. Biotinylated PCR products immobilised on streptavidin-sepharose beads which are consequently collected on the probe tips of the vacuum tool. Washes with 70% ethanol (position 1), 0.2M NaOH (position 2) and 10 mM Tris Acetate PH 7.5 (position 3) results in the desired single strand pyrosequencing template. Trough no 4 contains ddH_2O for the final wash of the tool after use. (http://www.pyrosequencing.com/DynPage.aspx?id=7267).

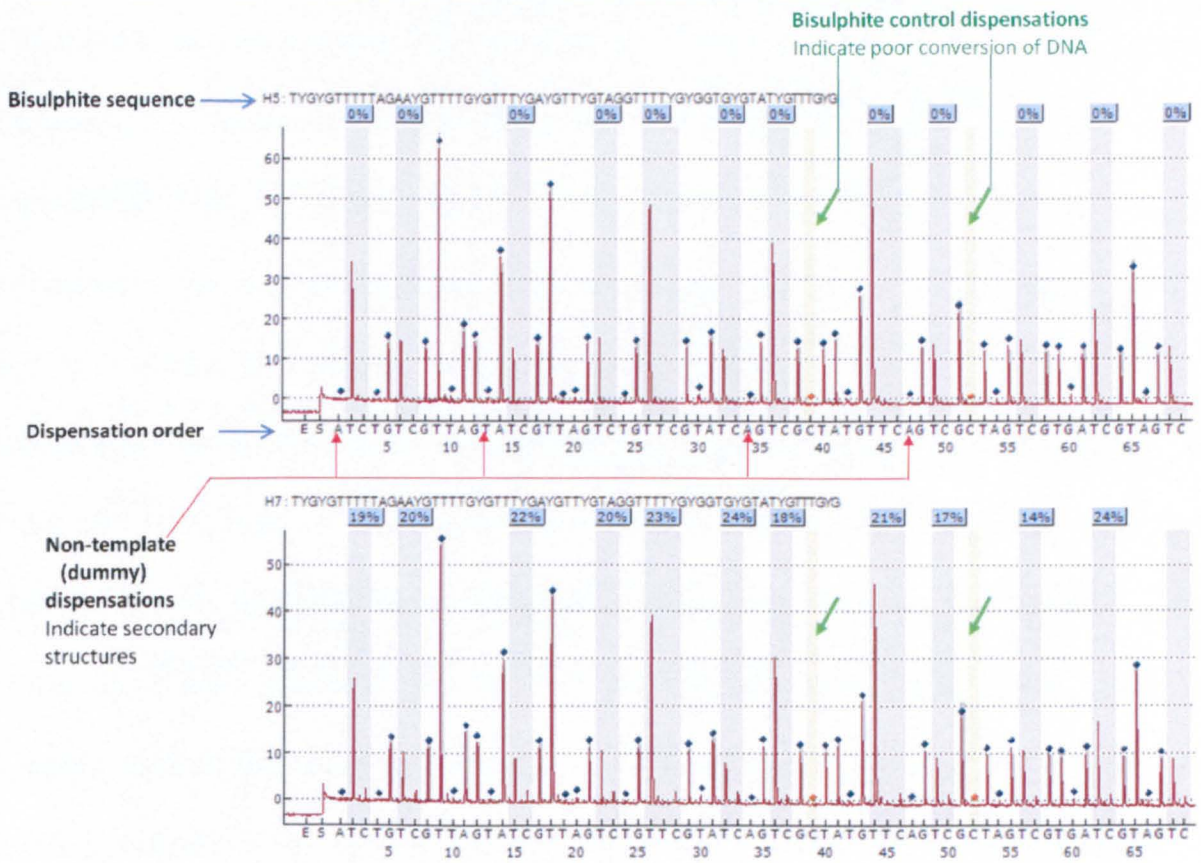


Figure 4.2.8.2. Example of MGMT pyrograms in a normal (upper) and tumour (lower) tissue. The sequence to be examined is shown on top of the pyrogram. The X axis demonstrates the nucleotide dispensation order and the Y axis shows the level of light produced due to the incorporation of each nucleotide. The gray shaded lanes indicate the examined CpGs. Bisulphite and secondary structure control dispensations are also indicated by green and red arrows respectively. The methylation level of each CpG is shown at the corresponding blue box. The percentage of methylation is calculated as the C/(C + T) peak ratio per CpG.

4.2.9. Immunohistochemistry (IHC)

The IHC for the two proteins, p16 and WT1, was undertaken in the Division of Pathology, Department of Molecular and Clinical Cancer Medicine by Mr. Andrew Dodson, using an Autostainer Link 48 (Dako, UK). Five μm tissue sections were used from ten adenocarcinomas and ten squamous carcinomas. Both p16 (CINtec, Roche mtm , Germany) and WT1 (Dako, UK) primary antibodies were used at a dilution 1:50 following high temperature antigen retrieval in high pH target retrieval buffer and detected using EnVision™ FLEX, High pH (Dako, UK). The secondary antibody was peroxidase conjugated and visualization was achieved by diaminobenzidine (DAB) staining. The positive control tissues used were spermatocytes for WT1 and cervival mucosa for p16 antibody. For negative control, the tissue was incubated with antibody diluents only, not including the primary antibody.

4.3. Results

4.3.1. Pyrosequencing Results for Methylation Analysis

The genes selected from the high throughput approaches and/or the hypothesis-based approaches were validated in an independent set of 48 NSCLC (24 adenocarcinomas and 24 squamous carcinomas) using Pyrosequencing. Particular attention was given to avoid by-products such as primer dimmers that may impair the subsequent pyrosequencing reaction(s) by competing for streptavidin binding positions.

Prior to analysis we set an arbitrary threshold of 20% as minimum accepted frequency in the primary disease. As it is well understood that multi-marker panels are required to cover the bulk of the spectrum of abnormal hypermethylation in lung cancer, this very stringent threshold was used to increase the probability of gaining successful panels with lower number of markers. This is very important as the available DNA in BWs is frequently of very low amount. Thus, keeping the number of markers (and thus assays) low, will facilitate the screening of samples from a wider range of DNA availability. The frequencies of tumour hypermethylation for each target are presented in Table 4.3.1.1. In the top part of this table, 10 genes demonstrating significant hypermethylation in the tumour tissue and absence of hypermethylation in the corresponding normal lung tissue are listed. Representative pyrograms are provided in Figure 4.3.1.1. In contrast, MGMT, TIMP3, ATM, STAT1, p14, p15 and MLH1 were below the 20% threshold we set. EDNRB, FHIT, SERPINB5 and ARHI provided some discrimination efficiency, however their relatively high methylation status in normal tissue excludes them from the pipeline.

Table 4.3.1.1. Target classification following pyrosequencing validation in the independent surgical tumour samples.

Gene	Initial selection route	Frequency (%)	Comment
TMEFF2	Epigenomics, Literature	54.2	pass
CYGB	Other	45.8	pass
RASSF1	Sequenom, Literature	41.7	pass
p73	Literature	39.6	pass
p16	Sequenom, Literature	37.5	pass
DAPK1	Sequenom, Literature	33.3	pass
TERT	Literature	31.3	pass
CDH13	Epigenomics, Sequenom	22.9	pass
RARb	Epigenomics, Literature	20.8	pass
WT1	Other	20.8	pass
MGMT	Epigenomics, Literature	16.7	Below 20%
FHIT	EU-ELC, Literature	32.3	High Backround in normal
EDRNB	AstraZeneca	48	High Backround in normal
SERPINB5	Sequenom	31	Tumour Hypomethylation
ARHI	Epigenomics	24	Imprinted
TIMP3	Epigenomics	0	No methylation in Ca
ATM	Literature	0	No methylation in Ca
STAT1	Literature	0	No methylation in Ca
p14	Literature	0	No methylation in Ca
p15	Literature	0	No methylation in Ca
MLH1	Literature	0	No methylation in Ca

Statistical analysis (Chi-square test; Bonferroni correction applied; significance set to $p<0.001$) demonstrated no significant differences in the hypermethylation frequencies of any of the 10 successful candidates between different age, sex, histology, pT/pN staging, differentiation and smoking (Table 4.3.2.). Moreover, we calculated the methylation index (MtI) as the average methylation of each sample. No statistically significant differences were observed among any of the above mentioned clinicopathological groups.

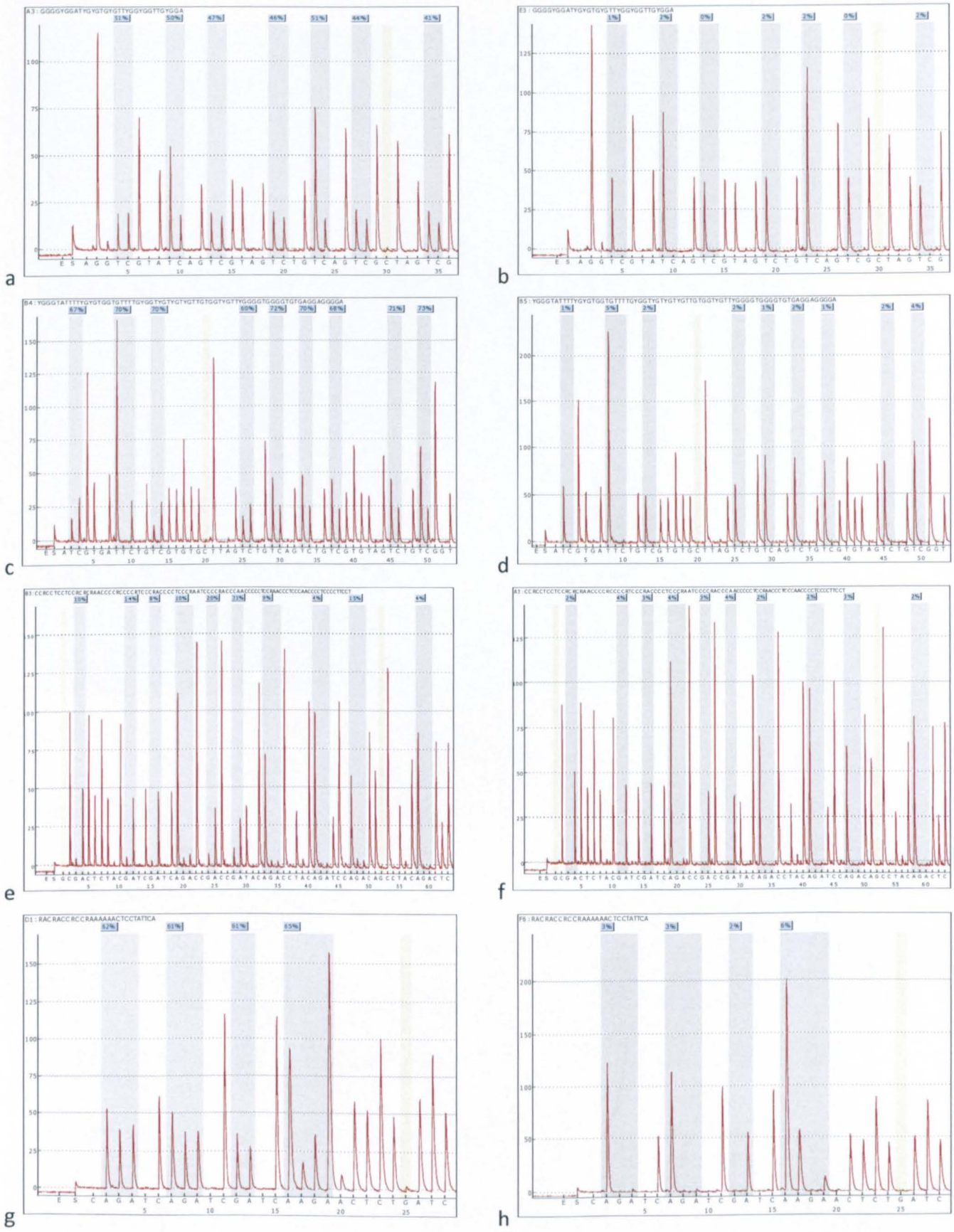


Figure 4.3.1.1. Representative pyrograms from samples with methylated and unmethylated p16 (a, b), RASSF1 (c, d), TERT (e, f) and CYGB (g, h).

Table 4.3.1.2. Pyrosequencing-based DNA methylation detection in the tissue validation set. Detailed results of the successful candidates.

Sample	Histology	Age	Gender	Smoking	pT	PN	Diff	p16	RASSF1	TMEFF2	TERT	CYGB	RARB	DAPK1	p73	WT1	CDH13
1	Ad	74	F	C	2	0	M/W										
2	Ad	50	F	C	2	0	M/W										
3	Ad	52	M	C	2	2	P										
4	Ad	68	F	C	2	1	M/W										
5	Ad	53	F	C	2	0	M/W										
6	Ad	58	F	C	2	0	M/W										
7	Ad	74	M	C	2	1	P										
8	Ad	69	F	C	2	2	P										
9	Ad	63	M	C	2	0	P										
10	Ad	64	F	C	2	2	P										
11	Ad	61	F	C	2	1	P										
12	Ad	67	M	C	2	1	P										
13	Ad	53	F	C	2	0	P										
14	Ad	67	F	C	2	0	M/W										
15	Ad	48	F	C	2	1	P										
16	Ad	68	M	X	2	0	P										
17	Ad	74	F	X	2	2	P										
18	Ad	67	F	X	2	0	M/W										
19	Ad	73	F	X	2	1	P										
20	Ad	66	F	X	2	0	M/W										
21	Ad	65	M	X	2	0	P										
22	Ad	68	F	X	2	2	P										
23	Ad	69	F	X	2	0	P										
24	Ad	77	F	X	2	1	M/W										
25	Sq	45	F	C	2	0	M/W										
26	Sq	78	M	C	2	1	P										
27	Sq	68	M	C	3	1	M/W										
28	Sq	70	F	C	2	1	P										
29	Sq	66	M	C	2	1	M/W										
30	Sq	65	M	C	3	1	P										
31	Sq	58	M	C	2	1	P										
32	Sq	66	M	C	3	1	M/W										
33	Sq	72	M	C	2	0	P										
34	Sq	47	M	C	2	2	M/W										
35	Sq	69	M	C	2	0	M/W										
36	Sq	68	M	C	3	1	P										
37	Sq	71	F	C	2	0	M/W										
38	Sq	69	F	C	2	1	M/W										
39	Sq	73	M	X	2	0	M/W										
40	Sq	60	M	X	2	1	M/W										
41	Sq	62	M	X	2	1	M/W										
42	Sq	60	M	X	2	1	P										
43	Sq	83	M	X	2	0	M/W										
44	Sq	83	M	X	2	0	M/W										
45	Sq	66	F	X	2	0	M/W										
46	Sq	63	M	X	2	0	M/W										
47	Sq	68	M	N	2	0	M/W										
48	Sq	70	M	N	4	1	P										

Ad=adenocarcinoma, Sq=squamous cell carcinoma, F=female, M=Male, C=current smoker, X= ex-smoker, N=non-smoker, Diff= differentiation M/W= moderade/good, P=poor. Green box=unmethylated, Red box=hypermethylated

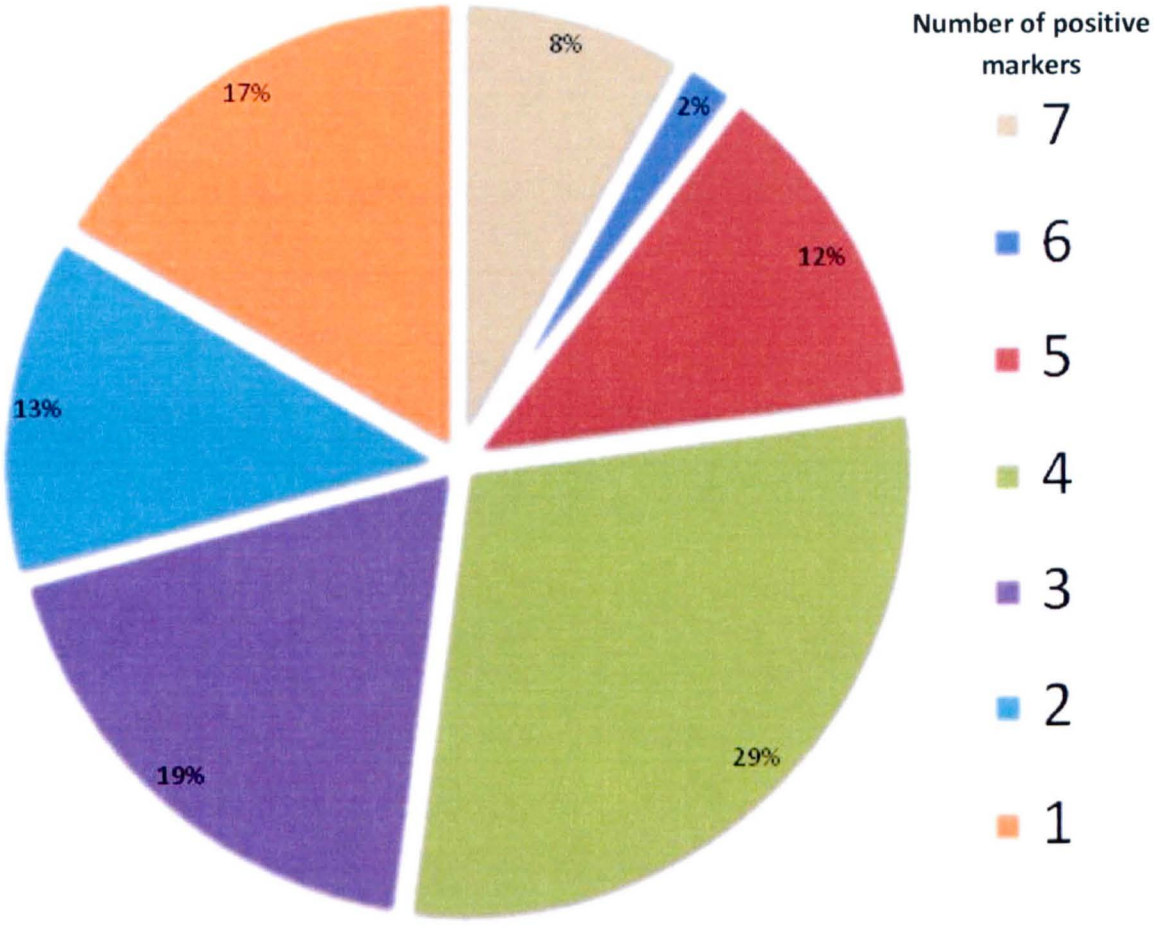


Figure 4.3.1.2. Pie chart demonstrating the distribution of the examined lung cancer tissue samples in relation to the number of positive methylation markers per sample (including only the 10 qualifying markers). The tumour sample coverage of the selected panel is 100%, on the basis of at least one methylated marker.

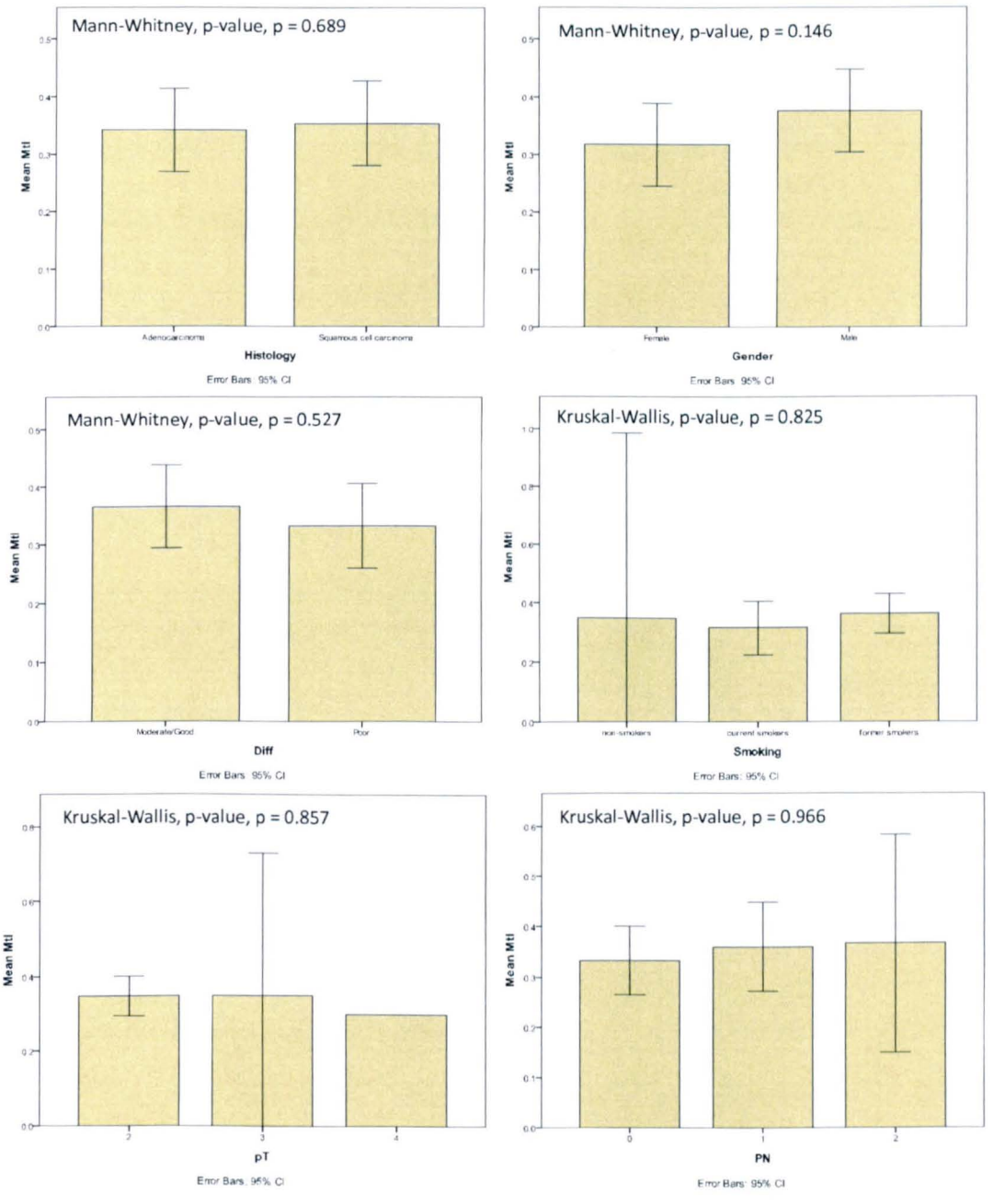


Figure 4.3.1.3. Histograms demonstrating the lack of association between the methylation index (MtI), reflecting the ratio of methylated marker per sample and the clinicopathological parameters such as histological diagnosis, gender, differentiation, smoking, pathological T stage (pT) and nodal metastasis (pN).

4.3.2. Immunohistochemistry for WT1 and p16

IHC was employed to confirm the expected downregulation of the encoded proteins of WT1 and p16 in a subset of lung cancer samples as a response to the high frequency of hypermethylation. Both genes demonstrated significant differences in DNA methylation between tumour and adjacent normal samples. Ten adenocarcinomas and ten squamous cell carcinomas were stained for the two proteins and evaluated by Professor J. Gosney. According to his interpretation, none of the tumours convincingly expressed WT1. Some membranous and/or cytoplasmic labelling was observed but this was considered to be non-specific. The IHC results are presented in Table 4.3.2.1. Characteristic examples of stained specimens are shown in Figure 4.3.2.1.

Table 4.3.2.1. IHC results. 10 adenocarcinomas and 10 squamous carcinomas stained for p16 and WT1 expression.

Sample ID	Diagnosis	WT1	p16
1	Adenocarcinoma	-	-*
2	Adenocarcinoma	-	-
3	Adenocarcinoma	-	-*
4	Adenocarcinoma	-*	-*
5	Adenocarcinoma	-	-
6	Adenocarcinoma	-	-
7	Adenocarcinoma	-	-*
8	Adenocarcinoma	-	-
9	Adenocarcinoma	-*	-
10	Adenocarcinoma	-*	-
11	Squamous Cell Ca	-	-
12	Squamous Cell Ca	-	-*
13	Squamous Cell Ca	-*	-
14	Squamous Cell Ca	-*	-
15	Squamous Cell Ca	-	-
16	Squamous Cell Ca	-	-
17	Squamous Cell Ca	-	++
18	Squamous Cell Ca	-	-
19	Squamous Cell Ca	-	-
20	Squamous Cell Ca	-	+

*Non specific (cytoplasmic) staining present, + weak nuclear staining, ++ strong nuclear staining

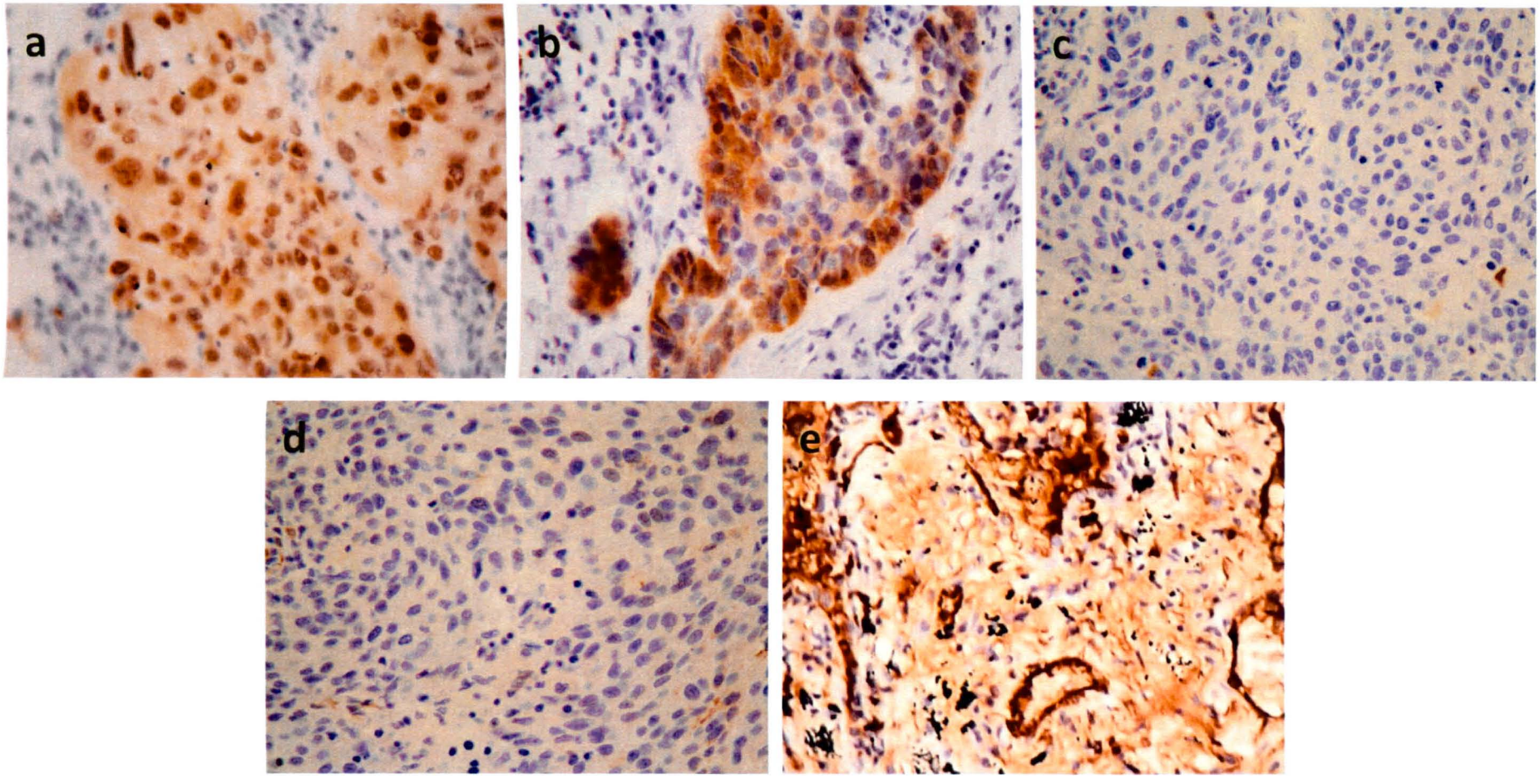


Figure 4.3.2.1. Immunohistochemichal staining a) Squamous cell carcinoma with positive nuclear p16 staining, b) Squamous cell carcinoma with p16 cytoplasmic/non-specific staining c) Squamous cell carcinoma negative for p16, d) Adenocarcinoma negative for WT1 with non-specific staining e) Adenocarcinoma with intense cytoplasmic but negative nuclear staining (magnification 400x)

4.4. Discussion

The discovery process took under consideration two lines of evidence: Data from previous high-throughput experimental analyses (expression and methylation microarrays) and hypothesis-based projects of our group as well as a literature review. The hypermethylation frequencies of the candidates were established by pyrosequencing in an independent group of NSCLC tissue. This step would ensure the selection of biomarkers with high abnormality ratio in the particular population. It is of note that a number of promoters previously reported to be hypermethylated in lung cancer demonstrated low or zero hypermethylation in our study. Examples include MLH1, p15 and ATM. Another gene which was eventually disqualified was MGMT, which demonstrated a relatively low frequency (16%) of hypermethylation in this set. These discrepancies may arise from the different origin of the samples. Ethnic origin as well as different lifetime exposures to environmental or lifestyle carcinogens may account partly for this difference. However, the most probable reason may be the different methodological approach. In particular, many of the early studies cited here employed end point MSP, nested MSP and MSRE. These techniques are well acknowledged for providing high numbers of false positives. In contrast, pyrosequencing is currently considered to be the gold standard technique for this step (Liloglou et al, 2012).

Pyrosequencing was chosen for frequency validation in primary lung cancer tissue while qMSP was the choice for the BWs. The reason for selecting Pyrosequencing as the validation method is that it provides quantitative information for CpG methylation over a long stretch of DNA. It thus combines mapping with quantitation, allowing for a greater resolution in pinpointing the target CpGs of interest. It is a very reliable method with built-in internal controls for the efficiency of bisulphite treatment.

Concerning the tumour sample coverage of the selected panel, it was evident that if we accept positives on the basis of at least one methylated marker, coverage is 100%. However, a more stringent approach of utilising at least two methylated markers for positive designation, coverage drops to 83% (Figure 4.3.1.2.).

It is of note that there was no difference in the frequency of hypermethylation detected in any of the 10 genes between adenocarcinomas and squamous cell carcinomas (Mann-Whitney test, Kruskal-Wallis test) (Figure 4.3.1.3). The clinicopathological characteristics used for the needs of the statistical analysis show no correlation with the methylation status of our target genes, for the construction of our candidate markers panel.

IHC staining for WT1 showed generally membranous and/or cytoplasmic labelling. This is considered to be non-specific. Previous studies have resulted in agreement with the current observation (Hwang et al, 2004; Ordonez) and the currently accepted consensus that WT1 is principally a DNA binding transcription factor mainly distributed in the nucleus (Oji et al, 2002). However, there is a published study using both monoclonal and polyclonal WT1 antibodies and claiming cytoplasmic staining of WT1 as positive (not background) (Nakatsuka et al, 2006). It has also been shown that aberrant cytoplasmic localisation of WT1 might alter the properties of tumour cells through the expressional regulation of variable genes (Ortega et al, 2003). It is clear that additional work is required to clarify this aspect of WT1 expression in lung tumours.

Furthermore, the same set of tissue samples were examined for p16 expression. Only two squamous cell carcinomas convincingly expressed nuclear p16. Sample 17 (Table 4.3.2.1) varies in intensity, but is diffusely expressed across the section. Sample 20 is weaker and more patchy. Another five samples showed some non-specific (cytoplasmic) staining and

were grouped with the remaining which were totally negative. Loss of p16 expression is very common in lung cancer and has been associated with multiple genetic and epigenetic aberrations such as loss of heterozygosity, homozygous deletions, mutations and hypermethylation (Blanco et al, 2007; Sterlacci et al, 2011).

In conclusion, the experimental work in this chapter resulted in the selection of ten promoters, namely p16, CYGB, TMEFF2, p73, WT1, CDH13, DAPK1, TERT, RASSF1 and RARb, which have been successfully validated for their frequency of hypermethylation in primary non-small cell lung carcinomas. These markers form the panel to be tested for their diagnostic efficiency in BWs.

Chapter 5 qMSP Development and Optimisation

5.1. Introduction

This chapter describes the experimental work undertaken to develop sensitive, specific and reproducible assays, enabling the reliable detection of hypermethylation in BWs.

5.1.1. qMSP assays for clinical use.

The high incidence rate of lung cancer, worldwide, directs the need for early diagnosis of the disease. Bronchoscopic examination following suspicious imaging results can reveal the presence of a bronchial lesion which is normally confirmed by biopsy and/or BWs. However, a significant number of cases remain clinically occult after bronchoscopy as cytological examination tends to miss almost half of the cases. Despite our constantly growing understanding of carcinogenesis, there is still an eager needs to design novel tools that can be applied as part of clinical practice (Kulis & Esteller). In the 1990s, the detection of abnormal promoter CpG island DNA hypermethylation emerged as a potential biomarker strategy for assessing cancer risk, early detection, prognosis and predicting therapeutic responses (Laird, 2003). Through the years, many techniques for the detection of DNA methylation have been discovered (Eads et al, 2000; Ehricht et al, 2006; Gonzalgo & Jones, 1997; Herman et al, 1996; Xiong & Laird, 1997). It is obvious that a molecular assay for clinical use must address the particularities of the specific disease and sample type. The particularity in this case is the high excess of normal contaminating DNA.

PCR-based methods that use sodium bisulphite treated DNA as a template are generally accepted as the most analytically sensitive and specific techniques for DNA methylation analysis (Kristensen & Hansen, 2009).

MSP was introduced in 1996 by Herman et al. to facilitate sensitive and specific methylation detection of any block of CpG sites in a CpG island. (Herman et al, 1996). This breakthrough in methylation analysis provided certain advantages comparing to Southern hybridisation approaches and bisulphite sequencing which were used until then. MSP needed only a small fraction of input DNA than Southern analysis and could detect significantly lower numbers of methylated alleles. Moreover, paraffin-embedded samples, which previously were excluded for Southern analysis, became a possible source for DNA methylation analysis. Another obstacle that MSP could bypass was that not only CpG sites that were recognised by methylation-sensitive restriction enzymes were available for analysis. Significant reduction of false positive results was noticed as partial digestion of the target sequence was a drawback in previous PCR approaches. Furthermore, sequencing-based methylation analysis (sometimes including cloning) was very costly in time and labour. Thus MSP provided a low-cost alternative which did not require specialised equipment. The major disadvantages of endpoint MSP are the lack of (a) quantitation ability and (b) internal bisulphite conversion controls and thus a weakness in recognising false positives.

Quantitative methylation specific PCR (qMSP) is the real-time modification of MSP and demonstrates particular advantages; The sensitivity of detection is orders of magnitude higher than its endpoint counterpart due to the use of fluorescence, especially when probes are employed. In addition, it provides the significantly higher specificity and quantitation ability. It is thus currently the method of choice for efficiently detecting methylated DNA copies in the presence of high numbers of unmethylated copies (Eads et al, 2000).

It is of great importance that sensitivity, specificity and reproducibility of qMSP is very well established prior to attempting biomarker analyses in clinical sample sets. Any given

biomarker study should combine a standardised assay with a well-characterized clinical cohort. The assays and the number of markers used should be able to compensate for the heterogeneity of origin of the nucleic acids found in biological fluids as well as the heterogeneity of epigenetic alterations within cancer cells.

5.2. Materials and Methods

5.2.1 Preparation of DNA controls and methylation standards.

Leukocyte DNA was extracted following a *phenol-chloroform* protocol. 950 µl lysis reagent (400 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 1% SDS) and 50 µl Proteinase K (Qiagen) were added on the frozen leukocyteLeukocyte pellet and the lysate was transfered into a 2 ml safelock tube (Eppendorf). Following an overnight incubation at 56 °C in an orbital shaker, 25 µl RNase A (20 mg/ml) was added and incubation continued at 37 °C for 1 h. 20 µl of fresh Proteinase K were added and incubation was continued for a further 2 hours. An equal volume of phenol (Fisher Scientific) was added and the lysates were mixed by inverting the tubes for 2 min. The tubes were centrifuged at 12000 g for 2 min at room temperature and the supernatant aqueous phase was transferred to a clean 2 ml safe-lock tube taking care not to disturb the interphase. Then an equal volume of chloroform was added to following continuous inversion of the tube for 2 min. After a similar centrifugation and transferring supernatant to a fresh tube, 1 ml of absolute isopropanol was added to precipitate the DNA, which was recovered by centrifugation at 14000 g for 15 min at 4 °C. The pellet was washed with 1 ml 70% ethanol (EtOH), span down for 5 min at 14000 g and dried before being resuspended in 200 µl TE pH=8.0. The samples were stored at 4 °C

overnight prior to $OD_{260/280}$ measurement . This was undertaken using a NanoDrop 2000 (Thermo Scientific) and DNA was normalised to 100 ng/μl.

For the preparation of methylated controls, leukocyteleukocyte DNA was *in vitro* methylated using CpG Methyltransferase (M.SssI) (NEB). This enzyme methylates all cytosine residues within the double stranded dinucleotide recognition sequence 5'CG3'. One μg DNA (10 μl) was combined with 5 μl of nuclease free water, 2 μl of 10x NEB buffer, 1 μl S-adenosylmethionine (SAM) and 1 μl (10 U) SssI methylase. After 1 hour incubation at 37 °C, the reaction was stopped by heating at 65 °C for 20 minutes.

For the preparation of serial dilutions leukocyte DNA (50ng/μl) and in vitro methylated leukocyte DNA at the same concentration were used. The following dilutions were tested: 5%, 1%, 0.5%, 0.25%, 0.125% of methylated : unmethylated DNA.

As DNA methylation demonstrates tissue specificity, it cannot be assumed that any given promoter is unmethylated in the various subpopulations of WBC. Thus, whole genome amplified DNA has been used as unmethylated technical control. The Repli-G screening kit (Qiagen) has been used for this purpose. Genomic DNA (100ng in 3μl) was mixed with 17μl SB1 buffer. After mixing using a vortex, the solution has been placed in a centrifuge for a brief spin. The tube has been place for 5 min at 65 °C. It is critical to leave the mixture to cool down in room temperature as the Repli-g Mini DNA Polymerase is thawing on ice. The preparation of master mix for the reaction described by the protocol contains 17 μl of SB2 buffer and 1 μl of Repli-g Mini DNA Polymerase. Thus, 18 μl of the master mix were added to 20 μl of denatured DNA. The mixture needed 16 hours incubation for maximum yield of DNA to be achieved. Inactivation of the polymerase was achieved by 3 minute incubation at 65°C.

5.2.2. Development of Quantitative Methylation Specific PCR (qMSP) assays

The qMSP assays were designed to specifically amplify bisulphite-converted methylated DNA target sequences in the presence of an excess of unmethylated counterpart sequences. Taqman technology uses fluorescence resonance energy transfer (FRET) to quantify *Taq* polymerase-based 5'–>3' exonuclease (displacement) activity on DNA primed-DNA substrates. Sequence-specific primers and an intervening probe are designed to cover an amplicon of approximately 100 bp in length. Besides increasing the specificity of the actual PCR, the probe is labelled with a fluorescent reporter dye on the 5' end and a quencher on the 3' end (Shames et al, 2007b). The principle of the qMSP approach that we used is illustrated in Figure 5.2.2.1.

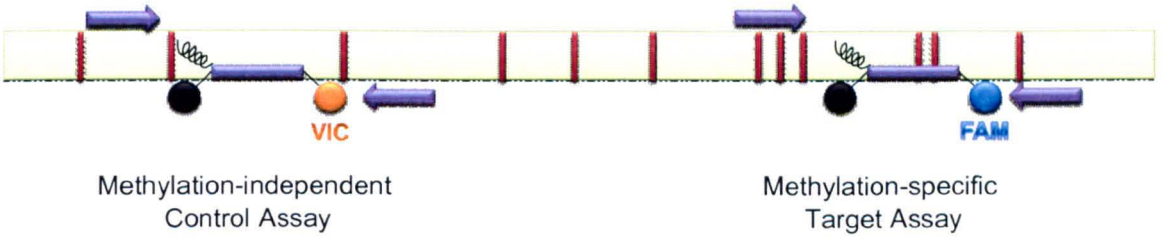


Figure 5.2.2.1. Multiplex qMSP primer/probe design principle. Target assays are designed on CpG groups with the 3'-end of the primer containing one or more CpGs. The MGB-probe is labelled with FAM which is the dye with the highest fluorescence levels at the particular excitation wavelength that AB platforms use, providing thus increased detection sensitivity. The control assay is designed on a CpG-free region providing unbiased amplification from both methylated and unmethylated DNA. The control probe is labelled with VIC. Purple arrows represent the primers; purple bars represent the MGB-probes with circles being indicative of the fluorophores (orange, blue) and the quenchers (black). Red vertical bars represent CpG positions

The methylation-specific primer and probe sequences are listed in Table 5.3.1. In the initial steps of assay development it became apparent that probes bearing minor groove binding moiety (Taqman MGB probes) provided significantly higher assay specificity. In addition, due to their smaller size, they allow for a more flexible assay design.

Table 5.2.2.1. Nucleotide sequences of methylation specific primers and probes for the qMSP assays utilised in the BW screening. The ACTB assay is methylation-independent acting as DNA input control.

Primer/probe name	Sequence 5'→ 3'	Modification
p16meth_F	GGAGGGGGTTTTTTCGTTAGTATC	
p16meth_R	CTACCTACTCTCCCCCTCTCCG	
p16meth_P	AACGCACGCGATCC	FAM-MGB
RASSF1meth_F	GTGGTGTTTTGCGGTCGTC	
RASSF1meth_R	AACTAAACGCGCTCTCGCA	
RASSF1_P	CGTTGTGGTCGTTCG	FAM-MGB
TMEFF2meth_F	GGAGAGTTAAGGCGTTTCGTAGTTC	
TMEFF2meth_R	CGTGGGAAGAGGTAGTCGGG	
TMEFF2meth_P	GTTTTTAGTTCGTTCG	FAM-MGB
TERTmeth_F	TTGGGAGTTCGGTTTGGTTTC	
TERTmeth_R	CACCCTAAAAACGCGAACGA	
TERTmeth_P	AGCGTAGTTGTTTCGG	FAM-MGB
CYGBmeth_F	GTGTAATTTCGTCGTGGTTTGC	
CYGBmeth_R	CCGACAAAATAAAAACTACGCG	
CYGBmeth_P	TGGGCGGGCGGTAG	FAM-MGB
RARbmeth_F	GATTGGGATGTCGAGAACGC	
RARbmeth_R	ACTTACAAAAAACCTTCCGAATACG	
RARbmeth_P	AGCGATTCGAGTAGGGT	FAM-MGB
DAPK1meth_F	CGAGCGTCGCGTAGAATTC	
DAPK1meth_R	ACCCTACAAACGAACTAACGACG	
DAPK1meth_P	AGCGTCGGTTTGGTAG	FAM-MGB
p73meth_F	TTGTTTTTTGGATTTTAAGCGTTTC	
p73meth_R	CACCCGAATCTCTCCTAACCG	
p73meth_P	TAACGCTAAACTCCTCG	FAM-MGB
WT1meth_F	GAGGAGTTAGGAGGTTCGGTC	
WT1meth_R	CACCCCAACTACGAAAACG	
WT1meth_P	AGTTCGGTTAGGTAGC	FAM-MGB
CDH13meth_F	CGTGTATGAATGAAAACGTCGTC	
CDH13meth_R	CACAAAACGAACGAAATTCTCG	
CDH13meth_P	CGTTTTTAGTCGGATAAAA	FAM-MGB
ACTBmgb_F	GGGTGGTGATGGAGGAGGTT	
ACTBmgb_R	TAACCACCACCCAACACACAAT	
ACTBmgb_P	TGGATTGTGAATTTGTGTTTG	VIC-MGB

The optimization process was long and focused on establishing the optimal primer/probe concentrations as well as thermal profiles to ensure maximum sensitivity, specificity and reproducibility of the assay. The qMSP reactions contained 1× TaqMan® Universal Master Mix II (Applied Biosystems) 250 nM probe, 300-900 nM primers (Table 5.2.2.2) and 2 µl eluate from the bisulphate treated DNA sample. The reactions were performed on a 7500 FAST real time cycler (Applied Biosystems) under the following thermal profile: 95 °C for 10 min and 50 cycles with time intervals of 95oC for 15 sec, 58 °C - 65 °C for 1 min (Table 5.2.2.3).

Table 5.2.2.2. Thermal profile for qMSP reactions.

Primer/probe mix	Fwd primer (nM in reaction)	Rev Primer (nM in reaction)	Probe (nM in reaction)
p16	700	700	250
TERT	250	250	250
RASSF1	700	700	250
TMEFF2	900	900	250
CYGB	300	300	250
RARb	500	500	250
DAPK1	250	250	250
p73	250	250	250
WT1	750	750	250
CDH13	250	250	250
ACTb	900	900	250

Table 5.2.2.3. Annealing information for qMSP optimised conditions.

Genes	Annealing temp (°C)	Time (sec)
p16	60	60
RASSF1	60	60
CYGB	64	5
	61	50
RARβ	65	5
	62	50
TERT	65	5
	62.5	50
WT1	62	60
ACTβ	58	20
	60	40
CDH13	64	5
	61	50
DAPK	65	5
	62.5	50
P73	65	5
	62.5	50
TMEFF	58	20
	60	40

5.3. Results from optimisation reactions

The sensitivity and specificity of the assays was tested on serial dilutions of artificially (SssI) methylated DNA in leukocyte DNA. This is because the bronchoalveolar lavages (BALs) can contain blood traces (sometimes a significant amount), thus leukocyte DNA is a "contaminating" source of unmethylated DNA. However, as not all genes are unmethylated in white blood cells, WGA DNA was constructed (Qiagen REPLI-g Screening kit) as a technical unmethylated DNA standard (Figure 5.3.1). Following multiple repetitions the sensitivity threshold was selected to 0.5% (1:200) as it provided total reproducibility, while higher dilutions (0.1%) proved less reliable (Figure 5.3.1). A methylation-independent assay with non-CpG bearing primers/probe was designed for the ACTB gene in order to normalize for input DNA, but also to be used as an exclusion criterion. We experimentally established that the cycle threshold (Ct) for the ACTB assay corresponding to 1000 diploid genomes (6.9 ng DNA per assay) was equal to 29. The latter cut-off was employed to ensure 5× coverage of the 1:200 sensitivity threshold.

It was important to set the threshold of reliable detection by performing multiple repetitions. It is also of note that sensitivity was expressed in relation to the dilution of the methylated to leukocyte DNA and not water. Dilution in water would be in this case clearly an invalid standard for such assays as it dilutes the target without increasing the competition by unmethylated counterpart copies.

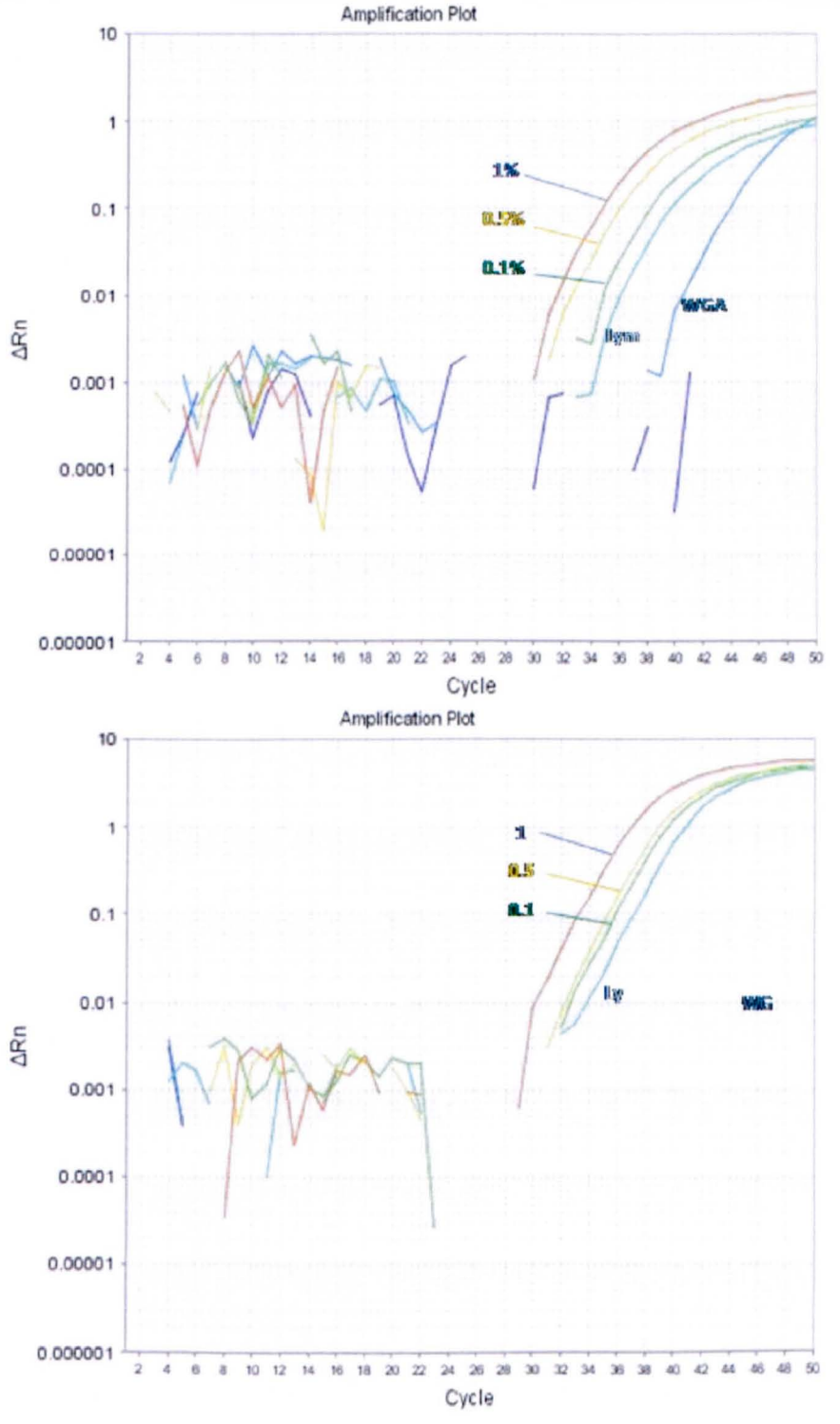


Figure 5.3.1. qMSP assay that includes standard curve of 1% to WGA DNA. Top diagram demonstrates suboptimal primer or/and reaction conditions. In the lower diagram the technical control that serves as experimental checkpoint shows no amplification where at the same time our standard curve DNA is amplified as expected.

Examples of triplicate reactions for the p16 and RASSF1 assays are shown in Figure 5.3.3. Two observations are the most import ones in there; firstly, the ΔCt response of the assay to the dilution of methylated: unmethylated DNA, is not linear. Secondly, the reproducibility drops gradually. We have additionally performed assays on target at higher dilution to 1:1000 and we rarely get signal from all 3 repetitions. Considering that these experiment were performed on control DNA, (i.e. high molecular weight, high purity), one can speculate that DNA from samples such as lavage and sputum, will certainly cause a higher variability. To secure the quantity of the copies, we have taken a conservative approach and set our threshold to 5:1000 or 0.5%, where we get totally reproducible results.

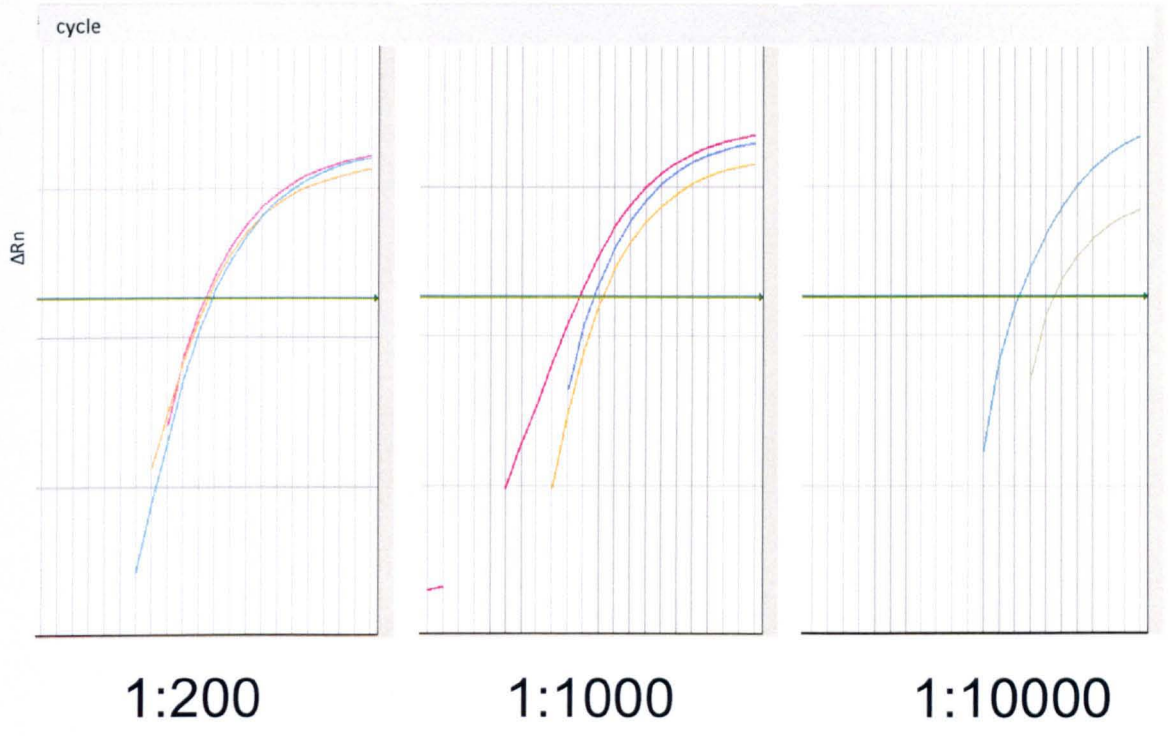


Figure 5.3.2. Experiments using gradually reduced DNA input amounts indicate the corresponding reproducibility. The reaction is extremely reproducible at the 1:200 and 1:1000 dilution with water but loses reproducibility at the 1:10000 dilution as the reaction reaches the region of stochastic amplification.

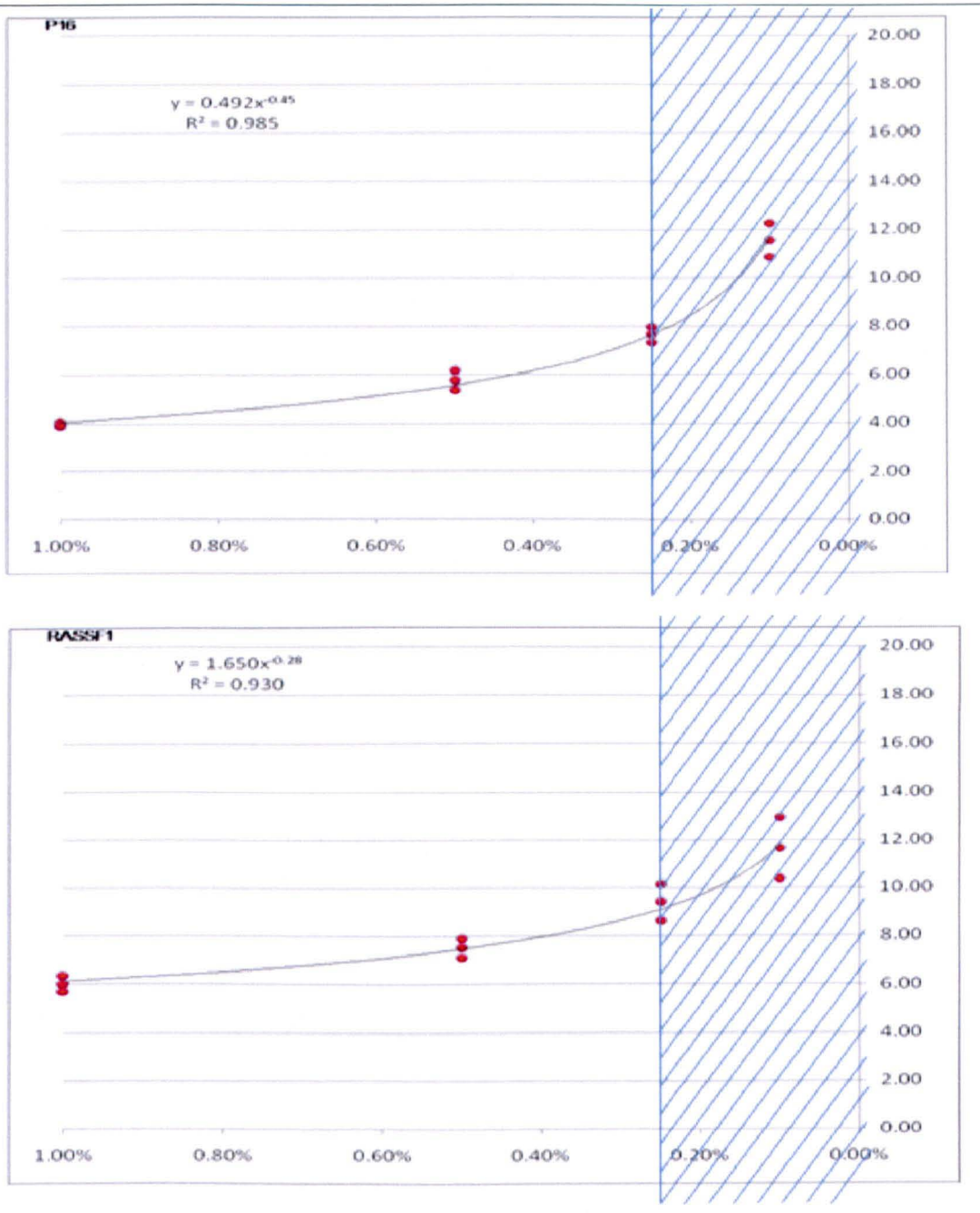


Figure 5.3.3. Reproducibility assays for target genes (p16 and RASSF1). The X axis represents the percentage of methylated DNA input, while the Y axis is the ΔCt (Target Ct- ACTB Ct). The reproducibility of both assays is reduced in response to dilution with unmethylated DNA. It is obvious that the response is not linear. The reproducibility of the assays in the shaded area would not be adequate for clinical use as it includes differences up to 3 ΔCt (= 8 -fold). These curves were used to set the accepted threshold of sensitivity without loss of reproducibility for the subsequent use of the assays in clinical samples.

5.4. Discussion regarding optimisation issues

Although, many studies have been undertaken to satisfy the hypothesis that hypermethylation detection is a promising marker, issues regarding assay sensitivity and specificity have been raised (Mulero-Navarro & Esteller, 2008). As the quantity of DNA in peripheral fluids is limited, the robustness of the assay is another important factor. In this study, it was decided to utilise different techniques for the different validation phases.

For the BWs validation phase(s), the use of some form of real-time PCR based assay is imperative. This is because as the hypermethylated promoter targets are expected to be present at very high dilution because of the high abundance of normal lung and blood cells in the specimens. Thus the target must be specifically amplified from a pool of DNA very rich in unmethylated targets. qMSP was selected for this phase as it has the sensitivity to pick very low copy number targets. During the course of optimisation we realised that the use of MGB probes greatly improves specificity (i.e. preferential amplification of methylated target) and facilitates probe design due to the small length requirements. The initial idea to perform multiplex target: control reactions, which would better correct for pipetting errors, was modified as multiplexing demonstrated loss of target detection sensitivity. This was not surprising; the abundant PCR target (in this case methylation-independent ACTB, present in 2 copies per genome) utilises the PCR resources eliminating the methylated target signal when below a specific threshold. In order overcome this obstacle, separate reactions were performed for each target, but DNA samples and PCR reagents (except primer/probes) were premixed and aliquoted, in order to reduce pipetting errors and to ensure a uniform spread of DNA input between target and control reactions. An additional quality control measure introduced was to ensure that the minimum amount of DNA input provided adequate

genome coverage, based on the 1:200 sensitivity threshold. This not only enhanced reproducibility but also served as sample an inclusion criterion.

Summarising this chapter, the developed qMSP assays were optimised using multiple technical controls and known positive/negative samples to ensure the maximum possible the sensitivity, specificity and reproducibility of detection. The results clearly indicate that these assays are robust enough to screen clinical samples with high reliability.

Chapter 6 DNA Methylation Panel Validation in the LLP Bronchial Washings sample set.

6.1 Introduction

As discussed in chapter 3, there has been a long-standing requirement for molecular biomarkers for application in BWs, in order to assist clinical diagnosis of lung cancer, has been a long-standing demand. Previous attempts to detect known molecular abnormalities in lung cancer included genomic instability (Ahrendt et al, 2001; Liloglou et al, 2001), DNA mutations, (Ahrendt et al, 1999) and more recently, DNA methylation (Schneider et al, 2011; de Fraipont et al, 2005). The latter has certain advantages regarding its biomarker applicability; it is a covalent modification resistant to post-sampling processing and stretches over a significant length allowing for flexible assay design.

The feasibility of DNA methylation detection in the BW of lung cancer patients has been demonstrated by a number of studies (Grote et al, 2006; Schmiemann et al, 2005) also reviewed in (Liloglou et al, 2012). However, few of the proposed biomarkers were further validated to date. One such validated biomarker that has recently received *Conformité Européenne In vitro* diagnostics (CE IVD) certification under the commercial name of Epi proLung® BL Reflex Assay (Epigenomics, AG) is *m*SHOX2 (Kneip et al, 2011).

This chapter describes the validation of the panel of DNA methylation biomarkers selected through lung tumour screening (Chapter 4) with assays developed for clinical use (chapter 5) in a large retrospective case-control BWs set from the Liverpool Lung Project. The study

design took under consideration the guidelines for biomarker studies issued by the National Cancer Institute - Early Detection Research Network (NCI-EDRN) (Pepe et al, 2001) as well as the CRUK Diagnostic roadmap (Figure 6.1.1).

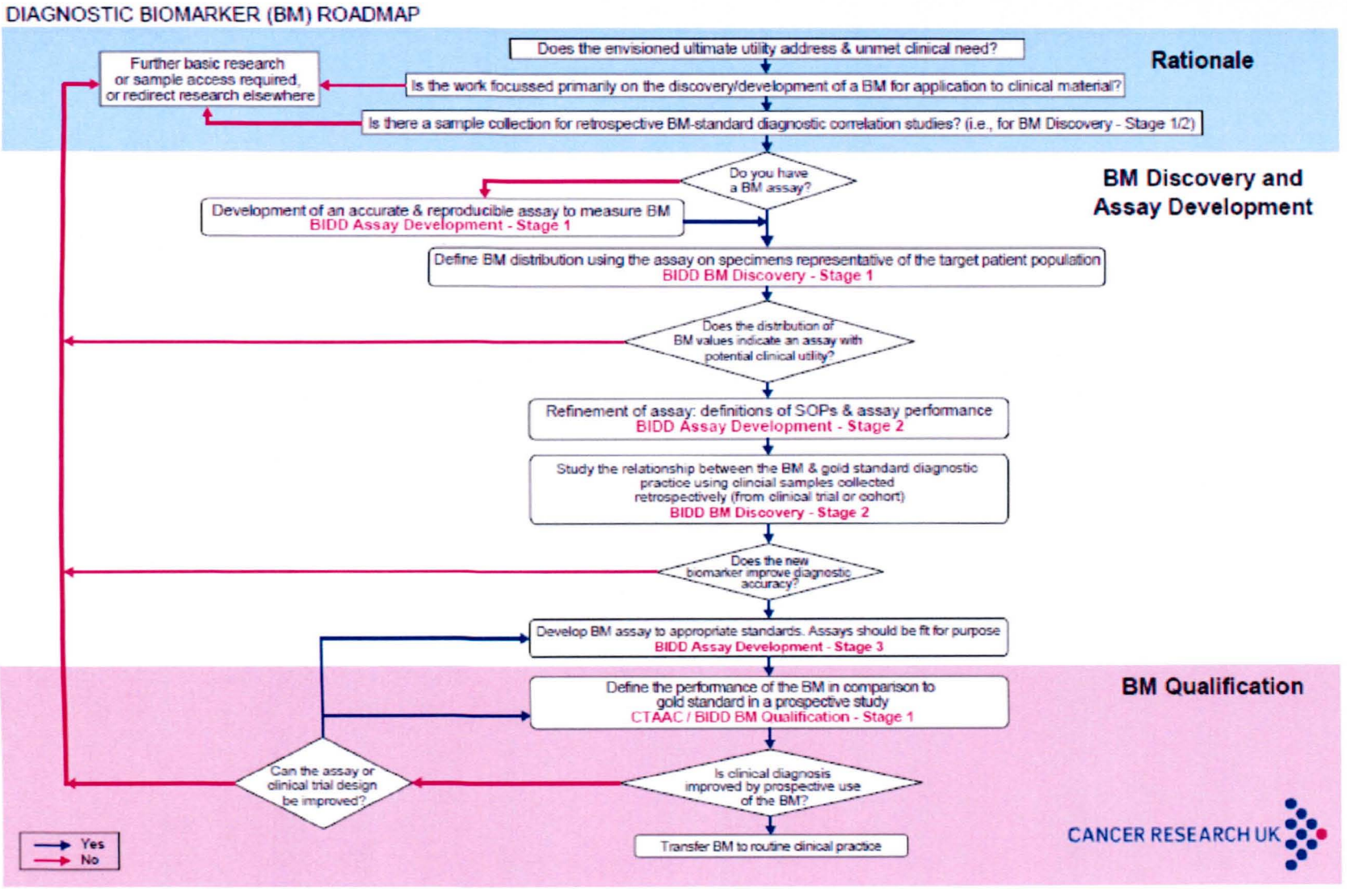


Figure 6.1.1. Cancer Research UK Diagnostic Roadmap indicating the important steps and decision points in biomarker discovery and validation (Adapted from CR-UK web site).

6.2 Materials and Methods

6.2.1 Patients and samples

A total of 655 individuals (333 lung cancer cases / 322 age/sex-matched controls) were included in the study (Table 6.2.1.1). All patients were recruited through the Liverpool Heart & Chest Hospital under the LLP umbrella. LLP has received ethical approval (LREC 97/141) and all the recruited patients provided informed consent.

Specimens were excluded if extracted DNA failed in quality control (see below in the qMSP description section). During the length of the study a number of control individuals developed lung cancer and were transferred to a "case" status. The case-control distributions of epidemiological and clinical characteristics for subjects in the training and testing datasets are shown in Table 6.2.1.1 demonstrating overall similar patterns between the two classes, with the exception of smoking.

BWs were stored in Saccomanno's fixative in an air-conditioned (18°C) room. Specimens' cytological adequacy was judged by the presence of alveolar macrophages.

Two ml of BWs were transferred into safelock tubes (Eppendorf) and 15 µl of 15% DTT was added. After 30 minutes incubation at room temperature, samples were centrifuged for 15 minutes at 14000 g, at 4 °C. DNA from 2 ml BWs was extracted using the Blood and Tissue kit (Qiagen). The ATL buffer was visually inspected to ensure there was no precipitate. If precipitate was present, incubation for 10 min at 55°C was appropriate to dissolve the precipitate. 180 µl ATL and 20 µl Proteinase K per sample were premixed and 200 µl were added in each sample.

Table 6.2.1.1. Frequency distribution of subjects' epidemiological & clinical characteristics

Subject characteristics	Training set (N=407)		Testing set (N=248)	
	Case (n=194)	Control (n=213)	Case (n=139)	Control (n=109)
Age group[↑]				
<60	33 (17.0)	57 (26.8)	18 (13.0)	19 (17.4)
60-79	150 (77.3)	144 (67.6)	110 (79.1)	84 (77.1)
80+	11 (5.7)	12 (5.6)	11 (7.9)	6 (5.5)
Age summary statistic[↑]				
mean ± sd	68.7±7.56	66.4± 8.56	68.4±8.07	67.6±8.78
Gender				
Male	114 (58.8)	115 (54.0)	80 (57.6)	63 (57.8)
Female	80 (41.2)	98 (46.0)	59 (42.5)	46 (42.2)
Smoking status[*]				
None smoker	8 (4.1)	40 (18.8)	4 (2.9)	25 (22.9)
Ex-smoker	103 (53.1)	91 (42.7)	63 (45.3)	65 (59.6)
Current smoker	74 (38.1)	42 (19.7)	72 (51.8)	18 (16.5)
Unknown	9 (4.6)	40 (18.8)	0 (0.0)	1 (0.9)
Summary of:				
Smoking duration[b]				
mean ±sd	44.7±12.06	39.0±13.73	43.9±13.14	34.6±14.58
median	46	41	45	37
Smoking pack years[¶]				
mean ±sd	45.0±26.93	42.4±29.66	50.7±34.54	32.0±19.82
median	42.1	39.4	45	28
Cytology[*‡]				
Negative	113 (58.3)	213 (100.0)	76 (54.7)	108 (99.1)
Positive	67 (34.5)	0 (0.0)	46 (33.1)	0 (0.0)
Suspicious	14 (7.2)	0 (0.0)	17 (12.2)	1 (0.9)
Histology Diagnosis				
Others[a]	3 (1.6)		20 (14.4)	
Large cell carcinoma	25 (12.9)		16 (11.5)	
Small cell carcinoma	4 (2.0)		39 (28.1)	
Squamous cell carcinoma	91 (46.9)		31 (22.3)	
Adenocarcinoma	68 (35.0)		22 (15.8)	
Unknown	3 (1.6)		11 (7.9)	
Sample duration (yrs) ‡				
<5	75 (38.7)	96 (45.1)	10 (7.2)	39 (35.8)
5+	119 (61.3)	117 (54.9)	129 (92.8)	70 (64.2)

[↑] boarderline significant in training set, * Statistically significant in training set ($p<0.05$)
‡ Statistically significant in testing set ($p<0.05$), ¶ Statistically significant in testing set with p-value from Mann-Whitney test. [b] Statistically significant in both dataset with p-value from Mann-Whitney test, [a] Others (adenocarcinoid, adenosquamous, Carcinoid, Carcinoma, NOS, Neoplasm, malignant, Tumour cells, malignant, Basal cell carcinoma)

Overnight incubation at 56°C with constant agitation took place. The tubes were centrifuged briefly, to collect lysate from the caps. Proper lysis of the samples was confirmed by visual inspection. Partially lysed samples were left for additional 6 hr incubation with 20 µl fresh Proteinase K. Subsequently, 410 µl of AL-ethanol (1:1 mix) buffer were added. After pulse-vortexing the tubes, to homogenize the mixture, a quick spin were needed to collect drops from the caps. The mixture was carefully applied to the QIAamp 96 well spin column plate, covered by AirPore tapes and spun for 10 min at 3,900 g. The flow-through lysates were stored in the fridge as a backup. According to protocol, 500 µl buffer AW1 were added on the columns and the plates were spun at 3,900 g for 15 min. The addition of 500 µl buffer AW2 was followed by another spin at 3,900 g for 15 min. The plate was transferred onto an elution rack and incubated at 70°C for 10 min to eliminate any potential ethanol traces. After incubation, 200 µl Buffer AE or 0.1×TE (pre-warmed at 50°C) were added and plates were incubated at room temperature for 5 min. DNA was recovered by centrifugation at 3,900 g for 2 min and stored at -20°C. DNA was quantified using Picogreen (Invitrogen) in a TECAN GENios Microplate Reader. Samples were normalized at 1 µg in 20 µl and bisulphite treatment was carried out using ZymoResearch 96-well Gold as previously described (Chapter 4).

6.2.3 Study Size and Power Calculations

Assuming a minimum of 87% positives for at least two markers in the lung cancer tissue set seen in chapter 4 (null hypothesis, TPR0=0.87) and an anticipated sensitivity of 95% for the markers combination (alternative hypothesis, TPR1 =0.95) we deduce power associated with different sample sizes, case-control ratios and acceptable false positive rates in a simulation

study (Janes & Pepe, 2006) as shown in Table 6.2.3.1. This indicated that a set of ≥200 cases is required in a 1:1 ratio with controls to achieve 86% power for a 5% false positive rate at the 95% confidence level.

Table 6.2.3.1. Statistical Power Simulation for the estimating the bronchial washing study size is presented. It was decided that the study will utilize 200 cases in a 1:1 case/control ratio design which provides 84% power at 5% false positive rate (bold letters)

No of cases	*Case: control ratio*	FPR_0	
		0.05	*0.10*
50	*0.5: 1*	*0.15*	*0.19*
	1:1	*0.14*	*0.27*
	1:2	*0.17*	*0.38*
100	*0.5:1*	*0.35*	*0.57*
	1:1	*0.49*	*0.77*
	1:2	*0.67*	*0.84*
138	*0.5:1*	*0.53*	*0.76*
	1:1	*0.61*	*0.92*
	1:1.75	*0.77*	*0.94*
	1:2	*0.84*	*0.94*
150	*0.5:1*	*0.49*	*0.85*
	1:1	*0.71*	*0.92*
	1:2	*0.88*	*0.95*
200	*0.5:1*	*0.64*	*0.94*
	1:1	**0.84**	**0.98**
	1:2	*0.96*	*0.98*

6.3 Results

As also pointed out in table 6.2.1.1 the case and control groups did not present significant demographic characteristics. The age range of cases was within the expected/published values for lung cancer and the control arm was matched (Figure 6.3.1).

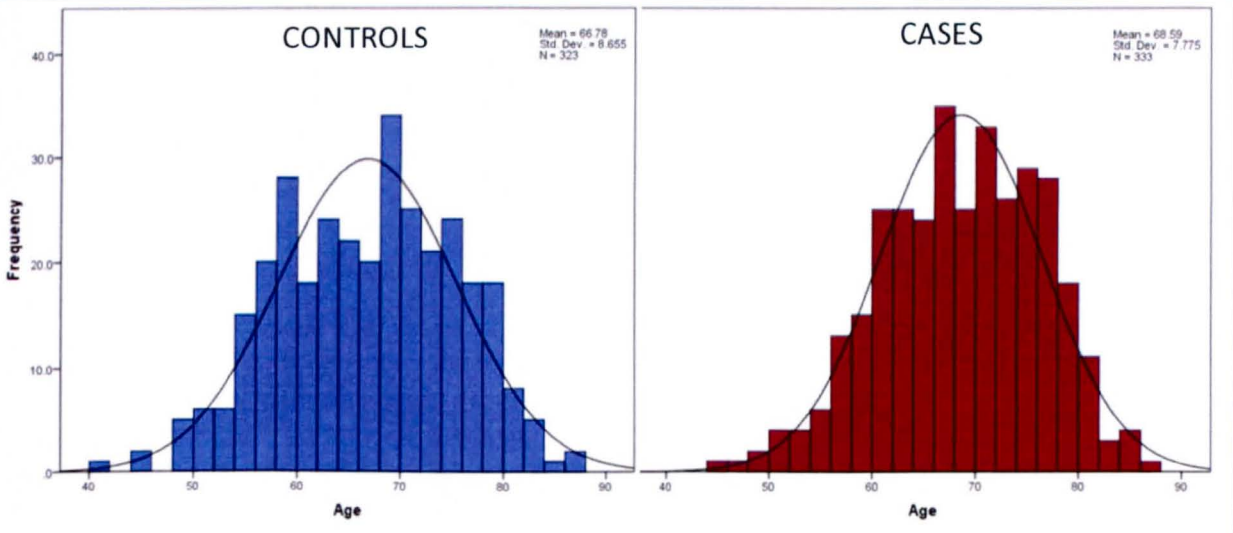


Figure 6.3.1. Age range of controls and cases utilised in this study. They both demonstrate a fairly normal distribution.

The 10 markers that qualified through the lung cancer tissue set (see chapter 4), i.e. TERT, RASSF1, WT1, p16, CYGB, RARβ, p73, DAPK, CDH13 and TMEFF were used to screen the training BW set. The detailed results are shown in Table 6.3.1 while the distribution of positives for each marker among cases and controls is depicted in Figure 6.3.2.

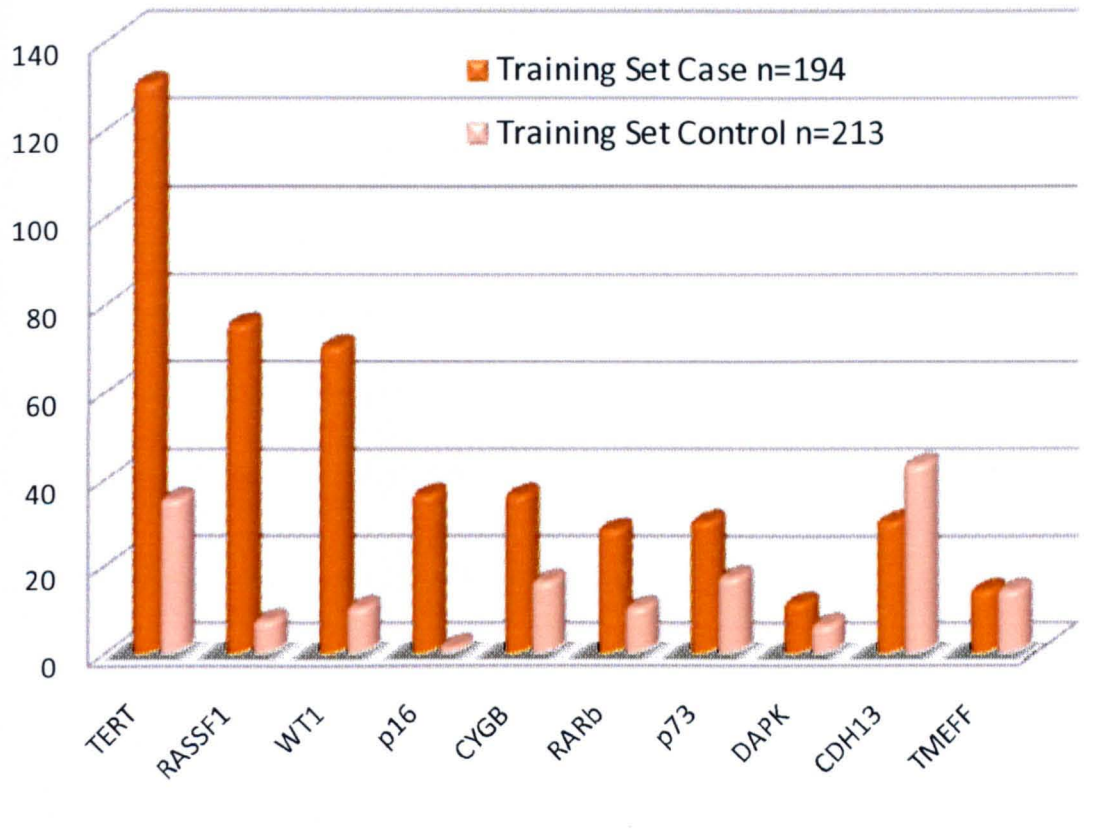


Figure 6.3.2. DNA methylation detection of the candidate biomarkers in the BWs training set. The Y axis represents numbers of positive subjects. Dark orange indicates DNA methylation positives in cases while light orange shows DNA methylation in lung cancer - free controls.

The training set served two purposes. Firstly, to assess the individual marker discriminating power between cases and controls and select the most specific ones for the subsequent screening the validation set. Secondly, to apply different statistical models and produce, in a self-training process, the optimal discriminating algorithm(s). Three statistical models combining markers in a panel were tested in order to identify an optimal algorithm for improved diagnostic efficiency. All of these pointed to six markers (CYGB, p16, RASSF1, TERT, RARβ and WT1).

The validation set was screened with these six markers. Detailed data for both sets are presented in table 6.3.1. The diagnostic accuracy, measured by Univariate association test, is presented for each independent marker for the two sets. The ROC analysis (AUC %) for the training set was calculated. The associations between the six markers with disease status in the training set were replicated in the validation data. The discriminatory ability and predictive accuracy of the top six markers (based on significant univariate associations in the training data) are also displayed in this table.

Three statistical models were employed to deliver optimal algorithms:

(a) Top six markers univariate model.

(b) Best Subset Regression (BSR).

(c) Multifactor Dimensionality Reduction (MDR).

Univariate association test

Each individual marker showed poor sensitivity and limited discriminatory ability in the training data except for TERT with modest sensitivity (67%) but is among markers with weak independent specificity. Overall, modest discriminations were only observed in TERT (AUC=0.75), RASSF1 (AUC=0.68) and WT1 (AUC=0.66). Also, only TERT, RASSF1, WT1 and p16 predicted disease class accurately in more than two-third subjects in the training data; the lowest accuracy of 52% was seen in CDH13 and TMEFF. Overall, all the six markers with good discriminatory and accuracy performances in the training dataset also had good predictions in the test dataset with slightly improved AUC and/or predictive accuracy. Only CYGB and p16 had reduction in performance for prediction in the validation data.

Marker combination by Best Subset Regression (BSR)

The BSR model demonstrated that 5 out of 6 markers (all but CYGB) with significant independent associations with the disease status were prominent in each of the best logit markers combination. The logit regression coefficients of markers by fitted models are shown in Table 6.3.2. The BICq and Cross-Validation criteria selected only four markers (p16, RASSF1, TERT and WT1) which had shown the most significant independent association with the disease status.

Markers combination by Multifactor Dimensionality Reduction (MDR)

The MDR suggested a 3-marker combination including TERT, WT1 and p16 as the best model for predicting subject's disease status (Table 6.3.3). This combination has the highest internal testing subset's balanced accuracy of 79.2% and was selected as the best in 10 out of 10 cross validations in the training dataset (Table 6.3.1). The results of potential 2- and 3-way interaction effects of the markers examined through the MB-MDR are shown in Table 6.3.4. The interaction of TERT, WT1 and p16 markers has the strongest association to disease status ($p < 0.001$). Also, the top most significant associations for 2-way interactions were observed among the three markers with interaction of WT1 and TERT having the strongest association followed by that of TERT and p16.

Table 6.3.1. Diagnostic accuracy of training and validation set.

Training Set

Markers	Positives		X^2	Model-based classification*	
	Case n=194	Control n=213	p-value	Accuracy (%)	AUC (95% CI)
TERT	130	35	$<10^{-4}$	75.7	0.75 (0.71, 0.79)
RASSF1	75	7	$<10^{-4}$	69	0.68 (0.64, 0.71)
WT1	70	10	$<10^{-4}$	67.1	0.66 (0.62, 0.69)
p16	36	1	$<10^{-4}$	60.9	0.59 (0.56, 0.62)
CYGB	36	16	$<10^{-3}$	57.3	0.56 (0.52, 0.59)
RARb	28	10	10^{-3}	56.8	0.55 (0.52, 0.58)
p73	30	17	0.08	53.8	0.52 (0.49, 0.55)
DAPK	11	6	0.15	53.6	0.51 (0.50, 0.53)
CDH13	30	43	0.22	52.3	0.52 (0.49, 0.56)
TMEFF	14	14	0.8	52.3	0.50 (0.48, 0.53)

Validation Set

Markers	Positives		X^2	Prediction using trained univariate logit model*	
	Case (n=139)	Control (n=109)	p-value	Accuracy (%)	AUC (95% CI)
TERT	75	2	$<10^{-3}$	73.4	0.76 (0.72, 0.80)
RASSF1	71	0	$<10^{-4}$	72.6	0.76 (0.71, 0.80)
WT1	73	8	$<10^{-3}$	70.2	0.73 (0.68, 0.77)
p16	18	0	$<10^{-4}$	51.2	0.57 (0.54, 0.59)
CYGB	15	0	$<10^{-4}$	50	0.55 (0.53, 0.58)
RARb	67	18	$<10^{-4}$	63.7	0.66 (0.60, 0.71)

* Disease class prediction based predicted $Pr(D) \geq 0.5$

Table 6.3.2. Coefficient and classification performance of top 6 univariate and best subset logit models in the training dataset.

Markers	Top 6 univariate markers	Best subset logit model		MDR markers
		AIC, BIC	BICq, CV {K=10, t = c(100, 1000)}	
Constant	-1.57	-1.46	-1.49	-1.37
CYGB	-0.16	-	-	-
p16	3.04	2.97	2.95	3.11
RASSF1	2.06	2.05	2.03	-
TERT	1.69	1.78	1.68	2.04
RARb	1.13	1.48	-	-
WT1	1.94	1.92	1.96	2.07

Table 6.3.3. Comparison of internal classification and prediction accuracies and cross-validation consistency of Best multi-marker MDR models identified using the training dataset.

Level	Marker	Classification Accuracy	Prediction accuracy	CV Consistency
1	TERT	75.29	75.64	10
2	TERT, WT1	77.80	78.01	9
3*	TERT, WT1, p16	78.83	79.24	10
4	TERT, WT1, p16, CDH13	79.39	78.87	5
5	TERT, WT1, p16, CDH13, DAPK	80.87	69.69	2
6	TERT, WT1, RASSF1, DAPK, CDH13, CYGB	80.72	76.02	2

* Overall best MDR combination

Table 6.3.4. MB-MDR top 3 associations for 3-and 2-way epistasis interactions among methylation markers.

Interactions	Markers	Risk combination	Number	P-value*
3-way interaction	WT1, TERT, p16	High risk Low risk Indeterminate	5 1 2	< 0.001
	WT1, TERT, RASSF1	High risk Low risk Indeterminate	4 1 3	< 0.001
	WT1, TERT,DAPK	High risk Low risk Indeterminate	3 1 4	< 0.001
2-way interaction	WT1, TERT	High risk Low risk Indeterminate	3 1 0	< 0.001
	TERT, p16	High risk Low risk Indeterminate	3 1 0	< 0.001
	TERT, RASSF1	High risk Low risk Indeterminate	2 1 1	< 0.001

Performance evaluations of discriminatory algorithms

The performance of the different discriminatory algorithms in training and validation data is shown in Table 6.3.5. All the logit discriminatory algorithms performed reasonably well in the training set. The performance of the top 6 univariate markers and the best subset with BICq or CV in the test data was similar, although the best subset algorithm was more sensitive but less specific in the training data. The MDR algorithm was slightly more specific but less sensitive than the best subset model with BICq or CV criteria in the training data, its performance in the test data was only similar to that of the best subset in terms of specificity (sp=0.98) and lower with regards to sensitivity (se=0.77). The addition of the top MB-MDR 2- and 3-way interactions into any of the best logit models did not alter their performance. Overall, the best subset logit model with BICq or CV criteria including TERT,

WT1, p16 and RASSF1, is the most parsimonious and best performed algorithm; thus selected for further evaluation.

The discriminatory performance of the "overall" best algorithm by clinical characteristics is shown in Table 6.3.5. This clinical stratification revealed that apart from cytology and histological diagnosis, the discriminatory algorithm showed similar AUCs in each level of age, gender, smoking status and time between specimen collection and diagnosis. The AUC value was, as expected, significantly higher for cytology positive compared to cytology negative subjects. The algorithm performed well in the different histological subtypes; however, discriminatory performance among small cell carcinoma subjects was significantly higher. The lowest performance was observed among patients with adenocarcinoma.

Table 6.3.5. Performance of best discriminatory algorithm by epidemiologic and clinical characteristics.

Clinical characteristics	% se (sp)	AUC (95% CI)	p- values*			
			vs. level 1	vs. level 2	vs. level 3	vs. level 4
Age						
<60	78.1 (89.5)	0.87 (0.80, 0.94)				
60-79	81.1 (80.3)	0.86 (0.83, 0.89)	0.81			
80+	86.4 (77.8)	0.86 (0.74, 0.97)	0.84	0.94		
Gender						
Male	81.3 (82.6)	0.87 (0.83, 0.91)				
Female	80.9 (82.0)	0.86 (0.82, 0.90)	0.77			
Smoking status						
None	83.3 (86.2)	0.87 (0.74, 0.99)				
Former	78.9 (83.3)	0.85 (0.81, 0.89)	0.78			
Current	82.9 (81.7)	0.88 (0.84, 0.92)	0.85	0.29		
Specimen –diagnosis date difference (yrs)						
<5	84.7 (81.5)	0.88 (0.84, 0.93)				
5+	79.8 (82.9)	0.86 (0.82, 0.89)	0.34			
Cytology						
Negative	73.9 (82.2)	0.82 (0.78, 0.86)				
Positive	91.3 (100.0)	0.96 (-)	<0.001			
Histology diagnosis^						
Adenocarcinoma	73.3 (82.3)	0.81 (0.76, 0.87)				
Large cell carcinoma	85.4 (82.3)	0.89 (0.83, 0.96)	0.06			
Small cell carcinoma	97.7 (82.3)	0.97 (0.94, 1.00)	<0.001	0.03		
Squamous cell carcinoma	84.4 (82.3)	0.88 (0.84, 0.92)	0.04	0.75	0.0002	
Others	78.3 (82.3)	0.84 (0.75, 0.94)	0.58	0.40	0.013	0.48

* DeLong test for AUC comparison extended for unpaired sample.

^all controls subjects used as control group for each histology subgroup.

The discriminatory performances of cytology alone, best subset algorithm and the two combined are shown in Table 6.3.6. Cytology alone has expectedly poor sensitivity (43%) and least predictive ability in terms of the ROC-AUC. The best subset discriminatory algorithm demonstrated almost double sensitivity (81%) but with a moderate specificity (82%). Inclusion of cytology in the algorithm improved specificity to 92% at a minute sensitivity cost (77%). Stratification of the results by cytology indicated that the best discriminatory algorithm is particularly useful for cytology negative subjects (Figure 6.3.3) as there was a tremendous increase in sensitivity among this group of subjects (3.1% to 74%) followed though by a moderate in specificity (100% to 82%).

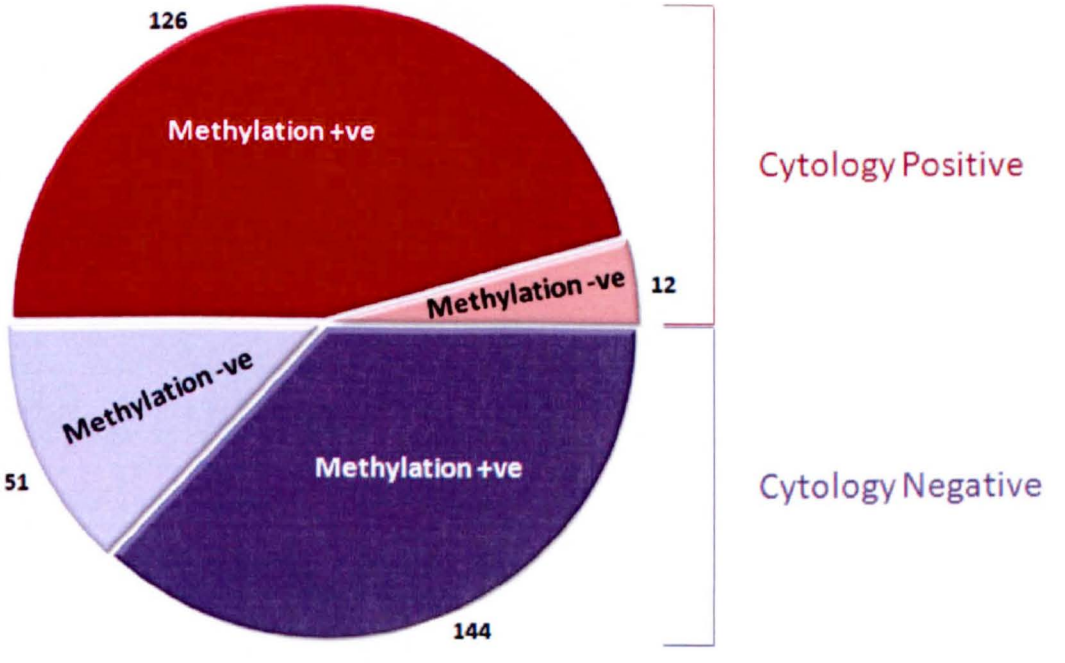


Figure 6.3.3. Distribution of DNA methylation positive cases in different cytology groups. The number of samples in each subgroup is shown next to the relevant slices. It is of note that a significant number of cytologically occult cases are methylation positive.

Table 6.3.6. Sensitivity, specificity and discriminatory accuracy of cytology classifier, best logit classifier and best logit classifier improved with cytology.

Prediction model	Cytology		Overall	AUC
	Positive	Negative		
Cytology only				
True positive	138/138 (100.0)	6/195 (3.1)	144/333 (**43.2%**)	
True Negative	0/1 (0.0)	321/321 (100.0)	321/322 (**99.7%**)	0.71 (0.68, 0.73)
Accuracy (%)	99.3	63.4	71.0	
Best logit classifier only				
True positive	126/138 (91.3)	144/195 (73.9)	270/333 (**81.1%**)	
True Negative	1/1 (100.0)	264/321 (82.2)	265/322 (**82.3%**)	0.86 (0.84, 0.89)
Accuracy (%)	91.4	79.1	81.7	
Best logit classifier & cytology combined				
True positive	138/138 (100.0)	118/195 (60.5)	256/333 (**76.9%**)	0.90 (0.87, 0.92)
True Negative	0/1 (0.0)	297/321 (92.5)	297/322 (**92.2%**)	
Accuracy (%)	99.3	80.4	84.4	

6.4. Discussion

6.4.1. Successful panel genes.

Following the analysis of the data provided by the qMSP assays, the optimal algorithm provided by the best fit model included a combination of cytology with p16, RASSF1, WT1 and hTERT. Below there is a brief description of the four genes.

p16, also known as CDKN2A (cyclin-dependent kinase inhibitor 2A) was early characterised as a human tumour suppressor gene (Shapiro & Rollins, 1996). It is located in the chromosome 9p21 region (Ohno, 1996). This locus encompasses ~42 kb and encodes three distinct tumour suppressor proteins, p14ARF, p15INK4b and p16INK4a (Witcher & Emerson, 2009). p16 is a key regulator of G1 phase cell-cycle arrest and senescence, achieved through inhibition of the cyclin dependent kinases CDK4 and CDK6 (Hara et al, 1996). p16 abnormalities in cancer include deletions and promoter hypermethylation. Inactivation of the gene by the latter mechanism is one of the earliest events leading to loss of function in numerous types of cancer such as lung, colorectal and breast (Belinsky et al, 1998; Foster et al, 1998).

RASSF1 is a Ras association domain family member 1. The gene encodes a protein similar to the RAS effector proteins. It is located in chromosomal region 3p21.3, in which loss of hetorozygosity is extremely common in most forms of cancer (Kok & Tilanus, 1996; Wistuba et al, 2000). Deletion of the short arm of chromosome 3 is the earliest and most common alteration which occurs in the pathogenesis of lung cancer (Hung et al, 1995). Epigenetic aberrations of RASSF1 have also a crucial role in cancer development. Promoter hypermethylation of RASSF1A was frequently detected in several tumour entities

(Dammann et al, 2005) leading to gene inactivation in advanced tumour stage (Lee et al, 2001). RASSF1 hypermethylation was also reported in BWs, showing association with smoking status (Kim et al, 2003).

hTERT (telomerase reverse transcriptase) is a ribonucleoprotein polymerase that maintains telomere ends by addition of the telomere repeat TTAGGG (Feng et al, 1995). Loss of telomerase activity leads to telomere attrition through multiple nuclear divisions (Harley, 1991). Telomerase is active in 70-90% of malignant tissues and immortal cell lines (Kim et al, 1994). The gene coding for hTERT is located in chromosome 5p15.33 (Feng et al, 1995). The activation of telomerase does not promote carcinogenesis but it does allow a cell to continue division and achieve immortality (Kirkpatrick & Mokbel, 2001). Various epigenetic regulatory phenomena related with the hTERT gene have been reported, e.g. hyperacetylation of core histones at the hTERT promoter (Xu et al, 2001). Among those epigenetic effects, DNA methylation has been observed early (Devereux et al, 1999).

Finally, WT1 encodes a transcription factor that contains four zinc-finger motifs at the C-terminus and a proline/glutamine-rich DNA-binding domain at the N-terminus. It is located at chromosome position 11p13, and its inactivation has been associated with a number of Wilms' tumours, as well as mutations has been found in germline of susceptible individuals (Haber & Housman, 1992). Loss of heterozygosity were combined with WT1 silencing driven by promoter hypermethylation to support a two-hit model theory (Satoh et al, 2003).

Searching within previous relevant literature, a number of our targets have appeared. p16, RASSF1, RARβ, MGMT and DAPK promoter methylation has been shown in BL (Ahrendt et al, 1999; Chan et al, 2002; de Fraipont et al, 2005; Grote, 2006; Kim et al, 2004; Schmiemann et al, 2005; Topaloglu et al, 2004; van der Drift et al, 2011). All the above studies rather

reported on the feasibility of detection, occasionally demonstrating differences between lung cancer cases and controls. However, their design did not facilitate biomarker validation as they suffered mainly by inadequate sample numbers and occasionally lack of appropriate set of controls. It is of note that CYGB, WT1 and hTERT, which demonstrated significant detection frequencies between cases and controls in our study, have never been previously shown in BWs.

It is also of note that TMEFF2 and CDH13 showed significant promoter hypermethylation in primary lung tumours, relatively to normal adjacent tissue, but provided no specificity at all in the training BWs set. Thus, they were excluded from the next phase. A significant number controls demonstrated methylated TMEFF2 and CDH13. The precise reason for this discrepancy remains unclear. One can speculate that TMEFF2 and CDH13 methylation is coming from inflammatory cells. Thus, the positive signal detected in the primary tumour may have originated from infiltrating leukocytes rather than the tumour cells themselves. In the BWs set, all controls were selected from a hospital cohort of individuals referred to the Rapid Access Clinic with severe loss of lung function. Many of these individuals were diagnosed with lung infections and chronic inflammatory conditions (bronchitis, emphysema, COPD etc). Inflammation-related methylation is already reported previously (Shivapurkar et al, 2004). Of course, this has to be experimentally confirmed for TMEFF2 and CDH13.

6.4.2. Study design: Statistical Power and biases.

The BWs screening phase is a nested retrospective case-control study, within the LLP hospital cohort. The study design process was led by our in-house statistician, Dr O. Raji and frequent consultation from Professor S. Duffy (Wolfson Institute of Preventive Medicine, London) who acts as the LLP statistical advisor. Power calculations taking into account the biomarker frequencies in the primary disease were undertaken to identify the minimum sample number (n=200 on each arm) for 80% power. Eventually, we ended up screening 1.5 times more samples (over 300 on each arm) boosting thus the overall power.

The potential biases were considered early in the study. As shown in chapter 6, the age range and male/female ratio follow the national figures demonstrating that our recruitment process did not impose any bias. Cases and controls were matched for age and sex. The vast majority are white British, residents of Merseyside, Cheshire and North Wales.

Two biases were recognised at the tissue validation phase: histology and T status. All the tissues screened were adenocarcinomas and squamous cell carcinomas. Small cell carcinomas were not included as they are largely inoperable, thus, we had no such frozen tissue available. In addition, less frequent histological types such as large cell carcinomas, carcinoids and carcinosarcomas, were available at very small numbers (n<10) in our tissue bank. It was therefore considered appropriate at that point that adenocarcinomas and squamous carcinomas, comprising approximately 75% of all lung neoplasms, would provide the basis for biomarker discovery. However, at the BWs phase it became evident that the small cell, large cell and other lung carcinomas were efficiently detected with this panel.
Most of the tissues in the validation step were of T status = 2. This is because T1s are usually too small to allow the pathologist sharing tissue for research, while T3s and T4s are usually

inoperable. Most of the T3/T4 samples in our tissue bank come from pre-operation understaging.

However, the over-representation of T2s in the tissue validation set did not seem to affect the panel's efficiency in detecting T1 carcinomas. This is not surprising as it is well established in the literature that hypermethylation of TSGs and DNA repair genes is an early event in carcinogenesis.

Chapter 7. Overall Study Appraisal and Final Conclusions

The current study is an integral part of the LLP, which aims to reduce mortality from lung cancer by facilitating early detection of the disease (Field & Youngson, 2002). Early detection of lung cancer is a well recognised unmet clinical need. Existing studies emphasize on the particular advantages for patient survival of stage T1 tumour resections (Brock et al, 2008). Spiral computed tomography (CT) screening trials most likely show the way forward in identifying small respectable lung lesions (Henschke et al, 2006). UKLS, the first such trial in the UK, has been just launched in September 2011. However, it is obvious that public health economics will almost certainly impose the requirement for a stratification tool to point the individuals in need for screening. This will reduce, or possibly eliminate, overdiagnosis, making screening feasible within a reasonable public health spending context. It is widely accepted to date that an epidemiological-molecular modelling approach is the way forward for stratification of high risk individuals (Field, 2008). There are numerous attempts utilising molecular markers in sputum and plasma in order to assist lung cancer diagnosis (Tsou et al, 2007). However, the number of high precision diagnostic biomarkers for early lung cancer detection is currently very low, despite the plethora of research articles on potential clinical biomarkers. The main reasons can be focused in the general tendency of previous studies not to follow phased approaches for biomarker discovery and validation. There is a high methodological diversity of detection techniques and lack of extensive assay validation. In addition, most studies are of inadequate statistical power and encompass unaccounted systematic biases.

Therefore, in this study, every possible effort was made to avoid such mistakes. The overall discovery and validation process was a careful and dynamic process ensuring compliance to

the EDRN and CR-UK biomarker guidelines. The selection of the target molecular abnormalities, the assay development and validation, the target population selection, inclusion/exclusion criteria and finally the statistical modelling of the results followed strict rules to allow a final product pursuing actual clinical use rather than just a publication in a respected journal.

7.1. DNA methylation as the target abnormality of choice.

The existing evidence on the universal character of epigenetic deregulation in human tumours is undisputable, with long lists of candidate biomarkers emerging for different aspects of cancer management such as risk modelling, early diagnosis, relapse prediction, prognosis, treatment stratification and treatment monitoring. The link between cancer and abnormal methylation has been known since 1983, with the demonstration that cancer genomes are relatively hypomethylated compared with normal counterparts (Feinberg & Tycko, 2004). In contrast, site specific hypermethylation of the often unmethylated CpG islands, mainly found in gene promoters is the most known and well characterised epigenetic modification in carcinogenesis (Bowman et al, 2006).

The very chemical nature and stability of DNA methylation makes it an attractive route for biomarkers development. The DNA methylation footprint of abnormal cells is very stable combining the fact that methylation is a covalent modification, resistant to sample fixation, and the fact that DNA is probably the most stable biological macromolecule. Not to be forgotten is the fact that DNA methylation changes would require cell duplication. This provides an enormous stability advantage in comparison to RNA and protein expression

which can be subjects to immediate change following chemical stimuli or even simple environmental changes. Thus, a given DNA methylation profile of fixed tissue (e.g. sputum or bronchial lavage) reflects in very high degree their real profile whilst still in the patient's body. This is of particular importance as it is compatible to specimen collection and storage in normal clinical practice, facilitating thus its potential implementation in a clinical environment.

7.2. Biomarker validation

The validation and clinical implementation of biomarkers is a lengthy and very expensive process, following very similar strategies to the drug pipelines. The EDRN of the National Cancer Institute in USA was the first to suggest discrete steps of biomarker validation (Figure 3.1). Taking into account the five phases suggested in there, this work covers phases 1-3, i.e. preclinical exploratory (Chapter 4), clinical assay validation (Chapter 5) and retrospective longitudinal (Chapter 6).

More recently, CR-UK has issued a number of biomarker road maps, which are in the same lines with the EDRN suggestions, probably more fit to the UK perspective. Although every single step is very important in this map (Figure 3.2), emphasis should be given to four points:

1. A biomarker addresses an unmet clinical need and must provide some patient benefit in a clinical setting.
2. The biomarker must demonstrate adequate representation in the target population.

3. A very accurate and reproducible assay is required. At the prospective phase(s) assays should meet GLP and GCP standards.
4. The biomarker should clearly demonstrate improved clinical outcome in comparison to the current gold standard(s).

Despite the number of existing reports on potential DNA methylation biomarkers in BWs (Liloglou & Field, 2010) none of them has progressed to clinically validating a product, with the exception of mSHOX (Schmidt et al, 2010). The main problems can be focused on the small numbers of patients used, the lack of training and validation sets, the limited follow-up and the diversity of methods employed. An additional consideration in biomarker studies is reporting. Inconsistent reporting in published literature is among the reasons why so many individual studies cannot be combined to produce more robust information. It is now accepted that biomarker studies reporting should comply with the STARD guidelines (Bossuyt & Reitsma, 2003). A similar set of guidelines is produced for prognostic biomarkers under the abbreviation REMARK (McShane et al, 2005). These provide a checklist of important aspects contributing to adequate reporting. This list includes study aims, study population, recruitment process, methods for diagnostic accuracy etc. It aims to impose the provision of all the important factors which could add biases and affect diagnostic accuracy.

Compliance to EDRN and CR-UK guidelines was a major element of our study design. This design proved a dynamic process frequently faced problem solving in both assay and patient inclusion aspects. Great emphasis was given to producing extremely robust DNA methylation assays which leave no space for subjective interpretations of the result. Every effort was made to avoid systematic biases that would provide misleading results.

The final product of this study reliably improves diagnostic efficiency of lung cancer in cytologically occult BWs. Current work in the lab examines the improvement of the algorithm by including additional promoters. In parallel discussions with an industrial partner are taking place in order to formulate the next step of transforming the assays into a CE certified clinical diagnostic kit. It is a strong belief within the group that a large prospective trial, which is currently being organised, will be completed in the next five years providing a clinical tool which will significantly assist in detecting lung cancer early with the aim of reducing its high mortality.

References

Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, Herman JG, Wu L, Decker PA, Jen J, Sidransky D (1999) Molecular detection of tumour cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J Natl Cancer Inst* **91:** 332-339

Ahrendt SA, Decker PA, Alawi EA, Zhu Yr YR, Sanchez-Cespedes M, Yang SC, Haasler GB, Kajdacsy-Balla A, Demeure MJ, Sidransky D (2001) Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer* **92:** 1525-1530

Alkhatib SG, Landry JW (2011) The Nucleosome Remodeling Factor. *FEBS Lett*

Antequera F, Boyes J, Bird A (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* **62:** 503-514

Aparicio T, Navazesh A, Boutron I, Bouarioua N, Chosidow D, Mion M, Choudat L, Sobhani I, Mentre F, Soule JC (2009) Half of elderly patients routinely treated for colorectal cancer receive a sub-standard treatment. *Crit Rev Oncol Hematol* **71:** 249-257

Ardizzoni A, Boni L, Tiseo M, Fossella FV, Schiller JH, Paesmans M, Radosavljevic D, Paccagnella A, Zatloukal P, Mazzanti P, Bisset D, Rosell R (2007) Cisplatin- versus carboplatin-based chemotherapy in first-line treatment of advanced non-small-cell lung cancer: an individual patient data meta-analysis. *J Natl Cancer Inst* **99:** 847-857

Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, G.J. D, Hoth DF, Oates JA, Peck CC, Schooley RT, Spilker BA, Woodcock J, Zeger SL, Do GJ (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics* **69:** 89 - 95

Ballestar E, Paz MF, Valle L, Wei S, Fraga MF, Espada J, Cigudosa JC, Huang TH, Esteller M (2003) Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J* **22:** 6335-6345

Bartolomei MS, Tilghman SM (1997) Genomic imprinting in mammals. *Annual review of genetics* **31:** 493-525

Baryshnikova E, Destro A, Infante MV, Cavuto S, Cariboni U, Alloisio M, Ceresoli GL, Lutman R, Brambilla G, Chiesa G, Ravasi G, Roncalli M (2008) Molecular alterations in spontaneous sputum of cancer-free heavy smokers: results from a large screening program. *Clin Cancer Res* **14:** 1913-1919

Baylin SB, Belinsky SA, Herman JG (2000) Aberrant methylation of gene promoters in cancer---concepts, misconcepts, and promise. *J Natl Cancer Inst* **92:** 1460-1461

Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. *Advances in cancer research* **72:** 141-196

Baylin SB, Ohm JE (2006) Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nature reviews* **6:** 107-116

Belinsky SA, Collins LG, Haines C, Perkel R, Enck RE, Gridelli C, Rossi A, Maione P, Hirsch FR, Fischer JR, Niklinski J, Zochbauer-Muller S, Mulshine JL, Hirsch FR, Ramalingam S, Belani CP, Macbeth F, Stephens R (2004) Gene-promoter hypermethylation as a biomarker in lung cancer. Lung cancer: diagnosis and management. Treatment of non-small-cell lung cancer: state of the art and development of new biologic agents. Future developments in the treatment of lung cancer. Lung cancer chemoprevention: moving from concept to a reality. State-of-the-art chemotherapy for advanced non-small cell lung cancer.Palliative treatment for advanced non-small cell lung cancer. *Nature reviews* **4:** 707-717

Belinsky SA, Grimes MJ, Casas E, Stidley CA, Franklin WA, Bocklage TJ, Johnson DH, Schiller JH (2007) Predicting gene promoter methylation in non-small-cell lung cancer by evaluating sputum and serum. *Br J Cancer* **96:** 1278-1283

Belinsky SA, Klinge DM, Dekker JD, Smith MW, Bocklage TJ, Gilliland FD, Crowell RE, Karp DD, Stidley CA, Picchi MA (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin Cancer Res* **11:** 6505-6511

Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K, Haney J, Kennedy TC, Hirsch FR, Miller Y, Franklin WA, Herman JG, Baylin SB, Bunn PA, Byers T (2006) Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer research* **66:** 3338-3344

Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG (1998) Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 11891-11896

Bellido ML, Radpour R, Lapaire O, De Bie I, Hosli I, Bitzer J, Hmadcha A, Zhong XY, Holzgreve W (2010) MALDI-TOF mass array analysis of RASSF1A and SERPINB5 methylation patterns in human placenta and plasma. *Biol Reprod* **82:** 745-750

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. *Genes Dev* **23:** 781-783

Bestor TH (2000) The DNA methyltransferases of mammals. *Human molecular genetics* **9:** 2395-2402

Bird AP (1996) The relationship of DNA methylation to cancer. *Cancer surveys* **28:** 87-101

Bird AP, Wolffe AP (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell* **99:** 451-454

Blanco D, Vicent S, Fraga MF, Fernandez-Garcia I, Freire J, Lujambio A, Esteller M, Ortiz-de-Solorzano C, Pio R, Lecanda F, Montuenga LM (2007) Molecular analysis of a multistep lung cancer model induced by chronic inflammation reveals epigenetic regulation of p16 and activation of the DNA damage response pathway. *Neoplasia* **9:** 840-852

Bossuyt PM, Reitsma JB (2003) The STARD initiative. *Lancet* **361:** 71

Bowie C, Bowie SH (1991) Radon and health. *Lancet* **337:** 409-413

Bowman RV, Yang IA, Semmler AB, Fong KM (2006) Epigenetics of lung cancer. *Respirology (Carlton, Vic* **11:** 355-365

Brait M, Ford JG, Papaiahgari S, Garza MA, Lee JI, Loyo M, Maldonado L, Begum S, McCaffrey L, Howerton M, Sidransky D, Emerson MR, Ahmed S, Williams CD, Hoque MO (2009) Association between lifestyle factors and CpG island methylation in a cancer-free population. *Cancer Epidemiol Biomarkers Prev* **18:** 2984-2991

Brambilla E, Travis WD, Colby TV, Corrin B, Shimosato Y (2001) The new World Health Organization classification of lung tumours. *Eur Respir J* **18:** 1059-1068

Brennan P, Hainaut P, Boffetta P (2011) Genetics of lung-cancer susceptibility. *Lancet Oncol* **12:** 399-408

Brock MV, Hooker CM, Ota-Machida E, Han Y, Guo M, Ames S, Glockner S, Piantadosi S, Gabrielson E, Pridham G, Pelosky K, Belinsky SA, Yang SC, Baylin SB, Herman JG (2008) DNA methylation markers and early recurrence in stage I lung cancer. *The New England journal of medicine* **358:** 1118-1128

Broker LE, Giaccone G (2002) The role of new agents in the treatment of non-small cell lung cancer. *Eur J Cancer* **38:** 2347-2361

Budgen A (2004) Asbestos: a clear and present danger--a UK perspective. *Lung cancer (Amsterdam, Netherlands)* **45 Suppl 1:** S77-79

Bullinger L, Ehrich M, Dohner K, Schlenk RF, Dohner H, Nelson MR, van den Boom D (2010) Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood* **115:** 636-642

Cairns BR (2001) Emerging roles for chromatin remodeling in cancer biology. *Trends in cell biology* **11:** S15-21

Carvalho L, Cardoso E, Nunes H, Baptista V, Gomes A, Couceiro P (2009) [The IASLC lung cancer staging project. Comparing the current 6(th) TNM edition with the proposed 7(th) edition]. *Rev Port Pneumol* **15:** 67-76

Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, Field JK (2008) The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* **98:** 270-276

Chan EC, Lam SY, Tsang KW, Lam B, Ho JC, Fu KH, Lam WK, Kwong YL (2002) Aberrant promoter methylation in Chinese patients with non-small cell lung cancer: patterns in primary tumours and potential diagnostic application in bronchoalevolar lavage. *Clin Cancer Res* **8:** 3741-3746

Chan TL, Yuen ST, Kong CK, Chan YW, Chan AS, Ng WF, Tsui WY, Lo MW, Tam WY, Li VS, Leung SY (2006) Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet* **38:** 1178-1183

Chedin F, Lieber MR, Hsieh CL (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 16916-16921

Chen H, Suzuki M, Nakamura Y, Ohira M, Ando S, Iida T, Nakajima T, Nakagawara A, Kimura H (2006) Aberrant methylation of RASGRF2 and RASSF1A in human non-small cell lung cancer. *Oncol Rep* **15:** 1281-1285

Cheung P, Allis CD, Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* **103:** 263-271

Chim SS, Jin S, Lee TY, Lun FM, Lee WS, Chan LY, Jin Y, Yang N, Tong YK, Leung TY, Lau TK, Ding C, Chiu RW, Lo YM (2008) Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test for fetal trisomy 21. *Clin Chem* **54:** 500-511

Cirincione R, Lintas C, Conte D, Mariani L, Roz L, Vignola AM, Pastorino U, Sozzi G (2006) Methylation profile in tumour and sputum samples of lung cancer patients detected by spiral computed tomography: a nested case-control study. *International journal of cancer* **118:** 1248-1253

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78:** 273-304

Clayton SJ, Scott FM, Walker J, Callaghan K, Haque K, Liloglou T, Xinarianos G, Shawcross S, Ceuppens P, Field JK, Fox JC (2000) K-ras point mutation detection in lung cancer: comparison of two approaches to somatic mutation detection using ARMS allele-specific amplification. *Clin Chem* **46:** 1929-1938

Colella S, Shen L, Baggerly KA, Issa JP, Krahe R (2003) Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *BioTechniques* **35:** 146-150

Collins LG, Haines C, Perkel R, Enck RE (2007) Lung cancer: diagnosis and management. *American family physician* **75:** 56-63

Dammann R, Schagdarsurengin U, Seidel C, Strunnikova M, Rastetter M, Baier K, Pfeifer GP (2005) The tumour suppressor RASSF1A in human carcinogenesis: an update. *Histol Histopathol* **20:** 645-663

Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, Deo H, Falk R, Forastiere F, Hakama M, Heid I, Kreienbrock L, Kreuzer M, Lagarde F, Makelainen I, Muirhead C, Oberaigner W, Pershagen G, Ruano-Ravina A, Ruosteenoja E et al (2005) Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ (Clinical research ed* **330:** 223

Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis VG, Field JK, Liloglou T (2008) Hypomethylation of transposable elements corelates with genomic instability in non-small cell lung cancer. *Int J Cancer* **In press**

de Fraipont F, Moro-Sibilot D, Michelland S, Brambilla E, Brambilla C, Favrot MC (2005) Promoter methylation of genes in bronchial lavages: a marker for early diagnosis of primary

and relapsing non-small cell lung cancer? *Lung cancer (Amsterdam, Netherlands)* **50:** 199-209

Deligezer U, Akisik EE, Erten N, Dalay N (2008) Sequence-specific histone methylation is detectable on circulating nucleosomes in plasma. *Clin Chem* **54:** 1125-1131

Deplus R, Brenner C, Burgers WA, Putmans P, Kouzarides T, de Launoit Y, Fuks F (2002) Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic Acids Res* **30:** 3831-3838

Devereux TR, Horikawa I, Anna CH, Annab LA, Afshari CA, Barrett JC (1999) DNA methylation analysis of the promoter region of the human telomerase reverse transcriptase (hTERT) gene. *Cancer research* **59:** 6087-6090

Divine KK, Pulling LC, Marron-Terada PG, Liechty KC, Kang T, Schwartz AG, Bocklage TJ, Coons TA, Gilliland FD, Belinsky SA (2005) Multiplicity of abnormal promoter methylation in lung adenocarcinomas from smokers and never smokers. *International journal of cancer* **114:** 400-405

Dobler CC, Crawford AB (2009) Bronchoscopic diagnosis of endoscopically visible lung malignancies: should cytologic examinations be performed routinely? *Intern Med J*

Doll R, Hill AB (1950) Smoking and carcinoma of the lung; preliminary report. *Br Med J* **2:** 739-748

Dunn BK, Verma M, Umar A (2003) Epigenetics in Cancer Prevention: Early Detection and Risk Assessment. Introduction. *Annals of the New York Academy of Sciences* **983:** 1-4

Dunn JR, Panutsopulos D, Shaw MW, Heighway J, Dormer R, Salmo EN, Watson SG, Field JK, Liloglou T (2004) METH-2 silencing and promoter hypermethylation in NSCLC. *Br J Cancer* **91:** 1149-1154

Dunwell T, Hesson L, Rauch TA, Wang L, Clark RE, Dallol A, Gentle D, Catchpoole D, Maher ER, Pfeifer GP, Latif F (2010) A genome-wide screen identifies frequently methylated genes in haematological and epithelial cancers. *Mol Cancer* **9:** 44

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV, Laird PW (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28:** E32

Ehrich M, Field JK, Liloglou T, Xinarianos G, Oeth P, Nelson MR, Cantor CR, van den Boom D (2006) Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. *Cancer research* **66:** 10911-10918

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, van den Boom D (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* **102:** 15785-15790

Ehricht R, Slickers P, Goellner S, Hotzel H, Sachse K (2006) Optimized DNA microarray assay allows detection and genotyping of single PCR-amplifiable target copies. *Mol Cell Probes* **20:** 60-63

Ehrlich M (2002) DNA methylation in cancer: too much, but also too little. *Oncogene* **21:** 5400-5413

Ehrlich S, Weiss D, Burghardt R, Infante-Duarte C, Brockhaus S, Muschler MA, Bleich S, Lehmkuhl U, Frieling H (2010) Promoter specific DNA methylation and gene expression of POMC in acutely underweight and recovered patients with anorexia nervosa. *J Psychiatr Res* **44:** 827-833

Erenpreisa J, Cragg MS (2007) Cancer: a matter of life cycle? *Cell Biol Int* **31:** 1507-1510

Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nature reviews* **6:** 259-269

Esteller M (2007a) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* **8:** 286-298

Esteller M (2007b) Epigenetic gene silencing in cancer: the DNA hypermethylome. *Human molecular genetics* **16 Spec No 1:** R50-59

Esteller M, Corn PG, Baylin SB, Herman JG (2001) A gene hypermethylation profile of human cancer. *Cancer Res* **61:** 3225-3229

Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison CD, Chan KK, Marcucci G, Calin GA, Huebner K, Croce CM (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 15805-15810

Fackler MJ, Rivers A, Teo WW, Mangat A, Taylor E, Zhang Z, Goodman S, Argani P, Nayar R, Susnik B, Sukumar S, Khan SA (2009) Hypermethylated genes as biomarkers of cancer in women with pathologic nipple discharge. *Clin Cancer Res* **15:** 3802-3811

Feigelson HS, Ross RK, Yu MC, Coetzee GA, Reichardt JK, Henderson BE (1996) Genetic susceptibility to cancer from exogenous and endogenous exposures. *Journal of cellular biochemistry* **25:** 15-22

Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nature reviews* **4:** 143-153

Feng J, Funk WD, Wang SS, Weinrich SL, Avilion AA, Chiu CP, Adams RR, Chang E, Allsopp RC, Yu J, et al. (1995) The RNA component of human telomerase. *Science (New York, NY* **269:** 1236-1241

Field JK (2008) Lung cancer risk models come of age. *Cancer Prev Res (Phila)* **1:** 226-228

Field JK, Duffy SW (2008) Lung cancer screening: the way forward. *Br J Cancer* **99:** 557-562

Field JK, Liloglou T, Warrak S, Burger M, Becker E, Berlin K, Nimmrich I, Maier S (2005a) Methylation discriminators in NSCLC identified by a microarray based approach. *Int J Oncol* **27:** 105-111

Field JK, Liloglou T, Xinarianos G, Prime W, Fielding P, Walshaw MJ, Turnbull L (1999) Genetic alterations in bronchial lavage as a potential marker for individuals with a high risk of developing lung cancer. *Cancer Res* **59:** 2690-2695

Field JK, Neville EM, Stewart MP, Swift A, Liloglou T, Risk JM, Ross H, Gosney JR, Donnelly RJ (1996) Fractional allele loss data indicate distinct genetic populations in the development of non-small-cell lung cancer. *Br J Cancer* **74:** 1968-1974

Field JK, Smith DL, Duffy S, Cassidy A (2005b) The Liverpool Lung Project research protocol. *Int J Oncol* **27:** 1633-1645

Field JK, Youngson JH (2002) The Liverpool Lung Project: a molecular epidemiological study of early lung cancer detection. *Eur Respir J* **20:** 464-479

Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, Schifano E, Booth J, van Putten W, Skrabanek L, Campagne F, Mazumdar M, Greally JM, Valk PJ, Lowenberg B, Delwel R, Melnick A (2010) DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17:** 13-27

Foster SA, Wong DJ, Barrett MT, Galloway DA (1998) Inactivation of p16 in human mammary epithelial cells by CpG island methylation. *Molecular and cellular biology* **18:** 1793-1801

Fraga MF, Esteller M (2002) DNA methylation: a profile of methods and applications. *BioTechniques* **33:** 632, 634, 636-649

Franklin RA, McCubrey JA (2008) A multipronged approach to prostate cancer. *Cancer Biol Ther* **7:** 594-595

Franklin WA (2000a) Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest* **117:** 80S-89S

Franklin WA (2000b) Pathology of lung cancer. *Journal of thoracic imaging* **15:** 3-12

Fryer AA, Nafee TM, Ismail KM, Carroll WD, Emes RD, Farrell WE (2009) LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. *Epigenetics* **4:** 394-398

Gao W, Kondo Y, Shen L, Shimizu Y, Sano T, Yamao K, Natsume A, Goto Y, Ito M, Murakami H, Osada H, Zhang J, Issa JP, Sekido Y (2008) Variable DNA methylation patterns associated with progression of disease in hepatocellular carcinomas. *Carcinogenesis*

Galasso M, Sandhu SK, Volinia S (2012) MicroRNA Expression Signatures in Solid Malignancies. *Cancer J* **18:** 238-243

Gazdar AF (2009) Personalized medicine and inhibition of EGFR signaling in lung cancer. *The New England journal of medicine* **361:** 1018-1020

Gazdar AF, Minna JD (2008) Deregulated EGFR signaling during lung cancer progression: mutations, amplicons, and autocrine loops. *Cancer Prev Res (Phila)* **1:** 156-160

Geiman TM, Muegge K (2010) DNA methylation in early development. *Mol Reprod Dev* **77:** 105-113

Gemma C, Sookoian S, Dieuzeide G, Garcia SI, Gianotti TF, Gonzalez CD, Pirola CJ (2010) Methylation of TFAM gene promoter in peripheral white blood cells is associated with insulin resistance in adolescents. *Mol Genet Metab* **100:** 83-87

Giaccone G (2002) Targeted therapy in non-small cell lung cancer. *Lung cancer (Amsterdam, Netherlands)* **38 Suppl 2:** S29-32

Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, Chajut A (2008) Serum microRNAs are promising novel biomarkers. *PLoS One* **3:** e3148

Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* **5:** 401-408

Godler DE, Slater HR, Amor D, Loesch DZ (2010) Methylation analysis of fragile X-related epigenetic elements may provide a suitable newborn screening test for fragile X syndrome. *Genet Med* **12:** 595

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science (New York, NY* **311:** 395-398

Gonzalgo ML, Jones PA (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res* **25:** 2529-2531

Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, Ditullio RA, Jr., Kastrinakis NG, Levy B, Kletsas D, Yoneta A, Herlyn M, Kittas C, Halazonetis TD (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434:** 907-913

Gorlov IP, Meyer P, Liloglou T, Myles J, Boettger MB, Cassidy A, Girard L, Minna JD, Fischer R, Duffy S, Spitz MR, Haeussinger K, Kammerer S, Cantor C, Dierkesmann R, Field JK, Amos CI (2007) Seizure 6-like (SEZ6L) gene and risk for lung cancer. *Cancer Res* **67:** 8406-8411

Gridelli C, Rossi A, Maione P (2003) Treatment of non-small-cell lung cancer: state of the art and development of new biologic agents. *Oncogene* **22:** 6629-6638

Grote HJ (2006) [Aberrant promoter methylation as biomarker for molecular cytological diagnosis of lung cancer]. *Verh Dtsch Ges Pathol* **90:** 216-226

Grote HJ, Schmiemann V, Geddert H, Bocking A, Kappes R, Gabbert HE, Sarbia M (2006) Methylation of RAS association domain family protein 1A as a biomarker of lung cancer. *Cancer* **108:** 129-134

Grote HJ, Schmiemann V, Geddert H, Rohr UP, Kappes R, Gabbert HE, Bocking A (2005) Aberrant promoter methylation of p16(INK4a), RARB2 and SEMA3B in bronchial aspirates from patients with suspected lung cancer. *International journal of cancer* **116:** 720-725

Haber DA, Housman DE (1992) Role of the WT1 gene in Wilms' tumour. *Cancer surveys* **12:** 105-117

Hall SJ, Chen SH, Woo SL (1997) The promise and reality of cancer gene therapy. *Am J Hum Genet* **61:** 785-789

Hara E, Smith R, Parry D, Tahara H, Stone S, Peters G (1996) Regulation of p16CDKN2 expression and its implications for cell immortalization and senescence. *Molecular and cellular biology* **16:** 859-867

Hargreaves DC, Crabtree GR (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21:** 396-420

Harley CB (1991) Telomere loss: mitotic clock or genetic time bomb? *Mutat Res* **256:** 271-282

Heard E, Clerc P, Avner P (1997) X-chromosome inactivation in mammals. *Annual review of genetics* **31:** 571-610

Hecht SS (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature reviews* **3:** 733-744

Heighway J, Knapp T, Boyce L, Brennand S, Field JK, Betticher DC, Ratschiller D, Gugger M, Donovan M, Lasek A, Rickert P (2002) Expression profiling of primary non-small cell lung cancer for target identification. *Oncogene* **21:** 7749-7763

Hemann MT, Zilfou JT, Zhao Z, Burgess DJ, Hannon GJ, Lowe SW (2004) Suppression of tumourigenesis by the p53 target PUMA. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 9333-9338

Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS (2006) Survival of patients with stage I lung cancer detected on CT screening. *The New England journal of medicine* **355:** 1763-1771

Herman JG (1999) Hypermethylation of tumour suppressor genes in cancer. *Semin Cancer Biol* **9:** 359-367

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America* **93:** 9821-9826

Hietanen P (1997) How to improve communication between cancer patients and care providers--patients' experiences. *The Tokai journal of experimental and clinical medicine* **22:** 267-270

Hirsch FR, Fischer JR, Niklinski J, Zochbauer-Muller S (2002) Future developments in the treatment of lung cancer. *Lung cancer (Amsterdam, Netherlands)* **38 Suppl 3:** S81-85

Ho L, Crabtree GR (2010) Chromatin remodelling during development. *Nature* **463:** 474-484

Hoffmann AC, Kaifi JT, Vallbohmer D, Yekebas E, Grimminger P, Leers JM, Izbicki JR, Holscher AH, Schneider PM, Metzger R, Brabender J (2009) Lack of prognostic significance of serum DNA methylation of DAPK, MGMT, and GSTPI in patients with non-small cell lung cancer. *J Surg Oncol* **100:** 414-417

Horio Y, Takahashi T, Kuroishi T, Hibi K, Suyama M, Niimi T, Shimokata K, Yamakawa K, Nakamura Y, Ueda R, et al., Slebos RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, Wagenaar SS, Vanderschueren RG, van Zandwijk N et al (1993) Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer

K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *Cancer research* **53:** 1-4

Horsfall MJ, Gordon AJ, Burns PA, Zielenska M, van der Vliet GM, Glickman BW (1990) Mutational specificity of alkylating agents and the influence of DNA repair. *Environ Mol Mutagen* **15:** 107-122

Hou L, Wang H, Sartori S, Gawron A, Lissowska J, Bollati V, Tarantini L, Zhang FF, Zatonski W, Chow WH, Baccarelli A (2010) Blood leukocyte DNA hypomethylation and gastric cancer risk in a high-risk Polish population. *International journal of cancer* **127:** 1866-1874

Hsu HS, Chen TP, Hung CH, Wen CK, Lin RK, Lee HC, Wang YC (2007) Characterization of a multiple epigenetic marker panel for lung cancer detection and risk assessment in plasma. *Cancer* **110:** 2019-2026

Hung J, Kishimoto Y, Sugio K, Virmani A, McIntire DD, Minna JD, Gazdar AF (1995) Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma. *JAMA* **273:** 1908

Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G et al (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452:** 633-637

Hwang H, Quenneville L, Yaziji H, Gown AM (2004) Wilms tumour gene product: sensitive and contextually specific marker of serous carcinomas of ovarian surface epithelial origin. *Applied immunohistochemistry & molecular morphology : AIMM / official publication of the Society for Applied Immunohistochemistry* **12:** 122-126

Issa JP (2004) CpG island methylator phenotype in cancer. *Nature reviews* **4:** 988-993

Issa JP, Gharibyan V, Cortes J, Jelinek J, Morris G, Verstovsek S, Talpaz M, Garcia-Manero G, Kantarjian HM (2005) Phase II study of low-dose decitabine in patients with chronic myelogenous leukemia resistant to imatinib mesylate. *J Clin Oncol* **23:** 3948-3956

Jacinto FV, Esteller M (2007) MGMT hypermethylation: a prognostic foe, a predictive friend. *DNA Repair (Amst)* **6:** 1155-1160

Jack RH, Davies EA, Moller H (2011) Lung cancer incidence and survival in different ethnic groups in South East England. *Br J Cancer* **105:** 1049-1053

Janes H, Pepe M (2006) The optimal ratio of cases to controls for estimating the classification accuracy of a biomarker. *Biostatistics* **7:** 456-468

Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. *CA: a cancer journal for clinicians* **60:** 277-300

Jenuwein T, Allis CD (2001) Translating the histone code. *Science (New York, NY* **293:** 1074-1080

Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3:** 415-428

Jones PA, Baylin SB (2007) The epigenomics of cancer. *Cell* **128:** 683-692

Kang S, Kim J, Kim HB, Shim JW, Nam E, Kim SH, Ahn HJ, Choi YP, Ding B, Song K, Cho NH (2006) Methylation of p16INK4a is a non-rare event in cervical intraepithelial neoplasia. *Diagn Mol Pathol* **15:** 74-82

Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, Yatabe Y, Takamizawa J, Miyoshi S, Mitsudomi T, Takahashi T (2005) Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer science* **96:** 111-115

Kates M, Swanson S, Wisnivesky JP (2011) Survival following lobectomy and limited resection for the treatment of stage I non-small cell lung cancer<=1 cm in size: a review of SEER data. *Chest* **139:** 491-496

Kazan-Allen L (2005) Asbestos and mesothelioma: worldwide trends. *Lung cancer (Amsterdam, Netherlands)* **49 Suppl 1:** S3-8

Keller AK, Jorgensen TM, Olsen LH, Stolle LB (2009) Detection of local metabolic changes after progressive and stepwise reduction of renal blood flow in pigs. *Transplant Proc* **41:** 44-48

Keohavong P, Lan Q, Gao WM, Zheng KC, Mady HH, Melhem MF, Mumford JL (2005) Detection of p53 and K-ras mutations in sputum of individuals exposed to smoky coal emissions in Xuan Wei County, China. *Carcinogenesis* **26:** 303-308

Kerr KM (2011) Personalized medicine for lung cancer: new challenges for pathology. *Histopathology*

Kerr KM, Galler JS, Hagen JA, Laird PW, Laird-Offringa IA (2007) The role of DNA methylation in the development and progression of lung adenocarcinoma. *Dis Markers* **23:** 5-30

Kim DH, Kim JS, Ji YI, Shim YM, Kim H, Han J, Park J (2003) Hypermethylation of RASSF1A promoter is associated with the age at starting smoking and a poor prognosis in primary non-small cell lung cancer. *Cancer research* **63:** 3743-3746

Kim H, Kwon YM, Kim JS, Lee H, Park JH, Shim YM, Han J, Park J, Kim DH (2004) Tumour-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer. *J Clin Oncol* **22:** 2363-2370

Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science (New York, NY* **266:** 2011-2015

King JL, Yang B, Sparks AE, Mains LM, Murray JC, Van Voorhis BJ (2010) Skewed X inactivation and IVF-conceived infants. *Reprod Biomed Online* **20:** 660-663

Kirkpatrick KL, Mokbel K (2001) The significance of human telomerase reverse transcriptase (hTERT) in cancer. *Eur J Surg Oncol* **27:** 754-760

Kneip C, Schmidt B, Seegebarth A, Weickmann S, Fleischhacker M, Liebenberg V, Field JK, Dietrich D (2011) SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer in plasma. *J Thorac Oncol* **6:** 1632-1638

Koivunen JP, Kim J, Lee J, Rogers AM, Park JO, Zhao X, Naoki K, Okamoto I, Nakagawa K, Yeap BY, Meyerson M, Wong KK, Richards WG, Sugarbaker DJ, Johnson BE, Janne PA (2008) Mutations in the LKB1 tumour suppressor are frequently detected in tumours from Caucasian but not Asian lung cancer patients. *Br J Cancer* **99:** 245-252

Kok TC, Tilanus HW (1996) Neoadjuvant treatment in oesophageal cancer: the needs for future trials. The Rotterdam Esophageal Tumour Study Group. *Eur J Surg Oncol* **22:** 323-325

Kondo M, Yokoyama T, Fukui T, Yoshioka H, Yokoi K, Nagasaka T, Imaizumi K, Kume H, Hasegawa Y, Shimokata K, Sekido Y (2005) Mutations of epidermal growth factor receptor of non-small cell lung cancer were associated with sensitivity to gefitinib in recurrence after surgery. *Lung Cancer* **50:** 385-391

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693-705

Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, Klotz JB, Letourneau EG, Lynch CF, Lyon JL, Sandler DP, Schoenberg JB, Steck DJ, Stolwijk JA, Weinberg C, Wilcox HB (2006) A combined analysis of North American case-control studies of residential radon and lung cancer. *Journal of toxicology and environmental health* **69:** 533-597

Kristensen LS, Hansen LL (2009) PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* **55:** 1471-1483

Kronenberg G, Colla M, Endres M (2009) Folic acid, neurodegenerative and neuropsychiatric disease. *Curr Mol Med* **9:** 315-323

Kulis M, Esteller M DNA methylation and cancer. *Adv Genet* **70:** 27-56

Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T (2007) Impaired microRNA processing enhances cellular transformation and tumourigenesis. *Nat Genet* **39:** 673-677

Kumar MS, Erkeland SJ, Pester RE, Chen CY, Ebert MS, Sharp PA, Jacks T (2008) Suppression of non-small cell lung tumour development by the let-7 microRNA family. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 3903-3908

Laird PW (2003) The power and the promise of DNA methylation markers. *Nature reviews* **3:** 253-266

Lee MG, Kim HY, Byun DS, Lee SJ, Lee CH, Kim JI, Chang SG, Chi SG (2001) Frequent epigenetic inactivation of RASSF1A in human bladder carcinoma. *Cancer research* **61:** 6688-6692

Levin ML, Goldstein H, Gerhardt PR (1950) Cancer and tobacco smoking; a preliminary report. *J Am Med Assoc* **143:** 336-338

Li J, Zhang Z, Dai Z, Plass C, Morrison C, Wang Y, Wiest JS, Anderson MW, You M (2003) LOH of chromosome 12p correlates with Kras2 mutation in non-small cell lung cancer. *Oncogene* **22:** 1243-1246

Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, Fang HB, Katz RL, Stass SA, Jiang F (2007) Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin Cancer Res* **13:** 482-487

Liggett T, Melnikov A, Tilwalli S, Yi Q, Chen H, Replogle C, Feng X, Reder A, Stefoski D, Balabanov R, Levenson V (2010) Methylation patterns of cell-free plasma DNA in relapsing-remitting multiple sclerosis. *J Neurol Sci* **290:** 16-21

Liloglou T, Bediaga NG, Brown BR, Field JK, Davies MP (2012) Epigenetic biomarkers in lung cancer. *Cancer Lett*

Liloglou T, Brambilla E, Nikolaidis G, Sozzi G, Thunnissen E, Brambilla C, Field JK (In preparation) p16 hypermethylation as a predicting factor for relapse of NSCLC.

Liloglou T, Field JK Detection of DNA methylation changes in body fluids. *Adv Genet* **71:** 177-207

Liloglou T, Field JK (2010) Detection of DNA methylation changes in body fluids. *Adv Genet* **71:** 177-207

Liloglou T, Maloney P, Xinarianos G, Fear S, Field JK (2000) Sensitivity and limitations of high throughput fluorescent microsatellite analysis for the detection of allelic imbalance: application in lung tumours. *International journal of oncology* **16:** 5-14

Liloglou T, Maloney P, Xinarianos G, Hulbert M, Walshaw MJ, Gosney JR, Turnbull L, Field JK (2001) Cancer-specific genomic instability in bronchial lavage: a molecular tool for lung cancer detection. *Cancer research* **61:** 1624-1628

Liloglou T, Scholes AG, Spandidos DA, Vaughan ED, Jones AS, Field JK (1997) p53 mutations in squamous cell carcinoma of the head and neck predominate in a subgroup of former and present smokers with a low frequency of genetic instability. *Cancer research* **57:** 4070-4074

Liloglou T, Walters M, Maloney P, Youngson J, Field JK (2002) A T2517C polymorphism in the GSTM4 gene is associated with risk of developing lung cancer. *Lung Cancer* **37:** 143-146

Ling C, Groop L (2009) Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes* **58:** 2718-2725

Lips EH, Gaborieau V, McKay JD, Chabrier A, Hung RJ, Boffetta P, Hashibe M, Zaridze D, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Field JK, Liloglou T, Xinarianos G, McLaughlin J et al (2010) Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol* **39:** 563-577

Liu Z, Zhao J, Chen XF, Li W, Liu R, Lei Z, Liu X, Peng X, Xu K, Chen J, Liu H, Zhou QH, Zhang HT (2008) CpG island methylator phenotype involving tumour suppressor genes located on chromosome 3p in non-small cell lung cancer. *Lung cancer (Amsterdam, Netherlands)*

Loeb LA, Bielas JH, Beckman RA (2008) Cancers exhibit a mutator phenotype: clinical implications. *Cancer Res* **68:** 3551-3557; discussion 3557

Macbeth F, Stephens R (2004) Palliative treatment for advanced non-small cell lung cancer. *Hematology/oncology clinics of North America* **18:** 115-130

Machida EO, Brock MV, Hooker CM, Nakayama J, Ishida A, Amano J, Picchi MA, Belinsky SA, Herman JG, Taniguchi S, Baylin SB (2006) Hypermethylation of ASC/TMS1 is a sputum marker for late-stage lung cancer. *Cancer research* **66:** 6210-6218

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* **311:** 195-198

Magewu AN, Jones PA (1994) Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Molecular and cellular biology* **14:** 4225-4232

Margot JB, Ehrenhofer-Murray AE, Leonhardt H (2003) Interactions within the mammalian DNA methyltransferase family. *BMC Mol Biol* **4:** 7

Mariatos G, Gorgoulis VG, Zacharatos P, Kotsinas A, Vogiatzi T, Rassidakis G, Foukas P, Liloglou T, Tiniakos D, Angelou N, Manolis EN, Veslemes M, Field JK, Kittas C (2000) Expression of p16(INK4A) and alterations of the 9p21-23 chromosome region in non-small-cell lung carcinomas: relationship with tumour growth parameters and ploidy status. *International journal of cancer* **89:** 133-141

Marsit CJ, Houseman EA, Christensen BC, Eddy K, Bueno R, Sugarbaker DJ, Nelson HH, Karagas MR, Kelsey KT (2006) Examination of a CpG island methylator phenotype and implications of methylation profiles in solid tumours. *Cancer Res* **66:** 10621-10629

Matzke MA, Mette MF, Kanno T, Matzke AJ (2003) Does the intrinsic instability of aneuploid genomes have a causal role in cancer? *Trends Genet* **19:** 253-256

McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S et al (2008) Lung cancer susceptibility locus at 5p15.33. *Nat Genet* **40:** 1404-1406

McKay MJ, Withana N, Davey DS, Bishop D, Chao M, Sprung CN (2011) Lymphoid and fibroblastic cell lineages from radiosensitive cancer patients: molecular analysis of DNA double strand break repair by major non-homologous end-joining sub-pathways. *Asia Pac J Clin Oncol* **7:** 17-26

McRonald FE, Liloglou T, Xinarianos G, Hill L, Rowbottom L, Langan JE, Ellis A, Shaw JM, Field JK, Risk JM (2006) Down-regulation of the cytoglobin gene, located on 17q25, in tylosis with oesophageal cancer (TOC): evidence for trans-allele repression. *Human molecular genetics* **15:** 1271-1277

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* **93:** 387-391

Milani L, Lundmark A, Kiialainen A, Nordlund J, Flaegstad T, Forestier E, Heyman M, Jonmundsson G, Kanerva J, Schmiegelow K, Soderhall S, Gustafsson MG, Lonnerholm G, Syvanen AC (2010) DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* **115:** 1214-1225

Miller YE (2005) Pathogenesis of lung cancer: 100 year report. *Am J Respir Cell Mol Biol* **33:** 216-223

Minna JD, Fong K, Zochbauer-Muller S, Gazdar AF (2002) Molecular pathogenesis of lung cancer and potential translational applications. *Cancer J* **8 Suppl 1:** S41-46

Mounawar M, Mukeria A, Le Calvez F, Hung RJ, Renard H, Cortot A, Bollart C, Zaridze D, Brennan P, Boffetta P, Brambilla E, Hainaut P (2007) Patterns of EGFR, HER2, TP53, and KRAS mutations of p14arf expression in non-small cell lung cancers in relation to smoking history. *Cancer Res* **67:** 5667-5672

Mulero-Navarro S, Esteller M (2008) Epigenetic biomarkers for human cancer: the time is now. *Crit Rev Oncol Hematol* **68:** 1-11

Mulshine JL, Hirsch FR (2003) Lung cancer chemoprevention: moving from concept to a reality. *Lung cancer (Amsterdam, Netherlands)* **41 Suppl 1:** S163-174

Mulshine JL, Sullivan DC (2005) Clinical practice. Lung cancer screening. *The New England journal of medicine* **352:** 2714-2720

Mulshine JL, van Klaveren RJ (2011) Lung cancer screening: what is the benefit and what do we do about it? *Lung cancer (Amsterdam, Netherlands)* **71:** 247-248

Nakatsuka S, Oji Y, Horiuchi T, Kanda T, Kitagawa M, Takeuchi T, Kawano K, Kuwae Y, Yamauchi A, Okumura M, Kitamura Y, Oka Y, Kawase I, Sugiyama H, Aozasa K (2006) Immunohistochemical detection of WT1 protein in a variety of cancer cells. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **19:** 804-814

Narsule CK, Ebright MI, Fernando HC (2011) Sublobar versus lobar resection: current status. *Cancer J* **17:** 23-27

Nigg EA (2002) Centrosome aberrations: cause or consequence of cancer progression? *Nat Rev Cancer* **2:** 815-825

Noma K, Allis CD, Grewal SI (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science (New York, NY* **293:** 1150-1155

Ohno A (1996) [p16/CDKN2/MTS1 gene abnormality in lung cancer]. *Nihon Rinsho* **54:** 503-506

Oji Y, Miyoshi S, Maeda H, Hayashi S, Tamaki H, Nakatsuka S, Yao M, Takahashi E, Nakano Y, Hirabayashi H, Shintani Y, Oka Y, Tsuboi A, Hosen N, Asada M, Fujioka T, Murakami M, Kanato K, Motomura M, Kim EH et al (2002) Overexpression of the Wilms' tumour gene WT1 in de novo lung cancers. *International journal of cancer* **100:** 297-303

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99:** 247-257

Ordonez NG (2000) Value of thyroid transcription factor-1, E-cadherin, BG8, WT1, and CD44S immunostaining in distinguishing epithelial pleural mesothelioma from pulmonary and nonpulmonary adenocarcinoma. *The American journal of surgical pathology* **24:** 598-606

Ortega A, Niksic M, Bachi A, Wilm M, Sanchez L, Hastie N, Valcarcel J (2003) Biochemical function of female-lethal (2)D/Wilms' tumour suppressor-1-associated proteins in alternative pre-mRNA splicing. *J Biol Chem* **278:** 3040-3047

Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, Belinsky SA (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer research* **60:** 5954-5958

Paul TA, Bies J, Small D, Wolff L (2010) Signatures of polycomb repression and reduced H3K4 trimethylation are associated with p15INK4b DNA methylation in AML. *Blood* **115:** 3098-3108

Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y (2001) Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* **93:** 1054-1061

Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* **28:** 1057-1068

Qiu C, Sawada K, Zhang X, Cheng X (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat Struct Biol* **9:** 217-224

Ramalingam S, Dobbs TW, Einzig AI, Wojtowicz-Praga S, Cascino M, Bonomi P, Belani CP (2004) Carboplatin and docetaxel in advanced non-small-cell lung cancer: results of a multicenter phase II study. *Cancer chemotherapy and pharmacology* **53:** 439-444

Rami-Porta R, Chansky K, Goldstraw P (2009) Updated lung cancer staging system. *Future Oncol* **5:** 1545-1553

Rauch T, Li H, Wu X, Pfeifer GP (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res* **66:** 7939-7947

Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, Kernstine KH, Riggs AD, Pfeifer GP (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A* **104:** 5527-5532

Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, Riggs AD, Pfeifer GP (2008) High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 252-257

Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science (New York, NY* **293:** 1089-1093

Ren J, Singh BN, Huang Q, Li Z, Gao Y, Mishra P, Hwa YL, Li J, Dowdy SC, Jiang SW (2011) DNA hypermethylation as a chemotherapy target. *Cell Signal* **23:** 1082-1093

Risch A, Plass C (2008) Lung cancer epigenetics and genetics. *International journal of cancer* **123:** 1-7

Robertson KD (2005) DNA methylation and human disease. *Nat Rev Genet* **6:** 597-610

Russo AL, Thiagalingam A, Pan H, Califano J, Cheng KH, Ponte JF, Chinnappan D, Nemani P, Sidransky D, Thiagalingam S (2005) Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer. *Clin Cancer Res* **11:** 2466-2470

Satoh Y, Nakagawachi T, Nakadate H, Kaneko Y, Masaki Z, Mukai T, Soejima H (2003) Significant reduction of WT1 gene expression, possibly due to epigenetic alteration in Wilms' tumour. *J Biochem* **133:** 303-308

Schmidt B, Liebenberg V, Dietrich D, Schlegel T, Kneip C, Seegebarth A, Flemming N, Seemann S, Distler J, Lewin J, Tetzner R, Weickmann S, Wille U, Liloglou T, Raji O, Walshaw M, Fleischhacker M, Witt C, Field JK (2010) SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer* **10:** 600

Schmiemann V, Bocking A, Kazimirek M, Onofre AS, Gabbert HE, Kappes R, Gerharz CD, Grote HJ (2005) Methylation assay for the diagnosis of lung cancer on bronchial aspirates: a cohort study. *Clin Cancer Res* **11:** 7728-7734

Schneider KU, Dietrich D, Fleischhacker M, Leschber G, Merk J, Schaper F, Stapert HR, Vossenaar ER, Weickmann S, Liebenberg V, Kneip C, Seegebarth A, Erdogan F, Rappold G, Schmidt B (2011) Correlation of SHOX2 gene amplification and DNA methylation in lung cancer tumours. *BMC Cancer* **11:** 102

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132:** 887-898

Schumacher A, Weinhausl A, Petronis A (2008) Application of microarrays for DNA methylation profiling. *Methods Mol Biol* **439:** 109-129

Shames DS, Minna JD, Gazdar AF (2007a) DNA methylation in health, disease, and cancer. *Curr Mol Med* **7:** 85-102

Shames DS, Minna JD, Gazdar AF (2007b) Methods for detecting DNA methylation in tumours: from bench to bedside. *Cancer Lett* **251:** 187-198

Shapiro GI, Rollins BJ (1996) p16INK4A as a human tumour suppressor. *Biochim Biophys Acta* **1242:** 165-169

Sharma P, Kumar J, Garg G, Kumar A, Patowary A, Karthikeyan G, Ramakrishnan L, Brahmachari V, Sengupta S (2008) Detection of altered global DNA methylation in coronary artery disease patients. *DNA Cell Biol* **27:** 357-365

Sharma S, Kelly TK, Jones PA (2010) Epigenetics in cancer. *Carcinogenesis* **31:** 27-36

Shaw RJ, Liloglou T, Rogers SN, Brown JS, Vaughan ED, Lowe D, Field JK, Risk JM (2006) Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. *Br J Cancer* **94:** 561-568

Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba, II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, Herz J, Minna JD, Gazdar AF (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* **97:** 339-346

Shivapurkar N, Stastny V, Suzuki M, Wistuba, II, Li L, Zheng Y, Feng Z, Hol B, Prinsen C, Thunnissen FB, Gazdar AF (2007) Application of a methylation gene panel by quantitative PCR for lung cancers. *Cancer Lett* **247:** 56-71

Shivapurkar N, Stastny V, Xie Y, Prinsen C, Frenkel E, Czerniak B, Thunnissen FB, Minna JD, Gazdar AF (2008) Differential Methylation of a Short CpG-Rich Sequence within Exon 1 of TCF21 Gene: A Promising Cancer Biomarker Assay. *Cancer Epidemiol Biomarkers Prev* **17:** 995-1000

Shivapurkar N, Takahashi T, Reddy J, Zheng Y, Stastny V, Collins R, Toyooka S, Suzuki M, Parikh G, Asplund S, Kroft SH, Timmons C, McKenna RW, Feng Z, Gazdar AF (2004) Presence of simian virus 40 DNA sequences in human lymphoid and hematopoietic malignancies and their relationship to aberrant promoter methylation of multiple genes. *Cancer research* **64:** 3757-3760

Sidransky D (2002) Emerging molecular markers of cancer. *Nature reviews* **2:** 210-219

Sidransky D, Messing E (1992) Molecular genetics and biochemical mechanisms in bladder cancer. Oncogenes, tumour suppressor genes, and growth factors. *Urol Clin North Am* **19:** 629-639

Sikking S, Liloglou T, Maloney P, Gosney JR, Field JK (2003) In-depth analysis of molecular alterations within normal and tumour tissue from an entire bronchial tree. *Int J Oncol* **22:** 589-595

Singal R, Ginder GD (1999) DNA methylation. *Blood* **93:** 4059-4070

Slebos RJ, Kibbelaar RE, Dalesio O, Kooistra A, Stam J, Meijer CJ, Wagenaar SS, Vanderschueren RG, van Zandwijk N, Mooi WJ, et al. (1990) K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *The New England journal of medicine* **323:** 561-565

Smith E, De Young NJ, Pavey SJ, Hayward NK, Nancarrow DJ, Whiteman DC, Smithers BM, Ruszkiewicz AR, Clouston AD, Gotley DC, Devitt PG, Jamieson GG, Drew PA (2008a) Similarity of aberrant DNA methylation in Barrett's esophagus and esophageal adenocarcinoma. *Mol Cancer* **7:** 75

Smith E, De Young NJ, Tian ZQ, Caruso M, Ruszkiewicz AR, Liu JF, Jamieson GG, Drew PA (2008b) Methylation of TIMP3 in esophageal squamous cell carcinoma. *World J Gastroenterol* **14:** 203-210

Smolarek I, Wyszko E, Barciszewska AM, Nowak S, Gawronska I, Jablecka A, Barciszewska MZ (2010) Global DNA methylation changes in blood of patients with essential hypertension. *Med Sci Monit* **16:** CR149-155

Sonpavde G, Goldman BH, Speights VO, Lerner SP, Wood DP, Vogelzang NJ, Trump DL, Natale RB, Grossman HB, Crawford ED (2009) Quality of pathologic response and surgery correlate with survival for patients with completely resected bladder cancer after neoadjuvant chemotherapy. *Cancer* **115:** 4104-4109

Sozzi G, Miozzo M, Pastorino U, Pilotti S, Donghi R, Giarola M, De Gregorio L, Manenti G, Radice P, Minoletti F, et al. (1995) Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. *Cancer research* **55:** 135-140

Sterlacci W, Tzankov A, Veits L, Zelger B, Bihl MP, Foerster A, Augustin F, Fiegl M, Savic S (2011) A comprehensive analysis of p16 expression, gene status, and promoter hypermethylation in surgically resected non-small cell lung carcinomas. *J Thorac Oncol* **6:** 1649-1657

Stewart DJ, Issa JP, Kurzrock R, Nunez MI, Jelinek J, Hong D, Oki Y, Guo Z, Gupta S, Wistuba, II (2009) Decitabine effect on tumour global DNA methylation and other parameters in a phase I trial in refractory solid tumours and lymphomas. *Clin Cancer Res* **15:** 3881-3888

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41-45

Strathdee G, Brown R (2002) Aberrant DNA methylation in cancer: potential clinical interventions. *Expert reviews in molecular medicine* **4:** 1-17

Sugimura H, Weston A, Caporaso NE, Shields PG, Bowman ED, Metcalf RA, Harris CC (1991) Biochemical and molecular epidemiology of cancer. *Biomed Environ Sci* **4:** 73-92

Sun S, Schiller JH, Gazdar AF (2007) Lung cancer in never smokers--a different disease. *Nature reviews* **7:** 778-790

Suzuki M, Shigematsu H, Iizasa T, Hiroshima K, Nakatani Y, Minna JD, Gazdar AF, Fujisawa T (2006) Exclusive mutation in epidermal growth factor receptor gene, HER-2, and KRAS, and synchronous methylation of nonsmall cell lung cancer. *Cancer* **106:** 2200-2207

Tessema M, Yu YY, Stidley CA, Machida EO, Schuebel KE, Baylin SB, Belinsky SA (2009) Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. *Carcinogenesis* **30:** 1132-1138

Tong AW (2006) Small RNAs and non-small cell lung cancer. *Curr Mol Med* **6:** 339-349

Topaloglu O, Hoque MO, Tokumaru Y, Lee J, Ratovitski E, Sidransky D, Moon CS (2004) Detection of promoter hypermethylation of multiple genes in the tumour and bronchoalveolar lavage of patients with lung cancer. *Clin Cancer Res* **10:** 2284-2288

Tost J, El Abdalaoui H, Gut IG (2006) Serial pyrosequencing for quantitative DNA methylation analysis. *BioTechniques* **40:** 721-722, 724, 726

Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* **2:** 2265-2275

Toyooka S, Tokumo M, Shigematsu H, Matsuo K, Asano H, Tomii K, Ichihara S, Suzuki M, Aoe M, Date H, Gazdar AF, Shimizu N (2006) Mutational and epigenetic evidence for independent pathways for lung adenocarcinomas arising in smokers and never smokers. *Cancer Res* **66:** 1371-1375

Travis WD, Giroux DJ, Chansky K, Crowley J, Asamura H, Brambilla E, Jett J, Kennedy C, Rami-Porta R, Rusch VW, Goldstraw P (2008) The IASLC Lung Cancer Staging Project: proposals for the inclusion of broncho-pulmonary carcinoid tumours in the forthcoming (seventh) edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* **3:** 1213-1223

Trichopoulos D, Li FP, Hunter DJ (1996) What causes cancer? *Scientific American* **275:** 80-87

Tsao AS, Tang XM, Sabloff B, Xiao L, Shigematsu H, Roth J, Spitz M, Hong WK, Gazdar A, Wistuba I (2006) Clinicopathologic characteristics of the EGFR gene mutation in non-small cell lung cancer. *J Thorac Oncol* **1:** 231-239

Tsou JA, Galler JS, Siegmund KD, Laird PW, Turla S, Cozen W, Hagen JA, Koss MN, Laird-Offringa IA (2007) Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. *Mol Cancer* **6:** 70

Tsui DW, Chan KC, Chim SS, Chan LW, Leung TY, Lau TK, Lo YM, Chiu RW (2007) Quantitative aberrations of hypermethylated RASSF1A gene sequences in maternal plasma in pre-eclampsia. *Prenat Diagn* **27:** 1212-1218

Turek-Plewa J, Jagodzinski PP (2005) The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett* **10:** 631-647

Ulivi P, Zoli W, Calistri D, Fabbri F, Tesei A, Rosetti M, Mengozzi M, Amadori D (2006) p16INK4A and CDH13 hypermethylation in tumour and serum of non-small cell lung cancer patients. *J Cell Physiol* **206:** 611-615

Vaissiere T, Cuenin C, Paliwal A, Vineis P, Hoek G, Krzyzanowski M, Airoldi L, Dunning A, Garte S, Hainaut P, Malaveille C, Overvad K, Clavel-Chapelon F, Linseisen J, Boeing H, Trichopoulou A, Trichopoulos D, Kaladidi A, Palli D, Krogh V et al (2009) Quantitative analysis of DNA methylation after whole bisulfitome amplification of a minute amount of DNA from body fluids. *Epigenetics* **4:** 221-230

Vallieres E, Shepherd FA, Crowley J, Van Houtte P, Postmus PE, Carney D, Chansky K, Shaikh Z, Goldstraw P (2009) The IASLC Lung Cancer Staging Project: proposals regarding the relevance of TNM in the pathologic staging of small cell lung cancer in the forthcoming (seventh) edition of the TNM classification for lung cancer. *J Thorac Oncol* **4:** 1049-1059

van der Drift M, Prinsen C, Knuiman J, Janssen J, Dekhuijzen PN, Thunnissen E (2011) Diagnosing peripheral lung cancer: the additional value of RASSF1A methylation and KRAS mutation analyses in washings in non-diagnostic bronchoscopy. *Chest*

van der Drift MA, Prinsen CF, Hol BE, Bolijn AS, Jeunink MA, Dekhuijzen PN, Thunnissen FB (2008) Can free DNA be detected in sputum of lung cancer patients? *Lung cancer (Amsterdam, Netherlands)* **61:** 385-390

van Zandwijk N, Mathy A, Boerrigter L, Ruijter H, Tielen I, de Jong D, Baas P, Burgers S, Nederlof P (2007) EGFR and KRAS mutations as criteria for treatment with tyrosine kinase inhibitors: retro- and prospective observations in non-small-cell lung cancer. *Ann Oncol* **18:** 99-103

Verri C, Roz L, Conte D, Liloglou T, Livio A, Vesin A, Fabbri A, Andriani F, Brambilla C, Tavecchio L, Calarco G, Calabro E, Mancini A, Tosi D, Bossi P, Field JK, Brambilla E, Sozzi G (2009) Fragile histidine triad gene inactivation in lung cancer: the European Early Lung Cancer project. *Am J Respir Crit Care Med* **179:** 396-401

Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM (2006) A microRNA expression signature of human solid tumours defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 2257-2261

Wang J, Sasco AJ, Fu C, Xue H, Guo G, Hua Z, Zhou Q, Jiang Q, Xu B (2008) Aberrant DNA methylation of P16, MGMT, and hMLH1 genes in combination with MTHFR C677T genetic polymorphism in esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* **17:** 118-125

Wang Q, Yan J, Chen X, Li J, Yang Y, Weng J, Deng C, Yenari MA (2011) Statins: multiple neuroprotective mechanisms in neurodegenerative diseases. *Exp Neurol* **230:** 27-34

Wang YC, Hsu HS, Chen TP, Chen JT (2006) Molecular diagnostic markers for lung cancer in sputum and plasma. *Ann N Y Acad Sci* **1075:** 179-184

Webber RJ, Rodriguez JG, Webber DS, Dunnebacke TH (2005) Development, characterization, and epitope mapping of a panel of twenty-four monoclonal antibodies specific for human inducible nitric oxide synthase. *Hybridoma (Larchmt)* **24:** 6-13

Westra WH (2000) Early glandular neoplasia of the lung. *Respir Res* **1:** 163-169

Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, Stegmaier C, Jacobs IJ, Brenner H (2008) Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS One* **3:** e2656

Wilhelm CS, Kelsey KT, Butler R, Plaza S, Gagne L, Zens MS, Andrew AS, Morris S, Nelson HH, Schned AR, Karagas MR, Marsit CJ (2010) Implications of LINE1 methylation for bladder cancer risk in women. *Clin Cancer Res* **16:** 1682-1689

Wistuba, II (2005) Histologic evaluation of bronchial squamous lesions: any role in lung cancer risk assessment? *Clin Cancer Res* **11:** 1358-1360

Wistuba, II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW, 3rd, Garner HR, McKay B, Latif F, Lerman MI, Lam S, Gazdar AF, Minna JD (2000) High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer research* **60:** 1949-1960

Witcher M, Emerson BM (2009) Epigenetic silencing of the p16(INK4a) tumour suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol Cell* **34:** 271-284

Wong MP, Fung LF, Wang E, Chow WS, Chiu SW, Lam WK, Ho KK, Ma ES, Wan TS, Chung LP (2003) Chromosomal aberrations of primary lung adenocarcinomas in nonsmokers. *Cancer* **97:** 1263-1270

Wynder EL, Graham EA (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. *J Am Med Assoc* **143:** 329-336

Xinarianos G, McRonald FE, Risk JM, Bowers NL, Nikolaidis G, Field JK, Liloglou T (2006) Frequent genetic and epigenetic abnormalities contribute to the deregulation of cytoglobin in non-small cell lung cancer. *Hum Mol Genet* **15:** 2038-2044

Xinarianos G, Scott FM, Liloglou T, Prime W, Turnbull L, Walshaw M, Field JK (2000) Evaluation of telomerase activity in bronchial lavage as a potential diagnostic marker for malignant lung disease. *Lung cancer (Amsterdam, Netherlands)* **28:** 37-42

Xiong Z, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* **25:** 2532-2534

Xu D, Popov N, Hou M, Wang Q, Bjorkholm M, Gruber A, Menkel AR, Henriksson M (2001) Switch from Myc/Max to Mad1/Max binding and decrease in histone acetylation at the telomerase reverse transcriptase promoter during differentiation of HL60 cells. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 3826-3831

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13:** 335-340

Yu IT, Chiu YL, Au JS, Wong TW, Tang JL (2006) Dose-response relationship between cooking fumes exposures and lung cancer among Chinese nonsmoking women. *Cancer research* **66:** 4961-4967

Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD (2001) Aberrant promoter methylation of multiple genes in non-small cell lung cancers. *Cancer Res* **61:** 249-255

Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, Minna JD, Gazdar AF (2003) Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *International journal of cancer* **107:** 612-616

Zochbauer-Muller S, Minna JD, Gazdar AF (2002) Aberrant DNA methylation in lung cancer: biological and clinical implications. *Oncologist* **7:** 451-457

Appendix 1. Raw data from qMSP in training and test set with clinicopathological data

MPI	Sex	Age	Set	Status	Cytology	Diagnosis	Type	T	N	Smoking	CYGB	RARb	CDH13	DAPK	p73	p16	RASSF1	TERT	TMEFF	WT1
53866	F	64	T	Case	Sus	AdenoCa	Sec	4	2	Cur	0	0	0	0	0	0	0	0	0	0
63128	M	81	T	Case	Pos	SqCCL	Pri	4	0	Cur	0	0	0	0	0	0	0	1	0	0
43071	F	76	T	Case	Pos	AdenoCa	Pri	4	2	Cur	0	0	0	1	0	0	0	0	0	1
68189	F	63	T	Case	Neg	AdenoCa	Pri	2	3	Cur	0	0	0	0	0	0	0	0	0	1
53918	F	72	T	Case	Neg	AdenoCa	Pri	1	0	Cur	0	0	0	0	0	0	0	1	0	0
59740	M	64	T	Case	Neg	AdenoCa	Pri	1	0	Cur	0	0	0	1	0	0	0	0	0	0
68196	M	68	T	Case	Sus	AdenoCa	Pri	1	0	Cur	0	0	0	0	0	0	0	1	0	0
73439	F	72	T	Case	Neg	AdenoCa	Pri	2	1	Cur	0	0	1	0	1	0	1	1	1	0
43971	F	75	T	Case	Neg	SqCCL	Sec	2	0	For	0	0	0	0	0	0	0	0	0	0
36338	M	59	T	Case	Neg	SqCCL	Sec	1	0	For	0	0	0	0	0	0	0	1	0	0
42415	M	79	T	Case	Neg	AdenoCa	Pri	4	1	For	0	0	0	0	0	0	0	0	0	0
68187	M	59	T	Case	Neg	AdenoCa	Pri	1		For	1	0	0	0	0	0	0	0	0	0
10544	M	60	T	Case	Pos	SqCCL	Pri	1	0	For	0	0	0	0	0	0	0	1	0	0
25103	F	61	T	Case	Pos	SqCCL	Pri	1	0	For	0	0	0	0	0	0	0	1	0	0
34914	M	72	T	Case	Sus	SqCCL	Met	1	2	For	0	1	0	0	0	0	0	0	0	0
60975	M	65	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	1	0	0	1	0	0
4043	F	85	T	Case	Neg	AdenoCa	Pri	4	2	For	0	0	0	0	0	0	0	0	0	0
37995	F	69	T	Case	Neg	AdenoCa	Pri	2	0	For	0	0	0	0	0	0	0	0	0	1
68232	F	78	T	Case	Pos	AdenoCa	Pri	4	1	For	0	0	0	0	0	0	0	1	0	0
40962	M	75	T	Case	Neg	NSCLC	Pri	2	2	For	0	0	0	0	0	0	0	1	0	0
25050	M	71	T	Case	Neg	SqCCL	Pri	2	0	For	0	0	0	0	0	0	0	1	0	0
65233	M	72	T	Case	Neg	Lung Ca	Met			For	0	0	1	0	1	0	0	0	0	0
68382	F	71	T	Case	Neg	NSCLC	Pri	3	2	For	0	0	0	0	0	0	0	1	0	0
68343	F	66	T	Case	Pos	SqCCL	Pri	2	3	For	0	0	0	0	0	0	0	0	0	0
53894	M	83	T	Case	Neg	Lung Ca	Sec			For	0	0	0	0	0	0	0	0	0	0
68373	M	50	T	Case	Pos	SqCCL	Pri	2	2	For	1	0	0	0	0	0	0	0	0	0
28654	M	61	T	Case	Neg	AdenoCa	Pri	1	1	For	0	0	0	0	0	0	0	0	0	0
65655	F	70	T	Case	Neg	AdenoCa	Met	2	0	For	0	0	0	0	0	0	0	0	0	0
24274	F	64	T	Case	Neg	AdenoCa	Pri	4	3	For	0	0	1	0	0	0	0	1	0	0

19874	M	82	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	1	0	0	1	0	0
45282	M	87	T	Case	Neg	AdenoCa	Pri	2	1	For	0	0	0	0	0	0	1	0	0	0
41192	F	78	T	Case	Neg	AdenoCa	Pri	1	1	For	0	0	0	0	0	0	0	0	0	0
19550	M	76	T	Case	Neg	SqCCL	Pri	2	0	For	0	1	1	0	0	0	0	0	0	0
783	F	64	T	Case	Neg	AdenoCa	Met			For	0	0	0	0	0	0	0	1	0	0
42419	M	78	T	Case	Neg	AdenoCa	Pri	1	1	For	0	0	0	0	0	0	0	1	0	0
22196	M	67	T	Case	Neg	AdenoCa	Pri	2	2	For	0	0	0	0	1	0	1	0	0	0
65663	F	63	T	Case	Neg	NSCLC	Pri	4	3	For	0	0	0	0	0	0	1	0	0	0
40968	F	58	T	Case	Neg	AdenoCa	Pri	2	0	For	0	0	0	0	0	0	0	0	0	0
68275	F	65	T	Case	Neg	AdenoCa	Met	2	2	For	0	0	0	0	0	0	0	0	0	0
43074	F	73	T	Case	Pos	SqCCL	Pri	1	0	For	0	0	0	0	0	0	0	0	0	0
73479	F	82	T	Case	Neg	AdenoCa	Pri	1	0	For	0	0	0	0	0	0	0	0	0	1
43080	M	77	T	Case	Neg	AdenoCa	Pri	4	0	For	0	0	0	1	0	0	0	0	0	0
42413	M	77	T	Case	Neg	SqCCL	Pri	3	0	For	0	0	0	0	0	0	0	0	0	0
53924	F	75	T	Case	Neg	AdenoCa	Pri	1	2	For	0	0	0	0	0	0	1	0	0	0
63191	M	67	T	Case	Neg	AdenoCa	Pri	1	0	For	0	0	0	0	0	0	0	1	0	0
40941	F	57	T	Case	Neg	SqCCL	Pri	1	0	For	0	0	0	0	0	0	0	0	0	0
19582	M	56	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	0	0	0	0	0	1
63149	M	79	T	Case	Neg	AdenoCa	Pri	2	0	For	0	0	0	0	0	0	0	0	0	0
42411	F	72	T	Case	Neg	AdenoCa	Pri	1	2	For	0	0	0	0	0	0	0	0	0	1
19873	M	69	T	Case	Neg	AdenoCa	Pri	2	0	For	0	0	0	0	0	0	0	1	0	0
63187	M	73	T	Case	Neg	Lung Ca	Met			For	0	0	0	0	0	0	0	0	0	0
68299	M	57	T	Case	Neg	SqCCL	Pri	1	0	For	0	0	0	0	0	0	0	0	0	0
1359	M	70	T	Case	Neg	SqCCL	Sec	2	0	For	1	0	0	0	0	1	1	1	0	0
58700	M	75	T	Case	Pos	SqCCL	Sec	1	1	For	0	0	0	0	0	0	1	1	0	0
16192	M	66	T	Case	Pos	AdenoCa	Sec	4	3	For	0	0	0	0	0	0	1	1	0	0
40950	M	85	T	Case	Neg	NSCLC	Pri	4	1	For	0	1	0	0	0	1	0	0	0	1
43965	F	60	T	Case	Neg	SqCCL	Pri	3	2	For	0	0	1	0	0	0	1	1	0	0
19865	F	68	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	0	1	1	1	0	1
19898	M	67	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	1	0	0	1	1	0	0

43076	F	77	T	Case	Pos	SqCCL	Pri	4	3	For	1	0	1	0	0	1	1	1	0	0
43978	M	72	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	1	1	0	1	0	1
47954	M	61	T	Case	Pos	AdenoCa	Pri	4	0	For	0	1	0	0	0	0	0	0	0	1
73521	M	72	T	Case	Pos	SqCCL	Pri			For	0	0	0	0	0	0	1	1	0	0
30684	M	71	T	Case	Neg	AdenoCa	Pri	2	1	For	0	1	0	0	0	0	1	1	0	0
9064	M	75	T	Case	Pos	SqCCL	Pri	4	0	For	0	0	0	0	0	0	0	1	0	1
45272	M	55	T	Case	Pos	SqCCL	Pri	2	2	For	0	0	0	0	0	1	0	1	0	0
63160	F	71	T	Case	Pos	SqCCL	Pri	2	2	For	0	0	0	0	0	0	1	1	0	0
19573	M	68	T	Case	Pos	AdenoCa	Pri	1	0	For	0	1	0	1	1	0	0	1	0	1
26202	M	61	T	Case	Pos	SqCCL	Pri	2	1	For	0	0	0	0	0	0	1	1	0	1
68251	M	72	T	Case	Sus	SqCCL	Pri	2	0	For	1	0	0	0	1	0	1	1	0	1
19612	F	75	T	Case	Neg	AdenoCa	Pri	2	2	For	0	0	0	0	0	0	0	1	0	1
39106	F	60	T	Case	Sus	SqCCL	Met	4	2	For	0	1	0	0	0	0	0	1	0	1
26211	F	68	T	Case	Pos	SqCCL	Pri	1	0	For	0	0	0	0	0	0	1	1	0	0
68214	M	72	T	Case	Pos	SqCCL	Met	2	1	For	1	0	0	0	0	0	1	1	0	1
65247	F	75	T	Case	Neg	AdenoCa	Pri	2	2	For	1	0	0	0	0	1	0	1	0	1
67928	M	73	T	Case	Neg	NSCLC	Met	2	0	For	0	0	0	0	0	1	0	1	0	0
28676	F	72	T	Case	Neg	NSCLC	Pri	4	2	For	0	1	0	0	0	0	1	0	0	1
19591	M	73	T	Case	Neg	SqCCL	Pri	2	1	For	0	0	0	0	1	0	1	1	0	0
45284	F	62	T	Case	Neg	AdenoCa	Pri	3	0	For	0	0	0	0	0	0	1	0	0	1
19533	F	79	T	Case	Neg	SqCCL	Pri	3	2	For	1	0	0	0	0	1	0	0	0	0
65656	M	79	T	Case	Pos	Lung Ca	Sec			For	0	1	1	0	0	0	1	1	0	1
63205	F	77	T	Case	Neg	AdenoCa	Pri	3	2	For	0	0	0	0	0	0	1	1	0	1
53839	M	72	T	Case	Neg	NSCLC	Pri	3	1	For	0	0	0	0	0	1	0	1	0	1
63186	M	62	T	Case	Neg	NSCLC	Pri	2	0	For	0	1	1	0	1	1	1	1	1	1
59704	M	56	T	Case	Neg	NSCLC	Pri	2	2	For	0	0	0	0	0	0	0	1	0	1
26217	F	76	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	0	0	1	1	0	0
22526	M	60	T	Case	Sus	SqCCL	Pri	2	1	For	1	0	1	0	0	1	1	1	1	1
28664	M	60	T	Case	Neg	SqCCL	Pri	2	0	For	0	0	0	0	1	1	0	1	0	0
19564	F	56	T	Case	Neg	AdenoCa	Pri	2	0	For	1	0	1	0	1	0	1	1	0	1

47947	M	78	T	Case	Neg	AdenoCa	Pri	3	1	For	0	0	0	0	0	0	0	1	0	1
58689	F	67	T	Case	Neg	SqCCL	Pri	2	0	For	0	0	0	0	0	0	1	1	0	0
43964	M	73	T	Case	Sus	SqCCL	Pri	2	1	For	0	0	1	0	0	0	1	1	1	0
7630	F	71	T	Case	Pos	SqCCL	Pri	2	0	For	0	0	0	0	0	1	0	1	0	0
53896	F	67	T	Case	Pos	AdenoCa	Pri	2	1	For	0	0	0	0	0	1	1	1	1	1
19856	M	75	T	Case	Neg	SCLC	Pri	2	1	For	0	0	0	0	0	0	1	1	0	0
68179	M	64	T	Case	Neg	SqCCL	Pri	2	0	For	0	0	1	0	0	1	1	1	0	1
19916	F	77	T	Case	Neg	SCLC	Met	4	2	For	0	0	0	0	1	0	1	1	0	1
47949	F	76	T	Case	Neg	NSCLC	Pri	4	0	For	0	0	0	0	0	0	1	1	0	1
68329	M	66	T	Case	Neg	AdenoCa	Pri	2	3	For	1	0	0	0	1	0	1	1	0	1
7078	F	67	T	Case	Neg	AdenoCa	Pri	2		For	0	0	0	0	0	0	0	1	0	1
53915	F	60	T	Case	Neg	AdenoCa	Pri	3	2	For	1	0	0	0	0	0	0	1	1	0
73443	M	75	T	Case	Neg	AdenoCa	Pri	2	0	For	0	1	0	0	0	0	0	0	0	1
7829	M	68	T	Case	Sus	SqCCL	Pri	2	1	For	1	0	0	0	0	1	1	1	0	1
60778	M	63	T	Case	Pos	Lung Ca	Pri	4	2	For	0	0	0	0	0	0	0	1	0	1
40981	M	68	T	Case	Sus	SqCCL	Pri	4	2	For	1	0	0	1	1	0	0	1	0	1
65661	F	64	T	Case	Neg	AdenoCa	Pri	1	0	For	0	1	0	0	0	0	0	1	0	0
63189	M	63	T	Case	Neg	AdenoCa	Pri	2	0	For	0	0	0	0	0	1	0	1	0	1
59710	M	70	T	Case	Sus	Lung Ca	Met	2	0	For	1	0	0	0	0	1	1	1	0	1
1420	M	56	T	Case	Pos	SqCCL	Pri	1	2	For	1	0	1	0	0	0	0	1	0	0
68304	F	78	T	Case	Neg	AdenoCa	Pri	2	1	For	1	0	0	0	1	0	0	1	0	1
68424	F	66	T	Case	Sus	SqCCL	Pri	2	2	For	1	0	0	0	0	1	0	0	0	1
63197	M	60	T	Case	Neg	AdenoCa	Sec	4	3	Non	0	0	0	0	0	0	0	0	0	0
6792	M	63	T	Case	Neg	SqCCL	Sec	3	2	Non	0	0	0	0	0	0	0	1	0	0
36011	M	71	T	Case	Neg	AdenoCa	Sec	2	0	Non	0	0	0	0	0	0	0	0	0	0
63190	F	68	T	Case	Neg	NSCLC	Pri	2	2	Non	0	0	0	0	0	0	0	0	0	0
5461	M	79	T	Case	Neg	SqCCL	Met	3	0	Non	0	0	0	0	1	0	0	0	0	0
14290	M	58	T	Case	Neg	SqCCL	Pri	4	0	Non	0	0	0	0	0	0	0	0	0	1
45287	F	60	T	Case	Neg	NSCLC	Pri	4	2	Non	0	0	0	0	0	0	0	1	0	0
53892	M	74	T	Case	Neg	SqCCL	Pri	2	0	Non	0	0	1	0	0	0	0	1	0	0

63173	F	66	T	Case	Neg	SqCCL	Pri			Non	0	0	0	0	0	1	0	0	0	0
68372	M	62	T	Case	Pos	SqCCL	Pri	4	2	Non	0	0	0	1	1	0	0	1	0	0
43077	M	61	T	Case	Pos	SqCCL	Pri	3	0	Non	0	0	0	0	0	0	1	0	0	0
73446	M	73	T	Case	Neg	SqCCL	Pri	3	2	Non	0	0	0	0	0	0	0	1	1	0
43979	M	46	T	Case	Neg	NSCLC	Pri	1	2	Non	0	0	0	0	0	0	0	0	0	1
45274	F	73	T	Case	Pos	SqCCL	Pri	4	0	Non	0	1	0	0	0	0	0	0	0	0
25037	M	67	T	Case	Pos	AdenoCa	Met	4	2	Non	0	0	0	0	0	0	0	1	0	0
67908	M	66	T	Case	Neg	SqCCL	Pri	4	2	Non	0	0	0	0	0	0	0	0	0	0
19592	M	60	T	Case	Pos	SqCCL	Pri	3	0	Non	0	0	0	0	0	1	0	0	0	0
59708	M	72	T	Case	Neg	AdenoCa	Pri	1	2	Non	0	0	0	0	0	0	0	0	0	0
40956	F	65	T	Case	Neg	SqCCL	Pri	2	0	Non	0	0	0	0	0	0	0	0	0	0
34893	M	73	T	Case	Neg	AdenoCa	Pri	1	0	Non	0	1	0	0	0	0	0	0	0	0
19549	M	69	T	Case	Neg	AdenoCa	Met	2	1	Non	0	0	0	0	0	0	0	1	0	0
53864	F	81	T	Case	Neg	SqCCL	Pri	2	0	Non	0	0	0	0	0	0	0	1	0	0
63157	F	59	T	Case	Neg	AdenoCa	Pri	1	0	Non	0	0	0	0	0	0	0	0	0	0
63152	F	76	T	Case	Pos	Lung Ca	Pri	4	0	Non	0	0	0	0	0	0	0	1	1	0
45275	F	78	T	Case	Pos	SqCCL	Pri	1	0	Non	0	0	0	0	0	0	0	0	0	1
68212	F	66	T	Case	Neg	SqCCL	Pri	2	1	Non	0	0	0	0	0	0	0	0	0	0
53889	F	65	T	Case	Pos	AdenoCa	Pri	4	2	Non	0	0	1	1	0	0	0	0	0	0
53875	M	69	T	Case	Neg	AdenoCa	Pri	2	0	Non	0	0	1	0	0	0	0	1	1	0
59709	F	51	T	Case	Neg	Adeno Sq	Pri	2	0	Non	0	0	0	0	0	0	0	0	0	0
14387	M	74	T	Case	Neg	SqCCL	Sec	4	0	Non	1	0	0	0	0	0	0	1	0	1
59720	F	57	T	Case	Neg	AdenoCa	Sec	2	0	Non	1	0	0	0	0	0	0	0	0	1
73445	F	73	T	Case	Neg	Lung Ca	Sec			Non	0	0	0	0	0	1	0	1	0	1
23412	F	76	T	Case	Neg	SqCCL	Pri	4	2	Non	0	0	0	0	1	0	1	1	0	0
73482	M	70	T	Case	Neg	NSCLC	Pri			Non	0	0	0	0	0	1	1	1	0	0
718	M	67	T	Case	Pos	SqCCL	Pri	4	2	Non	0	0	0	1	0	0	1	1	0	0
11436	M	63	T	Case	Sus	NSCLC	Pri	4	0	Non	0	0	0	0	1	0	1	1	0	0
19888	M	59	T	Case	Pos	SqCCL	Pri	3	0	Non	1	0	1	1	0	0	1	1	0	1
23426	M	60	T	Case	Pos	SqCCL	Pri	2	2	Non	0	0	0	0	0	1	1	1	0	0

53871	M	81	T	Case	Pos	NSCLC	Pri	4	2	Non	1	0	1	0	0	1	1	1	0	1
59707	M	65	T	Case	Pos	SqCCL	Pri	2	2	Non	0	0	0	0	0	1	0	1	0	0
65243	M	66	T	Case	Pos	SqCCL	Pri	3	0	Non	0	0	0	0	0	1	1	0	0	0
68186	M	74	T	Case	Pos	SqCCL	Pri	2	0	Non	1	0	1	0	0	1	1	1	1	1
68332	M	74	T	Case	Pos	SqCCL	Pri	4	3	Non	0	1	0	0	0	0	0	1	0	0
68383	M	68	T	Case	Pos	SqCCL	Pri	2	2	Non	1	1	0	0	0	0	1	1	0	1
68399	F	59	T	Case	Pos	SqCCL	Pri	4	0	Non	1	1	1	0	0	1	1	1	0	0
73538	M	69	T	Case	Pos	Lung Ca	Pri	4	2	Non	0	0	0	0	0	1	1	1	0	0
19575	F	76	T	Case	Pos	AdenoCa	Pri	4	0	Non	0	0	0	0	1	0	1	1	0	0
40951	F	64	T	Case	Pos	SqCCL	Met	2	0	Non	1	1	0	0	0	0	1	1	0	0
68230	M	76	T	Case	Pos	SqCCL	Pri	3	2	Non	1	1	0	0	0	0	1	1	0	1
73480	M	53	T	Case	Pos	AdenoCa	Pri	4	3	Non	0	0	0	0	1	0	1	1	0	1
58693	M	58	T	Case	Pos	NSCLC	Pri	2	2	Non	0	0	0	0	0	0	0	1	1	1
19913	M	75	T	Case	Pos	SqCCL	Pri	4	0	Non	0	0	1	0	1	0	0	1	0	1
23431	M	63	T	Case	Pos	SqCCL	Pri	1		Non	1	0	0	0	0	1	0	1	0	1
23411	F	80	T	Case	Neg	AdenoCa	Pri	2	0	Non	0	0	0	0	0	0	1	1	0	0
21643	M	81	T	Case	Pos	SqCCL	Pri	2	2	Non	0	1	0	0	0	0	1	1	0	1
68283	F	75	T	Case	Neg	AdenoCa	Pri	2	1	Non	0	0	0	0	1	0	1	1	0	1
68341	M	65	T	Case	Neg	SqCCL	Sec	2	1	Non	0	0	0	0	1	0	0	1	0	1
68200	M	66	T	Case	Pos	SqCCL	Pri	4	3	Non	0	0	1	0	0	0	1	1	0	0
68233	M	64	T	Case	Pos	SqCCL	Pri	2	0	Non	0	0	0	0	0	0	0	1	0	1
53834	F	68	T	Case	Neg	AdenoCa	Pri	2	0	Non	0	0	0	0	0	0	1	1	0	1
30719	M	57	T	Case	Neg	SqCCL	Pri	4	2	Non	0	1	0	0	0	0	1	0	1	1
68285	F	73	T	Case	Pos	AdenoCa	Pri	1	0	Non	0	1	1	0	1	0	1	1	1	1
65239	M	68	T	Case	Sus	SqCCL	Pri	2	1	Non	1	0	0	0	0	0	0	1	0	1
68213	F	56	T	Case	Neg	SCLC	Pri	4	2	Non	1	1	0	0	1	0	1	1	0	0
53939	F	77	T	Case	Neg	SqCCL	Pri	2	0	Non	1	0	0	1	0	0	1	1	0	0
58716	M	73	T	Case	Neg	SCLC	Pri	2	2	Non	1	0	0	0	0	1	1	1	0	1
19557	F	65	T	Case	Neg	AdenoCa	Pri	2	1	Non	0	0	0	0	0	0	1	1	0	0
6767	M	57	T	Case	Pos	SqCCL	Pri	3	0	Non	0	1	1	0	0	0	0	1	0	1

68293	F	70	T	Case	Neg	AdenoCa	Pri	1	0	Non	0	0	0	0	0	0	1	1	0	0
68352	F	71	T	Case	Pos	AdenoCa	Pri	2	2	Non	0	0	1	0	0	1	0	1	0	0
59732	M	71	T	Case	Neg	Lung Ca	Pri	4	2	Non	1	0	1	0	1	0	1	1	1	1
53925	F	71	T	Case	Neg	SqCCL	Pri	4	2	Non	0	1	1	0	0	0	0	1	0	0
68277	F	70	T	Case	Neg	AdenoCa	Pri	1	1	Non	0	0	0	0	0	0	0	1	0	1
73525	M	60	T	Case	Sus	SqCCL	Pri	4	2	Non	0	1	0	0	0	1	1	1	0	0
19867	F	64	T	Case	Neg	Lung Ca	Pri	2	2		0	0	0	0	0	0	0	1	0	0
1255	F	79	T	Case	Pos	SqCCL	Pri				0	1	0	0	0	0	0	0	0	0
68356	M	75	T	Case	Neg	SCLC	Pri				0	0	1	0	0	1	1	1	0	0
78661	M	71	T	Case	Neg	SqCCL	Pri				1	0	0	0	0	0	1	1	0	0
73558	M	76	T	Case	Pos	SqCCL	Pri	1	2		1	0	0	0	0	0	1	1	0	0
73559	M	67	T	Case	Neg	SqCCL	Pri	1	0		1	1	1	0	0	0	0	1	0	0
73553	M	68	T	Case	Pos	SqCCL	Pri				0	0	0	0	0	0	1	1	0	0
73530	M	70	T	Case	Neg	Carcinoid	Pri				0	0	1	0	0	0	1	1	0	1
4275	F	75	T	Case	Neg	Carcinoid	Pri				0	0	0	0	0	0	1	1	0	0
19611	M	75	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
27547	F	60	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
42958	F	59	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
53847	F	60	T	Control	Neg					Cur	0	0	1	0	1	0	0	0	0	0
58690	F	74	T	Control	Neg					Cur	0	0	0	0	1	0	0	0	0	0
68180	M	69	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
68376	M	70	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
73477	F	65	T	Control	Neg					Cur	0	0	1	0	0	0	0	0	0	0
1320	F	56	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
19581	F	67	T	Control	Neg					Cur	0	0	0	0	1	0	0	0	0	0
19587	M	66	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0
30706	M	79	T	Control	Neg					Cur	1	0	0	0	0	0	0	0	0	0
30712	M	76	T	Control	Neg					Cur	0	0	0	1	0	0	0	1	0	0
34906	F	73	T	Control	Neg					Cur	1	0	0	0	0	0	0	0	0	0
38005	M	59	T	Control	Neg					Cur	0	0	0	0	0	0	0	0	0	0

40955	F	55	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
42402	M	57	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
43067	F	51	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
53835	M	69	T	Control	Neg		Cur	1	0	0	0	0	0	0	0	0	0
53843	M	60	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
54752	M	63	T	Control	Neg		Cur	0	1	1	0	0	0	0	0	0	0
55841	F	79	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
58711	M	73	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	1
59718	M	80	T	Control	Neg		Cur	0	0	0	0	1	0	0	0	0	0
60773	F	56	T	Control	Neg		Cur	0	0	0	0	0	0	0	1	0	0
60774	F	59	T	Control	Neg		Cur	0	0	1	0	0	0	0	0	0	0
63199	F	68	T	Control	Neg		Cur	1	0	0	0	0	0	0	1	0	0
65227	F	62	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
65657	F	73	T	Control	Neg		Cur	0	0	1	0	1	0	0	0	0	0
68181	M	69	T	Control	Neg		Cur	1	0	1	0	0	0	0	1	0	0
68298	F	76	T	Control	Neg		Cur	0	0	0	0	0	0	0	1	0	0
68347	M	73	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
68351	M	71	T	Control	Neg		Cur	0	0	0	0	1	0	0	0	0	0
68375	M	64	T	Control	Neg		Cur	0	0	1	0	1	0	0	1	0	0
68396	M	73	T	Control	Neg	Anal carcinoma	Cur	0	0	1	0	0	0	0	0	0	0
68407	M	58	T	Control	Neg		Cur	0	0	1	0	1	0	0	0	0	0
73484	F	62	T	Control	Neg	Brochiectasis	Cur	0	0	0	1	0	0	0	0	1	0
73533	M	59	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
78657	F	54	T	Control	Neg		Cur	0	0	0	0	0	0	0	0	0	0
78662	F	54	T	Control	Neg		Cur	0	0	0	0	0	0	1	0	0	0
144	F	74	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
4935	M	76	T	Control	Neg		For	0	0	0	0	0	0	0	0	1	0
34920	M	55	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	1
38010	M	75	T	Control	Neg		For	0	0	1	0	0	0	0	1	0	0
40960	M	65	T	Control	Neg		For	0	0	0	1	1	0	1	1	0	0

53917	F	58	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
58694	M	77	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
63154	M	65	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
65219	F	51	T	Control	Neg		For	1	1	1	0	1	0	0	0	0	0
67910	F	70	T	Control	Neg		For	0	0	1	0	0	0	0	1	0	0
68229	F	54	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73441	M	59	T	Control	Neg		For	0	0	0	0	1	0	0	0	1	0
73458	M	64	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
73547	M	68	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
3443	M	58	T	Control	Neg	Sarcoid	For	0	0	0	0	0	0	0	0	0	0
3621	F	64	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
4508	M	66	T	Control	Neg		For	0	1	1	0	0	0	0	0	0	0
7400	M	76	T	Control	Neg		For	1	0	0	0	0	0	0	0	0	0
19556	F	56	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
19580	F	68	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
19589	M	76	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
24095	M	56	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
25055	F	60	T	Control	Neg		For	0	0	1	0	0	0	0	0	1	0
25058	M	74	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
25085	M	82	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
25101	M	69	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
28665	M	64	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
34916	M	64	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
37991	F	73	T	Control	Neg		For	0	0	0	0	0	0	0	0	1	0
38011	M	75	T	Control	Neg		For	0	0	0	0	0	1	0	1	0	1
38017	F	74	T	Control	Neg	Pulmary Nodules	For	0	0	0	0	0	0	0	0	0	0
38018	F	58	T	Control	Neg		For	0	0	0	0	0	0	0	0	1	0
40967	M	63	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
43068	M	81	T	Control	Neg	fibrosis	For	1	0	0	0	0	0	0	1	0	0
43976	M	74	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0

44130	F	82	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
47950	F	61	T	Control	Neg		For	0	0	0	0	0	0	1	0	0	0
53870	F	69	T	Control	Neg	Breast Cancer	For	0	0	1	0	0	0	0	0	0	1
53877	F	58	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
53879	F	62	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
53897	M	52	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
53917	F	58	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
53937	M	63	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
53938	M	68	T	Control	Neg	COPD	For	0	0	0	0	0	0	1	0	1	1
53940	F	72	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
54383	F	71	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
58712	F	76	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	1
58715	F	77	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
59705	F	75	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
60777	F	68	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
60780	F	67	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
60781	F	70	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
63145	M	62	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
63148	M	56	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
63202	F	49	T	Control	Neg		For	0	0	0	0	1	0	0	0	1	0
65662	F	67	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
65664	F	75	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
65666	M	70	T	Control	Neg		For	0	0	1	0	1	0	0	0	0	0
66134	M	59	T	Control	Neg		For	0	1	1	0	0	0	0	0	0	0
67895	M	76	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
67915	F	55	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
67927	M	62	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68182	M	72	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
68188	M	66	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
68204	M	57	T	Control	Neg		For	0	0	0	0	1	0	0	0	0	0

68258	M	77	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68264	F	71	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
68269	M	77	T	Control	Neg		For	0	0	1	0	1	0	0	1	0	0
68270	M	76	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68281	M	79	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
68289	F	77	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68307	F	57	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68313	M	66	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68330	F	64	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68331	M	68	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68370	M	78	T	Control	Neg		For	0	0	1	0	0	0	0	0	0	0
68378	M	74	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68425	F	69	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68427	F	58	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
68432	M	58	T	Control	Neg		For	0	0	0	0	0	0	0	1	0	0
73451	F	59	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73463	M	69	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73469	M	72	T	Control	Neg		For	0	1	0	0	0	0	0	0	0	0
73483	M	77	T	Control	Neg		For	0	1	0	0	0	0	0	0	0	0
73502	F	54	T	Control	Neg		For	1	0	0	0	0	0	0	0	0	0
73503	M	60	T	Control	Neg		For	1	0	1	0	0	0	0	0	0	0
73512	M	69	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73516	M	77	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73527	M	61	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
73551	M	81	T	Control	Neg	Mesothelioma	For	0	0	1	1	1	0	1	1	0	0
78645	M	63	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	0
78658	M	67	T	Control	Neg		For	0	0	0	0	0	0	0	0	0	1
53934	M	69	T	Control	Neg		Non	0	1	1	0	0	0	0	1	0	0
68321	F	64	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
19574	M	65	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0

24099	F	71	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
25038	F	57	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
25072	M	79	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
25105	F	58	T	Control	Neg		Non	0	0	0	0	1	0	0	0	0	0
34915	F	71	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
42416	F	60	T	Control	Neg		Non	0	0	0	0	0	0	0	1	0	0
42418	M	58	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
42420	M	72	T	Control	Neg		Non	0	0	0	0	0	0	1	0	0	0
43070	M	56	T	Control	Neg		Non	1	0	0	0	0	0	0	0	1	0
45273	F	69	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
47960	M	69	T	Control	Neg		Non	0	0	0	0	1	0	0	0	0	0
53840	M	76	T	Control	Neg		Non	0	0	0	0	0	0	0	0	1	0
53845	F	62	T	Control	Neg	Breast Cancer	Non	0	0	0	0	0	0	0	0	0	0
53846	F	75	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
53868	M	70	T	Control	Neg	bronchiectasis	Non	0	1	0	1	1	0	0	1	0	0
53890	M	58	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
53919	M	67	T	Control	Neg		Non	0	0	0	0	1	0	0	0	0	0
58706	M	51	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
58713	M	60	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
59728	F	71	T	Control	Neg		Non	1	0	0	0	0	0	0	0	0	0
60776	M	65	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
63126	F	59	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
63167	F	65	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
65230	M	69	T	Control	Neg		Non	0	0	0	0	0	0	0	1	0	0
68195	F	70	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68238	M	51	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68245	M	62	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
68248	M	56	T	Control	Neg		Non	0	0	1	0	0	0	0	0	0	0
68256	F	49	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68259	F	57	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0

68276	M	70	T	Control	Neg		Non	0	1	0	0	1	0	0	0	0	0
68335	F	64	T	Control	Neg		Non	0	0	0	0	0	0	0	1	0	0
68400	F	58	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68419	M	53	T	Control	Neg		Non	0	0	0	0	0	0	0	1	0	0
68426	M	63	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68428	M	57	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
68429	F	68	T	Control	Neg	consolid/infecti	Non	1	0	0	0	0	0	0	0	0	1
73475	F	56	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
78653	F	61	T	Control	Neg		Non	0	0	0	0	0	0	0	0	0	0
30691	F	69	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
40985	F	67	T	Control	Neg			0	0	0	0	0	0	0	1	0	0
73552	M	68	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
19561	M	63	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
19576	F	77	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
19897	F	57	T	Control	Neg			0	0	0	0	0	0	0	0	1	0
19910	M	70	T	Control	Neg	Col Ca		0	0	0	0	0	0	0	0	0	0
21879	M	86	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
25065	F	81	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
25084	M	59	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
25095	F	82	T	Control	Neg			0	0	0	0	0	0	0	0	1	0
25100	M	69	T	Control	Neg			0	0	0	0	0	0	0	1	0	0
26207	M	58	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
28681	M	70	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
30688	F	66	T	Control	Neg			0	0	0	0	0	0	0	1	1	0
30716	M	80	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
38002	M	68	T	Control	Neg			1	0	0	1	0	0	0	0	0	0
40948	M	70	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
40977	F	80	T	Control	Neg			0	0	1	0	0	0	0	1	0	1
41195	F	84	T	Control	Neg			0	0	0	0	0	0	0	0	0	0
43961	M	41	T	Control	Neg			0	0	0	0	0	0	0	0	1	0

43974	F	62	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
44137	F	84	T	Control	Neg						0	1	1	0	0	0	0	0	0	0	
45280	M	72	T	Control	Neg						0	0	0	0	0	0	0	1	0	0	
45283	F	69	T	Control	Neg						0	0	1	0	0	0	0	0	0	0	
47957	F	56	T	Control	Neg						0	0	1	0	0	0	0	1	0	0	
53872	F	78	T	Control	Neg						0	0	1	0	0	0	0	0	0	0	
59719	M	76	T	Control	Neg	Pleural plaques					1	0	1	0	1	0	0	0	0	1	
63118	M	46	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
63158	F	66	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
63195	F	76	T	Control	Neg	COPD					1	0	0	0	0	0	0	1	0	0	
65231	F	56	T	Control	Neg	Col Ca					0	0	0	0	0	0	0	1	0	0	
65244	F	55	T	Control	Neg						0	0	0	0	0	0	0	1	0	0	
68216	F	74	T	Control	Neg						0	0	1	0	0	0	0	0	0	0	
68268	M	77	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
68308	F	73	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
68309	M	64	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
73478	F	64	T	Control	Neg						0	0	0	0	0	0	1	0	0	0	
73546	M	55	T	Control	Neg						0	0	0	0	0	0	0	0	0	0	
73550	M	71	T	Control	Neg	Pleural Plaques					0	0	0	0	0	0	0	0	0	0	
73580	M	78	V	Case	Pos	SqCCL	Pri			Cur	0	0				0	1	0		0	
30707	F	80	V	Case	Pos	AdenoCa	Pri	4	2	Cur	0	1				0	0	1		0	
30687	M	79	V	Case	Neg	SCLC	Pri			Cur	0	1				0	0	0		1	
40972	F	69	V	Case	Neg	Lung Ca	Pri			Cur	1	0				0	0	1		1	
26208	F	70	V	Case	Neg	Lung Ca	Met			For	0	0				0	0	0		0	
40975	F	69	V	Case	Neg	Lung Ca	Sec			For	0	0				0	0	0		0	
53900	M	80	V	Case	Neg	Lung Ca	Met			For	0	1				0	0	0		0	
28662	M	78	V	Case	Pos	SqCCL	Pri			For	0	1				0	0	0		0	
34918	F	66	V	Case	Neg	AdenoCa	Pri	1	0	For	0	0				0	0	0		0	
19544	M	63	V	Case	Neg	Lung Ca	Pri	4	2	For	0	0				0	1	0		0	
53878	M	72	V	Case	Neg	AdenoCa	Pri	1	2	For	0	0				0	0	0		1	

53881	M	67	V	Case	Sus	Lung Ca	Pri			For	0	0	1	0	0	0
25059	M	68	V	Case	Pos	Lung Ca	Pri		3	For	0	0	0	0	0	1
8709	M	72	V	Case	Neg	AdenoCa	Pri	2	0	For	0	1	0	0	0	0
44136	M	57	V	Case	Pos	AdenoCa	Pri			For	0	0	0	0	0	0
43960	F	68	V	Case	Sus	AdenoCa	Pri			For	0	0	0	0	1	0
53831	M	66	V	Case	Neg	Lung Ca	Met			For	0	0	0	0	0	0
25104	M	78	V	Case	Neg	Lung Ca	Pri	1	0	For	0	1	0	0	0	0
47945	M	74	V	Case	Neg	Lung Ca	Met			For	0	0	0	0	0	0
44131	M	72	V	Case	Neg	Lung Ca	Pri	3	2	For	0	0	0	0	0	0
66140	F	74	V	Case	Neg	Lung Ca	Met			For	0	0	0	0	0	0
26214	M	68	V	Case	Sus	NSCLC	Sec			For	0	0	0	0	0	0
25078	F	68	V	Case	Neg	Lung Ca	Met			For	0	0	0	0	0	1
60461	M	61	V	Case	Neg	Lung Ca	Pri			For	0	0	1	0	0	0
65241	M	59	V	Case	Pos	SqCCL	Pri			For	0	0	0	0	1	0
53861	F	64	V	Case	Neg	Lung Ca	Pri	1	0	For	0	1	0	0	0	1
40982	M	57	V	Case	Neg	SqCCL	Pri	4	0	For	0	1	0	1	1	1
19519	M	74	V	Case	Neg	Lung Ca	Pri			For	0	0	0	0	1	1
37996	F	77	V	Case	Neg	SCLC	Pri			For	0	1	0	1	1	1
4537	M	62	V	Case	Pos	SCLC	Pri			For	0	0	0	1	1	0
25045	F	63	V	Case	Pos	SCLC	Pri			For	0	1	0	1	1	1
25054	M	68	V	Case	Pos	AdenoCa	Pri			For	0	1	0	0	0	1
36012	M	70	V	Case	Pos	SqCCL	Pri			For	0	1	0	0	0	1
47959	F	73	V	Case	Sus	SCLC	Pri			For	0	1	0	1	1	0
30693	F	68	V	Case	Neg	Lung Ca	Pri			For	0	0	0	1	1	1
43073	M	81	V	Case	Neg	AdenoCa	Pri			For	0	1	0	1	1	1
43975	F	50	V	Case	Neg	SCLC	Pri			For	0	0	0	1	1	1
37994	F	78	V	Case	Pos	AdenoCa	Pri	4	2	For	0	1	0	1	1	1
41193	M	76	V	Case	Neg	AdenoCa	Pri			For	0	1	0	1	0	0
24188	M	70	V	Case	Neg	Lung Ca	Pri			For	0	1	0	0	0	1
78863	F	72	V	Case	Pos	Lung Ca	Pri			For	0	0	0	1	1	0

19590	F	80	V	Case	Neg	Lung Ca	Pri		1	For	0	1		0	0	0	1
20117	F	60	V	Case	Neg	Lung Ca	Met			For	0	0		0	1	1	0
25068	F	71	V	Case	Neg	SCLC	Pri			For	0	1		0	1	1	1
53859	F	69	V	Case	Neg	SCLC	Pri			For	0	1		0	1	1	1
19551	M	73	V	Case	Pos	SqCCL	Pri	2	1	For	1	1		1	1	1	1
19872	M	84	V	Case	Pos	SqCCL	Pri			For	1	0		0	1	1	1
24101	M	68	V	Case	Pos	SCLC	Pri			For	0	1		0	1	1	0
26209	M	74	V	Case	Pos	AdenoCa	Pri			For	1	1		0	0	1	1
26224	M	60	V	Case	Pos	Lung Ca	Pri			For	1	1		0	1	1	0
28656	F	47	V	Case	Pos	AdenoCa	Pri	4	2	For	0	0		1	1	1	1
28680	M	77	V	Case	Pos	SqCCL	Pri			For	1	1		1	1	1	1
42404	M	69	V	Case	Pos	AdenoCa	Pri	2	1	For	1	1		0	0	0	1
43065	M	78	V	Case	Pos	SqCCL	Pri	4		For	0	0		1	0	0	1
19520	M	76	V	Case	Neg	Lung Ca	Pri			For	0	1		0	1	1	0
40949	F	76	V	Case	Sus	SCLC	Pri	2	0	For	0	0		0	1	1	1
30690	F	77	V	Case	Neg	AdenoCa	Pri			For	0	1		0	1	1	1
40946	M	76	V	Case	Sus	AdenoCa	Pri			For	1	1		0	0	0	1
19569	F	70	V	Case	Pos	SCLC	Pri			For	0	0		0	1	1	1
79017	F	75	V	Case	Neg	Lung Ca	Pri			For	0	1		0	0	1	0
26200	M	77	V	Case	Neg	SqCCL	Pri			For	0	1		0	0	0	1
73586	M	60	V	Case	Sus	Lung Ca	Pri			For	0	1		0	1	0	0
30717	M	53	V	Case	Neg	Carcinoid	Pri	2	0	For	0	0		0	1	0	1
21666	M	70	V	Case	Neg	Lung Ca	Pri			For	0	0		1	1	0	0
19570	M	64	V	Case	Neg	Cystic mucinous	Pri			For	0	1		0	0	0	1
1667	M	58	V	Case	Neg	Carcinoid	Met	2	1	For	0	1		0	1	0	1
19528	M	70	V	Case	Neg	Lung Ca	Pri			For	0	1		0	0	0	1
43957	F	63	V	Case	Neg	Lung Ca	Pri			Non	0	1		0	0	0	0
30689	M	52	V	Case	Neg	SqCCL	Pri			Non	0	0		0	0	0	0
26213	M	68	V	Case	Pos	AdenoCa	Pri			Non	0	0		0	0	0	0

59712	M	78	V	Case	Neg	Lung Ca	Met			Non	0	0		0	0	0	0
16423	F	82	V	Case	Neg	SCLC	Pri			Non	0	0		0	0	1	0
19588	F	76	V	Case	Neg	Lung Ca	Pri	1	0	Non	0	0		0	0	1	0
19895	M	77	V	Case	Neg	Lung Ca	Pri	1	1	Non	0	0		0	0	0	0
24097	F	72	V	Case	Neg	Lung Ca	Pri			Non	0	0		0	0	1	0
30701	F	83	V	Case	Pos	SqCCL	Pri			Non	0	0		0	0	1	0
42407	M	78	V	Case	Pos	SqCCL	Pri			Non	0	0		0	0	0	0
19628	M	53	V	Case	Neg	SCLC	Pri			Non	0	0		0	0	1	0
30715	F	62	V	Case	Neg	SqCCL	Pri			Non	0	0		0	0	0	0
19861	F	58	V	Case	Neg	Carcinoid	Pri	1	1	Non	0	0		0	0	0	1
43086	M	68	V	Case	Neg	LCLC	Pri	1	0	Non	0	0		0	0	0	0
40973	M	62	V	Case	Neg	Lung Ca	Pri			Non	0	0		0	0	0	0
53886	M	66	V	Case	Sus	SqCCL	Pri	1		Non	0	1		0	0	0	0
43977	F	62	V	Case	Neg	AdenoCa	Pri	1	0	Non	0	0		0	0	0	0
30692	F	67	V	Case	Neg	Lung Ca	Pri	3	0	Non	0	1		0	0	0	0
44134	F	54	V	Case	Neg	AdenoCa	Sec			Non	0	1		0	0	0	1
19623	M	73	V	Case	Neg	SCLC	Pri			Non	0	1		0	0	0	1
29142	F	72	V	Case	Sus	SCLC	Pri			Non	1	1		0	0	0	1
5458	M	75	V	Case	Neg	SCLC	Pri			Non	1	0		0	1	1	1
25053	M	65	V	Case	Neg	SqCCL	Pri			Non	0	0		0	1	1	0
25075	M	76	V	Case	Neg	SqCCL	Pri			Non	0	0		0	1	0	1
13534	M	74	V	Case	Pos	SqCCL	Pri			Non	0	0		0	0	1	1
21642	F	61	V	Case	Pos	Lung Ca	Pri			Non	0	0		0	1	1	1
25066	F	51	V	Case	Pos	AdenoCa	Pri			Non	0	1		0	1	1	1
26222	M	62	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	0
28660	F	66	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	1
28661	F	85	V	Case	Sus	SqCCL	Pri	2		Non	0	1		1	0	1	1
34895	M	82	V	Case	Pos	SqCCL	Pri			Non	0	0		0	1	1	1
34909	M	62	V	Case	Pos	SqCCL	Pri			Non	0	1		0	1	1	1
40971	F	68	V	Case	Sus	SCLC	Pri			Non	0	1		0	1	1	0

41197	M	62	V	Case	Sus	SCLC	Pri			Non	0	0		0	1	1	1
53841	F	66	V	Case	Pos	SqCCL	Pri			Non	0	1		1	0	1	1
25067	M	60	V	Case	Pos	Lung Ca	Pri	2	2	Non	1	0		1	1	1	1
40984	F	66	V	Case	Sus	SqCCL	Pri			Non	0	1		1	0	1	1
47961	M	51	V	Case	Neg	SqCCL	Pri			Non	1	1		0	0	1	1
19584	M	74	V	Case	Pos	SCLC	Pri			Non	0	0		0	1	1	1
19896	M	71	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	0
53854	F	65	V	Case	Sus	SCLC	Pri			Non	0	0		0	1	1	1
78627	M	61	V	Case	Pos	AdenoSq	Pri			Non	0	1		0	0	1	1
5785	F	73	V	Case	Neg	Lung Ca	Pri			Non	0	0		0	1	1	0
25073	M	63	V	Case	Neg	SCLC	Pri			Non	0	0		1	1	1	1
28663	F	80	V	Case	Neg	SCLC	Pri			Non	1	1		0	1	1	1
28675	M	70	V	Case	Neg	SCLC	Pri			Non	0	1		0	1	0	0
36014	M	81	V	Case	Neg	AdenoCa	Pri			Non	0	0		1	1	0	0
40954	F	67	V	Case	Neg	AdenoCa	Pri			Non	1	0		0	1	1	0
43079	M	66	V	Case	Neg	AdenoCa	Pri			Non	0	1		0	1	1	1
47952	F	63	V	Case	Neg	Lung Ca	Pri			Non	0	1		0	1	1	1
53882	M	67	V	Case	Neg	SCLC	Pri			Non	0	1		0	1	1	0
19517	M	61	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	0
19554	M	63	V	Case	Pos	SCLC	Pri			Non	0	0		0	1	1	1
19560	F	62	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	0
25046	F	63	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	0	0
25060	M	59	V	Case	Pos	SqCCL	Pri			Non	0	0		0	1	0	1
26221	M	55	V	Case	Sus	SqCCL	Pri			Non	1	0		1	1	1	1
26223	M	60	V	Case	Pos	SCLC	Pri			Non	0	1		0	1	1	1
34896	F	71	V	Case	Pos	AdenoSq	Pri			Non	0	0		1	1	1	1
34911	M	56	V	Case	Pos	AdenoCa	Pri			Non	0	0		1	1	0	1
40957	F	78	V	Case	Pos	SCLC	Pri			Non	0	0		1	1	0	1
40976	M	67	V	Case	Sus	AdenoCa	Pri	4		Non	0	1		0	0	0	1
43962	F	67	V	Case	Sus	Lung Ca	Pri			Non	0	0		0	1	1	0

19914	M	65	V	Case	Neg	Lung Ca	Pri			Non	0	1		1	0	0	0
25061	F	66	V	Case	Neg	SCLC	Pri			Non	0	1		0	1	1	1
26216	F	69	V	Case	Neg	Lung Ca	Pri	2	3	Non	0	0		0	1	1	1
53851	F	66	V	Case	Neg	SCLC	Pri			Non	0	0		0	1	1	1
30698	M	56	V	Case	Neg	SCLC	Pri	1	0	Non	0	0		0	1	1	0
67909	F	76	V	Case	Neg	Lung Ca	Pri			Non	0	0		0	1	1	0
19629	F	76	V	Case	Neg	LCLC	Pri	1	1	Non	0	1		0	1	0	0
44135	F	75	V	Case	Neg	Lung Ca	Pri	2	0	Non	0	1		0	0	1	0
42421	M	59	V	Case	Neg	SqCCL	Pri			Non	0	1		0	0	0	1
36416	M	67	V	Control	Neg					Cur	0	0		0	0	0	0
53862	M	79	V	Control	Neg					Cur	0	0		0	0	0	0
53867	F	66	V	Control	Neg					Cur	0	0		0	0	0	0
63140	F	75	V	Control	Neg					Cur	0	0		0	0	0	0
1391	F	48	V	Control	Neg					Cur	0	0		0	0	0	0
19868	M	68	V	Control	Neg	Asbestosis				Cur	0	1		0	0	0	1
21880	F	58	V	Control	Neg					Cur	0	0		0	0	0	0
22103	F	64	V	Control	Neg					Cur	0	0		0	0	0	0
23759	F	67	V	Control	Neg					Cur	0	0		0	0	0	0
25047	F	87	V	Control	Neg					Cur	0	1		0	0	0	0
25071	M	68	V	Control	Neg					Cur	0	1		0	0	0	0
26210	F	76	V	Control	Neg					Cur	0	0		0	0	0	0
31808	M	60	V	Control	Neg					Cur	0	1		0	0	0	0
53832	M	71	V	Control	Neg					Cur	0	1		0	0	0	0
53833	M	61	V	Control	Neg					Cur	0	1		0	0	0	0
53855	F	46	V	Control	Neg					Cur	0	0		0	0	0	0
58688	F	52	V	Control	Neg					Cur	0	0		0	0	0	0
65217	F	76	V	Control	Neg					Cur	0	0		0	0	0	0
66141	F	75	V	Control	Neg					Cur	0	0		0	0	0	0
68215	F	64	V	Control	Neg					Cur	0	0		0	0	0	0
68231	M	54	V	Control	Neg					Cur	0	0		0	0	0	0

73543	M	62	V	Control	Neg		Cur	0	0		0	0	0	0
78972	M	60	V	Control	Neg		Cur	0	0		0	0	0	0
79023	F	80	V	Control	Neg		Cur	0	0		0	0	0	0
25081	F	71	V	Control	Sus		Cur	0	1		0	0	0	0
1199	M	54	V	Control	Neg		For	0	0		0	0	0	0
18092	M	58	V	Control	Neg		For	0	0		0	0	0	0
65211	M	67	V	Control	Neg		For	0	0		0	0	0	0
73570	M	73	V	Control	Neg	bowel Ca	For	0	0		0	0	0	0
2900	M	78	V	Control	Neg		For	0	0		0	0	0	0
5357	M	64	V	Control	Neg		For	0	0		0	0	0	0
6869	F	75	V	Control	Neg		For	0	0		0	0	1	0
8388	F	63	V	Control	Neg		For	0	0		0	0	0	0
8738	M	66	V	Control	Neg		For	0	0		0	0	0	0
8960	M	70	V	Control	Neg		For	0	0		0	0	0	0
19527	M	55	V	Control	Neg		For	0	0		0	0	0	0
19541	F	59	V	Control	Neg		For	0	0		0	0	0	0
19543	M	78	V	Control	Neg		For	0	1		0	0	0	0
19627	M	69	V	Control	Neg		For	0	0		0	0	0	0
19894	F	71	V	Control	Neg		For	0	0		0	0	0	0
19911	M	49	V	Control	Neg		For	0	1		0	0	0	0
19915	F	60	V	Control	Neg		For	0	0		0	0	0	0
22123	F	64	V	Control	Neg		For	0	0		0	0	0	0
23489	F	64	V	Control	Neg		For	0	0		0	0	0	0
23757	F	62	V	Control	Neg		For	0	0		0	0	0	0
24100	M	70	V	Control	Neg		For	0	0		0	0	0	0
25074	M	73	V	Control	Neg		For	0	0		0	0	1	0
25094	M	77	V	Control	Neg	Mesothelioma	For	0	0		0	0	0	0
26205	M	71	V	Control	Neg		For	0	0		0	0	0	0
30039	M	60	V	Control	Neg		For	0	0		0	0	0	0
30694	M	78	V	Control	Neg		For	0	0		0	0	0	0

38004	F	82	V	Control	Neg		For	0	0		0	0	0	0
38016	M	75	V	Control	Neg		For	0	0		0	0	0	0
38021	M	71	V	Control	Neg		For	0	0		0	0	0	0
40940	M	78	V	Control	Neg		For	0	0		0	0	0	0
40974	F	65	V	Control	Neg		For	0	0		0	0	0	0
42403	M	64	V	Control	Neg		For	0	0		0	0	0	0
43066	F	82	V	Control	Neg		For	0	0		0	0	0	0
43081	M	58	V	Control	Neg		For	0	0		0	0	0	0
43973	F	73	V	Control	Neg		For	0	0		0	0	0	0
44128	F	62	V	Control	Neg		For	0	0		0	0	0	1
53876	M	69	V	Control	Neg		For	0	0		0	0	0	0
53885	M	75	V	Control	Neg		For	0	0		0	0	0	0
53923	M	79	V	Control	Neg		For	0	0		0	0	0	0
58704	F	63	V	Control	Neg		For	0	0		0	0	0	0
58705	M	66	V	Control	Neg		For	0	1		0	0	0	0
59716	F	80	V	Control	Neg		For	0	0		0	0	0	0
59723	F	71	V	Control	Neg		For	0	1		0	0	0	0
59734	M	58	V	Control	Neg		For	0	0		0	0	0	0
60772	M	60	V	Control	Neg		For	0	0		0	0	0	0
60775	F	63	V	Control	Neg		For	0	0		0	0	0	0
63131	M	74	V	Control	Neg		For	0	0		0	0	0	0
65238	F	60	V	Control	Neg		For	0	0		0	0	0	1
68202	F	69	V	Control	Neg		For	0	0		0	0	0	0
68234	F	70	V	Control	Neg		For	0	0		0	0	0	0
68255	F	78	V	Control	Neg	Carcinomatosis	For	0	0		0	0	0	0
68295	M	79	V	Control	Neg	Heart failure	For	0	1		0	0	0	1
68353	M	55	V	Control	Neg		For	0	0		0	0	0	0
68381	M	79	V	Control	Neg		For	0	0		0	0	0	0
68385	M	66	V	Control	Neg		For	0	0		0	0	0	0
68405	M	72	V	Control	Neg	Mesothelioma	For	0	0		0	0	0	0

73500	M	69	V	Control	Neg		For	0	0	0	0	0	0	
73528	F	69	V	Control	Neg		For	0	0	0	0	0	0	
73529	M	78	V	Control	Neg		For	0	1	0	0	0	1	
73531	M	70	V	Control	Neg		For	0	1	0	0	0	0	
78856	M	78	V	Control	Neg		For	0	0	0	0	0	0	
78860	M	73	V	Control	Neg		For	0	1	0	0	0	0	
78908	M	73	V	Control	Neg		For	0	0	0	0	0	0	
78941	M	82	V	Control	Neg		For	0	0	0	0	0	0	
78971	F	73	V	Control	Neg		For	0	0	0	0	0	0	
25051	F	55	V	Control	Neg		Non	0	0	0	0	0	0	
45271	M	65	V	Control	Neg		Non	0	0	0	0	0	0	
53852	M	51	V	Control	Neg		Non	0	0	0	0	0	0	
4765	M	58	V	Control	Neg		Non	0	0	0	0	0	0	
19625	F	50	V	Control	Neg		Non	0	1	0	0	0	0	
23418	M	60	V	Control	Neg		Non	0	1	0	0	0	0	
23430	M	50	V	Control	Neg		Non	0	0	0	0	0	0	
26215	M	63	V	Control	Neg	Mesothelioma	Non	0	0	0	0	0	0	
28653	F	59	V	Control	Neg		Non	0	0	0	0	0	0	
31480	M	65	V	Control	Neg		Non	0	0	0	0	0	0	
37998	M	71	V	Control	Neg		Non	0	0	0	0	0	0	
53884	M	78	V	Control	Neg		Non	0	0	0	0	0	1	
53927	M	78	V	Control	Neg		Non	0	0	0	0	0	1	
59743	F	69	V	Control	Neg		Non	0	0	0	0	0	0	
65654	F	66	V	Control	Neg		Non	0	0	0	0	0	0	
68201	F	72	V	Control	Neg		Non	0	0	0	0	0	0	
68435	F	63	V	Control	Neg		Non	0	0	0	0	0	0	
78683	M	69	V	Control	Neg	COPD	Non	0	1	0	0	0	1	
21003	F	78	V	Control	Neg			0	0	0	0	0	0	

Sex		Set		Cytology		Type		Smoking	
M	Male	T	Training	Neg	Negative	Pri	Primary	Cur	Current
F	Female	V	Validation	Sus	Suspisious	Sec	Secondary	For	Former
				Pos	Positive	Met	Metastatic	Non	Non-smoker

Appendix 2. List of Reagents

List of reagents

CHEMICAL REAGENTS

Name	Company	Cat. number
3130 POP-7™ polymer	Applied Biosystems (ABI)	p/n4352759
3730 10x Buffer with EDTA	Applied Biosystems	p/n4335613
5-Aza-2'-deoxycytidine	Sigma Aldrich	A3656
Acetic Acid	Fisher Chemicals	BP2401
Agarose	Fisher Bioreagents	BPE-1356-50
Pyrosequencing Annealing Buffer	Qiagen	40-0036
Pyrosequencing Binding Buffer	Qiagen	40-0033
Boric Acid (H_2BO_3)	Fluka	15663
Chloroform	BDH	100776B
EDTA	BDH	100935V
Ethanol	Department of Chemistry, UoL	UN1170
Glycerol	Sigma Aldrich	G6279
Hi-Di™ Formamide	Applied Biosystems	4311320
Hydrogen chloride (HCl)	BDH	101254H
Hydrogen peroxide (H_2O_2)	Sigma Aldrich	H1009
Isopropanol	Fisher Scientific	P17490/17
Low-melting point agarose	Sigma Aldrich	A9414
Phenol : chlofoform : isoamyl alcohol (50:49:1)	Fisher Bioreagents	BPE-1752p-400
S-adenosylmethionine (SAM)	NEB	B9003S
SafeView solution	NBS Biologicals	NBS-SV1
Sodium acetate anhydrous	BDH	102365R
Sodium chloride (NaCl)	Sigma Aldrich	S9625
Sodium hydroxide (NaOH)	Fluka	7169
Streptavidin Sepharose	GE Healthcare	17-5113-01
Trizma® Acetate	Fluka	93337
Trizma® base	Sigma Aldrich	T1503

ENZYMES

Name	Company	Cat. number
M.SssI CpG methyltransferase	NEB	M0226S
Proteinase K	Qiagen	19133
RNase A	Qiagen	19101

PAGE NUMBERING AS ORIGINAL

KITS AND ASSAYS

Name	Company	Cat. number
DNeasy® Blood and Tissue Kit	Qiagen	69506
EZ DNA Methylation™ Kit	Zymo Research	D5002
HotStarTaq Plus DNA Polymerase Master Mix Kit	Qiagen	203645
Oligo aCGH/ChIP-on-chip Hybridization Kit	Agilent	5188-5220
ProtoBlock Solution	National Diagnostics	CL-252
PyroGold™ SQA Reagents	Biotage	40-0045
QIAquick Gel Extraction Kit	Qiagen	28704
TaqMan® Gene Expression Master Mix	Applied Biosystems	4369542

MOLECULAR MASS LADDERS

Name	Company	Cat. number
Full Range Amersham™ Rainbow™ Marker	GE Healthcare	RPN800E
HyperLadder 1, 100 lanes	Bioline	BIO-33053

PRIMERS

Name	Company	Cat. number
MGB probes	Applied Biosystems	N/A
Oligonucleotides	MWG	N/A

Appendix 3. Manuscripts submitted.

Supporting manuscript accepted in Clinical Chemistry and Laboratory Medicine (CCLM)

DNA methylation biomarkers in biological fluids for early detection of respiratory tract cancer.

Soultana Markopoulou[1], Georgios Nikolaidis[2]and Triantafillos Liloglou[2].

[1]Aristotle University of Thessaloniki, Medical School, Department of Microbiology, Thessaloniki, Greece

[2]University of Liverpool, Institute of Translational Medicine, Department of Molecular and Clinical Cancer Medicine, Liverpool, United Kingdom.

Correspondence to

Dr. T. Liloglou
University of Liverpool
Cancer Research Centre
200 London Road,
Liverpool L3 9TA, UK
Tel 0044151 7948958
Fax 0044151 7948989
Email: tliloglo@liv.ac.uk.

Abstract

Cancers of the respiratory tract (lung and head and neck) share common aetiologies, risk factors and molecular characteristics. Epigenetic reprogramming is one of the hallmarks of cancer and DNA methylation is currently the best-studied form. There are a number of characteristics of DNA methylation, which seem advantageous in biomarker development. Early detection is still an unmet clinical care need, which guarantees to significantly reduce the mortality of patients with respiratory cancers. The application of such biomarkers in biological fluids being sampled in everyday clinical practice is a long term demand.

In this review we summarise the current literature on DNA methylation detection in bronchial washings, sputum, saliva, plasma and serum and discuss the potential of their clinical implementation. We also discuss important aspects of biomarker development and validation pointing to the appropriate route for a biomarker to reach clinical standards.

Introduction

Cancers of the respiratory tract (oral cavity, pharynx, larynx, trachea, lung,) share a significant number of molecular and epidemiological characteristics. Together they account for almost two million cases per year (1). Although the incidence and mortality figures differ among histological types and locations, respiratory tract cancers overall constitute a major public health threat. Their management suffers from the lack of modern molecular tools, which could assist in improving clinical outcomes by diagnosing malignant lesions earlier and/or stratifying them in appropriate therapeutic regimes with higher efficiency.

Lung cancer is by far the most frequent cause of cancer-related deaths in the world accounting for 1.38 million deaths per year worldwide. The International Agency for Research on Cancer (IARC) reported 1.61 million new cases in 2008 accounting for the 12.7% of the total cancer incidence worldwide (1). Despite the small increase in the trend for survival rates observed in the past few years the survival percentage remains dismal. This fact is mainly attributed to the advanced stage in which diagnosis is frequently made; only 15% of lung cancer patients are diagnosed in the early localised stage and hence have the most chances for successful treatment and long-term survival(2, 3).

The term head and neck cancer encompasses cancers of the upper aerodigestive tract, the paranasal sinuses, and the salivary glands. Over 90% of these cancers are histologically categorized as squamous cell carcinomas stemming mainly from the oral cavity, nasopharynx, oropharynx, hypopharynx, and the larynx. The worldwide incidence of head and neck cancers was over half a million in 2008 whereas mortality was estimated at 320,000. The survival rates have exhibited a slight improvement over the past thirty years but remain quite low.

DNA methylation and cancer.

When the word of Greek origin "epigenetics" was introduced to the scientific society at the start of the 1940s nobody could predict that it would give a name to a new field of great importance and consequence to biology and medicine (4). Epigenetic mechanisms include DNA methylation, covalent histone modifications, nucleosome remodelling and miRNAs (5, 6). Although distinct, all types of epigenetic regulation are at a constant interplay so to effectively regulate gene expression. To date, it is well understood that key normal processes including development, cell cycle and differentiation are dependent on a tightly regulated epigenetic programme. The impact of epigenetic regulation on the homeostasis of a multicellular organism can be realized by the various disease states that arise from the perturbation of its controlling mechanisms. The list of human diseases found to implicate some form of epigenetic disorder is constantly growing.(7)

DNA methylation is the most extensively explored epigenetic modification in humans. It is a chemical modification entailing the attachment of a methyl group at the 5' position of a cytosine ring within CpG dinucleotides. Three main DNA methytransferases (DNMTs) have been found to be involved in the generation and heritable preservation of DNA methylation patterns. DNMT1 is the maintenance methyltransferase which acts upon hemi-methylated DNA following replication whereas DNMT3a and DNMT3b are termed *de novo* methyltransferases acting independently of replication on both hemi-methylated and unmethylated CpGs (8, 9). Most prominent sites of CpGs in the human genome are regions of large repetitive sequences and CpG islands. The latter are defined as DNA stretches of over 200 bases with an exceptionally high G+C content (>50%) and are present in almost 60% of all human gene promoters (7).In human adult cells promoter associated CpG islands are normally unmethylated, with the exception of imprinted genes and X chromosome inactivated genes in females (10). Furthermore, a high percentage of methylated CpGs has been observed in repetitive elements contributing

tochromosomal integrity and genomic stability (11, 12). However, DNA methylation is also observed in areas with lower CpG content situated close to the CpG islands and termed as CpG shores. Methylation of CpG shores has been found to be mainly tissue specific and also conserved in human and mouse (13, 14). Other recent findings have revealed the existence of non-CG methylation in human stem cells and although it is believed to be a key element in pluripotency its mechanism is yet to be understood (15, 16).

Epigenetic deregulation is a common phenomenon in numerous human pathologies and is currently considered a hallmark of cancer. Tumorigenic cells suffer from extensive aberrations in their epigenetic profiles particularly in respect to methylation patterns. The cancer epigenome is characterised by global hypomethylation, especially at repetitive and parasitic sequences, leading to genomic instability (12, 17). Moreover, hypomethylation has been shown to cause loss of genomic imprinting and subsequent activation of certain proto-oncogenes (18, 19). Another distortion of methylation in cancer cells is the hypermethylation of CpG islands located in the promoters of tumor-suppressor and miRNA genes (20-22). A 2009 study on colorectal tissue has revealed that methylation is also augmented in CpG shores of cancer cells (14).

Abnormal promoter hypermethylation is a very frequent event in lung carcinogenesis(23),(24).Aberrant cellular functions are mediated by the methylation of promoter regions of numerous genes such as *FHIT* (25), *RASSF1A* (26), *APC* (26, 27), *DAPK* (26, 28, 29) and $p16^{INK4a}$. The latter is reported to have a hypermethylation frequency of 22-47% in non-small cell lung cancer NSCLC and is the best studied example of promoter hypermethylation in human lung cancer (26-28, 30, 31). In addition, global genome hypomethylation is a feature of large number of NSCLCs and correlates with genomic instability (12). Distorted methylation patterns have also been extensively observed in head and neck cancers, including hypermethylation of the promoter region of tumor suppressor genes, (32-39) as well as hypomethylation of repetitive sequences

(40, 41). It has also been shown that *p16* hypermethylation can predict malignant transformation of oral dysplasias (42).

The high frequency of abnormal DNA methylation in respiratory tract cancers is therefore an important aspect increasing the potential of DNA methylation-based markers in the clinical management of these cancers. However, the biomarker development and validation route is a long and costly one.

DNA methylation as a biomarker tool

DNA methylation appears advantageous in clinical oncology biomarker discovery, combining a number of attractive characteristics:

1. DNA methylation is a frequent event in human cancer. Discovery studies can point to potential panels of gene-targets providing virtually 100% coverage in the primary disease tissues.
2. It is a covalent modification and thus chemically stable to post sampling process.
3. DNA methylation patterns won't change due to the environmental shock samples are subjected prior to fixation/freezing.
4. DNA methylation changes affect a long stretch of DNA rather than a single nucleotide, facilitating therefore assay design.
5. Last but not least, DNA is the most stable biological macromolecule.

An additional advantage in early detection studies is that CpG island promoter hypermethylation is an early occurrence in the process of carcinogenesis therefore its detection can aid early diagnosis, especially in individuals with a high risk of developing a malignancy such as smokers (43, 44). Currently, however, there are no large longitudinal studies to demonstrate the extent of the clinical benefit for monitoring smokers.

DNA methylation biomarkers: Discovery and validation.

There is a wide spectrum of DNA methylation detection techniques. Based on the principle of 5m-C detection, they can be grouped in (a) affinity methods, utilising anti-5mC antibodies or methyl binding domain (MBD) moieties, (b) methylation sensitive restriction endonuclease (MSRE) methods and (c) sodium bisulphite conversion methods. There is a plethora of downstream applications and combinations of methods used, including microarrays, next generation sequencing (NGS), methylation specific PCR (MSP / qMSP / Methylight), genomic sequencing, pyrosequencing, MALDI-TOF, single strand conformation polymorphism (SSCP) analysis, high resolution melting (HRM), 2D gels and many more (45).

Pyrosequencing Methylation Analysis (PMA) and quantitative methylation specific PCR (qMSP) are currently considered as the gold standard methods, either as standalone methods or for the validation of DNA methylation results derived from high throughput platforms such as microarrays or next generation sequencing. Both methods utilise bisulphite converted DNA template. PMA provides the advantage of quantitative sequencing information over a longer stretch of DNA (Figure 1). The area of interest is amplified using methylation independent (non CpG containing) primers (35, 46). In the subsequent pyrosequencing reaction the methylation level of each CpG is calculated as the ratio of C/T incorporation at this position (Figure 2). The limitation of the method is that it can accurately detect DNA methylation levels of >5% (47). qMSP on the other hand can accurately detect minute DNA methylation levels down to 0.1%. However, this method utilises methylation-specific primers thus it essentially interrogates only the CpGs in the primer region (Figure 2). Therefore the selection method depends on the particularities of the questions asked in each research project.

It is well acknowledged that different phases of biomarker discovery and validation require the use of different techniques. In the early discovery phase, high throughput

techniques such as microarrays and NGS are frequently utilised. These can screen simultaneously thousands of targets reaching up to a genome-wide extent. The issues for consideration here are the signal/noise ratio hampering sensitivity the large amounts of input DNA required as well as the bioinformatic capacity required to reduce the false positives from the next phase. This next phase usually employs techniques such as sequencing, pyrosequencing and MALDI TOF, which can provide methylation information over a significant number of CpGs. Following technical (same set of tissues used in high throughput approach) and biological (independent set of tissues) validation in the primary disease tissue, the qualifying targets are validated in the body fluids. For this step, however, different types of assays are required; these should be able to detect methylated sequences in high dilution of unmethylated normal DNA. A short checklist of considerations for the clinical validation study includes the following:

1. Analytical sensitivity, specificity and reproducibility of the assay.
2. Consistency of technical and biological controls and standards
3. Consistency of specimen process, especially when clinical samples are received from multiple sources. This consistency should be experimentally established using appropriate controls.
4. Patient and sample inclusion/exclusion criteria should be clearly set.
5. Tissue specificity of DNA methylation: are there any normal cell types bearing methylated target copies expected in the clinical specimens? Normal blood exhibits low levels (i.e. in small cell sub-populations) of DNA methylation on many genes. This has to be carefully quantified and recorded.
6. The DNA input is an important issue. A targeted sensitivity requires a proportional number of DNA copies (or cell genomes) to be added in the reaction. This will statistically cover the experiment ensuring that lack of detection is not due to insufficient DNA amount. For example, with a technically established 1:1000 sensitivity, if the DNA input is less than 6.9 ng (1000 copies/genomes) a positive

detection could fail due to stochastic events of the PCR. .It is actually suggested that for increased statistical significance one should ensure at least 4-5 times genome coverage of the targeted sensitivity.

7. Power calculations are absolutely necessary to ensure statistical significance of the study results
8. Biases: identify and quantify all possible biases (age, sex, stage, histological types, lifestyle etc) that the recruitment strategy may confer.

Biomarker studies should comply to the guidelines set by the Early Detection Research Network (EDRN, http://edrn.nci.nih.gov/) and any national guidelines (e,g CR-UK biomarker roadmap, http://science.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@fun/documents/generalcontent/cr_027484.pdf). It is also recommended that diagnostic biomarker studies reporting should comply to the STARD (48) guidelines and/or REMARK guidelines (49) when reporting prognostic biomarkers.

Early detection of respiratory tract tumors by DNA methylation in biological fluids

The current evidence on the utility of DNA methylation as biomarker in respiratory cancer diagnosis is summarized in Table 1.

a. DNA methylation detection in sputum and bronchial washings

Currently, cytological examination of bronchial washings and sputum is routine practice followed for suspected lung cancer cases. However, cytology presents a poor efficiency (50, 51) missing virtually half of the cases. Thus there is a long standing demand for molecular biomarkers that can improve the rate of diagnosis of lung cancer.

Early studies in bronchial washings demonstrated abnormal hypermethylation of many promoters including *p16, RASSF1, RARβ*and *APC* (52-56). More recently, the diagnostic value of *RASSF1A* methylation and *KRAS* mutations in bronchial washings reported a 29% efficiency of detecting malignancy in false-negative or ambiguous

cytology outcomes (57). DNA hypermethylation of *HOXA9* was evaluated by pyrosequencing in 185 induced sputum specimens demonstrating 70.7% sensitivity but a very poor (55.1%) specificity (58). An alternative method to traditional bronchoscopy, bronchoscopic microsampling, is employed in the collection of epithelial lining fluid in the lungs. Analysis of the methylation profile of *APC, ESR1, p16* and *RARβ* genes in such material from 61 patients and resulted in 74% overall detection sensitivity and 96.9% specificity (59).

Although the information gained from these studies demonstrate the potential of using DNA methylation biomarkers in bronchial washings for the diagnosis of lung cancer, they were limited to a preclinical research environment and were not properly clinically validated. The only CE labeled DNA methylation biomarker for lung cancer diagnosis to date is mSHOX2 which has been commercialized by Epigenomics (Berlin) under the name EpiProLung BL Reflex Assay. The assay combines HeavyMethyl™ and TaqMan® technologies and determines of the relative amount of methylated SHOX2 (60, 61). It's analytical performance has been extensively tested demonstrating reproducible positive detection of 0.8% methylated copies (61). The assay has been shown to substantially improve diagnosis of lung cancer in bronchial washings with non-conclusive cytology/pathology results (60, 61). It has also shown to detect abnormal DNA methylation in plasma (62), however its clinical performance there is lower than in bronchial aspirates. An important issue, which is consistent with the very nature of the bronchoscopic examination, is that DNA methylation in bronchial lavage appears more efficient for tumors located at the main bronchi, usually squamous or SCLC, rather than peripheral tumors (63). It is also critical to mention that the sensitivity of any molecular assay in bronchial washings is subject to the high degree of variability of this type of sample. This is because there is no single standardized protocol; bronchoscopy varies significantly between hospitals and, frequently, even between patients under the same bronchoscopist, who tries to deal to his best with the particularities of each case.

b. DNA methylation detection in saliva/oral rinses

Exfoliated oral mucosal cells, both normal and malignant, can be easily and inexpensively obtained by the minimally invasive collection of saliva and oral rinses. Consequently such samples have been utilized in a number of studies intending to investigate the possibility of accurate biomarker-based diagnosis of cancer.

LINE-1 methylation levels are frequently used to evaluate the global genome methylation status (12). LINE-1 hypomethylation was reported in oral rinses derived from oral squamous cell carcinoma (OSCC) patient oral rinses (64). Improved sensitivity and specificity of LINE-1 methylation in both oral rinses and white blood cells (WBCs) was demonstrated from the same group, suggesting that LINE-1 methylation is a potential biomarker for OSCC under the condition that specific LINE-1 methylation patterns are taken into account (65). Promoter methylation of microRNA-137 (miR-137) in OSCC has been previously reported in tissue studies (66). More recently, mir-137 methylation in oral rinses of squamous cell carcinoma patients and of healthy volunteers was shown in 21.2% of all cases and in 3% of control specimens (67). *KIF1A* and *EDNRB* hypermethylation in salivary rinses correlated with the presence of head and neck squamous cell carcinoma (HNSCC) suggesting a potential non-invasive tool in HNSCC diagnosis (68). In the same study, the promoters of *CDH4, TERT, NISCH, PAK3, VGF* and *MAL* were methylated in rinses from healthy individuals, emphasizing on the exclusion criteria for clinically useful biomarkers. The diagnostic value of *KIF1A* and *EDNRB* hypermethylation in salivary rinses for diagnosing oral cavity cancers was also stressed in a different study from the same group (69) suggesting that *EDNRB* methylation can potentially be used in the discrimination among patients with premalignant and malignant oral lesions. *HOXA9* and *NID2* methylation were shown to have a high sensitivity and specificity in OSCC tissue, however their performance in saliva was much lower (70).

Apart from assisting in diagnosis, DNA methylation can also contribute in the surveillance of neoplastic disease. A seven-gene panel signature, (*DAPK, DCC, MINT-31, TIMP-3, p16, MGMT* and *CCNA1*) in salivary rinses is shown to be associated with local recurrence and survival in HNSCC patients (71, 72).

c. DNA methylation detection in blood, plasma and serum

There is an extensive list of papers exploiting blood and its components for the evaluation of DNA methylation biomarkers. Serum and plasma of cancer patients has been shown to contain irregularly high concentrations of DNA and is therefore another type of specimen in which methylation biomarkers can be identified (73). In addition, they are both routine non-invasive clinical samples that can be obtained at low cost.

Analysis of the methylation status of six genes on chromosome 3p (*hOGG1, RAR-B, SEMA3B, RASSF1A, BLU and FHIT*) in peripheral blood mononuclear cell specimens demonstrated that 97.5% of the NSCLC patients presented promoter methylation in at least one of the six genes and 43.8% had at least 3 methylated genes in comparison to 6.3% of normal samples (74). The *p16, RASSF1A, GSTP1, MTHFR* and *MGMT* methylation signature has also been evaluated in peripheral blood cells but no clear association was evident between hypermethylation and case/control status with the exception of *RASSF1A* (75). In addition, peripheral blood leucocytes have been used in the identification of DNA methylation biomarkers with the application of BeadChip technology in small-cell lung carcinoma (SCLC). The assay identified 62 CpG sites which were methylated in SCLC and 9 of these were further validated with pyrosequencing (76).

Following the identification of *SHOX2* methylation as a useful biomarker in bronchial aspirates, effort was given to develop a plasma-based assay for lung cancer detection. The study reported lower sensitivity and specificity compared with bronchial aspirates, and low methylation levels in patients with stage I disease suggesting the use of

additional biomarkers for early detection (62). The suitability of *DCC, KIF1A, NISCH*, and *RARβ* as methylation biomarkers in plasma has been assessed in the plasma of patients with stage I NSCLC suggesting these are promising candidates for early lung cancer detection (77). A different signature incorporating 9 tumor suppressor genes (*APC, CDH13, KLK10, DLEC1, RASSF1A, EFEMP1, SFRP1, RARβ* and *p16*) was evaluated in plasma samples as suitable methylation biomarkers for diagnosis of NSCLC in a Chinese population in which a five gene sub-set (*APC, RASSF1A, CDH13, KLK10, DLEC1*) reached asensitivity of 83.64% and a specificity of 74.0% for cancer diagnosis (78).

Serum samples from lung cancer patients were analyzed to initially examine a panel of 15 genes in respect to aberrant methylation and subsequently the six most promising markers (*APC, CDH1, MGMT, DCC, RASSF1A*, and *AIM1*) were evaluated in an independent set (79).

d. Other samples

Venturing to satisfy the need for non-invasive lung cancer screening procedures *Han et al.* (2009) tested the promoter methylation patterns of *DAPK, RASSF1A and PAX5*β in exhaled breath condensate. Although methylation was detectable in such samples they do not appear to be suitable for early lung cancer detection but may be of use in overall risk assessment (80).

Conclusions

The existing literature clearly demonstrates that (a) DNA methylation is very frequent in respiratory cancers and DNA methylation signatures can be of clinical significance (b) detection of DNA methylation is feasible in relevant biological liquids without changing the current clinical routine of their collection and (c) robust assays compiling the correct DNA methylation biomarkers can assist in early detection and surveillance of these tumors.

However, it is clearly evident that very few studies have moved further than a small preclinical level providing promising results. Large validation projects are rare as also are prospective studies measuring the diagnostic efficiency and the enhancement of current clinical gold standards. Such studies are expensive and require large cohorts of patients. Thus it seems imperative that consortia of research sites with relevant interests should be formed to address this problem.

References

1. Ferlay A. 2011. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10
2. McErlean A, Ginsberg MS. Epidemiology of lung cancer. Semin Roentgenol. 2011 Jul;46(3):173-7.
3. Conway DI, Hashibe M, Boffetta P, Wunsch-Filho V, Muscat J, La Vecchia C, et al. Enhancing epidemiologic research on head and neck cancer: INHANCE - The international head and neck cancer epidemiology consortium. Oral Oncol. 2009 Sep;45(9):743-6.
4. Waddington CH. The Epigenotype. Endeavour. 1942;1(18).
5. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010 Jan;31(1):27-36.
6. Tsai HC, Baylin SB. Cancer epigenetics: linking basic biology to clinical medicine. Cell Res. 2011 Mar;21(3):502-17.
7. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010 Oct;28(10):1057-68.
8. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999 Oct 29;99(3):247-57.
9. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998 Jul;19(3):219-20.
10. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002 Jan 1;16(1):6-21.
11. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007 Apr;8(4):286-98.
12. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int J Cancer. 2009 Jan 1;124(1):81-7.
13. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet. 2009 Dec;41(12):1350-3.
14. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009 Feb;41(2):178-86.
15. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009 Nov 19;462(7271):315-22.
16. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Res. 2010 Mar;20(3):320-31.
17. Esteller M. Epigenetics in cancer. N Engl J Med. 2008 Mar 13;358(11):1148-59.
18. Watt PM, Kumar R, Kees UR. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. Genes Chromosomes Cancer. 2000 Dec;29(4):371-7.
19. Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. Biochim Biophys Acta. 2007 Jan;1775(1):138-62.
20. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science. 1985 Apr 12;228(4696):187-90.
21. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res. 2007 Feb 15;67(4):1424-9.
22. Toyota M, Suzuki H, Sasaki Y, Maruyama R, Imai K, Shinomura Y, et al. Epigenetic silencing of microRNA-34b/c and B-cell translocation gene 4 is associated with CpG island methylation in colorectal cancer. Cancer Res. 2008 Jun 1;68(11):4123-32.
23. Field JK, Liloglou T, Warrak S, Burger M, Becker E, Berlin K, et al. Methylation discriminators in NSCLC identified by a microarray based approach. Int J Oncol. 2005 Jul;27(1):105-11.
24. Ehrich M, Field JK, Liloglou T, Xinarianos G, Oeth P, Nelson MR, et al. Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. Cancer Res. 2006 Nov 15;66(22):10911-8.

25. Verri C, Roz L, Conte D, Liloglou T, Livio A, Vesin A, et al. Fragile histidine triad gene inactivation in lung cancer: the European Early Lung Cancer project. Am J Respir Crit Care Med. 2009 Mar 1;179(5):396-401.
26. Fischer JR, Ohnmacht U, Rieger N, Zemaitis M, Stoffregen C, Manegold C, et al. Prognostic significance of RASSF1A promoter methylation on survival of non-small cell lung cancer patients treated with gemcitabine. Lung Cancer. 2007 Apr;56(1):115-23.
27. Kim DS, Cha SI, Lee JH, Lee YM, Choi JE, Kim MJ, et al. Aberrant DNA methylation profiles of non-small cell lung cancers in a Korean population. Lung Cancer. 2007 Oct;58(1):1-6.
28. Shivapurkar N, Sherman ME, Stastny V, Echebiri C, Rader JS, Nayar R, et al. Evaluation of candidate methylation markers to detect cervical neoplasia. Gynecol Oncol. 2007 Dec;107(3):549-53.
29. Tang X, Khuri FR, Lee JJ, Kemp BL, Liu D, Hong WK, et al. Hypermethylation of the death-associated protein (DAP) kinase promoter and aggressiveness in stage I non-small-cell lung cancer. J Natl Cancer Inst. 2000 Sep 20;92(18):1511-6.
30. Dammann R, Schagdarsurengin U, Seidel C, Strunnikova M, Rastetter M, Baier K, et al. The tumor suppressor RASSF1A in human carcinogenesis: an update. Histol Histopathol. 2005 Apr;20(2):645-63.
31. Chen C, Yin N, Yin B, Lu Q. DNA methylation in thoracic neoplasms. Cancer Lett. 2011 Feb 1;301(1):7-16.
32. Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, et al. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis. 2009 Mar;30(3):416-22.
33. Hasegawa H, Ninomiya M, Honbou T, Hamatsu T, Kudo K, Ushijima C, et al. [A case of long-term survival by combined modality therapy for liver and pulmonary metastasis of rectal cancer]. Gan to kagaku ryoho Cancer & chemotherapy. 2011 Nov;38(12):2310-2.
34. Dikshit RP, Gillio-Tos A, Brennan P, De Marco L, Fiano V, Martinez-Penuela JM, et al. Hypermethylation, risk factors, clinical characteristics, and survival in 235 patients with laryngeal and hypopharyngeal cancers. Cancer. 2007 Oct 15;110(8):1745-51.
35. Shaw RJ, Liloglou T, Rogers SN, Brown JS, Vaughan ED, Lowe D, et al. Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing. Br J Cancer. 2006 Feb 27;94(4):561-8.
36. Shaw RJ, Hall GL, Lowe D, Bowers NL, Liloglou T, Field JK, et al. CpG island methylation phenotype (CIMP) in oral cancer: associated with a marked inflammatory response and less aggressive tumour biology. Oral Oncol. 2007 Oct;43(9):878-86.
37. Shaw RJ, Hall GL, Woolgar JA, Lowe D, Rogers SN, Field JK, et al. Quantitative methylation analysis of resection margins and lymph nodes in oral squamous cell carcinoma. Br J Oral Maxillofac Surg. 2007 Dec;45(8):617-22.
38. Shaw RJ, Omar MM, Rokadiya S, Kogera FA, Lowe D, Hall GL, et al. Cytoglobin is upregulated by tumour hypoxia and silenced by promoter hypermethylation in head and neck cancer. Br J Cancer. 2009 Jul 7;101(1):139-44.
39. Schache AG, Hall G, Woolgar JA, Nikolaidis G, Triantafyllou A, Lowe D, et al. Quantitative promoter methylation differentiates carcinoma ex pleomorphic adenoma from pleomorphic salivary adenoma. Br J Cancer. 2010 Dec 7;103(12):1846-51.
40. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. Oncogene. 2004 Nov 18;23(54):8841-6.
41. Poage GM, Houseman EA, Christensen BC, Butler RA, Avissar-Whiting M, McClean MD, et al. Global hypomethylation identifies Loci targeted for hypermethylation in head and neck cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. [Research Support, N.I.H., Extramural]. 2011 Jun 1;17(11):3579-89.
42. Hall GL, Shaw RJ, Field EA, Rogers SN, Sutton DN, Woolgar JA, et al. p16 Promoter methylation is a potential predictor of malignant transformation in oral epithelial dysplasia. Cancer Epidemiol Biomarkers Prev. 2008 Aug;17(8):2174-9.

43. Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nat Med. 2011 Mar;17(3):330-9.
44. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer. 2004 Sep;4(9):707-17.
45. Martin-Subero JI, Esteller M. Profiling epigenetic alterations in disease. Adv Exp Med Biol. 2011;711:162-77.
46. Daskalos A, Oleksiewicz U, Filia A, Nikolaidis G, Xinarianos G, Gosney JR, et al. UHRF1-mediated tumor suppressor gene inactivation in nonsmall cell lung cancer. Cancer. 2011 Mar 1;117(5):1027-37.
47. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nat Protoc. [Research Support, Non-U.S. Gov't]. 2007;2(9):2265-75.
48. Bossuyt PM, Reitsma JB. The STARD initiative. Lancet. 2003 Jan 4;361(9351):71.
49. McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic factors: what are we missing? Journal of the National Cancer Institute. [Comment
Editorial]. 2005 Jul 20;97(14):1023-5.
50. Dobler CC, Crawford AB. Bronchoscopic diagnosis of endoscopically visible lung malignancies: should cytological examinations be carried out routinely? Intern Med J. 2009 Dec;39(12):806-11.
51. Van't Westeinde SC, van Klaveren RJ. Screening and early detection of lung cancer. Cancer J. 2011 Jan-Feb;17(1):3-10.
52. Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. J Natl Cancer Inst. 1999 Feb 17;91(4):332-9.
53. Kim H, Kwon YM, Kim JS, Lee H, Park JH, Shim YM, et al. Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer. J Clin Oncol. 2004 Jun 15;22(12):2363-70.
54. Topaloglu O, Hoque MO, Tokumaru Y, Lee J, Ratovitski E, Sidransky D, et al. Detection of promoter hypermethylation of multiple genes in the tumor and bronchoalveolar lavage of patients with lung cancer. Clin Cancer Res. 2004 Apr 1;10(7):2284-8.
55. Schmiemann V, Bocking A, Kazimirek M, Onofre AS, Gabbert HE, Kappes R, et al. Methylation assay for the diagnosis of lung cancer on bronchial aspirates: a cohort study. Clin Cancer Res. 2005 Nov 1;11(21):7728-34.
56. Grote HJ. [Aberrant promoter methylation as biomarker for molecular cytological diagnosis of lung cancer]. Verh Dtsch Ges Pathol. 2006;90:216-26.
57. van der Drift M, Prinsen C, Knuiman J, Janssen J, Dekhuijzen PN, Thunnissen E. Diagnosing peripheral lung cancer: the additional value of RASSF1A methylation and KRAS mutation analyses in washings in non-diagnostic bronchoscopy. Chest. 2011 Jun 23.
58. Hwang SH, Kim KU, Kim JE, Kim HH, Lee MK, Lee CH, et al. Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. Clin Chem Lab Med. 2011 Apr;49(4):699-704.
59. Yasuda H, Soejima K, Nakayama S, Kawada I, Nakachi I, Yoda S, et al. Bronchoscopic microsampling is a useful complementary diagnostic tool for detecting lung cancer. Lung Cancer. 2011 Apr;72(1):32-8.
60. Schmidt B, Liebenberg V, Dietrich D, Schlegel T, Kneip C, Seegebarth A, et al. SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. BMC Cancer. 2010;10:600.
61. Dietrich D, Kneip C, Raji O, Liloglou T, Seegebarth A, Schlegel T, et al. Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. Int J Oncol. 2012 Mar;40(3):825-32.
62. Kneip C, Schmidt B, Seegebarth A, Weickmann S, Fleischhacker M, Liebenberg V, et al. SHOX2 DNA Methylation Is a Biomarker for the Diagnosis of Lung Cancer in Plasma. J Thorac Oncol. 2011 Oct;6(10):1632-8.
63. de Fraipont F, Moro-Sibilot D, Michelland S, Brambilla E, Brambilla C, Favrot MC. Promoter methylation of genes in bronchial lavages: a marker for early diagnosis of primary and relapsing non-small cell lung cancer? Lung Cancer. 2005 Nov;50(2):199-209.

64. Subbalekha K, Pimkhaokham A, Pavasant P, Chindavijak S, Phokaew C, Shuangshoti S, et al. Detection of LINE-1s hypomethylation in oral rinses of oral squamous cell carcinoma patients. Oral Oncol. 2009 Feb;45(2):184-91.
65. Pobsook T, Subbalekha K, Sannikorn P, Mutirangura A. Improved measurement of LINE-1 sequence methylation for cancer detection. Clin Chim Acta. 2011 Jan 30;412(3-4):314-21.
66. Kozaki K, Imoto I, Mogi S, Omura K, Inazawa J. Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer. Cancer Res. 2008 Apr 1;68(7):2094-105.
67. Langevin SM, Stone RA, Bunker CH, Grandis JR, Sobol RW, Taioli E. MicroRNA-137 promoter methylation in oral rinses from patients with squamous cell carcinoma of the head and neck is associated with gender and body mass index. Carcinogenesis. 2010 May;31(5):864-70.
68. Demokan S, Chang X, Chuang A, Mydlarz WK, Kaur J, Huang P, et al. KIF1A and EDNRB are differentially methylated in primary HNSCC and salivary rinses. Int J Cancer. 2010 Nov 15;127(10):2351-9.
69. Pattani KM, Zhang Z, Demokan S, Glazer C, Loyo M, Goodman S, et al. Endothelin receptor type B gene promoter hypermethylation in salivary rinses is independently associated with risk of oral cavity cancer and premalignancy. Cancer Prev Res (Phila). 2010 Sep;3(9):1093-103.
70. Guerrero-Preston R, Soudry E, Acero J, Orera M, Moreno-Lopez L, Macia-Colon G, et al. NID2 and HOXA9 promoter hypermethylation as biomarkers for prevention and early detection in oral cavity squamous cell carcinoma tissues and saliva. Cancer Prev Res (Phila). 2011 Jul;4(7):1061-72.
71. Carvalho AL, Henrique R, Jeronimo C, Nayak CS, Reddy AN, Hoque MO, et al. Detection of promoter hypermethylation in salivary rinses as a biomarker for head and neck squamous cell carcinoma surveillance. Clin Cancer Res. 2011 Jul 15;17(14):4782-9.
72. Sun W, Zaboli D, Wang H, Liu Y, Arnaoutakis D, Khan T, et al. Detection of TIMP3 Promoter Hypermethylation in Salivary Rinse as an Independent Predictor of Local Recurrence-Free Survival in Head and Neck Cancer. Clin Cancer Res. 2012 Feb 15;18(4):1082-91.
73. Lo YM. Circulating nucleic acids in plasma and serum: an overview. Ann N Y Acad Sci. 2001 Sep;945:1-7.
74. Liu Z, Li W, Lei Z, Zhao J, Chen XF, Liu R, et al. CpG island methylator phenotype involving chromosome 3p confers an increased risk of non-small cell lung cancer. J Thorac Oncol. 2010 Jun;5(6):790-7.
75. Vineis P, Chuang SC, Vaissiere T, Cuenin C, Ricceri F, Johansson M, et al. DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study. Epigenetics. 2011 Feb;6(2):195-201.
76. Wang J, Wang B, Chen X, Bi J. The prognostic value of RASSF1A promoter hypermethylation in non-small cell lung carcinoma: a systematic review and meta-analysis. Carcinogenesis. 2011 Mar;32(3):411-6.
77. Ostrow KL, Hoque MO, Loyo M, Brait M, Greenberg A, Siegfried JM, et al. Molecular analysis of plasma DNA for the early detection of lung cancer by quantitative methylation-specific PCR. Clin Cancer Res. 2010 Jul 1;16(13):3463-72.
78. Zhang Y, Wang R, Song H, Huang G, Yi J, Zheng Y, et al. Methylation of multiple genes as a candidate biomarker in non-small cell lung cancer. Cancer Lett. 2011 Apr 1;303(1):21-8.
79. Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin Cancer Res. 2011 Jul 1;17(13):4494-503.
80. Han W, Wang T, Reilly AA, Keller SM, Spivack SD. Gene promoter methylation assayed in exhaled breath, with differences in smokers and lung cancer patients. Respir Res. 2009;10:86.

Figure legends

Figure 1. Pyrogram example from TERT promoter methylation analysis. The interrogated sequence (reverse strand sequencing) is shown at the top of the program while the nucleotide dispensation order is shown at the bottom of the graph. The incorporation of each nucleotide is calculated providing information on the DNA methylation level of ten separate CpGs on the sequence. Guanine dispensations 1 and 52 target non-CpG cytosines and are utilised for bisulphite efficiency control

Figure 2. Schematic diagram demonstrating PMA and qMSP assays for p16 promoter. Grey arrows represent qMSP primers targeting CpGs at their 3' end. The assay also includes a probe (grey rectangle) which also spans CpGs. PMA primers (dark arrows) do not span CpGs thus being methylation independent. CpGs are indicated by a capital C.

Figure 1

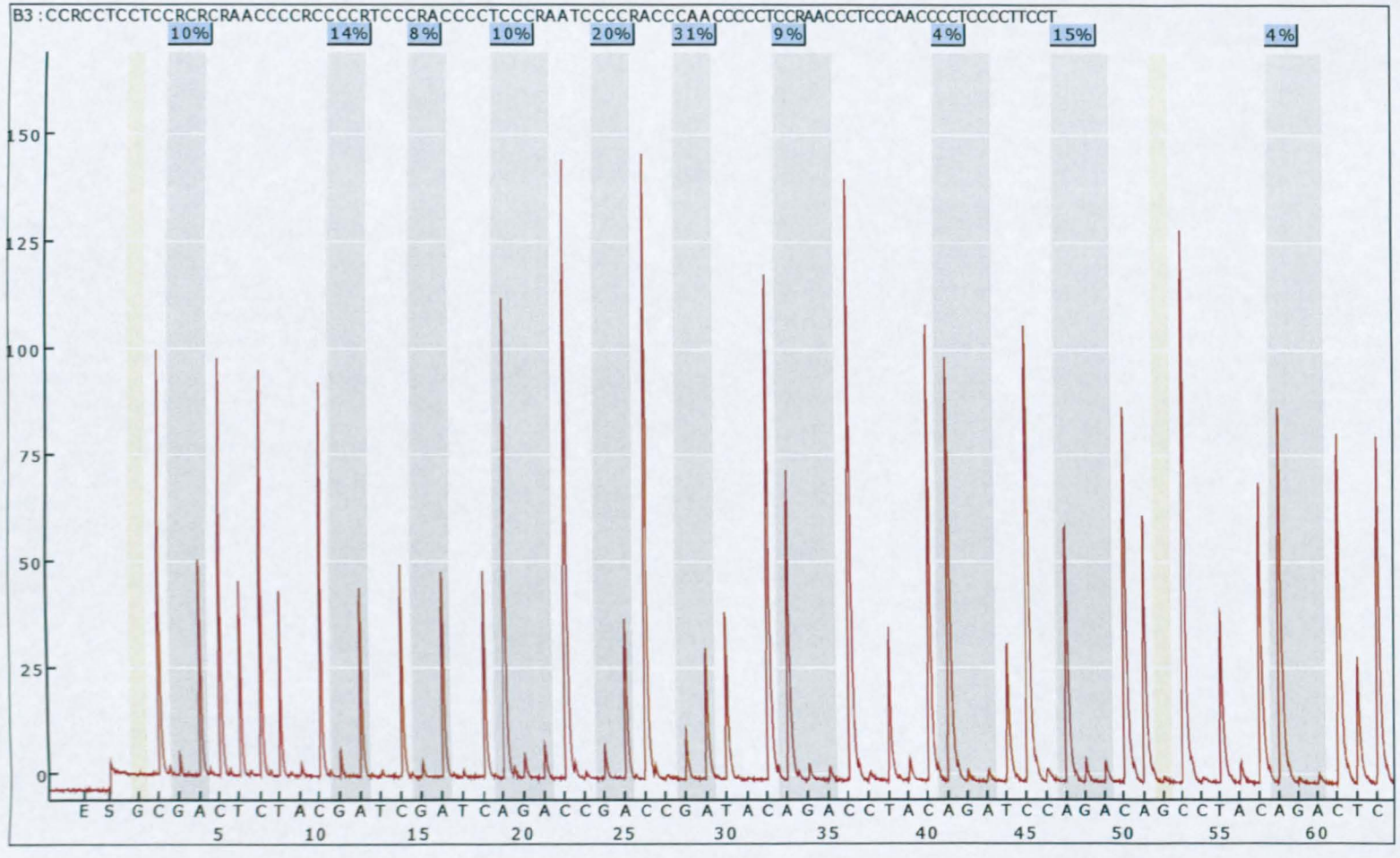
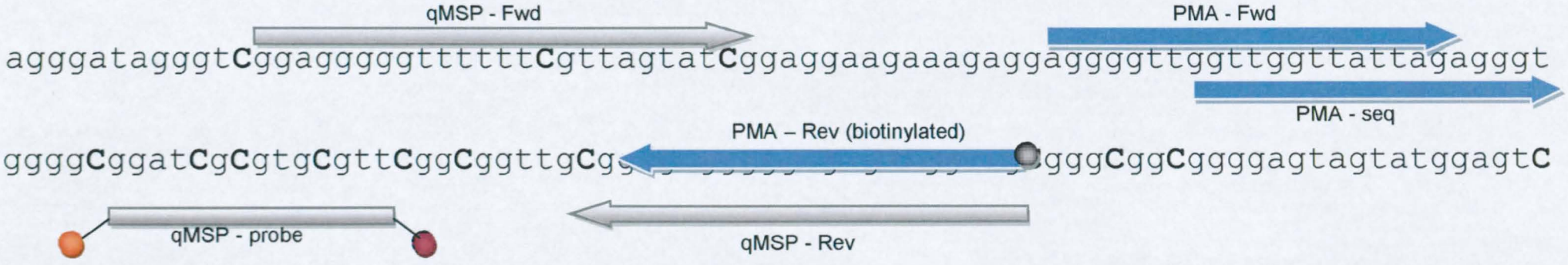

B3 : CCRCCTCCTCCRCRCRAACCCCRCCCCRTCCCRACCCCTCCCRAATCCCCRACCCAACCCCCTCCRAACCCTCCCAACCCCTCCCCTTCCT
10%
14%
8%
10%
20%
31%
9%
4%
15%
4%
150
125
100
75
50
25
0
E S G C G A C T C T A C G A T C G A T C A G A C C G A C C G A T A C A G A C C T A C A G A T C C A G A C A G C C T A C A G A C T C
5
10
15
20
25
30
35
40
45
50
55
60

FIGURE 2.


qMSP - Fwd
PMA - Fwd
agggatagggtCggaggggttttttCgttagtatCggaggaagaaagaggaggggttggttggttattagagggt
PMA - seq
PMA – Rev (biotinylated)
ggggCggatCgCgtgCgttCggCggttgCgggggCggCggggagtagtatggagtC
qMSP - probe
qMSP - Rev

Supporting manuscript submitted to Cancer Research

DNA methylation biomarkers improve diagnostic efficiency of lung cancer.

Georgios Nikolaidis[1]*, Olaide Y Raji[1]*, Soultana Markopoulou[1], John R Gosney[1], Julie Bryan[1], Chris Warburton[2], Martin Walshaw[3], John Sheard[2], John K Field[1], Triantafillos Liloglou[1‡]

[1] University of Liverpool, Department of Molecular & Clinical Cancer Medicine, Liverpool, UK

[2] University Aintree Hospital, Liverpool, UK

[3] Liverpool Heart and Chest Hospital, Liverpool, UK

* These authors had equal contribution to this study

‡ Corresponding author

Address for correspondence during manuscript evaluation

Triantafillos Liloglou
University of Liverpool
Cancer Research Centre
200 London Road
L3 9TA, Liverpool
United Kingdom
T: 0044 151 7948958
F: 0044 151 7948958
E: tliloglo@liv.ac.uk

ABSTRACT

The exceptional high mortality of lung cancer can be instigated to a high degree by late diagnosis. Despite the plethora of studies on potential molecular biomarkers for lung cancer diagnosis, very few have reached clinical implementation. In this study we developed a panel of DNA methylation biomarkers and validated their diagnostic efficiency in bronchial washings from a large retrospective cohort. Candidate targets from previous high-throughput approaches were examined by Pyrosequencing in an independent set of 48 lung tumor/normal paired. Ten promoters were selected and quantitative methylation-specific PCR (qMSP) assays were developed and used to screen 655 bronchial washings (BWs) from the Liverpool Lung Project (LLP) subjects divided into training (194 cases and 214 Controls) and validation (139 cases and 109 controls) sets. Three statistical models were employed to select the optimal panel of markers and evaluate the performance of the discriminatory algorithms. The final logit regression model incorporated hypermethylation at p16, TERT, WT1 and RASSF1.

The performance of this 4-gene methylation signature in the validation set demonstrated 82% sensitivity and 91% specificity. In comparison, cytology alone in this set provided 43% sensitivity at 100% specificity. The diagnostic efficiency of the panel did not show any biases with age, gender, smoking and the presence of a non-lung neoplasm. However, sensitivity was predictably higher in central (squamous and small cell) than peripheral (adenocarcinomas) tumors as well as in pT≥2 stage tumors. These findings clearly demonstrate the impact of DNA methylation-based assays in the diagnosis of cytologically occult lung neoplasms. The incorporation of additional targets will further increase diagnostic efficiency, while a prospective trial is currently imminent in the LLP study to provide data on the enhancement of the diagnostic accuracy in a clinical setting.

INTRODUCTION

Lung cancer causes more deaths than any other neoplasia in both the USA (1) and the UK (2); late detection is a major contributor to this high mortality rates (3). Bronchoscopic examination following suspicious imaging results can reveal the presence of a bronchial lesion, which is normally confirmed histologically by biopsy and/or bronchial washings (BWs - also referred to as bronchial lavage or bronchoalveolar lavage). However, a significant number of cases remain clinically occult after bronchoscopy as cytological examination tends to miss almost half of the cases (4, 5).

The implementation of molecular biomarkers in the early diagnosis of lung cancer has been a long standing goal. Particular focus was given in identifying such biomarkers in bronchial washings in individuals with a high risk of developing lung cancer. Previous attempts in bronchial washings to detect known molecular abnormalities in lung cancer, included genomic instability (6, 7), DNA mutations (8, 9) and more recently, DNA methylation (10, 11). The latter has certain advantages regarding its biomarker applicability; it is a covalent DNA modification, resistant to post-sampling processing and spans a significant nucleotide length, allowing for flexible assay design (12).

The feasibility of DNA methylation detection in the BW of lung cancer patients has been demonstrated in a number of studies (13-15) (reviewed in (12) and(16)). However, very few of the proposed biomarkers have been validated in large case control data sets. One such validated biomarker that has recently received CE IVD certification, under the commercial name of Epi proLung® BL Reflex Assay (Epigenomics, AG) is *m*SHOX2 (17)

In the current study, we describe the validation of a panel of DNA methylation biomarkers in a large retrospective case-control bronchial washings set (655 individuals) from the Liverpool Lung Project (18) and demonstrate a substantial gain in sensitivity of detection over standalone cytology.

METHODS

Study design

A brief outline of the study development is shown in Figure 1. The study extends over biomarker development phases 1 and 2, based on the EDRN guidelines (19). The promoter targets (*p16, RASSF1, TMEFF2, TERT, CYGB, RARB, DAPK1, p73, WT1* and *CDH13*) were identified from previous work of our group (18, 20-23) and others (24-29) and validated by pyrosequencing in an independent set of 48 primary NSCLC surgical tissues (Supplementary Table 1). Quantitative Methylation PCR (qMSP) assays were developed for these ten markers in order to screen the bronchial washing specimens. For this phase, two nested case-control bronchial washing sets were selected from the Liverpool Lung Project (LLP) retrospective cohort. Inclusion criteria were, specimens with>2 years post-sampling follow-up information obtained through hospital records, the Merseyside & Cheshire Cancer Registry (MCCR) and the Office of National Statistics (ONS). Specimens were excluded if extracted DNA failed in quality control (see below in the qMSP description section). The case-control distributions of epidemiological and clinical characteristics for subjects in the training and test datasets are shown in Table 1 demonstrating overall similar patterns between the two classes, with the exception of smoking. Samples were randomized in 96-well plates and tested in a blinded fashion.

Study size and power calculations

Power calculations were based on the target methylation frequencies found in the validation lung cancer tissue set (Supplementary Table 2). Assuming a minimum of 87% positives for at least two markers (null hypothesis, TPR0=0.87) and an anticipated sensitivity of 95% for the markers combination (alternative hypothesis, TPR1 =0.95) we deduce power associated with different sample sizes, case-control ratios and acceptable false positive rates in a simulation study (30) as shown in Supplementary Table 2. This indicated that a set of ≥200 cases is required in a 1:1 ratio with controls to achieve 86% power for a 5% false positive rate at the 95% confidence level.

Patients, samples and DNA.

The two study sets comprised a total of 655 individuals (333 lung cancer cases / 322 age/sex-matched controls) (Figure 1). Patients had been retrospectively recruited through the Liverpool Heart & Chest Hospital under the Liverpool Lung Project (LLP) umbrella. All patients were referred to the bronchoscopy clinics with a clinical suspicion of lung cancer. At the end of the clinical work up, the diagnoses for the majority of non lung cancer patients were, bronchitis, COPD, bronchiectasis and chest infections while at lower frequency heart conditions, sarcoidosis, asbestosis were diagnosed. It has to be noted that 36 individuals in the control group(s) had other (non-lung) cancers diagnosed such as colon, breast, prostate, skin, esophagus and oral as well as four mesotheliomas. The Liverpool Lung Project has received ethical approval and all the recruited patients provided informed consent.

DNA from frozen lung tumor and paired normal tissue was extracted as previously described (22). Bronchial washings were stored in Saccomanno's fixative in an air-

conditioned (18°C) room and the specimens' cytological adequacy was judged by the presence of alveolar macrophages. DNA was extracted using the Blood and Tissue kit (Qiagen), quantified using Picogreen (Invitrogen) and up to 1μg DNA was bisulphate converted using the EZ-96 DNA Methylation-Gold Kit (ZymoResearch).

Pyrosequencing Methylation Analysis (PMA)

Pyrosequencing assays were designed for early validation of targets in the lung tumor solid tissue. Standard protocols that have been previously described (22, 23) were used. The primers for the pyrosequencing analysis are provided in Supplementary Table 3.

Quantitative Methylation Specific PCR (qMSP)

The qMSP assays were designed to specifically amplify bisulphite-converted methylated DNA target sequences in the presence of an excess of unmethylated counterpart sequences. The methylation-specific primer and probe sequences are listed in Supplementary Table 4. During assay development, it became apparent that probes bearing minor groove binding moiety (Taqman MGB probes) provided a significant improvement in the assay specificity and, due to their smaller size, allowed for a more flexible assay design. A lengthy optimization process identified primer concentrations and thermal profiles ensuring reproducible specificity. The qMSP reactions contained 1× TaqMan® Universal Master Mix II (Applied Biosystems) 250 nM probe, 300-900 nM primers (Supplementary Table 2) and 2 μl eluate from the bisulphate treated DNA sample. The reactions were performed on a 7500 FAST real time cycler (Applied Biosystems) under the following thermal profile: 95°C for 10 min, 50 cycles consisted of 95°C for 15 sec, 60°C -62°C for 1 min.

The sensitivity/specificity of the assays was tested on serial dilutions of artificially (SssI) methylated DNA in lymphocyte DNA. In addition, whole genome amplified (WGA) DNA was constructed unsing the RepliG screening kit (Qiagen) as an absolute unmethylated DNA standard. Following multiple repetitions the sensitivity threshold was selected to 0.5% (1:200) as it provided total reproducibility, while higher dilutions (0.1%) proved less reliable. A methylation-independent assay with non-CpG bearing primers/probe was designed for the ACTB gene in order to normalize for input DNA, but also to be used as an exclusion criterion. We experimentally established that a cycle threshold (Ct)=29 for ACTB assays corresponded 6.9 ng DNA (1000 diploid genomes). This cut-off was employed to ensure 5× genome coverage at the 1:200 sensitivity threshold.

The training set was screened with CYGB, p16, RASSF1, TERT, CDH13, TMEFF2, p73, DAPK1, RARβ and WT1. Following statistical analysis, CYGB, p16, RASSF1, TERT, RARβ and WT1, which demonstrated the highest independent sensitivity/specificity or selected by various multivariate models, were evaluated in the independent validation set.

Statistical analysis

Exploratory univariate analysis

The distribution of subjects' epidemiological, clinical and methylation characteristics were described separately for training and testing datasets. Categorical characteristics were compared between cases and controls using Chi-square test and Fisher's exact test when less than 5 individuals were observed. Student t-test was used to investigate statistical significant case-control difference in quantitative characteristics.

The Mann-Whitney non-parametric test was employed where normality assumption failed.

Identification of optimum markers

Univariate exploratory analysis was used to provide insight into the marginal effect of each marker on subject status. The best generalized linear model (best GLM) was used to identify the best additive logit combination mostly predictive of subject status. The model was fitted using Akaike information criterion (AIC), Bayesian information criterion (BIC, BICq) and cross-validation (CV) as selection methods. Multifactor Dimensionality Reduction (MDR) was used to investigate non-additive combination of the markers, which provides an assessment of epistasis (non-linear interactions) among the markers (31). The significance of the association between subject's disease status and each marker interaction was tested based on the Model-based Multifactor Dimensionality Reduction permutation test (32).

Model-based logit algorithms were derived in the training dataset for discrimination and prediction of subject status and validated in the testing dataset. These were done separately for (a) the top 6 markers from the univariate analysis, (b) markers selected from the 'overall' best subset GLM and (c) markers from the 'overall' best MDR combination. Cytology was added as an additional factor to the best discriminatory/predictive model.

The predictive performance of each algorithm was evaluated in the test data. The disease probability (ranging from 0 to 1) was used to classify (training subjects) or predict (test subjects) as cases for probability ≥0.5 or controls otherwise. The classification and predictive accuracies were assessed using diagnostic measures including accuracy, sensitivity and specificity. The area under ROC curve (AUC) was

used to summarize performance over the range of predicted probabilities. The overall performance of the best discriminatory model and its extended version that incorporates cytology was evaluated in a combined training and testing data, stratified by epidemiological and clinical factors such as age, gender, smoking status, lung cancer histological subtype and time distance from specimen collection to diagnosis. The independent ROC-AUCs from the stratified analyses were compared using the DeLong test (33) extended for unpaired ROC curves.

RESULTS

Diagnostic efficiency of the DNA methylation panel.

Pyrosequencing methylation analysis of the set of 48 surgical NSCLC specimens resulted in a set of 10 promoters (CYGB, p16, RASSF1, TERT, CDH13, TMEFF2, p73, DAPK1, RARβ and WT1) demonstrating high frequency of methylation in tumor tissue and absence of methylation in the normal adjacent counterpart (Supplementary Table 1). The training BW case-control set was subsequently screened with the developed qMSP assays. Three statistical models (Univariate association test, Marker combination by Best Subset Regression (BSR) and Markers combination by Multifactor Dimensionality Reduction (MDR)) were tested in order to identify the optimal marker panel(s) and algorithm(s) for improved diagnostic efficiency. Univariate analysis of the ten examined markers is presented in Table 2. All three models pointed to six markers (CYGB, p16, RASSF1, TERT, RARβ and WT1) which were subsequently used to screen the validation set (Supplementary Table 5).

The performance of the different discriminatory algorithms in training and validation data is shown in Table 3. All the logit discriminatory algorithms performed well in the training set. The performance of the top 6 univariate markers and the best subset with BICq or CV in the test data was similar, although the best subset algorithm was more sensitive but less specific in the training data. The MDR algorithm was slightly more specific but less sensitive than the best subset model with BICq or CV criteria in the training data, its performance in the test data was only similar to that of the best subset in terms of specificity (sp=0.98) and lower with regards to sensitivity (se=0.77). The addition of the top MB-MDR 2- and 3-way interactions into any of the best logit models did not alter their performances (data not shown). Overall, the best subset logit

model with BICq or CV criteria including TERT, WT1, p16 and RASSF1, is the most parsimonious and best performed algorithm which was then applied in the validation set.

The diagnostic efficiency of this algorithm in the validation set is depicted in Table 4. The sensitivity of the panel was higher (90%) in the cytology positive cases than the cytology negative ones (75%). Overall the sensitivity was 82% while the specificity is very high (91%). Therefore the panel classified correctly 213/248 individuals of the validation set (diagnostic accuracy = 85.9%). When including the cytology result into the model the sensitivity was similar while specificity (92%) and diagnostic accuracy (86.3%). However, the diagnostic efficiency of the methylation panel is profoundly higher in comparison to the cytological evaluation alone, which demonstrates 45% sensitivity and 99% specificity.

Overall performance of the panel in clinical subgroups.

Following the validation of the 4-gene panel signature in the test set, an overall performance analysis of this panel was undertaken including both sets. The purpose of this was to identify possible biases in diverse epidemiological and clinical subgroups. Table 5 demonstrates the details of this analysis. The model performed equally among different age and gender groups. In addition, no differences in the sensitivity and specificity of detection were observed in relation to the age of the specimen in storage. Most importantly no significant sensitivity/specificity differences were observed among different smoking groups. Interestingly, the specificity of the panel was similar to the control sub-group bearing no malignant tumors at all (82.1%) and the control sub-group bearing tumors in other organs of the body except lung

(83.3%). As expected, the sensitivity of the methylation panel was higher in cytology positive bronchial washings. It was also evident that stage T1 tumors were less detectable (63%) than T2, T3 and T4 (over 80%) while no such difference was seen for nodal metastasis. When comparing sensitivities of cytology and DNA methylation in diverse pT groups (Figure 2), two points become obvious; (1) DNA methylation sensitivity is consistently higher in all groups than that of cytology and (2) that cytology demonstrates higher sensitivity in T4s as opposed to T1, T2, T3 (chi-square test, p=0.002) while DNA methylation has equally high sensitivity in T2, T3, T4 compared to T1 (chi-square test, p=0.004) . Concerning histology, the panel demonstrated a higher efficiency in detecting small cell (100%) and squamous cell (83%) lung tumors in comparison to adenocarcinomas (75%). Our cohort also included a few inoperable cases with unconfirmed pathology thus such analysis was not applicable to these specimens.

Discussion

The late diagnosis of lung cancer remains the major reason for the large number of deaths due to this disease. Earlier diagnosis with successful surgical intervention is currently the best way forward. The advent of early detection through CT screening holds future promise but still has to be implemented (34, 35). Cytological diagnosis of the disease remains one of the major investigative tools, unfortunately it misses up to half of the cases. Thus the diagnostic efficiency in cytologically occult bronchoscopic material is essential. Despite the number of articles suggesting potential biomarkers, very few have progressed to the next level towards clinical evaluation. The main reasons include low study size and thus statistical power, extensive diversity of methods and lack of assay optimization to reach clinical standards (12).

In this study, we undertook a retrospective case-control design to evaluate DNA methylation biomarkers utilizing a training and a validation sample set (overall 655 individuals) from the Liverpool Lung Project. The study was designed to maximize compliance to the EDRN guidelines (19, 36), the Cancer Research UK Diagnostic Biomarker Roadmap (37) and STARD (38) recommendations for reporting in this manuscript. We developed very robust qMSP assays and established sensitivity and specificity through thousands of repetitions. qMSP is currently considered the gold standard method for reliably detecting DNA methylation in high dilution (39, 40). It must be noted that white blood cells, which are frequently present in bronchial washings, are not *de facto* methylation-free for all genes. Thus we determined a positive control based cut-off (0.5% methylation dilution), which was *always* at least 4-fold higher (>2 ΔCt) from the lymphocyte methylation signal. We have also used a methylation-independent assay for the ACTB gene to quantify the DNA input and thus (a) be used as an exclusion criterion, indicative of inadequate amount of DNA and (b) provide normalization for the target gene signal.

Our biomarker qualification process through training and validation sets demonstrated that a panel of TERT, WT1, p16 and RASSF1 methylation markers provides a parsimonious and efficient algorithm for correctly predicting lung cancer status in 85.9% of tested bronchial washing specimens. We utilized three different models to identify a useful marker panel and develop the discriminatory/predictive algorithm utilizing them. The consistency of various analyses conducted supports the usefulness of the markers, providing further support to previous suggestions on the use of marker panels than single markers in order to improve sensitivity and specificity (41, 42).

RASSF1 methylation in bronchial washings has been recently shown to increase diagnostic sensitivity (40) Our study is also in agreement with a previous report on p16 and RASSF1 and RARβ methylation specificity in cancer cases (although RARβ was not eventually included in the final panel) (43). However, CDH13 appears as a cancer-specific marker in the latter while in our study had clearly no discrimination efficiency. The methodological approach (endpoint MSP vs qMSP) may be a source of this difference.

It is apparent that the DNA methylation panel reported in this manuscript has superior sensitivity (82%) compared to cytology alone (45%), while its specificity is marginally lower (92%). Cytology is currently the clinical gold standard for BW evaluation but it is known to have a low sensitivity of detection (4, 5). Therefore the use of DNA methylation biomarkers can be used in a clinical setting to improve the diagnostic efficiency for lung cancer. The incorporation of cytology into the model did not alter the diagnostic efficiency in our validation set. A larger cohort of specimens is currently being recruited in the LLP in an effort to confirm whether this DNA methylation marker panel can substitute or complement the cytological report in bronchial washings for lung cancer diagnosis. In any case the diagnostic benefit of this panel in cytologically occult specimens is profound.

It is important that this panel demonstrated no biases related to age and gender. Most importantly, its diagnostic efficiency is independent of the smoking status, suggesting that it detects cancer-specific alterations rather than tobacco-related field cancerization. It is also of note that correct classification was not influenced by the presence of other (non-lung) cancers in the control population. As RASSF1, TERT, WT1 and p16 are common epigenetic players in cancer development, this has to do

with the origin of the specimen (i.e. the lung) rather than the specificity of these four markers to lung cancer.

The better performance in central (small cell and squamous carcinomas) rather than peripheral adenocarcinomas was not surprising as bronchoscopy is expected to sample the latter at lower efficiency. It is of note that while our initial marker selection was based on adenocarcinoma and squamous carcinoma tissue only, their performance in the BW from patients with other histological subtypes (e.g small cell, carcinoids etc) was equally efficient. It is also of no surprise that lower sensitivity was achieved in smaller (T1) tumors, as these presumable seed less cells in the lung cavity. It is still important though that DNA methylation detected more than half of T1 tumors, group in which cytology has particularly low sensitivity.

Although the current sensitivity can be improved by expanding the existing panel, it is still almost double of the current gold standard, which is cytology. Thus clinical implementation could proceed provided that the diagnostic efficiency reported here is further validated in an independent cohort; preferably a multi-site case control study should be undertaken. One of the main problems appears to be the potential shortage of DNA from bronchial washings if higher numbers of markers need to be included. This can be overcome by the use of microfluidic PCR arrays that significantly reduce reaction volumes and thus required input DNA.

In this study, we utilized a training and a validation cohort to identify a panel of DNA methylation based biomarkers with potential diagnostic utility for lung cancer detection in bronchial washings specimens. This 4-marker panel significantly improves the diagnosis rate compared to cytological evaluation only clearly demonstrating that DNA methylation biomarkers can become a useful clinical tool for the diagnosis of lung

cancer, especially in cytologically occult bronchoscopic material. However, the timely delivery of such molecular diagnostic tools can only be accomplished through consortia which share samples and information and utilize common methodologies throughout the diagnostic process from sampling to reporting.

WORD COUNT: 3600

Acknowledgements

This research was supported by Cancer Research UK and the Roy Castle Lung Cancer Foundation, UK. Authors would like to thank Mrs Jenny Goggin for assisting with specimen processing.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: a cancer journal for clinicians. 2010;60:277-300.
2. Jack RH, Davies EA, Moller H. Lung cancer incidence and survival in different ethnic groups in South East England. Br J Cancer. 2011;105:1049-53.
3. Mulshine JL, van Klaveren RJ. Lung cancer screening: what is the benefit and what do we do about it? Lung cancer (Amsterdam, Netherlands). 2011;71:247-8.
4. Van't Westeinde SC, van Klaveren RJ. Screening and early detection of lung cancer. Cancer J. 2011;17:3-10.
5. Dobler CC, Crawford AB. Bronchoscopic diagnosis of endoscopically visible lung malignancies: should cytologic examinations be performed routinely? Intern Med J. 2009.
6. Liloglou T, Maloney P, Xinarianos G, Hulbert M, Walshaw MJ, Gosney JR, et al. Cancer-specific genomic instability in bronchial lavage: a molecular tool for lung cancer detection. Cancer research. 2001;61:1624-8.
7. Ahrendt SA, Decker PA, Alawi EA, Zhu Yr YR, Sanchez-Cespedes M, Yang SC, et al. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. Cancer. 2001;92:1525-30.
8. Somers VA, van Henten AM, ten Velde GP, Arends JW, Thunnissen FB. Additional value of K-ras point mutations in bronchial wash fluids for diagnosis of peripheral lung tumours. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology. 1999;13:1120-4.
9. Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. J Natl Cancer Inst. 1999;91:332-9.

10. Schneider KU, Dietrich D, Fleischhacker M, Leschber G, Merk J, Schaper F, et al. Correlation of SHOX2 gene amplification and DNA methylation in lung cancer tumors. BMC Cancer. 2011;11:102.
11. de Fraipont F, Moro-Sibilot D, Michelland S, Brambilla E, Brambilla C, Favrot MC. Promoter methylation of genes in bronchial lavages: a marker for early diagnosis of primary and relapsing non-small cell lung cancer? Lung cancer (Amsterdam, Netherlands). 2005;50:199-209.
12. Liloglou T, Field JK. Detection of DNA methylation changes in body fluids. Adv Genet. 2010;71:177-207.
13. Baryshnikova E, Destro A, Infante MV, Cavuto S, Cariboni U, Alloisio M, et al. Molecular alterations in spontaneous sputum of cancer-free heavy smokers: results from a large screening program. Clin Cancer Res. 2008;14:1913-9.
14. Hsu HS, Chen TP, Wen CK, Hung CH, Chen CY, Chen JT, et al. Multiple genetic and epigenetic biomarkers for lung cancer detection in cytologically negative sputum and a nested case-control study for risk assessment. J Pathol. 2007;213:412-9.
15. Grote HJ. [Aberrant promoter methylation as biomarker for molecular cytological diagnosis of lung cancer]. Verh Dtsch Ges Pathol. 2006;90:216-26.
16. Liloglou T, Bediaga NG, Brown BR, Field JK, Davies MP. Epigenetic biomarkers in lung cancer. Cancer letters. 2012.
17. Dietrich D, Kneip C, Raji O, Liloglou T, Seegebarth A, Schlegel T, et al. Performance evaluation of the DNA methylation biomarker SHOX2 for the aid in diagnosis of lung cancer based on the analysis of bronchial aspirates. International journal of oncology. 2012;40:825-32.
18. Field JK, Smith DL, Duffy S, Cassidy A. The Liverpool Lung Project research protocol. International journal of oncology. 2005;27:1633-45.
19. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93:1054-61.
20. Ehricht R, Slickers P, Goellner S, Hotzel H, Sachse K. Optimized DNA microarray assay allows detection and genotyping of single PCR-amplifiable target copies. Mol Cell Probes. 2006;20:60-3.
21. Xinarianos G, McRonald FE, Risk JM, Bowers NL, Nikolaidis G, Field JK, et al. Frequent genetic and epigenetic abnormalities contribute to the deregulation of cytoglobin in non-small cell lung cancer. Human molecular genetics. 2006;15:2038-44.
22. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. International journal of cancer. 2009;124:81-7.
23. Daskalos A, Logotheti S, Markopoulou S, Xinarianos G, Gosney JR, Kastania AN, et al. Global DNA hypomethylation-induced DeltaNp73 transcriptional activation in non-small cell lung cancer. Cancer Lett. 2011;300:79-86.
24. Palmisano WA, Divine KK, Saccomanno G, Gilliland FD, Baylin SB, Herman JG, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer research. 2000;60:5954-8.
25. Virmani AK, Rathi A, Zochbauer-Muller S, Sacchi N, Fukuyama Y, Bryant D, et al. Promoter methylation and silencing of the retinoic acid receptor-beta gene in lung carcinomas. J Natl Cancer Inst. 2000;92:1303-7.
26. Toyooka S, Toyooka KO, Maruyama R, Virmani AK, Girard L, Miyajima K, et al. DNA methylation profiles of lung tumors. Mol Cancer Ther. 2001;1:61-7.

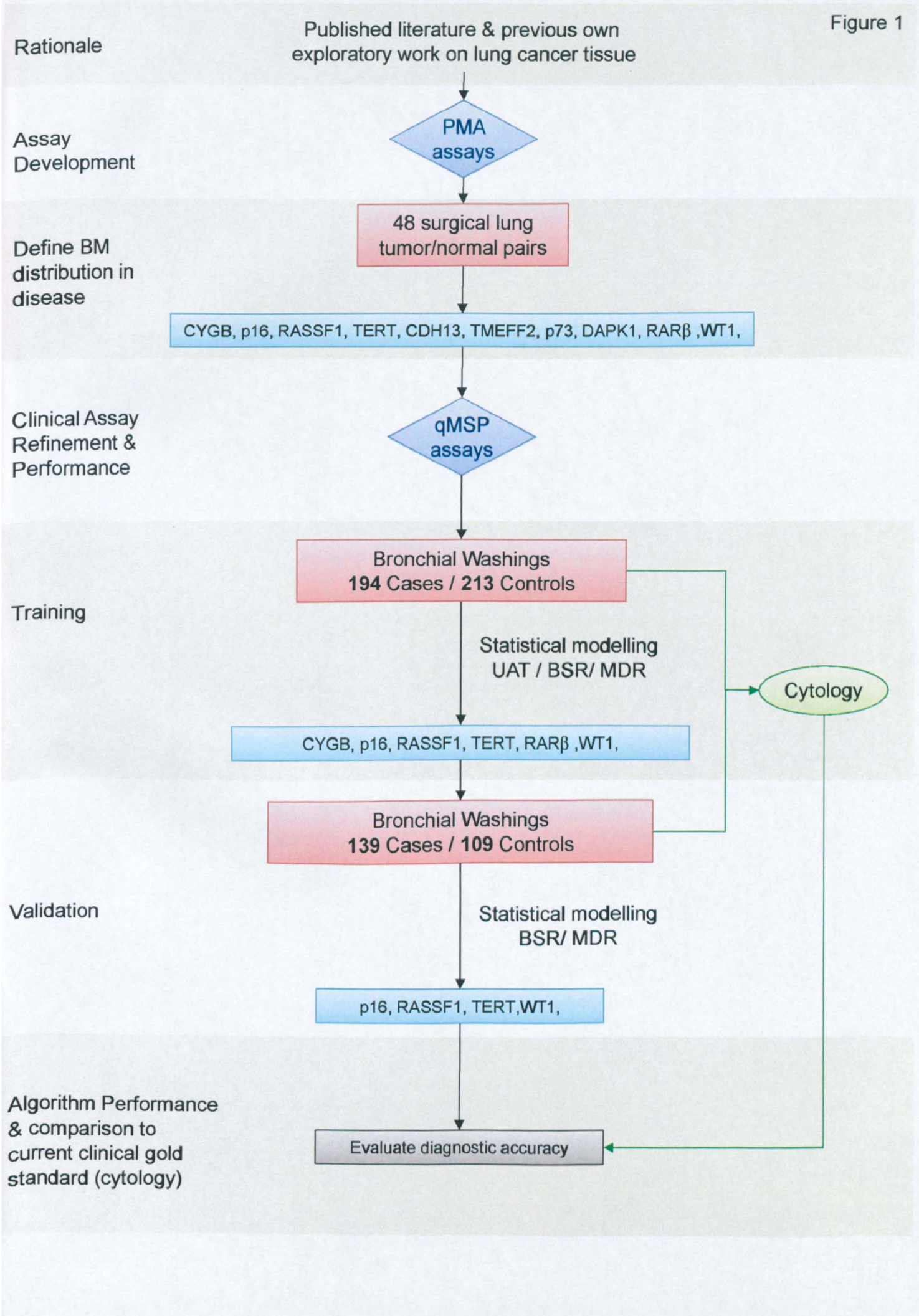
27. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer research. 2002;62:2370-7.
28. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. International journal of cancer. 2003;107:612-6.
29. Shivapurkar N, Stastny V, Suzuki M, Wistuba, II, Li L, Zheng Y, et al. Application of a methylation gene panel by quantitative PCR for lung cancers. Cancer Lett. 2007;247:56-71.
30. Janes H, Pepe M. The optimal ratio of cases to controls for estimating the classification accuracy of a biomarker. Biostatistics. 2006;7:456-68.
31. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001;69:138-47.
32. Calle ML, Urrea V, Vellalta G, Malats N, Steen KV. Improving strategies for detecting genetic patterns of disease susceptibility in association studies. Stat Med. 2008;27:6532-46.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.
34. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N Engl J Med. 2011.
35. Field JK, Smith RA, Aberle DR, Oudkerk M, Baldwin DR, Yankelevitz D, et al. International Association for the Study of Lung Cancer Computed Tomography Screening Workshop 2011 Report. J Thorac Oncol. 2012;7:10-9.
36. Baker SG, Kramer BS, Srivastava S. Markers for early detection of cancer: statistical guidelines for nested case-control studies. BMC Med Res Methodol. 2002;2:4.
37. CRUK. (http://science.cancerresearchuk.org/funding/find-grant/all-funding-schemes/biomarkers-imaging-discovery-development-project-grants/ssLINK/CR_027484. 2009.
38. Bossuyt PM, Reitsma JB. The STARD initiative. Lancet. 2003;361:71.
39. Hoque MO, Topaloglu O, Begum S, Henrique R, Rosenbaum E, Van Criekinge W, et al. Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects. J Clin Oncol. 2005;23:6569-75.
40. van der Drift M, Prinsen C, Knuiman J, Janssen J, Dekhuijzen PN, Thunnissen E. Diagnosing peripheral lung cancer: the additional value of RASSF1A methylation and KRAS mutation analyses in washings in non-diagnostic bronchoscopy. Chest. 2011.
41. Kumar S, Mohan A, Guleria R. Biomarkers in cancer screening, research and detection: present and future: a review. Biomarkers. 2006;11:385-405.
42. Manne U, Srivastava RG, Srivastava S. Recent advances in biomarkers for cancer diagnosis and treatment. Drug Discov Today. 2005;10:965-76.
43. Kim H, Kwon YM, Kim JS, Lee H, Park JH, Shim YM, et al. Tumor-specific methylation in bronchial lavage for the early detection of non-small-cell lung cancer. J Clin Oncol. 2004;22:2363-70.

Figure Legends

Figure 1. Outline of the study progress phases. The distribution of candidate biomarker (BM) targets is validated for by Pyrosequencing Methylation analysis (PMA) in an independent set of lung cancer tissues. Quantitative methylation-specific PCR (qMSP) assays are developed and evaluated for their robustness in clinical samples. These are used to screen the training bronchial washings (BW) set from lung cancer patients and age/sex matched controls. Statistical modeling demonstrates six markers with higher discriminating efficiency and these are used to screen the validation BW set. Further statistical modeling is applied to test the derived algorithms in the validation set. The qualifying 4-marker panel incorporates cytological data in order to construct the final algorithm. UAT: Univariate Association Test, BSR: Best Subset Regression, MDR: Multifactor Dimensionality Reduction.

Figure 2. Sensitivities of cytology and DNA methylation in different pathological stages of lung cancer. DNA methylation demonstrates superior sensitivity across all stages. D: DNA methylation panel, C: cytology.

Figure 1


Rationale
Published literature & previous own exploratory work on lung cancer tissue
Assay Development
PMA assays
Define BM distribution in disease
48 surgical lung tumor/normal pairs
CYGB, p16, RASSF1, TERT, CDH13, TMEFF2, p73, DAPK1, RARβ ,WT1,
Clinical Assay Refinement & Performance
qMSP assays
Training
Bronchial Washings
194 Cases / 213 Controls
Statistical modelling
UAT / BSR/ MDR
Cytology
CYGB, p16, RASSF1, TERT, RARβ ,WT1,
Bronchial Washings
139 Cases / 109 Controls
Validation
Statistical modelling
BSR/ MDR
p16, RASSF1, TERT,WT1,
Algorithm Performance & comparison to current clinical gold standard (cytology)
Evaluate diagnostic accuracy

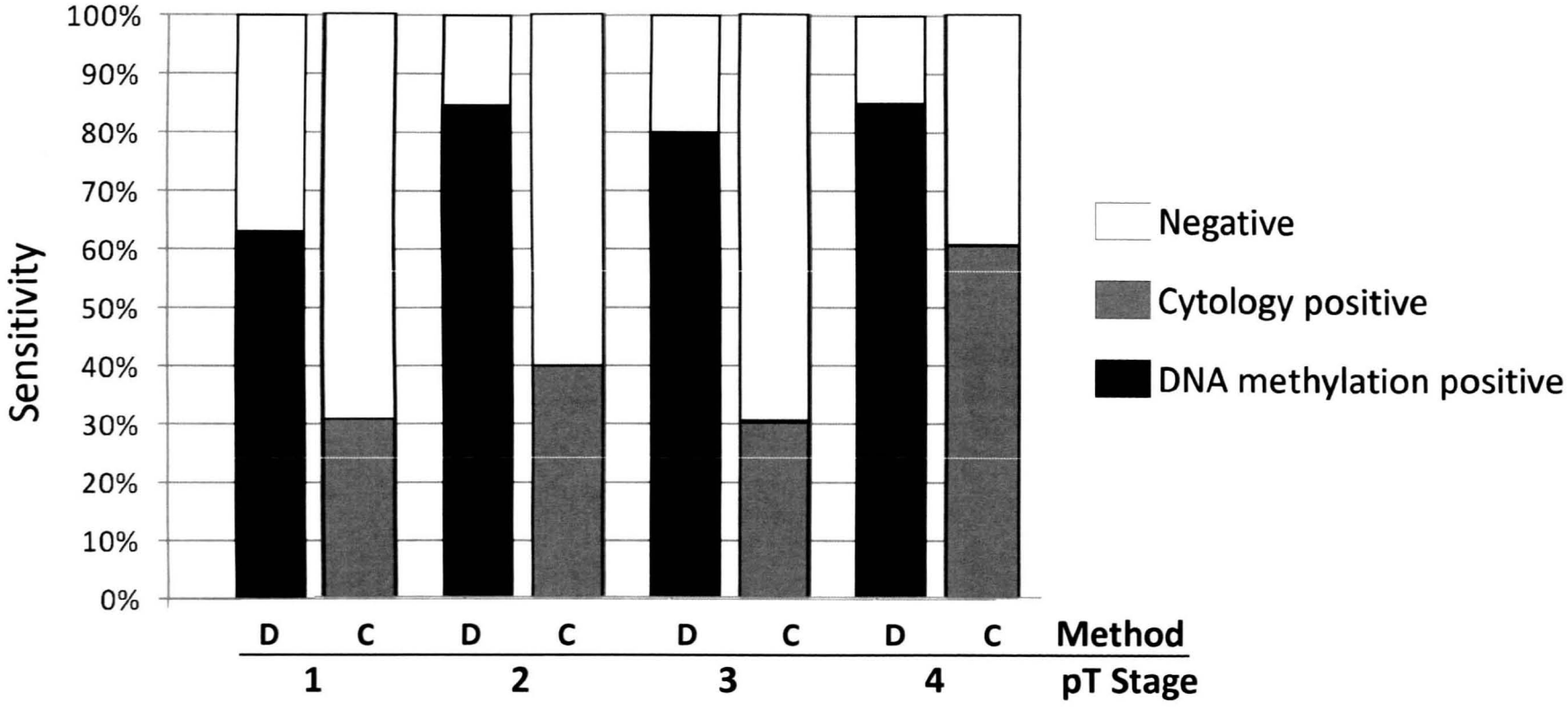

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Sensitivity
D
C
D
C
D
C
D
C
Method
1
2
3
4
pT Stage
Negative
Cytology positive
DNA methylation positive

Table 1: Frequency distribution of subject's epidemiological and clinical characteristics by case-control status

Subject characteristics	Training set (N=407)		Testing set (N=248)	
	Case (n=194)	Control (n=213)	Case (n=139)	Control (n=109)
Age group[↑]				
<60	33 (17.0)	57 (26.8)	18 (13.0)	19 (17.4)
60-79	150 (77.3)	144 (67.6)	110 (79.1)	84 (77.1)
80+	11 (5.7)	12 (5.6)	11 (7.9)	6 (5.5)
Age summary statistic[↑]				
mean ± sd	68.7±7.56	66.4± 8.56	68.4±8.07	67.6±8.78
Gender				
Male	114 (58.8)	115 (54.0)	80 (57.6)	63 (57.8)
Female	80 (41.2)	98 (46.0)	59 (42.5)	46 (42.2)
Smoking status[*]				
None smoker	8 (4.1)	40 (18.8)	4 (2.9)	25 (22.9)
Ex-smoker	103 (53.1)	91 (42.7)	63 (45.3)	65 (59.6)
Current smoker	74 (38.1)	42 (19.7)	72 (51.8)	18 (16.5)
Unknown	9 (4.6)	40 (18.8)	0 (0.0)	1 (0.9)
Summary of:				
Smoking duration[b]				
mean ±sd	44.7±12.06	39.0±13.73	43.9±13.14	34.6±14.58
median	46	41	45	37
Smoking pack years[¶]				
mean ±sd	45.0±26.93	42.4±29.66	50.7±34.54	32.0±19.82
median	42.1	39.4	45	28
Cytology[*‡]				
Negative	113 (58.3)	213 (100.0)	76 (54.7)	108 (99.1)
Positive	67 (34.5)	0 (0.0)	46 (33.1)	0 (0.0)
Suspicious	14 (7.2)	0 (0.0)	17 (12.2)	1 (0.9)
Histology Diagnosis				
Others [a]	3 (1.6)		20 (14.4)	
Large cell carcinoma	25 (12.9)		16 (11.5)	
Small cell carcinoma	4 (2.0)		39 (28.1)	
Squamous cell carcinoma	91 (46.9)		31 (22.3)	
Adenocarcinoma	68 (35.0)		22 (15.8)	
Unknown	3 (1.6)		11 (7.9)	
Sample duration (yrs) ‡				
<5	75 (38.7)	96 (45.1)	10 (7.2)	39 (35.8)
5+	119 (61.3)	117 (54.9)	129 (92.8)	70 (64.2)

[↑] boarderline significant in training set

[*] Statistically significant in training set ($p<0.05$)

‡ Statistically significant in testing set ($p<0.05$)

¶ Statistically significant in testing set with p-value from Mann-Whitney test

[b] Statistically significant in both dataset with p-value from Mann-Whitney test

[a] Others (adenocarcinoid, adenosquamous, Carcinoid, Carcinoma, NOS, Neoplasm, malignant, Tumor cells, malignant, Basal cell carcinoma)

Table 2. Univariate association tests for the examined biomarkers in the training and validation bronchial washing sets.

Markers	Training Set					Validation Set				
	Positives		X^2 p-value	Model-based classification*		Positives		X^2 p-value	Prediction using trained univariate logit model*	
	Case n=194	Control n=213		Accuracy (%)	AUC (95% CI)	Case (n=139)	Control (n=109)		Accuracy (%)	AUC (95% CI)
TERT	130	35	$<10^{-4}$	75.7	0.75 (0.71, 0.79)	75	2	$<10^{-3}$	73.4	0.76 (0.72, 0.80)
RASSF1	75	7	$<10^{-4}$	69.0	0.68 (0.64, 0.71)	71	0	$<10^{-4}$	72.6	0.76 (0.71, 0.80)
WT1	70	10	$<10^{-4}$	67.1	0.66 (0.62, 0.69)	73	8	$<10^{-3}$	70.2	0.73 (0.68, 0.77)
p16	36	1	$<10^{-4}$	60.9	0.59 (0.56, 0.62)	18	0	$<10^{-4}$	51.2	0.57 (0.54, 0.59)
CYGB	36	16	$<10^{-3}$	57.3	0.56 (0.52, 0.59)	15	0	$<10^{-4}$	50.0	0.55 (0.53, 0.58)
RARb	28	10	10^{-3}	56.8	0.55 (0.52, 0.58)	67	18	$<10^{-4}$	63.7	0.66 (0.60, 0.71)
p73	30	17	0.08	53.8	0.52 (0.49, 0.55)					
DAPK	11	6	0.15	53.6	0.51 (0.50, 0.53)					
CDH13	30	43	0.22	52.3	0.52 (0.49, 0.56)					
TMEFF	14	14	0.80	52.3	0.50 (0.48, 0.53)					

* Disease class prediction based predicted $Pr(D) \geq 0.5$

Table 3: Evaluation of classification and predictive accuracies of discriminatory algorithms in training and testing dataset

Performance measure	Discriminatory algorithms			
	Top 6 Univariate	Best subset logit (AIC or BIC)	(BICq or CV)	MDR
Classification performance in training dataset				
Se/Sp (%)*	79.4/79.8	80.4/80.3	80.4/77.9	77.8/79.8
DA (%)*	79.6	80.3	79.1	78.9
AUC (95% CI)	0.85 (0.82, 0.89)	0.86 (0.83, 0.90)	0.85 (0.81, 0.88)	0.82 (0.78, 0.86)
Predictive performance in test dataset				
SE/Sp/sp (%)*	82.0/90.8	-	82.0/90.8	77.0/90.8
DA (%)*	85.9		85.9	83.1
AUC (95% CI)	0.90 (0.87, 0.94)		0.89 (0.85, 0.93)	0.85 (0.81, 0.89)

* Evaluated at probability of disease =0.5, Se: sensitivity, Sp: specificty, DA: discriminatory accuracy
AIC: Akaike Information Criterion, BICq: Bayesian Information Criterion, CV: Cross-validation

Table 4. Validation of the best subset logit model in the bronchial washings test set. Comparative efficiency of the models including DNA methylation (p16, RASSF1, WT1, TERT) only and DNA methylation with incorporated cytology versus cytology only.

		Cytology	Negative	Positive	Sensitivity	Specificity
Methylation Panel Model	Lung Cancer	Negative	19	57	75%	
		Positive	6	57	90%	
		Overall	25	114	**82%**	
	Controls	Negative	98	10		91%
		Positive	1	0		100%
		Overall	99	10		**91%**
Methylation Panel +Cytology model	Lung Cancer		25	114	**82%**	
	Controls		100	9		**92%**
Cytology only	Lung Cancer		76	63	**45%**	
	Controls		108	1		**99%**

Table 5: Overall performance of best discriminatory algorithm by epidemiologic and clinical characteristics in both training and validation sets.

Clinical characteristics	Number of specimens	% se	% (sp)	Chi-square test p- value
Diagnosis				
Lung Cancer	333	81.1		
Other (non-lung) cancer	36		(83.3)	
No malignancy	286		(82.1)	
Age				
<60	117	78.1	(89.5)	
60-79	498	81.1	(80.3)	0.81
80+	40	86.4	(77.8)	
Gender				
Male	372	81.3	(82.6)	
Female	335	80.9	(82.0)	0.77
Smoking status				
None	77	83.3	(86.2)	
Former	222	78.9	(83.3)	0.78
Current	206	82.9	(81.7)	
Specimen in storage (yrs)				
<5	220	84.7	(81.5)	
≥5	435	79.8	(82.9)	0.34
Lung cancer cases only				
Cytology				
Negative	189	74.1		
Positive	144	90.3		**<0.001**
Stage (pT)				
1	46	63.0		
2	91	84.6		
3	20	80.0		**0.018**
4	53	84.9		
Nodal status (pN)				
0	94	74.5		
1	35	85.7		
2	63	84.1		0.34
3	13	84.6		
Histology diagnosis				
Adenocarcinoma	92	75.0		
Squamous cell carcinoma	118	83.1		
Small cell carcinoma	41	100.0		**0.003**
Others*	82	75.6		

* Includes adenosquamous carcinomas (n=3), large cell carcinomas (n=2) carcinoids (n=5), lung carcinomas of non confirmed pathology (n=72)