1 General introduction
INTRODUCTION TO THIS THESIS

Lung cancer

Lung cancer imposes a major health care problem in many countries in the world. It is the most common solid cancer, with an estimated 1.6 million new cases annually worldwide (13% of total cancer diagnoses), and the leading cause of cancer death in western countries.\textsuperscript{1,2} In the Netherlands, 11,699 new cases of lung cancer were diagnosed in 2011, which reflects an increase of 30% since the beginning of the 21st century.\textsuperscript{3} A further increase is expected, mainly due to relative enlargement of the ageing population.

Smoking is the predominant cause of lung cancer.\textsuperscript{4} A lifetime smoker has a 20-30 fold increased risk of lung cancer compared to a lifetime non-smoker\textsuperscript{5}; of all smokers, 15% will ultimately develop lung cancer.\textsuperscript{6} Incidence numbers of lung cancer follow smoking trends in the population over time with a lag time of about 30 years. Tobacco consumption peaked by the middle of the 20th century, but declined at the end of the 20th century, which is now reflected in decreasing lung cancer incidence in males. However, as females started smoking several decades later than men, a 75% increase of lung cancer diagnoses among females was observed between 2000-2010.\textsuperscript{7,8} Quitting smoking results in risk reduction for lung cancer, but a former smoker will never reach baseline levels.\textsuperscript{9,10}

Although smoking is the major cause of lung cancer, in never-smokers (15% of lung cancers in males, 53% in females) lung cancer ranks the seventh cause of cancer death worldwide (before cervical cancer and prostate cancer).\textsuperscript{11} Gender and geographical variations may play a role as etiologic risk factor, as well as genetic predisposition, exposure to occupational carcinogens, and hormonal and environmental determinants.\textsuperscript{12,13}

Staging and treatment options

Lung cancer prognosis is defined by stage of disease at time of presentation (Table 1).\textsuperscript{14} Staging classification is based on the 7th edition of the TNM system\textsuperscript{15}: T is determined by tumour size/location (T1-4), N defines involvement of lymph nodes (N0-3), and M denotes spread of the tumour in lymph nodes beyond N3 station and/or distant metastases (M0,1a-b). More than 60% of patients present with metastatic disease, mainly due to lack of symptoms in early stage lung cancer. In addition, symptoms are often non-specific and attributed to other causes. Limited treatment options are available for metastatic disease, resulting in an overall poor 5-year survival
rate of 16%. When lung cancer is diagnosed in stage I/II, potential curative treatment can be offered, of which surgical radical resection or (stereotactic) radiotherapy are the cornerstones of treatment.

Lung cancer classification: histology and genomics

Lung cancer comprises a heterogeneous spectrum of disease, pathologically classified according to the World Health Organization (WHO) standards. Recently, a modification of lung cancer classification was published, with emphasis on histological diagnosis and molecular testing for driver mutations, relevant in clinical decision-making and choice of treatment (further discussed below).

Histologically, lung cancer is divided broadly into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). SCLC occurs less common (~15%), has a strong association with smoking (98%) and is characterized by an aggressive tumour behaviour. It responds initially well on chemotherapy, but has a high recurrence rate and overall has a similar dismal prognosis as NSCLC. NSCLC accounts for 85% of all lung cancers, and 85% is related to smoking. Before molecular characterization of tumours became clinically relevant, NSCLC subtypes were lumped, because of similar natural history and therapy response to cytotoxic chemotherapy. Major subtypes are adenocarcinoma (AdC) and squamous cell carcinoma (SqCC); large cell carcinoma and miscellaneous other types (such as neuroendocrine tumours) are less frequently diagnosed. The diagnosis ‘NSCLC not otherwise specified’ should be avoided, as since the beginning of this century, the distinction between AdC and SqCC with specific mutational changes in subgroups provide markers for prediction of therapy response.

AdCs are mostly peripherally located and arise from progenitor cells of the
bronchioles (Clara cells), alveoli (Type II pneumocytes) or mucin-producing cells. It is the most common histologic type in never-smokers. A proportion of AdCs harbour specific mutations in the DNA of the cancer cells, which may play a key role in personalization of treatment. In this context, identification of genetic aberrations in \( \text{EGFR} \) and \( \text{ALK} \) genes are relevant for daily practice. For example, treatment with \( \text{EGFR} \) tyrosine kinase inhibitors (TKIs) as compared to classical chemotherapy has shown to increase progression-free survival in patients with lung cancer characterized by activating \( \text{EGFR} \) mutations. \( \text{EGFR} \) mutations are (in particular) associated with terminal respiratory unit type AdC, women, never-smokers and Asians, and have shown to be mutually exclusive with various other mutations in primary lung cancer (e.g., \( \text{KRAS} \), \( \text{ROS1} \), \( \text{BRAF} \), \( \text{HER}-2 \), \( \text{RET} \)), though not with \( \text{PIK3CA} \) mutation.

Some SqCC variants were not taken into account in the WHO classification. Traditionally, SqCCs were thought to arise centrally in the lung-airways, originating from metaplastic changes of the respiratory epithelium. However, peripheral SqCC have been described. Whether peripheral SqCC arises from metaplastic changes during peripheral AdC development or not, remains to be established. A comprehensive study was performed by the Cancer Genome Atlas Research Network, investigating aberrations on genomic level in squamous cell carcinomas: \( \text{TP53} \) mutation was nearly universal and in 69% of tumours an altered \( \text{PI3K/RTK/RAS} \) signalling pathway was observed. In 7%, \( \text{EGFR} \) amplification was demonstrated. Promising therapeutic targets are \( \text{FGFR1} \) amplification and \( \text{DDR2} \) mutation. Interestingly, a molecular classification of lung cancer runs parallel with the conventional histologic classification.

**Lung cancer development**

During tumour development, cancer cells acquire several biological capabilities, as described by Weinberg and Hanahan: sustained cell proliferation, insensitivity to growth inhibitors, escape from apoptosis, limitless replication, angiogenesis, and tissue invasion and metastasis. In 2011, four more ‘hallmarks’ were added: deregulated metabolic pathways, evasion of the immune system, genome instability and mutation, and tumour-induced inflammation.

Lung cancer evolves, depending on the growth rate of the tumour cells and supporting stroma, over a period over approximately 10-30 years before it becomes clinically detectable. Chronic inflammation induced by exposure to an endogenous or exogenous agent results in the accumulation of molecular and morphological changes, leading to multicentric preinvasive lesions in the respiratory tract. This phenomenon
is referred to as field cancerization and is related to airway epithelium damage mainly caused by tobacco smoking.\textsuperscript{35,36} Lung cancer development in smokers is a chemical carcinogenesis due to multiple acquired changes in DNA of certain epithelial cells. There are at least 20 known lung cancer specific carcinogens in tobacco smoke; these and their metabolites bind to DNA, resulting in adduct formation and double strand breaks, while also the DNA repair system is affected.\textsuperscript{37} These may eventually give rise to gene mutations that predispose to lung cancer.

The identification of molecular alterations and their timing in lung carcinogenesis is a difficult task. Here, it is relevant to differentiate so-called driver mutations from passenger mutations to understand in what way the mutation contributes to the cancer process. Driver mutations are causally linked to (initiation of) cancer development, whereas passenger mutations reflect incidental mutation(s) during progression, possibly giving rise to higher chance for acquired resistance against therapy. In the clinical context, both types of mutations can be used as supportive diagnostic and possibly also prognostic markers; furthermore, driver mutations may serve as target for therapy.

Several research groups have constructed models proposing the sequence of molecular alterations in respiratory cells that give rise to AdC (Figure 1A) or SqCC (Figure 1B).\textsuperscript{38–42} As current knowledge on exact timing is still limited and alterations are often present in only part of tumours, the models are only indicative and not complete. Nevertheless, these may be used as basal framework for our understanding of the underlying molecular biology in (pre)malignant cells.

Molecular aberrations have been demonstrated in histologically normal cells in smokers that increases during successive precursor stages of lung cancer.\textsuperscript{36} Biologically significant events in lung carcinogenesis are, amongst others, mutations in genes $TP53$ and $KRAS$,\textsuperscript{34} which are explained in detail below. Furthermore, epigenetic modifications (see separate paragraph below) play a major role and have shown to occur throughout all developmental cancer stages.

The tumour-suppressor gene $TP53$ encodes the protein $p53$ (‘guardian of the genome’), which is a key regulator of the cell cycle in reaction to DNA damage and cell stress, by inducing cell cycle arrest, DNA repair, and if the latter fails, apoptosis. A deletion of or mutation in $TP53$ is a common (40-60\%) and possibly early event in lung cancers,\textsuperscript{44} and is associated with smoking.\textsuperscript{45} In sputum of lung cancer patients, but also of healthy controls, $TP53$ mutations have been detected.\textsuperscript{36,46,47}

$KRAS$ is part of the $RAS$ protein family and promotes cell growth in the $EGFR$
FIGURE 1. A simplified scheme, based on literature, showing histological and molecular changes in respiratory epithelium during development of A) adenocarcinoma, and B) squamous cell carcinoma. Histological images of hematoxylin and eosin stained paraffin samples of lung tissue (20x objective magnification) are displayed. Note that although the changes are shown as sequential in one direction, most preneoplastic morphologic changes are reversible and not all depicted molecular changes are obligatory for development of invasive cancer. Also, the list of markers is incomplete.
### B)

<table>
<thead>
<tr>
<th>Early</th>
<th>Intermediate</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal lung</td>
<td>Step 1 Hyperplasia</td>
<td>Step 2 Metaplasia and Dysplasia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Step 4 Invasive Carcinoma</td>
</tr>
</tbody>
</table>

- Small telomeric deletions
- Telomerase dysregulation
- Telomerase reactivation
- LOH: 3p
- 9p21 (p16)
- 17p
- 8p21-23
- p53
- 3p
- 5q21-22
- 13q14

**Microsatellite alterations**

- MYC overexpression
- Loss of FHit
- VEGF overexpression
- Cyclines D1 and E overexpression
- Bcl-2 overexpression

- MAGE A1, A3, B2
- MAGE, BAGE, LAGE, CT17, SSX

**Neoangiogenesis**

**Aneuploidy**

- Mutation: KRAS
- TP53
When mutated, KRAS becomes constitutionally active. KRAS mutations are most prevalent in AdCs (20–30% in western countries and 10% in eastern countries) and are mutually exclusive with EGFR and ALK mutations.24,48 KRAS mutation may be detected in sputum one year before lung cancer diagnosis.49 In general, KRAS mutation is not detected in cancer-free subjects,36,46 which implies that KRAS mutation testing is suitable for diagnostic purposes.

DNA hypermethylation

Epigenetics refers to functionally relevant changes in the genome that are not caused by changes in the DNA sequence, but are encoded by modifications of DNA chromatin components. One of these mechanisms involves gene methylation, which consists of the addition of a methylgroup (-CH3) to the 5-carbon position of cytosine within a CpG (cytosine-guanine) dinucleotide, mediated by DNA methyltransferases (DNMTs). Methylation is a physiological process within cells, essential for normal development and regulation of gene expression to maintain genomic stability, for example in X-chromosome inactivation.50,51
In cancer cells, DNA hypermethylation is observed in regions described as CpG islands, which are frequently located in promoter regions of genes (Figure 2). Methylation affects the binding of transcription factors. When a promoter becomes hypermethylated, transcription factors are unable to bind to the promoter and thereby mRNA expression is prevented. This may result in transcriptional silencing of genes encoding proteins that play a role in inhibiting both cell growth and tumour formation, so-called tumour suppressor genes.

Paradoxically, the overall level of gene methylation can be reduced in cancer cells compared to normal cells in parallel with hypermethylation in CpG islands. A global hypomethylation is generally observed in the ‘body’ (coding regions and introns) of genes and repetitive DNA sequences. This may result in genetic instability and the activation of (tumour-promoting) genes that are thereby not expressed normally. DNA hypomethylation is involved in benign to malignant tumour transformation; this topic lies beyond the scope of this thesis. Where the term ‘methylation’ is used, it refers to DNA hypermethylation.

Smoking is associated with aberrant DNA hypermethylation of certain genes. The tobacco-specific carcinogen NNK indirectly prevents DNMT1 degradation and induces nuclear accumulation of DNMT1, resulting in DNA promoter hypermethylation of tumour-suppressor genes. Also, double strand break repair capacity has shown to be reduced in smokers with a high methylation index compared to smokers with no genes methylated. An imbalance in the DNA repair process may result in genomic instability. Genomic instability has shown to be a pivotal event in progressive squamous metaplasia. Even after smoking cessation, methylation may persist, which is illustrated by similar prevalences of p16 hypermethylation in sputum of current versus former smokers.

Aberrant DNA methylation takes place early during lung carcinogenesis. The degree of hypermethylation increases during progression of lung cancer development. Thus, in cancer, hypermethylation markers may be useful for both (early) detection and prognosis, monitoring of disease during and after treatment, and could serve as targets for epigenetic therapy.

Lung cancer diagnosis

When a patient presents with symptoms, the diagnostic workup to establish lung cancer diagnosis involves different steps. Imaging techniques (chest X-ray (CXR), computed tomography (CT), positron emission scanning) are required to establish tumour location and assess tumour burden (staging, cTNM) and provides information
on where to obtain tissue/cells for (molecular) pathological examination. Importantly, radiological imaging is non-specific for lung cancer. For confirmation of lung cancer diagnosis, additional invasive procedures as bronchoscopy with or without ultrasound/CT guidance are performed to procure tumour material for histological/cytological/molecular diagnosis. Limitations of these techniques are that not all patients are physically able to undergo these procedures and pulmonary AdCs are difficult to reach with bronchoscopy, as these are usually located in the periphery of the lung. In this case, other procedures such as transthoracic needle biopsy, mediastinoscopy, video-assisted thoracoscopic surgery (VATS), or even thoracotomy aid in procurement of tumour tissue.

Less or non-invasive tools to establish lung cancer diagnosis and/or to examine mutational status of the tumour may be the use of bronchoalveolar lavage and sputum. Sputum is an easily accessible biological specimen that can be obtained by simple and inexpensive means, and consists of a mixture of saliva and epithelial lining fluid. The thin layer on the surface epithelium of the respiratory tract has a volume of approximately 2 ml, and is composed of mucus secreted by goblet cells and proteins secreted by Clara cells and other cells. The lining fluid is continuously moved by the ciliary system in proximal direction towards the larynx, and swallowed into the digestive tract. The airways are thus protected by the removal of inhaled and local debris. In general, this process occurs unnoticed. In the majority of smokers, hyperplasia of goblet cells in the respiratory epithelium leads to excess production of sputum. In order to prevent airway blockage, sputum is often expectorated. This is illustrated by the typical smoker’s morning cough, to get rid of sputum that has accumulated during the night. Spontaneous sputum is usually easily obtained from smokers. For former and non-smokers it is more difficult to produce sputum, although this issue can be overcome by sputum induction. When the cells are collected in Saccomanno’s fixative, the sputum can be stored at room temperature for years.

Sputum can be used for lung cancer diagnosis (reviewed in ). Sputum cytology reveals inflammatory and epithelial cells mostly from oropharynx. In lung cancer, less than 1% of sputum cells are exfoliated tumour cells. Sputum cytology has limited value for lung cancer diagnosis with a pooled sensitivity rate of 66% (95% confidence interval 42-97%), though with high specificity (pooled rate 99%, 95%CI: 68-100%) and is currently not part of routine diagnostics in western countries. It is most commonly successful in central endobronchial location of the tumour, SqCC, number of sputum samples collected per patient, large tumour size and/or advanced stage of tumour. Besides proteins, sputum may also contain fragments of tumour and inflammatory cell
Derived DNA. Interestingly, molecular tumour aberrations may occur in sputum before morphological changes are observed by cytological analysis, and can be detected using advanced molecular techniques with high sensitivity. DNA mutations, DNA hypermethylation and microsatellite aberrations have been detected in sputum. Prevalence for gene hypermethylation in sputum is higher than in serum.

**Screening**

Ideally, lung cancer is detected before metastases are present and curative treatment is feasible. The detection of lung cancer at an early stage, where treatment may be curative, may reduce lung cancer mortality. The latency period before lung cancer becomes clinically evident offers an opportunity to screen high risk individuals for the presence and treatment of asymptomatic lung cancer. Therefore, a major research effort is directed to development of screening tools.

In 1968, the WHO presented guidelines in which conditions for a screening test for a certain disease were defined. In short, the test should be able to identify unrecognized disease in asymptomatic individuals who are at risk of having the disease, can be systematically applied, is patient friendly, cost-effective and should lead to a reduced mortality. In the 1970s, several large randomized controlled trials for lung cancer screening were performed in heavy smokers, starting with CXR screening, later in combination with sputum cytology. The additive diagnostic value of sputum cytology was examined in a combined analysis of the Johns Hopkins Lung Project and Memorial Sloan Kettering Study (dual screen versus CXR only) with nine years of follow-up, revealing a slightly lower lung cancer mortality rate in the dual screen arm (~10%; not statistically significant) in particular in SqCCs. A recent Cochrane review concluded that no gain in patient mortality was observed. Thus, there is no future role for chest X-ray and sputum cytology in the screening setting.

In this century, higher analytical sensitivity was demonstrated with the usage of low dose spiral CT (LDCT) (nodule detection >2 mm; suspicious of malignancy ≥5 mm) by diverse single arm studies and randomized controlled trials. The National Lung Screening Trial (NLST) in the United States reported a relative mortality reduction of 20% for high risk individuals who were screened with LDCT, as compared to those who were screened by CXR. However, their methodology resulted in i) a high false positivity rate, because many small benign nodules were detected, ii) questionable cost-efficiency and iii) potential harmful radiation exposure, leaving room for improvement. The final outcomes on lung cancer mortality of on-going LDCT randomized controlled screening trials in Europe (NELSON, DLCST) are awaited, in which the control groups receive...
no screening.\textsuperscript{87,88} Apart from different entry criteria and duration of interval between screening rounds (NELSON), in these studies also a different CT-detected pulmonary nodule management (based on volumetry) is proposed as employed in the NLST, which will lead to difference in positive predictive values between the trials.\textsuperscript{89}

OUTLINE OF THE THESIS

The above indicates that a non-invasive tool for complementary use to radio-imaging to confirm lung cancer diagnosis and/or to examine mutational status of the tumour is currently lacking. In addition, screening may be improved as it currently suffers from low specificity. Sputum may comprise such non-invasively collected specimen and the detection of (epi)genetic alterations associated with lung cancer (development), such as DNA hypermethylation analysis. Sputum may be of interest for detection and/or screening of lung cancer. From technical point of view, DNA is easily extracted from sputum. DNA modification by bisulphite converts unmethylated cytosine into uracil, whereas methylated cytosine remains unaffected, thereby enabling the differentiation of methylated and unmethylated DNA presence in a sample. Targeted primers and probes can be devised to detect hypermethylation in specific regions of genes. Also, it is possible to combine complementary biomarkers, as it is unlikely that one biomarker is able to detect all cancers, due to the heterogeneous nature of lung cancer.

The main focus of this thesis is to examine the value of the molecular biomarkers in sputum for the diagnosis of lung cancer.

Chapter 2 first addresses the current state of sputum analysis for lung cancer diagnosis. It provides a review of all studies on sputum examination in the past decennium, with special attention to DNA hypermethylation markers, and a particular focus on the possibility to distinguish diagnostic biomarkers from risk biomarkers.

Chapter 3 evaluates DNA hypermethylation status of selected biomarkers \textit{RASSF1A}, \textit{APC} and cytoglobin (\textit{CYGB}) in sputum samples from symptomatic lung cancer patients and controls, in comparison with sputum cytology. In addition to the use of conventional statistical analysis methods to assess biomarker performance, we design a novel analytical method, based on the capability of the biomarker to act as either diagnostic marker or risk marker by constructing a risk classification model with post-test probabilities.
Chapter 4 describes the outcome of a study which examined whether the duration of sputum sampling influences the detection rate of DNA hypermethylation in sputum. This is evaluated using the biomarkers RASSF1A, APC and CYGB in a subset of lung cancer patients and controls, who all participated in the study described in Chapter 5.

Chapter 5 addresses the external validation of previous mentioned biomarkers with additional discovered biomarkers 3OST2, PRDM14, PHACTR3 and FAM19A4 for the detection of lung cancer in sputum samples in a learning and validation set. The study was performed among symptomatic lung cancer patients and controls. Furthermore, the risk classification model as introduced in Chapter 3 was appraised.

To improve sensitivity and specificity of the molecular sputum test for lung cancer diagnosis, we hypothesize that another diagnostic test using a different approach may act complementary to the molecular sputum test. In Chapter 6, the effect of combining the molecular sputum test with exhaled breath analysis in a cohort of lung cancer patients and controls is examined.

Currently, personalized treatment with EGFR TKIs is preferred above conventional chemotherapy in patients with a confirmed activating EGFR mutation in their primary tumour tissue. In Chapter 7, the feasibility of EGFR mutation analysis in sputum of lung cancer patients is investigated. The sensitivity of different molecular assays is examined in relation to the mutation status of the tumour tissue.

Chapter 8 examines the molecular sputum test in a cohort of asymptomatic high risk individuals aiming to assess its performance for screening purposes. Towards this goal, sputum of subjects (suspicious of) having lung cancer and controls who participated in the Dutch-Belgian NELSON study, a randomized controlled trial for screening with LDCT, was used.

Finally, in Chapter 9, we put the data from all chapters into perspective and describe in more detail our findings in relation to possible implementation in lung cancer diagnosis and/or screening.
REFERENCES

3. IKNL. Nederlandse Kankerregistratie [Internet]. [cited 2013 Sep 22];Available from: http://cijfersoverkanker.nl/
22. Gridelli C, Peters S, Sgambato A, Casaluce F, Adjei AA, Ciardiello F. ALK inhibitors in the treatment of...
General Introduction


