9

Summary and general discussion
BACKGROUND OF THE THESIS

At time of initiation of the studies described in this thesis (2009), non-invasive methods to diagnose lung cancer were inadequate (invasive techniques were needed for pathological assessment) and no practical screening tool existed to identify early stage lung cancer, hence establishing an unmet need for a novel test for lung cancer detection.

The work presented in this thesis evaluated potential molecular biomarkers in sputum for the diagnosis of lung cancer. In this chapter, the main findings of these studies are summarized and critically discussed, and suggestions for future research are provided.

SUMMARY OF FINDINGS

Chapter 2 provides a systematic overview of all studies that were published between 2003-2013 on sputum examination for lung cancer diagnosis. A rationale is presented for the distinction between a ‘diagnostic marker’ and ‘risk marker’. A positive ‘diagnostic marker’ denotes the presence of disease, whereas a positive ‘risk marker’ reflects the increased probability of acquiring the disease. If a particular marker shows a positive test result in a larger proportion of the control population than can be explained by maximized true positive assumption, this marker is considered to be a ‘risk marker’. The specificity of the marker hence determines its clinical application.

Most research that has been conducted to investigate biomarkers in sputum involved small study populations. In the literature, none of the investigated biomarkers yields an individual sensitivity of 100%. Furthermore, the majority of the proposed markers do not show diagnostic capacity, when considering specificity levels. A few studies constructed panels of biomarkers that were complementary to each other, but these have not been validated in independent study cohorts.

A meta-analysis of studies examining DNA hypermethylation in matched sputum and tumour samples revealed a concordance in approximately 78% of the cases.

In Chapters 3-5, DNA hypermethylation biomarkers were studied in sputum of symptomatic lung cancer patients and controls, and compared to sputum cytology as reference test. According to the Early Detection Research Network (EDRN) guidelines, these studies involved phase II trials: the ability of biomarkers to distinguish lung
cancer patients from controls was examined and thresholds for a positive (diagnostic) test were defined.

In Chapter 3, DNA hypermethylation status of biomarkers \textit{RASSF1A}, \textit{APC} and cytoglobin (\textit{CYGB}) was examined in two independent sets (i.e., Set 1 and Set 2) of sputum samples from symptomatic lung cancer patients and controls, collected in Nijmegen. In addition to the use of conventional statistical analysis methods to assess biomarker performance, we constructed and evaluated a risk classification model with post-test probabilities, based on the capability of the biomarker to act as either ‘diagnostic marker’ or ‘risk marker’.

\textit{RASSF1A} demonstrated the best discriminatory capacity between lung cancer patients and controls with consistent results in both sets. Regarding individual specificity levels, \textit{RASSF1A} was considered as ‘diagnostic marker’ (specificity ≥96%), and \textit{APC} and \textit{CYGB} as ‘risk markers’. The panel of \textit{RASSF1A}, \textit{APC} and \textit{CYGB} showed a sensitivity of 63% with specificity of 78% for lung cancer diagnosis in Set 1. In the risk model, 36% of lung cancer patients were defined as high risk (>60% chance on lung cancer) with two false-positives in Set 1. The model was reproducible in Set 2. Overall cytological analysis was positive in 22% of cases; when combined with \textit{RASSF1A} hypermethylation, sensitivity increased to 52% with similar specificity of 94%.

In a methodological study, we demonstrated in Chapter 4 that prolonged sputum sampling (6-9 days) improves sensitivity of DNA hypermethylation biomarkers for lung cancer diagnosis. This study was performed in a subset of lung cancer patients and controls who participated in the study described in Chapter 5. DNA hypermethylation of biomarkers \textit{RASSF1A}, \textit{APC} and \textit{CYGB} was assessed. The results of the markers combined showed positive hypermethylation in 43% of canisters I (i.e., day 1-3) with a corresponding specificity of 96%. Cumulative 6- and 9-day methylation frequencies showed a gain in sensitivity for lung cancer diagnosis (53% and 64%, respectively) at the expense of small loss in specificity (94% and 91%, respectively). An additional 37% of lung cancer patients (who had tested negative in the first canister) were detected in either canister II or III. For comparison, triplicate testing of a single canister showed high concordance and no significant differences.

Chapter 5 describes the results of independent validation of the above mentioned biomarkers with additional novel biomarkers \textit{3OST2}, \textit{PRDM14}, \textit{PHACTR3} and \textit{FAM19A4} in prospectively collected sputum samples of symptomatic lung cancer patients and
controls. Eight hospitals in the Amsterdam and Utrecht regions participated. The risk classification model as introduced in Chapter 3 was evaluated. A learning set (73 cases, 86 controls) and validation set (159 cases, 154 controls) were constructed from the sputum bank. RASSF1A showed the best diagnostic performance in both sets, further substantiating RASSF1A hypermethylation as ‘diagnostic marker’. Two panels were constructed consisting of either RASSF1A, 3OST2 and PRDM14 (sensitivity 82%, specificity 66%; based on Youden’s J index), or RASSF1A, 3OST2 and PHACTR3 (sensitivity 67%, specificity 90%; based on high specificity). The risk model incorporated RASSF1A (high risk classification), 3OST2 and PRDM14 (lower risk classifications). In the learning set, 40% of lung cancer patients were classified as high risk, with 2.3% false-positivity. These results were found to be reproducible in the validation set. Furthermore, prolonged sputum testing was re-evaluated in this cohort with hypermethylation testing of either RASSF1A, 3OST2 or PRDM14. An additional 47% of lung cancer patients (who tested negative in canister I) was detected in canisters II or III, substantiating results obtained in Chapter 4. Molecular sputum analysis proved to be superior over cytology (sensitivity 14%; P<0.001).

In Chapter 6, we reported on the results of combining the molecular sputum test with exhaled breath analysis for lung cancer diagnosis in a small cohort of lung cancer patients and controls. eNose was able to detect 80% of lung cancer patients, with a specificity of 48%. In a second group of ‘eNose only’ samples, sensitivity was similar with lower specificity (13%), though not significant, indicating reproducibility. Multivariate logistic regression (based on Youden’s J index, as defined in Chapter 5) included RASSF1A and 3OST2, and not eNose. However, sensitivity for lung cancer diagnosis increased to 100% when RASSF1A was combined with eNose (specificity 42%). Of interest, the additional complementary diagnostic value of RASSF1A hypermethylation to eNose and vice versa was significant.

In Chapter 7, we demonstrated that EGFR mutation analysis is feasible in sputum of lung cancer patients. DNA of a similar cohort of subjects and serial dilutions of EGFR mutated cell lines were examined using different molecular techniques in a multicenter collaboration. All techniques were capable of detecting EGFR mutation in sputum (relative sensitivity 30-50% compared to tissue testing), with Pangaea Biotech SL Technology (PST) showing the best performance. PST also had highest analytical sensitivity (detection of 0.5% mutated DNA). None of the COPD controls and lung cancer patients without EGFR mutation-positive tumours were tested positive in sputum,
indicating high specificity. Chance of positive test results was higher in (related) sputum samples in case of positive sputum cytology.

Finally, in a phase III study, we investigated DNA hypermethylation biomarkers in sputum of a cohort of asymptomatic individuals at high risk for lung cancer in Chapter 8. Participants were enrolled in the NELSON trial (Nederlands Leuvens Longkanker Screenings Onderzoek) and collected sputum at baseline. Lung cancer patients were identified at follow-up. DNA hypermethylation of RASSF1A, APC, CYGB, 3OST2, FAM19A4, PHACTR3 and PRDM14 was assessed in sputum samples. All biomarkers showed moderate sensitivity, of which APC was able to make a significant distinction for the whole group between individuals with LDCT detected lung cancer and cancer-free controls. However, for APC, specificity was low. RASSF1A hypermethylation in cases (sensitivity 11%, specificity 92%) was most likely to occur within two years prior to diagnosis (5 of 6 cases; \( P = 0.03 \)). The risk model introduced in Chapters 3 and 5 showed to be insufficient in predicting lung cancer in subjects at risk. Sputum cytology did not detect any lung cancers.

GENERAL DISCUSSION OF STUDY RESULTS

Study design

The development of a biomarker for use in a clinical and/or screening setting is a stepwise process, involving biomarker discovery, assessment of biomarker accuracy and validation in clinical and screening context. The studies in this thesis were constructed according to the guidelines of the Early Detection Research Network (EDRN) and the ‘prospective-specimen-collection, retrospective-blinded-evaluation (ProBE)’ design (Chapter 5).\(^1,2\) In addition, studies were reported as recommended in the Statement for Reporting Studies of Diagnostic Accuracy (STARD).\(^3\) These guidelines pursue optimal quality of diagnostic accuracy studies. The EDRN guidelines are a lead for studies to facilitate biomarker development for early detection of cancer and consist of five consecutive phases (Appendix A).\(^1\) The concept of the PRoBE design provides a solid base for particularly phase II studies.\(^2\) Strong points of this study design are the prospective sample collection, representativeness of the study population for the target population, the use of standardized protocols, training-validation set approach with randomized selection of subjects, generalizability of test results through multicenter involvement, and blinding of investigators for test outcome.
EDRN recommends to define minimally acceptable ‘true positive rate (TPR)’ and ‘false positive rate (FPR)’ prior to (evaluation of data of) a phase II study. This requires the dichotomization of test results in order to enable a binary positive or negative test result and depends on the threshold of the test (Appendix B). Once a threshold is chosen, this should be tested in an independent study set.

Published studies on hypermethylation of biomarkers in sputum consist mainly of phase II studies. Apart from one study, the selected markers have not been further evaluated in samples from asymptomatic high risk individuals (phase III). In these studies, usually sensitivity and specificity estimates are calculated. However, one may doubt whether this adequately reflects the diagnostic capability of a biomarker. A disadvantage of dichotomization is the loss of additional information. From this point of view, a classification model in which test results are stratified in different categories may yield more information regarding the risk status for the subject (i.e., chance on lung cancer development).

Analysis of biomarker performance

A general perspective of biomarker evaluation is explained in appendix B. For our studies, we formulated three statistical approaches for biomarker analysis (Figure 1):

A) The first method, the Youden's J statistic, is a commonly used measure for maximal diagnostic effectiveness of a biomarker to discriminate between cases and controls. Youden's J reflects the difference between the true positive rate and the false positive rate (\(J = \text{sensitivity} + \text{specificity} - 1\)). On the receiver operating characteristic (ROC) curve, this point is the farthest situated from the line of equality (diagonal line). J lies between 0 and 1, with 1 indicating that complete discrimination is possible between cases and controls (maximum potential effectiveness), and 0 when values are completely overlapping.

B) The second approach takes into account the diagnostic capability of a biomarker. By setting a threshold for each marker at high specificity, the number of false-negative test results is kept low. PPV will be high in case of a high prevalence setting (e.g., diagnostic assessment in symptomatic patients in the hospital). Based on the rationale introduced in Chapter 2, a cut-off value is established that yields a specificity of at least 96% (assuming that a maximum of 4% positive test results in the population under study is related to undetected lung cancer). As a consequence, sensitivity is low. However, in combination with complementary ‘diagnostic markers’, a diagnostic panel with sufficiently high sensitivity may possibly be constructed. Looking at the down side of the
'diagnostic markers', false positivity may occur in practice. In our definition this is limited to 4%. A hypothetical combination of 2 or 3 ‘diagnostic markers’ should result in a minimum specificity of 92 and 88%, respectively. This was in part supported by our findings in Chapter 5, the panel of three ‘diagnostic markers’ RASSF1A, 3OST2 and PHACTR3 showed a specificity of 89.5% in the learning set, but 80.5% in the validation set.

C) The risk classification model consists of categories of increasing post-test probabilities (0-20%, 20-40%, 40-60%, ≥60%) for lung cancer diagnosis, with different cut-off values per category. ‘Diagnostic markers’ can be used to diagnose lung cancer in high risk individuals, whereas ‘risk markers’ designate subjects in lower risk categories and may be indicative of lung cancer development in some of the test positives. In these cases, follow-up with monitoring may be required. This analysis differs from the first two approaches as the markers are now included as continuous variables. The model may provide directions for a clinical management strategy, similar to the BCRAT model in breast cancer.
Most studies reported in the literature examined the presence of methylation by qualitative analysis with electrophoresis only. In contrast, quantitative analysis enables normalization to an internal control, yielding ratio values that provide information about the relative frequency of methylation in a sample. The definition of ‘hypermethylation’ then depends on the threshold. By setting a high threshold with high specificity, the possible diagnostic value of a marker is determined. Hence, markers that have been defined as a ‘risk marker’ based on qualitative analysis, may still have diagnostic value, if a high threshold is chosen in quantitative analysis. A re-examination of those markers in this setting may be useful in future research.

DNA hypermethylation in sputum: technical approach

In our studies, we investigated DNA hypermethylation status of gene promoter regions. This analysis has several advantages, but also limitations, which will both be discussed below.

**Issue: Marker selection**

In the search for a putative biomarker, DNA hypermethylation biomarkers are appealing targets for lung cancer assessment. DNA methylation alterations are usually localized at CpG islands in promoter regions of genes, which are relatively small parts of genes. DNA hypermethylation allows for simple design of probes and primers that either bind to methylated or unmethylated DNA to exert selective DNA amplification in quantitative methylation specific PCR (qMSP). The choice for the markers investigated in this thesis was based on available literature and in-house development.

**Issue: DNA stability**

Genomic DNA is readily isolated from sputum after mucus-dissolving treatment with dithiothreitol (DTT). Directly upon arrival, sputum was DTT-treated and cell pellets were stored in Saccomanno’s fixative. In Chapter 3, we report that DNA hypermethylation can be reliably examined in sputum that has been stored for more than 5 years. Leng and coworkers reported that storage of untreated sputum in Saccomanno’s fixative over 3 years could lead to DNA degradation. Therefore, DTT-treatment, as was prospectively performed in our studies, within a short period after collection apparently preserves DNA integrity during long-term storage.

The subsequent step involves modification of the DNA. During this process, unmethylated cytosine is converted to uracil, whereas methylated cytosine remains unaffected. Although this procedure is necessary to obtain templates for qMSP that
enable the distinction between methylated and unmethylated DNA, 75-90% of the DNA is degraded by the bisulfite treatment (and unpublished data). This degradation needs to be taken into account in samples with low concentration of hypermethylated tumour DNA relative to non-hypermethylated normal DNA. To examine the magnitude of the degradative effect, the methylation negative samples of Chapter 5 were re-examined for RASSF1A hypermethylation with a twofold increase in DNA input, revealing only 1-2% gain in sensitivity. We also observed that individual pooling of genomic DNA of the three canisters (instead of testing separately), followed by DNA modification and qMSP, resulted in decreased sensitivity (from 71% to 38%; data unpublished), while specificity remained the same. Our explanation is that possible dilution of the tumour DNA results in a concentration below detection limit of the qMSP assay. These observations imply that DNA hypermethylation basically appears to be present in low concentrations in sputum, near the limit of detection (LOD; analytical sensitivity) of the assay.

**Issue: false-positive test results**

Another concern may be incomplete bisulfite conversion, leading to false-positive methylation results.\(^9\) Chance on this phenomenon decreases with standardized bisulfite treatment protocols and design of primer pairs that comprise several cytosine nucleotides in their sequence (at least 30%).\(^10\)

In general, qMSP is a robust and simple method. In multiplex qMSP, different primers can be used simultaneously to assess methylation status of more than one gene.\(^11\) However, caution is needed in the number of PCR cycles. In a small interlaboratory study, described in Chapter 2, we showed that consistent results were present up to 55 PCR cycles. Importantly, false-positive outcome was more likely when more than 55 PCR cycles were used. This observation was supported by a meta-analysis of studies that investigated p16 hypermethylation in sputum.\(^7,12–24\) In Chapter 4, we showed that qMSP results were reproducible with duplicate and even triplicate testing (45 amplification cycles). It is thus necessary to limit qMSP to maximally 55 PCR cycles to minimize false positivity.

**Issue: sputum representativeness**

In a meta-analysis, we observed concordance in 78% of matched tumour and sputum samples (Chapter 2), indicating that DNA hypermethylation in sputum is frequently representative for hypermethylation in the primary tumour and that sputum can be used as surrogate biological specimen for tumour methylation. However, this also
implies that sputum samples in the missing fraction (~20%) are not representative.

All sputum samples were cytologically scored for presence of alveolar macrophages and bronchial epithelial cells. In the studies of Chapter 3 and 5, these were present in 86% and 96% of samples, respectively, indicating adequate sputum quality. No association was observed with DNA hypermethylation, which suggests that presence of alveolar macrophages or bronchial epithelial cells is not obligatory to ensure that sputum is suitable for hypermethylation analysis.

DNA hypermethylation in sputum of symptomatic lung cancer patients

Our studies showed that tumour-related DNA characteristics were present in sputum. In studies conducted both in Nijmegen and Amsterdam, RASSF1A hypermethylation demonstrated to be the best (diagnostic) marker (Table 1). The high specificity of this marker in symptomatic lung cancer patients is in line with previous research\textsuperscript{16,25–27} and indicates that RASSF1A is a strong predictor for lung cancer with high positive predictive value. We could not provide an explanation for the lower specificity of RASSF1A in the validation set in Chapter 5, except for chance, since the selection process for learning and validation set was random. Nonetheless, RASSF1A may facilitate lung cancer diagnosis and treatment. For example, in patients with a CT-visible tumour in whom invasive diagnostic procedures are not conclusive, the presence of RASSF1A hypermethylation may be decisive for a malignant over a benign tumour.

The remaining markers APC, CYGB, 3OST2, PRDM14, PHACTR3 and FAM19A4 were considered to be 'risk markers', based on lower specificity in the control population. None of the biomarkers showed high sensitivity without compromising specificity. Noteworthy is that no single molecular alteration can be expected to be present in every tumour cell, due to the heterogeneous nature of lung cancer.\textsuperscript{28} DNA hypermethylation of a given biomarker would likely allow detection of a proportion of lung cancers. Therefore, assembly of a complementary panel of biomarkers may increase sensitivity. In Chapter 3, a panel was created consisting of RASSF1A, APC and CYGB (based on Youden's J index). This combination seemed rewarding (sensitivity 63%, specificity 78%), but the panel did not show consistent results in the second set, thus reducing reliability of this combination. In Chapter 5, multivariate analysis yielded a combination rule with higher sensitivity, consisting of RASSF1A, 3OST2 and PRDM14 (sensitivity 82%, specificity 66%). When the biomarkers were included as ‘diagnostic markers’, a panel was constructed consisting of RASSF1A, 3OST2 and PHACTR3 with sensitivity of 67% and specificity of 90%. For clinical use, the latter approach limits false-positivity rate and deserves further exploration. Overall, it remains a challenge to construct a panel
Summary and general discussion

In Chapter 3, a risk model based on different post-test probabilities was designed on Set 1 and subsequently evaluated in Set 2. Of note, if high risk was defined as >90% chance on lung cancer, 33% of lung cancer patients were correctly identified with one false-positive. As more lung cancer patients were detected with a lower threshold and number of additional false-positives was relatively small, we chose to define high risk as >60% chance on lung cancer, as >90% and >60% were within this data set only minimally different. Interestingly, the risk model showed to be statistically reproducible in Set 2. However, the model had moderate capability to discriminate lung cancer patients from controls, which implicates that this combination of markers was not useful for lung cancer diagnosis. In Chapter 5, the risk model consisted of RASSF1A, 3OST2 and PRDM14. The model was more accurate when compared to the model in Chapter 3, but still many lung cancers (30%) remained undetected or were classified as ‘low risk’. We concluded that the concept of the risk model is appealing, but further refinement with stronger (diagnostic) markers is needed.

Sputum cytology in symptomatic patients

Our data revealed improved sensitivity and specificity of hypermethylation markers for lung cancer diagnosis in sputum compared to cytological examination alone. However, in the clinical setting, the combination of cytology and RASSF1A hypermethylation analysis may still be useful to detect primary lung cancer in symptomatic subjects. This may be of benefit especially in low-budget countries, where sputum cytology is still part of diagnostic routine and invasive procedures are too expensive or not available.

An explanation for the lower sensitivity of sputum cytology in Chapter 5 as compared to Chapter 3 is that in Nijmegen only symptomatic lung cancer patients were

| TABLE 1. Sensitivity and specificity of RASSF1A hypermethylation in symptomatic lung cancer patients and controls in two study cohorts. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| RASSF1A Sensitivity (%) | 95% CI          | Specificity (%) | 95% CI          | P-value         |
| CWZ                      |                 |                 |                 |                 |
| Set 1                    | 41              | 31 - 51         | 96              | 89 - 99         | <0.001          |
| Set 2                    | 52              | 38 - 65         | 94              | 91 - 96         | <0.001          |
| VUmc                     |                 |                 |                 |                 |
| Learning set             | 43              | 31 - 55         | 97              | 90 - 99         | <0.001          |
| Validation set           | 37              | 29 - 45         | 88              | 82 - 93         | <0.001          |

with high sensitivity while maintaining specificity.
included in whom a primary tumour was detected, whereas in Amsterdam also lung cancer patients with tumour recurrence were enrolled. Recurrent tumours are usually detected at smaller size when compared to primary lung cancers (due to intensive clinical follow-up of lung cancer patients). The likelihood of positive sputum cytology is reduced in smaller tumours.\textsuperscript{30}

**Prolonged sputum sampling**

As tumour cells only make up <1\% of sputum composition, they may not be present in some sputum samples. Sputum sampling during more than three consecutive days yielded an increase in sensitivity for lung cancer diagnosis. The observed gain was not related to repetitive testing. It seems worthwhile to test multiple canisters, to avoid missed diagnoses if only the first canister is examined. This is in consonance with previous studies evaluating sputum cytology in sputum samples that were sequentially collected over several days.\textsuperscript{30,31}

One older crossover study suggested a learning effect in time, resulting in higher cytological representativeness of sputum (i.e., increased yield of airway-derived cells) after collecting for a longer period.\textsuperscript{32} We could not confirm this, as sensitivity rates of methylation markers for lung cancer did not differ significantly between the three canisters. This may be explained by detailed patient instructions at enrollment in our study, ensuring adequate sputum expectoration bypassing the learning effect. This is also supported by the high frequency of alveolar macrophages and bronchial epithelial cells in our sputum samples, confirming that the sputum is indeed derived from the respiratory tract.

Interestingly, a higher increase in sensitivity was observed for ‘risk markers’ relative to the ‘diagnostic marker’ RASSF1A. But, this is (by definition) at the expense of a greater loss in specificity for the ‘risk markers’ as compared to the ‘diagnostic marker’ (Chapter 5). To minimize loss of specificity, it may be considered to perform only additional testing of ‘diagnostic markers’, although the actual gain in sensitivity for RASSF1A hypermethylation per extra canister in this study is not high: 7.4\%.

Overall, prolonged sputum sampling may be a means to increase representativeness of sputum samples and hence improve concordance of hypermethylation between tumour tissue and sputum samples. However, if the addition of a second marker would lead to \(\geq 7.4\%\) increase in sensitivity, addition of one (or more) markers may be preferred over prolonged sampling, as sampling during three days instead of nine days would be less demanding for both patient and clinician.
DNA hypermethylation in sputum combined with exhaled breath analysis in symptomatic patients

When used in concert, DNA hypermethylation in sputum and exhaled breath analysis were able to detect all symptomatic lung cancer patients in our relatively small series (n=20). Both methods act on different biological levels and it is remarkable that they complemented each other. Still, main critical points are the small study population and the lack of availability and standardization of these technologies.

To gain detailed insights in the composition of volatile organic compounds in exhaled breath, for instance, for use as candidate biomarker for lung cancer or to unravel lung cancer pathophysiology, breath analysis using gas chromatography mass spectrometry (GC-MS) is considered the gold standard. However, as GC-MS is a sophisticated technique (it is expensive and involves complex analysis), it is (currently) not suitable for high-throughput application.

A biological natural version of the eNose technology is the canine olfactory system. One study with specially trained ‘sniffer’ dogs reported that dogs are indeed able to detect lung cancer patients from controls with a moderate sensitivity of 71% and specificity of 93%. Dogs were not hindered by signal interference such as smoking, but one has to bear in mind that the dogs need intensive training and service is not infinite. Development of reproducible sensitive electromechanical devices may be preferred.

Theoretical advantages of eNose technology are that it is non-invasive, inexpensive, portable, easy to use and allows for quick on-the-spot analysis. Up to 30% of smokers have no sputum production, even when induced with saline. eNose may be a suitable means for lung cancer detection in these subjects. A limitation of the eNose is that test results may differ essentially between devices, even if the eNoses are from the same manufacturer. Therefore, it is currently difficult to generalize results between the eNoses. On the other hand, breathprint measurements are highly reproducible using the same device.

Our findings may hold promise for pre-selection of high risk individuals to undergo LDCT screening for the early detection of lung cancer. LDCT is a costly method and has a high false-positive rate reported in the National Lung Screening Trial (NLST). If high risk individuals are screened first by exhaled breath and sputum examination, 40% will be ruled out of disease, resulting in a smaller group of subjects to undergo LDCT.

Since the low combined specificity is attributable to the eNose, refinement of eNose technology is desirable. Future technical improvements may greatly facilitate the eNose analysis. Further validation of both methods combined in a larger cohort and
the feasibility of the tests in a screening setting needs to be examined.

Similar to the approach of combining two different modalities of DNA hypermethylation in sputum and exhaled breath analysis, it may be interesting to examine a possible complementary diagnostic value with other technologies as well (for instance, blood-based markers), as further discussed below.

**EGFR mutation analysis in sputum**

Detection of *EGFR* mutation in sputum may be used for lung cancer prediction or to assess whether targeted treatment with tyrosine kinase inhibitors is indicated.

All examined techniques were capable of demonstrating *EGFR* mutation in sputum, with a sensitivity range similar to that of *EGFR* mutation detection in blood. The chance on detecting the mutation was significantly higher when tumour cells were present in the sample, or in a related sputum sample, as judged by cytological examination. This indicates that tumour load may be of influence. Nonetheless, *EGFR* mutation was also detected in one sample that did not show tumour cells, nor in a related sample, at cytological examination, implying that presence of tumour cells is not a prerequisite for a positive *EGFR* mutation test. This was in line with a previous study on *KRAS* mutation analysis in sputum and with the hypermethylation analysis of this thesis.

Only exon deletions 19 were identified, but no L858R and T790M mutations, which may be a coincidence due to the small sample size. It is probably not related to technical specifications of the assays, as analytical features were highly similar between the assays. Since T790M mutation is associated with secondary resistance towards TKI treatment, it is likely to become manifest at a later stage during lung cancer progression (clonal selection). It may be initially present in a subset of lung cancer cells, below LOD.

The findings of this study can be promising for patients of whom insufficient tumour material is available or if bronchoscopy cannot be performed due to physical limitations. Our results warrant further evaluation of the diagnostic value of *EGFR* mutation analysis in a larger cohort of subjects and the examination of sputum in patients who are not considered for bronchoscopy. Conceptually, it may be interesting to investigate whether combinations of assays, such as COLD-PCR with PNA, will lead to higher sensitivity without compromising specificity.

**DNA hypermethylation in sputum of asymptomatic high risk individuals**

For which purpose may biomarkers be used in lung cancer screening? We hypothesized the following three scenarios:
1. Pre-selection of high risk individuals for LDCT screen, to reduce the number of participants to undergo LDCT. In this case, a high sensitivity is necessary, but specificity of the biomarker test is of less importance (because LDCT will be performed in those with a positive test).

2. Diagnosis of lung cancer after detection of a solid nodule on LDCT screen. In the NLST, 96% of positive screening results were false-positive.\textsuperscript{37} Both high sensitivity and specificity are desirable. In particular, the ‘diagnostic markers’ may be relevant in this setting. Furthermore, a number of stage I cancers detected by LDCT screening may not or very slowly progress to advanced stage in a certain time frame.\textsuperscript{46} This phenomenon, resulting in the epidemiologic term ‘overdiagnosis’ (more lung cancer cases are diagnosed than expected due to the fact that slow growing tumours give rise to symptoms much later in time), may in future be tackled by biomarker analysis, ideally identifying those cancers that will truly progress and require treatment, or postponing treatment in patients with slow growing tumours.

3. Risk assessment after LDCT screen that showed a ground glass lesion. Theoretically, a ground glass lesion may remain in situ for a while. In this context, biomarkers predicting progression to invasive lung cancer may be used, possibly guiding clinical follow-up policy.

Preferably, non-smokers should also be included in LDCT screening trials, although it is obvious that not the whole population of age 50-75 years can be tested. Therefore, further elucidation of risk profiles is needed, taking into account risk factors and lung cancer characteristics in non-smokers. In current LDCT screening trials, only heavy smokers ($>$15 pack years (PY)) are included. This implies that 15% of lung cancers, occurring in never-smokers and smokers $<$15 PY, will be missed.\textsuperscript{47} In Chapter 5, we reported that most of selected hypermethylation biomarkers can also detect lung cancer in non-smokers with particular high specificity for the combination of RASSF1A and 3OST2. In never smokers, usually adenocarcinomas are detected. \textit{EGFR} mutations, and to a lesser extent \textit{KRAS} mutations, may therefore be attractive as biomarker as well, as these predominantly occur in adenocarcinomas.\textsuperscript{48,49}

We assumed, based on the rationale, that a marker with moderate specificity in phase II studies was not suitable for use as ‘diagnostic marker’, but may still be interesting as ‘risk marker’ with relevant complementary value. Such a marker should therefore not be excluded from testing in phase III studies. However, it turned out that our ‘risk markers’ are very weak predictors in screening setting. The risk model as
composed in Chapters 3 and 5 was inadequate in identifying subjects that developed lung cancer during follow-up, indicating that this model is not suitable for screening when incorporating the selected markers. We therefore conclude that ‘risk markers’ only have potential in screening when they complement ‘diagnostic markers’, provided that specificity is maintained.

‘Diagnostic marker’ RASSF1A demonstrated a low sensitivity (11%). Nevertheless, despite the observation that RASSF1A behaves as a ‘risk marker’ with a specificity below 96%, the marker still showed a high specificity (90%). Additional markers with similar high specificity may be useful in the screening setting.

Current research focuses on defining the optimal set-up of LDCT screening. The NELSON trial uses slightly different inclusion criteria than the NLST (regarding age and smoking PY), employs varying screening intervals and lung nodule management is based upon a different strategy.\textsuperscript{50,51} First reports show a lower false-positive rate as compared to NLST. Furthermore, McWilliams \textit{et al} defined predictive factors that are associated with lung nodule malignancy, demonstrating reduction of false-positivity in a cohort of Canadian high risk individuals.\textsuperscript{52} In another study, various screening models were compared to designate optimal inclusion criteria and screening intervals.\textsuperscript{53} Thus, sputum analysis seems not to be of additive value when the NELSON screening strategy is used, but it is not excluded that sputum analysis may be supporting in other lung cancer screening models with, for example, high false-positivity rate (e.g., NLST).

**Sputum cytology for lung cancer screening**

Sputum cytology in the NELSON trial only detected minor cellular aberrations in 13% of cases, which does not correspond to reported sensitivity of sputum cytology in CXR trials in the past (20-30%).\textsuperscript{54,55} This is explained by the smaller size of LDCT detected tumours (nodules starting from 2 mm can be detected) as compared to CXR detected tumours (minimally 2 cm) in asymptomatic high risk individuals. Thus, there is no role for sputum cytology in LDCT screening.

**FUTURE EFFORTS**

**Biomarkers in sputum**

Taking all results of our studies into consideration, we conclude that ‘risk markers’ are not suitable for lung cancer risk stratification both in the clinical and screening setting. Hence, novel, preferably diagnostic, non-invasive markers are needed. These
new biomarkers may conceptually not only be used for (early) lung cancer detection, but also may have potential to monitor disease during and after treatment, and provide information on prognosis. Specific sputum-based biomarkers for histological subtypes (especially to enable distinction between squamous cell carcinomas and adenocarcinomas) may be useful in assisting choice of treatment. To this end, EGFR mutation analysis in sputum is possible (as mentioned above).

High-throughput epigenome profiling of the different lung cancer subtypes may facilitate comparison of DNA methylation profiles and may result in the identification of additional, complementary, lung-cancer specific DNA hypermethylation markers for examination in sputum.\textsuperscript{56} Of note, the reversibility of hypermethylation suggests that targeted epigenetic therapy may be feasible to reactivate expression of tumour-suppressor genes,\textsuperscript{57} also requiring identification of predictive markers.

Promising novel technologies have been developed in recent years, among which targeted next generation sequencing with, for instance, the Truseq Amplicon Cancer Panel (TSACP). This is a multiplexed assay enabling comprehensive screening of mutations in genes that are often affected in cancer pathways. In house validation of this technique showed good performance on DNA from formalin-fixed paraffin-embedded tumour samples, similar to a routinely used HRM-sequencing assay (Sie \textit{et al}, submitted), yet formalin artifacts may negatively impact mutation detection sensitivity.\textsuperscript{58–60} However, the claimed high sensitivity is not confirmed yet. We applied TSACP specifically to investigate \textit{EGFR} mutations in sputum, having DNA of higher quality compared to formalin-fixed samples. In the context of diagnosing lung cancer, it may be appealing to analyze sputum samples for mutations in the other genes as well. Possibly mutation analysis (TSACP or larger similar panels) may lead to higher complementary sensitivity. It is worthwhile to examine this on the NELSON sputum samples.

MicroRNAs (miRs) in sputum have emerged as putative novel lung cancer biomarkers.\textsuperscript{61–63} One study performed concurrent analysis of miRs in sputum and LDCT in symptomatic stage I lung cancer patients,\textsuperscript{64} reporting enhanced specificity of LDCT with testing of miRs. This study needs validation in a large cohort and examination of the microRNAs in sputum of a screening population. We detected miRs in both the cellular fraction of sputum and in extracellular vesicles (EVs; data not published) of symptomatic patients and controls. EVs transport miRs, messenger RNAs (mRNAs) and proteins, which play a role in cell-to-cell communication.\textsuperscript{65} They are considered to exert diverse biological functions in physiological and pathological conditions. So far, EVs detection in sputum has not been reported. Interestingly, in one lung cancer patient
with confirmed KRAS mutation in the primary tumour, we identified the mutation in mRNA derived from EVs in sputum of this patient. This may be further explored in future research to gain more insights in lung carcinogenesis.

Biomarkers in blood

In this thesis, our focus was on sputum-based biomarkers for lung cancer detection. A similar, relatively non-invasive methodology would be the examination of biomarkers in blood. At start of our studies, standardization methods were lacking and blood-based biomarkers seemed to be suitable only for advanced stage lung cancer. Our group demonstrated that cpDNA levels were only higher over controls in patients with invasive lung cancer, and not in subjects with high- or low-grade pre-invasive lesions, suggesting that cpDNA levels are not increased during the pre-invasive stages of lung squamous carcinogenesis.\(^{66}\)

However, in recent years novel developments have been reported on the use of blood-based genomic tumour markers for lung cancer detection, for instance miRs (reviewed in \(^{67}\)). Interestingly, Sozzi and coworkers validated a miR signature classifier in plasma of high risk individuals in the Multicenter Italian Lung Detection (MILD) trial and reported a five-fold reduction of CT-false positivity with combined use of LDCT and the classifier.\(^{68}\) In advanced stage lung cancer, blood markers have potential to monitor disease during and after treatment, and to provide information on dissemination.\(^{69,70}\) Likewise, EGFR and KRAS mutation analysis is feasible in plasma DNA of lung cancer patients and may have a role in monitoring treatment,\(^{71–74}\) but in routine clinical practice tumour tissue is still required for mutation analysis. Protein analysis in blood has also shown opportunities for lung cancer prediction and prognosis,\(^{75}\) but at individual level most markers have limited diagnostic accuracy (reviewed in \(^{76}\)). However, since most concern early phase studies, validation and standardization of methods is needed.

CONCLUSION

In short, RASSF1A hypermethylation has been validated as ‘diagnostic marker’ and may play a role in clinical lung cancer diagnosis. Feasibility of EGFR mutation analysis in sputum should be further examined, and may have therapeutical consequences. In the screening setting with LDCT, exploration of exhaled breath analysis combined with new DNA hypermethylation markers may conceptually be rewarding in the
pre-selection of high risk individuals.
REFERENCES


58. Tuononen K, Mäki-Nevala S, Sarhadi VK, et al. Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-


APPENDIX A.
EARLY DETECTION RESEARCH NETWORK (EDRN) GUIDELINES FOR BIOMARKER DEVELOPMENT STUDIES.

The EDRN guidelines describe five consecutive phases:

1. Preclinical exploratory studies to identify genes (biomarkers) that show relative overexpression or underexpression in tumour tissue compared to non-tumour tissue.

2. Development of a clinical assay using biological specimens that are procured non-invasively from cases and controls, in which the biomarkers (or related products) can be measured. Clinical evaluation should demonstrate that the biomarker(s) is able to discriminate patients with cancer from controls without cancer.

3. In retrospective longitudinal repository studies, the capacity of the biomarker to detect preclinical disease is examined in biological specimens sampled from subjects who were asymptomatic at time of collection. Follow-up enables the identification of subjects who developed cancer. Thus, samples collected prior to disease onset can be compared with controls. If the biomarker indeed appears to be able to predict disease, criteria are formulated for a positive screening test based on these studies.

4. In prospective screening studies, accuracy of the test is measured by the screening of at risk individuals. Those with a positive screening test will undergo diagnostic procedures. Detection rate and false referral rate are assessed. A large study cohort is required, representing the target screening population.

5. The final phase consist of cancer control studies. In these studies, the focus lies on whether the test leads to a reduction of cancer mortality in the screened population.
APPENDIX B. EVALUATION OF TEST ACCURACY FOR LUNG CANCER DIAGNOSIS.

The diagnostic accuracy of a test reflects the level of agreement between the (binary) outcome of the index test (test under evaluation) and the reference standard. The reference standard consists of another test or combination of methods, including clinical follow-up.

Most studies examining a novel test describe diagnostic accuracy in statistical terms of sensitivity and specificity. Ideally, a diagnostic test is able to identify all subjects having the condition, with no false-positive and false-negative test results (i.e., 100% sensitivity and 100% specificity) in all populations under study, regardless of age, gender, ethnicity, risk factors and tumour stage. However, in practice this is difficult to achieve, as usually overlapping outcomes exist in cases and controls. The application of the test then determines what the ‘optimal’ threshold is.

A high threshold leads to a high specificity of the test, with the advantage that false-positivity rate will be low. In clinical practice, this means (with ‘certain diagnosis’) that no additional (invasive) diagnostic procedures have to be performed with associated morbidity and mortality and subjects are spared unnecessary psychological distress. However, due to a low sensitivity, true cancer diagnoses may be missed. Alternatively, a low threshold can be considered in order to detect as many lung cancers as possible. A high sensitivity is yielded, at the expense of more false-positive test results: an essential part of the controls will also be positive.

The setting in which a test is used is thus of relevance. To establish lung cancer diagnosis in symptomatic patients, different test performance criteria are required as compared to a test that is used for risk assessment in asymptomatic high risk individuals. A high sensitivity is recommended for ‘ruling out’ disease, in case of a negative test result, whereas a high specificity ‘rules in’ disease, when the test is positive.

For each biomarker, a series of cut-off (threshold) values are available, each yielding a certain sensitivity and specificity. In a receiver operating characteristic (ROC) curve, the true positive rate (sensitivity) as function of the false-positive rate (1-specificity) across all possible cut-off values are displayed. The ability of an index test to discriminate between presence and absence of a condition, for a given cut-off value, is assessed by a contingency table (Table 1). The performance of the index test is measured by several intrinsic parameters, of which most relevant are explained below:
Sensitivity of the index test: the proportion of subjects who are correctly identified as having the condition (i.e., ‘true positive’ rate). Sensitivity = $\frac{a}{a + c}$

Specificity of the index test: the proportion of subjects who are correctly rejected by the index test as not having the condition (i.e., ‘true negative’ rate). Specificity = $\frac{d}{b + d}$.

Positive predictive value (PPV): the proportion of subjects with a positive test result that is correctly diagnosed having the condition; hence, this is affected by the prevalence of the condition in the population under study. PPV increases in a clinical setting with a high prevalence. A synonym is ‘post-test probability of disease’. PPV = $\frac{a}{a + b}$

Negative predictive value (NPV): the proportion of subjects with a negative test result that is correctly rejected having the condition. NPV = $\frac{d}{c + d}$

Accuracy of the index test: the proportion of correct test results (both true positive and true negative) in the population under study. Accuracy = $\frac{a + d}{a + b + c + d}$

<table>
<thead>
<tr>
<th><strong>Condition</strong>*</th>
<th><strong>Present</strong></th>
<th><strong>Absent</strong></th>
<th><strong>PPV</strong></th>
<th><strong>NPV</strong></th>
<th><strong>Sensitivity</strong></th>
<th><strong>Specificity</strong></th>
<th><strong>Accuracy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Index test</strong></td>
<td><strong>Positive</strong></td>
<td>a - true positive</td>
<td>b - false positive</td>
<td>$PPV$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Negative</strong></td>
<td>c - false negative</td>
<td>d - true negative</td>
<td>$NPV$</td>
<td><strong>Sensitivity</strong></td>
<td><strong>Specificity</strong></td>
<td><strong>Accuracy</strong></td>
</tr>
</tbody>
</table>

* Presence of the condition is as defined by reference test (gold standard).