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 RASSF1A, APC and cytoglobin (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’.

RASSF1A demonstrated the best discriminatory capacity between lung cancer patients and controls with consistent results in both sets. Regarding individual specificity levels, RASSF1A was considered as ‘diagnostic marker’ (specificity ≥96%), and APC and CYGB as ‘risk markers’. The panel of RASSF1A, APC and 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 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 RASSF1A, APC and 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 3OST2, PRDM14, PHACTR3 and 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. \textit{RASSF1A} showed the best diagnostic performance in both sets, further substantiating \textit{RASSF1A} hypermethylation as ‘diagnostic marker’. Two panels were constructed consisting of either \textit{RASSF1A}, \textit{3OST2} and \textit{PRDM14} (sensitivity 82%, specificity 66%; based on Youden’s J index), or \textit{RASSF1A}, \textit{3OST2} and \textit{PHACTR3} (sensitivity 67%, specificity 90%; based on high specificity). The risk model incorporated \textit{RASSF1A} (high risk classification), \textit{3OST2} and \textit{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 \textit{RASSF1A}, \textit{3OST2} or \textit{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%; \textit{P}<0.001).

In \textbf{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 \textit{RASSF1A} and \textit{3OST2}, and not eNose. However, sensitivity for lung cancer diagnosis increased to 100% when \textit{RASSF1A} was combined with eNose (specificity 42%). Of interest, the additional complementary diagnostic value of \textit{RASSF1A} hypermethylation to eNose and vice versa was significant.

In \textbf{Chapter 7}, we demonstrated that \textit{EGFR} mutation analysis is feasible in sputum of lung cancer patients. DNA of a similar cohort of subjects and serial dilutions of \textit{EGFR} mutated cell lines were examined using different molecular techniques in a multicenter collaboration. All techniques were capable of detecting \textit{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 \textit{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 Leuven 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.