Molecular sputum analysis for the diagnosis of lung cancer

A J Hubers1, C F M Prinsen2, G Sozzi3, B I Witte4 and E Thunnissen*,1

1Department of Pathology, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands; 2Department of Pathology, Canisius-Wilhelmina Hospital, Weg door Jonkerbos 100, 6532 SZ Nijmegen, The Netherlands; 3Tumor Genomics Unit, Department of Experimental Oncology, Fondazione IRCCS Istituto Nazionale Tumori, Via Venezian 1, 20133 Milan, Italy and 4Department of Epidemiology and Biostatistics, VU University Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

Lung cancer is the leading cause of cancer mortality rate worldwide, mainly because of the presence of metastatic disease at the time of diagnosis. Early detection of lung cancer improves prognosis, and towards this end, large screening trials in high-risk individuals have been conducted since the past century. Despite all efforts, the need for novel (complementary) lung cancer diagnostic and screening methods still exists. In this review, we focus on the assessment of lung cancer-related biomarkers in sputum in the past decennium. Besides cytology, mutation and microRNA analysis, special attention has been paid to DNA promoter hypermethylation, of which all available literature is summarised without time restriction. A model is proposed to aid in the distinction between diagnostic and risk markers. Research on the use of sputum for non-invasive detection of early-stage lung cancer has brought new insights and advanced molecular techniques. The sputum shows a promising potential for routine diagnostic and possibly screening purposes.

In theory, a biomarker sputum test for early detection may be developed for three possible applications: (i) identification of at-risk individuals, who may be screened with LDCT after a positive biomarker test; (ii) after the first LDCT screen shows a solid lesion, a sputum biomarker may be developed as a diagnostic test for malignancy; and (iii) after the first LDCT screen shows a ground glass lesion, we can determine whether the lesion has a high or low chance of becoming malignant. In the setting of patients with symptomatic lung cancer, a sputum test may be useful for diagnostic workup of malignancy and if diagnosed with lung cancer to perform predictive analysis.

Biomarker screening may be categorised into (i) risk markers, which identify individuals at high risk of developing lung cancer, and (ii) diagnostic markers, which uncover invasive lung cancer. The biomarker must meet several conditions, such as being superior to conventional detection methods in terms of sensitivity and specificity, before it is considered suitable for clinical implementation (Box 1). In this review, diagnostic markers are defined as markers recognising (the transition to) invasive lung cancer. At this stage, the disease may be measurable but still
asymptomatic. A risk marker is able to identify subjects at risk without measurable disease. In pathobiological terms, this marker may be associated with several conditions, such as measure of exposure to carcinogen and development of carcinoma in situ (Selamat et al, 2011).

In 2003, a review summarised the status of mutation analysis and initial methylation findings in sputum (Thunnissen, 2003). This manuscript provides an overview of developments in sputum analysis for lung cancer diagnosis in the past 10 years. The PubMed terms ‘lung cancer’ and ‘sputum’ were used. In addition, special attention is paid to DNA hypermethylation.

### SPUTUM CYTOLOGY

By means of cytology, tumour cells can be identified in sputum through aberrant cell morphology. Status of the diagnostic value of sputum cytology has not changed in the past decennium. In the clinical diagnostic setting, the sensitivity of sputum cytology is 60%, which also depends on the number of sputum samples examined (Risse et al, 1985). Although in developed countries the procurement of tumour biopsies/tumour cytology replaced the use of sputum cytology as standard for lung cancer diagnosis (Rivera and Mehta, 2007), in lower budget countries sputum cytology is an affordable diagnostic instrument and still clinically implemented (Ammanagi et al, 2012).

### MOLECULAR ANALYSIS OF SPUTUM

#### DNA mutation analysis. **For DNA mutation analysis, currently the relevant part of the gene of interest is amplified, usually using polymerase chain reaction (PCR) technology. This is a very sensitive, low-cost, rapid and simple method.** Disadvantages are that contamination may be an issue, as well as that the enzyme DNA polymerase has a small error. In about 0.1% of amplicons, an incorrect nucleotide may be incorporated (Eckert and Kunkel, 1991). If this error happens early in the PCR procedure, it will propagate and may lead to a false-positive signal, thus reducing specificity.

Mutations of tumour-suppressor gene p53 and oncogene KRAS have been identified to have a role in lung carcinogenesis (Hanahan and Weinberg, 2011). In 50% of lung cancer cases, mutations or deletions are present in the p53 gene (Greenblatt et al, 1994). KRAS mutations mostly occur in adenocarcinomas (20–30% in western countries and 10% in eastern countries) (Shigematsu et al, 2005).

Various KRAS mutation detection techniques have been investigated on sputum specimens (Table 1). Peptid nucleic acid–PCR–restriction fragment length polymorphism (PNA–PCR–RFLP) and Point-EXACCT were described as methods of choice (Thunnissen, 2003). Destro et al (2004) confirmed KRAS mutation in 79% of the sputum samples from lung cancer patients with a KRAS mutation in their tumour tissue (n = 14). In controls, none tested positive. Keohavong and co-workers (2004, 2005) conducted studies in Xuan Wei County (China), where lung cancer rates were fivefold higher than the Chinese national average. Mutation detection was optimised by application of cell cytocentrifugation and laser capture microdissection, enabling detection of low fraction mutations, even in morphologically benign bronchial epithelial cells (Keohavong et al, 2004, 2005). With this approach, examination of cytology is still needed for dissecting abnormal or benign epithelial cells for enrichment. In a cancer-free population, mutations in both genes were identified (15 out of 92) (Keohavong et al, 2005). These mutations occurred in none of the matched buccal epithelial cells, indicating that the latter cells are not suitable as a surrogate marker for lung cancer (risk).

Until recently, most sputum studies have been performed on patients with symptomatic lung cancer. Research conducted before 2003 show that KRAS mutations may be detected in sputum at least 1 year before clinical diagnosis of lung cancer (Somers et al, 1998). Baryshnikova et al (2008) were the first to investigate sputum from a large LDCT screening cohort (n = 803) consisting of asymptomatic heavy smokers, assessing frequency of KRAS and p53 mutations, next to DNA promoter hypermethylation of p16, NORE1A and RASSF1A (Supplementary Table 1). KRAS mutation analysis was performed by restriction endonuclease-mediated selective PCR with a reported sensitivity of one mutant per 1000 wild-type genes. No KRAS mutation was identified, especially in the 18 subjects who developed lung cancer during the follow-up period. None of these patients had molecular alterations at baseline. In 15 out of 803 (2%) participants, a p53 mutation was found, of whom one patient was diagnosed with early-stage lung cancer in follow-up without confirmation of the p53 mutation in the tumour.

These studies suggested that KRAS might be more suitable as a diagnostic marker than for risk assessment in precancerous stages. Future studies with further follow-up of participants are needed to elucidate whether molecular alterations of KRAS and p53 are indeed suggestive for lung cancer development.

Mutations in the tyrosine kinase domain of the epidermal growth factor receptor (EGFR) have been identified in parts of lung adenocarcinomas, and are associated with high response rates to treatment with EGFR tyrosine kinase inhibitors (Sharma et al, 2007). Epidermal growth factor receptor mutation analysis has been performed in some sputum samples as part of larger series of other cytological samples, mostly without detailed information and not compared with the original tumour (Boldrini et al, 2007; Takano et al, 2007; Tanaka et al, 2010). In a total of three publications, 3 out of 25 sputum samples were positive in cases with cytologically proven malignant cells.

EML4-ALK is a lung cancer fusion oncogene that is estimated to be expressed in 3–6% of lung adenocarcinomas (Takeuchi et al, 2008), showing marked response to treatment with ALK inhibitors (Kwk et al, 2010). Recently, Soda et al (2012) reported the development of a multiplex RT–PCR system that was able to detect EML4-ALK mutations in 4 out of 35 sputum samples, which were part of a prospective screening cohort of NSCLC patients.

Optimisation of EGFR and EML4-ALK mutation detection in sputum may, in the future, contribute to minimise the use of invasive bronchoscopy or transthoracic needle biopsies to secure tumour biopsies for mutation testing, a clinical need in monitoring personalised treatment.

#### DNA hypermethylation. **Aberrant DNA promoter methylation is a cell control mechanism in lung carcinogenesis** (Selamat et al, 2011),
involving the addition of a methyl group at the carbon 5 position of cytosines at CpG sites in DNA. A widely used approach to distinguish methylated DNA from unmethylated DNA is exposing genomic DNA to bisulphite before PCR. In this process, unmethylated cytosine is converted into uracil, whereas methylated cytosine remains unchanged. The templates are next subjected to denaturing gradient gel electrophoresis before PCR. In this process, bisulphite conversion may be incomplete. In that disadvantage that bisulphite conversion may be incomplete. In that

| Study | Subjects | Gene | Method | PCR cycles | Cases | Controls | Remarks |
|-------|----------|------|--------|------------|-------|----------|---------|
| Baryshnikova et al (2008) | Smokers | KRAS, p53 | PCR-RFLP, PCR and SSCP | 40 | 0/506 (0%) | 15/803 (2%) | Follow-up 2–6 years; 18 patients developed lung cancer without molecular alterations at baseline. One patient with p53 mutation at baseline developed SqCC, but was not confirmed in resected tumour tissue. Also DNA promoter hypermethylation tested of p16, NORE1A and RASSF1A (Supplementary Table 1). |
| Destro et al (2004) | NSCLC | Smokers | KRAS | PCR-RFLP | 40 | 11/50 (22%) | 0/100 (0%) | Fourteen of 50 tumour tissue samples tested KRAS mutation positive. In three cases, concomitant p16 hypermethylation (Supplementary Table 1) |
| Keohavong et al (2003) | All | KRAS | MAE, PCR and DGGE | 47 | XW: 23/102 (23%), BH: 7/50 (14%) | Data of both tumour and sputum were presented together. Two study populations: Xuan Wei County (XW) and Beijing and Henan (BH), respectively. XW subjects were exposed to coal smoke |
| Keohavong et al (2004) | Lung cancer NS | KRAS, p53 | Cell centrifugation, laser capture microdissection, PCR and DGGE (KRAS), SSCP (p53) | 42 | 2/15 (13%) | Subjects were exposed to coal smoke. KRAS mutation status of primary tumour unknown |
| Keohavong et al (2005) | (Non) Smokers | KRAS, p53 | Cell centrifugation, laser capture microdissection, PCR and DGGE (KRAS), SSCP (p53) | 42 | 2/92 (2%) | Subjects were exposed to coal smoke |
| Zhang et al (2003) | NSCLC | KRAS | MAE, PCR and DGGE | 55 | 10/22 (46%) | In 12 out of 22 matched tumour–sputum samples, KRAS mutation was identified using the same method (κ = 0.64, 95% confidence interval: 0.32–0.95, P < 0.01). One patient tested negative in tumour, but positive in sputum |

Abbreviations: All = all types of lung cancer included; DGGE = denaturing gradient gel electrophoresis; MAE = mutant allele enrichment; NS = not specified; NSCLC = non-small-cell lung cancer; PCR = polymerase chain reaction; RFLP = restriction fragment length polymorphism; SqCC = squamous cell carcinoma; SSCP = single-strand conformational polymorphism.
non-small-cell lung cancer, others also examined small-cell lung cancer and unspecified lung cancer cases. The examined populations consisted usually of more male subjects than female subjects (average 75% in cases, 70% in controls, respectively).

From the summary table, it is apparent that only for a limited number of genes published data were available for categories 4 or 5 cycles, respectively. Concerning p16 gene, five studies (Belinsky et al, 1998; Destro et al, 2004; Olaussen et al, 2005; Cirincione et al, 2006; Shivapurkar et al, 2007) with ≤55 PCR cycles and nine studies with >55 PCR cycles (Kersting et al, 2000; Palmisano et al, 2000; Konno et al, 2004; Belinsky et al, 2006; Hsu et al, 2007; Liu et al, 2008; Guzmán et al, 2012; Leng et al, 2012; Shin et al, 2012) with sensitivity and specificity data were available for bivariate analysis (Reitsma et al, 2005). Interestingly, mean specificity was shown to be significantly lower in the group of studies with ≤55 PCR cycles (74% vs 87%, P<0.001), whereas

---

**Table 2. Studies investigating presence and/or expression of RNA and tumour-related proteins in sputum**

| Study            | Cases          | Controls        | Method         | Protein/ gene | Results cases | Results controls | Se (%) | Sp (%) | Positive cytology | Remarks                                                                 |
|------------------|----------------|-----------------|----------------|---------------|---------------|-----------------|--------|--------|-------------------|------------------------------------------------------------------------|
| Sun et al (2009) | All Benign pulmonary disease | RTQ-PCR, immunocytochemistry | APRIL*         | 58/71 (82%)   | 2/62 (3%)     | 82              | 97     |        |                   | Cases 10/71 (14%); all SCC Immunocyto: 11/71 (16%); Healthy subjects: 1/65 (2%); Cutoff value: mean ± 2 s.d. of mRNA expression in healthy subjects |
| Mecklenburg et al (2004)b | All Benign pulmonary disease | RT-PCR | MAGE-1 | 2/14 (14%) | 0/2 (0%) | 14 | 100 |        | cases 1/8 (13%); Positive cytology sample was not tested with RT-PCR; Cytology of remaining samples not performed |
|                  |                |                 | MAGE-2 | 1/14 (7%) | 0/2 (0%) | 7 | 100 |        |                                                                 |
|                  |                |                 | MAGE-3/6 | 0/14 (0%) | 0/2 (0%) | 0 | 100 |        |                                                                 |
|                  |                |                 | MAGE-4 | 2/14 (14%) | 0/2 (0%) | 14 | 100 |        |                                                                 |
|                  |                |                 | MAGE-12 | 2/14 (14%) | 0/2 (0%) | 14 | 100 |        |                                                                 |
|                  |                |                 | All combined | 5/14 (36%) | 0/2 (0%) | 36 | 100 |        |                                                                 |
| Jheon et al (2004)b | All Benign pulmonary disease | RT-N-PCR | MAGE A1–6 | 72/134 (54%) | 3/140 (2%) | 54 | 98 |        | Cases 6/31 (19%); Also spontaneous sputum collected. Data of lung cancer patients from group I (collection at the day of thoracotomy) and II (lung cancer in clinical workup) combined. Follow-up (1 year): no cancer in controls |
|                  |                |                 | TRAP method | Telomerase | 8/27 (30%) | 30 |    |        |                                                                 |
| Pasrija et al (2007)b | All Cancer-free subjects | TRAP method | Telomerase | 23/34 (68%) | 3/30 (10%) | 68 | 90 |        |                                                                 |
| Pio et al (2010)b | All Cancer-free subjects | Anti-factor H antibodies | Complement factor H* | 80 | 88 |    |    |        | Also spontaneous sputum collected. Se and Sp based on cutoff ROC curve |
| Kalomenidis et al (2004) | All Benign pulmonary disease | IRMA | CEA* | 57 | 95 |    |    |        | Se and Sp based on cutoff ROC curves |
|                  |                | IRMA | NSE* | 19 | 95 |    |    |        |                                                                 |
|                  |                | IRMA | CYFRA 21-1* | 36 | 95 |    |    |        |                                                                 |
| Hillas et al (2008)b | All COPD | IRMA | CEA | NS | NS | cases 4/50 (8%); CEA median concentration. Cases: 713 ng ml⁻¹, controls 518 ng ml⁻¹ |    |    |        |                                                                 |
|                  |                | IRMA | NSE | NS | NS | NSE median concentration. Cases: 12 ng ml⁻¹, controls 13.7 ng ml⁻¹ |    |    |        |                                                                 |
|                  |                | RIA | CYFRA 21-1* | 86 | 75 |    |    |        |                                                                 |

Abbreviations: All = all types of lung cancer included; CEA = carcinoembryonic antigen; COPD = chronic obstructive pulmonary disease; IRMA = immunoradiometric assay; NSE = neuron-specific enolase; NS = not specified; RIA = radioimmunoassay; ROC = receiver operating characteristic; RT–(Q)(N)–PCR = reverse transcriptase (quantitative) (nested)–polymerase chain reaction; Se = sensitivity; Sp = specificity; TRAP = telomeric repeat amplification protocol.

*a*P < 0.05 significance level between cases and controls.

*b*Induced sputum.
sensitivity was higher, but not significantly different (49% vs 33%, \( P > 0.13\)). This literature analysis supports the above-mentioned theoretical notion that a high number of PCR cycles leads to a higher chance of false-positive results. It is not excluded that a diagnostic marker may be looked upon as a risk marker (as defined in Box 1), when >55 PCR cycles are run with possible induced false positivity. Moreover, when comparing the number of PCR cycles (>55 or \( \leq 55\)) with marker classification (diagnostic vs risk), a biomarker is more likely to be classified as a risk marker if >55 PCR cycles were applied compared with at most 55 PCR cycles (85% vs 58%, \( P = 0.002\)).

In 10 studies (Kersting et al, 2000; Palmisano et al, 2000; Chen et al, 2002; Liu et al, 2003; Wang et al, 2003; Destro et al, 2004; Olaussen et al, 2005; Cirincione et al, 2006; Belinsky et al, 2007; Hsu et al, 2007) matched tumour and sputum samples were examined. The median frequency of gene hypermethylation was higher in tumour than in sputum samples: 48% (interquartile range 36–64%) vs 38% (interquartile range 31–57%), respectively. A meta-analysis on exact data (Kersting et al, 2000; Liu et al, 2003; Wang et al, 2003; Olaussen et al, 2005; Shivapurkar et al, 2007; Shin et al, 2012) showed that this observed tendency was not significant (\( P = 0.09\); Durkalski et al, 2003). Median concordance of methylation between tumour and matched sputum, calculated from the same studies, is 78% (interquartile range 73–91%), indicating that the use of sputum as non-invasive biological fluid for detection of aberrant methylation is representative of the methylation status of primary tumour tissue.

Still, none of the biomarkers yield 100% sensitivity. The multidimensional character of lung cancer, in which various genes might be involved (Hansen et al, 2011), requires a panel of markers that can complement each other in lung cancer detection. Several studies calculated combined sensitivity and specificity for hypermethylated genes (Zochbauer-Müller et al, 2003; Belinsky et al, 2005, 2006, 2007; Hsu et al, 2007), revealing higher performance when compared with the markers individually. These algorithms seem promising, but are scarcely validated in independent study cohorts. Interestingly, one study (Leng et al, 2012) replicated a panel of previously published hypermethylation markers (Belinsky et al, 2007) in two independent slightly different cohorts: case–control vs asymptomatic stage I lung cancer patients. They showed a slightly higher sensitivity and specificity in the second cohort. However, the methylation panels were not exactly similar between the study cohorts. Also, as sputum samples were stored in Saccamanno after collection without further treatment, DNA quality may be reduced, possibly affecting the study data. Therefore, at this point in time, it is difficult to define an unambiguous biomarker signature panel for lung cancer risk based on these results.

Patient selection, sputum collection and procedure methods might explain the differences in rates of methylation between studies investigating the same biomarker.

Research into additional novel markers remains necessary.

**Loss of heterozygosity.** Microsatellite alterations present as loss of heterozygosity (LOH), or as microsatellite instability (MSI). Conceptually, LOH is essentially different from previous markers, because it explores the absence of the allele that is present in the normal situation, whereas the other above-mentioned biomarkers look for the presence of a specific abnormality. Because the fraction of tumour cells in sputum is usually <1%, the majority of the cells will not have LOH. Therefore, looking for tumour-related LOH has a disadvantage: requiring a difference that is higher than the threshold of the test based on signal-to-noise ratio. For example, when LOH is present in 1% of tumour cells, the proportion of missing alleles is 0.5%. To demonstrate this, a test is required that is able to make a distinction between 100% (normal reference DNA; e.g. lymphocytes) and 99.5% (mixed sample with 99% normal and 1% heterozygous tumour DNA). It is difficult to perceive a clinical assay with such a low variation coefficient that this small difference can be reliably detected in sputum.

Using polymorphic DNA markers in PCR-based assays, LOH and MSI has been reported in sputum of lung cancer patients. These polymorphic DNA markers are non-informative in individuals who are homozygotic for these markers. Therefore, several markers need to be examined to cover the general population.

Four studies have been conducted on LOH and lung cancer, of which the most recent ones were published in 2007 (Arvanitis et al, 2003; Wang et al, 2003; Castagnaro et al, 2007; Hsu et al, 2007). No studies have followed since. All studies report comparable results with LOH in 26–55% in cases and 0–11% in cancer-free controls. Prevalence of MSI was low in all studies, ranging from 4 to 35% in cases and 0 to 5% in controls. Arvanitis et al (2003) tested 48 markers in sputum and bronchial washings (analysed together), in which non-cancer-specific markers were also included. Looking at informative loci, they calculated fractional allele loss values. Significant variations were observed for the markers, which may be related to non-neoplastic genetic alterations. This kind of results needs to be confirmed by others. Taking these data and the technical considerations into account, there is room for debate whether LOH and MSI by themselves are suitable as sputum biomarkers for lung cancer.

**MicroRNA.** MicroRNAs (miRNAs) are a class of small non-coding RNA molecules, which are associated with a spectrum of biological and pathological processes.

In a small feasibility study, Xie et al (2010) demonstrated that endogenous miRNAs are stably present in sputum specimens. Using real-time RT–PCR, miR-21 and miR-155 were detected, of which miR-21 was significantly overexpressed in sputum of lung cancer patients as compared with cancer-free subjects. Furthermore, elevated miR-21 expression was more sensitive (70%) than conventional sputum cytology (48%) in diagnosing lung cancer.

The same research group defined miRNA signatures for different histologic types of lung cancer in studies of similar design (Xing et al, 2010; Yu et al, 2010). Sensitivity increased when complementary miRNAs were combined in a panel as compared with single miRNAs. For the diagnosis of squamous cell lung cancer, the combination of miR-205, miR-210 and miR-708 yielded 73% sensitivity and 96% specificity. A panel consisting of miR-21, miR-200b, miR-375 and miR-486 produced 81% sensitivity and 92% specificity in discriminating sputum of lung adenocarcinoma patients from controls. The authors found no association between miRNA expression and stage of lung cancer, suggesting that the miRNA signatures can be used as a tool in the detection of early lung cancer. Overall, miRNA analysis has recently become available and more studies in sputum seem useful.

**Messenger RNA.** From practical point of view, there is a disadvantage using messenger RNA (mRNA). In contrast to miRNA (see above), mRNA is rapidly degraded in the sputum. Therefore, it is necessary to process sputum after collection as soon as possible.

Several studies investigated aberrant mRNA profiles in sputum (Jheon et al, 2004; Mecklenburg et al, 2004; Sun et al, 2009) (Table 2). Reverse-transcriptase quantitative PCR (RTQ–PCR) was more sensitive than sputum cytology (14%) and immunocytochemistry (16%). In short, two studies revealed high specificity and reasonable sensitivity (Jheon et al, 2004; Sun et al, 2009). Confirmation of these results is needed.

**Protein.** Several studies explored the presence and/or expression of tumour-related proteins in sputum of lung cancer patients and controls (Table 2). Sun et al (2009) reported significantly elevated expression of a proliferation-inducing ligand (APRIL) in sputum.
of lung cancer patients compared with controls (82% vs 3%, respectively).

Pio et al (2010) demonstrated increased levels of complement factor H in sputum of lung cancer patients, and suggested that large plasma proteins as factor H reflects hyperpermeability in tumour circulation. Factor H quantification may aid in improving sensitivity of sputum cytology for lung cancer diagnosis, but is not proof of malignancy similar to hemoptysis.

**Fluorescence in situ hybridisation.** Fluorescence in situ hybridisation (FISH) assay allows detection of chromosomal aneuxomy, rearrangements and copy number changes in interphase cells, but usually requires the cytological or automated detection of abnormal cells. Fluorescence in situ hybridisation by itself is not superior to sputum cytology, but can improve sensitivity of lung cancer detection when used in conjunction with sputum cytology or as a confirmatory test (Romero et al, 2003; Katz et al, 2008). Li et al (2007) showed that FISH analysis of both HYAL2 and FHIT deletions was more sensitive than cytology alone (sensitivity: 76%; specificity: 92%). Kettunen et al (2006) did not find significant differences in copy number gain between high-risk subjects and healthy never-smokers, indicating that copy number gain is not useful as a risk marker. Liu et al (2008) used enrichment procedure based on anti-CD14 and anti-CD16 antibody beads before FISH and cytology. However, sensitivity of FISH and cytology results remained comparable (58% vs 53%).

No internationally standardised method exists for cytometry (Thunnissen et al, 1996). So far, the data are useful for analysis on group level, but its relevance is questionable for the individual patient.

**Other markers.** Free DNA exists in higher concentration in the serum of lung cancer patients than in the serum of controls (Sozzi et al, 2003). Van der Drift et al (2008) found that the amount of free DNA in sputum was related to severity of inflammation, but not in the presence of lung cancer.

In a small study, sequence variants in mitochondrial DNA (mtDNA) were investigated in specimens (no sputum) of lung, bladder and kidney cancer patients, and sputum from 12 cancer-free heavy smokers (Jakupciak et al, 2008). Tumours were found to contain significantly more mtDNA mutations compared with matched body fluids and blood, and sputum of controls. Biological relevance of mitochondrial mutations yet needs to be clarified.

Fourier transform infrared (FTIR) spectroscopy is a non-invasive method that visualises biochemical changes in sputum by determination of absorbance levels of infrared wavenumbers. In a small feasibility study, Lewis et al (2010) reported that a panel of wavenumbers was able to distinguish cancer sputum from healthy control sputum. Fourier transform infrared might have the potential as a high-throughput method for screening.

Black matter deposition (anthracosis) was assessed in sputum by Konno et al (2004), next to DNA hypermethylation (Supplementary Table 1). Mean anthracotic index of lung cancer patients was significantly higher than that in controls and might thus be suitable for identifying a population at risk for lung cancer development. Remarkably, this index was not correlated with smoking or with detection of lung cancer cells in the sputum samples.

**CONCLUSION**

Ten years of additional research on the use of sputum in risk assessment or the early detection of lung cancer has brought new insights and more advanced molecular techniques. Polymerase chain reaction-based assays made detection of low fraction mutations feasible in sputum, although one has to be cautious for false-positivity induced by high number of PCR cycles. More biomarkers have been identified in sputum, such as DNA hypermethylation markers, miRNAs and tumour-related proteins, which show the potential for screening purposes. A rational for the distinction of a risk from a diagnostic marker was provided.

Although in recent years many markers have been examined in sputum, they are currently not sufficiently validated for clinical application. These studies, comparing sensitivity and specificity of cytology with molecular analysis, respecting technical limitations, should be reported in future studies.

**ACKNOWLEDGEMENTS**

This work was supported by Dutch Cancer Society (KWF 2008–4220).

**REFERENCES**

Ammanagi AS, Dombale MD, Miskin AT, Dandagi GL, Sangoli S.S. (2012) Sputum cytology in suspected cases of carcinoma of lung (Sputum cytology a poor man’s bronchoscopy?). Lung India 29: 19–23.

Arvanitis DA, Papadakis E, Zafiropoulos A, Spandidos DA (2003) Fractional allelic loss is a valuable marker for human lung cancer detection in sputum. Lung Cancer 40: 55–66.

Baryshnikova E, Destro A, Infante MV, Cavuto S, Cariboni U, Allossio 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.

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 JH, 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 Res 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. Proc Natl Acad Sci USA 95: 11891–11896.

Boldrini L, Gissifredi S, Ursino S, Camacci T, Baldini E, Meli F, Fontanini G (2007) Mutational analysis in cytological specimens of advanced lung adenocarcinoma: a sensitive method for molecular diagnosis. J Thorac Oncol 2: 1086–1090.

Castagnaro A, Marangio E, Verduri A, Chetta A, D’Ippolito R, Del Donno M, Olivieri D, Di Cola G (2007) Microsatellite analysis of induced sputum DNA in patients with lung cancer in heavy smokers and in healthy subjects. Exp Lung Res 33: 289–301.

Chen J-T, Chen Y-C, Wang Y-C, Tseng R-C, Chen C-Y, Wang Y-C (2002) Alterations of the p16(INK4a) gene in resected nonsmall cell lung tumors and exfoliated cells within sputum. Int J Cancer 98: 724–731.

Ciricincione R, Lintas C, Conte D, Mariani L, Roz L, Vignola AM, Pastorino U, Sozzi G (2006) Methylation profile in tumor and sputum samples of lung cancer patients detected by spiral computed tomography: a nested case–control study. Int J Cancer 118: 1248–1253.

Destro A, Bianchi P, Allossio M, Laghi L, Di Gioia S, Malesci A, Cariboni U, Gribaudi G, Bulfamante G, Marchetti A, Bosari S, Infante M, Ravasi G, Roncalli M (2004) K-ras and p16(INK4a) alterations in sputum of NSCLC patients and in heavy asymptomatic chronic smokers. Lung Cancer 44: 23–32.

Durskalski VL, Palesch YY, Lipsitz SR, Rust PF (2003) Analysis of clustered matched-pair data. Stat Med 22: 2417–2428.

Eckert KA, Kunkel TA (1991) DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl 1: 17–24.
Ferlay J, Shin H-R, Bray F, Forman D, Mathers C, Parkin DM (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127: 2893–2917.

Field JK, Liloglou T, Niaz A, Bryan J, Gosney JR, Giles T, Brambilla C, Brumbilla E, Vesin A, Timsit J-F, Hainaut P, Martinet Y, Vignaud JM, Thunnissen FB, Prinzen C, Snijders PJ, Smit EF, Sozzi G, Roz L, Ibach A, Becker HD, Elbourn JS, Magee ND, Montuenga LM, Pajares MJ, Lozano MD, O’Byrne KJ, Harrison DJ, Niklinski J, Cassidy A (2009) EUEL project: a multi-centre, multipurpose study to investigate early stage NSCLC, and to establish a biobank for ongoing collaboration. Eur Respir J 34: 1477–1486.

Greenblatt MS, Bennett WP, Hollstein M, Harris CC (1994) Mutations in the K-ras proto-oncogene in human pancreatic and other epithelial cancers. Cancer Res 54: 4855–4878.

Guzman L, Depix MS, Salinas AM, Roldan R, Aguayo F, Silva A, Vinet R (2012) Analysis of aberrant methylation on promoter sequences of tumor suppressor genes and total DNA in sputum samples: a promising tool for early detection of COPD and lung cancer in smokers. Diagn Pathol 7: 87.

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646–674.

Hansen KD, Timp W, Bravo HC, Sabuncyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Brien E, Zhang K, Irizarry RA, Feinberg AP (2011) Increased methylation variation in epigenetic domains across cancer types. Nat Genet 43: 768–775.

Hillas G, Moschos C, Dimakou K, Vlastos F, Avgieropoulou S, Christakopoulou I, Rasidakis A, Bakakos P (2008) Carcinomembrine antigen, neuron-specific enolase and cytokeratin fragment 19 (CYFRA 21-1) levels in induced sputum of lung cancer patients. Lab Invest 88: 542–547.

Hirsch FR, Franklin WA, Gazdar AF, Bunn PA (2001) Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. Clin Cancer Res 7: 5–22.

Hsu H-S, Chen T-P, Wen C-K, Hung C-H, Chen C-Y, Chen J-T, Wang Y-C (2007) Multiple genetic and epigenetic biomarkers for lung cancer detection in cytologically negative sputum and a nested case–control study for risk assessment. J Pathol 213: 412–419.

Hwang S-H, Kim KU, Kim J-E, Kim H-M, Lee MK, Lee CH, Oh T, An S (2011) Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. Clin Chem Lab Med 49: 699–704.

Jakupciak JP, Maragh S, Markowitz ME, Greenberg AK, Hoque MO, Malitra A, Barker PE, Wagner PD, Rom WN, Srivastava S, Sidransky D, O’Connell CD (2008) Performance of mitochondrial DNA mutations detecting early stage cancer. BMC Cancer 8: 285.

Jhon S, Hyun D-S, Lee D-S, Lee Y-S, Yoon G-S, Jeon C-H, Park J-W, Park C-K, Jung M-H, Lee K-D, Chang H-K (2004) Lung cancer detection by a RT-nested PCR using MAGE A1–6 common primers. Lung Cancer 43: 29–37.

Kalomenidis I, Dimakou K, Kolintza A, Vlami K, Papadakis M, Sotiropoulou E, Pasrija T, Srinivasan R, Behera D, Majumdar S (2007) Telomerase activity in lung cancer patients using laser capture microdissection microscope and nested PCR using MAGE A1–6 common primers. Mod Pathol 20: 1232–1239.

Kettunen E, Salmenvikki K, Vuopala K, Toljamo T, Kuosma E, Norppa H, Knuttila S, Kaleda S, Huuskonen MS, Arstila S (2002) Copy number gains on 5p15, 6p11–q11, 7p12, and 8q24 are rare in sputum cells of individuals at high risk of lung cancer. Lung Cancer 54: 169–176.

Konno S, Morishita Y, Fukasawa M, Shu Y, Wang D, Tanaka R, Minami Y, Iijima T, Noguchi M (2004) Anticancer antibody and DNA methylation status of sputum contents can be used for identifying the population at risk of lung carcinoma. Cancer 102: 348–354.

Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou S-HI, Dezhube BJ, Janne PA, Costa DR, Varella-Garcia M, Kim W-H, Lynch TJ, Fidias P, Stubbs H, Engelmann JA, Soquist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Seliga R, Shapiro GL, Clark JW, Iafrate AJ (2010) Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med 363: 1693–1703.

Leng S, Do K, Yingling CM, Picci Ma, Wolf HJ, Kennedy TC, Feser WJ, Baron AE, Franklin WA, Brock MW, Herman JG, Baylin SB, Byers T, Stidley CA, Belinsky S (2012) Defining a gene promoter methylation signature in sputum for lung cancer risk assessment. Clin Cancer Res 18: 3387–3395.

Lewis PD, Lewis KE, Ghosal R, Baillie S, Lloyd AJ, Wills J, Godfrey R, Kloor P, Mur LAI (2010) Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. BMC Cancer 10: 640.

Li R, Todd NW, Qiu F, Fan T, Zhao YY, Rodgers WH, Fang H-B, Katz RL, Stass SS, 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.

Liu Y, An Q, Li L, Zhang D, Huang J, Feng X, Cheng S, Gao Y (2003) Hypermethylation of p16INK4a in Chinese lung cancer patients: biological and clinical implications. Carcinogenesis 24: 1897–1901.

Liu Y, Lan Q, Shen M, Jin J, Mumford J, Ren D, Keohavong P (2008) Aberrant gene promoter methylation in sputum from individuals exposed to smoky coal emissions. Anticancer Res 28: 2061–2066.

Martinet N, Alla F, Farray G, Labit B, Drouot H, Vidali R, Picard E, Gaube MP, Le Faoou D, Siat J, Borelly J, Vannymen P, Bazarbachi T, Vignaud JM, Martinet Y (2000) Retinoic acid receptor and retinoid X receptor alterations in lung cancer precursor lesions. Cancer Res 60: 2869–2875.

McKeilburg I, Stratakis DF, Huber RM, Haussinger K, Morresi-Hauf A, Riethmueller G, Kufer P (2004) Detection of melanoma antigen-A expression in sputum and bronchial lavage fluid of patients with lung cancer. Chest 125: 1645–1665.

Melamed MR, Fiehinger BJ, Zaman MB, Heelan RT, Perchick WA, Martini N (1984) Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York. Chest 86: 44–53.

National Lung Screening Trial Research Team: Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatzonis C, Marcus PM, Sicks J (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. New Engl J Med 365: 395–409.

Olausen KA, Soria JC, Park YW, Kim KJ, Kim SH, Ro JY, Andrè F, Jang SJ (2005) Assessing abnormal gene promoter methylation in paraffin-embedded sputum from patients with NSCLC. Eur J Cancer 41: 2112–2119.

Palacios 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 Epidemiol Biomarkers Prev 9: 303–308.

Pasrija T, Srinivasan R, Behera D, Majumdar S (2007) Telomerase activity in sputum and telomerase and its components in biopsies of advanced lung cancer. Eur J Cancer 43: 1476–1482.

Patz EF, Rossi S, Harpole DH, Herndon JE, Goodman PC (2000) Correlation of tumor size and survival in patients with stage IA non-small cell lung cancer. Chest 117: 1568–1571.

Pio R, Garcia J, Corrales L, Ajona D, Fleischhammer M, Pajares MJ, Cardenal F, Seijo L, Zulueta JJ, Nadal E, Witt C,Lozano MD, Schmidt B, Montuenga LM (2010) Complement factor H is elevated in bronchoalveolar lavage fluid and sputum from patients with lung cancer. Cancer Epidemiol Biomarkers Prev 19: 2665–2672.

Qiu F, Todd NW, Li R, Peng H, Liu Z, Yantis HG, Katz RL, Stass SS, Jiang F (2008) Magnetic enrichment of bronchial epithelial cells from sputum for lung cancer diagnosis. Cancer 114: 275–283.

Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58: 982–990.
Rivera MP, Mehta AC (2007) Initial diagnosis of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 132: 1315–148S.

Risco E, Arribas L, Esteban A, Buyse M, Keynton H, Guinot P, Kurieme T, Ishikawa Y, Tanaka Y, Tanaka T, Inamura K, Togashi Y, Ito Y, Sakamoto H, Yoshida T, Tateishi U, Nokihara H, Yamamoto N, Sekine I, Kunitoh H, Matsuoso Y, Furuta K, Tamura T (2007) Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. Clin Cancer Res 13: 5385–5390.

Risse EK, Van 't Hof MA, Laurini RN, Vooijs PG (2003) Chromosomal abnormalities in non-small cell lung carcinomas and in bronchial epithelia of high-risk smokers detected by multi-target interphase fluorescence in situ hybridization. J Mol Diagn 5: 103–112.

Selamat SA, Galler JS, Joshi AD, Fyfe MN, Campan M, Siegmund KD, Romeo MS, Sokolova IA, Morrison LE, Zeng C, Baro´n AE, Hirsch FR, Miller SH, Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7: 169–181.

Selamat SA, Galler JS, Joshi AD, Fyfe MN, Campan M, Siegmund KD, Romeo MS, Sokolova IA, Morrison LE, Zeng C, Baro´n AE, Hirsch FR, Miller SH, Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7: 169–181.

Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Shin K-C, Lee K-H, Lee C-H, Shin I-H, Suh H-S, Jeon C-H (2012) MAGE-A1 and -A3: a potential panel of microRNA markers for early detection of lung cancer. Lung Cancer 77: 286–291.

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 gene expression panel for quantitative PCR for lung cancers. Cancer Lett 247: 56–71.

Siegel R, Naishadham D, Jemal A (2012) Cancer statistics, 2012. CA Cancer J Clin 62: 10–29.

Soda M, Isobe K, Inoue A, Maemondo M, Oizumi S, Fujita Y, Gemma A, Yamashita Y, Ueno T, Takeuchi K, Choi YL, Miyazawa H, Tanaka T, Hagiwara K, Mano H (2012) A prospective PCR-based screening for the EML4-ALK oncogene in non-small cell lung cancer. Clin Cancer Res 18: 5682–5689.

Somers VA, Pietersen AM, Theunissen PH, Thunnissen FB (1998) Detection of K-ras point mutations in sputum from patients with adenocarcinoma of the lung by point-EXACCT. JClin Oncol 16: 3061–3068.

Suzuki M, Wistuba II, Miyazawa H, Ikebuchi K, Morita S, Kobayashi K, Hagiwara K, Mano H (2012) A prospective PCR-based screening for the EML4-ALK oncogene in non-small cell lung cancer. Clin Cancer Res 18: 5682–5689.

Sun B, Wang H, Wang X, Huang H, Ding W, Jing R, Shi G, Zhu L (2009) A proliferation-inducing ligand: a new biomarker for non-small cell lung cancer. Exp Lung Res 35: 486–500.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)