Conventional sputum cytology can be used for the detection of lung cancer, but has shown a low yield in prospective screening trials. This review focuses on the technical aspects relevant to the outcome of DNA and image analysis in sputum. Published articles are discussed in the light of the technical background. Recent developments in DNA analysis and nuclear image analysis show a clear potential to improve or refine diagnosis beyond that achieved with conventional sputum cytology examination. The challenge for future studies in DNA and nuclear analysis of sputum is to ensure high levels of quality control and to confirm these initial encouraging results.

The aim of this paper is to provide a state of the art critical review of sputum analysis and its role in the detection of lung cancer. This review is based on an appraisal of the peer reviewed literature published up to January 2003 on this topic. The paper is divided into two sections. The first deals with the analysis of whole cells by means of routine cytological examination, nuclear image analysis, and protein analysis. The second concerns the examination of extracted nucleic acids, namely: DNA and RNA.

The molecular events that occur during lung carcinogenesis have been reviewed recently.1–4 The presence and persistence of specific molecular changes in the bronchial epithelium are considered to be high risk markers in exposed patients and predictive for invasion.5–7 Squamous metaplasia and p53 expression in bronchial epithelium is smoking related.8–10 Severe dysplasia and carcinoma in situ are associated with mutations in p53. Recently, the natural course of carcinoma in situ was reported to involve rapid progressive to invasion.11–12 These observations have increased interest in the role of the analysis of such markers in routine clinical practice, and are leading to critical assessment of laboratory methodology to determine which markers can reliably be detected in sputum samples.

CYTOLOGICAL EXAMINATION

Many years ago, Saccomanno et al defined the cytological changes that occur during the development of lung cancer.13 These changes were mainly documented in squamous metaplastic cells, and represent cellular aspects of toxic damage of respiratory tract epithelium as a result of—for example, smoking or radon gas. The transition from mild, moderate, and marked atypia, to carcinoma in situ and then to invasive carcinoma has been described. The transition time varies considerably between patients, but on average the transition from mild to marked atypia takes five years, and the change from moderate atypia to carcinoma in situ and from marked atypia to invasive carcinoma takes an additional five years. Interestingly, patients developing squamous cell and small cell carcinoma show the same cytological changes. The point of development into one of the different types of carcinoma is unclear, but the transition time seems to be slightly shorter for patients developing small cell lung cancer.15–17

Sputum can either be induced or collected spontaneously. Three day pooled sputum increases the chance of detection.14 Saccomanno’s fixative (50% ethyl alcohol and 2% carbowax) is recommended for collection, transport, and fixation.15 In a crossover study comparing induced and spontaneous sputum a learning phase was shown. The second technique used for collection yielded the best results.16 However, for peripheral cancer (not visible endoscopically) induced sputum is most informative.17

A sputum sample is considered representative if alveolar macrophages or bronchial epithelial cells are present because this shows that the samples originate from deep within the lung.18–20 The minimum number of macrophages reported varies from 150 to recently at least five.21

Usually, a minor part of the sputum sample is analysed cytologically for the presence of cancerous cells, and only a small proportion of cells (< 1%) are tumour cells. In previous reviews of sputum examination in lung cancer, an average sensitivity of 65% (range, 22–98%) was shown.19,22,23 The chance of detecting abnormal cells increased with: (1) centrally located tumours; (2) large and/or high stage tumours; (3) squamous cell carcinomas rather than adenocarcinomas (probably because squamous cell carcinomas are frequently more centrally located); and (4) increasing number of sputum samples.24–25

“A sputum sample is considered representative if alveolar macrophages or bronchial epithelial cells are present because this shows that the samples originate from deep within the lung”

Sputum cytology plays a limited role in prospective lung cancer screening studies.24–27

Keywords: DNA; image analysis; nuclear image analysis; cytology; sputum; lung cancer; DNA analysis; nuclear image analysis; cytological examination.25

Abbreviations: LOH, loss of heterozygosity; MSP, methylation specific polymerase chain reaction; PCR, polymerase chain reaction; RT, reverse transcription.
Malignant cells were present in the sputum of 11 of the 39 lung cancer cases detected (5226 sputum examinations, 11 squamous cell carcinomas and an additional 14 cases with marked atypia). However, compared with $x$ ray screening, sputum cytology was associated with a higher chance of detecting tumours at early stages. Similar figures were found in a recent Japanese screening sputum study: eight of 36 cases were also detected in the sputum, and four of these were detected by sputum only (all squamous cell carcinomas). In an older study, probably using less up to date $x$ ray technology than is available today, sputum cytology preceded the radiological diagnosis by 18–36 months. For the limited group of patients where lung cancer was detected by sputum cytology, the five year survival rate was 80%, and the resectability was 83%, suggesting that cancers detected by sputum cytology were either detected early or were slower growing. Nevertheless, during the last century the mortality rate in screening studies ($x$ ray and cytology) did not improve.

In a recent SPRORE programme study, with a high risk cohort of patients with chronic obstructive pulmonary disease and > 40 pack years, 1.7% of the subjects had invasive carcinoma or carcinoma in situ. Moderate dysplasia was present in 25% of the subjects. Carcinoma was detected in 9% of these patients after fluorescent bronchoscopy, leading to a total 3.9% prevalence in this high risk group. This emphasises once more the usefulness of sputum cytology abnormalities as markers for lung cancer development in a population at high risk for lung cancer.

NUCLEAR IMAGE ANALYSIS

Nuclear image analysis is a relatively old approach that has been used to distinguish between small cell and non-small cell lung cancer. Basically, nuclei are stained stochiometrically with a Feulgen reaction (in which there is a linear relationship between the degree of staining and the amount of DNA) suitable for image acquisition and digitisation of the chromatin pattern (euchromatin and heterochromatin). Each pixel of the image has a certain value that relates to DNA density. Mathematical calculations on the distribution of the pixel values describe certain characteristics, such as “chromatin clumping”, “chromatin density”, and “homogeneity”. When this technique is used for the determination of possible differences between normal and (pre)malignant changes, the term malignancy associated changes is used.

Recently, the group of Palic et al performed malignancy associated change analysis on non-malignant cells obtained from buccal smears and sputum. Thus, possible changes in these cells are determined and related as surrogate markers for the presence of lung cancer or a high chance for the development of lung cancer. For adenocarcinoma, a 60% sensitivity and a 90% specificity for the prediction of malignancy was reported. For stage 0 and 1 lung cancer, a sensitivity of 45% and a specificity of 90% can be reached, compared with 14% for conventional cytology at a 99% specificity. In another study, Marek et al performed a multi-institutional analysis in five countries, using semi-automated sputum cytometry, and found a sensitivity of 75% and a specificity of 98%. The findings of these two groups were similar, so that this type of analysis might be useful for lung cancer screening in the future. However, confirmation of these studies in an independent test set, where the technical procedure is kept identical, is necessary.

PROTEINS

The measurement of a heterogeneous nuclear riboprotein, A2/B1, appears to be promising for the detection of early lung cancer. The reactivity of heterogeneous nuclear riboprotein B1 was classified as strong in the sputum of patients with lung, oral, and oesophageal squamous cancer. However, the affinity of the antibody raised to this protein is not high enough for immediate large scale application and awaits further refinement.

p53 has been demonstrated in squamous cells in sputum with concomitant low molecular weight cytokeratins, a pattern frequently found in lung cancer. p53 was demonstrated in an immunofluorescence assay with repeated incubation of the second step in nine of 16 patients with lung cancer, 11 of 25 in coal exposed individuals, and in 0 of 17 controls.

The presence of guanidinobenzoatase on the cell surface was first associated with malignancy in the early 1980s, and in 1998 interest in this molecule was renewed. It has been suggested that this cell surface protease is induced on the surface of mature epithelial cells (“surrogate marker”) in patients with lung cancer.

Overall protein analysis in sputum may be a practical approach for early lung cancer detection. Heterogeneous nuclear riboprotein seems to be the best candidate, but better antibodies are needed before this approach can be put into practice.

DNA EXAMINATION

The target cells for early detection in lung cancer are tumour cells, but they comprise only a minor fraction of the sputum sample (usually less than 1%). In addition to the vital cell component, it is assumed that dead cells are shed into the surrounding epithelial lining fluid, thus providing sufficient DNA for detection. The amount of dead cells and/or naked DNA is not known, but again it is probably only a minor component.

The relatively low amount of tumour DNA (< 1%, see above) in a background of wild-type DNA imposes specific requirements on the detection techniques to be used. Because techniques such as sequencing or mutation specific oligonucleotide hybridisation have a technical sensitivity of around 10–20% these cannot be used on isolated crude DNA. Ideally, internal and external negative and positive controls should be incorporated. The internal positive control provides information about the maximum signal and its related target frequency, whereas the external positive control(s) provide information about the daily variation. The internal negative control(s) provide a threshold above which a test signal is positive. The external negative control(s) monitor possible contamination, as a result of the infidelity of Taq polymerase. They should be used from the start of the test onwards and also in the nested polymerase chain reaction (PCR), if used. In addition, with any PCR an error may be introduced during amplification. The chance of this error occurring is approximately 1/8000 and may lead to a false positive test result. This PCR bias may be circumvented by repeating the whole procedure in those cases with a positive outcome in the first test, starting with the isolated DNA. If the second test is also positive, then this sample should be called positive for mutation, otherwise it should be considered negative. In this way, the chance of a sample being falsely labelled as positive because of a PCR error is less than $64 \times 10^{-6}$. To estimate the technical sensitivity, a model system with low copy numbers may be obtained in three ways, namely: (1) dilution of the PCR product, (2) mixing mutant DNA with increasing amounts of wild-type DNA, and (3) mixing vital mutant cells with wild-type cells. This last option is preferable because it approaches the conditions found in biological samples like sputum.

In cases where crude DNA or RNA is not used because of a low target fraction, cytological analysis may be done first to identify the specific cells of interest so that subsequently
molecular analysis can be performed. This last approach is worthwhile for the detection of abnormalities, but not practical for early detection.

‘The relatively low amount of tumour DNA in a background of wild-type DNA imposes specific requirements on the detection techniques to be used’

K-ras and methylation have been reported relatively frequently in sputum, whereas the demonstration of p53 and loss of heterozygosity (LOH) abnormalities, DNA adducts, and RNA is still limited. Therefore, the main emphasis will be on the first two approaches.

**K-ras**

In general, K-ras mutations occur in adenocarcinomas. Table 1 summarises the articles describing K-ras mutation analysis in sputum along the general outlines described above. In general, it appears that (1) when sputum from K-ras mutated adenocarcinomas is examined, the sputum may be positive (showing the same mutation) or negative; (2) a surprisingly high proportion of K-ras mutation positive sputum samples has been reported in squamous cell carcinomas; and (3) K-ras mutation positive sputum samples have been found in patients without lung carcinoma. These last two points are at variance with the biological phenomena known in lung cancer: K-ras is rarely detected in squamous cell carcinoma of the lung. This may be explained by the assumption that K-ras mutations are present in non-malignant sputum cells or alternatively that there is a technical explanation for the observations. Studies on adenocarcinomas used different techniques to those on squamous cell carcinomas and non-cancerous tissues. A possible technical explanation may be that with allele specific amplification, mismatched alleles amplify 10^-2 to 10^-4 less than perfectly matched templates. This implies that when more cycles are performed with PCR the chance of false positivity increases proportionally. This phenomenon, which has been called ‘breakthrough’, implies that with enough cycles of amplification even negative samples will give rise to a positive signal. The use of PNA clamping reduces this chance.

Overall, of the three techniques reporting results consistent with the biology of lung cancer, the cloning approach is impractical for large scale lung cancer screening. Therefore, peptic nucleic acid-PCR-restriction fragment length polymorphism and Point-EXACCT are the methods of choice.

The question of whether K-ras mutations are present in non-malignant epithelial cells is influenced in the same way as sputum samples, because a similar variety of techniques has been used. Based on techniques used to test sputum from K-ras mutated adenocarcinomas, the DNA target fraction is around 1/150 (range, 1/50–3000) in sputum. Importantly, K-ras mutations may be detected in sputum more than one year before clinical diagnosis (range, one to 46 months). This time period may provide an opportunity for early detection. In addition, mutations were found in cases where sputum cytology was negative.

**Methylation**

DNA methylation at the 5’ position of the pyrimidine ring is an important epigenetic alteration in eukaryotes. 5-Methylcytosine is found mainly in CpG islands and, when found in the promoter region of genes, the binding of transcriptional factors is altered and other proteins such as methyl-DNA binding proteins are able to bind, resulting in gene silencing. The different methods used to demonstrate methylation have recently been reviewed by Fraga and Esteller. Methylation specific PCR (MSP) is a sensitive method.

Over 40 genes involved in methylation in lung cancer have been reviewed by Tsou et al. Recently, several reports on methylation analysis in sputum have been published. Between 55% and 100% of sputum samples from patients with lung cancer were methylated positive, whereas for controls the figure varied from 3% to 35%. This range seems to be gene dependent.

‘5-Methylcytosine is found mainly in CpG islands and, when found in the promoter region of genes, the binding of transcriptional factors is altered and other proteins such as methyl-DNA binding proteins are able to bind, resulting in gene silencing’

The following technical considerations are useful for a better understanding of the results: a nested PCR is performed (two sets of 40 cycles), where the first PCR has primer binding sites outside the CpG islands and the second uses PCR primers that are methylation specific. Bisulfite modification is a step that requires time for the modification of cytosine to uracil. However, if the modification is incomplete, then unmethylated cytosine will still be present in the template DNA, which can be detected with methylated primers in MSP, leading to a (false) positive test result. A control for the bisulfite modification may be sequencing of

---

Table 1: Detailed technical analysis of different methods used for K-ras detection in sputum

<table>
<thead>
<tr>
<th>Method</th>
<th>Sens.</th>
<th>Int. pos. ctrl</th>
<th>Int. neg. ctrl</th>
<th>Ext. pos. ctrl</th>
<th>Ext. neg. ctrl</th>
<th>Cancer</th>
<th>Pos. neg. tumour</th>
<th>Pos. non-cancer/mutation</th>
<th>R test</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR cloning ASO*</td>
<td>nm</td>
<td>nm</td>
<td>Yes</td>
<td>Yes</td>
<td>8/10</td>
<td>0/5</td>
<td></td>
<td></td>
<td>No</td>
<td>54</td>
</tr>
<tr>
<td>PCR cloning ASO†</td>
<td>nm</td>
<td>nm</td>
<td>Yes</td>
<td>Yes</td>
<td>0/12</td>
<td>–</td>
<td></td>
<td></td>
<td>–</td>
<td>54</td>
</tr>
<tr>
<td>Enriched-PCR‡</td>
<td>np</td>
<td>np</td>
<td>Yes</td>
<td>Yes</td>
<td>18/37</td>
<td>5/40</td>
<td></td>
<td></td>
<td>yes</td>
<td>56</td>
</tr>
<tr>
<td>MASAg</td>
<td>np</td>
<td>np</td>
<td>Yes</td>
<td>Yes</td>
<td>18/100</td>
<td>3/15</td>
<td></td>
<td></td>
<td>No</td>
<td>56</td>
</tr>
<tr>
<td>PNA-PCR-RFUP*</td>
<td>1/1000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5/51</td>
<td>0/25</td>
<td></td>
<td></td>
<td>–</td>
<td>57</td>
</tr>
<tr>
<td>Point-EXACCT**</td>
<td>1/15 000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>5/11</td>
<td>0/11</td>
<td></td>
<td></td>
<td>Yes</td>
<td>58</td>
</tr>
</tbody>
</table>

*ASO, allele specific oligonucleotide hybridisation; case selection: only K-ras positive adenocarcinomas; threshold for number of spots to be called positive not mentioned. †Sputum samples with PAP III–V, no direct relation to K-ras positive cancer known. ‡Twenty cycles of PCR, then digestion with BstN1 for a cut in wild-type alleles and subsequently another ‘x’ number of cycles; all squamous cell carcinomas. §Repeated samples performed three times, no further details, one first analysis requires 2 µg DNA. MA: MutaScreen; mutation specific amplification; PCR 1 x 32 cycles; Positive cancer cases include 11/59 adenocarcinomas and 5/32 squamous cell carcinoma (SCC) cases; 5 cases in sputum without K-ras positivity of primary tumour; Int. neg. ctrl nm implies not possible because a mixture of mutant primers was used. ¶From primary tumour K-ras status not known; 4/22 SCC, 1 adenocarcinoma, 2 large cell carcinoma cases. Quote: use of 2 PCR steps with PNA-PCR-RFUP instead of 3 of PCR-RFUP reduces Taq polymerase borne errors, 54 cycles; concordant with washing sample of same patient. **Paraffin wax embedded sputum from patients with adenocarcinoma: 5/11 cases with K-ras positive and 0/11 cases with K-ras negative tumours. *Sens., minimum target frequency (also called sensitivity of the technique); Int. pos. ctrl, internal positive control; Int. neg. ctrl, internal negative control; Ext. pos. ctrl, external positive control; Ext. neg. ctrl, external negative control; Pos. cancer, number and percentage of positive lung cancer cases; Pos. non-cancer, number and percentage of positive cases in patients without lung cancer; R test, test repeated?; nm, not mentioned; np, not possible; –, not done.
the PCR product. If cytosines can still be detected by sequencing, modification is incomplete. This does not necessarily need to be a cytosine at a CpG island, but if this is the case, MSP may give rise to false positive signals at CpG sites. Thus, total conversion is crucial to the success of the analysis. Recently, Grunau et al described crucial experimental parameters for genomic bisulfite sequencing. They mention that, depending on the conditions, for each 1000 cytosines a few will not be deaminated to uracil. When performing a total of 80 PCR cycles, the chance is very high that even one unconverted cytosine in a CpG may be detected. As a control for the presence of methylation a restriction enzyme that digests after MSP has been reported, but this control does not take incomplete bisulfite modification into account. Occasional sequencing of a random selection of PCR products after the first PCR cycle may reveal insight in the conversion process. In addition, external negative controls that react with the same primers should be performed during each experiment. The level (number of cycles or intensity of product) where this control becomes positive is the threshold for detection. In addition to the problem of incomplete cytosine conversion, the possible in fidelity of Taq polymerase can also cause false positives in methylation specific PCR.

An alternative to bisulfite modification is digestion with a restriction enzyme, as used by Chen et al. As a check for incomplete digestion, they repeated the digestion procedure before 31 cycles of amplification and ran digested and control DNA on an agarose gel. A ratio of > 0.5 was determined to be the threshold for aberrant methylation. Using this threshold, 13 of 21 sputum samples showed methylation, with matched specimens of lung cancer tissue showing concordance in 11 of 12 of the cases. This study awaits confirmation.

Overall, it is still unclear whether the presence of DNA methylation in sputum is a true biological phenomenon, or partly the result of the technical procedures used. The two different technical approaches for methylation detection should be compared in a crossover study.

p53
Mutation detection for three hot spots in the p53 gene was performed on DNA from sputum with the mutation allelic specific amplification approach (table 1). Sputum samples from four of 51 patients with lung carcinomas and one of 25 smokers were positive. Another paper describes the use of reverse transcription (RT)-PCR on RNA from archival sputum samples in 10 patients in whom the primary tumour showed p53 mutation. One of the 10 samples showed a mutation in a sputum sample 84 days before diagnosis. The number of cases and procedures used are too limited for any strong conclusions. Overall, p53 mutation detection in sputum may be an option for early lung cancer detection; however, the technique is laborious to perform and is therefore not a practical possibility.

LOH
To date, microsatellite alterations have been detected in specific cells scraped from a sputum sample: LOH was detected in the cytologically abnormal cells. This approach shows that chromosomal abnormalities are present in sputum. However, this approach is not practical for screening, unless these abnormalities in sputum cells with metaplasia denote an essentially higher chance for lung cancer development than in metaplasia alone.

DNA adducts
It is possible to determine lipophilic DNA adducts in the sputum of smokers. The association of these adducts with risk for the development of lung cancer is not yet known, but the method is laborious and impractical for large scale screening.

RNA examination
The preprogastrin releasing peptide was one of seven neuroendocrine marker genes analysed by means of RT-PCR that showed differences in expression between normals and controls. This is noteworthy because five of 23 sputum samples from patients with small cell lung cancer were positive, whereas most of the other nucleic acid markers are specific for non-small cell lung carcinoma. Working with RNA requires specific precautions to avoid the breakdown of RNA and this can be costly. In theory, RNA analysis may be useful in the detection of lung cancer, although these tests are still in the early days of development.

CONCLUSIONS
Sputum cytology examination is useful for early detection in populations at high risk for lung cancer. In addition, DNA analysis and nuclear image analysis show clear potential to improve or refine diagnosis beyond the use of conventional sputum cytology examination. When technical quality control is performed between different laboratories and the reproducibility and robustness of the techniques have been demonstrated, a major improvement in early lung cancer detection could be achieved. To this end, testing of relatively straightforward samples may reveal opportunities for technical clarification. The challenge for both DNA analysis and nuclear image analysis in sputum is to ensure high degrees of quality control and to confirm these initial encouraging results.

REFERENCES
cytometry on induced sputum become a screening tool for lung cancer.


Sputum in patients with airflow obstruction and significant smoking histories. Cooperative early lung cancer study.


Cancer screening.

A community hospital program.

Adenocarcinoma.

Acta Cytol 1984;28:399–408.

Carcinoma.


Diagnosing intra-abdominal malignancy

Cytology testing of ascitic fluid may be of limited usefulness in the detection and management of malignancy. Initial assessment should instead concentrate on a thorough history with examination, which should then direct any further laboratory assessment, with cell count and differential, albumin, and culture being mandatory. Additionally serum albumin should be obtained to calculate the serum-ascites albumin gradient as an important determinate of the differential count.

A literature search and retrospective audit was conducted for a major English teaching hospital. Of 276 samples sent for assessment, 45 female and three male specimens were found to contain malignant cells. Of these, 85% (n = 41) were ovarian in origin. Of the remaining 228 specimens (from 84 male and 108 female patients), 220 were reported to have no evidence of malignancy and eight to be inadequate. In this group, the results of 35 (19%) patients were found to be false negatives, and their cancer was detected by radiological investigations.

In the management of ovarian cancer, ascitic fluid collected during the operation is important for staging and is of prognostic value. For other intra-abdominal malignancies ascites is associated with a poorer prognosis.

The study results have led to the following recommendations. Firstly, cytology samples should be sent only for women as this is the only group of patients to show a sizeable detection rate and a positive result that could alter management.

Secondly, if history and examination are not conclusive or suspicious of malignancy, ascitic and serum albumin should be measured in the initial paracentesis to aid diagnosis, rather than sending “routine” cytology.

Thirdly, early ultrasound should be undertaken in suspected malignant ascites, rather than waiting for the cytology result.

Sputum examination for early detection of lung cancer

F B J M Thunnissen

doi: 10.1136/jcp.56.11.805

Updated information and services can be found at:
http://jcp.bmj.com/content/56/11/805

These include:

**References**
This article cites 69 articles, 18 of which you can access for free at:
http://jcp.bmj.com/content/56/11/805#BIBL

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/