Detection of Lung Cancer by Automated Sputum Cytometry

Roger A. Kemp, PhD, Daniel M. Reinders, MSc, and Bojana Turic, MD

Introduction: Biomarkers may prove to be valuable tools to manage those at risk of lung cancer. Sputum analysis using DNA cytometry has shown promise, but an automated, objective sputum analysis test has yet to be developed. This study evaluated the performance characteristics of the LungSign test for lung cancer and compared them to conventional cytology.

Methods: A multicenter validation trial was conducted in which sputum specimens were prospectively collected from subjects suspected of having lung cancer during diagnostic workup. Specimens were placed on slides, DNA stained using Feulgen thionin, and analyzed using an automated cytometry-based scoring system. Smears were also prepared from the sputum specimens, stained by the Papanicolaou procedure, and analyzed using conventional cytology. LungSign scores and conventional cytology results were compared with the subject diagnoses.

Results: A total of 1235 high-risk subjects were enrolled at nine clinical sites. Of 1123 subjects included for analysis, 370 were found to have lung cancer—a 33% prevalence. The a priori selected LungSign score threshold detected 40% of all lung cancers and 35% of stage I lung cancers with 91% specificity. Test performance was statistically equivalent across cancer stages, histologic types, and localizations for 330 analyzable lung cancer subjects. LungSign receiver operating characteristic area under the curve measure for the test was 0.692. Conventional cytology detected 16% of lung cancers with 99% specificity.

Conclusions: DNA cytometry of sputum using the LungSign test detects stage I lung cancer and may provide a new tool to manage high-risk individuals.

Key Words: Early diagnosis, Lung neoplasms, Image cytometry, Cytology.

(J Thorac Oncol. 2007;2: 993–1000)
and generates a single score for each specimen. The specimen scoring function uses a 13-feature linear discriminant function incorporating seven statistical measures of properties of diploid and near-diploid cells (MAC features) and six of hyperdiploid cells. Feature selection and weight estimation was performed using a set of preclinical data collected before 2003. The preclinical results suggested that the test could achieve nearly 50% sensitivity for stage I lung cancer at 90% specificity.

PATIENTS AND METHODS

The validation trial was designed to evaluate the ability of LungSign to detect lung cancer as compared with standard clinical workup for patients suspected of having lung cancer. Patient accrual began in March 2003 with three clinical centers. Additional centers joined during the trial for a total of nine sites enrolling subjects and collecting specimens by the time accrual ended in September 2004.

The target population was patients referred for evaluation of the presence of lung cancer based on their medical history and clinical symptoms. Subjects were eligible for inclusion in the study if they met all of the following criteria: male/female aged 40 years or older; suspected of having lung cancer based on symptoms or radiographic abnormality; about to undergo more extensive diagnostic tests such as bronchoscopy, CT scan, fine-needle aspiration biopsy, surgery, etc.; capable of undergoing sputum induction. Subjects were excluded if they were unable to cough due to chest pain, had unstable angina, or had received radiation therapy to the chest or systemic chemotherapy in the past 6 months.

The recruitment goal was 200 analyzable sputum specimens from subjects confirmed to have lung cancer. This number was selected to set an acceptable standard error for the sensitivity values measured for the test. Similarly, a recruitment goal of 800 analyzable sputum specimens was set on the number of negative specimens to address test specificity estimates.

Review of the clinical protocol was completed by Institutional Review Board Services of Canada, and approval for investigational testing of the device was obtained from the Health Canada Therapeutic Products Directorate. Each clinical site obtained approval of the protocol from their local ethics review board.

Sputum specimens were collected prospectively from subjects undergoing diagnostic investigations for lung cancer. Informed consent was obtained from each subject. Specimens were collected through induction with an ultrasonic nebulizer using 3% hypertonic saline. The procedure was conducted for 15 to 30 minutes depending on the patient’s ability to produce sputum. Most induction procedures required less than 15 minutes and 99% required less than 30 minutes. No adverse events were reported during the trial. Sputum specimens were collected in specimen collection jars into which 25 ml of cytologic fixative was added. Specimens were refrigerated at 4°C for an allowed maximum of 8 days before either processing or shipping.

Specimens were processed to slides either at the clinical site according to the study protocol or at the Perceptronix laboratory in Vancouver, Canada. Slides were prepared at four sites (Slovenia, United Kingdom, Korea, Russia), and the others couriered liquid specimens directly to Vancouver. Mucolysis of the specimens was performed chemically using dithiothreitol and a mechanical shaker. A total of five slides were prepared from each specimen. Two smears were created for conventional cytological analysis, and three cytospin slides were prepared for LungSign analysis. The slides were allowed to air dry overnight before being shipped to Vancouver for staining.

The staining and analysis of all slides were done in Vancouver. Smears were stained using the Papanicolaou procedure, and cytospin slides were stained using Feulgen thionin. Slides were given blinded barcodes to ensure no linkage could be made between the results of LungSign and cytologic analysis during the course of the trial.

Site investigators established the reference diagnoses for all subjects in the trial. Each subject classified as negative had a minimum of 3 months of diagnostic follow-up. Clinical investigators submitted case report forms that included patient demographic data and relevant information used to determine the subjects’ lung cancer status. These reports were held by the trial monitor; study personnel were blinded as to the diagnostic outcomes of the patients until the conclusion of the trial.

Conventional smears were analyzed by a cytotechnologist and then by an experienced pathologist from the British Columbia Cancer Agency, who made the final determination for each case. Slides were graded into six categories: normal, benign, mild dysplasia, moderate dysplasia, severe dysplasia, or carcinoma present. A grading of severe dysplasia or higher was treated as positive by cytology in the analysis in the results section. The adequacy criterion for conventional cytology was the presence of alveolar macrophages.

Feulgen thionin–stained slides were scanned using an automated image cytometer (AcCell-Savant, Molecular Diagnostics Inc., Chicago, IL) to collect images of cell nuclei for analysis. A decision tree automatically categorized all imaged objects according to a set of photometric, morphometric, and texture features. Epithelial cell nuclei were collected and specimens were considered scorable if at least 400 nuclei were imaged. A score threshold of 5.0, which had corresponded to 90% specificity for the preclinical data, was used to identify specimens as suspicious for lung cancer.

The performance of LungSign was analyzed for the study threshold and compared with conventional cytology. Receiver operating characteristic (ROC) analysis was performed to examine the test performance for other score thresholds. Lung cancer sensitivities were broken down by cancer stage, histology, and localization. Binomial confidence limits were calculated by the Clopper-Pearson method. Tests of independence for contingency tables were conducted using the Fisher exact test. ROC area under the curve (AUC) was calculated by the trapezoidal rule and the variance estimate of the ROC AUC was estimated by the method of Delong et al.
RESULTS

A total of 1235 subjects were enrolled in the study from nine sites (Table 1). From this total, 112 were excluded from the final analysis. The reasons for exclusion were primarily the presence of other cancers (34 subjects), lack of final diagnosis by site investigators (31), and chemo/radiation treatment before induction (20). Of 1123 subjects included for analysis, 370 were diagnosed with lung cancer (33%). This prevalence was significantly higher than the expected prevalence of 20% to 25%. Figure 1 shows the flow of subjects from enrollment to the result of their diagnostic workups. Included subjects were predominantly male (69%) and were primarily current or former smokers (82%), as shown in Table 2. Among those with a smoking history, both lung cancer positives and negatives tended to be heavy smokers, with a median of 40 and 33 pack-years, respectively. There were no explicit smoking or minimum pack-year inclusion requirements in the trial. Despite this, both positives and negatives consisted mostly of smokers with a significant smoking history.

The TNM stage, histologic type, and localization of the tumors are presented in Table 3. Early-stage cases represented a significant fraction of the total number of cancers with stage I comprising 30% of tumors. Because only one stage 0 was diagnosed, it is included with stage I for subsequent analysis. Squamous cell carcinoma (SCC) and adenocarcinoma were the dominant histologic types, representing 39% and 33% of the cancers, respectively. There was an approximately even mix of central and peripheral tumors (41% and 51%, respectively). Adenocarcinoma was significantly more common ($p < 0.0001$) for peripheral cancers than for central ones, representing 44% and 19% of each localization, respectively.

Figure 2 shows the overall ROC performance of LungSign in the validation trial. The study threshold of 5.0 resulted in an empirical specificity of 91% (95% confidence interval [CI]: 89%–93%) and sensitivity of 40% (95% CI: 35%–46%). A more general measure of diagnostic performance, the ROC AUC, was found to be 0.692 (95% CI: 0.655–0.729). The ROC curve is skewed to toward higher specificities, suggesting that the test does better at identifying subjects at higher risk of lung cancer than for assuring negativity. This is directly illustrated in Figure 3, which shows the positive likelihood ratio (LR+) and negative likelihood ratio (LR–) as functions of score threshold. These curves show how the LungSign test result modifies the subject’s odds of

<table>
<thead>
<tr>
<th>TABLE 1. Clinical Sites in the Validation Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Slovenia</td>
</tr>
<tr>
<td>Russia</td>
</tr>
<tr>
<td>United Kingdom</td>
</tr>
<tr>
<td>Canada</td>
</tr>
<tr>
<td>Korea</td>
</tr>
<tr>
<td>Spain</td>
</tr>
<tr>
<td>United States</td>
</tr>
<tr>
<td>Austria</td>
</tr>
<tr>
<td>The Netherlands</td>
</tr>
</tbody>
</table>

Volunteer:
- Met inclusion criteria
- Subject consent
- Sputum induction

n=1235

Subjects included for analysis:

n=1123

Excluded subjects:

n=112

Sputum adequate for LungSign analysis:

n=986

Sputum inadequate for LungSign analysis:

n=137

Negative diagnosis:

n=656

Lung cancer diagnosis:

n=330

Negative diagnosis:

n=97

Lung cancer diagnosis:

n=40

FIGURE 1. Breakdown of participants by inclusion, specimen adequacy and diagnostic workup.
harboring lung cancer. The LR+ (Figure 3A) for a score threshold of 5.0 is 4.5 (95% CI: 3.4–6.0), increasing further with higher thresholds. In contrast, the lowest LR− (Figure 3B) is 0.41 (95% CI: 0.2–0.8), meaning that a negative test result corresponded with a reduction in the subject’s odds of disease by around a factor of two.

Whereas different thresholds can be used with LungSign, conventional cytology records the observance of cellular atypia in the specimen. Its performance is therefore represented by a single point in the ROC graph. The sensitivity of conventional cytology was 16% (95% CI: 11%–22%) with a specificity of 99.1% (95% CI: 97.6%–99.7%), which corresponds with an LR+ of 17.2 (95% CI: 6.5–46.0). LungSign sensitivity is similar to conventional cytology at a correspondingly high specificity (Figure 2). For the subset of specimens (n_pos = 194, n_neg = 383) analyzable by both methods, combining results using the logical odds ratio increased the LungSign sensitivity from 37.1% to 42.3% while decreasing specificity slightly from 94.3% to 93.5%.

The difference in adequacy criteria between LungSign and cytology led to significant differences in the adequacy rates observed in the trial. Whereas LungSign requires 400 analyzable cells, cytology required the presence of macrophages as evidence of a deep cough. Sputum specimens contained a median of 4500 cells (range, 0–10,000) and the inadequacy rate of specimens for LungSign analysis was 12% (95% CI: 10%–14%). The cytology inadequacy rate was the same for negative and lung cancer specimens at 43% (95% CI: 40%–45%) and was significantly higher than that of LungSign. Although nearly half of the sputum specimens were cytology inadequate, LungSign diagnostic performance was not statistically lower for these specimens (p = 0.195). LungSign ROC AUC for specimens considered adequate by conventional cytology was 0.71 (95% CI: 0.67–0.76), while LungSign ROC AUC for cytology-inadequate specimens was 0.66 (95% CI: 0.60–0.72). Similarly, no statistical difference in ROC AUC was seen for sites that prepared sputum slides versus those that sent liquid specimens for analysis (p = 0.567).

Table 4 shows the sensitivity by stage for LungSign using the study threshold and for conventional cytology. The number of positive subjects available for analysis by LungSign was higher because slides inadequate for cytology are often scorable by LungSign. The specimen scoring function was developed using a preclinical data set with positive cases that consisted predominantly of specimens from stage I can-
It was anticipated that this would yield a function the sensitivity of which was relatively independent of tumor stage. It can be seen that for the study threshold, LungSign was as sensitive for early-stage as late-stage cancers, with no statistically significant difference in sensitivity among tumor stages ($p = 0.49$). Conventional cytology sensitivity varied between 12% and 20%. There was no statistically significant difference in sensitivity of adequate specimens observed for conventional cytology ($p = 0.59$).

The LungSign sensitivity by histology varied between 33% for adenocarcinoma and 52% for small-cell cancer (Table 5). However, there was no statistically significant difference in sensitivity for the LungSign study threshold among the histologic classes ($p = 0.32$). This was not the case for conventional cytology, in which a difference in sensitivity among histologic classes was observed ($p = 0.015$), attributed to the low sensitivity (5%) observed for adenocarcinoma. Both detection approaches had slightly higher sensitivities for central lesions versus peripheral ones, but the differences are not statistically significant for either LungSign ($p = 0.40$) or for conventional cytology ($p = 0.33$).

To investigate the performance of LungSign for alternative score thresholds, ROC analysis was performed. Figure 4A shows the ROC performance of the test for stage 0/I versus stage II+ lung cancer. The ROC curves are comparable over the range of possible score thresholds. The cumulative distributions of scores for stage 0/I and stage II+ lung cancer specimens exhibit no statistically significant difference ($p = 0.38$ using the two-sample Kolmogorov-Smirnov test). The performance of conventional cytology is also shown on the figure.

Figure 4B shows the ROC performance of LungSign for SCC versus adenocarcinoma, the two dominant histologic types. Figure 4C shows the ROC performance for lesions designated as central versus peripheral. Both pairs of ROC curves are comparable, and the cumulative score distributions show no significant statistical difference using the two-sample Kolmogorov-Smirnov test ($p = 0.79$ and $p = 0.80$, respectively). As before conventional cytology performances are shown. The discrepancy in the sensitivity of conventional cytology for adenocarcinoma versus SCC appears notably in this figure.

**DISCUSSION**

The purpose of the validation trial was to establish LungSign performance characteristics for lung cancer detection in a very high risk population. The ROC curve in Figure 2 shows that the test provides significant discrimination for lung cancers over a wide range of score thresholds. The a priori threshold evaluated in the trial provided 40% sensitivity and 91% specificity overall, corresponding to a 4.5-fold

### TABLE 4. Test Sensitivity Results by Cancer Stage for Adequate Specimens

<table>
<thead>
<tr>
<th>Score</th>
<th>Total positive subjects</th>
<th>Sensitivity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/I</td>
<td>35</td>
<td>35% (26–45)</td>
</tr>
<tr>
<td>II</td>
<td>21</td>
<td>41% (28–56)</td>
</tr>
<tr>
<td>III</td>
<td>35</td>
<td>38% (28–49)</td>
</tr>
<tr>
<td>IV</td>
<td>33</td>
<td>46% (35–59)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Score</th>
<th>Total positive subjects</th>
<th>Sensitivity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>69</td>
<td>12% (5–22)</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>15% (5–32)</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
<td>20% (11–32)</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>13% (4–27)</td>
</tr>
</tbody>
</table>

Cancer cases with unspecified staging are not shown.
increase in the odds of lung cancer. Test sensitivity for early versus later stage disease (Figure 4A) was statistically equivalent over the whole ROC curve. We believe that this results from the LungSign development process, where 135 of 280 preclinical specimens used to develop the scoring function were stage I cancer. Because the observation of DNA aneuploidy is less common in specimens from stage I lung cancer, the test is primarily a MAC detector; systematic changes to the nuclear conformation of surrounding cells occur early in the development of the cancerous lesion and are detectable through its progression.

Although previous DNA cytometry studies have shown increasing levels of DNA aneuploidy in later stage cancers, this was not observed in the LungSign trial. Xing et al. observed an increase in the rate of 5c-ploidy exceeding cells with lung cancer stage. In our study, only 26 of the 330 lung cancer specimens contained any 5c exceeding cells, and there was no significant statistical difference in their frequency across stages. The sensitivity of the hyperdiploid component and MAC components of the LungSign score did not vary by cancer stage.

Although the presence of severe dysplasia in sputum is highly indicative of lung cancer, it is not necessarily diagnostic. Nonetheless, it was decisive in that it provided a high positive likelihood multiplier, albeit with low sensitivity. In a clinical context, both decisiveness and sensitivity are desirable; consider that even a test that produces no false-positive results may not be accepted if the number of cancers detected does not justify the cost. There is a tradeoff between being able to detect enough cancers while maintaining the decisiveness of the test outcome. An advantage of a continuous test is that it may be optimized for this tradeoff by adjusting the score threshold.

Although conventional cytology had a lower sensitivity for adenocarcinoma, LungSign did not demonstrate a sensitivity difference based on histologic type. This may be due to the design methodology used in the development of the test. The specimen scoring function was designed to detect changes indicative of all histologic types of lung malignancy with the least complexity (i.e., least number of parameters). Nuclear chromatin features that may be good predictors of a particular histologic type of malignancy may have been excluded in favor of predictors that are associated with malignancy in general. The design decision to use a single function to detect malignancy rather than scoring functions for each histologic type was based on the number of lung cancer cases available to train the system.

A simple cytometric model for lung cancer detection has less ability to represent the differences between negatives and lung cancers, but is also less prone to overfit and is likely to be more generalizable to new patient populations. In the future, it may be possible to develop separate classifiers for different cancer types. Because SCC tends to be centrally located and adenocarcinoma tends to be peripheral, histology-specific classifiers could provide additional information about the location of potential malignancy. LungSign showed no significant sensitivity difference for central versus peripheral localizations of lesions. It was expected that cytology would be twice as sensitive for central lesions, and although sensitivity was higher for central versus peripheral lesions (19% versus 14%), the difference was not great enough to be statistically significant.

The prevalence of lung cancer in the trial was 33%, which was higher than expected and higher than that studied in the recent DNA cytometry studies. The high prevalence appears to reflect the referral practices at the institutions that participated in the trial. Several of the physicians in the study specialize in dealing with patients who are likely to have lung cancer and received referrals of suspected cases from other physicians. These patients typically had other forms of lung disease, and Figure 2 therefore represents the ROC for LungSign applied in a differential diagnosis context.

The cytology inadequacy rate ranged from 20% in Kelowna to 65% in Korea, averaging 43% overall. This is high given that other studies have reported inadequacy rates <25%. The reason for the high inadequacy rate is not known. All sites followed the same induction protocol and used the same equipment. Sites that were less experienced with sputum induction tended to have higher inadequacy rates, but the inadequacy rates for all sites other than Kelowna exceeded 30%.

Although the presence of macrophages ensures that a sputum specimen derives from the lower respiratory tract, the premise of the design of the LungSign test was that subtle
DNA chromatin changes could be detected in hundreds or thousands of epithelial cells in sputum from cancer patients; frankly malignant cells are not required to detect the disease. Consequently, induction specimens that do not contain the guarantee of quality by the presence of macrophages may yet contain enough bronchial epithelial cells to produce a valuable score. This point is demonstrated by the fact that LungSign overall performance was comparable for cytology-inadequate versus adequate specimens.

One of the advantages of sputum analysis using a biomarker approach in contrast to sputum cytology is that the operating performance point of the test is adjustable. This property makes it possible to optimize the performance of the test for use in conjunction with standard methods used in early lung cancer detection. There have been two suggested roles for such biomarker tests to support early detection using CT and bronchoscopy: before19,30 and after 30 diagnostic workup.

As a prescreener, a biomarker test would be employed at a low specificity/high sensitivity threshold to find a higher risk population suitable for CT screening. The test would eliminate a significant number of unproductive CTs with as few as possible false negatives. McWilliams et al.19 developed a simple sputum cytometry measure using the observation of five hyperdiploid nuclei (DI/H11022/H11022 >1.2) in a specimen, which they referred to as atypia, as an indication for more intensive follow-up. In their study, the measure detected 13 of 14 cancers with a specificity of 25% for the 547 negative subjects. Although the number of cancers in their study was small, the result suggests that the detection of DNA atypia alone, when applied to a screening population, may be able to eliminate a significant fraction of unproductive CT scans.

As the resolution of CT scanners improves, more and more nodules of undetermined significance are being found. More than half of patients will be found to have nodules,31 and many of these nodules will require follow-up consisting of biopsy or repeated CT scans over a 2-year period. In a Mayo Clinic trial, only 7.1% of prevalence scan nodules >4 mm were found to be malignant. Biomarker tests can be used to find a higher risk subset of patients among those found with nodules of insufficient size to be immediately actionable. As seen in Figure 3A, subjects with high LungSign scores are at a significantly elevated risk of harboring lung cancer. A high score for patients with indeterminate nodules by CT may potentially be used to identify those who should be sent for follow-on investigations more quickly. Alternatively, low scores could help to reduce the anxiety of those who are less likely to harbor a malignancy during the period in which nodule growth is being assessed.

In conclusion, the measurement of nuclear chromatin features through fully automated DNA cytometry may be a valuable tool for the detection of lung cancer in high-risk individuals. The LungSign test was able to resolve lung cancers from negatives in a difficult context—a diagnostic population with 33% disease prevalence. To our knowledge, the validation trial, comprising analyzable sputum specimens from 330 lung cancer patients, represents the largest lung cancer population studied prospectively with a cytometric

![Figure 4](https://example.com/figure4.png)
test. Although the trial protocol measured the performance of a single application of LungSign, alone, it is the test’s performance in conjunction with CT that is of interest. Commenting on the recent exciting results from the ELCAP CT trial,32 Michael Unger\textsuperscript{3} wrote that although no single test has the ability “to provide unequivocal information about the biology of the tumor. . . the combination of molecular and radiographic approaches can enrich the information on which physicians base their decisions.” LungSign, as a continuous measure of abnormality that is sensitive for early-stage cancer, may provide clinicians with additional information to help guide their management decisions.

ACKNOWLEDGMENTS

The authors acknowledge the following investigators and institutions for their participation in the study: Dr. Marjeta Tercelj, Ljubljana University Medical Centre, Slovenia; Dr. Alexei Barchuk, Dr. Denis Vassiliev, and Dr. Andrey Arseniev, N. N. Petrov Research Institute of Oncology, St. Petersburg, Russia; Dr. David Snead and Dr. Paul Dhillon, University Hospitals Coventry, Coventry, UK; Dr. Lorne Fox and Dr. Islam Mohamed, B.C. Cancer Agency, Vancouver, Canada; Dr. Young Whan Kim, Seoul National University College of Medicine, Seoul, Korea; Dr. Andreas Gschwendtner, Hospital Germans Trias I Pujol, Barcelona, Spain; Dr. York Miller and Dr. Wilbur Franklin, University of Colorado Health Sciences Center, Denver, CO; Dr. Andreas Gschwendtner, Dr. Herbert Jamnig, and Dr. Martin Hackl, Hospital Natters, University of Innsbruck, Austria; Dr. Thomas Sutedja, Vrije University Hospital, Amsterdam, The Netherlands.