Automated detection of genetic abnormalities combined with cytology in sputum is a sensitive predictor of lung cancer

Ruth L Katz*[1], Tanweer M Zaidi[1], Ricardo L Fernandez[1], Jingpin Zhang[1], Weigong He[1], Charisse Acosta[1], Michal Daniely[2], Lea Madi[2], Mary A Vargas[3], Qiong Dong[3], Xiaoying Gao Feng Jiang[4], Nancy P Caraway[1], Ara A Vaporciyan[5], Jack A Roth[5] and Margaret R Spitz[3]

[1]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [2]Bioview Ltd, Rehovoth, Israel; [3]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [4]Department of Pathology, University of Maryland, Baltimore, MD, USA and [5]Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Detection of lung cancer by sputum cytology has low sensitivity but is noninvasive and, if improved, could be a powerful tool for early lung cancer detection. To evaluate whether the accuracy of diagnosing lung cancer by evaluating sputa for cytologic atypia and genetic abnormalities is greater than that of conventional cytology alone, automated scoring of genetic abnormalities for 3p22.1 and 10q22.3 (*SP-A*) by fluorescence *in situ* hybridization (FISH) and conventional cytology was done on sputa from 35 subjects with lung cancer, 25 high-risk smokers, and 6 healthy control subjects. Multivariate analysis was performed to select variables that most accurately predicted lung cancer. A model of probability for the presence of lung cancer was derived for each subject. Cells exfoliated from patients with lung cancer contained genetic aberrations and cytologic atypias at significantly higher levels than in those from control subjects. When combined with cytologic atypia, a model of risk for lung cancer was derived that had 74% sensitivity and 82% specificity to predict the presence of lung cancer, whereas conventional cytology achieved only 37% sensitivity and 87% specificity. For diagnosing lung cancer in sputum, a combination of molecular and cytologic variables was superior to using conventional cytology alone.
Modern Pathology (2008) **21**, 950–960; doi:10.1038/modpathol.2008.71; published online 23 May 2008

Keywords: surfactant protein A gene; 3p22.1; FISH; cytology; field cancerization effect; sputum

Early detection of lung cancer is mandatory to reduce its extremely high mortality rate.[1] Numerous trials, worldwide, have compared the roles of sputum cytology, conventional chest X-ray, and helical computed tomography (CT) scanning in detecting early lung cancer.[2–4] Until now, however, the results of such screening tests have been controversial because of either low accuracy or the uncertain significance of their findings.[2,3]

Although sputum abnormalities may reflect neoplastic events in the central airways,[3] cytologic screening of sputum lacks sensitivity and has generally been abandoned in favor of tests, such as

fine needle aspiration or bronchoscopically procured brush, wash, or lavage specimens. However, because sputum collection is noninvasive, it would be advantageous to develop robust biomarkers that could be tested on cells in sputum that indicate the presence or increased risk of lung cancer.

We previously demonstrated, through extensive fluorescence *in situ* hybridization (FISH) mapping studies of NSCLC and adjacent normal bronchial epithelial cells, consistently high levels of deletions in malignant cells and adjacent bronchial epithelial cells in genes located within 3p22.1 (containing the *GC20* gene, *RPL 14*, *CD39A*, and *PMGB*) and 10q22.3 (containing the surfactant protein A gene, *SP-A*).[5–7] These genetic deletions have also been observed in brush specimens of bronchial epithelial cells on the same side as the tumor (TBB), normal lung tissue (NTP) distal to the tumor, and in the main bronchus on the opposite side of the tumor (NBB).[6,7]

This 'so-called' molecular field cancerization effect likely results from multiple clonal abnormalities arising within respiratory epithelial cells exposed to carcinogenic substances from tobacco smoke and other pollutants and reflects genetic predisposition for reduced DNA repair capacity in such patients.[8–11] The presence of concurrent cytologic atypia, especially moderate and severe dysplasia, in sputum cells is also believed to reflect this field effect and has been shown to be substantially associated with an increased risk of developing lung cancer.[12] This observed field effect led us to hypothesize that sputum from a patient with lung cancer or at high risk for it would contain upper airway cells showing these molecular abnormalities, even if these cells showed minimal morphologic atypia.

Thus, the goals of our study were to evaluate the role of cytologic atypia and genetic abnormalities detected by FISH to predict the presence of lung cancer. We also wished to develop a quantitative model that could be used to predict an individual patient's risk for lung cancer.

Materials and methods

Study Design and Patient Population

In this prospective clinical trial, we evaluated sputum samples collected between April 1, 2004 and November 30, 2006, from consecutive patients with lung cancer and from age-matched healthy (nonsmokers) and high-risk (history of heavy smoking) control subjects. None of the participants had received prior radiotherapy or chemotherapy.

The University of Texas MD Anderson Cancer Center Investigational Review Board approved this study, and all study participants signed a consent form detailing the research methods.

Trained staff interviewers from MD Anderson Cancer Center administered an epidemiologic questionnaire to all study participants. Data collected included demographic characteristics and history of tobacco use.[9] In the cancer patients who underwent resection, sputum samples were collected before surgery. All healthy and high-risk smoker control subjects had helical CT scans negative for the detection of lung cancer at the time of study entry and for the following 2 years.

All participants (except for the healthy control subjects, who underwent induced sputum production after saline inhalation with a nebulizer) were instructed to cough into a container that was filled with Sacommano's fixative (90% alcohol, 5% acetic acid, and 5% polyethylene glycol) on 3 consecutive days of arising. Mailed in sputa were cytocentrifuged and filtered through gauze. The sediment was used for at least 10 preparations. Eight cytospin preparations were generated on positively charged glass slides using a Shandon Cytospin 2 Cytocentrifuge (Thermo Fisher Scientific Inc., Waltham,

MA) and fixed in a 4:1 mixture of methanol and acetone. The remaining two preparations were fixed in 95% alcohol for Papanicolaou's staining. Scoring of two Papanicolaou-stained sputa was performed by a cytotechnologist and two senior cytopathologists, none of whom had knowledge of the patients' clinical history. Discrepant diagnoses were scored by consensus over a multiheaded microscope by all three observers (CA, NPC, and RLK). Slides were screened and classified according to a seven-tiered scoring system as follows: negative, squamous metaplasia, mild dysplasia, moderate dysplasia, severe dysplasia, carcinoma, or insufficient for diagnosis. Cytologic classification was based on histologic criteria for squamous metaplasia, dysplasia, and carcinoma *in situ* according to the WHO classification.[13] Slides were considered insufficient for diagnosis, if they had zero to three histiocytes; excessive cellular degeneration; obscuring bacterial, fungal or neutrophilic contamination; or limited number of epithelial cells.

FISH Analysis

A two-color FISH assay using bacterial artificial chromosome probes for 3p22.1 and 10q22.3 (*SP-A*), combined with commercial centromeric probes for chromosomes 3 (cep 3) and 10 (cep 10; Vysis Inc., Downers Grove, IL) was performed on two separate slides. A bacterial artificial chromosome clone of 180 000 bp, located at 10q22.3 and confirmed to contain genomic sequences of *SP-A* comprising both *SFTPA1* and *SFTPA2* (RP11-506MI3; Invitrogen Corp., Carlsbad, CA), was isolated and labeled with Spectrum Green dUTP (Vysis).[5] The chromosomal location of the clone was confirmed on a normal metaphase spread in combination with a centromeric 10 probe, directly fluorescence labeled with Spectrum Orange (Vysis).

The BAC clone for 3p22.1, containing almost 200 kb of genomic sequences and consisting of four known genes—*RPL 14, CD39A, GC20 translation factor sui1 homolog,* and *PMGB*—was obtained commercially from Invitrogen and grown in *Escherichia coli*. It was subsequently isolated, linearized, and labeled with Spectrum Green dUTP according to the manufacturer's directions. Localization of the bacterial artificial chromosome clone on chromosome 3 was confirmed by using normal metaphase FISH. One hundred nanograms of each labeled probe was mixed with an equal quantity of human Cot-1 DNA (Life Technologies, Rockville, MD) in 10 μl of LSI hybridization buffer (Vysis) and mounted on a slide together with 1 μl of either cep 3 (for 3p22.1) or cep 10 (for the *SP-A* gene on 10q22.3). Hybridization and post-washing were carried out as described previously.[5] Counterstaining of nuclei was performed with 4′, 6-diamidino-2-phenylindole and evaluated under a fluorescence microscope equipped with the appropriate filter

combinations. If the hybridization signals were deemed satisfactory, the slides were sent for automated FISH scanning.

Automated FISH Scanning

Slides were automatically scanned by the Duet™ (BioView Ltd Rehovot, Israel). The Duet is based on a fully automated microscope (Olympus BX61, Japan), a motorized 8-slides stage (Marzhauser, Wetzler, Germany) and a 3CCD progressive scan color camera (JVC KYF75U, Japan). FISH scanning was performed using ×60 oil objective in fluorescent illumination and by using appropriate filters and a software program specifically designed to capture and analyze the orange and green fluorescent signals generated from our probes. While scanning, the system produces images of all captured cells that can be further reviewed by the operator.

Two slides were scanned for each sputum samples: one slide was hybridized to the 3p22.1 and its control centromere and the second slide was hybridized to the 10q22.3 probe and centromere 10. At least 100 epithelial cells per slide were scanned and evaluated for FISH abnormalities. For each cell, the presence or absence of centromeric probes (orange signals) or locus-specific probes (green signals) was recorded. Cells were classified into four major subgroups: (a) Normal—displaying two centromeric and two locus-specific signals; (b) Deletion—displaying more centromeric than locus specific signals (ie, two centromeric and one locus-specific signals, three and two, etc); (c) Polysomy—displaying multiple gains of both centromere and locus-specific signals; (d) Monosomy—displaying only one centromeric signal.

Cells that could not be scored were discarded into an unclassified category.

Only cells that were clearly nonoverlapping and complete, with well-preserved nuclei, with high-quality fluorescence signals, and without background fluorescence were scored. At the end of each scan, the images of all cells that were scanned and analyzed were displayed and a pie chart summarizing the chromosomal abnormalities found in each scan was generated. Two experienced observers, who were trained on the system and blinded to the patients' clinical status confirmed the automatic classification and reclassified cells that were misinterpreted by the system.

Normal controls consisted of pooled human lymphocytes hybridized and quantitated automatically in the same batches as sputa for 3p22.1 and 10q22.3 as the mean number of cells with summed abnormalities of 3p22.1, cep 3, 10q22.3, and cep 10 ± 1 s.d.

Statistical Analyses

The mean, s.d., median, and range for continuous variables were analyzed using Wilcoxon's rank-sum

Table 1 Characteristics of the subject populations

	Cancer (N = 35)	No cancer (N = 31)
Sex		
Female	21	16
Male	14	15
History of smoking (pack year)		
0	7	6
<20	7	1
20–50	12	16
50–100	6	5
>100	2	1
Age (mean, median)		
Mean (range)	65 (47–81)	59 (27–75)
Stage		
I/IA/IB	19	0
II	7	0
III/IIIA	6	0
IV	2	0
Location of tumor		
Central tumor	6	0
Peripheral tumor	29	0
Histology		
Adenocarcinoma	22	0
Squamous cell carcinoma	10	0
Neuroendocrine carcinoma	2	0
Non-small cell carcinoma	1	0

test to assess the differences in the distribution of genetic abnormalities between cancer patients and control subjects. For categorical variables such as sex, smoking history, cytologic diagnosis, and disease stage, Fisher's exact test was used to assess the association between the different variables and cancer status.

Univariate and multivariate logistic regression models for estimating cancer status were also performed. From the multivariate models, receiver–operator-characteristic (ROC) curves were produced to estimate each individual's predicted probability of having cancer.

Results

Patient Population

A total of 71 patients were enrolled in the trial, but five were excluded, because they had poor-quality sputum specimens that did not produce valuable cells for FISH analysis. Of the 66 patients whose sputa could be evaluated, 35 had predominantly early-stage lung cancer and 31 were control patients, of whom 6 were healthy and 25 were at high risk for lung cancer because of their history of heavy smoking (Table 1). The patients with cancer had NSCLC, classified predominantly as adenocarcinoma and squamous cell carcinoma. Most of the patients had peripherally based tumors. Cytologic

diagnoses of squamous metaplasia and mild, moderate, and severe dysplasia were significantly associated with cancer status (Table 2). Of 31 patients without cancer, 19/31(70.37%) were negative by cytology; conversely 27/35 patients (77%) of patients with cancer had cytology diagnosis of squamous metaplasia or worse levels of atypia ($P = 0.009$). A diagnosis of moderate or severe dysplasia was associated with cancer in 42% of cases, while only 4/31 or 13% of negative for cancer high-risk smokers had moderate or severe dysplasia.

Table 2 Association between cytologic diagnosis on sputum and cancer[†]

Cytology	Cancer patients	%	Control	%
Negative	8	29.63	19	70.37
Squamous metaplasia	10	83.33	2	16.67
Mild dysplasia	6	50.00	6	50.00
Moderate dysplasia	7	77.78	2	22.22
Severe dysplasia	4	66.67	2	33.33
Total	35		31	

*1 = negative, 2 = squamous metaplasia, 3 = mild dysplasia, 4 = moderate dysplasia, and 5 = severe dysplasia.
[†]$P = 0.009$ (Fisher's exact test).

Hybridization Efficiency and Cutoff Values for FISH

Each gene-specific probe was confirmed on metaphases from normal lymphocytes to hybridize to the appropriate centromeric and locus-specific regions for cen 3, 3p22.1, cen 10, and 10q22.3. Cancer patients had an average of 183 cells (median, 131; range, 49–589) for 3p22.1 and 158 cells (median, 139; range, 35–455) for 10q22-23 evaluated. Control subjects had an average of 204 cells for 3p22.1 (median, 183; range, 40–673) and 170 for 10q22-23 (median, 189; range, 13–474) evaluated.

Similarly, diploid signals were noted for each probe in interphase nuclei from five batches of normal lymphocytes. The mean ($\pm$ s.d.) deletion and polysomy value for 10q 22.3 was 1.14 ± 0.59, and that for deletions and polysomies of cen 3 and 3p22.1 was 3.02 ± 1.73. For all chromosomal abnormalities of 3 and 10, the mean and s.d. were 4.91 ± 2.50.

Comparison of Patient Characteristics, Genetic Changes in Sputum and Cancer Status

There was no significant difference between cancer status for patients' age or smoking history in pack years (Table 1). There were, however, significant differences in the percentages of chromosomal abnormalities in epithelial cells in relationship to
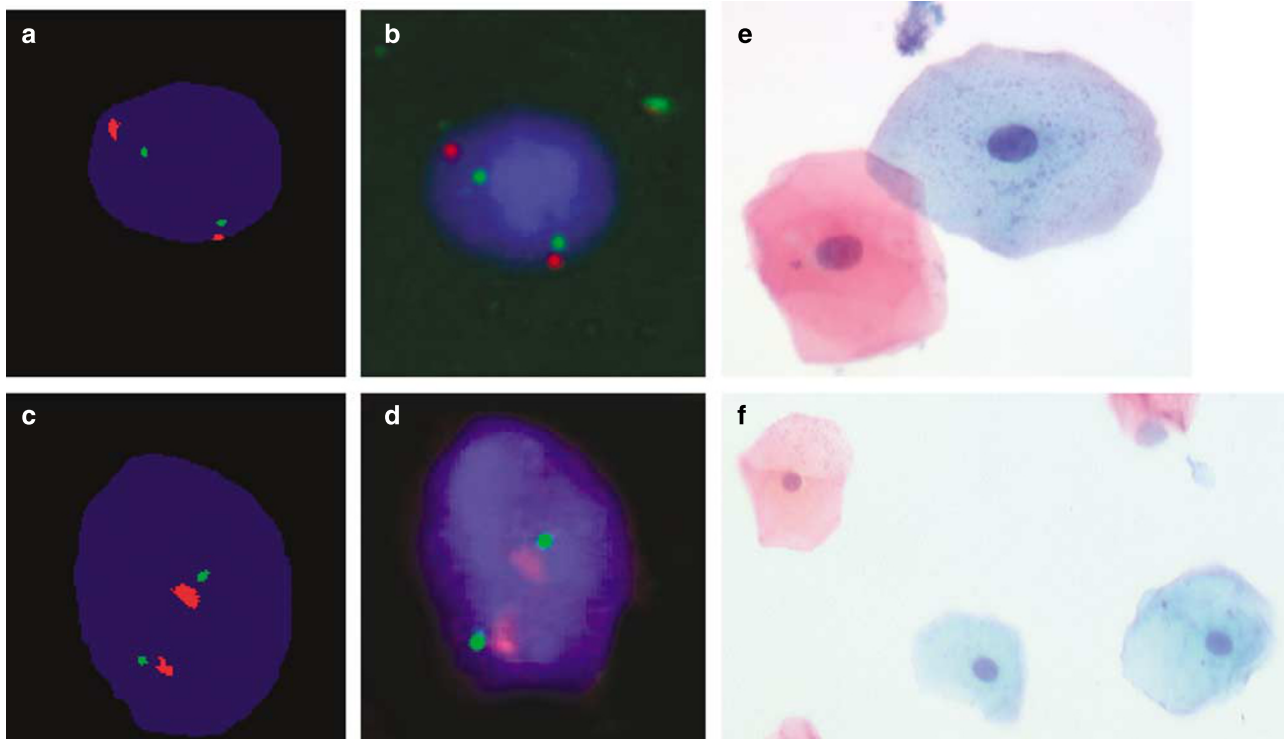

Figure 1 Sputum sample from a healthy control patient, a 52-year-old never smoker, showing (**a** and **b**) squamous epithelial cells with diploid signals for 3p22.1 (green) and centromeric 3 (orange) and (**c** and **d**) SP-A gene (green) and centromeric 10 (red). (**e** and **f**) Normal squamous cells in sputum with unremarkable nuclear features and thin homogenous cytoplasm (Papanicolaou's stain; original magnification × 400). Total FISH abnormalities of 3p22.1, 3, 10q22-23, and centromeric 10 were 6.7. Probability of cancer risk = 0.064.
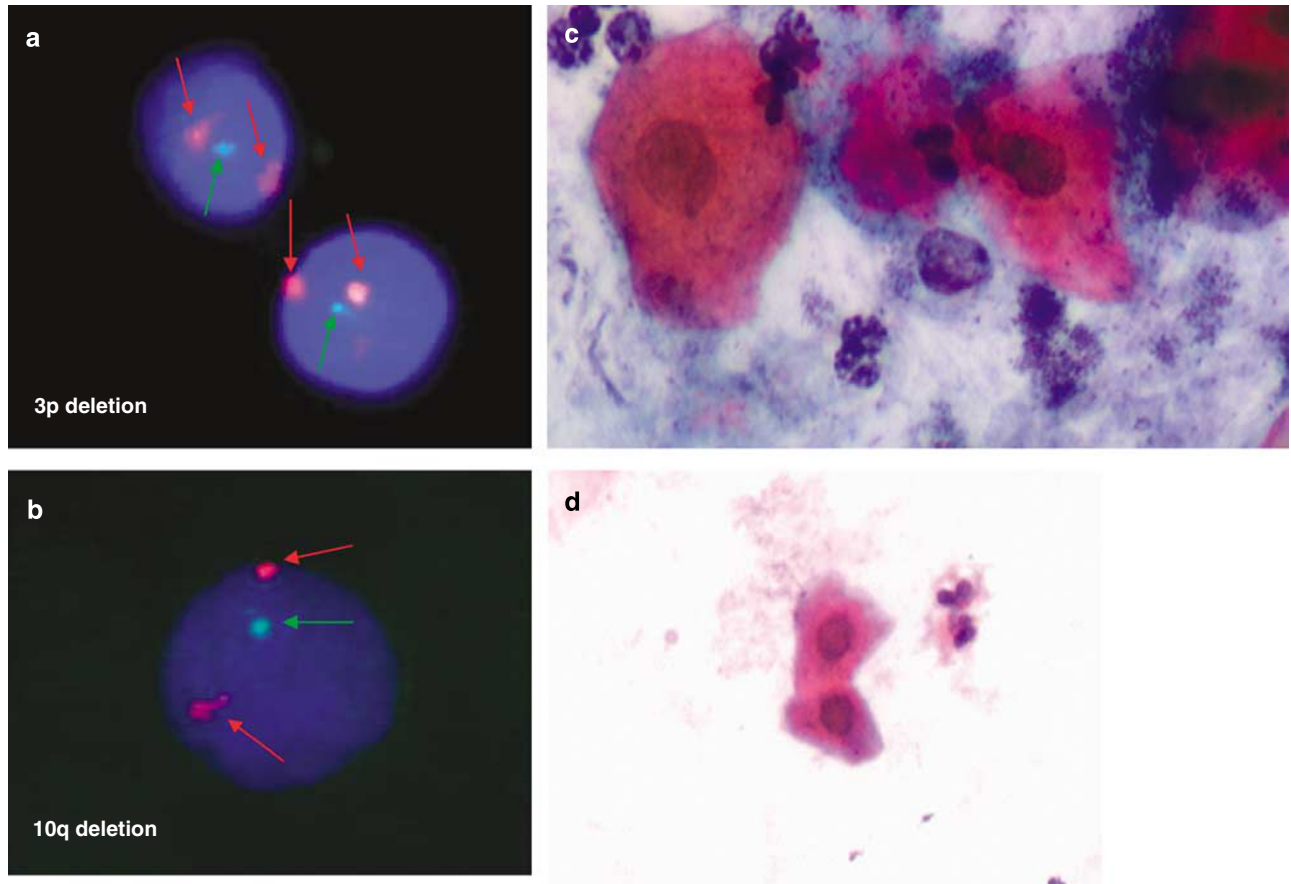
Figure 2 Sputum specimen from a high-risk 61-year-old man with 100 pack years of smoking history and a CT scan negative for lung cancer. FISH assay shows (**a**) deletion of 3p22.1 green signal (one green signal) relative to centromeric 3 (two red signals). Similarly, (**b**) depicts loss of green signal at 10q22.3 *(SP-A)* relative to two red signals of centromeric 10 (oil immersion objective, original magnification × 630). The total number of chromosomal abnormalities of chromosomes 3p22.1, centromeric 3 and 10, and 10q22.3 was 15.5. (**c** and **d**) Cytologic examination shows cells of (**c**) moderate to severe dysplasia with hyperchromatic irregular nuclei and keratinized thickened eosinophillic cytoplasm and (**d**) atypical squamous metaplasia (Papanicolaou's stain; original magnification × 400). Probability of cancer risk = 0.844.

the patients' cancer status. (Table 2). Significantly, more abnormalities in epithelial cells of 3p, deletions of 10q, all abnormalities of 3, deletions and polysomies of 3p, all abnormalities of 10, and all 3 and 10 abnormalities (P-values <0.018, <0.013, <0.033, <0.026, <0.018, and <0.008, respectively) were present in the cancer patients than in the control subjects (Figures 1–4).

In a logistic regression model estimating cancer status (Table 3), the most significant parameters were age, deletions of 3p22.1 and 10q22.3, and a variety of abnormalities of 3p22.1 and 10q22.3, and chromosomes 3 and 10 as well as both squamous metaplasia and moderate dysplasia *vs* negative cytologic results. There was no correlation between sex and smoking history (including pack years, current smoker, former smoker, or never smoker), and cancer status.

The multivariate logistic regression model to estimate cancer status (Table 4) selected six variables—two genetic and four cytologic—as the most predictive parameters for estimating cancer status where the P-value cutoff point was chosen as 0.10.

The variable with the highest odds ratio (OR) was moderate dysplasia (OR 17.96) followed by squamous metaplasia (OR 14.84), severe dysplasia (OR 5.39), mild dysplasia (OR 3.63), deletion and polysomy of 10q22.3 greater or less than 2 (OR 4.38), and deletions and polysomies of 3p22.1 and cen 3 greater or less than 5 (OR 3.01). The ROC curve (Figure 5), using a cutoff point of estimated $P = 0.004$, showed the area under the curve to be 0.822, when using both the cytologic and FISH parameters. Using only the selected FISH variables of deletions of 10q and 3 and 3p abnormalities resulted in an ROC curve of 0.682 ($P = 0.065$), whereas using only the cytologic diagnosis resulted in an ROC curve of 0.742 ($P = 0.040$) (Figure 6).

If just moderate and severe dysplasia were considered to be the predictors of the presence of lung cancer, then the sensitivity and specificity of the sputum cytology test were 32% and 87%, respectively. If only severe dysplasia was considered, the sensitivity of sputum cytology was 11% and the specificity, 99%.
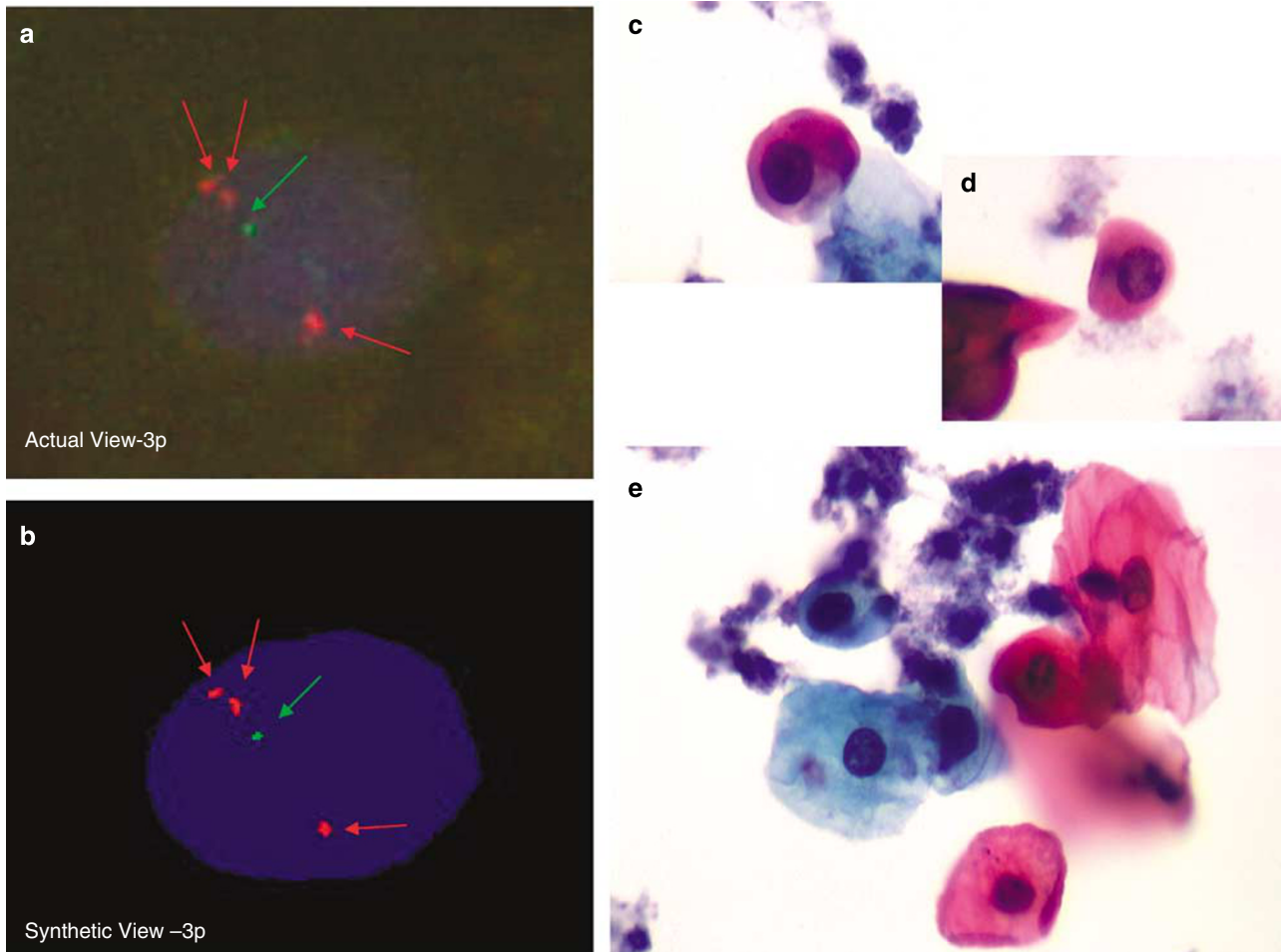
Figure 3 Sputum specimen from a 67-year-old woman nonsmoker with a family history of lung cancer, who had stage III adenocarcinoma (peripheral lesion). FISH analysis showed (**a** and **b**) deletions of 3p22.1 as indicated by the loss of one green signal relative to two red centromeric 3 signals. Total deletion of 10q 22.3 was 5.69%, centromeric10 monosomy was 2.43%, of and deletion 3p22.1 was 6.79%. Total number of FISH abnormalities involving chromosomes 3 and 10 were 15.89. Cytologic examination showed (**c–e**) extensive moderately and severely dysplastic cells (arrows; Papanicolaou's stain; original magnification ×400). This patient died with metastatic disease 24 months later. Probability of cancer risk = 0.817.

Table 5a shows the risk score, either: low, intermediate, or high, based on both cytology and FISH variables.

Table 5b (Supplementary Information) shows the actual probability for the presence of cancer in the sputum for each study participant using a combination of genetic variables and cytologic diagnosis. Assuming a cutoff of $P > 0.600$ to indicate high risk for cancer, then six of the high-risk control subjects appear to be at high risk for developing cancer compared to 21 of the cancer patients. Similarly, there were two patients with cancer that had low probability scores of 2 or < 2.

Discussion

Because of CT's high sensitivity and lack of specificity, it would be desirable to develop a minimally invasive test for genetic susceptibility that may assist in identifying those individuals at

the highest risk for developing lung cancer. However, conventional cytologic sputum screening lacks sensitivity for various reasons, including its inability to detect small atypical squamous cells, the fact that abnormal cells may not be shed from peripherally based lesions, the patient's inability to produce an adequate cough specimen, and contamination of the specimen by neutrophils and bacteria.

The uninduced sputum from healthy subjects contains for the most part squamous cells from oral contamination and upper airways, a few alveolar macrophages, and a few bronchial epithelial cells.[14] In patients who have smoked and are at risk to develop bronchogenic carcinoma, the precursor lesions are squamous metaplasia, squamous dysplasia and carcinoma *in situ* that occur in the bronchi and large airways. Sputum cytology is the only noninvasive way to assess these lesions. The sputum classification schemes consist of gradations of microscopic abnormalities similar to those observed in histologic sections from lower airways
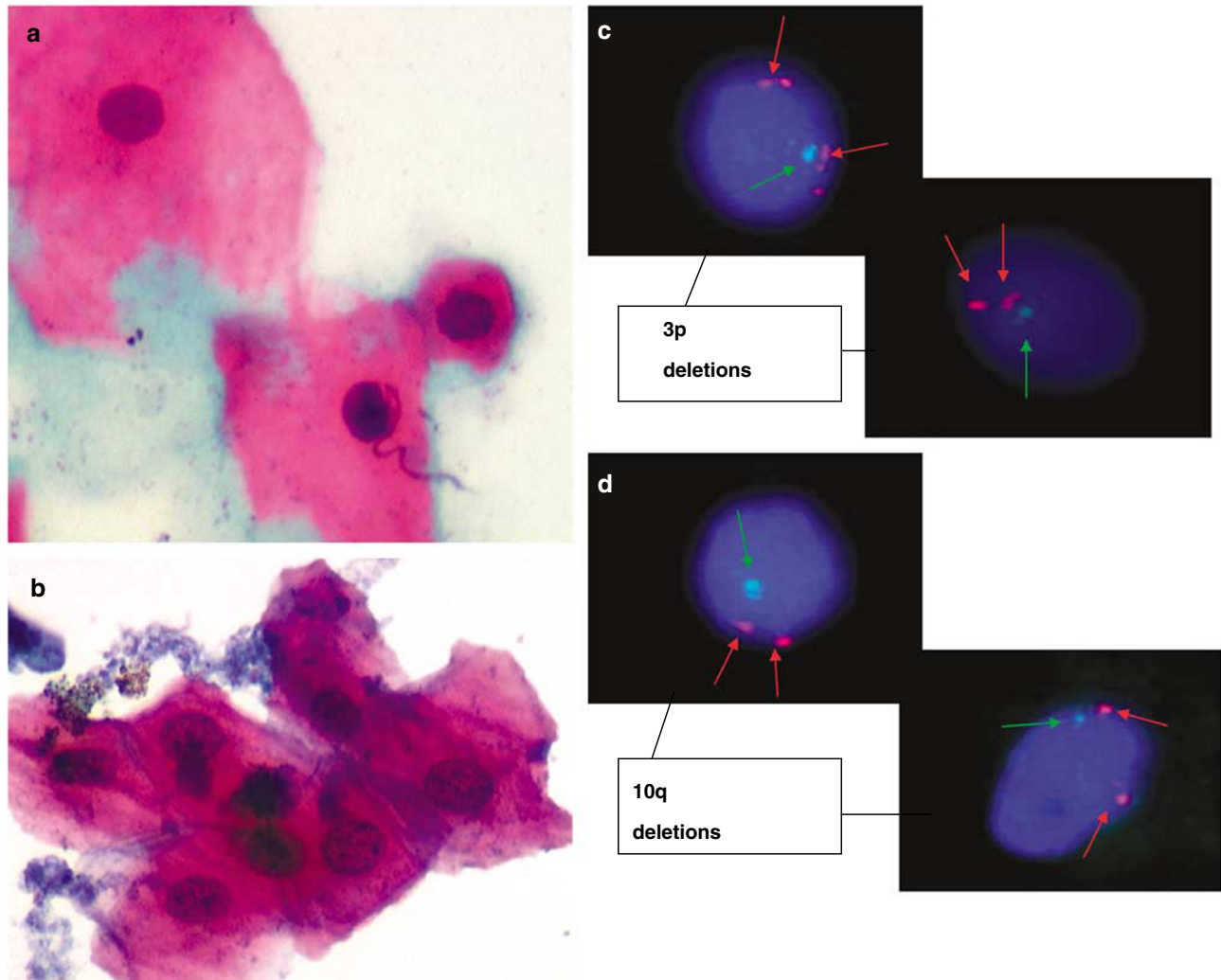
Figure 4 Sputum sample from a 63-year-old man with a history of 50 pack years of smoking and no evidence of cancer shows (**a** and **b**) cells of moderate and severe dysplasia (arrows) on cytologic examination (Papanicolaou's stain; original magnification × 400) FISH showed deletions of 3p22.1 (**c**) and 10q22.3 (**d**) as indicated by the loss of one green signal relative to two orange signals. 3p22.1 deletions were 5.63% and 10q22.3 deletions were 3.78%. Total FISH abnormalities of chromosomes 3 and 10 were 9.4. Probability of cancer risk = 0.830.

of smokers. The cytologic manifestations of dysplasia occur as increasingly severe cellular changes involving both: cell size and nuclear size, chromatin density, and nuclear membrane irregularity.[13]

Molecular changes occur early in neoplastic development including LOH at one or more 3p regions and 9p21, and may be found in histologically normal epithelium[15] similar to the molecular abnormalities of 3p22.1 and *SP-A,* and are the manifestations of a field cancerization effect.[6,7]

Recently, Wang *et al*[16] detailed four types of bronchial epithelial dysplasia (basal cell dysplasia, columnar cell dysplasia, bronchial epithelial dysplasia with transitional differentiation, and squamous dysplasia) that occurred throughout the bronchial tree in patients who had their lungs resected for bronchogenic cancer. For both squamous and adenocarcinomas, there existed large numbers of various types of bronchial epithelial dysplasias regardless of the histological subtype of the

bronchogenic carcinoma. Hence, it is not surprising that sputum from patients with adenocarcinoma should manifest cellular atypias in the form of squamous metaplasia, mild, moderate, and severe dysplasia, representative of the field effect of atypias secondary to tobacco usage, passive smoking, or industrial or asbestos exposure.[15]

In a recent prospective study of smokers with COPD, risk for incident lung cancer was increased among those with sputum cytologic atypia graded as moderate or worse.[17] Because of the weaker association between adenocarcinoma and sputum atypia, it was suggested that other biomarkers in sputum such as FISH assays of chromosomal changes may be complementary to diagnosis.[17]

The findings in this study showing squamous metaplasia and all degrees of dysplasia to be present in patients with lung cancer regardless of histologic subtype are consistent with the findings of recent investigators.[12,16,17]

If only the abnormalities detected on FISH analysis were used, the sensitivity was higher but the specificity was lower. However, a combination of FISH and sputum cytology resulted in a considerably higher sensitivity rate than obtained with morphology alone. We concluded that subjects in a high-risk group, with high probability scores derived from cytologic and FISH analyses according to our model, should undergo CT scanning. If the CT findings were negative, these subjects would be ideal candidates to undergo fluorescence broncho-scopy to exclude the presence of central airway pre-invasive malignant lesions.

In this study, there were six high-risk patients without CT evidence of lung cancer that might qualify for bronchoscopy based on their probability scores. The finding of a low probability score in several of the lung cancer patients no doubt reflected an inadequate sputum sample. In the future, use of induced sputa and stricter criteria for adequacy, such as the presence of bronchial epithelial cells and a greater number of histiocytes may improve the accuracy of the test.

The results of this study validated our approach to measuring and quantitating molecular abnormalities in consecutive fields of epithelial cells that were not necessarily cytologically abnormal. The presence of chromosomal abnormalities in both normal and atypical epithelial cells exfoliated in sputum confirmed our previous findings that cellular genetic abnormalities of 3p22.1 and 10q22.3 reflect a field cancerization effect within the bronchial cells of individuals at high risk for developing cancer.[6,7]

Other investigators tested epithelial cells in sputum by FISH using a commercial probe set for four different chromosomal regions (5p15, 6p11-q11, 7p12 (including epidermal growth factor receptor) and 8q24 (including *C-myc*), and required

Table 3 Univariate logistic regression model for diagnosis of cancer

Variable	Coefficient	s.e.	P-value	Odds ratio
Age	0.05	0.03	0.047	1.05
History of smoking (pack year)	−0.09	0.23	0.68	0.91
Deletions of 3p	0.17	0.1	0.081	1.185
Abnormalities of 3p	−40.03	235.97	0.865	0
Monosomy of cep 3	0.16	0.22	0.459	1.175
Polysomy of cep 3	0.33	0.37	0.377	1.386
Polysomy of 3p	0.72	0.5	0.146	2.063
Deletions of 10q	0.44	0.18	0.014	1.546
Abnormalities of 10q	−0.14	0.59	0.806	0.866
Monosomy of cep 10	0.15	0.14	0.277	1.159
Polisomy of cep 10	−0.41	0.53	0.444	0.666
Polisomy of 10q	0.03	0.43	0.944	1.031
All abnormalities of 3	0.17	0.09	0.052	1.183
Deletion and Polysomy of 3p	0.19	0.1	0.049	1.214
Aneusomy 3	0.22	0.2	0.268	1.249
All abnormalities of 10	0.18	0.09	0.045	1.197
Deletion and Polysomy of 10q	0.35	0.15	0.02	1.426
Aneusomy 10	0.11	0.13	0.388	1.114
All abnormalities of 3 and 10	0.15	0.06	0.01	1.166
Sex (male *vs* female)	−0.34	0.5	0.494	0.711
Cytological diagnosis* (2 *vs* 1)	2.47	0.88	0.005	11.873
Cytological diagnosis* (3 *vs* 1)	0.87	0.71	0.226	2.375
Cytological diagnosis* (4 *vs* 1)	2.12	0.91	0.019	8.312
Cytological diagnosis* (5 *vs* 1)	1.56	0.96	0.106	4.75

*Cytologic diagnosis: 1 = negative, 2 = squamous metaplasia, 3 = mild dysplasia, 4 = moderate dysplasia, and 5 = severe dysplasia.
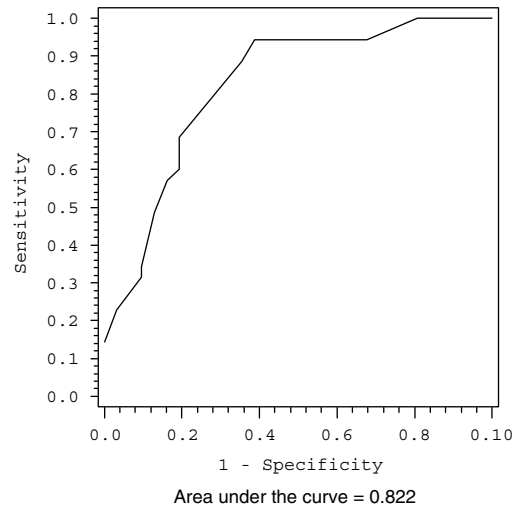

Figure 5 ROC curve based on fish variables and cytology diagnosis.

Table 4 Multivariate logistic regression model for diagnosis of cancer

Variable	Coefficient	P-value	Odds ratio (95%CI)
Intercept	−2.677	0.002	
All abnormalities of 3/3p	1.103	0.083	3.01 (0.87–10.50)
Deletions and polysomies of 10/10q	1.477	0.033	4.38 (1.13–17.06)
Cytological diagnosis (2 *vs* 1)*	2.697	0.005	14.84 (2.23–98.53)
Cytological diagnosis (3 *vs* 1)*	1.289	0.117	3.63 (0.72–18.20)
Cytological diagnosis (4 *vs* 1)*	2.888	0.005	17.96 (2.36–136.92)
Cytological diagnosis (5 *vs* 1)*	1.685	0.118	5.39 (0.65–44.68)

Cytological diagnosis: *1 = negative, 2 = squamous metaplasia, 3 = mild dysplasia, 4 = moderate dysplasia, and 5 = severe dysplasia.

958

a positive sputum diagnosis to be based on DNA copy number gains for at least two probes in a minimum of two or three cells.[18,19] When using this approach, the sensitivity (50%) and specificity (81%) of FISH did not exceed the sensitivity of sputum cytology to detect lung cancer.[18] Furthermore, heavy tobacco smokers and asbestos-exposed workers had FISH results similar to those seen in never smokers, suggesting that the composition of these probes was not optimal for detecting early lung cancer in high-risk populations.[18] In contrast, others reported that FISH combined with cytology led to an improved diagnosis of malignancy.[20] In a recent study, combined genetic aberrations for genes HYAL2 and FHIT were found with FISH in 76% of sputa from patients with cancer, but in only 47% of cases that considered positive on cytology, demonstrating that with an appropriate choice of probes, FISH can detect abnormal cells that may be undetectable by cytology.

Our sputum probe set of 3p22.1 and *SP-A* was selected on the basis of results of high-resolution comparative genomic hybridization analysis of cDNA microarrays in adenocarcinomas and squamous cell carcinomas that showed significantly higher levels of these deletions, relative to those in normal human bronchial epithelial cells, in almost all tumors tested.[21] These probes were subsequently tested by FISH in adenocarcinomas and squamous cell carcinomas and found to correlate significantly with the results of the comparative genomic hybridization.[21]

Deletion of *SP-A* are frequent in lung cancer cells as well as in adjacent bronchi, normal lung, and bronchial cells from main stem bronchi on the normal and tumor sides.[5–7] Deletions of *SP-A* in lung cancer cells are associated with polysomy of the genes for centromeric 10/10q22.3, EGFR, 5p15, 6p11-q11, and *C-MYC,* which are genes that drive cell cycle and proliferation.[22] *SP-A* deletions also correlated with lymph node metastases, higher stage and poor prognosis.[5,22]

Chromosome 3p deletion is currently the most common finding in lung cancer. Allelic losses at one or more chromosome 3p21.3 locus are the most frequent chromosomal abnormalities detected in the bronchial epithelia of smokers and are detected even in normal bronchial mucosae of smokers.[15]

The use of FISH for diagnostic purposes has increased considerably in the last few years, primarily because FISH permits visualization and examination of genetic aberrations as rare events in a large number of cells that may have normal genetic composition. FISH is ideally suited for cytologic specimens such as sputum, which may be obtained spontaneously by coughing or induced by inhalation of nebulized saline. The major value of sputum biomarkers is to identify patients at high risk for cancer-related events, such as the development of premalignant lesions or early cancers, so that these patients may be subject to intense surveillance either by fluorescent bronchoscopic examination with removal of neoplastic lesions[23,24] or by regular helical CT scanning of lungs to detect peripheral carcinomas. Additionally, this is an ideal population to benefit from the use of chemopreventive agents and smoking-cessation counseling.

In summary, we used an automated quantitative system to score FISH abnormalities in epithelial cells from noninduced sputum specimens, which resulted in a gallery of cells that could then be interactively classified in conjunction with morphologic findings. Software programs specific to the sputum application and the size of the probes and specific filter sets were used to maximize the accuracy of the testing. Correlating our sputum findings with disease state per individual, we discovered that epithelial cells in sputum from patients with NSCLC were cytologically and genetically abnormal relative to those from a high-risk control group with no CT evidence of lung cancer as well as healthy controls. On the basis of the results of our model for risk of lung cancer, which should be
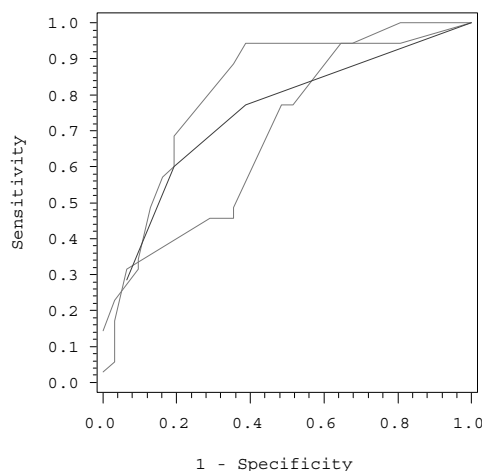
---------- Area under the curve = 0.822, combining FISH variables with cytology

---------- Area under the curve = 0.682, based only on FISH variables

---------- Area under the curve = 0.742, based only on cytological diagnosis

Figure 6 ROC curves based on different models.

Table 5a Risk score of cancer by FISH and cytology score[a]

Risk score	Cancer N (%)	No cancer N (%)
0–1 (low)	3 (9.0)	16 (52.0)
2 (intermediate)	15 (43.0)	11 (36.0)
3–4 (high)	17 (49.0)	4 (13.0)

Cytology diagnosis: 1 coded as 0, 2–3 coded as 1, 4–5 coded as 2.
All abnormalities of 3/3p22.1 < 5 coded as 0, > = 5 coded as 1.
Deletions and polysomies of 10/10q22.3 < 2 coded as 0, > = 2 coded as 1.
N = number of patients, (%) = percentage.
[a]All abnormalities of 3/3p22.1 plus deletions and polysomies plus cytology diagnosis.

Table 5b Estimates of probability of cancer for each patient by sputum evaluation of genetic and cytologic variables

Patient*	All abnormalities of 3/3p22.1[a]	Deletions and polysomies of 10/10q22.3[b]	Cytologic[c] diagnosis	Cancer status	Probability of cancer
1	1	0	3	Absent	0.429
2	0	0	5	Absent	0.271
3	0	1	1	Absent	0.232
4	0	0	1	Absent	0.064
5	0	1	1	Absent	0.232
6	1	1	1	Absent	0.476
7	1	1	3	Absent	0.767
8	1	1	1	Absent	0.476
9	1	0	1	Absent	0.172
10	0	0	3	Absent	0.200
11	1	1	1	Absent	0.476
12	1	1	5	Absent	0.830
13	0	1	1	Absent	0.232
14	1	0	4	Absent	0.788
15	1	0	3	Absent	0.429
16	0	1	2	Absent	0.817
17	0	0	1	Absent	0.064
18C	0	1	1	Absent	0.232
19C	0	0	1	Absent	0.064
20C	1	0	1	Absent	0.172
21C	0	0	1	Absent	0.064
22C	0	0	1	Absent	0.064
23C	0	1	1	Absent	0.232
24	1	1	1	Absent	0.476
25	0	0	1	Absent	0.064
26	1	1	1	Absent	0.476
27	0	1	4	Absent	0.844
28	0	1	1	Absent	0.232
29	1	0	3	Absent	0.429
30	0	0	3	Absent	0.200
31	0	1	2	Absent	0.817
32	1	0	2	Present	0.755
33	1	1	2	Present	0.931
34	1	1	2	Present	0.931
35	1	1	2	Present	0.931
36	1	1	1	Present	0.476
37	1	1	1	Present	0.476
38	1	1	1	Present	0.476
39	0	1	4	Present	0.844
40	1	0	1	Present	0.172
41	0	1	3	Present	0.522
42	1	1	3	Present	0.767
43	1	1	3	Present	0.767
44	1	1	5	Present	0.830
45	1	1	3	Present	0.767
46	1	0	4	Present	0.788
47	0	1	4	Present	0.844
48	1	1	1	Present	0.476
49	0	1	3	Present	0.522
50	0	0	2	Present	0.505
51	0	1	5	Present	0.619
52	1	1	5	Present	0.830
53	0	1	2	Present	0.817
54	0	0	4	Present	0.553
55	0	1	5	Present	0.619
56	0	1	2	Present	0.817
57	1	1	1	Present	0.476
58	1	1	2	Present	0.931
59	0	1	2	Present	0.817
60	0	1	4	Present	0.844
61	1	1	1	Present	0.476
62	1	1	1	Present	0.476
63	1	1	4	Present	0.942
64	0	0	4	Present	0.553
65	1	1	2	Present	0.931
66	0	0	3	Present	0.200

Patients 1–31 are controls without cancer, healthy controls (18–23C), and high-risk controls; patients 32–66 have cancer.
[a]All abnormalities of 3/3p22.1, if $<5 = 0$, if $>5 = 1$.
[b]All deletions and polysomies of 10/10q22.3, if $<2 = 0$, if $>2 = 1$.
[c]Cytologic diagnosis: 1 = negative, 2 = squamous metaplasia, 3 = mild dysplasia, 4 = moderate dysplasia, and 5 = severe dysplasia.

validated in a larger study, we conclude that the best predictive sputum assay for lung cancer will be a combination of morphologic characteristics and molecular abnormalities in both atypical cells and morphologically normal cells.

Disclosure/Conflict of interest

Drs Katz and Jiang are holders of an issued USA patent (US Pat. Appl. No. 20060078885) for FISH probes to 3p22.1 and 10q22-23. Dr Katz is a consultant to Bioview Company, Rehovoth, Israel, and performed this study on a Bioview system on loan from Bioview to the Pathology Department, MD Anderson Cancer Center. Bioview company has an option through the MD Anderson Technology Office to license these probes.

References

1 Jemal A, Thomas A, Murray T, et al. Cancer statistics, 2002. CA Cancer J Clin 2002;52:23–47.

2 Swensen SJ, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. Radiology 2005;235:259–265.

3 Melamed MR. Lung cancer screening results in the National Cancer Institute New York Study. Cancer 2000;89:2356–2362.

4 Henschke CI, Yankelevitz DF, Libby DM, et al. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 2006;26:1763–1771.

5 Jiang F, Caraway NP, Nebiyou Bekele B, et al. Surfactant protein A gene deletion and prognostics for patients with stage I non-small lung cancer. Clin Cancer Res 2005;1:5417–5424.

6 Fernandez RL, Zaidi T, Caraway N, et al. Field Cancerization I Non-Small Cell Lung Cancer Demonstrated by Fluorescence in situ Hybridization (FISH) for 3p22.1 and 10q22-23: Correlation of Molecular Abnormalities with Clinical Variables. presented at Lung SPORE Winter Meeting: Los Angeles, CA, 2006.

7 Barkan GA, Caraway NP, Jiang F, et al. Comparison of molecular abnormalities in bronchial brushings and tumor touch preparations. Cancer 2005;105:35–43.

8 Cheng L, Spitz MR, Hong WK, et al. Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. Carcinogenesis 2000;21:1527–1530.

9 Wu X, Lin J, Etzel CJ, et al. Interplay between mutagen sensitivity and epidemiological factors in modulating lung cancer risk. Int J Cancer 2007;120:2687–2695.

10 Wu X, Zhao Y, Hohn SE, et al. Benzo[a]pyrene diol epoxide-induced 3p21.3 aberrations and genetic predisposition to lung cancer. Cancer Res 1998;58:1605–1608.

11 Hollander MC, Philburn RT, Patterson AD, et al. Deletions of XPC leads to lung tumors in mice and is associated with early events in human lung carcinogenesis. Proc Natl Acad Sci USA 2005;102:13200–13205.

12 Prindiville SA, Byers T, Hirsch FR, et al. Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. Cancer Epidemiol Biomarkers Prev 2003;12:987–993.

13 Franklin WA, Wistuba II, Geisinger K, et al. Squamous dysplasia and carcinoma in situ. in: Travis W et al. (eds). WHO Classification of Tumors of the Lung, Pleura, Thymus and Heart. IARC Press: Lyon, 2004, pp 68–72.

14 Melamed M. The lower respiratory tact in the absence of cancer: conventional and aspiration cytology, chapter 19. in: Koss LG (ed). Koss' Diagnostic Cytology, Volume I, 5th edn, Lippincott Williams & Wilkins Press: Philadelphia, PA, USA, 2005, pp 568–642.

15 Park IW, Wistuba II, Maitra A, et al. Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. J Natl Cancer Inst 1999;91:1863–1868.

16 Wang GF, Lai MD, Yang RR, et al. Histological types and significance of bronchial epithelial dysplasia. Modern Pathology 2006;19:429–437.

17 Byers T, Wolf HJ, Franklin WA, et al. Sputum cytologic atypia predicts incident lung cancer: defining latency and histologic specificity. Cancer Epidemiology Biomarkers Prev 2008;17:158–162.

18 Kettunen E, Salmenkivi K, Vuopala K, et al. Copy number gains on p515, 6p11q11, 7p12, and 8q24 are rare in sputum cells of individuals at high risk of lung cancer. Lung Cancer 2006;54:169–176.

19 Variella-Garcia M, Kittelson J, Schulte AP, et al. Multi-target interphase fluorescence in situ hybridization assay increases sensitivity of sputum cytology as a predictor of lung cancer. Cancer Detect Prev 2004;28:244–251.

20 Li R, Todd NW, Qiu Q, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. Clin Cancer Res 2007;13:482–487.

21 Jiang F, Yin Z, Caraway NP, et al. Genomic profiles in state I primary non small cell lung cancer using comparative genomic hybridization analysis of cDNA microarrays. Neoplasia 2004;6:623–635.

22 Katz RL, Zaidi T, Fernandez R, et al. Deletion of gene for Surfactant Protein A Correlates with Over-expression of EGFR, C-MYC, 6p11-q11 and Polysomy 10/10q22.3 in Non-small cell Lung Cancer(NSCLC). A Fish Based Study of 45 Patients. Mod Pathol 2008;21(Suppl 1):344.

23 Shibuya K, Fujisawa T, Hoshino H, et al. Fluorescence bronchoscopy in the detection of preinvasive bronchial lesions in patients with sputum cytology suspicious or positive for malignancy. Lung Cancer 2001;32:19–25.

24 Loewen G, Natarajan N, Tan D, et al. Autofluorescence bronchoscopy for lung cancer surveillance based on risk assessment. Thorax 2007;62:335–340.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)