Evaluation of a Serum Lung Cancer Biomarker Panel

Peter J Mazzone[1], Xiao-Feng Wang[2], Xiaozhen Han[2], Humberto Choi[1], Meredith Seeley[1], Richard Scherer[3] and Victoria Doseeva[3]

[1]Respiratory Institute, Cleveland Clinic, Cleveland, OH, USA. [2]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. [3]20/20 GeneSystems, Rockville, MD, USA.

ABSTRACT

BACKGROUND: A panel of 3 serum proteins and 1 autoantibody has been developed to assist with the detection of lung cancer. We aimed to validate the accuracy of the biomarker panel in an independent test set and explore the impact of adding a fourth serum protein to the panel, as well as the impact of combining molecular and clinical variables.

METHODS: The training set of serum samples was purchased from commercially available biorepositories. The testing set was from a biorepository at the Cleveland Clinic. All lung cancer and control subjects were >50 years old and had smoked a minimum of 20 pack-years. A panel of biomarkers including CEA (carcinoembryonic antigen), CYFRA21-1 (cytokeratin-19 fragment 21-1), CA125 (carbohydrate antigen 125), HGF (hepatocyte growth factor), and NY-ESO-1 (New York esophageal cancer-1 antibody) was measured using immunoassay techniques. The multiple of the median method, multivariate logistic regression, and random forest modeling was used to analyze the results.

RESULTS: The training set consisted of 604 patient samples (268 with lung cancer and 336 controls) and the testing set of 400 patient samples (155 with lung cancer and 245 controls). With a threshold established from the training set, the sensitivity and specificity of both the 4- and 5-biomarker panels on the testing set was 49% and 96%, respectively. Models built on the testing set using only clinical variables had an area under the receiver operating characteristic curve of 0.68, using the biomarker panel 0.81 and by combining clinical and biomarker variables 0.86.

CONCLUSIONS: This study validates the accuracy of a panel of proteins and an autoantibody in a population relevant to lung cancer detection and suggests a benefit to combining clinical features with the biomarker results.

KEYWORDS: carcinoembryonic antigen, carbohydrate antigen 125, cytokeratin-19 fragment 21-1, New York esophageal cancer-1 antibody, hepatocyte growth factor

Introduction

Early detection and diagnostic biomarkers have the potential to improve lung cancer care. An accurate early detection biomarker could improve the efficiency of lung cancer screening, allowing more lung cancers to be found while minimizing the potential harms to those with a low likelihood of having lung cancer. An accurate diagnostic biomarker has the potential to improve the management of imaging findings, expediting therapy for early stage cancers while minimizing the risks from evaluating those with benign disease.

Biomarker development proceeds through a series of phases toward the goal of proving clinical utility. Clinical utility suggests that the outcomes of patient management decisions improve with the use of the biomarker when compared with current standard practice. As even biomarkers that are felt to be accurate can lead to clinical decisions that harm some while benefiting others, it is crucial that clinical utility is demonstrated. Prior to embarking on a clinical utility study, a discovered biomarker proceeds through phases of analytical validation, to ensure that the assay is robust and precise, and clinical validation, to determine the accuracy of the biomarker in the intended use population.[1]

A handful of molecular biomarker panels have reached the phase of clinical validation.[2–4] The reported accuracies of these panels have been modest, leading to choices about whether to optimize sensitivity (rule-out test) or specificity (rule-in test). Recently, a panel of serum proteins and an autoantibody was developed based on prior evidence of an association between each of the components of the panel and the presence of lung cancer. The panel includes the following proteins: carcinoembryonic antigen (CEA), carbohydrate antigen 125 (CA125), cytokeratin-19 fragment 21-1 (CYFRA 21-1), and New York esophageal cancer-1 antibody (NY-ESO-1). Assays developed to measure the panel have been reported to have acceptable metrics of analytical validity, and a clinical validation study showed that the panel had a sensitivity of 74% at a specificity of 80%, using an algorithm that combines data from all 4 assays.[5] This accuracy and the relatively low cost of the assays required to measure this panel relative to most omics-based platforms suggest a potential role as a diagnostic marker for lung cancer. It is unclear whether this accuracy would be improved by adding additional tumor-related antigens to the panel. Hepatocyte

growth factor (HGF) levels have been associated with the presence and prognosis of lung cancer.[6,7] An inexpensive assay is available to measure HGF making it a reasonable antigen to explore for optimizing the accuracy of the panel.

The goals of this study were to confirm the accuracy of this panel of biomarkers on an independent data set, to explore the impact of adding a fourth serum protein (HGF) and to explore the accuracy relative to and in combination with clinical risk predictors with a focus on patients at risk of having lung cancer.

Methods

Training set serum samples

All of the cancer and normal control samples used in the training set were Institutional Review Board (IRB)-approved, consented serum samples that were purchased from the Clinical Research Center of Cape Cod, Inc. (Cape Cod, MA, USA), Asterand (Detroit, MI, USA), Indivumed (Hamburg, Germany), and Bioreclamation IVT (New York, NY, USA). All of the lung cancer samples were collected at physicians' offices or hospitals.

All lung cancer and control serum samples were from patients 50 years of age or older who were current or former smokers with a smoking history of greater than 20 pack-years and less than 15 years of smoking cessation. Details about the actual number of pack-years were not available. Diagnosis of the lung cancer cohort was confirmed from surgical pathology reports. The control group had no clinical evidence of current or prior cancer.

Testing set serum samples

All of the cancer and normal control samples used in the testing set were obtained from an IRB-approved blood biorepository at the Cleveland Clinic (IRB #10-521). All patients had provided written informed consent. All lung cancer cases were confirmed by biopsy. All samples were obtained prior to treatment of the cancer. Control samples were obtained from patients attending the lung cancer screening clinic or general pulmonary clinic. The control group had no clinical evidence of current or prior cancer.

Sample analysis

Multiplex magnetic bead–based immunoassay of CEA, CYFRA21-1, CA125, and HGF in patients' serum samples was performed using reagents from EMD Millipore, Inc. (Temecula, CA, USA), as previously described.[5] The MILLIPLEX MAP Human Circulating Cancer Biomarker Magnetic Bead Panel 1 was used. These 4 tumor proteins (CEA, CYFRA21-1, CA125, and HGF) were measured using the MAGPIX instrument (Luminex Corporation, Austin, TX, USA) as previously described.[5] Using median fluorescence intensity (MFI) values and a 5-parameter logistic curve fitting

method (xPONENT software for the MAGPIX), the concentrations of each tumor protein in the samples were calculated. The calculated protein concentration values were used for the subsequent analysis.

The NY-ESO-1 autoantibody detection was performed using an immunoassay developed at 20/20 GeneSystems (Rockville, MD, USA) and the MAGPIX reader, as previously described.[5] Background-subtracted MFI values were used for the subsequent analysis.

Statistical analysis

The study cohort was divided into 2 groups based on the outcome of cancer or control. The demographics, comorbidities, and cancer characteristics were described using sample mean with standard deviation or proportion as appropriate.

Multiple of the median analysis. The 5 biomarkers were tested separately and combined in 4- or 5-biomarker panels using a multiple of the median (MoM) approach where data from each biomarker were converted to a multiple of a population median value by dividing by the median value of the control group.[8] An area under the receiver operator characteristic (ROC) curve (AUC) analysis was used to estimate the diagnostic power of each panel and to determine the sensitivity at an MoM cutoff value that yielded 80% specificity in the training set. The accuracy of the test was validated using this cutoff value to determine the sensitivity and specificity in the testing set. Additional cutoff values were explored.

Multivariate logistic regression analysis. To determine the direction and statistical significance of the effect of each biomarker on the outcome, we performed multivariate logistic regression (MLR) analysis for the full training set and applied the model developed to the testing set. The AUC was calculated for the ROC curve that was constructed based on the model. In addition, exploratory MLR analyses were performed on the testing set, divided by stage and histology, and after including clinical variables. Clinical variables included age, sex, a clinical diagnosis of chronic obstructive pulmonary disease (COPD), and smoking history.

Random forest analysis. Random forest (RF) models were used to identify the variables that were associated with and predictive of cancer.[9] To avoid the possible overfitting of the MLR models, we used the repeated random-split cross-validation procedure.[10] Specifically, we randomly split the data into training (70%) and testing (30%) sets 100 times. The RF model was built on each training set and then evaluated on the corresponding test set. The validation results were reported as the average performance over all 100 test sets. Exploratory RF analyses were performed on the testing set, divided by stage and histology, and after including clinical variables (as described above).

Table 1. Clinical characteristics of the patients with cancer and controls in the training and testing sets.

| | TRAINING (604) | | TESTING (400) | |
	CANCER (268)	CONTROL (336)	CANCER (155)	CONTROL (245)
Age	64.0	64.5	65.3	68.3
Sex (% F)	43.7	39.9	40	51.9
Smoking (C/F/N)	NA	NA	20/129/6	95/142/7
Pack-years	>20	>20	43	35
Adenocarcinoma, %	56.3		47.7	
Squamous, %	33.2		39.4	
Stage I, %	53.7		33.5	
Stage II, %	24.3		12.3	
Stage III, %	17.9		37.4	
Stage IV, %	4.1		16.8	

Abbreviation: NA, not applicable.

Table 2. Performance of the individual BMs and panels of BMs in the training and testing sets.

| BM | TRAINING SET | | TESTING SET | |
	AUC	*P* VALUE	AUC	*P* VALUE
CEA	0.71	<.0001	0.70	<.0001
CA125	0.69	<.0001	0.67	<.0001
CYFRA	0.55	.06214	0.68	<.0001
HGF	0.65	<.0001	0.66	<.0001
NY-ESO-1	0.64	<.0001	0.45	.09453
4 BM: CEA, CA125, CYFRA, NY-ESO-1	0.75	<.0001	0.77	<.0001
5 BM (+HGF)	0.76	<.0001	0.80	<.0001

Abbreviations: AUC, area under the receiver operator characteristic curve; BM, biomarker; CA125, carbohydrate antigen 125; CEA, carcinoembryonic antigen; CYFRA 21-1, cytokeratin-19 fragment 21-1; HGF, hepatocyte growth factor; NY-ESO-1, New York esophageal cancer-1 antibody.
The listed panel accuracies are from the multiple of the median method calculations.

Results

The training set consisted of 604 patient samples (268 with lung cancer, 336 controls). Of the 268 patients with lung cancer, 151 (56.3%) were adenocarcinoma and 144 (53.7%) were stage I. The testing set consisted of 400 patient samples (155 with lung cancer, 245 controls). Of the 155 patients with lung cancer, 74 (47.7%) were adenocarcinoma and 52 (33.5%) were stage I (Table 1).

The combination of the biomarkers studied was more accurate than the individual biomarkers considered alone (AUC: 0.75-0.80 vs 0.45-0.71) (Table 2). Using the MoM method on the training set, a test result threshold with a specificity of 80% was established. This threshold was applied to the testing set to reveal a sensitivity of 49% and a specificity of 96% for both the 4-biomarker panel (without HGF) and the 5-biomarker panel (with HGF). In exploratory analysis, adjusting the threshold to obtain 80% specificity in the testing set led to a sensitivity of 65%. The highest combination of sensitivity and specificity for the 4-biomarker panel was 59% and 90% and for the 5-biomarker panel was 64% and 85%. Evaluation of the accuracy by stage and histology in the testing set showed improved sensitivity as the stage increased and relatively consistent accuracy across histologies (Table 3).

An MLR model was built on the training set then applied to the testing set. The AUC of the MLR model built on the training set was 0.77. This MLR model applied to the testing set yielded an AUC of 0.74 with a sensitivity of 39% and a specificity of 97%. In exploratory analysis, an MLR model built on the testing set alone had an AUC of 0.81 with a sensitivity of 53% and a specificity of 93%. Random forest modeling of

the testing set alone yielded an average AUC of 0.84 with an average sensitivity and specificity of 66% and 86%. An MLR model built from clinical variables in the testing set (age, sex, COPD, smoking history) had an AUC of 0.68. When clinical variables were combined with the 5-biomarker panel, the AUC was 0.86 (Table 4). The addition of clinical variables to biomarker results showed the largest improvement in AUC in patients with stage I (0.64-0.81 in MLR modeling and 0.67-0.79 in RF modeling).

Discussion

This study attempted to validate the accuracy of a combined protein and antibody panel in a population at risk of having lung cancer, determine whether the inclusion of an additional protein to the panel could improve the accuracy, and explore the impact of combining clinical and biomarker variables on test accuracy. The results suggest that the combination of markers is more accurate than any of the markers alone. The accuracy of the panel was slightly lower than in the previous report, and the addition of HGF to the panel did not substantially improve its accuracy. The biomarker was more accurate in

late-stage disease than early-stage disease. In exploratory analysis, the highest accuracy was achieved by combining clinical features and biomarker results.

In previous work, a training set of 230 patients and testing set of 150 patients was used to assess the 4-biomarker panel. The overall sensitivity was reported to be 74% at a specificity of 80%.[5] This is compared with the current validation results, where the sensitivity was 49% at 96% specificity in the testing set when using the MoM method. Accuracies for early-stage disease in the prior trial are superior to those in the current report (71% sensitivity at 83% specificity for stage I-II compared with 52%-63% sensitivity at 80% specificity in this study). The reasons for these differences, as well as differences in the CYFRA and HGF levels between the 2 sets, are not clear but could represent variability inherent to the sources of the samples. This highlights the need for multiple rounds of validation of biomarker accuracies in different settings. Other panels of established protein cancer biomarkers have reported similarly promising results.[11,12] A recent large study using a different panel of proteins showed a sensitivity of 89% at a specificity of 82%. The patient population of that study differed from our study in that all controls were symptomatic, controls had a lower risk of having lung cancer, and 52% of the population had stage IV disease.[11] Our study also highlights the challenge of optimizing a biomarker panel by trying to add potentially useful markers to the panel. Although reasonably accurate as a stand-alone marker, HGF was not able to substantially add to the accuracy of the panel, suggesting overlap with the performance of other markers. This is a common experience of biomarker developers.[13]

The intended use population for this study was patients at risk of having lung cancer. All participants in the training set were at least 50 years old and had smoked 20 pack-years or more. The age range of the testing set was similar and the group had a substantial smoking history. To pursue further validation and clinical utility testing, it should be determined whether the results of this study support further development of this biomarker as a lung cancer diagnostic. To justify moving forward, the accuracy of the test should support the potential

Table 3. Accuracy of BM panels by stage and histology.

STAGE	SENSITIVITY (%) AT 80% SPECIFICITY	
	4-BM	5-BM
I (n = 52)	52	52
II (n = 19)	63	53
III (n = 58)	67	71
IV (n = 26)	85	85
Adenocarcinoma (n = 74)	70	68
Squamous (61)	57	59
SCLC (n = 13)	62	69

Abbreviations: BM, biomarker; SCLC, small cell lung cancer.

Table 4. Accuracy of a clinical model built using the testing set of 400 subjects (including age, sex, chronic obstructive pulmonary disease, smoking history) compared with and combined with the 5-BM panel.

	LR MODEL	RANDOM FOREST 70:30 SPLIT			
	AUC	SENSITIVITY, % (SD)	SPECIFICITY, % (SD)	ACCURACY, % (SD)	AUC (SD)
Clinical	0.68	34 (0.08)	85 (0.07)	65 (0.03)	0.66 (0.04)
5-BM test	0.81	66 (0.06)	86 (0.04)	78 (0.03)	0.84 (0.03)
Combined	0.86	67 (0.07)	88 (0.04)	79 (0.03)	0.86 (0.03)

Abbreviations: AUC, area under the receiver operating characteristic curve; BM, biomarker; LR, logistic regression.
Sensitivity, specificity, accuracy, and AUC are averages of 100 iterations of a 70:30 training:testing split.

clinical application. To estimate the accuracy required to justify investment in a clinical utility study, a formula has been suggested that incorporates the accepted benefit:harm balance of current standard practice (the formula states that sensitivity/$(1 - \text{specificity}) \geq [(1 - \text{prevalence})/\text{prevalence}] \times \text{harm/benefit}$; harm/benefit can be expressed as $1/N$, where N is the number of control subjects testing positive that is tolerated to benefit one case subject testing positive).[14] Currently, we accept the balance of benefit:harm in lung cancer screening for a population with a 0.83% incidence of lung cancer (ie, the incidence of lung cancer during the screening years of the National Lung Screening Trial[15]; 120 control subjects test positive for each positive case subject). With this accepted standard, we can use this formula to determine a test accuracy that would allow us to use the results of a biomarker to expand the eligibility criteria for lung cancer screening to a population with a 0.4% incidence of lung cancer. The calculation reveals that the positive likelihood ratio (PLR) (ie, sensitivity/$(1 - \text{specificity})$) of the test would have to be at least 2.1 (eg, a sensitivity of 50% at a specificity of 88% or a sensitivity of 70% at a specificity of 83%). Based on this estimate, the accuracy of the biomarker panel in this study approaches values that could justify further validation of the accuracy of the test in the context of lung cancer screening (overall sensitivity of 49% at a specificity of 96%, PLR: 12.3; stage I sensitivity of 52% at a specificity of 80%, PLR: 2.6). In addition, the cost of this test is much lower than most omics-based testing platforms currently available. Cost is particularly important to consider when developing a screening test. The next steps in the development of this panel include larger clinical validation studies in the intended use population, followed by clinical utility studies if supported by the validation results. Similarly, one might consider applying this test to help with the evaluation of indeterminate lung nodules. Although not directly assessed in this project, accuracies of the biomarker alone for stage I disease may approximate the accuracy within a nodule population. Combining the biomarker with clinical and imaging variables may optimize a nodule risk prediction tool.

The strengths of this study include a reasonably large number of samples from a cohort relevant to potential clinical applications, with samples obtained from more than one source. The sample sets included a substantial portion of cases with early-stage disease, and a diverse set of relevant patient comorbidities, supporting the robustness of the method. The results were compared with and were more accurate than clinical prediction, and the combination of the marker results with clinical features improved the accuracy of both. A weakness of the study was that there was less information available about the quality of the training samples and there was less metadata available for these samples, so exploratory analysis was performed only on the testing set.

This study validates the accuracy of a panel of proteins and an autoantibody in a population relevant to lung cancer detection and suggests a benefit to combining clinical features with the biomarker results.

Author Contributions
PJM takes responsibility for the content of the manuscript, including the data and analysis. PJM, X-FW, XH, HC, MS, RS, and VD substantially contributed to the conception or design of the work or the acquisition, analysis, or interpretation of data for the work. PJM, X-FW, XH, HC, and VD contributed to drafting the work or revising it critically for important intellectual content. PJM, X-FW, XH, HC, MS, RS, and VD finally approved the version submitted for publication and contributed to the accountability for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES
1. Pennello GA. Analytical and clinical evaluation of biomarkers assays: when are biomarkers ready for prime time? *Clin Trials*. 2013;10:666–676.
2. Silvestri GA, Vachani A, Whitney D, et al; AEGIS Study Team. A bronchial genomic classifier for the diagnostic evaluation of lung cancer. *N Engl J Med*. 2015;373:243–251.
3. Jett JR, Peek LJ, Fredericks L, Jewell W, Pingleton WW, Robertson JF. Audit of the autoantibody test, EarlyCDT®-Lung, in 1600 patients: an evaluation of its performance in routine clinical practice. *Lung Cancer*. 2014;83:51–55.
4. Vachani A, Hammoud Z, Springmeyer S, et al. Clinical utility of a plasma protein classifier for indeterminate lung nodules. *Lung*. 2015;193:1023–1027.
5. Doseeva V, Colpitts T, Gao G, Woodcock J, Knezevic V. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *J Transl Med*. 2015;13:55.
6. Siegfried JM, Weissfeld LA, Luketich JD, Weyant RJ, Gubish CT, Landreneau RJ. The clinical significance of hepatocyte growth factor for non-small cell lung cancer. *Ann Thorac Surg*. 1998;66:1915–1918.
7. Tsuji T, Sakamori Y, Ozasa H, et al. Clinical impact of high serum hepatocyte growth factor in advanced non-small cell lung cancer. *Oncotarget*. 2017;8: 71805–71816.
8. Bishop JC, Dunstan FD, Nix BJ, Reynolds TM, Swift A. All MoMs are not equal: some statistical properties associated with reporting results in the form of multiples of the median. *Am J Hum Genet*. 1993;52:425–430.
9. Breiman L. Random forests. *Machine learning*. 2001;45:5–32.
10. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res*. 2008;14:5977–5983.
11. Molina R, Marrades RM, Auge JM, et al. Assessment of a combined panel of six serum tumor markers for lung cancer. *Am J Respir Crit Care Med*. 2016;193: 427–437.
12. Daly S, Rinewalt D, Fhied C, et al. Development and validation of a plasma biomarker panel for discerning clinical significance of indeterminate pulmonary nodules. *J Thorac Oncol*. 2013;8:31–36.
13. Macdonald IK, Murray A, Healey GF, et al. Application of a high throughput method of biomarker discovery to improvement of the EarlyCDT®-Lung test. *PLoS ONE*. 2012;7:e51002. doi:10.1371/journal.pone.0051002.
14. Pepe MS, Janes H, Li CI, Bossuyt PM, Feng Z, Hilden J. Early-phase studies of biomarkers: what target sensitivity and specificity values might confer clinical utility? *Clin Chem*. 2016;62:737–742.
15. Aberle DR, Adams AM, Berg CD, et al; The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.