Development of a Protein Biomarker Panel to Detect Non–Small-Cell Lung Cancer in Korea

Young Ju Jung,[1,2] Evaldas Katilius,[3] Rachel M. Ostroff,[3] Youndong Kim,[4] Minkyoung Seok,[4] Sujin Lee,[4] Seongsoo Jang,[5] Woo Sung Kim,[1] Chang-Min Choi[1]

Abstract

We investigated candidate biomarkers for the Korean population using a new aptamer-based proteomic technology and developed a 7-protein panel that discriminates lung cancer from controls. This 7-marker panel was able to discriminate lung cancer from controls with an area under the curve of 0.82/0.77 in the train/verification set respectively. This panel may have clinical utility in risk-stratifying screen-detected lung nodules.

Background: Lung cancer screening using low-dose computed tomography reduces lung cancer mortality. However, the high false-positive rate, cost, and potential harms highlight the need for complementary biomarkers. We compared the diagnostic performance of modified aptamer-based protein biomarkers with Cyfra 21-1. **Patients and Methods:** Participants included 100 patients diagnosed with lung cancer, and 100 control subjects from Asan Medical Center (Seoul, Korea). We investigated candidate biomarkers with new modified aptamer-based proteomic technology and developed a 7-protein panel that discriminates lung cancer from controls. A naive Bayesian classifier was trained using sera from 75 lung cancers and 75 controls. An independent set of 25 cases and 25 controls was used to verify performance of this classifier. The panel results were compared with Cyfra 21-1 to evaluate the diagnostic accuracy for lung nodules detected by computed tomography. **Results:** We derived a 7-protein biomarker classifier from the initial train set comprising: EGFR1, MMP7, CA6, KIT, CRP, C9, and SERPINA3. This classifier distinguished lung cancer cases from controls with an area under the curve (AUC) of 0.82 in the train set and an AUC of 0.77 in the verification set. The 7-marker naive Bayesian classifier resulted in 91.7% specificity with 75.0% sensitivity for the subset of individuals with lung nodules. The AUC of the classifier for lung nodules was 0.88, whereas Cyfra 21-1 had an AUC of 0.72. **Conclusion:** We have developed a protein biomarker panel to identify lung cancers from controls with a high accuracy. This integrated noninvasive approach to the evaluation of lung nodules deserves further prospective validation among larger cohorts of patients with lung nodules in screening strategy.

Clinical Lung Cancer, Vol. 18, No. 2, e99-107
Keywords: Aptamer, Biomarker, Cyfra 21-1, Lung cancer screening, Lung nodule

Introduction

Lung cancer remains the leading cause of cancer-related deaths in Korea and the world.[1] It is estimated that, in 2015, there will be 25,640 new cases and over 17,559 deaths from lung cancer in Korea alone.[2] Although lung cancer death rates are declining in developed countries, lung cancer incidence and death rates are rapidly rising in the developing world, where the smoking prevalence continues to increase.[1]

Because most people with early stage lung cancer are asymptomatic, more than 60% of patients are diagnosed at advanced stages when a cure is unlikely.[3,4] The 5-year survival rate for patients with advanced disease can be less than 10%, whereas the 5-year survival rate in patients with stage I disease may be greater than 70%.[5] Because early detection of lung cancer is critical to reducing mortality and morbidity, significant research efforts are focused on the development of accurate lung cancer screening.

Despite the 20% relative reduction in lung cancer-related mortality shown by low-dose chest computed tomography (LDCT)

[1]Department of Pulmonary and Critical Care Medicine
[2]Health Promotion Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea
[3]SomaLogic, Boulder, CO
[4]Aptamer Sciences Inc., Pohang, Korea
[5]Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Submitted: Apr 7, 2016; Revised: Sep 6, 2016; Accepted: Sep 6, 2016; Epub: Oct 5, 2016

Address for correspondence: Chang-Min Choi, MD, Department of Pulmonary and Critical Care Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-Gil, Songpa-gu, Seoul, Korea 05505
E-mail contact: ccm9607@gmail.com

1525-7304/$ - see frontmatter
http://dx.doi.org/10.1016/j.cllc.2016.09.012

screening in the National Lung Screening Trial (NLST),[6] 24.2% of screening LDCT scans were positive, with 96.4% of observed nodules determined to be false positives.[6,7] In addition to detecting aggressive tumors, more than 18% of all lung cancers detected by LDCT in the NLST appear to be indolent, and over-diagnosis and over-treatment are a concern when evaluating the risks of LDCT screening for lung cancer.[8]

Sensitive and specific lung cancer biomarkers, measured in a noninvasively collected biological specimen such as serum, could help guide clinical decision-making in high-risk subjects, particularly in patients with CT-detected pulmonary nodules. Published data on clinically available serum biomarkers, most notably cytokeratin 19 fragment 21.1 (Cyfra 21-1), carcinoembryonic antigen, and tissue plasminogen activator in non−small-cell lung cancer (NSCLC), show limited sensitivity and specificity, particularly in early-stage disease.[9-11] The poor performance of univariate analyses in this study limits the use of single protein biomarkers in the lung cancer diagnostic field. It is widely thought that using multi-marker algorithm biomarkers may accelerate the development of protein-based diagnostic tools. Advances in molecular diagnostics and understanding of genomics have led to the discovery of several lung cancer biomarkers with potential to complement the current screening standards.[12-15] However, developing multi-protein biomarkers is limited owing to coverage, precision throughput, pre-analytical variability, and cost.[16] Aptamers, defined as single chain nucleotide ligands, were developed as a novel capture array that shares many features with antibodies. Owing to its innate property, aptamer technology enables developing a robust, highly multiplexed proteomic assay to analyze multiple proteins in complex biological samples. Recently, Gold and colleagues have developed an aptamer-based high-throughput proteomics assay, SOMAscan.[17] Their team has discovered 44 candidate biomarkers using SOMAscan technology for NSCLC and constructed the 12-marker naive Bayesian classifier.[18] However, in the following validation study, Mehan and colleagues discovered that some of these biomarkers were variable, and instead selected 15 robust biomarkers using SOMAscan and built a 7-marker Random Forest classifier.[19] These studies have supported the application of aptamers in the development of lung cancer biomarkers in the US population with the potential to complement the current screening standards.

However, the epidemiology and molecular biology of lung cancer may differ according to ethnic or geographic backgrounds. This is the first report of a modified aptamer-based assay developed to detect NSCLC in the context of CT screening for the Korean population.

The objective of our study was to develop and verify a blood-based protein biomarker panel for lung cancer detection and to compare its diagnostic performance with conventional Cyfra 21-1. In addition, we evaluated the ability of the panel to distinguish benign from malignant lung nodules.

Patients and Methods

Study Design

The objectives of this study were to discover biomarkers that discriminate NSCLC from benign lung diseases and healthy controls, to train a multi-biomarker classifier of NSCLC to meet pre-specified performance criteria, and to verify the performance of this classifier in a separate set of blinded samples.

The results from these diagnostic case-control studies were compared with Cyfra 21-1, and evaluated for the ability to enhance the diagnostic accuracy of lung nodules detected on CT. For the initial training study, sera from 75 patients with NSCLC and 75 controls were tested. A verification set consisted of an independent set of 25 primary lung cancer cases and 25 controls. A verification set was blinded and randomly shuffled by using the 'RAND()' function in Excel 2013 (Microsoft) and unblinded after the classifier fixed by train set.

Lung Cancer Cases

Between March 2012 and September 2012, we enrolled 100 patients with biopsy-proven primary lung cancer. Study participants were enrolled in the Asan Medical Center (Seoul, Korea) lung research registry, after obtaining informed consent under institutionally approved clinical research protocols. The case population included patients who had positive findings via CT imaging and who were diagnosed with pathologic or clinical stage I to IV NSCLC. Demographic data, including gender, age at diagnosis, smoking status, and other clinical information, were provided by medical record. Blood samples were collected within 4 weeks of the first biopsy-proven lung cancer diagnosis and before treatment or removal of the tumor by standard surgical procedures.

Controls

The control population was selected to represent subjects who received a lung health screening exam in the health promotion center of the Asan Medical Center. During August 2012, 100 controls consenting to the study were recruited. All participants received a chest CT scan, and provided the demographic data collected by self-report questionnaires. Peripheral blood sampling was collected simultaneously with the CT scan. All participants were determined to be cancer-free after a minimum of 1-year follow-up.

Sample Collection and Processing

All serum specimens were collected following uniform protocols recommended by the National Cancer Institute's Early Detection Research Network.[20] All samples were allowed to clot, and serum was recovered by centrifugation within 2 to 8 hours of collection and stored at −80°C. Samples were thawed once for aliquoting prior to proteomic analysis.

Candidate Biomarker Measurement

The modified aptamer-based multiplex assay was performed essentially as described.[19] Briefly, a biotin moiety is linked via a photocleavable linker to each modified aptamer. The modified aptamer mixture is mixed with diluted serum, and the protein binding step is performed by incubating at 37°C for 3 hours. Then the mixture is transferred to a streptavidin-coated microtiter plate and incubated for 30 minutes to capture aptamer-protein complexes via the biotin tag on the aptamers. A series of wash steps are performed to remove unbound proteins in the sample. Then an amine-reactive biotin reagent is added to the microtiter plate to tag the captured proteins with biotin. Then aptamer-protein complexes are released to solution by photocleaving the linker using long-wavelength UV irradiation (~365 nm). The solution containing aptamer-protein complexes is then transferred to fresh streptavidin-coated plates,