



Published in final edited form as:

*J Proteome Res.* 2013 August 2; 12(8): 3689–3696. doi:10.1021/pr400274w.

## Glycoproteomic Analysis of Bronchoalveolar Lavage (BAL) Fluid Identifies Tumor-associated Glycoproteins from Lung Adenocarcinoma

Qing Kay Li, MD PhD<sup>1</sup>, Punit Shah, PhD<sup>1</sup>, Yan Li, PhD<sup>1</sup>, Paul O Aiyetan, MD MS<sup>1</sup>, Jing Chen, PhD<sup>1</sup>, Rex Yung, MD<sup>2</sup>, Daniela Molena, MD<sup>3</sup>, Edward Gabrielson, MD<sup>1</sup>, Frederic Askin, MD<sup>1</sup>, Daniel W Chan, PhD<sup>1</sup>, and Hui Zhang, PhD<sup>1</sup>

<sup>1</sup>Department of Pathology, The Johns Hopkins Medical Institutions, Baltimore, MD21287

<sup>2</sup>Department of Medicine, The Johns Hopkins Medical Institutions, Baltimore, MD21287

<sup>3</sup>Department of Surgery, The Johns Hopkins Medical Institutions, Baltimore, MD21287

### Abstract

Cytological examination of cells from bronchoalveolar lavage (BAL) is commonly used for the diagnosis of lung cancer. Proteins released from lung cancer cells into BAL may serve as biomarkers for cancer detection. In this study, *N*-glycoproteins in 8 cases of BAL fluid, as well as 8 lung adenocarcinoma tissues and 8 tumor-matched normal lung tissues, were analyzed using the solid-phase extraction of *N*-glycoprotein (SPEG), iTRAQ labeling and liquid chromatography tandem mass spectrometry (LC-MS/MS). Of 80 glycoproteins found in BAL specimens, 32 were identified in both cancer BAL and cancer tissues with levels of 25 glycoproteins showing at least a 2-fold difference between cancer and benign BAL. Among them, 8 glycoproteins showed greater than 2-fold elevations in cancer BAL, including Neutrophil elastase (NE), Integrin alpha-M, Cullin-4B, Napsin A, Lysosome-associated membrane protein 2 (LAMP2), Cathepsin D, BPI fold-containing family B member 2, and Neutrophil gelatinase-associated lipocalin. The levels of Napsin A in cancer BAL were further verified in an independently collected 39 BAL specimens using an ELISA assay. Our study demonstrates that potential protein biomarkers in BAL fluid can be detected and quantified.

### Keywords

glycoproteomics; bronchoalveolar lavage (BAL); glycoproteins; lung adenocarcinoma; non-small cell lung cancer (NSCLC)

---

Address Correspondence to: Qing Kay Li MD PhD, Assistant Professor, Department of Pathology, The Johns Hopkins Bayview Medical Center, 4940 Eastern Ave., Baltimore, MD 21224, Tel: (410)550-0671, Fax: (410)550-0075, qli23@jhmi.edu.

### Conflict of interest

The authors declare no conflict of interest.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Introduction

Lung cancer is the leading cause of cancer-related death in the United States and worldwide<sup>1,2</sup>. Currently, the diagnosis of lung mass involves the combination of radiological and histological evaluation of the lesion. Bronchoscopy is a minimally invasive procedure and commonly used for obtaining lesional tissue for histological examination<sup>3-6</sup>. Several types of specimens can be obtained during the bronchoscopic procedure, such as transbronchial fine needle aspiration biopsy (TBNA) with ultrasound guidance (EBUS-TBNA) or without ultrasound guidance, bronchial brushing and bronchoalveolar lavage (BAL) specimens<sup>3-6</sup>. These specimens/tests are not always conclusive in differentiating small benign nodules from cancer<sup>3-6</sup>. Molecular markers in the sputum, bronchial epithelium and blood are being evaluated<sup>7-11</sup>, and although several recent studies show promise of biomarkers in the improvement of the diagnostic specificity of lung cancer<sup>12-14</sup>, these biomarkers are still not being used clinically. Therefore, aggressive surgical procedures, such as video-assisted thoracoscopic biopsy and thoracotomy, are frequently performed in patients with suspicious lung lesions. These invasive diagnostic procedures are associated with a 15.9 to 32.4% complication rate and are very costly<sup>15, 16, 17</sup>.

There is evidence that proteins in BAL fluid reflect the physiological and pathological status of the lung<sup>18, 19</sup>. Quantitative proteomic analysis of BAL fluid has been used to study certain benign lung diseases such as asthma, cystic fibrosis, and interstitial lung disease<sup>18, 19</sup>, but the protein profile in lung cancers has not been well studied<sup>20, 21</sup>. The majority of proteins either secreted from cells or located on extracellular surface of cells are glycoproteins. Glycoproteins play important roles in the regulation of cellular functions, including cell differentiation, proliferation, cellular interactions with their surrounding environment, and invasion or metastasis of tumor cells<sup>22</sup>. In addition, most FDA approved clinical tumor markers are glycoproteins, such as carcinoembryonic antigen (CEA), CA125, CA19-9 and prostate specific antigen (PSA). Glycoproteomic analysis of lung cancer tissue for discovery of tumor-associated glycoproteins is now quite approachable<sup>12, 23-26</sup>.

In the past decade, several highly sensitive and quantitative proteomic techniques, such as high-content quantitative proteomic using liquid chromatography tandem mass spectrometry (LC-MS/MS) and spectral count, isobaric tags for relative and absolute quantitation (iTRAQ) labeling, and the multiple-reaction monitoring (MRM) have been developed<sup>27-29</sup>. These techniques have an increased sensitivity and throughput capability of accurate analysis of biological samples. For example, iTRAQ labeling technique, using stable isotope to label samples, allows for an accurate quantitation of the peptide abundance in biological samples by direct comparison of light and heavy tags from peptides in different samples in the same spectrum, whereas, the LC-MS/MS method determines the peptide abundance by spectral count based on the number of redundant spectra for each protein. The combination of these techniques could increase the detection of low abundance proteins in biological samples.

In this study, we analyzed the glycoprotein profile in discarded BAL fluid specimens collected from primary lung adenocarcinoma and benign lung disease patients using glycoproteomic approaches. We compared the relative abundance of glycoproteins in cancer

BAL with those in lung cancer tissue to identify potential tumor-associated glycoproteins. We also developed a targeted ELISA assay to show the detection of Napsin A in individual BAL specimens, as example to verify our observation in proteomic study of BAL. Although a large scale study is still needed, our data highlight the potential utility of discarded BAL fluid from cytological laboratory in the discovery of lung cancer biomarkers and potential utility of BAL for lung cancer diagnosis.

## Material and Methods

### Collection of BAL specimens and lung tumor tissues

BAL specimens were collected during bronchoscopic procedure from patients with primary lung adenocarcinomas (ADC) or benign lung diseases. During each procedure, 10 cc of normal saline was used to rinse the lesional area; then fluids were collected through the bronchoscopy into a sterilized test tube and sent to the cytological laboratory for the evaluation of tumor cells. After the examination of lung cells, discarded BAL fluids were collected from cytological laboratory. Pooled samples were used for the glycoproteomic analysis. Of 8 BAL specimens, 4 cases were cytologically diagnosed as “benign respiratory epithelium, negative for malignancy”; and 4 cases were cytologically diagnosed as “adenocarcinoma or poorly differentiated adenocarcinoma”. In our study, caution was taken not to use BAL specimens after fine needle aspiration (FNA) biopsy, to minimize blood contamination.

For a validation study, 39 independently collected BAL specimens, including 18 cases of primary lung ADCs, 6 cases of benign lung diseases, 9 cases of primary lung squamous cell carcinomas (SQCC), 6 cases of small cell lung carcinomas (SCLC), were used for the ELISA assay. All BAL specimens were centrifuged to remove cellular debris and were kept in a  $-80^{\circ}\text{C}$  freezer prior to proteomic analysis.

In addition, lung adenocarcinoma and tumor-matched normal lung tissues were obtained from surgical resected specimens and included in our study. They were 4 cases of early stage (pT1 and pT2), 4 cases of late stage (pT3 and pT4) cancers, and tumor-matched normal lung tissues. All tissues were fixed in formalin and embedded in paraffin (FFEP). Tumor cells and tumor-matched normal tissue were microdissected for *N*-glycoproteomic analysis. The pathological stage and subtype of lung carcinoma were classified according to American Joint Committee on Cancer (AJCC) staging manual<sup>30</sup>, World Health Organization (WHO) and International Association for the Study of Lung Cancer/American Thoracic Society classification<sup>31</sup>.

All samples were annotated with available clinical information in a manner that protected patient identity. Our study was approved by the Johns Hopkins Medical Institution Review Board (IRB).

### Materials

Hydrazide resin (Bio-Rad, Hercules, CA), Sodium periodate (Bio-Rad, Hercules, CA), Tris (2-carboxyethyl) phosphine (TCEP) (Pierce, Rockford IL), PNGase F (New England Biolabs, Ipswich, MA), Sequencing grade trypsin (Promega, Madison, WI), C18 columns

(Waters, Sep-Pak Vac),  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) (Agilent, Palo Alto, CA), MALDI 4700 mass calibration standards (Applied Biosystems, Foster City, CA), and other chemicals were purchased from Sigma-Aldrich. iTRAQ reagent and mass calibration standards were from Applied Biosystems (Foster City, CA); BCA assay kit was from Pierce (Rockford, IL); SCX columns and C18 resin were from Sepax (Newark, DE).

### Peptide extraction from FFPE tumor tissues and BAL specimens

The formalin-fixed paraffin-embedded (FFPE) tumor and tumor-matched normal lung tissue were cut into 10 micron in thickness, placed on a glass slide, and rehydrated prior to microdissection. The H&E stained sections were used to estimate the tumor content. The rehydrated tissue was microdissected and collected into a centrifuge vial with 25  $\mu$ l of PBS buffer. Tissues were sonicated for 5 min on an ice bath and centrifuged at 13,200 rpm for 5 min. 100 $\mu$ l of trifluoroethanol (TFE) was added to each sample and incubated for 2 hrs. Supernatant was collected by the centrifugation. The protein concentration in the supernatant and BAL specimens were measured using BCA protein assay kit (Thermo Fisher Scientific Inc., Rockford, Illinois). 100  $\mu$ g of proteins from pooled cancer and non-cancer tissues as well as pooled cancer and non-cancer BAL specimens were incubated with the trypsin digestion buffer (100 mM Tris-HCl, pH 7.5) containing 2  $\mu$ g of trypsin at 37°C overnight. The digested peptides were used for glycopeptide isolation.

### Isolation of formerly *N*-linked glycopeptides

The deglycosylated peptides of formerly *N*-linked glycopeptides were isolated using the solid-phase extraction of *N*-glycoprotein (SPEG) method as described previously<sup>26</sup>. Briefly, tryptic peptides from each sample of tissues and BAL specimens were cleaned with C18 columns. Samples were eluted from C18 column in 50% of acetonitrile and oxidized by sodium periodate to a final concentration of 10 mM and incubated at room temperature for 1 hr in the dark, then, samples were conjugated to hydrazide resin at room temperature for 4 hr in 80% acetonitrile and 0.1 % TFA (trifluoroacetic acid). Non-glycosylated peptides were removed by washing the resin three times each with 800 $\mu$ l of 1.5M NaCl, H<sub>2</sub>O, and 100mM NH<sub>4</sub>HCO<sub>3</sub>. Formerly *N*-linked glycopeptides were released from the resin by incubation overnight with 1 $\mu$ l of PNGase F in 100 mM NH<sub>4</sub>HCO<sub>3</sub> at 37°C. The purified peptides were dried and resuspended into 40 $\mu$ l of 0.4% acetic acid and 10  $\mu$ l of formerly *N*-glycopeptides from cancer and non-cancer BAL specimens were used for LC-MS/MS analysis.

### iTRAQ labeling

The formerly *N*-linked glycopeptides from tissues and the remaining peptides from BAL samples were labeled by iTRAQ. iTRAQ labeling of peptides was performed as described previously<sup>32</sup>. Briefly, the iTRAQ 8-plex reagent was dissolved in 70 $\mu$ l of methanol. Dried tryptic peptides from each sample was resuspended into 20  $\mu$ l of iTRAQ dissolution buffer, then mixed with 70  $\mu$ l of iTRAQ 8-plex reagent and incubated for 1 hour at room temperature. After iTRAQ labeling, the reaction solutions were combined and cleaned by SCX column. Then, labeled peptides were dried and resuspended into 10  $\mu$ l of 0.4% acetic acid solution prior to mass spectrometry analysis.

## LC-MS/MS analysis

Label-free peptides and the iTRAQ labeled peptides were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS) using a Orbitrap Velos (Thermo Fisher Scientific Inc) interface with a nanoLC system (Eksigent, Dublin, California). 10  $\mu$ l of label-free peptides or iTRAQ labeled-peptides were loaded on a self-packed C18 column (75  $\mu$ m ID  $\times$  10 cm, Magic C18 5  $\mu$ m, 100A), and gradient eluted over 100 minutes at 300 nL/minute into the mass spectrometer. The HPLC mobile phase A and B were 0.2% formic acid in HPLC grade water and 0.2% formic acid in HPLC grade acetonitrile, respectively. The mobile phase B was increased from 5% to 40% in 90 min. LC-MS/MS data was obtained using a data dependent analysis of the top ten precursors and a dynamic exclusion of 30 seconds. Mass range for precursor ion acquisition was set at 350–1700 m/z with mass resolution of 30,000. All MS/MS were acquired using Orbitrap after HCD. In label free experiment, ion was isolated with 1.9 mass window and HCD collision energy for fragmentation was set at 35. In iTRAQ experiment, ion was isolated at 1.2 mass window with collision energy set at 45.

## LC-MS/MS data analysis

The acquired LC-MS/MS data was searched against the Homo sapiens taxonomy of the IPI Human database V3.52 (73,928 sequences) using Sequest (search algorithm within Proteome Discoverer, Thermo Scientific version 1.3) with following parameters: two missed cleavages allowed, trypsin as cleavage enzyme, a tolerance of 50 ppm on precursors and 0.8 daltons on the fragment ions. Modifications allowed include: carbamidomethylation of cysteines sets to static, oxidation of methionine and deamidation of asparagine and glutamine set to variable. Peptides were further filtered for consensus sequence of *N*-linked glycosylation motif NXS/T. Data was also searched against a decoy database and filtered with a 5% false discovery rate (FDR).

Formerly *N*-glycopeptides identified from individual BAL samples were further analyzed to obtain semi-quantitative data using spectral counts. The relative ratio of the abundance of glycoproteins between cancer and benign BAL samples was calculated. The relative ratio of iTRAQ-labeled tissues and BAL was from Proteome Discoverer, Thermo Scientific version 1.3.

## Enzyme-linked immunosorbent assay (ELISA)

Napsin A levels in BAL specimen were determined using an ELISA assay kit per manufacturer's protocol. Napsin A assay kit was purchased from IBL international (Toronto, ON, Canada). Briefly, 100  $\mu$ l of BAL specimens were incubated in a 96-well microtiter plate pre-coated with Napsin A capture antibody overnight at 4 °C, followed by incubation with a HRP conjugated mouse anti-human Napsin A antibody for 30 min at 4 °C. Substrate solution containing tetra methyl benzidine was then added and allowed to react for 20 min at room temperature (RT). The wells were thoroughly washed between incubations. Absorbance readings were acquired at 450 nm using an absorbance reader (Biotek, Winooski, VT). Standard curves were generated from a four-parameter logistic curve fit using recombinant human Napsin A (concentration range: 0–250 ng/mL). BAL samples with Napsin A concentrations over the range of 250 ng/mL were diluted at 1:20 or 1:50 and

re-measured. All specimens were assayed as duplicates. The mean of two values was obtained. Total protein concentration was determined for all BAL specimens using BCA protein assay (Thermo Fisher Scientific Inc., Rockford, Illinois). The levels of Napsin were expressed as ng/mg total proteins in BAL samples.

## Results

### Identification of glycoproteins in BAL specimens

In the study, we collected discarded BAL fluid specimens from cytological laboratory, including both cancer and benign cases. The morphological features of ADC, SQCC and SCLC were summarized in Figure 1. The clinical information of BAL specimens for LC-MS/MS analysis was summarized in Table 1. We analyzed the glycoprotein profiles in BAL specimens, using the following steps: 1) recover tryptic peptides from BAL specimens, 2) isolate formerly *N*-linked glycopeptides using SPEG method, 3) quantitatively analyze and compare isolated peptides between cancer and benign BAL specimens.

Formerly glycosylated peptides obtained using SPEG approach were deglycosylated and subsequently analyzed by LC-MS/MS. The acquired LC-MS/MS data was searched against IPI Human database as well as a decoy database and filtered with a 5% FDR. These identifications contained the consensus sequence of *N*-linked glycosylation N-X-S/T motif (X is any amino acid except proline). In BAL specimens, a total of 123 unique *N*-linked glycosites that contained at least one N-X-S/T motif were identified (Supplementary Table 1).

Peptides identified from BAL samples were further analyzed to obtain semi-quantitative data using spectral counts to determine the relative abundance for each protein. The MS/MS spectra were also used to search protein database SEQUEST. The 123 unique *N*-glycosites from 80 glycoproteins were identified in cancer or benign BAL specimens (Supplementary Table 1). For Napsin A, we were able to identify 31 peptide to spectra matches (PSMs) in cancer samples and 15 PSMs for benign samples.

### Further determine the lung cancer-derived glycoproteins in BAL using iTRAQ labeling

To further determine which of the glycoproteins in BAL specimens originated from lung cancer tissues, we analyzed glycoproteins from BAL specimens and lung cancer tissues using iTRAQ labeling approach, and compared the protein abundance in BAL specimens with the abundance in lung tumor tissues. It is well known that cellular proteins from lung tissues are expressed at different stage of the tumor; therefore, we included both early and late stage of lung adenocarcinomas tissues and their matched non-tumor tissues in our study. In the study, we analyzed pooled specimens from 6 groups: benign BAL, cancer BAL, early stage lung tumor tissues, late stage lung tumor tissues, and early and late stage of tumor-matched non-tumor lung tissues. Formerly *N*-glycopeptides from these 6 groups were isolated from each pooled specimens and labeled with iTRAQ. The iTRAQ labeled peptides mixture was analyzed by LC-MS/MS. A total of 32 glycoproteins were identified in both BAL specimens and tumor tissues (Table 2 and Supplement Table 2).

The analysis of glycoproteins using the combined approaches of SPEG capturing and iTRAQ labeling demonstrated a specific correlation of *N*-glycoproteins in cancer tissue and the presence of those tumor glycoproteins in cancer BAL specimens.

In spite of the fact that the same amount of protein from each specimen was analyzed using the same procedure, the levels of each glycoproteins were different between cancer and benign BAL as well as in early and late stage of lung tumor tissues. Of 32 glycoproteins identified in both BAL specimens and tumor tissues, 8 proteins showed greater than 2-fold elevations in cancer BAL, including Neutrophil elastase (NE), Integrin alpha-M, Cullin-4B, Napsin A, Lysosome-associated membrane protein 2 (LAMP2), Cathepsin D, BPI fold-containing family B member 2, and Neutrophil gelatinase-associated lipocalin (Table 2).

### Detection of Napsin A in BAL specimens by the ELISA assay

Of these 8 elevated proteins in cancer BAL, a particularly interesting one is Napsin A, which has been established as a marker for the identification of lung adenocarcinoma in tumor tissue<sup>30, 31</sup>. Identification of Napsin A in the BAL fluid using LC-MS/MS analysis is shown in Figure 2. We further investigated the expression of Napsin A in an independently collected BAL sample set using an ELISA assay, which included BAL samples from both NSCLC and SCLC as well as benign lung diseases. The clinical information of patients was summarized in Table 3. The protein concentration of individual specimen was also determined and used for the normalization of Napsin A levels. The mean levels of Napsin A in benign disease, ADC, SQCC and SCLC were 55.48±39.13 (mean ± SE), 295.23±104.56, 102.16±34.50 and 50.41±28.71 ng/mg total BAL proteins (Figure 3). There is a significant difference between benign and adenocarcinoma groups ( $P < 0.05$ ,  $P = 0.041$ , the Student's *T* test).

We also performed a receiver operating curve (ROC) analysis comparing the level of Napsin A in benign lung disease with ADC group (Figure 4), that revealed an area under the curve (AUC) of 0.846. Using an arbitrary cutoff at the level of 55 ng/mg total BAL protein (the mean level of Napsin A in benign lung diseases) resulted in a sensitivity of 84.21%, specificity of 66.67%, apparent positive predictive value (PPV) of 84.21% and apparent negative predictive value (NPV) of 66.67% for distinguishing benign and ADC in BAL fluids.

### Discussion

In the study, we compared *N*-glycoprotein profile in BAL specimens with that of tumor tissues and tumor-matched normal lung tissues. There are several advantages of our strategy of focusing on the identification and verification of glycoproteins using both tumor tissue and BAL specimen. First, the glycopeptides-capturing technique can enrich low abundance extracellular tumor-derived proteins in the BAL. Second, SPEG glycopeptide capturing reduces the complexity of the BAL proteome by eliminating the dominant serum proteins, which are typically non-glycoproteins. Third, iTRAQ isotopic labeling of peptides from BAL and tissue can further verify glycoproteins from both samples and facilitate the detection of low abundance tumor-derived proteins within complex BAL. We were able to identify 80 unique *N*-glycoproteins in both cancer and benign BAL specimens. Our study

provides evidence that BAL specimen represents a lung compartment-specific biofluid and can be analyzed for biomarkers discovery. Furthermore, in the United States, 94 million individuals who are current or former smokers are at high risk of developing lung cancer, and the majority of smokers harbor small lung nodules<sup>17</sup>. Thus, it is becoming increasingly important to distinguish small benign nodules from cancer with minimally invasive procedures. Our quantitative proteomic analysis of BAL fluid was useful for characterizing the complex alveolar microenvironment in lung cancers and facilitating the identification of tumor-associated protein biomarkers; and for the separation of benign from malignant lesions.

Of 80 *N*-glycoproteins identified in BAL, a particularly interesting one is Napsin A. Our previously published study and others have shown that Napsin A is a novel lung tissue marker and expressed in lung adenocarcinomas<sup>33, 34</sup>. We conducted an ELISA assay using an independent set of BAL samples; and found an elevated Napsin A level in adenocarcinoma BAL in comparison to benign lung disease BAL. The ROC analysis comparing the level of Napsin A in benign lung disease with ADC group, revealed an AUC of 0.846. Using an arbitrary cutoff at the level of 55 ng/mg total BAL protein resulted in the sensitivity of 84.21% and the specificity of 66.67%, respectively. Our analysis was only based on a small number of cases. Obviously, a larger scale of study is necessary to further evaluate the utility of these potential protein biomarkers in the identification of cancers. Our data demonstrated a correlation of specific glycoproteins in cancer tissue/cells and the presence of those glycoproteins in cancer BAL specimens. The fact that tumor-specific proteins were detectable in cancer BAL also suggests that potential biomarkers could be used for the detection of cancer using BAL<sup>35</sup>.

Currently, the diagnosis/classification of small non-calcified solitary pulmonary nodule (SPN) is difficult and involves combinations of radiological surveillance, cytological examination of malignant cells in either conventional sputum or bronchoscopic specimens such as bronchoalveolar lavage (BAL), bronchial brushing, transbronchial fine needle aspiration biopsy (TBNA). These tests are not always conclusive in differentiating benign nodules from cancer. Over the past decade, several clinical trials using highly sensitive image technology, such as low-dose computed tomography (LDCT), have improved our ability to detect early lung cancers, that are usually with small sizes of tumors, from benign nodules. However, lung cancer screening trials in both the United States and Europe have also shown high false-positive rates<sup>17, 36</sup>. For example, in the European NELSON trial, 1451 nodules (19.2%) were diagnosed as indeterminate and 119 nodules (1.6%) diagnosed as positive for carcinoma, but only 70 cases (0.9%) were confirmed to be lung cancers after further pre-surgical work-up<sup>36</sup>. In recent large scale United States National Lung Cancer Screening Trial (NLST), any non-calcified solitary pulmonary nodule (SPN) measuring at least 4 mm in any diameter by LDCT was classified as “suspicious for lung cancer”<sup>17</sup>. Among the 26,309 individuals enrolled in the trial, 39.1% had at least one positive screening result during the entire screening phase<sup>17</sup>. The total false positive rate among positive tests with LDCT was 96.4%, and only 3.6% of the patients (649 cases) were confirmed to actually have lung cancers<sup>17</sup>. Thus, aggressive surgical procedures, including mediastinoscopy, thoracoscopy and thoracotomy are frequently performed in patients with



suspicious nodules. These invasive diagnostic procedures are associated with a 15.9 to 32.4% complication rate among patients<sup>17</sup> and are very costly<sup>15, 16</sup>. It is becoming increasingly important to distinguish small benign nodules from cancer with minimally invasive procedures and reduce the use of more invasive second-line procedures. Our data indicate that Napsin A and other proteins in airway fluid have a potential to be used as biomarker to develop a surrogate test for the detection of lung adenocarcinoma (which is the most common type of lung cancers), therefore, to improve our ability of separation cancer from benign lung nodules in high risk patients. Furthermore, recent study also demonstrated that Napsin A is not only a marker of adenocarcinoma but also a potential prognostic marker, using shotgun LC-MS/MS studies and selected-reaction monitoring (SRM) mass spectrometry<sup>24</sup>.

A recent study by Kikuchi T, et al has shown that 3621 proteins were identified from pooled lung tumor tissue using in-depth proteomic analysis<sup>12</sup>. In the study, they analyzed both ADC and SQCC tumor tissue using the shotgun proteomic method, and further validate their findings using multiple reactions monitoring MS, IHC staining and Western blotting analysis. In comparison to their study, we have also identified protein lipocalin in cancer BAL using SPEG and iTRAQ approaches. Furthermore, Zeng X, et al studied serum glycoproteins in ADC and SQCC patients using SPEG and LC-MS/MS approach<sup>37</sup>. They found that 22 glycoproteins showed significant changes in sera from lung cancer patients. Interestingly, we also identified 7 of those glycoproteins in cancer BAL fluid, including Cathepsin D, Complement factor I, Kininogen-1, Kallistatin, Leucine-rich alpha-2-glycoprotein, Multimerin-1 and plasma protease C1 inhibitor. Taken together, these findings are important and suggest that unique tumor-derived proteins can be identified in lung parenchyma and in airway fluid, and some of them could also be further detected in the serum.

Some of the glycoproteins identified in BAL have been reported to play critical roles in lung cancers. For example, NE has been suggested to play an important role in NSCLC<sup>38-40</sup>. Early study has shown that the immunoreactive NE produced by lung cancer cell could facilitate the invasion of the cancer either by directly dissolving the tumor matrix or indirectly by activating a protease cascade<sup>38</sup>. Recent study has also found that patients with certain mutations in the promoter region of NE gene have an increased risk for developing lung cancers<sup>39</sup>. NE may be a potential therapeutic target in lung cancer<sup>40</sup>. Furthermore, the expression of Cathepsin D has been detected in lung adenocarcinomas<sup>41</sup>, and its expression has been suggested to correlate with poor differentiation of tumor and to be a potential prognostic marker for lung adenocarcinoma<sup>42</sup>. Oumeraci T, et al also demonstrated that several proteins were differentially expressed between cancer and non-cancer BAL fluid detected by MALDI (matrix-assisted laser desorption/ionization) approach<sup>43</sup>. Kahn N, et al. has studied molecular markers in endobronchial epithelial-lining fluid and found that 4 genes were upregulated in NSCLC patients, including tenascin-C, [c-X-c motif] ligand 14, S100 A9 and keratin 17<sup>44</sup>. All these studies indicate that glycoproteins have the potential to be used as biomarkers to improve the diagnosis of lung cancer and monitoring disease progression in lung cancer patients.

In our study, the expression of Napsin A was detected in more than 80% of pT1 and pT2 tumors using IHC approach, however, LC-MS/MS data showed a decrease level of Napsin A in early tumor in the comparison to the tumor-matched normal tissue. This could be due to several reasons. It is well known that the presence of alveolar macrophages in the normal lung parenchyma may potentially affect the level of Napsin A<sup>33</sup>. Secondly, the sampling error might occur, since we only had a few cases in the LC-MS/MS study. Obviously, a larger scale of study is necessary to further validate our findings.

In summary, our results demonstrate that glycoproteins in BAL fluid can be detected and quantified. Among them, potential protein biomarkers can be identified. A large scale study is necessary to further investigate the role of tumor-associated proteins in lung cancer.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work is partially supported by Drs. Ji&Li Family Cancer Research Foundation (QKL); National Cancer Institute, the Early Detection Research Network (EDRN grant U01CA152813) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC U24CA160036) (HZ and DWC). Authors thank Dr. Erika Rodriguez and Ms. Ming-Hui Ao for their technique helps.

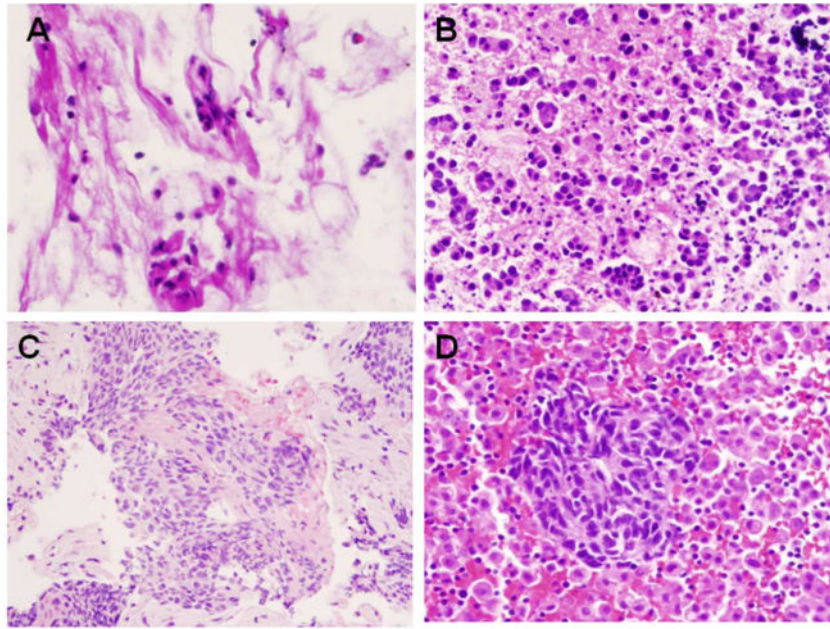
## References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012; 62:10–29. [PubMed: 22237781]
2. Travis WD, IASLC Staging Committee. Reporting lung cancer pathology specimens. Impact of the anticipated 7th Edition TNM classification based on recommendations of the IASLC Staging Committee. *Histopathology.* 2009; 54:3–11. [PubMed: 19187176]
3. Feller-Kopman D, Yung RCW, Burroughs F, Li QK. Cytology of endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA). A study of 135 cases with histology correlation. *Cancer Cytopathol.* 2009; 117:482–490.
4. Stoll LM, Yung RCW, Clark DP, Li QK. Cytology of endobronchial ultrasound-guided transbronchial needle aspiration verse conventional transbronchial needle aspiration. *Cancer Cytopathol.* 2010; 118:278–286. [PubMed: 20740503]
5. Sing A, Freudenberg N, Kortsik C, Wertzel H, Klosa B, Hasse J. Comparison of the sensitivity of sputum and brush cytology in the diagnosis of lung carcinomas. *Acta Cytol.* 1997; 41:399–408. [PubMed: 9100773]
6. Hasanovic A, Rekhman N, Sigel CS, Moreira AL. Advances in fine needle aspiration cytology for the diagnosis of pulmonary carcinoma. *Patholog Res Int.* 2011; 2011:897292. [PubMed: 21776344]
7. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, Singh S, Chen WJ, Chen JJW, Yang PC. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med.* 2007; 356:11–20. [PubMed: 17202451]
8. Coate LE, John T, Tsao MS, Shepherd FA. Molecular predictive and prognostic markers in non-small-cell lung cancer. *Lancet Oncol.* 2009; 10:1001–1010. [PubMed: 19796752]
9. Showe MK, Vachani A, Kossakov AV, Yousef M, Nichols C, Nikonova EV, Chang C, Kucharczuk J, Tran B, Wakeam E, Yie TA, Speicher D, Rom WN, Albelda S, Showe LC. Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res.* 2009; 69:9202–9210. [PubMed: 19951989]

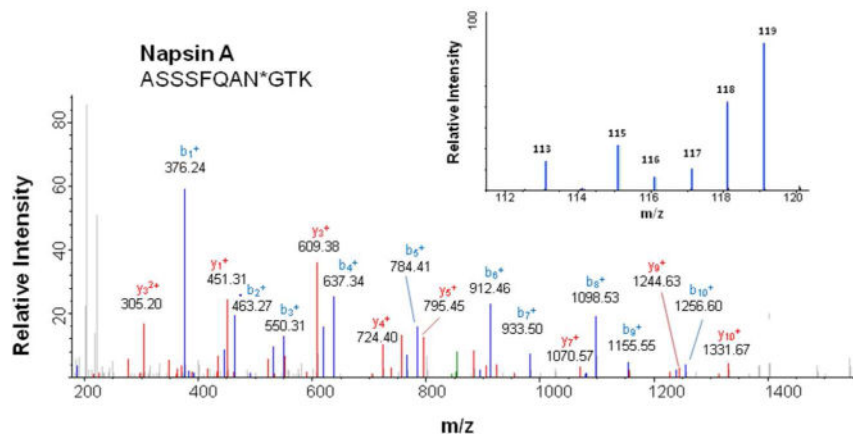
10. Taguchi A, Politi K, Pitteri SJ, Lockwood WW, Faça VM, Kelly-Spratt K, Wong CH, Zhang Q, Chin A, Park KS, Goodman G, Gazdar AF, Sage J, Dinulescu DM, Kucheralapati R, Depinho RA, Kemp CJ, Varmus HE, Hanash SM. Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell*. 2011; 20:289–299. [PubMed: 21907921]
11. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, Naidich DP, Crawford B, Eylers E, Liu D, Tan EM. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer*. 2010; 10:234. [PubMed: 20504322]
12. Kikuchi T, Hassanein M, Amann JM, Liu Q, Slebos RJ, Rahman SM, Kaufman JM, Zhang X, Hoeksema MD, Harris BK, Li M, Shyr Y, Gonzalez AL, Zimmerman LJ, Liebler DC, Massion PP, Carbone DP. In-depth Proteomic Analysis of Nonsmall Cell Lung Cancer to Discover Molecular Targets and Candidate Biomarkers. *Mol Cell Proteomics*. 2012; 11:916–32. [PubMed: 22761400]
13. Nolen BM, Langmead CJ, Choi S, Lomakin A, Marrangoni A, Bigbee WL, Weissfeld JL, Wilson DO, Dacic S, Siegfried JM, Lokshin AE. Serum biomarker profiles as diagnostic tools in lung cancer. *Cancer Biomark*. 2011–2012; 10:3–12. [PubMed: 22297547]
14. Bigbee WL, Gopalakrishnan V, Weissfeld JL, Wilson DO, Dacic S, Lokshin AE, Siegfried JM. A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening. *J Thorac Oncol*. 2012; 7:698–708. [PubMed: 22425918]
15. Goldberg-Kahn B, Healy JC, Bishop JW. The cost of diagnosis: a comparison of four different strategies in the workup of solitary radiographic lung lesions. *Chest*. 1997; 111:870–876. [PubMed: 9106563]
16. Kutikova L, Bowman L, Chang S, Long SR, Obasaju C, Crown WH. The economic burden of lung cancer and the associated costs of treatment failure in the United States. *Lung Cancer*. 2005; 50:143–154. [PubMed: 16112249]
17. National Lung Screening Trial Research Team. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
18. Magi B, Bargagli E, Bini L, Rottoli P. Proteome analysis of bronchoalveolar lavage in lung diseases. *Proteomics*. 2006; 6:6354–6369. [PubMed: 17133372]
19. Meyer KC. Bronchoalveolar lavage as a diagnostic tool. *Semin Respir Crit Care Med*. 2007; 28:546–560. [PubMed: 17975782]
20. Poletti V, Poletti G, Murer B, Saragoni L, Chilosi M. Bronchoalveolar lavage in malignancy. *Semin Respir Crit Care Med*. 2007; 28:534–545. [PubMed: 17975781]
21. Li QK, Gabrielson E, Askin F, Chan DW, Zhang H. Glycoproteomics using fluid based specimens in the discovery of lung cancer protein biomarkers. Promise and challenge. *Proteomics Clinical App*. 2013; 7:55–69.
22. Drake PM, Cho W, Li B, Prakobphol A, Johansen E, Anderson NL, Regnier FE, Gibson BW, Fisher SJ. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem*. 2010; 56:223–236.
23. Li QK, Gabrielson E, Zhang H. Application of glycoproteomics for the discovery of biomarkers in lung cancer. *Proteomics: Clinical App*. 2012:5–6. 244–256.
24. Nishimura T, Nomura M, Tojo H, Hamasaki H, Fukuda T, Fujii K, Mikami S, Bando Y, Kato H. Proteomic analysis of laser-microdissected paraffin-embedded tissues: (2) MRM assay for stage-related proteins upon non-metastatic lung adenocarcinoma. *J Proteome*. 2010; 73:1100–1110.
25. Lehtio J, De Petris L. Lung cancer proteomics, clinical and technological considerations. *J Proteome*. 2010; 73:1851–1863.
26. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*. 2003; 21:660–666. [PubMed: 12754519]
27. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76:4193–4201. [PubMed: 15253663]

28. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004; 3:1154–1169. [PubMed: 15385600]
29. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics*. 2006; 5:573–588. [PubMed: 16332733]
30. Edge, SB.; Byrd, DR.; Compton, CC.; Fritz, AG.; Greene, FL.; Trotti, A. *AJCC Cancer Staging Manual*. 7. Springer; New York: 2010. p. 311-337.
31. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, Garg K, Austin JH, Asamura H, Rusch VW, Hirsch FR, Scagliotti G, Mitsudomi T, Huber RM, Ishikawa Y, Jett J, Sanchez-Cespedes M, Sculier JP, Takahashi T, Tsuboi M, Vansteenkiste J, Wistuba I, Yang PC, Aberle D, Brambilla C, Flieder D, Franklin W, Gazdar A, Gould M, Hasleton P, Henderson D, Johnson B, Johnson D, Kerr K, Kuriyama K, Lee JS, Miller VA, Petersen I, Roggli V, Rosell R, Saijo N, Thunnissen E, Tsao M, Yankelewitz D. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011; 6:244–285. [PubMed: 21252716]
32. Tian Y, Tan AC, Sun X, Olson MT, Xie Z, Jinawath N, Chan DW, Shih IeM, Zhang Z, Zhang H. Quantitative proteomic analysis of ovarian cancer cells identified mitochondrial proteins associated with Paclitaxel resistance. *Proteomics Clin Appl*. 2009; 3:1288–1295. [PubMed: 21113235]
33. Stoll LM, Johnson MW, Gabrielson E, Askin F, Clark DP, Li QK. The utility of Napsin-A in the identification of primary and metastatic lung adenocarcinoma among cytologically “poorly differentiated carcinoma”. *Cancer Cytopathol*. 2010; 118:441–449. [PubMed: 20830690]
34. Hirano T, Fujioka K, Franzèn B, Okuzawa K, Uryu K, Shibamura H, Numata K, Konaka C, Ebihara Y, Takahashi M, Kato H, Auer G. Relationship between TA01 and TA02 polypeptides associated with lung adenocarcinoma and histocytological features. *BMJ*. 1997; 75:978–985.
35. Steiling K, Kadar AY, Bergerat A, Flanigon J, Sridhar S, Shah V, Ahmad QR, Brody JS, Lenburg ME, Steffen M, Spira A. Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One*. 2009; 4:e5043. [PubMed: 19357784]
36. van den Bergh KA, Essink-Bot ML, Bunge EM, Scholten ET, Prokop M, van Iersel CA, van Klaveren RJ, de Koning HJ. Impact of computed tomography screening for lung cancer on participants in a randomized controlled trial (NELSON trial). *Cancer*. 2008; 113(2):396–404. [PubMed: 18484588]
37. Zeng X, Hood BL, Sun M, Conrads TP, Day RS, Weissfeld JL, Seigfried JM, Bigbee WL. Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *J Proteome Res*. 2010; 9:6440–6449. [PubMed: 20931982]
38. Yamashita J, Tashiro K, Yoneda S, Kawahara K, Shirakusa T. Local increase in polymorphonuclear leukocyte elastase is associated with tumor invasiveness in non-small cell lung cancer. *Chest*. 1996; 109:1328–1334. [PubMed: 8625686]
39. Topic A, Ljubic M, Nikolic A, Petrovic-Stanojevic N, Dopudja-Pantic V, Mitic-Milicic M, Radojkovic D. Alpha-1-antitrypsin phenotypes and neutrophil elastase gene promoter polymorphisms in lung cancer. *Pathol Oncol Res*. 2011; 17:75–80. [PubMed: 20521180]
40. Moroy G, Alix AJ, Sapi J, Hornebeck W, Bourguet E. Neutrophil elastase as a target in lung cancer. *Anticancer Agents Med Chem*. 2012; 12:565–579. [PubMed: 22263788]
41. Higashiyama M, Doi O, Kodama K, Yokouchi H, Kasugai T, Ishiguro S. Influence of cathepsin D expression in lung adenocarcinoma on prognosis: possible importance of its expression in tumor cells and stromal cells, and its intracellular polarization in tumor cells. *J Surg Oncol*. 1997; 65:10–19. [PubMed: 9179261]
42. Mimae T, Tsuta K, Maeshima AM, Okada M, Asamura H, Kondo T, Tsuda H. Cathepsin D as a potential prognostic marker for lung adenocarcinoma. *Pathol Res Pract*. 2012; 208:534–540. [PubMed: 22824147]
43. Oumeraci T, Schmidt B, Wolf T, Zapatka M, Pich A, Brors B, Eils R, Fleischhacker M, Schlegelberger B, von Neuhoff N. Bronchoalveolar lavage fluid of lung cancer patients: Mapping

- the uncharted waters using proteomics technology. *Lung Cancer*. 2011; 72:136–138. [PubMed: 21334092]
44. Kahn N, Meister M, Eberhardt R, Muley T, Schnabel PA, Bender C, Johannes M, Keitel D, Sültmann H, Herth FJ, Kuner R. Early detection of lung cancer by molecular markers in endobronchial epithelial-lining fluid. *J Thorac Oncol*. 2012; 7:1001–1008. [PubMed: 22588153]

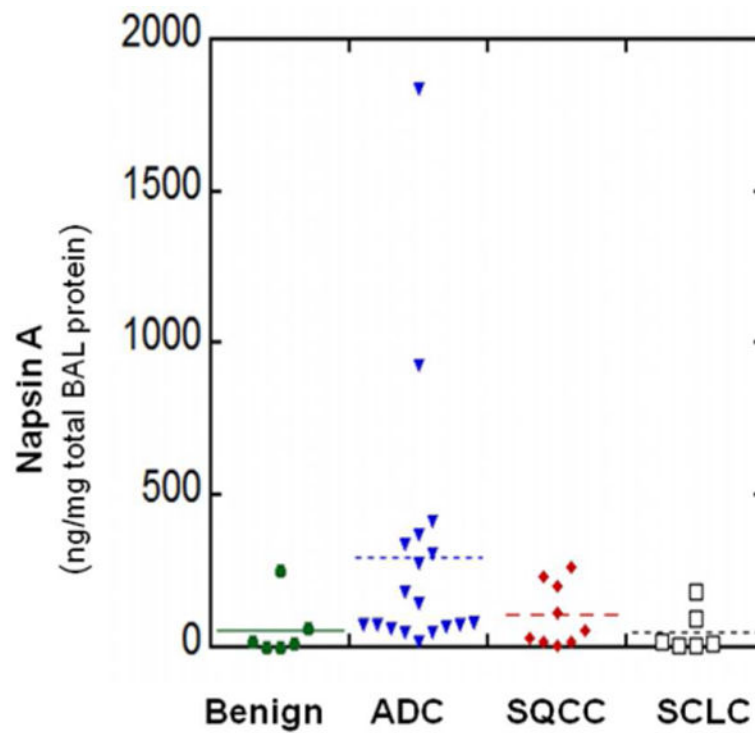


**Figure 1.** Cytological diagnosis of cells in BAL specimens. (A) benign reactive bronchial epithelium, (B) adenocarcinoma (ADC), (C), Squamous cell carcinoma (SQCC), and (D) Small cell lung carcinoma (SCLC). Papanicolaou stain X40.



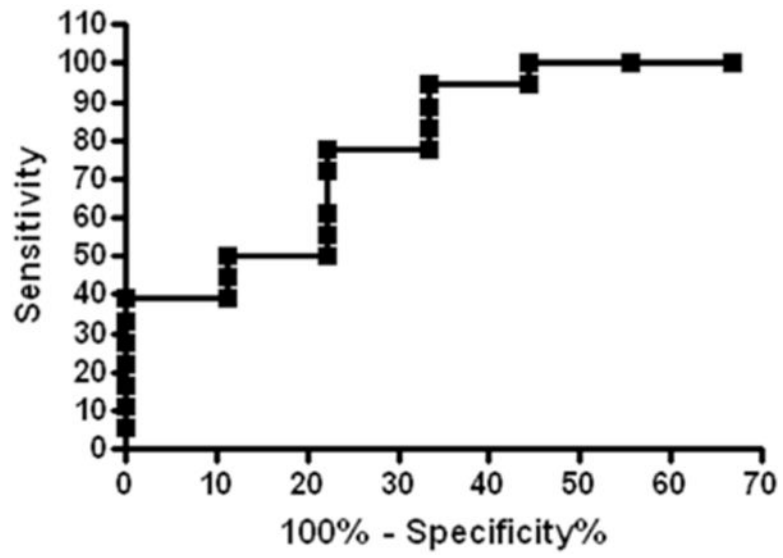
**Figure 2.** Identification of Napsin A in BAL and lung adenocarcinoma tissues. The MS/MS spectrum corresponding to Napsin A peptide [ASSSFQAN\*GTK] was detected in both BAL specimen and tumor tissue. N\* is the N-linked glycosylation site and deaminated to Aspartic Acid (D) when N-glycans were released from the site by PNGase F. Insert: iTRAQ report ions for Napsin A peptide from different tissue and BAL samples. The report ions are:

- 113 Pooled benign BAL
- 115 Pooled non-tumor tissue from early stage tumor
- 116 Pooled tumor tissue from early stage tumor
- 117 Pooled non-tumor tissue from late stage tumor
- 118 Pooled tumor tissue from late stage tumor
- 119 Pooled cancer BAL



**Figure 3.** Detection of glycoproteins in BAL specimens by ELISA assay. BAL fluid specimens were collected from lung cancer and benign lung disease patients. Napsin A levels in benign lung disease, adenocarcinoma (ADC), Squamous cell carcinoma (SQCC), and Small cell lung carcinoma (SCLC) were determined.





**Figure 4.** ROC analysis of Napsin A concentrations in BAL fluid from benign lung disease (n=7) and adenocarcinoma (n=18). AUC=0.846. The sensitivity and specificity are 84.21 % and 66.67%, respectively.

**Table 1**

Summary of clinical information of BAL specimen for the LC-MS/MS analysis.

Cases with diagnoses of benign lesions					
Case	Age/Sex	Clinical History	CT findings	Cytological Diagnosis	Pathologic Stage
1	28/F	Pneumonia	Right lung density	Benign respiratory epithelium and pulmonary alveolar macrophages	N/A
2	68/F	Lung nodules	1.0 cm non-calcified nodule	Benign respiratory epithelium and pulmonary alveolar macrophages	N/A
3	18/M	pneumonia	Atelectasis and consolidation	Benign respiratory epithelium	N/A
4	62/F	Fever	Bilateral density	Benign respiratory epithelium	N/A
Cases with diagnoses of primary lung adenocarcinomas					
5	69/F	Lung mass	6.5 cm mass	Adenocarcinoma	pT3
6*	83/F	Lung mass	1.7 cm mass	Adenocarcinoma	pT1
7	61/F	Lung mass	1.5 cm mass	Adenocarcinoma	pT1
8	69/F	Lung mass	6.0 cm mass	Adenocarcinoma	pT3

N/A: not applicable.

\* the patient developed brain metastasis later.

**Table 2**

Quantitative analysis of proteins from BAL and lung tissues using iTRAQ

<b>Description</b>	<b>116/115</b>	<b>118/117</b>	<b>119/113</b>
Neutrophil elastase	0.646	1.356	29.382
Integrin alpha-M	0.713	2.146	20.499
Cullin-4B	0.883	0.921	15.264
Napsin-A	0.286	4.252	5.145
LAMP2	0.479	1.772	4.675
Cathepsin D	0.436	61.061	3.897
BPI fold-containing family B member 2	0.789	1.585	3.064
Neutrophil gelatinase-associated lipocalin	0.873	0.961	2.695
Ig alpha-2 chain C region	0.665	5.955	1.799
Myeloperoxidase	0.816	0.884	1.596
Immunoglobulin J chain	0.592	10.353	1.456
Galectin-3-binding protein	0.560	8.286	0.981
Hemopexin	0.711	4.157	0.920
Alpha-1-antitrypsin	0.986	0.998	0.623
Beta-2-glycoprotein 1	0.407	11.946	0.504
Sushi domain-containing protein 2	0.030	0.256	0.487
Ceruloplasmin	0.986	0.998	0.458
Clusterin	0.639	3.438	0.419
Plasma protease C1 inhibitor	0.561	8.115	0.397
Plasma kallikrein	0.951	4.237	0.371
Complement factor B	0.400	9.984	0.297
Afamin	0.549	13.601	0.297
Complement C4-A	0.347	4.982	0.262
Kininogen-1	0.584	5.449	0.225
Ig mu chain C region	0.390	3.613	0.184
Haptoglobin	0.574	22.625	0.142
Ig gamma-2 chain C	0.671	10.092	0.104
Ig gamma-1 chain C region	0.742	5.833	0.098
Ig gamma-4 chain C region	0.649	5.405	0.084
Ig gamma-3 chain C region	0.739	4.995	0.067
Complement C3	0.989	5.559	0.060
Fibrinogen beta chain	0.302	0.998	0.036

113	Pooled benign BAL
115	Pooled non-tumor tissues from early stage tumor
116	Pooled tumor tissues from early stage tumor
117	Pooled non-tumor tissues from late stage tumor
118	Pooled tumor tissues from late stage tumor
119	Pooled cancer BAL

Table 3

Patient characteristics for BAL ELISA assay

Category	Benign	ADC	SQCC	SCLC
Gender	F	11	5	2
	M	7	4	4
Race	African American	2	3	3
	Caucasian	4	4	3
	Other	0	2	0
Smoking status	Current smoker	2	3	2
	Ex-smoker	2	2	4
	Never smoker	1	0	0
	Unknown	1	3	0
Median age at diagnosis	54 yr	63.5yr	68.5yr	70yr
Range	18–85 yr	51–83 yr	55–85yr	65–73yr
Pathological stage	pT1	3	1	0
	pT2	N/A	4	2
	pT3	N/A	5	1
	pT4	N/A	2	1
Unknown	N/A	4	4	2
Total number of samples	6	18	9	6

ADC:adenocarcinoma

SQCC: squamous cell carcinoma

SCLC: small cell lung carcinoma

yr: years

pT: pathological stage

N/A: not applicable.