

Proteomic Analysis of Human Saliva From Lung Cancer Patients Using Two-Dimensional Difference Gel Electrophoresis and Mass Spectrometry*

Hua Xiao[‡], Lei Zhang[‡], Hui Zhou[‡], Jay M. Lee^{§**}, Edward B. Garon[¶],
and David T. W. Wong^{‡**†§§¶¶}

Lung cancer is often asymptomatic or causes only non-specific symptoms in its early stages. Early detection represents one of the most promising approaches to reduce the growing lung cancer burden. Human saliva is an attractive diagnostic fluid because its collection is less invasive than that of tissue or blood. Profiling of proteins in saliva over the course of disease progression could reveal potential biomarkers indicative of oral or systematic diseases, which may be used extensively in future medical diagnostics. There were 72 subjects enrolled in this study for saliva sample collection according to the approved protocol. Two-dimensional difference gel electrophoresis combined with MS was the platform for salivary proteome separation, quantification, and identification from two pooled samples. Candidate proteomic biomarkers were verified and prevalidated by using immunoassay methods. There were 16 candidate protein biomarkers discovered by two-dimensional difference gel electrophoresis and MS. Three proteins were further verified in the discovery sample set, prevalidation sample set, and lung cancer cell lines. The discriminatory power of these candidate biomarkers in lung cancer patients and healthy control subjects can reach 88.5% sensitivity and 92.3% specificity with AUC = 0.90. This preliminary data report demonstrates that proteomic biomarkers are present in human saliva when people develop lung cancer. The discriminatory power of these candidate biomarkers indicate that a simple saliva test might be established for lung cancer clinical screening and detection. *Molecular & Cellular Proteomics* 11: 10.1074/mcp.M111.012112, 1–12, 2012.

Lung cancer has long been a significant worldwide public health issue. In the United States, lung cancer is the most

From the [‡]Dental Research Institute, [§]Department of Surgery, David Geffen School of Medicine, [¶]Department of Medicine, David Geffen School of Medicine, ^{||}School of Dentistry, ^{**}Jonsson Comprehensive Cancer Center, ^{††}Division of Head and Neck Surgery/Otolaryngology, ^{§§}Henry Samueli School of Engineering, University of California, Los Angeles, Los Angeles, California 90095

Received June 17, 2011, and in revised form, October 25, 2011

Published, MCP Papers in Press, November 17, 2011, DOI 10.1074/mcp.M111.012112

common cause of cancer-related death in men and women. According to the American Cancer Society, the major type of lung cancer is non-small-cell lung cancer (NSCLC)¹, which comprises ~85% of lung cancers (1, 2). Small cell lung cancer (SCLC) is another type of lung cancer, which represents about 15% of cases (3). Cigarette smoking causes lung cancer, which is by far the major risk factor (4, 5, 6). Lung cancer has a high mortality rate among malignancies, in part because symptoms are frequently absent until disease is already metastatic and is, therefore, incurable (7, 8, 9). Early detection represents a very promising approach to reduce lung cancer incidence and mortality. However, conventional diagnostic methods for lung cancer are unsuitable for widespread screening because they are either expensive and occasionally miss tumors or invasive cancer (10, 11, 12, 13). Computed tomography has been widely used for lung cancer early screening, although it often produces a high rate of false positives (7, 14, 15). Better diagnostic methods are urgently needed to improve the detection of lung cancer.

Tissue and blood have been extensively used in attempts to detect lung cancer earlier (1, 16, 17, 18). Sputum has also been used to predict lung cancer by detecting its aberrant promoter methylation (19, 20) or metal ions (21); however, the screening might be limited by the sample availability or composition variability. Different proteomic technologies have been engaged for this type of biomarker discovery. Quantitative proteomic technologies, like iTRAQ (22, 23, 24) and two-dimensional gel electrophoresis (2-DE) (5, 25) have been extensively used for proteome analysis in different types of lung cancer samples.

Human saliva is an attractive early detection biofluid (26, 27, 28) because its collection is noninvasive, and it contains a

¹ The abbreviations used are: NSCLC, non-small-cell lung cancer; 2D-DIGE, two-dimensional difference gel electrophoresis; SCLC, small cell lung cancer; MALDI-TOF MS, matrix assisted laser desorption ionization/time of flight mass spectrometry; LC-MS/MS, Liquid chromatography tandem mass spectrometry; HP, haptoglobin hp2; AZGP1, zinc α 2-glycoprotein; ANXA1, annexin A1; ROC, receiver operating characteristic; AUC, area under ROC curve; COPD, chronic obstructive pulmonary disease.

large array of proteins, many of which have been shown to be informative for the detection of oral (29, 30) and systemic diseases (31, 32, 33). Systemic diseases, like lung cancer, may affect salivary glands directly or indirectly, and may influence the quantity of saliva that is produced, as well as the composition (26, 27, 34). A recent study has shown that salivary biomarkers are significantly and consistently altered in a mouse lung cancer model, which suggests that the salivary glands may be regulated by mediators released from remote tumors (35).

In this study, human saliva samples were collected from lung cancer patients and matched healthy control subjects. Salivary proteins were analyzed and compared in the two pooled samples from each group. We hypothesized that lung cancer related proteins exist in human saliva, which could be clinically used to discriminate lung cancer patients from healthy control subjects. With the discovery and validation of discriminatory proteomic markers from saliva, lung cancer will be noninvasively detected with high specificity and sensitivity.

EXPERIMENTAL PROCEDURES

Patients and Study Design—This study consisted of two phases, including a discovery phase and a confirmation phase (Fig. 1). In total, 72 saliva samples were used. All these saliva samples were collected under a protocol approved by our institutional review board (IRB) and all patients provided written informed consent. The samples were all collected, processed, and stored in a similar fashion. Unstimulated saliva samples were consistently collected, stabilized, and preserved as previously described (31). The sample supernatants were kept at -80°C prior to assay. Identified proteins were first verified in the discovery sample set (10 lung cancer samples and 10 healthy control samples). A prevalidation sample set (26 lung cancer samples and 26 healthy control samples) was used in the biomarker confirmation phase.

Lung Cancer Cell Lines—NSCLC cell line (NCI-H1299 and NCI-H460) and healthy lung cell line (MRC-5) were obtained from ATCC (Manassas, VA). They were cultured in DMEM medium in duplicate and lysed by 62.5 mM Tris-HCL (pH = 6.8) containing 2% SDS (w/v), 10% Glycerol, and 0.2 mM dithiothreitol.

Proteomics Platform—Two dimensional difference gel electrophoresis (2D-DIGE) was used to separate and quantify the salivary proteins from lung cancer patients and matched healthy control subjects. MALDI-TOF MS and LC-MS/MS were used to identify the proteins in the selected gel spots.

2D-DIGE—Saliva protein concentration was determined by BCA Protein Assay Kit (Thermo Scientific Pierce Protein Research Products, Rockford, IL). Pooled saliva samples were made from 10 lung cancer samples and 10 healthy control samples, respectively, by taking equal amount of protein from each individual sample. The 200 μg of proteins in each pooled saliva sample were precipitated by ethanol and then resuspended in the 2D cell lysis buffer (30 mM Tris-HCl, pH 8.8, containing 7 M urea, 2 M thiourea, and 4% 3-[(3-cholamidopropyl)dimethylammonio]propanesulfonate). The total proteins of the two pooled saliva samples, lung cancer patients and healthy control subjects, were labeled by minimal CyDye Cy3 and Cy5, respectively. The two labeled samples were then combined and subjected to 2D-DIGE. Briefly, after loading the labeled samples, isoelectric focusing (pH3–10) was run following the protocol provided by Amersham BioSciences. The 17 cm immobilized pH gradient strip was rinsed in the SDS-gel running buffer before transferring into

13.5% SDS-gels. The SDS-gels were run at 15°C until the dye front ran out of the gels. Gel images were scanned immediately by using Typhoon TRIO (Amersham BioSciences, GE healthcare, Waukesha, WI) following running the SDS-PAGE. The fold change of the protein expression levels was manually obtained from in-gel analysis by using software Progenesis SameSpots (Version 3.3, Nonlinear Dynamics Limited). After in-gel image analysis, 253 spots were located on the gel with fold change ranging from 1.1 to 8. Based on their fold change (greater than or equal to 1.5), abundance, and relative location on the gel, 30 spots were further excised and in-gel digested by trypsin.

In-gel Trypsin Digestion and Mass Spectrometry Identification—For matrix-assisted laser desorption ionization/time of flight (MALDI-TOF) MS analysis, selected spots were cut and washed multiple times to remove staining dye and other inhibitory chemicals. Dried gel spots were rehydrated in the digestion buffer containing sequencing grade modified trypsin (Promega, Madison, WI). Proteins were digested in-gel at 37°C overnight. Digested peptides were extracted from the gel with trifluoroacetic acid extraction buffer and shaking. The digested tryptic peptides were desalted by using C-18 Zip-tips (Millipore, Billerica, MA). The desalted peptides were mixed with CHCA matrix (alpha-cyano-4-hydroxycinnamic acid) and spotted into wells of a MALDI plate for MALDI-TOF MS/MS (ABI4800) identification. Protein identification was based on peptide fingerprint mass mapping (using MS spectra) and peptide fragmentation mapping (using MS/MS spectra). Combined MS and MS/MS spectra were submitted for database search using Global Protein Server Explorer software version 3.6 (Applied Biosystems, Foster City, CA) equipped with the MASCOT search engine (version 2.2.0, <http://www.matrixscience.com>) to identify proteins from the National Center for Biotechnology Information nonredundant *Homo sapiens* amino acid sequence database (224815 sequences out of 9937670 sequences). The parameters for searching were enzyme of trypsin, 1 missed cleavage, fixed modifications of carbamidomethyl (C), variable modifications of oxidation (M), peptide mass tolerance: ± 0.5 Da, fragment mass tolerance: ± 0.5 Da, peptide charge of 1+ and monoisotopic. Only significant hits, as defined by the MASCOT probability analysis ($p < 0.05$), were accepted. The criteria of two peptides and C.I. $\% > 95$ were used for protein identification.

For LC-MS/MS analysis, the proteins in the selected spots were digested as the same as for MALDI-TOF MS analysis and loaded to LC-MS/MS (Eksigent NanoLC-2D with Thermo LTQ XL). Peptides were first enriched on a reverse phase trap column (ProteoPep II, 100 $\mu\text{m} \times 2.5$ cm, C18, 5 μm , 300Å, New Objective) and then eluted to analytical column (ProteoPep II, 75 $\mu\text{m} \times 10$ cm, C18, 5 μm , 300Å, New Objective). The mobile phase consisted of buffer (A) 5% acetonitrile and 0.1% formic acid in water, and buffer (B) 0.1% formic acid in acetonitrile. A flow rate of 400 nL/min and 60 min of the gradient from 15% B to 95% B was applied for the separation of peptides. Spectra were collected and processed by Xcalibur software version 3.3.0 (Thermo Scientific, Waltham, MA). Combined MS and MS/MS spectra were converted from RAW to mzXML (ReAdW version 4.3.1) and submitted for database search again Human Swissprot by using X!Tandem (version 2010.04.21). The parameters for searching were enzyme of trypsin, 1 missed cleavage, fixed modifications of carbamidomethyl (C), variable modifications of oxidation (M), parent ion tolerance of 4 Da and fragment mass tolerance: ± 0.4 Da. The criteria of two peptides and log (E-value) < -10 were used for protein identification.

Western Blot—Reduced saliva protein or cellular protein (15 μg of total proteins per lane) was loaded into a 10% Bis-Tris gel and run at 150V in MES SDS Running Buffer for one hour. Prestained protein standard (Invitrogen, Carlsbad, CA) was used to track protein migration. The proteins were transferred to nitrocellulose membrane by using iBlot (Invitrogen). The membrane was then washed in wash

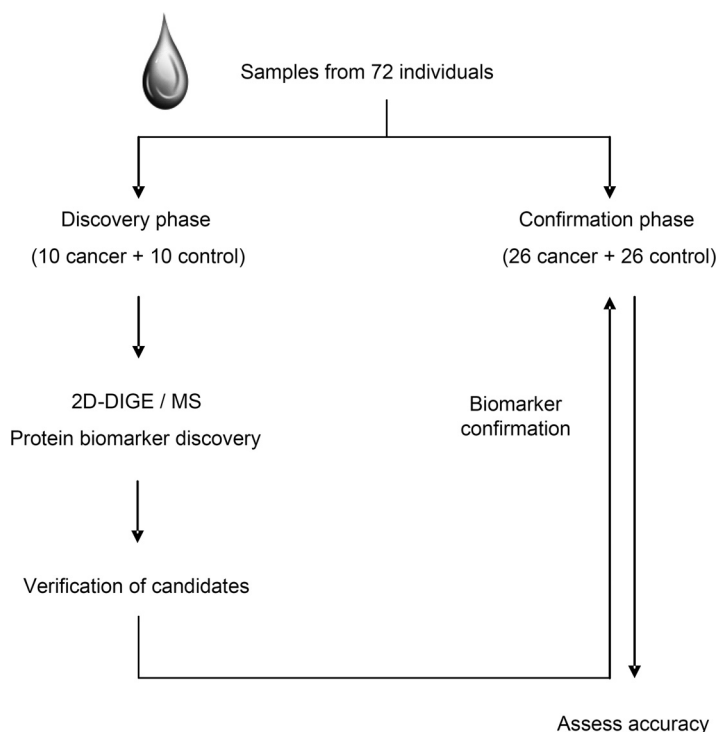


FIG. 1. Study design for lung cancer salivary proteomic biomarker discovery.

buffer (10 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 0.1% (v/v) Tween-20 (Sigma-Aldrich, St. Louis, MI)) before blocking for one hour in wash buffer containing 5% of nonfat dry milk (Santa Cruz, Santa Cruz, CA). After further washes in wash buffer, the membrane was incubated with the primary antibody (Mouse monoclonal to annexin A1 (ANXA1) (ab2487, Abcam, Cambridge, MA), mouse monoclonal to haptoglobin hp2 (HP) (LS-B2863, Lifespan Bioscience), mouse monoclonal to zinc α 2-glycoprotein (AZGP1) (sc-13585, Santa Cruz Biotechnology), mouse monoclonal to β -actin (A1978, Sigma-Aldrich)) according to the manufacturers' instructions in blocking buffer at room temperature for two hours. The membrane was then washed before applying the secondary antibody (Anti-mouse IgG, peroxidase-linked species-specific whole antibody from sheep, GE Healthcare) according to manufacturer instructions, for one hour at room temperature. Finally, the membrane was washed and visualized using ECL Plus detection kit (GE Healthcare). The signal intensity of the bands was measured by using Image J software (National Institutes of Health, Bethesda, MD) and the corresponding β -actin expression was used as reference.

ELISA—The ELISA test for human calprotectin was performed according to the manufacturer's instructions (Cell Sciences). All saliva samples were diluted 100 times with sample diluents. All standards and saliva samples were loaded in duplicate. For cell lysate, 3 μ g of total protein was loaded to each well in duplicate.

Data Analysis—The software Graphpad Prism (Version 5.01) and R were used for all data analysis. p value was calculated based on Wilcoxon test and $p < 0.05$ was used as cutoff for significance. Logistic regression method was used in the prediction model building. For all the validated biomarkers, we constructed the receiver operating characteristic (ROC) curve and computed the area under the curve (AUC) value by numerical integration of the ROC curve. The confirmed lung cancer related proteins were analyzed by logistic regression and stepwise backward selection was performed to determine final combinations of biomarkers. The sensitivity and specificity for the biomarker combinations were estimated by identifying

the threshold of the predicted probability that yielded the highest sum of sensitivity and specificity.

RESULTS

Study Design—The study design is briefly shown in Fig. 1. All saliva samples were collected under a protocol approved by our IRB (IRB#10-000505), and all enrolled subjects provided written informed consent. The 36 lung cancer saliva samples were collected from patients who have been diagnosed as lung cancer by using biopsy at the Ronald Reagan UCLA Medical Center and will proceed to surgery; most of them are NSCLC (Table I). The saliva samples from 36 healthy control subjects were collected as controls by matching their age, sex, and ethnicity with the cancer group ($p > 0.05$). Their smoking history was matched generally by whether they are current or former smokers and their duration and intensity. Patient demographics and clinical profiles are presented in Table I.

Lung Cancer Saliva Proteome Analysis—For the proteomic biomarker discovery, 10 untreated lung cancer patients and 10 matched healthy control subjects were recruited for saliva sample collection. Pooled saliva of the cancer and control group was made by pooling equal amount of protein from each individual sample for the following analysis. Then the two pooled saliva samples were analyzed by 2D-DIGE. The merged 2D-DIGE image is shown in Fig. 2. After in-gel image analysis, 253 spots were located on the gel with fold change ranging from 1.1 to 8. Based on their fold changes, abundance and relative location on the gel, 30 spots were further excised, in-gel tryptic digested and then analyzed by mass spectrometry.

TABLE I
Patient demographics and clinical profiles

Demographic	Discovery set		Confirmation set	
	Cancer (n = 10)	Control (n = 10)	Cancer (n = 26)	Control (n = 26)
Age, years	59.2 ± 6.36	60.4 ± 4.72	63.15 ± 11.63	61.65 ± 9.4
Range	45–69	45–71	43–90	45–89
Sex				
Male	5	5	12	11
Female	5	5	14	15
Ethnicity (USA)				
Caucasian	8	8	22	19
Others	2	2	4	7
Smoke History				
Yes	6	7	16	17
Lung Cancer				
NSCLC	9		24	
SCLC	1		2	

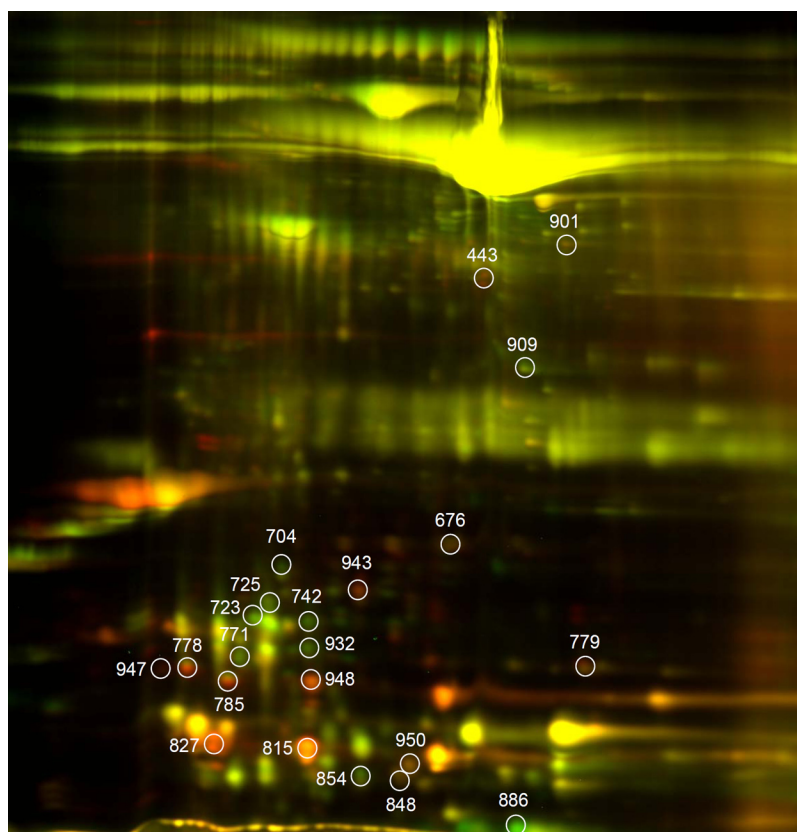


FIG. 2. **Salivary proteomic biomarker discovery.** The merged 2D-DIGE image of proteins from lung cancer patients (Red) and healthy control subjects (Green).

Of these 30 spots, 19 had a fold change greater than or equal to 2. Their digested peptides were loaded to MALDI-TOF plate, and the data was searched by MASCOT search engine against National Center for Biotechnology Information nonredundant *Homo sapiens* amino acid sequence database. Fourteen of these spots were identified as 11 unique proteins. The identified proteins are shown in Table IIA. Eleven of the 30 spots had fold change greater than or equal to 1.5 but less than 2, their digested peptides were loaded to LC-MS/MS, and the data was searched against X!Tandem database. The

results (Table IIB) show that eight spots were identified as seven unique proteins.

In total, 22 spots were identified as proteins with high confidence according to database searching and 16 unique proteins were identified. All the spots are marked on the merged 2D-DIGE image in Fig. 2. The detailed information about these identified proteins is listed in Table II. According to their fold changes, biological function and the availability of validation reagents, four proteins were selected for verification, including HP, ANXA1, AZGP1, and human calprotectin.

TABLE II
Identified human salivary proteins with significant changes between lung cancer patients and healthy control subjects

(A)

Spots ID	Protein name	Accession no.	Protein MW (kDa)	Protein PI	Peptides count	Protein score	Protein score C. I. %	Fold change
443	Annexin A1	gi 119582950	40.2	6.6	10	336	100.0	4.5
704	Interleukin 1 receptor antagonist	gi 27894317	19.9	4.7	5	242	100.0	-2.1
742	Prolactin-induced protein	gi 116642259	9.1	5.3	6	359	100.0	-2.1
771	Lipocalin 1	gi 4504963	19.2	5.4	4	115	100.0	-2.0
778, 785, 947, 948	Zinc Alpha2-Glycoprotein	gi 145579641	13.5	5.5	8, 4, 3, 2	356, 89, 85, 62	100.0	(2.5), (2.0), (3.0), (4.3)
779	Cyclophilin A	gi 1633054	17.9	7.8	2	174	100.0	2.0
815	Cystatin S	gi 4503109	16.2	5.0	5	115	100.0	2.0
827	Human calprotectin	gi 82407447	13.1	5.7	3	182	100.0	5.7
886	S100 calcium-binding protein A8	gi 21614544	10.8	6.5	7	232	100.0	-5.3
943	Haptoglobin hp2	gi 223976	41.7	6.2	6	132	100.0	3.6
950	Cystatin D	gi 19882256	16.1	6.7	4	305	100.0	2.2

(B)

Spots ID	Protein name	Accession no.	Protein MW (kDa)	Protein PI	Unique peptides	log(e)	log(I)	Fold change
676	HBA2	ENSP00000251595	15.2	7.3	2	-18.3	4.6	1.6
723	Lipocalin 1	ENSP00000360846	19.2	5.4	7	-74.3	5.4	-1.5
725, 932	Prolactin-induced protein	ENSP00000291009	16.6	5.3	6,7	(-89.1), (-111.1)	(5.6), (5.8)	(-1.8), (-1.8)
848	CST5	ENSP00000307132	16.1	6.7	4	-34.1	5.1	1.6
854	S100 calcium-binding protein A9	ENSP00000357727	13.2	5.7	7	-91.2	6.2	-1.5
901	SERPINB1	ENSP00000370115	42.7	5.9	6	-57.5	4.5	1.9
909	Carbonic anhydrase VI	ENSP00000366661	28.7	6.6	2	-12.6	3.8	-1.5

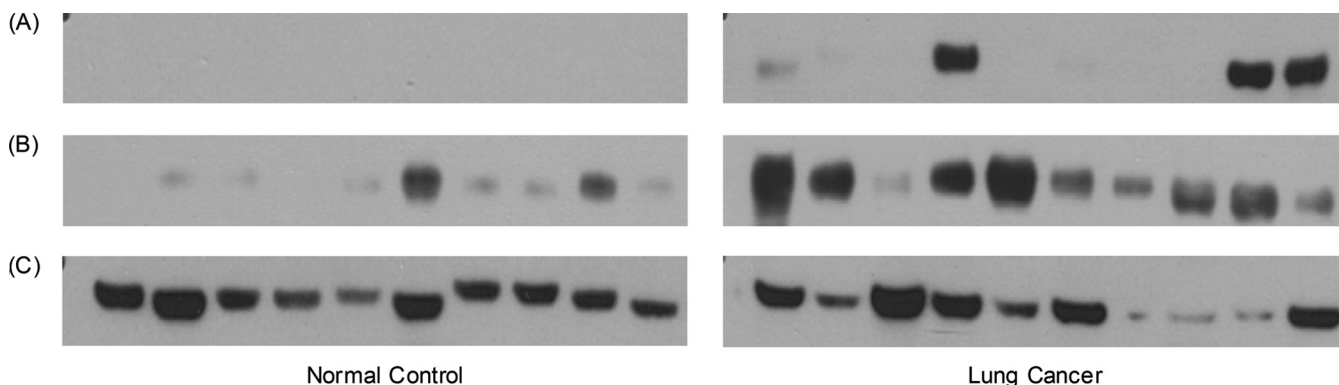


FIG. 3. The Western blot of HP (A), AZGP1 (B), and β-actin (C) in the saliva of lung cancer patients and healthy control subjects. Equal amount of protein were loaded to each lane.

In terms of the relative standard deviation of current immunoassay methods, only those spots with fold change ≥ 2 were selected for biomarker verification.

Verification of Candidates in Discovery Sample Set—Immunoassays were used to verify these four proteins. Western blot was used to test ANXA1, HP, and AZGP1. Human calprotectin was tested by ELISA kits. Of these four proteins, all of them could be well detected, except ANXA1. The Western blot of HP, AZGP1, and the corresponding β-actin are shown in Fig. 3. The distribution of HP, AZGP1 and human calprotectin in lung can-

cer and healthy control groups showed the significant differences with *p* value of 0.0041, 0.0040, and 0.015, respectively.

Confirmation in a Prevalidation Sample Set—To further confirm these proteins' presence in human saliva, a prevalidation sample set (26 lung cancer samples and 26 healthy control samples) was used. By using the same immunoassay methods mentioned above, all the three salivary proteins (HP, AZGP1, and human calprotectin) still show the significant difference in lung cancer patients and healthy control subjects with *p* value of 1.48E-4, 1.05E-4, and 4.56E-5, respectively.

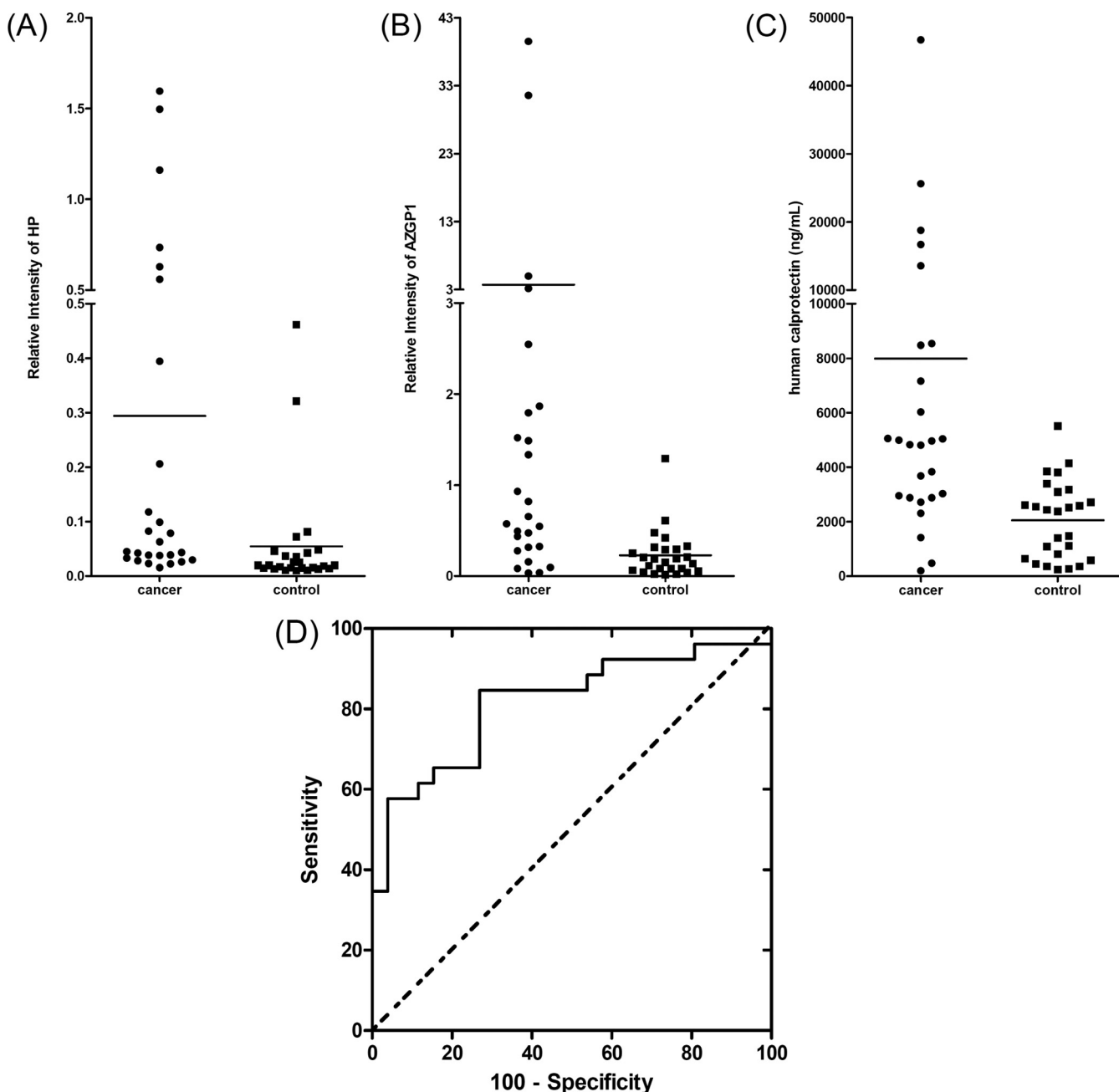


FIG. 4. **Salivary proteomic biomarker evaluation and confirmation.** The dot plots of HP (A), AZGP1 (B), and human calprotectin (C) in the saliva of lung cancer patients and healthy control subjects of prevalidation sample set. D, ROC curve for human calprotectin (ROC = 0.817).

Biomarker Performance—The performance of these validated protein markers was further evaluated for the detection of lung cancer in the prevalidation sample set. The dot plots for all the three proteins are shown in Figs. 4A, 4B, 4C. The ROC curve for human calprotectin is shown in Fig. 4D. The corresponding AUC value for HP, AZGP1, and human calprotectin are 0.807, 0.813, and 0.817, respectively.

Logistic regression was used to combine different markers through software R. The biomarker performance of all combinations of the three prevalidated salivary proteomic bio-

markers were listed in Table III. The corresponding threshold was chosen based on the maximum sum of sensitivity and specificity. In Table IIIA, the threshold was from the prevalidation sample set. Whereas in Table IIIB, the threshold was from the discovery sample set. Even though they have the same AUC value, the threshold in Table IIIB is larger than that in Table IIIA (except the combination of HP and human calprotectin), which lead to different sensitivity and specificity. The dot plot of logistic regression combined biomarkers of human calprotectin, HP, and AZGP1 is shown in Fig. 5A. The

TABLE III
Performance of different biomarker combination in the prevalidation sample set

(A) Biomarker performance based on the threshold from the pre-validation sample set

	AUC	Threshold	Sensitivity%	Specificity%	PPV%	NPV%
Calprotectin+AZGP1	0.88	0.4234	84.6	92.3	91.7	85.7
Calprotectin+HP	0.85	0.4946	73.1	88.5	86.4	76.7
HP+AZGP1	0.82	0.5038	65.4	92.3	89.5	72.7
Calprotectin+AZGP1+HP	0.90	0.4277	88.5	92.3	92.0	88.9

(B) Biomarker performance based on the threshold from the discovery sample set

	AUC	Threshold	Sensitivity%	Specificity%	PPV%	NPV%
Calprotectin+AZGP1	0.88	0.6911	65.4	100.0	100.0	74.3
Calprotectin+HP	0.85	0.2451	92.3	46.2	63.2	85.7
HP+AZGP1	0.82	0.6515	53.9	96.2	93.3	67.6
Calprotectin+AZGP1+HP	0.90	0.6446	73.1	96.2	95.0	78.1

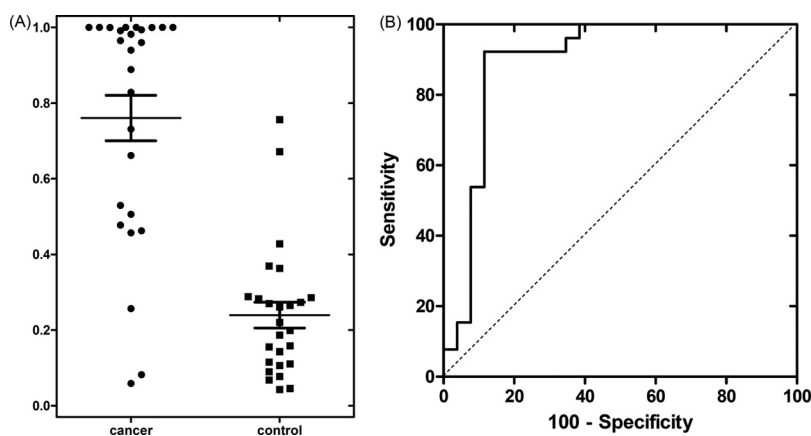


FIG. 5. Performance of biomarker panel. A, the dot plots of logistic regression combined biomarkers of human calprotectin, HP and AZGP1; B, ROC curve for the combined three biomarkers.

corresponding ROC curve is shown in Fig. 5B. As listed in Table IIIA, the combined performance of HP, AZGP1, and human calprotectin can reach 88.5% sensitivity and 92.3% specificity with AUC = 0.90. The corresponding PPV (positive predictive value) and NPV (negative predictive value) were also shown.

Confirmation in Lung Cancer Cells—In order to determine whether the three salivary proteins were lung cancer related, all three proteins were tested in NSCLC cells (NCI-H1299 and NCI-H460 cell line) and healthy lung cells (MRC-5 cell line). Results show that the concentration of human calprotectin in MRC-5 cells is 0.156 ng/(μ g total protein). It up-regulated to 0.235 ng/(μ g total protein) and 0.336 ng/(μ g total protein) in NCI-H460 cells and NCI-H1299 cells, respectively. The Western blot of HP, AZGP1 and the corresponding β -actin are shown in Fig. 6A. Their quantification data (Fig. 6B) shows that these two proteins were also elevated in lung cancer cells. Collectively, the concentration of all the three proteins was increased in the two lung cancer cell lines when compared with the healthy lung cells. These findings were consistent with our findings in the human saliva, which might help confirm why these three proteins were elevated in lung cancer patients' saliva.

Influence of Smoking—Among the 52 samples in the pre-validation sample set, 33 of them have smoking history and 19 of them are nonsmokers. Besides, 17 lung cancer patients have smoking history and nine lung cancer patients are nonsmokers. The three biomarkers' distribution in these non-smoking and smoking subjects is comprehensively compared. The data is shown in Fig. 7. The significant difference between each group is labeled.

DISCUSSION

In this manuscript, we have demonstrated that discriminatory proteomic biomarkers are present in human saliva when lung cancer developed. Although a proof of concept study, this is the first study engaged a *de novo* biomarker discovery approach in saliva for lung cancer detection. Three proteins were first discovered by 2D-DIGE and MS and then further verified in the discovery sample set, prevalidation sample set and lung cancer cell lines. The discriminatory power of these candidate biomarkers in lung cancer patients and healthy control subjects can reach 92.0% PPV and 88.9% NPV (Table III). These data collectively demonstrated that these three salivary proteins could potentially be used as discriminatory biomarkers to differentiate lung cancer patients from healthy control subjects.

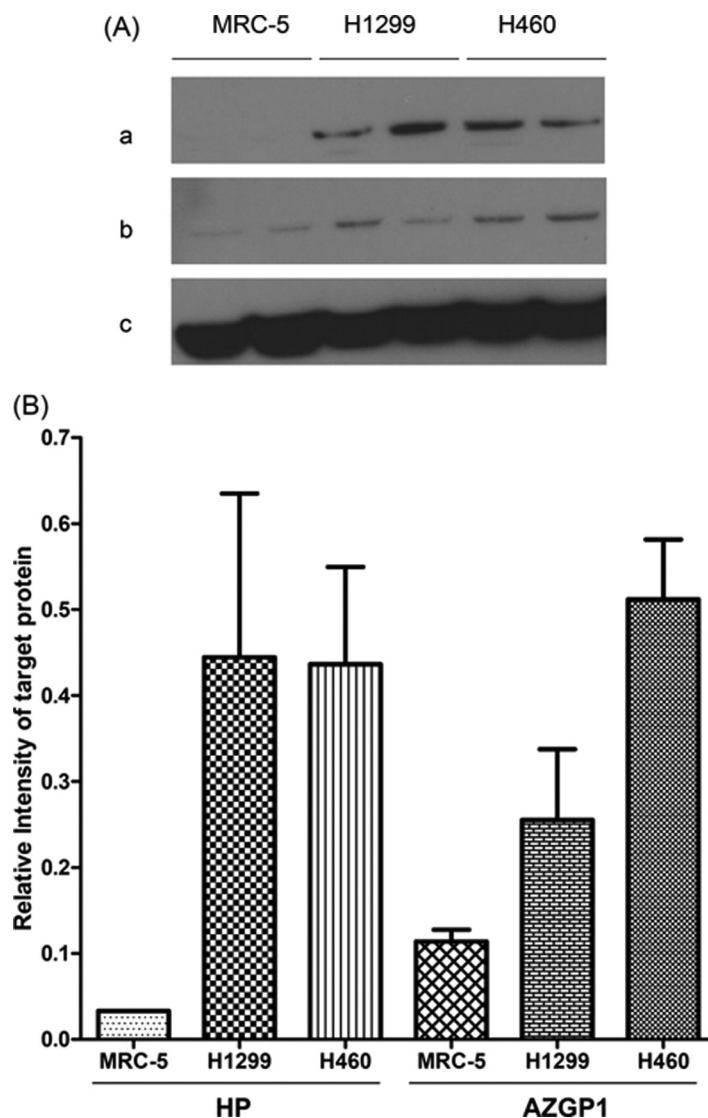


FIG. 6. **Biomarker verification in lung cancer cell lines.** A, The Western blot of HP (a) and AZGP1 (b) in different lung cancer cell lines. Their corresponding β -actin is shown in (c); B, The corresponding quantification of the Western blot in (A).

Candidate Biomarkers—In this study, our goal is to develop discriminatory proteomic biomarkers for the early detection of lung cancer. Although there were 253 spots located on the gel after in-gel image analysis, only spots with relatively large fold change (≥ 1.5) were exercised for further identification and verification. It is of interests that most of these identified proteins were lung cancer related. Of note is HP, which has been found in serum and used as biomarker for the diagnosis of NSCLC (2, 36). The overexpression of this acute phase protein in human lung cancer tissues has been confirmed by immunohistochemistry and Western blot (37). Notably, primary tumors of lung could release soluble factors that induce the expression human calprotectin, which could facilitate the survival and proliferation of metastasizing cancer cells (38). Human calprotectin is the complex of S100A8 and S100A9. It has antibacterial, antifungal, immunomodulating, and antipro-

liferative effects in human saliva. Whereas S100A8 alone did not inhibit fungal growth, S100A9 by itself had a moderate antifungal effect (39). One possible reason is that in case of disease, more S100A8 and S100A9 may conjugate together, which result in the up-regulation of human calprotectin and the down-regulation of S100A8 and S100A9 (Table II). The up-regulation of S100A8 and S100A9 in lung cancer tissue has been tested by immunohistochemical staining (40). The mRNA levels of AZGP1 in human lung tissue has also been reported to correlate with lung cancer disease status (41). The expression of AZGP1 protein in lung tissue is evidenced by immunohistochemical staining (6). The presence of AZGP1 serum autoantibody may be used as a prognostic marker in patients with lung adenocarcinoma. The up-regulation of AZGP1 mRNA in lung adenocarcinoma may be affected by chromatin remodeling by means of histone acetylation (42).

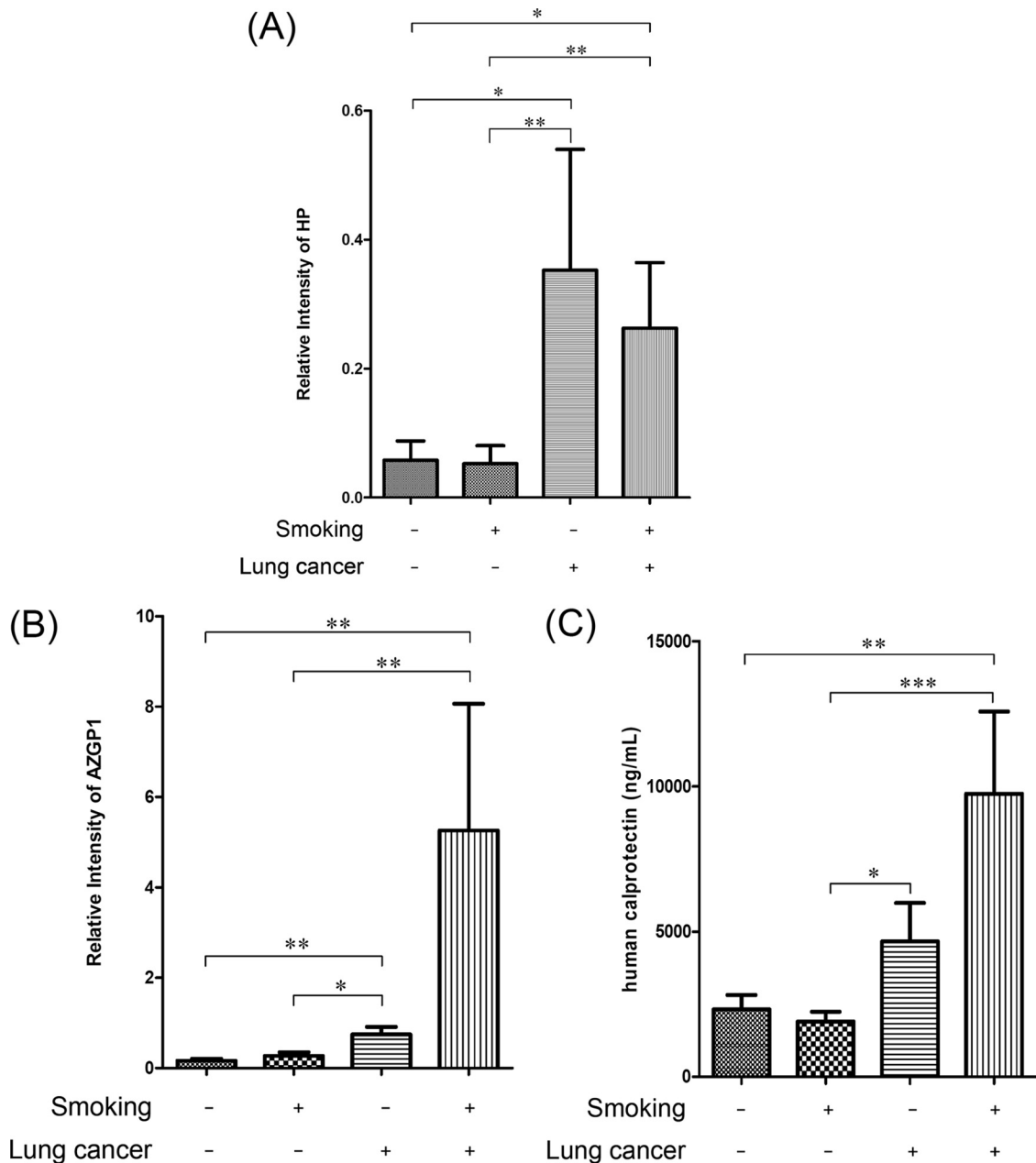


FIG. 7. The quantification data of HP (A), AZGP1 (B), and human calprotectin (C) in smoking and nonsmoking groups (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

Besides, annexins are cytosolic proteins that can be associate with cell membranes in a calcium-dependent manner and ANXA1 has been found up-regulated in lung tumors (43, 44). Of these selected candidates, three of them could be well detected. The negative result of ANXA1 may be because of its low abundance in saliva or the limited specificity of the antibody.

Based on the biomarker performance in the prevalidation sample set (Table III), different sensitivity and specificity have been calculated for this sample set based on the selected thresholds. In Table IIIA, the threshold was from the prevalidation

sample set that based on the maximum sum of sensitivity and specificity. In order to validate the model built in the discovery sample set, its threshold was applied to the prevalidation sample set (Table IIIB). Compared with that in Table III A, the sum of sensitivity and specificity in Table IIIB is generally weaker because of the selected thresholds. For the human calprotectin and HP combination, the threshold from the discovery sample set is less than that from the prevalidation sample set. This caused the ideal sensitivity (92.3%) while with frustrating specificity (46.2%). For the combination of all three proteins, even though a much higher threshold was

applied, good sensitivity (73.1%) and specificity (96.2%) were achieved. This data demonstrated the robust power of our discovered biomarkers in the differentiation of lung cancer patients from normal control subjects.

The Effects of Smoking on Salivary Biomarkers—Smoking is by far the leading risk factor for lung cancer (3, 4, 17, 45). Its effects on salivary biomarkers were comprehensively compared. In this study, smoking has elevated the expression of AZGP1 and human calprotectin and decreased the expression of HP in the cancer group (Fig. 7). However, our data show that smoking did not demonstrate a significant influence on these markers. It is worth mentioning that the distribution of HP and AZGP1 show the significant difference between the healthy control and lung cancer groups, no matter whether they are smokers or nonsmokers. In the healthy control subjects group, smoking does not have a significant effect on the expression of these proteins. These results demonstrate that these verified protein biomarkers were mostly elevated by lung cancer.

The Specificity of Salivary Proteomic Biomarkers to Lung Cancer—The clinical samples used in this study were matched by their age, sex, and ethnicity between the cancer group and control group. Their smoking history was matched by whether they are current or former smokers and their smoking duration and intensity.

Most lung cancer patients will have chronic obstructive pulmonary disease (COPD) because both conditions are mainly caused by smoking. Our verified salivary biomarkers also have been reported to be related with smoking (5, 6). References show that the three verified biomarkers also related with COPD in tissue (37), bronchoalveolar lavage fluids (46), and blood (16, 47). However, only 1 or 2% of COPD patients will go on to develop lung cancer. Though COPD is considered the “transition state” between normal and lung cancer, its protein profiling may not likely be the median. The goal of this study is to discover biomarkers that can differentiate normal subjects from lung cancer patients. Thus COPD was not included in this study for lung cancer biomarker discovery. However, to further verify the specificity of the discovered lung cancer biomarkers, COPD should be included in further definitive validation.

The Proof of Concept—Through the two phases of the pilot study of salivary biomarker development for lung cancer detection, the preliminary data show that three proteins were identified and preclinically validated. All of them were also confirmed to be elevated in lung cancer cells. The performance of these proteins was further evaluated. The distributions of these proteins in the saliva of lung cancer patients and matched healthy control subjects have demonstrated their discriminatory power for lung cancer detection. To the best of our knowledge, the present study is the first proof-of-concept report on *de novo* salivary proteomic biomarker discovery and prevalidation for lung cancer detection. Although further validation in a larger sample set is necessary for definitive vali-

ation, these discovered and confirmed salivary protein biomarkers have the potential to be used for the detection of lung cancer. A saliva test could be easily conducted in the clinic and would have the potential to detect disease at an earlier time point, when the likelihood of curative therapy would be greater.

Acknowledgments—We thank Dr. Jieping Yang for the kind help in lung cancer cell experiments. We thank David Akin for assistance in clinical sample collection, process, and storage.

* This work was supported by National Institutes of Health grants U01-DE16275 and R21-CA0126733 to DTW.

¶¶ To whom correspondence should be addressed: Dental Research Institute, School of Dentistry, University of California-Los Angeles, 10833 Le Conte Ave, CHS 73-024, Los Angeles, CA 90095. Tel.: 1-310-206-1138; Fax: 1-310-825-7609; E-mail: dtww@ucla.edu.

** To whom correspondence should be addressed: Dental Research Institute, School of Dentistry, University of California-Los Angeles, 10833 Le Conte Ave, CHS 73-024, Los Angeles, CA 90095. Tel.: 1-310-206-1138; Fax: 1-310-825-7609; E-mail: xiaohuacn@gmail.com.

||| DTW is also supported by the Felix & Mildred Yip Endowed Professorship.

REFERENCES

1. Ilie, M., Mazure, N. M., Hofman, V., Ammadi, R. E., Ortholan, C., Bonnetaud, C., Havet, K., Venissac, N., Mograbi, B., Mouroux, J., Pouyssegur, J., and Hofman, P. (2010) High levels of carbonic anhydrase IX in tumour tissue and plasma are biomarkers of poor prognostic in patients with non-small cell lung cancer. *Br. J. Cancer* **102**, 1627–1635
2. Hoagland, L. F., 4th, Campa, M. J., Gottlin, E. B., Herndon, J. E., 2nd, and Patz, E. F., Jr. (2007) Haptoglobin and Posttranslational glycan-modified derivatives as serum biomarkers for the diagnosis of nonsmall cell lung cancer. *Cancer* **110**, 2260–2268
3. Pleasance, E. D., Stephens, P. J., O’Meara, S., McBride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C., Varela, I., Nik-Zainal, S., Davies, H. R., Ordoñez, G. R., Mudie, L. J., Latimer, C., Edkins, S., Stebbings, L., Chen, L., Jia, M., Leroy, C., Marshall, J., Menzies, A., Butler, A., Teague, J. W., Mangion, J., Sun, Y. A., McLaughlin, S. F., Peckham, H. E., Tsung, E. F., Costa, G. L., Lee, C. C., Minna, J. D., Gazdar, A., Birney, E., Rhodes, M. D., McKernan, K. J., Stratton, M. R., Futreal, P. A., and Campbell, P. J. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190
4. Hecht, S. S. (2002) Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention. *Lancet Oncol.* **3**, 461–469
5. Jessie, K., Pang, W. W., Haji, Z., Rahim, A., and Hashim, O. H. (2010) Proteomic Analysis of Whole Human Saliva Detects Enhanced Expression of Interleukin-1 Receptor Antagonist, Thioredoxin and Lipocalin-1 in Cigarette Smokers Compared to Non-Smokers. *Int. J. Mol. Sci.* **11**, 4488–4505
6. Vanni, H., Kazeros, A., Wang, R., Harvey, B. G., Ferris, B., De, B. P., Carolan, B. J., Hübner, R. H., O’Connor, T. P., and Crystal, R. G. (2009) Cigarette Smoking Induces Overexpression of a Fat-Depleting Gene AZGP1 in the Human. *Chest* **135**, 1197–1208
7. Gohagan, J. K., Marcus, P. M., Fagerstrom, R. M., Pinsky, P. F., Kramer, B. S., Prorok, P. C., Ascher, S., Bailey, W., Brewer, B., Church, T., Engelhard, D., Ford, M., Fouad, M., Freedman, M., Gelmann, E., Gierada, D., Hocking, W., Inampudi, S., Irons, B., Johnson, C. C., Jones, A., Kucera, G., Kvale, P., Lappe, K., Manor, W., Moore, A., Nath, H., Neff, S., Oken, M., Plunkett, M., Price, H., Reding, D., Riley, T., Schwartz, M., Spizarny, D., Yoffie, R., and Zylak, C. (2005) Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. *Lung Cancer* **47**, 9–15
8. Hirsch, F. R., Franklin, W. A., Gazdar, A. F., and Bunn, P. A., Jr. (2001) Early detection of lung cancer: Clinical perspectives of recent advances in

- biology and radiology. *Clin. Cancer Res.* **7**, 5–22
9. Spira, A., Beane, J. E., Shah, V., Steiling, K., Liu, G., Schembri, F., Gilman, S., Dumas, Y. M., Calner, P., Sebastiani, P., Sridhar, S., Beamis, J., Lamb, C., Anderson, T., Gerry, N., Keane, J., Lenburg, M. E., and Brody, J. S. (2007) Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* **13**, 361–366
 10. Brower, V. (2009) Biomarker Studies Abound for Early Detection of Lung Cancer. *J. Natl. Cancer Inst.* **101**, 11–13
 11. Peng, G., Tisch, U., Adams, O., Hakim, M., Shehada, N., Broza, Y. Y., Billan, S., Abdah-Bortnyak, R., Kuten, A., and Haick, H. (2009) Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat. Nanotechnol.* **4**, 669–673
 12. Brambilla, C., Fievet, F., Jeanmart, M., de Fraipont, F., Lantuejoul, S., Frappat, V., Ferretti, G., Brichon, P. Y., and Moro-Sibilot, D. (2003) Early detection of lung cancer: role of biomarkers. *Eur. Respir. J.* **21**, 36S–44S
 13. Hirsch, F. R., Merrick, D. T., and Franklin, W. A. (2002) Role of biomarkers for early detection of lung cancer and chemoprevention. *Eur. Respir. J.* **19**, 1151–1158
 14. Yau, G., Lock, M., and Rodrigues, G. (2007) Systematic review of baseline low-dose CT lung cancer screening. *Lung Cancer* **58**, 161–170
 15. van Klaveren, R. J., Oudkerk, M., Prokop, M., Scholten, E. T., Nackaerts, K., Vernhout, R., van Iersel, C. A., van den Bergh, K. A., van 't Westeinde, S., van der Aalst, C., Thunnissen, E., Xu, D. M., Wang, Y., Zhao, Y., Gietema, H. A., de Hoop, B. J., Groen, H. J., de Bock, G. H., van Ooijen, P., Weenink, C., Verschakelen, J., Lammers, J. W., Timens, W., Willebrand, D., Vink, A., Mali, W., and de Koning, H. J. (2009) Management of lung nodules detected by volume CT scanning. *N. Engl. J. Med.* **361**, 2221–2229
 16. Yee, J., Sadar, M. D., Sin, D. D., Kuzyk, M., Xing, L., Kondra, J., McWilliams, A., Man, S. F., and Lam, S. (2009) Connective Tissue-Activating Peptide III: A Novel Blood Biomarker for Early Lung Cancer Detection. *J. Clin. Oncol.* **27**, 2787–2792
 17. Ding, L., Getz, G., Wheeler, D. A., Mardis, E. R., McLellan, M. D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D. M., Morgan, M. B., Fulton, L., Fulton, R. S., Zhang, Q., Wendl, M. C., Lawrence, M. S., Larson, D. E., Chen, K., Dooling, D. J., Sabo, A., Hawes, A. C., Shen, H., Jhangiani, S. N., Lewis, L. R., Hall, O., Zhu, Y., Mathew, T., Ren, Y., Yao, J., Scherer, S. E., Clerc, K., Metcalf, G. A., Ng, B., Milosavljevic, A., Gonzalez-Garay, M. L., Osborne, J. R., Meyer, R., Shi, X., Tang, Y., Koboldt, D. C., Lin, L., Abbott, R., Miner, T. L., Pohl, C., Fewell, G., Haipek, C., Schmidt, H., Dunford-Shore, B. H., Kraja, A., Crosby, S. D., Sawyer, C. S., Vickery, T., Sander, S., Robinson, J., Winckler, W., Baldwin, J., Chiriac, L. R., Dutt, A., Fennell, T., Hanna, M., Johnson, B. E., Onofrio, R. C., Thomas, R. K., Tonon, G., Weir, B. A., Zhao, X., Ziaugra, L., Zody, M. C., Giordano, T., Orringer, M. B., Roth, J. A., Spitz, M. R., Wistuba, II, Ozenberger, B., Good, P. J., Chang, A. C., Beer, D. G., Watson, M. A., Ladanyi, M., Broderick, S., Yoshizawa, A., Travis, W. D., Pao, W., Province, M. A., Weinstock, G. M., Varmus, H. E., Gabriel, S. B., Lander, E. S., Gibbs, R. A., Meyerson, M., and Wilson, R. K. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075
 18. Patz, E. F., Campa, M. J., Gottlin, E. B., Kusmartseva, I., Guan, X. R., and Herndon, J. E., 2nd (2007) Panel of serum biomarkers for the diagnosis of lung cancer. *J. Clin. Oncol.* **25**, 5578–5583
 19. Palmisano, W. A., Divine, K. K., Saccomanno, G., Gilliland, F. D., Baylin, S. B., Herman, J. G., and Belinsky, S. A. (2000) Predicting lung cancer by detecting aberrant promoter methylation in sputum. *Cancer Res.* **60**, 5954–5958
 20. Belinsky, S. A., Klinge, D. M., Dekker, J. D., Smith, M. W., Bocklage, T. J., Gilliland, F. D., Crowell, R. E., Karp, D. D., Stidley, C. A., and Picchi, M. A. (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin. Cancer Res.* **11**, 6505–6511
 21. Gray, R. D., Duncan, A., Noble, D., Imrie, M., O'Reilly, D. S., Innes, J. A., Porteous, D. J., Greening, A. P., and Boyd, A. C. (2010) Sputum Trace Metals Are Biomarkers of Inflammatory and Suppurative Lung Disease. *Chest* **137**, 635–641
 22. Haura, E. B., Muller, A., Breitwieser, F. P., Li, J. N., Grebien, F., Colinge, J., and Bennett, K. L. (2011) Using iTRAQ Combined with Tandem Affinity Purification to Enhance Low-Abundance Proteins Associated with Somatic Mutated EGFR Core Complexes in Lung Cancer. *J. Proteome Res.* **10**, 182–190
 23. Bortner, J. D., Richie, J. P., Das, A., Liao, J., Umstead, T. M., Stanley, A., Stanley, B. A., Belani, C. P., and El-Bayoumy, K. (2011) Proteomic Profiling of Human Plasma by iTRAQ Reveals Down-Regulation of ITI-HC3 and VDBP by Cigarette Smoking. *J. Proteome Res.* **10**, 1151–1159
 24. Pernemalm, M., De Petris, L., Eriksson, H., Branden, E., Koyi, H., Kanter, L., Lewensohn, R., and Lehtiö, J. (2009) Use of narrow-range peptide IEF to improve detection of lung adenocarcinoma markers in plasma and pleural effusion. *Proteomics* **9**, 3414–3424
 25. Lou, X. M., Xiao, T., Zhao, K., Wang, H., Zheng, H. W., Lin, D., Lu, Y., Gao, Y., Cheng, S., Liu, S., and Xu, N. (2007) Cathepsin D is secreted from M-BE cells: Its potential role as a biomarker of lung cancer. *J. Proteome Res.* **6**, 1083–1092
 26. Kaufman, E., and Lamster, I. B. (2002) The diagnostic applications of saliva - A review. *Crit. Rev. Oral Biol. Med.* **13**, 197–212
 27. Hu, S., Loo, J. A., and Wong, D. T. (2007) Human saliva proteome analysis and disease biomarker discovery. *Expert Rev. Proteomics* **4**, 531–538
 28. Hu, S., Yen, Y., Ann, D., and Wong, D. T. (2007) Implications of salivary proteomics in drug discovery and development: a focus on cancer drug discovery. *Drug Discov. Today* **12**, 911–916
 29. Hu, S., Arellano, M., Boontheung, P., Wang, J. H., Zhou, H., Jiang, J., Elashoff, D., Wei, R., Loo, J. A., and Wong, D. T. (2008) Salivary Proteomics for Oral Cancer Biomarker Discovery. *Clin. Cancer Res.* **14**, 6246–6252
 30. Hu, S., Wang, J. H., Meijer, J., Leong, S., Xie, Y. M., Yu, T., Zhou, H., Henry, S., Vissink, A., Pijpe, J., Kallenberg, C., Elashoff, D., Loo, J. A., and Wong, D. T. (2007) Salivary proteomic and genomic biomarkers for primary Sjogren's syndrome. *Arthritis Rheum.* **56**, 3588–3600
 31. Zhang, L., Xiao, H., Karlan, S., Zhou, H., Gross, J., Elashoff, D., Akin, D., Yan, X., Chia, D., Karlan, B., and Wong, D. T. (2010) Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. *PLoS ONE* **5**, e15573
 32. Zhang, L., Farrell, J. J., Zhou, H., Elashoff, D., Akin, D., Park, N. H., Chia, D., and Wong, D. T. (2010) Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology* **138**, 949–957, e941–947
 33. Xiao, H., and Wong, D. T. (2010) Proteomics and its applications for biomarker discovery in human saliva. *Bioinformatics* **5**, 294–296
 34. Vitorino, R., Lobo, M. J. C., Ferrer-Correia, A. J., Dubin, J. R., Tomer, K. B., Domingues, P. M., and Amado, F. M. (2004) Identification of human whole saliva protein components using proteomics. *Proteomics* **4**, 1109–1115
 35. Gao, K., Zhou, H., Zhang, L., Lee, J. W., Zhou, Q., Hu, S., Wolinsky, L. E., Farrell, J., Eibl, G., and Wong, D. T. (2009) Systemic Disease-Induced Salivary Biomarker Profiles in Mouse Models of Melanoma and Non-Small Cell Lung Cancer. *PLoS ONE* **4**, e5875
 36. Bharti, A., Ma, P. C., Maulik, G., Singh, R., Khan, E., Skarin, A. T., and Salgia, R. (2004) Haptoglobin alpha-subunit and hepatocyte growth factor can potentially serve as serum tumor biomarkers in small cell lung cancer. *Anticancer Res.* **24**, 1031–1038
 37. Abdullah, M., Schultz, H., Kahler, D., Branscheid, D., Dalhoff, K., Zabel, P., Vollmer, E., and Goldmann, T. (2009) Expression of the acute phase protein haptoglobin in human lung cancer and tumor-free lung tissues. *Pathol. Res. Practice* **205**, 639–647
 38. Ghavami, S., Chitayat, S., Hashemi, M., Eshraghi, M., and Chazin, W. J. (2009) S100A8/A9: A Janus-faced molecule in cancer therapy and tumorigenesis. *Eur. J. Pharmacol.* **625**, 73–83
 39. Sroussi, H. Y., Köhler, G. A., Agabian, N., Villines, D., and Palefsky, J. M. (2009) Substitution of methionine 63 or 83 in S100A9 and cysteine 42 in S100A8 abrogate the antifungal activities of S100A8/A9: potential role for oxidative regulation. *FEMS Immunol. Med. Microbiol.* **55**, 55–61
 40. Su, Y. J., Xu, F., Yu, J. P., Yue, D. S., Ren, X. B., and Wang, C. L. (2010) Up-regulation of the expression of S100A8 and S100A9 in lung adenocarcinoma and its correlation with inflammation and other clinical features. *Chin. Med. J. (Engl.)* **123**, 2215–2220
 41. Falvella, F. S., Spinola, M., Pignatiello, C., Noci, S., Conti, B., Pastorino, U., Carbone, A., and Dragani, T. A. (2008) AZGP1 mRNA levels in normal human lung tissue correlate with lung cancer disease status. *Oncogene* **27**, 1650–1656
 42. Albertus, D. L., Seder, C. W., Chen, G., Wang, X. J., Hartojo, W., Lin, L., Silvers, A., Thomas, D. G., Giordano, T. J., Chang, A. C., Orringer, M. B.,

- Bigbee, W. L., Chinnaiyan, A. M., and Beer, D. G. (2008) AZGP1 Autoantibody Predicts Survival and Histone Deacetylase Inhibitors Increase Expression in Lung Adenocarcinoma. *J. Thorac. Oncol.* **3**, 1236–1244
43. Brichory, F. M., Misek, D. E., Yim, A. M., Krause, M. C., Giordano, T. J., Beer, D. G., and Hanash, S. M. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9824–9829
44. Oh, P., Li, Y., Yu, J. Y., Durr, E., Krasinska, K. M., Carver, L. A., Testa, J. E., and Schnitzer, J. E. (2004) Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* **429**, 629–635
45. Church, T. R., Anderson, K. E., Caporaso, N. E., Geisser, M. S., Le, C. T., Zhang, Y., Benoit, A. R., Carmella, S. G., and Hecht, S. S. (2009) A Prospectively Measured Serum Biomarker for a Tobacco-Specific Carcinogen and Lung Cancer in Smokers. *Cancer Epidemiol. Biomarkers Prev.* **18**, 260–266
46. Merkel, D., Rist, W., Seither, P., Weith, A., and Lenter, M. C. (2005) Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. *Proteomics* **5**, 2972–2980
47. Engstrom, G., Segelstorm, N., Ekberg-Aronsson, M., Nilsson, P. M., and Lindgärde, F. (2009) Plasma markers of inflammation and incidence of hospitalisations for COPD: results from a population-based cohort study. *Thorax* **64**, 211–215