Identification of Novel Low Molecular Weight Serum Peptidome Biomarkers for Non-Small Cell Lung Cancer (NSCLC)

Juan Yang,^{1†} Yong-Chun Song,^{2†} Tu-Sheng Song,¹ Xiao-Yan Hu,¹ You-Min Guo,³
Zong-Fang Li,⁴ Cheng-Xue Dang,^{2*} and Chen Huang^{1*}

¹Department of Genetics and Molecular Biology, Medical school of Xi'an Jiaotong University/Key Laboratory of Environment and Disease-Related Gene, Ministry of Education, PR China

²Department of Tumor Surgery, the First Affiliated Hospital, Xi'an Jiaotong University College of Medicine, PR China

³Department of Medical Image, the First Affiliated Hospital of Xi'an Jiaotong University, PR China

⁴Department of Respiratory Medicine, the Second Affiliated Hospital of Xi'an Jiaotong University/Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education, PR China

Aim: To identify discriminating protein patterns in serum samples among non-small cell lung cancer (NSCLC), chronic obstructive pulmonary disease (COPD), pneumonia, and healthy controls. To discover specific low molecular weight (LMW) serum peptidome biomarkers and establish a diagnostic pattern for NSCLC by using proteomic technology. **Methods:** We used magnetic bead-based separation followed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify patients with NSCLC, COPD, and pneumonia. A total of 154 serum samples were analyzed in this study, among which there were 60 serum samples from NSCLC patients, 30 from patients with other lung-related diseases (16 pneumonia patients and 14 patients with

COPD) as disease controls, and 64 from healthy volunteers as healthy control. The mass spectra, analyzed using ClinProTools software, distinguished between cancer patients and healthy individuals based on GA algorithm model. **Results:** In this study, we generated numerous discriminating m/z peaks as well as disease-specific discrimination peaks. A set of five potential biomarkers (m/z : 7,763.24, 1,012.61, 4,153.16, 1,450.55, and 2,878.89) could be used as the diagnostic biomarkers to distinguish NSCLC patients from healthy controls. In the training set, patients with NSCLC could be identified with sensitivity of 97.5% and specificity of 98.8%. Similar results were obtained in the testing set, showing 80.7% sensitivity and 91.2% specificity.

[†]These authors contributed equally to this work.

Grant sponsor: National Science Foundation for Postdoctoral Scientists of China; Grant number: 20090461301; Grant sponsor: Young Scientist Foundation from the Medical School in Xi'an Jiaotong University; Grant number: YQN0809; Grant sponsor: Scientific Research Support Program for New Teachers; Grant number: 0116-081410-05; Grant sponsor: Postdoctoral Science Foundation of Xi'an Jiaotong University; Grant number: 2116-04212217; Grant sponsor: Guang Hua Medical Innovation Research Foundation; Grant number: 0203407; Grant sponsor: International Sci-tech Cooperation Program; Grant number: 2009DFA31420; Grant sponsor: National Natural Science Foundation of China; Grant number: 30900364.

*Correspondence to: Chen Huang, Department of Genetics and Molecular Biology, Medical School of Xi'an Jiaotong University/Key Laboratory of Environment and Disease-Related Gene, Ministry of Education, 76 Western Yanta Road, Xi'an, Shaanxi 710061, PR China. E-mail: hchen@mail.xjtu.edu.cn; or Cheng-Xue Dang, Department of Tumor Surgery, the First Affiliated Hospital, Xi'an Jiaotong University College of Medicine, 277 Western Yanta Road, Xi'an, Shaanxi, 710004, PR China. E-mail: Dangchengxue1@yahoo.com.cn

Received 1 September 2011; Accepted 4 February 2012

DOI 10.1002/jcla.21502

Published online in Wiley Online Library (wileyonlinelibrary.com).

Conclusion: Our study demonstrated that a combined application of magnetic beads with MALDI-TOF MS technique was suitable

for identification of serum biomarkers for NSCLC. *J. Clin. Lab. Anal.* 26:148–154, 2012. © 2012 Wiley Periodicals, Inc.

Key words: MALDI-TOF MS; NSCLC; COPD; pneumonia; magnetic beads; serum biomarkers

INTRODUCTION

Lung cancer is a challenging global clinical problem, and is the first killer among all types of cancer in China. Non-small cell lung cancer (NSCLC) accounts for 80% of all lung cancer cases (1, 2). In early stage, NSCLC surgery is almost curative. However, the majority of patients are diagnosed in advanced stage, so treatment is aimed at improving quality of life rather than curing the disease. To date, no screening tool has been widely available for the early detection of lung cancer (3). Although NSCLC has been found to express several serum markers, none has been adequately sensitive and specific to NSCLC (4–7). An ideal screening test should be noninvasive with high sensitivity and specificity (8). Thus, identifying novel biomarkers with high specificity and sensitivity are necessary for improving the early detection of NSCLC.

Low molecular weight (LMW) proteins/peptidome contain an unexplored archive of histological information, and are expected to yield useful biomarkers for early disease detection (9). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) is an important proteomic technology, which has been most widely used to analyze the serum LMW proteome (10, 11). Magnetic bead-based purification approaches can capture large amount of LMW proteins/peptides, and make serum profiling suitable for general MS analyses (11, 12).

Chronic obstructive pulmonary disease (COPD) and pneumonia are major sources of inflammation in lung tissues, and are associated with an increased risk of developing lung cancer (13). Thus, both of them are interesting comparators in NSCLC proteomics biomarker discovery studies. In the present study, MALDI-TOF-MS analysis coupled with MB (Magnetic Beads) weak cation exchange (MB-WCX) has been used to detect serum proteome of 60 patients with NSCLC, 30 patients with other lung-related diseases (16 pneumonia patients and 14 COPD patients), and 64 healthy volunteers. Based on these data, we aimed to: (1) identify the LMW serum proteomic profiles of patients with NSCLC, pneumonia, and COPD; (2) screen specific-disease discrimination peaks; (3) discover novel LMW serum protein/peptide biomarkers for NSCLC; and (4) ultimately construct a diagnostic model for improving diagnostic efficiency of NSCLC.

MATERIALS AND METHODS

Patient Specimens and Sample Collection

This study was approved by the Ethics Committee and the Human Research Review Committee of Xi'an Jiaotong University. A total of 154 serum samples were included in this study, of which 60 were collected from NSCLC patients (28 for training set and 32 for testing set), 30 from patients with other lung-related diseases (16 pneumonia patients and 14 patients with COPD) as disease controls, and 64 from healthy volunteers as healthy control (32 for training set and 32 for testing set). The details of serum samples and the data set were listed in Table 1. The sample collection was conducted in the Second Affiliated Hospital of Xi'an Jiaotong University between August 2008 and May 2009. All the patients were recently diagnosed. All blood samples were drawn while the patients or healthy controls were seated and nonfasting. The samples were collected in 10-cc serum separator tubes at 4°C for 1 hr, and then centrifuged at 3,000 rpm at 4°C for 20 min. The serum samples were distributed into 500- μ L aliquots and stored at –80°C until subsequent uses.

Sample Preparation and MALDI-TOF MS Analysis

We used MB-WCX chromatography (ClinProt™ purification reagent sets of Bruker Daltonics) for LMW protein separation of samples. MB-WCX purifications were performed according to the manufacturer's protocol for serum and with the Bruker Magnetic Separator (8 well, #65554). To prepare the MALDI target, we spotted 1 μ L of a mixture containing 10 μ L 0.3 g/L α -cyano-4-hydroxy cinnamic acid in 2:1 ethanol/acetone (volume/volume) and 1 μ L of the eluted peptide fraction onto the MALDI AnchorChip™ (Bruker Daltonics, Germany) sample target platform (384 spots). To evaluate the reproducibility, each serum sample was spotted for three repeats.

Air-dried targets were measured immediately using a calibrated Autoflex III MALDI-TOF MS (Bruker Daltonics, Germany), FlexControl software (version 3.0; Bruker), and optimized measuring protocols. Matrix suppression up to 1,000 Da, with a mass range of 1,000–10,000 Da was set as the default. Instrument calibration parameters were determined using standard peptide and

TABLE 1. Details of Serum Samples and Data Set

Type	Sample size	Gender	Mean age	Age range	Training set	Testing set
NSCLC	60	18 F/42 M	62.47	31–84	28	32
Pneumonia	16	5 F/11 M	54.17	39–68	16	0
COPD	14	4 F/10 M	65.34	53–74	14	0
Healthy controls	64	32 F/32 M	58.375	34–76	32	32

protein mixtures. All measurements were performed in a blinded manner, including patient and control sera in one mixed approach.

Data Processing with ClinProTools Software

Data analyses were performed using the programs Flex analysis v3.0 and ClinProTools v2.2 (Bruker Daltonics, Germany). ClinProTools v2.2 uses a standard data preparation workflow including spectra pretreatment, peak picking, and peak calculation operation, and was employed to recognize peptide patterns in this study. For statistical analysis, a *k*-nearest neighbor genetic algorithm as implemented in this software suite was used to identify statistically significant differences in protein peaks among the groups analyzed. The protein fingerprint data were analyzed by ClinProTools v2.2. Comparisons among NSCLC, pneumonia, COPD, and healthy controls were performed with the Wilcoxon test; statistical significance was assumed when *P* value was < 0.001.

RESULTS

Identification of Serum LMW Protein Profile

LMW protein profile spectra of 154 serum samples were detected by MALDI-TOF MS combined with MB-WCX magnetic beads, which has proved particularly effective in resolving LMW protein/peptides in human serum (Fig. 1). In this study, we analyzed the LMW protein profile spectra of all 154 serum samples (60 NSCLC patients, 16 pneumonia patients and 14 patients with COPD, and 64 healthy controls) (Table 1). We evaluated LMW protein changes in the serum samples of 28 NSCLC, 16 pneumonia, and 14 COPD patients, and then compared them with one another and 32 healthy controls in the training set. By analyzing the spectra (screened from any two groups in the training set) using the ClinProTools software 2.2, we were able to identify proteomic patterns that can clearly distinguish between NSCLC patients and healthy controls as well as pneumonia or COPD patients and healthy controls.

Comparison of Mass Spectra among NSCLC, Pneumonia, COPD Patients, and Controls

The comparison of mass spectra (1–10 kDa) generated by NSCLC, pneumonia, COPD, and healthy control was shown in Figure 1. Within this mass range, large numbers of differentially expressed proteins or peptides could be detected. By comparing three patient groups with the health controls, a combined application of MB-WCX magnetic beads and MALDI-TOF MS detected a total of 108, 113, and 115 *m/z* peaks, with 64, 35, and 46 of them being significant (*P* < 0.001), in NSCLC, pneumonia, and COPD groups, respectively. Accordingly, 20, 5, and 14 significant peaks were identified as specific for each group (Figs. 2, 3 and 4). Moreover, 30 significant peaks were shared between NSCLC and pneumonia groups, and 32 significant peaks between NSCLC and COPD groups (Table 2). A total of 18 significant peaks were detected across all three groups (Table 2).

Screening of LMW Protein Biomarkers of NSCLC and Construction of Diagnostic Model

All detected peaks were used with a *k*-nearest neighbor genetic algorithm in ClinProt system to generate a cross-validated classification model. The model comprised a set of five potential biomarkers (*m/z*: 7,763.24, 1,012.61, 4,153.16, 1,450.55, and 2,878.89) with their AUC (Area Under Curve) values of 0.89, 0.97, 0.88, 0.93, and 0.95, respectively. While the peaks with *m/z* 7,763.24, 1,012.61, and 4,153.16 were downregulated in NSCLC group, the other two biomarkers (*m/z*: 1,450.55 and 2,878.89) were upregulated in NSCLC group. Analyzing the training set based on the GA algorithm model, NSCLC patients could be distinguished from healthy controls with 97.5% sensitivity and 98.8% specificity. Analysis of spectra from the completely blinded test set (32 NSCLC patients and 32 healthy controls) accurately discriminated NSCLC patients from healthy controls with sensitivity of 80.7% and specificity of 91.2%.

DISCUSSION

Human LMW serum peptidome contains low-abundance peptide biomarkers produced from specific and ongoing tumorigenic processes. Our study showed

Fig. 1. Mass spectra (1,000–10,000 Da) obtained from NSCLC, pneumonia and COPD patients, and healthy controls. m/z , mass-to-charge ratio.

that the approach for LMW serum peptidome, coupled with a combined use of MALDI-TOF MS analysis and simple preparations using magnetic beads, were suitable for yielding useful biomarkers for NSCLC detection. In

this study, we generated numerous discriminating m/z peaks as well as disease-specific discrimination peaks; those peaks could accurately distinguish NSCLC patients from healthy individuals. Our findings confirmed that a

Fig. 2. Comparison of the average expression levels of 20 NSCLC-specific m/z peaks ($P < 0.001$) between NSCLCs and controls.

Fig. 3. Comparison of the average expression levels of five pneumonia-specific m/z peaks ($P < 0.001$) between pneumonias and controls.

combined application of Magnetic beads with MALDI-TOF MS technique and ClinProTools v2.2 is suitable for LMW serum proteomic analyses.

In previous reports, the cancer or disease group was frequently compared with healthy controls, and no cancer-related disease control was included (14–17). Thus, their identified biomarkers were not considered as disease specific. Brenne et al. (2011) revealed that previous lung diseases (COPD, pneumonia, and tuberculosis) are associated with an increased risk of lung cancer with the evidence among nonsmokers supporting a direct relationship between previous lung diseases and lung cancer (13). Therefore, in the present study, we selected COPD and pneumonia lung-related diseases for disease controls. To our knowledge, our study represents the first effort to describe serum profiling in NSCLC in comparison with

COPD and pneumonia based on MALDI-TOF MS and magnetic beads. We found that almost half the significant peaks identified between NSCLC and healthy controls were also shared among COPD and pneumonia groups, further indicating the necessity of including cancer-related disease controls in such studies.

Using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) technique, Yang et al. (2009) reported that the peak with m/z of 6,628 (identified as apolipoprotein C-I) was down-regulated in NSCLC patients, and thus could be used as specific biomarkers to NSCLC. While in our study, a similar significant discriminating m/z of 6,629.53 ($P < 0.001$) was found downregulated not only in NSCLC but also in pneumonia (Table 2). Du et al. (2005) identified five mass peaks that correlated with SCLC, with

Fig. 4. Comparison of the average expression levels of 14 COPD-specific m/z peaks ($P < 0.001$) between COPDs and controls.

TABLE 2. The 64 Discriminating m/z Peaks ($P < 0.001$) between NSCLC and Healthy Controls and Their Average Expression Levels in NSCLC, COPD, Pneumonia, and Healthy Controls

Mass	Control	NSCLC	Pneumonia	COPD	Mass	Control	NSCLC	Pneumonia	COPD
1,546.43	7.66	3.71	####	4.3	2,953.31	14.42	19.73	####	27.32
7,763.24	9.82	4.77	7.97	####	1,450.55	2.81	4.21	####	####
2,883.76	2.43	3.97	####	3.94	1,779.31	2.12	3.73	####	####
2,093.19	2.19	1.49	####	1.35	7,007.25	0.72	0.62	####	####
7,563.83	0.49	0.36	####	####	4,194.12	7.64	5.84	####	4.49
6,047.64	0.85	1.28	####	####	7,651.87	0.6	0.49	####	####
7,920.5	0.68	0.48	####	####	8,562.86	0.47	0.64	####	0.88
8,139.21	1.14	0.76	####	####	1,082.44	1.64	3.42	####	####
2,932.99	5.37	9.06	####	8.9	2,769.86	4.07	6.38	####	####
1,887.15	2.15	2.99	####	####	5,753.12	1.76	2.93	####	####
1,897.93	1.09	2.66	####	####	4,281.54	1.27	2.03	####	3.05
3,883.45	7.01	4.26	6.72	####	1,207.13	2.9	2.38	####	2.28
1,741.38	2.11	3.35	####	####	4,169.93	4.02	3.29	####	2.64
3,952.19	8.86	5.94	####	5.74	8,762.73	0.48	0.33	####	####
7,468.37	0.54	0.69	####	0.88	1,563.43	2.04	2.42	3.49	####
5,292.77	0.91	1.23	####	####	1,331.12	2.14	2.55	####	####
5,079.68	0.89	1.35	####	####	2,545.83	1.73	1.95	####	####
1,350.83	1.75	2.37	1.91	####	1,866.42	3.02	12.22	####	####
2,554.64	2.14	3.23	####	3.38	8,812.29	0.47	0.39	0.32	0.36
4,529.7	0.96	1.18	####	1.35	1,945.53	2.41	1.45	1.16	####
1,012.61	3.66	1.57	1.18	1.55	2,281.08	1.47	3.3	3.77	####
1,519.98	2.72	1.9	1.78	1.51	1,563.11	2.14	2.42	3.49	####
1,945.53	2.41	1.45	1.16	####	2,281.08	1.47	3.3	3.77	####
2,669.09	4.77	7.25	7.3	9.23	2,682.65	1.67	3.42	4.92	2.92
2,878.89	1.65	4.8	6.59	3.46	2,900.91	1.63	2.58	3.03	2.43
3,208.49	1.56	2.64	2.34	####	3,279.09	1.87	4.27	3.58	####
3,934.92	2.9	1.99	1.57	1.51	4,054.17	7.91	5.34	3.69	3.57
4,153.16	2.9	1.79	1.41	1.43	5,247.81	1.02	1.46	1.54	1.57
5,264.02	0.82	1.21	1.47	1.57	5,336.3	8.92	13.12	15.11	18.91
6,387.9	0.89	0.55	0.4	0.59	6,431.4	4.32	2.16	1.34	2.53
6,529.26	0.82	0.61	0.43	0.56	6,629.53	10.42	5.98	3.54	####
6,937.21	0.74	0.55	0.51	0.6	8,686.87	0.69	0.43	0.41	0.43

####: No expression identified.

peaks corresponding to m/z values of 4,172.21 and 6,336.6 showing downregulation, and peaks corresponding to m/z values of 5,336.83, 7,469.23, and 1,211.27 showing upregulation, by comparing spectra generated from SCLC samples between 30 SCLC patients and 44 healthy controls based on MALDI-TOF MS (19). Among these peaks, the 5,336 m/z was also identified in our study (Table 2) with $P < 0.001$ and also showing upregulation in NSCLC, COPD, and pneumonia. And from our study, it showed that 5,336 m/z was not NSCLC specific biomarker, it also expressed in COPD, pneumonia, and SCLC. And the expression levels were different in different lung-related disease. Therefore, these m/z peaks, which co-expressed in lung cancer and other lung-related diseases, are worth further investigation. In addition, Du et al. (2005) reported two significant discrimination peaks of m/z 1,865.79 and 1,778.67(19), which are similar to 1,866.42 and 1,779.31 in the present study (Table 2).

A single biomarker has an inherent specificity and sensitivity that cannot be improved, but multiple biomarkers can be combined to achieve improved clinical performances. In this study, a set of five potential biomarkers (m/z : 7,763.24, 1,012.61, 4,153.16, 1,450.55, and 2,878.89) can be used as the diagnostic biomarkers to distinguish NSCLC patients from healthy controls. Values of 97.5% sensitivity and 98.8% specificity were obtained in the training set and values of 80.7% sensitivity and 91.2% specificity in the testing set, both of which showed high predictive accuracy. The origin and full identity of these five protein peaks for the purpose of differential diagnosis is not required. Nevertheless, knowing their exact identities would be essential for understanding what biological role these proteins/peptides may take in the pathogenesis of NSCLC, potentially leading to novel therapeutic targets. Therefore, in future research, we would attempt to identify and characterize these five protein biomarkers and then validate them by western blot or Enzyme-linked immunosorbent assay (ELISA).

CONFLICT OF INTEREST

The authors confirm that they have no financial or personal relationships that cause a conflict of interest regarding the work in the manuscript.

REFERENCES

- Goldstraw P, Crowley J, Chansky K, et al. International Association for the Study of Lung Cancer International Staging Committee; Participating institutions. The IASLC Lung Cancer Staging Project: Proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *J Thorac Oncol* 2007;2:706–714.
- Vivian B, Stefan H, Benjamin N, et al. Relevance of circulating biomarkers for the therapy monitoring and follow-up investigations in patients with non-small cell lung cancer. *Cancer Biomark* 2010;6:191–196.
- Patz EF, Campa MJ, Gottlin EB, et al. Panel of serum biomarkers for the diagnosis of lung cancer. *J Clin Oncol* 2007;25:5578–5583.
- Tarro G, Perna A, Esposito C. Early diagnosis of lung cancer by detection of tumor liberated protein. *J Cell Physiol* 2005;203:1–5.
- Sung HJ, Cho JY. Biomarkers for the lung cancer diagnosis and their advances in proteomics. *BMB Rep* 2008;41:615–625.
- Heo SH, Lee SJ, Ryoo HM, et al. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC–MS/MS. *Proteomics* 2007;7:4292–4302.
- Haleem JI, Timothy JW, Timothy DV. Cancer biomarker discovery: Opportunities and pitfalls in analytical methods. *Electrophoresis* 2011;32:967–975.
- Rodríguez-Piñeiro AM, Blanco-Prieto S, Sánchez-Otero N, et al. On the identification of biomarkers for non-small cell lung cancer in serum and pleural effusion. *J Proteome* 2010;73:1511–1522.
- Kawashima Y, Fukutomi T, Tomonaga T, et al. High-yield peptide-extraction method for the discovery of subnanomolar biomarkers from small serum samples. *J Proteome Res* 2010;9:1694–1705.
- Albrethsen J. Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clin Chem* 2007;53:852–858.
- Wang QT, Li YZ, Liang YF, et al. Construction of a multiple myeloma diagnostic model by magnetic beads-based MALDI-TOF Mass spectrometry of serum and pattern recognition software. *Anat Rec* 2009;292:604–610.
- Fiedler GM, Baumann S, Leichtle A, et al. Standardized peptidome profiling of human urine by magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem* 2007;53:421–428.
- Brenne DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: A systematic review and meta-analysis. *PLoS One* 2011;6:e17479.
- Lam KWK, Lo SCL. Discovery of diagnostic serum biomarkers of gastric cancer using proteomics. *Proteomics Clin Appl* 2008;2:219–228.
- Schaub NP, Jones KJ, Nyalwidhe JO, et al. Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. *J Am Coll Surg* 2009;1–9.
- Schwamborn K, Krieg R'C, Grosse J, et al. Serum proteomic profiling in patients with bladder cancer. *Eur Urol* 2009;56:989–997.
- Guo RY, Pan CQ, Shen JM, et al. New serum biomarkers for detection of esophageal carcinoma using Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Cancer Res Clin Oncol* 2011;137:513–519.
- Yang Y, Zhao S, Fan YX, et al. Detection and identification of potential biomarkers of non-small cell lung cancer. *Technol Cancer Res Treat* 2009;6:455–465.
- Du J, Yang SY, Lin XL, et al. Use of anchochip-time-of-flight spectrometry technology to screen tumor biomarker proteins in serum for small cell lung cancer. *Diagn Pathol* 2010;5:60.