[Frontiers in Bioscience E3, 1-10, January 1, 2011]

Biomarker discovery and identification from non-small cell lung cancer sera

Jie Du[1], Shuan-ying Yang[1], Xiu-li Lin[1], Wen-li Shang[1], Wei Zhang[1], Shu-fen Huo[1], Li-na Bu[2], Wei Zhang[3], Bin Zhou[3], Yan-dong Nan[1], Hua-dong Zheng[1], Yan-feng Liu[1]

[1]Department of Respiratory Medicine, Second Affiliated Hospital of Medical School, Xi'an Jiaotong University, Xi'an, China, [2]Second Hospital of Xi'an, Xi'an, China, [3]Department of General Thoracic Surgery, Second Affiliated Hospital of Medical School, Xi'an Jiaotong University, Xi'an, China

TABLE OF CONTENTS

1. ABSTRACT

Currently, serum biomarkers might usually be thought not to be used for early detection of lung cancer by some researchers. In this study, we used a highly optimized ClinProt-matrix-assisted laser desorption/ionization time-of flight mass spectrometer (MALDI-TOF-MS) to screen non-small cell lung carcinoma (NSCLC) markers in serum. A training set of spectra derived from 45 NSCLC patients, 24 patients with benign lung diseases (BLDs) and 21 healthy individuals, was used to develop a proteomic pattern that discriminated cancer from non-cancer effectively. A test set, including 74 cases (29 NSCLC patients and 45 controls), was used to validate this pattern. After cross-validation, the classifier showed sensitivity and specificity, 86.20% and 80.00%, respectively. Remarkably, 100% of early stage serum samples could be correctly classified as lung cancer. Furthermore, the differential peptides of 1865Da and 4209Da were identified as element of component 3 and eukaryotic peptide chain release factor GTP-binding subunit ERF, respectively. The patterns we described and peptides we identified may have clinical utility as surrogate markers for detection and classification of NSCLC.

2. INTRODUCTION

Recent scientific progress of sequencing of the genome (1) and new approaches to modeling complex biological systems (2), may ultimately lead to improved anticancer therapy. However, prior to these findings, the best anticancer strategies still rely on early detection followed by close monitoring for early relapse so that therapies can be appropriately adjusted (3). Proteomic expression profiles generated from sera have been suggested as potential tools for the early diagnosis of cancer and other diseases. Appropriate biomarker based screens should be minimally invasive and reproducible. A simple blood test that detects molecules specific to a tumor would be ideal. In addition, screening technology must be sufficiently sensitive to detect early cancers and specific enough to classify individuals without cancer as being free of disease.

Up to now, lung cancer has been the leading cause of malignancy-related deaths in China, and the five-year survival rate of patients is about 15.2%, despite diagnostic imaging and therapeutic improvements over the past decade (4). Some tumor markers, including CEA, p53, Cyfra21-1, and CA19-9, have been investigated and are

commonly used as lung cancer biomarkers (5). However, few biomarkers have been accepted as indicators of clinical diagnosis, progression and/or prognosis, either because of lacking specificity or because of conflicting reports. Therefore, the discovery of specific or novel biomarkers of lung cancer is urgently needed. Recently, new strategies that facilitate proteomic analysis by magnetic beads dramatically simplifying the preanalytical sample separation and coupling with MS have been introduced for biomarker discovery research. One such MALDI profiling has been successfully used to differentiate breast (6), and bladder cancer from controls. Similar studies of lung cancer have not been reported yet. Although additional research may identify a single highly predictive marker of NSCLC, a comprehensive list of potential tumor markers will greatly facilitate the development of an assay with high predictive value.

In this study, we compared serum proteomic profiles between untreated NSCLC patients and the non-cancer population to discover NSCLC-specific protein markers and establish the pattern for diagnosis of NSCLC. The patterns we described may have clinical utility as surrogate markers for detection and classification of NSCLC, and meanwhile, two peptides (1865Da and 4209Da) were identified as element of component 3 and eukaryotic peptide chain release factor GTP-binding subunit ERF, respectively. These findings may have important implications for future peptide biomarker discovery efforts.

3. MATERIALS AND METHODS

3.1. Patients Serum Samples

Serum samples from each group were collected simultaneously dependent on the patient's availability in the Department of Respiratory Medicine, Second Affiliated Hospital of Medical School of Xi'an Jiaotong University between September 2003 and May 2008. The samples of NSCLC patients were characterized according to the International Union Against Cancer (UICC, 2002) staging system of lung cancer. Diagnoses were pathologically confirmed, and specimens were obtained before treatment. The control group was collected from 46 benign lung diseases (BLDs) and 44 healthy individuals. NSCLC patients and controls were further divided into profiling (training set) and validation (test set) groups.

3.2. Sample preparation and treatment with magnetic beads

All serum samples were collected early in the morning before breakfast. The sera were left at 4~6 degrees C for 2h, centrifuged at 10000 rpm, 4 degrees C for 10 min, and then aliquoted, and stored at -80 degrees C. 10μL MB-WCX binding solution and 5μL serum sample were added to the beads and mixed completely, which was then placed on the magnetic beads separation device (MPC-auto96, Dynal, Oslo, Norway). The beads were pulled to the side by magnetic force, and the supernatant was removed and discarded. The magnetic beads were washed three times with MB-WCX washing solution by shaking the beads up and down as needed. The supernatant was removed and the beads remained in place. 5μL elution solvent was added to the bead pellet and mixed by pipeting up/down, and the beads were pulled to the side and a fraction of the elution was transferred to another tube. 10μL a-cyano-4-hydroxycinnamic acid (0.3 g/l in ethanol: acetone 2:1) was added to the 1μL elution in a 348-well microliter plate and mixed carefully. A 1μL mixture was spotted in quadruplicate on a MALDI AnchorChipTM (Bruker Daltonics, Bremen, Germany).

3.3. Samples analyzed by MALDI-TOF

Samples were assayed randomly and blinded to the operator. The protein fingerprint data were analyzed by FlexAnalysis 3.0 (Bruker Daltonics, Bremen, Germany). Profile spectra were acquired from an average of 400 laser shots. ClinProt software version 2.0 was employed to subtract baseline, normalize spectra (using total ion current), and determine protein peaks with the m/z value and peak intensities in the mass range (500-10000Da). The peaks were filtered to maintain an S/N of more than 3. To establish new diagnostic models for NSCLC, A genetic algorithm (GA) contained in this software suite was used to identify statistically significantly differential protein peaks in the analyzed training set. The peaks input to the model with highest accuracy were selected as the set of potential biomarkers.

3.4. Differential peptides identification

Peptide extracts were dried and resuspended in 15–20 mL 5% formic acid for further MS/MS analysis by an LTQ Orbitrap mass spectrometer (ThermoFisher, USA) and typically 5μL of peptide extracts were actually injected for analysis. Analysis steps by LTQ-Orbitrap were as follows: The peptide extracts were loaded at 15 $mL.min^{-1}$ for 6 min on a nanoAcquity™Column, followed by eluting and separating on a nanoAcquity™UPLC™Column, using 90-min gradients with 95% water, 5% acetonitrile (ACN), 0.1% formic acid (solvent A); and 95% ACN, 5% water, 0.1% formic acid (solvent B) at a flow rate of 300 nL/min. The samples were run in data-dependent mode, where each full MS scanning was followed by three consecutive MS/MS scans of the 3 most abundant peptide molecular ions (typically doubly and triply charged ions), which were selected consecutively for CID. The MS survey scans (300-2,000Da) were carried out and the acquisition cycle consisted of a survey scan at the highest resolving power (100,000). Dynamic exclusion was used with a series of parameters and the acquired MS/MS data were processed using BioworksBrowser 3.3.1. A sequence database search was performed with the International Protein Index (IPI Human3.45) (4).

3.5. Statistical Analysis

We imported these data to Clinprotools software and employed several statistical ways, including ANOVA, clustering, KNN, Algorithm and genetic algorithm to get the differential proteins. We used the Wilcoxon non-parametric test to analyze between the two groups, the Kruskal-Wallis test among the three groups to find differential proteins, and the Benjamini-Hochberg method for correction of P values, with a threshold of $P < 0.05$. Then, we used Matlab to apply 164*95 unsupervised

Table 1. Characteristics of serum samples

Characteristics	No. of samples
Lung cancer group	68
Male	46
Female	22
Mean age in years (range)	58 (31 to78)
Disease stage	
I/II	20
III/IV	48
Tumor Histology	
Squamous cell carcinoma	34
Adenocarcinoma	34
Benign lung disease	46
Male	28
Female	18
Mean age in years (range)	60 (31 to76)
penumonia	21
pulmonary tuberculosis	25
Healthy individuals	44
Male	26
Female	18
Mean age in years (range)	59 (44 to76)

Table 2. The M/Z and CV of the reference selected peaks between samples

M/Z (Da)	CV (%)	M/Z (Da)	CV (%)
2461	0.239559	4126	0.241348
2981	0.245415	5019	0.187579
3066	0.176466	5571	0.287579
3363	0.158828	5738	0.304054
3613	0.162182	6646	0.174836
4061	0.261492	7148	0.247252

Table 3. Specificity and sensitivity for each model

	NNet	Tree	KNN	GA+KNN
Accuracy	83/90	80/90	83/90	86/90
Sensitivity	41/45	42/45	40/45	43/45
Specificity	42/45	38/45	43/45	43/45
cross validation (k-fold) mean±SD	80.0%±0.11	85.6%±0.129	78.9%±0.097	86.67%±0.11

clustering, with the horizontal labeled as samples and vertical as protein peaks. We used genetic algorithms (GA) of Clinprotools software combined with a *k* neighbors (KNN) classification algorithm for the establishment of Category Forecast. We set the K value as 3 in the KNN.

4. RESULTS

4.1. System reproducibility

The mass accuracy was achieved by external calibration. The reproducibility of mass spectrum generation was determined with respect to the relative peak intensities. Visual comparison of 13 reference samples/spectra was described. The CV of the selected peaks' mass was always less than 30% and did not differ statistically between the different sample and laser settings (Table 2).

4.2. Unsupervised analysis of peptide ion signals from MS-based serum profiling differentiates NSCLC from controls.

We analyzed the serum peptide profiles of 68 patients with NSCLC, as well as 90 control sera from healthy volunteers and BLDs using new high-resolution MALDI-TOF MS coupled with bead fractionation. Samples from patients with NSCLC and from control individuals were then randomly distributed during processing and analysis. A total of 100 distinct m/z values were resolved in the 800–10000Da range. A spreadsheet (peak list) containing the normalized intensities (i.e., signal intensities, after baseline subtraction, were divided by the total ion current of the corresponding spectrum and multiplied by a scaling factor of 10^7) of all 100 peaks for each of the 164 samples was then taken for unsupervised, average-linkage hierarchical clustering using a standard correlation (Figure 1). This resulted in clear, distinct patterns that differentiate disease from control in binary comparisons. 20 peptides (4209, 4054, 4172, 4072, 1865, 1778, 4090, 1691, 2104, 3954, 3935, 4963, 4123, 4267, 6632 only 15 here) were detected out, which were different in expression of the sera of NSCLC and health control. They also have significant confidence ($P<10^{-6}$) in the identity of the above proteins by a Wilcoxon validation, and they were shown in Figure 2. Within them, the top two discriminating peaks of 4209Da and 4054Da got an accuracy of 73.9% (34/46), 69.2% (27/39) and 81% (64/79) respectively, when using an algorithm method of KNN. Therefore, they could separate cancer sera from normal and benign groups (Figure 3A).

As can be seen, the classification was achieved primarily through a contrast in peak intensities. This can be seen from the scatter plot shown in Figure3A. Both of the peaks intensities decreased in the sera of NSCLC compared to the health and benign controls. Using the 4209Da mass peak as an example, illustrates the analysis of proteins in the molecular weight range of 4200–4220Da with the ClinProt. As illustrated by Fig.2, there is an increase in the quantity of proteins in the range of 4200–4220Da in the serum among controls when compared with cancer patients, especially at the molecular weight of approximately 4209Da. To analyze the actual discriminative power of the two markers, we produced an ROC-curve visualizing the performance of the two-class classifier in Figure 3B; the AUC of the classifier was 0.911 and 0.889, respectively.

4.3. Feature selection and model establishment

In this approach, we randomly selected 90 samples (including 45 cancers, 21 normal and 24 benign sera) as the training set and the remaining samples as the test set. In the first test, we applied the KNN (*k*-nearest neighbor) method, single-hidden-layer neural network (NNet) and Decision Tree algorithm with all detected peaks, and each model got a specificity and sensitivity (Table 3, Figure 4A). In order to evaluate the generalization ability of the classifier, the method of cross validation named *k*-fold was used. In the method, the data set was divided into *k* equal parts of data points. Then *k* models were generated where each time a different one of the *k* parts was omitted. The omitted part was classified against the model calculated from the remaining *k*-1 parts. The procedure was repeated for *k* times and the average classification result was returned as the prediction capability. KNN had shown better classified results than NNet and Tree, but the cross-validation rate was less than the two algorithms. Secondly, all detected peaks were used

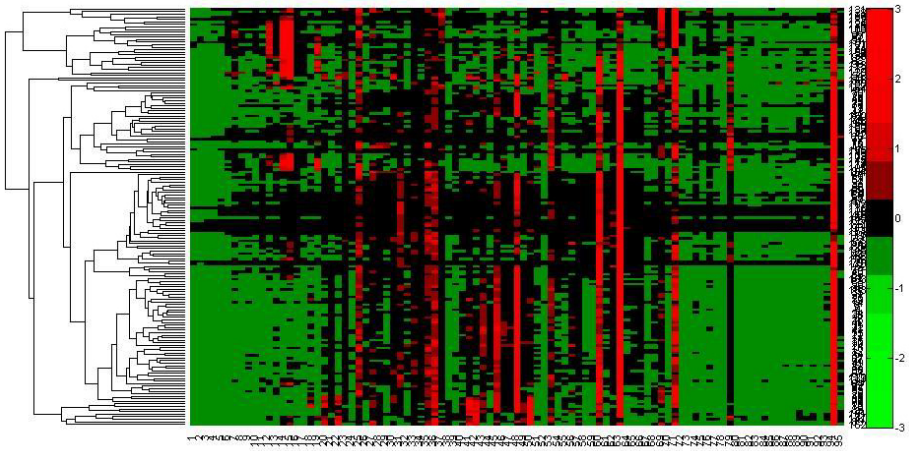

Figure 1. Un-supervised hierarchical clustering analysis of MS-based serum peptide profiling data derived from 2 groups of cancer patients and healthy controls. Serum samples from healthy volunteers and patients with NSCLC were prepared following the standard protocol. The 2 groups were randomized before automated solid-phase peptide extraction and MALDI-TOF MS. Spectra were processed and aligned using the Qcealign script.

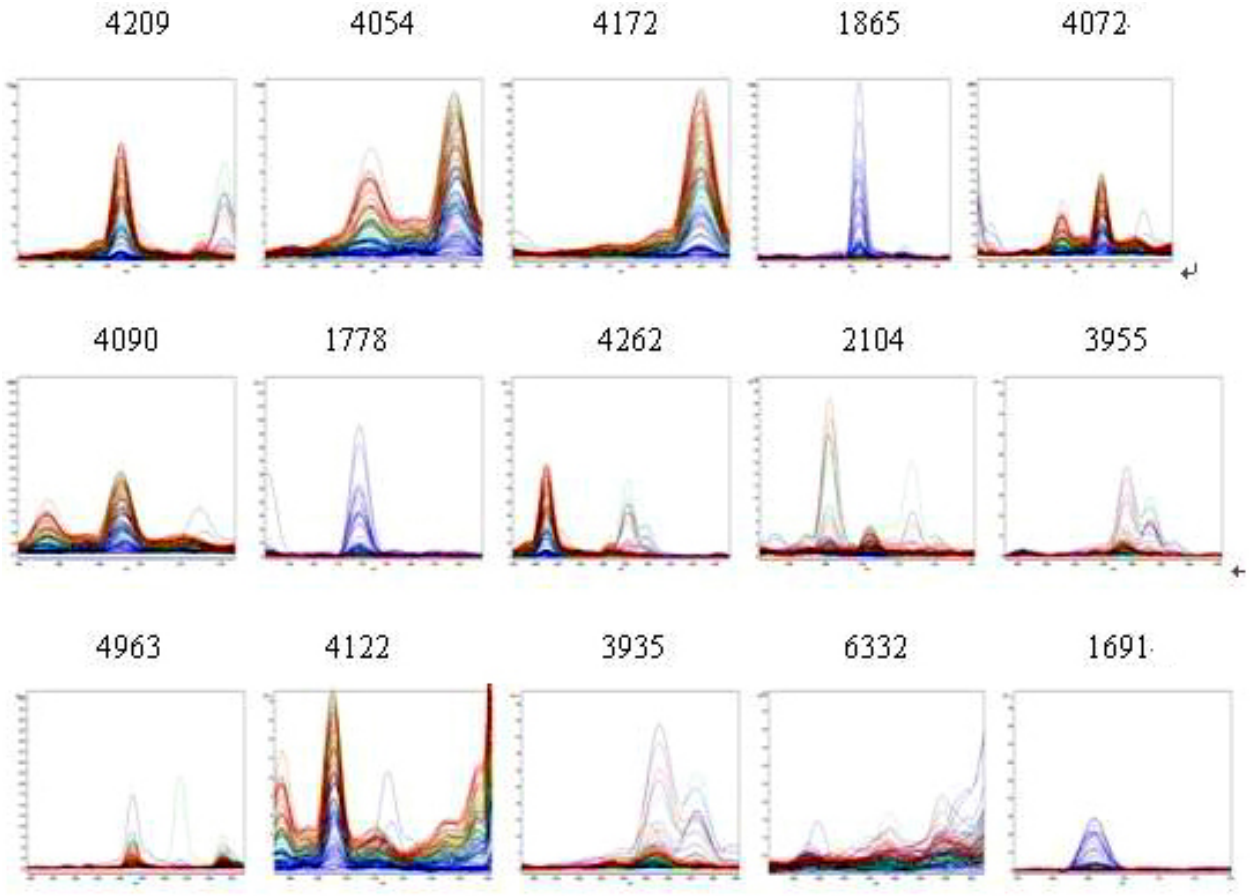

Figure 2. MALDI-TOF mass spectral overlays of selected peaks derived from serum peptide profiling of 3 groups of cancer patients and healthy controls. Spectra were obtained, aligned, and normalized as described in Methods and were displayed using the mass spectra viewer. The 15 overlays are arrayed so that an identical mass range window is shown for each of the 3 binary comparisons in which spectral intensities have been normalized and scaled to the same size. The monoisotopic mass (m/z).

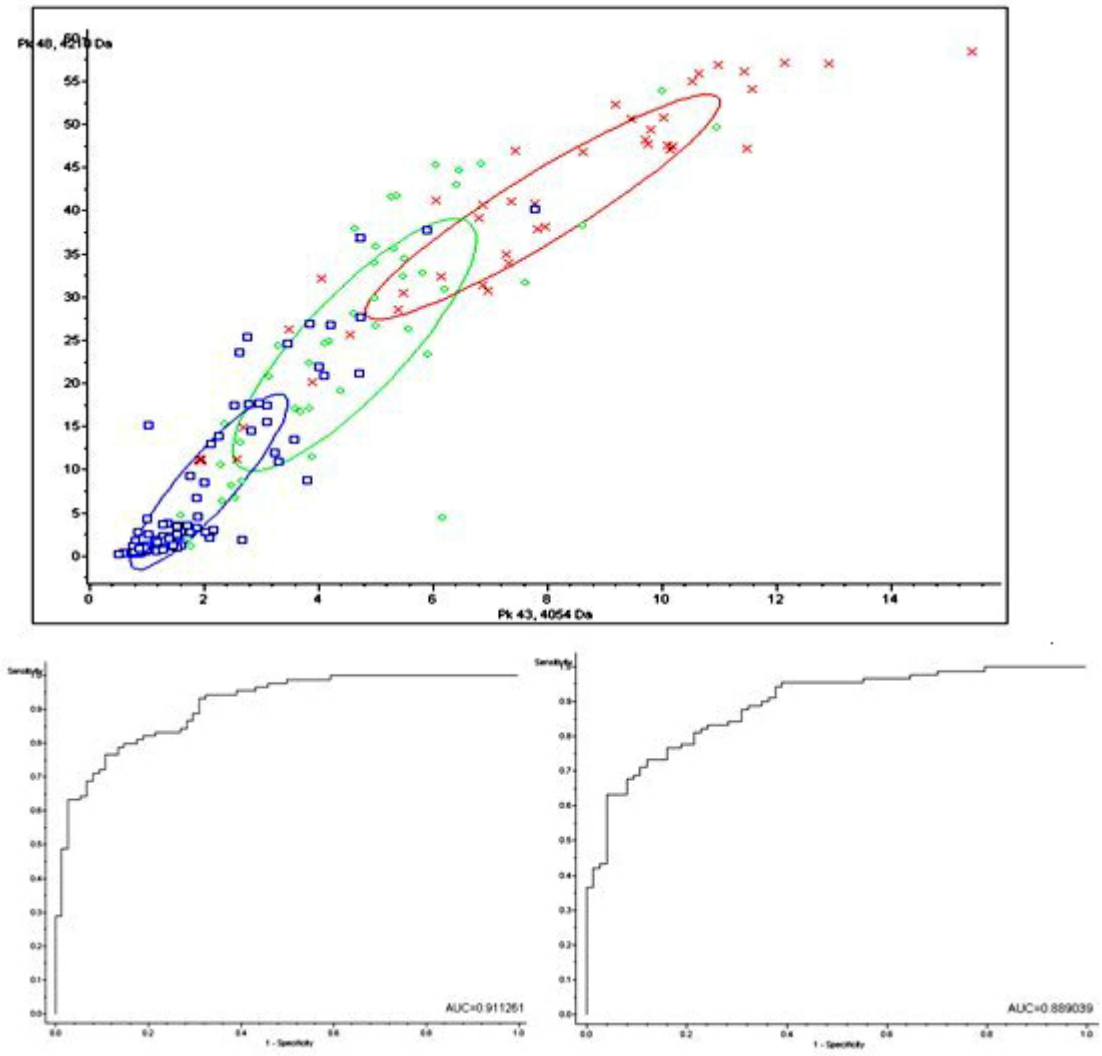

Figure 3. The character of the two differential peaks, 4209 Da and 4054 Da, in cancer and other control sera (A) Scatter plots of the top two peaks on the basis of which the classification patient-control group was made The x-axis represented 4054Da and the y-axis represented 4209Da in order to observe their ability of samples distribution (□ for normal control, for benign control and × for cancer) (B) The ROC of peak 4054 (left) and 4209Da (right), with an AUC value of 0.911 and 0.889, respectively.

with a KNN genetic algorithm to generate classification model. The genetic algorithm can search for peaks from the peak space to get a highly accurate classification rate automatically. A model consisting of 7 peptides (1865, 2951, 2932, 4090, 4643, 3262 and 3191Da) was generated. Classification rate of the optimal model was 95.56% (control) and 95.56% (NSCLC) on the training data and was 80.00% (controls) and 86.20% (NSCLC) on the test set. Specifically, the cross validation rate of the method was 86.67% (Table 3). Other algorithms also had a result on the dataset with the 7 peptides (Table 4, Figure 4B).

4.4. Identification of peaks of proteins by MS/MS

As we know, identification of a biomarker which is closely related with a given cancer is not only helpful for investigation of carcinogenesis and prognosis, but also useful for therapy of cancer and drug development. In addition, if the markers could be identified and specific high affinity antibodies would be generated to them, then more direct and potentially less expensive methods for analysis could be developed. So, we employed LTQ-Orbitrap analysis to identify 3 out of the above 15 peptides, which were differently expressed between NSCLC and health control groups. MS/MS analysis of 1865Da detected most b and y ions (Figure 5A). 2 ions (peptides 1778 and 1865Da) detected from a sample of No.41 serum, were counted as one identification because they were described by the similar set of peptides. Both peptides gave consistently higher MS ion signals in the cancer patient sera than in the matched controls. These two peptides were SKITHRIHWESASLL and SSKITHRIHWESASLL, corresponded to complement C3 or complement C3 precursor. Peptide 4209Da was analyzed by MS/MS and identified here for the first time from No.204 expressing serum. Its sequence was identified as EQSDFCPWYTGLPFIPYLDNLPNFNRSIDGPIRLPI

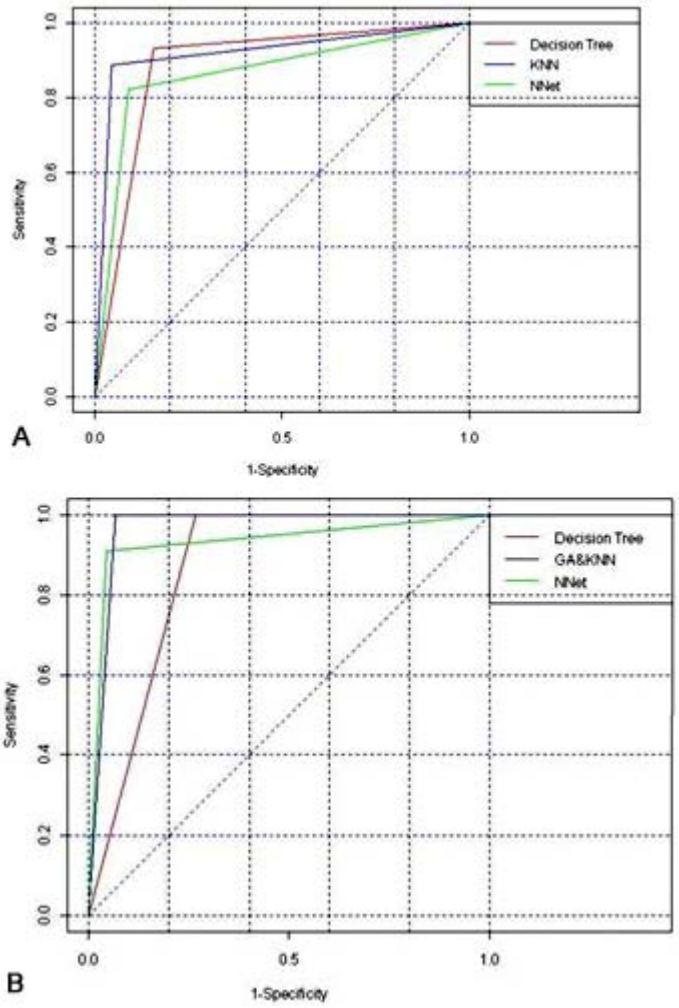

Figure 4. A. ROC-curve for the three algorithms of Decision Tree, KNN and NNet. The true positive recognition rate (sensitivity) is demonstrated on the y-axis against the false negative recognition rate (1-sepecificity) on the x-axis. ROC_KNN=0.922, ROC_Tree=0.889, ROC_NNet=0.867; 4B. ROC-curve for the three algorithms of Decision Tree, GA&KNN and NNet on the dataset with 7 peptides. The true positive recognition rate (sensitivity) is demonstrated on the y-axis against the false negative recognition rate (1-sepecificity) on the x-axis. ROC_KNN=0.967, ROC_NNet=0.933, ROC_Tree=0.867.

which was unique to a eukaryotic peptide chain release factor GTP-binding subunit ERF3b (Figure 5B).

5. DISCUSSION

Lung cancer accounts for the most cancer deaths in all malignancy tumors. As lung cancer is a heterogeneous, multistage and multifactor disease, combinations of markers that cover a broad clinical phenotype for enhancing diagnostic value and therapeutical applications are necessary. Due to the heterogeneity of lung cancer and the lack of sensitivity and specificity of individual markers, the current positive rate is as low as 14% for lung cancer early detection, so there is a growing consensus that panels of markers can improve screening, diagnosis, prognosis, or monitoring responses to therapy (8). Thus, we predict that a combination of tumor-expressed and host response proteins can be used to develop a profile of cancer for clinical screening. So in this study, we first tried to construct a combination model of biomarkers for discriminating lung cancer from a control, and then we used LTQ-Orbitrap to identify some of the markers. Our further work will try to explore multiplex immunoassay for this markers' clinical application.

Human biological fluids contain various types of markers and these biomarkers are useful for diagnosis, prognosis and drug development for many diseases. Blood is a human fluid that can be easily obtained clinically. The easily accessible serum proteome consists of only a lot number of abundant proteins such as albumin and immunoglobulin, which complicates the detection of the many low-abundance proteins. Many methods were employed to remove the abundant proteins such as albumin, a-amylase, and antibody fragments from serum to further reveal low-abundance proteins, which may take

Table 4. The results on the dataset with the 7 peptides

	Nnet	Tree
Accuracy	86/90	78/90
Sensitivity	42/45	45/45
Specificity	44/45	33/45
cross validation(k-fold) mean±SD	83.3%±0.08	83.3%±0.14

many useful low-abundance proteins and low-mass peptides away as a consequence, however. Furthermore, even if one possesses these proteins, traditional techniques such as 2-DE couldn't separate low-abundance proteins and low-mass peptides effectively.

Here, we describe a new technology platform ClinProt (9) for the simultaneous measurement of large numbers of serum polypeptides by using magnetic bead-based, sample processing, and MALDI-TOF MS. The technology is automated on a liquid handling robot for high throughput and reproducibility. It was reported that the use of porous particles for sample pretreatment is more sensitive than surface capture on chips because spherical particles have larger combined surface areas than small-diameter spots. The application of magnetic bead fractionation in combination with MALDI-TOF is appropriate for the detection of low concentrations of proteins and peptides in serum (10). High sensitivity and resolution allow detection, within a molecular mass range of 800-15,000Da, of over 400 polypeptides in a single droplet of serum (11). Sven et al (12) found that protein profiling by MALDI-TOF MS after proteome fractionation with magnetic beads is a robust, precise, and rapid method for the investigation of such complex samples as blood.

The high sensitivity and reproducibility of the novel platform for proteome profiling in our study is supported by the recent studies. Villanueva et al (6) stated that a limited subset of serum peptides (a signature) provides accurate class discrimination between patients with 3 types of solid tumors and controls without cancer. There are 14 potential biomarkers for prostate cancer, 14 for breast cancer, and 58 for bladder cancer. They obtained a sensitivity of 100% (41/41) by class prediction of a prostate cancer validation set using a support vector machine (SVM). Freed et al (13) used a MALDI-TOF bead-based analysis in the study of 27 normal, 25 healthy smoker, and 24 head and neck squamous cell carcinoma (HNSCC) (pretreatment and matched posttreatment) serum samples. All detected peaks were used with a k-nearest neighbor genetic algorithm in ClinProt and their classification rates were 89% normal, 93% benign, and 98% for HNSCC.

Circulating tumor markers could impact cancer patient outcomes, resulting in improved screening, diagnosis, staging, and management. Until now, we have not seen any similar study of lung cancer with the sera peptides of NSCLC. We analyzed serum samples from 74 NSCLC and 90 non-cancer controls using ClinProt-MALDI-TOF-MS. A panel of serum markers consisting of seven peptides (1865, 2951, 2932, 4090, 4643, 3262 and 3191Da) was established by applying a genetic algorithm. For the cross-validation with test set, the classifier had a sensitivity of 86.20%□25/29□, a specificity of 80.00 % (36/45), positive predictive value of 73.53%, and negative predictive value of 90%. The mistaken classification was possibly related to amalgamation with other diseases, such as hypertension, bronchitis, and ileus, which influenced the serum proteins. Moreover, all the patients of the early stage were accurately diagnosed, which means our study may be great potential ability for early detection of NSCLC at the clinical level. In this study, we adopted LTQ Orbitrap high-resolution mass spectrometer for protein identification. The LTQ Orbitrap instrument can operate at a mass resolution of up to 100,000 (13), which is useful for maximizing the number of metabolites that can be separated. In addition, three confirmed biomarkers were identified. Both peptide ions (1778 and 1865Da) that were of higher intensity in NSCLC derived from complement C3 or complement C3 precursor, were indistinguishable by MS/MS. Identification of the 4209Da ion confirmed its identity as a eukaryotic peptide chain release factor GTP-binding subunit ERF3b.

It has been reported that complement-independent pathways are critical for leukocyte recruitment into neoplastic tissue and leukocyte-mediated potentiation of tumorigenesis. There are 25 known soluble complement proteins. Among these complement proteins, C3 is the most versatile and multifunctional molecule identified to date, having evolved structural features that allow it to interact in a specific manner with at least 25 different proteins (14). C3 is also the most abundant complement protein in serum (1.2 g/L) and supports the activation of all 3 pathways of complement activation, the classic, alternative, and lectin pathways (15). The functional significance of C3 is as a regulator of inflammatory cell infiltration and activation during malignant progression. De Visser et al (16) found that complement-independent pathways are critical for leukocyte recruitment into neoplastic tissue and leukocyte-mediated potentiation of tumorigenesis. Similar complement C3f peptides and kininogen peptides have also been reported in other cancer serum samples (6). With the use of ClinProt-MALDI-TOF-MS analysis, both Villanueva and Freed et al. identified the complement C3f (Da) in the 4 different types of solid tumors studies. This showed that the complement C3f (1865m/z) would likely become a potential diagnostic biomarker for malignancy. The precursor of the third component of human complement C3 (pro-C3), is a disulfide-linked two-chain protein. The pro-C3 is converted by limited proteolysis to C3. The relationship between pro-C3 and C3 was established with the use of HepG2, a cell line derived from a human hepatocellular carcinoma, which synthesizes at least 17 plasma proteins. Since complement C3f is unlikely to be released by tumor cells, the up-regulated complement C3f in NSCLC patients may be due to dysfunction of relative tissue, and such acute-phase reactants may still represent only epiphenomena because of the presence of the tumor and may not be specific to a particular type of cancer (17). Translation termination is regulated by a heterodimeric release factor consisting of eRF1 and eRF3 (18-19). The eRF3 stimulates the termination reaction in a GTP-dependent manner and is essential for cell growth. It has been reported that eRF3 has GTP-binding and hydrolysis activities, which is stimulated by GTPase-

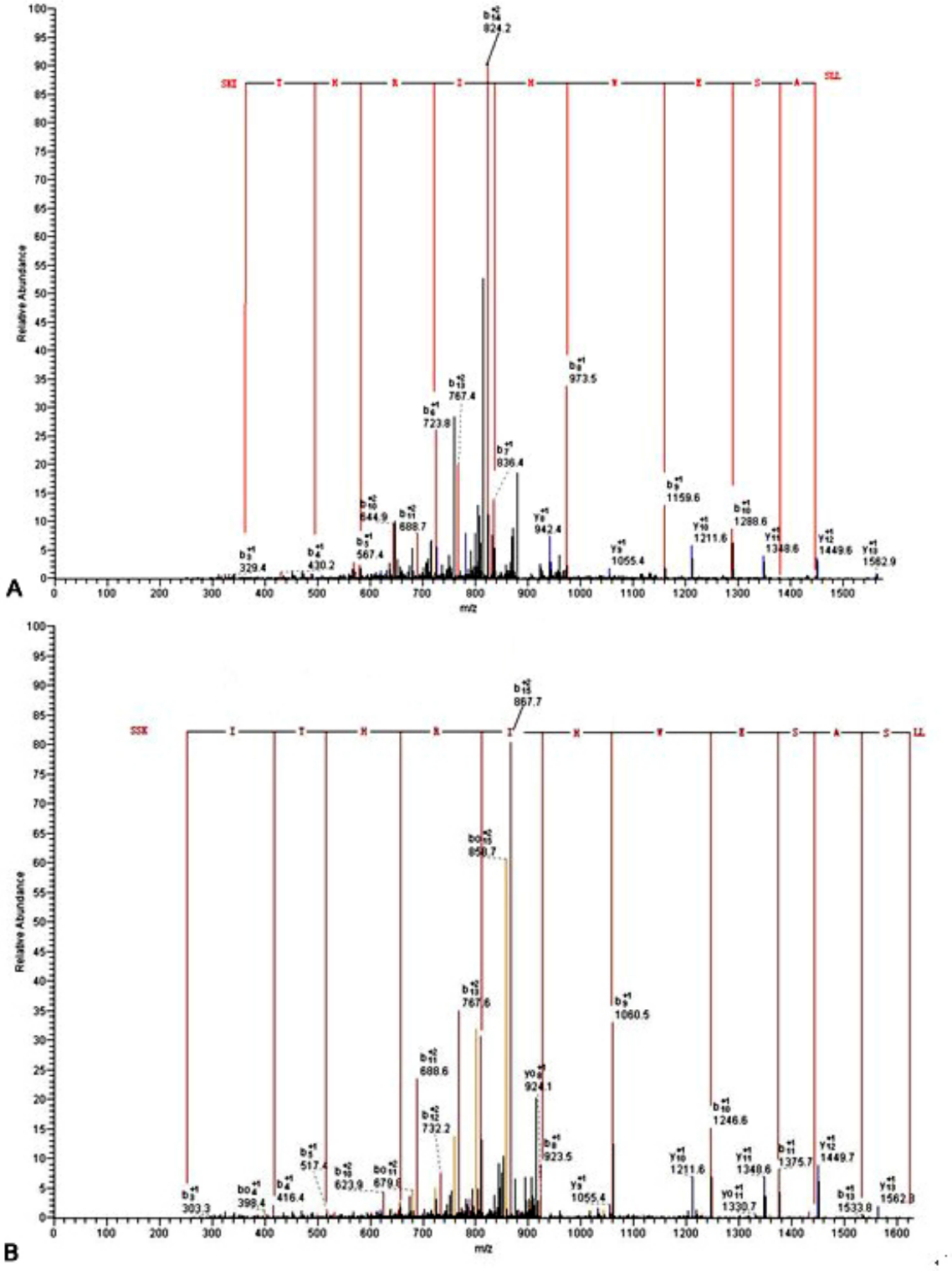

Figure 5. MS/MS identification of serum peptides as fragment of complement C3 (A) and eukaryotic peptide chain release factor GTP-binding subunit ERF (B). The fragment ion spectrum shown here was taken for a MS/MS ion search of the International Protein Index (IPI Human3.45) database (http://www.ebi.ac.uk/IPI/IPIhelp.html). b and y fragment ion series are indicated together with the limited sequences.

activating proteins consisting of eRF1 and ribosome (20). The GTP-binding protein eRF3 interacts with eRF1, which directly recognizes the termination codons to perform translation termination (21). A human protein of 628 amino acids was identified, named eRF3b, which is highly homologous to the known human eRF3 henceforth named eRF3a. The most important difference in the nucleotide sequence is that eRF3b lacks a GGC repeat close to the initiation codon in eRF3a. Jakobsen et al (22) found that the protein is active in vitro as a potent stimulator of the release factor activity of human eRFl. Like eRF3a, eRF3b exhibits GTPase activity, which is ribosome-and eRFl-dependent. In vivo assays show that the human eRF3b is able to enhance the release factor activity of endogenous and overexpressed eRFl with all three stop codons.

The current study shows that the selection of a combination of multiple proteins resolved by MALDI may become a potential diagnostic approach. We are encouraged by the fact that the combined model successfully classified all 6 stage one/two NSCLCs from the test set, which suggest that our approach is useful for early NSCLC detection. However, these results are limited by a very few stage one/two cases in our study. Further research is needed to confirm our current findings by using larger amounts of study samples.

6. ACKNOWLEDGEMENT

This work was supported by grants from the National Natural Science Foundation of China (No. 30570795) and Program for New Century Excellent Talents in University (No. NCET06-0845) and the Key Program in Science and Technology of Shaanxi Province Shaanxi [No. 2007K09-01(3)].

7. REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. International Human Genome Sequencing Consortium.. Initial sequencing and analysis of the human genome. *Nature,* 409:860–921 (2001)

2. Benedict Anchang, Mohammad J. Sadeh, Juby Jacob, Achim Tresch, Marcel O. Vlad, Peter J. Oefner, Rainer Spang. Special Feature: Complex Systems: From Chemistry to Systems Biology Special Feature: Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *PNAS*, 106: 6447-52 (2009)

3. Qiu F, Liu HY, Zhang XJ, Tian YP. Optimization of magnetic beads for MALDI-TOF MS analysis. *Frontiers in Bioscience,* 14: 3712-23 (2009)

4. Chen GA, Wang XJ, Yu JJ, Sooryanarayana Varambally, Yu JD, Dafydd G. Thomas, Lin MY, Prakash Vishnu, Wang ZW, Wang R, Jeff Fielhauer, Debashis Ghosh, Thomas J. Giordano, Donald Giacherio, Andrew C. Chang, Mark B. Orringer, Talal El-Hefnawy, William L. Bigbee, David G. Beer, Arul M. Chinnaiyan. Autoantibody profiles reveal ubiquilin 1 as a humoral immune response target in lung adenocarcinoma. *Cancer Res,* 67: 3461-7 (2007)

5. I. Horváth, Z. Lázár, N. Gyulai, M. Kollai, G. Losonczy. Exhaled biomarkers in lung cancer. *Eur Respir J,* 34: 261-75 (2009)

6. Lynn S. Adams, Shiuan Chen. Phytochemicals For Breast Cancer Prevention By Targeting Aromatase. *Frontiers in Bioscience*, 14: 3846-63 (2009)

7. Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, Paul Tempst. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest*, 116: 271–84 (2006)

8. Yang SY, Zhang WG, Sun XZ, He JY, Li YL, Liu Y, Zhang J, Dong XL, Yang DC. A preliminary evaluation of diagnostic value of five serum tumor markers for lung cancer. *Journal of Xian Jiaotong University (Medical Sciences)*, 26(4): 352-5 (2005)

9. Ketterlinus, R., Hsieh, SY., Teng, SH., Lee, H., Pusch, W. Fishing for biomarkers: analyzing mass spectrometry data with the new ClinProTools software. *Biotechniques,* Suppl: 37-40 (2005)

10. Tirumalai, RS., Chan, KC., Prieto, DRA., Issaq HJ. Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics*, 2: 1096-103 (2003)

11. Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P.. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem,* 76: 1560-70 (2004)

12. Sven Baumann, Uta Ceglarek, Georg Martin Fiedler, Jan Lembcke, Alexander Leichtle, Joachim Thiery. Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem*, 51: 973-80 (2005)

13. Dunn, WB. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys Biol,* 5: 11001 (2008)

14. Sahu, A., Sunyer, JO., Moore, WT., Sarrias, MR. Soulika, AM., Lambris, JD.. Structure, functions, and evolution of the third complement component and viral molecular mimicry. *Immunol Res*, 17: 109-21 (1998)

15. Bohana-Kashtan, O., Ziporen, L., Donin, N., Kraus, S., Fishelson, Z. Cell signals transduced by complement. *Mol Immunol,* 41: 583-97 (2004)

16. de Visser, KE., Korets, LV., Coussens, LM. Early neoplastic progression is complement independent. *Neoplasia*, 6: 768-76 (2004)

17. Freed, GL., Cazares, LH., Fichandler, CE., Fuller, TW., Sawyer, CA., Stack BC., Schraff S., Semmes OJ., Wadsworth JT., Drake RR.. Differential capture of serum proteins for expression profiling and biomarker discovery in pre- and posttreatment head and neck cancer samples. *Laryngoscope*, 118: 61-8 (2008)

18. Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R. Ingevechtomov S. Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J*, 14: 4065-72 (1995)

19. Stansfield, I., Jones, KM., Kushnirov, VV., Dagkesamanskaya, AR., Poznyakovski. The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in Saccharomyces cerevisiae. *EMBO J*, 14: 4365-73 (1995)

20. Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., Kisselev, L.. Eukaryotic polypeptide chain release factor eRF3 is an eRF1-and ribosome-dependent guanosine triphosphatase. *RNA*, 2: 334-41 (1996)

21. Kobayashi, T., Funakoshi, Y., Hoshino, SI., Katada, T.. The GTP-binding Release Factor eRF3 as a Key Mediator Coupling Translation Termination to mRNA Decay. *J Biol Chem,* 279: 45693-700 (2004)

22. Jakobsen, CG., Segaard, TM., Jean-Jean, O., Frolova, L., Justesen, J.. Identification of a novel termination release factor eRF3b expressing the eRF3 activity in vitro and in vivo. *Mol Biol (Mosk)*, 35: 672-81 (2001)

Key Words: Biomarker, Proteomics, ClinProt, NSCLC, Sera

Send correspondence to: Shuang-ying Yang, Department of Respiratory Medicine, Second Affiliated Hospital of Medical School, Xi'an Jiaotong University, Xi'an, China. Tel: 86-13991392919, Fax: 86-02987678599, E-mail: yangshuanying66@163.com

http://www.bioscience.org/current/volE3.htm