Original Article

Protein Profiles Correlated with Recurrence of Early Stage Non-Small Cell Lung Cancer

Zhe-ming Lu[1,2], Shao-lin Hong[2*]

[1]*Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Etiology, Peking University School of Oncology, Beijing Cancer Hospital & Institute, Beijing 100142, China*
[2]*Duke Comprehensive Cancer Center, Duke University Medical Center, Durham, NC 27710, USA*

CLC number: R734.2 Document code: A Article ID: 1000-9604(2010)01-0021-06
DOI: 10.1007/s11670-010-0021-z

ABSTRACT

Objective: To elucidate protein markers that differentiate stage I non-small cell lung cancer (NSCLC) that subsequently develop metastatic disease to those that do not develop metastasis by protein expression profiles.

Methods: Matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and two dimensional difference gel electrophoresis (2D-DIGE) platforms were used to separate proteins in whole tumor specimens. Quantitating mRNA expression was used for validation.

Results: Twelve proteins were identified as expressed differentially between two groups from protein expression platforms. But from gene expression platform no marker could distinguish patients with recurrent vs. nonrecurrent disease.

Conclusion: Analysis of multiple protein markers may be more informative to predict prognosis of early stage lung cancer.

Key words: Non-small cell lung cancer; Recurrence; Biomarker

INTRODUCTION

Lung cancer currently accounts for 14% of all deaths in the United States and 28% of all cancer deaths[1]. The overall survival rate is 15%, and only 20% of cases present at an early stage. Although the standard care for stage I NSCLC is surgical resection, post-resection survival is only approximately 50% at 5 years[2]. This suggests that at least 50% of all stage I patients have undetectable metastatic disease at the time of presentation. In addition to highlighting the prognostic inadequacy of our current method of pathologic staging, the above statistics underscore pathologic staging, the above statistics underscore the difficulty of optimizing treatment in patients with early stage disease. To address issues of diagnosis, prognosis, and therapeutic management of patients with lung cancer, investigators have put great effort into identifying potential biomarkers[3–6]. Although a number of potential markers of metastasis in NSCLC have been investigated, no clinically useful biomarker specific for metastasis has been identified[7].

Two dimensional difference gel electrophoresis (2D-DIGE) followed by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) platforms have been well documented to separate proteins in either whole tumor specimens or laser capture microdissected cells[6, 8]. We propose to use these platforms to compare protein expression profiles of primary lung tumor tissues resected from patients with stage I lung tumors who subsequently developed metastatic disease to those who do not

Received 2009–08–27; **Accepted** 2009–12–17
This work was supported by NIH grant (1RO1-CA-109384-01A).
*Corresponding author.
E-mail: slhong@duke.edu

develop metastasis, to elucidate protein markers that differentiate these two phenotypes.

MATERIALS AND METHODS

Patients and Tissue Collection

A total of 46 diagnosed primary NSCLC patients during the period between 2003 and 2005 were obtained from the gross specimen at the time of surgery at Duke University Medical Center. Two groups of lung cancer patients were selected: those with no recurrence after at least four years and those with recurrence within two years after primary therapy. All patients selected for entry into the study met the following criteria: (1) a histologically confirmed diagnosis of non-small cell lung cancer, (2) stage I pathology at the time of diagnosis, (3) no prior chemotherapy or radiation for lung cancer, and (4) complete surgical resection of the tumor. Tumors were diagnosed based on medical history, physical examination and computed tomography (CT) of the chest and upper abdomen. The Institutional Review Board obtained informed consent from patients before the collection of biospecimens for medical research. Tissue specimens were frozen in liquid nitrogen and stored at -80°C. Clinical information for all patients is shown in Table 1. All samples were examined by two pathologists, to assure the pathologic diagnosis and the integrity of the samples prior to lysate preparation. The same selected region of the tumor was used for protein and mRNA isolation.

MALDI-TOF MS Analysis

We selected 15 patients with no recurrence more than 48 months after resection and 15 matched patients with recurrence of their malignancy and metastatic disease confirmed within 24 months post resection. Lung cancer tissues were prepared for protein expression profiles by MALDI-TOF MS as previously described[3]. Briefly, the samples were prepared for analysis using a conventional dried-droplet protocol with matrix sinapinic acid. A total of about 500 ng lysate protein was deposited on the MALDI sample stage. For each crystallized sample spot, two MALDI-TOF mass spectra were collected. The raw data in each mass spectrum were archived using proteins of known mass as internal or external calibrants. Signal: noise ratios (S/N) were evaluated using the standard software provided by the instrument manufacturer. MALDI mass spectra were collected at several different positions from each spot with a mass window from 2,000 to 80,000 m/z. The m/z values of all the peaks were used to generate a composite peak list of all the ion signals from each Rotofor run. A T^2 test-based statistical pattern recognition was used to determine the most valuable parts of the spectra, organized into bins, in terms of their ability to separate the two sets of data (non-recurrent and recurrent).

2D DIGE and Mass Spectrometry

In addition to MALDI-TOF MS analysis on total protein samples, we pursued an alternative strategy of performing 2D DIGE to identify proteins. Altogether, 10 patients without recurrence and 10 patients with recurrence were selected. Each tissue sample (14 mg) was solubilized in a lysis buffer containing 7 mmol/L urea, 2 mmol/L thiourea, 4% (w/v) CHAPS, 30 mmol/L Tris-HCl, pH 8.5, and 1X protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). The supernatants were collected and impurities were removed using a 2-D Clean-Up Kit (GE Healthcare). A protein assay was performed on each sample using a 2-D Quant Kit (GE Healthcare).

Aliquots of protein from each sample were labeled individually with N-hydroxy succinimidyl ester derivatives of the cyanine dyes Cy3 or Cy5 (GE Healthcare). Dye labeling efficiency was normalized by assigning 5 lysates from patients with no recurrence and 5 lysates from patients with recurrence to Cy3, and the remaining lysates to Cy5. A pooled internal standard consisting of equal amounts of protein from the 20 lysates was labeled with Cy2. We assembled 10 pools of protein, each pool consisting of an equal quantity of tumor protein (125 μg) from a patient without recurrence, a patient with recurrence, and the pooled internal standard.

Labeled samples were subjected to isoelectric focusing (IEF) on a multiPhor II electrophoresis unit using 13-cm, pH 3–10 linear gradient IPG strips (GE Healthcare) according to the manufacturer's recommendations. After the first dimension separation, the IPG strips were equilibrated in 6 mmol/L urea, 2% (w/v) SDS, 0.375 mmol/L Tris-HCl, pH 8.8, and 20% (v/v) glycerol (equilibration buffer, EB) containing 20 mg DTT/ml for 10 min. This was followed by a 10 min incubation in EB containing 25 mg iodoacetamide/ml. Second dimension

electrophoresis was carried out on 12% polyacrylamide gels. The gels were imaged on a Typhoon 9400 Variable Mode Imager (GE Healthcare) at the 3 excitation and emission wavelengths specific for each of the CyDyes and then stained with Coomassie Blue to enable visualization of spots for picking. Intra-gel analysis was performed using DeCyder DIA (Difference In-gel Analysis) v6.0 software (GE Healthcare). Inter-gel matching and statistical analyses were performed using DeCyder BVA (Biological Variance Analysis) v6.0. The gels were matched by the software using at least 25 manual landmarks for each gel. The matches were then confirmed visually and any errors made by the software were corrected.

Table 1. The Clinical and Pathological Parameters of Patients

	Non-recurrent	Recurrent
Number of patients	21	25
Mean age	66.3	60.5
Gender		
Male	16	8
Female	5	10
Unspecified	0	7
Histology		
SCC*	11	10
ADC	8	6
BAC	1	0
Large cell carcinoma	1	1
Unspecified NSCLC	0	8

SCC: Squamous cell carcinoma; ADC: Adenocarcinoma; BAC: Bronchioloalveolar carcinoma.

Protein spots of interest were excised manually using a One Touch 2D Gel Spot Picker from gels that had been aligned to corresponding Cy5 images. After trypsin digestion and extraction from the gel pieces, the protein fragments were immediately spotted onto the target plate, and the peptide mass fingerprint data for each protein spot were obtained with a reflection MALDI-TOF mass spectrometer. Proteins were identified by the search engine Protein Prospector against NCBI and MSDB databases.

Quantitative Real-time PCR

Tissues from 25 patients (12 without recurrence and 13 with recurrence) were selected for marker validation by quantitating mRNA expression corresponding to the identified proteins. Total RNA was extracted from 30 mg of tissue using the RNeasy Protect Mini Kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions. The concentration and purity of RNA was verified by measurements of absorption at 260 and 280 nm. cDNA was produced using random primers and the High-Capacity cDNA Archive Kit (Applied Biosystems) according to the manufacturer's instructions.

We evaluated the expression levels of GAPDH and 18s rRNA as internal controls. Real-time PCR was performed by using TaqMan Gene Expression Assays in the ABI Prism 7900 sequence detector (Applied Biosystems, Foster City, CA). Each amplification was performed in triplicate, in parallel with a negative control (absence of cDNA). Each transcript was detected in multiplex reactions with target and control primer-probe sets. Each reaction contained 1.25 µl of the target primer-probe set, 1.25 µl of an internal control primer-probe set, 12.5 µl of Taqman PCR Master Mix and sterile water to a final volume of 25 µl. Reactions were incubated for 10 min at 95°C followed by 40–45 cycles at 95°C for 15 sec and 60°C for 60 s (2-step reaction). Threshold cycle (Ct) was defined by the SDS 2.2.1 Relative Quantification (ΔΔCt) Study program (ABI). The relative expression of mRNA in the tumor samples normalized to the internal control was analyzed as described by Livak and Schmittgen (Ref)[9].

Statistical Analysis

Each of the variables was analyzed independently and as part of a multi-variant analysis to determine the correlation with clinical

outcomes. Student's T-tests were used to identify the difference in relative gene expression values between comparison groups. *P* value <0.05 was considered significant.

RESULTS

Protein Identification by MALDI-TOF MS

We have developed a mass spectrometry-based platform to generate protein expression profiles from lung tissue lysates in order to discover and identify markers predictive of lung cancer recurrence. We compared the protein profiles of 15 non-recurrent tumors and 15 recurrent tumors by MALDI-TOF MS analysis. Comparison of spectra initially revealed 15 peaks that showed varying degrees of differential representation between the two groups of specimens. Upon visual inspection, we found that 14 of the m/z values either did not correspond to valid peaks or, if peaks were present, were equally distributed between non-recurrent and recurrent specimens. One peak however, at m/z about 16.8 kDa, was elevated in approximately 50% of the recurrent cases. Using a combination of solution-phase isoelectric focusing (Rotofor, BioRad, Hercules, CA) and 2D-GE, we identified the protein responsible for this peak as calmodulin (Accession Q13942, MW 16.706 kDa).

Protein Identification by 2D DIGE and MS Protein Identification

Protein expressed by 10 non-recurrent and 10 recurrent lung cancer tissues were analyzed by 2D gel electrophoresis for differential protein expression. On the basis of Decyder software analysis, a total of 2500 protein spots were identified across the three control samples.

To identify the protein spots that were differentially expressed (P<0.05) between non-recurrent and recurrent lung cancers, 17 spots which showed different expression between the two groups were picked from the preparative gel for digestion with trypsin and subjected to peptide mass fingerprint analysis by MALDI-MS. The resulting spectrometric data, including molecular mass, PI and Ion Score in the experiment were analyzed by searching the NCBI and MSDB databases with the search program Protein Prospector. Table 2 lists the molecular masses and isoelectric points of 12 proteins that can be identified.

Table 2. Proteins Associated with Lung Cancer Recurrence

NCBI protein ID	Protein	Protein MW	MS & MS/MS score	Gene	Genebank accession number
AAK74072	Heat shock protein gp96 precursor (Tumor rejection antigen 1)	90138.1	201	TRA1	NM_003299
S68363	Protein disulfide isomerase family A, member 3	56746.8	238	PDIA3	NM_005313
Q6ZW57	Hypothetical protein	23707.2	51	N/A	
Q68DN5	Hypothetical protein DKFZp779N1935	69357.5	52	N/A	
Q6N0B3	Hypothetical protein DKFZp686P03159, annexin family	40327.7	124	ANXA2	
Q6PIQ7	Hypothetical protein	25005.2	78	IGLC1	
1LYWB	Cathepsin D (EC 3.4.23.5), chain B	26229.4	379	CTSD	NM_001909
SNHUC9	Proteasome endopeptidase complex (EC 3.4.25.1) Chain C9	29465.2	164	PSMA4	NM_002789
1HUH	Carbonate dehydratase (EC 4.2.1.1)	28466.2	364	CA1	NM_001738
AAA35831	Ferritin light subunit	20037.1	630	FTL	NM_000146
Q6PI71	Keratin 5 - Homo sapiens (Human)	62340	123	KRT5	NM_000424
Q13942	Calmodulin	16,800	294	CALM	NM_006888

Quantitative Analyses of Gene Expression

To determine the expression levels of the genes encoding putative recurrence marker proteins, we used the real-time quantitative RT-PCR with the ABI PRISM 7900 Sequence Detection System

(Perkin-Elmer Applied Biosystems, Foster City, CA, USA) to analyze a panel of 12 non-recurrent and 13 recurrent NSCLC lung tissues. The mRNA expression of identified proteins were analyzed.

To find an appropriate normalization control for differing amounts of RNA, we assayed and compared the expression levels of the putative housekeeping genes GAPDH and 18S ribosomal RNA. The levels of expression of each housekeeping gene were invariant between recurrent and non-recurrent samples. For 18S rRNA, the mean Ct was 7.99±2.02 for non-recurrent and 7.90±1.07 for recurrent. For GAPDH, the mean Ct was 22.32±2.25 for non-recurrent and 22.99±2.52 for recurrent. Since GAPDH is expressed at roughly the same level as all the marker genes under study (data not shown), we decided to use GAPDH for normalization. Three independent analyses of each cDNA sample were performed for each housekeeping gene and comparisons of these results showed a very high degree of reproducibility (data not shown).

Next, we determined the relative differences in putative marker gene expressions between non-recurrent and recurrent lung cancers. The results of these assays are summarized in Table 3. RNAs from all 25 lung cancer samples were positive for each marker assayed except that CA1 was too low to be detected in most samples, which was similar to other reports[10]. Univariate analysis of relative expression levels showed no marker alone could distinguish patients with recurrent vs. nonrecurrent disease (P>0.1). We next asked whether some combination of these proteins was predictive of recurrence. With a large number of predictors, a common first step to multivariate analyses is to screen predictors and the consider for multivariate analyses those factors that are statistically significant at the 0.25 level of significance in univariate analyses. Unfortunately, none of these variables was statistically significant.

We compared the expression levels of these genes to the known pathologic, clinical, and molecular parameters of the primary and metastatic lung cancer in this study. The marker KRT5 is highly expressed in SCC compared to other types, while FTL and PHKD are expressed higher in ADC (P<0.05). However, no association was observed between the metastatic disease and primary tumors. No correlation was observed between tumor size, histology and the levels of other of these markers.

Table 3. Relative expression ($2^{-\Delta\Delta Ct}$) of isolated genes in NR and R lung cancer tissue

	PSMA4	PHKD	TRA1	ANXA5	PDIA3	FTL	KRT5	CTSD	ANXA2	IGLC1
NR	14.36	12.09	14.55	11.45	12.27	14.45	13.45	12.00	13.73	13.73
	6.99	6.35	6.29	6.86	7.16	6.31	6.68	6.13	7.30	5.48
R	10.92	12.85	10.77	13.38	12.69	10.85	11.69	12.92	11.46	11.46
	7.02	7.87	7.46	7.40	7.28	7.49	7.55	8.00	7.00	8.26
P value	0.24	0.80	0.20	0.52	0.89	0.22	0.55	0.76	0.45	0.45

DISCUSSION

Most patients with non-small cell lung cancer (NSCLC) go on to develop advanced disease. Tumor stage is an important predictor of patient outcome; however, it is currently not possible to specifically identify high-risk patients[11]. The spectrum of proteins expressed by a cell is a reflection of its phenotype; thus, elucidation of protein profiles should allow us to gain more insight into the biology of the disease and, hence, lead to advances in tumor classification, stratification, and prognosis[12]. Based on the distinctly disparate biology of non-recurrent and recurrent lung cancer tissues we reasoned that each could possess different protein profiles. In order to examine these differences, we chose two protein expression platforms, 2D DIGE and MALDI-TOF MS in the hope that relevant phenotypic characteristics of cancer may be discovered that are overlooked by investigations of gene expression alone[13].

Using these protein expression platforms, we have identified eight known proteins (or subunits thereof) and four hypothetical proteins that are differentially expressed in stage I non-small cell lung cancer. Six proteins – carbonic anhydrase, heat shock protein gp96, cathepsin D, ferritin, keratin 5, and calmodulin – have been reported previously to have an association with lung cancer, either as tumor specific markers or as markers of tumor progression[10, 14–18]. Two of the proteins –

protein disulfide isomerase and proteasome chain 9, or PSMA4 – are correlated for the first time with lung tumor progression. Most of these proteins participate in biological pathways including cellular growth and differentiation, transcriptional control, cell signaling, protein disulfide formation and phosphorylation, cell metabolism and apoptosis.

Following protein identification, we validated, by real time PCR, the gene expressions on a tissue set of 25 early stage lung tumors. Unfortunately, we can not find a significant correlation between non-recurrence and recurrence of any gene. However, there are differences in gene expression among NSCLC subtypes. KRT5 is usually expressed in SCC, but not often in lung adenocarcinoma, about 0%–19%[5]. In our study, KRT5 also showed higher expression in SCC. FTL and PHKD, on other hand, were expressed at higher levels in ADC compared to other subtypes.

If protein expression data is consistent with the gene expression profile (which was not always the case here), this can help to identify a subgroup containing high-risk stage I patients. Because of the morphologic and molecular heterogeneity of lung carcinomas and the complex nature of treatment responses, it is still difficult to predict the prognosis of early stage lung cancer by using a combination of multiple factors. But analysis of multiple biologic or molecular markers may be more informative than any single marker.

Acknowledgement

We thank the help of Dr. Edwards F. Patz for research design and data analysis and his support from NIH grant 1RO1-CA-109384-01A for this paper.

REFERENCES

[1] Edwards BK, Howe HL, Ries LA, et al. Annual report to the nation on the status of cancer, 1973–1999, featuring implications of age and aging on U.S. cancer burden [J]. Cancer 2002; 94:2766–92.

[2] Mountain CF. Revisions in the international system for staging lung cancer [J]. Chest 1997; 111:1710–7.

[3] Howard BA, Zheng Z, Campa MJ, et al. Translating biomarkers into clinical practice: prognostic implications of cyclophilin A and macrophage migratory inhibitory factor identified from protein expression profiles in non-small cell lung cancer [J]. Lung Cancer 2004; 46:313–23.

[4] Nikliński J, Niklińska W, Laudanski J, et al. Prognostic molecular markers in non-small cell lung cancer [J]. Lung Cancer 2001; 34:S53–8.

[5] Blackhall FH, Wigle DA, Jurisica I, et al. Validating the prognostic value of marker genes derived from a non-small cell lung cancer microarray study. Lung Cancer [J]. 2004; 46:197–204.

[6] Chen G, Gharib TG, Wang H, et al. Protein profiles associated with survival in lung adenocarcinoma [J]. Proc Natl Acad Sci USA 2003; 100: 13537–42.

[7] Osaki, T Oyama T, Inoue M, Gu CD, et al. Molecular biological markers and micrometastasis in resected non-small-cell lung cancer. Prognostic implications [J]. Jpn J Thorac Cardiovasc Surg 2001; 49:545–51.

[8] Celis JE, Gromov P. 2D protein electrophoresis: can it be perfected [J]? Curr Opin Biotechnol 1999; 10:16–21.

[9] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method [J]. Methods 2001; 25:402–8.

[10] Chiang WL, Chu SC, Yang SS, et al. The aberrant expression of cytosolic carbonic anhydrase and its clinical significance in human non-small cell lung cancer [J]. Cancer Lett 2002; 188:199–205.

[11] Martins SJ, Pereira JR. Clinical factors and prognosis in non-small cell lung cancer [J]. Am J Clin Oncol 1999; 22:453–7.

[12] Voss T, Ahorn H, Haberl P, et al. Correlation of clinical data with proteomics profiles in 24 patients with B-cell chronic lymphocytic leukemia [J]. Int J Cancer 2001; 91:180–6.

[13] Gygi SP, Rochon Y, Franza BR, et al. Correlation between protein and mRNA abundance in yeast [J]. Mol Cell Biol 1999; 19:1720–30.

[14] Wang Q, He Z, Zhang J, et al. Overexpression of endoplasmic reticulum molecular chaperone GRP94 and GRP78 in human lung cancer tissues and its significance [J]. Cancer Detect Prev 2005; 29:544–51.

[15] Higashiyama M, Doi O, Kodama K, et al. Influence of cathepsin D expression in lung adenocarcinoma on prognosis: possible importance of its expression in tumor cells and stromal cells, and its intracellular polarization in tumor cells [J]. J Surg Oncol 1997; 65:10–9.

[16] Fracchia A, Ubbiali A, EI Bitar O, et al. A comparative study on ferritin concentration in serum and bilateral bronchoalveolar lavage fluid of patients with peripheral lung cancer versus control subjects [J]. Oncology 1999; 56:181–8.

[17] Miettinen M, Sarlomo-Rikala M. Expression of calretinin, thrombomodulin, keratin 5, and mesothelin in lung carcinomas of different types: an immunohistochemical analysis of 596 tumors in comparison with epithelioid mesotheliomas of the pleura [J]. Am J Surg Pathol 2003; 27:150–8.

[18] Liu GX, Sheng HF, Wu S. A study on the level of calmodulin and DNA in human lung cancer cells [J]. Br J Cancer 1996; 73:899–901.