REGULAR ARTICLE

Proteomic signatures for histological types of lung cancer

Masahiro Seike[1, 3], Tadashi Kondo[1], Kazuyasu Fujii[1], Tetsuya Okano[1], Tesshi Yamada[1], Yoshihiro Matsuno[2], Akihiko Gemma[3], Shoji Kudoh[3] and Setsuo Hirohashi[1]

[1] Cancer Proteomics Project, National Cancer Center Research Institute, Tokyo, Japan
[2] Clinical Laboratory Division, National Cancer Center Hospital, Tokyo, Japan
[3] Fourth Department of Internal Medicine, Nippon Medical School, Tokyo, Japan

We performed proteomic studies on lung cancer cells to elucidate the mechanisms that determine histological phenotype. Thirty lung cancer cell lines with three different histological backgrounds (squamous cell carcinoma, small cell lung carcinoma and adenocarcinoma) were subjected to two-dimensional difference gel electrophoresis (2-D DIGE) and grouped by multivariate analyses on the basis of their protein expression profiles. 2-D DIGE achieves more accurate quantification of protein expression by using highly sensitive fluorescence dyes to label the cysteine residues of proteins prior to two-dimensional polyacrylamide gel electrophoresis. We found that hierarchical clustering analysis and principal component analysis divided the cell lines according to their original histology. Spot ranking analysis using a support vector machine algorithm and unsupervised classification methods identified 32 protein spots essential for the classification. The proteins corresponding to the spots were identified by mass spectrometry. Next, lung cancer cells isolated from tumor tissue by laser microdissection were classified on the basis of the expression pattern of these 32 protein spots. Based on the expression profile of the 32 spots, the isolated cancer cells were categorized into three histological groups: the squamous cell carcinoma group, the adenocarcinoma group, and a group of carcinomas with other histological types. In conclusion, our results demonstrate the utility of quantitative proteomic analysis for molecular diagnosis and classification of lung cancer cells.

1 Introduction

Lung cancer is a leading cause of cancer mortality worldwide and its incidence continues to increase [1]. Lung cancers are classified as small cell lung carcinoma (SCLC) or non-small cell lung carcinoma (NSCLC). NSCLC consists of three major histological subtypes: squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma (LCC) [2]. The histological typing of lung cancer correlates with its clinical features. SCLC is a high-grade neuroendocrine tumor characterized by its propensity for early metastasis and a short doubling time. Therefore, most patients with SCLC present at an advanced stage and, despite chemotherapy and radiotherapy, the prognosis is generally poor [3]. In contrast, NSCLC is often localized at the time of diagnosis and is surgically resectable. However, prognosis for patients with NSCLC is variable, in part because lung cancers frequently show histological heterogeneity such as AC with SCC component. Although the histology of lung cancer is important in establishing a therapeutic strategy, the molecu-

Correspondence: Tadashi Kondo, MD, PhD, Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
E-mail: takondo@gan2.res.ncc.go.jp
Fax: +81-3-3547-5298

Abbreviations: AC, adenocarcinoma; **LCC**, large cell carcinoma; **LCNEC**, large cell neuroendocrine carcinoma; **NSCLC**, non-small cell lung carcinoma; **PCA**, principal component analysis; **SCC**, squamous cell carcinoma; **SCLC**, small cell lung carcinoma

lar backgrounds determining particular histological phenotypes are obscure.

The development of lung cancer is a multi-step process that includes activation of oncogenes such as ras, myc, EGFR, and c-kit and inactivation of tumor suppressor genes such as p53, p16, Bcl-2, and FHIT [4]. Such genetic alterations can affect the entirety of mRNA and protein expression in an interactive function-related manner and result in complex cancer phenotypes. Therefore, the development of lung cancer cannot be attributed to aberration in any single gene or protein and, in order to understand the mechanisms underlying cancer biology and to develop effective therapeutic strategies, comprehensive approaches to multiple genes and proteins are required. To study the biology of lung cancer, proteome technology has been used to establish the profile of protein expression in lung cancer and to identify novel patterns of aberrant protein expression [5–12].

Here, we used 2-D DIGE to study the protein expression patterns associated with the histology of lung cancer cells. Quantitative protein expression was assessed by multivariate analysis and statistical learning methods. As the majority of lung cancer tissues contain mixtures of different cell types, we utilized well-characterized lung cancer cell lines to capture the protein expression patterns associated with particular histological types of lung cancer. The patterns were then used to classify lung cancer cells isolated from tumor tissues by laser microdissection. We identified by MS the proteins corresponding to the informative protein spots.

2 Materials and methods

2.1 Cell lines, clinical materials and protein extraction

The lung cancer cell lines used had a histological background of: (i) squamous cell carcinoma (PC-1, PC-10, RERF-LC-AI, SQ-5, LC-1/Sq, LC-1F, LK-2, EBC-1, QG-56, and VMRC-LCP); (ii) small cell carcinoma (Lu-130, Lu-134, Lu-135, Lu-139, Lu-140, Lu-165, PC-6, MS-1, SBC-3, and SBC-5); and (iii) adenocarcinoma (A549, PC-3, PC-9, PC-14, RERF-LC-KJ, RERF-LC-MS, RERF-LC-OK, LC-2/ad, ABC-1, and VMRC-LCD). The lung cancer cell lines PC-1, PC-3, PC-6, PC-9, PC-10, and QG-56 were obtained from Immuno-Biological Laboratories (Gunma, Japan). The cell lines A549, PC-14, RERF-LC-KJ, LC-2/ad, SQ-5, LC-1/Sq, LC-1F, RERF-LC-AI, Lu-130, Lu-134, Lu-135, Lu-139, Lu-140, Lu-165, and MS-1 were obtained from RIKEN Cell Bank (Ibaraki, Japan). The cell lines ABC-1, RERF-LC-MS, RERF-LC-OK, LK-2, EBC-1, VMRC-LCD, VMRC-LCP, SBC-3, and SBC-5 were purchased from Health Science Research Resources Bank (Osaka, Japan). All cell lines were maintained in the optimal medium until use. When the cells reached 80–90% confluence, they were washed twice with PBS, scraped off into a tube, and briefly centrifuged. The cell pellets were incubated in a lysis buffer containing 6 M urea, 2 M thiourea, 3% CHAPS, and 1% Triton X-100 for 30 min on ice. After centrifugation at

15 000 rpm for 30 min, the supernatant (cellular protein fraction) was recovered and the protein concentration was measured with a Protein Assay Kit (Bio-Rad, Hercules, CA, USA). The protein sample was adjusted to pH 8.0 with 30 mM Tris.

The tissue specimens were obtained from tumors surgically resected at National Cancer Center Hospital in 2002 and 2003. This study was approved by the institutional review board of National Cancer Center. All of the patients provided informed consent. The tissue samples were from 13 ACs, 13 SCCs, 2 large cell neuroendocrine carcinomas (LCNECs), 1 LCC, and 1 SCLC. The mean age of patients was 68 years (range 48–81 years). A detailed description of the specimens is presented in Table 1. Laser microdissection followed by 2-D DIGE was performed according to our previous report [13]. The pathological diagnosis was established by experienced pathologists. Briefly, O.C.T.-embedded frozen tissue blocks were cut into 10 μm thick tissue sections with a Leica CM 3050 S (Leica, Milton Keynes, UK). The tissue sections were placed on a membrane-coated slide glass (Leica), fixed with 95% ethanol for 30 s and washed in water. After being soaked in 10% Mayer's hematoxylin (Muto Pure Chemicals, Tokyo, Japan) for 1 min, they were washed twice with 95% ethanol and once with water, each for 10 s. The neighboring section was occasionally stained with a standard

Table 1. Clinical variables of lung cancer patients

Variable	Number
Gender	
Male	22
Female	8
Mean age (range)	68 (48–81)
Histological types	
Adenocarcinoma (AC)	13
Squamous cell carcinoma (SCC)	13
Large cell neuroendocrine carcinoma (LCNEC)	2
Large cell carcinoma (LCC)	1
Small cell carcinoma (SCLC)	1
Stage	
I (IB)	10
II (IIB)	13
III (IIIA)	7
Differentiation	
Well	4
	(AC 4/SCC 0)
Moderate	13
	(AC 6/SCC 7)
Poor	9
	(AC 3/SCC 6)
Background lung	
Usial interstitial pneumonia	5
Emphysema	1
Normal lung	4

hematoxylin and eosin method to support the diagnosis. All staining procedures were performed on ice. The area for microdissection was determined by microscopic observation and recorded with Laser Microdissection Version 3.1.0.0 (Leica). Laser microdissection was then performed with a AS LMD (Leica). Cancer cells were collected directly into lysis buffer at a rate of 1 mm^2 of area microdissected *per* 2-D PAGE image required. The protein sample was adjusted to pH 8.0 with 30 mM Tris.

2.2 Fluorescence labeling of protein samples

An internal control mixture was made by mixing portions of the 30 cell line samples. The labeling reaction was performed according to the manufacture's instruction and our previous report [13]. In brief, 30 µg protein sample from the cell lines, or protein lysate corresponding to a 3 mm^2 area of microdissected cancer cells, was reduced by incubation with tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) (Sigma) for 60 min at 37°C. The reduced samples were then labeled with Saturation Cysteine Dye (Amersham Biosciences, Buckinghamshire, UK) for 30 min at 37°C. The characteristics of Saturation Cysteine Dye have been described elsewhere [14]. The internal control sample, which was a mixture of equal amounts of the cell lines, was labeled with Cy3 and the samples from individual cell lines or from microdissected tissues accounting for 3 mm^2 area were labeled with Cy5. The labeling reaction was terminated with an equal volume of lysis buffer containing 130 mM DTT and 2.0% Pharmalyte (Amersham Biosciences). Then Cy3-labeled internal control sample and Cy5-labeled experimental samples were mixed. The volume of mixture was adjusted to 1460 µL with lysis buffer containing 65 mM DTT and 1.0% Pharmalyte. All labeling procedures were performed in the dark.

2.3 2-DE

2-D PAGE was performed as described elsewhere with some modifications [13]. Briefly, the fluorescence-labeled proteins were separated by 2-D PAGE, with the first separation by isoelectric point with IEF and the second separation by molecular weight with SDS-PAGE. Each labeled protein sample, volume of 1460 µL was divided into triplicate IPG dry strip gels (24 cm length, p*I* range between 3.0 and 10; Amersham Biosciences); one gel was rehydrated with 420 µ protein sample and each sample was separated in triplicate gels. After rehydration for 12 h, IEF was performed with an IPGphor (Amersham Biosciences) for a total of 80 kVh at 20°C. After IEF, the IPG gels were equilibrated with equilibration buffer containing 6 M urea, 50 mM Tris-HCl (pH 8.8), 30% glycerol, and 1.0% SDS for 15 min at room temperature. The equilibrated IPG gels were applied onto 9–15% polyacrylamide gradient gels and sealed with low melting temperature agarose (Amersham Biosciences), and the proteins were separated at 20°C for

15 h at 17 W *per* 12 gels with an EttanDalt II (Amersham Biosciences). All electrophoresis procedures were performed in the dark.

2.4 Image acquisition and quantification of protein spots

After electrophoresis, the gels were scanned at appropriate wavelengths for Cy3 and Cy5 dyes with a MasterImager 2640 (Amersham Biosciences). The DIA mode of DeCyder software (Amersham Biosciences) was used to determine the margins of the spots, quantify the spot intensities, and calculate relative spot intensity as the ratio between the total intensity of the gel and the intensity of each individual spot. The BVA mode of DeCyder software was used to standardize the relative spot intensity of the Cy5 image to that of the Cy3 image in the same gel. The standardized spot intensity was then averaged across the triplicate gels. Standardized intensity was integrated and exported as an xml file to the data-mining software.

2.5 Multivariate analysis of protein expression profiles

Hierarchical clustering was performed by calculating Pearson correlations to determine the distances between the samples and by using the algorithm of Ward to construct the tree with GeneMaths software (Applied Maths, Sint-Martens-Latem, Belgium). Principal component analysis (PCA) was used as a dimension-reduction technique with Impressionist software (GeneData, Basel, Switzerland).

To identify the informative protein sets for classification, we used a leave-one-out cross-validation method with Impressionist software (GeneData). We developed a classification rule by applying a support-vector-machine algorithm, where a linear hyperplane in the multi-dimensional protein expression space separates the samples according to the existing group structure with a maximal margin for each sample. The performance of the classification rule is evaluated by a leave-one-out cross-validation. In this study, three groups of lung cancer cell lines, the SCC group, the AC group and the SCLC group, were used to train the support vector machine. A spot ranking method was used to rank the spots according to their contribution to the classification on the basis of the expected alteration of cross-validation error rate by removing the spot. The classification performance of the developed patterns was further validated using the surgical specimens of lung cancer.

2.6 Identification and functional classification of proteins corresponding to protein spots

To identify the proteins corresponding to the spots, the preparative gel containing 500 µg labeled-protein was prepared. As the fluorescence labeling changed p*I* and molecular weight of protein spots, all proteins had to be labeled for a

preparative gel. The gel image of the preparative gel was analyzed by BVA-mode of DeCyder software and the spots of interest were recorded in a text file. The automated spot recovery robot, SpotPicker (Amersham Biosciences), recovered the spots in a 96-well plate. In-gel digestion was performed as described previously [15] and the tryptic peptides were subjected to mass spectrometric study. PMF analysis was performed with a Q-Star Pulser-i equipped with the oMALDI ion source (Applied Biosystems, Framingham, CA, USA). The eluted peptides were mixed with saturated dihydroxybenzoic acid (DHB) in 50% ACN/0.1% TFA and spotted onto a target plate. All mass spectra were externally calibrated with a mixture of three peptides included in the Sequenzyme Peptide Mass Standards kit (Applied Biosystems): des-arg1-bradykinin (M_r 904.4681), angiotensin I (M_r 1296.6853) and glu1-fibrinopeptide B (M_r 1570.6774). Mass spectra were processed with the Analyst QS and MASCOT program and a search of the Swiss-Prot database was performed with a mass tolerance of less than 100 ppm. The protein ranked at top in Analyst QS and/or MASCOT program was considered to be the corresponding one. The identified proteins were classified functionally on the basis of category in GeneCards (http://genecards.bcgsc.ca//index.html).

3 Results

3.1 Clustering of 30 lung cancer cell lines and identification of important spot sets for histological classification

We used 2-D DIGE to generate the protein expression profiles of 30 lung cancer cell lines and 30 lung cancer specimens isolated from lung cancer tissue by laser microdissection. To select reproducible spots and to avoid spots specific to *in vitro* or *in vivo* situations, we selected 131 protein spots present in all Cy3 and Cy5 images. We used hierarchical clustering to interpret the pattern of protein expression. A dendrogram created on the basis of similarities of protein expression profiles across the 30 lung cancer cell lines showed that they were broadly divided into two groups corresponding to their histological background (Fig. 1A). Tree (a) consisted of ten SCLC cell lines, and the remaining cell lines formed the other tree (b), suggesting that the protein expression pattern of SCLC cell lines might be substantially different from those of the other cell lines. All SCC cell lines belonged to branch (e). In contrast, nine of the AC cell lines were clustered in two branches (c) and (d), and one AC cell line (PC-3) was located in branch (e) with the SCC cell lines. AC cell lines seem to have greater heterogeneity compared with cell lines of other tissue types. We attempted to validate the results of clustering by using another unsupervised classification method, PCA. PCA visualizes the relatedness of protein expression, avoiding the deterministic and arbitrary nature of hierarchical clustering. Visual

assessment of relationships between the cell lines indicated that all lung cancer cell lines, except the AC cell line PC-3, formed groups according to their histological type of origin. Consistent with the results of hierarchical clustering analysis, SCLC cell lines formed a distinct group with a wide margin separating SCLC cells from the other cells. Overall, both unsupervised classification methods demonstrated that the histological groups of lung cancer cell lines have certain protein expression patterns that distinguish them from the other groups.

We selected the informative spots for the classification by use of a spot ranking method. The classification error rate was calculated as a function of the number of top-scoring spots used for discrimination. We found that spot sets consisting of the 11, 32 or 64 best-scoring spots minimized the classification error rate (20%), and the error rate did not change until all spots were used (data not shown). These three sets of protein spots appear to be representative of the histological background of lung cancer cells and are candidates as markers for histological classification.

The discrimination performance of the three best-scoring spot sets was evaluated by unsupervised classification methods. Figure 1C shows the results of hierarchical clustering of the lung cancer cell lines on the basis of the expression profile of the 32 selected protein spots. The dendrogram shows that all cell lines were clearly divided according to their histological type of origin (Fig. 1C). In contrast to the results of clustering analysis using all spots (Fig. 1A), the SCC cell line group formed a separate major tree and the SCLC cell line group was clustered close to the AC cell line group. This change was probably a result of the spot ranking method removing spots distinguishing SCLC cell lines from the other cell lines; as a consequence, spots with unique expression patterns in SCC cell lines would have more significant effects on clustering. PCA with the 32 spots also showed better discrimination of the three cell line groups than when all spots were used for the analysis: the three cell line groups were separated from each other by wider margins, and the AC cell line PC-3 was located together with the other AC cell lines (Fig. 1D). We also performed hierarchical clustering and PCA of the cell lines with the spot sets consisting of the 11 or 64 best-scoring protein spots. The cell lines were generally well grouped according to their original histology, but several cell lines were clustered with groups of different histological background (data not shown). Therefore, we selected the 32-spot set for further studies.

3.2 Localization of the 32 protein spots on 2-D gels and identification of proteins corresponding to the spots

Figure 2A shows the localization of the 32 protein spots on the 2-D gels. The spots were distributed over the entire gel image. The intensity of some spots was differentially regu-
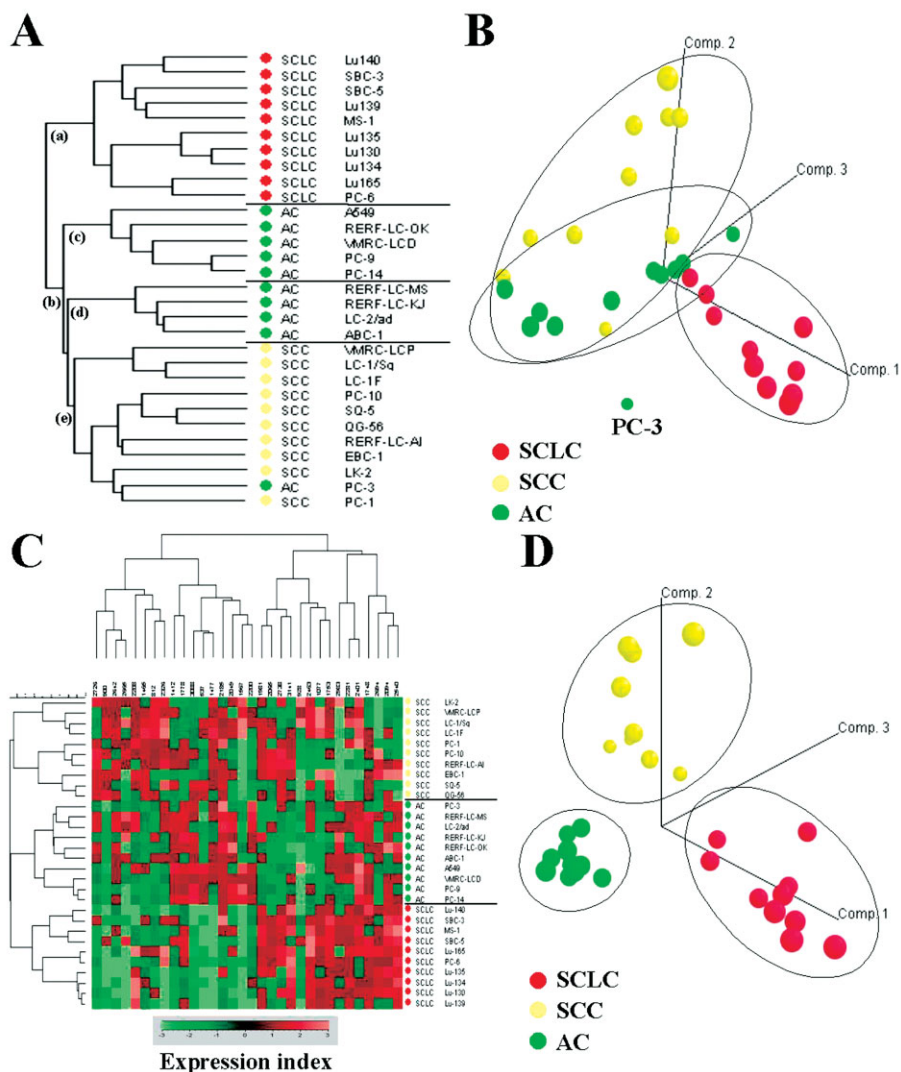
Figure 1. Statistical analysis of 30 lung cancer cell lines as a function of their protein expression profiles. (A) Dendrogram of hierarchical clustering analysis. The cell lines and their histology of origin are listed with color-coding on the left. (B) Three-dimensional plot of principal component analysis (PCA). The apparent groups yielded by PCA are enclosed in circles. Note that the PC-3 cell line, which was located with the SCC cell lines in branch (e) in hierarchical clustering analysis, is located separately from any other cell line group. (C) Two-way hierarchical clustering analysis using the intensity of the selected 32 spots. (D) Three-dimensional plot of PCA using the intensity of the selected 32 spots. The apparent groups yielded by PCA are enclosed in circles.

lated in the various cell lines. For example, spots 3141, 812 and 2463 had higher intensities in AC, SCC and SCLC cell lines, respectively (Fig. 2B). Because the classification is based on standardized spot intensities, which were generated by taking the ratio between Cy5 intensity and Cy3 intensity, visual differences in spot intensity on the Cy5 image between the cell line groups do not necessarily exactly match the numerical data used for the classification.

Mass spectrometric studies were performed on all 32 protein spots and identified 14 of them. The results of mass spectrometric identification are summarized in Table 2.

3.3 Laser microdissection of lung cancer tissues and protein expression profile

We examined whether the 32 spots could be used to classify lung cancer cells *in vivo* according to their histological phenotype. Lung cancer tissues contain many types of

non-tumor cells, including normal counterpart cells, fibroblasts, various inflammatory cells and proliferating vascular structures. Such cellular heterogeneity of tumor tissues could prevent accurate quantitative expression analysis, because each cell population has its own proteome. For more accurate proteomic analysis, we isolated lung cancer cells by laser microdissection and then extracted the proteins from the cells. Figure 3 shows the process of laser microdissection. The diagnosis was performed with a 10 μm thick tissue section stained with conventional HE staining (left panel), and laser microdissection was used to isolate cells from the neighboring sections stained with hematoxylin. The proteins were extracted from the isolated cells, labeled with Cy5, mixed with the Cy3-labeled internal control mixture and then separated by 2-D PAGE. An area of approximately 1 mm^2 of cancer cells was collected for each 2-D image on a large format gel.

Table 2. List of spots informative for the classification of lung cancer cells

Spot no.[a]	Access. no.[b]	Protein name	MS score[c]	Match peptides	MS/MS score[d]	Co-verage (%)	Observed mass (kDa)	Observed pI	Theoretical mass (kDa)	Theoretical pI	Spot ran-king[e] AC/SCC	Function[f]
537	–	–	–	–	–	–	–	–	–	–	16	–
812	P30101	Protein disulfite iso-merase A3	634	14	–	33.7	65.9	5.6	56.8	6.0	23	isomerase activity
900	P05209	Tubulin-alfa-1	407	6	–	18.8	62.0	5.1	50.2	4.9	17	major constituent mi-crotubules
928	P00352	Aldehyde dehydro-genase 1A1	283	6	–	16.8	59.4	6.5	54.7	6.3	18	free retinal binding
1077	P50395	Rab GDP dissociation inhibitor beta	266	5	–	15.1	51.1	6.4	50.7	6.1	21	GDP/GTP exchange reaction
1412	–	–	–	–	–	–	–	–	–	–	14	–
1465	O75874	Isocitrate dehydrogenase 3 alfa	546	10	–	21.1	44.7	6.2	46.7	6.5	7	isocitrate/isopropylmalate dehydrogenase
1477	P08865	40S ribosomal protein SA	844	9	–	33.2	40.6	4.8	32.9	4.8	13	laminin receptor
1567	–	–	–	–	–	–	–	–	–	–	5	–
1748	P04406	Glyceraldehyde 3-phosphate dehydrogenase	73	3	–	6.0	37.2	9.5	35.9	8.6	12	glycolysis/gluconeogene-sis
1753	–	–	–	–	–	–	–	–	–	–	2	–
1778	P00359	Glyceraldehyde 3-phosphate dehydrogenase	80	3	–	11.0	37.2	9.7	35.9	8.6	15	glycolysis/gluconeogene-sis
1981	P06753	Tropomyosin alfa3	733	12	–	33.8	32.8	4.6	32.8	4.7	25	cytoskeleton actin filament stabilization
2049	–	–	–	–	–	–	–	–	–	–	32	–
2065	–	–	–	–	–	–	–	–	–	–	24	–
2094	O00299	Chloride intracellular channel protein 1	134	2	–	8.0	31.7	5.2	26.9	5.1	27	chloride ion channel
2185	–	–	–	–	–	–	–	–	–	–	29	–
2200	P00938	Triosephosphate isomerase	124	2	–	7.0	29.9	6.7	26.5	6.5	10	triosephosphate isomer-ase
2208	–	–	–	–	–	–	–	–	–	–	28	–
2281	–	–	–	–	–	–	–	–	–	–	19	–
2326	–	–	–	–	–	–	–	–	–	–	11	–
2401	–	–	–	–	–	–	–	–	–	–	9	–
2463	P32119	Peroxiredoxin 2	234	5	–	23.2	26.7	5.6	21.9	5.7	3	redox regulation
2540	–	–	–	–	–	–	–	–	–	–	20	–
2642	–	–	–	–	–	–	–	–	–	–		–
2665	Q01469	Fatty acid-binding protein	221	4	–	35.6	23.7	6.4	15.2	6.6	1	lipid metabolism
2694	–	–	–	–	–	–	–	–	–	–	8	–
2726	–	–	–	–	–	–	–	–	–	–	22	–
2738	–	–	–	–	–	–	–	–	–	–	26	–
2983	–	–	–	–	–	–	–	–	–	–	6	–
3088	–	–	–	–	–	–	–	–	–	–	30	–
3141	P09382	Galectin-1	259	5	47	23.7	23.0	5.0	14.6	5.3	4	Carbohydrate binding
											23 spots[g] 97%[h]	

a) Spot numbers correspond to those in Fig. 2

b) Accession no. according to Swiss-Prot

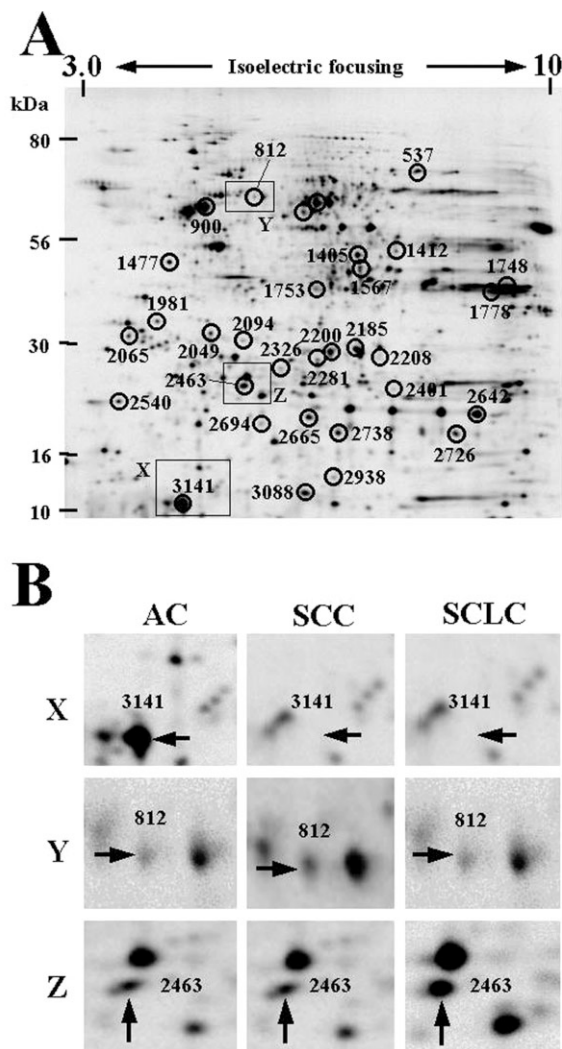c) MS score was generated by Analyst QS

d) MS/MS score was generated by MASCOT

e) Spots were ranked according their contribution to the classification

f) Proteins were functionally classified according to Amigo ontology

g) Number of spots with which the classification error rate was minimal

h) Average classification accuracy of cross-validation analysis

Figure 2. (A) Representative 2-D image of a Cy3-labeled protein mixture from 30 lung cancer cell lines. The 32 best-scoring protein spots for classification are circled; spot numbers correspond to those in Table 2. (B) Differential expression of proteins between cells of different origin is shown.

The cells were categorized according to the expression pattern of the 32 spots. In the dendrogram of hierarchical clustering, the microdissected lung cancer cells were divided into two major trees (Fig. 4A). One tree (a) consisted of nine ACs (AC group 1) and another tree (b) was formed by the remaining cells, including the other four ACs (AC group 2). Each AC group contained lung cancer cells from tumors with various clinical stages and degrees of differentiation, indicating that the expression patterns of the 32 spots were not able to distinguish the ACs on the basis of their clinical stage and differentiation. All SCC cells were clustered in two branches, (d) and (g). All SCC samples with clinical stage III and poor differentiation were located in branch (d) (SCC group 1), whereas all samples in branch (g) were in the early clinical stage, and all except one were moderately differ-

entiated (SCC group 2). Although these observations suggest the possible association of proteomic pattern with clinical stage, a larger number of samples would be required to confirm the correlation.

We also performed PCA of lung cancer cells *in vivo* on the basis of the expression levels of 32 spots (Fig. 4B). The lung cancer cells formed three groups: the AC group, the SCC group, and a group of carcinomas with other histological types. Because the variances due to histological differences might be greater than those due to clinical stage or differentiation, SCCs of late clinical stage and with poorly differentiated histology were not distinguished from other SCCs in PCA. Consistent with the results of the hierarchical clustering study, SCLC, LCNEC and LCC seemed to be distinguishable from the SCC and AC groups. However, because the sample size was not sufficiently large, it was not clear whether they belonged to a certain distinctive group. The spots were ranked according to their contribution to the separation, and the results are summarized in Table 2.

4 Discussion

Histological type is one of the important clinical features of lung cancer. Although the histological differentiation of lung cancer can be assessed by monitoring the expression of tumor markers such as CEA, CA 125, CYFRA 21–1, SCC, and NSE [16], the molecular background corresponding to histological variation is largely obscure. Here, we analyzed protein expression profiles generated by 2-D DIGE by applying multivariate methods and statistical-learning analyses, and found protein groups highly associated with the histological types of lung cancer. Lung cancer tissues are heterogeneous to various extents, and the majority of NSCLCs contain lung cancer cells with different histological types. In addition, lung cancer tissues include non-tumor cells, and laser microdissection may not be able to remove all of them. Therefore, we began our experiments with well-characterized lung cancer cell lines and used protein spots present in cells both *in vitro* and *in vivo*. A similar strategy was employed by Virtanen *et al.* [17] in an mRNA expression study to integrate expression data from lung cell lines and tumors; the genes differentially regulated between lung cancer cells *in vitro* and *in vivo* were removed to dissect away the influence of contaminating non-tumor cells. In this study, to utilize the common image of 2-D PAGE between *in vitro* and *in vivo* study, Saturation Cysteine Dye was used to label protein samples. As Saturation Cysteine Dye has high-sensitivity for spot detection, small amount of proteins from laser microdissected tissues can generate the gels of large-scale 2-D PAGE [13]. Previously, we identified the protein expression patterns corresponding to the histology of lung cancer tissues using 2-D DIGE with the other type of fluorescent dye, Minimal Dye (Amersham Biosciences) [18]. As the 2-D profiles generated by Saturation Cysteine Dye and those by Minimal Dye are different [14], the protein expression pat-
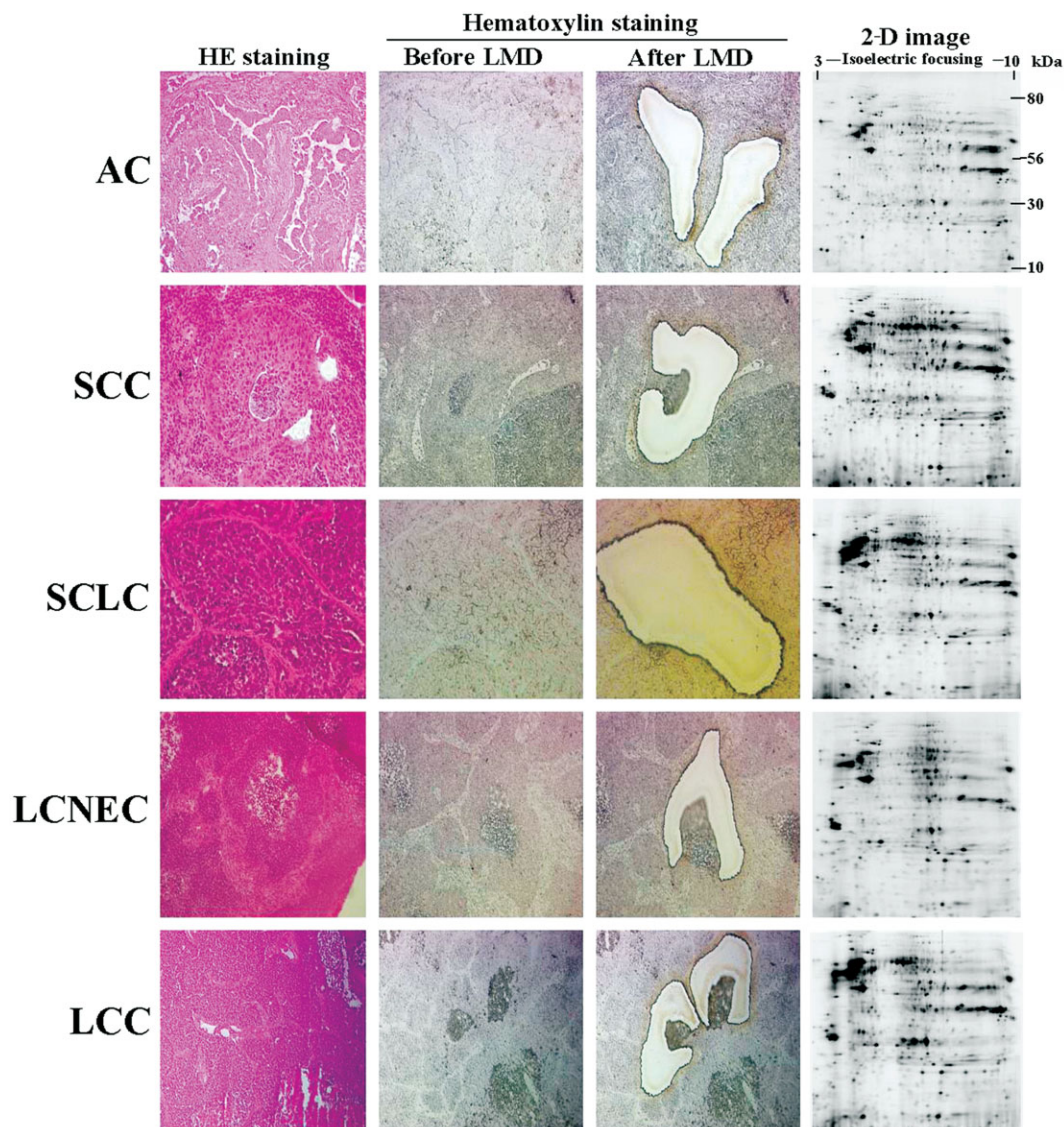
Figure 3. 2-D-DIGE of lung cancer cells isolated by laser microdissection. A frozen section of lung cancer tissue was stained with conventional hematoxylin-eosin staining (left panel). The neighboring section was stained with hematoxylin (middle left) and the tumor cells were recovered from that section by means of laser microdissection (middle right). The proteins were extracted from the microdissected lung cancer cells, labeled with the fluorescent dyes for quantitative expression study and separated by 2-D-PAGE (right panel).

terns corresponding to the histological type of lung cancer tissues were examined using Saturation Cysteine Dye in this report. We found that the proteins identified as informative for the histological classification of cells *in vitro* also classified cells *in vivo* according to their histology. These results demonstrate that the expression patterns of these proteins capture certain histological characteristics that are maintained in cell lines after long-term culture. In addition, our findings suggested that a pattern developed in cell lines can be applied to tumor tissue samples, giving more credence to the applicability of intervention experiments in cell lines to human tissues.

We found that the AC cell line, PC-3, was not classified with the other AC cell lines. A transcriptomic study has also revealed that this cell line had a different mRNA expression pattern from the other AC cells [17]. These results suggest that the cells either might dedifferentiate toward the characteristics of SCC or SCLC, or that SCC or SCLC subcomponents in AC tumors might clonally expand.

Laser microdissection removed the surrounding stromal cells, which would have affected the protein content of the lung cancer cells. We considered that the effects of the stromal components on the tumor cells would result in alterations of the proteome and that such alterations would remain
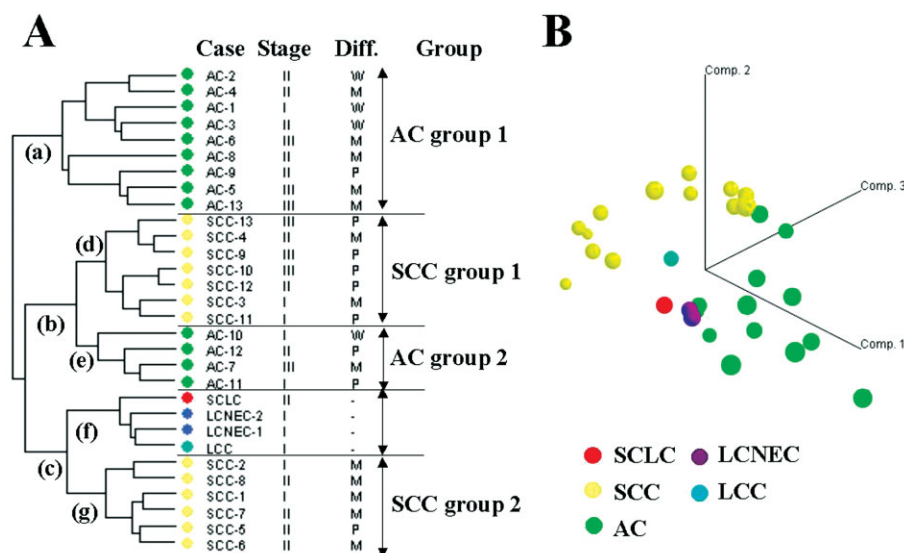
Figure 4. Multivariate studies on lung cancer cells obtained by laser microdissection. (A) Dendrogram of hierarchical clustering analysis of all lung cancer samples on the basis of 32 protein spots. The cell samples and prior information about their histology of origin are listed with color-coding on the left. (B) PCA of all lung cancer samples on the basis of 32 protein spots. The cell samples were plotted in three-dimensional space as a function of the similarity of their expression profiles.

in the frozen tissues. Thus we were able to observe the effects of surrounding stromal cells on the tumor cells. To study the proteome of stromal cells, we may be able to recover the stromal cells using laser microdissection.

Of the 32 informative spots, mass spectrometry identified 14 of the corresponding proteins: three enzymes, two structural proteins and two redox regulators, with the others involved in glycogenesis, small molecule transportation, acting as a receptor or ion channel. The list includes interesting proteins in terms of squamous cell differentiation or cancer progression. Fatty acid-binding protein 5 (FABP5) was considered as the most informative spot for the discrimination of ACs from SCCs in our study, and a previous report showed that FABP5 is associated with epidermal cell differentiation [19]. Thus, FABP5 may also play an important role in the differentiation of lung cancer cells. We also identified proteins involved in cancer progression. The MGr1 antigen was previously reported to be up-regulated in multidrug-resistant gastric cancer cells [20, 21] and was later found to be identical to the human 37 kDa laminin receptor precursor [22]. Further studies of these proteins will refine standard pathologic analysis and give a new insight into lung cancer phenotypes and their differentiation.

Proteomic classification of lung cancer cells resulted in the unexpected identification of a subgroup of SCC with advanced clinical stage, suggesting that the subgroups of SCCs reflect their malignancy. However, the sample size we used was not sufficient for statistical evaluation of our speculation, and further large-scale studies will be required to confirm these possibilities. Recently, proteomic approaches have been employed to develop prognostic tumor markers for lung cancer. Using 2-D PAGE, Hanash's group reported that a set of 20 protein spots could predict the survival of patients with lung adenocarcinoma [10]. MALDI-TOF MS has been used to identify a peptide expression pattern from which the survival of NSCLC patients could be predicted [12].

Our results could support the idea that current proteomic technologies can capture protein expression patterns corresponding to the clinical features of lung cancer and that such patterns will be useful to establish therapeutic strategies. The protein expression patterns corresponding to the subgroups with poor survival or different therapeutic responses should be considered in future studies. The patterns of tumors after chemotherapy, with and without preceding radiotherapy, should also be studied. As the proteins involved in these patterns are strongly associated with certain clinical features of lung cancer, studies on those proteins will lead to further understanding of the biology of this disease.

5   References

[1] Ginsberg, R. J., Vokes, E. E., Rosenzweig, K., *Non-small Cell Lung Cancer*. DeVita, V. T., Jr., Hellman, S., Rosenberg, S. A., (Eds.), *Cancer: Principles and Practice of Oncology*, Ed. 6, Lippincott Williams and Wilkins, Philadelphia 2001, pp. 917–983.

[2] Travis, W. D., Colby, T. V., Corrin, B., Shimosato, Y., Brambilla, E., *Histopathological Typing of Lung and Pleural Tumors. International Histological Classification of Tumors*, Ed. 3, World Health Organization. Springer-Verlag Berlin, Heidelberg 1999, pp. 31–40.

[3] Spira, A., Ettinger, D. S., *N. Engl. J. Med.* 2004, *350*, 379–392.

[4] Massion, P. P., Carbone, D. P., *Respir. Res.* 2003, *4*, 12.

[5] Hanash, S., Brichory, F., Beer, D., *Dis. Markers* 2001, *17*, 295–300.

[6] Chen, G., Gharib, T. G., Huang, C. C., Thomas, D. G. *et al.*, *Clin. Cancer Res.* 2002, *8*, 2298–2305.

[7] Gharib, T. G., Chen, G., Wang, H., Huang, C. C. *et al.*, *Neoplasia* 2002, *4*, 440–448.

[8] Chen, G., Wang, H., Gharib, T. G., Huang, C. C. *et al.*, *Mol. Cell. Proteomics* 2003, *2*, 107–116.

[9] Chen, G., Gharib, T. G., Thomas, D. G., Huang, C. C. *et al.*, *Proteomics* 2003, *3*, 496–504.

[10] Chen, G., Gharib, T. G., Wang, H., Huang, C. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *11*, 13537–13542.

[11] Cuezva, J. M., Chen, G., Alonso, A. M., Isidoro, A., *et al.*, *Carcinogenesis*, 2005, in press.

[12] Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., *et al.*, *Lancet* 2003, *362*, 433–439.

[13] Kondo, T., Seike, M., Mori, Y., Fujii, K. *et al.*, *Proteomics* 2003, *3*, 1758–1766.

[14] Shaw, J., Rowlinson, R., Nickson, J., Stone, T., *Proteomics* 2003, *3*, 1181–1195.

[15] Seike, M., Kondo, T., Mori, Y., Gemma, A. *et al.*, *Cancer Res.* 2003, *63*, 4641–4647.

[16] Molina, R., Filella, X., Auge, J. M., Fuentes, R. *et al.*, *Tumour Biol.* 2003, *24*, 209–218.

[17] Virtanen, C., Ishikawa, Y., Honjoh, D., Kimura, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 12357–12362.

[18] Seike, M., Kondo, T., Fujii, K., Gemma, A. *et al.*, *Proteomics* 2004, *4*, 2776–2788.

[19] Siegenthaler, G., Hotz, R., Chatellard-Gruaz, D., Didierjean, L. *et al.*, *Biochem. J.* 1994, *302*, 363–371.

[20] Fan, K., Fan, D., Cheng, L. F., Li, C., *Anticancer Res.* 2000, *20*, 4809–4814.

[21] Shi, Y., Han, Y., Wang, X., Zhao, Y. *et al.*, *Gastric Cancer* 2002, *5*, 154–159.

[22] Shi, Y., Zhai, H., Wang, X., Wu, H. *et al.*, *Cell Mol. Life Sci.* 2002, *59*, 1577–1583.