Erratum: Application of highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dyes) to laser microdissection and two-dimensional difference gel electrophoresis (2D-DIGE) for cancer proteomics

Tadashi Kondo & Setsuo Hirohashi

Nat. Protoc. doi:10.1038/nprot.2006.421; published online 25 January; corrected online 22 February 2007.

In the PDF version of this article initially published online, the publication date was shown as 25 January 2006 instead of 25 January 2007. The error has been corrected in the PDF version of the article.

Proteomic Signature Corresponding to the Response to Gefitinib (Iressa, ZD1839), an Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor in Lung Adenocarcinoma

Tetsuya Okano,[1,6] Tadashi Kondo,[1] Kiyonaga Fujii,[2] Toshihide Nishimura,[2] Toshimi Takano,[4] Yuichiro Ohe,[4] Koji Tsuta,[5] Yoshihiro Matsuno,[5] Akihiko Gemma,[6] Harbumi Kato,[2,3] Shoji Kudoh,[6] and Setsuo Hirohashi[1]

Abstract **Purpose:** We aimed to identify candidate proteins for tumor markers to predict the response to gefitinib treatment.

Experimental Design: We did two-dimensional difference gel electrophoresis to create the protein expression profile of lung adenocarcinoma tissues from patients who showed a different response to gefitinib treatment. We used a support vector machine algorithm to select the proteins that best distinguished 31 responders from 16 nonresponders. The prediction performance of the selected spots was validated by an external sample set, including six responders and eight nonresponders. The results were validated using specific antibodies.

Results: We selected nine proteins that distinguish responders from nonresponders. The predictive performance of the nine proteins was validated examining an additional six responders and eight nonresponders, resulting in positive and negative predictive values of 100% (six of six) and 87.5% (seven of eight), respectively. The differential expression of one of the nine proteins, heart-type fatty acid–binding protein, was successfully validated by ELISA. We also identified 12 proteins as a signature to distinguish tumors based on their *epidermal growth factor receptor* gene mutation status.

Conclusions: Study of these proteins may contribute to the development of personalized therapy for lung cancer patients.

Non-small cell lung carcinoma (NSCLC) accounts for ~85% of lung cancer cases (1). Biomarker(s) that predict the response to gefitinib (Iressa; AstraZeneca, Macclesfield, United Kingdom), an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, may help to improve the choice of therapeutic strategy in patients with NSCLC. Gefitinib improves NSCLC-related symptoms and quality of life in some patients with advanced NSCLC who do not respond to platinum-based chemotherapy. However, the response rate for gefitinib remains <20% in patients with NSCLC (2–4), and treatment with gefitinib is associated with serious adverse effects, such as severe acute interstitial pneumonia in 5.4% of the patients who received the treatment (5, 6). Thus, it is imperative to select appropriate patients for treatment with gefitinib and exclude patients in whom gefitinib is unlikely to exhibit any clinical benefit. Women, patients who have never smoked, patients with adenocarcinoma, and East Asians are major subgroups of responders (3, 4, 6–8). Recently, gain-of-function somatic mutation in the tyrosine kinase domain of the EGFR has been correlated with the response to gefitinib (9, 10). However, other studies have revealed that correction of the phenotype arising from EGFR mutation may not account for all of the clinical benefits of gefitinib (11, 12), and both preclinical and clinical studies have reported that the efficacy of gefitinib is independent of EGFR expression level (11, 13–15). Although molecular features of the *EGFR* gene, including mutation and high copy number, (16, 17) are associated with response to gefitinib, other molecular markers in the tumor, such as HER2 overexpression (18), Akt phosphorylation (19), and other EGFR downstream molecules (20), also correlate with response. These observations suggest a role for unknown, but important, factors in gefitinib sensitivity. Identification and elucidation of such factors will improve existing therapeutic protocols and contribute to further understanding of the mechanisms of gefitinib sensitivity.

Authors' Affiliations: [1]Proteome Bioinformatics Project, National Cancer Center Research Institute; [2]Clinical Proteome Center and [3]Department of Surgery, Tokyo Medical University; [4]Department of Internal Medicine and [5]Clinical Laboratory Division, National Cancer Center Hospital; and [6]The Fourth Internal Department of Medicine, Nippon Medical School, Tokyo, Japan

Received 7/7/06; revised 11/17/06; accepted 11/29/06.

Grant support: 'Third-Term Comprehensive Control Research for Cancer' conducted by the Ministry of Health, Labor, and Welfare and by the Program for Promotion of Fundamental Studies in Health Sciences in the National Institute of Biomedical Innovation of Japan. Tetsuya Okano is the recipient of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research (Japan).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Current address for K. Fujii: Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan.

Requests for reprints: Tadashi Kondo, Proteome Bioinformatics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Phone: 81-3-3542-2511, ext. 3004; Fax: 81-3-3457-5298; E-mail: takondo@gan2.res.ncc.go.jp.

doi:10.1158/1078-0432.CCR-06-1654

To identify the gene products correlated with the efficacy of gefitinib, genome-wide screening was done recently for NSCLC. A global mRNA expression study using DNA microarrays and biopsy samples identified 51 genes associated with the sensitivity to gefitinib and established a numerical scoring system to predict the response (21). This expression study also led to the establishment of ELISA assays for the identified gene products in serum. Preclinical studies involving mRNA profiling of NSCLC xenografts resulted in the identification of a set of genes that were differentially expressed between tumors that were sensitive and insensitive to gefitinib treatment (22, 23). These studies will lead to the identification of novel biomarkers to predict the response to gefitinib treatment. However, mRNA expression does not necessarily correlate with protein level, and posttranslational modifications, such as phosphorylation, cannot be predicted from the amount of RNA or from the DNA sequence (24). With this background, comprehensive expression studies at the protein level, an approach called proteomics, have been conducted in patients with lung cancer to develop biomarkers that predict clinical outcomes (25). However, no global protein expression study has yet been done on the mechanism of response to gefitinib.

To identify the proteomic signature for sensitivity to gefitinib and to use that signature as a tumor marker to predict the response to gefitinib, we analyzed global protein expression levels in lung adenocarcinoma tissues for whom we have detailed information on *EGFR* gene status. The surgical specimens were obtained at the time of surgery from patients who subsequently had recurrence and received gefitinib monotherapy. We then used two-dimensional difference gel electrophoresis (2D-DIGE) covering ~2,000 proteins to identify a set of proteins of which expression was associated with sensitivity to gefitinib and with EGFR mutation. The predictive performance of the protein set was validated with an independent data set and compared with that of EGFR mutation.

Materials and Methods

Patients and tissue samples. We examined tumor tissues from patients who relapsed after surgery and received gefitinib monotherapy. Two hundred seventy-nine patients who received gefitinib at the National Cancer Center Hospital from July 2002 to December 2004 were evaluated for inclusion in this study. Ninety-two patients relapsed after surgical resection of primary NSCLC and started to receive monotherapy with gefitinib 250 mg/d for 14 days (n = 92). We used tumor tissues obtained at the time of surgery and stored in vapor nitrogen. Fifteen patients were excluded from our study for the following reasons: frozen tissues were not available (n = 10) and tumor histology showed squamous cell carcinoma (n = 4) or pleomorphic carcinoma (n = 1). The histologic features of the tissues were reviewed by two board-certified pathologists (Y.M and K.T.) and diagnosis was based on the latest WHO classification of lung adenocarcinoma (8, 26–28). The tumor responses were classified into complete response (CR), partial response (PR), and progressive disease (PD) using standard bidimensional measurements (29). In this study, patients without a marked reduction of tumor size were subdivided into minor response (MR) and stable disease (SD) groups. MR was defined as a 25% decrease in the sum of the products of perpendicular diameters of all measurable lesions at any point during gefitinib treatment. SD was defined as a <25% decrease in tumor size after treatment. The clinical information is summarized in Table 1, and further information, including EGFR mutation status, is summarized in Supplementary Table S1. Consent was obtained from all patients and the protocol was approved by the institutional review board of the National Cancer Center.

To identify the proteins associated with response to gefitinib, we compared the protein expression profiles of responders (CR and PR) and nonresponders (PD). Of 77 samples available, the effects of gefitinib treatment were not examined for six cases because the treatment was not completed. These six samples were excluded from this study. We constructed two sample sets in the following way (Table 2): a training sample set comprising 31 responders (2 CRs + 29 PRs) and 16 nonresponders (16 PDs) and a test set comprising six responders (6 PRs) and 8 nonresponders (8 PDs) from whom samples were obtained between June and December 2004 (Table 2). As no significant differences were observed between CRs and PRs (Supplementary Fig. S1A), we grouped CRs and PRs together in the responder group.

Protein extraction and protein expression profiling. The frozen tumor tissues were crushed to frozen powder with a Multi-Beads Shocker (Yasui-kikai, Osaka, Japan) under cooling with liquid nitrogen. The frozen powder was then treated with urea lysis buffer (7 mol/L urea, 2 mol/L thiourea, 3% CHAPS, 1% Triton X-100) for 30 min on ice. After centrifugation at 15,000 rpm for 30 min, the supernatant was recovered as cellular protein for the protein expression study.

Protein samples were labeled with CyDye DIGE Fluor saturation dye (GE Healthcare Amersham Biosciences, Uppsala, Sweden) according to

Table 1. Patient characteristics

	No. patients	%
Gender		
Female	33	43
Male	44	57
Age (y)		
Median (range)	62.2 (32-80)	—
Histologic type		
Adenocarcinoma		100
Papillary/acinar/bronchioloalveolar/solid	30/16/9/6	49/26/15/10
Smoking history*		
Never smokers	37	48
Former smokers	12	16
Current smokers	28	36
ECOG performance status†		
0/1/2/3	24/39/9/5	31/51/12/6
Prior chemotherapy		
Yes	30	39
No	47	61
Response to gefinitib		
CR/PR/MR/SD/PD/NE	2/35/2/8/24/6	3/45/3/10/31/8
EGFR gene status		
Mutation L858R	18	23.4
DEL‡	18	23.4
G719§	2	2.6
Wild-type	35	45.4
Unknown	4	5.2

Abbbreviation: NE, not evaluated.
*Never-smokers: those who had never had a smoking habit; former smokers: those who had stopped smoking at least 1 yr before diagnosis; and current smokers: active smokers at diagnosis of NSCLC or those who had stopped smoking less than 1 yr before diagnosis.
†ECOG performance status was monitored according to the previous report (44).
‡Deletional mutations in exon 19.
§G719S and G719C.

Table 2. Training and test sets to develop the classifier for the response to gefitinib

	Training set			Test set		
	Responders, n = 31 (%)	Nonresponders, n = 16 (%)	P	Responders, n = 6 (%)	Nonresponders, n = 8 (%)	P
Age						
Mean ± SD	64.0 ± 8.9	60.5 ± 12.0	0.330	57.5 ± 12.8	62.8 ± 6.1	0.386
Gender						
Male	17 (55)	9 (56)	0.927	3 (50)	5 (62.5)	0.640
Female	14 (45)	7 (44)		3 (50)	3 (37.5)	
Smoking history						
Never smokers	17 (55)	9 (56)	0.286	4 (67)	4 (50)	0.054
Former smokers	7 (22.5)	1 (6)		2 (33)	0 (0)	
Current smokers	7 (22.5)	6 (38)		0 (0)	4 (50)	
EGFR gene status						
Mutation	27 (87)	1 (6)	<0.001	4 (66)	0 (0)	0.006
Wild type	3 (10)	13 (81)		1 (17)	8 (100)	
Unknown	1 (3)	2 (13)		1 (17)	0 (0)	
Prior chemotherapy						
(+)	12 (39)	5 (31)	0.614	6 (100)	0 (22)	<0.001
(−)	19 (61)	11 (69)		0 (0)	8 (100)	
Performance status						
0	11 (35.5)	6 (37.5)	0.945	2 (33)	1 (12.5)	0.347
1	11 (35.5)	10 (62.5)		4 (67)	7 (87.5)	
2	6 (19)	0 (0)		0 (0)	0 (0)	
3	3 (10)	0 (0)		0 (0)	0 (0)	

our previous report (30). We prepared an internal control consisting of a mixture of small portions of all protein samples obtained before May 2004 (31). The internal control sample and the individual experimental samples were labeled with Cy3 and Cy5 CyDye DIGE Fluor saturation dyes, respectively. Five micrograms of Cy3- or Cy5-labeled protein were mixed and coseparated by two-dimensional PAGE. The first-dimension separation was achieved on an Immobiline pH gradient gel (isoelectric point range, 4-7; 24 cm length) with a Multiphor II (GE Healthcare Amersham Biosciences). The second-dimension separation was done with an EttanDalt II (GE Healthcare Amersham Biosciences) with a 9% to 15% gradient polyacrylamide gel. After electrophoresis, the gels were scanned at appropriate wavelengths for Cy3 and Cy5 (Supplementary Fig. S2A). The ratio between Cy5 and Cy3 intensity was calculated for all protein spots in identical gels by the use of DeCyder software (GE Healthcare Amersham Biosciences; ref. 31). The standardized spot intensities were then logarithmically transformed and subjected to a data-mining package (Impressionist; GeneData, Basel, Switzerland). We ran triplicate gels for each sample and calculated the averaged standardized spot intensity.

To assess the reproducibility of the proteomic data with the internal control in our analyses, we generated triplicate protein profiles from identical samples (case 9; Supplementary Table S1) and compared the standardized intensity of the paired spots (Supplementary Fig. S2B). Scattergrams with 1,980, 1,646, and 1,873 spots showed that the intensities of 1,916 (93.7%), 1,599 (94.7%), and 1,770 (94.5%) spots, respectively, were scattered within a 2-fold difference, and the correlation values were also high (r values > 0.93; Supplementary Fig. S2B).

Data analysis. A bioinformatic approach based on a support vector machine (SVM) algorithm and a leave-one-out cross-validation was used to identify proteins of which expression was associated with tumor characteristics, including therapeutic response to gefitinib and the presence of EGFR mutation (32).

Protein identification. Proteins corresponding to the protein spots of interest were identified by mass spectrometry (30). The proteins were recovered in a gel plug by using an automated spot collector (SpotPicker; GE Healthcare Amersham Biosciences) and digested with sequence grade trypsin (Promega, Madison, WI; ref. 30). Trypsin digests were applied to liquid chromatography coupled with tandem mass spectrometry (LTQ, Thermo, Waltham, MA). A database search against Swiss-Prot was done with Mascot software. Patients with a Mascot score of 35 or more were used for protein identification. When multiple proteins were identified in a single spot, the proteins with the highest number of peptides were considered as those corresponding to the spot.

Mutations in the EGFR gene. EGFR mutations in the samples obtained between July 2002 and May 2004 were examined as described in our previous report (8). Analysis of samples obtained between June 2004 and December 2004 was done by high-resolution melting analysis with a LightCycler HR-1 system (Idaho Technology Inc., Salt Lake City, UT).

ELISA. The expression level of heart-type fatty acid–binding protein (H-FABP) in protein samples from 55 lung adenocarcinoma patients (2 CRs, 28 PRs, 6 SDs, 1 MR, and 18 PDs) was measured in a clinical laboratory (SRL, Tokyo, Japan) with a commercially available ELISA kit (MARKIT-M H-FABP, Dainippon Pharmaceutical, Tokyo, Japan) according to the manufacturer's instructions (Supplementary Table S1). All these 55 samples were included in a 2D-DIGE analysis set in this study.

Results

Proteomic signature for the response to gefitinib. We first selected 1,685 protein spots that appeared in at least 80% of the images of Cy3-labeled internal control. We further selected 87 protein spots that showed different intensities between responder and nonresponder groups ($P < 0.05$, Wilcoxon test). Although potentially resulting in a loss of information, this trimming process decreased the possibility that the classifier would be significantly influenced by irrelevant expression data. We selected protein sets for which expression was associated with response to gefitinib by using a SVM algorithm. Accuracy, plotted as a function of spot number, was constant until the number of spots decreased to less than nine, showing that accurate classification did not require all protein spots (Fig. 1A). The location on the two-dimensional map is shown for the selected nine spots (Fig. 1B; Supplementary Fig. S3).

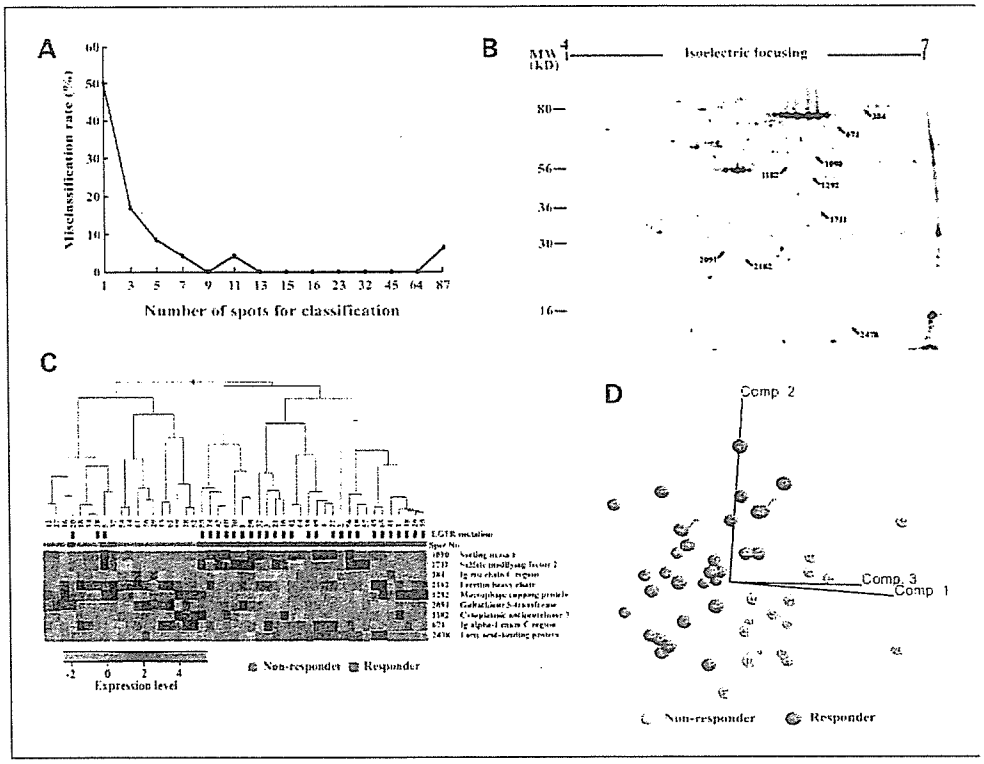

Fig. 1. Data-mining procedure to develop the prediction model for the response to gefitinib. *A*, a spot ranking method selected a few protein spots by which the cumulative error rate of a leave-one-out cross-validation became minimal. The spot ranking method indicated that the error rate was minimal when the prediction model was constructed by a particular nine protein spots. *B*, localization of the selected nine protein spots on the two-dimensional map. An enlarged two-dimensional image is shown in Supplementary Fig. S2. *C*, hierarchical clustering analysis of the samples in the learning set using the selected nine protein spots. Black bars, the presence of EGFR mutations within exons 18 to 21. *D*, principal component analysis of the samples in the learning set using the selected nine protein spots. Comp.1, 2, and 3, the first component 1, 2, and 3, respectively.

Mass spectrometry revealed that these nine spots corresponded to nine gene products (Table 3). Overall similarity of the selected spots is shown in Supplementary Fig. S1B and C. As the responder group in the training set consisted mainly of PRs, the obtained proteomic signature would presumably be more reflective of PR than CR.

The classification performance of the selected nine protein spots was validated by unsupervised classification. Hierarchical clustering showed that all tumor samples in the training set, except for cases 5, 20, and 37, were grouped according to their sensitivity to gefitinib based on the expression pattern of the nine proteins (Fig. 1C). In principal component analysis, all 47 samples seemed to be separated into two groups, although the border between these groups was not clear (Fig. 1D). Although hierarchical clustering and principal component analysis are crude methods of validation of classification, the results obtained using them were consistent.

To validate the predictive performance of the nine protein spots, we investigated a newly enrolled test sample set that was completely independent of the learning set. Based on the expression level of the nine protein spots, the distance of each sample from the hyperplane created by the SVM algorithm, defined as the SVM value, was calculated. The samples with a positive SVM value were grouped as responders and the samples with a negative SVM value were grouped as nonresponders. As a consequence, all training set samples were correctly classified in accordance with their clinical response to gefitinib (Fig. 2). All responders (six PRs) and seven of eight nonresponders (eight PDs) in the test set were also correctly classified. The expression pattern of the nine protein spots in the nonresponder patient (case 75) was more similar to that of the responder group, for unknown reasons. We also validated the results using the samples from patients who showed MR and SD. We found that the two patients showing MR were categorized as responders and that among the eight patients showing three SDs were classified into the responder group and five SDs into the nonresponder group. We did a leave-one-out cross validation for all 47 samples in the training set and the test set using nine protein spots with 1,000 times random permutation. All but two cases, cases 37 and 75, were correctly classified according to their status of response to the treatment. The overall misclassification error rate was 3.3%. Consequently, the model predicted the response to gefitinib in 13 of the 14 (92.8%) newly enrolled samples from the responders and nonresponders and may be useful for disease monitoring.

Proteomic signature for EGFR gene mutation. We studied the spots on the prediction for EGFR mutation. We set a training sample set, including 58 samples (34 mutation-positive samples and 24 mutation-negative samples; Supplementary Table S2). We found that the 12 protein spots showed the high correlation with the EGFR mutation (Supplementary Data; Supplementary Figs. S4-6). The classification and prediction performance of the selected 12 protein spots was successfully validated using the external validation sample set, including four mutation-positive samples and 11 mutation-negative samples (Supplementary Fig. S7). Only one protein, sulfate modifying factor 2, was shared between the signatures for the response and for the mutation (Table 3; Supplementary Table S3).

Expression of H-FABP measured by ELISA. We validated the differential expression of the identified proteins by the use of a widely available clinical assay. The expression level of H-FABP in the same tumor samples as those used in 2D-DIGE was measured with a commercially available ELISA kit intended for serum assays (Fig. 3). H-FABP expression measured by ELISA was highly correlated with that measured by 2D-DIGE (Pearson correlation, 0.76295; $P < 0.0001$). The ELISA study also showed

that the expression level of H-FABP was significantly different between the responder (PR and CR) and nonresponder (PD) groups (P = 0.0031, Mann-Whitney U test) and also between the patients with MR or SD and the nonresponder group (P = 0.0047, Mann-Whitney U test). These results indicate that up-regulation of H-FABP in tumor tissues can be monitored by routine clinical methods.

Discussion

We identified 87 protein spots of which the intensity was statistically significantly different between samples from the responder (CR and PR) and nonresponder (PD) groups in the training set. Application of a data-mining procedure allowed identification of a set of nine protein spots that accurately distinguished between responders and nonresponders. The different expression levels of these nine protein spots allowed classification of 13 of 14 of our test PR and PD cases in accordance with their clinical response to gefitinib. These protein spots classified cases showing a MR to gefitinib (MR) into the responder group. The intermediate cases, SD, were categorized into both responder and nonresponder groups. The usefulness of our findings will be validated in a larger clinical data set.

Table 3. List of proteins for the response to gefitinib

Spots no.*	Rank	Accession no.†	Identtified protein†	MW (DA)‡	pI‡	Ion charge state (+)	MZ (obs)§	Mass‖	δ¶	Miss**	Mascot ions score††	Peptide sequence
384	5	Q96RP9	Ig mu chain C region	49,557	6.35	2	810.3	1,617.7	0.91	0	74	QVGSGVTTDQVQAEAK
						2	640.1	1,277.5	0.63	0	47	YAATSQVLLPSK
671	1	P01876	Ig α-1 chain C region	37,655	6.08	2	919.2	1,836.0	0.32	0	68	QEPSQGTTTFAVTSILR
						2	771.8	1,540.7	0.91	0	54	DASGVTFTWTPSSGK
1090	7	Q9UNH7	SNX 6	46,649	5.81	2	636.5	1,270.5	0.55	0	73	NLVELAELELK
						2	577.0	1,152.2	−0.33	0	39	SLVDYENANK
1182	8	P50453	Cytoplasmic antiproteinase 3	42,404	5.61	2	816.4	1,629.8	0.95	0	82	IEELLPGSSIDAETR
						2	626.6	1,249.4	1.66	0	75	AFQSLLTEVNK
						2	591.0	1,179.5	0.47	0	63	LVLVNAIYFK
						2	757.5	1,513.6	−0.56	0	47	LQEDYDMESVLR +Oxidation (M)
1292	6	P40121	Macrophage capping protein	38,518	5.88	2	633.8	1,264.4	1.18	0	85	VSDATGQMNLTK
						2	676.8	1,351.4	0.05	0	79	YQEGGVESAFHK
						2	932.1	1,861.1	1.11	0	50	MQYAPNTQVEILPQGR +Oxidation (M)
						2	659.8	1,317.3	0.23	0	41	EGNPEEDLTADK
1711	3	Q8NBJ7	Sulfatase modifying factor 2	33,857	7.78	2	792.5	1,581.7	1.32	0	112	MGNTPDSASDNLGFR
						2	779.9	1,557.6	0.15	0	95	GASWIDTADGSANHR
						2	740.0	1,477.6	0.36	0	83	LPTEEEWEFAAR
						2	613.2	1,224.4	−0.02	0	66	FLMGTNSPDSR
						2	629.9	1,256.5	1.27	0	55	SVLWWLPVEK
						2	818.0	1,633.8	0.12	1	55	RLPTEEEWEFAAR
						2	837.7	1,672.9	0.48	0	47	LEHPVLHVSWNDAR
2091	9	P09211	Glutathione *S*-transferase P	23,225	5.44	2	647.5	1,292.5	0.44	0	36	MLLADQGQSWK +Oxidation (M)
2182	4	P02794	Ferritin heavy chain	21,094	5.30	2	823.4	1,643.8	1.04	0	91	MGAPESGLAEYLFDK +Oxidation (M)
						2	648.3	1,294.5	0.03	0	53	NVNQSLLELHK
2478	2	P05413	Fatty acid-binding protein, heart	14,727	6.34	2	735.2	1,467.5	0.81	0	103	LGVEFDETTADDR
						2	798.7	1,595.7	−0.32	1	73	LGVEFDETTADDRK
						2	603.3	1,204.3	0.26	0	70	WDGQETTLVR
						2	455.0	907.0	1.04	0	67	SLGVGFATR
						2	774.7	1,546.8	0.56	0	61	QVASMTKPTTIIEK
						2	438.0	873.0	0.88	0	54	NGDILTLK
						1	889.6	889.0	−0.41	0	45	SIVTLDGGK

Abbreviation: pI, isoelectric point.
*Spot numbers refer to those in Fig. 1B (Supplementary Fig. S3).
†Accession nos. of proteins were derived from Swiss-Prot and National Center for Biotechnology Information nonredundant databases.
‡Theoretical molecular weight and isoelectric point were obtained from Swiss-Prot and the ExPASy database (http://au.expasy.org).
§Experimental m/z value.
‖Relative molecular mass calculated from the peptide sequence.
¶Difference (error) between the experimental and calculated masses.
**Number of missed cleavage sites.
††Mascot ions score (http://www.matrixscience.com/search_form_select.html).

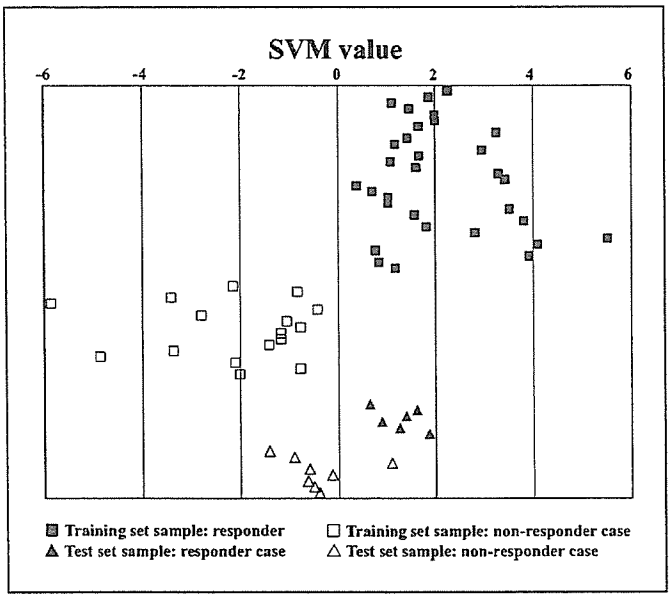

Fig. 2. Predictive performance of the nine spots was validated by examining the SVM value of each sample in the group.

We identified the proteins whose expression was correlated with response to gefitinib and found associations with the EGFR signal pathway and with the biology of lung cancer. Sorting nexin (SNX) 6 is a member of a SNX family that functions in the intracellular trafficking of plasma membrane receptors (33). SNXs form complexes with other SNXs and with plasma membrane receptors. In complexes with SNX1, SNX2, and SNX4, SNX6 interacts with the intercellular portion of the EGFR as well as with transforming growth factor-β receptor, insulin receptor, leptin receptor, and platelet-derived growth factor receptor (34). By binding to the kinase domain of the transforming growth factor-β receptor, SNX6 perturbs transforming growth factor-β signal transduction (34). The other SNX family, SNX1, decreases the expression of EGFR by activating the endosome-to-lysosome pathway with enterophilin-1 (35), although the functions of the complex of SNX6 and EGFR have not yet been reported. The functional association of SNX6 with oncogene product Pim-1, which has been implicated in the development of hematopoietic (36), gastric (37), and prostatic (38) malignancies, suggests the involvement of SNX6 in cancer biology. Kakiuchi et al. (21) reported that another SNX family member, SNX13, was correlated with the response to gefitinib in patients with NSCLC. These reports suggest that SNX6 might play an important role in signal transduction pathways that affect the phenotypes of lung cancer.

We tried to identify the proteins whose expression was associated with EGFR mutation. Because gefitinib is a specific inhibitor of EGFR and mutation of EGFR is considered to be a predictive marker for gefitinib sensitivity, we had expected some similarity between the set of proteins predicting sensitivity to gefitinib and the set of proteins reflecting EGFR mutation status. However, only sulfate modifying factor 2 was common to the two sets. Search of the PubMed database revealed no association of sulfate modifying factor 2 with the EGFR pathway and no evidence for its involvement in resistance to chemotherapy. Similarly, the other proteins correlated with EGFR mutation status had no obvious involvement in the EGFR pathway. Functional studies on these proteins will contribute to further understanding of EGF signaling in cells and to discovery of novel therapeutic targets in lung cancer.

2D-DIGE is a high-performance proteomic technology and a powerful tool to develop candidate biomarkers. However, 2D-DIGE requires expensive fluorescent dyes and well-trained operators to run the gels. Thus, routine clinical studies with multiple large-format two-dimensional gels and a 2D-DIGE protocol are unlikely to be practical. Application of our results requires a simple and cost-effective method that can be used routinely in the clinic. In addition, as we need to examine the expression of multiple proteins, a practical tool for simultaneously measuring the amount of the other proteins is required. With that in mind, we validated measurement of the differential expression of H-FABP by the use of a commercially available ELISA kit (MARLIT-M H-FABP) that is routinely used in hospitals for the early diagnosis of acute myocardial infarction using serum samples. The expression level of H-FABP in tumor tissues as monitored by the ELISA assay was highly correlated with that by 2D-DIGE, and a significant difference in H-FABP expression was observed between responders (CR + PR), minor responders (MR + SD), and nonresponders (PD). Thus, our results can provide a simple and direct method to predict the response to gefitinib.

H-FABP functions in intracellular lipid transport, storage, and metabolism. As H-FABP is highly expressed in heart and released into plasma after myocardial injury, it has been used as a plasma marker for early diagnosis of acute myocardial infarction and stroke. However, many lines of evidence also suggest an association of H-FABP with cancer biology. Higher expression of H-FABP was observed in a more tumorigenic small-cell lung cancer cell line (39) compared with its counterpart. Increased expression of H-FABP is associated with tumor aggressiveness, metastasis, and poor prognosis of gastric cancer (40). In contrast, H-FABP is known to have growth-inhibitory activity in breast cancer cells (41), and breast cancer does not express H-FABP because of gene silencing by hypermethylation (42). These observations suggest complexity in the way that H-FABP is involved in the progression of cancer. Recently, Loeffler-Ragg et al. (43) reported that another FABP family member, E-FABP, is up-regulated in gefitinib-resistant colon cancer cell lines compared with gefitinib-sensitive cell

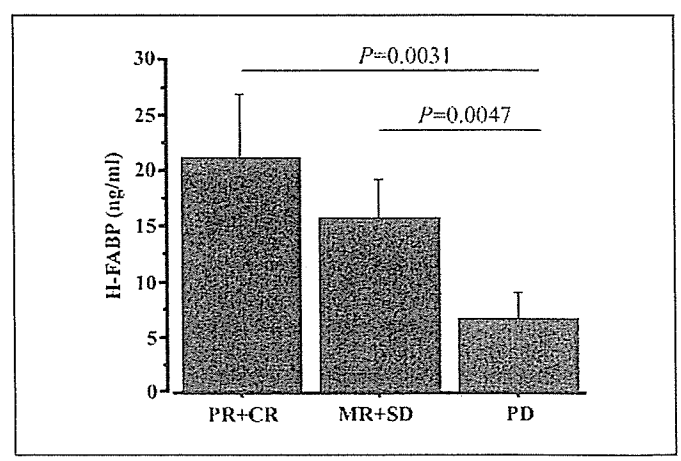

Fig. 3. ELISA assay for H-FABP. The differential expression level of H-FABP was validated by ELISA assay.

lines. Further study on the contribution of the FABP family to cancer phenotypes, including resistance to chemotherapy, will provide novel insights into cancer biology.

In conclusion, our proteomic study has identified proteins whose expression can predict the response to gefitinib in patients with recurrence of lung adenocarcinoma. Large-scale validation of the present results and functional analysis to elucidate the contribution and synergies of the identified proteins in the response to gefitinib will assist in developing novel therapeutic strategies for lung cancer.

References

1. Breathnach OS, Freidlin B, Conley B, et al. Twenty-two years of phase III trials for patients with advanced non-small-cell lung cancer: sobering results. J Clin Oncol 2001;19:1734–42.
2. Herbst RS. Dose-comparative monotherapy trials of ZD1839 in previously treated non-small cell lung cancer patients. Semin Oncol 2003;30:30–8.
3. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (The IDEAL 1 Trial). J Clin Oncol 2003; 21:2237–46.
4. Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. JAMA 2003;290:2149–58.
5. Inoue A, Saijo Y, Maemondo M, et al. Severe acute interstitial pneumonia and gefitinib. Lancet 2003;361: 137–9.
6. Takano T, Ohe Y, Kusumoto M, et al. Risk factors for interstitial lung disease and predictive factors for tumor response in patients with advanced non-small cell lung cancer treated with gefitinib. Lung Cancer 2004;45:93–104.
7. Miller VA, Kris MG, Shah N, et al. Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to gefitinib in advanced non-small-cell lung cancer. J Clin Oncol 2004;22:1103–9.
8. Takano T, Ohe Y, Sakamoto H, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. J Clin Oncol 2005;23:6829–37.
9. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
10. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350:2129–39.
11. Cappuzzo F, Gregorc V, Rossi E, et al. Gefitinib in pretreated non-small-cell lung cancer (NSCLC): analysis of efficacy and correlation with HER2 and epidermal growth factor receptor expression in locally advanced or metastatic NSCLC. J Clin Oncol 2003; 21:2658–63.
12. Pao W, Miller VA. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors, and non-small-cell lung cancer: current knowledge and future directions. J Clin Oncol 2005;23:2556–68.
13. Baselga J, Rischin D, Ranson M, et al. Phase I safety, pharmacokinetic, and pharmacodynamic trial of ZD1839, a selective oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. J Clin Oncol 2002;20: 4292–302.
14. Moasser MM, Basso A, Averbuch SD, Rosen N. The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER2-overexpressing tumor cells. Cancer Res 2001; 61:7184–8.
15. Janmaat ML, Kruyt FA, Rodriguez JA, Giaccone G. Response to epidermal growth factor receptor inhibitors in non-small cell lung cancer cells: limited antiproliferative effects and absence of apoptosis associated with persistent activity of extracellular signal-regulated kinase or Akt kinase pathways. Clin Cancer Res 2003; 9:2316–26.
16. Cappuzzo F, Hirsch FR, Rossi E, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. J Natl Cancer Inst 2005;97:643–55.
17. Hirsch FR, Varella-Garcia M, Bunn PA, Jr., et al. Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol 2003;21:3798–807.
18. Hirata A, Hosoi F, Miyagawa M, et al. HER2 overexpression increases sensitivity to gefitinib, an epidermal growth factor receptor tyrosine kinase inhibitor, through inhibition of HER2/HER3 heterodimer formation in lung cancer cells. Cancer Res 2005;65:4253–60.
19. Cappuzzo F, Magrini E, Ceresoli GL, et al. Akt phosphorylation and gefitinib efficacy in patients with advanced non-small-cell lung cancer. J Natl Cancer Inst 2004;96:1133–41.
20. Han SW, Hwang PG, Chung DH, et al. Epidermal growth factor receptor (EGFR) downstream molecules as response predictive markers for gefitinib (Iressa, ZD1839) in chemotherapy-resistant non-small cell lung cancer. Int J Cancer 2005;113:109–15.
21. Kakiuchi S, Daigo Y, Ishikawa N, et al. Prediction of sensitivity of advanced non-small cell lung cancers to gefitinib (Iressa, ZD1839). Hum Mol Genet 2004;13: 3029–43.
22. Zembutsu H, Ohnishi Y, Daigo Y, et al. Gene-expression profiles of human tumor xenografts in nude mice treated orally with the EGFR tyrosine kinase inhibitor ZD1839. Int J Oncol 2003;23:29–39.
23. Jain A, Tindell CA, Laux I, et al. Epithelial membrane protein-1 is a biomarker of gefitinib resistance. Proc Natl Acad Sci U S A 2005;102:11858–63.
24. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–30.
25. Chen G, Gharib TG, Wang H, et al. Protein profiles associated with survival in lung adenocarcinoma. Proc Natl Acad Sci U S A 2003;100:13537–42.
26. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E. Histological typing of lung and pleural tumours. World Health Organization International Classification of Tumors. New York (NY): Springer-Verlag; 1999.
27. Colby TV, Noguchi M, Henschke C, et al. Adenocarcinoma. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. Pathology and genetics: tumors of the lung, pleura, thymus, and heart. Lyon (France): IARC; 2004. p. 35–44.
28. Ebright MI, Zakowski MF, Martin J, et al. Clinical pattern and pathologic stage but not histologic features predict outcome for bronchioloalveolar carcinoma. Ann Thorac Surg 2002;74:1640–6.
29. Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992;10:239–53.
30. Fujii K, Kondo T, Yokoo H, Yamada T, Iwatsuki K, Hirohashi S. Proteomic study of human hepatocellular carcinoma using two-dimensional difference gel electrophoresis with saturation cysteine dye. Proteomics 2005;5:1411–22.
31. Alban A, David SO, Bjorkesten L, et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 2003;3:36–44.
32. Brown MP, Grundy WN, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci U S A 2000;97:262–7.
33. Worby CA, Dixon JE. Sorting out the cellular functions of sorting nexins. Nat Rev Mol Cell Biol 2002;3: 919–31.
34. Parks WT, Frank DB, Huff C, et al. Sorting nexin 6, a novel SNX, interacts with the transforming growth factor-β family of receptor serine-threonine kinases. J Biol Chem 2001;276:19332–9.
35. Pons V, Peres C, Teulie JM, et al. Enterophilin-1 interacts with focal adhesion kinase and decreases β1 integrins in intestinal Caco-2 cells. J Biol Chem 2004;279:9270–7.
36. Hammerman PS, Fox CJ, Birnbaum MJ, Thompson CB. Pim and Akt oncogenes are independent regulators of hematopoietic cell growth and survival. Blood 2005;105:4477–83.
37. Chen CN, Lin JJ, Chen JJ, et al. Gene expression profile predicts patient survival of gastric cancer after surgical resection. J Clin Oncol 2005;23: 7286–95.
38. Xie Y, Xu K, Dai B, et al. The 44 kDa Pim-1 kinase directly interacts with tyrosine kinase Etk/BMX and protects human prostate cancer cells from apoptosis induced by chemotherapeutic drugs. Oncogene 2006;25:70–8.
39. Zhang L, Cilley RE, Chinoy MR. Suppression subtractive hybridization to identify gene expressions in variant and classic small cell lung cancer cell lines. J Surg Res 2000;93:108–19.
40. Hashimoto T, Kusakabe T, Sugino T, et al. Expression of heart-type fatty acid-binding protein in human gastric carcinoma and its association with tumor aggressiveness, metastasis, and poor prognosis. Pathobiology 2004;71:267–73.
41. Huynh HT, Larsson C, Narod S, Pollak M. Tumor suppressor activity of the gene encoding mammary-derived growth inhibitor. Cancer Res 1995;55: 2225–31.
42. Huynh H, Alpert L, Pollak M. Silencing of the mammary-derived growth inhibitor (MDGI) gene in breast neoplasms is associated with epigenetic changes. Cancer Res 1996;56:4865–70.
43. Loeffler-Ragg J, Skvortsov S, Sarg B, et al. Gefitinib-responsive EGFR-positive colorectal cancers have different proteome profiles from non-responsive cell lines. Eur J Cancer 2005;41:2338–46.
44. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5: 649–55.

 DOI 10.1002/pmic.200600488 *Proteomics* 2006, *6*, 6300–6316

RESEARCH ARTICLE

Protein clusters associated with carcinogenesis, histological differentiation and nodal metastasis in esophageal cancer

Hiromitsu Hatakeyama[1, 4], Tadashi Kondo[1], Kiyonaga Fujii[1], Yukihiro Nakanishi[2], Hoichi Kato[3], Satoshi Fukuda[4] and Setsuo Hirohashi[1]

[1] Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan
[2] Pathology Division, National Cancer Center Research Institute, Tokyo, Japan
[3] Esophageal Surgery Division, Department of Surgery, National Cancer Center Hospital, Tokyo, Japan
[4] Department of Otolaryngology, Head and Neck Surgery, Hokkaido University Graduate School of Medicine, Hokkaido, Japan

We examined the proteomic background of esophageal cancer. We used laser microdissection to obtain tumor tissues from 72 esophageal squamous cell carcinoma cases and adjacent normal tissues in 57 of these cases. The 2D-DIGE generated quantitative expression profiles with 1730 protein spots. Based on the intensity of the protein spots, unsupervised classification distinguished the tumor tissues from their normal counterparts, and subdivided the tumor tissues according to their histological differentiation. We identified 498 protein spots with altered intensity in the tumor tissues, which protein identification by LC-MS/MS showed to correspond to 217 gene products. We also found 41 protein spots that were associated with nodal metastasis, and identified 33 proteins corresponding to the spots, including cancer-associated proteins such as alpha-actinin 4, hnRNP K, periplakin, squamous cell carcinoma antigen 1 and NudC. The identified cancer-associated proteins have been previously reported to be individually involved in a range of cancer types, and our study observed them collectively in a single type of malignancy, esophageal cancer. As the identified proteins are involved in important biological processes such as cytoskeletal/structural organization, transportation, chaperon, oxidoreduction, transcription and signal transduction, they may function in a coordinate manner in carcinogenesis and tumor progression of esophageal cancer.

Received: July 6, 2006
Accepted: August 12, 2006

Keywords:
2D-DIGE / Bioinformatics / Carcinogenesis / Esophageal cancer / Lymph node metastasis

1 Introduction

Esophageal cancer is the eighth most common cancer [1] and the sixth leading cause of cancer death worldwide [2]. Despite the use of modern surgical techniques in combination with radio- and chemotherapy, early recurrence is common and the overall 5-year survival rate remains below approximately 40% [3–5]. Poor prognosis of esophageal cancer is attributed

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

Abbreviation: SCCA, squamous cell carcinoma antigen

to extensive local invasion and frequent regional lymph node metastasis even at initial diagnosis, however, the mechanisms of esophageal cancer progression remain largely obscure. Previous investigations reported that aberrant regulation of PGP9.5 [6], epidermal growth factor receptor [7], p21 [8] and stromelysin [9] is associated with short postoperative survival. However, as multiple genetic and epigenetic events have been observed in cancer progression, the development of complex malignant phenotypes is unlikely to be solely attributable to any single gene. Thus, comprehensive and integrative studies that will link the individual molecular aberrations observed is required to broaden our understanding of esophageal carcinogenesis and subsequently lead to the development of better treatment options for esophageal cancer patients.

Global mRNA expression studies conducted using array-based methodologies identified the gene clusters responsible for the carcinogenesis and progression of esophageal cancer [10–12]. However, practical biomarkers for novel therapeutic strategies and genes to act as molecular targets have not been identified yet, indicating some of the limitations of this approach. The proteins contained in the cells are the functional translations of the genome and directly control the malignant phenotypes of tumor cells. However, studying DNA sequences and measuring the amount of RNA do not predict post-translational aberrations resulting from phosphorylation, glycosylation or proteolysis that occur in cancer progression. In addition, global comparison between transcriptome and proteome revealed that the expression level of mRNAs is not always parallel with that of corresponding proteins [13-15]. Previous studies using proteomic approaches detected proteins that were aberrantly regulated in esophageal cancer tissues [16-18]. In this study, we were able to detect candidate biomarkers for early diagnosis and monitoring cancer progression by improving the existing proteomics strategies.

In this report, we used laser microdissection to recover tumor cells and the matched neighboring normal epithelial cells from the surgical specimens of 72 esophageal cancer cases, and subjected the recovered cells to proteomic analysis using 2D-DIGE. The 2D-DIGE has previously been applied to laser-microdissected esophageal tumor tissues [19, 20]. However, the fluorescent dyes used to label the proteins had low sensitivity, equivalent to that of silver staining, and the number of samples examined in these studies was not sufficient to allow meaningful bioinformatics analyses. Here, we used highly sensitive fluorescent dyes for 2D-DIGE, integrated proteomic and clinico-pathological information using bioinformatics methods on a large clinical sample set, and found protein clusters associated with carcinogenesis, histological differentiation and lymph node metastasis of esophageal cancer. We identified the proteins included in the clusters by MS, and validated the proteomic results using specific antibodies. The functional classification and chromosomal localization of the identified proteins were also examined. This is the first detailed, comprehensive and quantitative proteomic study on esophageal cancer using a large clinical sample set.

2 Materials and methods

2.1 Patients

Tumor cells and their adjacent normal mucosal cells were collected from 72 of 185 cases of esophageal squamous cell carcinoma, which were surgically resected in 1998 and 1999 at the National Cancer Center Hospital. Two or three tissue fragments, less than 10 mm^3 in volume, were grossly obtained from the 72 cases. Matched normal mucosal tissues were also obtained from 57 of these cases. The resected tissues were snap-frozen in liquid nitrogen and stored at -80°C until use. All 72 enrolled cases were newly diagnosed as squamous cell carcinoma of the esophagus. The patients did not receive anticancer treatment prior to surgery and the cases were followed up for at least five years after surgery. The summary of the clinicopathological data of the cases analyzed is shown in Table 1, while individual case informa-

Table 1. Clinicopathological data of 72 esophageal cancer cases analyzed

Gender	
Male	62
Female	10
Age (mean ± SD)	61.9 ± 7.2
Location	
Cervical	4
Upper thoracic	8
Middle thoracic	40
Lower thoracic	18
Abdominal	2
Histological differentiation	
Well differentiated	16
Moderately differentiated	29
Poorly differentiated	27
Prognosis (5-year survival)	
No evidence of disease	30
Dead of disease	39
Dead of another disease	3

Table 2. TNM classification of 72 esophageal tumors analyzed[a)]

	Lymph node metastasis		
	Positive	Negative	Total
Depth of invasion			
pT1	2	0	2
pT2	6	1	7
pT3	41	12	53
pT4	6	4	10
	55	17	72
	Stage II A : 13	Stage II B : 8	Stage III : 51

a) Tumors were staged according to the WHO Tumor-Node-Metastasis classification scheme [21].

tion is available in Supplementary Table 1. The tumors were staged based on WHO guidelines [21] and the results are summarized in Table 2. This study was approved by the ethics committee of the National Cancer Center and written informed consent was obtained from the patients.

Based on the WHO classification scheme for esophageal tumors [22], tumors in this study were well differentiated in 16 (22%) cases, moderately differentiated in 29 (40%), and poorly differentiated in 27 (38%) (Table 1) [22]. Based on the same scheme, the depth of invasion was variable, although most cases were pT3 stage (n = 53, 74%). Lymph node involvement was observed in 55 cases (76%) (Table 2). Tumors with distant metastases were not included in this study. As histological differentiation is a significant variable in predicting overall survival [23] and proteins affecting differentiation may be prognostic biomarker candidates, particular emphasis was placed on data analysis in relation to the degree of tumor differentiation.

2.2 Laser microdissection and protein extraction

We specifically recovered tumor cell populations to the exclusion of non-cancerous cells using laser microdissection for the subsequent proteomic study as in our previous report [24]. In brief, the frozen tissues were embedded in OCT compound (Sakura Finetechnical, Tokyo, Japan), 8-μm thick sections were prepared using a cryostat (Leica CM 3050S, Leica Microsystems, Wetzlar, Germany) and stained with hematoxylin and eosin to confirm the pathological diagnosis. Eight-micron thick neighboring sections were mounted on a thin supporting polyethylene membrane pretreated with a tissue-adhesive solution (0.1% poly-L-lysine, Sigma Aldrich, St. Louis, MO). The sectioned tissues were routinely stained with Mayer Hematoxylin; all staining procedures were performed on ice. Mayer Hematoxylin-stained sections were subjected to laser microdissection (Leica Laser Microdissection version 3.1.0.0, Leica Microsystems). Tumor cells were recovered from non-necrotic tissues using a pulsed ultraviolet laser beam, avoiding sample contamination with infiltrating inflammatory cells, stromal cells and vascular components (Fig. 1). Corresponding morphologically normal esophageal epithelium samples, located at least 5 cm away from the cancerous tissues, were also obtained from 57 of the 72 patients. Hematoxylin and eosin-stained sections were examined to confirm the diagnosis and following sections were stained only with hematoxylin for the proteomic study, as eosin staining has been found to hinder 2D-DIGE [24]. Protein corresponding to 1 mm^2 of microdissected area, recorded during microdissection, was recovered from hematoxylin-stained tissues for each 2D-DIGE gel.

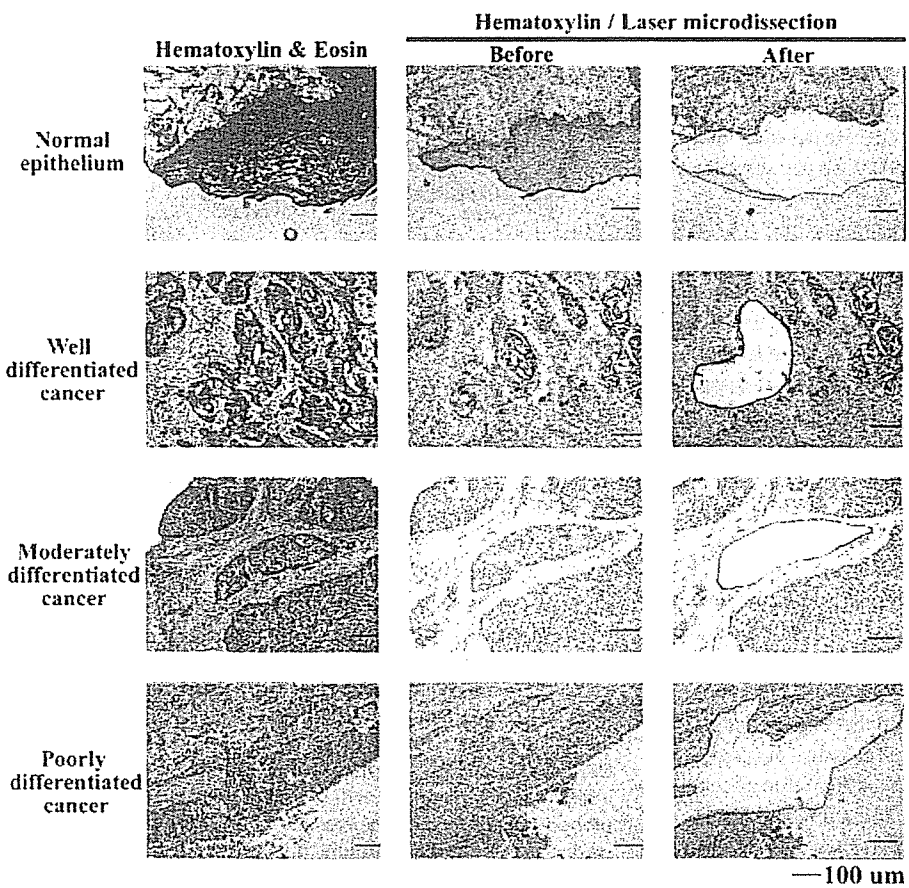

Figure 1. Representative histology of normal esophageal epithelium and esophageal tumor with well, moderate, and poor differentiation. Tissue sections were stained with hematoxylin and eosin for histological observation and hematoxylin alone for the proteomic study. The microscopic appearance of the tissues before and after laser microdissection is demonstrated.

2.3 Preparation of fluorescence-labeled protein samples

The microdissected tissues were immediately treated with urea lysis buffer, consisting of 6 M urea, 2 M thiourea, 3% CHAPS and 1% Triton X-100. The total area of microdissected tissue was 1 mm^2 per 2D-DIGE gel [24]. Protein labeling was carried out as in our previous report [25]. In brief, microdissected tissues with an area of 3 mm^2 were incubated with 50 μL of the urea lysis buffer with 40 mM Tris-HCl (pH 8.0). The protein samples were then reduced with 8 nmol tris-(2-carboxyethyl) phosphine hydrochloride (TCEP; Sigma Aldrich) at 37°C for 60 min and were fluorescence labeled by incubation with 12 nmol of Cy5 (CyDye DIGE Fluor saturation dye, GE Healthcare Biosciences, Uppsala, Sweden) at 37°C for 30 min. The labeling reaction was terminated by addition of urea lysis buffer containing DTT and Pharmalyte (pH 4–7, GE Healthcare Biosciences) so that their final concentrations were 65 mM and 2.0%, respectively.

We created an internal control sample by mixing a small portion of an individual protein samples. The protein concentration was measured with a Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). Five micrograms of the internal control sample was incubated with 2 nmol TCEP at 37°C for 60 min and labeled with 3 nmol of Cy3 dye (CyDye DIGE Fluor saturation dye, GE Healthcare Biosciences) at 37°C for 30 min. After terminating the labeling reaction, individual Cy5-labeled samples corresponding to 1 mm^2 of laser-microdissected area were mixed with 5 μg of Cy3-labeled internal control sample and urea lysis buffer containing 35 mM DTT and 1.0% Pharmalyte (GE Healthcare Biosciences) was added to a final volume of 420 μL per sample.

2.4 2D-PAGE and image acquisition

Protein expression profiles were created as in our previous report, with some modifications [25]. In brief, for first dimension separation, IPG gels (p*I* range 4 to 7, 24 cm length) were rehydrated with a mixture of the Cy3-labeled internal control sample with the Cy5-labeled individual samples at room temperature overnight. IEF was performed using Multiphor II (GE Healthcare Amersham Biosciences) at 20°C. Following equilibration in a buffer containing 6 M urea, 2% SDS, 50 mM Tris-HCl pH 8.8, 30% glycerol and 32 mM DTT, IPG gels were transferred, in batches of 12, onto 12.5% homogenous polyacrylamide gels and embedded in agarose between low-fluorescent glass plates. Proteins were then subjected to a second dimension separation at 17 W for 15 h at 20°C on a 40-cm long SDS-PAGE gel using a vertical electrophoresis apparatus equipped with a cooling system (Bio-Craft, Tokyo, Japan). For preparative purposes, 200 μg of proteins were first labeled with a single fluorescent dye (CyDye DIGE Fluor saturation dye) and subjected to 2D-PAGE. Gels were scanned at the appropriate wavelengths for Cy3 and Cy5 with Typhoon Trio (GE Healthcare Biosciences) to obtain the images of labeled proteins. Spot detection, quantification, and standardization of spot intensities were carried out using the DeCyder 5.0 software (GE Healthcare Biosciences).

Figure 2A demonstrates the protocol of the 2D-DIGE experiments. Both Cy3- and Cy5-images were generated from single gels by laser scan. All Cy3-images contained all spots that were detected on the Cy5-images because they represented the common internal control sample, which was a mixture of all individual samples. Thus, the gel-to-gel variations were canceled, because the Cy5 to Cy3-intensity ratio was standardized for every spot and every gel.

To assess the reproducibility of our 2D-DIGE system, we examined the similarity of the protein expression profiles of identical tissue samples (Fig. 2B). Microdissection was performed twice from the same tumor tissue of case No. 50 (Supplementary Table 1) and protein samples were independently prepared. Protein expression profiles were created from duplicate gels for each protein sample using 2D-DIGE.

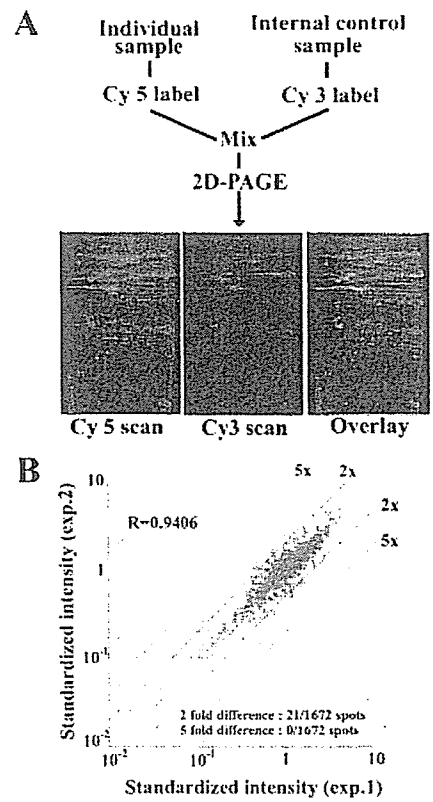

Figure 2. (A) The 2D-DIGE protocol is illustrated. Each individual sample and a pooled reference sample were labeled with Cy5 and Cy3, respectively, mixed, and separated on a 2D-PAGE gel. Gels were scanned with laser, and a set each of Cy3- and Cy5-images was obtained from each gel. (B) The reproducibility of 2D-DIGE for the quantitative study was evaluated by comparing two independent separations of the same protein sample. Spots were plotted based on their expression level. More than 98% spots were scattered within twofold differences and the correlation value was high (r = 0.9406).

A pair wise comparison revealed that more than 98% spots were scattered within twofold differences and the correlation value was high (r = 0.9406). Visual inspection of the gel images revealed that the protein spots that showed higher than twofold differences between the profiles were shadowed by the streaking of the other spots or located near the end of the gel (data not shown). Such spots were filtered out during the bioinformatics analyses.

2.5 Data analysis

As previously [25], we used bioinformatics to link quantitative proteomic data with clinicopathological parameters in order to identify the fraction of the proteome most relevant to esophageal cancer progression. In brief, standardized spot intensities were exported to Expressionist (GeneData, Basel, Switzerland), a data mining program. The standardized spot intensities were averaged between the duplicate gels and analyzed using scatter plotting, self-organizing map (SOM) [26, 27], hierarchical clustering and principal component analysis. Survival curves were calculated by the Kaplan-Meier method [28] and differences in survival probabilities were examined with the log-rank test.

2.6 Protein identification by MS

In-gel digestion and MS protein identification were described as in our previous report [29]. In brief, the target protein spots were recovered from the gels using an automated spot recovering machine, ProHunter (AsOne, Osaka, Japan) into a 96-well PCR plate. Gel plugs were washed with methanol, ammonium bicarbonate, and ACN three times, dried and treated with TPCK-treated trypsine overnight. The peptides were extracted from the gel by treating the gel with ACN.

The tryptic peptides resulting from the in-gel digestion were subjected to analysis by nano-scale microcapillary RP LC-ESI MS/MS. Paradigm MS4 HPLC dual solvent delivery system (Michrom BioResource., Auburn, CA) for micro-flow HPLC, an HTS PAL auto sampler (CTC Analytics, Zwingen, Switzerland) and a Finnigan LTQ linear IT mass spectrometer (Thermo Electron, San Jose, CA) equipped with a nano-ESI (NSI) source (AMR, Tokyo, Japan) were used for protein identification as described [30]. Digested peptide mixtures were separated on a microcapillary RP Magic C18 column (3 μm, 200 Å, 50 × 0.2 mm i.d.; Michrom). Peptides were eluted through 10 to 80% linear gradient buffer B (10% water and 0.1% formic acid in ACN v/v) in buffer A (2% ACN and 0.1% formic acid in water v/v) over 10 min. The effluent solvent from the HPLC was placed into the mass spectrometer through an NSI needle at a flow rate of 1.0–1.2 μL/min (FortisTip; OmniSeparo-TJ, Hyogo, Japan). The voltage was 1.8 kV and the capillary was heated to 200°C. No sheath or auxiliary gas was used. The mass spectrometer was operated in a data-dependent acquisition mode in which MS acquisition with a mass range of *m/z* 450–1800 would automatically switch to MS/MS acquisition under the automated control of the Xcalibur software (version 2.0, Thermo Electron). The full MS scan was acquired by the following MS/MS experiments with an isolation width of *m/z* 2.0; the activation amplitude parameter was set at 35%, on the three most abundant ions detected in the survey scan. Data were acquired with dynamic mass-exclusion windows that had a 30-s exclusion duration and exclusion mass widths of −1.0 and +2.0 Da.

Raw data were converted to dta format (peak list file) using ExtractMS version 2.11 (ThermoElectron), the software supplied with the instrument, with peptide mass range set at 450 to 600 Da and MS/MS minimum peak list set at 25, prior to launching MASCOT searches. MASCOT (version 2.1, Matrix Science, London, UK) searches were performed against *Homo sapiens* subsets of the sequences in the Swiss-Prot (12867 sequences in the Sprot_47.8 fasta file) and NCBI (131447 sequences in the NCBInr_20050422 fasta file) non-redundant protein sequence databases. The following search parameters were used in all MASCOT searches: tolerance of two missed trypsin cleavages, variable modification on the methionine residue (oxidation, +16 Da), and a maximum error tolerance of ±2.0 Da in the MS data and ±1.0 Da in the MS/MS data. Protein hits with more than two significant matched peptides with the distinct sequences (p <0.05, which with our search parameters equals a MASCOT ions score of 35 or more for the Swiss-Prot database and 42 or more for the NCBI database) were statistically considered to estimate the confidence of protein identifications. In addition, the MS/MS spectra of the identified peptides were manually inspected.

2.7 Western blotting

Proteins were separated by SDS-PAGE on a 10–20% polyacrylamide gradient gel and transferred onto an NC membrane. The differential expression of the identified proteins was monitored using antibodies against cytokeratin 14 (1:500, Neo Markers, Fremont, CA), periplakin (1:200, Santa Cruz Biotechnology, Santa Cruz, CA), annexin I (1:5000, BD Bioscience, San Jose, CA), squamous cell carcinoma antigen 1/2 (SCCA1/2) (1:200, Santa Cruz Biotechnology), calgulanulin B (1:200, Santa Cruz Biotechnology), HSP60 (1:5000, BD Biosciences), and beta-actin (1:1000, Abcam, Cambridge, CB, UK) as controls. The secondary antibodies against mouse IgG, rabbit IgG (both GE Healthcare Biosciences) and goat IgG (Santa Cruz) were used for the antibodies against cytokeratin 14, annexin I, HSP60 and beta-actin; against SCCA1/2 and calgulanulin B; and against periplakin, respectively. Antibody-antigen complexes were visualized with an ECL system (GE Healthcare Biosciences) using LAS 1000 (Fuji Film, Tokyo, Japan). The intensity of the protein bands was quantified and the relative intensity for the examined proteins was calculated by standardizing the intensity of the beta-actin bands on the same membrane.

2.8 Chromosomal location of the genes corresponding to the identified proteins

The chromosome location of the identified proteins was studied by searching the NCBI database using a database search software, Annotation Tracker (GE Healthcare Biosciences).

3 Results

3.1 Hierarchical clustering analysis of tumor samples based on their protein expression profiles

Seventy-two tumor tissues and 57 normal tissues were classified by hierarchical clustering analysis according to their protein expression profiles (Fig. 3). Protein spots that appeared in at least 80% of the Cy3-images were used for the analysis. The numbers of spots analyzed per sample ranged between 1414 and 1730. Based on the overall similarity of protein expression, the samples were divided into two groups: tumor tissues (Tree I) and normal epithelial tissues (Tree II) (Fig. 3). The tumor samples in Tree I were further grouped reflecting their histological differentiation; all but one well-differentiated tumors were located in one branch together with the moderately differentiated tumors, and all poorly differentiated tumors were distinguished from the well-differentiated tumors. The presence or absence of lymph node metastases did not appear to be associated with the proteomic classification of tumors (Tree I) or normal epithelial tissues (Tree II); both were grouped independently of their lymph node metastasis status. The anatomic site was not associated with the proteomic profile classification either. Enlarged trees with the patient ID and the number of spots in the individual samples are demonstrated in Supplementary Figs 1 and 2, respectively. These observations suggested that the proteomic profiles most dominantly reflected the malignant transformation, and secondly the histological differentiation.

The protein spots were also clustered according to their expression level across the 129 samples (72 tumor tissues and 57 normal counterparts). Of 1730 protein spots identified, the protein spots in categories A (544 spots, 31%) and C (705 spots, 41%) showed decreased or increased intensity, respectively, in many of the tumor tissues compared with their normal tissues. There did not seem to be any obvious or consistent differences between the protein expression levels of the normal and the tumor tissues in category B (481 spots, 28%). Selection of spots with different signal intensity levels was later achieved, however, taking into account statistical significance.

3.2 Comparison of the protein expression profiles

We examined the similarity of the protein expression profiles of the 129 laser-microdissected tissue samples. Pair wise correlation coefficiencies across all samples were performed

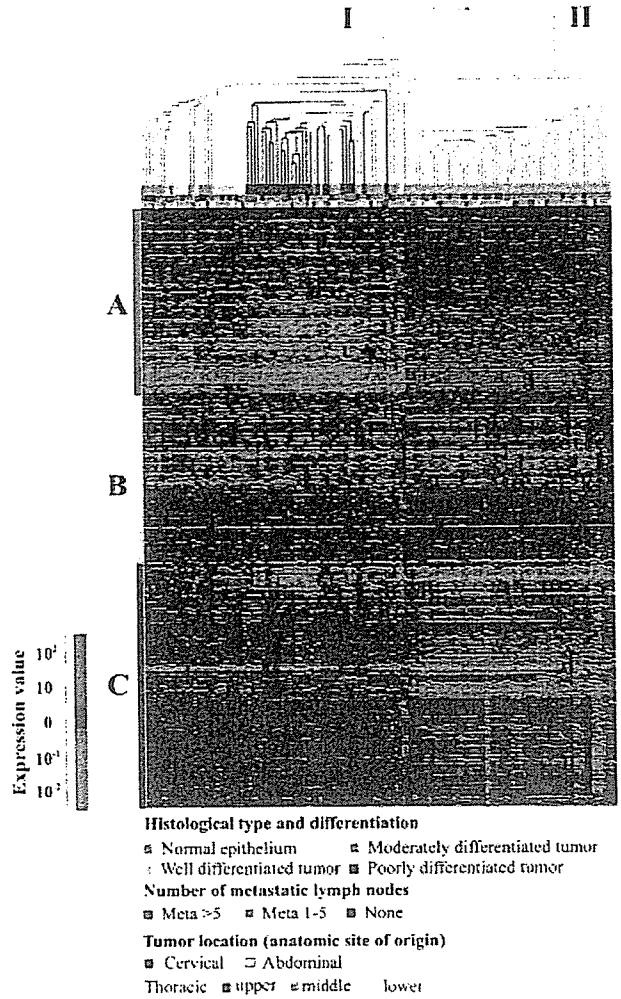

Figure 3. Unsupervised hierarchical clustering classified 129 laser microdissected tissue samples based on the fluorescence intensity of 1730 identified protein spots. The histological differentiation of the samples was demonstrated by the colored nodes as indicted in the panel. Note that the two dominant trees (Tree I and II) accurately partition the tumor tissues and normal tissues, respectively. Tree I is subdivided into branches correlating with histological differentiation. The tumor location or the presence or absence of lymph node metastases did not appear to be associated with the proteomic classification of tumors (Tree I) or normal epithelial tissues (Tree II). Proteins were categorized according to their preferential expression in the normal and tumor tissues. Protein spots in category A (544 spots) and C (705 spots) showed increased or decreased intensity in tumor samples. Protein spots in category B (481 spots) did not show consistent differences between normal and tumor tissues. Patient ID for the samples are demonstrated in Supplementary Fig. 1.

and the results are summarized in the correlation matrix (Fig. 4A). The correlation matrix demonstrated that samples in the same category, either normal epithelial tissues or tumor tissues, had similar protein expression profiles. By visual inspection, normal epithelial tissues showed more homogeneous protein expression profiles compared with

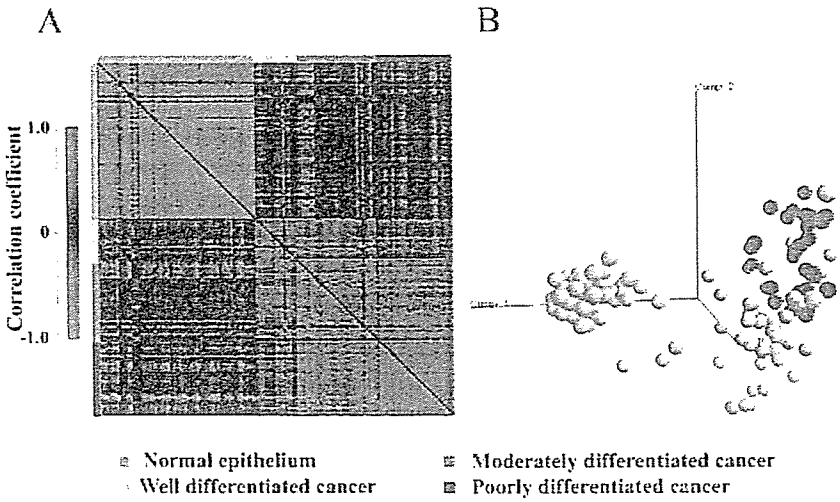

Figure 4. (A) Correlation matrix summarizing the overall similarity of the expression profiles of unselected protein spots for all sample pairs. The red and green colors indicate that the protein expression in the paired samples had high and low similarity, respectively. Samples in the same category, either normal epithelial tissues or tumor tissues, had similar protein expression profiles. (B) Principal component analysis grouped the tissue samples based on unselected protein spots. The normal tissues and the status of histological differentiation of tumor tissues are color-coded as indicated in the panel. Normal epithelial tissues were distinguished from tumor tissues, and well-differentiated tumors from poorly differentiated ones. The distances between the normal epithelial tissue samples were smaller than the distances among the tumor tissue samples, suggesting that the protein expression profiles of the normal tissues may be more homogeneous than the profiles of the tumor tissues. The patient ID and the number of spots used in the study are shown in Supplementary Fig. 2.

tumor tissues. Indeed, the average correlation coefficiency between the normal tissues was higher than that between the tumor tissues (0.54 and 0.43, respectively). In contrast, the average correlation coefficiency of pairs of normal and tumor tissues was low ($r = -0.088$). The differences in the protein expression profiles between the tumor tissues with different histological differentiation, and those with and without lymph node metastases were less obvious in this study.

The similarity of protein expression profiles was also examined by non-hierarchical classification. Principal component analysis defines the directions of maximum variance between the samples and represents the samples in a multidimensional space constructed by the resulting dimensions. The 129 samples were represented in the 3-D space made by the first three major directions generated (Fig. 4B). In the principal component analysis, the normal epithelial tissues were distinguished from the tumor tissues, and the tumor tissues with well-differentiated histology were separated from those with poorly differentiated histology. In addition, the distances between the normal epithelial tissue samples were smaller than those among the tumor samples, suggesting that the protein expression profiles of the normal tissues may be more homogeneous than those of the tumor tissues. The tumor tissues were not divided according to the status of lymph node metastasis or the anatomical site of origin in the multidimensional space (data not shown).

Taken together, the greatest differences were, predictably, observed between the proteomic profile of normal and tumor cells. Histological differentiation appeared to be the second most dominant factor affecting the proteome. Although differences in the lymph node metastasis status affected patient survival significantly, they were not reflected in the overall proteomic profile or individual proteomic characteristics of the cases examined in this study.

3.3 Proteins differentially expressed between tissue groups

Hierarchical clustering analysis suggested the presence of protein spots whose intensity seemed to be different between normal epithelial tissues and tumor tissues (Fig. 3). To identify the proteins associated with carcinogenesis and histological differentiation, we compared protein expression levels between sample groups. We selected spots that corresponded to proteins whose average expression level showed more than twofold differences between sample groups that were statistically significant (Wilcoxon test, p-value <0.01). The comparison between normal and tumor tissue groups resulted in the identification of 338 such proteins, while the comparison of the protein expression profiles of normal tissues with those of highly-, moderately- and poorly differentiated tumors resulted in the selection of 326, 316, and 389 spots, respectively. Many protein spots appeared repeatedly in the comparisons, and as a total, 498 distinct proteins were selected. The number of proteins differentially expressed, including the overlapping ones, is summarized in Supplementary Fig. 3. Of the 498 spots, 221 had more and 227 had less intensity in the tumor tissues compared with normal tissues.

The relations between the fold differences and the number of spots is summarized in Fig. 5A. Regarding the spots whose intensity was increased in the tumor tissue samples, the differences were mostly less than threefold. In contrast, many of the spots with lower intensity in the tumor tissues showed more than threefold difference compared with the normal tissues.

The pair wise similarity of the expression profiles of the 498 selected protein spots was examined in all samples, and the results are summarized in Fig. 5B. The normal and tumor tissues shared similar expression patterns with each other as a whole, the average r value being 0.65 and 0.49, respectively. In contrast, the average r value was only −0.14 in the pair of normal and tumor tissues, clearly showing that the expression pattern of the 498 selected protein spots is significantly different between normal and tumor tissues.

Similarly, principal component analysis based on the expression profiles of the 498 spots also distinguished between the normal and tumor tissues in the multi-dimensional space (Fig. 5C). Within tumor tissues, the well-differentiated tissues were separated from the poorly differentiated ones.

3.4 Self-organizing map of laser-microdissected tissues using 498 selected protein spots

We constructed a self-organizing map that demonstrated that the average spot intensity of the up- or down-regulated proteins, which were categorized in cluster A and B, respectively, was constant in all tumor tissues without obvious correlation with histological differentiation (Figs. 6A and B, upper panel). However, visual inspection of the heat-map suggested that the intensity of some of the 498 spots correlates with histological differentiation (Fig. 6A and B, lower panel). To further examine this point, we created separate self-organizing maps for the 221 and 277 spots with increased and decreased intensity, respectively (Fig. 6C and D, upper panel). We found that the up-regulated protein

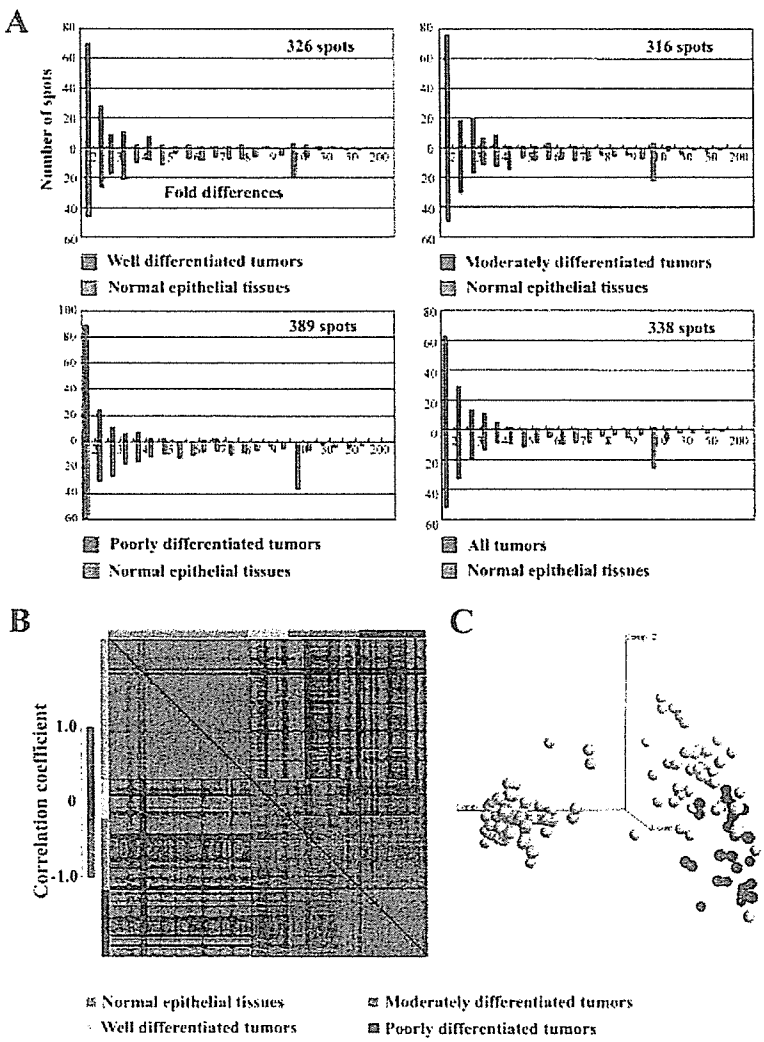

Figure 5. (A) The number of spots showing significantly different degrees of intensity between the normal and tumor tissues. The x-axis represents the fold differences and the y-axis shows the number of spots. The red-coded bars indicate the frequency of the protein spots, which had increased intensity in tumor tissues. The green bars show the frequency of the spots with higher intensity in the normal tissues. (B) The overall similarity of tissue samples was monitored on the basis of the intensity of 498 selected protein spots, showing significant intensity differences between the normal and tumor tissues. (C) Principal component analysis distinguished sample groups on the basis of the expression of 498 selected spots.

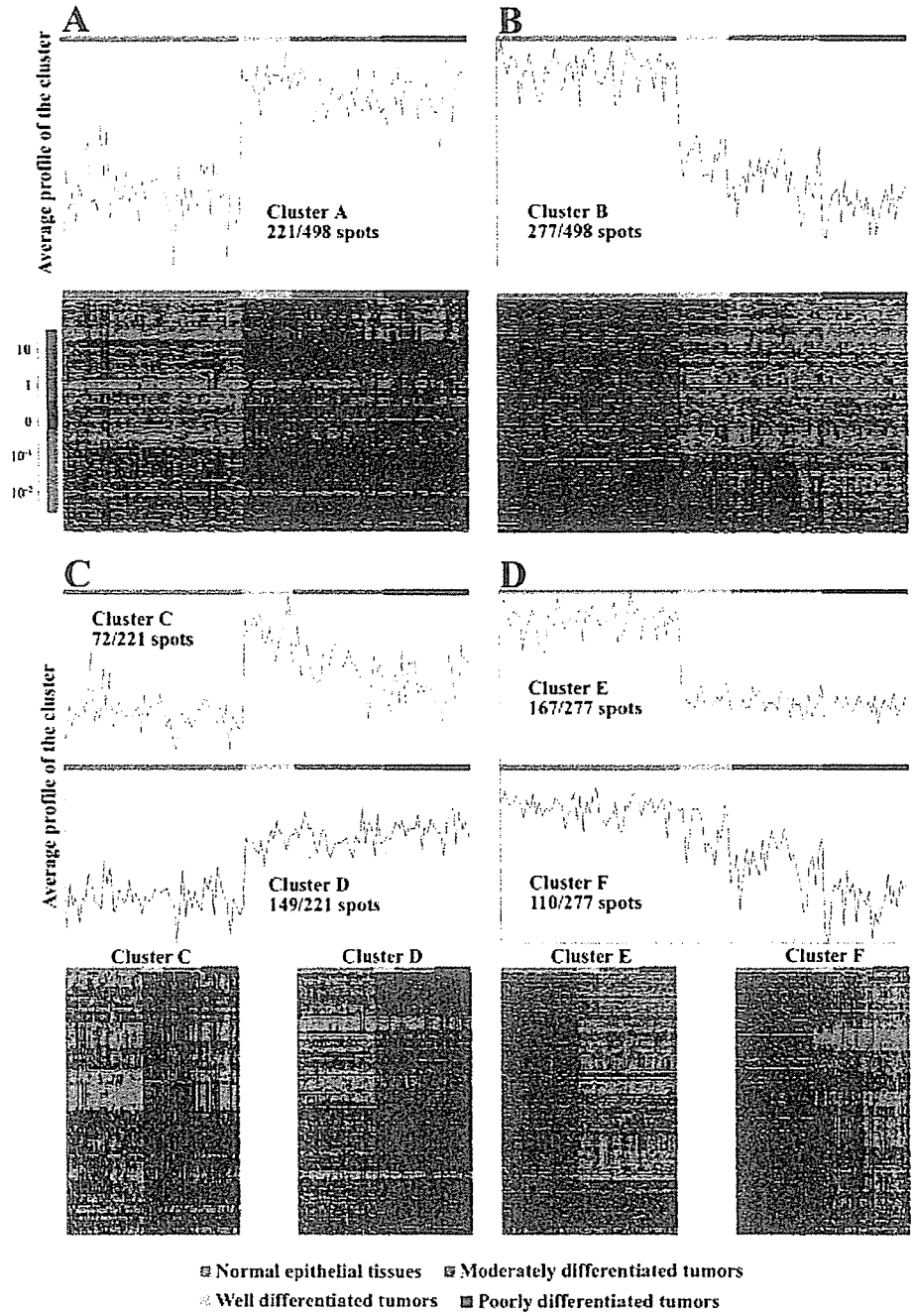

Figure 6. Self-organizing map of the protein spots. The plots in the upper panels of (A) and (B) demonstrate the average profile of the 221 up-regulated (cluster A) and the 277 down-regulated proteins (cluster B) in tumors. Average profiles of the subcluster of the up- and down-regulated proteins in tumors are shown in clusters C and D in the upper panel of (C), and in clusters E and F in the upper panel of (D), respectively. Individual spot intensity of the individual samples is demonstrated as heat-maps in the lower panels of (C) and (D). Rows in the heat-maps represent protein spots, while columns represent individual samples. Tissue types are color-coded. The averaged expression levels of the proteins in cluster A and B show constant up- and down-regulation in tumors without obvious correlation with histological differentiation (A and B). The subclusters, clusters C and F, show the histology-dependent regulation.

spots were divided into two subgroups. Seventy-two spots in cluster C showed notably higher expression levels in the well-differentiated tumor tissues compared with moderately and poorly differentiated ones (Fig. 6C, upper panel). The expression levels of the 72 spots in the individual samples are demonstrated in the heat-map panels (Fig. 6C, lower left panel). In contrast, the average level of the 149 spots in cluster D was constant in the tumor tissues irrespective of histological differentiation (Fig. 6C, upper panel and lower right panel). Proteins down-regulated in tumors were also divided into two clusters, clusters E and F, reflecting the correlation between expression level and histological differentiation. Spot intensity in cluster E was consistently lower in all tumors. Spot intensity in cluster F was higher in well- and moderately differentiated tumors (Fig. 6D, upper panel). Spot intensity in the individual samples is demonstrated in the heat-map panels (Fig. 6D, lower right panel). The localization of the spots in clusters C, D, E and F on representative 2-D gels is shown in Supplementary Figs. 4, 5, 6 and 7, respectively.

The 3-D views of selected spots belonging to clusters C, D, E, and F and varying based on tissue histology are shown in Fig. 7. The 3-D images derived from the Cy5-images. As the intensity of the Cy5-image was standardized based on that of the Cy3-image in the identical gel before statistical analysis, this 3-D view does not always reflect the results of the expression study precisely. However, the different intensities of the spots across the samples were obvious even in the Cy5-images.

3.5 Proteins associated with lymph node metastasis

We examined whether the number of lymph node metastases is associated with patient outcome in our sample set. A previous study correlated nodal metastasis with dismal prognosis in esophageal cancer [31]. The frequency of lymph node metastases in this study is summarized in Table 3, and the data concerning the individual patients are described in Supplementary Table 1. Kaplan-Meier analysis showed that the patients without lymph node metastasis had a significantly higher survival rate compared with the patients with more than five lymph node metastases ($p = 0.003$, Fig. 8). In contrast, patients with one to five lymph node metastases did not show significant differences in terms of survival rate compared with patients without lymph node metastases (Fig. 8). Thus, we assumed that the spots that had distinct intensity between the tumors without nodal metastases and those with more than five lymph node metastases would correspond to proteins that may be tumor marker candidates to predict patient survival, and we examined these differences in more detail. We found 41 such protein spots that had significantly different intensity between the two groups (Wilcoxon test, $p < 0.01$). The localization of these spots on a representative 2-D gel is shown in Supplementary Fig. 8. The 41 spots corresponded to 32 up- and 9 down-regulated proteins in the lymph node positive groups. Expression patterns of selected spots in all tumors are shown in Fig. 9A. By visual inspection, the intensity of the 32 spots corresponding to up-regulated proteins appeared to correlate with the number of lymph node metastases as a whole (cluster G, Fig. 9A). In contrast, the intensity of the 9 spots corresponding to down-regulated proteins was consistent among tumors with four or less lymph node metastases (cluster H, Figs. 9A and B). In principal component analysis, tumor tissues were distinguished from normal tissues on the basis of the expression pattern of the 41 protein spots (Fig. 9C), suggesting that the expression pattern of these proteins changes during the course of carcinogenesis.

3.6 Protein identification by MS and confirmation of the identification using specific antibodies

MS protein identification revealed that the 498 spots showing different intensity between normal and cancer tissues corresponded to proteins generated from 217 distinct genes, and the 41 spots associated with nodal metastasis corresponded to the protein products of 33 distinct genes. The results of identification and data supporting the protein identification are shown in Supplementary Tables 2 and 3.

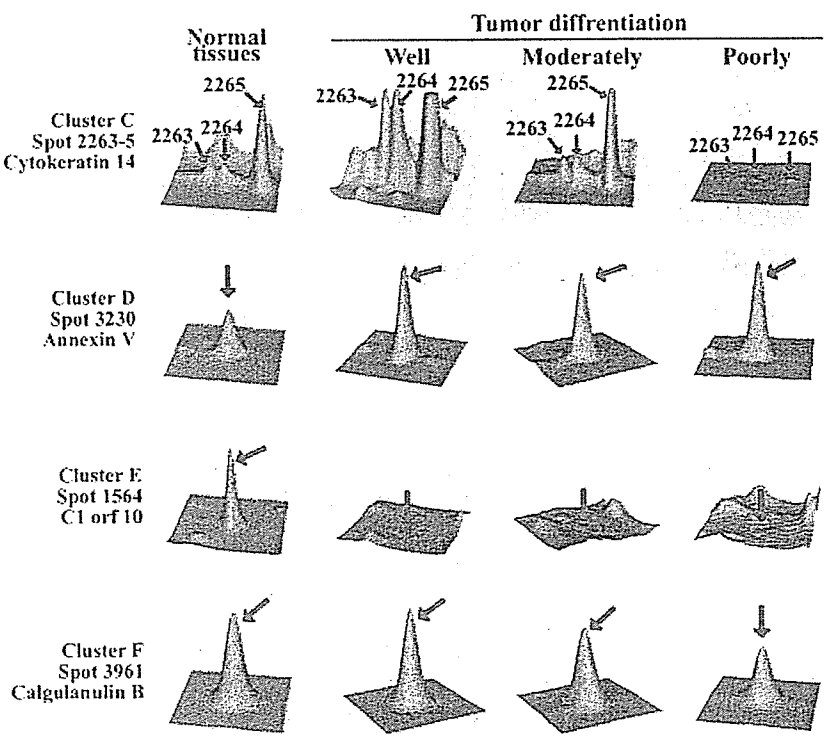

Figure 7. Screen shots using the DeCyder software in BVA mode showing 3-D views of the Cy5-images of representative protein spots in clusters C, D, E and F. Cluster names correspond to those in Fig. 6. The spot numbers correspond to those in Supplementary Figs. 4–7 and Supplementary Tables 2 and 3. The proteins corresponding to the spots were later identified by MS and are demonstrated in the left side of the panels.

Table 3. Frequency of lymph node metastasis in the esophageal cancer cases studied

Number of lymph node metastases	Number of patients
0	17
1	14
2	9
3	9
4 or 5	8
More than 5	15

We validated the results of 2D-DIGE using specific antibodies. We selected the proteins that showed a representative expression pattern for each cluster. The proteins were separated by SDS-PAGE, transferred onto a membrane and incubated with specific antibodies. The intensity of each band was quantified and standardized by that of beta-actin in the same membrane (Fig. 10). A number of different protein spots for cytokeratin 14 were observed in clusters C and D, and cytokeratin 14 expression in Western blotting was similar to that in cluster C (Fig. 10A). This observation may reflect the fact that a higher number of cytokeratin 14 spots with similar molecular weight was present in cluster C (12 spots) than cluster D (2 spots) (Supplementary Table 2). Western blotting showed that periplakin, annexin I and SCCA1/2 expressions were lower in tumor tissues

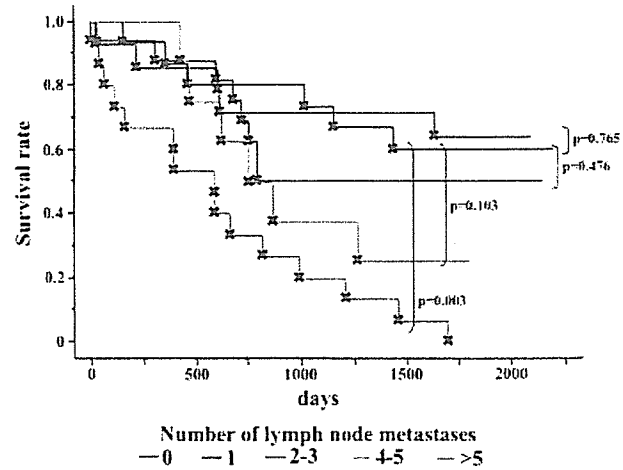

Figure 8. Survival curves of esophageal cancer patients in our sample set, subgrouped according to the number of lymph node metastases present. Survival of the patients with more than five lymph node metastases was significantly shorter than that of patients without lymph node metastases. Survival curves were calculated by the Kaplan–Meier method and statistical differences were calculated by log-rank test.

(Fig. 10A). These results were consistent with those obtained by 2D-DIGE; these proteins were included in cluster E, in which proteins were consistently down-regulated in tumor

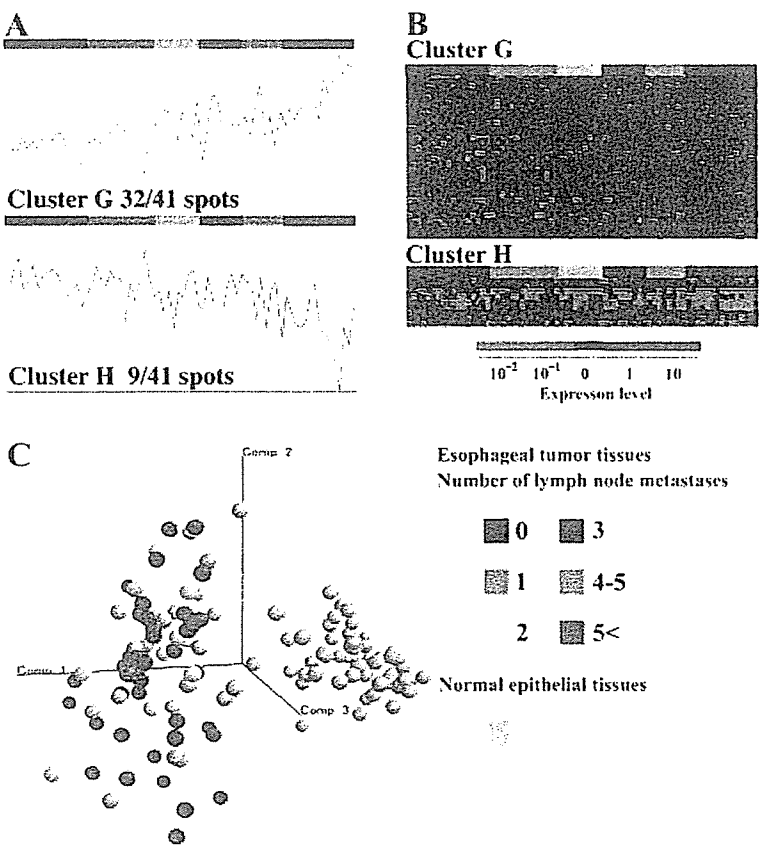

Figure 9. Protein spots associated with lymph node metastasis. Average profile of the 41 protein spots the intensity of which was statistically significantly different between tumors without lymph node metastases and those with more than five lymph node metastases are demonstrated (A). Spots with increased or decreased intensity in the tumor tissues compared with the normal tissues are separately demonstrated as clusters G and H, respectively. The intensity of the 41 protein spots in the individual samples is shown in the heat-map (B). Principal component analysis grouped all tissue samples according to the expression level of the 41 protein spots (C). The localization of the protein spots on a representative 2-D gel is shown in Supplementary Fig. 8.

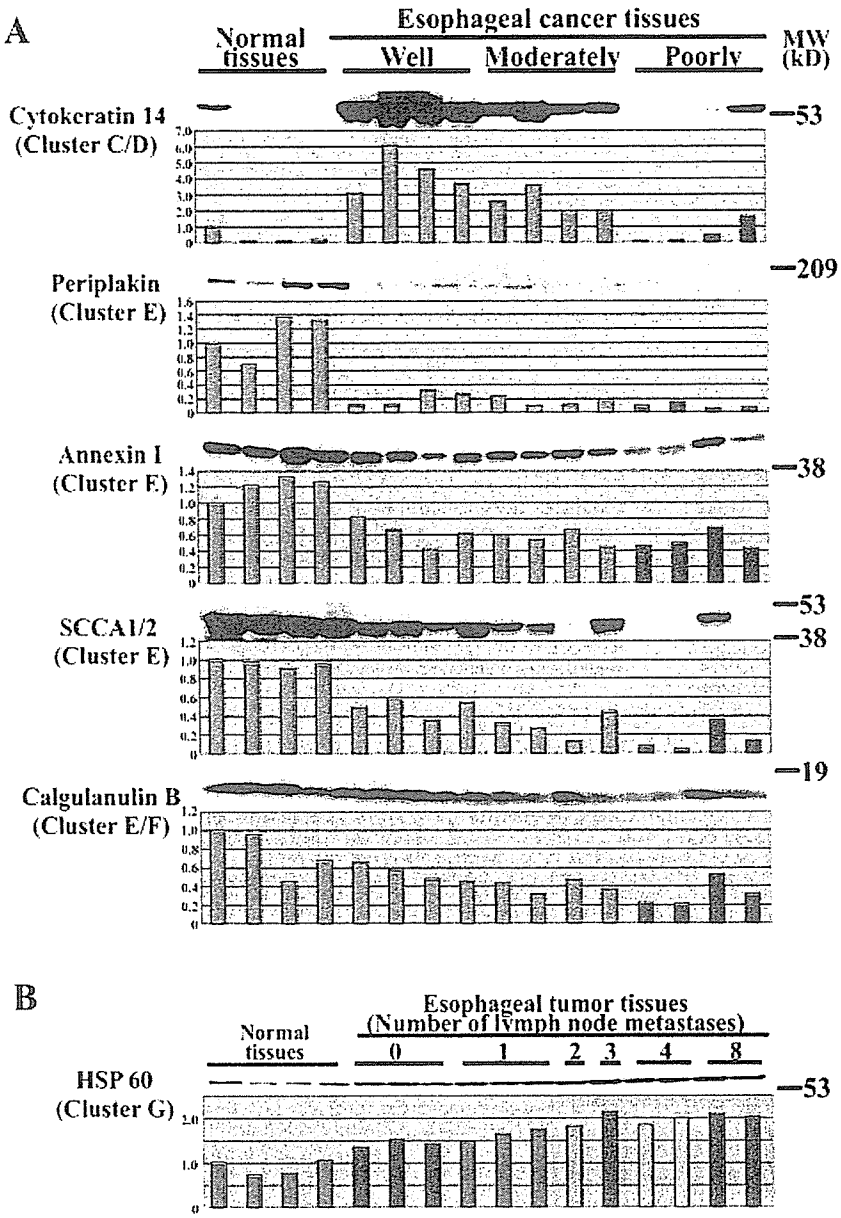

Figure 10. Western blotting results for selected proteins for each cluster validated the 2D-DIGE results. The intensity of each band was quantified and standardized by that of beta-actin in the same membrane (upper panel). The intensity of the bands was standardized by that of beta-actin and compared between the samples (lower panel). (A) The expression levels of selected proteins belonging to clusters C, D, E and F were examined. (B) The expression level of HSP 60 in cluster G was monitored. The color codes correspond to those in Fig. 3. Western blotting successfully validated the results of 2D-DIGE.

tissues irrespective of histological differentiation. Both clusters E and F included two calgulanulin B spots. In Western blotting, although the expression of calgulanulin B was lower in tumor tissues, a correlation between its expression level and histological differentiation was not obvious, probably because Western blotting detected the total amount of calgulanulin B, including expression of its variant which was averaged in the analysis (Fig. 10A). HSP 60 was selected as a representative protein from cluster G; the Western blotting results revealed that the expression of HSP 60 was higher in tumor tissues almost in parallel with the number of lymph node metastases, again being consistent with the results of 2D-DIGE (Fig. 10B).

3.7 Functional characterization of the proteins identified and chromosomal localization of the corresponding genes

We classified the identified proteins based on their function according to their classification in Gene Ontology and the literature curation (Fig. 11). The proteins frequently observed in clusters A and B were categorized as cytoskeletal/structural proteins, transporters, chaperones/heat shock proteins, proteins in the signal transduction pathway, and proteins involved in proteolysis. The transporter proteins appeared more frequently in cluster A than in cluster B. Proteins involved in proteolysis processes were preferentially down-regulated in the tumor tissues.