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.

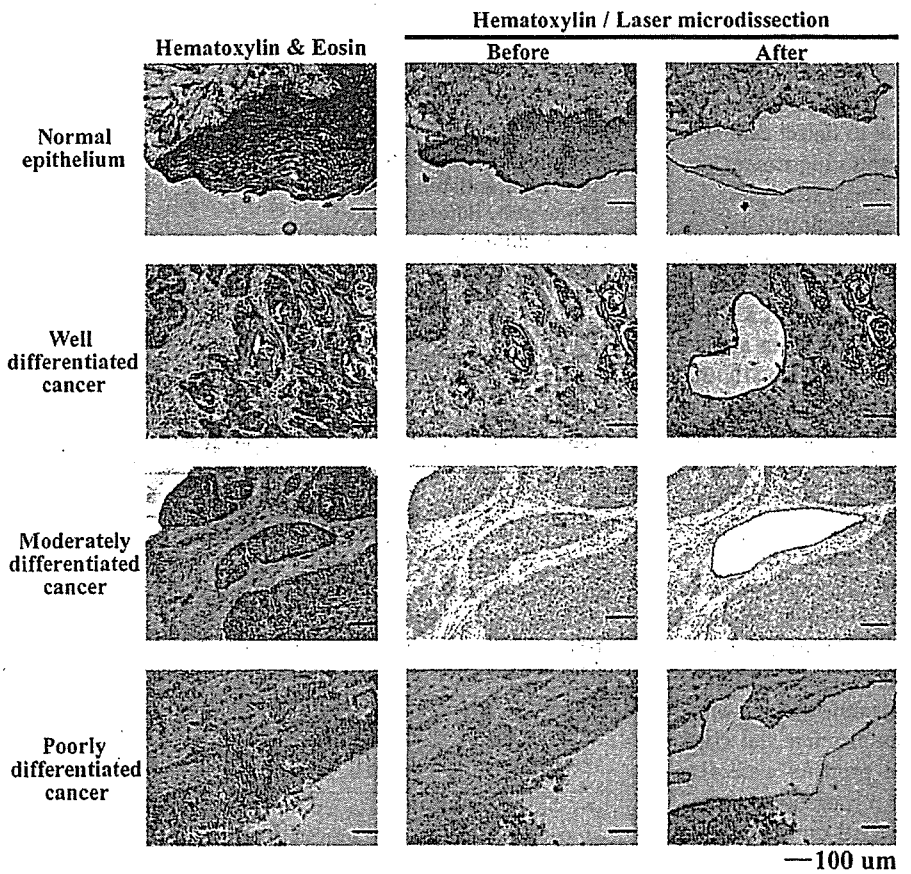

Figure 1. Representative histology of normal esophageal epithelium and esophageal tumor with well, moderate, and poor differentiation. Tissue sections were stained with hematoxylin and eosin for histological observation and hematoxylin alone for the proteomic study. The microscopic appearance of the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

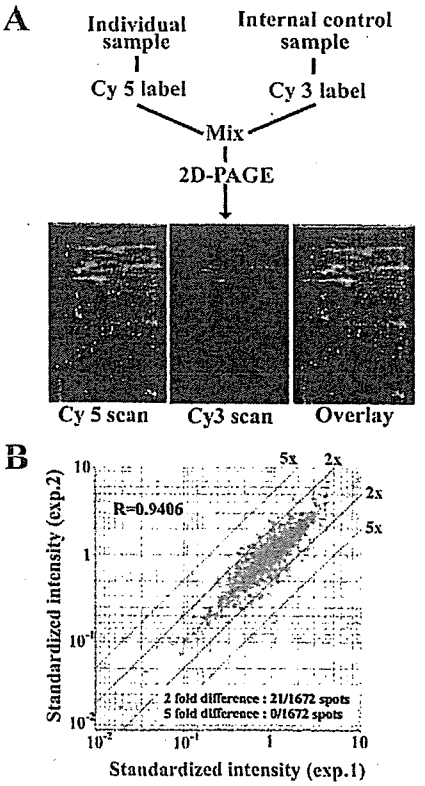

Figure 2. (A) The 2D-DIGE protocol is illustrated. Each individual sample and a pooled reference sample were labeled with Cy5 and Cy3, respectively, mixed, and separated on a 2D-PAGE gel. Gels were scanned with laser, and a set each of Cy3- and Cy5-images was obtained from each gel. (B) The reproducibility 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

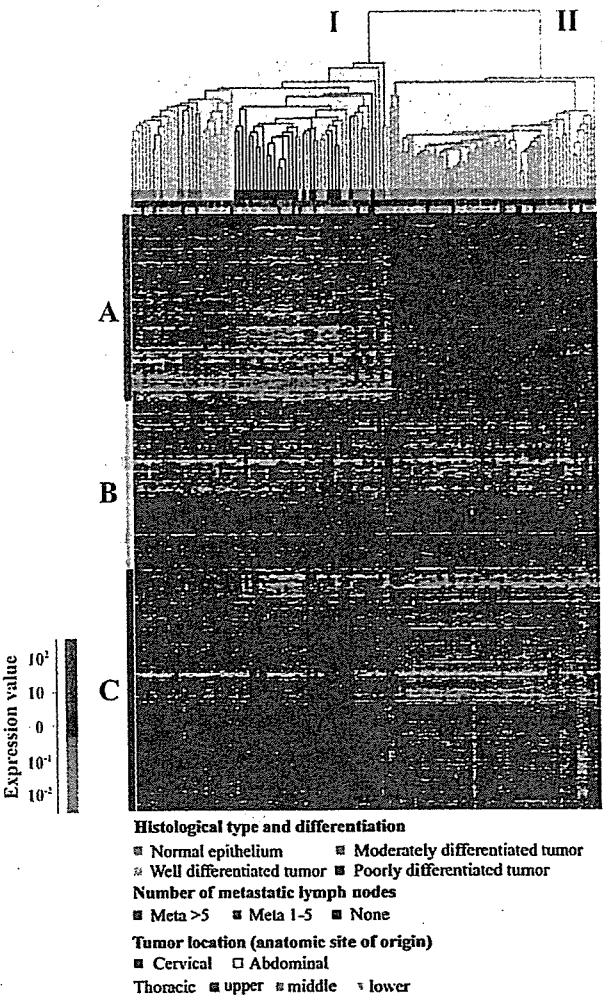

Figure 3. Unsupervised hierarchical clustering classified 129 laser microdissected tissue samples based on the fluorescence intensity of 1730 identified protein spots. The histological differentiation of the samples was demonstrated by the colored nodes as indicted in the panel. Note that the two dominant 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

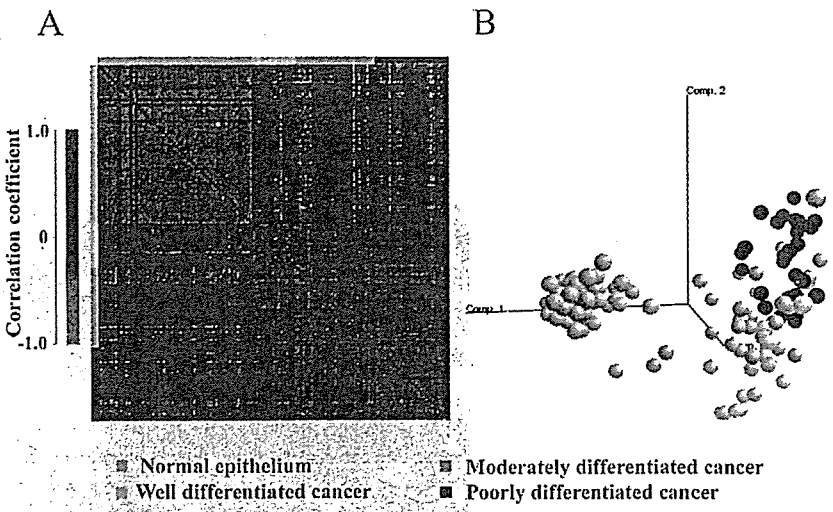

Figure 4. (A) Correlation matrix summarizing the overall similarity of the expression profiles of unselected protein spots for all sample pairs. The red and green colors indicate that the protein expression in the paired samples had high and low similarity, respectively. Samples in the same category, 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

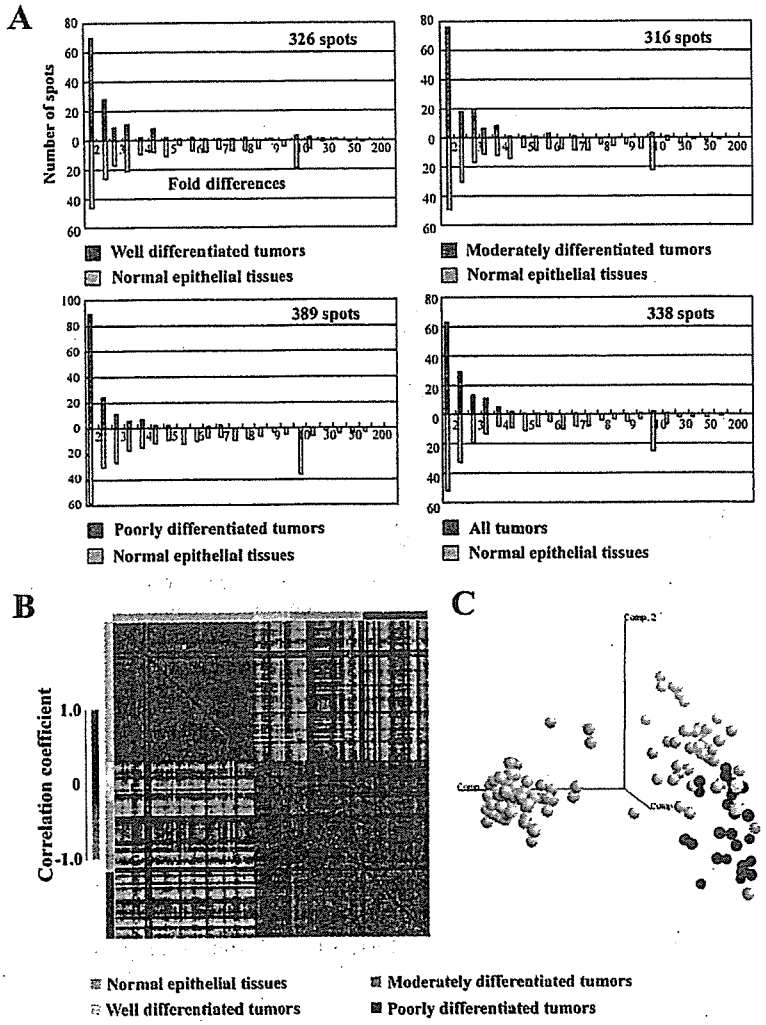

Figure 5. (A) The number of spots showing significantly different degrees of intensity between the normal and tumor tissues. The x-axis represents the fold differences and the y-axis shows the number of spots. The red-coded bars indicate the frequency of the protein spots, which had increased intensity 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.

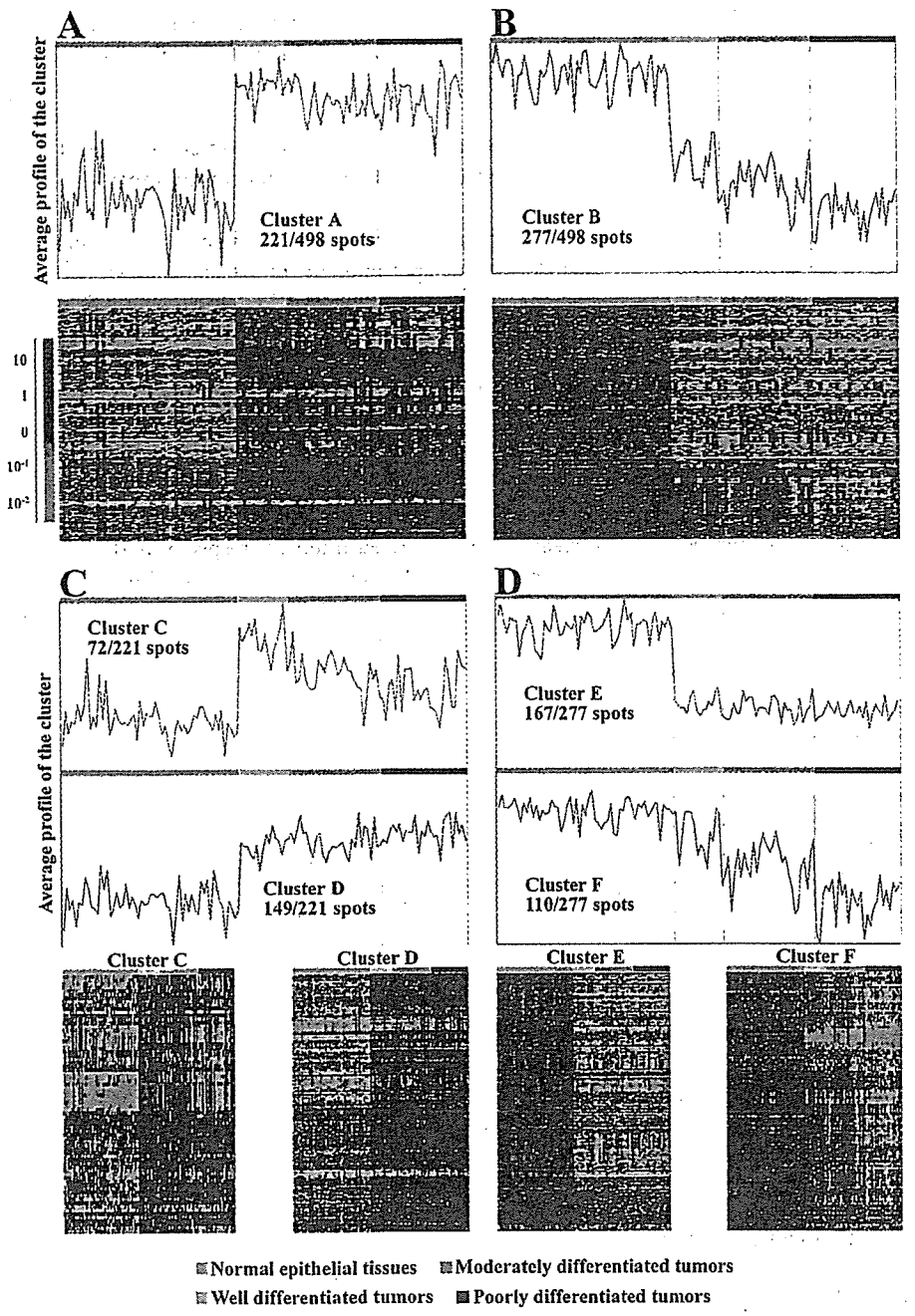

Figure 6. Self-organizing map of the protein spots. The plots in the upper panels of (A) and (B) demonstrate the average profile of the 221 up-regulated (cluster A) and the 277 down-regulated proteins (cluster B) in tumors. Average profiles of the subcluster of the up- and down-regulated proteins in 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.

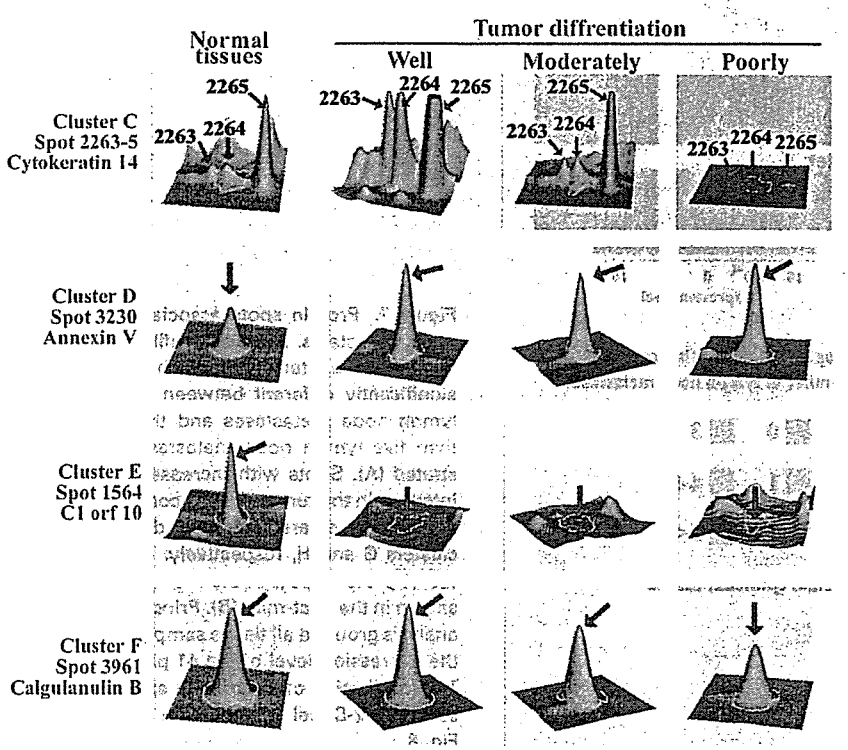

Figure 7. Screen shots using the DeCyder software in BVA mode showing 3-D views of the Cy5-images of representative protein spots in clusters C, D, E and F. Cluster names correspond to those in Fig. 6. The spot numbers correspond to those in Supplementary Figs. 4–7 and Supplementary Tables 2 and 3. The 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

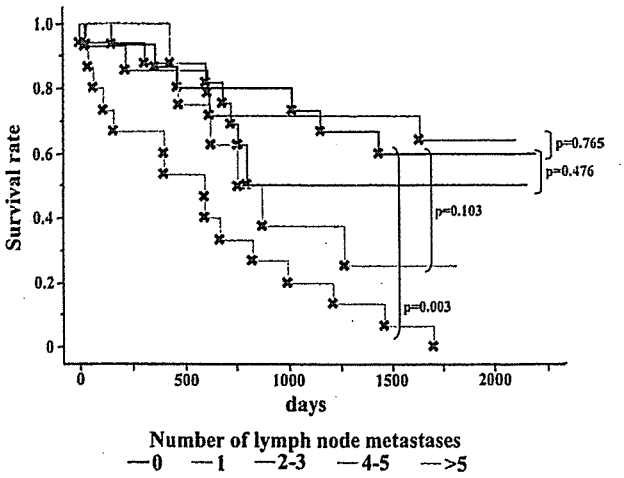

Figure 8. Survival curves of esophageal cancer patients in our sample set, subgrouped according to the number of lymph node metastases present. Survival of the patients with more than five lymph node metastases was significantly shorter than that of patients without lymph node metastases. Survival curves 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

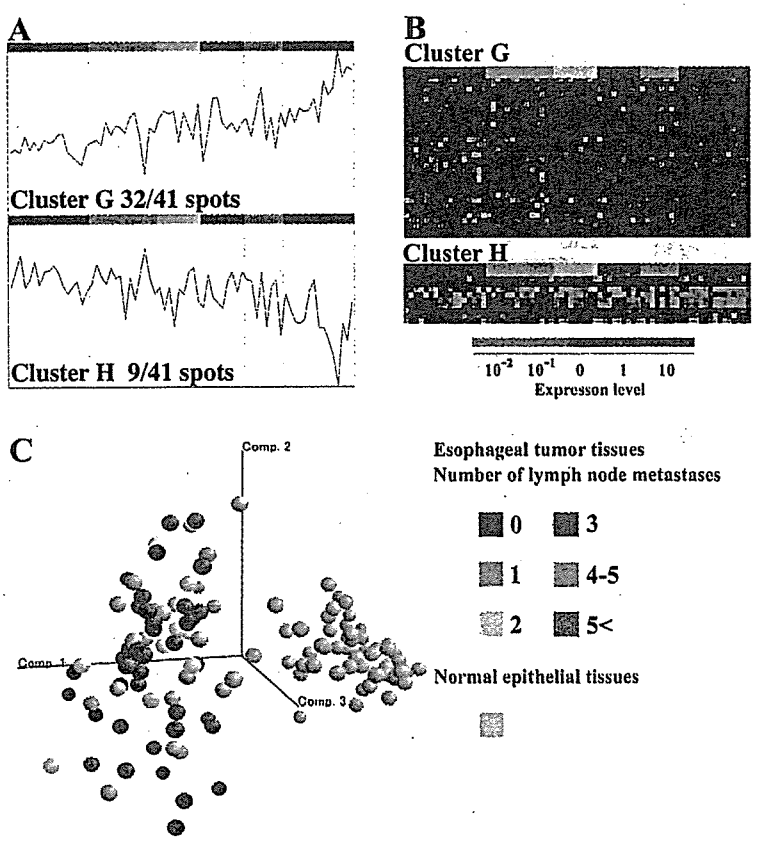

Figure 9. Protein spots associated with lymph node metastasis. Average profile of the 41 protein spots the intensity of which was statistically significantly different between tumors without lymph node metastases and those with more than five lymph node metastases are demonstrated (A). Spots with increased 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.

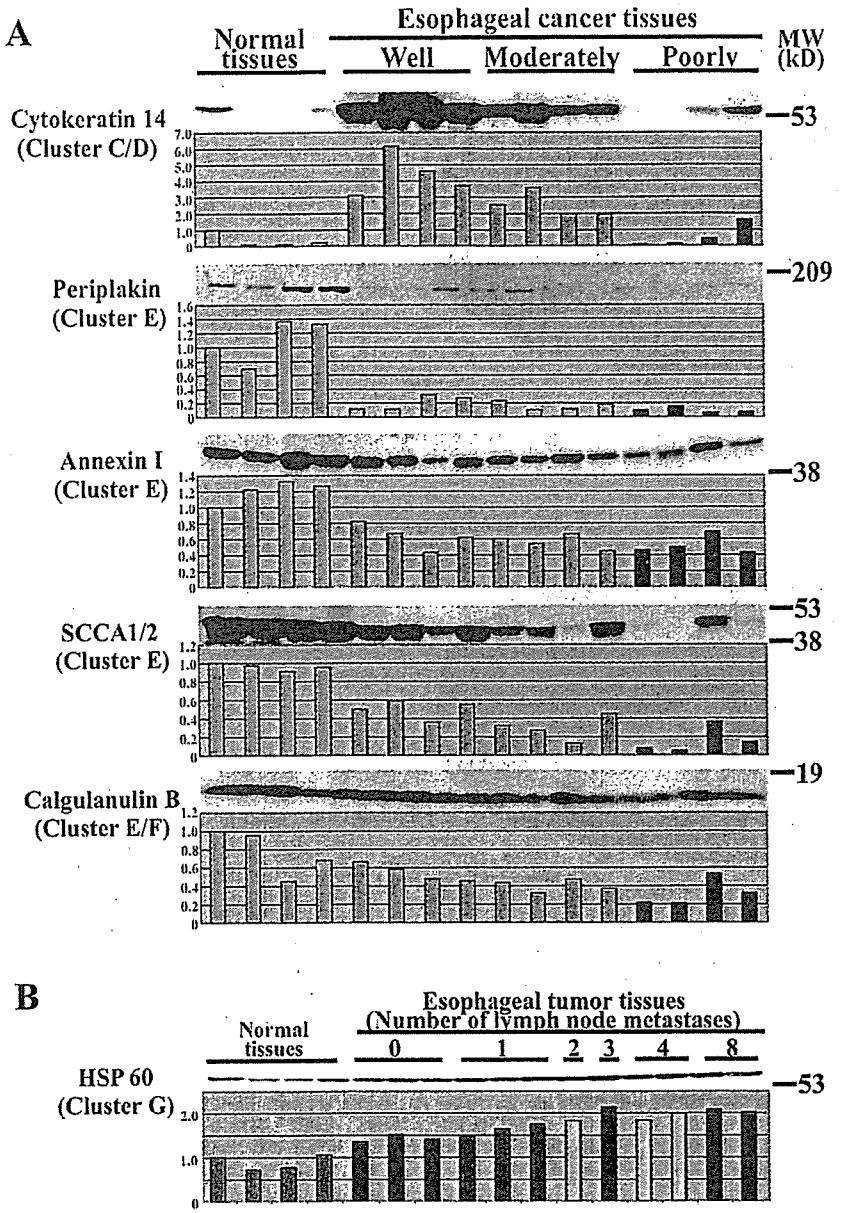

Figure 10. Western blotting results for selected proteins for each cluster validated the 2D-DIGE results. The intensity of each band was quantified and standardized by that of beta-actin in the same membrane (upper panel). The intensity of the bands was standardized by that of beta-actin and compared 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.

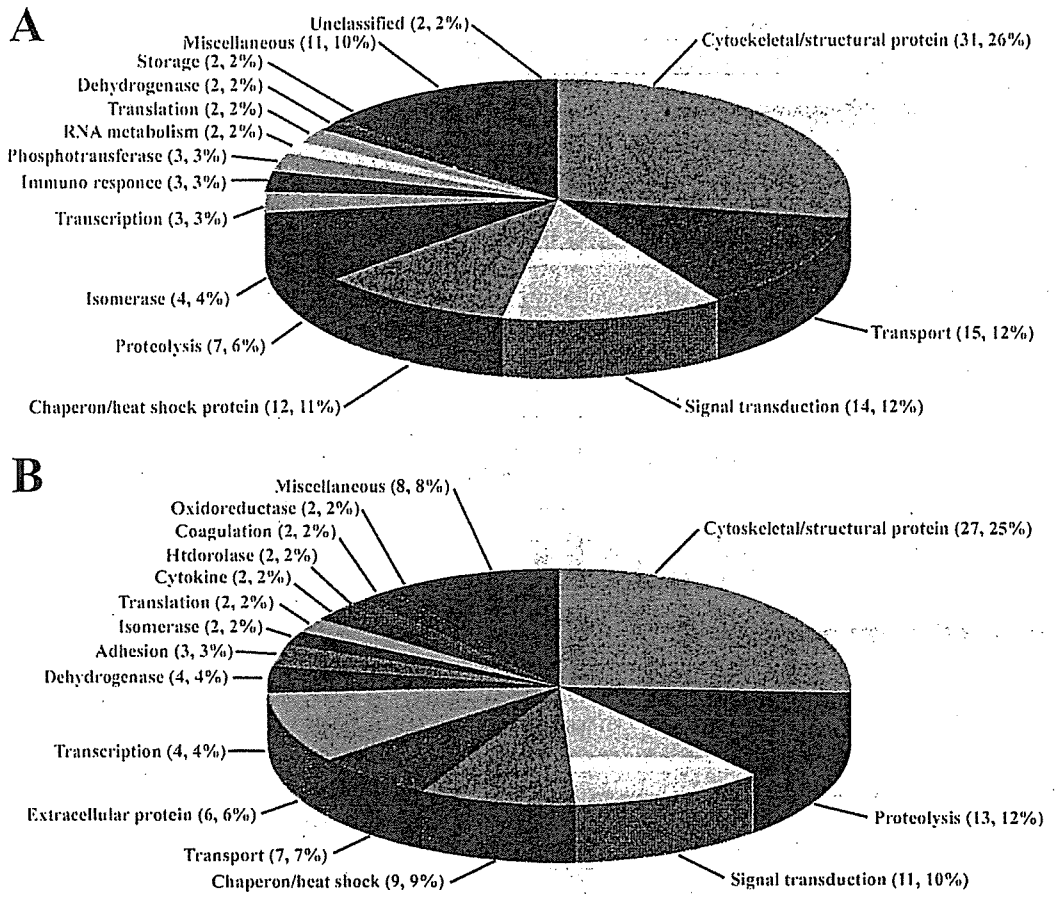

Figure 11. Functional classification of the 217 proteins identified corresponding to the 498 protein spots detected. Proteins function is based on their classification in Gene Ontology and the literature curation. Proteins up- (n = 113) and down- (n = 104) regulated in tumor tissues are shown in panels (A) and (B), respectively. The number of proteins in each group and the percentage are indicated in the parentheses. Categories containing only one protein were combined in the 'miscellaneous' group. The classification is shown in more detail in Supplementary Table 2.

We investigated the chromosomal localization of the genes corresponding to the identified 223 proteins by searching the NCBI database (Supplementary Table 2). We plotted the number of the identified proteins as a function of chromosome location in Supplementary Fig. 8, and found that a relatively large number of the genes identified were located on chromosomes 1, 12, and 17. We also found that only one gene, L-plastin, derived from chromosome 13, and no gene from the Y chromosome, although multiple genes were located on all the other chromosomes.

4 Discussion

In this study, all cases were treated in a consistent manner by a single hospital-based surgeon and no patient received preoperative therapy. In addition, all tumors were staged pathologically according to strict clinicopathological criteria, and complete follow-up and outcomes data were available for all patients. The tumors studied were matched with normal epithelial tissues from the same patients, avoiding differences due to genetic variations. Moreover, the cells were collected using laser microdissection to avoid contamination from other tissue components. The use of this advantageous sample set allowed us to investigate proteomic differences correlating with the clinicopathological parameters.

We first conducted unsupervised classification of the tissue samples to explore which clinicopathological characters most dominantly corresponded to the proteomic data. The proteome of tumors was obviously distinct from that of normal tissues, and corresponded to the histological differentiation of the tumors. The presence or otherwise of nodal metastases and the anatomic site did not seem to affect the global protein expression pattern. These observations may be consistent with our current understanding of cancer biology. During the course of carcinogenesis, a small number of genetic and epigenetic alterations may cause chromosomal instability and drastic changes in entire regions of the ge-

nome. Indeed, global genomic studies using array comparative genomic hybridization technique showed that amplifications might concern entire chromosomal regions in esophageal cancer [32–34]. The proteomic differences between normal and tumor tissues might reflect these genomic alterations. Tumor tissues were separated according to their differentiation status on the basis of their proteomic pattern, because histological differentiation is one of the most dominant factors affecting the morphology and behavior of esophageal tumors. Although the lymph node metastasis status and the tumor location are also clinically important features, tumors could not be distinguished into groups based on these features with histological observation alone. Thus, the overall proteomic profiles of tissues may reflect their major phenotypes, as they appear histologically. Proteomic studies may have the potential to identify the mechanisms underlying histopathological differentiation in tumors.

Another interesting observation in the unsupervised study was that the degree of heterogeneity of protein expression profiles depended on the tissue types; the protein expression profiles of the normal tissues were more homogenous compared with those of the tumor tissues. The proteomic pattern of the normal cells was similar between the different samples, probably because the global protein expression in the normal epithelial cells is strictly regulated. In contrast, the proteomic pattern of the tumor tissues was less homogenous than the normal tissues, given the fact that tumor cells accumulate alterations that are random and/or advantageous for survival during the transformation process. This hypothesis is supported by the observation that the DNA copy number and mRNA expression pattern were more homogenous in the normal mucosa compared with that in colorectal tumor tissues [35].

Of 1730 protein spots studied, 498 showed statistically significant differences in intensity between normal and tumor tissues. MS study revealed that 113 and 104 distinct proteins corresponded to spots with increased and decreased intensity in the tumor tissues, respectively. The proteins in clusters A, C and D, that were up-regulated in the tumor tissues, included several cancer-associated proteins such as alpha-actinin 4 [36], heterogeneous nuclear ribonucleoprotein K (hnRNP K) [37–39]. RNA-binding protein FUS [40], N-myc downstream-regulated gene 1 protein [41], sorcin [42, 43], DNA replication licensing factor MCM7 [44], N-myc interactor [45], cathepsin D [46], annexin A5 [47], annexin A1 [48], breast carcinoma amplified sequence 2 [49], tumor protein D54 [50], tumor protein D52 [51], DJ-1 [52], transgelin-2 [53], galectin-1 [54], and nucleophosmin [55]. The proteins in clusters B, E and F, that were down-regulated in the tumor tissues, included proteins known to be associated with cancer, such as periplakin [16], gelsolin [56], caldesmon [57], src-substrate cortactin [58], squamous cell carcinoma antigen 1 [59], annexin A1 [48], hepatoma-derived growth factor [60], calgranulin B [61], and thioredoxin [62]. These proteins are involved in a variety of biological processes including signal transduction, apoptosis, RNA processing, apoptosis, transcription, DNA replication, multi-drug resistance, vesicle trafficking, cytoskeletal organization, and intracellular redox regulation. Notably, although these proteins have been reported to be aberrantly expressed individually in various cancer types, our comprehensive study enabled a panoramic view of their expression in a single type of cancer, namely esophageal cancer. Although these proteins may function in an aberrant but coordinate manner in esophageal cancer, it is presently difficult to map their functional network. Our findings may, therefore, be used as a basis for biomarker development. Proteins found to be up-regulated in esophageal cancer in our study include prognostic markers that are established in other types of cancer; these were sorcin, implicated in acute myeloid leukemia [42], MCM7 in neuroblastoma [63], cathepsin D in endometroid adenocarcinoma [64], breast carcinoma [65], glioma [66], and colon cancer [67], breast carcinoma amplified sequence 2 in breast cancer [68], gelsolin in lung cancer [69], and galectin-1 in lung cancer [70]. Sorcin is also implicated in multidrug-resistance in leukemia cells [71]. These results prompt future studies aimed at further elucidating the prognostic relevance or otherwise of the above proteins, individually or in combination, to develop improved prognostic biomarkers, which in turn may lead to the establishment of, better therapeutic strategies for esophageal cancer patients.

Lymph node metastasis is one of the critical parameters affecting prognosis. In our study, the number of lymph node metastases correlated with patient survival (Fig. 8). We found 41 protein spots the intensity of which was statistically different between the patients without nodal metastasis and those with more than five lymph node metastases. Although only 4 of the 41 protein spots had significantly different intensity between the normal and tumor tissue groups, principal component analysis distinguished the cancer tissues from the normal ones on the basis of the intensity of the 41 protein spots. These observations suggest that the overall expression pattern of this group of proteins becomes aberrant during the course of carcinogenesis, and changes further during cancer progression. Molecular targeting therapy to revert the aberrant expression of these proteins to their normal status may be effective in preventing lymph node metastasis.

The 41 protein spots associated with nodal metastasis corresponded to 33 distinct proteins. Of these 33 proteins, 10 were identified as corresponding to spots that showed increased or decreased intensity in the tumor tissues compared with their normal counterparts. However, apart from spots 791 and 1584, the protein spots whose intensity differed between normal and tumor tissues were different from those associated with nodal metastasis for the same protein. For example, the intensity of spot 1084, identified as corresponding to gelsolin, was diminished in tumors with more than five lymph node metastases. Spot 1084 did not show significant differences in intensity between normal and tumor tissues, but the intensity of the other gelsolin spots, spots 1223 and 1230, was decreased in the tumor

tissues. Diminished expression of gelsolin has been observed in many types of cancer [56, 72, 73] and considered as a negative prognostic factor [69]. Our observation suggests that identical proteins may contribute to carcinogenesis and cancer progression in varying ways depending on the status of PTM.

Our study revealed that tumors with nodal metastases had higher amount of hnRNP K than those without lymph node metastases. The expression of hnRNP K was also up-regulated in esophageal cancer tissues compared with the normal tissues. The overexpression of hnRNP K has been linked to a range of cancers including breast cancer [38], hepatocellular carcinoma [39], and colorectal cancer [37]. In breast cancer, grade III tumors contained more hnRNP K protein than grade II tumors, suggesting that hnRNP K is involved in the mechanisms of metastasis [38]. The hnRNP K functions as a transcriptional coactivator of p53 [74] and regulates the expression of genes involved in mitogenic responses [75, 76]. Therefore, hnRNP K overexpression may cause various genes to be aberrantly expressed and subsequently contribute to the malignant transformation of normal cells. Previous studies using comparative genomic hybridization revealed that gains of 5p15 correlated with distant organ metastasis in esophageal cancer [77]. As hnRNP K is located in this region, it may be one of the target genes for amplification in esophageal cancer. We also found an inverse correlation between NudC expression levels and nodal metastasis. NudC is a microtubules-associating protein and functions in mitosis and cytokinesis [78]. Lin *et al.* [79] demonstrated that adenovirus expressing NudC inhibited the growth of prostate tumors by blocking cell division in a mouse xenograft model. Both these observations suggest that NudC may potentially be used as a molecular therapy target in esophageal cancer.

Functional classification of the proteins detected in our study demonstrated that more than 60% of the up- and down-regulated proteins in esophageal cancer were categorized as cytoskeletal/structural proteins, transporter proteins, proteins involved in signal transduction, chaperone/heat shock proteins, and proteins controlling proteolysis. Proteins involved in proteolysis were observed more frequently in the down-regulated protein group than in the up-regulated one. These proteins included both proteases and protease inhibitors, and others functioning in the ubiquitin-proteasome pathway. We found that three serine protease inhibitors such as SCCA1, maspin, and serpin B13 were down-regulated in esophageal cancer. All these serpins are located on 18q21 and were down-regulated in head and neck squamous cell carcinoma because of loss of heterozygosity (LOH) [80, 81]. *In vitro* experiments revealed that maspin [82] and SCCA1 [83] are involved in suppression of tumor growth and invasion in head and neck tumors. Taken together, our observations suggest that decreased expression of maspin and SCCA1 in esophageal cancer may promote cancer cell growth, probably due to genomic alterations. The other 17 proteins, which function in proteolysis processes and were aberrantly expressed in tumors in our study, may also play an important role in tumor progression.

Cancer is a disease of the DNA and all proteomic changes are directly or indirectly attributable to genetic and epigenetic lesions. In this proteomic study, although we found possible association of the changes with reported chromosomal abnormalities in a limited number of proteins, the correlation of proteomic and genomic alterations was largely obscure. Of 223 genes, we found that only 1 gene was coded on chromosome 13 and no gene on chromosome Y. It is unlikely that 2D-DIGE allows the observation of gene products only from chromosomes other than chromosome 13 and Y, as 2D-DIGE identifies proteins with reduced cysteine residues with some expression level. We also noted many products of genes on chromosomes 1, 12 and 17. To understand the molecular background of these observations, we may have to identify the proteins for all spots on the 2-D gel used and examine the chromosomal distribution of the genes. Genomic factors including DNA copy number, somatic mutations, single-nucleotide-polymorphisms, and DNA methylation may influence the mRNA and protein expression levels, and a more comprehensive study may be able to explain the mechanisms that lead to proteomic aberrations at the DNA level.

In conclusion, we combined the application of laser microdissection, 2D-DIGE and MS to a large set of well-characterized cases to construct the proteomic profile of esophageal cancer and identified 240 proteins the expression level of which was associated with carcinogenesis, histological differentiation and the number of lymph node metastases. Notably, proteins previously reported to be associated with cancer individually in different types of tumors, were shown to be aberrantly expressed in a single type of malignancy, esophageal cancer. Integrated and comprehensive studies on such proteins will be beneficial for biomarker development and molecular targeting therapy for esophageal cancer. Further studies that will correlate proteomic and genomic alterations will give us clues to understand the mechanisms controlling the overall protein expression pattern in tumor cells, and such understanding will likely lead to novel therapeutic strategies.

This work was supported by a grant from the Ministry of Health, Labour and Welfare and by the Program for Promotion of Fundamental Studies in the National Cancer Institute of Biomedical Innovation of Japan. Hiromitsu Hatakeyama is a recipient of Research Resident Fellowship from the Foundation for Promotion of Cancer Research (Japan) within the framework of the 3rd Term Comprehensive 10-Year Strategy for Cancer Control.

5 References

[1] Parkin, D. M., Pisani, P., Ferlay, J., *Int. J. Cancer* 1999, *80*, 827–841.

[2] Pisani, P., Parkin, D. M., Bray, F., Ferlay, J., *Int. J. Cancer* 1999, *83*, 18–29.

[3] Ilson, D. H., *Cancer Treat. Rev.* 2003, *29*, 525–532.

[4] Enzinger, P. C., Mayer, R. J., *N. Engl. J. Med.* 2003, *349*, 2241–2252.

[5] Mariette, C., Balon, J. M., Piessen, G., Fabre, S. *et al.*, *Cancer* 2003, *97*, 1616–1623.

[6] Mandelker, D. L., Yamashita, K., Tokumaru, Y., Mimori, K. *et al.*, *Cancer Res.* 2005, *65*, 4963–4968.

[7] Gibault, L., Metges, J. P., Conan-Charlet, V., Lozac'h, P. *et al.*, *Br. J. Cancer* 2005, *93*, 107–115.

[8] Nakamura, T., Hayashi, K., Ota, M., Ide, H. *et al.*, *Dis. Esophagus* 2004, *17*, 315–321.

[9] Sharma, R., Chattopadhyay, T. K., Mathur, M., Ralhan, R., *Oncology* 2004, *67*, 300–309.

[10] Tamoto, E., Tada, M., Murakawa, K., Takada, M. *et al.*, *Clin. Cancer Res.* 2004, *10*, 3629–3638.

[11] Kan, T., Shimada, Y., Sato, F., Ito, T. *et al.*, *Ann. Surg. Oncol.* 2004, *11*, 1070–1078.

[12] Gomes, L. I., Esteves, G. H., Carvalho, A. F., Cristo, E. B. *et al.*, *Cancer Res.* 2005, *65*, 7127–7136.

[13] Chen, G., Gharib, T. G., Huang, C. C., Taylor, J. M. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 304–313.

[14] Varambally, S., Yu, J., Laxman, B., Rhodes, D. R. *et al.*, *Cancer Cell* 2005, *8*, 393–406.

[15] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., *Mol. Cell. Biol.* 1999, *19*, 1720–1730.

[16] Nishimori, T., Tomonaga, T., Matsushita, K., Oh-Ishi, M. *et al.*, *Proteomics* 2006, *6*, 1011–1018.

[17] Qi, Y., Chiu, J. F., Wang, L., Kwong, D. L., He, Q. Y., *Proteomics* 2005, *5*, 2960–2971.

[18] Zhang, L. Y., Ying, W. T., Mao, Y. S., He, H. Z. *et al.*, *World J. Gastroenterol.* 2003, *9*, 650–654.

[19] Zhou, G., Li, H., DeCamp, D., Chen, S. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 117–124.

[20] Zhou, G., Li, H., Gong, Y., Zhao, Y. *et al.*, *Proteomics* 2005, *5*, 3814–3821.

[21] Sobin, L. H., Wittekind, C., *TNM Classification of Malignant Tumours*, 6th Edn., John Wiley & Sons, New York 2002.

[22] Gabbert, H. E. S. T., Hainaut, P., Nakamura, Y., Field, J. K., Inoue, H., *Squamous cell carcinoma of the oesophagus*, IARC Press, Lyon 2000, pp. 10–16.

[23] Lieberman, M. D., Shriver, C. D., Bleckner, S., Burt, M., *J. Thorac. Cardiovasc. Surg.* 1995, *109*, 130–139.

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

[25] Yokoo, H., Kondo, T., Fujii, K., Yamada, T. *et al.*, *Hepatology* 2004, *40*, 609–617.

[26] Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q. *et al.*, *Proc. Natl. Acad. Sci. USA* 1999, *96*, 2907–2912.

[27] Kohonen, T., *Self-organizing Maps*, Springer-Verlag, New York 1997.

[28] Kaplan, E. L., Meier, P., *J. Am. Stat. Assoc.* 1958, 457–481.

[29] Okano, T., Kondo, T., Kakisaka, T., Fujii, K. *et al.*, *Proteomics* 2006, *6*, 3938–3948.

[30] Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S. *et al.*, *Proteomics* 2005, *5*, 1150–1159.

[31] Hsu, C. P., Chen, C. Y., Hsia, J. Y., Shai, S. E., *Eur. J. Cardiothorac. Surg.* 2001, *19*, 10–13.

[32] Heiskanen, M. A., Bittner, M. L., Chen, Y., Khan, J. *et al.*, *Cancer Res.* 2000, *60*, 799–802.

[33] Ishizuka, T., Tanabe, C., Sakamoto, H., Aoyagi, K. *et al.*, *Biochem. Biophys. Res. Commun.* 2002, *296*, 152–155.

[34] Tanabe, C., Aoyagi, K., Sakiyama, T., Kohno, T. *et al.*, *Genes Chromosomes Cancer* 2003, *38*, 168–176.

[35] Tsafrir, D., Bacolod, M., Selvanayagam, Z., Tsafrir, I. *et al.*, *Cancer Res.* 2006, *66*, 2129–2137.

[36] Honda, K., Yamada, T., Hayashida, Y., Idogawa, M. *et al.*, *Gastroenterology* 2005, *128*, 51–62.

[37] Klimek-Tomczak, K., Mikula, M., Dzwonek, A., Paziewska, A. *et al.*, *Br. J. Cancer* 2006, *94*, 586–592.

[38] Mandal, M., Vadlamudi, R., Nguyen, D., Wang, R. A. *et al.*, *J. Biol. Chem.* 2001, *276*, 9699–9704.

[39] Li, C., Hong, Y., Tan, Y. X., Zhou, H. *et al.*, *Mol. Cell. Proteomics* 2004, *3*, 399–409.

[40] Perez-Mancera, P. A., Sanchez-Garcia, I., *Semen. Cancer Biol.* 2005, *15*, 206–214.

[41] Wang, Z., Wang, F., Wang, W. Q., Gao, Q. *et al.*, *World J. Gastroenterol.* 2004, *10*, 550–554.

[42] Tan, Y., Li, G., Zhao, C., Wang, J. *et al.*, *Leuk. Res.* 2003, *27*, 125–131.

[43] Parekh, H. K., Deng, H. B., Choudhary, K., Houser, S. R., and Simpkins, H., *Biochem. Pharmacol.* 2002, *63*, 1149–1158.

[44] Ren, B., Yu, G., Tseng, G. C., Cieply, K. *et al.*, *Oncogene* 2006, *25*, 1090–1098.

[45] Li, H., Lee, T. H., Avraham, H., *J. Biol. Chem.* 2002, *277*, 20965–20973.

[46] Sebzda, T., Saleh, Y., Gburek, J., Andrzejak, R. *et al.*, *J. Exp. Ther. Oncol.* 2005, *5*, 145–150.

[47] Carcedo, M. T., Iglesias, J. M., Bances, P., Morgan, R. O., Fernandez, M. P., *Biochem. J.* 2001, *356*, 571–579.

[48] Shen, D., Chang, H. R., Chen, Z., He, J. *et al.*, *Biochem. Biophys. Res. Commun.* 2005, *326*, 218–227.

[49] Nagasaki, K., Maass, N., Manabe, T., Hanzawa, H. *et al.*, *Cancer Lett.* 1999, *140*, 219–226.

[50] Nourse, C. R., Mattei, M. G., Gunning, P., Byrne, J. A., *Biochim. Biophys. Acta* 1998, *1443*, 155–168.

[51] Byrne, J. A., Tomasetto, C., Garnier, J. M., Rouyer, N. *et al.*, *Cancer Res.* 1995, *55*, 2896–2903.

[52] Le Naour, F., Misek, D. E., Krause, M. C., Deneux, L. *et al.*, *Clin. Cancer Res.* 2001, *7*, 3328–3335.

[53] Shi, Y. Y., Wang, H. C., Yin, Y. H., Sun, W. S. *et al.*, *Br. J. Cancer* 2005, *92*, 929–934.

[54] Kayser, K., Hauck, E., Andre, S., Bovin, N. V. *et al.*, *Anticancer Res.* 2001, *21*, 1439–1444.

[55] Boissel, N., Renneville, A., Biggio, V., Philippe, N. *et al.*, *Blood* 2005, *106*, 3618–3620.

[56] Tanaka, M., Mullauer, L., Ogiso, Y., Fujita, H. *et al.*, *Cancer Res.* 1995, *55*, 3228–3232.

[57] Watanabe, K., Kusakabe, T., Hoshi, N., Saito, A., Suzuki, T., *Hum. Pathol.* 1999, *30*, 392–396.

[58] Lagarkova, M. A., Boitchenko, V. E., Mescheryakov, A. A., Kashkarova, U. A., Nedospasov, S. A., *Oncogene* 2000, *19*, 5204–5207.

[59] Kano, M., Shimada, Y., Kaganoi, J., Sakurai, T. *et al.*, *Br. J. Cancer* 2000, *82*, 429–435.

[60] Yamamoto, S., Tomita, Y., Hoshida, Y., Takiguchi, S. *et al.*, *Clin. Cancer Res.* 2006, *12*, 117–122.

[61] Hermani, A., Hess, J., De Servi, B., Medunjanin, S. *et al.*, *Clin. Cancer Res.* 2005, *11*, 5146–5152.

[62] Biaglow, J. E., Miller, R. A., *Cancer Biol. Ther.* 2005, *4*, 6–13.

[63] Vasudevan, S. A., Nuchtern, J. G., Shohet, J. M., *World J. Surg.* 2005, *29*, 317–324.

[64] Dvalishvili, I., Charkviani, L., Charkviani, T., Turashvili, G., Burkadze, G., *Georgian Med. News* 2005, 27–31.

[65] Nikolic-Vukosavljevic, D., Markicevic, M., Grujic-Adanja, G., Petrovic, A. *et al.*, *Clin. Exp. Metastasis* 2005, *22*, 363–368.

[66] Fukuda, M. E., Iwadate, Y., Machida, T., Hiwasa, T. *et al.*, *Cancer Res.* 2005, *65*, 5190–5194.

[67] Sis, B., Sagol, O., Kupelioglu, A., Sokmen, S. *et al.*, *Pathol. Res. Pract.* 2004, *200*, 379–387.

[68] Regitnig, P., Moser, R., Thalhammer, M., Luschin-Ebengreuth, G.. *et al.*, *J. Pathol.* 2002, *198*, 190–197.

[69] Yang, J., Tan, D., Asch, H. L., Swede, H. *et al.*, *Lung Cancer* 2004, *46*, 29–42.

[70] Szoke, T., Kayser, K., Baumhakel, J. D., Trojan, I. *et al.*, *Oncology* 2005, *69*, 167–174.

[71] Zhou, Y., Xu, Y., Tan, Y., Qi, J. *et al.*, *Leuk. Res.* 2006, *30*, 469–476.

[73] Asch, H. L., Winston, J. S., Edge, S. B., Stomper, P. C., Asch, B. B., *Breast Cancer Res. Treat.* 1999, *55*, 179–188.

[74] Lee, H. K., Driscoll, D., Asch, H., Asch, B., Zhang, P. J., *Prostate* 1999, *40*, 14–19.

[75] Moumen, A., Masterson, P., O'Connor, M. J., Jackson, S. P., *Cell* 2005, *123*, 1065–1078.

[76] Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I., Levens, D., *Mol. Cell. Biol.* 1996, *16*, 2350–2360.

[77] Ostrowski, J., Kawata, Y., Schullery, D. S., Denisenko, O. N., Bomsztyk, K., *Nucleic Acids Res.* 2003, *31*, 3954–3962.

[78] Ueno, T., Tangoku, A., Yoshino, S., Abe, T. *et al.*, *Clin. Cancer Res.* 2002, *8*, 526–533.

[79] Aumais, J. P., Williams, S. N., Luo, W., Nishino, M. *et al.*, *J. Cell. Sci.* 2003, *116*, 1991–2003.

[80] Lin, S. H., Nishino, M., Luo, W., Aumais, J. P. *et al.*, *Oncogene* 2004, *23*, 2499–2506.

[81] Schneider, S. S., Schick, C., Fish, K. E., Miller, E. *et al.*, *Proc. Natl. Acad. Sci. USA* 1995, *92*, 3147–3151.

[82] Nakashima, T., Pak, S. C., Silverman, G. A., Spring, P. M. *et al.*, *Biochim. Biophys. Acta* 2000, *1492*, 441–446.

[83] Yasumatsu, R., Nakashima, T., Hirakawa, N., Kumamoto, Y. *et al.*, *Head Neck* 2001, *23*, 962–966.

[84] Nakashima, T., Yasumatsu, R., Kuratomi, Y., Masuda, M. *et al.*, *Head Neck* 2006, *28*, 24–30.

Int. J. Cancer: **119**, 1491–1494 (2006)

SHORT REPORT

Frequent *EGFR* mutations in brain metastases of lung adenocarcinoma

Shingo Matsumoto[1,2], Kenji Takahashi[1], Reika Iwakawa[1], Yoshihiro Matsuno[3], Yukihiro Nakanishi[4], Takashi Kohno[1], Eiji Shimizu[2] and Jun Yokota[1]*

[1]*Biology Division, National Cancer Center Research Institute, Tokyo, Japan*
[2]*Division of Medical Oncology and Molecular Respirology, Faculty of Medicine, Tottori University, Tottori, Japan*
[3]*Diagnostic Pathology Division, National Cancer Center Hospital, Tokyo, Japan*
[4]*Pathology Division, National Cancer Center Research Institute, Tokyo, Japan*

Lung adenocarcinomas often metastasize to the brain, and the prognosis of patients with brain metastases is still very poor. The *epidermal growth factor receptor* (*EGFR*) gene is mutated in a considerable fraction of primary lung adenocarcinomas, in particular those with drastic response to EGFR tyrosine kinase inhibitors. The present study was designed to elucidate the prevalence of *EGFR* mutations in brain metastases and the timing of their occurrence during cancer progression. *EGFR* mutations were detected in 12 of 19 metastatic lung adenocarcinomas to the brain (63%). This frequency was higher than those in previous studies for *EGFR* mutations at various stages of lung adenocarcinoma in East Asia, including Japan (*i.e.*, 20–55%). In 6 cases with *EGFR* mutations, the corresponding primary lung tumors were also examined for the mutations, and in all of them, the same types of *EGFR* mutations were detected also in the primary tumors. In 2 of them, second metastatic brain tumors in addition to the first ones were also available for analysis, and the same types of *EGFR* mutations were detected in both the first and second ones in both cases. These results indicate that *EGFR* mutations are present frequently in brain metastases and occur preceding brain metastasis. These findings will be highly informative for treatment of metastatic lung adenocarcinoma to the brain.

Key words: *EGFR* mutation; metastatic brain tumor; lung adenocarcinoma

Epidermal growth factor receptor (EGFR) is a member of a family comprised of 4 homologous receptors, EGFR (ERBB1), HER-2/neu (ERBB2), HER-3 (ERBB3) and HER-4 (ERBB4). Ligand binding to EGFR leads to receptor tyrosine kinase (TK) activation and a series of downstream signaling activation that mediates proliferation, migration, invasion and suppression of apoptosis.[1] Recently, it was revealed that most lung adenocarcinoma patients who were responsive to gefitinib, an EGFR TK inhibitor, had somatic mutations in the kinase domain of the *EGFR* gene in their tumor cells.[2,3] Subsequently, it was reported that *EGFR* mutations are present in a considerable fraction of lung adenocarcinoma and occur more frequently in East Asian patients, including Japanese, than in Caucasian patients.[4–10] Furthermore, the incidence of *EGFR* mutations was significantly high in female patients and patients without smoking histories. In our previous study, *EGFR* mutations were detected frequently in noninvasive bronchioloalveolar carcinomas, suggesting that *EGFR* mutations occur early in the development of adenocarcinoma, and those with the mutations further progress to invasive and metastatic carcinomas.[11] However, to our knowledge, there is no report showing the prevalence of *EGFR* mutations in metastatic lung adenocarcinomas, and thus, it remains unclear whether *EGFR* mutations are indeed retained in metastatic lung adenocarcinoma or not. Elucidation of this issue will be implicative for treatment with EGFR TK inhibitors against advanced lung adenocarcinomas, which often metastasize systemically to diverse sites, such as brain, bone, adrenal glands and liver.[12] Therefore, we examined metastatic lung adenocarcinomas to the brain for *EGFR* mutations and compared the mutation status in the metastatic brain tumors with the corresponding primary tumors, if they were available for the analysis. We also examined these tumors for *KRAS* mutations, which have been reported as being mutually exclusive for *EGFR* mutations.[4,7]

A total of 21 metastatic brain tumor tissues were obtained from 19 patients who were treated during the period from 1986 to 2001 at the National Cancer Center Hospital, Tokyo, Japan. These tumor tissues were obtained at surgery or at autopsy. In 2 of the 19 cases, the second brain surgery was performed against the second recurrence in the brain 15 months and 24 months after the first brain surgery, respectively, and thus brain tumor tissues were obtained twice during their clinical courses. In 8 of the 19 cases, the corresponding primary lung tumors were obtained at lung surgery preceded by brain surgery. None of the patients were treated with gefitinib through all clinical courses. In 16 of the 19 cases, primary and metastatic tumors were macrodissected and were subjected to genomic DNA extraction by the method described previously.[13] In the remaining 3 cases, from which 3 primary tumors and 5 metastatic tumors were obtained, cancer cells were microdissected using the Pixcell Laser Capture Microdissection system (Arcturus Engineering, Mountain View, CA). Their genomic DNAs were isolated by SDS/proteinase K digestion and phenol/chloroform extraction as described previously.[14] Exons 18–21 of the *EGFR* gene and exons 1 and 2 of the *KRAS* gene were examined for mutations by genomic PCR amplification and direct sequencing of PCR products. PCR primer sequences and PCR conditions are described previously.[11]

Table I shows the result of all cases examined. In 12 of the 19 cases (63%), *EGFR* mutations were detected in their metastatic brain tumors. In 2 cases (cases 2 and 3), for which both the first and second metastatic brain tumors were available for the analysis, *EGFR* mutations were detected in both the metastatic tumors, and the type of mutation in the second metastatic tumor was the same as that in the first one in both cases. In 8 cases, for which primary lung tumors were also available for the analysis, 6 cases had *EGFR* mutations in their primary tumors. All metastatic tumors from the 6 cases with *EGFR* mutations in the primary tumors had the same types of mutations as those in the primary tumors, respectively. Figure 1 shows representative sequence chromatograms of case 1, for which cancer cells of primary and metastatic tumors were collected by microdissection. There was no case that *EGFR* mutations were not detected in primary tumors and were detected in the corresponding metastatic brain tumors. Additionally, there was also no case that *EGFR* mutations were detected only in primary tumors and were not detected in the corresponding metastatic brain tumors. Thus, it was shown that *EGFR* mutations in primary adenocarcinomas are retained in their metastatic brain tumors.

The first two authors contributed equally to this work.

Grant sponsors: Ministry of Health, Labor and Welfare; National Institute of Biomedical Innovation (NiBio).

***Correspondence to:** Biology Division, National Cancer Center Research Institute, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan. Fax: +81-3-3542-0807. E-mail: jyokota@gan2.ncc.go.jp

Received 21 October 2005; Accepted 7 February 2006
DOI 10.1002/ijc.21940
Published online 26 April 2006 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – *EGFR* AND *KRAS* MUTATIONS IN LUNG ADENOCARCINOMA WITH BRAN METASTASIS

Case		Clinical characteristics					Mutation		
No.	Tumor	Age	Gender	Smoking	Stage[1]	Interval time (months)[2]	Gene	Exon	Amino acid change
1	P, M	48	F	–	IIIA	26	*EGFR*	19	E746-A750 del
2	P, M1, M2	43	F	–	IIA	M1, 18; M2, 33	*EGFR*	19	E746-E749 del, AT750-751VA
3	P, M1, M2	59	M	+	IIB	M1, 40; M2, 64	*EGFR*	19	L747-T751 del
4	P, M	49	F	–	IIIB	0.5	*EGFR*	19	E746-A750 del
5	P, M	54	M	+	IIIB	6	*EGFR*	19	E746-A750 del
6	P, M	70	M	+	IIIA	14	*EGFR*	21	L858R
7	P, M	64	M	+	IB	16			
8	P, M	58	M	+	IIIA	22			
9	M	56	M	+			*EGFR*	19	E746-A750 del
10	M	51	F	–			*EGFR*	19	E746-A750 del
11	M	56	F	–			*EGFR*	19	E746-A750 del
12	M	49	M	+			*EGFR*	19	E746-A750 del, T751A
13	M	59	M	+			*EGFR*	19	E746-T751del, SP752-753IS
14	M	48	F	–			*EGFR*	21	L858R
15	M	53	M	+			*KRAS*	1	G12C
16	M	59	M	+			*KRAS*	1	G12C
17	M	62	M	+					
18	M	74	M	+					
19	M	67	M	+					

P, primary tumor; M, metastatic brain tumor; M1, first metastatic brain tumor; M2, second metastatic brain tumor.
[1]Pathological stage according to the TNM classification at the time of lung surgery for the primary tumor.–[2]Interval time from lung surgery to brain surgery.

Types of *EGFR* mutations detected in the present study were 10 in-frame deletions (83%) in exon 19 and 2 point mutations (17%) in exon 21. The most frequent mutation was a simple deletion of 5 amino acid residues from codon 746 to 750 (6/12, 50%). Both of the 2 point mutations were the leucine to arginine mutation at codon 858 (L858R). These 2 types of mutation are known to be the most common ones in lung cancer, especially in lung adenocarcioma. The remaining types of *EGFR* mutations detected were a simple deletion of 5 amino acid residues from codon 747 to 751 and 3 deletions coupled with 1 or 2 amino acid substitutions. No mutation was detected in exons 18 and 20 in the present study, although several point mutations and in-frame insertions have been identified in those exons in primary lung adenocarcinomas in previous studies.[4,7]

KRAS mutations were detected in 2 of 7 metastatic brain tumors without *EGFR* mutations. Both of the 2 mutations were the glycine to cysteine mutation at codon 12 (G12C). We also analyzed the association of *EGFR* mutations with clinicopathological characteristics, such as age, gender and smoking history (Table I). All female patients, who were never-smokers, had *EGFR* mutations in their tumors and the mutations were significantly more frequent in female patients (6/6, 100%) than in male patients (6/13, 46%) (Fisher's exact test, $p = 0.0436$). Therefore, the mutual exclusiveness of *EGFR* and *KRAS* mutations as well as frequent mutations in female nonsmokers was consistent with previous findings.[4,6–8,10]

We demonstrated here that *EGFR* mutations were frequently present in metastatic brain tumors of lung adenocarcinoma. In previous studies for *EGFR* mutations in various stages of lung adenocarcinomas in East Asia, the frequency of the mutations were 20–55%.[4,5,8,15] The higher incidence of *EGFR* mutations in our study raises a possibility that the mutations may be associated with metastasis of lung adenocarcinoma. In recent studies, it was suggested that *EGFR* mutations occur early in the development of lung adenocarcinoma.[11,16–18] Yoshida *et al.*[17] showed that *EGFR* mutations were present in 3% of atypical adenomatous hyperplasia (AAH), which is considered to be a precursor lesion of lung adenocarcinoma, and the presence of the mutations was increasingly frequent during sequential progression from AAH to invasive adenocarcinoma through bronchioloalveolar carcinoma (BAC). Our previous study also demonstrated that a majority of BACs had *EGFR* mutations.[11] Moreover, Tang *et al.*[18] reported that *EGFR* mutations identical to those in tumors were present in the histologically normal respiratory epithelium in 9 of 21 patients with lung adenocarcinoma carrying *EGFR* mutations in the tumors. In the present study, we also showed that all 6 cases with *EGFR* mutations in their metastatic brain tumors had the identical mutations in the corresponding primary tumors. Thus, *EGFR* mutations are likely to be an early genetic alteration in multistage carcinogenic processes of lung adenocarcinoma, and additional genetic alterations responsible for brain metastasis may occur in cancer cells with *EGFR* mutations. Indeed, our previous study, using the same samples as those from 16 of the 19 cases, showed the sequential accumulation of allelic losses during tumor progression. In particular, in case 6, which has *EGFR* mutations in both the primary and metastatic tumors, loss of heterozygosity (LOH) on chromosomes 2q, 13q and 18q was shown to accumulate during tumor progression.[19] In the present study, we showed that *EGFR* mutations, which had been present in the primary tumors at stages IIB-IIIB, were all retained in their brain metastases. Although *EGFR* mutations may be associated with the genesis and/or early progression of lung adenocarcinoma, our results showing the retention and frequent presence of the mutations in metastatic brain tumors indicate that lung adenocarcinomas with *EGFR* mutations may also have a higher potential of metastasizing to the brain. For this reason, it should be noted that *EGFR* gene alterations occur frequently in gliomas, a common brain tumor.[20,21] However, since there is no information on the prevalence of *EGFR* mutations in metastatic lung adenocarcinomas to sites other than the brain at present, further studies will be needed to clarify the association of *EGFR* mutations with metastatic sites of lung adenocarcinoma.

The brain is one of the most frequent metastatic sites of lung adenocarcinoma. Since traditional chemotherapy is not so effective against metastatic brain tumors of lung adenocarcinoma, radiotherapy is, to date, the main treatment for patients with them. Nevertheless, the prognosis of those patients is poor and median survival is only 3–6 months.[22] Interestingly, recent reports demonstrated that metastatic brain tumors of lung adenocarcinoma frequently and drastically responded to gefitinib, an EGFR TK inhibitor,[23–25] although *EGFR* mutations were not examined in these cases. Those reports could have an impact on treatment for metastatic brain tumors of lung adenocarcinoma; however, the role of gefitinib in therapeutic strategies against metastatic lung adenocarcinoma has not yet been established. This is the first report, to our knowledge, to analyze a considerable number of metastatic brain tumors of lung adenocarcinoma for *EGFR* mutations and to demonstrate the frequent presence of them. These results will provide us with a rationale for the use of this drug for treatment against metastatic brain tumors of lung adenocarcinoma.

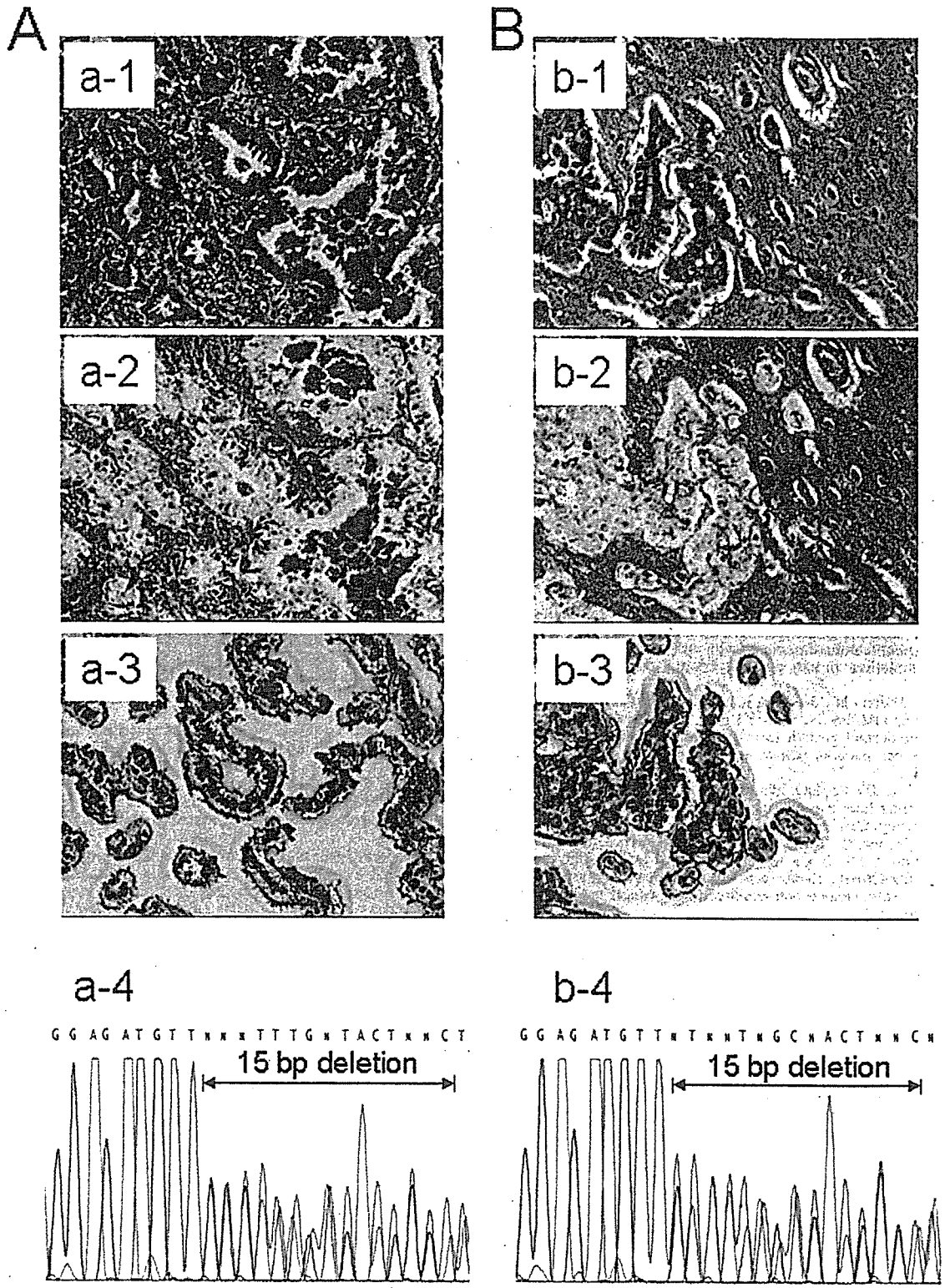

FIGURE 1 – Representative photographs of laser capture microdissection and sequence data obtained from a primary lung adenocarcinoma (*a*) and a corresponding metastatic brain tumor (*b*) in the same patient (case 1 in Table I; magnification of photographs, ×100). (*a-1*) and (*b-1*) shows the tumors in hematoxylin-stained tissue sections before microdissection. (*a-2*) and (*b-2*) shows the same sections after microdissection. (*a-3*) and (*b-3*) shows the cells captured on the transfer films. The sequence chromatogram from the primary tumor is shown in (*a-4*), and that from the metastatic brain tumor is shown in (*b-4*). Both show sequence chromatograms in exon 19 using antisense sequencing primer. A heterozygous in-frame 15 bp deletion was detected from both the samples, demonstrating a deletion of 5 amino acid residues from codon 746 to 750.

Acknowledgements

T.K. was the recipient of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan during the study. This work was supported by the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control, for Cancer Research (16S-1), and the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio).

References

1. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signalling. Exp Cell Res 2003;284:31–53.
2. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC 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.
3. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497–500.
4. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 2004;64: 8919–23.
5. Huang SF, Liu HP, Li LH, Ku YC, Fu YN, Tsai HY, Chen YT, Lin YF, Chang WC, Kuo HP, Wu YC, Chen YR et al. High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. Clin Cancer Res 2004;10:8195–203.
6. Marchetti A, Martella C, Felicioni L, Barassi F, Salvatore S, Chella A, Camplese PP, Iarussi T, Mucilli F, Mezzetti A, Cuccurullo F, Sacco R et al. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 2005;23:857–65.
7. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005; 97:339–46.
8. Han SW, Kim TY, Hwang PG, Jeong S, Kim J, Choi IS, Oh DY, Kim JH, Kim DW, Chung DH, Im SA, Kim YT et al. Predictive and prognostic impact of epidermal growth factor receptor mutation in non-small-cell lung cancer patients treated with gefitinib. J Clin Oncol 2005;23:2493–501.
9. Qin BM, Chen X, Zhu JD, Pei DQ. Identification of EGFR kinase domain mutations among lung cancer patients in China: implication for targeted cancer therapy. Cell Res 2005;15:212–17.
10. Cortes-Funes H, Gomez C, Rosell R, Valero P, Garcia-Giron C, Velasco A, Izquierdo A, Diz P, Camps C, Castellanos D, Alberola V, Cardenal F et al. Epidermal growth factor receptor activating mutations in Spanish gefitinib-treated non-small-cell lung cancer patients. Ann Oncol 2005;16:1081–6.
11. Matsumoto S, Iwakawa R, Kohno T, Suzuki K, Matsuno Y, Yamamoto S, Noguchi M, Shimizu E, Yokota J. Frequent EGFR mutations in non-invasive bronchioloalveolar carcinoma. Int J Cancer 2006;118:2498–504.
12. Colby TV, Noguchi M, Henschke C, Vazquez MF, Geisinger K, Yokose T, Chori P, Rami-Porta R, Franks T. Adenocarcinoma. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, eds. Pathology and genetics: tumours of the lung, pleura, thymus and heart, Lyon: IARC, 2004.35–44.
13. Yokota J, Wada M, Shimosato Y, Terada M, Sugimura T. Loss of heterozygosity on chromosomes 3, 13, and 17 in small-cell carcinoma and on chromosome 3 in adenocarcinoma of the lung. Proc Natl Acad Sci USA 1987;84:9252–6.
14. Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. Nat Med 1996;2:1035–7.
15. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. Int J Cancer 2006; 118:257–62.
16. Yatabe Y, Kosaka T, Takahashi T, Mitsudomi T. EGFR mutation is specific for terminal respiratory unit type adenocarcinoma. Am J Surg Pathol 2005;29:633–9.
17. Yoshida Y, Shibata T, Kokubu A, Tsuta K, Matsuno Y, Kanai Y, Asamura H, Tsuchiya R, Hirohashi S. Mutations of the epidermal growth factor receptor gene in atypical adenomatous hyperplasia and bronchioloalveolar carcinoma of the lung. Lung Cancer 2005;50:1–8.
18. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, Gazdar AF, Wistuba II. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. Cancer Res 2005;65:7568–72.
19. Shiseki M, Kohno T, Nishikawa R, Sameshima Y, Mizoguchi H, Yokota J. Frequent allelic losses on chromosomes 2q, 18q, and 22q in advanced non-small cell lung carcinoma. Cancer Res 1994;54:5643–8.
20. Nagane M, Lin H, Cavenee WK, Huang HJ. Aberrant receptor signaling in human malignant gliomas: mechanisms and therapeutic implications. Cancer Lett 2001;162 (Suppl):S17–S21.
21. Maher EA, Furnari FB, Bachoo RM, Rowitch DH, Louis DN, Cavenee WK, DePinho RA. Malignant glioma: genetics and biology of a grave matter. Genes Dev 2001;15:1311–33.
22. Zabel A, Debus J. Treatment of brain metastases from non-small-cell lung cancer (NSCLC): radiotherapy. Lung Cancer 2004;45 (Suppl 2): S247–S252.
23. Chiu CH, Tsai CM, Chen YM, Chiang SC, Liou JL, Perng RP. Gefitinib is active in patients with brain metastases from non-small cell lung cancer and response is related to skin toxicity. Lung Cancer 2005;47: 129–38.
24. Cappuzzo F, Ardizzoni A, Soto-Parra H, Gridelli C, Maione P, Tiseo M, Calandri C, Bartolini S, Santoro A, Crino L. Epidermal growth factor receptor targeted therapy by ZD 1839 (Iressa) in patients with brain metastases from non-small cell lung cancer (NSCLC). Lung Cancer 2003;41:227–31.
25. Namba Y, Kijima T, Yokota S, Niinaka M, Kawamura S, Iwasaki T, Takeda Y, Kimura H, Okada T, Yamaguchi T, Nakagawa M, Okumura Y et al. Gefitinib in patients with brain metastases from non-small-cell lung cancer: review of 15 clinical cases. Clin Lung Cancer 2004;6:123–8.

JOURNAL OF BIOSCIENCE AND BIOENGINEERING
Vol. 102, No. 1, 46–52. 2006
DOI: 10.1263/jbb.102.46

Classification of Intramural Metastases and Lymph Node Metastases of Esophageal Cancer from Gene Expression Based on Boosting and Projective Adaptive Resonance Theory

Hiro Takahashi,[1,2,3] Kazuhiko Aoyagi,[3] Yukihiro Nakanishi,[4] Hiroki Sasaki,[3] Teruhiko Yoshida,[3] and Hiroyuki Honda[2]*

Research Fellow of the Japan Society for the Promotion of Science (JSPS), 8 Ichibancho, Chiyoda-ku, Tokyo 102-8472, Japan,[1] Department of Biotechnology, School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan,[2] Genetics Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan,[3] and Pathology Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan[4]

Received 22 February 2006/Accepted 15 April 2006

Esophageal cancer is a well-known cancer with poorer prognosis than other cancers. An optimal and individualized treatment protocol based on accurate diagnosis is urgently needed to improve the treatment of cancer patients. For this purpose, it is important to develop a sophisticated algorithm that can manage a large amount of data, such as gene expression data from DNA microarrays, for optimal and individualized diagnosis. Marker gene selection is essential in the analysis of gene expression data. We have already developed a combination method of the use of the projective adaptive resonance theory and that of a boosted fuzzy classifier with the SWEEP operator denoted PART-BFCS. This method is superior to other methods, and has four features, namely fast calculation, accurate prediction, reliable prediction, and rule extraction. In this study, we applied this method to analyze microarray data obtained from esophageal cancer patients. A combination method of PART-BFCS and the U-test was also investigated. It was necessary to use a specific type of BFCS, namely, BFCS-1,2, because the esophageal cancer data were very complexity. PART-BFCS and PART-BFCS with the U-test models showed higher performances than two conventional methods, namely, *k*-nearest neighbor (kNN) and weighted voting (WV). The genes including *CDK6* could be found by our methods and excellent IF-THEN rules could be extracted. The genes selected in this study have a high potential as new diagnosis markers for esophageal cancer. These results indicate that the new methods can be used in marker gene selection for the diagnosis of cancer patients.

[**Key words:** cancer classification, boosting, projective adaptive resonance theory, esophageal cancer, intramural metastases]

Cancer is a major cause of human deaths in the many countries. Esophageal cancer is the eighth most common cancer and the sixth most common cause of cancer-related mortality in the world (1). This cancer is a well-known cancer with poorer prognosis than other cancers. Lymph node metastasis is one of the reasons for its poor prognosis in potentially resectable solid epithelial tumors. Furthermore, intramural metastasis (skip metastasis) has poorer prognosis than lymph node metastasis (2). From such situations, the prognosis of cancer patients with the same clinical diagnosis can differ, frequently. Therefore, it is important that the prognosis of cancer patients is made accurately and that an adequate treatment is proposed. However, the diagnosis of cancer patients is determined by a complex causality involving multiple factors because the mechanisms of cancer development (or malignancy) are extremely complex. Gene expression data from DNA microarrays are individualized and useful in the diagnosis and prognosis of diseases (3). To conduct this analysis, it is necessary to select genes significantly expressing mRNA and strongly related to the diagnosis or prognosis of disease, because the performance of classification analysis can decline owing to such large quantities of data.

Feature selection has been performed to screen candidate genes for modeling. There are two types of approach: the wrapper and filter approaches. In the former, features (genes) are selected as a part of mining algorithms, such as support vector machines (SVMs) (4), a fuzzy neural networks (FNNs) combined with the SWEEP operator method (FNN-SWEEP) (3), and a boosted fuzzy classifier with the SWEEP operator method (BFCS) (5, 6). On the other hand, in the filter approach, features are selected by filtering methods, such as the U-test, the t-test, signal-to-noise statistic (S2N)

* Corresponding author. e-mail: honda@nubio.nagoya-u.ac.jp
phone: +81-(0)52-789-3215 fax: +81-(0)52-789-3214