**Table 1.** Characters of cell lines, culture conditions, and number of analyzed spots

| Cell line | | Cell line | | Culture medium[d] | Number of spots[e] | Spots in 80% | |
|---|---|---|---|---|---|---|---|
| number[a] | Cell origin[b] | name | Source[c] | | | Cy3-image[f] | References |
| | HL | | | | | | |
| 1 | HL | KM-H2 | DSMZ | 1 | 2006 | 1597 | 17 |
| 2 | HL | HDLM2 | DSMZ | 2 | 1925 | 1538 | 18 |
| 3 | HL | L-428 | DSMZ | 1 | 1944 | 1571 | 19 |
| 4 | HL | HD-MY-Z | DSMZ | 1 | 2018 | 1579 | 20 |
| | B cell | | | | | | |
| 5 | BL | Namalwa | Human-Science | 1 | 2016 | 1595 | 21 |
| 6 | BL | EB-3 | Human-Science | 1 | 1986 | 1543 | 22 |
| 7 | BL | RAMOS | Human-Science | 1 | 2044 | 1611 | 23 |
| 8 | BL | Raji | Tohoku | 1 | 1850 | 1474 | 24 |
| 9 | BL | TL-1 | Tohoku | 1 | 2002 | 1574 | 25 |
| 10 | BL | Daudi | Tohoku | 1 | 1942 | 1509 | 26 |
| 11 | BL | HS-sultan | Human-Science | 1 | 1998 | 1579 | 27, 28 |
| 12 | B-ALL | RS4;11 | DSMZ | 3 | 2013 | 1587 | 29 |
| 13 | B-ALL | REH | DSMZ | 1 | 2066 | 1597 | 30 |
| 14 | B-ALL | NALM-6 | DSMZ | 1 | 1989 | 1568 | 31 |
| 15 | DLBCL | DB | ATCC | 4 | 1984 | 1575 | 32 |
| 16 | DLBCL | Toledo | ATCC | 4 | 1994 | 1598 | 33 |
| 17 | DLBCL | Pfeiffer | ATCC | 4 | 2042 | 1605 | 33 |
| 18 | DLBCL | KARPAS-422 | DSMZ | 1 | 1930 | 1538 | 34 |
| 19 | DLBCL | OCI-LY-19 | DSMZ | 3 | 1958 | 1578 | 35 |
| 20 | FL | DOHH-2 | DSMZ | 1 | 2485 | 1650 | 36 |
| 21 | FL | SU-DHL-4 | DSMZ | 1 | 1979 | 1577 | 37 |
| 22 | PCL | KARPAS-620 | DSMZ | 2 | 2029 | 1570 | 38 |
| 23 | PCL | SK-MM2 | DSMZ | 1 | 2068 | 1592 | 39 |
| | T cell | | | | | | |
| 24 | T-ALL | JURKAT | Tohoku | 1 | 1913 | 1493 | 40 |
| 25 | T-ALL | PEER | Tohoku | 1 | 2018 | 1591 | 41 |
| 26 | T-ALL | CCRF-CEM | Tohoku | 1 | 2012 | 1589 | 42 |
| 27 | T-ALL | Molt3 | Human-Science | 1 | 1948 | 1570 | 43 |
| 28 | T-ALL | Molt4 | Human-Science | 1 | 1937 | 1556 | 43 |
| 29 | T-ALL | TALL-1 | Human-Science | 1 | 1996 | 1609 | 44 |
| 30 | T-ALL | CCRF-HSB2 | Human-Science | 1 | 2000 | 1609 | 45 |
| 31 | ALCL | SU-DHL-1 | DSMZ | 1 | 1952 | 1563 | 46 |
| 32 | ALCL | Karpas299 | DSMZ | 1 | 1938 | 1568 | 47 |
| 33 | ALCL | SR-786 | DSMZ | 5 | 1982 | 1595 | 48 |
| 34 | ALCL | SUP-M2 | DSMZ | 1 | 2014 | 1593 | 49 |
| 35 | ATL | ILT-Mat | Tohoku | 6 | 1983 | 1565 | 50 |
| 36 | ATL | TL-SU | Tohoku | 1 | 2018 | 1550 | 40 |
| 37 | ATL | TL-Hir | Tohoku | 1 | 2006 | 1594 | 51 |
| 38 | CTCL | Hut78 | Tohoku | 1 | 1912 | 1537 | 52 |
| 39 | CTCL | Hut102 | Tohoku | 6 | 1978 | 1597 | 52 |
| | NK cell | | | | | | |
| 40 | NK cell lymphoma | KHYG-1 | Human-Science | 6 | 2085 | 1592 | 53 |
| 41 | NK cell lymphoma | NK92 | ATCC | 7 | 1977 | 1569 | 54 |
| 42 | NK cell lymphoma | KAI3 | Human-Science | 6 | 2040 | 1598 | 55 |

a) Cell line numbers refer to those in Figs. 3 and 4

b) Cell origin as described in the indicated references. HL, Hodgkin's lymphoma; BL, Burkitt's lymphoma; B-ALL, B cell acute lymphoblastic leukemia; DLBCL, diffuse large B cell lymphoma; FL, follicular lymphoma; PCL, plasmacytoma; T-ALL, T cell acute lymphoblastic leukemia; ALCL, anaplastic large cell lymphoma; ATL, adult T cell leukemia; CTCL, cutaneous T cell lymphoma

c) Source of cell lines: DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (Braunschweig, Germany); Human Science (Osaka, Japan); Tohoku, Institute of Development, Aging and Cancer, Tohoku University (Miyagi, Japan); ATCC, American Type Culture Collection (Manassas, VA)

d) Culture media are those recommended by suppliers. Culture media: 1, RPMI 1640 90% + FBS 10%; 2, RPMI 1640 80% + FBS 20%; 3, alpha-MEM 90% + FBS 10%; 4, RPMI 1640 90%, FBS 10%, 10 mM HEPES, 1 mM sodium pyruvate, 4.5 g/L glucose, 1.5 g/L sodium bicarbonate; 5, RPMI 1640 85% + FBS 15%; 6, RPMI 1640 90%, FBS 10%, IL-2 100 U/mL; 7, alpha-MEM 75%, FBS 12.5%, horse serum 12.5%, IL-2 100 U/L, 0.2 mM inositol, 0.1 mM 2-ME, 0.02 mM folic acid

e) Total number of spots detected by DeCyder software

f) Number of spots present in more than 80% of Cy3 images

### 2.5 Data-mining process

For scatter plot analysis, proteins were extracted three times from KAI3 cells and labeled with Cy5 fluorescent dye as described in Section 2.2. They were mixed with the Cy3-labeled reference sample and separated on quadruplicate 2-D PAGE gels. The normalized and averaged spot intensities were transformed logarithmically and subjected to scatter plot analysis. Overall correlation of the expression patterns across the 42 cell lines was monitored by a correlation matrix.

The Wilcoxon or Kruskal–Wallis test was employed to identify the protein spots that discriminated between the cell line groups with high confidence. The class prediction system, having learned the expression features of the groups, was then used to classify samples. Classification accuracy of various spot sets was assessed by leave-one-out cross-validation. In the leave-one-out cross-validation, each cell line sample was left out in turn, and the prediction model with the discriminator spot set was constructed using the remaining cell line samples. The cell line sample left out was then used as the test case to evaluate the accuracy of the class prediction. This process was repeated by reducing lower-ranked spots for all samples. For spot ranking, various algorithms were used. The lower-ranked spots were removed and the classification accuracy was calculated again. This process was repeated and the cross-validation error rate was plotted as a function of spot numbers. These analyses were performed using Impressionist (Gene Data, Basel, Switzerland).

GeneMaths (Applied Maths, Sint-Martens-Latem, Belgium) was used for hierarchical clustering analysis by measuring Euclidian distance as a similarity coefficient and by Ward's algorithm as a tree-building method.

### 2.6 In-gel digestion and MS study

Protein identification was performed as described previously [57]. In brief, preparative gels were made by loading 500 μg of non-labeled protein sample and the gels were stained with SYPRO Ruby. The gel image was acquired by scanning the gels at the appropriate wavelength for SYPRO Ruby with a MasterImager 2640 (Amersham Biosciences). Spots were marked with the DIA mode of DeCyder software and recovered with an automated spot collector (Spot Picker, Amersham Biosciences). The recovered gel plugs were washed extensively with 50 mM ammonium bicarbonate and air-dried. The protein in the gel was treated with 200 ng of TPCK-treated trypsin at 37°C overnight. The trypsin-digested peptides were recovered by incubation with 50% ACN/0.1% TFA and mixed with an equal volume of matrix solution, dihydroxybenzoic acid. PMF and MS/MS analysis were performed on a Q-Star Pulser-i equipped with the oMALDI ion source (Applied Biosystems, Foster City, CA, USA). The results of identification by PMF and MS/MS were scored with the Analyst QS and Mascot programs, respectively, and the top-scoring gene products were considered to be the corresponding proteins.

## 3 Results

### 3.1 Protein expression profiles of lymphoma cell lines

We designed the experiment so that the 2-D image of each cell line was normalized with respect to the common image in the same gel to avoid gel-to-gel variations resulting from electrophoresis. The reference sample and the experimental samples were labeled with Cy3 and Cy5, respectively. The Cy3-labeled reference sample and each Cy5-labeled sample were mixed together and co-separated by 2-D PAGE. As all gels produce a common Cy3 image of the reference sample, standardization of the spot intensity of the Cy5 image to that of the Cy3 image can compensate for gel-to-gel variation, allowing quantitative and reproducible study of protein expression. This approach enables semi-quantitative comparison of proteins between the samples. Based on the relative intensity of the spots, we estimated the degrees of both up- or down-regulation of proteins among samples. Because the absolute amounts of proteins could be measured from the fluorescence intensity using the standard curve, quantitative inter-protein comparison would have been possible. However, we did not attempt this in the present study. An example of a two-channel 2-D image is shown in Fig. 1A. The blue Cy5 image of KAI3 cells is merged with the red Cy3 image of the reference sample in the same gel. As the reference sample contains protein extracts from all cell lines, including KAI3, the spots on the Cy3 image include those of KAI3 cells. The location of the spots is almost the same in the two images because the two samples were co-separated in the same gel and the electrophoretic properties of Cy3- or Cy5-labeled proteins are designed to be almost identical.

The reproducibility of the protein expression profile was evaluated by scatter plot analysis. Three independent 2-D DIGE separations of the protein extract of KAI3 cells were performed, and the correlation of spot intensity was examined. The scattergram in Fig. 1B shows the high reproducibility of spot intensity: in all pairs of experiments, the intensity of at least 98.23% of spots was distributed within a two-fold difference and the correlation coefficient was at least 0.7176.

The overall correlation of spot intensities across the 42 cell lines is shown in Fig. 1C. The degree of correlation of protein expression patterns was demonstrated by measuring correlation coefficients, indicating that cell lines derived from lymphoid neoplasms with similar phenotypes had common protein expression profiles. In particular, the cell lines in the B cell group showed more homogeneous protein expression profiles, perhaps reflecting the biological variations of lymphocyte lineages: B cells differentiate into plasmacytes and are recognized by CD19 throughout the B cell lineage until plasma cell differentiation. On the other hand, a correlation matrix revealed that cell lines of the T cell group were relatively heterogeneous in terms of their protein expression profiles, perhaps reflecting the clinical features of

 

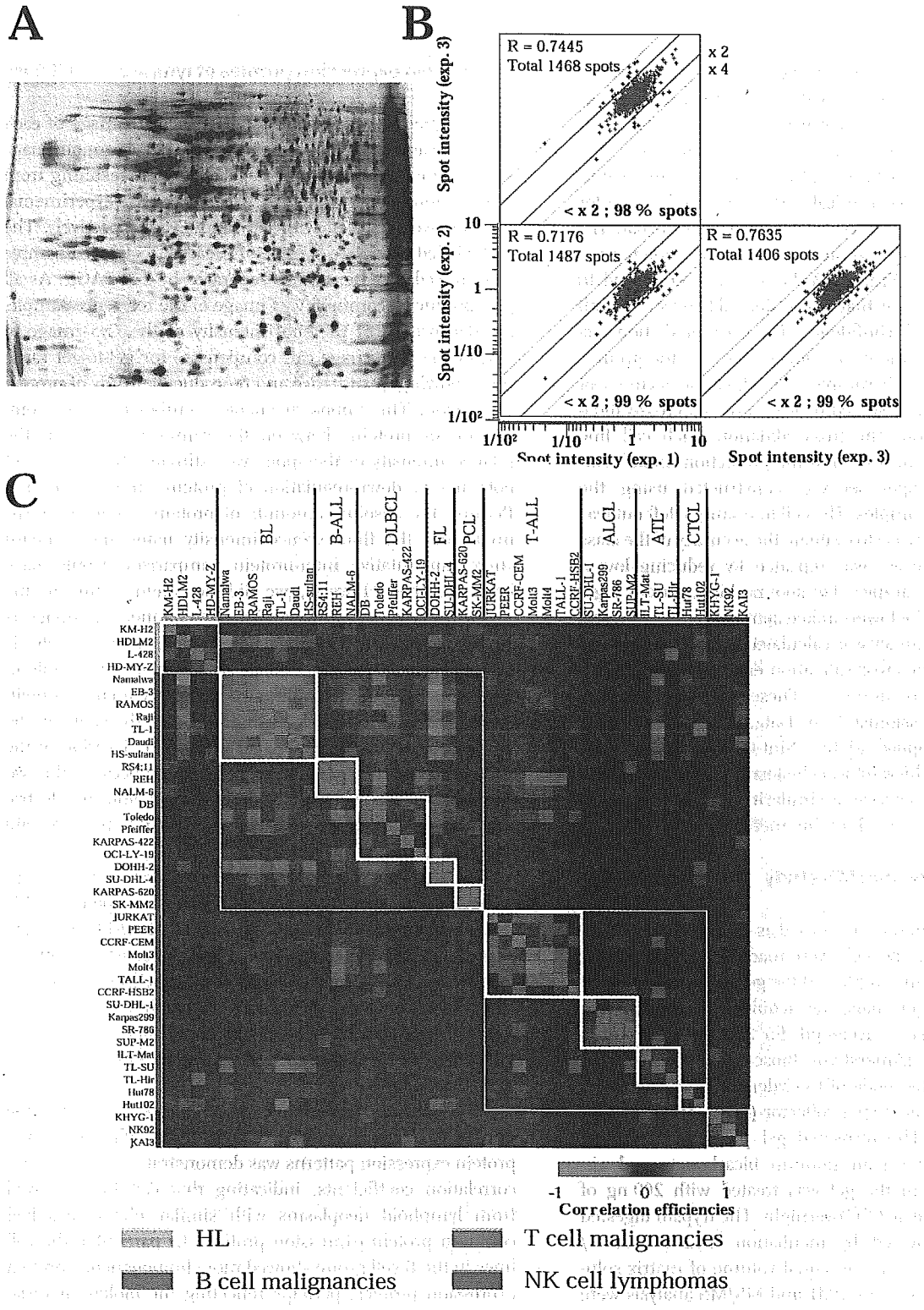


**Figure 1.** Quantitative and reproducible protein expression profiles. (A) Two-color image of a 2-D profile. A Cy5 image of KAI3 cells (blue) was merged with a Cy3 image of the reference sample (red). (B) Scattergram of expression profile of KAI3 cells. Comparison of data from three independent experiments revealed the high reproducibility of protein expression profiling. (C) Correlation matrix of cell lines based on protein expression profiles. The similarity of profiles is shown by color. The boxed matrixes indicate the combinations of cells having the same histological subtype.

 

T cell lymphoma; one-third of T cell lymphoma arise, spread to, or relapse at extranodal sites, and the site of disease is important for disease definition [58].

### 3.2 Strategy of proteomic analysis

The DeCyder software merged and quantified 83 977 protein spots in 168 gels representing the 42 cell lines (Fig. 2). Among them, 66 143 spots having valid values in at least 80% of Cy3 images were retained for further analysis. The spots with an expression level statistically and significantly different between particular cell line groups were selected with the Wilcoxon or Kruskal–Wallis tests ($p < 0.05$). We grouped and compared the 42 cell lines as follows: (i) Hodgkin's lymphoma (HL) cells *versus* other cells, (ii) cells from B cell malignancies *versus* cells from T cell malignancies *versus* cells from natural killer (NK) cell lymphoma, (iii) HL-cells *versus* anaplastic large cell lymphoma (ALCL) cells. Various numbers of spots were selected depending on the sample sets: 197 spots for discrimination between HL cells and other cells, 251 spots for discrimination between the cells from B cell malignancies, T cell malignancies and NK cell lymphoma, and 84 spots for discrimination between HL cells and ALCL cells. On the basis of the leave-one-out cross-validation error rate, we prioritized the protein spots according to their contribution to the classification: the protein spot sets that minimized the classification error rate were chosen as the best. The robustness of the discrimination potential of these spot sets was further evaluated by unsupervised classification methods, including principal component analysis and hierarchical clustering analysis. Among the selected spot sets, those that clearly categorized the cell lines into groups by these methods were finally considered as the most informative for the classification. The proteins corresponding to these spots were identified by MS.

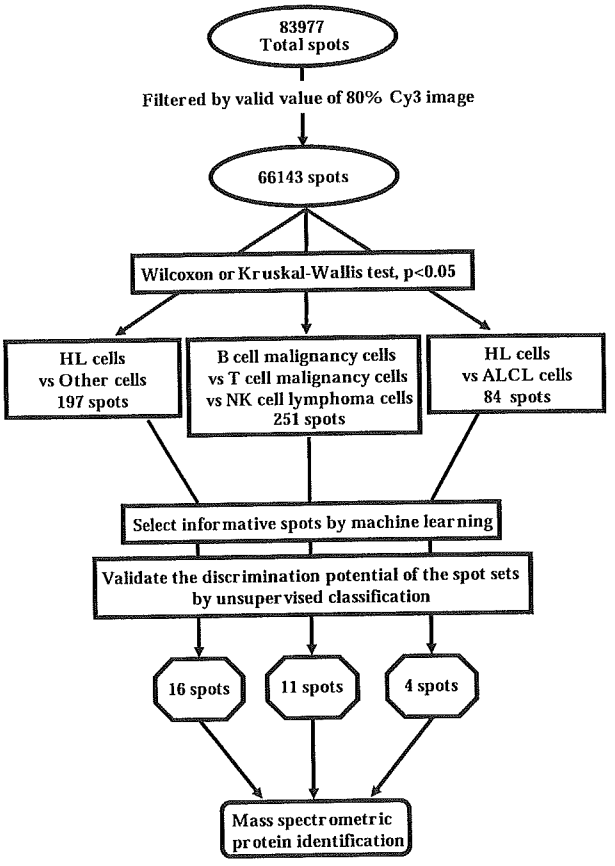


**Figure 2.** Schematic diagram of the data analysis strategy in this study. The total number of spots recognized and quantified by the DeCyder software was initially 83 977. Spots were filtered out if they appeared in less than 80% of Cy3 images. The Wilcoxon or Kruskal–Wallis test ($p < 0.05$) was used to identify the differentially expressed spots, and machine-learning methods were used to select the informative spot sets. Unsupervised classification methods were used to evaluate the discrimination potential of the spots. The proteins corresponding to those spots were identified by MS.

### 3.3 Multivariate analysis of cell lines

We illustrate in Fig. 3 the process of spot selection using multivariate analysis and statistical-learning methods for discrimination between HL cells and other cells. The Wilcoxon test selected 197 spots whose intensity was significantly different ($p < 0.05$) between the two groups. Among the spot ranking methods, the plot of the leave-one-out cross-validation error rate revealed that Fisher linear discriminant analysis identified minimal spot sets including 16 or 32 spots by which HL cells could be discriminated from other cells with the lowest cross-validation error rates (Fig. 3A). Although the other three algorithms also yielded spot sets that could segregate the cell line groups, more spots were required to obtain the lowest cross-validation error rate or the minimal error rate with those algorithms was higher than that by sparse linear discriminant analysis (Fig. 3A). The robustness of the discrimination potential of the two spot sets was examined with unsupervised methods. Using the 16 spots, principal component analysis (Fig. 3B) and hierarchical clustering analysis (Fig. 3C) clearly divided the cells into two groups. On the other hand, the unsupervised study using 32 spots resulted in ambiguous classification (data not shown). Therefore, we concluded that the set containing 16 spots was most informative for discrimination between HL cells and other cells.

We performed the same procedure for the other sets of the cell line groups, comparing protein expression profiles between (i) B cell malignancy cells, T cell malignancy cells and NK cell lymphoma cells (Fig. 4A) and (ii) HL cells *versus* ALCL cells (Fig. 4B). The Wilcoxon test was used for discrimination between two groups and the Kruskal–Wallis test was used for three groups. The results of unsupervised classification analysis of the cell lines using the selected spots are illustrated in Fig. 4. In principal component analysis, all cell lines were categorized into the expected groups. In hierarchical clustering analysis, T cell leukemia cell lines PEER and CCRF-HSB2 were localized in the group of B cell

 

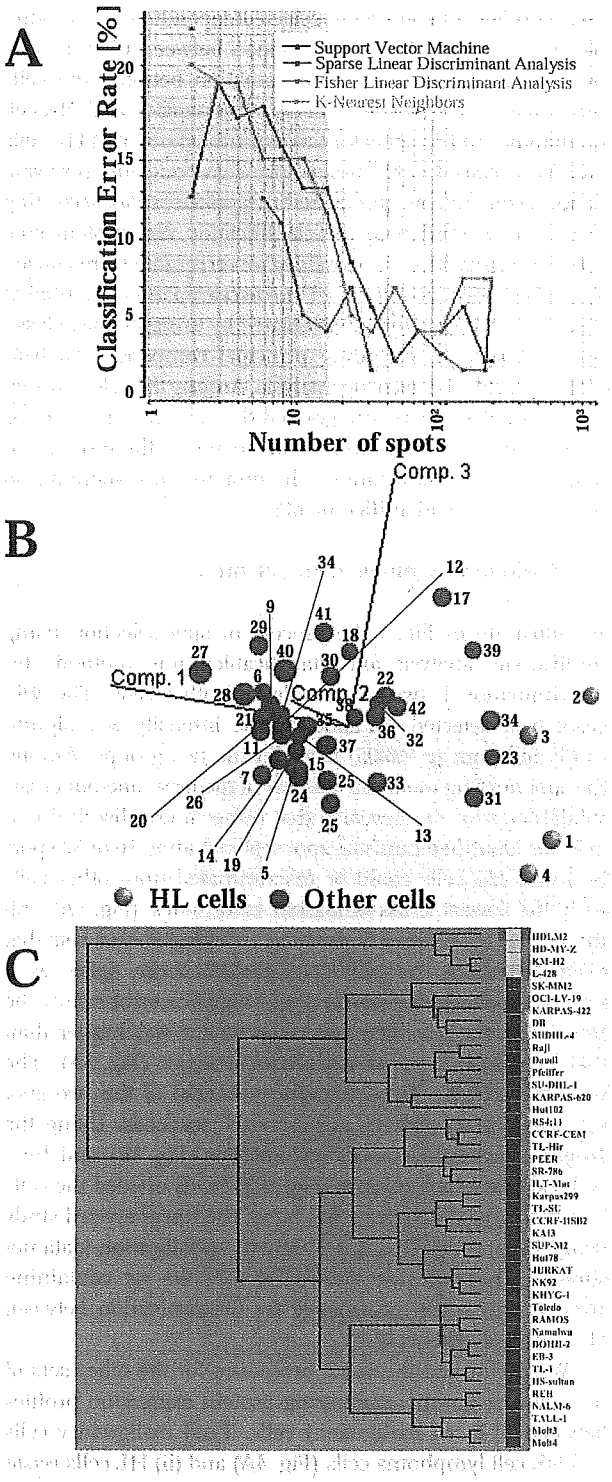


**Figure 3.** Multivariate analysis of protein expression profiles of HL cells and other cells. (A) Spot ranking methods were used to select the spot set that could discriminate the HL cell group from other cell group with minimal cross-validation error rate. Cross-validation error rate is plotted as a function of the number of spots that best discriminate the two groups. (B) Principal component analysis of the cell lines using the selected 16 spots. The cell lines are numbered according to Table 1. (C) Hierarchical clustering analysis of the cell lines using the selected 16 spots.

malignancy cells and B cell lymphoma cell line KARPAS-422 was clustered with T cell malignancy cells, but the other cell lines were grouped according to their origin (Fig. 4A).

### 3.4 Localization of the informative spots on 2-D gels and identification of proteins corresponding to the spots

The location of the informative spots on the 2-D images is shown in Fig. 5. The proteins corresponding to the numbered spots were later identified by MS.

Figure 6A illustrates the process of protein identification, using spot 21 as an example. The mass spectrogram of the tryptic digest of spot 21 was subjected to PMF analysis, resulting in the identification of galectin-1 with an identification score of 1009 and protein coverage of 58.5% by amino acid count. The identification was also confirmed by MS/MS data for the tryptic peptide with an $m/z$ value of 1800.6786 (Fig. 6B–D), identifying spot 21 as galectin-1 with a MASCOT score of 86. Figure 6E shows the amino acid sequence of galectin-1, indicating that seven peptides were used for the identification. Similar procedures were performed for the other protein spots. Of 31 protein spots, 23 protein spots were identified. The results of identification are summarized in Table 2.

## 4 Discussion

HL is characterized by the presence of Hodgkin and Reed-Sternberg (HRS) cells, which usually comprise less than 1% of the cellular infiltrate in the lymphoma tissues [59]. We found that HL cells had a distinct overall proteomic profile: the HL cells were more similar to each other than to any other cell groups (Fig. 1C). It is now generally accepted that HL cells in most cases are derived from germinal center B cells (or, rarely, T cells), although the expression of many B cell markers is lost in HL cells [60]. Our findings suggest that HL cells develop their own protein expression pattern during the course of transformation and that the loss of B cell markers reflects this overall alteration. These concepts are consistent with recent transcriptomic studies on HRS cells, where on the basis of mRNA expression profiles the cells clustered as a distinct entity irrespective of their B or T cell origin [61].

In the comparison among B cell malignancies, T cell malignancies and NK cell lymphoma, both principal component analysis and hierarchical clustering analysis divided these cells according to their original phenotypic groups. In addition, the protein expression pattern of NK lymphoma cells was closer to that of T cell malignancies than that of B cell malignancies. These findings are consistent with recent studies showing that although B cells, T cells, and NK cells are derived from a common lymphoid progenitor [62], NK cells are biologically more related to T cells than to B cells [58]. Although NK cells and T cells can be dis-

 

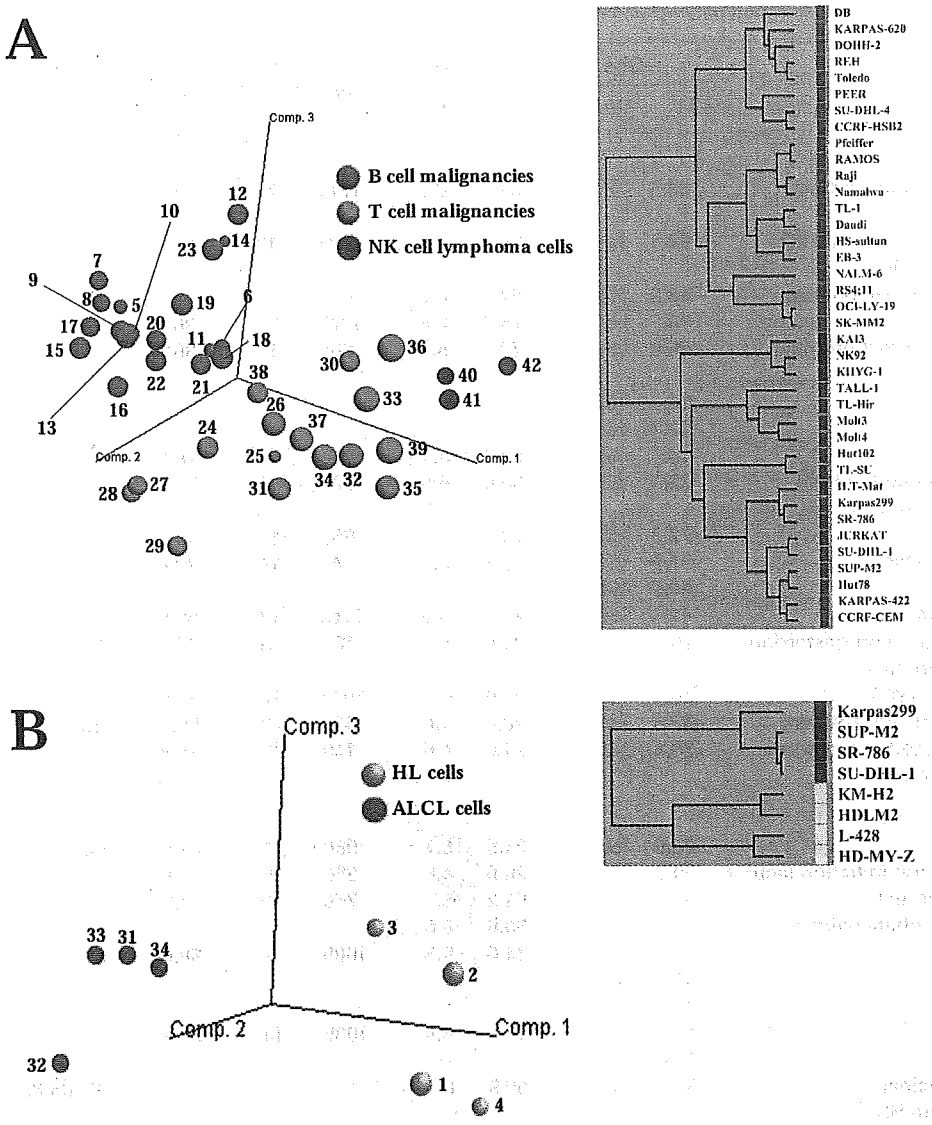


**Figure 4.** Validation of the selected spots for classification. The cell lines were grouped by principal component analysis and hierarchical clustering analysis using the selected spots. (A) Cell lines from B cell malignancies, T cell malignancies, and NK cell lymphoma. (B) Cell lines from HL and ALCL. The color code in the right panel of hierarchical clustering analysis corresponds to that in principal component analysis. The cell lines are numbered according to Table 1.

tinguished by immunophenotype and molecular genetic studies, there is overlap in NK cell and T cell antigen expression, function, and patterns of disease [58]. We identified seven proteins that were considered to distinguish these three histological subtypes on the basis of their expression pattern. However, linkage of these proteins to histological differentiation was not found by literature validation. Further studies of these proteins may provide unique findings suggestive of novel concepts in lymphocyte ontogeny.

We also examined whether a proteomic approach could contribute to the development of biomarkers for the diagnostically gray area at the interface between HL and other lymphoid neoplasms such as ALCL. ALCL and HL are biologically distinct entities: the majority of ALCL and HL cells are of T cell and B cell origin, respectively [63]. Therefore, the expression of T cell or B cell markers suggests an appropriate diagnosis. However, in some cases, the morphological and immunological features of these neoplasms may overlap considerably, and tumors in this gray area may present clinical problems [64]. Recently, transcriptomic studies, based on cDNA array technology, have revealed that HL cells and ALCL cells can be distinguished from each other by the expression of a small number of genes, including clusterin [65], c-*jun*, calmodulin, growth factor receptor-bound protein (GRB2), and S100A4 [66]. Our study included the protein spots corresponding to Jun-activated domain binding protein 1, calmodulin, and GRB2 (data not shown). However, they were not detected as important proteins for the classification in our study. There are four possible explanations for this discrepancy. First, many lines of evidence suggest that the expression of mRNA does not necessarily reflect that of protein [3–6]. Therefore, transcriptomic and proteomic studies may produce discrepant results by their nature. Second, 2-D DIGE is less sensitive than DNA microarray and does

 

**Table 2.** Identified proteins informative for the classification of lymphoid neoplasms

| Spot number[a] | Accession number[b] | Protein name | Obs.[c] | | Theor.[d] | | MS score[e] | Match[f] | Coverage[g] | MS/MS score[h] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $M_r$ | p*I* | $M_r$ | p*I* | | | | |
| HL cells *vs.* other cells | | | | | | | | | | |
| 1 | P31939 | Bifunctional purine biosynthesis protein PURH | 66.5 | 6.7 | 64.6 | 6.3 | 1771 | 23 | 47.2 | – |
| 2 | P14314 | Protein kinase C substrate, 80 kDa protein, heavy chain | 94.1 | 4.4 | 59.3 | 4.3 | 639 | 13 | 21.4 | – |
| 3 | Q96KP4 | Cytosolic nonspecific dipeptidase | 48.8 | 5.9 | 52.9 | 5.7 | – | – | – | 61 |
| 4 | P08133 | Annexin A6 | 75.8 | 5.6 | 75.7 | 5.4 | 1077 | 22 | 36.1 | – |
| | P38646 | Stress-70 protein, mitochondrial precursor | 75.8 | 5.6 | 73.7 | 5.9 | 659 | 17 | 30.8 | – |
| 5 | – | – | 26.4 | 5.2 | – | – | – | – | – | – |
| 6 | – | – | 29.1 | 5.1 | – | – | – | – | – | – |
| 7 | P48637 | Glutathione synthetase | 48.3 | 5.7 | 52.4 | 5.7 | 724 | 13 | 28.5 | – |
| 8 | Q15691 | Microtubule-associated protein RP/EB family member 1 | 32.1 | 5.1 | 30.0 | 5.0 | 794 | 11 | 50.7 | – |
| 9 | P13796 | L-plastin | 72.3 | 5.2 | 70.3 | 5.2 | 666 | 10 | 14.8 | – |
| 10 | P41250 | Glycyl-tRNA synthetase | 81.9 | 6.2 | 83.1 | 6.6 | 746 | 14 | 22.6 | – |
| 11 | – | – | 54.9 | 5.2 | – | – | – | – | – | – |
| 12 | P50453 | Cytoplasmic antiproteinase 3 | 41.8 | 5.8 | 42.4 | 5.6 | 1148 | 17 | 50.8 | – |
| 13 | O00264 | Membrane associated progesterone receptor component 1 | 26.4 | 4.5 | 21.5 | 4.6 | 185 | 5 | 26.2 | 36 |
| 14 | P31146 | Coronin-like protein p57 | 55.4 | 6.7 | 51.0 | 6.3 | 1944 | 26 | 44.9 | – |
| 15 | P12004 | Proliferating cell nuclear antigen | 33.6 | 4.6 | 28.8 | 4.6 | 1121 | 12 | 55.2 | 113 |
| 16 | O00264 | Membrane associated progesterone receptor component 1 | 26.4 | 4.4 | 21.5 | 4.6 | 149 | 3 | 15.4 | 41 |
| B cells *vs.* T cells *vs.* NK cells | | | | | | | | | | |
| 17 | P32119 | Peroxiredoxin 2 | 25.5 | 5.6 | 21.9 | 5.7 | 980 | 11 | 47.5 | 94 |
| 18 | P56537 | Eukaryotic translation initiation factor 6 | 29.0 | 4.4 | 26.6 | 4.6 | 350 | 5 | 27.6 | 66 |
| 19 | Q13451 | FK506-binding protein 5 | 52.6 | 6.1 | 51.2 | 5.7 | 758 | 12 | 24.9 | – |
| 20 | Q9NVS9 | Pyridoxine-5′-phosphate oxidase | 28.2 | 6.2 | 30.0 | 6.6 | – | – | – | 33 |
| 21 | P09382 | Galectin-1 | 19.3 | 5.1 | 14.6 | 5.3 | 1009 | 7 | 58.5 | 86 |
| 22 | – | – | 53.6 | 6.8 | – | – | – | – | – | – |
| 23 | – | – | 29.7 | 5.6 | – | – | – | – | – | – |
| 24 | P09104 | Gamma enolase | 46.9 | 4.9 | 47.1 | 4.9 | 1035 | 14 | 37.1 | – |
| 25 | – | – | 46.0 | 4.6 | – | – | – | – | – | – |
| 26 | P31942 | Heterogeneous nuclear ribonucleoprotein H3 | 36.5 | 6.8 | 36.9 | 6.4 | – | – | – | 83, 65,60 |
| 27 | – | – | 64.4 | 5.1 | – | – | – | – | – | – |
| HL cells *vs.* ALCL cells | | | | | | | | | | |
| 28 | P18206 | Vinculin | 108.9 | 6.2 | 123.7 | 5.5 | 766 | 28 | 33.5 | – |
| 29 | P13796 | L-plastin | 69.1 | 5.4 | 70.3 | 5.2 | 1346 | 24 | 44.3 | – |
| 30 | Q16555 | Dihydropyrimidinase-related protein-2 | 65.6 | 6.2 | 62.3 | 6.0 | 420 | 16 | 41.6 | – |
| 31 | – | – | 109.4 | 6.1 | – | – | – | – | – | – |

a) Spot numbers refer to those in Fig. 5
b) Accession number in Swiss-Prot
c) Observed $M_r$ (kDa) and p*I*
d) Theoretical $M_r$ (kDa) and p*I* from the ExPASy database
e) Analyst QS score indicates the confidence of the protein identification
f) Number of peptides used for identification
g) Amino acid coverage of matched peptides
h) MASCOT score indicates the confidence of the protein identification

not reveal a proteome that corresponds exactly to the transcriptome revealed by DNA microarray technology. In addition, the data obtained by 2-D DIGE contain information on protein isoforms which result from PTM. The spots observed by 2-D DIGE often represent one of several possible protein isoforms, and the up- or down-regulation of spot intensity may not always reflect the total amount of corresponding proteins. Therefore, generally, transcriptomic data and pro-

 

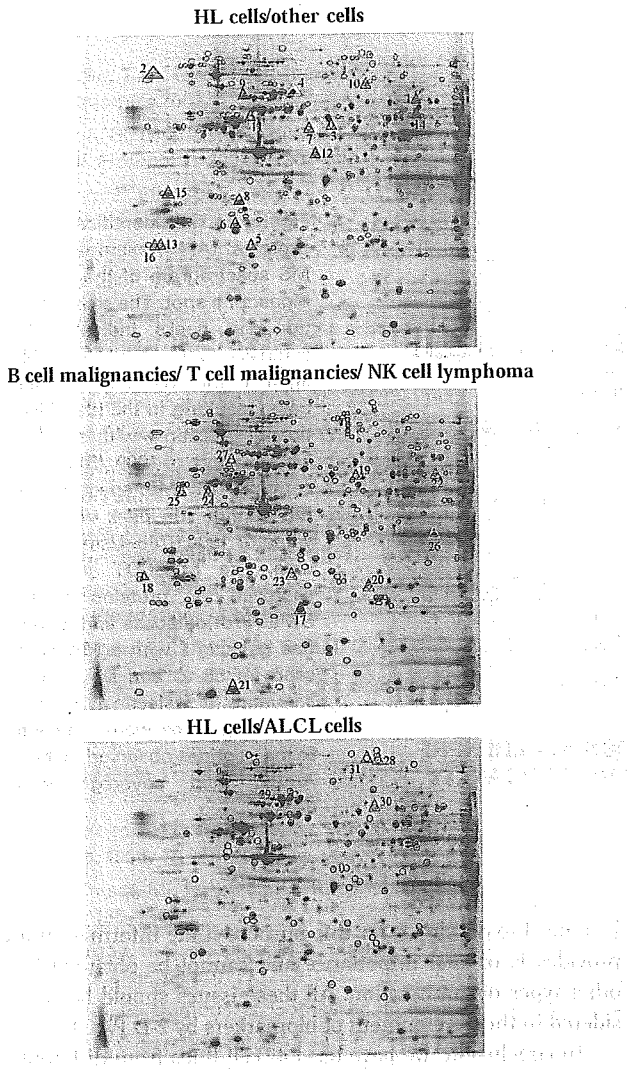


**Figure 5.** Localization of the informative spots on the 2-D PAGE gels. The spots selected by the Wilcoxon or Kruskal–Wallis tests are circled. Spots in a triangle are those identified as informative by the machine-learning methods. (A) HL cells *versus* other cells; (B) B cell malignancies *versus* T cell malignancies *versus* NK cell lymphoma; (C) HL cells *versus* ALCL cells. The spots are numbered according to Table 2.

teomic data obtained by 2-D DIGE may be essentially different and cannot be matched. Third, different statistical-learning methods used in these studies prioritized the genes or proteins in different ways. This may explain why the genes involved in both studies were not necessarily selected in a consistent manner, and also why the results were not matched even between transcriptomic studies. Fourth, as we used tissue culture cells instead of *in vivo* cells, their expression profiles might have changed during establishment of the cell lines. The classification potential of the identified proteins will need to be validated using clinical specimens and other quantification methods such as ELISA and immunohistological examination.

We identified protein expression patterns that can be used to discriminate between cells of lymphoid neoplasms. However, the proteins did not include those classically used in immunohistochemical studies to characterize these cells. For example, CD19, CD20, and CD79 for B cells, CD3 for T cells, and CD56 for NK cells were not identified in the present study. We suggest that the expression level of such proteins is too low to be observed in our 2-D system, and in fact we did not identify them on 2-D images by global protein identification experiments (data not shown). The sensitivity of the fluorescent dyes used here is slightly less than that of silver staining [67, 68] and, to our knowledge, the identification of conventional biomarkers for lymphoma cells, such as the CD series, in silver-stained 2-D gels has not been reported. Proteomic tools with higher sensitivity, such as 2-D DIGE with sensitive fluorescent dyes [69] or larger format 2-D PAGE [70], may identify the known biomarkers and link them to the novel ones.

In this report, we used cell lines rather than primary tumor specimens, because primary tumors are composed of various types of cells, including tumor cells, non-neoplastic lymphoid cells, blood vessels, and stromal components, and the use of cell lines avoids the problem of contamination with non-tumor cells. Therefore, we started our experiment from *in vitro* cells to capture the essential proteomic signature corresponding to each histological subtype. However, cell lines may not retain all the phenotypes of the primary tumor, and expression patterns might be altered by cell culture conditions. We found that the present proteomic data did not group two T cell leukemia cell lines (PEER and CCRF-HSB2) and one B cell lymphoma cell line (KARPAS-422) according to their original phenotypes, probably because the expression of informative proteins was altered during long-term culture in these cell lines. In addition, it is difficult to relate clinical information such as patient survival and response to chemotherapy to expression studies on cell lines. Therefore, the discrimination potential of the proteins identified by *in vitro* studies should be examined using clinical materials. DIGE technology with highly sensitive fluorescent dyes, CyDye DIGE Fluor saturation dye (Amersham Biosciences), enables the direct use of small amounts of protein obtained from laser-microdissected tissues with a high throughput [71]. Recently, we found the protein spots that classified tissue-cultured lung cancer cell lines according to their histological subtypes based on their expression profile using 2-D DIGE [72]. The identified spots categorized the lung cancer cells isolated by laser microdissection according to their histology [72]. A similar strategy for data integration will be achieved in a study of lymphoid neoplasms.

DIGE technology using a common internal standard sample enables the integration of data obtained from different sources such as *in vitro* and *in vivo* cells. Integration of the data will generate various possible applications. For example, xenograft experiments, in which the antitumor effects of reagents are evaluated in hematologic cells trans-

 

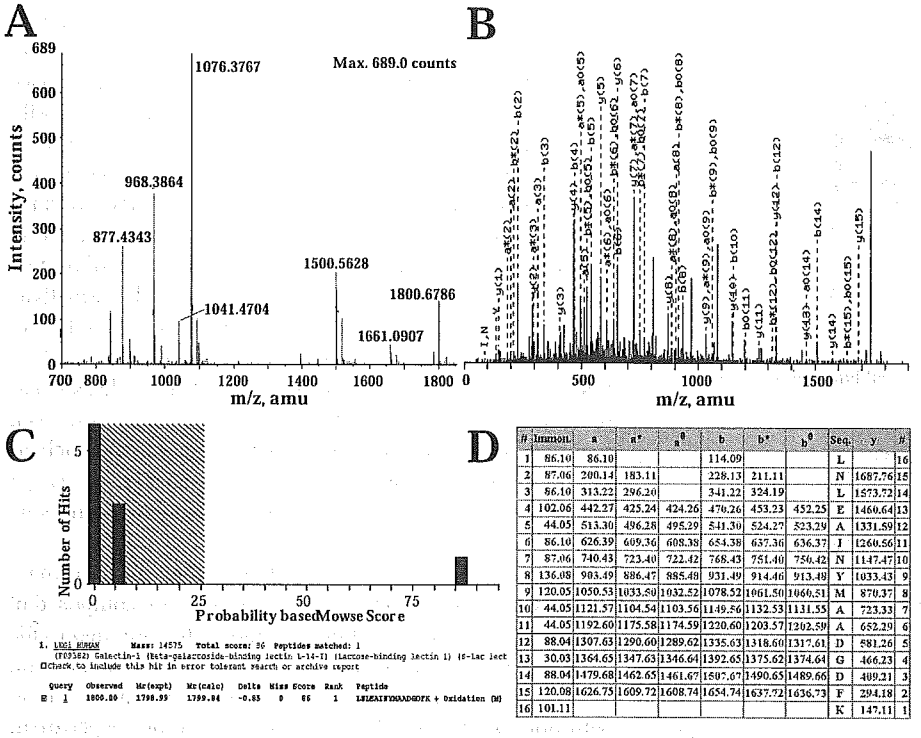


**Figure 6.** MS identification of proteins. (A) An example of an MS spectrogram of the trypsin digest of a spot. The seven peptide ion peaks indicated by asterisks were used for PMF against the Swiss-Prot database, resulting in the identification of galectin-1 with an Analyst QS score of 1009. (B) Representative MS/MS spectrum of the peptide ion peak with *m/z* value of 1800.6786. Using these data, the MASCOT program searched the proteins in the Swiss-Prot database and identified galectin-1 with a MASCOT score of 86 (C and D). (E) The sequence of galectin-1. Underlined sequences were matched to the peptides observed by MS. Amino acid coverage was 58.5%.

planted into model animals [73], can also be combined with such proteomic methods. These applications should contribute to the development of clinical markers. The advantage of using 2-D DIGE as a common platform for data integration is that we will be able to monitor the expression level of protein isoforms, which reflect the status of PTMs, without the need for specific antibodies against each isoform. ELISA may enable high-throughput screening to confirm the results of proteomics using a large number of samples, and immunohistochemistry may provide information on the localization of proteins in tissues and cells to give novel insights into the functional roles of the target proteins. However, these methods require antibodies against specific isoforms, and it is not always easy to generate such antibodies. 2-D DIGE enables the identification of spots from different origins by image matching even if the proteins corresponding to the spots are unknown. However, as 2-D PAGE has not yet been fully automated and requires well-trained operators, it is presently a challenge to use the 2-D platform in a clinical setting. There have been efforts to develop an automated 2-D PAGE system (NextGenSciences Ltd, Cambridgeshire, UK), although the current application was developed for extensive analytical purposes and not for screening, and the existing laboratory examinations still require experienced staff. Although the use of 2-D PAGE as a tool for routine clinical examination is unfamiliar, there will be a need to develop it for this purpose if the information it provides is of vital importance and cannot be obtained by other types of examination. All these issues should be considered in the development of biomarkers by 2-D PAGE.

In conclusion, we have used 42 cell lines from different lymphoid neoplasms to demonstrate the usefulness of proteomic data obtained by quantitative 2-D PAGE. Multivariate analysis and statistical-learning methods identified unique patterns of protein expression corresponding to the histological subtypes. These results suggest the potential of proteomic studies in the development of clinical biomarkers for lymphoid neoplasms.

*This study was supported by a grant from Pharmaceuticals and Medical Devices Agency of Japan.*

## 5 References

[1] Rosenblatt, K. P., Bryant-Greenwood, P., Killian, J. K., Mehta, A. *et al.*, *Annu. Rev. Med.* 2004, *55*, 97–112.

[2] Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. *et al.*, *Nature Genet.* 1999, *21*, 10–14.

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

[4] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, *18*, 533–537.

[5] Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B. *et al., Mol. Cell. Proteomics* 2002, *1*, 323–333.

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

[7] Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. *et al., J. Proteome Res.* 2003, *2*, 43–50.

[8] Gygi, S. P., Rist, B., Griffin, T. J., Eng, J. *et al., J. Proteome Res.* 2002, *1*, 47–54.

[9] Pawlak, M., Schick, E., Bopp, M. A., Schneider, M. J. *et al., Proteomics* 2002, *2*, 383–393.

[10] Chen, G., Gharib, T. G., Wang, H., Huang, C. C. *et al., Proc. Natl. Acad. Sci. USA* 2003, *100*, 13537–13542.

[11] Chan, J. K., *Hematol. Oncol.* 2001, *19*, 129–150.

[12] Browne, P., Petrosyan, K., Hernandez, A., Chan, J. A., *Am. J. Clin. Pathol.* 2003, *120*, 767–777.

[13] Harris, N. L., Jaffe, E. S., Diebold, J., Flandrin, G. *et al., J. Clin. Oncol.* 1999, *17*, 3835–3849.

[14] Hans, C. P, Weisenburger, D. D., Greiner, T. C., Randy, D. *et al., Blood* 2004, *103*, 275–282.

[15] Rudiger, T., Jaffe, E. S., Delsol, G., deWolf-Peeters, C. *et al., Ann. Oncol.* 1998, S31–S38, 31–38.

[16] Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M. *et al., N. Engl. J. Med.* 2002, *346*, 1937–1947.

[17] Kamesaki, H., Fukuhara, S., Tatsumi, E., Uchino, H. *et al., Blood* 1986, *68*, 285–292.

[18] Drexler, H. G., Gaedicke, G., Lok, M. S., Diehl, V. *et al., Leuk. Res.* 1986, *10*, 487–500.

[19] Schaadt, M., Fonatsch, C., Kirchner, H., Diehl, V., *Blut* 1979, *38*, 185–190.

[20] Bargou, R. C., Mapara, M. Y., Zugck, C., Daniel, P. T. *et al., J. Exp. Med.* 1993, *177*, 1257–1268.

[21] Klein, G., Dombos, L., Gothoskar, B., *Int. J. Cancer* 1972, *10*, 44–57.

[22] Epstein, M. A., Barr, Y. M., *Lancet* 1964, *41*, 252–253.

[23] Klein, G., Giovanella, B., Westman, A., Stehlin, J. S. *et al., Intervirology* 1975, *5*, 319–334.

[24] Epstein, M. A., Achong, B. G., Barr, Y. M., Zajac, B. *et al., J. Natl. Cancer Inst.* 1966, *37*, 547–559.

[25] Hayashi, Y., Matsumura, Y., Nishihira, T., Watanabe, I. *et al., Jpn. J. Exp. Med.* 1980, *50*, 423–434.

[26] Klein, E., Klein, G., Nadkarni, J. S., Nadkarni, J. J. *et al., Cancer Res.* 1968, *28*, 1300–1310.

[27] Harris, N. S., *Nature* 1974, *250*, 507–509.

[28] Drexler, H. G., MacLeod, R. A., Dirks, W. G., *Blood* 2001, *98*, 3495–3496.

[29] Stong, R. C., Korsmeyer, S. J., Parkin, J. L., Arthur, D. C., Kersey, J. H., *Blood* 1985, *65*, 21–31.

[30] Rosenfeld, C., Goutner, A., Choquet, C., Venuat, A. M. *et al., Nature* 1977, *267*, 841–843.

[31] Hurwitz, R., Hozier, J., LeBien, T., Minowada, J. *et al., Int. J. Cancer* 1979, *23*, 174–180.

[32] Beckwith., M., Longo, D. L., O'Connell, C. D., Moratz, C. M., Urba, W. J., *J. Natl. Cancer Inst.* 1990, *82*, 501–509.

[33] Gabay, C., Ben-Bassat, H., Schlesinger, M., Laskov, R., *Eur. J. Haematol.* 1999, *63*, 180–191.

[34] Dyer, M. J., Fischer, P., Nacheva, E., Labastide, W., Karpas, A., *Blood* 1990, *75*, 709–714.

[35] Chang, H., Blondal, J. A., Benchimol, S., Minden, M. D., Messner, H. A., *Leuk. Lymphoma* 1995, *19*, 165–171.

[36] Kluin-Nelemans, H. C., Limpens, J., Meerabux, J., Beverstock, G. C. *et al., Leukemia* 1991, *5*, 221–224.

[37] Epstein, A. L., Herman, M. M., Kim, H., Dorfman, R. F., Kaplan, H. S., *Cancer* 1976, *37*, 2158–2176.

[38] Nacheva, E., Fischer, P. E., Sherrington, P. D., Labastide, W. *et al., Br. J. Haematol.* 1990, *74*, 70–76.

[39] Eton, O., Scheinberg, D. A., Houghton, A. N., *Leukemia* 1989, *3*, 729–735.

[40] Sugamura, K., Fujii, M., Kannagi, M., Sakitani, M. *et al., Int. J. Cancer* 1984, *34*, 221–228.

[41] Ravid, Z., Goldblum, N., Zaizov, R., Schlesinger, M. *et al., Int. J. Cancer* 1980, *25*, 705–710.

[42] McCarthy, R. E., Junius, V., Farber, S., Lazarus, H., Foley, G. E., *Exp. Cell Res.* 1965, *40*, 197–200.

[43] Minowada, J., Ohnuma, T., Moore, G. E., *J. Natl. Cancer Inst.* 1972, *49*, 891–895.

[44] Hiraki, S., Miyoshi, I., Kubonishi, I., Matsuda, Y. *et al., Gann* 1978, *69*, 115–118.

[45] Adams, R. A., Flowers, A., Davis, B. J., *Cancer Res.* 1968, *28*, 1121–1125.

[46] Epstein, A. L., Kaplan, H. S., *Cancer* 1974, *34*, 1851–1872.

[47] Fischer, P., Nacheva, E., Mason, D. Y., Sherrington, P. D. *et al., Blood* 1988, *72*, 234–240.

[48] Su, I. J., Balk, S. P., Kadin, M. E., *Am. J. Pathol.* 1988, *132*, 192–198.

[49] Morgan, R., Smith, S. D., Hecht, B. K., Christy, V. *et al., Blood* 1989, *73*, 2155–2164.

[50] Takeshita, T., Goto, Y., Nakamura, M., Fujii, M. *et al., J. Cell. Physiol.* 1988, *136*, 319–325.

[51] Takeshita, T., Goto, Y., Tada, K., Nagata, K. *et al., J. Exp. Med.* 1989, *169*, 1323–1332.

[52] Gazdar, A. F., Carney, D. N., Bunn, P. A., Russell, E. K. *et al., Blood* 1980, *55*, 409–417.

[53] Yagita, M., Huang, C. L., Umehara, H., Matsuo, Y. *et al., Leukemia* 2000, *14*, 922–930.

[54] Gong, J. H., Maki, G., Klingemann, H. G., *Leukemia* 1994, *8*, 652–658.

[55] Tsuge, I., Morishima, T., Morita, M., Kimura, H. *et al., Clin. Exp. Immunol.* 1999, *115*, 385–392.

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

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

[58] Greer, J. P., Kinney, M. C., Loughran, T. P. Jr., *Hematology* (Am Soc Hematol Educ Program) 2001, 259–281.

[59] Weiss, L. M., Chan, J. K. C., MacLennan., K., Warnke, R. A., in: Mauch, P. M., Armitage, J. O., Diehl, V., Hoppe, R. T., Weiss, L. M. (Eds.), *Hodgkin's Disease.* Lippincott Williams & Wilkins, Philadelphia 1999, pp. 101–120.

[60] Kuppers, R., *Adv. Cancer Res.* 2002, *84*, 277–312.

[61] Kuppers, R., Klein, U., Schwering, I., Distler, V. *et al., J. Clin. Invest.* 2003, *111*, 529–537.

[62] Kondo, M., Weissman, I. L., Akashi, K., *Cell* 1997, *28*, 661–672.

 

[63] Harris, N. L., Jaffe, E. S., Diebold, J., Flandrin, G. *et al.*, *Mod. Pathol.* 2000, *13*, 193–207.

[64] Chittal, S. M., Delsol, G., *Cancer Surv.* 1997, *30*, 87–105.

[65] Wellmann, A., Thieblemont, C., Pittaluga, S., Sakai, A. *et al.*, *Blood* 2000, *96*, 398–404.

[66] Thorns, C., Gaiser, T., Lange, K., Merz, H. *et al.*, *Pathol. Int.* 2002, *52*, 578–585.

[67] Patton, W. F., *J. Chromatogr. B* 2002, *771*, 3–31.

[68] Tonge, R., Shaw, J., Middleton, B., Rowlinson, R. *et al.*, *Proteomics* 2001, *1*, 377–396.

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

[70] Klose, J., Nock, C., Herrmann, M., Stuhler, K. *et al.*, *Nat. Genet.* 2002, *30*, 385–393.

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

[72] Seike, M., Kondo, T., Fujii, K., Okano, T. *et al.*, *Proteomics* 2005, *5*, 2931–2948.

[73] Mitsiades, C. S., Mitsiades, N. S., McMullan, C. J., Poulaki, V. *et al.*, *Cancer Cell* 2004, *5*, 221–230.

 

*Proteomics* 2005, *5*, 2939–2948 DOI 10.1002/pmic.200401166



REGULAR ARTICLE

# Proteomic signatures for histological types of lung cancer

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

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

Received: August 11, 2004
Revised: November 22, 2004
Accepted: November 30, 2004

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

**Keywords:**
Bioinformatics / Laser microdissection / Lung cancer / Two dimensional difference gel electrophoresis

## 1 Introduction

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

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

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

lar backgrounds determining particular histological phenotypes are obscure.

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

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

## 2 Materials and methods

### 2.1 Cell lines, clinical materials and protein extraction

The lung cancer cell lines used had a histological background of: (i) squamous cell carcinoma (PC-1, PC-10, RERF-LC-AI, SQ-5, LC-1/Sq, LC-1F, LK-2, EBC-1, QG-56, and VMRC-LCP); (ii) small cell carcinoma (Lu-130, Lu-134, Lu-135, Lu-139, Lu-140, Lu-165, PC-6, MS-1, SBC-3, and SBC-5); and (iii) adenocarcinoma (A549, PC-3, PC-9, PC-14, RERF-LC-KJ, RERF-LC-MS, RERF-LC-OK, LC-2/ad, ABC-1, and VMRC-LCD). The lung cancer cell lines PC-1, PC-3, PC-6, PC-9, PC-10, and QG-56 were obtained from Immuno-Biological Laboratories (Gunma, Japan). The cell lines A549, PC-14, RERF-LC-KJ, LC-2/ad, SQ-5, LC-1/Sq, LC-1F, RERF-LC-AI, Lu-130, Lu-134, Lu-135, Lu-139, Lu-140, Lu-165, and MS-1 were obtained from RIKEN Cell Bank (Ibaraki, Japan). The cell lines ABC-1, RERF-LC-MS, RERF-LC-OK, LK-2, EBC-1, VMRC-LCD, VMRC-LCP, SBC-3, and SBC-5 were purchased from Health Science Research Resources Bank (Osaka, Japan). All cell lines were maintained in the optimal medium until use. When the cells reached 80–90% confluence, they were washed twice with PBS, scraped off into a tube, and briefly centrifuged. The cell pellets were incubated in a lysis buffer containing 6 M urea, 2 M thiourea, 3% CHAPS, and 1% Triton X-100 for 30 min on ice. After centrifugation at 15 000 rpm for 30 min, the supernatant (cellular protein fraction) was recovered and the protein concentration was measured with a Protein Assay Kit (Bio-Rad, Hercules, CA, USA). The protein sample was adjusted to pH 8.0 with 30 mM Tris.

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

**Table 1.** Clinical variables of lung cancer patients

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

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

### 2.2 Fluorescence labeling of protein samples

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

### 2.3 2-DE

2-D PAGE was performed as described elsewhere with some modifications [13]. Briefly, the fluorescence-labeled proteins were separated by 2-D PAGE, with the first separation by isoelectric point with IEF and the second separation by molecular weight with SDS-PAGE. Each labeled protein sample, volume of 1460 μL was divided into triplicate IPG dry strip gels (24 cm length, p*I* range between 3.0 and 10; Amersham Biosciences); one gel was rehydrated with 420 μ protein sample and each sample was separated in triplicate gels. After rehydration for 12 h, IEF was performed with an IPGphor (Amersham Biosciences) for a total of 80 kVh at 20°C. After IEF, the IPG gels were equilibrated with equilibration buffer containing 6 M urea, 50 mM Tris-HCl (pH 8.8), 30% glycerol, and 1.0% SDS for 15 min at room temperature. The equilibrated IPG gels were applied onto 9–15% polyacrylamide gradient gels and sealed with low melting temperature agarose (Amersham Biosciences), and the proteins were separated at 20°C for 15 h at 17 W *per* 12 gels with an EttanDalt II (Amersham Biosciences). All electrophoresis procedures were performed in the dark.

### 2.4 Image acquisition and quantification of protein spots

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

### 2.5 Multivariate analysis of protein expression profiles

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

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

### 2.6 Identification and functional classification of proteins corresponding to protein spots

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

 

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

## 3 Results

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

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

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

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

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

Figure 2A shows the localization of the 32 protein spots on the 2-D gels. The spots were distributed over the entire gel image. The intensity of some spots was differentially regu-

 

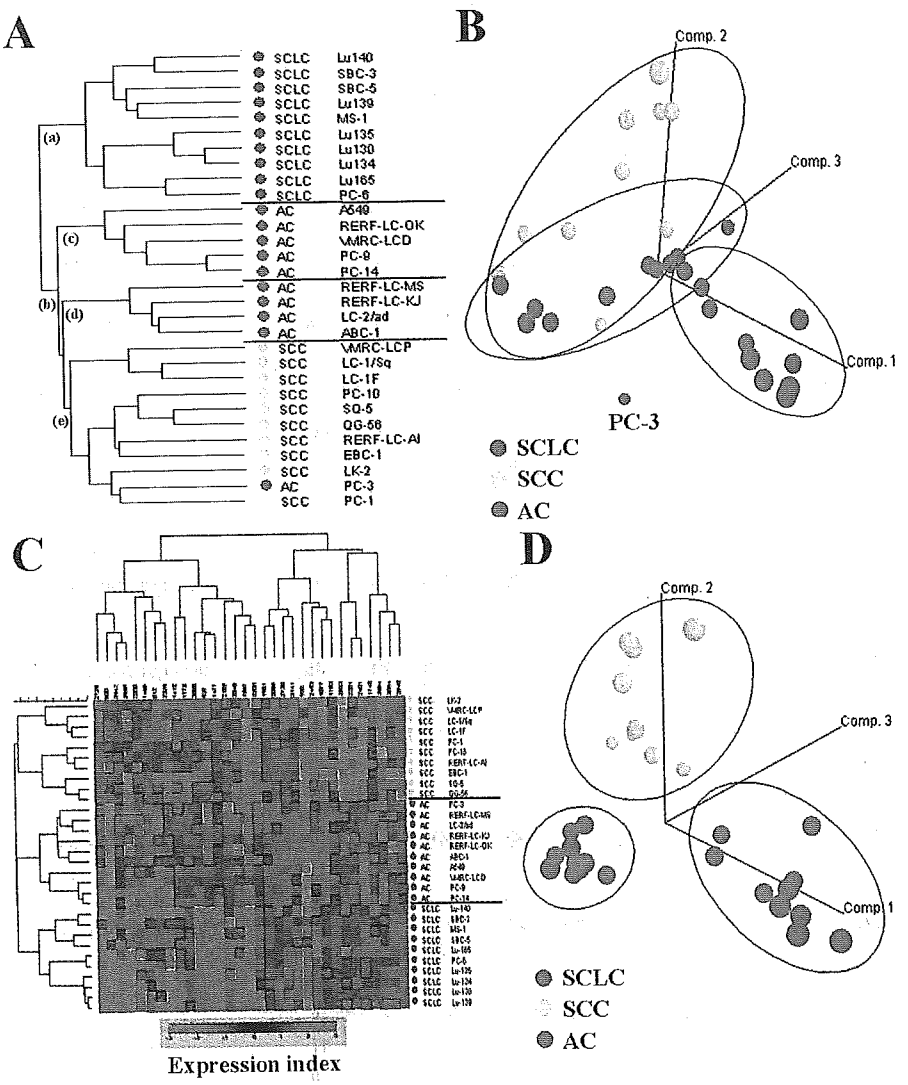


**Figure 1.** Statistical analysis of 30 lung cancer cell lines as a function of their protein expression profiles. (A) Dendrogram of hierarchical clustering analysis. The cell lines and their histology of origin are listed with color-coding on the left. (B) Three-dimensional plot of principal component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

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

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

### 3.3 Laser microdissection of lung cancer tissues and protein expression profile

We examined whether the 32 spots could be used to classify lung cancer cells *in vivo* according to their histological phenotype. Lung cancer tissues contain many types of non-tumor cells, including normal counterpart cells, fibroblasts, various inflammatory cells and proliferating vascular structures. Such cellular heterogeneity of tumor tissues could prevent accurate quantitative expression analysis, because each cell population has its own proteome. For more accurate proteomic analysis, we isolated lung cancer cells by laser microdissection and then extracted the proteins from the cells. Figure 3 shows the process of laser microdissection. The diagnosis was performed with a 10 μm thick tissue section stained with conventional HE staining (left panel), and laser microdissection was used to isolate cells from the neighboring sections stained with hematoxylin. The proteins were extracted from the isolated cells, labeled with Cy5, mixed with the Cy3-labeled internal control mixture and then separated by 2-D PAGE. An area of approximately 1 $mm^2$ of cancer cells was collected for each 2-D image on a large format gel.

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

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

a) Spot numbers correspond to those in Fig. 2
b) Accession no. according to Swiss-Prot
c) MS score was generated by Analyst QS
d) MS/MS score was generated by MASCOT
e) Spots were ranked according their contribution to the classification
f) Proteins were functionally classified according to Amigo ontology
g) Number of spots with which the classification error rate was minimal
h) Average classification accuracy of cross-validation analysis

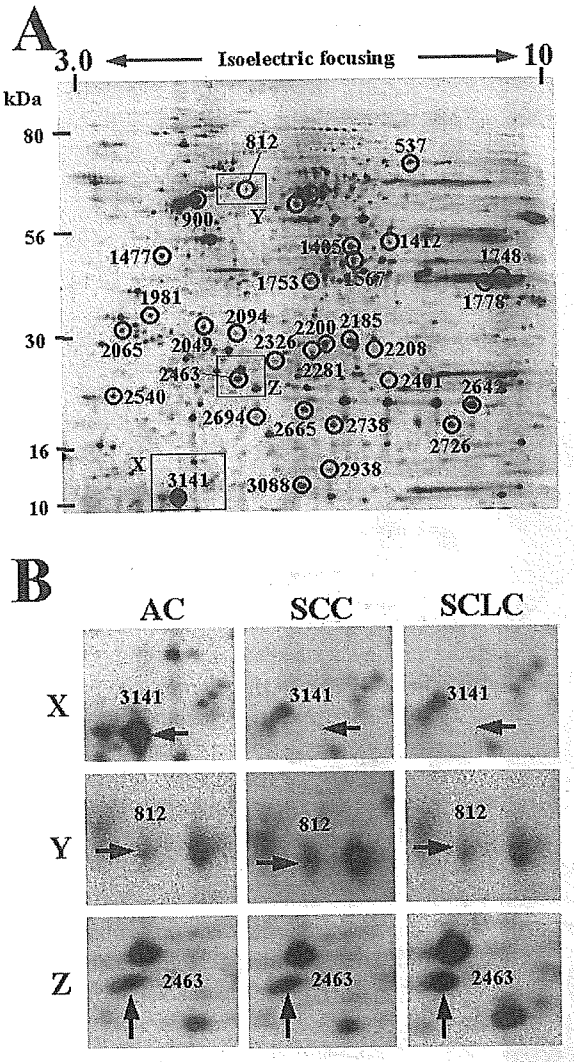


**Figure 2.** (A) Representative 2-D image of a Cy3-labeled protein mixture from 30 lung cancer cell lines. The 32 best-scoring protein spots for classification are circled; spot numbers correspond to those in Table 2. (B) Differential expression of proteins between cells of different origin is shown.

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

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

## 4 Discussion

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

 

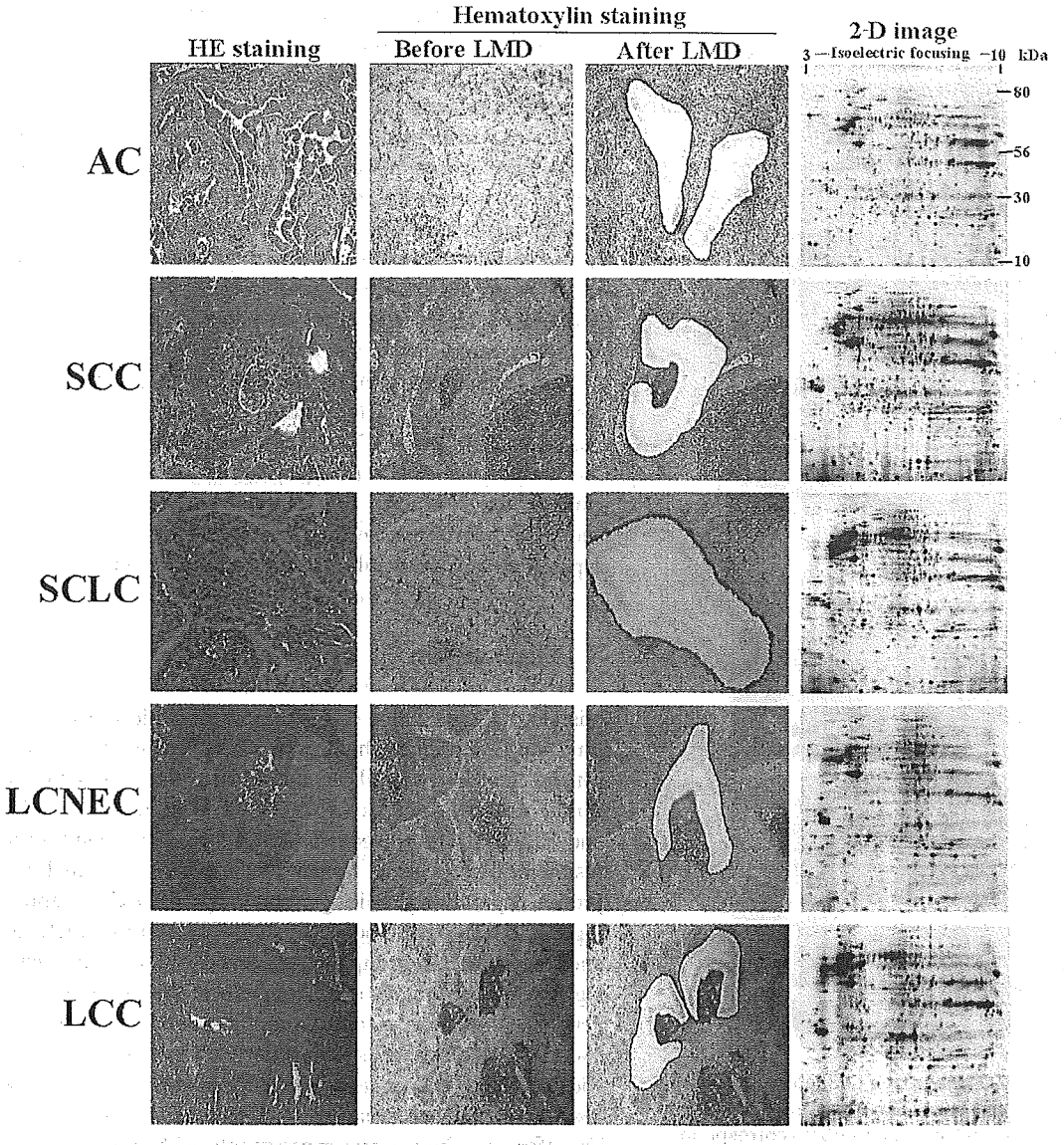


**Figure 3.** 2-D-DIGE of lung cancer cells isolated by laser microdissection. A frozen section of lung cancer tissue was stained with conventional hematoxylin-eosin staining (left panel). The neighboring section was stained with hematoxylin (middle left) and the tumor cells were recovered from that section by means of laser microdissection (middle right). The proteins were extracted from the microdissected lung cancer cells, labeled with the fluorescent dyes for quantitative expression study and separated by 2-D-PAGE (right panel).

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

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

Laser microdissection removed the surrounding stromal cells, which would have affected the protein content of the lung cancer cells. We considered that the effects of the stromal components on the tumor cells would result in alterations of the proteome and that such alterations would remain

 

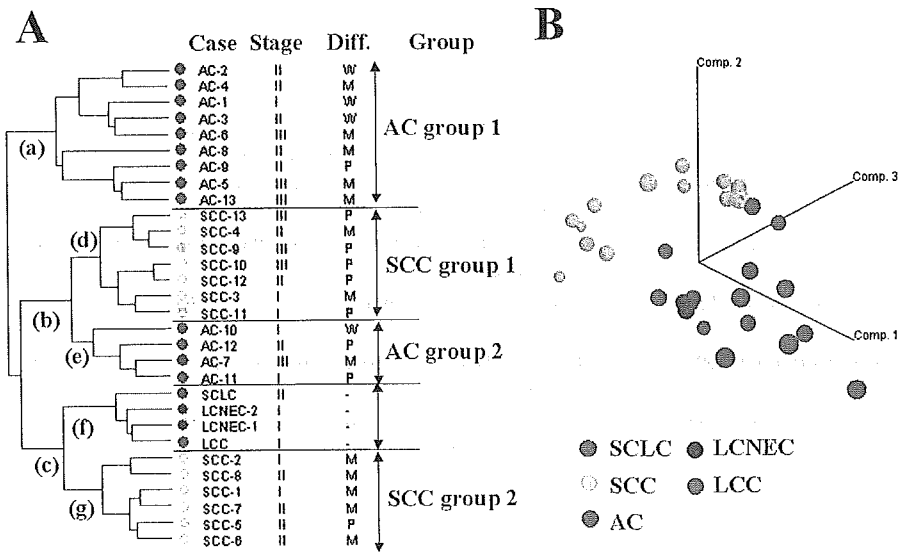


**Figure 4.** Multivariate studies on lung cancer cells obtained by laser microdissection. (A) Dendrogram of hierarchical clustering analysis of all lung cancer samples on the basis of 32 protein spots. The cell samples and prior information about their histology of origin are listed with color-coding on the left. (B) PCA of all lung cancer samples on the basis of 32 protein spots. The cell samples were plotted in three-dimensional space as a function of the similarity of their expression profiles.

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

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

Proteomic classification of lung cancer cells resulted in the unexpected identification of a subgroup of SCC with advanced clinical stage, suggesting that the subgroups of SCCs reflect their malignancy. However, the sample size we used was not sufficient for statistical evaluation of our speculation, and further large-scale studies will be required to confirm these possibilities. Recently, proteomic approaches have been employed to develop prognostic tumor markers for lung cancer. Using 2-D PAGE, Hanash's group reported that a set of 20 protein spots could predict the survival of patients with lung adenocarcinoma [10]. MALDI-TOF MS has been used to identify a peptide expression pattern from which the survival of NSCLC patients could be predicted [12]. Our results could support the idea that current proteomic technologies can capture protein expression patterns corresponding to the clinical features of lung cancer and that such patterns will be useful to establish therapeutic strategies. The protein expression patterns corresponding to the subgroups with poor survival or different therapeutic responses should be considered in future studies. The patterns of tumors after chemotherapy, with and without preceding radiotherapy, should also be studied. As the proteins involved in these patterns are strongly associated with certain clinical features of lung cancer, studies on those proteins will lead to further understanding of the biology of this disease.

*This study was supported by a grant from Pharmaceuticals and Medical Devices Agency of Japan.*

## 5 References

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

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

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

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

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

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

 