is illustrated in Fig. 6, which is published as supporting information on the PNAS web site.

Biopsy specimens obtained from 20 rheumatoid arthritis patients who had been taking nonsteroidal anti-inflammatory drugs (NSAIDs) were evaluated according to a system (26) modified from Dixon *et al.* (27) for grading of chemical gastritis. Five histological features, which are composed of (*i*) foveolar hyperplasia, (*ii*) edema and prominence of muscle fibers in the lamina propria, (*iii*) vasodilatation and congestion, (*iv*) inverse scale for polymorphonuclear neutrophil activity, and (*v*) inverse scale for chronic lymphocyte infiltration, were graded from 0 (normal or absent) to 3 (marked). Each grade of five categories was added to provide a total chemical score that could range from 0 to 15. Scores from 11 to 15 were considered diagnostic of chemical gastritis. The Ethical Committee of Shinshu University School of Medicine approved these study plans.

**Antibodies and Immunohistochemistry.** The antibodies used in this study were CSLEX-1 (mouse IgM, Pharmingen), HECA-452 (rat IgM, Pharmingen), MECA-79 (rat IgM, Pharmingen), NCC-ST-439 (mouse IgM, Nippon Kayaku, Tokyo), mouse IgG anti-human CD31 (DAKO), QBEND10 (mouse IgG directed to human CD34, Immunotech, Luminy, France), and rabbit anti-*H. pylori* polyclonal (DAKO). Immunostaining was performed by using the EnVision$^+$ system (DAKO). Briefly, serial sections were deparaffinized and rehydrated, and endogenous peroxidase activity was blocked by soaking in 0.3% $H_2O_2$ methanol solution. Before immunostaining, the antigens were retrieved by incubating sections in a microwave in 10 mM Tris·HCl buffer (pH 8.0) containing 1 mM EDTA for CD31 and CD34 or by 5 min of treatment with 20 $\mu$g/ml proteinase K for *H. pylori*. The tissue sections were blocked with 3% FBS in PBS and incubated with primary antibody. After being washed in PBS, the slides were incubated with horseradish peroxidase- and secondary antibody-conjugated polymer, EnVision$^+$ (DAKO). The color reaction was developed with 3,3′-diaminobenzidine containing 0.02% $H_2O_2$ (Vector). Sections were briefly counterstained with hematoxylin. Negative controls were obtained by omitting the primary antibodies. Tonsil tissue was used as a positive control (Fig. 7, which is published as supporting information on the PNAS web site).

The numbers of CD34$^+$, MECA-79$^+$, and HECA-452$^+$ vessels in 10 high-power fields of view with ×400 magnification were obtained. The numbers of MECA-79$^+$ and HECA-452$^+$ vessels each were divided by the number of CD34$^+$ vessels, yielding the percentages of MECA-79$^+$ and HECA-452$^+$ vessels, respectively, as described in ref. 28.

**Construction of L- and E-Selectin·IgM Chimeras.** The Fc region of human IgM was amplified with the oligonucleotides 5′-CGGGATCCTGTGATTGCTGAGCTGCCTCCCA-3′ and 5′-GCTCTAGATCAGTAGCAGGTGCCAGCTGTGT-3′ by using pcDNA1-mouse P-selectin·IgM (29) as a template and subcloned into the *Bam*HI/*Xba*I site of pcDNA1.1 (pcDNA1.1-IgM). For construction of pcDNA1.1-human L-selectin·IgM, the 5′ end of L-selectin was excised from pCDM8-human L-selectin·IgG by *Eco*RI digestion and blunted with Klenow fragment (Roche). After digestion of the 3′ end with *Bam*HI, the excised cDNA was subcloned into the blunted *Hin*dIII site and *Bam*HI site of pcDNA1.1-IgM, to form pcDNA1.1-human L-selectin·IgM. Similarly, the PCR product obtained by using pcDNA1-human E-selectin·IgG (30) as a template, was subcloned into pcDNA1.1-IgM, forming pcDNA1.1-human E-selectin·IgM.

**Selectin·IgM Chimeric Protein-Binding Assay.** Human L-selectin·IgM and E-selectin·IgM chimeric proteins were obtained from the cultured medium of COS-1 cells after transient transfection of pcDNA1.1-L-selectin·IgM and pcDNA1.1-E-selectin·IgM as described in ref. 31.

After blocking endogenous peroxidase activity as described above, we incubated tissue sections with selectin·IgM chimeric protein and rinsed them with PBS containing $Ca^{2+}$. The sections were then incubated with horseradish peroxidase-conjugated goat anti-human IgM antibody (Pierce). After being rinsed in PBS containing $Ca^{2+}$, the color reaction was developed with 3,3′-diaminobenzidine as described above. The sections were briefly counterstained with hematoxylin. A control experiment was done with PBS containing 1 mM EDTA.

**Statistical Analysis.** The statistical difference in the percentages of positive vessels in different chronic infiltrate grades was analyzed by the Kruskal–Wallis test followed by Dunn's post hoc test. The difference of proportions (percentage of patients) was analyzed by Fisher's exact test. In both tests, $P$ values at $<0.05$ are considered significant.

**Stable Expression of P-Selectin Glycoprotein Ligand 1 (PSGL-1), Fucosyltransferase VII (FucT-VII), Core 1 Extension β1,3-*N*-acetylglucosaminyltransferase (Core1-β3GlcNAcT), Core 2 β1,6-*N*-acetylglucosaminyltransferase I (Core2GlcNAcT-I), and L-Selectin Ligand Sulfotransferase (LSST) in CHO Cells.** CHO cells stably expressing PSGL-1 and Core1-β3GlcNAcT, Core2GlcNAcT-I, or FucT-VII were established as described in ref. 32. CHO cells stably expressing PSGL-1, Core1-β3GlcNAcT, and FucT-VII, CHO-PSGL-1·C1·F7 (32) were further stably transfected with pcDNA1.1-LSST (30) and pCMV/Bsd (Invitrogen) and selected with Blasticidin S (Invitrogen). The cells were sorted and cloned for MECA-79$^+$ staining to obtain CHO-PSGL-1·C1·F7·LSST.

CHO-PSGL-1·C2·F7 cells (32) were transfected with pcDNA1.1-LSST. Cloned cells positive for HECA-452 and negative for CSLEX-1 were chosen as CHO-PSGL-1·C2·F7·LSST. Previously, it was shown that 6-sulfo sialyl Lewis X is recognized by HECA-452 antibody but not by CSLEX-1 antibody (33). Similarly, CHO-PSGL-1·C2·F7·LSST cells were stably transfected with Core1-β3GlcNAcT, yielding MECA-79$^+$ CHO-PSGL-1·C1·C2·F7·LSST.

## Results

**HEV-Like Vessels Are Induced in *H. pylori*-Induced Inflammation.** Because it has been reported that *de novo* formation of HEV-like vessels, which express PNAd, is associated with various chronic inflammatory diseases, we determined whether chronic inflammation caused by *H. pylori* infection is also associated with the formation of HEV-like vessels. To do so, gastric mucosa from patients infected with *H. pylori* was stained with MECA-79 antibody, which reacts with 6-sulfo sialyl Lewis X on extended core 1, and HECA-452 antibody, which reacts equally well with sialyl Lewis X and 6-sulfo sialyl Lewis X-capping structure on extended core 1 and core 2 branches (Fig. 1; see also Fig. 8, which is published as supporting information on the PNAS web site). As shown in Fig. 2, gastric mucosa derived from *H. pylori*-infected patients displayed HEV-like vessels expressing MECA-79 and HECA-452 antigens, as well as CD31 and CD34, which are markers for vascular endothelial cells. Moreover, these HEV-like vessels can potentially recruit L-selectin-expressing lymphocytes, because L-selectin·IgM chimeric protein bound the same vessels in a $Ca^{2+}$-dependent manner. Indeed, a large number of B and T lymphocytes were recruited in the gastric mucosa infected with *H. pylori* (Fig. 9, which is published as supporting information on the PNAS web site). These results indicate that *H. pylori*-induced inflammation is associated with the formation of PNAd present on HEV-like vessels.

The above results demonstrated that 6-sulfo sialyl Lewis X on extended core 1 O-glycans is present based on positive staining by MECA-79 and HECA-452 antibodies. To elaborate further the chemical nature of L-selectin ligands on these HEV-like

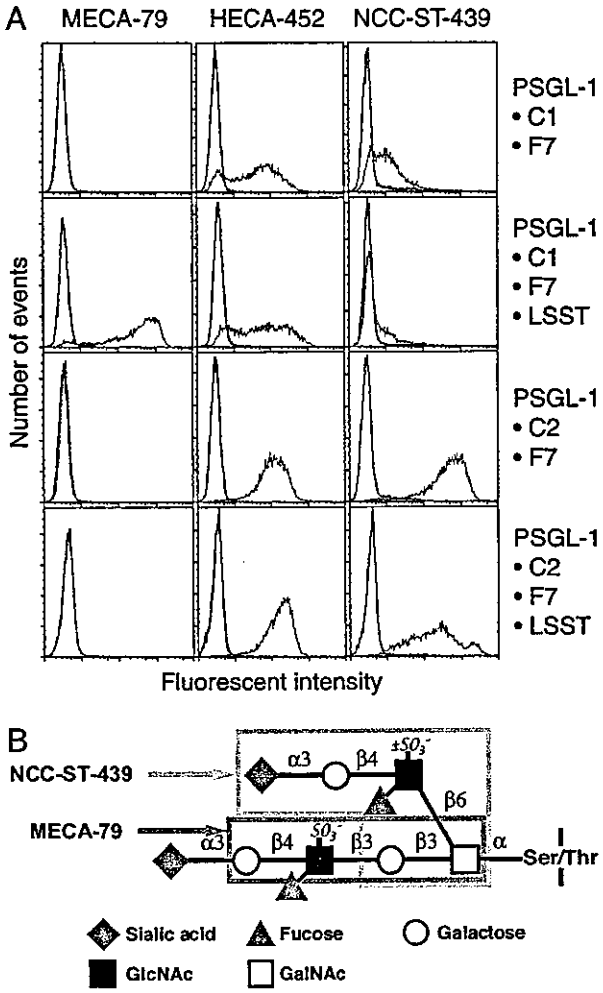


Fig. 1. FACS analysis of transfected CHO cells expressing sialyl Lewis X or 6-sulfo sialyl Lewis X on extended core 1 or core 2 branched O-glycans (32) and structures of MECA-79 and NCC-ST-439 epitopes are shown. (*A*) MECA-79 binds CHO cells expressing 6-sulfo sialyl Lewis X on extended core 1 (CHO-PSGL-1·C1·F7·LSST) but not sialyl Lewis X (CHO-PSGL-1·C1·F7). NCC-ST-439 binds CHO cells expressing sialyl Lewis X (CHO-PSGL-1·C2·F7) and 6-sulfo sialyl Lewis X (CHO-PSGL-1·C2·F7·LSST) on core 2 branched O-glycans but barely binds those expressing sialyl Lewis X or 6-sulfo sialyl Lewis X on extended core 1 O-glycans. HECA-452 binds all cells tested, but its reactivity apparently depends on the expression level of extended core 1 O-glycans. (*B*) L-selectin ligand containing 6-sulfo sialyl Lewis X on core 2 branch and extended core 1 structure is shown. The epitopes for NCC-ST-439 and MECA-79 are shown in boxes.

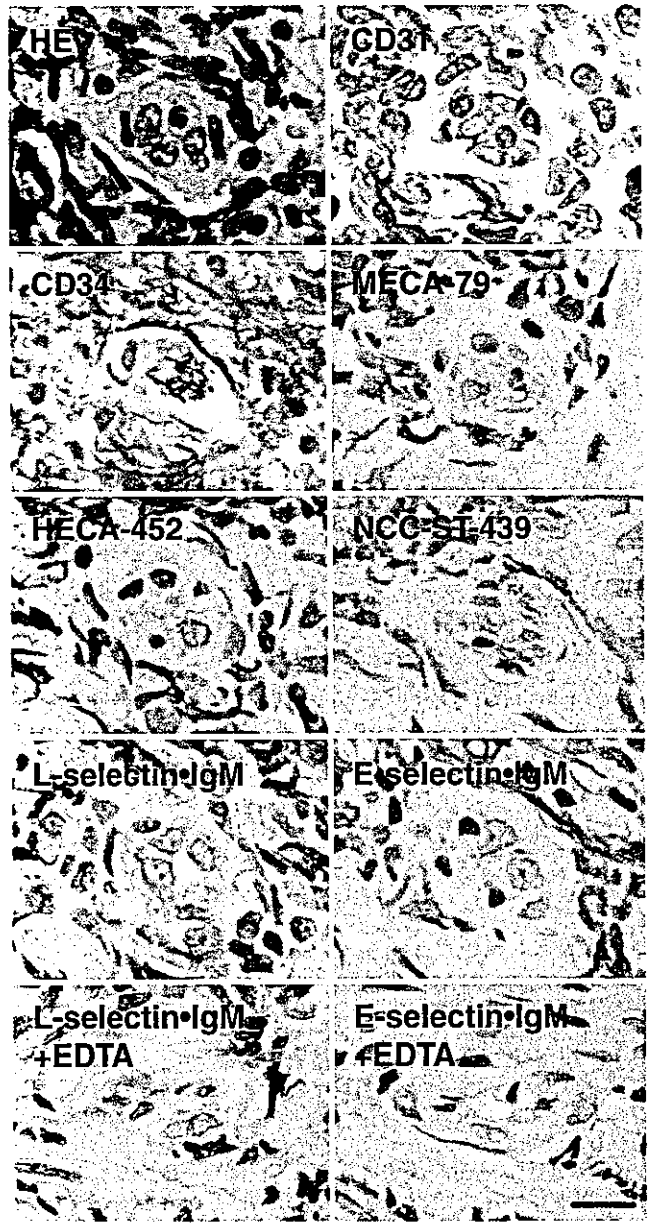


Fig. 2. MECA-79, HECA-452, and NCC-ST-439 antigens in a HEV-like vessel in the gastric mucosa with *H. pylori*-associated chronic active gastritis. Serial sections were subjected to immunostaining with anti-CD31, anti-CD34, MECA-79, HECA-452, and NCC-ST-439 antibodies and to a binding assay with L-selectin-IgM and E-selectin-IgM chimeric proteins in the absence or presence of 1 mM EDTA. HE, hematoxylin and eosin staining. (Bar, 10 $\mu$m.)

vessels, the NCC-ST-439 monoclonal antibody was used. NCC-ST-439 antibody binding has been tested for sialyl Lewis X-capping structure on Gal$\beta$1→4GlcNAc$\beta$1→6GalNAc$\alpha$1→R but not on natural core 2 branched O-glycan Gal$\beta$1→4GlcNAc$\beta$1→6(**Gal$\beta$1→3**)GalNAc$\alpha$1→R (missing Gal$\beta$1→3 shown in bold) (24). Moreover, it has not been determined whether 6-sulfo sialyl Lewis X is also recognized by this antibody. To test these possibilities, we made CHO cells expressing various types of O-glycans and stained the cells with NCC-ST-439 antibody. Fig. 1 illustrates that NCC-ST-439 antibody binds to CHO cells expressing nonsulfated and 6-sulfo sialyl Lewis X on core 2 branched O-glycans but barely to CHO cells expressing those capping structures on extended core 1 O-glycans. Fig. 2 shows that NCC-ST-439 antibody can also stain HEV-like vessels formed in the gastric mucosa. These results combined suggest that PNAd induced by *H. pylori* infection expresses 6-sulfo sialyl Lewis on both extended core 1 and core 2 branched structures in the same manner as PNAd expressed in the secondary lymphoid organs (21).

**Increased Formation of HEV-Like Vessels Is Correlated with Progression of Inflammation.** Based on morphological examination, progression of inflammation initiated by *H. pylori* infection can be roughly divided into four stages i.e., normal, mild, moderate, and marked (25). In the moderate and marked stages, intestinal metaplasia frequently takes place, which indicates an advanced stage of the disease.

Fig. 3*A Lower* indicates a marked stage of inflammation in which recruitment of mononuclear cells obscured proper glands in gastric mucosa compared with glands shown in mucosa at mild

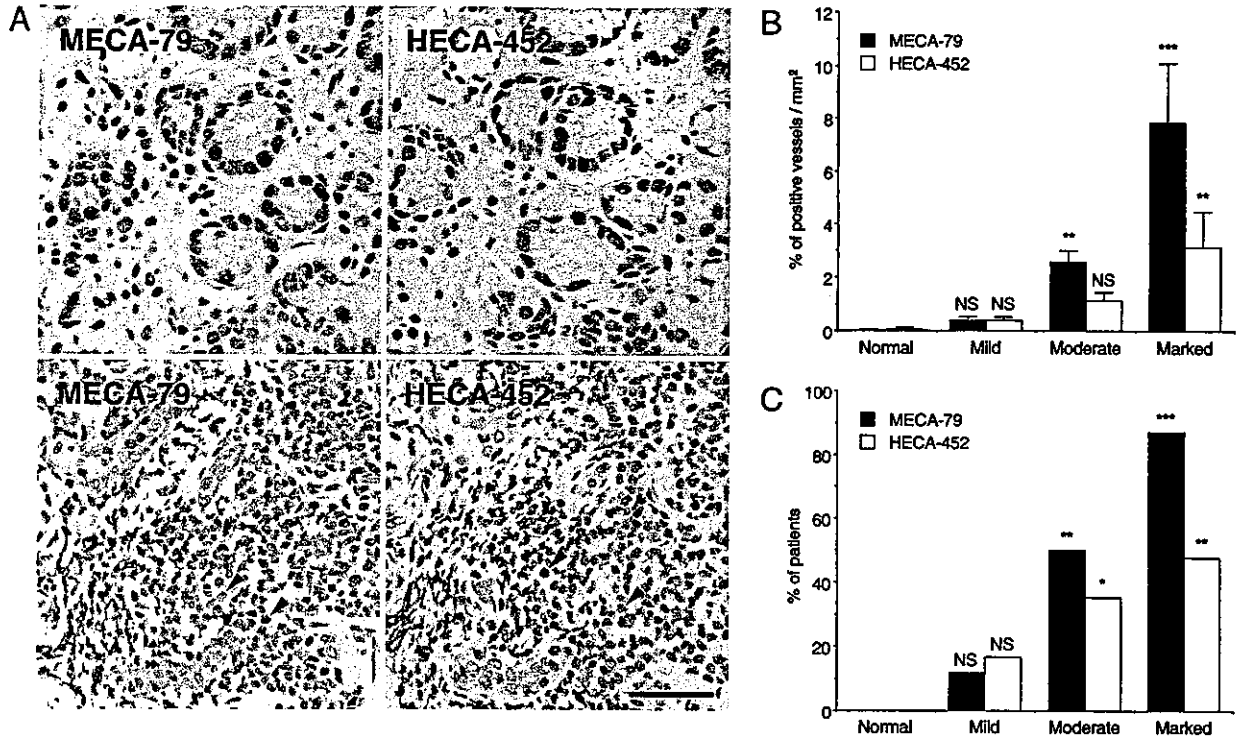


Fig. 3. Gastric mucosa of different degrees of chronic inflammation and association of HEV-like vessels with progression of inflammation. (*A*) (*Upper*) Gastric mucosa at a mild stage barely expresses HEV-like vessels with minimum recruitment of lymphocytes. (*Lower*) Gastric mucosa at a marked stage express a significant number of recruited lymphocytes (arrowheads) around HEV-like vessels. (*B*) The number of MECA-79$^+$ or HECA-452$^+$ vessels is positively correlated with the progression of chronic inflammation. Each group consists of 11 (normal), 42 (mild), 67 (moderate), and 23 (marked) patients. (*C*) The number of patients exhibiting ≥1% MECA-79$^+$ or HECA-452$^+$ vessels is highly correlated with the progression of chronic inflammation. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant. (Bar, 50 μm.)

stage (Fig. 3*A Upper*). These observations demonstrate that lymphocyte infiltration is more prominent when HEV-like vessels are more abundant.

After examining >140 human specimens, we found that the number of HEV-like vessels, as detected by MECA-79 or HECA-452 antibody, is positively correlated with the progression of inflammation (Fig. 3*B* and Table 1, which is published as supporting information on the PNAS web site). Fig. 3*C* illustrates that more patients display HEV-like vessels as the inflammation progresses. *H. pylori* was detected in 0%, 21%, 82%, and 87% of patients in normal, mild, moderate, and marked stages of the inflammation, respectively. Overall, HEV-like vessels were found in 79.2% of *H. pylori*-infected patients. These results combined indicate that progression of inflammation, initiated by *H. pylori* infection, is highly correlated with formation of HEV-like vessels at the gastric mucosa.

**Formation of HEV-Like Vessels Depends on Continuous *H. pylori* Infection.** To determine whether the formation of HEV-like vessels is correlated with the presence of *H. pylori* infection, gastric biopsies were obtained from 17 patients with chronic active gastritis before and after eradication of *H. pylori* by treatment with antibiotics and a proton pump inhibitor. Patients with a moderate inflammation stage displayed both *H. pylori* and HEV-like vessels detected by MECA-79 and HECA-452 antibodies (Fig. 4*A*). After eradication of *H. pylori*, the gastric mucosa of all of the patients examined no longer displayed HEV-like vessels as assessed by MECA-79 and HECA-452 staining and showed minimum lymphocyte infiltration (Fig. 4*B*). These results indicate that continuous infection of *H. pylori* is necessary for formation and maintenance of HEV-like vessels expressing PNAd.

**HEV-Like Vessels in NSAIDs-Induced Gastritis.** To determine whether HEV-like vessels are induced in chronic inflammatory response by causes other than *H. pylori* infection, gastric mucosa obtained from long-standing rheumatoid arthritis patients taking NSAIDs were examined, because it is known that continuous use of NSAIDs results in chemical gastritis. The majority of 20 patients examined exhibited chemical gastritis phenotype and were devoid of HEV-like vessels (Fig. 5*A*; see also Fig. 10*A*, which is published as supporting information on the PNAS web site). HEV-like vessels were found in the specimens from 6 of 20 patients, but 3 of these patients were also infected with *H. pylori*. All of those three patients display lower scores for chemical gastritis, and HEV-like vessels were likely formed by inflammation caused by *H. pylori* infection. In three *H. pylori*-free patients, HEV-like vessels were found in only 0.68%, 0.67%, and 0.21% of CD34$^+$ vessels, and two of them displayed intestinal metaplasia, suggesting a possible prior infection of *H. pylori*. Fig. 5*B* illustrates that the remaining patient displayed both chemical gastritis phenotype and lymphocyte recruitment. Interestingly, the intensity of MECA-79 staining was much less for this patient than for those infected with *H. pylori* (compare Fig. 5*B* with Figs. 2–4). These results indicate that chemical gastritis negligibly induces PNAd in the gastric mucosa.

## Discussion

The present studies demonstrate that *H. pylori*-induced inflammation is associated with the recruitment of lymphocytes through *de novo* formation of PNAd on HEV-like vessels. HEV-like vessels are absent in human gastric mucosa under normal conditions. Once chronic inflammation occurs, these vessels, which bind L-selectin-IgM chimeric protein in a $Ca^{2+}$-dependent manner, appear. Our observations of *de novo* for-

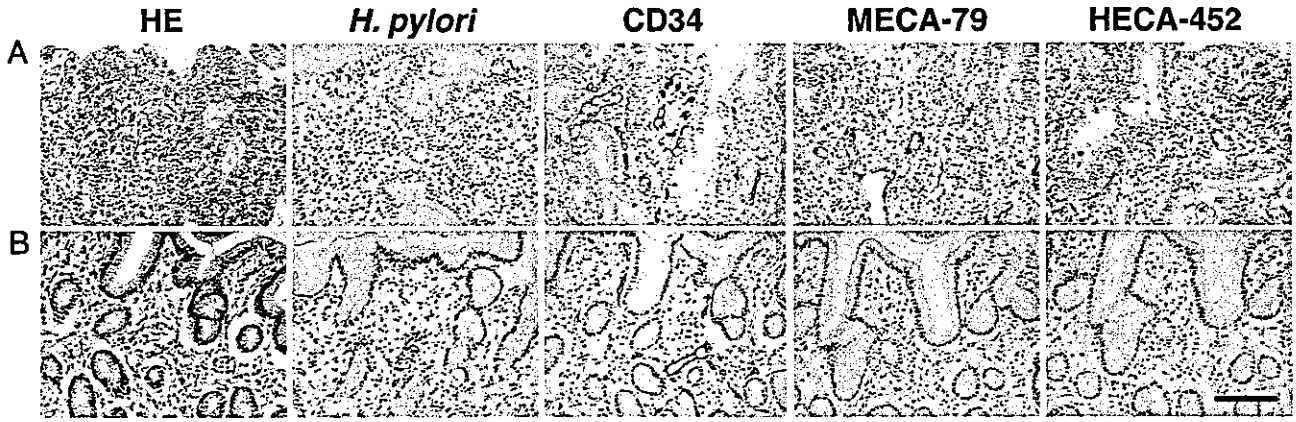


Fig. 4. Disappearance of HEV-like vessels in the gastric mucosa after eradication of *H. pylori*. Gastric mucosa infected by *H. pylori* was examined before and 2 months after a treatment to eradicate *H. pylori*. (*A*) Before the treatment, HEV-like vessels detected by MECA-79 and HECA-452 antibodies were abundant, and large numbers of mononuclear cells (lymphocytes) were present around these vessels. (*B*) After eradication of *H. pylori*, HEV-like vessels were no longer present and few mononuclear cells were present. (Bar, 100 $\mu$m.)

mation of HEV-like vessels are similar to those described previously in various human chronic inflammatory diseases, such as rheumatoid arthritis, lymphocytic thyroiditis, inflammatory bowel diseases, and in acute heart allograft rejection (14–17, 34). However, none of the previous studies showed disappearance of HEV-like vessels by removing causes that led to their formation. We have established here that the abundance of HEV-like vessels is highly correlated with the progression of *H. pylori*-associated chronic active gastritis, and that HEV-like vessels are no longer expressed after *H. pylori* is eradicated. By contrast, NSAIDs-induced gastritis patients display only very few HEV-like vessels, if any.

In the present study, we showed that HEV-like vessels are positive for MECA-79 and HECA-452 staining, suggesting that 6-sulfo sialyl Lewis X on the extended core 1 structure is present on PNAd. We also found that HEV-like vessels were stained by NCC-ST-439, the intensity of which was similar to that on neutrophils expressing sialyl Lewis X on core 2 branched O-glycans (35, 36). These results suggest that HEV-like vessels in the gastric mucosa should contain 6-sulfo sialyl Lewis X and sialyl Lewis X on extended core 1 and core 2 branched O-glycans. In previous studies, MECA-79 has been frequently used to characterize PNAd on HEV-like vessels. Although HECA-452 antibody was also used in one study, no structural basis for differential staining was inferred (28). By using HECA-452 and NCC-ST-439 antibodies in addition to MECA-79 antibody, we have obtained more detailed knowledge of the chemical nature of L-selectin ligands on HEV-like vessels induced by *H. pylori* infection. Because it is difficult to isolate a sufficient number of cells expressing PNAd from many pathological specimens, a combination of MECA-79, HECA-452, and NCC-ST-439 staining would be useful to determine the chemical nature of HEV-like vessels formed in these chronic inflammatory diseases.

It has been shown that the *H. pylori* adhesin Sab A binds sialyl dimeric Lewis X-bearing glycosphingolipids in the surface mucous cells (37). The expression of L-selectin ligands including 6-sulfo sialyl Lewis X may thus enhance adhesion of *H. pylori* to mucosa expressing HEV-like vessels. At the same time, colonization of *H. pylori* on the gastric mucosa may contribute to gastritis by producing autoantibodies, because both *H. pylori* and acid-secreting parietal cells share Lewis X antigen (38). Core 2 branched O-glycans are capped by $\alpha$1,4-linked *N*-acetylglucosamine in the gastric mucosa, and those O-glycans are secreted from gland mucous cells (39). Strikingly, our recent studies demonstrate that this glandular mucin containing $\alpha$1,4-GlcNAc-capped O-glycans inhibits *H. pylori* growth and thus acts as a natural antibiotic against *H. pylori* infection, whereas MUC5AC secreted from surface mucous cells stimulates *H. pylori* growth

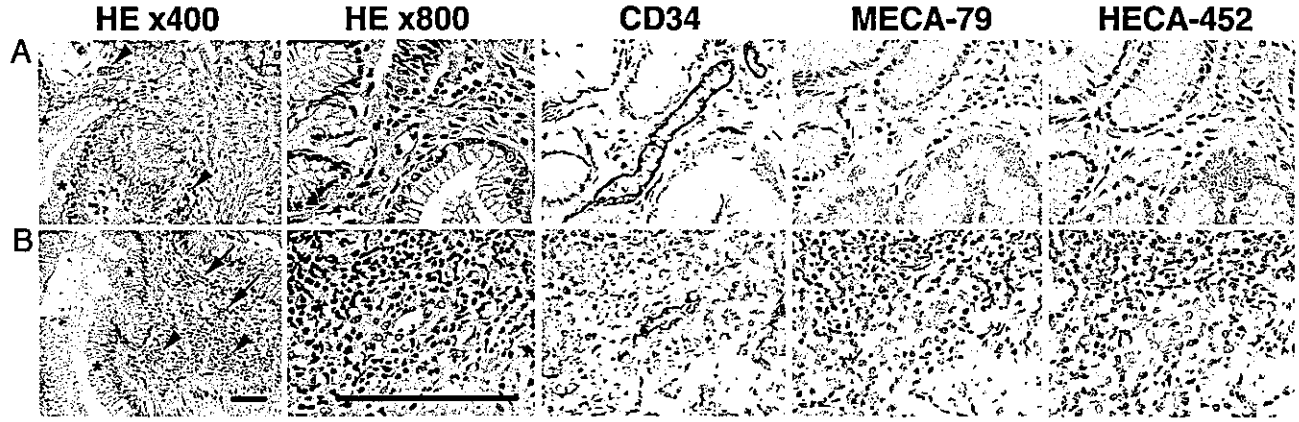


Fig. 5. Few HEV-like vessels are associated with chemical gastritis caused by NSAIDs. (*A* and *B*) Gastric mucosa obtained from two rheumatoid arthritis patients taking NSAIDs. In both patients, typical chemical gastritis phenotype, such as foveolar hyperplasia (asterisks) and vasodilatation and congestion (arrowheads), is evident. (*B*) MECA-79$^+$ HEV-like vessels are associated only with substantial lymphocyte recruitment (arrows), which is rather atypical for chemical gastritis. Images for HE ×800, CD34, MECA-79, and HECA-452 are further enlarged in the same scale to show details. (Bar, 50 $\mu$m.)

(39). The results are consistent with the observation that *H. pylori* rarely colonizes the layer where gland mucous cell-derived mucins are present (2). These combined results suggest that the expression of sialyl Lewis X may facilitate *H. pylori* infection and resultant inflammation, whereas $\alpha$1,4-GlcNAc-capped O-glycans in the gland mucous cells inhibits this process.

The present study also demonstrated that HEV-like vessels disappear once *H. pylori* is eradicated by antibiotic treatment. The results indicate that the formation of HEV-like vessels and thus recruitment of lymphocytes to chronic inflammatory sites in the gastric mucosa can be reversed once *H. pylori* infection is abolished. It is tempting to speculate that lipopolysaccharides and/or other components from the organism, acting through Toll-like receptor-dependent pathways in the gastric epithelium, exerts the elaboration of cytokines, i.e., lymphotoxin $\alpha$. This effect might in turn modulate gene expression in postcapillary venules in ways that cause their biochemical, functional, and morphological transformation by up-regulating chemokines, such as CCL19 and CCL21, that act on CCR7 receptors.

Several enzymes play major roles in the biosynthesis of PNAd at the secondary lymphoid organs, including LSST (also called GlcNAc6ST-2 or HEC-GlcNAc6ST) (30, 40), Core2GlcNAcT-I (21, 31), and Core1-$\beta$3GlcNAcT (21, 32). Mice that are mutant in any of these enzymes exhibit reduced lymphocyte homing activity compared with wild-type mice (31, 32, 40). Future studies that determine whether inhibition of L-selectin ligand synthesis can inhibit the chronic inflammation induced by *H. pylori* infection and, hence, the formation of gastric carcinoma would be of great interest. Mice with targeted mutations in various glycosyltransferases and sulfotransferases responsible for L-selectin ligand synthesis will likely provide critical tools for those studies.

We thank Yasuyo Shimojo and Akiko Ishida for preparation of tissue sections, Dr. John Lowe (University of Michigan Medical School, Ann Arbor) for providing pcDNA1-P-selectin-IgM, Dr. Elise Lamar for critical reading of the manuscript, Yoav Altman for assisting in FACS analysis, and Aleli Morse for organizing the manuscript. This work was supported by the National Institutes of Health Grant R37 CA33000 (to M.F.) and by Ministry of Education, Culture, Sports, Science, and Technology of Japan Priority Area 14082201 and the Ministry of Health, Labor, and Welfare of Japan (Third Term Comprehensive Control Research for Cancer) (to J.N.).

1. Marshall, B. J. & Warren, J. R. (1984) *Lancet* **i**, 1311–1315.
2. Hidaka, E., Ota, H., Hidaka, H., Hayama, M., Matsuzawa, K., Akamatsu, T., Nakayama, J. & Katsuyama, T. (2001) *Gut* **49**, 474–480.
3. Sipponen, P. & Hyvarinen, H. (1993) *Scand. J. Gastroentrol.*, Suppl., **196**, 3–6.
4. Peek, R. M., Jr. & Blaser, M. J. (2002) *Nat. Rev. Cancer* **2**, 28–37.
5. Yuasa, Y. (2003) *Nat. Rev. Cancer* **3**, 592–600.
6. Freund, J. N., Domon-Dell, C., Kedinger, M. & Duluc, I. (1998) *Biochem. Cell. Biol.* **76**, 957–969.
7. Hinoi, T., Lucas, P. C., Kuick, R., Hanash, S., Cho, K. R. & Fearon, E. R. (2002) *Gastroenterology* **123**, 1565–1577.
8. Bayerdorffer, E., Neubauer, A., Rudolph, B., Thiede, C., Lehn, N., Eidt, S. & Stolte, M. (1995) *Lancet* **345**, 1591–1594.
9. Nomura, A., Stemmermann, G. N., Chyou, P. H., Kato, I., Perez-Perez, G. I. & Blaser, M. J. (1991) *N. Engl. J. Med.* **325**, 1132–1136.
10. Parsonnet, J., Freidman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N. & Sibley, R. K. (1991) *N. Engl. J. Med.* **325**, 1127–1131.
11. Higashi, H., Tsutsumi, R., Muto, S., Sugiyama, T., Azuma, T., Asaka, M. & Hatakeyama, M. (2002) *Science* **295**, 683–686.
12. Rosen, S. D. (2004) *Annu. Rev. Immunol.* **22**, 129–156.
13. von Andrian, U. H. & Mackay, C. R. (2000) *N. Engl. J. Med.* **343**, 1020–1034.
14. Duijvestijn, A. M., Kerkhove, M., Bargatze, R. F. & Butcher, E. C. (1987) *J. Immunol.* **138**, 713–719.
15. Kabel, P. J., Voorbij, H. A., de Haan-Meulman, M., Pals, S. T. & Drexhage, H. A. (1989) *J. Clin. Endocrinol. Metab.* **68**, 744–751.
16. van Dinther-Janssen, A. C., Pals, S. T., Scheper, R., Breedveld, F. & Meijer, C. J. (1990) *J. Rheumatol.* **17**, 11–17.
17. Salmi, M., Granfors, K., MacDermott, R. & Jalkanen, S. (1994) *Gastroenterology* **106**, 596–605.
18. Streeter, P. R., Rouse, B. T. & Butcher, E. C. (1988) *J. Cell. Biol.* **107**, 1853–1862.
19. Mazzucchelli, L., Blaser, A., Kappeler, A., Schärli, P., Laissue, J. A., Baggiolini, M. & Uguccioni, M. (1999) *J. Clin. Invest.* **104**, R29–R54.
20. Dogan, A., Du, M., Koulis, A., Briskin, M. J. & Isaacson, P. G. (1997) *Am. J. Pathol.* **151**, 1361–1369.
21. Yeh, J. C., Hiraoka, N., Petryniak, B., Nakayama J., Ellies, L. G., Rabuka, D., Hindsgaul, O., Marth, J. D., Lowe, J. B. & Fukuda, M. (2001) *Cell* **105**, 957–969.
22. Hemmerich, S., Leffler, H. & Rosen, S. D. (1995) *J. Biol. Chem.* **270**, 12035–12047.
23. Duijvestijn, A. M., Horst, E., Pals, S. T., Rouse, B. N., Steere, A. C., Picker, L. J., Meijer, C. J. & Butcher, E. C. (1988) *Am. J. Pathol.* **130**, 147–155.
24. Kumamoto, K., Mitsuoka, C., Izawa, M., Kimura, N., Otsubo, N., Ishida, H., Kiso, M., Yamada, T., Hirohashi, S. & Kannagi, R. (1998) *Biochem. Biophys. Res. Commun.* **247**, 514–517.
25. Dixon, M. F., Genta, R. M., Yardley, J. H. & Correa, P. (1996) *Am. J. Surg. Pathol.* **20**, 1161–1181.
26. Taha, A. S., Nakshabendi, I., Lee, F. D., Sturrock, R. D. & Russel, R. I. (1992) *J. Clin. Pathol.* **45**, 135–139.
27. Dixon, M. F., O'Connor, H. J., Axon, A. T., King, R. F. & Johnston, D. (1986) *J. Clin. Pathol.* **39**, 524–530.
28. Renkonen, J., Tynninen, O., Hayry, P., Paavonen, T. & Renkonen, R. (2002) *Am. J. Pathol.* **161**, 543–550.
29. Maly, P., Thall, A., Petryniak, B., Rogers, C. E., Smith, P. L., Marks, R. M., Kelly, R. J., Gersten, K. M., Cheng, G., Saunders, T. L., *et al.* (1996) *Cell* **86**, 643–653.
30. Hiraoka, N., Petryniak, B., Nakayama, J., Tsuboi, S., Suzuki, M., Yeh, J. C., Izawa, D., Tanaka, T., Miyasaka, M., Lowe, J. B. & Fukuda, M. (1999) *Immunity* **11**, 79–89.
31. Hiraoka, N., Kawashima, H., Petryniak, B., Nakayama, J., Mitoma, J., Marth, J. D., Lowe, J. B. & Fukuda, M. (2004) *J. Biol. Chem.* **279**, 3058–3067.
32. Mitoma, J., Petryniak, B., Hiraoka, N., Yeh, J. C., Lowe, J. B. & Fukuda, M. (2003) *J. Biol. Chem.* **278**, 9953–9961.
33. Mitsuoka, C., Kawakami-Kimura, N., Kasugai-Sawada, M., Hiraiwa, N., Toda, K., Ishida, H., Kiso, M., Hasegawa, A. & Kannagi, R. (1997) *Biochem. Biophys. Res. Commun.* **230**, 546–551.
34. Toppila, S., Paavonen, T., Nieminen, M. S., Hāyry, P. & Renkonen, R. (1999) *Am. J. Pathol.* **155**, 1303–1310.
35. Fukuda, M., Carlsson, S. R., Klock, J. C. & Dell, A. (1986) *J. Biol. Chem.* **261**, 12796–12806.
36. Wilkins, P. P., McEver, R. P. & Cummings, R. D. (1996) *J. Biol. Chem.* **271**, 18732–18742.
37. Mahdavi, J., Sonden, B., Hurtig, M., Olfat, F. O., Forsberg, L., Roche, N., Angstrom, J., Larsson, T., Teneberg, S., Karlsson, K. A., *et al.* (2002) *Science* **297**, 573–578.
38. Guruge, J. L., Falk, P. G., Lorenz, R. G., Dans, M., Wirth, H. P., Blaser, M. J., Berg, D. E. & Gordon, J. I. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3925–3930.
39. Kawakubo, M., Ito, Y., Okimura, Y., Kobayashi, M., Sakura, K., Kasama, S., Fukuda, M. N., Fukuda, M., Katsuyama, T. & Nakayama, J. (2004) *Science* **305**, 1003–1006.
40. Hemmerich, S., Bistrup, A., Singer, M. S., van Zante, A., Lee, J. K., Tsay, D., Peters, M., Carminati, J. L., Brennan, T. J., Carver-Moore, K., *et al.* (2001) *Immunity* **15**, 237–247.

 DOI 10.1002/pmic.200300795 *Proteomics* 2004, *4*, 2776–2788

# Proteomic signature of human cancer cells

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

[1]Cancer Proteomics Project, National Cancer Center Research Institute, Tokyo, Japan
[2]The Fourth Department of Internal Medicine, Nippon Medical School, Tokyo, Japan

We assessed proteomic profiles as biomarkers for monitoring cell phenotypes. Protein expression profiles were obtained by fluorescence two-dimensional difference gel electrophoresis (2-D-DIGE), in which quantitative ability is improved by labeling proteins with fluorescent dyes prior to electrophoresis. Integrated protein spot intensities were analyzed by a statistical approach. The proteomic data of two groups of cell lines: (1) adenocarcinoma (AC) cell lines derived from lung, pancreas and colon tissues and (2) lung cancer cell lines with different histological backgrounds, including AC, squamous cell carcinoma and small cell carcinoma, were assessed on the basis of prior biological information. Hierarchical clustering analysis and principal component analysis were used to divide the cell lines into subgroups on the basis of similarities between their protein expression profiles. The majority of cell lines were grouped according to their organ of origin or histological background. A machine-learning algorithm selected 32 protein spots that were responsible for the classification. The results indicate that proteomic data generated by 2-D-DIGE can provide a signature of essential cell phenotypes, suggesting that it might be possible to apply this technique to developing tumor markers that could identify the organ of origin of metastatic tumors and contribute to the differential diagnosis of lung cancer.

| | |
|---|---|
| Received | 2/9/03 |
| Revised | 26/1/04 |
| Accepted | 16/2/04 |

**Keywords:** Adenocarcinoma / Fluorescence two-dimensional difference gel electrophoresis / Lung cancer / Support vector machine

## 1 Introduction

Genetic alterations during tumor progression affect essential phenotypes of the cell, such as signal transduction pathways, cell proliferation, cell cycle, apoptosis and cell mobility, and ultimately cause normal cells to develop into fully malignant tumor cells. Recent advances in expression profiling at the mRNA and protein levels have made it possible to observe such complex changes on a genome-wide scale, and global expression data for large sets of samples combined with biological and clinical data have been providing useful information for cancer studies. Studies of global mRNA expression have been performed to better understand the mechanisms of tumor progression [1], to perform differential diagnosis [2], to predict patient outcome [3–5], and to find treatment targets [6] in many types of cancer. The statistical algorithms used in these studies selected sets of mRNAs as tumor markers on the basis of prior clinical information and expression level. However, a global comparative study of mRNA and protein expression revealed that measurement of mRNA abundance is not an optimal method to monitor alterations in biological systems as a whole, because mRNA abundance often does not correspond to protein abundance [6–8]. Such discrepancies are a result of post-transcriptional events, including post-transcriptional control of protein translation and protein degradation, that cannot be predicted from mRNA sequence or abundance. Moreover, the proteins in cancer cells undergo many types of aberrant post-translational modifications, including phosphorylation and glycosylation, most of which cannot be monitored by information on mRNA expression. Therefore, to better understand the

**Correspondence:** Dr. Tadashi Kondo, 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; **Cy3**, 1-(5-carboxypentyl)-1'-propylindocarbocyanine halide *N*-hydroxy-succinimidylester; **Cy5**, 1-(5-carboxypentyl)-1'-methylindodi-carbocyanine halide *N*-hydroxy-succinimidyl ester; **PCA**, principal component analysis; **SCC**, squamous cell carcinoma; **SCLC**, small cell lung carcinoma; **SVM**, support vector machine

 

complex mechanisms of tumor progression and to identify tumor markers for better diagnosis, proteomic studies should be carried out as a complementary strategy to the transcriptomic approach.

Recent developments in proteome technology have allowed the description of several hundred features *per* sample, and future progress will lead to visualization of a larger proportion of the proteome. Statistical algorithms have been adapted for data mining in proteomic research in order to identify proteomic signatures for clinical use. For example, hierarchical clustering analysis and principal component analysis (PCA) have been performed on data from 2-DE of samples from breast cancer cell lines [9], lung cancer tissues [7], and ovarian cancer tissue [10–12]. MS-based technology (ProteinChip; Ciphergen Biosystems, Fremont, CA, USA) has been used in an attempt to identify multiple biomarkers in serum for early detection of ovarian cancer [13], prostate cancer [14] and breast cancer [15] on the basis of statistical algorithms. Although application of statistical methods to protein expression has remained limited compared with its application to mRNA expression, these studies have yielded candidate protein sets for tumor markers.

In the present study, we applied statistical learning methods to the data generated by fluorescence 2-D-DIGE. Although 2-DE is a standard method for monitoring global protein expression, it is hampered by low reproducibility and time-consuming gel staining procedures. In 2-D-DIGE, the proteins are labeled with fluorescent dyes, mixed with an internal control sample labeled with a different fluorescent dye, and then separated in the same gel by 2-DE. By normalizing the spot intensity of the sample to that of the internal control in the same gel, electrophoretic differences can be minimized for more accurate quantitative analysis. We applied this method to 30 human adenocarcinoma (AC) cell lines derived from different organs and to 30 human lung cancer cell lines having different histological backgrounds to determine whether the 2-D pattern classified the cell lines and whether known phenotypes could be the factors responsible for classification.

## 2 Materials and methods

### 2.1 Cell culture and protein extraction

The AC cell lines used were derived from: (1) lung cancer (A549, PC-3, PC-9, PC-14, RERF-LC-KJ, RERF-LC-MS, RERF-LC-OK, LC-2/ad, ABC-1, and VMRC-LCD); (2) colon cancer (DLD-1, HCT-8, COLO320, PMCO-1, LoVo, T34, Caco2, SW948, SW1417, and SW1463); and (3) pancreatic cancer (BxPC-3, AsPC-1, Mpanc-96, HPAF-II, Hs766T, MIAPaca-2, Capan-1, Capan-2, and PANC-1). The lung cancer cell lines used had a histological background of: (1) squamous cell carcinoma (SCC; PC-1, PC-10, RERF-LC-AI, SQ-5, LC-1/Sq, LC-1F, LK-2, EBC-1, QG-56, and VMRC-LCP); (2) small cell lung carcinoma (SCLC; Lu-130, Lu-134, Lu-135, Lu-139, Lu-140, Lu-165, PC-6, MS-1, SBC-3, and SBC-5); and (3) AC (A549, PC-3, PC-9, PC-14, RERF-LC-KJ, RERF-LC-MS, RERF-LC-OK, LC-2/ad, ABC-1, and VMRC-LCD). All pancreatic cancer cell lines and the colon cancer cell lines DLD-1, HCT-8, LoVo, SW948, SW1417, and SW1463 were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). 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 lung cancer 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, and the colon cancer cell lines COLO320, T34, and Caco2, were obtained from the RIKEN Cell Bank (Ibaraki, Japan). The lung cancer cell lines ABC-1, RERF-LC-MS, RERF-LC-OK, LK-2, EBC-1, VMRC-LCD, VMRC-LCP, LK-2, SBC-3, and SBC-5 were purchased from the Health Science Research Resources Bank (Osaka, Japan). The cell line PMCO-1 was established in the National Cancer Center Research Institute (Tokyo, Japan). The cell lines were maintained in their optimal medium according to the recommendatation 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 for 30 min in a lysis buffer (LB) containing 6 M urea, 2 M thiourea, 1% Triton X-100, and 3% CHAPS. The sample was centrifuged at 15 000 rpm for 30 min. Then the supernatant containing the cell proteins was recovered and the protein concentration was measured with a Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, USA) and adjusted to 1 mg/mL with LB. The pH of the protein sample was adjusted to 8.5 with 30 mM Tris-HCl.

### 2.2 Fluorescence labeling of protein samples

The proteins were labeled with fluorescent dyes developed for the 2-D-DIGE system (Amersham Biosciences, Little Chalfont, Buckinghamshire, UK). Protein samples (50 μg) were labeled with 200 pmol of 1-(5-carboxypentyl)-1'-propylindocarbocyanine halide *N*-hydroxy-succinimidyl ester (Cy3 fluorescent dye) or 1-(5-carboxypentyl)-1'-methylindodi-carbocyanine halide *N*-hydroxy-succinimidyl ester (Cy5 fluorescent dye) in separate tubes for 30 min. The labeling reaction was quenched with 0.2 mM lysine and an equal volume of LB containing 130 mM DTT and 2.0% Pharmalyte (Amersham Biosciences). Cy3-

 

labeled samples were mixed together for use as an internal control mixture and added to each Cy5-labeled sample. The volume of the Cy3- and Cy5-labeled protein mixture was adjusted to 420 μL by adding LB containing 65 mM DTT and 1.0% Pharmalyte. All labeling procedures were carried out in the dark. The procedure for protein labeling and sample preparation is illustrated in Fig. 1A.

### 2.3 2-DE

The fluorescence-labeled proteins were separated by 2-DE, with the first separation performed according to p*I* with IPG gels (24 cm length, p*I* range between 3.0 and 10; Amersham Biosciences) and the second separation performed according to $M_r$ with the EttanDalt II system (Amersham Biosciences). After rehydration of the IPG gels with a labeled protein sample, IEF was performed with an IPGphor system (Amersham Biosciences) for a total of 80 kVh at 20°C. After equilibration, the IPG gels were applied to second-dimension gels, and the proteins were separated at 20°C for 15 h at 17 W *per* 12 gels. All electrophoresis procedures were performed in the dark.

### 2.4 Image acquisition and data mining of 2-D patterns

The gels were scanned at appropriate wavelengths for Cy3 and Cy5 dyes with MasterImager 2640 (Amersham Biosciences). The spots were detected and quantified

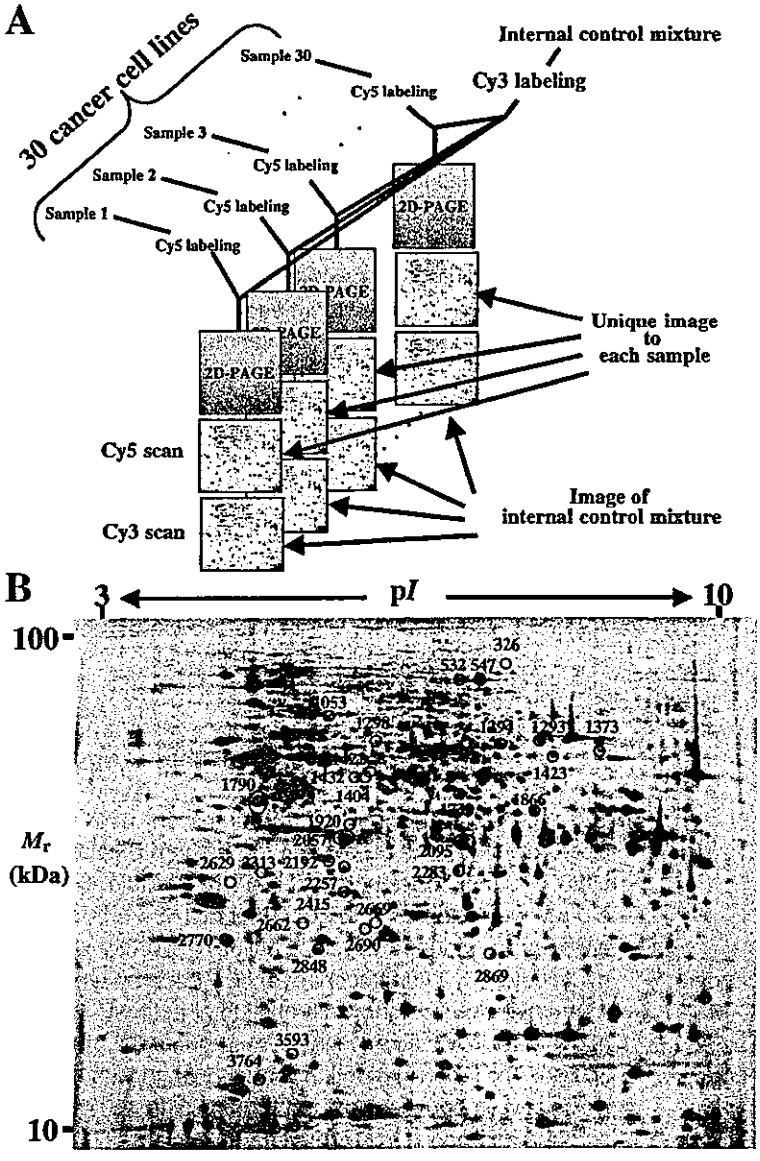


**Figure 1.** Proteomic profiling with the 2-D-DIGE system. (A) Procedures for fluorescence labeling and sample preparation are illustrated. An internal control sample was created by mixing all cell line samples and was labeled with Cy3. Cy5-labeled individual samples were then mixed with the Cy3-labeled internal control sample and separated in the same gel. Gel scanning at dual wavelengths produced Cy3 and Cy5 images from one gel. Normalization of the spot intensity of the Cy5 image to that of the Cy3 image in the same gel compensates for electrophoresis artifacts. (B) Representative 2-D image of AC cells generated from the Cy3-labeled internal control sample. The 32 circled spots were later identified as informative spots for the classification of AC cell lines.

 

with the DIA mode of the DeCyder software (Amersham Biosciences). The relative spot intensity was calculated as a ratio between total intensity of the gels and the intensity of each spot. The relative spot intensity of the Cy5 image was normalized to that of the Cy3 image in the same gel. Since the protein sample generating the Cy3 image was the same in all gels, any differences observed between the Cy3 images were due to electrophoretic artifacts, and could be compensated for by normalization of Cy5 intensity to Cy3 intensity. The BVA mode of the DeCyder software (Amersham Biosciences) was used to perform the normalization procedures. We separated each protein sample into three gels and averaged the normalized spot intensities in the triplicate gels. We selected the protein spots present in all Cy3 and Cy5 images for further analysis.

### 2.5 Statistical analysis of proteomic data

The set of spot intensities and prior biological information were used for the statistical analysis. Hierarchical clustering was performed for the cell lines by calculating Pearson correlations to determine the distances between the cell lines and by using the algorithm of Ward to construct the tree. GeneMaths (Applied Maths, Sint-Martens-Latem, Belgium) was used for the clustering analysis. PCA was performed with Impressionist (Gene Data, Basel, Switzerland). PCA is a multidimension-reduction technique to visualize the similarities and differences between the samples. Together with hierarchical clustering, it belongs to the class of unsupervised analysis, *i.e.* the algorithm does not make use of any existing group structure. Impressionist (Gene Data) was also used for machine learning methods, a leave-one-out cross-validation, and a spot ranking method on the basis of the support vector machine algorithm (SVM). In SVM, a linear hyperplane in the multidimensional protein expression space was determined between the training groups with a maximal margin for each sample. The spot ranking method based on the SVM algorithm was used to determine the subset of protein spots that was best suited for discrimination. On the basis of the SVM algorithm, a leave-one-out cross-validation was performed and the set of spots with the minimal cross-validation error rate was selected as the spot set that characterizes the groups. Fold differences of the spot intensities between the groups were calculated with the BVA mode of the DeCyder software (Amersham Biosciences).

### 2.6 In-gel proteolytic digestion and extraction of peptides from gels

DeCyder software was used to record the position of the selected spots as a text file, and an automated spot collector, SpotPicker (Amersham Biosciences), was used to excise the spots. In-gel proteolytic digestion was carried out to generate tryptic peptides from the protein spots. Briefly, the spots were excised and the gel plugs were washed with water and dehydrated with 100% ACN. The gel plugs were dried completely and incubated overnight with 100 ng of *N*-tosyl-*L*-phenylalanine chloromethylketone-treated trypsin (Promega, Southampton, UK) in 50 mM ammonium bicarbonate. Trypsin digests were recovered in 10 μL of 70% ACN/0.1% TFA. The purified peptides were mixed with an equal volume of matrix solution, dihydroxybenzoic acid, and subjected to mass spectrometric analysis. PMF was performed with a Q-Star Pulser-*i* equipped with the oMALDI ion source (Applied Biosystems, Foster City, CA) followed by a database search against Swiss-Prot with the Analyst QS program (Applied Biosystems). When the results of PMF were ambiguous, MS/MS analysis was performed on the peptides.

## 3 Results

### 3.1 Procedure of sample preparation and a representative 2-D image

Figure 1A illustrates the schematic workflow of protein labeling and sample preparation used to achieve quantitative protein expression analysis. As the Cy3 images in all gels were generated from the same internal control mixture, any differences between Cy3 images are attributable to electrophoretic artifacts. Normalization of the Cy5 image to the Cy3 image in the same gel will compensate for such artifacts. Figure 1B shows a representative 2-D image of the Cy3-labeled internal control mixture generated from the protein samples from 30 AC cell lines. We selected a reproducible 205 spots that were present in all Cy3 and Cy5 images. The 32 circled spots were later identified as informative spots for the classification of AC cell lines.

### 3.2 Unsupervised classification of 30 AC cell lines

To estimate the ability of proteomic profiling to distinguish between known types of AC cell lines, the cell lines were first analyzed by hierarchical clustering based on the intensity of 205 protein spots. A dendrogram of the 30 AC cell lines is shown in Fig. 2A. The 30 AC cell lines were divided into two major groups. One tree (a) consisted of a branch (c) of seven colon cancer cell lines and a branch (d) of six pancreatic cancer cell lines, and another tree (b) was formed by all 10 lung cancer cell lines, three colon cancer cell lines, and four pancreatic

 

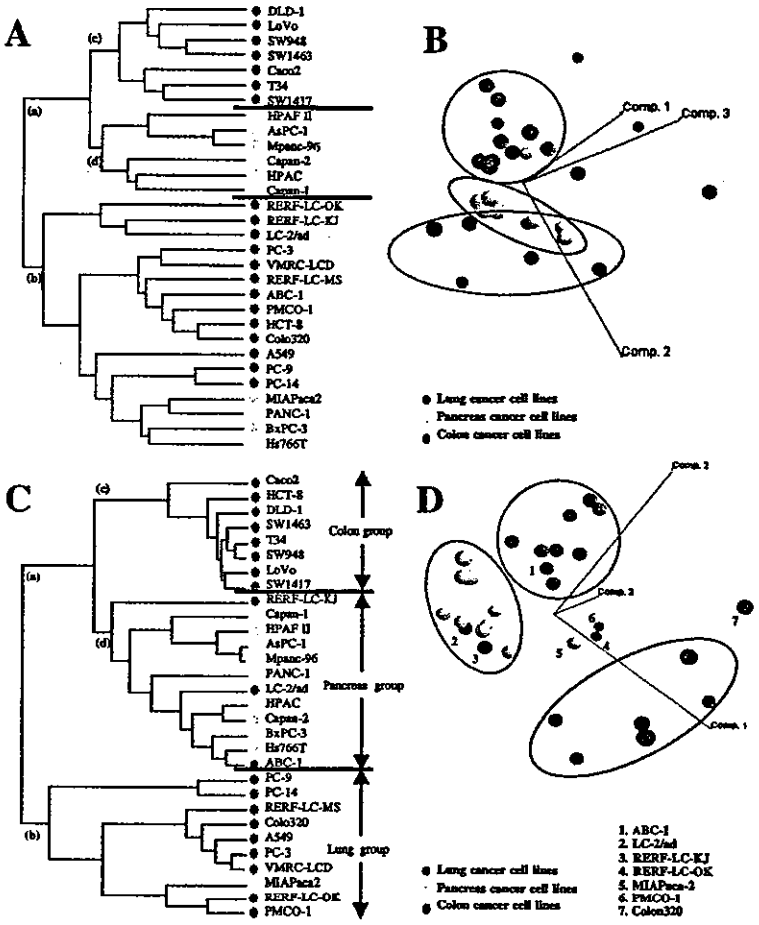


**Figure 2.** Classification of AC cell lines. (A) Hierarchical clustering analysis with 205 nonselected spots. The dendrogram was created on the basis of similarities between protein expression patterns. The cell line names and the organ of origin are listed with color coding on the right. (B) Three-dimensional plot of the results of PCA using 205 nonselected spots. The apparent groups yielded by PCA are enclosed in circles. (C) Hierarchical clustering analysis with the 32 selected spots. Most of the cell lines are clearly divided into groups according to their organ of origin. (D) PCA with the 32 selected spots. The cell lines that clustered with other cell lines with different organs of origin are numbered and listed.

cancer cell lines. In tree (b), the lung cancer cell lines were on three branches with different types of cell lines. Although many cell lines clearly tended to form groups according to their original organ, pancreatic cell lines were clustered in two different trees. PCA was performed using the same data set of protein expression profiles as used for hierarchical clustering. As shown in Fig. 2B, six lung cancer cell lines, nine pancreatic cancer cell lines, and eight colon cancer cell lines formed distinct groups. The other cell lines were separated from their main group and scattered in groups of cell lines having a different origin. These observations suggested that the AC cell lines had certain common protein expression patterns that distinguished them from other cell lines according to their organ of origin.

### 3.3 Classification of 30 AC cell lines according to 32 selected protein spots

We used a machine-learning method based on an SVM algorithm to identify the optimal set of protein spots for classification. The SVM model was trained by three groups of AC cell lines established from lung, pancreas, and colon. We found that the best 32 protein spots minimized the classification error (28%), and the error rate did not change until the number of spots reached 100 (data not shown). These 32 protein spots were selected for further statistical analysis. Figure 2C shows the results of hierarchical clustering of the cell lines according to the expression profile of the 32 selected protein spots. The dendrogram shows that the cell lines were divided into two trees. One tree (a) consisted of a branch (c) of eight colon cancer cell lines and a branch (d) of nine pancreatic cancer cell lines and three lung cancer cell lines. Another tree (b) consisted of seven lung cancer cell lines, two colon cancer cell lines, and one pancreatic cancer cell line. Overall, 80% of the colon cancer cell lines, 90% of the pancreatic cancer cell lines, and 70% of the lung cancer cell lines formed a group according to their organ of origin on the basis of the expression profiles of the 32 protein spots.

We also used the 32 protein spots for PCA of the cell lines. As shown in Fig. 2D, six lung cancer cell lines, nine pancreatic cancer cell lines, and eight colon cancer cell lines

 

tended to form separate groups. We compared the results of PCA and hierarchical clustering analysis and found that there were grouping features common to both PCA and hierarchical clustering. In the PCA, two lung cell lines, LC-2/ad and RERF-LC-KJ, were located close to the pancreatic cancer cell line group, and the same cell lines were clustered with the pancreatic cell lines in the hierarchical clustering analysis (Fig. 2C). The lung cancer cell line RERF-LC-OK, the pancreatic cancer cell line MIAPaca-2, and the colon cancer cell line PMCO-1 were positioned close to each other in the PCA. This was consistent with the results of the hierarchical clustering analysis, showing that they were more similar to each other than to other cell lines (Fig. 2C). The colon cancer cell line Colo320 localized near the lung cancer cell line group in the PCA, and it was localized in tree (b) with seven lung cancer cell lines in the hierarchical clustering analysis (Fig. 2C). In contrast, the lung cancer cell line ABC-1 localized in the colon cancer cell line group in PCA, but its protein expression pattern was similar to that of the pancreatic cancer cell lines in the hierarchical clustering analysis (Fig. 2C).

### 3.4 Spot ranking and characterization of the 32 selected spots

The spot ranking method ranked the spots according to their contribution to the classification. The SVM model was trained by a set of two cell line groups: Seven lung cancer cell lines (PC-9, PC-14, RERF-LC-MS, A549, PC-3, VMRC-LCD, RERF-LC-OK) and nine pancreatic cancer cell lines (Capan-1, HPAF II, AsPC-1, Mpanc-96, PANC-1, HPAC, Capan-2, BxPC-3, Hs766T). Since these cell lines formed groups corresponding to their organ of origin in the hierarchical clustering analysis and the PCA, we expected to find common protein profiles in the groups (Figs. 2C and D). The same procedure was performed using a set of seven lung cancer cell lines and eight colon cancer cell lines, and a set of nine pancreatic cancer cell lines and eight colon cancer cell lines. In each case, the cross-validation error rate was plotted as a function of number of top scoring spots and the spots were ranked according to their ability to be used as a basis for assigning the cell line to its organ of origin. The results of spot ranking are summarized in Table 1.

We performed mass spectrometric analysis of the 32 spots. Tryptic peptides of each spot were subjected to MS and the tryptic pattern of the spots was searched against the Swiss-Prot database. The 32 spots are shown in the 2-D image in Fig. 1B, and the identification of 19 spots is summarized in Table 1. The fold differences of spot intensities between the cell line groups are also listed in Table 1.

### 3.5 Unsupervised classification of 30 lung cancer cell lines

We investigated whether the proteomic signature in the 2-D patterns could also be used to classify lung cancer cell lines according to their histological background. The 174 spots expressed in all Cy3 and Cy5 images of lung cancer cell lines were selected for statistical analysis. Hierarchical clustering analysis divided the 30 lung cancer cell lines into subgroups on the basis of similarity of the protein profiles (Fig. 3A). The resulting dendrogram shows that all SCLC cells lines clustered together with three AC cell lines in tree (a). Another tree (b) consisted of 10 SCC cell lines and seven AC cell lines. In tree (b), all AC cell lines were on the same branch, but the SCC cell lines were distributed between two branches. PCA was also performed for 30 lung cancer cell lines with the 174 spots (Fig. 3B). All of the SCLC cell lines formed a distinct group, and seven AC cell lines formed a group along the first principal component. The SCC cell lines seemed to be divided into two different groups, and their distribution corresponded to the results of hierarchical clustering, in which the SCC cell lines were divided into two groups in the same way (Fig. 3A). There were three AC cell lines distributed in the SCLC group and the SCC group. Hierarchical clustering analysis showed that the expression profiles of these three cell lines were similar to those of the SCLC cell line group (Fig. 3A).

### 3.6 Classification of 30 lung cancer cell lines according to 71 selected protein spots

The spot ranking method selected the best spots for assigning the cell lines to the group corresponding to their histological background. Three cell line groups, the SCLC group, the SCC group and the AC group, were used to train the SVM model, with the result that the 71 best-scoring protein spots minimized the cross-validation error rate (data not shown). The expression profile of these 71 spots was used for further classification of the cell lines. Hierarchical clustering analysis based on the 71 selected spots divided the lung cancer cell lines into two groups (Fig. 3C). One tree (a) was composed of the 10 SCLC cell lines, and another tree (b) included a branch (c) of seven AC cell lines and a branch (d) of 10 SCC cell lines and three AC cell lines. The cell lines were also analyzed by PCA using the 71 spots (Fig. 3D). The majority of cell lines were divided into three groups: a SCLC cell line group, a SCC cell line group, and an AC cell line group. AC cell lines PC-3, VMRC-LCD, and RERF-LC-OK, which were distributed in the SCC cell line group in the hierarchical clustering analysis (Fig. 3C), were located close to the SCLC cell line group in the PCA (Fig. 3D). These results

 

**Table 1.** Protein spots which discriminated one cell line group from the others

| Spot no.[a] | Swiss-Prot access no.[b] | Protein descripton[c] | Score[d] | Number of peaks[e] | Protein coverage (%)[f] | Lung/Pancreas Spot ranking[g] | Lung/Pancreas Fold differences[h] | Lung/Colon Spot ranking[g] | Lung/Colon Fold differences[h] | Pancreas/Colon Spot ranking[g] | Pancreas/Colon Fold differences[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2662 | – | Not identified | – | – | – | 1 | 29 | 1 | 33.5 | 28 | 1.16 |
| 2283 | – | Not identified | – | – | – | 2 | 0,55 | 5 | 0.51 | 29 | 0.93 |
| 1730 | O75874 | Isocitrate dehydrogenase 3 alfa | 1439 | 19 | 49.8 | 20 | 1.16 | 2 | 1,4 | 23 | 1.2 |
| 3764 | – | Not identified | – | – | – | 15 | 1.71 | 9 | 6.48 | 1 | 3.8 |
| 1373 | – | Not identified | – | – | – | 18 | 0.77 | 28 | 1.55 | 2 | 2.01 |
| 2057 | P07195 | L-lactate dehydrogenase B chain | 1652 | 13 | 41.3 | 3 | 2.85 | 29 | 1 | 3 | 0.35 |
| 1790 | P08727 | Keratin 19 | 2581 | 20 | 45.3 | 8 | 0.23 | 7 | 0.59 | 4 | 2.61 |
| 3593 | – | Not identified | – | – | – | 29 | 0.64 | 12 | 1.58 | 5 | 2.45 |
| 2848 | P32119 | Peroxiredoxin 2 | 815 | 10 | 44.4 | 14 | 1.38 | 31 | 0.62 | 6 | 0.45 |
| 1053 | – | Not identified | – | – | – | 17 | 1.22 | 17 | 2.6 | 7 | 2.14 |
| 2770 | P13693 | Translationally controlled tumor protein | 424 | 5 | 26.2 | 26 | 1.02 | 24 | 0.65 | 8 | 0.64 |
| 2869 | – | Not identified | – | – | – | 10 | 1.23 | 10 | 0.85 | 9 | 0.69 |
| 1423 | P00390 | Glutathione reductase | 366 | 8 | 15.3 | 5 | 1.78 | 25 | 1.58 | 10 | 0.89 |
| 2690 | – | Not identified | – | – | – | 28 | 1.15 | 19 | 0.75 | 11 | 0.65 |
| 2257 | P09525 | Annexin A4 | 1118 | 16 | 51.4 | 11 | 0.39 | 22 | 0.68 | 12 | 1.74 |
| 2629 | – | Not identified | – | – | – | 6 | 1.34 | 4 | 3.48 | 13 | 2.59 |
| 1404 | P26641 | Elongation factor 1-gamma | 332 | 7 | 11 | 7 | 0.61 | 14 | 0.68 | 14 | 1.11 |
| 1298 | – | Not identified | – | – | – | 30 | 0.89 | 21 | 1.41 | 15 | 1.59 |
| 1866 | Q9Y617 | Phosphoserine aminotransferase | 965 | 12 | 33 | 9 | 2.03 | 16 | 0.78 | 16 | 0.38 |
| 2669 | – | Not identified | – | – | – | 19 | 1.02 | 32 | 1.57 | 17 | 1.54 |
| 532 | P13639 | Elongation factor-2 | 646 | 10 | 10.8 | 24 | 0.96 | 13 | 1.86 | 18 | 1.94 |
| 2313 | Q99426 | Tubulin-specific chaperone B | 310 | 7 | 11 | 21 | 0.99 | 20 | 1.33 | 19 | 1.33 |
| 547 | P13639 | Elongation factor-2 | 884 | 13 | 14.2 | 16 | 0.94 | 3 | 0.44 | 20 | 0.47 |
| 1294 | O60701 | UDP-glucose 6-dehydrogenase | 799 | 11 | 25.3 | 31 | 1.39 | 23 | 1.45 | 21 | 1.04 |
| 2415 | Q06323 | Proteasome activator complex subunit 1 | 278 | 6 | 23.3 | 13 | 0.5 | 11 | 0.59 | 22 | 1.18 |
| 326 | – | Not identified | – | – | – | 22 | 0.93 | 30 | 0.98 | 24 | 1.05 |
| 2095 | P40925 | Malate dehydrogenase 1,NAD | 692 | 11 | 34.4 | 4 | 4.26 | 8 | 4.58 | 25 | 1.08 |
| 1422 | – | Not identified | – | – | – | 32 | 0.74 | 26 | 0.44 | 26 | 0.59 |
| 1920 | O75874 | Isocitrate dehydrogenase 3 alfa | 38 | – | – | 23 | 2.92 | 18 | 1.78 | 27 | 0.61 |
| 1293 | P14618 | Pyruvate kinase, M1 isozyme | 1944 | 22 | 42.2 | 12 | 0.73 | 27 | 0.97 | 30 | 1.33 |
| 1432 | P35998 | 26s protease regulatory subunit 7 | 347 | 13 | 27.5 | 27 | 0.5 | 15 | 0.46 | 31 | 0.92 |
| 2192 | Q15181 | Inorganic pyrophosphatase | 314 | 5 | 20.4 | 25 | 0.55 | 6 | 0.51 | 32 | 0.92 |
| Number of spots that yielded 0% classification error rate | | | | | | 2 spots | | 2 spots | | 16 spots | |

a) Spot number refers to Fig. 1
b) Swiss-Prot accession number was recorded as a reference for the identification
c) Protein description in Swiss-Prot database
d) Reliability of spot identification was shown by Analyst QS score. Spot 1920 was identified by MS/MS analysis and the MASCOT score is indicated
e) The number of peaks used for spot identification is shown
f) Coverage of protein sequence by the peptides used for spot identification is shown
g) Spots are ranked according to their contribution to the classification
h) Fold differences were calculated between two groups

 

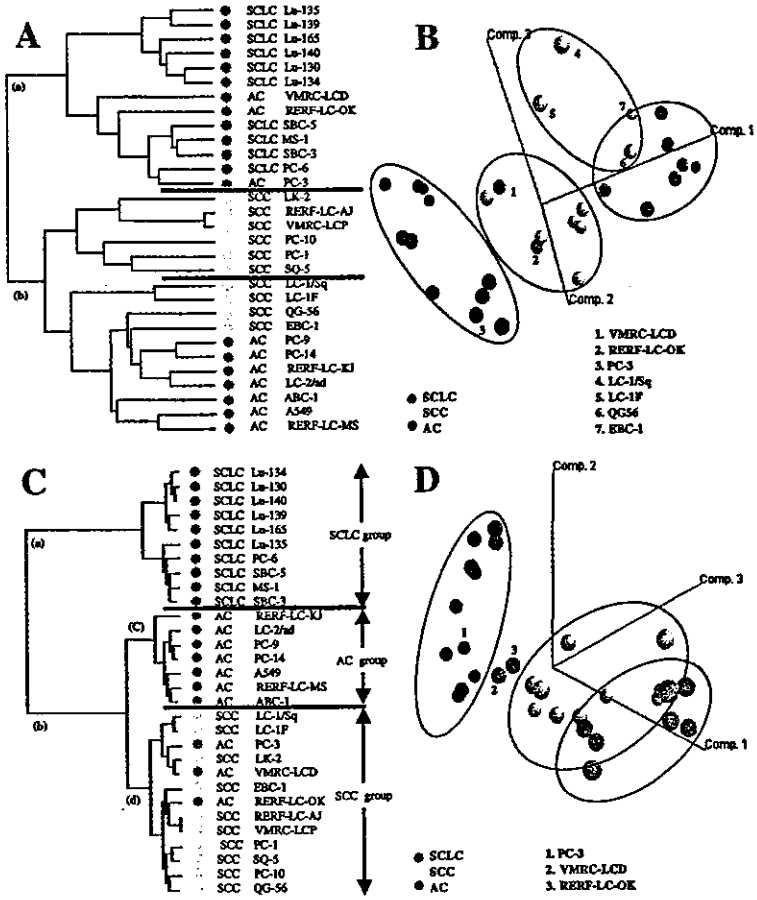


**Figure 3.** Classification of lung cancer cell lines. (A) Hierarchical clustering analysis with 174 nonselected spots. The dendrogram was created on the basis of similarities between protein expression patterns. The cell line names and the histological background are listed with color coding on the right. (B) Results of PCA using 174 nonselected spots. The apparent cell line groups are circled. The cell lines that did not localize with the other cell lines with the same histological background are numbered and listed. (C) Hierarchical clustering analysis with the 71 selected spots. Most of the cell lines are clearly divided into groups according to their original histology. (D) PCA with the 71 selected spots. The cell lines that did not cluster with the other cell lines with the same histological background are numbered and listed.

indicate that the cell lines could be divided into groups according to the expression pattern of the 71 spots, and that histological background could be the primary factor in the classification. However, neither hierarchical clustering nor PCA classified the three AC cell lines PC-3, RERF-LC-OK and VMRC-LCD into the AC cell line group.

### 3.7 Spot ranking and characterization of the 71 selected spots

We attempted to rank the 71 spots according to their contribution to the classification. The SVM model was trained with three cell line groups. As the 71 selected spots failed to assign AC cell lines PC-3, VMRC-LCD and RERF-LC-OK to the groups corresponding to their original histology (Figs. 3C and D), these three cell lines were not used to train the SVM model. The training procedures were the same as those for the classification of the 30 AC cell lines. We listed the top-scoring 20 spots that could distinguish one cell line group from the others with a minimal cross-validation error rate. The results of spot ranking and the minimal number of spots for which the cross-validation error rate was 0% are listed in Tables 2, 3 and 4. We identified the listed spots by MS. The identification of the protein spots with their Analyst QS scores are summarized in Tables 2, 3 and 4. The fold differences in spot intensities between the cell line groups of different original histology are also listed. The localization of these spots on representative 2-D images of the lung cancer cell proteins is shown in Fig. 4. There were 18 up-regulated and 16 down-regulated spots in the SCLC cell line group compared with the other groups (Fig. 4A). The AC cell line group was characterized by 22 up-regulated and 11 down-regulated spots (Fig. 4B). In addition to having 11 up-regulated and 20 down-regulated spots, the SCC cell line group had three spots whose intensity was higher than in one group and lower than in the other group (Fig. 4C).

## 4 Discussion

### 4.1 Proteomic signature retained during long-term culture

Our results demonstrate that the protein profile generated by 2-DE could classify the AC cell lines according to their organ of origin and divide the lung cancer cell lines

 

**Table 2.** Protein spots with different intensity between SCLC and AC

| Spot no.[a] | Access no.[b] | Protein description[c] | Score[d] | Number of peaks[e] | Protein coverage (%)[f] | Spot ranking[g] | Fold differences[h] |
|---|---|---|---|---|---|---|---|
| High in SCLC | | | | | | | |
| 2848 | P32119 | Peroxiredoxin 2 | 815 | 10 | 44.4 | 4 | 4.12 |
| 1218 | P40227 | T-complex protein 1, zeta subunit | 1374 | 18 | 38.2 | 8 | 1.28 |
| 3325 | – | Not identified | – | – | – | 12 | 6.12 |
| 1729 | P20073 | Annexin A7 | 323 | 8 | 18.2 | 14 | 1.36 |
| 547 | P13639 | Elongation factor-2 | 1265 | 19 | 23.7 | 15 | 1.44 |
| 995 | P20700 | Lamin B1 | 404 | 7 | 14.7 | 18 | 1.72 |
| 932 | P11142 | Heat shock protein 71 kDa protein | 574 | 12 | 25.4 | 19 | 1.32 |
| 1361 | P78371 | T-complex protein 1, beta-subunit | 2139 | 28 | 59.6 | 20 | 1.35 |
| High in AC | | | | | | | |
| 1681 | P05783 | Keratin 18 | 1783 | 22 | 46.5 | 1 | 0.04 |
| 1437 | P05787 | Keratin 8 | 1655 | 23 | 54.7 | 2 | 0.02 |
| 2105 | P07355 | Annexin II | 997 | 14 | 47.8 | 3 | 0.13 |
| 1411 | P05787 | Keratin 8 | 1686 | 23 | 51.6 | 5 | 0.06 |
| 3358 | P18282 | Destrin | 313 | 6 | 40.6 | 6 | 0.48 |
| 1435 | – | Not identified | – | – | – | 7 | 0.92 |
| 2405 | – | Not identified | – | – | – | 9 | 0.39 |
| 2225 | – | Not identified | – | – | – | 10 | 0.21 |
| 1338 | – | Not identified | – | – | – | 11 | 0.37 |
| 3322 | P23528 | Cofilin | 334 | 4 | 31.9 | 13 | 0.77 |
| 2669 | – | Not identified | – | – | – | 16 | 0.51 |
| 1340 | P50995 | Glucose-6-phosphate 1-dehydrogenase | 958 | 12 | 20 | 17 | 0.83 |

a) Spot number was refered to in Fig. 4
b) Swiss-Prot accession number was recorded as a reference for the identification
c) Protein description in Swiss-Prot database
d) Reliability of spot identification is shown by the Analyst QS score
e) Number of peaks used for spot identification is shown
f) Coverage of protein sequence by the peptides used for spot identification is shown
g) Spots are ranked according to their contribution to the classification
h) Fold differences were calculated between two groups

according to their histological background. These findings indicate that most tissue-cultured cells retain characteristics that reflect their *in vivo* origin and differentiation phenotype after long-term culture. Our results are consistent with previous reports that tissue-cultured cells retain the gene expression signature of their original tissues [16–19]. The protein spots used for the classification may include proteins whose expression pattern *in vitro* is similar to that *in vivo*. These proteins may be essential for the fundamental characteristics of the cell. In our study, known biomarker proteins, including the cytokeratin family, that allow differentiation between different epithelial cancers [20–22] were not identified as important for the classification, probably because our 2-D gels did not reveal these proteins or because we only studied protein spots present in all samples. Our results suggest that discovery of single proteins specific to individual organs or histological types might not be necessary to develop biomarkers to differentiate cancer cells, because each organ or tissue might have a unique quantitative expression pattern of common proteins. Normal and cancer cells might also be distinguishable by such a set of common protein spots, and disease-specific proteins might not be required. This hypothesis will be examined in further research.

### 4.2 Development of biomarkers for tumors of unknown origin and for AC subclasses

We demonstrated that AC cells share common proteomic signatures corresponding to their organ of origin. The proteomic signature of the original organ may yield novel

 

**Table 3.** Protein spots with different intensity between AC and SCC

| Spot no.[a] | Access no.[b] | Protein description[c] | Score[d] | Number of peaks[e] | Protein coverage (%)[f] | Spot ranking[g] | Fold differences[h] |
|---|---|---|---|---|---|---|---|
| High in AC | | | | | | | |
| 1681 | P05783 | Keratin 18 | 1783 | 22 | 46.5 | 1 | 4.28 |
| 3322 | P23528 | Cofilin | 334 | 4 | 31.9 | 2 | 2.11 |
| 1422 | – | Not identified | – | – | – | 3 | 3.18 |
| 532 | P13639 | Elongation factor-2 | 646 | 10 | 10.8 | 4 | 1.11 |
| 3269 | P16587 | ADP-ribosylation factor 3 | 910 | 9 | 54.1 | 8 | 1.43 |
| 3358 | P18282 | Destrin | 313 | 6 | 40.6 | 9 | 1.84 |
| 1760 | P02570 | Actin | 704 | 9 | 30.4 | 10 | 1,45 |
| 1437 | P05787 | Keratin 8 | 1655 | 23 | 54.7 | 11 | 4.02 |
| 2837 | – | Not identified | – | – | – | 12 | 1.51 |
| 2538 | – | Not identified | – | – | – | 13 | 1.61 |
| 2105 | P07355 | Annexin II | 997 | 14 | 47.8 | 14 | 1.49 |
| 2425 | – | Not identified | – | – | – | 15 | 2.12 |
| 2751 | – | Not identified | – | – | – | 16 | 1.11 |
| 1666 | P49411 | Elongation factor-Tu | 1095 | 14 | 37.2 | 17 | 1.11 |
| 2793 | – | Not identified | – | – | – | 18 | 1.3 |
| 1592 | P04765 | Eukaryotic initiation factor 4A-I | 382 | 7 | 19.7 | 19 | 1.53 |
| High in SCC | | | | | | | |
| 547 | P13639 | Elongation factor-2 | 1265 | 19 | 23.7 | 5 | 0.65 |
| 2771 | P78417 | Glutathione S-transferase-P1 | 214 | 4 | 11.3 | 6 | 0.9 |
| 1435 | – | Not identified | – | – | – | 7 | 0.7 |
| 1920 | O75874 | Isocitrate dehydrogenase 3 alpha | 38 | – | – | 20 | 0.92 |

For legend see Table 2.

clinical biomarkers. About 0.5% to 9% of all cancers diagnosed are described as cancers of unknown primary site, and the most frequent diagnosis has been AC [23–26]. The identification of biomarkers for the primary site of AC will lead to the development of new diagnostic techniques and novel therapeutic approaches for cancers of unknown origin [27–29]. Currently, molecular signatures unique to the primary organ have been extensively studied using DNA microarray data and statistical learning methods [30, 31]. Our results indicate that 2-DE could also be a platform for the development of protein markers for organs of unknown origin.

Protein expression patterns will also provide a basis for a novel classification of lung AC. Our results demonstrated that a subclass of AC cell lines, PC-3, VMRC-LCD and RERF-LC-OK, had a common proteomic pattern distinct from that of the majority of AC cell lines. These observations are consistent with those in a recent transcriptome study [32]. Virtanen *et al.* [32] used 6671 unique genes to study lung tumors and cell lines and found that the samples were essentially clustered according to pathological type, but some tumor cell lines, including the three AC cell lines PC-3, VMRC-LCD and RERF-LC-OK, were not classified as AC cells on the basis of their mRNA expression profiles. Although we cannot exclude the possibility that SCC or SCLC cells present in AC tissue might be selected during establishment of the cell lines, it is more likely that these observations indicate that a subpopulation of AC cells could differentiate into unique pathological types. Subclassification of AC has clinical significance, be-cause certain subclasses, based on mRNA or protein expression patterns, correlate with patient survival [33, 34].

## 5 Concluding remarks

There are two technical challenges to the development of clinical markers based on 2-DE of cell line proteins. The first point to consider is that, as the protein spots involved in the proteomic signature may contain specific protein isoforms resulting from post-translational modifications, the assay system should be able to quantify each isoform in a high-throughput manner. Currently, 2-DE is the only proteomic technique able to quantify individual protein

 

**Table 4.** Protein spots with different intensity between SCLC and SCC

| Spot no.[a] | Access no.[b] | Protein description[c] | Score[d] | Number of peaks[e] | Protein coverage (%)[f] | Spot ranking[g] | Fold differences[h] |
|---|---|---|---|---|---|---|---|
| High in SCLC | | | | | | | |
| 2848 | P32119 | Peroxiredoxin 2 | 815 | 10 | 44.4 | 2 | 3.46 |
| 1422 | – | Not identified | – | – | – | 4 | 1.52 |
| 1404 | P26641 | Elongation factor 1-gamma | 332 | 7 | 11 | 7 | 1.22 |
| 532 | P13639 | Elongation factor-2 | 646 | 10 | 10.8 | 8 | 1.56 |
| 1340 | P50995 | Glucose-6-phosphate 1-dehydrogenase | 958 | 12 | 20 | 9 | 1.2 |
| 1739 | P02570 | Actin | 791 | 8 | 27.7 | 11 | 1.19 |
| 1760 | P02570 | Actin | 704 | 9 | 30.4 | 12 | 1.09 |
| 2837 | – | Not identified | – | – | – | 15 | 1.22 |
| 3269 | P16587 | ADP-ribosylation factor 3 | 910 | 9 | 54.1 | 16 | 2.22 |
| 2425 | – | Not identified | – | – | – | 18 | 1.4 |
| 1175 | – | Not identified | – | – | – | 19 | 1.36 |
| 932 | P11142 | Heat shock protein 71 kDa protein | 574 | 12 | 25.4 | 20 | 1.28 |
| High in SCC | | | | | | | |
| 1437 | P05787 | Keratin 8 | 1655 | 23 | 54.7 | 1 | 0.09 |
| 1411 | P05787 | Keratin 8 | 1592 | 29 | 43.5 | 5 | 0.17 |
| 2105 | P07355 | Annexin II | 997 | 14 | 47.8 | 3 | 0.2 |
| 1681 | P05783 | Keratin 18 | 1783 | 22 | 46.5 | 6 | 0.19 |
| 1305 | – | Not identified | – | – | – | 10 | 0.48 |
| 1415 | P00367 | Glutamate dehydrogenase 1 | 384 | 11 | 21.1 | 13 | 0.32 |
| 2771 | P78417 | Glutathione S-transferase-P1 | 214 | 4 | 11.3 | 14 | 0.44 |
| 1432 | P35998 | 26s protease regulatory subunit 7 | 347 | 13 | 27.5 | 17 | 0.52 |

For legend see Table 2.

isoforms without specific antibodies. However, as 2-DE is hampered by time-consuming and labor-intensive procedures, use of this technique in routine clinical examinations requires modifications enabling high-throughput sample processing without user intervention. Small 2-D gels can already be run in hours on smaller IPGs (Zoom GelRunner System; Invitrogen, Carlsbad, CA, USA). As conventional time-consuming silver-staining procedures can be replaced by fluorescence labeling prior to electrophoresis (2-D-DIGE; Amersham Biosciences), spot detection can be achieved by gel scanning in a high-throughput manner. Automated 2-DE systems such as a2DE (NextGen Sciences, Cambridgeshire, UK) and ProTeam 2D (Tecan, Research Triangle Park, NC, USA) have been developed. Although these systems are not presently combined, automated 2-D systems based on the existing technologies can be created to run multiple miniaturized 2-D gels in a high-throughput manner.

Other high-throughput proteomic technologies such as protein array systems can also be used to create automated 2-D systems if specific antibodies or reagents to each protein isoform are available. However, it is not always easy to create antibodies to specific protein isoforms, and reagents to detect proteins, such as aptamers [35], cannot yet replace antibodies. Another point to bear in mind is that, as tumor tissues contain many types of nontumor cells, including normal epithelial counterparts, inflammatory cells and stromal cells, accurate proteomic study requires that tumor cells be isolated from tissues before protein extraction, especially for the purpose of applying the results obtained from studies on cell lines directly to clinical examination. To isolate specific populations of cells, laser capture microdissection (LCM) has been developed and coupled with various proteomic technologies, including 2-DE antibody arrays, reverse phase arrays, SELDI-TOF MS and immunoassays of biomarkers. As application of LCM technology to proteomics also allows histopathological observations, global protein expression data can be integrated with biological information based on pathology. The problem of this strategy is that the amount of protein obtained from LCM is limited, and use of LCM to provide adequate amounts of sample for proteomic analysis is labor-intensive. For 2-DE, ultra high sensitivity fluorescent dyes for protein labeling have

 

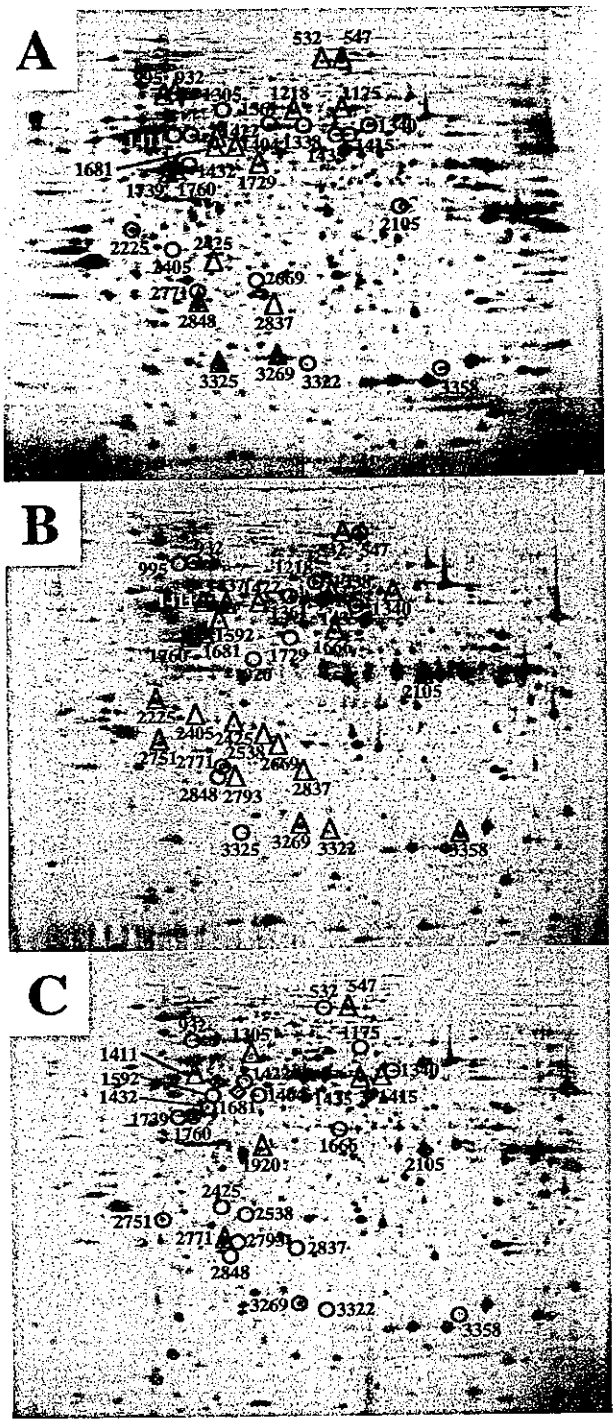


**Figure 4.** Localization of the spots informative for the classification of lung cancer cells are indicated in the representative 2-D images of each cell line group; (A) SCLC cell line group, (B) AC cell line group, (C) SCC cell line group. Spots outlined with circles and by triangles are uniformly down- or up-regulated, respectively, compared with the other cell line groups. Spots outlined with diamonds are higher than in one group and lower than in the other group. The identity of the spots and fold differences are summarized in Tables 2, 3 and 4.

solved these problems [36]. In conclusion, because 2-DE has the unique advantage of being able to measure each protein isoform without specific antibodies, and its drawbacks may be overcome by applying modern technologies, it will be quite feasible to develop clinical examination systems based on 2-DE.

*This study was supported by a grant from the Ministry of Health, Labor and Welfare, by a grant from the Ministry of Education, Culture, Sports, Science and Technology, and by the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research of Japan. M. Seike and K. Fujii are awardees of Research Resident Fellowships from the Foundation for Promotion of Cancer Research in Japan.*

## 6 References

[1] Okabe, H., Satoh, S., Kato, T., Kitahara, O. *et al.*, *Cancer Res.* 2001, *61*, 2129–2137.

[2] Ramsay, M., *Trends Mol. Med.* 2002, *8*, 159.

[3] Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J. *et al.*, *Nat. Med.* 2002, *8*, 816–824.

[4] Ramaswamy, S., Ross, K. N., Lander, E. S., Golub, T. R., *Nat. Genet.* 2003, *33*, 49–54.

[5] van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D. *et al.*, *Nature* 2002, *415*, 530–536.

[6] Qin, L. F., Lee, T. K., Ng, I. O., *Life Sci.* 2002, *70*, 1677–1690.

[7] Oh, J. M., Brichory, F., Puravs, E., Kuick, R. *et al.*, *Proteomics* 2001, *1*, 1303–1319.

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

[9] Harris, R. A., Yang, A., Stein, R. C., Lucy, K. *et al.*, *Proteomics* 2002, *2*, 212–223.

[10] Alaiya, A. A., Franzen, B., Fujioka, K., Moberger, B. *et al.*, *Int. J. Cancer* 1997, *73*, 678–683.

[11] Alaiya, A. A., Franzen, B., Hagman, A., Silfversward, C. *et al.*, *Int. J. Cancer* 2000, *86*, 731–736.

[12] Alaiya, A. A., Franzen, B., Hagman, A., Dysvik, B. *et al.*, *Int. J. Cancer* 2002, *98*, 895–899.

[13] Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J. *et al.*, *Lancet* 2002, *359*, 572–577.

[14] Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D. *et al.*, *Cancer Res.* 2002, *62*, 3609–3614.

[15] Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y., Chan, D. W., *Clin. Chem.* 2002, *48*, 1296–1304.

[16] Myers, T. G., Waltham, M., Li, G., Buolamwini, J. K. *et al.*, *Electrophoresis* 1997, *18*, 647–653.

[17] Scherf, U., Ross, D. T., Waltham, M., Smith, L. H. *et al.*, *Nat. Genet.* 2000, *24*, 236–244.

[18] Staunton, J. E., Slonim, D. K., Coller, H. A., Tamayo, P. *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, *98*, 10787–10792.

[19] Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M. *et al.*, *Nat. Genet.* 2000, *24*, 227–235.

[20] Moll, R., Franke, W. W., Schiller, D. L., *Cell* 1982, *31*, 11–24.

 

[21] Shimonishi, T., Miyazaki, K., Nakamura, Y., *Hepatology* 2000, *37*, 55–63.

[22] Goldstein, N. S., Bassi, D., *Anat. Pathol.* 2001, *115*, 695–702.

[23] Foggi, P., *Clin. Ter.* 1999, *150*, 301–306.

[24] Hainsworth, J. D., Greco, F. A., *Oncology* 2000, *14*, 563–574.

[25] Muir, C., *Cancer* 1995, *75*, 353–356.

[26] Altman, E., Cadman, E., *Cancer* 1986, *57*, 120–124.

[27] Markman, M., *Cleve. Clin. J. Med.* 1997, *64*, 73–75.

[28] Hainsworth, J. D., Greco, F. A., *N. Engl. J. Med.* 1993, *329*, 257–263.

[29] Hainsworth, J. D., Greco, F. A., *Important Adv. Oncol.* 1991, 173–190.

[30] Shedden, K. A., Taylor, J. M. G., Giordano, T. J., Kuick, R. *et al.*, *Am. J. Pathol.* 2003, *163*, 1985–1995.

[31] Nishizuka, S., Chen, S. T., Gwardry, F. G., Alexander, J. *et al.*, *Cancer Res.* 2003, *63*, 5243–5250.

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

[33] Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, *98*, 13784–13789.

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

[35] Hamaguchi, N., Ellington, A., Stanton, M., *Anal. Biochem.* 2001, *294*, 126–131.

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

 

# Proteomic Signature Corresponding to Alpha Fetoprotein Expression in Liver Cancer Cells

Hideki Yokoo,[1,2] Tadashi Kondo,[1] Kazuyasu Fujii,[2] Tesshi Yamada,[1] Satoru Todo,[2] and Setsuo Hirohashi[1]

**Alpha fetoprotein (AFP) has been implicated in the development of hepatocellular carcinoma and is considered to be a diagnostic and prognostic tumor marker. Because elevated expression of AFP is associated with many characteristics of hepatocellular carcinoma tissues, we hypothesized that multiple proteins may function in a coordinated manner with AFP. To identify such proteins, we performed global protein expression analysis, namely a proteomic study. The protein expression profiles of 9 AFP-producing liver cancer cell lines (JHH-5, HuH-1, PLC/PRL/5, Hep3B, HT-17, JHH-7, HuH-7, HepG2, Li-7) and 7 nonproducing liver cancer cell lines (HLE, JHH-6, Sk-Hep-1, JHH-4, HLF, RBE, SSP-25) were generated by fluorescence 2-dimensional difference gel electrophoresis. In fluorescence 2-dimensional difference gel electrophoresis, proteins are labeled with fluorescent dyes before electrophoresis for more accurate quantitative expression analysis. We identified 11 protein spots that distinguished AFP-producing cell lines from nonproducing cell lines by multivariate studies. The spots showed consistent alterations in amount in AFP-producing cell lines (6 up-regulated and 5 down-regulated). An additional 5 liver cancer cell lines (KIM-1, KYN-2, KYN-3, PH5-CH, PH5-T) also were correctly grouped with respect to their AFP production on the basis of the intensity of the 11 protein spots. The proteins corresponding to the 11 selected spots were identified by mass spectrometry and were categorized into 4 groups based on their known role in apoptosis, glucose metabolism, cytoskeletal organization, or translation. In conclusion, we found a novel association of AFP with other proteins. Their interaction should provide insight into the biology of AFP-producing hepatocellular carcinoma cells.** (HEPATOLOGY 2004;40:609–617.)

Hepatocellular carcinoma (HCC) represents the vast majority of liver cancer, accounting for more than 90% of all primary liver cancers, and ranks fifth in frequency among all malignancies worldwide.[1,2] Infection with hepatitis B or C virus has been identified as an etiological factor, and the subsequent cellular and histological changes leading to HCC have been extensively studied.[3] Genetic alterations, including loss of heterozygosity[4,5] and aberrant DNA methylation,[6] also have been implicated in the carcinogenesis of HCC. Nevertheless, the outcome for HCC patients still remains dismal, partly because of the difficulty in establishing an early diagnosis and the frequent recurrence and intrahepatic metastasis after surgery.[7] Although systematic chemotherapy for unresectable HCC has been widely used, its efficacy remains low and complications such as significant myelosuppression are observed during the course of treatment.[8]

Serum alpha fetoprotein (AFP) has been considered to be a hallmark of development of HCC.[9] Serum AFP alone has a limited role in the early diagnosis of HCC, because a considerable proportion of HCC patients do not have elevated serum AFP, and serum AFP can increase in patients with diseases other than HCC; it may, however, be a useful prognostic indicator, because the median survival rate of HCC patients with markedly elevated AFP was significantly shorter than that of patients with normal or moderately elevated AFP.[10] AFP is a multifunctional glycoprotein belonging to the family of albumin-like proteins.[11] AFP has been shown to serve as a dual regulator of growth in a multitude of cell types and cancers involving

*Abbreviations: HCC, hepatocellular carcinoma; AFP, alpha fetoprotein; MS, mass spectrometry; SVM, Support Vector Machine; SLDA, Sparse Linear Discriminant Analysis.*

*From the [1]Cancer Proteomics Project, National Cancer Center Research Institute, Tokyo, Japan, and the [2]Department of General Surgery, Hokkaido University Graduate School of Medicine, Sapporo, Japan.*

*Received February 22, 2004; accepted May 10, 2004*

*Supported by a grant from Pharmaceuticals and Medical Devices Agency of Japan.*

*Address reprint requests to: Tadashi Kondo, M.D., Ph.D., 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.*



*Published online in Wiley InterScience (www.interscience.wiley.com).*

*DOI 10.1002/hep.20372*

several mechanisms, including apoptotic regulation.[12] AFP also functions as a transporter of estrogens,[13] fatty acids,[14] and bilirubin[15] and modulates immune response in macrophages and T lymphocytes.[16] Although the structure and function of AFP have been extensively studied,[17] its exact role in the development of HCC has not been defined in detail.

Recently, global gene-expression studies have been shown to have great value in identifying the genes associated with various clinical states of liver cancer, such as early intrahepatic recurrence,[18] and have been used to develop diagnostic and prognostic biomarkers. Kawai et al[19] performed a genome-wide gene expression study and found that AFP-producing HCC cell lines shared a distinct expression profile of genes, including those related to apoptosis, cell cycle, cell–cell interaction and oncogenesis. Similarly, classification of HCC cell lines based on unbiased gene expression profiles also resulted in 2 major subgroups corresponding to AFP expression level, suggesting that the elevated AFP may be a part of the molecular signature that determines the major subtypes of HCC.[20] These studies provide clues to the particular genetic pathways in which AFP is involved and contribute to further understanding of the biological and clinical significance of AFP in HCC.

In this report, we performed global protein expression analysis, a proteomic study, to identify the proteins that function in a common network with AFP. Because higher expression of AFP often is associated with greater tumor size, histological undifferentiation, and portal vein thrombosis,[10] and it is unlikely that any single protein is directly responsible for such a variety of characteristics, we hypothesized that multiple proteins may function in a coordinate manner with AFP. Although discordance of messenger RNA and protein expression has been reported,[21–24] the DNA microarray data for AFP-producing cells were not examined at the protein level, and a proteomic study has not yet been performed. In this study, we generated protein expression profiles using fluorescence 2-dimensional difference gel electrophoresis, in which proteins are labeled with fluorescent dyes before electrophoresis, and used a multivariate method to capture a proteomic signature corresponding to AFP production.

## Materials and Methods

***Cell Lines.*** The cell lines JHH-4, JHH-5, JHH-6, JHH-7, HuH-1, PLC/PRF/5, HuH-7, HLE, and HLF were obtained from the Human Science Research Resources Bank (Sennai, Osaka, Japan). The cell line SK-Hep-1 was obtained from the American Type Culture Collection (Rockville, MD). Two cholangiocellular carcinoma cell lines, SSP-25 and RBE, and the HepG2 cell line were obtained from the Riken Gene Bank (Ibaragi, Tsukuba, Japan). The HT17 and Hep3B cell lines were from the Cell Resource Center for Biomedical Research, Tohoku University (Sendai, Miyagi, Japan). The KYN-2, KYN-3, and KIM-1 cell lines were provided by Dr. Masamichi Kojiro of Kurume University. Li-7 and 2 immortalized hepatocyte cell lines, PH5-T and PH5-CH, were previously established in the National Cancer Center Research Institute.[25,26] All cell lines were HCC ones, unless otherwise specified. To minimize the effects of culture condition on protein expression, the cells were maintained with the recommended culture media, and proteins were extracted when the cells reached approximately 80% confluence.

***Preparation of Fluorescence-Labeled Protein Samples.*** The cells were washed with phosphate-buffered saline and treated with 10% trichloroacetic acid for 30 minutes on ice. Then, the cells were collected into a tube, were pelleted by a brief centrifugation, and were washed with ice-cold phosphate-buffered saline. The cells were then incubated with urea lysis buffer, consisting of 7 M urea, 2 M thiourea, 3% CHAPS, and 1% Triton X-100 for 30 minutes on ice. After centrifugation at 15,000 rpm for 30 minutes, the protein concentration of recovered supernatant was measured with a Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA) and adjusted to 1 mg/mL with urea lysis buffer. We then made an internal control sample by mixing the protein samples of all cell lines used. After the pH of the protein samples had been adjusted to 8.0 with 30 mM Tris-HCl, the individual samples and the internal control sample were labeled for 30 minutes with 200 nmol of 1-(5-carboxypentyl)-1′-propylindocarbocyamine halide (Cy3) and 1-(5-carboxypentyl)-1′-methylindocarbocyamine halide (Cy5; Amersham Biosciences, Little Chalfont, Buckinghamshire, UK), respectively. The labeling reaction was terminated by incubation with 0.2 mM lysine for 10 minutes on ice. The labeled samples were treated with an equal volume of urea lysis buffer containing 130 mM DTT and 2% ampholine for 15 minutes on ice. The labeled internal sample and the samples of the individual cell lines then were mixed together. The final volume was adjusted to 430 μL with urea lysis buffer containing 65 mM DTT and 1.0% ampholine. The procedure of sample labeling and mixing is illustrated in Fig. 2.

***Two-Dimensional Polyacrylamide Gel Electrophoresis.*** For first-dimension separation, immobilized pH gradient gels (24 cm length; pI range, 3–10; Amersham Biosciences) were rehydrated with labeled protein samples for 12 hours at 20 °C. Isoelectric focusing was performed with IPGphor (Amersham Biosciences) for a