total of 80 kVh at 20°C. The immobilized pH gradient gels were equilibrated for 15 minutes in equilibration buffer containing 6 M urea, 50 mM Tris (pH 8.8), 30% glycerol, 1.0% sodium dodecyl sulfate, and 16 mM DTT, and then for another 15 minutes in the same buffer containing 122 mM iodoacetamide instead of DTT. Equilibrated immobilized pH gradient gels were transferred onto 9% to 15% gradient polyacrylamide gels and embedded with agarose. The second-dimension electrophoresis was performed at 17 W for 15 hours at 20°C. We divided the Cy3-Cy5 mixture in 3 gels and calculated the average spot intensity.

After electrophoresis, the gels were scanned at the appropriate wavelengths for Cy3 and Cy5 with a 2920 2D Master Imager (Amersham Biosciences). The exposure time was determined to ensure that the maximum spot intensity was not saturated. For every spot, the Cy5 intensity was normalized to the Cy3 intensity in the same gel. The BVA mode of DeCyder software (Amersham Biosciences) was used to normalize the intensity of all spots and calculate the average spot intensity among 3 gels.

Machine-Learning Method and Multivariate Analysis. The 1334 protein spots that appeared in at least 80% of Cy3 images were selected and subjected to statistical analysis. The spot intensity was standardized by subtraction of the average value and by division by the root mean square, and then transformed logarithmically. Hierarchical clustering analysis and principal component analysis were performed using GeneMaths software (Applied Maths, Sint-Martens-Latem, Belgium). Cross-validation and spot ranking were performed with Impressionist software (Gene Data, Basel, Switzerland). The prediction model for AFP production was created with the Visual Data Mining system (Suuri-ken, Tokyo, Japan).

Western Blotting for AFP. To explore the expression of intracellular AFP, 30 μg of protein lysate was labeled with Cy3 as described in Preparation of Fluorescence-Labeled Protein Samples and separated by sodium dodecyl sulfate -polyacrylamide gel electrophoresis with 12% polyacrylamide or 2D-polyacrylamide gel electrophoresis. The gels were scanned at appropriate excitation and emission wavelengths for Cy3, and the amount of protein loaded onto the gels was normalized by the sum of band intensities of the entire lane. The proteins then were transferred to polyvinylidene difluoride membranes and incubated with anti-AFP antibody (1:200 dilution; ZYMED, San Francisco, CA). The blots were reacted with anti-mouse immunoglobulin G antibody conjugated with peroxidase (1:1,000; Amersham Biosciences), and the AFP signal was monitored by an enhanced chemiluminescence system (Amersham Biosciences). The intensity of the AFP signal was quantified with an LAS-1000 image analyzer (Fujifilm, Tokyo, Japan).

Mass Spectrometric Protein Identification. For preparative purposes, 500 μg of unlabeled protein sample was separated by 2D-polyacrylamide gel electrophoresis and stained with SYPRO Ruby dye (Molecular Probes, Eugene, OR). The image of the preparative gel was matched to that of the analytical gels and the spots of interest were excised. The gel plugs were washed twice for 5 min with water, then with 100% vol/vol acetonitrile. After complete drying, the protein in the gel was digested at 37°C overnight with trypsin (Promega, Southampton, UK) in 50 mM ammonium bicarbonate, with gentle agitation. Peptides then were extracted from the gel plugs with 50 μL of 50% acetonitrile in 0.1% trifluoroacetic acid and were concentrated to approximately 10 μL. A 1-μL sample of extracted peptides was mixed with saturated dihydroxybenzoic acid in 50% acetonitrile/0.1% trifluoroacetic acid and spotted onto a target plate. Peptide mass fingerprinting analysis was performed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MS) with a Q-Star Pulser-i equipped with the orthogonal injection/matrix-assisted laser desorption/ionization ion source (Applied Biosystems, Framingham, CA). Mass spectra were processed with the Analyst QS program (Applied Biosystems), and a search of the Swiss-Prot database was performed for peptide mass fingerprinting with a mass tolerance of less than 50 ppm.

Results

Expression of AFP in Liver Cancer Cell Lines. We examined the expression of AFP in 19 HCC cell lines and 2 cholangiocellular carcinoma cell lines. Western blotting revealed that 11 HCC cell lines expressed AFP, whereas the other cell lines did not (Fig. 1A). The expression level of AFP as a function of relative expression ratio between the AFP signal generated by enhanced chemiluminescence and the total Cy3 intensity of bands is presented as a histogram of the 21 cell lines (Fig. 1B). The degree of AFP expression varied even among the AFP-producing cell lines.

Proteomic Pattern of Fluorescence-Labeled Proteins. The spot intensities were analyzed by fluorescence 2-dimensional difference gel electrophoresis in a quantitative way. Protein samples of cell lines were mixed together to create the internal control sample and labeled with Cy3. Protein samples of individual cell lines were labeled with Cy5. Cy3-labeled and Cy5-labeled protein samples were mixed and coseparated by 2D-polyacryl-

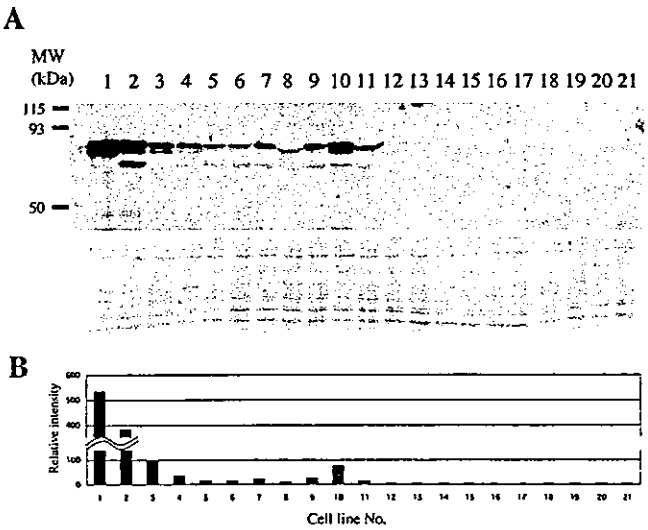

Fig. 1. Expression of alpha fetoprotein (AFP) in the cell lines. (A) Cy3-labeled proteins separated by sodium dodecyl sulfate - polyacrylamide gel electrophoresis were reacted with anti-AFP antibody (**upper panel**). The Cy3-labeled proteins were visualized by scanning the gels at the appropriate wavelength for Cy3 (**lower panel**). (B) Histograms of AFP expression level as a function of the ratio between AFP signal and the total intensity of Cy5-labeled protein bands. 1, HuH-7; 2, JHH-7; 3, JHH-5; 4, HepG2; 5, HT17; 6, HuH-1; 7, Hep3B; 8, Li-7; 9, PLC/PRL/5; 10, KIM-1; 11, KYN-2; 12, HLE; 13, HLF; 14, JHH-4; 15, JHH-6; 16, SK-Hep-1; 17, KYN-3; 18, PH5-CH; 19, PH5-T; 20, RBE; 21, SSP-25.

amide gel electrophoresis in the same gel. Two-dimensional profiles of the internal control sample and individual cell line samples were obtained by scanning the same gel with wavelengths unique to Cy3 or Cy5. The fluorescent dyes are engineered so that the migration of a given protein labeled with Cy3 or Cy5 is practically identical in 2-dimensional gels. Because the Cy3 image is produced from the common internal control sample, normalization of Cy5 image spot intensity to the intensity of the corresponding Cy3 image spot in the same gel avoided electrophoretic artifacts and enabled quantitative analysis. The 2-color image of Cy3-labeled or Cy5-labeled proteins showed that the location of spots was identical between the 2 images and that the intensity of each spot was distinct, reflecting the differential expression level in the internal control sample and individual cell line (Fig. 2A).

Figure 2B shows a representative protein expression profile with approximately 2,000 protein spots. For multivariate analysis, we selected 1,334 spots that were present in the Cy3 images of at least 80% of gels. The number of selected spots in each cell line is summarized in Fig. 2D. We did not filter the spots with respect to expression level. The 11 spots circled in Fig. 2B were selected later by a machine-learning method as essential spots for classification of liver cancer cell lines according to their AFP production.

We examined the expression of AFP on 2-dimensional gel. Western blotting identified AFP as 6 protein spots on the membrane. However, the spots corresponding to AFP were not observed on the Cy3 image (Fig. 2C), because the expression level of AFP was below the detection limit of fluorescence 2-dimensional difference gel electrophoresis, and 1 of the spots was behind the spot for heat-shock protein 70 (HSP 70). Therefore, the AFP spots were not included in our study.

To observe the degree to which the protein expression profiles were characteristic to each cell line, we observed the overall correlation between the protein expression profiles of the cell lines (Fig. 2D). Correlation matrix revealed that the protein expression profiles of AFP-negative cell lines were more homogeneous than those of AFP-positive cell lines. This different degree of correlation may reflect the differentiation status of the HCC cell lines; the AFP expression level was higher in poorly differentiated HCC,[27] and the protein expression profile of well-differentiated cells may be more uniform than that of poorly differentiated cells.

Supervised Classification and Spot Identification. We used a machine-learning method to find the spots that best discriminated AFP-producing cells from nonproduc-

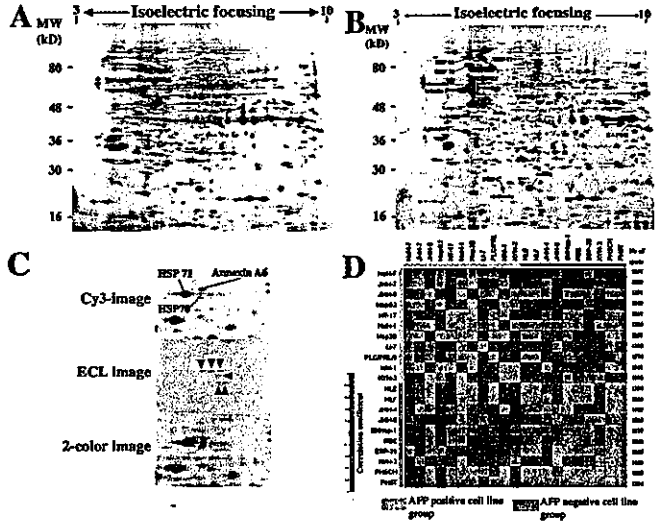

Fig. 2. Protein expression profile of liver cancer cell lines. (A) Cy3-labeled and C5-labeled proteins are overlaid for comparison. (B) Representative 2D image of the Cy3-labeled internal control sample. The 11 circled spots later were selected as informative for AFP production and the boxed area was enlarged in Fig. 6. (C) Localization of alpha-fetoprotein (AFP) spots on the 2-dimensional (2D) image. The Cy3-labeled protein sample was separated by 2D-polyacrylamide gel electrophoresis (Cy3 image), transferred to a membrane, and reacted with anti-AFP antibody (enhanced chemiluminescence image [ECL]). In the 2-color image, AFP spots were localized in the Cy3 image. (D) Correlation matrix of protein expression profiles of hepatocellular cancer cell lines. The similarity of the profiles between the cell lines is illustrated by color.

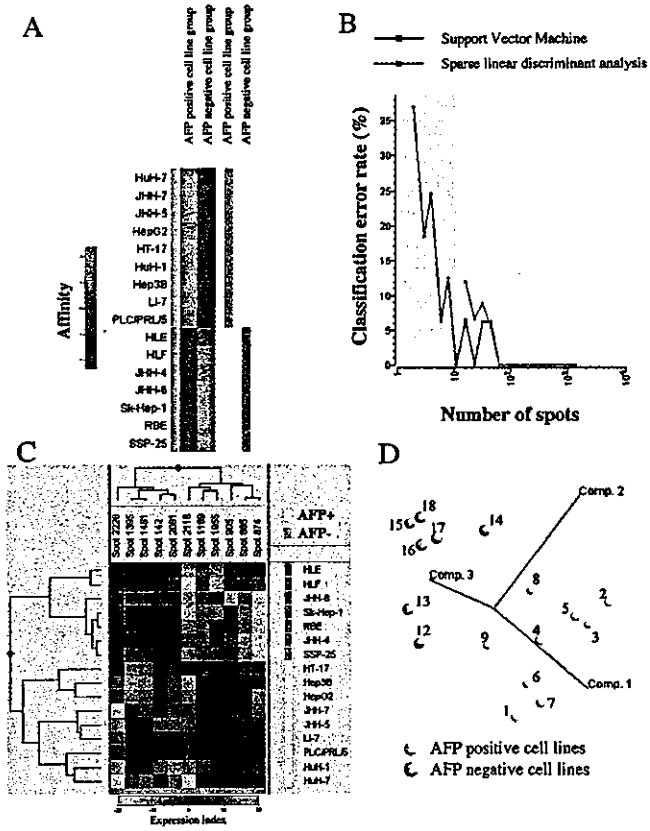

Fig. 3. Multivariate analysis of proteomic data. (A) Crossvalidation study of liver cancer cell lines. The colors in the left sidebar indicate the affinity of the samples to the groups in the training set; cell lines that were classified into a particular group with low crossvalidation error rate were defined as having high affinity to the group. Green and blue indicate the highest and the lowest affinity to the groups in the training set, respectively. (B) Spot ranking analysis. The Support Vector Machine model was trained with 2 groups: the alpha fetoprotein (AFP)-producing and the nonproducing cell line group. Classification error rates were plotted with a particular number and combination of spots. (C) Hierarchical clustering analysis. The dendrogram of the cell lines was created on the basis of the expression level of the 11 selected spots. (D) Principal component analysis. The cell lines were plotted in a 3-dimensional space using the expression pattern of the 11 spots. The cell line numbers correspond to those in Fig. 1.

ing cells. In the leave-one-out crossvalidation, 9 AFP-producing cell lines and 7 nonproducing cell lines were used as test samples. Fifteen cell lines were randomly selected and were used to generate the classifier, and the classification performance was validated using 1 cell line that was not used to generate the classifier. The probability of obtaining crossvalidation error by chance was obtained by repeating this crossvalidation procedure using 1,000 random permutation trials. We first examined 4 classifier algorithms, including a linear Support Vector Machine (SVM), Sparse Linear Discriminant Analysis (SLDA), Fisher Linear Discriminant Analysis, and K-Nearest Neighbors. Figure 3A illustrates the results of crossvalidation based on the SVM algorithm, indicating that the classifier categorized the cell lines into the groups in the training set with low crossvalidation error rate. SLDA also resulted in correct classification (data not shown). The other algorithms, Fisher Linear Discriminant Analysis and K-Nearest Neighbors, misclassified the cell lines (data not shown).

To determine the set of spots that best distinguished the 2 groups, we applied a spot ranking method in which the crossvalidation was performed with all combinations of spots with multiple independent classifier algorithms, and the misclassification error rate was plotted as a function of the number of best spots. We tested linear SVM and SLDA as classifiers, and the algorithms of support vector machine weight, recursive feature elimination, sparse linear ranking, supervised gene shaving, ANOVA, and Kruskal Wallis test to rank the spots according to the degree of contribution to the classification. Figure 3B shows the results of spot ranking using Support Vector Machine Weight as a ranking method. The crossvalida-

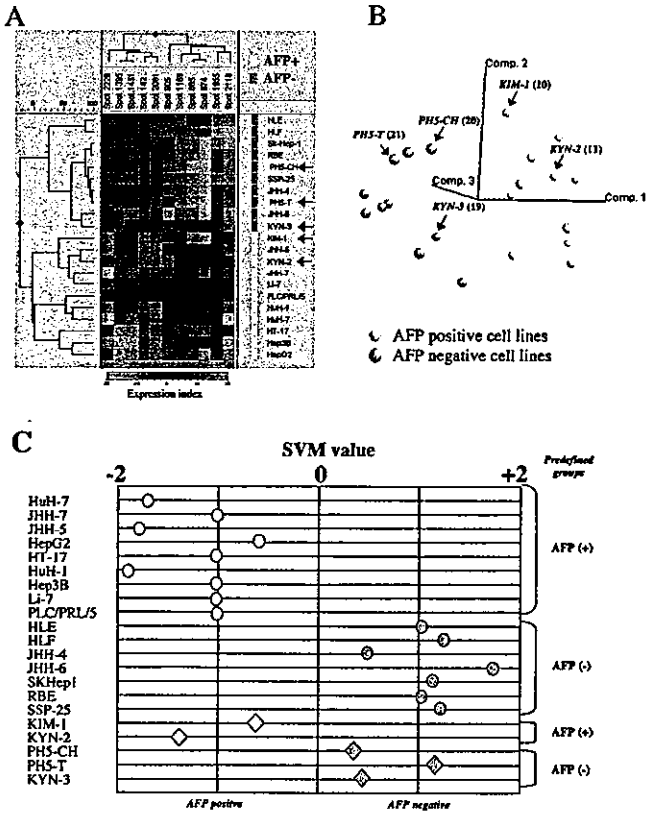

Fig. 4. Validation of the predictive performance for alpha-fetoprotein (AFP) production. (A) Dendrogram created by hierarchical analysis using the 11 spots. The arrows indicate the additional 5 liver cancer cell lines. (B) Principal component analysis by the 11 spots with the additional 5 liver cancer cell lines. The additional 5 liver cancer cell lines were shown by arrows. (C) AFP production predictor model. The predictor uses the 11 protein spots that distinguish AFP-producing and nonproducing cell lines. The Support Vector Machine (SVM) model with AFP-producing cell lines yields negative values and that with nonproducing cell lines yields positive values.

tion error rate reached 0% when particular sets of 11 and 67 spots were used with SVM and SLDA, respectively. The other ranking algorithms did not identify spot sets that had a classification error rate of 0%. Because the 67 spots selected by SLDA included the 11 spots selected by SVM, we concluded that the combination of SVM as a classifier and Support Vector Machine Weight as a ranking method could best select the essential features of the expression profile representing AFP production.

We confirmed the performance of the 11 spots as a classifier using unsupervised classification methods. Figure 3C shows the dendrogram of cell lines created by hierarchical clustering analysis using the 11 spots. The cells were clustered in 2 major trees according to their AFP productivity. The 11 protein spots also were divided into 2 major trees: the intensity of 6 spots was up-regulated and the intensity of 5 spots was down-regulated in AFP-producing cell lines (Fig. 3C). Principal component analysis showed that the 16 cell lines were also divided into 2 major groups corresponding to their AFP production on the basis of the 11-spot profile (Fig. 3D).

Use of the 11-Spot Set to Predict AFP Production by HCC Cell Lines. We further examined the predictive performance of the 11 spots using the cell lines that were not used to create the classifier. Those included 2 AFP-producing cell lines (KIM-1 and KYN-2) and 3 nonproducing cell lines (KYN-3, PH5-CH, and PH5-T). Using the selected 11-spot set, hierarchical clustering (Fig. 4A) and principal component analysis (Fig. 4B) grouped these additional cell lines correctly into either the AFP-producing or the nonproducing cell line group. We then used these 11 spots to construct a scoring model to predict AFP production and applied this model to an independent test set of 5 HCC cell lines. All training samples and the additional 5 cell lines were classified according to their AFP production (Fig. 4C).

These results indicate that the profiles of the 11 spots may represent the cellular phenotypes differently shared between the AFP-producing cell lines and the nonproducing cell lines.

Mass Spectrometric Identification of the 11 Spots and Their Expression in the Cell Lines. A database search with an MS and tandem mass spectrometry (MS/MS) study of tryptic peptides of spot 2,226 (Fig. 5A, B) resulted in identification of galectin-I (Fig. 5C–E). MS analysis was performed on the other protein spots, and the Mascot scores obtained allowed positive protein identification (Table 1).

We demonstrated the expression levels of the identified proteins between the cell lines, and the 2-dimensional images for AFP-producing and nonproducing cell lines are shown in Fig. 6. The Cy5 intensity standardized by Cy3 intensity also was shown for all the cell lines used. It is important to note that global messenger RNA expression studies on liver cancer cell lines did not identify these proteins.[21,22]

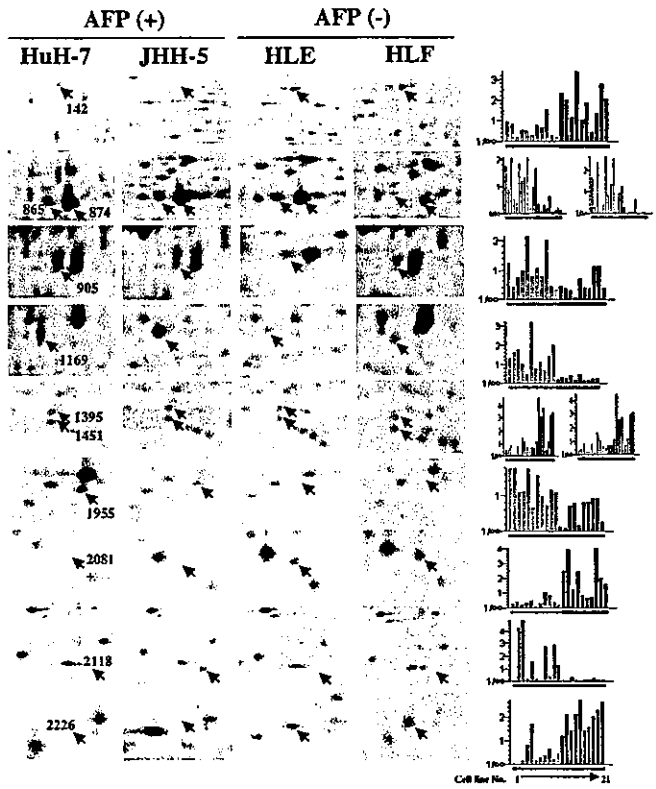

Fig. 5. Expression analysis of the 11 identified protein spots. (A) Differential expression of the 11 spots is shown in alpha fetoprotein (AFP)–producing (HuH-7 and JHH-5) and nonproducing (HLE and HLF) cells. The spot numbers correspond to those in Fig. 2B. (B) Histograms of expression of the 11 spots in all cell lines. The cell line numbers are referred to in Fig. 1.

Discussion

Elevated expression of serum AFP has been shown to be associated with various malignant characteristics of HCC (greater tumor size, bilobar involvement, undifferentiated tissue types, massive or diffuse types, and portal vein thrombosis) and poor prognosis.[10] These pathological and clinical features could be explained by a common protein network with AFP expression. Taking annexin A1 as an example, this protein is a major substrate for tyrosine kinases such as epidermal growth factor receptor and serine/threonine kinases such as protein kinase C[43,44] and has been implicated in cellular signal transduction.[45] Enhanced expression of annexin A1 was observed in poorly differentiated HCC tissues compared with their well-differentiated counterparts, and the nontumorous region contained significantly lower amounts of annexin A1, suggesting that annexin A1 is related to the histological

Table 1. Characterization of Protein Spots Informative for Classification

Spot Ranking*	Spot No.	Fold Difference† (*P* Value)	Identification (Protein Name, Accession No.)‡	Score§	Function‖	Known Connection to HCC/Other Cancer	Reference¶
1	874	5.12 (1.2e-008)	Glucose-6-phosphate 1-dehydrogenase (P11413)	1131	G	Yes/yes	28, 29
2	2118	9.19 (0.00032)	Ubiquitin-like protein SUMO-1 conjugating enzyme (P50550)	202	A	No/no	–
3	2226	0.27 (7.4e-008)	Galectin 1 (P09382)	240	A	No/yes	–, 30
4	2081	0.29 (5.3e-009)	Mitochondrial dicarboxylate carrier (Q9UBX3)	1141	G	No/no	–
5	1955	1.97 (7.4c-006)	BH3 interacting domain death agonist (P55957)	427	A	Yes/yes	31, 32
6	865	2.64 (0.00014)	Aldehyde dehydrogenase 1A1 (P00352)	251	G	Yes/no	33, –
7	1169	3.15 (5.4e-010)	Isocitrate dehydrogenase (O75874)	228	G	Yes/yes	34, 35
8	1395	0.38 (0.00038)	Annexin A1 (P04083)	364	A	Yes/yes	36, 37
9	905	1.59 (0.00013)	Keratin, type II cytoskeletal 7 (P08729)	440	C	Yes/yes	38, 34
10	1451	0.45 (0.00014)	60S acidic ribosomal protein P0 (P05388)	242	T	Yes/yes	40, 41
11	142	0.43 (8.5e-006)	Vinculin (P18206)	597	C	No/yes	–, 42

*Spot ranking was calculated on the basis of contribution of each spot to the classification.
†Fold differences: average intensity of AFP-producing cells/nonproducing cells.
‡Identification was performed by mass spectrometric study followed by database search against: Swiss-Prot.
§Reliability of spot identification was shown by Analyst QS score.
‖Functional classification of proteins: G, glucose metabolism; A, apotosis; C, cytoskeleton; T, translation.
¶References for the connection of protein to HCC/other cancer.

grade of HCC and is involved in the malignant transformation process.[36] The histological undifferentiation observed in AFP-positive HCC may be explained by the enhanced expression of annexin A1.

In summary, we identified the differential expression of 4 apoptosis-related proteins in AFP-positive cells. Two of them are known to be associated with HCC (BH3-interacting domain death agonist and annexin A1) and the other 2 have not been implicated in the mechanisms of HCC (ubiquitin-like protein SUMO-1 conjugating enzyme and galectin I). It has been reported that AFP has growth-suppressive effects on cells by inducing apopto-

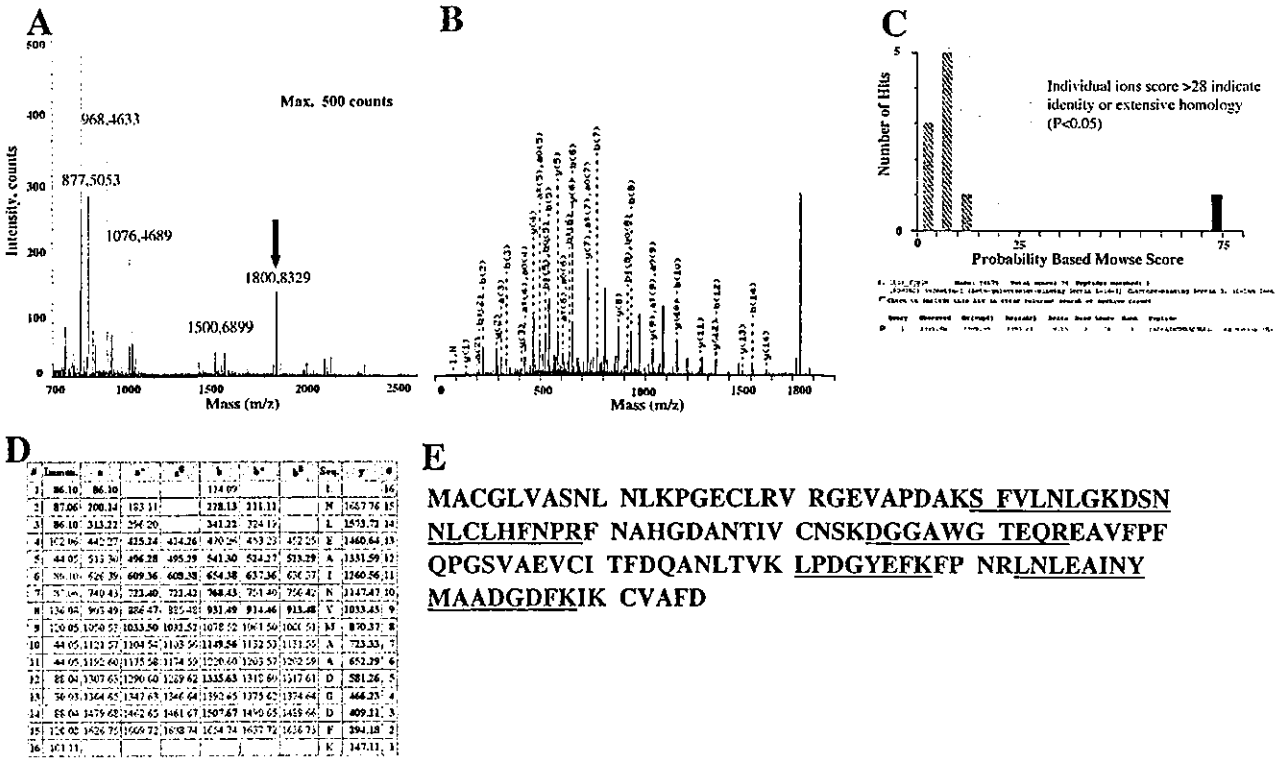

Fig. 6. Mass spectrometry (MS) analysis. (A) MS spectrogram of tryptic digests from spot 2,226 for peptide mass fingerprinting (PMF). The peptide with the asterisk was subjected to MS/MS analysis. (B) MS/MS spectrogram of peptide ion. (C, D) Output of database search by the Mascot program using MS/MS data. (E) Tryptic peptide sequences of galectin-1 matched with the spectral peaks are underlined. m/2, mass-to-change ratio.

sis.[46] The expression of apoptosis-related proteins in AFP-producing cells may explain the apoptotic effects of AFP on the cells. Because the aberrant regulation of apoptosis is known to contribute to the development of HCC,[47] clarification of the association of apoptosis-related proteins with AFP will be useful for understanding the mechanisms of HCC development.

Aberrant glucose metabolism also has been reported in HCC,[28] and we identified 4 proteins involved in glucose metabolism. The higher expression of glucose-6-phosphate 1-dehydrogenase, aldehyde dehydrogenase 1A1, and isocitrate dehydrogenase has been reported in liver cancer and hepatoma cells.[33,34] However, mitochondrial dicarboxylate carrier in liver has not been previously considered.[29]

The other proteomic approach, such as MS-based separation techniques, which may be complement to fluorescence 2-dimensional difference gel electrophoresis, will identify the other proteins associating with AFP. Those proteins will also provide novel insights into the biology of HCC.

References

1. Okuda K. Hepatocellular carcinoma. J Hepatol 2000;32(Suppl 1):225–237.
2. Schafer DF, Sorrell MF. Hepatocellular carcinoma. Lancet 1999;353: 1253–1257.
3. Beasley RP, Hwang LY, Lin CC, Chien CS. Hepatocellular carcinoma and hepatitis B virus. A prospective study of 22,207 men in Taiwan. Lancet 1981;2:1129–1133.
4. Kuroki T, Fujiwara Y, Tsuchiya E, Nakamori S, Imaoka S, Kanematsu T, et al. Accumulation of genetic changes during development and progression of hepatocellular carcinoma: loss of heterozygosity on chromosome arm 1p occurs at an early stage of hepatocarcinogenesis. Genes Chromosomes Cancer 1995;13:163–167.
5. Tsuda H, Zhang WD, Shimosato Y, Yokota J, Terada M, Sugimura T, et al. Allele loss on chromosome 16 associated with progression of human hepatocellular carcinoma. Proc Nad Acad Sci U S A 1990;87:6791–6794.
6. Kondo Y, Kanai Y, Sakamoto M, Mizokami M, Ueda R, Hirohashi S. Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis—a comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma. HEPATOLOGY 2000;32:970–979.
7. Poon RT, Fan ST, Lo CM, Liu CL, Wong J. Intrahepatic recurrence after curative resection of hepatocellular carcinoma: long term results of treatment and prognostic factors. Ann Surg 1999;229:216–222.
8. Leung TW, Patt YZ, Lau WY, Ho SK, Yu SC, Chan AT, et al. Complete pathological remission is possible with systemic combination chemotherapy for inoperable hepatocellular carcinoma. Clin Cancer Res 1999;5: 1676–1681.
9. Fujiyama S, Tanaka M, Maeda S, Ashihara H, Hirata R, Tomita K. Tumor markers in early diagnosis, follow-up and management of patients with hepatocellular carcinoma. Oncology 2002;62:57–63.
10. Tangkijvanich P, Anukulkarnkusol N, Suwangool P, Lertmaharit S, Hanvivatvong O, Kullavanijaya P, et al. Clinical characteristics and prognosis of hepatocellular carcinoma: analysis based on serum alpha-fetoprotein levels. J Clin Gastroenterol 2000;31:302–308.
11. Gillespie JR, Uversky VN. Structure and function of alpha-fetoprotein: a biophysical overview. Biochim Biophys Acta 2000;1480:41–56.
12. Semenkova LN, Dudich EI, Dudich IV, Shingarova LN, Korobko VG. Alpha-fetoprotein as a TNF-resistance factor for human hepatocarcinoma cell line HepG2. Tumor Biol 1997;18:30–40.
13. Soloff MS, Swartz SK, Pearlmutter AF, Kithier K. Binding of 17beta-estradiol by variants of alpha-fetoprotein in rat amniotic fluid. Biochim Biophys Acta 1976;427:644–651.
14. Parmelee DC, Evenson MA, Deutsch HF. The presence of fatty acids in human alpha-fetoprotein. J Biol Chem 1978;253:2114–2119.
15. Ruoslahti E, Estes T, Seppala M. Binding of bilirubin by bovine and human alpha-fetoprotein. Biochim Biophys Acta 1979;578:511–519.
16. Lu CY, Changelian PS, Unanue ER. Alpha-fetoprotein inhibits macrophage expression of Ia antigens. J Immunol 1984;132:1722–1727.
17. Mizejewski GJ. Alpha-fetoprotein structure and function: relevance to isoforms, epitopes, and conformational variants. Exp Biol Med 2001;226: 377–408.
18. Iizuka N, Oka M, Yamada-Okabe H, Nishida M, Maeda Y, Mori N, et al. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet 2003;361:923–929.
19. Kawai HF, Kaneko S, Honda M, Shirota Y, Kobayashi K. Alpha-fetoprotein-producing hepatoma cell lines share common expression profiles of genes in various categories demonstrated by cDNA microarray analysis. HEPATOLOGY 2001;33:676–691.
20. Lee JS, Thorgeirsson SS. Functional and genomic implications of global gene expression profiles in cell lines from human hepatocellular cancer. HEPATOLOGY 2002;35:1134–1143.
21. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720–1730.
22. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 1997;18:533–537.
23. Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, et al. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol Cell Proteomics 2002;1:323–333.
24. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardia SL, et al. Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics 2002;1:304–313.
25. Genda T, Sakamoto M, Ichida T, Asakura H, Kojiro M, Narumiya S, et al. Cell motility mediated by rho and Rho-associated protein kinase plays a critical role in intrahepatic metastasis of human hepatocellular carcinoma. HEPATOLOGY 1999;30:1027–1036.
26. Noguchi M, Hirohashi S. Cell lines from non-neoplastic liver and hepatocellular carcinoma tissue from a single patient. In Vitro Cell Dev Biol Anim 1996;32:135–137.
27. Okuda H, Nakanishi T, Takatsu K, Saito A, Hayashi N, Yamamoto M, et al. Comparison of clinicopathological features of patients with hepatocellular carcinoma seropositive for alpha-fetoprotein alone and those seropositive for des-gamma-carboxy prothrombin alone. J Gastoenterol Hepatol 2001;16:1290–1296.
28. Pepinsky RB, Sinclair LK. Epidermal growth factor-dependent phosphorylation of lipocortin. Nature 1986;321:81–84.
29. Khanna NC, Tokuda M, Chong SM, Waisman DM. Phosphorylation of p36 in vitro by protein kinase C. Biochem Biophys Res Commun 1986; 137:397–403.
30. Ullrich A, Schlessinger J. Signal transduction by receptors with tyrosine kinase activity. Cell 1990;61:203–212.
31. Masaki T, Tokuda M, Ohnishi M, Watanabe S, Fujimura T, Miyamoto K, et al. Enhanced expression of the protein kinase substrate Annexin 1 in human hepatocellular carcinoma. HEPATOLOGY 1996;24:72–81.
32. Dudich E, Semenkova L, Dudich I, Gorbatova E, Tochtamisheva N, Tatulov E, et al. Alpha-fetoprotein causes apoptosis in tumor cells via a pathway independent of CD95, TNFR1 and TNFR2 through activation of caspase-3-like proteases. Eur J Biochem 1999;266:750–761.
33. Yamamoto K, Takenaka K, Kajiyama K, Shimada M, Shirabe K, Taketomi A, et al. Cell proliferation and cell loss in nodule-in-nodule hepatocellular carcinoma. Hepatogastroenterology 1999;46:813–819.

34. Taketa K, Shimamura J, Ueda M, Shimada Y, Kosaka K. Profiles of carbohydrate-metabolizing enzymes in human hepatocellular carcinomas and preneoplastic livers. Cancer Res 1988;48; 467–474.
35. Yanagawa Y, Chen JC, Hsu LC, Yoshida A. The transcriptional regulation of human aldehyde dehydrogenase I gene. The structural and functional analysis of the promoter. J Biol Chem 1995;270:17521–17527.
36. Seow TK, Ong SE, Liang RC, Ren EC, Chan L, Ou K, et al. Two-dimensional electrophoresis map of the human hepatocellular carcinoma cell line, HCC-M, and identification of the separated proteins by mass spectrometry. Electrophoresis 2000;21:1787–1813.
37. Fiermonte G, Dolce V, Arrigoni R, Runswick MJ, Walker JE, Palmieri F. Organization and sequence of the gene for the human mitochondrial dicarboxylate carrier: evolution of the carrier family. Biochem J 1999;344: 953–960.
38. Poirier F, Bourin P, Bladier D, Joubert-Caron R, Caron M. Effect of 5-azacytidine and galectin 1 on growth and differentiation of the human B lymphoma cell line bl36. Cancer Cell Int 2001;1:1–12.
39. Chen GG, Lai PB, Chan PK, Chak EC, Yip JH, Ho RL, et al. Decreased expression of Bid in human hepatocellular carcinoma is related to hepatitis B virus X protein. Eur J Cancer 2001;37:1695–1702.
40. Ravi R, Bedi A. Sensitization of tumor cells to Apo2 ligand/TRAIL-induced apoptosis by inhibition of casein kinase II. Cancer Res 2002;62: 4180–4185.
41. Huang CJ, Severin E, Blum M. Flow-cytometric determination of dehydrogenase activities in primary human gastrointestinal tumor cell lines. Anal Cell Pathol 1994;6:93–103.
42. Hippo Y, Yashiro M, Ishii M, Taniguchi H, Tsutsumi S, Hirakawa K, et al. Differential gene expression profiles of scirrhous gastric cancer cells with high metastatic potential to peritoneum or lymph nodes. Cancer Res 2001; 61:889–895.
43. Uenishi T, Kubo S, Yamamoto T, Shuto T, Ogawa M, Tanaka H, et al. Cytokeratin 19 expression in hepatocellular carcinoma predicts early postoperative recurrence. Cancer Sci 2003;94:851–857.
44. Taguchi J, Nakashima O, Tanaka M, Hisaka T, Takazawa T, Kojiro M. A clinicopathological study on combined hepatocellular and cholangiocarcinoma. J Gastroenterol Hepatol 1996;11:758–764.
45. Kondoh N, Wakatsuki T, Ryo A, Hada A, Aihara T, Horiuchi S, et al. Identification and characterization of genes associated with human hepatocellular carcinogenesis. Cancer Res 1999;59:4990–4996.
46. Barnard GF, Staniunas RJ, Mori M, Puder M, Jessup MJ, Steele GD Jr, et al. Gastric and hepatocellular carcinomas do not overexpress the same ribosomal protein messenger RNAs as colonic carcinoma. Cancer Res 1993;53:4048–4052.
47. Ben-Ze'ev A. The use of two-dimensional gel electrophoresis in studies on the role of cytoskeletal plaque proteins as tumor suppressors. Electrophoresis 1996;17:1752–1763.

〔シンポジウム：医療プロテオミクス―基礎から臨床へ― 総説〕

二次元電気泳動を用いた臨床マーカーの開発
―必要とされる技術開発と臨床研究への応用―

近藤 格

SUMMARY

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) has a great utility to develop clinical markers, as it uncovers many types of multiple proteins in quantitative, reproducible and high throughput manner. Current technological progress around 2D-PAGE, such as fluorescence dyes, bioinformatics, laser microdissection, enable more sophisticated and accurate proteomic study. For instance, the tumor marker for unknown-origin tumor will be developed by identifying the expression pattern of spots. Although numerous challenges still lie ahead, the future of proteomics, especially medicine-oriented proteomics, by 2D-PAGE looks promising.

Key words: 2D-PAGE, bioinformatics, laser microdissection, mass spectrometry, proteomics.

はじめに：臨床マーカー開発ツールとしての二次元電気泳動

二次元電気泳動を用いたタンパク質の発現解析は1970年代に始まり，今もなおプロテオミクスではもっとも普及している．一度の泳動で数1000種類のタンパク質が観察できること[1]，一度に多検体を泳動できること，初期投資が少額であること，結果をデータベース化できること，などを考えると，これからも二次元電気泳動の需要は衰えることはないだろう．DNA塩基配列のデータベースが整備され質量分析が進歩し，スポットに対応するタンパク質の同定は以前に比べると圧倒的に簡単にできるようになった．また二次元電気泳動用の蛍光色素（2D-DIGE）によって定量性やスループット性は著しく向上した[2]．二次元電気泳動では数1000種類の多種多様なタンパク質が観察されるので，多くの臨床検体について二次元電気泳動でタンパク質発現プロファイルを作成すれば，臨床情報とリンクするようなタンパク質を見つけることができるだろう．またそのようなタンパク質は治療のターゲットや臨床マーカーになりうる．

臨床検体のプロテオミクスには，一般的に，培養細胞を使ったプロテオミクスにはない課題がある．1つ目は，膨大な発現情報をどのように臨床情報と結び付けて考えるか，2つ目は目的とする細胞からいかに正確な発現プロファイルを作成するか，という2点である．二次元電気泳動ではこれらはどのような技術で解決されるかについて述べる．

バイオインフォーマティクスとレーザーマイクロダイセクション

二次元電気泳動の解析には専用の解析ソフトが必要不可欠である．しかし，多くの二次元電気泳動解析ソフトではスポットの濃度を個別に一つずつサンプル間で比較しており，これでは二次元電気泳動から得られる情報を十分調べているとは言えない．タンパク質の発現量（スポットの濃度）の相互関係を調べたり，相互関係の背後にあるルールを既知の臨床情報と関連づけて発見することも二次元電気泳動でできるはずである．そのような目的では多変量解析や機械学習法の手法が一般に用いられる．トランスクリプトームの分野では発現情報をさまざまな手法で解析し，がんの個性を解明しようとする試みが盛んに行われている[3]．

Clinical marker development by two-dimensional polyacrylamide gel electrophoresis.

Tadashi Kondo; 国立がんセンター研究所腫瘍プロテオミクスプロジェクト
Correspondence address: Tadashi Kondo; Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.

第54回日本電気泳動学会総会・シンポジウム
（受付2004年7月21日，受理2004年8月10日，刊行2004年9月15日）

その影響でさまざまなデータマイニングソフトがたいへん安価に（あるものはフリーで）入手でき，誰でも比較的手軽に高度な解析を行うことができるようになっている．二次元電気泳動のデータを解析ソフトから外にアウトプットすることでこれらを利用することができる[4]．このようにすれば複雑な発現情報を臨床情報とリンクさせて解析することが可能になる．

腫瘍組織にはがん細胞だけでなく，対応する正常細胞，血管内皮細胞，間質細胞，血球細胞などが含まれている．また，組織の外側や血管中には血液に由来するタンパク質も豊富に含まれている．これらをまとめてすりつぶしていたのでは何のプロテオームを見ているのか分からなくなってしまう．したがって，正確なプロテオーム解析のためにはサンプリング段階でがん細胞だけを回収する必要がある．そのような目的ではレーザーマイクロダイセクション法が用いられる．レーザーマイクロダイセクションではたくさんの細胞を回収することは難しいが，超高感度な蛍光色素を用いてタンパク質を標識すれば，ごく少数の細胞からでも二次元電気泳動を行うことができる[5]．蛍光色素は2色利用できるので，2D-DIGE法による定量性に優れた解析が可能である[2]．

臨床マーカーの開発例
―原発不明がんのための腫瘍マーカー―

原発不明がんとは，転移性腫瘍だけが見つかって原発臓器がどこにあるのか分からないがんのことである．原発巣が見つからない症例は全悪性腫瘍の数％に存在し，頻度として多くないとは言え，生命予後に関わる重要な問題である．原発巣を同定できるような，言い換えれば臓器特異的なタンパク質の発現パターンを二次元電気泳動で発見できないだろうかと考え，次ぎのような実験を行った．まず，由来する臓器が異なるヒト腺がん細胞株を30株用意した．内訳は大腸がん10株，肺がん10株，そして膵がん10株である．それら細胞株からタンパク質を抽出し，2D-DIGE法でタンパク質発現プロファイルを作成した．スポットの濃度に基づき，機械学習法（サポートベクターマシーン）を用いて30株をうまく分けることができるスポットを32個決定した．この32個のスポットを用いた階層的クラスター解析や主成分分析では，がん細胞を由来する3つの臓器に一致して分類できることを確認した．対応するタンパク質を質量分析で同定してみると，個々のタンパク質としては臓器特異的なものは含まれていなかった．つまり，普遍的に存在するようなタンパク質であっても，その発現量のバランス（発現パターン）には臓器の特異性が存在し，それを同定することで由来する臓器を特定できるということが分った[6]．現在この実験をさらに発展させ，腺がんや扁平上皮がんなど200種類以上の培養がん細胞を用いて，臓器特異的なタンパク質およびタンパク質発現パターンを同定しようとしている．二次元電気泳動を用いて臓器特異タンパク質を発見する試みは治療法の開発のためにも有望であることが報告されており，臨床応用を目指した研究へと発展させたい[7]．

二次元電気泳動で開発された腫瘍マーカーを実用化するためには

がんの再発，転移，抗癌剤の感受性などの予測に有用であるとされたスポットを実際に腫瘍マーカーとして使用するためには，多検体をスクリーニングできるような工夫が必要である．一つの方法はスポットに対する特異抗体を使用し，ELISAやマイクロアレイの系を用いることである．二次元電気泳動で観察されているスポットのほとんどは何らかのタンパク質のアイソフォームである．一つのタンパク質が複数のスポットを生み出し，その中のいくつかの増減ががんの形質に関連している．あるタンパク質に対する抗体はそのすべてのアイソフォームに反応してしまうことが多いのだが，アイソフォーム全体の量としてはがんの形質に関連しない例がある．そのような場合は，個々のスポットに対する特異抗体を作成する必要がある．しかし，現在のところそれは必ずしも容易ではない．また，複数のスポットの組み合わせで診断を行おうとする場合はそれらスポット全てに対する特異抗体が必要となるのだが，一つずつ抗体を作っていくというのはあまり現実的ではない．もっとも，組み合わせに用いられる1つか2つのスポットだけを調べてもうまくがんの性格を判定できている例もあるので，まずは市販抗体で試してみる価値はあるだろう[8]．

もう一つの方法は，二次元電気泳動自体をスクリーニングのために使用するというものである．この方法は小さなサイズのものであればかなり現実的である．まずタンパク質の蛍光色素標識は自動化できるだろう．一次元目のゲルは泳動装置の台数に比例して一度に何本でも泳動することができるし，平衡化のステップは自動化できる．二次元目のゲルの作成も一度に可能であるが，一次元目のゲルを二次元目に乗せるところだけは手動である．泳動後のスポットの検出は，サンプルを蛍光標識しておけばスキャナーの台数に応じてハイスループットに行うことができる．ゲルの解析も測定するスポットが限られていればある程度は自動化できるだろう．きわめて重要なスポットの組み合わせを同定できたときに抗体の作成が困難であれば，このようなシステムも考えられる．

二次元電気泳動を用いた疾患プロテオミクスの展望

本稿では二次元電気泳動を用いた臨床マーカーの開発について述べた．定量性・再現性に優れた二次元電気泳動は，一般に疾患プロテオミクスにおいてたいへん有用なツール

である．転写因子や受容体，細胞周期に関連したタンパク質は二次元電気泳動では観察されないというのが定説だったが，質量分析を用いてスポットの同定を網羅的に行ってみると，実はそれらはスポットとして観察されていたということが分ってきた．従ってこれからは疾患のメカニズムの解明にも期待が持てる．二次元電気泳動の基本原理はこの30年でまったく変わっていない[9,10]．このことは二次元電気泳動のアイデアがいかに秀逸であったかを表している．現代ではタンパク質サンプルの分画をとる装置（Rotophor, BioRad社／Zoom IEF Fractionator, Invitrogen社），ユニークな染色試薬（Minimal dye, Saturation dye, Amersham Bioscience社／ProQ Diamond, ProQ Emerald. SYPRO-Ruby, Molecular Probe社），さまざまな画像解析（DeCyder／Amersham Biosciences社／Progenesis, Nonlinear Dynamic社／PDQUEST, BioRad社）や統計のためのソフト（Impressionist, GeneData社／GeneMaths, Applied Maths社／S-Plus, 数理システム社），そして質量分析などの周辺機器・試薬が発展してきた．それらを駆使すると何が分ってくるのか，実に楽しみである．

文　献

1) Klose J, Nock C, Herrmann M, Stuhler K, Marcus K, Bluggel M, Krause E, Schalkwyk LC, Rastan S, Brown SD, Bussow K, Himmelbauer H, Lehrach H. Genetic analysis of the mouse brain proteome. Nat Genet 2002; 30:385–393.
2) 近藤　格，広橋説雄．二次元電気泳動法による発現解析．プロテオーム解析マニュアル．東京：羊土社，2004: 26–43.
3) Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, Li Y, Robles AI, Chen Y, Ma ZC, Wu ZQ, Ye SL, Liu YK, Tang ZY, Wang XW. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat Med 2003;9:416–423.
4) Yokoo H, Kondo T, Yamada T, Todo S and Hirohashi S. Proteomic signature corresponding to AFP expression in human liver cancer cells. Hepatology 2004;40:609–617.
5) Kondo T, Seike M, Mori Y, Fujii K, Yamada T, Hirohashi S. Application of sensitive fluorescence dyes in a linkage of laser microdissection and two-dimensional gel electrophoresis as a tool for cancer proteomic study. Proteomics 2003;3:1758–1766.
6) Seike M, Kondo T, Fujii K, Gemma A, Kudoh S, Yamada T, Hirohashi S. Proteomic signature of human cancer cells. Proteomics 2004;4:2776–2788.
7) Oh P, Li Y, Yu J, Durr E, Krasinska KM, Carver LA, Testa JE, Schnitzer JE. Nature 2004;429:629–635.
8) Chen G, Gharib TG, Wang H, Huang CC, Kuick R, Thomas DG, Shedden KA, Misek DE, Taylor JM, Giordano TJ, Kardia SL, Iannettoni MD, Yee J, Hogg PJ, Orringer MB, Hanash SM, Beer DG. Protein profiles associated with survival in lung adenocarcinoma. Proc Natl Acad Sci USA 2003;100:13537–13542.
9) Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 1975;26:231–243.
10) O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 1975;25:4007–4021.