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

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

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

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

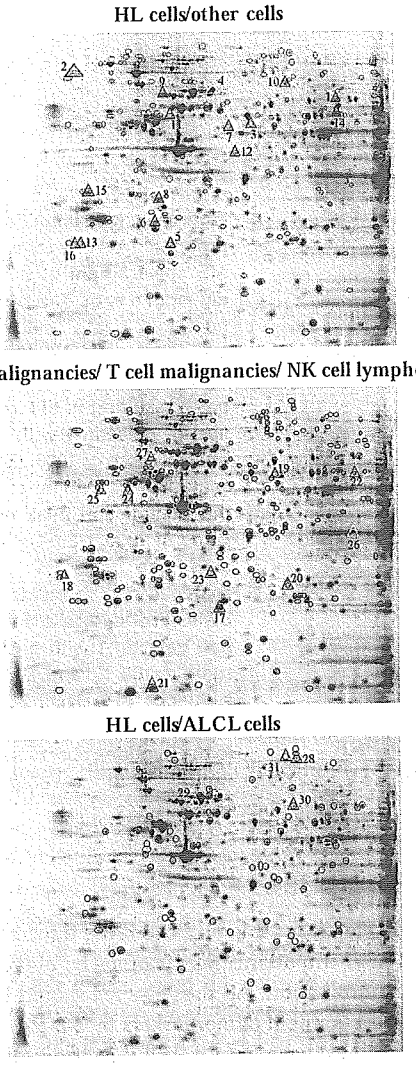


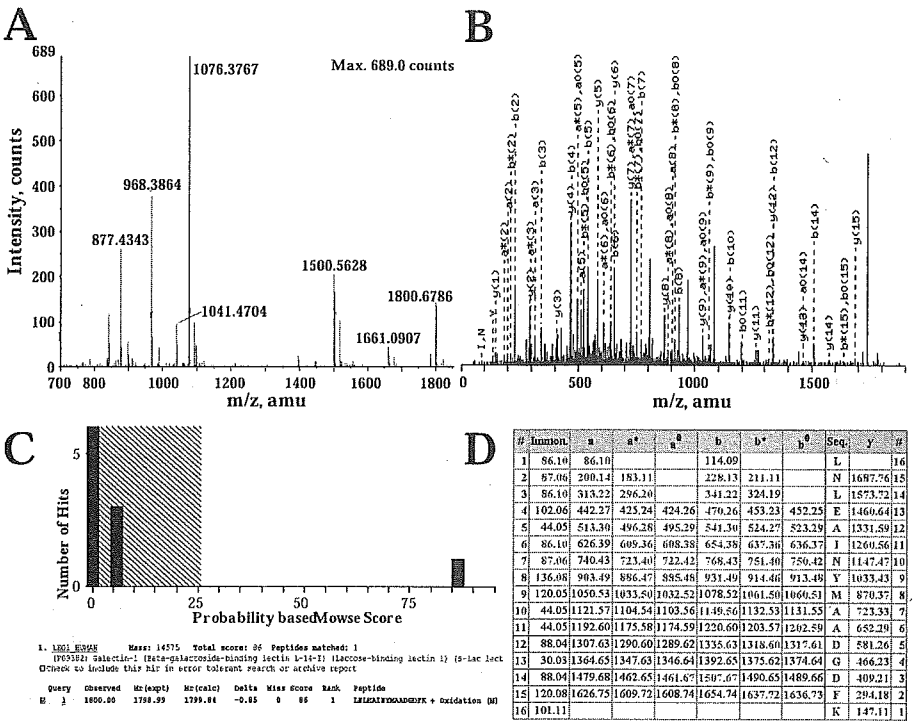
Figure 5. Localization of the informative spots on the 2-D PAGE gels. The spots selected by the Wilcoxon or Kruskal–Wallis tests are circled. Spots in a triangle are those identified as informative by the machine-learning methods. (A) HL cells *versus* other cells; (B) B cell malignancies *versus* T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

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

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

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

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



#	Ammon.	a	a*	a°	b	b*	b°	Seq.	y	#
1	86.10	86.10			114.09			L		16
2	87.06	200.14	183.11		228.13	211.11		N	1687.76	15
3	86.10	313.22	296.20		341.22	324.19		L	1573.72	14
4	102.06	442.27	425.24	424.26	470.26	453.23	452.25	E	1460.64	13
5	44.05	513.30	496.28	495.29	541.30	524.27	523.29	A	1331.59	12
6	86.10	626.39	609.36	608.38	654.38	637.36	636.37	I	1260.56	11
7	87.06	740.43	723.40	722.42	768.43	751.40	750.42	N	1147.47	10
8	136.08	903.49	886.47	885.48	931.49	914.46	913.48	Y	1033.43	9
9	120.05	1050.53	1033.50	1032.52	1078.52	1061.50	1060.51	M	870.37	8
10	44.05	1121.57	1104.54	1103.56	1149.56	1132.53	1131.55	A	723.33	7
11	44.05	1192.60	1175.58	1174.59	1220.60	1203.57	1202.59	A	652.29	6
12	88.04	1307.63	1290.60	1289.62	1335.63	1318.60	1317.61	D	581.26	5
13	30.03	1364.65	1347.63	1346.64	1392.65	1375.62	1374.64	G	466.23	4
14	88.04	1479.68	1462.65	1461.67	1507.67	1490.65	1489.66	D	409.21	3
15	120.08	1626.75	1609.72	1608.74	1654.74	1637.72	1636.73	F	294.18	2
16	101.11							K	147.11	1

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

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

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

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

5 References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

REGULAR ARTICLE

Proteomic signatures for histological types of lung cancer

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

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

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

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

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

1 Introduction

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

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

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

lar backgrounds determining particular histological phenotypes are obscure.

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

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

2 Materials and methods

2.1 Cell lines, clinical materials and protein extraction

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

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

Table 1. Clinical variables of lung cancer patients

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

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

2.2 Fluorescence labeling of protein samples

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

2.3 2-DE

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

2.4 Image acquisition and quantification of protein spots

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

2.5 Multivariate analysis of protein expression profiles

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

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

2.6 Identification and functional classification of proteins corresponding to protein spots

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

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

3 Results

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

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

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

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

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

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

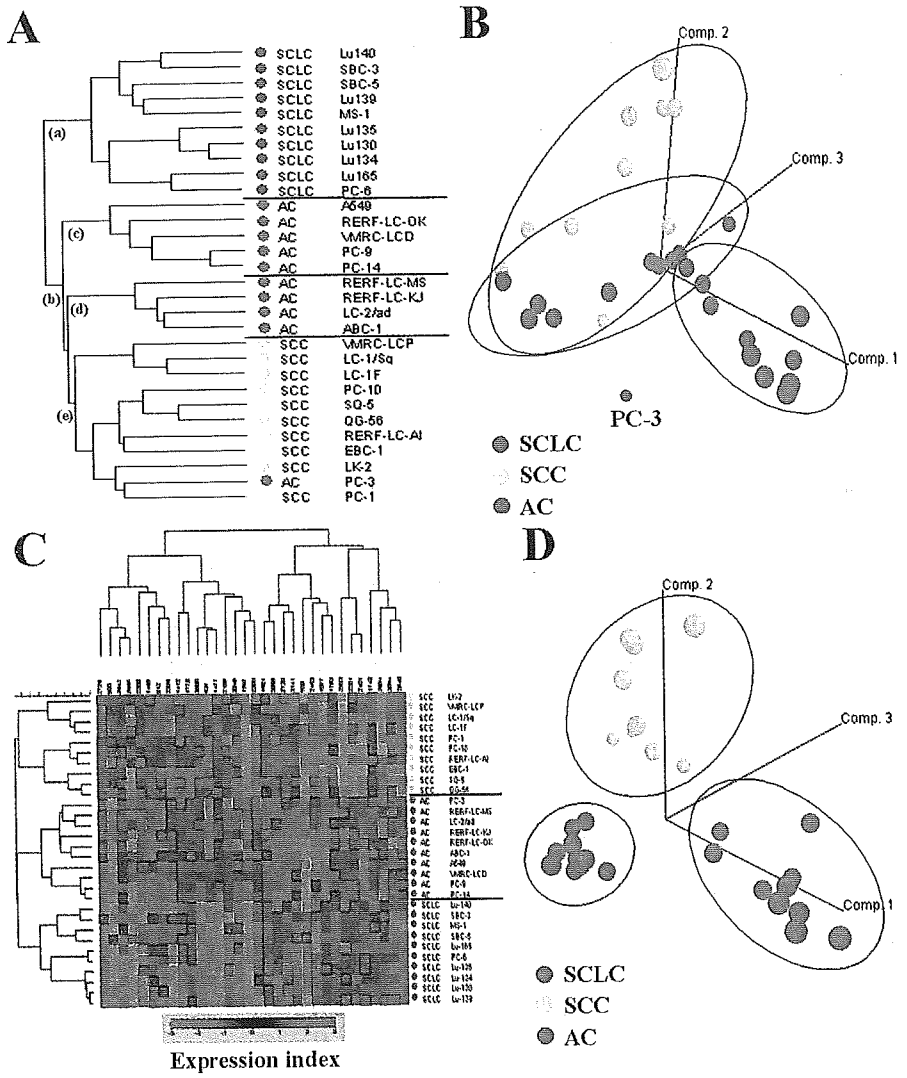


Figure 1. Statistical analysis of 30 lung cancer cell lines as a function of their protein expression profiles. (A) Dendrogram of hierarchical clustering analysis. The cell lines and their histology of origin are listed with color-coding on the left. (B) Three-dimensional plot of principal component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

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

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

3.3 Laser microdissection of lung cancer tissues and protein expression profile

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

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

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

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

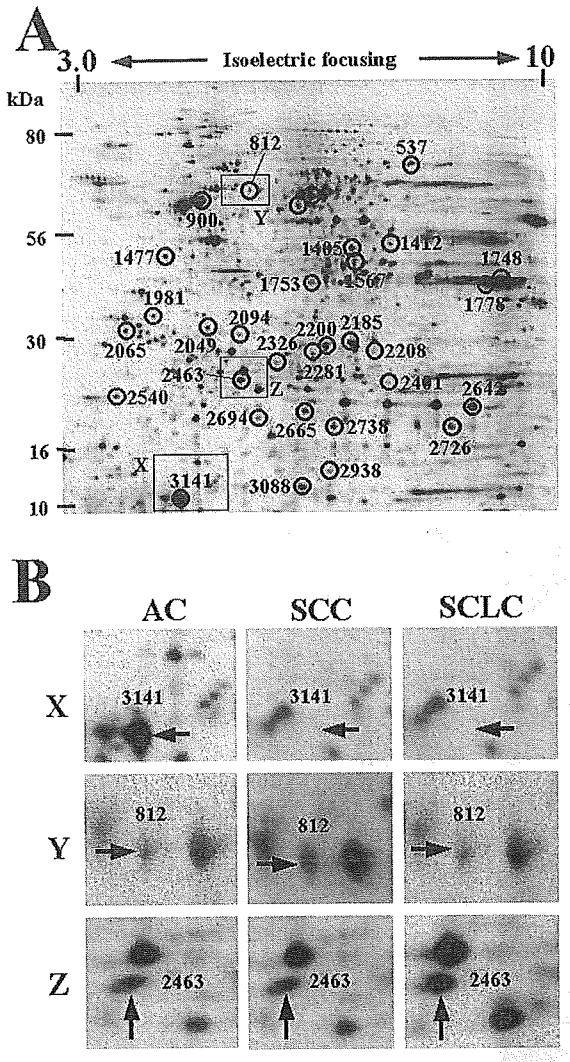


Figure 2. (A) Representative 2-D image of a Cy3-labeled protein mixture from 30 lung cancer cell lines. The 32 best-scoring protein spots for classification are circled; spot numbers correspond to those in Table 2. (B) Differential expression of proteins between cells of different origin is shown.

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

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

4 Discussion

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

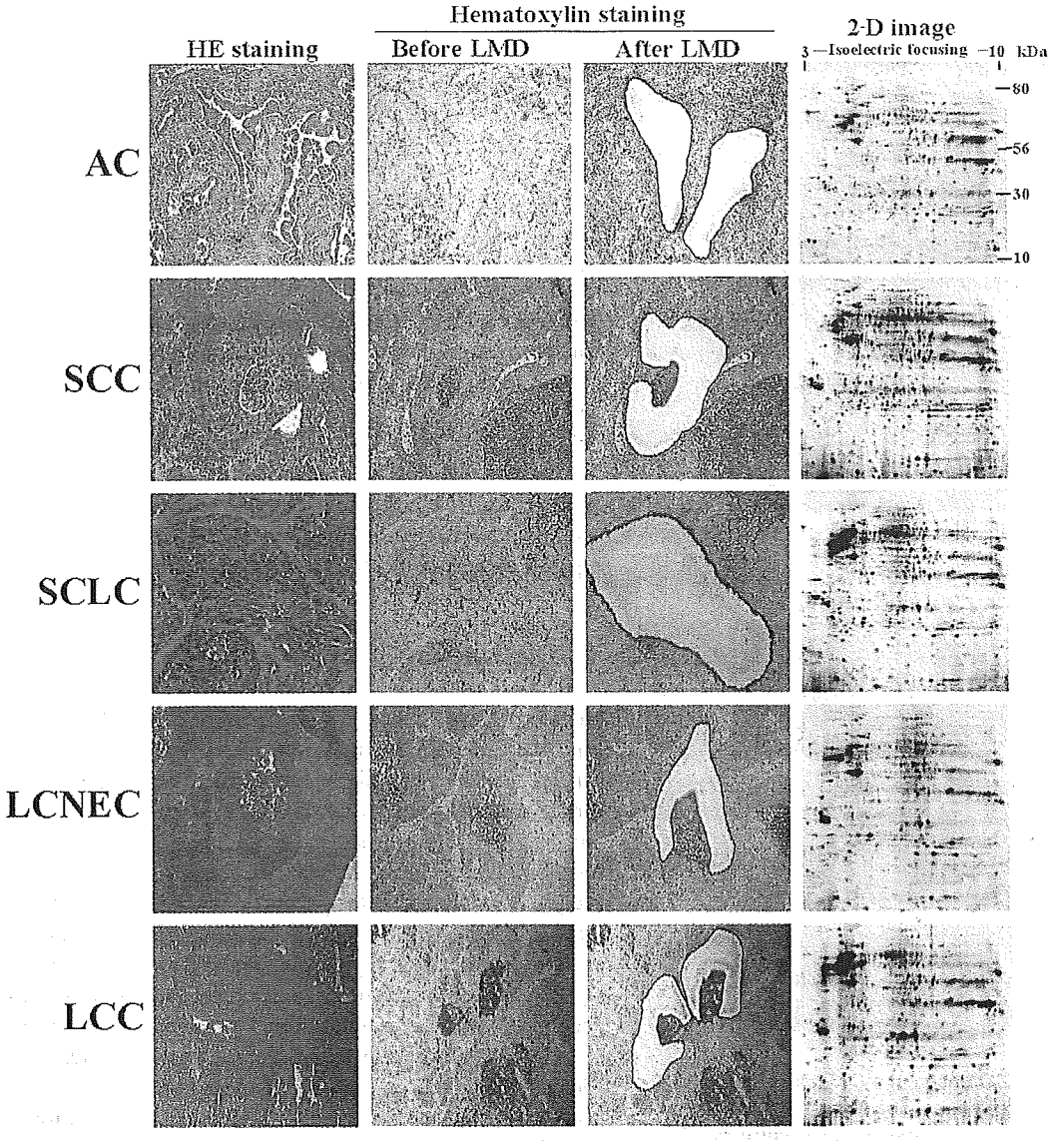


Figure 3. 2-D-DIGE of lung cancer cells isolated by laser microdissection. A frozen section of lung cancer tissue was stained with conventional hematoxylin-eosin staining (left panel). The neighboring section was stained with hematoxylin (middle left) and the tumor cells were recovered from that section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

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

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

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

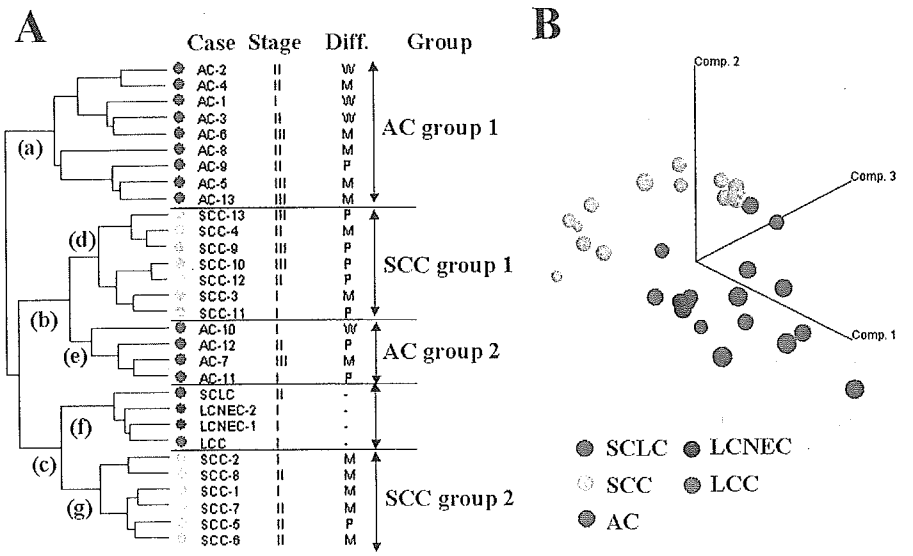


Figure 4. Multivariate studies on lung cancer cells obtained by laser microdissection. (A) Dendrogram of hierarchical clustering analysis of all lung cancer samples on the basis of 32 protein spots. The cell samples and prior information about their histology of origin are listed with color-coding on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

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

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

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

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

5 References

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

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

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

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

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

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

[7] Gharib, T. G., Chen, G., Wang, H., Huang, C. C. *et al.*, *Neoplasia* 2002, *4*, 440–448.

[8] Chen, G., Wang, H., Gharib, T. G., Huang, C. C. *et al.*, *Mol. Cell. Proteomics* 2003, *2*, 107–116.

[9] Chen, G., Gharib, T. G., Thomas, D. G., Huang, C. C. *et al.*, *Proteomics* 2003, *3*, 496–504.

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

[11] Cuezva, J. M., Chen, G., Alonso, A. M., Isidoro, A., *et al.*, *Carcinogenesis*, 2005, in press.

[12] Yanagisawa, K., Shyr, Y., Xu, B. J., Massion, P. P., *et al.*, *Lancet* 2003, *362*, 433–439.

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

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

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

[16] Molina, R., Filella, X., Auge, J. M., Fuentes, R. *et al.*, *Tumour Biol.* 2003, *24*, 209–218.

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

[18] Seike, M., Kondo, T., Fujii, K., Gemma, A. *et al.*, *Proteomics* 2004, *4*, 2776–2788.

[19] Siegenthaler, G., Hotz, R., Chatellard-Gruaz, D., Didierjean, L. *et al.*, *Biochem. J.* 1994, *302*, 363–371.

[20] Fan, K., Fan, D., Cheng, L. F., Li, C., *Anticancer Res.* 2000, *20*, 4809–4814.

[21] Shi, Y., Han, Y., Wang, X., Zhao, Y. *et al.*, *Gastric Cancer* 2002, *5*, 154–159.

[22] Shi, Y., Zhai, H., Wang, X., Wu, H. *et al.*, *Cell Mol. Life Sci.* 2002, *59*, 1577–1583.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 823 (2005) 82–97
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Two-dimensional electrophoresis database of fluorescence-labeled proteins of colon cancer cells

Yasuharu Mori [a], Tadashi Kondo [a,*], Tesshi Yamada [a], Akihiko Tsuchida [b], Tetsuya Aoki [b], Setsuo Hirohashi [a]

[a] *Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan*
[b] *Third Department of Surgery, Tokyo Medical University, Japan*

Received 7 May 2004; accepted 29 May 2005
Available online 11 July 2005

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 823 (2005) 82–97
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Two-dimensional electrophoresis database of fluorescence-labeled proteins of colon cancer cells

Yasuharu Mori [a], Tadashi Kondo [a,*], Tesshi Yamada [a], Akihiko Tsuchida [b], Tetsuya Aoki [b], Setsuo Hirohashi [a]

[a] *Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan*
[b] *Third Department of Surgery, Tokyo Medical University, Japan*

Received 7 May 2004; accepted 29 May 2005
Available online 11 July 2005

Abstract

We constructed a novel database of the proteome of DLD-1 colon cancer cells by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of fluorescence-labeled proteins followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) analysis. The database consists of 258 functionally categorized proteins corresponding to 314 protein spots. The majority of the proteins are oxidoreductases, cytoskeletal proteins and nucleic acid binding proteins. Phosphatase treatment showed that 28% of the protein spots on the gel are phosphorylated, and mass spectrometric analysis identified 21 of them. Proteins of DLD-1 cells and of laser-microdissected colon cancer tissues showed similar distribution on 2D gels, suggesting the utility of our database for clinical proteomics.

Keywords: Proteome; Two-dimensional gel electrophoresis; 2D database; 2D-DIGE; colon cancer

1. Introduction

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is an established method of protein separation [1,2], and has been used in many types of biological studies to monitor protein expression in a global manner [3,4]. However, analysis by 2D-PAGE is hampered by low reproducibility due to electrophoretic artifacts and by the limited dynamic range of spot intensity. Recently, novel fluorescent dyes have been developed and used for fluorescence two-dimensional difference gel electrophoresis (2D-DIGE) (Amersham Biosciences, Buckinghamshire, UK) with the aim of solving the problems inherent to 2D-PAGE. In 2D-DIGE, multiple samples are differentially labeled with different fluorescent dyes, mixed together, and co-separated in the same gel. Multiple 2D images are obtained by scanning the gel with appropriate wavelengths specific for each fluorescent dye [5–8]. Because samples are run concomitantly in the same gel, artifactual differences produced by electrophoresis can be distinguished from biological alterations, and spot matching can be performed in a less labor-intensive manner. 2D-DIGE also allows more accurate quantitative analysis of protein expression, since, by their nature, fluorescent dyes have a broader and more quantitative dynamic range than do colorimetric-based staining systems. In addition, the replacement of labor-intensive staining procedures by a simple laser scan enables high-throughput spot detection. With these advantages, 2D-DIGE has been applied to the study of breast cancer [9], esophageal cancer [10], and colon cancer [11]. However, global identification of the proteins corresponding to protein spots has been performed only on 2D images visualized by colorimetric methods, including

Abbreviations: 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; 2D-DIGE, fluorescence two-dimensional difference gel electrophoresis; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; CIAP, calf intestine alkaline phosphatase; IPG, immobilized pH gradient; DHB, dihydroxybenzoic acid; DTT, dithiothreitol; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone

* Corresponding author. Tel.: +81 3 3542 2511; fax: +81 3 3547 5298.
E-mail address: takondo@gan2.res.ncc.go.jp (T. Kondo).

1570-0232/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2005.05.024

silver staining [12], SYPRO Ruby [13] and Coomassie Brilliant Blue [14]. Because 2D databases have been created using only these images, the potential of 2D-DIGE for differential expression analysis has not been fully explored.

In this study, we studied the 2D image of fluorescence-labeled proteins. We identified by mass spectrometry (MS) 258 proteins corresponding to 314 fluorescence-labeled protein spots and functionally grouped the identified proteins according to the Panther Category classification in the Celera Discovery System. We discuss the characteristics of the proteins observed by 2D-DIGE, and the possible application of this technology for cancer research.

2. Materials and methods

2.1. Materials

1-(5-Carboxypentyl)-1′-propylindocarbocyanine halide *N*-hydroxy-succinimidyl ester (Cy3 fluorescent dye), 1-(5-carboxypentyl)-1′-methylindodi-carbocyanine halide *N*-hydroxysuccinimidyl ester (Cy5 fluorescent dye), and Decyder software Version 4.0 were purchased from Amersham Biosciences (Little Chalfont, Buckinghamshire, UK).

2.2. Cell culture, clinical material, and protein extraction

The human colon cancer cell line DLD-1 was obtained from the American Type Culture Collection (ATCC; Rockville, MD) and maintained in a humidified CO_2 incubator with RPMI-1640 medium supplemented with 10% FBS. DLD-1 cell protein samples were prepared as follows. After washing the cells with PBS and incubating them with 10% trichloroacetic acid (TCA) for 30 min on ice, they were scraped off and collected by brief centrifugation. The cell pellet was then suspended with lysis buffer (6 M urea, 2 M thiourea, 3% CHAPS and 1% Triton X-100) for 30 min on ice. The protein sample was centrifuged at 15,000 rpm for 30 min at 4 °C, and the supernatant was recovered.

Colon cancer tissue was obtained from a surgical specimen from a patient with colon cancer at Tokyo Medical University after obtaining the patient's informed consent. Laser microdissection was performed as described previously [15]. Briefly, the surgically resected colon cancer tissue was snap-frozen in liquid nitrogen immediately after resection and embedded in optimal cutting temperature (O.C.T.) compound (Sakura Finetechnical Co., Ltd., Tokyo, Japan). The O.C.T. embedded tissue blocks were cut into 10 μm thick sections with a Leica cryostat CM 3050 S (Leica, Milton Keynes, UK). The sectioned tissues were placed on a membrane-coated slide glass (Leica) pretreated with tissue adhesive solution, 0.1% poly-L-lysine (Sigma, St. Louis, MO). The sectioned tissues were fixed with 95% ethanol for 30 s [DD1] and washed in water. After being soaked in Mayer's hematoxylin (Muto Pure Chemicals, Tokyo, Japan) for 1 min, they were washed twice in 95% ethanol and once in water, each for 10 s. The neighboring section was occasionally stained with standard hematoxyline and eosine to confirm the diagnosis. All staining procedures were carried out on ice. Under microscopic observation, colon cancer cells were isolated with the Leica Microdissection System (Leica). The isolated tumor cells were treated with lysis buffer as described above.

Protein concentration was measured with a Protein Assay Kit (Bio-Rad Laboratories, Inc., Hercules, CA) and adjusted to 1 mg/ml with lysis buffer. The pH of the samples was adjusted to 8.5 with 30 mM Tris–HCl, a 50 μg protein sample was labeled with Cy3 or Cy5 fluorescent dye for 30 min, and the labeling reaction was terminated by adding 0.2 mM lysine for 10 min. The labeled sample was then incubated for 15 min with an equal volume of lysis buffer containing 130 mM DTT and 2.0% ampholine (Amersham Biosciences), and the total volume of the sample was adjusted to 420 μl with lysis buffer containing 65 mM DTT and 1.0% ampholine. All procedures after the start of the labeling reaction were carried out on ice in the dark.

2.3. 2D-PAGE

2D-PAGE was performed as described previously [1]. Briefly, IPG strip gels (24 cm long, p*I* range 3.0 and 10.0; Amersham Biosciences) were allowed to rehydrate with a labeled protein sample for 12 h at 20 °C. Isoelectric focusing was carried out with an IPGphor (Amersham Biosciences) for a total of 80 kVh at 20 °C. The IPG gel was equilibrated with equilibration buffer containing 3 M urea, 50 mM Tris–HCl (pH 8.8), 30% glycerol, 1.0% SDS, and 16 mM DTT for 15 min with gentle agitation. Another incubation was then carried out with the equilibration buffer in which DTT was replaced by 122 mM iodoacetamide. A 9–15% gradient polyacrylamide gel between low-fluorescence glass plates measuring 200 mm × 240 mm × 1 mm was used in the second separation. The equilibrated IPG gel was placed on the top of the second-dimension gel with embedding agarose (Amersham Biosciences), and the second-dimension electrophoresis was performed at 20 °C for 15 h at 18 W per gel with the EttanDalt II system (Amersham Biosciences). For preparative purposes, we applied 550 μg of protein sample. Although several hundreds of protein samples were required for preparative purpose, it is very costly to label whole proteins with fluprescent dyes. Because labeled and unlabeled proteins had almost identical electrophoretic mobility and 50 μg protein was enough for spot detection, we mixed 500 μg of unlabeled protein sample with 50 μg of labeled protein and separated them together in the same gel for preparative purpose. The second-dimension separation was performed as described above, except that the concentration of DTT was increased to 162 mM for the first equilibration, and iodoacetamide was replaced by 0.05% acrylamide for the second equilibration. The second-dimension gel was bonded to the inner plate

coated with bind-silane solution (Amersham Biosciences) according to the manufacturer's protocol. All electrophoresis procedures were performed in the dark.

2.4. Gel imaging and spot collection

The gels containing fluorescence-labeled proteins were scanned directly between the glass plates with a laser scanner (2D Master Gel Imager; Amersham Biosciences). After scanning, the same gel was stained with SYPRO Ruby dye (Molecular Probes, Eugene, OR) as follows. Briefly, the gel was fixed in 30% (v/v) methanol and 7.5% (v/v) acetic acid for 1 h, and after being washed with water it was incubated with SYPRO Ruby for 3 h at room temperature. The gel image was then read with a 2D Master Gel Imager. The images were exported to the DeCyder software as 16-bit tiff files. Molecular masses were determined by running a Broad Range Protein Marker (Bio-Rad, Hercules, CA) at the sides of selected gels. p*I* values were calculated according to the instructions of the manufacturer of the Immobiline DryStrips. Observed molecular mass and isoelectric point were determined with PDQUEST (Bio-Rad). Scatter plot analysis for pair-wise overall comparison was also performed with PDQUEST (Bio-Rad).

2.5. In-gel proteolytic digestion and mass spectrometry analysis for protein identification

The position of selected spots was recorded as a text file by use of DeCyder software. An automated spot collector, SpotPicker (Amersham Biosciences), excised the spots according to the content of the text file and transferred the gel plugs to 96-well plates. In-gel proteolytic digestion was carried out as described previously, with some modifications [16]. Briefly, the gel plugs were washed twice for 5 min with Milli Q water (Millipore, Bedford, MA) and then incubated with 100% acetonitrile for 10 min. After completely drying the gels with a SpeedVac (Thermo Savant, Holbrook, NY), the protein in the gel was digested overnight with 100 ng of TPCK-treated trypsin (Promega, Southampton, UK) in 50 mM ammonium bicarbonate at 37 °C for 3 h with gentle agitation. Following digestion, peptides were extracted with 50 μl of 50% acetonitrile/0.1% TFA, twice, and concentrated with the SpeedVac. A 1 μl sample of extracted peptides was then mixed with an equal volume of matrix solution composed of saturated dihydroxybenzoic acid (DHB) in 50% acetonitrile/0.1% TFA. The mixture was spotted onto a target plate and subjected to mass spectrometry analysis.

Protein identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) was performed with a Q-Star Pulser-i equipped with the oMALDI ion source (Applied Biosystems, Framingham, CA). We optimized various parameters including pulse rate, power level, collision energy/gas, depending on the samples and the ion peaks. All mass spectra were externally calibrated with a mixture of the three peptides included in the Sequenzyme Peptide Mass Standards kit (Applied Biosystems): des-arg1-bradykinin (Mr 904.4681), angiotensin I (Mr 1296.6853) and glu1-fibrinopeptide B (Mr 1570.6774). Mass spectra were processed with the Analyst QS program (Applied Biosystems). A minimum of a 5-peptide match, a maximum of 1 miscleavage, and cysteine modifications by acrylamide were used for peptide mass fingerprinting (PMF). A search of the SWISS-PROT database was performed with a mass tolerance of 100 ppm by using the Analyst QS program (Applied Biosystems). When PMF did not result in an unambiguous identification, selected peptides were subjected to MS/MS analysis. When the results by both MS and MS/MS analysis were consistent, the identification was considered as positive. The theoretical Mr and p*I* were obtained from the SWISS-PROT database. Results of identification by PMF and MS/MS data were scored by the Analyst QS and Mascot programs, and the top-scoring gene products were considered to be the corresponding proteins.

2.6. Functional classification of identified proteins

The identified proteins were functionally categorized according to Panther Category. In brief, the SWISS-PROT entry name of the identified proteins was entered into the Celera Discovery System (http://www.celera.com). We searched how the identified proteins were categorized in the database and the results were summarized in Table 2.

2.7. Dephosphorylation of proteins by phosphatase treatment

Proteins were dephosphorylated by treatment with calf intestinal alkaline phosphatase (CIAP; New England Biolabs, Beverly, MA). Briefly, a 750 μg protein sample was diluted with a reaction buffer consisting of 1 M NaCl, 0.1 M $MgCl_2$, 10 mM DTT and 0.5 M Tris (pH 7.9) and divided into two equal volumes. Then, 10 μl of CIAP (10 U/μl) diluted in CIAP buffer, consisting of 50 mM KCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$ and 10 mM Tris (pH 8.2), was added to one sample, and 10 μl of the CIAP buffer lacking CIAP was added to the other sample. After incubation at 37 °C for 15 min, the reaction was terminated by addition of urea lysis buffer. The samples were labeled with Cy3 and Cy5 fluorescence dyes, respectively, mixed together and separated by 2D-PAGE in the same gel as described above. Pair-wise overall comparison of spot intensity was performed by scatter plot analysis using PDQUEST (Bio-Rad).

2.8. Pair-wise comparison of 2D images from DLD-1 cells and colon cancer cells

Protein samples from DLD-1 cells and from microdissected colon cancer tissue were labeled with Cy3 and Cy5 fluorescent dyes, respectively, and subjected to 2D-PAGE as described above. Pair-wise overall comparison was performed by scatter plot analysis as described above.

3. Results

3.1. Identification of fluorescence-labeled protein spots by MALDI-TOF

Among the protein spots present in both the quantitative and analytical gels, approximately 500 spots were chosen without regard to their intensity and subjected to in-gel proteolytic digestion followed by mass spectrometry analysis. Protein identification was first achieved by PMF and, for some proteins, the results were further confirmed by MS/MS analysis. We finally identified the proteins corresponding to 314 protein spots listed in Table 1. Fig. 1 shows a representative 2D pattern of fluorescence-labeled proteins of DLD-1 cells with the spot numbers of the identified proteins. A representative mass spectrometric pattern of trypsin digests from spot 216 is shown in Fig. 2A. Processing of ion peaks and a database search resulted in the identification of spot 216 as Rho GDP-dissociation inhibitor 1 (Fig. 2B). Mass spectrometry analysis was performed on the other protein spots, and the spots with sufficient peptide coverage to obtain positive protein identification are listed in Table 1.

3.2. Classification of identified proteins

We grouped the identified protein spots according to Panther Category by accessing the Celera Discovery System (Table 2). The majority of the proteins belonged to the oxidoreductase (10.5%), cytoskeletal protein (9.4%), nucleic acid binding protein (8.9%), chaperone (7.0%) and isomerase (6.7%) families (Table 2). In this study, members of certain protein families, including transcription factors, transporters, membrane traffic proteins and viral proteins,

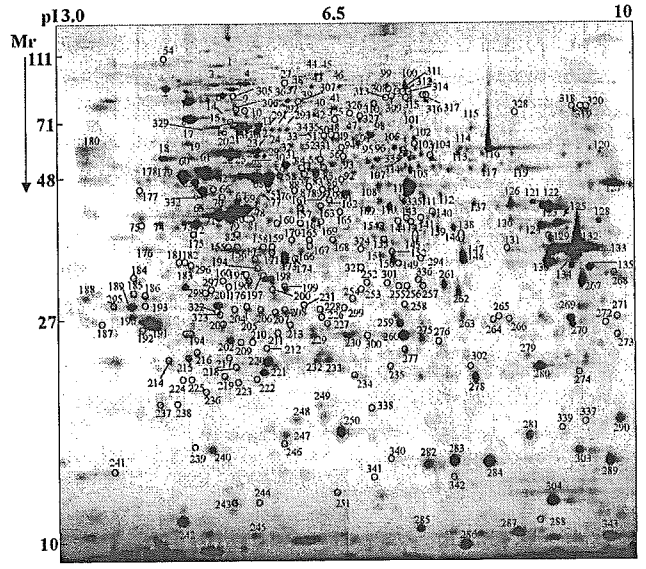
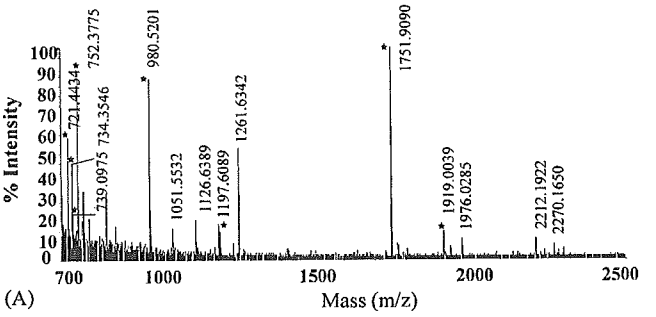


Fig. 1. A representative 2D image of Cy5-labeled proteins extracted from DLD-1 cells. The total number of spots observed is approximately 1500. The numbers assigned to the protein spots are those shown in Table 1. Proteins not identified by mass spectrometry analysis are not numbered in this figure.



(A)

Rho GDP-dissociation inhibitor 1

(B)

MAEQEPTAEQ LAQIAAENEE DEHSVNYKPP AQKSIQEIQE
LDKDDESLRK YKEALLGRVA VSADPNVPNV VVTGLTLVCS
SAPGPLELDL TGDLESFKKQ SFVLKEGVEY RIKISFRVNR
EIVSGMKYIQ HTYRKGVKID KTDYMVGSYG PRAEEYEFLT
PVEEAPKGML ARGSYSIKSR FTDDDKTDHL SWEWNLTIKK
DWKD

Fig. 2. (A) Mass spectrogram of spot no. 216 in Fig. 1. A total of eight peptides (indicated by asterisks) are assigned to those of Rho GDP-dissociation inhibitor 1. (B) The peptide sequences corresponding to the peptide peaks in the mass spectrogram are underlined. The amino acid sequence coverage is 49.5%.

were not observed. However, these proteins may exist in the range of 2D-PAGE and the reason why they were not observed is probably because their expression level was lower than the detection limit of Cy5 fluorescent dye. Based on the genome database, Medjahed et al. constructed the virtual 2D image; they calculated p*I* and molecular weight of gene products and plotted the products in two-dimensional space [18]. According to their data, p*I* and molecular weight of most of proteins distributed within the range of 2D-PAGE. More extensive study such as sample fractionation prior to electrophoresis may visualize these lower abundant proteins. Our data showed that some regulatory proteins such as kinases and receptors, which are generally expressed in small amounts, were already identified in 2D gel. Therefore, 2D-PAGE has a potential to uncover most of proteome and increasing the sensitivity for spot detection should be the critical issue of developing 2D-DIGE. Since 2D imaging of fluorescence-labeled proteins allows monitoring of proteins from a broad spectrum of functions, it represents a powerful tool for initial study of the proteome and will be useful for biological studies in which quantitative expression profiles are required.

3.3. Redundancy of identified proteins and prediction of total number of proteins

We looked into Table 1 and found that, among the 314 protein spots identified, 52 gene products appeared more than once (Table 3). Post-translational modifications, including phosphorylation, splicing variation, and glycation, can result in more than one protein spot from the same gene. To estimate the contribution of phosphorylation to spot redundancy in our 2D gels, the protein sample was dephosphorylated by phosphatase treatment and subjected to 2D separation. 2D images of proteins with or without phosphatase treatment