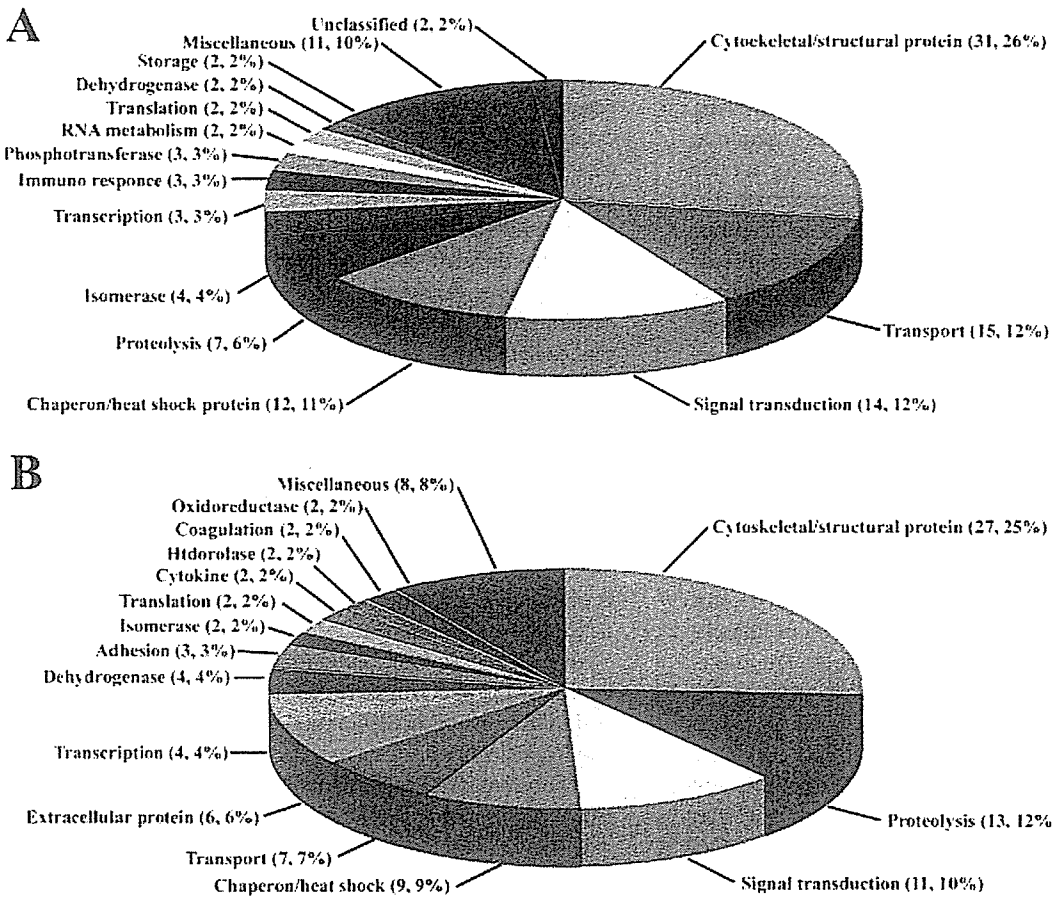

Figure 11. Functional classification of the 217 proteins identified corresponding to the 498 protein spots detected. Proteins function is based on their classification in Gene Ontology and the literature curation. Proteins up- (n = 113) and down- (n = 104) regulated in tumor tissues are shown in panels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

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

4 Discussion

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

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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

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

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

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

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

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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

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

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

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

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

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

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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

5 References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

DOI 10.1002/pmic.200500883 *Proteomics* 2006, *6*, 3938–3948

RESEARCH ARTICLE

Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis

Tetsuya Okano[1, 4], Tadashi Kondo[1], Tatsuhiko Kakisaka[1], Kiyonaga Fujii[1], Masayo Yamada[1], Harubumi Kato[2, 3], Toshihide Nishimura[2], Akihiko Gemma[4], Shoji Kudoh[4] and Setsuo Hirohashi[1]

[1] Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan
[2] Clinical Proteome Center, Tokyo Medical University, Tokyo, Japan
[3] Department of Surgery, Tokyo Medical University, Tokyo, Japan
[4] Fourth Department of Internal Medicine, Nippon Medical School, Tokyo, Japan

To investigate aberrant plasma proteins in lung cancer, we compared the proteomic profiles of serum from five lung cancer patients and from four healthy volunteers. Immuno-affinity chromatography was used to deplete highly abundant plasma proteins, and the resulting plasma samples were separated into eight fractions by anion-exchange chromatography. Quantitative protein profiles of the fractionated samples were generated by two-dimensional difference gel electrophoresis, in which the experimental samples and the internal control samples were labeled with different dyes and co-separated by two-dimensional polyacrylamide gel electrophoresis. This approach succeeded in resolving 3890 protein spots. For 364 of the protein spots, the expression level in lung cancer was more than twofold different from that in the healthy volunteers. These differences were statistically significant (Student's *t*-test, *p*-value less than 0.05). Mass spectrometric protein identification revealed that the 364 protein spots corresponded to 58 gene products, including the classical plasma proteins and the tissue-leakage proteins catalase, clusterin, ficolin, gelsolin, lumican, tetranectin, triosephosphate isomerase and vitronectin. The combination of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis provides a valuable tool for serum proteomics in lung cancer.

Received: December 6, 2005
Revised: February 19, 2006
Accepted: March 12, 2006

Keywords:
Lung cancer / Multi-dimensional liquid chromatography / Two-dimensional difference gel electrophoresis

1 Introduction

Lung cancer is a leading cause of cancer death in Japan, claiming 55 000 lives annually, and is a major health problem in many countries. The prognosis of patients with lung cancer is generally poor, with an overall 5-year survival rate for patients receiving treatment of only 14%. In contrast, overall 5-year survival for patients diagnosed with stage I adenocarcinoma approaches 63% [1]. As non-small-cell lung cancer accounts for almost 80% of lung cancers, of which 40% are adenocarcinoma, a substantial number of patients with lung cancer have the potential to be cured successfully by early treatment. However, the majority of lung tumors have reached locally advanced stage III (33%) or metastatic stage IV (41%) by the time of diagnosis [2]. Therefore, early diagnosis of lung cancer is necessary to improve patient survival. Plasma is a preferred specimen for the early diagnosis of lung cancer because samples are easily available by non-

Correspondence: Professor Tadashi Kondo, Proteome Bioinformatics 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

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

invasive methods. However, the currently available plasma tumor markers such as CEA, NSE, TPA, chromogranin, CA125, CA19–9, Cyfra 21–1, and ProGRP have limited sensitivity and specificity for early diagnosis [3], and novel plasma markers are required.

Multi-dimensional separation techniques based on the combination of LC and gel electrophoresis have been applied for plasma proteomics [4]. Typically, LC with immuno-affinity, size-exclusion, ion-exchange and RP columns is used to fraction the proteins, which are then subjected to high-resolution 2-D-PAGE for quantitative expression studies. Sample fractionation prior to 2-D-PAGE can separate gene products present in low copy numbers from highly abundant proteins, increasing the number of observable proteins. Indeed, less abundant proteins, including tissue-leakage proteins and regulatory proteins, have been identified in plasma by the combination of LC and 2-D-PAGE [4]. As the results of 2-D-PAGE can be quantified and stored in a database, 2-D-PAGE is a powerful tool for biomarker development. However, the intrinsic limitations of 2-D-PAGE resulting from gel-to-gel variations can hinder accurate comparisons of protein expression levels. To cancel such experimental differences, 2-D-DIGE has been developed [5–7]. In 2-D-DIGE, the experimental samples and an internal control sample are labeled with different fluorescent dyes, mixed together and co-separated in identical gels. The fluorescent dyes are designed so that the electrophoretic migration of proteins labeled with the different dyes is almost identical. Therefore, the intensity of the spots of the experimental sample can be normalized to the intensity of the corresponding spots of the internal control sample in the same gel. In 2-D-DIGE the amounts of proteins are measured as fluorescence signals, thus the dynamic range is wider than with conventional silver staining and spot detection can be achieved by simple laser scanning in a high-throughput manner. 2-D-DIGE has been used to characterize the proteome of plasma from patients with disease [8]. Recently, novel highly sensitive fluorescent dyes, CyDye DIGE Fluor saturation dyes (GE Healthcare Amersham Biosciences, Uppsala, Sweden), referred to here as 'saturation dye', have been developed [9]. The high sensitivity of saturation dye has enabled proteomic studies of primary cultured human hepatocytes [10] and laser-microdissected tumor tissues [11–14], where only limited amounts of proteins were available.

In this study, we used the combination of multi-dimensional chromatography and 2-D-DIGE with saturation dye to conduct a proteomic comparison of serum from patients with lung cancer and from healthy volunteers. The high sensitivity of saturation dye generated 3890 protein spots from only 30 μL of plasma. We identified 364 spots as aberrantly regulated plasma proteins in lung cancer. MS revealed that these spots corresponded to 58 gene products. The identified proteins included previously uncharacterized tissue-leakage proteins in addition to acute phase plasma proteins. These results demonstrate the utility of multi-dimensional LC and 2-D-DIGE for plasma proteomics in cancer.

2 Materials and methods

2.1 Serum samples

Blood samples by venipuncture and informed consent were obtained from five patients with lung cancer and four healthy volunteers at Nippon Medical School. Clinical information for the donors of serum proteins is summarized in Table 1. A 10-mL blood sample was obtained with a VENOJECT II (10 mL; TERUMO, Tokyo, Japan) and allowed to clot for 2 h at 4°C. The clotted material was removed by centrifugation at 3000 rpm for 10 min. The supernatant sera obtained from the blood samples were recovered and stored at −80°C until use.

Table 1. Patients' characteristics

	Histology [29]	Gender	Age (years)	Stage [30]
1.	Adenocarcinoma	Male	48	T2N2M0 IIIA
2.	Adenocarcinoma	Male	62	T4N2M0 IIIB
3.	Squamous cell carcinoma	Female	75	T4N3M1 IV
4.	Squamous cell carcinoma	Male	76	T4N1M1 IV
5.	Small cell carcinoma	Male	40	T2N2M0 IIIA
Four healthy volunteers		4 Males	29~38	

2.2 Immuno-affinity and anion-exchange chromatography

Serum proteins were separated by immuno-affinity and ion-exchange chromatography with the AKTA Explore system (GE Healthcare Amersham Biosciences). The immuno-affinity column (4.6 × 100 mm; Agilent Technologies) contained anti-albumin, anti-transferrin, anti-haptoglobin, anti-alpha-1-anti-trypsin, anti-IgA and anti-IgG resins. A serum sample (30 μL) was diluted with 120 μL of a neutral buffer (buffer A; Agilent Technologies), and filtered with a spin filter (pore size 0.22 μm; Agilent Technologies) by centrifugation at 16 000 × *g* for 1 min prior to use. The filtered sample was applied to the immuno-affinity column at a flow rate of 0.5 mL/min for 10 min. The flow-through fraction was recovered and the proteins retained in the column were eluted with a low-pH urea buffer (buffer B; Agilent Technologies) at a flow rate of 1.0 mL/min for 10 min. The column was recycled for further use by washing with buffer A at a flow rate of 1.0 mL/min for 12 min. The flow-through fractions were concentrated to 500 μL in a Spin Concentrator, 5K MWCO (4 mL capacity, Agilent Technologies).

The concentrated sample was diluted with 4.5 mL of 25 mM Tris-HCl, pH 9.0, and applied to a Resource Q column (1.0 mL resin, 6.4 mm id × 30 mm; GE Healthcare

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Amersham Biosciences) at a flow rate of 4.0 mL/min. The separations were performed with a step-wise gradient of NaCl as follows: 0 mM for 12.5 min, 100, 150, 200, 250, 300, 350, and1000 mM for 2.5 min each. All elution buffers contained 25 mM Tris-HCl, pH 9.0. The anion-exchange column was then recycled for further use by thorough washing with 25 mM Tris-HCl, pH 9.0, containing 2 M NaCl. In the interval between the elution steps, the pump system was washed with 10 mL of the next elution buffer. Protein peaks were monitored at 280 nm. The eluted proteins were concentrated with Amicon Ultra PL-10 (a molecular mass cut-off 10 kDa) in an Amicon Ultra-15 filter unit (Amicon Bredford, MA).

2.3 2-D DIGE

The fractionated proteins were precipitated by addition of four volumes of acetone at −20°C for 20 min. After centrifugation at 13 000 rpm for 10 min, the supernatant was discarded and the pellet was air-dried for 10 min. The dried pellet was dissolved in 50 μL of lysis buffer containing 6 M urea, 2 M thiourea, 3% CHAPS, 1% Triton X-100 and 40 mM Tris (pH 8.0). The protein concentration was measured with a Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA). The fluorescence labeling was performed as described previously with some modifications [10]. In brief, samples (50 μg of protein) were reduced by incubation with 2 μM Tris-(2-carboxethyl)phosphine hydrochloride (TCEP; Sigma, St. Louis, MO) at 37°C for 60 min. The protein samples were fluorescence labeled by incubation with 40 nM of saturation Cy3 or Cy5 dye (GE Healthcare Amersham Biosciences) at 37°C for another 30 min. For 2-D-PAGE, the labeling reaction was terminated by addition of an equal volume of the lysis buffer containing 130 mM DTT and 2.0% Pharmalyte (Amersham Biosciences). In the fraction eluted with 1000 mM NaCl, the amount of protein was so small that its concentration could not be measured. Therefore, all the protein in these fractions was labeled and used for the subsequent studies. As the trains of spots in the vertical dimension were not observed on 2-D images, we concluded that the proteins were saturatedly labeled with fluorescent dyes.

The 2-D-PAGE was carried out as described previously in our reports [11]. In brief, the first dimension separation was carried out with Immobiline Drystrips (24 cm, pH 3–10; GE Healthcare Amersham Biosciences). Each strip was rehydrated for 12 h at 30 V with 420 μL of protein sample, and IEF was performed in an IPGphor unit (GE Healthcare Amersham Bioscience). After electrophoresis, the strips were equilibrated in equilibration buffer (6 M urea, 2% SDS, 50 mM Tris HCl, pH 8.8, 30% glycerol w/v) for 20 min. The second dimension separation was performed on homemade 9–15% gradient polyacrylamide gels with an EttanDalt II system (GE Healthcare Amersham Biosciences) at a constant wattage of 17 W at 20°C for 17 h. The image was acquired by scanning the gels with a laser scanner (2D MasterImager; GE Healthcare Amersham Biosciences). Spot detection, quantification and image matching were performed with DeCyder software (GE Healthcare Amersham Biosciences). One portion of the labeled proteins was separated by SDS-PAGE using 10% polyacrylamide gel to monitor the contents of the fractions. Gel electrophoresis was performed in the dark.

2.4 Protein identification by MS

In-gel digestion was performed for protein spots excised by an automated spot collector (SpotPicker; GE Healthcare Amersham Biosciences), according to our previous report [10]. Then, the gel pieces were extensively washed with ammonium bicarbonate. The protein in the dried gel plug was digested overnight at 37°C with sequencing-grade modified trypsin (Promega, Madison, WI). The tryptic digests were recovered by incubation with 50% ACN/0.1% TFA, and the isolated peptides were concentrated under nitrogen gas and subjected to LC-MS/MS. The MS study was carried out as described previously [15]. In brief, the LC-MS/MS system comprised a Paradigm MS4 dual solvent delivery system (Michrom BioSciences, Auburn, CA, USA) for HPLC, an HTS PAL auto sampler with two 10-port injector valves (CTC Analytics, Zwingen, Switzerland), and a Finnigan LTQ linear ITMS (Thermo Electron, San Jose, CA, USA) equipped with NSI sources (AMR, Tokyo, Japan). A database search against Swiss-Prot was performed with MASCOT software. When multiple proteins were identified in a single spot, the proteins with the highest numbers of peptides were considered as those corresponding to the spots. When multiple protein candidates were listed with an equal number of the identified peptides, the proteins with a higher MASCOT score were selected.

2.5 Western blotting

Protein samples were separated by SDS-PAGE and transferred onto NC membranes. Differential expression of plasma proteins was examined by specific antibodies against leucine-rich alpha 2 glycoprotein (x1000 dilution; Abnova, Taiwan), inter-alpha-trypsin inhibitor heavy chain H4 (x1000 dilution; Santa Cruz Biotechnology, Santa Cruz, CA), plasma retinol binding protein (x50 dilution; Biomeda), haptoglobin (x2000 dilution; Sigma Aldrich, St. Louis, MO, USA), complement component C4 (x200 dilution; Antibody Shop, Gentofte, Denmark), complement component C3 (x200 dilution; Antibody Shop), and prothrombin (x100 dilution; BD Transduction laboratories). The anti-prothrombin antibody could recognize both thrombin and prothrombin. A second antibody against goat IgG (Santa Cruz Biotechnology) was used for inter-alpha-trypsin inhibitor heavy chain H4 at a dilution of 1:2000. A second antibody against mouse IgG (GE Healthcare Amersham Biosciences) was used for leucine-rich alpha 2 glycoprotein, plasma retinol-binding protein, haptoglobin, complement component C4, complement component C3 and prothrombin at a dilution of 1:1000. Immune complexes were detected with an enhanced

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

chemiluminescence system (GE Healthcare Amersham Biosciences) and monitored with an LAS-1000 (Fuji Film, Tokyo, Japan).

3 Results

3.1 Evaluation of the fractionation process

In the first dimension chromatography, we used a recently developed immuno-affinity column to deplete the most abundant plasma proteins, including albumin, immunoglobulin, transferrin, anti-trypsin, and haptoglobin. This immuno-affinity column has been employed for plasma proteomics to enrich the less abundant plasma proteins [4, 16–18]. Whole serum was separated into two fractions: the flow-through fraction containing low abundance serum proteins and the bound fraction containing the abundant proteins listed above (Fig. 1A). Proteins in the flow-through fraction were subjected to anion-exchange chromatography (Fig. 1B). To ensure reproducibility of protein overlap between neighboring fractions, we employed step-wise separation instead of gradient separation. In addition, we washed the column with the elution buffer at the end of each fraction step to reduce the carry-over of elution buffer between the steps. To evaluate the reproducibility of the fractionation process, the 250-mM NaCl fraction of three independently prepared protein samples was separated by RP chromatography (Fig. 1C). The three chromatograms showed that the fractionation process was highly reproducible.

To examine the effects of fractionation, the protein samples were labeled with fluorescent dyes and separated by SDS-PAGE (Fig. 2). The major bands in the unfractionated serum proteins, as indicated by arrows (lane 1), were observed in the bound fraction of the immuno-affinity column (lane 3). Proteins in the flow-through fraction of the immuno-affinity column generated a higher number of bands (lane 2). The other portion of this fraction was further separated by anion-exchange chromatography, labeled with fluorescent dyes and subjected to SDS-PAGE (lanes 4–11). The fractions from anion-exchange chromatography showed distinct protein migration patterns (lanes 4–11) compared with the unfractionated serum (lane 1) and the flow-through fraction of the immuno-affinity column (lane 2). Anion-exchange chromatography combined with immuno-affinity chromatography reduced the complexity and dynamic range of the protein content and separated the less abundant proteins from the neighboring highly abundant proteins, resulting in an increase in the number of observable proteins.

We conducted 2-D-PAGE to separate further the fractionated proteins. The 2-D-PAGE allows detection of post-translational differences, observed as a shift in p*I* that are difficult to recognize by SDS-PAGE. To validate the effects of fractionation on the number of observable proteins, we labeled a whole-serum sample and a fractionated protein

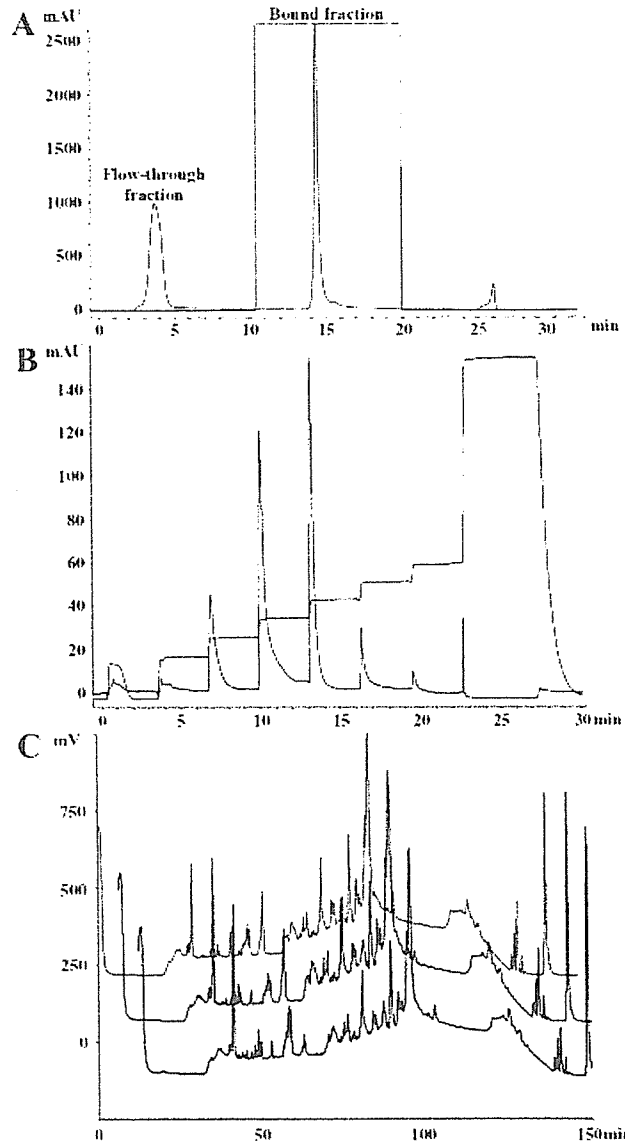

Figure 1. Process of multi-dimensional chromatography separation. (A) The immuno-affinity column separated plasma proteins into two fractions: flow-through and bound fraction. The column was recycled by washing after elution of the bound proteins. (B) The flow-through fraction from the immuno-affinity column was subjected to anion-exchange chromatography and separated into eight fractions by elution with a step-wise gradient of NaCl. (C) Three independent samples of the 250 mM NaCl fraction were prepared and separated by RP chromatography to examine the reproducibility of the immuno-affinity and anion-exchange procedures.

sample with Cy3 and Cy5, respectively, mixed them together and co-separated them by 2-D-PAGE. Figure 3 shows the two-color images of whole-serum (red) and fractionated proteins (green). It is apparent that most of the protein spots that bound to, and were recovered from the immuno-affinity column, matched the spots with high intensity in whole

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

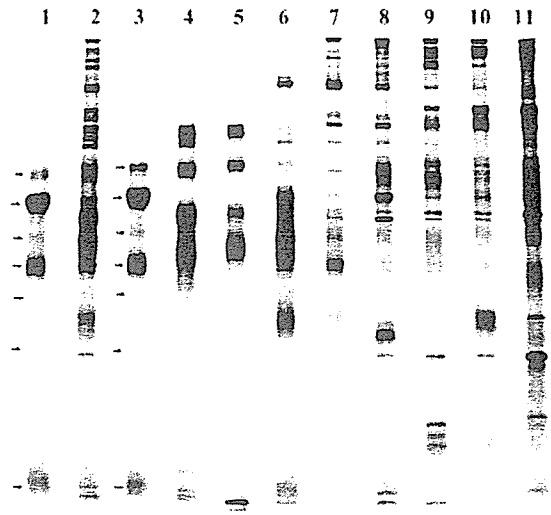

Figure 2. Separation of whole and fractionated plasma samples by SDS-PAGE. Lane 1, whole serum sample; lane 2, flow-through fraction from the immuno-affinity column; lane 3, bound fraction from the immuno-affinity column; lanes 4–11, the fractions eluted by 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

serum (panel A). In contrast, immuno-depletion of these proteins resulted in a distinct 2-D image; most of the protein spots in the flow-through fraction of the immuno-affinity column were not observed in the 2-D image of whole serum (panel B). These results are consistent with previous reports [17, 19]. We also separated the Cy5-labeled fractions from anion-exchange chromatography by 2-D-PAGE (panels C–J). Visual comparison of the location and intensity of spots through the series of 2-D images can be achieved easily by using the Cy3 image of whole serum as a common internal standard. It is clear that the number of spots was increased by fractionation with anion-exchange chromatography. However, it is unclear how many proteins repeatedly appeared as different spots in the gels.

3.2 Comparison of serum from patients with lung cancer and healthy volunteers

We compared serum proteins in patients with lung cancer and healthy volunteers (Table 1). We prepared a mixture of the paired protein samples to make an internal standard and labeled it with Cy3. The individual samples were labeled with Cy5 and mixed with the Cy3-labeled internal standard sample. The mixture of labeled protein samples was separated by 2-D-PAGE. Figure 4 shows a representative gel image of Cy5-labeled proteins. The number of spots observed is described in the lower right corner of the 2-D image panel. The number of spots in the mixture of unfractionated serum of patients with lung cancer and of healthy volunteers was 124, and fractionation increased the total number of spots to 3766. The arrows indicate the 364 spots whose expression in lung cancer differed significantly by more than twofold from that in the healthy volunteers (Student's *t*-test, *p*-value less than 0.05). MS identified the proteins from the spots. We observed multiple proteins from single spots in several cases. These observations were consistent with the previous report [20]. Tentatively, the proteins with the highest number of peptides were considered to be the proteins corresponding to the spots, although this is not always the case. Based on this criterion, we concluded that MS study of these 364 spots resulted in the identification of 58 distinct gene products. The results of protein identification are summarized in Suppl. Table 1.

We found that certain proteins appeared on the 2-D images as multiple protein spots with aberrantly controlled expression in lung cancer. In addition, different protein spots translated from the same gene showed different regulation, so that isoforms of a particular plasma protein could be up- or down-regulated in lung cancer. Figure 5 summarizes these observations. Among 58 proteins identified, 18 were detected as a single spot. These were alpha-1-acid glycoprotein, apolipoprotein D, apolipoprotein M, carbonic anhydrase II, catalase, ceruloplasmin, complement C5, corticosteroid-binding globulin, Ig alpha-1 chain C region, Ig gamma-1 chain C region, Ig kappa chain V-III region SIE, inter-alpha-trypsin inhibitor heavy chains H1, H2 and H4, kininogen, tetranectin, triosephosphate isomerase, and vitronectin. Spots for 28 proteins showed consistent up- or down-regulation (Fig. 5A) and the spots for 12 proteins showed inconsistent regulation (Fig. 5B). For example, the intensity of all 67 haptoglobulin spots was increased in lung cancer, whereas for the 38 spots of complement component C3 the intensity of 20 spots was up-regulated and that of 18 spots was down-regulated in lung cancer. Thus, isoform-specific regulation may exist for these proteins.

We then used specific antibodies to validate the differential expression of the plasma proteins whose isoforms showed consistent aberrations in lung cancer (Fig. 6), because clinical application of our results would require differential expression to be monitored by methods more convenient than the combination of multi-dimensional LC and 2-D-DIGE. We examined individual samples by SDS-PAGE followed by Western blotting with specific antibodies. The higher intensity of the lower bands of leucine-rich alpha-2-glycoprotein in lung cancer was consistent with the results of multi-dimensional chromatography followed by 2-D-PAGE (Fig. 6A). The intensity of the upper bands of leucine-rich alpha-2-glycoprotein did not differ between the two groups and the spots corresponding to these bands were not recognized as those showing aberrant intensity in lung cancer (Fig. 6A). The higher intensity of haptoglobulin bands was also consistent with the results of 2-D-PAGE (Fig. 6D). However, the intensities of the bands of inter-alpha-trypsin inhibitor heavy chain H4 (Fig. 6B) and plasma retinol binding protein (Fig. 6C) were also higher in lung cancer, and these results were inconsistent with those from multi-

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

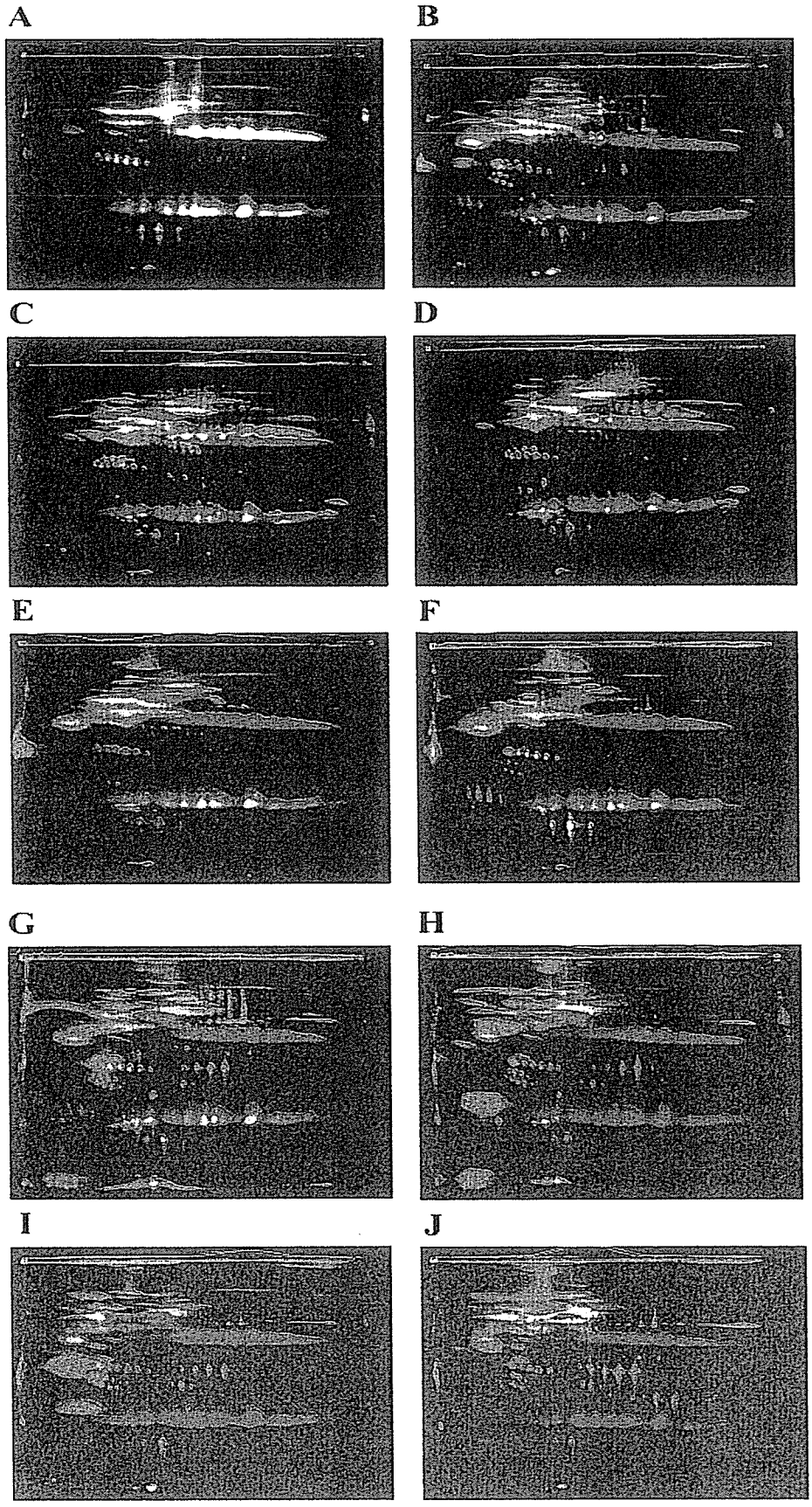

Figure 3. 2-D-PAGE of plasma proteins. Whole serum and fractionated proteins were labeled with Cy3 (red) and Cy5 (green), respectively. The labeled protein samples were mixed together and co-separated by 2-D-PAGE. The fractionated proteins showed distinguishable protein profiles, suggesting the efficient enrichment of low abundance proteins by fractionation. The sources of the protein samples are as follows: (A) bound fraction from the immuno-affinity column; (B) flow-through fraction from the immuno-affinity column; (C–J) 2-D images of the fractions eluted with 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

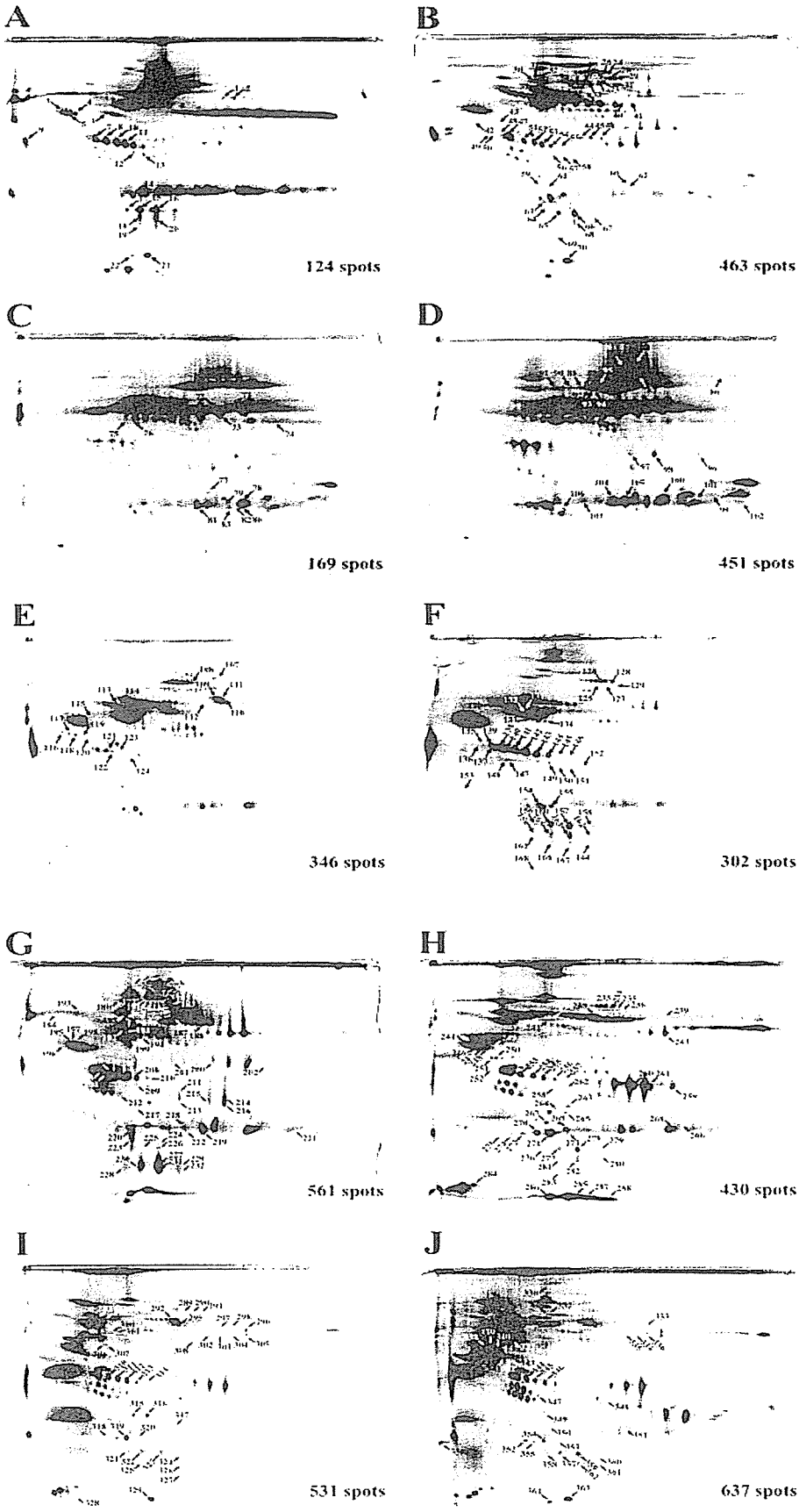

Figure 4. Localization of spots on the 2-D map of Cy5-labeled proteins. The arrows indicate the spots showing aberrant expression in lung cancer. The spot numbers correspond to those in Suppl. Table 1. The sources of the protein samples are as follows: (A) unfractionated serum; (B) flow-through fraction from the immuno-affinity column; (C–J) 2-D images of the fractions eluted with 0, 100, 150, 200, 250, 300, 350 and 1000 mM NaCl from the anion-exchange column.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

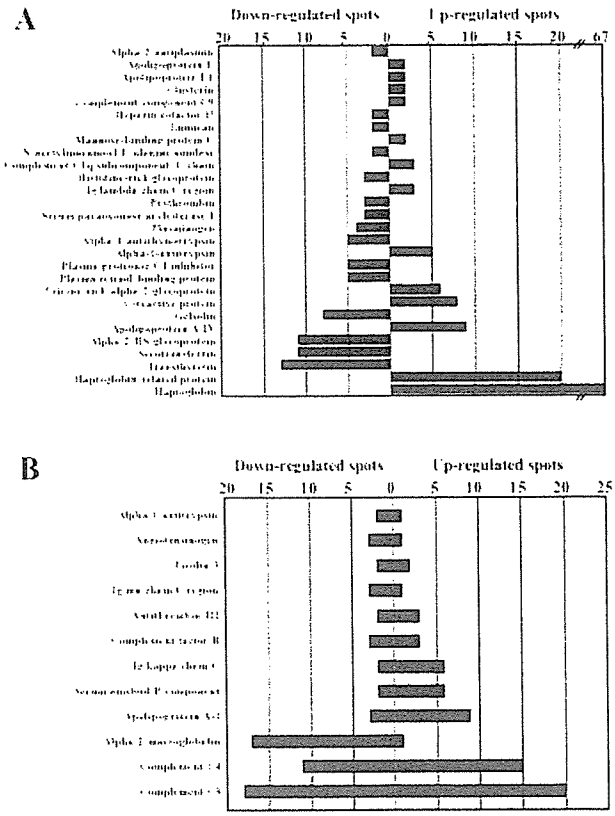

Figure 5. Analysis of spots showing aberrant expression levels in lung cancer. (A) Proteins for which the corresponding spots showed consistent differences in lung cancer; (B) proteins for which the corresponding spots showed inconsistent differences in lung cancer. X-axis representss the number of protein spots identified as those showing a significantly aberrant expression level in lung cancer. The 18 proteins identified as single spots were described in Section 3.

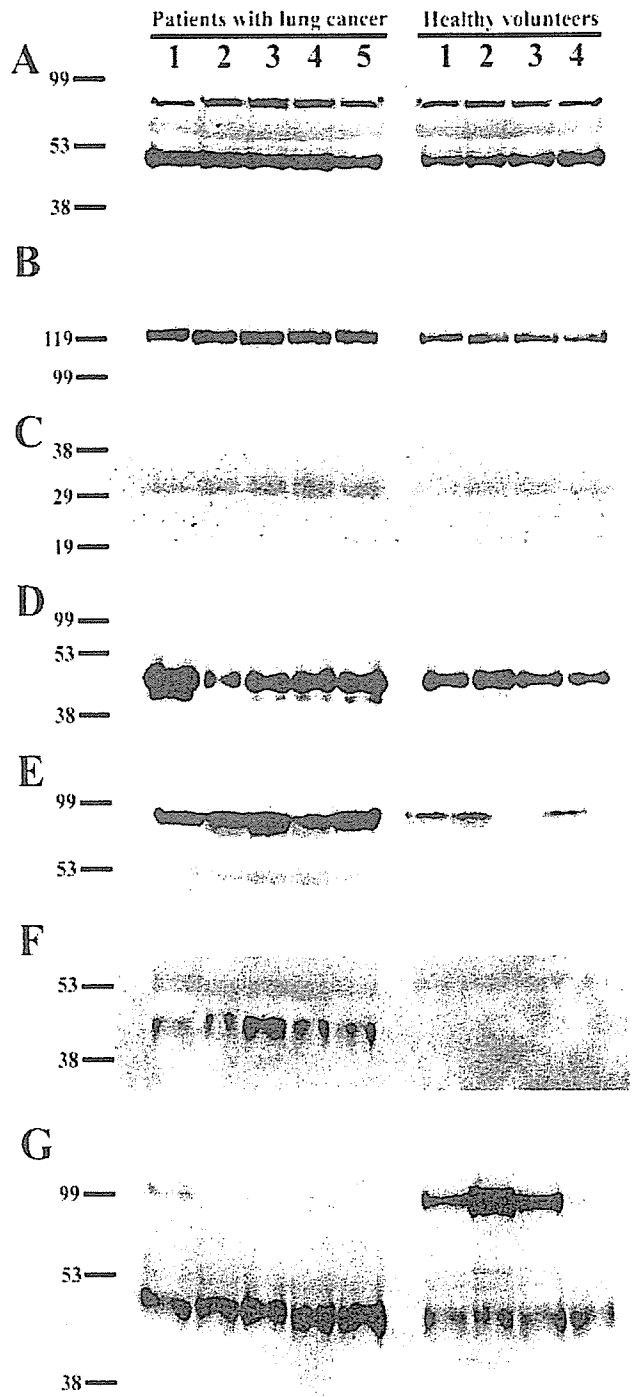

Figure 6. Validation of the differential expression of plasma proteins between individuals. Western blotting using (A) anti-leucine-rich alpha 2 glycoprotein, (B) anti-inter-alpha-trypsin inhibitor heavy chain H4, (C) anti-plasma retinol-binding protein, (D) anti-haptoglobin, (E) anti-complement component C4, (F) anti-complement component C3, and (G) anti-prothrombin antibody.

dimensional chromatography followed by 2-D-PAGE. We also studied the expression of complement component C4, whose isoforms showed differential expression. Complement component C4 consists of three subunits with molecular masses of 97, 75 and 33 kDa. The antibody we used visualized the 97-kDa band (Fig. 6E). On 2-D gel images, the expression level of the C4 spots of molecular mass approximately 97 kDa was increased in lung cancer. Western blotting identified increased expression of complement component C3, with a molecular mass of 45 kDa, though the C3 spots with this molecular weight showed both up- and down-regulated intensity in lung cancer (Fig. 6F). These results may indicate that complement component C3 is up-regulated as a whole in lung cancer and that 2-D-DIGE detects minor isoforms that are regulated in a different way from the major (abundant) isoforms. Western blotting also revealed an increased expression of prothrombin, with a molecular mass of 50 kDa, in lung cancer, and this result was inconsistent with the data obtained by 2-D-DIGE (Fig. 6G). Western blotting demonstrated increased expression of prothrombin with a molecular mass of 99 kDa in three of four healthy volunteers (Fig. 6G), and this increase was not observed by 2-D-DIGE (Fig. 4).

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

4 Discussion

We applied a highly sensitive fluorescent dye-labeling technique to multi-dimensional chromatography followed by 2-D-PAGE for lung cancer plasma proteomics. In this study, the proteins were separated in three dimensions according to their net charge, p*I* and molecular weight. Pieper *et al.* [4] reported a similar approach to the plasma proteome. In their report, human serum samples were separated into 74 fractions by immuno-affinity, anion-exchange, and size-exclusion chromatography. The fractionated samples were then subjected to 2-D-PAGE, yielding approximately 3700 protein spots. They identified by MS 325 proteins corresponding to about 1800 spots. In our study, we observed 3890 spots and identified 58 proteins corresponding to 364 spots. As we focused on proteins with aberrant expression in lung cancer and did not perform global protein identification, it is difficult to compare the performance of these two methods, although most proteins we identified were also identified in [4]. However, our use of highly sensitive fluorescent dyes has several advantages over the method used in this report, as exemplified by the following four points. First, Pieper *et al.* [4] used 20 mL of serum as a starting material, whereas we needed only 30 μL of plasma to obtain a similar number of spots. The different amount of plasma protein required for the analysis was probably a result of the sensitivity of the spot detection method, as Pieper *et al.* visualized the proteins by staining with CBB G-250, which is less sensitive than saturation dye. The amount of plasma sample available may be limited, and a method requiring less sample may be more suitable for clinical proteomics. Alternatively, our protocol may have the potential to visualize even less abundant plasma proteins by increasing the sample volume. Secondly, gel staining with CBB G-250 is time-consuming (more than 3 days in the study of Pieper *et al.*), whereas spot detection of fluorescence-labeled proteins can be completed within 30–60 min by a laser scanner. This feature is especially advantageous because several 2-D gels are required for each of the fractionated samples. Thirdly, the fluorescent dye-labeling method may allow more quantitative and reproducible protein expression profiling by running an internal control sample and generating multiplex images. Forthly, the linear dynamic range of protein expression level measured as fluorescent intensity in 2-D-DIGE is wider than that in conventional 2-D-PAGE based on a colorimetric method such as CBB G-250.

Wang *et al.* [17] reported the utility of the fluorescent dye-labeling technique in the linkage of IEF, LC and SDS-PAGE for the study of intact plasma proteins. They identified plasma proteins associated with acute graft-versus-host disease. In their protocol, the highly abundant plasma proteins were depleted with an immuno-affinity column. Low abundance proteins were then labeled with fluorescent dyes, mixed, and subjected to IEF and RP chromatography. Finally, the fractionated proteins were separated by SDS-PAGE and detected by laser scanning. To label the proteins, they used minimal dye, which labels lysine residues [9]. Although minimal dye is less sensitive than saturation dye, minimal dye is advantageous when three samples are to be compared because Cy2 is available in addition to Cy3 and Cy5 [18].

The results of this study are consistent with previous reports of aberrant expression of plasma proteins in lung cancer. Alpha-1-acid glycoprotein in plasma has been shown to be highly sensitive and specific for the detection of lung cancer [21], and its levels are correlated with treatment effects and prognosis in patients with non-small-cell lung cancer treated with docetaxel [22]. The utility of retinol-binding protein for the diagnosis of lung cancer has also been described [23]. We also found aberrant expression of tissue-leakage plasma proteins whose expression in tumor cells is correlated with lung cancer. For example, we found that the expression level of clusterin was increased in lung cancer. Clusterin is a component of membrane-coated vesicles and has a protective function in cells. Previous studies have revealed that inhibition of clusterin expression enhances the effects of paclitaxel or gemcitabine in delaying the growth of A549 tumors [24]. We also found decreased expression of gelsolin in lung cancer. Down-regulation of gelsolin in tumor cells was correlated with poor survival of patients with non-small-cell lung cancer [25]. Although the pathophysiological significance of these tissue-leakage proteins is largely obscure, their clinical application as candidate tumor markers should be considered.

In this study, we did not identify proteins showing very low abundance such as growth factors, suggesting that further development of the current protocol is required for a comprehensive understanding of the lung cancer proteome. Block *et al.* [26] reported that they observed low abundance tissue-leakage proteins as spots on 2-D images when they fractionated plasma samples with a lectin-affinity column. Thus, additional variations of multi-dimensional separation including multiple specific immuno-affinity columns will facilitate the study of lower abundance plasma proteins.

The proteins retained in the immuno-affinity column were not limited to the six target plasma proteins (Fig. 2). Albumin is known to interact with a variety of proteins [27], and as the proteins were separated as intact forms in our study, depletion of albumin may cause the removal of unexpected proteins. We also identified alpha-1-anti-trypsin, haptoglobin, and transferrin in the fractionated samples as aberrantly regulated plasma proteins in lung cancer, even though our procedure was designed to deplete these proteins at the initial step of fractionation by use of the immuno-affinity column. Together, these observations suggest that the immuno-depletion procedure requires

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

further optimization. For example, it may be possible to dissociate protein complexes prior to the immuno-depletion procedure so that albumin-bound proteins are recovered in the flow-through fraction. For better depletion of target proteins, the flow-through fraction could be applied repeatedly to the immuno-affinity column or two immuno-affinity columns could be connected in tandem. These variations are compatible with our current protocol based on 2-D-DIGE. An alternative explanation of the failure of the immuno-affinity procedure to completely deplete the target plasma proteins is that these proteins display aberrant features in lung cancer and are unable to bind to an immuno-affinity column containing antibodies raised against normal plasma proteins. These issues should be considered further when immuno-affinity procedures are applied to disease proteomics.

We found isoform-specific aberrations of plasma protein expression in lung cancer: the protein spots corresponding to 12 gene products were both up- and down-regulated. In contrast, the spots of 28 proteins showed consistent up- or down-regulation. Because the spots of haptoglobin, which showed the largest identified number of spots from a single gene product in our study, were consistently up-regulated in lung cancer, a high number of spots does not always result in inconsistent regulation, suggesting isoform-specific regulation of certain proteins. These isoform-specific aberrations may not be identified by analysis of peptide subsets from complex digests as accomplished using multidimensional protein identification technology, suggesting the importance of analysis of intact plasma proteins. Although isoform-specific alterations could be a potential source of biomarkers, it might be difficult to use such isoforms as biomarkers in a clinical setting if their monitoring requires time-consuming proteomic technology such as the combination of multidimensional chromatography and 2-D-PAGE. More conventional and high-throughput methods will be required for clinical application. Although use of a specific antibody is always an early consideration when developing a high-throughput method, we demonstrated that the results of Western blotting following SDS-PAGE did not always match those from 2-D-PAGE. This discrepancy probably arises from the facts that the antibodies used were not isoform specific and the samples were prepared differently. To solve this problem, a novel technique such as the use of DNA aptamers [28] may be required to generate molecules with high affinity for each protein variant. This issue is a generic problem using 2-D-PAGE, in which various isoforms are visualized, quantified and selected as candidate biomarkers. However, this does not mean that a 2-D-PAGE approach is impractical. We believe that 2-D-PAGE is a powerful tool for providing detailed proteomic information that cannot be obtained from other proteomic methods, and that novel technologies are required to utilize the output of 2-D-PAGE for clinical application.

We are grateful to Dr. Teruhiko Yoshida for his collaboration and to Kano Nishiyama for technical assistance. This work was supported by a grant from the Ministry of Health, Labor and Welfare and by the Program for Promotion of Fundamental Studies at the National Institute of Biomedical Innovation of Japan. Tetsuya Okano and Tatsuhiko Kakisaka are awardees of Research Resident Fellowships from the Foundation for Promotion of Cancer Research (Japan) for the 3rd Term Comprehensive 10-Year Strategy for Cancer Control.

5 References

[1] Hoffman, P. C., Mauer, A. M., Vokes, E. E., *Lancet* 2000, *355*, 479–485.

[2] Fry, W. A., Phillips, J. L., Menck, H. R., *Cancer* 1999, *86*, 1867–1876.

[3] Tarro, G., Perna, A., Esposito, C., *J. Cell Physiol.* 2005, *203*, 1–5.

[4] Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S. *et al.*, *Proteomics* 2003, *3*, 1345–1364.

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

[6] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 91–98.

[7] Unlu, M., Morgan, M. E., Minden, J. S., *Electrophoresis* 1997, *18*, 2071–2077.

[8] Chen, J. H., Chang, Y. W., Yao, C. W., Chiueh, T. S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 17039–17044.

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

[10] Fujii, K., Kondo, T., Yokoo, H., Yamada, T. *et al.*, *Proteomics* 2005, *5*, 1411–1422.

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

[12] Sitek, B., Luttges, J., Marcus, K., Kloppel, G. *et al.*, *Proteomics* 2005, *5*, 2665–2679.

[13] Greengauz-Roberts, O., Stoppler, H., Nomura, S., Yamaguchi, H. *et al.*, *Proteomics* 2005, *5*, 1746–1757.

[14] Seike, M., Kondo, T., Fujii, K., Okano, T. *et al.*, *Proteomics* 2005, *5*, 2939–2948.

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

[16] Pieper, R., Su, Q., Gatlin, C. L., Huang, S. T. *et al.*, *Proteomics* 2003, *3*, 422–432.

[17] Wang, H., Clouthier, S. G., Galchev, V., Misek, D. E. *et al.*, *Mol. Cell. Proteomics* 2005, *4*, 618–625.

[18] Misek, D. E., Kuick, R., Wang, H., Galchev, V. *et al.*, *Proteomics* 2005, *5*, 3343–3352.

[19] Bjorhall, K., Miliotis, T., Davidsson, P., *Proteomics* 2005, *5*, 307–317.

[20] Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C. *et al.*, *Proteomics* 2005, *5*, 2385–2395.

[21] Ganz, P. A., Baras, M., Ma, P. Y., Elashoff, R. M., *Cancer Res.* 1984, *44*, 5415–5421.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[22] Bruno, R., Olivares, R., Berille, J., Chaikin, P. *et al.*, *Clin. Cancer Res.* 2003, *9*, 1077–1082.

[23] Edes, T. E., McDonald, P. S., *Cancer Detect. Prev.* 1991, *15*, 341–344.

[24] July, L. V., Beraldi, E., So, A., Fazli, L. *et al.*, *Mol. Cancer Ther.* 2004, *3*, 223–232.

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

[26] Block, T. M., Comunale, M. A., Lowman, M., Steel, L. F. *et al.*, *Proc. Natl. Acad. Sci. USA* 2005, *102*, 779–784.

[27] Mehta, A. I., Ross, S., Lowenthal, M. S., Fusaro, V. *et al.*, *Dis. Markers* 2003, *19*, 1–10.

[28] Fang, X., Sen, A., Vicens, M., Tan, W., *Chembiochem.* 2003, *4*, 829–834.

[29] Travis, W., Colby, T., Corrn, B., Shimosato, Y., Brambilla, E., *Histological Typing of Tumours of Lung and Pleura: World Health Organization International Classification of Tumors*, Springer Verlag, New York 1999.

[30] Sobin L. H., W. C., *UICC: TNM Classification of Malignant Tumours*, John Wiley & Sons, New York 1997, pp. 91–100.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

 DOI 10.1002/pmic.200600196 *Proteomics* 2006, *6*, 4402–4409

RESEARCH ARTICLE

Proteomic signatures corresponding to histological classification and grading of soft-tissue sarcomas

Yoshiyuki Suehara[1, 2], Tadashi Kondo[1], Kiyonaga Fujii[3], Tadashi Hasegawa[4]**, Akira Kawai[5], Kunihiko Seki[4], Yasuo Beppu[5], Toshihide Nishimura[3], Hisashi Kurosawa[2] and Setsuo Hirohashi[1]*

[1] Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan
[2] Department of Orthopedic Surgery, Juntendo University School of Medicine, Tokyo, Japan
[3] Clinical Proteome Center, Tokyo Medical University, Tokyo, Japan
[4] Pathology Division, National Cancer Center Research Institute, Tokyo, Japan
[5] Orthopedic Surgery Division, National Cancer Center Hospital, Tokyo, Japan

We performed a global protein expression study on soft-tissue sarcoma in order to develop novel diagnostic biomarkers and allow molecular classification. 2-D difference gel electrophoresis was used to generate the global protein expression profiles of 80 soft-tissue sarcoma samples with seven different histological backgrounds. We found that 67 protein spots distinguished the subtypes of soft-tissue sarcoma. Hierarchical clustering with these 67 protein spots resulted in the grouping of all 80 sarcoma samples corresponding to the histological classification. We found that the expression pattern of tropomyosin isoforms was different in conventional and pleomorphic leiomyosarcomas. We also identified five proteins, including alpha-1-antitrypsin, alpha-actinin 1, HSP27, and elongation factor 2, that could differentiate between malignant fibrous histiocytomas and leiomyosarcomas in grade III into low-risk and high-risk groups, which differed significantly with respect to survival. These results establish proteomics as a powerful tool to develop novel biomarkers for diagnosis and molecular classification of soft-tissue sarcomas. Identification of proteins associated with survival in grade III sarcoma will allow delineation of a high-risk group that may benefit from adjuvant therapy and the exclusion of low-risk patients in whom additional therapies are unlikely to exhibit clinical benefit.

Received: March 20, 2006
Accepted: May 11, 2006

Keywords:
2D-DIGE / Histological classification / Histological grading / Soft-tissue sarcoma / Toropomyosin

1 Introduction

Soft-tissue sarcoma is a rare malignant tumor. As the therapeutic response often depends on the histological subclass [1], accurate histological diagnosis is crucial to predict the biological behavior of the tumor and establish a therapeutic strategy. The subtypes of soft-tissue sarcoma are diagnosed by morphological findings, immunophenotyping, and karyotyping assisted by molecular analysis of specific-gene rearrangements. However, the wide range of histological appearance of soft-tissue tumors occasionally results in low validity and reproducibility of histological diagnosis and grading [2]. Molecular aberrations are not always consistent

Correspondence: Dr. Tadashi Kondo, Proteome Bioinformatics 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: +81-3-357-5298

Abbreviations: GIST, gastrointestinal stromal tumors; **MFH**, malignant fibrous histiocytomas

* Present address: Proteome Bioinformatics Project, National Cancer Center Research Institute, Tokyo, Japan
** Present address: Department of Clinical Pathology, Sapporo Medical University School of Medicine, Sapporo, Japan

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

in the same tumor type, and not all tumors have typical gene rearrangements [3]. Thus, novel approaches are needed to overcome some of the limitations of the present studies.

Recent development in cancer biology has been the study of global protein expression, an approach known as proteomics. DNA microarray technologies have also been applied at a genome-wide scale to the development of biomarkers for soft-tissue tumors, and several candidates have been identified [4–6]. However, RNA analysis alone cannot predict the functional features of the products such as PTMs, subcellular location, and association with ligands, as well as the rate of change with time of such properties, and these factors play important roles in the malignant behavior of tumor cells. Further, many lines of evidence have indicated discordance between mRNA expression and protein expression [7–9]. These difficulties underline the potential advantages of a proteomic rather than a transcriptomic approach. In addition, as the findings at the protein level will be more applicable to the subsequent immunohistochemical studies, the results can be validated using numerous paraffin-embedded tissues with clinical information by specific antibody. Proteomic technologies have been used to develop novel molecular subclassifications and diagnostic biomarkers for lung cancer [10, 11], ovarian tumors [12], breast cancer [13, 14], leukemia [15], and prostate cancer [9].

We performed a proteomic study of a panel of 80 soft-tissue sarcomas representing different histological subtypes and grading. We identified protein clusters that define histological subclasses and that differentiate low-risk from high-risk grade III groups, which differed significantly with respect to survival, of malignant fibrous histiocytomas (MFH) and leiomyosarcomas. This is the first systematic large-scale proteomic study to identify proteins associated with the histological subtypes and grading of soft-tissue sarcomas.

2 Materials and methods

The procedures that are only briefly described here are described in detail in Supplementary Text.

2.1 Patient population and protein extraction

We used tissue specimens from 80 soft-tissue sarcomas that had been surgically resected at the National Cancer Center Hospital in the period between 1996 and 2004. This project was approved by the Institutional Review Board of the National Cancer Center. The tumors comprised seven histological subgroups: 6 clear cell sarcomas, 10 myxoid liposarcomas, 9 gastrointestinal stromal tumors (GIST), 5 malignant peripheral nerve sheath tumors (MPNST), 12 synovial sarcomas, 28 MFH (10 myxoid and 18 pleomorphic), and 10 leiomyosarcomas (7 conventional and 3 pleomorphic) (Supplementary Fig. 1 and Supplementary Table 1).

The frozen samples were crushed to powder with a CryoPress (Microtech Nichion, Chiba, Japan) under cooling with liquid nitrogen. The frozen powder was then homogenized with urea lysis buffer (6 M urea, 2 M thiourea, 3% CHAPS, 1% triton X-100). After centrifugation at 15 000 rpm for 30 min, the supernatant was used as the source of cellular proteins for protein expression studies.

2.2 Protein expression profiling

Protein expression profiles were obtained by 2-D difference gel electrophoresis (2-D-DIGE). The internal control consisted of a mixture of all the protein samples used in this study. The internal control sample and the individual sample were labeled with the CyDye DIGE Fluor saturation dyes Cy3 and Cy5 (GE Healthcare Amersham Biosciences), respectively. The Cy3-labeled control sample and the Cy5-labeled individual sample were mixed and coseparated by 2-D PAGE. After electrophoresis, the gels were scanned at appropriate wavelengths for Cy3 and Cy5 to generate the images of the reference sample and the individual sample, respectively. We ran triplicate gels for each sample, and in total 240 gels were run and 480 images were created for the 80 samples. The ratio between Cy5 intensity and Cy3 intensity was calculated for all spots in each gel with DeCyder software (GE Healthcare Amersham Biosciences) to obtain the standardized spot intensities. As the Cy3 image represents the internal standard sample, this standardization procedure eliminates gel-to-gel differences. The standardized spot intensities were logarithmically transformed, averaged for triplicate gels and analyzed with the data-mining package Impressionist (GeneData, Basel, Switzerland).

To assess the reproducibility of the proteomic data in our analyses, we compared the protein profiles obtained from three independent separations of sample L-10 (Fig. 1A, Supplementary Table 1). Scatter plot analysis revealed that the standardized intensity of more than 99% of the spots was scattered within a 2.0-fold difference. We used the averaged spot intensity from the triplicate gels for further data analysis.

2.3 Data analysis

We identified the proteins that were informative for classification by using a support vector machine algorithm and a leave-one-out crossvalidation [16, 17]. First, we selected the spots that appeared in more than 75% of Cy3 images. Second, the multiclass problem was solved by a series of seven one-*versus*-all comparisons (*e.g.*, "clear cell sarcoma" *vs.* "not clear cell sarcoma") and the spots with expression levels significantly different between two groups of different tumor types were selected on the basis of the Wilcoxon method. The pairwise similarity of expression profiles was examined for the spots between all samples to exclude irrelevant samples. Third, we developed classifiers by using the support vector

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

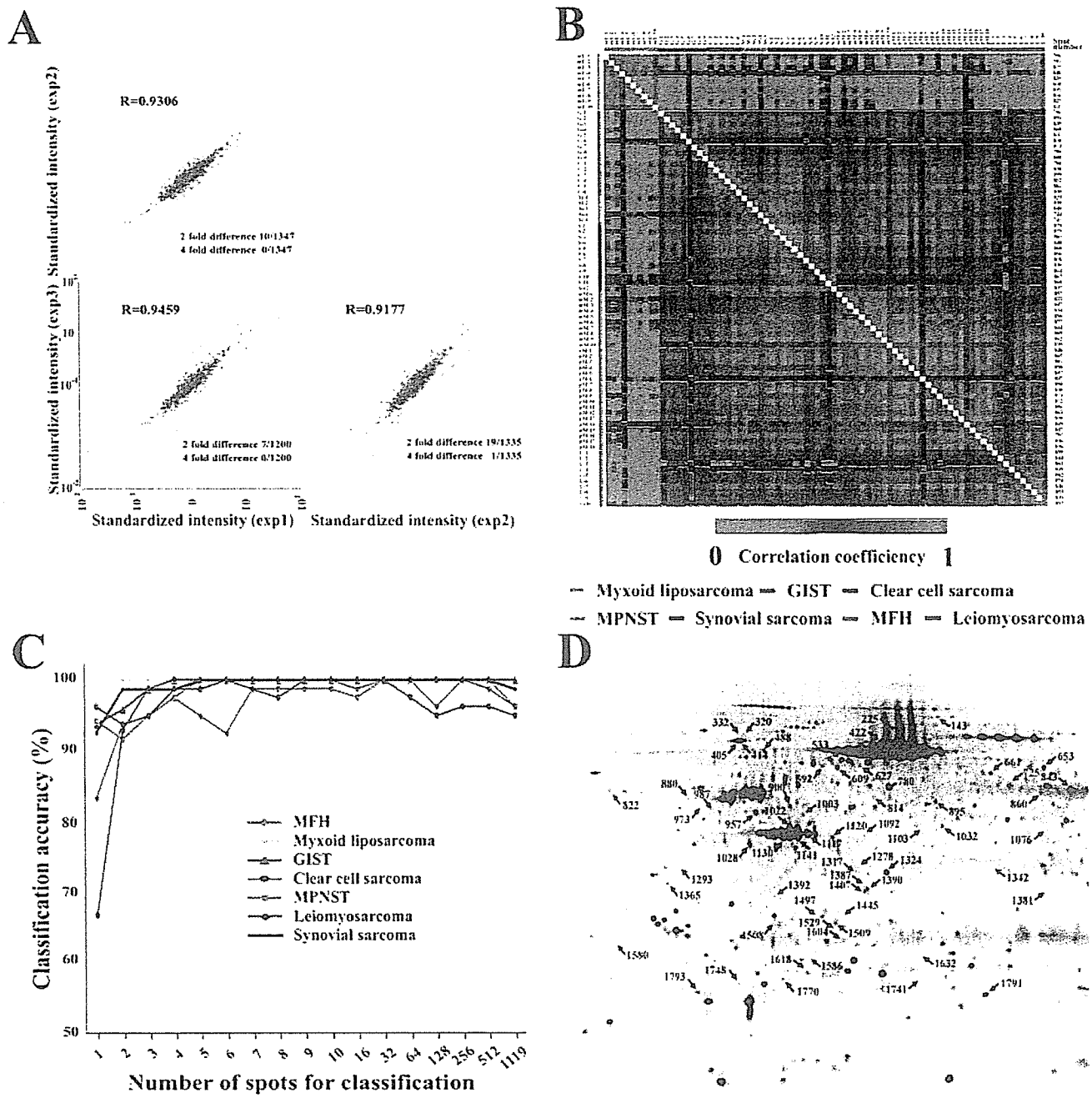

Figure 1. Spot selection for histological classification. (A) Evaluation of reproducibility of 2-D-DIGE by scattergram. (B) Correlation matrix of the 80 sarcomas on the basis of the expression profiles of the selected spots. The sarcomas were ordered by histological class and color-coded. The patient ID corresponds to that in Supplementary Table 1. The number of spots examined is listed on the right (Supplementary Fig. 3A). (C) Mean leave-one-out crossvalidation accuracy was plotted as a function of the number of spots used by each classifier. (D) Localization of the selected 67 spots on the image of the 2-D PAGE. Protein identification is summarized in Supplementary Table 2.

machine and the accuracy of the classifiers was evaluated by a leave-one-out crossvalidation. The process was repeated for each sample, and the cumulative crossvalidation error rate was calculated. Fourth, we ranked all protein spots according to their relative contribution to the classification on the basis of support vector machine weight. By calculating the cumulative error rate while eliminating proteins in the order of increasing rank, we examined the relationship between the number of proteins and classification accuracy. Fifth, the classification performance of a candidate classifier was evaluated by multivariate analysis including principal component analysis and hierarchical clustering.

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

2.4 Protein identification by MS and Western blotting

Proteins corresponding to the spots of interest were identified by MS. Cy5-labeled proteins separated by 2-D PAGE were recovered in gel plugs and digested with modified trypsin (Promega, Madison, WI, USA). The trypsin digests were subjected to LC coupled with MS/MS on a machine equipped with a nanoelectrospray ion source. A search of the mass of the peptide ion peaks against the Swiss-Prot database was performed with MASCOT software. Protein identification and differential expression were confirmed by Western blotting using specific antibodies against tropomyosin isoforms.

2.5 Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

3 Results

Protein expression profiling was performed for the 80 soft-tissue sarcomas by 2-D-DIGE. We selected protein spots that appeared in at least 75% of the 240 images of the Cy3-labeled internal control sample and that showed different spot intensities between one histological subgroup and the other tumors (Wilcoxon test, $p < 0.01$). Although potentially resulting in the loss of information, this trimming process decreased the possibility that the classifier would be significantly influenced by irrelevant expression data. We observed the pairwise similarity of the protein expression profiles of the 80 samples. A correlation matrix showed that the myxoid liposarcoma samples had a distinct protein expression profile that differed from that of the rest of the tumor samples, which shared a similar overall protein expression profile (Fig. 1B and Supplementary Fig. 3A). However, hierarchical clustering using all protein spots did not result in the classification of soft-tissue sarcomas according to their histology (Supplementary Fig. 2). At this stage we tried to omit irrelevant samples that could hinder the proper classification. However, as we could not identify obviously anomalous samples (Fig. 1B and Supplementary Fig. 3A), we used all 80 samples in the subsequent study. We selected protein sets of which the expression was associated with the existing histological classification by using a support vector machine. The accuracy, plotted as a function of spot number, was constant until the number of spots decreased to less than five, showing that accurate classification did not require all protein spots (Fig. 1C). We decided to study the top ten ranked protein spots for each classification and evaluate their classification performance. As three spots overlapped among the 70 spots, 67 spots were selected (Fig. 1D and Supplementary Fig. 4). The results of classification, shown as the distance of a sample from the hyper-plane, demonstrated that all samples except one were classified correctly with the trained support vector machine (Supplementary Fig. 5). Mass spectrometric study identified the proteins corresponding to these 67 spots, revealing that the 67 spots were products of 67 different genes (Supplementary Table 2). The pairwise correlation matrix of the 80 samples showed that sarcomas in the same histological group had a common expression profile of the 67 proteins (Fig. 2A and Supplementary Fig. 3B). To evaluate the classification performance of the 67 protein spots, we performed principal component analysis of the 80 samples with the 67 spots (Fig. 2B). Liposarcoma and leiomyosarcoma showed a distinct protein expression profile and the other tumors were also grouped with those of the same histological background. In contrast, the profiles of three pleomorphic leiomyosarcoma samples (L-1, L-5, and L-7) were separated from that of conventional leiomyosarcoma, and close to that of MFH.

To examine correlations among tumors with different histological backgrounds, we performed hierarchical clustering. Hierarchical clustering with the 67 spots yielded two clusters of tumors (Fig. 2C and Supplementary Fig. 6). All samples of leiomyosarcoma and MFH fell within Cluster 1. Cluster 2 contained two major groups – one consisting of with myxoid liposarcomas, and the other of clear cell sarcomas, synovial sarcomas, MPSNTs, and GISTs. In Cluster 1, pleomorphic leiomyosarcoma (L-1, L-5, and L-7) and conventional leiomyosarcoma (L-2, 3, 4, 6, 8, 9, and 10) were divided into two branches; all three pleomorphic leimyosarcomas were separated from the conventional leiomyosarcomas and associated with the MFH samples, consistent with the results of principal component analysis (Fig. 2B). We identified the proteins responsible for leiomyosarcoma subtypes. Supervised classification of pleomorphic and conventional leiomyosarcomas identified ten proteins as the most different features of the leiomyosarcoma subclasses (the process of spot selection is demonstrated in Supplementary Fig. 7 and the results of protein identification are summarized in Supplementary Table 2). Hierarchical classification using the selected ten spots divided leiomyosarcoma samples into pleomorphic and conventional types, validating the results of the support vector machine classification (Fig. 3A). Use of specific antibodies confirmed the differential expression of tropomyosin isoforms between these two groups of leiomyosarcoma, also showing that relative protein amounts measured by 2-D-DIGE and Western blotting were highly correlated (Fig. 4 and Supplementary Fig. 8). Pleomorphic leiomyosarcoma and MFH shared the similar expression pattern of tropomyosine isoforms (Supplementary Fig. 9).

Histological grading has been correlated with patient outcome [18]. Using protein expression profiles of 10 leiomyosarcomas and 28 MFHs, we investigated the proteins that could distinguish between these tumors of grades I/II and those of grade III. Supervised classification identified a set of five proteins by which the crossvalidation error rate was minimal for these two tumor groups (the process of spot

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim