またその測定できる質量領域は40kDa以下の低分子量タンパク質・ペプチドに限定されている。さらに，抽出されたペプチドのアミノ酸同定の難易性や機械間誤差の問題など克服すべき課題も存在する。これらの問題点が克服できれば，多施設で使用できる全国規模でのデータベース作成が可能であると思われる。臨床応用に向けたさらなる基盤整備が望まれる。

文　献

1）Petricoin EF, Ardekani AM, Hitt BA, et al : Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359 : 572-577, 2002
2）Honda K, Hayashida Y, et al : Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res (in press)
3）Hara T, Honda K, Ono M, et al : Identification of two serum biomarkers of renal cell carcinoma by surface-enhanced laser desorption/ionization mass spectrometry. J Urol **174** : 1213-1217, 2005
4）Hayashida Y, Honda K, Osaka Y, et al : Possible prediction of chemoradiosensitivity of esophageal cancer by serum protein profiling. Clin Cancer Res (in press)

本田　一文
平成３年 日本大学松戸歯学部卒業
現在，国立がんセンター研究所化学療法部・腫瘍プロテオミクスプロジェクト活性物質選別研究室室長
専門分野：腫瘍病理学，口腔外科
E-mail : khonda@gan2.res.ncc.go.jp

0022-5347/05/1744-1213/0
THE JOURNAL OF UROLOGY®

Vol. 174, 1213–1217, October 2005
Printed in U.S.A.
DOI: 10.1097/01.ju.0000173915.83164.87

IDENTIFICATION OF 2 SERUM BIOMARKERS OF RENAL CELL CARCINOMA BY SURFACE ENHANCED LASER DESORPTION/IONIZATION MASS SPECTROMETRY

TOMOHIKO HARA,* KAZUFUMI HONDA, MASAYA ONO, KATSUSUKE NAITO, SETSUO HIROHASHI AND TESSHI YAMADA†

From the Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute (TH, KH, MO, SH, TY), Tokyo, and the Department of Urology, Yamaguchi University School of Medicine (TH, KN), Ube, Japan

ABSTRACT

Purpose: Surface enhanced laser desorption/ionization mass spectrometry can generate robust information from a small amount of clinical samples such as serum and plasma. In this study we identified novel diagnostic biomarkers of renal cell carcinoma (RCC) by large-scale serum protein profiling using surface enhanced laser desorption/ionization mass spectrometry.

Materials and Methods: Proteomic spectra were generated by a time of flight mass spectrometer from a set of training samples (21 patients with RCC and 24 healthy volunteers) and another set of validation samples (19 patients with RCC, 20 healthy volunteers and 5 patients with pyelonephritis). Information on the peaks (intensity and m/z) was extracted from the mass spectra using newly developed algorithms, and the Mann-Whitney's U test and linear support vector machine were used to identify the peaks distinguishing RCC samples from the controls.

Results: Two peaks with molecular masses of 4,151 and 8,968 m/z were selected as significantly more prominent in RCC samples ($p < 0.01$) among the 3,539 peaks in the range of 3,000 to 30,000 m/z obtained from the training samples. Simultaneous recognition of these 2 biomarkers was shown to have a sensitivity of 89.5% for the diagnosis of RCC and an overall specificity of 80.0% (95% [19 of 20] of healthy volunteers and 20% [1 of 5] of patients with pyelonephritis) in the blinded validation samples, and to allow detection of RCC in stage I (UICC) in 88.9% (16 of 18) of the cases.

Conclusions: We identified 2 serum biomarkers potentially useful for the early diagnosis of RCC. This finding warrants a further large-scale multi-institutional analysis for clinical evaluation of the diagnostic significance of these biomarkers.

KEY WORDS: serum, proteomics; carcinoma, renal cell; protein array analysis

Malignant tumors of the kidney and renal pelvis comprise more than 2% of newly diagnosed cancers in Western countries and Japan, and account for approximately 11,900 deaths a year in the United States.[1] Renal cell carcinoma (RCC) is a major histological subtype of malignancies of the kidney. Since RCC can become quite large before causing any symptoms, the appearance of symptoms such as gross hematuria and abdominal pain or discomfort often indicates an advanced stage of the disease. Surgery is an effective method for treating localized RCC, but the treatment of metastatic disease still remains to be established. Consequently 5-year survival has been reported to be 91% to 100%, 74% to 96%, 59% to 70% and 16% to 32%, respectively, in patients with RCC stages I, II, III and IV (according to the revised TNM classification [1997, UICC]).[2]

The introduction of computerized tomography and ultrasonography in routine medical practice has dramatically increased the accidental discovery of early RCC (from approximately 10% in the early 1970s to 61% in 1998).[2] However, mass screening by computerized tomography, magnetic resonance imaging or positron emission tomography is not cost-effective because of the relatively low incidence of RCC, and small RCC nodules can easily be overlooked by a not-so-thorough ultrasound examination. Furthermore, diagnostic radiography has been estimated to contribute to the cumulative risk of cancer.[3] All of these indicate the necessity for the development of a new alternative safe modality for efficient detection of early RCC.

Recent advances in mass spectrometry (MS) based proteomic technologies, of which surface enhanced laser desorption/ionization mass spectrometry (SELDI-MS) is the key technology, may revolutionize medical diagnosis and cancer screening. SELDI chips can selectively tether, concentrate and purify various biomolecules such as serum proteins on a chemical surface in a highly reproducible way. The molecular mass and quantity of the biomolecules trapped on the chips can be measured with high sensitivity and resolution by laser desorption/ionization MS. Serum proteomics by SELDI-MS has been reported to be highly successful for the diagnosis of malignancies of various organs, including those of the ovary,

Submitted for publication November 26, 2004.
Study received Yamaguchi University School of Medicine (Ube, Japan) and the National Cancer Center (Tokyo, Japan) ethical committee approval.
Supported by grants from the Ministry of Health, Labor and Welfare of Japan, and by the Program for the Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research of Japan.
Nothing to disclose.
* Recipient of the Research Resident Fellowship from the Foundation for the Promotion of Cancer Research.
† Correspondence: Chemotherapy Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan (telephone: 81-3-3542-2511 ext. 4270; FAX: 81-3-3547-6045; e-mail: tyamada@ncc.go.jp).

***Editor's Note*: This article is the first of 5 published in this issue for which category 1 CME credits can be earned. Instructions for obtaining credits are given with the questions on pages 1508 and 1509.**

For another article on a related topic see page 1460.

prostate and pancreas.[4–7] In particular it has been reported that early ovarian cancer can be detected by pattern recognition of multiple serum proteins.[4]

This study identifies diagnostic biomarkers of RCC by comprehensive serum protein profiling using SELDI-MS. We increased the chances of identification of new biomarkers by increasing the number of detectable protein peaks using fractionation of crude serum samples, using SELDI arrays with different chemical surface characteristics and conditions, and introducing newly developed peak detection algorithms. We report on the identification of a combination of 2 biomarkers that allows discrimination of RCC from healthy control samples with high specificity, and detection of early RCC with high sensitivity.

MATERIALS AND METHODS

Serum samples. Blood samples were collected from 89 individuals (40 untreated patients with RCC, 44 healthy volunteers and 5 untreated patients with pyelonephritis) seen in the Urology Department of Yamaguchi University Hospital (Ube, Japan) between January 1992 and December 2003. Informed consent was obtained from all of the donors. The samples were collected in 9 ml glass tubes and allowed to coagulate at 4C. Serum (1 to 2 ml) was separated under sterile conditions and cryopreserved at -30C until the analyses. The study was reviewed and approved by the ethical committees of Yamaguchi University School of Medicine (Ube, Japan) and the National Cancer Center (Tokyo, Japan).

SELDI-MS. To increase the number of detectable protein peaks, the serum samples were fractionated by step-wise anion exchange chromatography (pH 9 plus flow through, pH 7, pH 5, pH 4, pH 3 and organic wash, fig. 1) using QHyper DF resin (Ciphergen Biosystems, Inc., Fremont, California) and Biomek® 2000 laboratory automation workstation robots (Beckman Coulter, Fullerton, California) as previously described.[7] Each sample was randomly assigned in duplicate to spots among 48 spots of 6 allied ProteinChip arrays (Ciphergen). Three types of ProteinChip arrays with different surface chemistries, namely immobilized metal affinity capture coupled with copper (IMAC-$Cu2^+$), weak hydrophobic (H50) and cationic (CM10) arrays, were used. The CM10 arrays were used under either low stringent (pH 4) or high stringent (pH 7) conditions as instructed by the supplier.

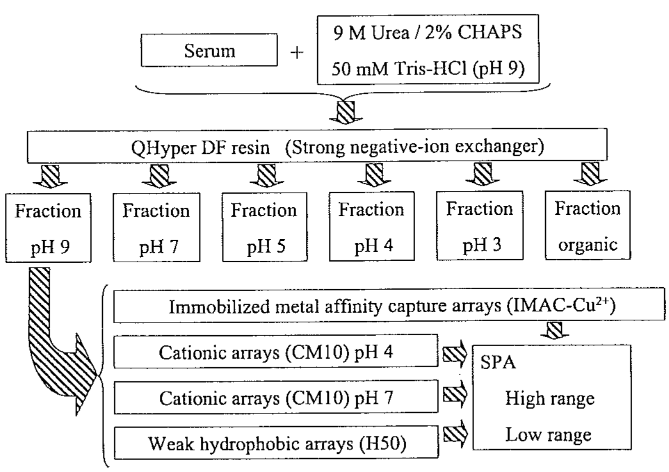

FIG. 1. Schema of serum protein fractionation and detection. Serum samples were denatured with 9 M Urea, 2% CHAPS, 50 mM Tris-HCl (pH 9) and fractionated by step-wise anion exchange chromatography. Six fractions were applied to ProteinChip arrays with 4 different surface characteristics (IMAC-Cu^{2+}, CM10 at pH 4, CM10 at pH 7 and H50). Bound proteins were ionized and detected using 2 different settings optimal for low (3,000 to 12,000 m/z) and high mass ranges (10,000 to 30,000 m/z). SPA, sinapinic acid.

Soon after they were air-dried, matrix (50% sinapinic acid in 50% acetonitrile/0.1% trifluoroacetic acid) was applied to the chips (fig. 1). All arrays were read on a Protein Biological System IIc ProteinChip reader (Ciphergen). Two types of time of flight-MS settings were applied for each spot. The low or high mass range was focused on the 3,000 to 12,000 or 10,000 to 30,000 m/z range, respectively. The lag time was set automatically for fitting the center of the optimal ranges. The laser intensity level was individually adjusted, set at 186 to 204 in the low range and 190 to 216 in the high range. Mass accuracy was calibrated on the day of the measurements using the All-in-1 Peptide molecular mass standard (Ciphergen). The day-to-day mass variation was determined to be less than 0.3%.

The coefficient of variance (CV) for the peak intensity was calculated using 4 to 22 randomly chosen peaks obtained from each fraction (Fr) of the plasma mixture prepared from the healthy volunteers. The mean CV was 0.26 in the higher mass range and 0.42 in the lower mass range (CM10 at pH 4).

Peak detection and quantification. The raw mass spectra were processed in several steps. The value of m/z was calculated from the turn of intensity data in conformity with the calibration formula of Ciphergen Biosystems Inc. The smoothing of the intensity data was performed using the Gaussian window function. Baseline and noise levels were calculated for every fixed interval of m/z[8] and the baseline level was subtracted from the intensity data. All apexes of the intensity data were defined as peaks. When the valley between 2 adjoining peaks was small, the lower peak was included in the higher peak. Peak alignment was carried out so that the number of samples which had peaks with a high signal-to-noise ratio within the permissible limits of error would be maximized. For samples with no peaks, a peak with a low signal-to-noise ratio was added during peak alignment. Peak intensity was normalized so that the total intensity after baseline subtraction was the same for all the samples. Intensities of the selected peaks were averaged among duplicates and peak intensities were compared.

Statistical analysis. Statistical differences were evaluated by Student's t test. Discriminatory peaks were selected by the nonparametric Mann-Whitney U test. A separating borderline for each peak that distinguished cancer samples from controls in the training set was calculated using the linear support vector machine (SVM) algorithm.[9]

RESULTS

Identification of 2 serum biomarkers that were significantly increased in RCC. We first analyzed 45 serum samples (21 RCC, and 24 sex and age matched healthy controls) as the training set (table 1). A total of 3,539 independent peaks were recognized in the range of 3,000 to 30,000 m/z (table 2). The intensity of some peaks was found to be significantly associated with the storage period of the samples and these peaks were excluded from further analysis (data not shown). The intensity of 2 peaks was significantly increased in RCC samples compared with that in the controls ($p < 0.01$). The first peak was at 4,151 ± 8 m/z ($p=0.0085$). Peaks in this range were observed in the spectra of the IMAC-Cu^{2+} (Fr 5), CM10 at pH 4 (Fr 4) as well as CM10 at pH 7 (Fr 4) arrays, and behaved similarly in all. The second peak was at 8,968 ± 16 m/z ($p=0.0065$). Peaks in this range were observed in the CM10 at pH 4 (Fr 4) and CM10 at pH 7 (Fr 4) arrays. Figure 2 depicts the representative 4,151 and 8,968 m/z peaks appearing in the CM10 arrays at pH 4 from 2 RCC and 2 control samples.

The receiver operator curve (ROC) areas of the 4,151 ± 8 m/z peaks in Fr 5 (4,144 m/z, IMAC-Cu^{2+}), Fr 4 (4,151 m/z, CM10 at pH 4) and Fr 4 (4,150 m/z, CM10 at pH 7) were 0.609, 0.775 and 0.635, respectively. The ROC areas of the 8,968 ± 16 m/z peaks in Fr 4 (8,968 m/z, CM10 at pH 4) and

TABLE 1. *Clinical characteristics and sex/age distribution of the 89 individuals whose samples were used for SELDI-MS analyses*

Characteristic	Training Set		Validation Set		
	RCC	Healthy	RCC	Healthy	PN
Case No.	21	24	19	20	5
No. male (%)	13 (62)	14 (58)	13 (68)	14 (70)	2 (40)
Mean age (± SD)*	65.14 (± 12.7)	56.29 (± 11.3)	60.37 (± 15.0)	50.80 (± 13.3)	57.6 (± 31.7)
Median age	69	57	65	53.5	75
Age range	32–79	40–77	36–84	38–85	17–85
No. UICC stage:					
I	11		7		
II	0		3		
III	3		3		
IV	5		6		
Unknown	2		0		

* The 2-tailed p value for RCC vs healthy controls in the training set was 0.64 (Student's t test). The p values in the validation set for RCC vs healthy controls, RCC vs PN and healthy controls vs PN were 0.61, 0.93 and 0.89, respectively.

TABLE 2. *Numbers of peaks detected in each fraction and array condition*

	CM10 at pH 4	CM10 at pH 7	H50	IMAC-Cu^{2+}
Fr 1	114	168	166	156
Fr 2	96	85	109	155
Fr 3	166	178	155	163
Fr 4	143	142	182	167
Fr 5	45	113	141	169
Fr 6	108	180	181	257
Totals	672	866	934	1,067

Fr 4 (8,968 m/z, CM10 at pH 7) were 0.911 and 0.926, respectively. The CV of the 4,151 m/z peak was 0.1612, and that of the 8,968 m/z peak was 0.3073 in the 45 samples of the training set (CM10 at pH 4, Fr 4).

Combination of the 2 serum biomarkers discriminates RCC from controls. Borderlines distinguishing cancer samples from controls were calculated using the linear SVM algorithm (fig. 3). The application of these borderlines to the 2-dimensional plotting of the intensity of the 4,151 and 8,968 m/z peaks (fig. 4) in Fr 4 (CM10 at pH 4) was found to allow correct diagnosis in 41 of the 45 samples (91.1%) from the training set (90.5% sensitivity and 91.7% specificity, fig. 4, *A*).

Because the 4,151 and 8,968 m/z peaks appeared in Fr 4 (CM10 at pH 4), we applied these hybridization conditions exclusively to the validation set. The same borderlines of the 4,151 and 8,968 m/z peaks were applied to the validation set consisting of 39 cases (19 patients with RCC and 20 sex/age matched healthy controls, table 1). Again, in comparison of patients with RCC and healthy controls, 36 of 39 (92.3%) could be diagnosed correctly (89.5% sensitivity and 95.0% specificity, fig. 4, *B*). The ROC areas of the 4,151 and 8,968 m/z peaks in these 84 RCC and healthy cases were 0.848 and 0.660, respectively. Of the 5 cases 4 (80.0%) with pyelonephritis (PN) in the validation set were classified into the RCC category.

Simultaneous recognition of the 2 biomarkers can detect early RCC. Among the total of 40 RCC cases (training and validation sets), 18, 3, 6 and 11 cases were classified as TNM stage I, II, III and IV, respectively (table 3). The detection rate of stage I RCC was 27.8% and was 61.1% when the 2 aforementioned peaks were recognized separately. However, the simultaneous recognition of the 2 biomarkers allowed detection of RCC in stage I in 88.9% (16 of 18) of the cases.

All 40 cases of RCC were considered to represent sporadic cases in terms of medical history. The histological diagnoses of the 40 cases were classified according to World Health Organization histological classification[10] as clear cell carcinoma, granular cell carcinoma, spindle cell carcinoma, cyst associated renal cell carcinoma or mixed type (table 4). The histological subtype of all 4 misclassified RCC cases was clear cell carcinoma. There were no significant differences in detection rates among the histological subtypes of RCC.

DISCUSSION

The protein composition of the serum is dominated by a handful of proteins such as albumin, transferrin, haptoglobin, immunoglobulins and lipoproteins.[11–13] In addition, the serum proteins are present across an extraordinarily dynamic range of concentrations.[11] This large dynamic range exceeds the analytical capabilities of conventional proteomic methods, making the detection of smaller amounts of serum proteins extremely challenging. Therefore, reduction of sample complexity was the first step in the current study of serum proteomics (fig. 1). We also developed a new algorithm for peak detection and were able to recognize more than 3,000 peaks from only 20 μl of a serum sample (table 2).

In addition to high sensitivity, clinical proteomics also requires high reproducibility. White et al suggested that the relatively low CV value of the spectra obtained by SELDI-MS may make quantitative analysis difficult.[14] We randomized cases and controls, and spotted the samples onto the ProteinChip arrays in duplicate using a robot system to minimize human errors. We performed careful quality control to ensure constant array hybridization and detection conditions. As a result we achieved CV values in our SELDI-MS system of 0.26 to 0.42, although even values in the range of 0.5 to 0.6 are considered acceptable.[14]

SELDI-MS can generate robust information from small amounts of clinical samples such as serum and plasma. Bioinformatics has a critical role in the analysis of SELDI-MS data.[14] SVM is a kind of machine learning that can perform binary classification such that the distance (the margin) from the closest training examples (the support vector) to the separating hyperplane is maximized in a much higher dimensionality feature space transformed from input original space with a kernel function.[9] Machine learning algorithms such as SVM, decision tree[5] and unified maximum separability analysis[6] may accidentally identify a complete binary classification in a certain closed sample set, especially when a large volume data set is incorporated. Thus, the classification must be validated in an independent sample set. Two reports have been published recently describing the successful application of SELDI-MS in the serological diagnosis of RCC.[15,16] However, the number of patients with RCC examined in these studies was rather small and it appeared that validation of the results in a separate cohort was necessary.

In this study we identified 2 biomarkers of RCC with a high diagnostic potential through comprehensive proteomic analysis of a training set consisting of 45 cases and controls (fig. 4, *A*). The accuracy of the classification was further verified using an independent validation set consisting of 44 blinded individuals (fig. 4, *B* and data not shown). The com-

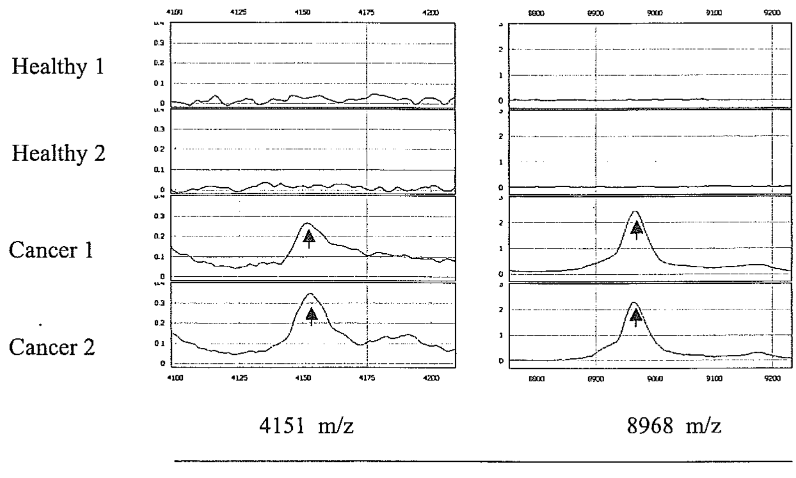

FIG. 2. Representative 4,151 and 8,968 m/z peaks (CM10 at pH 4 Fr 4) detected in serum samples of 2 patients with RCC and 2 controls in biomarker discovery training set.

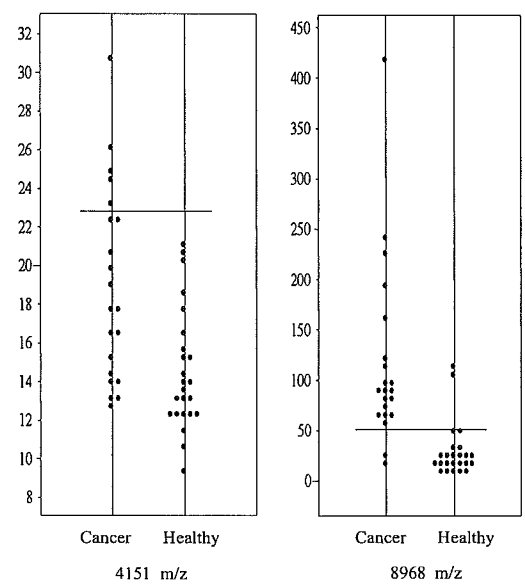

FIG. 3. Plots of intensity of 4,151 and 8,968 m/z peaks (CM10 at pH 4 Fr 4) detected in serum samples of biomarker discovery training set (21 patients with RCC and 24 healthy controls). Borderlines were calculated using linear SVM algorithm.

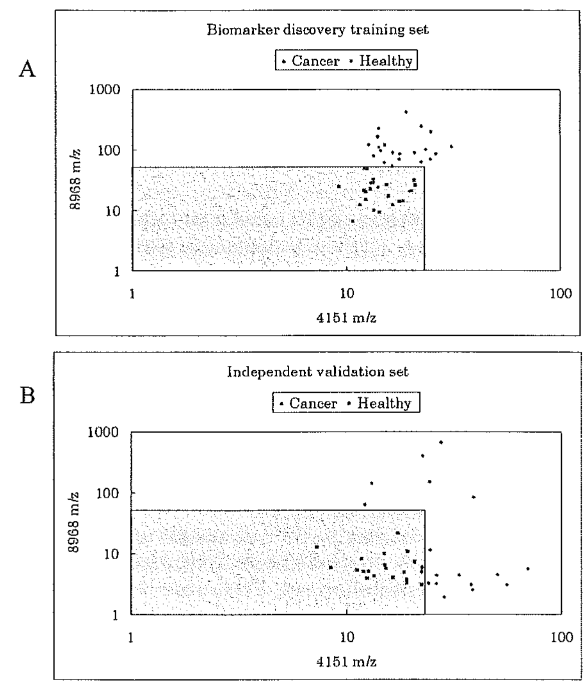

FIG. 4. Two-dimensional plots of intensity of 4,151 (x-axes) and 8,968 m/z (y-axes) peaks (CM10 at pH 4 Fr 4) detected in serum samples of biomarker discovery training set (21 patients with RCC [blue dots] and 24 healthy controls ([red dots], *A*) and independent validation set (19 patients with RCC [blue dots] and 20 healthy controls [red dots], *B*). Cases inside gray areas were diagnosed as healthy and cases outside areas were diagnosed as RCC. Note that same boundary was used in training and validation sets.

bination of a minimal number of proteins at 4,151 and 8,968 m/z appeared to allow serum samples of patients with RCC to be discriminated from those of age and sex matched healthy controls (table 1). Simultaneous recognition of the 2 proteins also seemed to have the potential to allow early detection of RCC (table 3).

Machine learning was performed with the training set which did not include any cases of benign kidney disease. As a result most cases of PN in the validation set were classified into the RCC category. However, that does not compromise the usefulness of our biomarkers for screening for RCC in a large population made up mostly of healthy individuals because pyelonephritis can be distinguished from RCC clinically. The 4,151 and 8,968 m/z peaks in this study are not considered attributable to production or secretion directly by cancer cells. In fact, there was no significant difference in detection rates among RCC cases in different stages or of different histological subtypes (tables 3 and 4). The many diagnostic low molecular weight proteins detected by SELDI-MS in the serum samples of patients with cancer have been reported to include cytokines, peptides, metabolic products and proteolytic fragments produced as a result of host reactions to diseases.[12,17] Liquid chromatography based purification and de novo sequencing may be necessary for protein identification of these metabolic or cleaved protein fragments.

Our biomarkers also may be useful for evaluating the efficacy of therapy and for monitoring the postoperative recurrence of RCC. It would be necessary to conduct large-scale prospective studies to conclude the clinical significance of the biomarkers identified in this study. However, before embarking on such a large-scale study in the general population, it may be more productive to perform the initial analysis in

TABLE 3. *Detection rates of the 2 biomarkers according to RCC clinical stage*

UICC Stage	4,151 M/z (%)	8,968 M/z (%)	Combination (%)*	Case No.
I	27.8	61.1	88.9	18
II	66.7	66.7	100	3
III	83.3	66.7	83.3	6
IV	72.7	45.5	90.9	11

* Detection rates of the 2-peak combination.

TABLE 4. *Detection rate according to the histological classification of RCC*

Histological classification*	Training Set	Validation Set
	Pos/Total No. (%)	Pos/Total No. (%)
Clear cell Ca	16/18 (88.9)	10/12 (83.3)
Clear + granular mixed type	1/1 (100)	2/2 (100)
Clear + spindle mixed type	1/1 (100)	0/0
Clear + granular + spindle mixed type	0/0	1/1 (100)
Granular cell Ca	1/1 (100)	2/2 (100)
Spindle cell Ca	0/0	1/1 (100)
Cyst associated RCC	0/0	1/1 (100)

* According to the WHO International Histological Classification of Tumours (1997).[10]

families affected with known hereditary forms of RCC,[18] and in individuals with a past or family history of multifocal and/or bilateral RCC.

CONCLUSIONS

One of the challenges of current urological practice is the need to discover reliable biomarkers for the early detection of RCC. We have reported on the identification and validation of novel biomarkers for the detection of RCC. Comparative proteomic profiling coupled with computerized machine learning without actual identification of the specific proteins seems to be a rapid and promising alternative to the traditional single protein assays. However, it should be borne in mind that this approach must always be endorsed by high reproducibility of the measurements.

Ms. T. Umaki and Y. Ishiyama provided technical assistance, and Dr. K. Aoshima, Mrs. H. Kuwabara, T. Isobe and H. Matsuzuki of the Mitsui Knowledge Industry Co., Ltd. provided the statistical analyses.

REFERENCES

1. Greenlee, R. T., Murray, T., Bolden, S. and Wingo, P. A.: Cancer statistics, 2000. CA Cancer J Clin, **50:** 7, 2000
2. Pantuck, A. J., Zisman, A. and Belldegrun, A. S.: The changing natural history of renal cell carcinoma. J Urol, **166:** 1611, 2001
3. Berrington de Gonzalez, A. and Darby, S.: Risk of cancer from diagnostic x-rays: estimates for the UK and 14 other countries. Lancet, **363:** 345, 2004
4. Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M. et al: Use of proteomic patterns in serum to identify ovarian cancer. Lancet, **359:** 572, 2002
5. Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H. et al: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res, **62:** 3609, 2002
6. Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S. et al: Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. Clin Cancer Res, **10:** 860, 2004
7. Zhang, Z., Bast, R. C., Jr., Yu, Y., Li, J., Sokoll, L. J., Rai, A. J. et al: Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res, **64:** 5882, 2004
8. Gras, R., Muller, M., Gasteiger, E., Gay, S., Binz, P. A., Bienvenut, W. et al: Improving protein identification from peptide mass fingerprinting through a parameterized multilevel scoring algorithm and an optimized peak detection. Electrophoresis, **20:** 3535, 1999
9. Byvatov, E. and Schneider, G.: Support vector machine applications in bioinformatics. Appl Bioinformatics, **2:** 67, 2003
10. Mostofi, F. K. and Davis, C. J.: Histological Typing of Kidney Tumours. World Health Organization International Histological Classification of Tumours, 2nd ed. Berlin: Springer, 1997
11. Anderson, N. L. and Anderson, N. G.: The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics, **1:** 845, 2002
12. Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P. and Veenstra, T. D.: Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics, **2:** 1096, 2003
13. Diamandis, E. P.: Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst, **96:** 353, 2004
14. White, C. N., Chan, D. W. and Zhang, Z.: Bioinformatics strategies for proteomic profiling. Clin Biochem, **37:** 636, 2004
15. Won, Y., Song, H. J., Kang, T. W., Kim, J. J., Han, B. D. and Lee, S. W.: Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. Proteomics, **3:** 2310, 2003
16. Tolson, J., Bogumil, R., Brunst, E., Beck, H., Elsner, R., Humeny, A. et al: Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. Lab Invest, **84:** 845, 2004
17. Liotta, L. A., Ferrari, M. and Petricoin, E.: Clinical proteomics: written in blood. Nature, **425:** 905, 2003
18. Richard, S., Lidereau, R. and Giraud, S.: The growing family of hereditary renal cell carcinoma. Nephrol Dial Transplant, **19:** 2954, 2004

Possible Prediction of Chemoradiosensitivity of Esophageal Cancer by Serum Protein Profiling

Yasuharu Hayashida,[1,2] Kazufumi Honda,[1] Yoshiaki Osaka,[2] Tomohiko Hara,[1] Tomoko Umaki,[1] Akihiko Tsuchida,[2] Tatsuya Aoki,[2] Setsuo Hirohashi,[1] and Tesshi Yamada[1]

Abstract **Purpose:** Establishment of a reliable method of predicting the efficacy of chemotherapy and radiotherapy is necessary to provide the most suitable treatment for each cancer patient. We investigated whether proteomic profiles of serum samples obtained from untreated patients were capable of being used to predict the efficacy of combined preoperative chemoradiotherapy against esophageal cancer.

Experimental Design: Proteomic spectra were obtained from a training set of 27 serum samples (15 pathologically diagnosed responders to preoperative chemoradiotherapy and 12 non-responders) by surface-enhanced laser desorption and ionization coupled with hybrid quadrupole time-of-flight mass spectrometry. A proteomic pattern prediction model was constructed from the training set by machine learning algorithms, and it was then tested with an independent validation set consisting of serum samples from 15 esophageal cancer patients in a blinded manner.

Results: We selected a set of four mass peaks, at 7,420, 9,112, 17,123, and 12,867 *m/z*, from a total of 859 protein peaks, as perfectly distinguishing responders from nonresponders in the training set with a support vector machine algorithm. This set of peaks (i.e., the classifier) correctly diagnosed chemoradiosensitivity in 93.3% (14 of 15) of the cases in the validation set.

Conclusions: Recent mass spectrometric approaches have revealed that serum contains a large volume of information that reflects the microenvironment of diseased organs. Although a multi-institutional large-scale study will be necessary to confirm each component of the classifier, there is a subtle but definite difference in serum proteomic profile between responders and nonresponders to chemoradiotherapy.

Esophageal carcinoma frequently metastasizes to lymph nodes and directly invades neighboring major organs, including the lung, trachea, bronchi, and large vessels, often making complete tumor resection difficult. Concurrent preoperative chemoradiotherapy has been done to increase the resectability of advanced esophageal cancer (1 – 4), and its advantages are (*a*) increased resectability, (*b*) reduction of surgical stress caused by extensive resection of major organs, and (*c*) control of micrometastasis. Our previous study showed that preoperative chemoradiotherapy for stage III and IV squamous cell carcinoma of the esophagus significantly reduced postoperative recurrence and improved survival (4). However, preoperative chemoradiotherapy does not always provide these benefits in patients with advanced esophageal cancer. Retrospective analyses of resected specimens have revealed that survival was prolonged only when preoperative chemoradiotherapy was shown to be effective pathologically. We reported previously a statistically significant more unfavorable outcome in cases in which preoperative chemoradiotherapy was pathologically ineffective (grade 1) than in cases in which it was effective (grade 2 and 3; ref. 5), and similar results have been reported by other investigators (6 – 8). If preoperative chemoradiotherapy were not effective, the patients would not only receive no benefit from the preoperative chemoradiotherapy but also experience relatively severe side effects of preoperative chemoradiotherapy. They might also miss the chance for curative surgery because of metastasis that might occur during the 2-month period of preoperative chemoradiotherapy. Thus, there is an urgent need for new diagnostic modalities that can reliably predict the efficacy of preoperative chemoradiotherapy in advance.

The surface-enhanced laser desorption and ionization (SELDI) chip can uniformly capture, concentrate, and purify various complicated biological materials on its small chemical surface (9), which represents the most fundamental improvement of SELDI over matrix-assisted laser desorption and ionization (10). Comprehensive profiles of proteins captured

Authors' Affiliations: [1]Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute, and [2]Third Department of Surgery, Tokyo Medical University, Tokyo, Japan
Received 3/24/05; revised 5/30/05; accepted 6/29/05.
Grant support: Ministry of Health, Labor and Welfare of Japan, Program for the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation of Japan, and Foundation for the Promotion of Cancer Research fellowship (Y. Hayashida and T. Hara).
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).
Requests for reprints: Tesshi Yamada, Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. Phone: 81-3-3542-2511; Fax: 81-3-3547-6045; E-mail: tyamada@ncc.go.jp.

doi:10.1158/1078-0432.CCR-05-0656

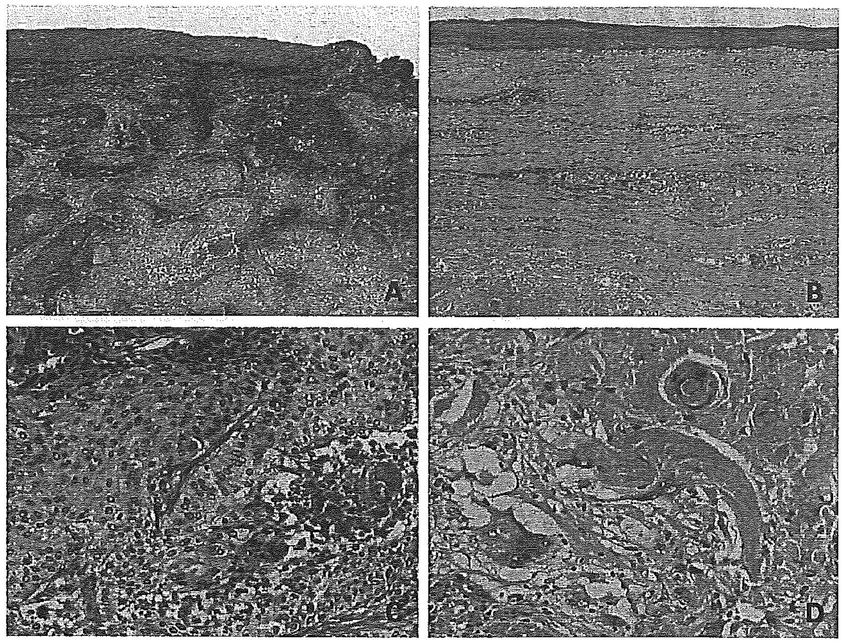

Fig. 1. Histopathology of esophageal cancers treated by preoperative chemoradiotherapy. Histologic findings in the surgical specimens from two representative cases of esophageal cancer resected 4 weeks after preoperative chemoradiotherapy. *A* and *C*, histopathologic appearance of the tumor in a case in which efficacy was classified as grade 1 (slightly effective), showing tumor nests with minimal sign of degeneration. *B* and *D*, histopathologic appearance of the tumor in a case in which efficacy was classified as grade 3 (markedly effective), showing granulomatous inflammation with multinucleated foreign body giant cells, a degenerative cancer pearl, and tumor necrosis. Original magnification, ×40 (*A* and *B*) and ×200 (*C* and *D*).

on SELDI chips can be created from samples of body fluid, such as serum, as small as a few microliters. SELDI-mass spectrometry (MS) analysis of the serum proteome has been reported to be capable of being used to diagnose malignancies of various organs, including the ovary, prostate, and pancreas (11–15), because serum protein profiles precisely reflect the microenvironmental conditions in diseased organs (16). This high diagnostic potential of SELDI-MS prompted us to investigate whether the proteomic profiles obtained from serum samples of untreated esophageal cancer patients could be used to predict the efficacy of preoperative chemoradiotherapy.

Materials and Methods

Patients and serum samples. Forty-two stage II to IV esophageal cancer patients treated with preoperative chemoradiotherapy and surgical resection at the Third Department of Surgery, Tokyo Medical University Hospital, between January 1998 and December 2001 were entered into this study. Preoperative chemoradiotherapy consisted of low-dose cisplatin (10 mg/m^2/d, 5 days weekly for 2 weeks, biweekly, total 100 mg/m^2), 5-fluorouracil (350 mg/m^2/d, 5 days weekly for 4 weeks, total 7,000 mg/m^2), and radiation (10-MV linear accelerator, 2 Gy/d, 5 days weekly for 4 weeks, total 40 Gy) administered concurrently. The histologic diagnosis of squamous cell carcinoma

Table 1. Clinical characteristics of 42 patients in this study

	Training set (n = 27)			Validation set (n = 15)		
	Responder (n = 15)	Nonresponder (n = 12)	P	Responder (n = 10)	Nonresponder (n = 5)	P
Age (mean ± SD)	61.7 ± 8.83	61.5 ± 5.27	0.95*	62.2 ± 12.0	62.5 ± 4.17	0.52*
Gender (%)						
Male	15 (100)	12 (100)	1†	6 (60)	4 (80)	1†
Female	0	0		4 (40)	1 (20)	
Tumor location‡ (%)						
Ce	0	1 (8.3)	0.44†	1 (10)	0	1†
Te	15 (100)	11 (91.7)		8 (80)	5 (100)	
Ae	0	0		1 (10)	0	
Clinical stage‡ (%)						
II	4 (26.7)	2 (16.7)	0.24†	1 (10)	0	0.43†
III	10 (66.7)	7 (58.3)		8 (80)	4 (80)	
IV	1 (6.7)	3 (25)		1 (10)	1 (20)	

*Analyzed using Student's t test.
†Analyzed using Fisher's exact t test.
‡Tumor location and clinical stage were classified according to the Guidelines for Clinical and Pathologic Studies on Carcinoma of the Esophagus (17).

Table 2. Reproducibility of 859 mass peaks detected by SELDI-QqTOF-MS

Array/wash condition	No. detectable peaks	Correlation of coefficient (mean ± SD)
H50	235	0.980 ± 0.016
CM10/pH 4	207	0.955 ± 0.040
CM10/pH 7	181	0.972 ± 0.030
Immobilized metal affinity capture coupled with copper	236	0.942 ± 0.079
Total	859	0.960 ± 0.019

was confirmed in endoscopic biopsy specimens before preoperative chemoradiotherapy, and surgical resection was done ~ 4 weeks after the completion of preoperative chemoradiotherapy. The effects of preoperative chemoradiotherapy were classified into grade 0 (ineffective), grade 1 (slightly effective; Fig. 1A and C), grade 2 (moderately effective), and grade 3 (markedly effective; Fig. 1B and D) according to the criteria described previously (ref. 17; Supplementary Material 1). Informed consent was obtained from all patients. Blood samples were collected 3 to 5 days before preoperative chemoradiotherapy and allowed to coagulate at room temperature. Serum was obtained by centrifugation at 3,000 rpm for 30 minutes and cryopreserved at −80°C until analyzed. The study was reviewed and approved by the Ethics Committee of Tokyo Medical University and the Ethics Committee of the National Cancer Center (Tokyo, Japan).

Surface-enhanced laser desorption and ionization/hybrid quadrupole time-of-flight mass spectrometry analysis. To denature serum proteins, 90 μL U9 buffer [9 mol/L urea, 2% CHAPS, and 50 mmol/L Tris-HCl (pH 9)] was added to 10 μL of each sample and vortexed for 20 minutes. To increase the number of detectable protein peaks, we used four different ProteinChip (Ciphergen Biosystems, Inc., Fremont, CA)

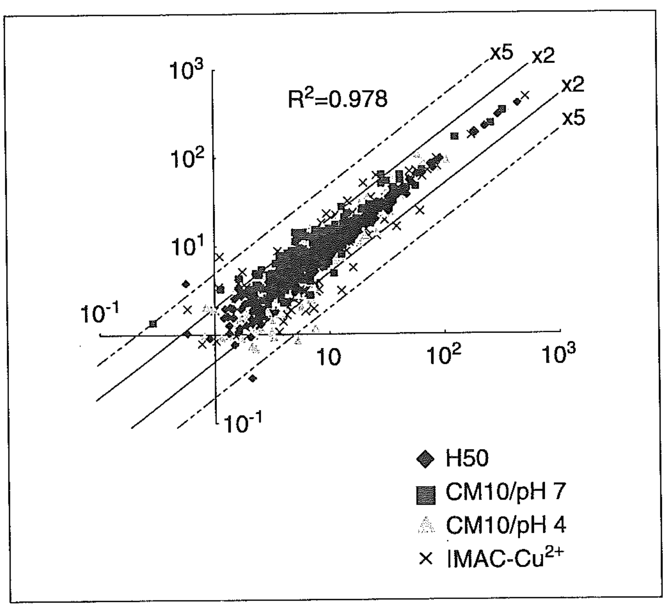

Fig. 2. Correlation analysis of quantitative data obtained from a serum sample. Two-dimensional plot of mass intensities of corresponding duplicated peaks that appeared in H50, CM10/pH 7, CM10/pH 4, and immobilized metal affinity capture coupled with copper (*IMAC-Cu^{2+}*) arrays. Note that >90% of the ionized protein peaks are plotted within 2-fold differences.

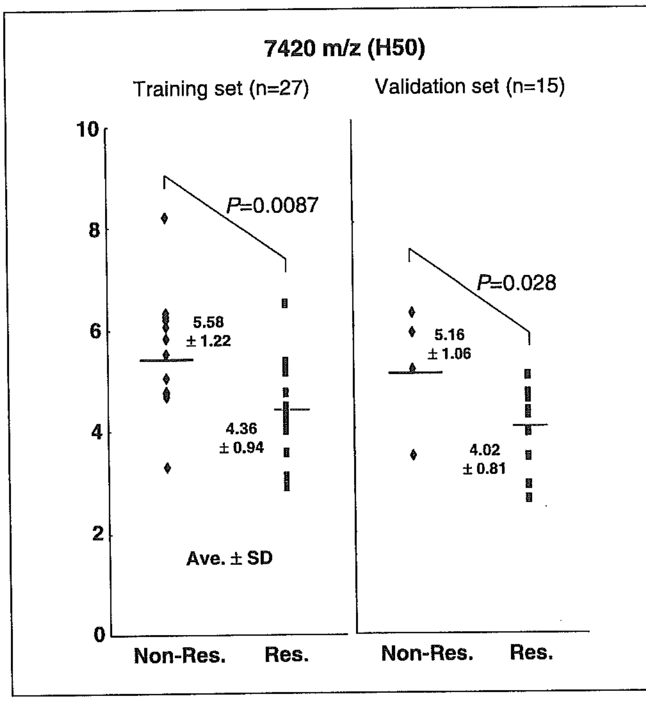

Fig. 3. Distribution pattern of the 7,420 *m/z* peak. Average intensity (average of duplicate measurements in arbitrary units) of the 7,420 *m/z* peak in the 42 serum samples of the training set (*left*) and validation set (*right*). The difference in distribution of intensity between nonresponders (*Non-Res.*) and responders (*Res.*) in both sets was significant (Mann-Whitney *U* test). The distribution of all other peaks is available in Supplementary Material 2.

array/wash conditions [i.e., reversed phase (H50), weak cation exchanger with low stringent wash (CM10/pH 4), cation exchanger with high stringent wash (CM10/pH 7), and immobilized metal affinity capture coupled with copper] as instructed by the supplier. Each sample was randomly assigned in duplicate to 1 of 96 spots of 12 allied ProteinChip arrays with a Biomek 2000 laboratory workstation (Beckman Coulter, Inc., Fullerton, CA). Sinapinic acid solution was prepared in 50% (v/v) acetonitrile and 5% (v/v) trifluoroacetic acid as an energy-absorbing matrix, and saturated solution (1 μL) was applied to each spot on the chips.

Low molecular weight proteins in the 2,000 to 40,000 *m/z* range were read on a high-resolution performance hybrid quadrupole time-of-flight MS (QqTOF-MS) Q-star XL (Applied Biosystems, Foster City, CA) equipped with a PCI 1000 ion source (Ciphergen Biosystems). The laser intensity, frequency, and accumulation time of the instrument were set at 60%, 25 Hz, and 90 seconds, respectively. Mass accuracy was externally calibrated on the day of the measurements by using the all-in-one-peptide molecular mass standard (Ciphergen Biosystems).

Peak detection and quantification. The mass data were converted to text files that consist of *m/z* and intensity after mass calibration by Analyst QS (Applied Biosystems) and processed by the procedures described in Supplementary Material 1 (18). Peak images were generated with Analyst QS and MassNavigator (Mitsui Knowledge Industry Co. Ltd., Tokyo, Japan) software.

Statistical analysis. Statistical differences were evaluated by Fisher's exact probability *t* test, Student's *t* test, and Mann-Whitney *U* test. We did binary classification using the support vector machine (SVM) algorithm such that the distance between the closest training example (the support vector) and the separating hyperplane would be maximized in a much higher dimensionality feature space transformed from the original input space with a kernel function (19). We used SVM with the Gaussian kernel to search to discriminate combinations of peaks. The SVM distance was calculated from the formula described in Supplementary Material 1.

Results

Serum protein profiling by surface-enhanced laser desorption and ionization-hybrid quadrupole time-of-flight mass spectrometry. The clinicopathologic characteristics of the 42 patients enrolled in this study are summarized in the Table 1. We defined cases with the grade 2 or 3 effects in response to preoperative chemoradiotherapy as responders and cases with the grade 1 effects as nonresponders, because there was a significant difference in outcome between these two groups (5). No cases were classified as grade 0. Serum samples were obtained before preoperative chemoradiotherapy with informed consent, and they were stored frozen until the SELDI-QqTOF-MS analysis. Peaks were detected and quantified from raw mass spectra by using custom algorithms specially designed for SELDI-QqTOF-MS (described in Materials and Methods). It was possible to extract a total of 859 mass peaks that were allied all across 42 duplicate serum samples in the 2,000 to 40,000 *m/z* range (Table 2). Mass deviation was within 0.05% throughout the study. The average intensities of the duplicate samples for these 859 mass peaks were used for the subsequent analyses. The mean correlation coefficient of 859 peaks between the duplicates of the 42 serum samples was 0.960 ± 0.019 (0.942-0.980; Table 2; Fig. 2).

Selection of a candidate classifier in the training set by machine learning. We selected 15 responders and 12 nonresponders with no statistically significant difference in age, sex, tumor location, or clinical stage as the training set from the total of 42 cases (Table 1). The remaining 15 cases were maintained apart as a blinded validation set. Based on the 27 samples in the training set, we estimated that not more than five markers would be necessary to achieve 100% discrimination between

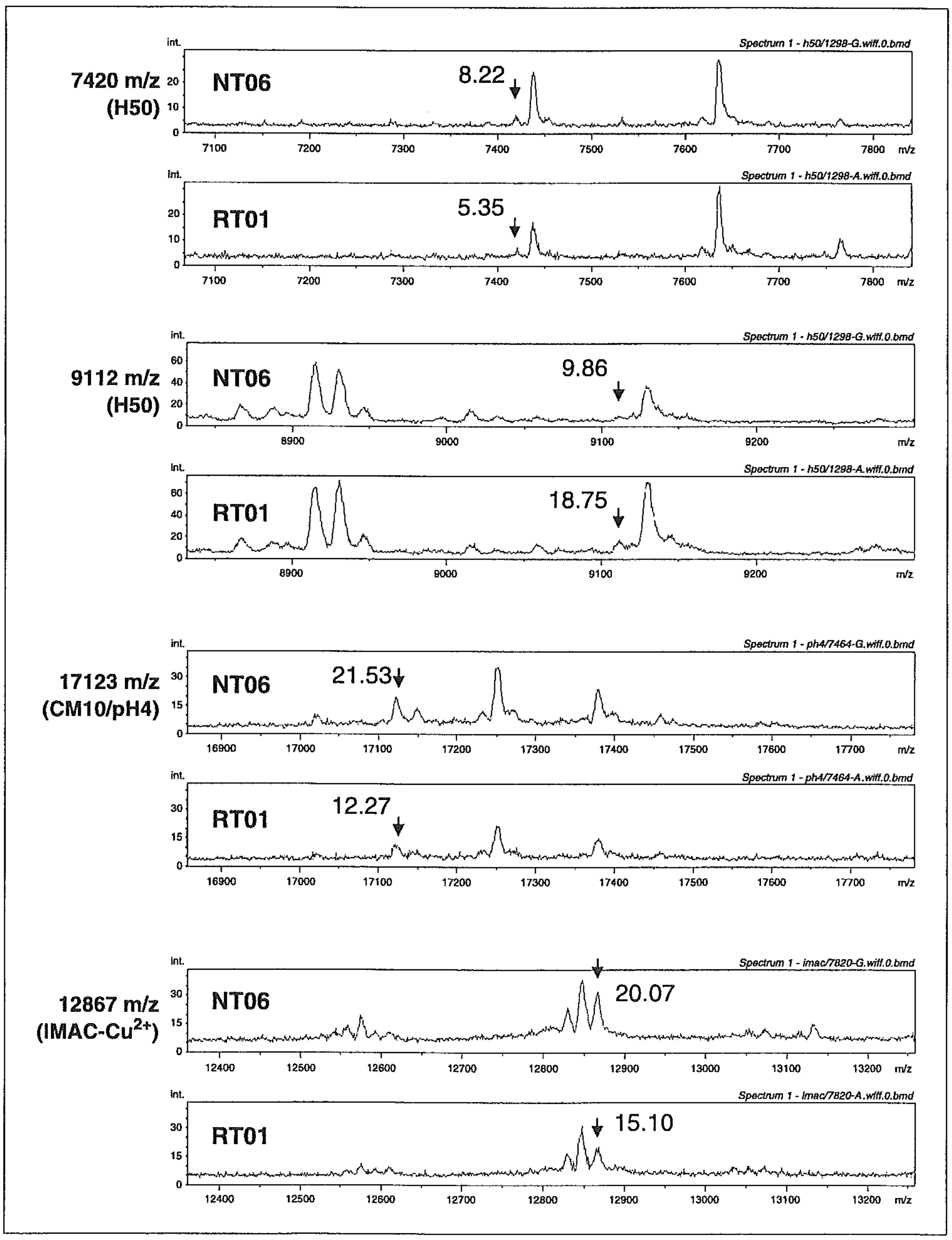

Fig. 4. Representative mass spectra showing the four peaks of the classifier at 7,420 (H50), 9,112 (H50), 17,123 (CM10/pH 4), and 12,867 (immobilized metal affinity capture coupled with copper) *m/z* in a nonresponder (NT06) and a responder (RT01). Arrows and numbers, classifier peaks and their relative intensities (in arbitrary units).

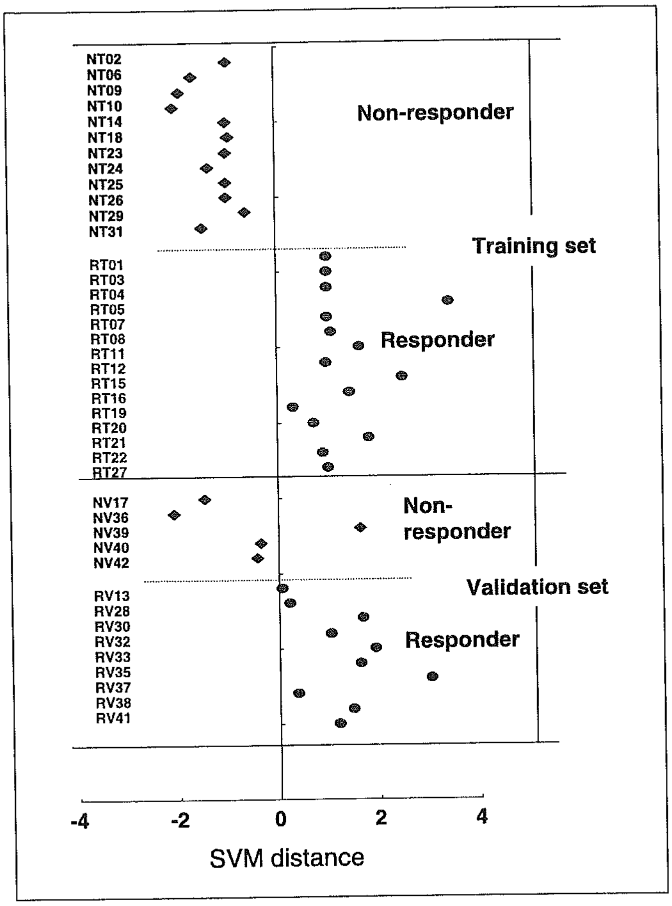

Fig. 5. Classification by the fixed discriminating hyperplane. Scatter graph showing the SVM distance of 42 cases in the training set and validation set. Cases with positive distances were diagnosed as "responders" and cases with negative distances were diagnosed as "nonresponders." NT, nonresponders in the training set; RT, responders in the training set; NV, nonresponders in the validation set; RV, responders in the validation set.

responders and nonresponders based on recursive feature elimination (ref. 20; data not shown). Because it is impossible to check all possible combinations of five markers in time, we first selected a peak at 7,420 *m/z* (H50) that was the most significantly different between responders and nonresponders in the training set ($P = 0.0087$, Mann-Whitney U test; Fig. 3). We then searched for supplementary peaks step-by-step until 100% selection of responders was achieved. Although marker sets that do not contain 7,420 *m/z* may be overlooked by this strategy, the probability of that occurring was thought to be low despite the significant reduction of labor. We found that a minimal set consisting of four peaks at 7,420 (H50), 9,112 (H50), 17,123 (CM10/pH 4), and 12,867 (immobilized metal affinity capture coupled with copper) *m/z* (Fig. 4) diagnosed 15 responders in the training set with 100% sensitivity.

The selection of the set of four peaks (i.e., the classifier) was evaluated by leave-one-out cross-validation. The classifier was able to distinguish the responders in the training set with 100% (15 of 15) sensitivity and the nonresponders with 91.7% (11 of 12) specificity.

Validation of the classifier. As shown in Fig. 5, the sera of the 15 responders and 12 nonresponders in the training set were perfectly separated into positive and negative directions by the fixed hyperplane determined by the classifier. We then used the same fixed hyperplane to calculate the SVM distances of the 15 samples in the validation set that had been set apart from the training set in a blinded manner. All 10 samples from responders in the validation set and 4 of the 5 samples from the nonresponders were classified correctly, yielding a positive predictive value for the diagnosis of responders of 90.9% (10 of 11), a negative predictive value of 100% (4 of 4), and an overall accuracy of 93.3% (14 of 15).

Discussion

Comparative proteomic profiling coupled with computerized machine learning without actual identification of the specific proteins seems to be a rapid and promising alternative method to traditional single protein assays. Multivariate analyses using proteomic data obtained by SELDI-MS have been reported to be highly successful for serum detection of malignancies of various organs (11–16), and this approach may revolutionize medical practice and cancer diagnosis. Resistance to chemotherapy and radiotherapy is probably mediated by a variety of molecular pathways, and it seems reasonable to assume that it is impossible to predict sensitivity to chemotherapy and radiotherapy by measuring any single biomarker. However, it should be borne in mind that multivariate discrimination is dependent on stacks of small differences between cases and controls and must always be corroborated by high accuracy of each measurement. To improve mass accuracy and resolution of MS measurements and minimize day-to-day and machine-to-machine drift, we used a high-resolution QqTOF-MS instrument. Our quality-control experiments revealed that the correlation coefficients between three independent measurements of a pool of plasma samples done every other day were 0.972 to 0.992 (data not shown). The advantage of high-resolution machines was originally predicted by Petricoin and Liotta (16). They divided raw mass spectra into 7,084 "bins" with identical spaces and a fixed *m/z*, whereas we developed custom algorithms to detect and align peaks in a rather conventional way and searched for markers using SVM and leave-one-out cross-validation rather than a genetic algorithm. The number of detectable peaks reached 859 per sample, thanks to the high-resolution mass separation. High reproducibility and low day-to-day mass variation (<0.05%) were achieved by using this high-resolution instrument in this study. We will describe our data for the comparison between low-resolution and high-resolution instruments elsewhere (21).

One nonresponder was misclassified as a responder (Fig. 5). We tried to select the classifier that would identify 100% of the responders with the smallest percentage of misdiagnosis of nonresponders. Predicting responders who are actually nonresponders is more problematic than predicting nonresponders who are responders, because the former type of misclassification may deprive patients whose tumors are sensitive to preoperative chemoradiotherapy of the benefit of preoperative chemoradiotherapy.

Baggerly et al. (22) cautioned about the reproducibility of conventional SELDI-MS analysis. In this study, we selected a classifier consisting of four peaks that distinguished responders to preoperative chemoradiotherapy from nonresponders in the training set. None of the four peaks were in the <2,000 *m/z* range, where chemical noise derived from the matrix often hamper correct measurement (22). We also confirmed that they were not derived from mass noise by visual inspection (Fig. 4). Most errors in biological measurements occur because of preanalytic bias derived from variables in the sample population as well as the sample collection, handling, and processing procedures (22, 23). We examined the serum samples of all consecutive patients with advanced esophageal cancer treated during a certain period and matched various clinical variables of the cases to be compared (Table 1). All serum samples were obtained from untreated patients and processed in the same manner without knowing the results of preoperative chemoradiotherapy; thus, preanalytic bias is unlikely in our study.

There are pros and cons to SELDI-MS with high-resolution QqTOF instruments. One of the two limitations of this study is the unavailability of protein identification. QqTOF-MS can detect peaks with one-digit or two-digit smaller intensities than low-resolution TOF-MS. The intensity of many peaks detected by SELDI-QqTOF-MS seems to be below the sensitivity of tandem MS. Furthermore, it was difficult to purify proteins from these low-intensity peaks without contamination by neighboring high-intensity peaks. Fourier transform MS may overcome these problems, because it does not depend on enzymatic digestion as a precondition for protein identification (24). However, no interface to the SELDI chip is currently available for Fourier transform MS. Many low molecular weight proteins detected by SELDI-MS in serum samples have been reported to be metabolic products, proteolytic fragments, cytokines, and peptides produced as a result of the host-tumor interaction in the microenvironment of the tumor (16, 25, 26). Shimada et al. (27) reported that expression of angiogenic factors predicted the response of esophageal cancer to chemoradiotherapy. We may have to consider the possibility that host factors have an effect on the chemoradiosensitivity of cancer. Actually, an active host reaction is often seen after killing of cancer cells by chemoradiotherapy as shown in Fig. 1D. The second limitation is the number of cases examined. It would be not easy to obtain a larger series of advanced esophageal cancers treated with preoperative chemoradiotherapy at one institution. However, the novel findings in this study warrant a large-scale multi-institutional analysis to confirm the feasibility of serum proteomic profiling as a means of predicting sensitivity to chemotherapy and radiotherapy.

Acknowledgments

We thank Dr. K. Aoshima, H. Kuwabara, T. Isobe, and H. Matsuzuki (Mitsui Knowledge Industry) for the statistical analyses.

References

1. Nygaard K, Hagen S, Hansen HS, et al. Pre-operative radiotherapy prolongs survival in operable esophageal carcinoma: a randomized, multicenter study of pre-operative radiotherapy and chemotherapy. The second Scandinavian trial in esophageal cancer. World J Surg 1992;16:1104–10.
2. Walsh TN, Noonan N, Hollywood D, Kelly A, Keeling N, Hennessy TP. A comparison of multimodal therapy and surgery for esophageal adenocarcinoma. N Engl J Med 1996;335:462–7.
3. Fiorica F, Di Bona D, Schepis F, et al. Preoperative chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis. Gut 2004;53:925–30.
4. Osaka Y, Takagi Y, Tsuchida A, et al. Concurrent preoperative chemoradiotherapy for stage III or IV esophageal squamous carcinoma. Oncol Rep 2004; 12:1121–6.
5. Osaka Y, Takagi T, Hoshino S, et al. Effective of preoperative chemoradiotherapy for advanced esophageal carcinoma. Nihon Gekakei Rengogakkai Zasshi 2004; 29:6–12.
6. Vogel SB, Mendenhall WM, Sombeck MD, Marsh R, Woodward ER. Downstaging of esophageal cancer after preoperative radiation and chemotherapy. Ann Surg 1995;221:685–95.
7. Heath EI, Burtness BA, Heitmiller RF, et al. Phase II evaluation of preoperative chemoradiation and postoperative adjuvant chemotherapy for squamous cell and adenocarcinoma of the esophagus. J Clin Oncol 2000;18:868–76.
8. Ancona E, Ruol A, Santi S, et al. Only pathologic complete response to neoadjuvant chemotherapy improves significantly the long term survival of patients with resectable esophageal squamous cell carcinoma: final report of a randomized, controlled trial of preoperative chemotherapy versus surgery alone. Cancer 2001;91:2165–74.
9. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. Biochem Biophys Res Commun 2002;292:587–92.
10. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumor subsets in on-small-cell lung cancer. Lancet 2003;362:433–9.
11. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.
12. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62: 3609–14.
13. Kozak KR, Amneus MW, Pusey SM, et al. Identification of biomarkers for ovarian cancer using strong anion-exchange ProteinChips: potential use in diagnosis and prognosis. Proc Natl Acad Sci U S A 2003;100: 12343–8.
14. Zhang Z, Bast RC, Jr., Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.
15. Koopmann J, Zhang Z, White N, et al. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. Clin Cancer Res 2004;10:860–8.
16. Petricoin EF, Liotta LA. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr Opin Biotechnol 2004;15:24–30.
17. The Japanese Society for Esophageal Disease. Guidelines for the clinical and pathologic studies on carcinoma of the esophagus. In: Isono K, editor. 9th ed. Tokyo: Kanehara Shuppan; 2001.
18. Gras R, Muller M, Gasteiger E, et al. Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. Electrophoresis 1999;20:3535–50.
19. Byvatov E, Schneider G. Support vector machine applications in bioinformatics. Appl Bioinformatics 2003;2:67–77.
20. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine Learning 2002;46: 389–422.
21. Honda K, Hayashida Y, Umaki T, et al. Possible detection of pancreatic cancer by plasma protein profiling. Cancer Res, in press.
22. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20:777–85.
23. Boguski MS, McIntosh MW. Biomedical informatics for proteomics. Nature 2003;422:233–7.
24. Gershon D. Proteomics technologies: probing the proteome. Nature 2003;424:581–7.
25. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. Nature 2003;425:905.
26. Tolson J, Bogumil R, Brunst E, et al. Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid α in renal cancer patients. Lab Invest 2004;84:845–56.
27. Shimada H, Hoshino T, Okazumi S, et al. Expression of angiogenic factors predicts response to chemoradiotherapy and prognosis of esophageal squamous cell carcinoma. Br J Cancer 2002;86:552–7.

Possible Detection of Pancreatic Cancer by Plasma Protein Profiling

Kazufumi Honda,[1] Yasuharu Hayashida,[1,2] Tomoko Umaki,[1] Takuji Okusaka,[4] Tomoo Kosuge,[5] Satoru Kikuchi,[1,2] Mitsufumi Endo,[2] Akihiko Tsuchida,[2] Tatsuya Aoki,[2] Takao Itoi,[3] Fuminori Moriyasu,[3] Setsuo Hirohashi,[1] and Tesshi Yamada[1]

[1]Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute; [2]Third Department of Surgery and [3]Fourth Department of Internal Medicine, Tokyo Medical University; and [4]Hepatobiliary and Pancreatic Oncology Division and [5]Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Tokyo, Japan

Abstract

The survival rate of pancreatic cancer patients is the lowest among those with common solid tumors, and early detection is one of the most feasible means of improving outcomes. We compared plasma proteomes between pancreatic cancer patients and sex- and age-matched healthy controls using surface-enhanced laser desorption/ionization coupled with hybrid quadrupole time-of-flight mass spectrometry. Proteomic spectra were generated from a total of 245 plasma samples obtained from two institutes. A discriminating proteomic pattern was extracted from a training cohort (71 pancreatic cancer patients and 71 healthy controls) using a support vector machine learning algorithm and was applied to two validation cohorts. We recognized a set of four mass peaks at 8,766, 17,272, 28,080, and 14,779 *m/z*, whose mean intensities differed significantly (Mann-Whitney *U* test, $P < 0.01$), as most accurately discriminating cancer patients from healthy controls in the training cohort [sensitivity of 97.2% (69 of 71), specificity of 94.4% (67 of 71), and area under the curve value of 0.978]. This set discriminated cancer patients in the first validation cohort with a sensitivity of 90.9% (30 of 33) and a specificity of 91.1% (41 of 45), and its discriminating capacity was further validated in an independent cohort at a second institution. When combined with CA19-9, 100% (29 of 29 patients) of pancreatic cancers, including early-stage (stages I and II) tumors, were detected. Although a multi-institutional large-scale study will be necessary to confirm clinical significance, the biomarker set identified in this study may be applicable to using plasma samples to diagnose pancreatic cancer. (Cancer Res 2005; 65(22): 10613-22)

Introduction

The 5-year survival rate of pancreatic cancer sufferers is the lowest among patients with common solid tumors. Pancreatic cancer is the fifth leading cause of cancer-related mortality in Japan and the fourth in the United States, with >19,000 estimated annual deaths in Japan and >28,000 in the United States (1–3). Pancreatic cancer is characterized by massive local invasion and early metastasis to the liver and regional lymph nodes. Because surgical resection is the only reliable curative treatment, early detection is essential to improve the outcomes of pancreatic cancer patients. However, the clinical symptoms of pancreatic cancer, except for obstructive jaundice, are often unremarkable until the advanced stages of the disease, and the anatomic location of the pancreas deep in the abdomen makes physical and ultrasonic detection of pancreatic cancer difficult. As a result, only 20% to 40% of pancreatic cancer patients undergo surgical resection (1, 4). Mass screening by computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) may not be cost-effective because of the relatively low incidence of pancreatic cancer, and the long-term safety of these modalities has not been established (5). Thus, new diagnostic modalities allowing early detection of pancreatic cancer in a safe/noninvasive and cost-effective way are needed.

Recently, mass spectrometry (MS)–based proteomic approaches have gained considerable attention as effective modalities for identifying new biomarkers of various diseases because of their high sensitivity, but proteomic analysis of blood samples has been hampered by the marked dominance of a handful of particularly abundant proteins, including albumin, immunoglobulins, and transferrins (6). Surface-enhanced laser desorption/ionization (SELDI)-MS was developed to resolve these problems and is considered to be among the most useful tools available for the analysis of serum and plasma (7–9). Proteins are captured, concentrated, and purified on the small chemical surface of a SELDI chip, and the molecular weight (*m/z*) and relative intensity of each protein captured on the chip are measured with sensitive time-of-flight (TOF)-MS. As a result, a comprehensive proteomic profile can be created from as little as 20 μL serum/plasma samples. Combined with multivariate bioinformatical analysis, serum proteomics by SELDI-TOF-MS has been reported be successfully applied to the diagnosis of ovarian and prostate cancers (10–13).

The ProteinChip system is a sophisticated commercial platform designed for SELDI-TOF-MS. This system has been widely used because of its high-throughput automated measurements. However, relatively low resolution and poor mass accuracy have been recognized as drawbacks of the TOF-MS instrument of this system, and the reproducibility of SELDI-MS data has been controversial (14–16). Multivariate discrimination is dependent on stacks of small differences between cases and controls. Recently, Petricoin and Liotta reported the use of high-resolution performance hybrid quadrupole TOF-MS (QqTOF-MS) instruments to significantly improve the resolution and mass accuracy of SELDI-MS compared with results obtained with low-resolution instruments (17, 18).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Tesshi Yamada, Chemotherapy Division, National Cancer Center Research Institute, 5-1-1 Tsukiji Chuoh-ku, Tokyo 104-0045, Japan. Phone: 81-3-3547-5201, ext. 4270; Fax: 81-3-3547-6045; E-mail: tyamada@ncc.go.jp.

doi:10.1158/0008-5472.CAN-05-1851

Table 1. Clinicopathologic characteristics of the 220 cases seen at NCCH

	Training cohort			Validation cohort		
	Cancer (n = 71)	Healthy (n = 71)	P	Cancer (n = 33)	Healthy (n = 45)	P
Age (mean ± SD)	61.3 ± 9.06	62.1 ± 10.0	0.6*	62.0 ± 9.06	63.2 ± 11.7	0.6*
Gender						
Male	37	33	0.5†	18	24	0.92†
Female	34	38		15	21	
Tumor location						
Head	34			17		
Body or tail	37			10		
Unknown	0			6		
Clinical stage						
I	1			1		
II	6			4		
III	10			1		
IV	54			27		

*Student's t test.
†Fisher exact probability test.

Koopmann et al. (19) identified a set of biomarkers for pancreatic adenocarcinoma using the ProteinChip system. They increased the number of detectable peaks using stepwise anion-exchange chromatography, but only two of the six fractions were used for subsequent analyses. The two protein peaks that most effectively discriminated between pancreatic cancer patients and healthy controls reportedly achieved a sensitivity of 78% and a specificity of 97%, but this sensitivity was below the level necessary for clinical application. More importantly, diagnostic performance was not validated in an independent cohort. We reviewed and refined various aspects of SELDI-MS. In this study, we first compared the results obtained using low-resolution TOF-MS and high-resolution QqTOF-MS instruments and confirmed the high reproducibility of data obtained using the latter. Computerized machine learning may identify even a perfect multivariate classifier within a closed sample set in a nonbiological/mathematical way (16). Erroneous identification by machine learning must be eliminated by validation experiments using an independent sample set. Herein, we report the identification and validation of a set of biomarkers that can detect pancreatic cancer with high accuracy.

Materials and Methods

Patients and plasma samples. Plasma samples (n = 245) were obtained from two institutes, the National Cancer Center Hospital (NCCH; Tokyo, Japan) between August 2002 and October 2003 and the Tokyo Medical University Hospital (TMUH; Tokyo, Japan) between February 2004 and February 2005. The 220 NCCH cases included untreated pancreatic ductal adenocarcinoma patients (n = 104) and healthy controls (n = 116), whereas the 25 TMUH cases included untreated pancreatic ductal adenocarcinoma patients (n = 9), individuals with pancreatic tumors and/or cysts (n = 6), chronic pancreatitis patients (n = 5), and healthy controls (n = 5). The pancreatic tumor and/or cyst category included two pathologically unproven mucinous cystic tumors, two pathologically unproven serous

Table 2. Comparison of low-resolution and high-resolution instruments

	High-resolution QqTOF-MS		Low-resolution TOF-MS			
	Unfractionated		Unfractionated		Fractionated	
	No. unique peaks*	Correlation coefficient (r), mean ± SD	No. unique peaks*	Correlation coefficient (r), mean ± SD	No. unique peaks*	Correlation coefficient (r), mean ± SD
H50	263	0.96 ± 0.03	64	0.96 ± 0.04	214	0.76 ± 0.35
CM10 pH 4	124	0.99 ± 0.01	53	0.90 ± 0.11	219	0.73 ± 0.33
CM10 pH 7	73	0.98 ± 0.01	48	0.89 ± 0.09	168	0.61 ± 0.46
IMAC-Cu^{2+}	177	0.95 ± 0.04	61	0.87 ± 0.13	271	0.70 ± 0.44
Total	637		226		872	

*Number of unique peaks detectable in plasma samples from 24 pancreatic cancer patients and 24 healthy controls.

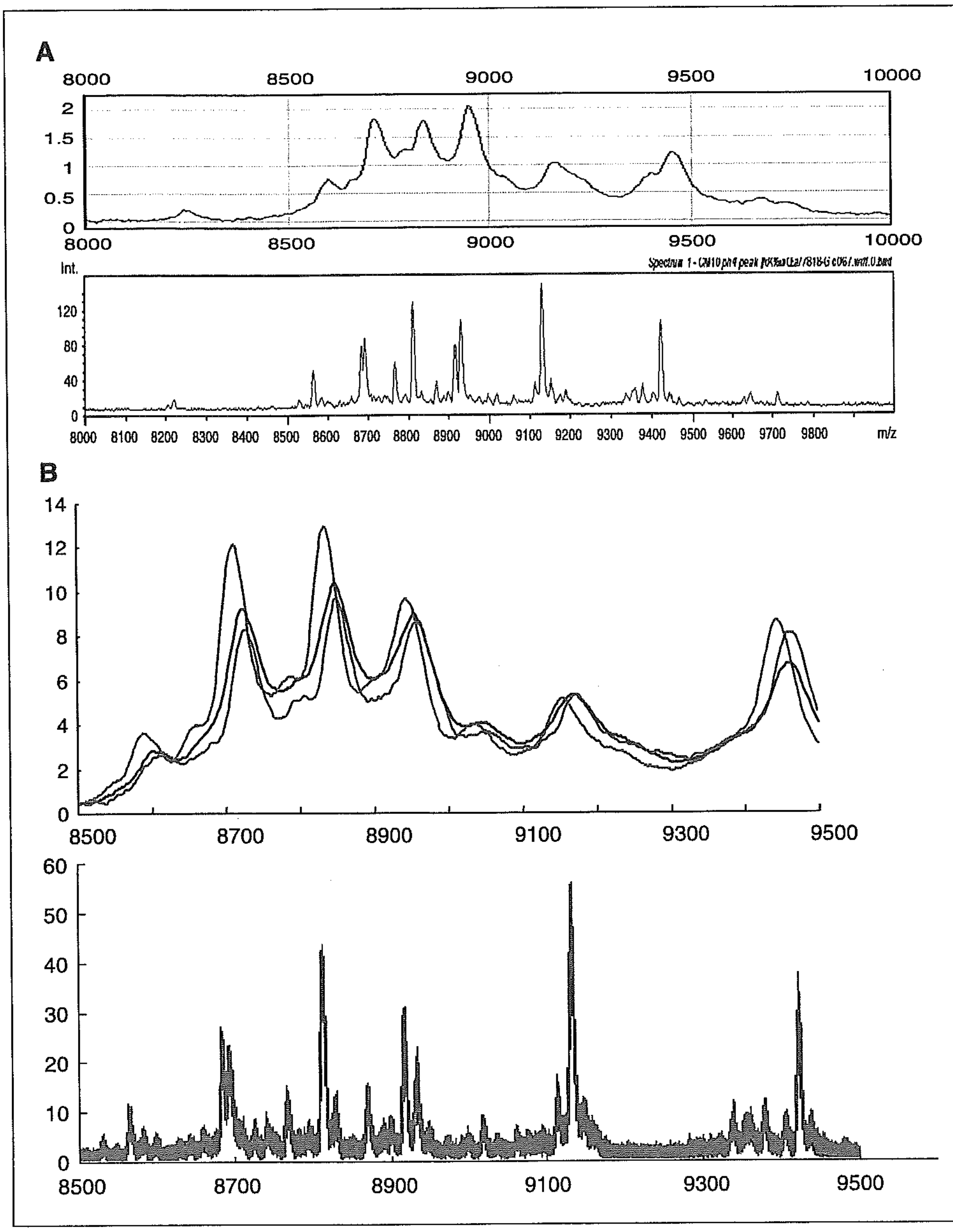

Figure 1. Comparison of low-resolution and high-resolution instruments. *A*, representative spectra of an unfractionated plasma sample in the range of 8,000 to 10,000 *m/z* obtained using a low-resolution TOF instrument (*top*) and a high-resolution QqTOF instrument (*bottom*). *B*, spectra of an unfractionated plasma sample in the range of 8,500 to 9,500 *m/z* obtained thrice every other day using a low-resolution TOF instrument (*top*) and a high-resolution QqTOF instrument (*bottom*). The spectra (*green*, *blue*, and *red lines*) were superimposed to allow visualization of the day-to-day variations. Note that only the *green line* is visible in the *bottom* because of the high reproducibility of results obtained with the QqTOF instrument.

papillary tumors, and two clinically diagnosed nonmalignant mass lesions in the pancreas. These cases are currently being followed, and a final diagnosis has not been obtained to date. The patients in the chronic pancreatitis category had no detectable mass lesions in the pancreas. Written informed consent was obtained from all of the subjects. Blood samples were collected in EDTA glass tubes. The supernatant was separated by centrifugation and cryopreserved at −80°C until analysis. All samples were processed in the same manner. The study was reviewed and approved by the ethics committees of the National Cancer Center (Tokyo, Japan; authorization nos. 16-36 and 16-71) and Tokyo Medical University (Tokyo, Japan; authorization no. 341).

The clinical characteristics of the patients are summarized in Table 1. Patients were classified as having clinical disease stage I, II, III, or IV according to the Fifth Edition of the General Rules for the Study of Pancreatic Cancer (Japanese Pancreas Society; ref. 20).

Surface-enhanced laser desorption/ionization. Ninety microliters of U9 buffer [9 mol/L urea, 2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid, and 50 mmol/L Tris-HCl (pH 9)] were added to 10 μL of each plasma sample and vortexed for 20 minutes. Parts of the denatured plasma samples were fractionated using stepwise anion-exchange chromatography (pH 9 plus flow trough, pH 7, pH 5, pH 4, pH 3, and organic wash) with QHyper DF resin (Ciphergen Biosystems, Inc., Fremont, CA) using a Biomek 2000 Laboratory Automation Robot (Beckman Coulter, Fullerton, CA) according to a previously described method (12, 21).

Each sample was randomly assigned, with a 96-spot format, to 12 ProteinChip arrays (8 spots per array; Ciphergen) in duplicate using the Biomek 2000 Robot. Three types of ProteinChip arrays with different surface chemistries [i.e., immobilized metal affinity capture coupled with copper (IMAC-Cu^{2+}), weak hydrophobic (H50), or cationic (CM10) arrays] were used (21). The CM10 arrays were used under either low-stringent (pH 4) or high-stringent (pH 7) conditions as instructed by the supplier. The arrays were air-dried and applied to the matrix (50% sinapinic acid in 50% acetonitrile/0.1% trifluoroacetic acid).

Time-of-flight mass spectrometry. TOF-MS analysis was done using two types of mass spectrometers, a low-resolution TOF-MS (PBS IIc, Ciphergen) and a high-resolution QqTOF-MS [Q-star XL (Applied Biosystems, Framingham, CA) equipped with a PCI 1000 (Ciphergen)]. Peak detection for the low-resolution instrument was done using CiphergenExpress software version 2.1 (Ciphergen). All of the spectra were compiled and normalized to the total ion currents, and the baselines were subtracted. Peaks between 3,000 and 30,000 *m/z* were autodetected using a signal-to-noise ratio of >3, and the peaks were clustered using second-pass peak selection with a signal-to-noise ratio of >2 and 0.3% mass windows. The permissible range of *m/z* drift between samples was set at 0.3% (21).

The high-resolution instrument was set to measure the range between 2,000 and 40,000 *m/z*. The laser intensity, laser frequency, and accumulation time were set to 60%, 25 Hz, and 90 seconds, respectively. The mass data obtained using the high-resolution instrument were converted to text files consisting of *m/z* and intensity after mass calibration by Analyst QS (Applied Biosystems) and were processed using newly developed in-house peak detection, normalization, and quantification software (22).

The peak data were visualized using Mass Navigator software (Mitsui Knowledge Industry, Tokyo, Japan). Mass accuracy was calibrated externally on the day of the measurements using an all-in-one-peptide molecular mass standard (Ciphergen).

Statistical analysis. Statistically significant differences were detected using the Fisher exact probability test, the Student's *t* test, and the Mann-Whitney *U* test. Receiver operator characteristics (ROC) curves were generated and the area under the curve (AUC) values were calculated using StatFlex software version 5.0 (Artech, Osaka, Japan; ref. 23).

We compiled the multivariate intensity data of the mass peaks into the distance from a support vector machine (SVM) hyperplane using the following formula (details in Supplementary Data; ref. 24):

$$dis(x_i) = \sum_{j=1}^{N} \lambda_j y_j \{k(x_j, x_i) + \alpha\}$$

where y_i is label (1 or −1), $k(x_j,x_i)$ is Gaussian kernel function, and λ_i is a value that maximizes [1] target function under [2] constrained conditions, where $L = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j K(x_i, x_j)$ is the [1] target function, $0 \leq \lambda_i \leq C \sum_{i=1}^{N} \lambda_i y_i = 0$ are the [2] constrained conditions, and α and C are constants 0.25 and 10, respectively.

Immunoradiometric assay of CA19-9. Plasma (100 μL) was analyzed using a commercially available immunoradiometric assay kit (Fujirebio Diag-nostic, Inc., Malvern, PA) according to the manufacturer's recommendations.

Results

Comparison between low-resolution and high-resolution instruments. The reproducibility of data obtained using the low-resolution TOF-MS instrument of the ProteinChip system has been a concern. We compared the number of detectable peaks and the reproducibility of data obtained using low-resolution TOF-MS and high-resolution QqTOF-MS instruments. From unfractionated plasma samples (24 pancreatic cancer patients and 24 healthy controls), a total of 226 unique peaks were detected using the low-resolution instrument and 637 unique peaks were detected using the high-resolution instrument (Table 2). This difference seems to be attributable to the mass resolutions of the instruments (Fig. 1*A*). In addition, we noticed significant mass drifts (<0.3%) in the data obtained with the low-resolution instrument (Fig. 1*B*). In contrast, the mass deviation was <0.05% for the high-resolution instrument (Fig. 1*B*). As a result, the correlation coefficients for three independent measurements of a pooled plasma sample done every other day with the high-resolution instrument reached 0.97 to 0.99 (data not shown).

Chromatographic fractionation reduced the reproducibility of measurements. Fractionation via stepwise anion-exchange chromatography has been widely done to increase the number of detectable peaks obtained with low-resolution instruments. Actually, the total number of detectable peaks increased from 226 to 872 with fractionation of the same plasma samples (Table 2). However, the fractionation procedure seemed to compromise the reproducibility of the measurements. Forty-eight plasma samples (24 pancreatic cancer patients and 24 healthy controls) were analyzed in duplicate, and the mean correlation coefficient of all the peaks calculated between the duplicates was 0.87 to 0.96 for the unfractionated samples and 0.61 to 0.76 for the fractionated samples (Table 2). Fig. 2*A* (unfractionated) and Fig. 2*B* (fractionated) show the results of duplicate assays of a representative plasma sample.

Based on these quality-control experiments, we decided to measure unfractionated plasma samples using the high-resolution QqTOF-MS instrument. More than 90% of the duplicate

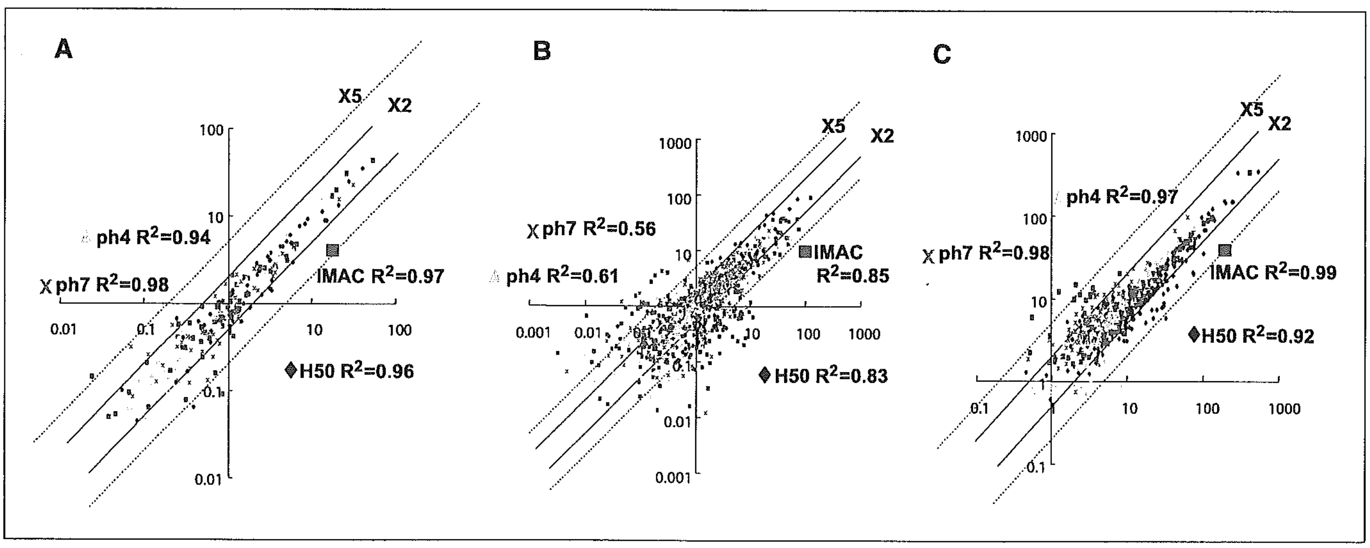

Figure 2. Reproducibility of data from the low-resolution and high-resolution instruments. Two-dimensional plot analyses of the mass intensities corresponding to the duplicated peaks that appeared in the H50 (*blue diamonds*), IMAC-Cu^{2+} (*red squares*), CM10 pH 4 (*yellow triangles*), and CM10 pH 7 (*light blue crosses*) arrays. Unfractionated (*A* and *C*) or fractionated (*B*) samples of the same plasma were measured using a low-resolution TOF instrument (*A* and *B*) and a high-resolution QqTOF instrument (*C*).

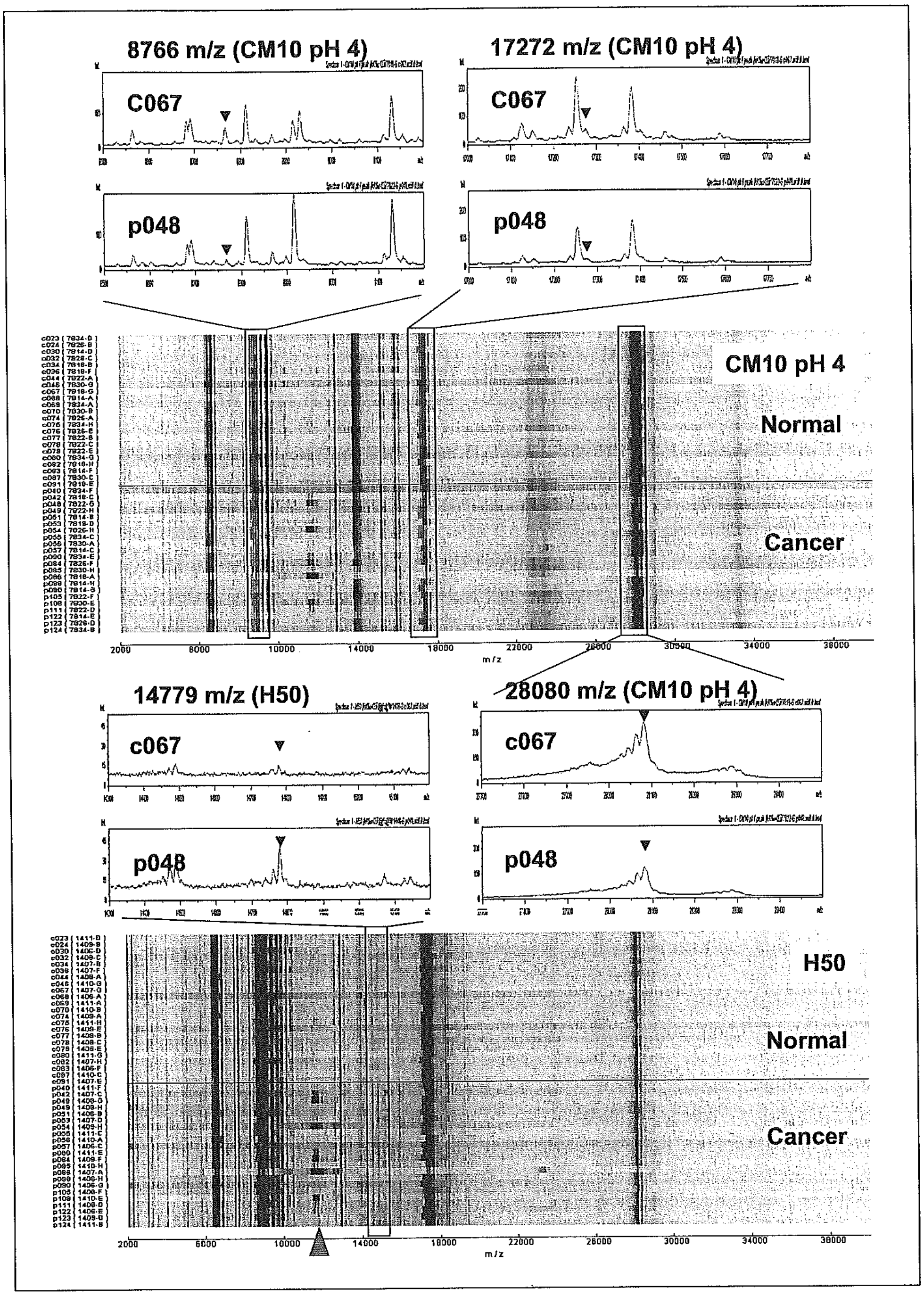

Figure 3. Representative mass spectra [a healthy control (*c067*) and a pancreatic cancer patient (*p048*)] and converted gel-like images [23 healthy controls (*c023-c091*) and 22 pancreatic cancer patients (*p040-p124*)] showing the peaks at 8,766, 17,272, 28,080 (CM10 pH 4), and 14,779 (H50) *m/z*. *Red arrowhead*, peak at 11,516 *m/z*, which was extracted using the Akaike information criterion (25).

protein peaks measured with the QqTOF-MS instrument were plotted within a 2-fold difference (Fig. 2*C*), and the mean correlation coefficient between duplicate assays was at least 0.95 (Table 2).

Identification of a candidate classifier in the training cohort by machine learning. From the total of 220 samples obtained at the NCCH, we selected 71 pancreatic cancer patients and 71 healthy controls with no statistically significant differences in age or sex distribution as a training cohort (Table 1). The remaining 78 cases served as a validation cohort. The clinicopathologic characteristics of these pancreatic cancer patients in the training and validation cohorts are summarized in Table 1.

The acquired MS peak information was stored in a large-capacity server computer, and the data set that most accurately discriminated pancreatic cancer patients from healthy controls was extracted using a rbf SVM learning algorithm (24). The set, or classifier, was composed of four protein peaks at 17,272 *m/z* (CM10 pH 4), 8,766 *m/z* (CM10 pH 4), 28,080 *m/z* (CM10 pH 4), and 14,779 *m/z* (H50). The selection of these four peaks was evaluated by leave-one-out (LOO) cross-validation. Representative spectra profiles and pseudo-gel images of the four peaks are shown in Fig. 3. Akaike information criterion procedure (25) selected another peak at 11,516 *m/z* (H50; indicated by a *red arrowhead* in Fig. 3). Although the 11,516 *m/z* peak was only

Table 3. Intensities of the 17,272, 8,766, 14,779, and 28,080 *m/z* peaks

Peaks (arrays)	Training cohort (n = 142)			Validation cohort (n = 78)		
	Cancer (n = 71)	Healthy (n = 71)	P*	Cancer (n = 33)	Healthy (n = 45)	P*
17,272 *m/z* (CM10 pH 4)	9.49 ± 2.88†	14.6 ± 2.29†	0.0000	9.74 ± 4.22†	14.5 ± 2.29†	0.0000
8,766 *m/z* (CM10 pH 4)	7.65 ± 3.53†	12.1 ± 5.55†	0.0000	7.04 ± 4.39†	13.4 ± 5.81†	0.0000
14,779 *m/z* (H50)	11.8 ± 4.43†	7.85 ± 3.68†	0.0000	10.4 ± 3.85†	6.46 ± 1.63†	0.00000
28,080 *m/z* (CM10 pH 4)	113 ± 36.7†	132 ± 33.5†	0.0022	92.4 ± 24.3†	110 ± 21.6†	0.0078

*Mann-Whitney *U* test.
†Mean ± SD intensities in arbitrary units.

detected in 1 of the 71 (1.4%) healthy controls, it was not included in the above discriminating data set generated by machine learning because of its low-positive rate in pancreatic cancer patients [19.7% (14 of 71)].

Statistical differences in all four peaks were recognized between the pancreatic cancer patients and the healthy controls (Mann-Whitney *U* test, $P < 0.0022$; Table 3). The ROC and AUC values of each peak and their combination in the 142 cases of the training cohort are shown in Fig. 4.

The intensity data of the four peaks obtained in each individual were compiled into a single value, the distance from a fixed SVM hyperplane, using the formula described in Materials and Methods and Supplementary Data. When the distance was positive, the individual was classified as having pancreatic cancer and vice versa. This classifier correctly diagnosed 97.2% (69 of 71) of the cancer patients and 94.4% (67 of 71) of the healthy controls in the training cohort (Fig. 5*A*).

Confirmation of the classifier in the first validation cohort. We next validated the discriminating performance of the classifier in a blinded manner using an independent cohort consisting of 78 individuals (NCCH) who had not been included in the training cohort (Table 1). Again, statistically significant differences in the mean intensities of every peak were observed between the 33 pancreatic cancer patients and the 45 healthy controls (Mann-Whitney *U* test, $P < 0.0078$; Table 3).

The SVM hyperplane determined in the training cohort was applied to the diagnosis of the 78 cases in the validation set. The same SVM hyperplane separated 90.9% (30 of 33) of the pancreatic cancer patients into the positive direction group and 91.1% (41 of 45) of the healthy controls into the negative direction group (Fig. 5*B*). The overall accuracy of the classification was 91.0% (71 of 78) in the validation cohort.

Combination of the surface-enhanced laser desorption/ionization classifier and CA19-9. Overall, the classifier was able to detect 95.2% (99 of 104) of the pancreatic cancer patients in the training and validation cohorts (Table 4). Although the number of cases was small, 83.3% (10 of 12) of stage I and II cases were detected (training and first validation cohorts). No statistically significant differences in detection rates were seen among cases with different tumor locations or different clinical stages (Table 4). To improve the detection rate, we measured plasma CA19-9 levels in all individuals whose residual samples were sufficient (29 pancreatic cancer patients and 39 healthy controls; Table 5). The sensitivity of CA19-9 (cutoff value of 37 units/mL) was 86.2% (25 of 29) and specificity was 94.9% (37 of 39). The SELDI classifier and the CA19-9 level were complementary. Combining CA19-9 and the SELDI classifier detected 100% (29 of 29) of cancer patients, but this combination yielded six false-positive cases [15.4% (6 of 39); Table 5].

Confirmation of the classifier in a second validation cohort obtained at a different institution. Finally, we did a second confirmatory experiment using samples collected prospectively at another institution. In total, 25 plasma samples from pancreatic cancer patients, individuals with other pancreatic diseases, and healthy volunteers were obtained from TMUH and analyzed in a blinded manner. Although the discovery of biomarkers useful for the differential diagnosis of pancreatic diseases was not the primary goal of this study, the classifier was able to discriminate pancreatic cancer patients and individuals with pancreatic tumors/cysts from healthy controls and pancreatitis patients (Table 4; Fig. 6). Four of the six patients with pathologically unproven pancreatic tumors/cysts were classified into the positive direction group. A close follow-up of these patients has been undertaken, because they may have premalignant or preclinical conditions. The SELDI classifier correctly identified 88.9% (8 of 9)

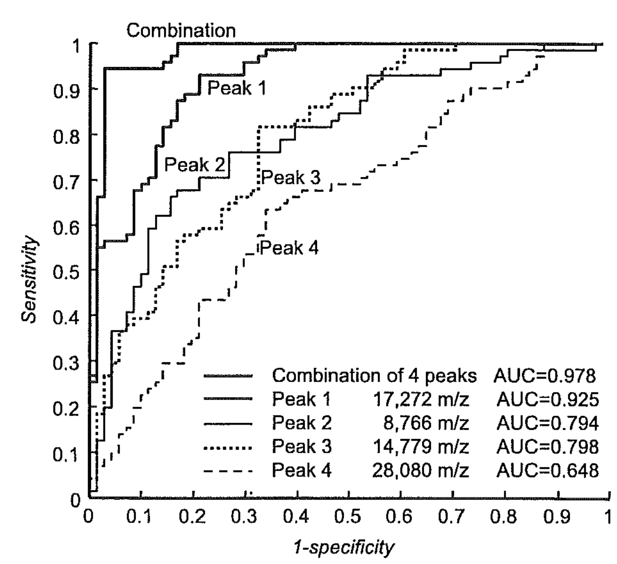

Figure 4. ROC curves and AUC values showing the discriminating capacities of the 17,272, 8,766, 28,080 (CM10 pH 4), and 14,779 (H50) *m/z* peaks individually and in combination.

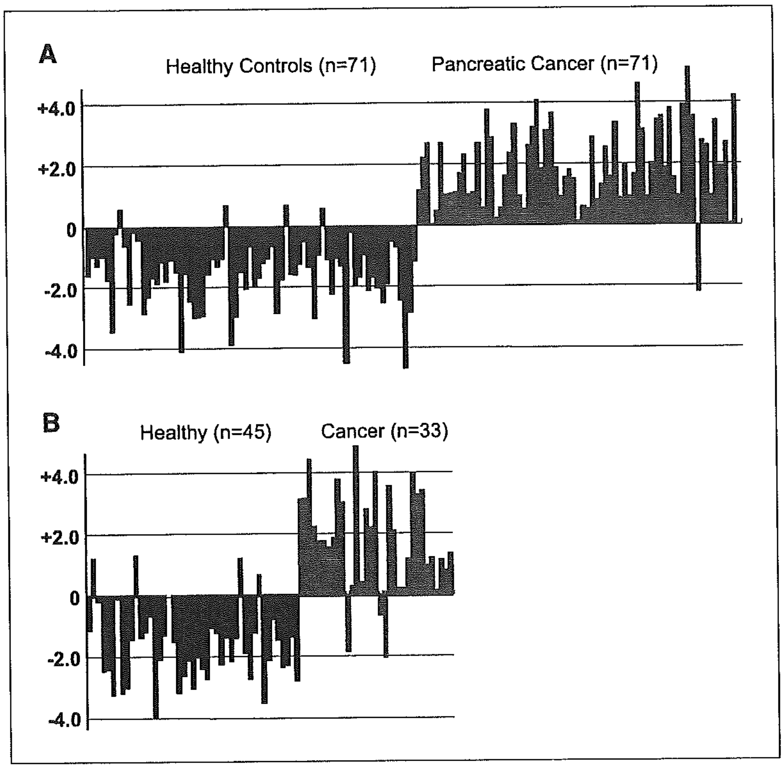

Figure 5. Calculated SVM distances of healthy controls (*black columns*) and pancreatic cancer patients (*gray columns*) in the training (*A*) and first validation (*B*) cohorts. Cases separated into the positive direction from the SVM hyperplane were classified as having "cancer" and those separated into the negative direction were classified as being "healthy."

of the pancreatic cancer patients and 80% (4 of 5) of the healthy controls, whereas the CA19-9 level correctly identified 66.7% (6 of 9) of the pancreatic cancer patients and 100% (5 of 5) of the healthy controls (Fig. 6). Again, in all the pancreatic cancer patients (9 of 9), the SELDI classifier and the CA19-9 level provided complementary results, even in this second validation cohort.

Discussion

Comparative proteomic profiling coupled with a computerized machine learning approach may revolutionize medical practice and cancer diagnosis. We compared the plasma protein profiles of a large number of pancreatic cancer patients and healthy controls with identical age and gender distributions (Table 1) to identify a biomarker for detecting pancreatic cancer patients in a large

Table 4. Diagnostic accuracy of the SELDI classifier

	Training cohort		Validation cohort (NCCH)		Validation cohort (TMUH)	
	No. cases	No. correctly classified samples* (%)	No. cases	No. correctly classified samples* (%)	No. cases	No. correctly classified samples* (%)
Healthy	71	67 (94.4)	45	41 (91.1)	5	4 (80)
Pancreatitis					5	4 (80)
Tumor/cyst†					6	4 (66.6)
Cancer	71	69 (97.2)	33	30 (90.9)	9	8 (88.9)
Cancer location						
Head	34	33 (97.1)	17	14 (82.4)	7	7 (100)
Body or tail	37	36 (97.3)	10	10 (100)	2	1 (50)
Unknown	0	0	6	10 (100)		
Clinical stage						
I	1	0 (0)	1	1 (100)	0	0
II	6	6 (100)	4	3 (75)	0	0
III	10	9 (90)	1	1 (100)	3	3 (100)
IV	54	54 (100)	27	25 (92.6)	6	5 (83.3)

*Number of healthy and chronic pancreatitis cases, considered to be "healthy," and number of pancreatic tumor/cyst and cancer cases given a diagnosis of "cancer."
†Pathologically unproven pancreatic tumor and/or cyst.

Table 5. Detection rates with CA19-9, the SELDI classifier, and their combination

	No. cases*	CA19-9, *n* (%)	SELDI, *n* (%)	Combination, *n* (%)
Healthy	39	3 (7.7)	4 (10.3)	6 (15.4)
Cancer	29	25 (86.2)	26 (89.7)	29 (100)
Clinical stage				
I	1	1 (100)	1 (100)	1 (100)
II	3	3 (100)	2 (66.6)	3 (100)
III	1	0 (0)	1 (100)	1 (100)
IV	24	21 (87.5)	22 (91.7)	24 (100)

*Cases whose plasma samples were available for CA19-9 measurement in the NCCH validation cohort.

population composed mainly of healthy individuals. The reproducibility of data obtained using the low-resolution instrument of the ProteinChip system has been a concern, but employing a high-resolution QqTOF instrument was found to significantly improve mass accuracy and minimize day-to-day variations (Fig. 1*B*). The high reproducibility of measurements was confirmed not by using a few high-intensity peaks selected intentionally but rather by using all the peaks detectable in the entire range (intensity and *m/z*) of mass spectra (Fig. 2). We also eliminated fractionation procedures, which increased the number of detectable peaks but significantly decreased reproducibility (Table 2; Fig. 2). A minimal set of four low-molecular weight proteins (Fig. 3) was found to be sufficient for discriminating pancreatic cancer patients with a sensitivity of 97.2% (69 of 71) and a specificity of 94.4% (67 of 71; Fig. 5*A*). This high discriminating capacity was confirmed by LOO cross-validation and ROC analysis (Fig. 4). We confirmed the discriminating capacity of our classifier in two independent validation cohorts (Figs. 5*B* and 6) to eliminate accidental identification of nonbiological/mathematical multivariate classifiers within a closed cohort by overfitting.

We noticed that a peak at 11,516 *m/z* (H50) was detected in 19.4% of the pancreatic cancer patients in the training cohort but in only 1.4% of the healthy controls (the peaks are indicated by a *red arrowhead* in Fig. 3). Tolson et al. (26) reported that an 11.5-kDa protein was detected in 32% of renal cell carcinoma patients but in none of the normal controls. Howard et al. (27) identified 11,682 *m/z* proteins in the sera of lung cancer patients as a diagnostic biomarker using matrix-assisted laser desorption/ionization (MALDI)-TOF-MS. Both groups identified the proteins as fragments of serum amyloid A. Serum amyloid A is an acute-phase reactant and a biomarker for inflammatory disease. The serum amyloid A level is elevated up to 1,000-fold during tissue damage and inflammation and is also increased in patients with various solid tumors and hematopoietic malignancies. However, serum amyloid A has not been recognized as a tumor marker because of its low positive rate (28, 29). Consistently, the 11,516 *m/z* peak was not incorporated into our classifier. The discovery of a single biomarker differing markedly between cancer patients and controls as well as having a high positive rate in cancer patients would be ideal but is perhaps not realistic. Since the discovery of CA19-9 in

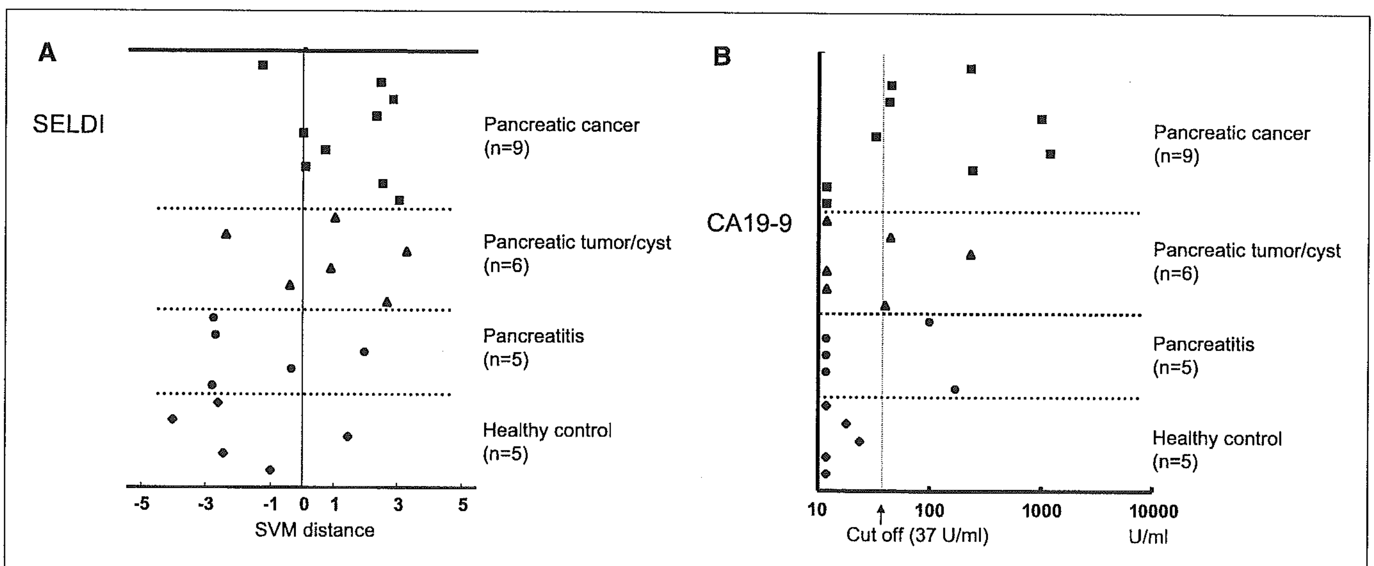

Figure 6. Confirmation in a second cohort treated at a different institution. *A*, calculated SVM distances of nine pancreatic cancer patients, six individuals with pancreatic tumors and/or cysts, five chronic pancreatitis patients, and five healthy controls seen at TMUH. *B*, plasma CA19-9 levels in nine pancreatic cancer patients, six individuals with pancreatic tumors and/or cysts, five chronic pancreatitis patients, and five healthy controls seen at TMUH. The cutoff value was set at 37 units/mL.