

A Serum Biomarker Model to Diagnose Pancreatic Cancer Using Proteomic Fingerprint Technology

Chunlin Ge¹
Ning Ma¹
Dianbo Yao¹
Fengming Luan¹
Chaojun Hu²
Yongzhe Li²
Yongfeng Liu¹

¹ The Department of General Surgery, The First Affiliated Hospital of China Medical University, Shenyang 110001, Liaoning Province, China.

² Department of Clinical Laboratory, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100730, China.

Correspondence to: Chunlin Ge
E-mail: chunlinge@yahoo.com.cn

This work was supported by a grant from the Key Project of Science and Technology Commission of Liaoning Province, China (No. 2005225003-3).

Received January 17, 2008; accepted May 19, 2008.

CJCO <http://www.cjco.cn>
E-mail: 2008cocrcr@gmail.com
Tel (Fax): 86-22-2352 2919

OBJECTIVE To establish a serum protein pattern model for screening pancreatic cancer.

METHODS Twenty-nine serum samples from patients with pancreatic cancer were collected before surgery, and an additional 57 serum samples from age and sex-matched individuals without cancer were used as controls. WCX magnetic beads and a PBS II-C protein chip reader (CIPHERGEN Biosystems Inc) were employed to detect the protein fingerprint expression of all serum samples. The resulting profiles comparing serum from cancer and normal patients were analyzed with the Biomarker Wizard system, to establish a model using the Biomarker Pattern system software. A double-blind test was used to determine the sensitivity and specificity of the model.

RESULTS A group of 4 biomarkers (relative molecular weights were 5,705 Da, 4,935 Da, 5,318 Da, 3,243 Da) were selected to set up a decision tree to produce the classification model to effectively screen pancreatic cancer patients. The results yielded a sensitivity of 100% (20/20), specificity of 97.4% (37/38). The ROC curve was 99.7%. A double-blind test used to challenge the model resulted in a sensitivity of 88.9% and a specificity of 89.5%.

CONCLUSION New serum biomarkers of pancreatic cancer have been identified. The pattern of combined markers provides a powerful and reliable diagnostic method for pancreatic cancer with high sensitivity and specificity.

KEY WORDS: pancreatic cancer, proteomic, SELDI-TOF-MS, classification model, biomarkers.

Copyright © 2008 by Tianjin Medical University Cancer Institute & Hospital and Springer

Introduction

The majority of patients with pancreatic cancer die within one year after a final diagnosis, with a surgical success rate of only 10–20%. The 5-year survival rate is only 3%–4%, a level which is similar to the rate 15 years ago^[1,2]. Following a radical operation, the 5-year survival may reach 19%–41%, if the tumor is ≤ 2 cm, and if the diameter is ≤ 1.0 cm it may attain 67%^[3], suggesting a remarkable difference. So an important way to increase the survival rate after pancreaticoduodenectomy and to improve patient prognosis is early diagnosis^[4]. But at the present time the sensitivity and specificity of clinical diagnostic methods are low, which presents the biggest obstacle for the effective diagnosis and treatment of these patients^[5–9].

Recently, it has become possible to identify new tumor markers to diagnose and monitor development of tumors as proteomics research develops^[10–12]. Surface-enhanced laser desorption and ionization time of flight mass spectrometry (SELDI-TOF-MS) is a new proteomics research technique developed over the past few years^[13–15]. This tech-

nique is better than 2-DIMENSIONAL electrophoresis and other mass-spectrum analysis methods, and it has been extensively applied for tumor marker screening^[16,17] of prostate cancer^[18–20], breast carcinoma^[21] bladder cancer^[22] hepatocellular carcinoma^[23], etc.^[24] However there have been only a few reports related to pancreatic cancer^[25–29]. The purpose of our study was to screen a series of specific marker proteins, in order to establish early diagnostic criteria for pancreatic cancer patients, and to provide more sensitive and more reliable indexes by finger printing of serum proteins from patients with pancreatic cancer and healthy controls.

Materials and Methods

Patients and serum samples

Preoperative blood was collected from 29 pancreatic cancer patients (17 male and 12 female, aged 46–70 years) who had been diagnosed by histopathology analysis. 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 UICC classification. Fifty-seven control blood samples were collected from sex- and age-matched individuals, including 30 healthy controls, 20 cases with chronic pancreatitis and 7 cases with pancreatic cystic tumors, as a composite control group. All sera were obtained according to a standardized sampling and coding protocol. After sample collection, the blood samples were collected in 5 mL BD Vacutainers without anticoagulant, and allowed to clot at room temperature up to 1 h, followed by centrifugation at 4°C for 30 min at 3,000 rpm. The sera were frozen and stored at -80°C for future analysis. All serum samples were obtained between June 2006 and July 2007 in the department of General Surgery, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning province, China. The study was approved by the ethics committee of the China Medical University, and written informed consent was obtained from all individuals from whom serum samples were collected.

Table 1. Clinicopathologic characteristics of the 29 pancreatic cancers.

Stage	Tumor location	
	Head	Body or tail
I	3	4
II	2	2
III	8	6
IV	3	1

Serum pretreatments and magnetic bead binding

Serum samples were pretreated with WCX magnetic beads (Beijing SED Science and Technology, Ltd.). In brief, 10 µl of each serum sample was mixed with 20 µl of U9 solution (9 mol/L urea, 2% CHAPS) in a 0.5

ml centrifuge tube. After incubating for 30 min at 4°C, the samples were diluted 40 fold by adding 370 µl of binding buffer (50 mmol/L sodium acetate, pH 4.0–4.3). Then, 50 µl of WCX magnetic beads (Beijing SED Science and Technology, Ltd.) was added to a PCR tube, and the tube placed in a magnet separator for 1 min, after which the supernatant was removed carefully using a pipette. The magnetic beads were then washed twice with 100 µl of binding buffer. The diluted serum sample (100 µl) was added and mixed with the activated magnetic beads carefully by pipetting up and down several times, followed by incubation for 1 h at 4°C and washing twice with 100 µl binding buffer. Following binding and washing, the bound proteins were eluted from the magnetic beads using 10 µl of 0.5% TFA. Then 5 µl of the eluted sample was diluted 1:2 with 5 µl of SPA (50% CAN + 0.5% TFA). One µl of the resulting mixture was aspirated and spotted onto an 8 spot pre-structured sample chip (Au-chip, Ciphergen Biosystems Inc). After air-drying for approximately 5 min at room temperature, the protein crystal on the chip was examined by SELDI-TOF-MS (Ciphergen, PBS-IIc).

Detection and statistical analysis

Data were collected by averaging 80 laser shots per spot with an intensity setting of 205, and a detector sensitivity setting of 8. The spectra from all samples were processed first using Biomarker Wizard Version 3.1 as follows: selection mass 5901 peak to normalize the dimension, subtraction of the baseline, auto-detect peaks to cluster (First 5, Min Peak 10%, Cluster Mass 0.3%, and Second Pass 2), sample group statistics. The PBS-II (c) Protein Chip reader was calibrated with the 'ALL-IN-ONE' peptide standard (Ciphergen Biosystems). The CV was under 10% for all the selected mass peaks.

Bioinformatics and biostatistics

Fifty-eight serum samples were used to build up a classification model, based on 20 from patients with pancreatic cancer and 38 from age- and sex-matched controls, including 20 healthy controls, 14 cases of chronic pancreatitis and 4 cases of pancreatic cystic tumors, an important disease control group. The whole protein profiling spectra obtained from serum samples were first normalized with total ion current normalization, a feature of Ciphergen's ProteinChip Software 3.1. Peak labeling was performed with the Biomarker Wizard feature of the Software. The data were processed with the *t*-test. A *P* value < 0.01 was considered statistically significant. The intensities of the selected peaks were then transferred to Biomarker Patterns Software Version 5.0 (BPS). The classification model was built up with BPS. A classification tree was set up to divide the training dataset into either the cancer group or the control group through multiple rounds of decision-making.

Blind test

Twenty-eight serum samples (patients with pancreatic cancer, chronic pancreatitis, pancreatic cystic tumors and healthy controls) were analyzed with blind prediction by using the same flowchart and procedure of data collection. The results were compared to clinical data for model evaluation.

Results

We obtained mass spectra for all the serum samples from SELDI-TOF-MS analysis using WCX magnetic beads. As depicted in Fig. 1, up to 150 protein peaks per spot were detected between m/z 1,000 and m/z 200,000, showing that the SELDI technology separation of low molecular weight proteins ($< 10,000$) was particularly effective.

The comparability among different samples showed that the serum profiles from cancer patients and control

individuals were very similar in spite of a few inter sample variations. Therefore, the few variations that consistently differentiated these 2 different groups could be considered as potential disease biomarkers. Here a training serum sample group set, including 20 cases of cancer patients and 38 cases of age- and sex-matched controls were used. All protein profiling spectra were combined and normalized with total ion current normalization, a feature of Ciphergen's ProteinChip Software 3.1. Peak labeling was performed with the Biomarker Wizard feature of the Software. The intensities of selected peaks were then transferred to BPS to build up a classification model.

There were 26 peaks from mass/charge ratio (m/z), in the 2,000~20,000 range where the protein content differences showed a statistical significance ($P \leq 0.01$). Then the intensities of these peaks were transferred to the Biomarker Patterns Software Version 5.0 (BPS) to construct the classification model. Table 2 lists the most

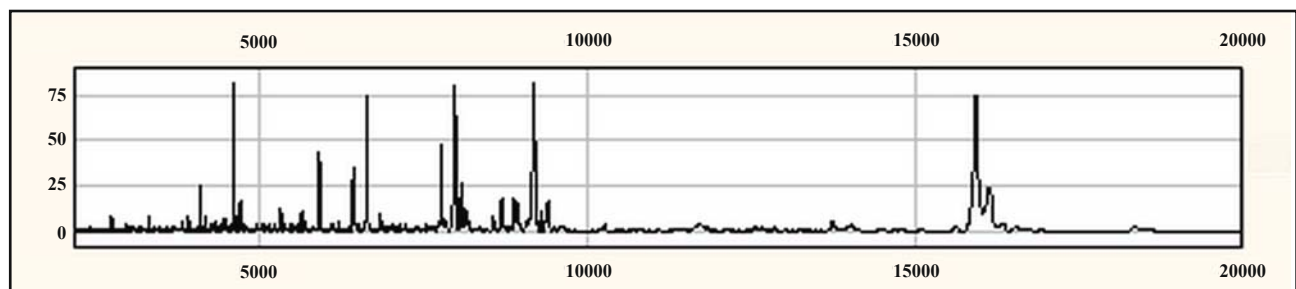


Fig. 1. Representative spectra of a plasma sample in the range of 2,000 to 10,000 m/z obtained using SELDI-TOF-MS.

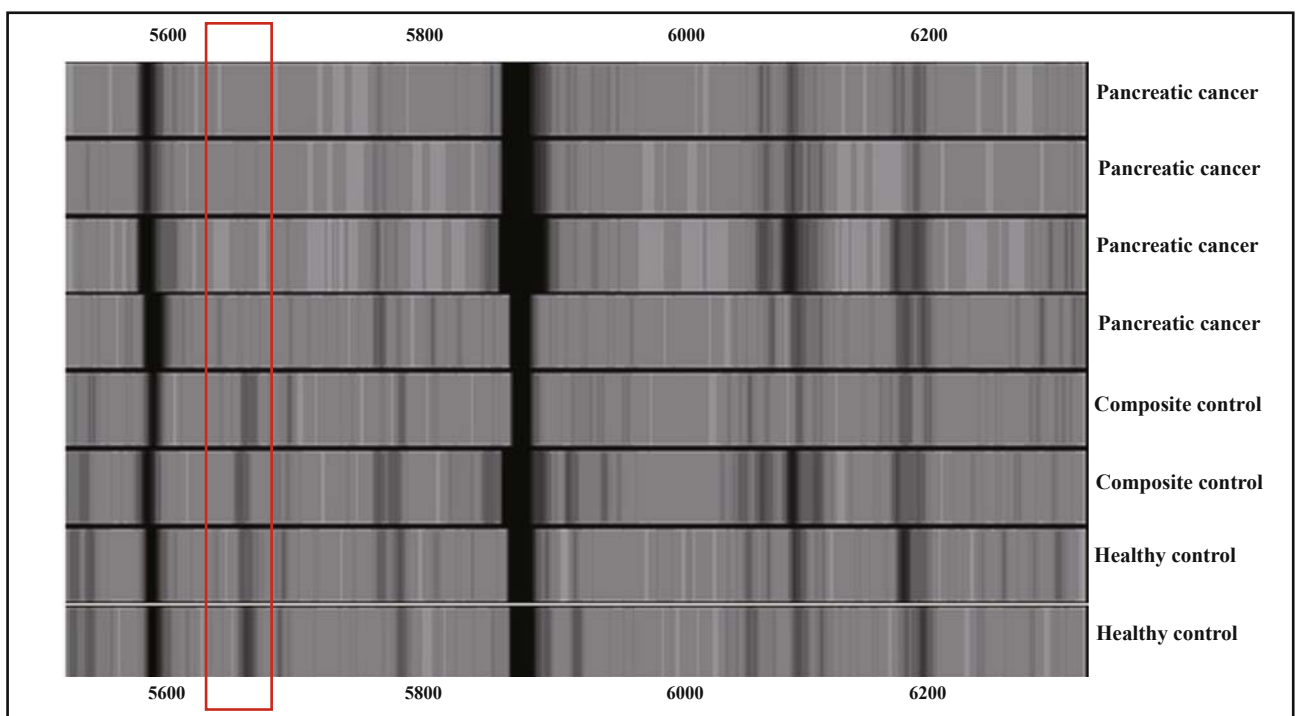


Fig. 2. Differential expression of surface-enhanced laser desorption and ionization peak m/z 5705 in the comparison of pancreatic cancer (top 4) and control (bottom 4) sera. Profile shown in peak display (top) and gel view (bottom) mode (Biomarker Wizard 3.0).

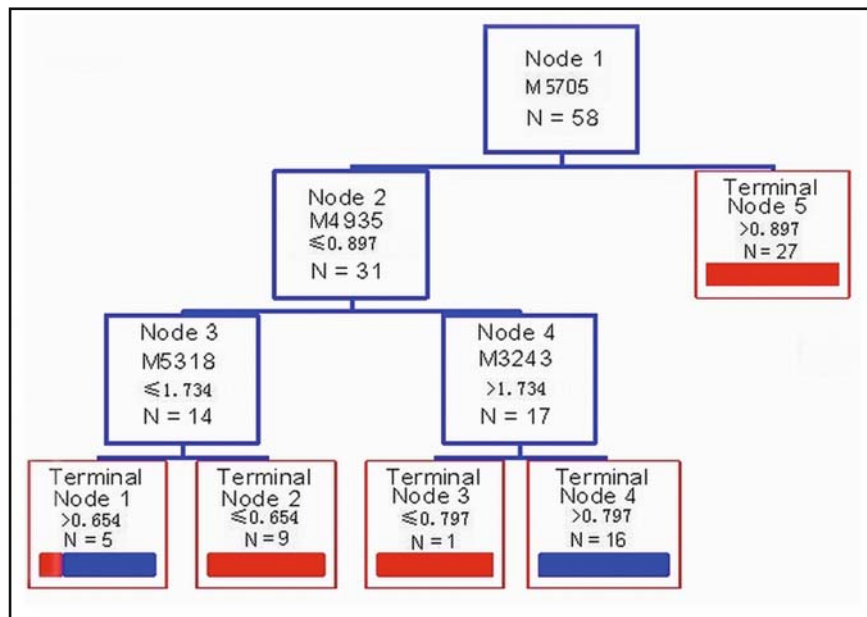


Fig.3. BPS decision tree model. The decision tree detailed the decision-making procedure and sample distribution of the BPS model. Each square is a node, which was labeled by a node number. Blue frames represented parent nodes and red frames represented terminal nodes.

Table 2. Important protein peaks that can be used for comparison of patients with pancreatic cancer and controls selected by BPS.

Peak (m/z)	P	Score*
5705**	9.21E-07	100
11692	1.00E-05	94.37
3319	9.51E-04	72.98
7762	2.04E-03	69.29
3818	2.11E-05	68.31
5479	2.62E-05	68.31
5318**	2.46E-06	32.80
4935**	0.001	30.99
7968	1.00E-04	26.12
2486	3.75E-05	18.17
8073	2.04 E-03	16.21
3243**	3.02E-05	9.25

Peaks were named by their m/z. *, the most important peak was assigned an importance index of 100; **, these were the proteins that were selected to construct the decision tree model.

important peaks determined by BPS when the software built up the decision tree model. To set up the decision model, four peaks with m/z of 5,705 Da, 4,935 Da, 5,318 Da, 3,243 Da, respectively, were chosen (Figs.2 and 3). Based on the model, peak 5,705 was the most important peak and was allocated to Node 1. If the intensity of peak 5,705 was lower than or equal to 0.897, 31 samples went to Node 2, which may be pancreatic cancer patients. Otherwise, samples entered Terminal Node 5, which were diagnosed for health. At Node 2, samples with intensities of peak 4,935 lower than or equal to 1.734 went to node 3, while the other samples

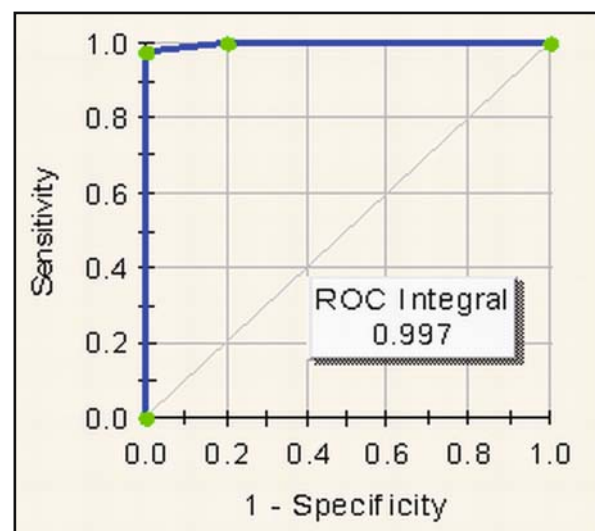


Fig.4. ROC curve of the decision tree model.

of Node 2 entered Node 4. At Node 3, if the intensity of the peak 5,318 was higher than 0.654, five cases of samples went to the left side of Terminal Node 1, which were diagnosed for pancreatic cancer patients, the other samples went to the right side of the Terminal Node 2 in the diagnosis of health. At nodes Node 4, samples with intensities of peak 3,243 lower than or equal to 0.797, one case of samples went to the left side of the Terminal Node 3, which was diagnosed for health, while the other 16 samples entered Terminal Node 4 in the diagnosis of pancreatic cancer patients. When all samples reached Terminal Nodes through this decision-making process, the model yielded a sensitivity of 100.0% (20/20) and a specificity of 97.4% (37/38) with a ROC curve of 99.7% (Fig.4).

To test the model, another 28 serum samples (9 of

pancreatic cancer patients and 19 of controls) were selected to perform the double-blind test. All samples were analyzed for the protein fingerprint with the preset method. The data from each sample were input into the previously built diagnostic model to obtain the prediction value. When the double-blind sample data set was used to challenge the model, the model predicted a sensitivity of 88.9% and a specificity of 89.5%.

Discussion

At present, the most commonly applied serum tumor marker used in the clinic for pancreatic cancer is CA19-9, but its sensitivity and specificity are not very satisfactory. False positives are particularly prevalent in obstructive jaundice, liver cirrhosis, pancreatitis and other malignant tumor patients. The sensitivity of CA19-9 for early pancreatic cancer is low, with a positive rate for detection of a tumor with a diameter of less than 2 cm being only 37.5%, so the value of serum CA19-9 in the diagnosis of early pancreatic cancer alone is limited^[9]. Therefore, new early diagnostic methods with high sensitivity and specificity for pancreatic cancer are needed.

The occurrence and development of pancreatic cancer is a multi-gene, multi-step complex progress. In the view of the function of proteins as the determinants of the phenotype of cells, there must be some subtle but important changes in the level of proteins in the early steps of pancreatic cancer formation. Such changes can not be in one kind of protein, but perhaps in a few or a group of related proteins or polypeptides, or in their metabolites or soluble membrane antigens that are released into the blood^[30,31]. Therefore if we apply some separate biological marker for the early diagnosis of pancreatic cancer, there will inevitably be problems with the sensitivity and specificity. Proteomics research can identify cells, tissues or the whole proteins of the body, can provide information concerning the function of a set of proteins, and can simultaneously reflect the results determined by the genetic characteristics and the external factors^[15].

SELDI-TOF-MS is a new proteomics research method developed in recent years. It can rapidly analyze and compare a protein atlas expressed by a tumor and normal cells, to find some unique combination pattern of protein markers for some type of tumor, and to build a model for early cancer diagnosis. SELDI-TOF-MS reduces the reliance on a single indicator, is convenient, can directly analyze the original biological samples (blood, urine, cerebrospinal fluid, etc.), requires only a small amount of samples, and has the characteristic of high-throughput. In particular it can effectively capture small molecule proteins with low abundance, reflect the real in vivo protein and permit an overall display of protein^[32].

At present, by the clinical use of SELDI technology, more results have been forthcoming in studies of early

detection and early diagnosis for a variety of tumors, including prostate^[18,19], breast^[21] and bladder cancers^[22], hepatocellular carcinoma^[23], etc. In our study we use the SELDI technique, setup various parameters, and analyzed the sera from patients with pancreatic cancer and controls. We found 66 discrepancy protein peaks, in which there were 26 protein content differences showing a statistical significance ($P \leq 0.01$). We classified, distinguished, analyzed and screened 4 protein peaks to produce the most excellent diagnosis, used the Biomarker Wizard, Biomarker Patterns system software, to establish a diagnostic model. By blind test verification, the results were satisfactory, resulting in a sensitivity and specificity which were much higher than the existing kinds of tumor markers^[7].

But if we only use the m/z 5,705 Da in the experimental results, the correct resolution rate was only 71.1%, which precisely illustrates that the SELDI detection technique combined with an artificial intelligence classification algorithm method to select the best effective combination of the protein peaks, rather than evaluate the diagnosis efficacy of a single protein peak. The combination we obtain is not a simple superposition of the peaks whose diagnostic weights are higher, but the optimal selection on the bases of the multi-analysis of the proteins differently expressed. So the overall diagnosis of proteomics is better than the conventional single tumor marker detection.

In order to reduce the confusion with other potential diseases, in the process of collecting specimens we strictly controlled standards, reduced biases to the greatest extent, to ensure the reliability of the experimental results. In this study our sample number was relatively limited, so the clinical significance of the results needs to be confirmed by increasing the number of cases, in particular cases of pancreatitis and pancreatic benign tumors as controls, In addition follow-up studies and verification of the clinical condition of the patients should be conducted, to reduce possible misdiagnosis or missed diagnosis. An objective evaluation for the application of SELDI technology needs to be made as well as to do research to provide new ideas and methods for the clinical molecular diagnosis of pancreatic cancer.

Acknowledgment

We thank our medical staff for help in collecting and processing blood samples and in documenting clinical data. We greatly appreciate the help from the Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, for providing access to the SELDI-TOF-MS laboratory and sharing of their expertise and for helpful comments.

References

- 1 Wray CJ, Ahmad SA, Matthews JB, et al. Surgery for pancreatic cancer: recent controversies and current practice. *Gastroenterology* 2005; 128: 1626–1641.
- 2 Li D, Xie K, Wolff R, et al. Pancreatic cancer. *Lancet* 2004; 363: 1049–1057.
- 3 Hawes RH, Xiong Q, Waxman I, et al. A multispecialty approach to the diagnosis and management of pancreatic cancer. *Am J Gastroenterol* 2000; 95: 17–31.
- 4 Ishikawa O, Ohigashi H, Imaoka S, et al. Minute carcinoma of the pancreas measuring 1 cm or less in diameter—collective review of Japanese case reports. *Hepatogastroenterology* 1999; 46: 8–15.
- 5 Nakao A, Fujii T, Sugimoto H, et al. Oncological problems in pancreatic cancer surgery. *World J Gastroenterol* 2006; 12: 4466–4472.
- 6 Kalra MK, Maher MM, Mueller PR, et al. State-of-the-art imaging of pancreatic neoplasms. *Br J Radiol.* 2003; 76: 857–865.
- 7 Carpelan-Holmström M, Louhimo J, Stenman UH, et al. CEA, CA 19-9 and CA 72-4 improve the diagnostic accuracy in gastrointestinal cancers. *Anticancer Res.* 2002; 22: 2311–2316.
- 8 Zhao XY, Yu SY, Da SP, et al. A clinical evaluation of serological diagnosis for pancreatic cancer. *World J Gastroenterol.* 1998; 4:147–149.
- 9 Kim JE, Lee KT, Lee JK, et al. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol* 2004; 19: 182–186.
- 10 Liotta LA, Kohn EC. The microenvironment of the tumour–host interface. *Nature* 2001; 411: 375–9.
- 11 Hanash SM. Global profiling of gene expression in cancer using genomics and proteomics. *Curr Opin Mol Ther* 2001; 3: 538–545.
- 12 Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. *Nature* 2003; 425:905.
- 13 Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization–time of flight–mass spectrometry. *Electrophoresis* 2000; 21:1164–1177.
- 14 Zhu H, Snyder M. Protein chip technology. *Curr Opin Chem Biol* 2003; 7: 55–63.
- 15 Issaq HJ, Veenstra TD, Conrads TP, et al. The SELDI–TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Comm* 2002; 292:587–592.
- 16 Conrads TP, Hood BL, Issaq HJ, et al. Proteomic patterns as a diagnostic tool for early-stage cancer: a review of its progress to a clinically relevant tool. *Mol Diagn* 2004; 8:77–85.
- 17 Petricoin EF, Liotta LA. SELDI–TOF–MS based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol* 2004; 15: 24–30.
- 18 Petricoin EF 3rd, Ornstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002; 94: 1576–1578.
- 19 Bañez LL, Prasanna P, Sun L, et al. Diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol* 2003; 170: 442–446.
- 20 Malik G, Ward MD, Gupta SK, et al. Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer. *Clin Cancer Res* 2005; 11: 1073–1085.
- 21 Somiari RI, Somiari S, Russell S, et al. Proteomics of breast carcinoma. *J Chromatogr B Analyt Technol Biomed Life Sci* 2005; 815: 215–225.
- 22 Langbein S, Lehmann J, Harder A, et al. Protein profiling of bladder cancer using the 2D–PAGE and SELDI–TOF–MS technique. *Technol Cancer Res Treat* 2006; 5: 67–72.
- 23 Paradis V, Degos F, Dargère D, et al. Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. *Hepatology* 2005; 41: 40–47.
- 24 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: 12343–12348.
- 25 Rosty C, Goggins M. Identification of differentially expressed proteins in pancreatic cancer using a global proteomic approach. *Methods Mol Med* 2005; 103: 189–197.
- 26 Bhattacharyya S, Siegel ER, Petersen GM, et al. Diagnosis of pancreatic cancer using serum proteomic profiling. *Neoplasia* 2004; 6: 674–686.
- 27 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: 860–868.
- 28 Honda K, Hayashida Y, Umaki T, et al. Possible detection of pancreatic cancer by plasma protein profiling. *Cancer Res* 2005; 65: 10613–10622.
- 29 Ehmann M, Felix K, Hartmann D, et al. Identification of potential markers for the detection of pancreatic cancer through comparative serum protein expression profiling. *Pancreas* 2007; 34: 205–214.
- 30 Srinivas PR, Verma M, Zhao Y, et al. Proteomics for cancer biomarker discovery. *Clin Chem* 2002; 48: 1160–1169.
- 31 Hanash S. Disease proteomics. *Nature* 2003; 422: 226–232.
- 32 Rodland KD. Proteomics and cancer diagnosis: the potential of mass spectrometry. *Clin Biochem* 2004; 37: 579–583.