Journal of Cellular Biochemistry 103:740–752 (2008)

Toward the Proteomic Identification of Biomarkers for the Prediction of HBV Related Hepatocellular Carcinoma

Qing-Yu He,[1,2]* Rui Zhu,[3] Ting Lei,[1,7] Mandy Y.M. Ng,[4] John M. Luk,[3] Pak Sham,[5] George K.K. Lau,[4] and Jen-Fu Chiu[6]**

[1]Department of Chemistry and Open Laboratory of Chemical Biology, University of Hong Kong, Hong Kong, China
[2]Institute of Life and Health Engineering, Jinan University, Guangzhou 510632, China
[3]Department of Surgery, University of Hong Kong, Hong Kong, China
[4]Department of Medicine, University of Hong Kong, Hong Kong, China
[5]Department of Psychiatry, University of Hong Kong, Hong Kong, China
[6]Department of Anatomy, University of Hong Kong, Hong Kong, China
[7]Department of Pathology, College of Medicine, Xi'an Jiaotong University, Xi'an 710061, China

Abstract Early detection is a key step for effective intervention of hepatocellular carcinoma (HCC), the lack of sensitive and specific biomarkers is a major reason for the high rate of HCC-related mortality. This report described an integrated strategy by combining SELDI-ProteinChip, sophisticated algorithm analysis, acetonitrile (ACN) pre-treatment and two-dimensional electrophoresis (2DE)–peptide mass fingerprinting (PMF) techniques to identify serological markers for the prediction of HBV-related HCC. Proteomic profiling of three groups of serum specimens from HBV-related HCC (50 cases), HBV infection (45 cases), and normal subjects (30 cases) was conducted by using SELDI-ProteinChip system and the resulting different protein peaks were subjected to stepwise statistical analyses. Three most discriminatory peaks at 5890, 11615, and 11724 Da, respectively, were screened out from the statistical algorithm and a predictive model based on the three peaks was constructed and tested using the newly enrolled serum samples. 2DE was applied to separate and compare the serum samples that were pre-treated by ACN precipitation. The protein spots obviously intensified in HCC sera in the 2DE region of 12 kDa were identified by PMF to be serum SAA, which was validated by SELDI-TOF spectra of HCC sera after immunoprecipitation using anti-SAA antibody and by Western blot experiments. Given the fact that SAA is not a specific biomarker, further attempt is being made to identify the other two most discriminatory peaks to realize the possibility of using the predictive model for HCC surveillance and prediction. J. Cell. Biochem. 103: 740–752, 2008.

Key words: HBV; HCC; tumor biomarkers; serum markers; protein profiling; SELDI-TOF ProteinChip; 2DE; MALDI-TOF MS

Hepatocellular Carcinoma (HCC) is one of the most common malignant tumors in the world. Chronic infections of hepatitis B & C virus (HBV & HCV) are two major risk factors leading to HCC incidence. Surgical resection is still the major potentially curative form of therapy, however only 10% of patients were at operable stages of disease at presentation. This is due in part to the lack of reliable tools for early diagnosis and thus most of patients were at a late stage at the time of clinical presentation. Currently, liver biopsy is the only reliable method for definite diagnosis but it can be associated with expense and risk of patient injury. The widely used serological tumor marker for HCC, α-fetoprotein (AFP) lacks specificity and sensitivity [Johnson, 1999; Johnson, 2001].

Abbreviations used: HBV, hepatitis B virus; HCV, hepatitis C virus; HCC, hepatocellular carcinoma; SELDI, surface enhanced laser desorption/ionization; MALDI, matrix-assisted laser desorption/ionization; TOF, time of flight; MS, mass spectrometry; MW, molecular weight; PMF, peptide mass fingerprinting; SAA, serum amyloid A; ACN, acetonitrile; IP, immunoprecipitation.

Grant sponsor: Hong Kong Research Grants Council; Grant numbers: HKU 7227/02M, HKU 7512/05M, HKU 7218/02M, HKU 7395/03M.

*Correspondence to: Prof. Qing-Yu He, Institute of Life and Health Engineering, Jinan University, Guangzhou 510632, China. E-mail: tqyhe@jnu.edu.cn

**Correspondence to: Prof. Jen-Fu Chiu, Department of Anatomy, University of Hong Kong, Pokfulam, Hong Kong, China. E-mail: jfchiu@hku.hk

Received 18 April 2007; Accepted 9 May 2007

DOI 10.1002/jcb.21443

This presents a great need of continued effort to identify better and more accurate biomarkers for designing novel diagnostic tools and therapeutic strategies for HCC.

As an emerging powerful biological research technology, proteomics has recently become an important tool in the identification of biomarkers and disease targets [He and Chiu, 2003; Petricoin and Liotta, 2003; Zhu et al., 2003]. One apparent advantage of proteomics is its ability to simultaneously inspect the whole proteome or sub-proteomes so that correlated proteins altered in expression and modification corresponding to a disease condition can be identified. This makes it possible to combine several protein markers together to form an index with higher sensitivity and specificity in the detection and monitoring of the disease. In recent years, proteomics has been extensively used to identify biomarkers for HCC with both HBV [Le Naour et al., 2002; Poon et al., 2003, 2005; Comunale et al., 2004; Li et al., 2004, 2005; Zhu et al., 2004; Feng et al., 2005; Tan and Chen, 2005] and HCV [Le Naour et al., 2002; Takashima et al., 2003, 2005; Paradis et al., 2005; Schwegler et al., 2005; Lee et al., 2006] infections and promising progresses have been reported. In particular, serum samples were directly used as a source of identifying biomarkers for the great potential of clinical application for early detection of HCC.

Protein profiling based on two-dimensional electrophoresis (2DE) or SELDI-ProteinChip system is the major strategy in proteomics for biomarker discovery. 2DE can separate thousands of proteins in a single experiment and then potential biomarkers can be identified by comparing the protein profiles between control and disease samples. Protein spots from 2DE are usually identified through mass spectroscopy (MS)-based peptide-mass fingerprinting (PMF) or MS/MS partial sequencing. However, protein separation by 2DE cannot be well performed for those proteins with molecular weights (MW) less than 6 kDa and isoelectric points (p*I*) less than 3 or higher than 11. Particularly, high abundant proteins in serum can shadow low abundant ones in 2DE and thus sample pre-treatment is usually required. As a complementary tool, ProteinChip SELDI-TOF technology is much useful in analyzing proteins with low MWs covering the entire p*I* range. Other advantages of SELDI-ProteinChip system include its high sensitivity and straightforwardness especially in examining body fluids such as serum samples. However, this system mainly aims at providing differences between specimens in low MW proteins and it is difficult for it to identify proteins directly.

In the present study, we focused on identifying potential biomarkers in the serum samples of HBV and HCC by employing an integrated strategy combining both SELDI-ProteinChip and 2DE to take the complementary advantages. Serum samples (125 cases in training set) were first profiled in SELDI-ProteinChip system and the resulting peak differences were analyzed by sophisticated statistical programs. The significant peaks were correlated to the 2DE gel spots generated by the protein profiling of acetonitrile (ACN) -pretreated sera, followed by the sequential protein identification through PMF and confirmation with immunoprecipitation and Western blotting. Differing from other similar studies that were related to liver fibrosis and cirrhosis [Poon et al., 2003, 2005; Zhu et al., 2004], we included HBV-infection samples that are at early stages of the disease progression to HCC. A predictive model containing three most discriminatory protein peaks was constructed to predict the probability of HCC incidence in test set (39 samples) with high sensitivity and specificity. The potential application of these biomarkers and the predictive model is discussed.

MATERIALS AND METHODS

Patient Materials

With patients' consent, we collected blood specimens from HBV and HCC patients at the Queen Mary Hospital of The University of Hong Kong. The study was approved by the research ethics committee of the University of Hong Kong. A total of 164 serum specimens were divided into training and test sets (Table I). The training set contains 125 samples including 45 HBV which are affirmed of absent from

TABLE I. Information of Subjects

Training group	Age (years)	Male/female
Normal	48 ± 9	16/14
HBV	45 ± 6	28/17
HBV-HCC	52 ± 12	30/20
Test group	**Age (years)**	**Male/female**
Normal	50 ± 10	7/7
HBV	53 ± 8	5/3
HBV-HCC	56 ± 15	10/7

malignancy or cirrhosis, 50 HBV-positive HCC, and 30 normal control sera. The test set contains 39 samples including 8 HBV, 17 HCC, and 14 normal controls. All serum samples were distributed into 50 μl aliquots and were stored at −80°C before assay.

SELDI-TOF ProteinChip Analysis of Serum Samples

Five types of protein chips including IMAC3, WCX2, SAX2, H50, and NP20 from Ciphergen Biosystems (Frement, CA) were used in the protein profiling of the serum samples. These protein chips have different surface affinity properties with IMAC3 for metal-binding proteins, WCX2 (weak cation exchanger) for positive-charge molecules, SAX2 (strong anion exchanger) for negative charge molecules, H50 for hydrophobic peptides, and NP20 (normal phase) for general binding of proteins. Based on the experiments, the IMAC3-Cu metal binding chip was found to give the most signals and thus its detection results were chosen for the sequential statistical analysis.

The experimental protocol was similar to that described previously [He et al., 2005]. Each serum fraction was thawed and pre-treated before loading on the protein chips. Serum sample of 3 μl was diluted into 6 μl of U9 solution (9 mol/L urea, 2% CHAPS, 50 mmol/L Tris–HCl, pH 9.0), incubated for 30 min, and mixed every 5 min in between when incubating on ice or shaken gently and continually in a cold room. After vortexing, each sample was diluted in 1:36 with corresponding binding buffers. Then 10 μl mixed serum fraction was spotted onto ProteinChip arrays with variety of chemical surfaces in duplicate.

Mass spectrometric analysis was performed by SELDI-TOF mass spectrometry in a PBS-II ProteinChip reader (Ciphergen Biosystems) according to an automated data collection protocol. This includes an average of 255 laser shots to each spot with a laser intensity of 240 and 265, respectively, depending on the measurement region (low = 2–20 kDa and high = 20–200 kDa, respectively). The detector was run at a sensitivity of 9. Spectral analysis was carried out using the ProteinChip software version 3.1 (Ciphergen Biosystems). All mass spectra were normalized to have the same total ion current. Preliminary cluster analysis of the detected signals and the determination of respective *P*-values were carried out with the Biomarker Wizard Program (Version 3.0; Ciphergen Biosystems). The significant protein peaks with their normalized intensities were subjected to the following statistical analysis.

Statistical Analysis

Three selection steps were performed to identify the most discriminatory set of protein peaks from the training set. Firstly a hierarchical clustering program, CLUSTAG (http://hkumath.hku.hk/web/link/CLUSTAG/CLUSTAG.html) was used to merge the protein peaks into a number of clusters, such that each cluster contains a "representative" peak that has squared correlation $R^2 \geq 0.8$ with all other members of the same cluster. Secondly, a series of ordinal logistic regression was performed on each protein peak, with disease status as the dependent variable and the protein peak as the predictor variable. Only protein peaks with $P < 0.001$ were retained in their clusters. Finally, a series of forward stepwise ordinal logistic regression, with a 0.001 probability of entry and 0.005 probability of removal, was performed on the clusters in order to screen out the most discriminatory protein peaks. All the logistic regression analyses were carried out by Stata version 7 (Stata Corporation, College Station, TX). A predictive model, based on the most discriminatory protein peaks, was used to discriminate the disease status of subjects (see Table II). The performance of this predictive model was assessed by its sensitivity and

TABLE II. Predictive Model Based on the Discriminatory Analysis

Probability of being a HCC patient $= 1/(1 + \exp(y - \text{delta}_i))$
Probability of carrying HBV $= 1/(1 + \exp(y - \text{delta}_{i+1})) - 1/(1 + \exp(y - \text{delta}_i))$
Probability of being a control $= 1 - 1/(1 + \exp(y - \text{delta}_{i+1}))$

where,
$y = \sum_{j=1\ldots n} S_j\beta_j$
S_j = the most discriminatory protein peaks selected from the training set
β_j = the beta-coefficients of S_j
n = number of the most discriminatory protein peaks selected from the training set
delta_i = the cut values of the ordinal logistic regression on the training set

specificity in identifying HCC and HBV subjects in the training and test sets.

Acetonitrile Precipitation

Briefly, the protocol involved adding 200 μl of serum to 400 μl of high-performance liquid chromatography (HPLC) grade ACN, votexing vigorously for 5 s, and allowing standing at room temperature for 30 min. Samples were then spun for 10 min at 14,000*g*, room temperature. After 1D SDS–PAGE confirmation for the effectiveness of ACN precipitation to remove high molecular weight proteins, an aliquot of supernatant (550 μl) was then lyophilized to 200 μl in a SpeedVac vacuum centrifuge (Thermo Savant) for 2DE analysis. The molecular marker used in the SDS–PAGE gel was purchased from Bio-Rad.

2DE Protein Separation

2DE gel electrophoresis was carried out with IPGphor II isoelectric focusing (IEF) Unit and Hoefer SE 600 Ruby electrophoresis unit (Amersham Biosciences) according to the method reported previously [He et al., 2004]. Briefly, 60 μg of proteins from ACN precipitated serum were mixed up to 250 μl of rehydration solution containing 8 M Urea, 4% CHAPS, 1 mM PMSF, 20 mM DTT, and 0.5% IPG buffer. The rehydration step was carried out with pre-cast 13 cm IPG strips for more than 10 h at low voltage of 30 V. IEF was run following a stepwise voltage increase procedure: 500 and 1,000 V for 1 h each and 5,000–8,000 V for about 10 h with a total of 64 kV/h. After IEF, the strips were subjected to two-step equilibration in equilibration buffers (6 M urea, 30% glycerol, 2% SDS, and 50 mM Tris–HCl pH 6.8) with 1% DTT for the first step, and 2.5% iodoacetamide for the second step. The strips were then transferred onto the 2D SDS–PAGE (13 cm) that was run on 1.0 mm thick 12.5% polyacrylamide gels at 10°C. Protein samples were run in pair side by side and triplicate electrophoresis was performed to ensure reproducibility. All gels were visualized by silver staining [He et al., 2003].

Image Analysis, In-Gel Tryptic Digestion and Protein Identification

As described previously [He et al., 2003], the stained gels were scanned and the images were comparatively analyzed using ImageMaster 2D Elite software (Amersham Biosciences). The protein spots with altered expression were cut off from the gels and subjected to tryptic in-gel digestion. The resulting peptide mixtures were applied to a Voyager DE-STR MALDI-TOF mass spectrometer to generate peptide mass spectra. PMF was performed by matching these mass spectra to the NCBInr protein database using MS-Fit (http://prospector.ucsf.edu/) to produce protein IDs. Strict criteria were followed to ensure accurate matching [He et al., 2003]. Species search was limited to *Homo sapiens*.

Depletion of Serum Amyloid A (SAA) by Immunoprecipitation (IP)

HCC serum sample (10 μl) was diluted into 490 μl IMAC3 buffer. After pre-clearing with protein-A agarose, 50 μl diluted sample was saved as the pre-IP control. The left 450 μl diluted serum was incubated with 3 μg anti-SAA antibody at 4°C with rotation overnight. The resulted immunocomplex was captured by adding 50 μl protein-A agarose and gently rotating at 4°C for 3 h. After centrifugation at 14,000*g* for 10 s, the supernatant was directly used for analysis by IMAC3-Cu protein chip and SELDI-TOF MS according to the protocol described previously.

Western Blotting

Serum proteins (10 μg) were separated on 12% SDS–PAGE, and then transferred onto a PVDF membrane. After incubating with blocking solution (TBS-T containing 5% non-fat milk) at 4°C overnight, the membrane was probed with antihuman SAA sheep polyclonal antibody (Calbiochem) at a 1:3,000 dilution for 2 h at room temperature. After washing three times of 15 min each with TBS-T, antigoat IgG antibody was added for 1 h at room temperature, and the antigen–antibody interaction was finally detected by ECL detection kit (Amersham-Pharmacia Biotech) and then exposed to the X-ray film. The molecular marker used here was obtained from Bio-Rad Company.

RESULTS

SELDI-TOF ProteinChip Profiling

As an initial screening to decide which protein chip is suitable for the serological protein analysis, five types of protein chips, H50, IMAC3, NP20, SAX2, and WCX2, were used for the protein profiling of the serum samples.

IMAC3-Cu was found to produce most significant protein peaks and thus was selected for this study. Figure 1 shows representative protein profiles from 1.5 to 20 kDa detected by the IMAC3 chips for the HCC, HBV, and normal sera side by side. The obvious differences in protein profiles among these three groups of serum samples can be easily observed even by naked eyes. ProteinChip software version 3.1 (Ciphergen Biosystems) that is pre-installed in the SELDI system was used to preliminarily analyze the protein profiles. With the selective criteria at $P \leq 0.05$ and fold-difference of peak intensity ≥ 2.0, 135 protein peaks in the training set of samples were found to be significant. These 135 signals were included into the following sophisticated discriminatory statistical analysis.

Discriminatory Statistical Analysis

The discriminatory statistical analysis was carried out by a step-wise selection approach to select the most discriminatory peaks for constructing a predictive model. The detail of the screening process is summarized in Scheme 1. In the first step, a total of 104 clusters were formed by CLUSTAG with a criterion of $R^2 \geq 0.8$. The distribution frequency of the protein peaks within clusters is shown in Table III. Secondly, a series of ordinal logistic regression was performed on each protein peak. A total of 65 protein peaks failed to reach the significance level of $P \leq 0.001$, and this resulted in a total of 49 clusters remaining. Among these clusters, 42 contained just one protein peak and these were grouped together to form one "cluster." Each of the eight clusters was subjected to a stepwise ordinal logistic regression, and this led to removal of all but 10 protein peaks. Table IV gives out the specific information of these 10 peaks. Finally, these 10 peaks were subjected to a stepwise ordinal logistic regression, in which only 3 protein peaks were retained (highlighted in Table IV). This set of three peaks, which located at 5890, 11615, and 11724 Da, respectively, was regarded as the most discriminatory protein peaks. The corresponding beta-coefficients, standard errors, and pseudo R^2 are shown in Table V. Figure 2 is the diagrams showing the serum levels of these three proteins that gradually increase from normal control to HBV infection and then to HCC.

The predictive model, after substituting the corresponding parameters, is shown in Scheme 2. The performance of the predictive model on both the training and test sets are shown in Table VI. For the training set, the model has above 90% sensitivity and specificity to predict the HCC subjects. The performance to predict subjects carrying HBV was reasonable, with 76% sensitivity and 85% specificity. For the test set, the predicative model has 100% sensitivity of identifying both the HCC subjects and subjects with HBV infection. The specificity is 100% for the HCC group and 87% for the HBV group.

Serum Pre-treatment by Acetonitrile Precipitation

We then attempt by using 2DE/MALDI-TOF MS technology to identify the last three proteins selected by the discriminatory statistical analysis described above. Since the high abundant proteins such as albumin and transferrin in serum shadow the detection of low abundant proteins in 2DE and the proteins in the low

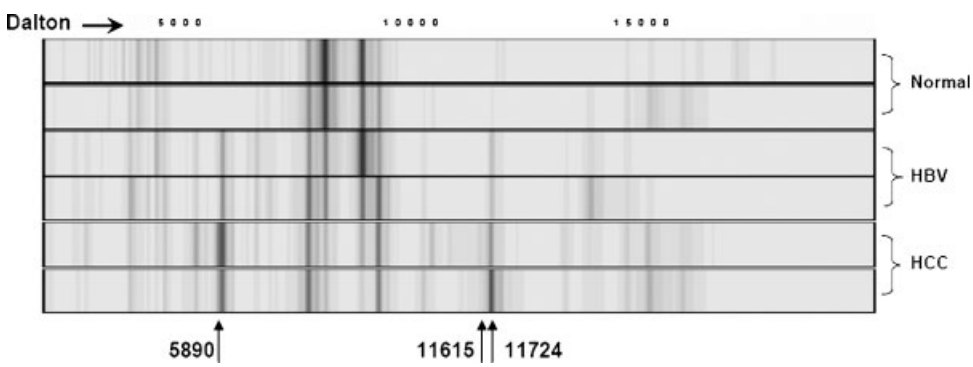
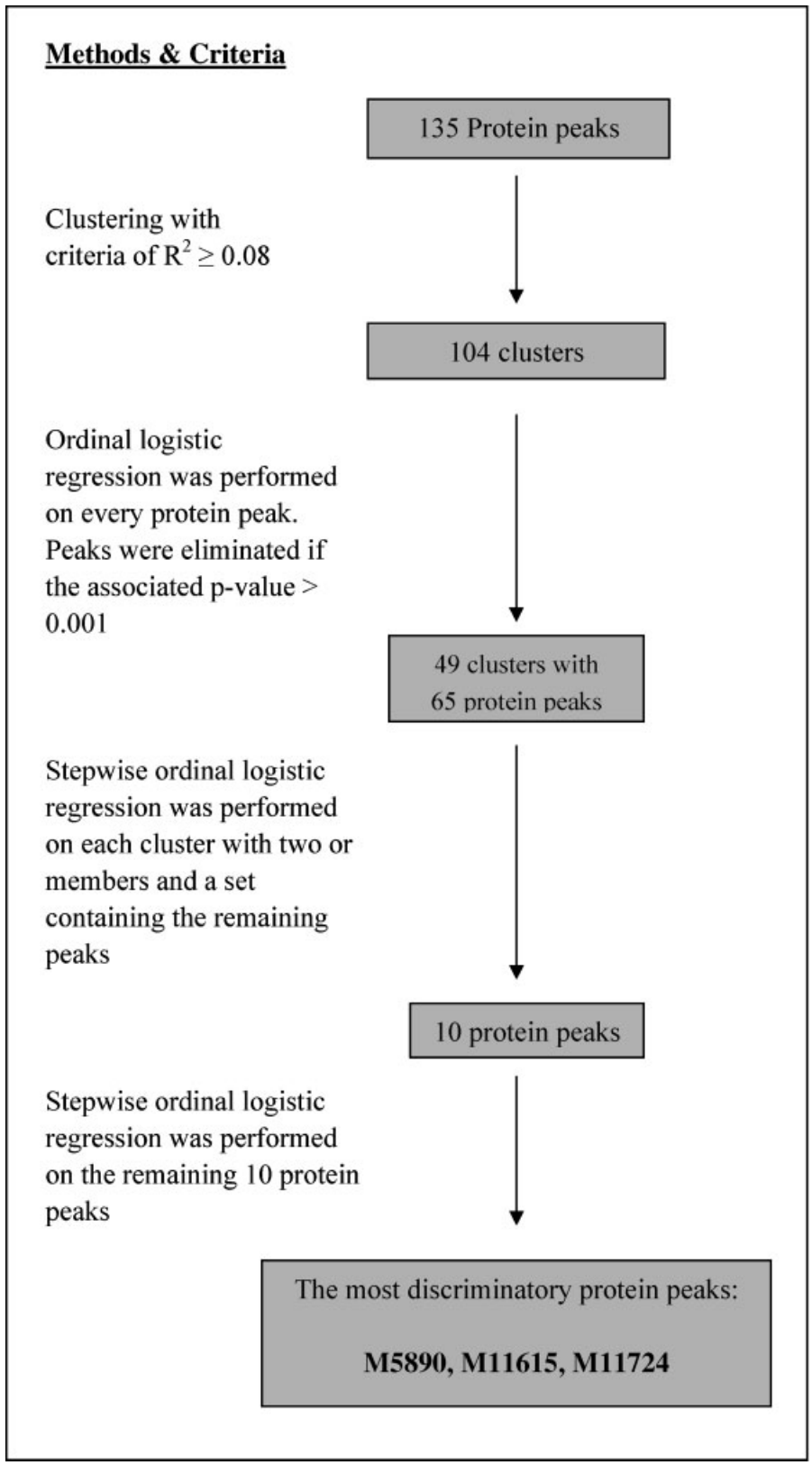


Fig. 1. IMAC3-Cu chip-SELDI spectral comparison among normal, HBV, and HCC serum samples. Three most significant differences are shown by arrows with molecular weight (MW) of 5890, 11615, and 11724 Da, respectively.



Scheme. 1. Discriminatory protein peaks selection steps.

TABLE III. Distribution of Protein Peaks Within Clusters That Formed by CLUSTAG

Number of protein peaks within a cluster	Frequency
1	92
2	5
3	3
4	2
7	1
9	1

molecular weight are usually in the low content region in serum, ACN was used to precipitate the high abundant proteins with high MW. Figure 3 shows 1D PAGE gel images for serum samples before and after treatment with ACN precipitation. Obviously, when compared to the whole serum sample (Lane 2), most of high abundant proteins with high MW were removed and the low abundant proteins with low MW were detected (Lane 3). These samples after ACN treatment were used for the following 2DE separation and comparison.

2DE and Protein Identification

Figure 4 shows 2DE images for the normal control (A) and HCC (B) pooled serum samples after ACN treatment. In the region of low molecular weight ~12 kDa, a train of protein spots at p*I* 5–8 were found to be significantly over-expressed in HCC samples. These differences may not be distinguished from the comparison with normal 1D gels since the similar molecular weight proteins in around p*I* = 4 shade the differentiation. These altered protein spots were then subjected to in-gel digestion and protein identification by PMF and they were found to be three isoforms of SAA. Figure 5 shows a representative MS spectrum of the SAA protein samples. Table VII lists corresponding protein IDs, parameters and residues of identified peptides in peptide fingerprinting matching for protein identification. Spots 1 and 2 are SAA1 (13,532 Da) with AAs 1–19 truncated and spot 3 may be SAA2 (10,167 Da).

Immunoprecipitation with Anti-SAA Antibody

Given the facts that the SAA spots have very close MW to two of the three most discriminatory protein peaks derived from the statistical analysis, the over-expression protein peaks in HCC from the ProteinChip profiling are probably SAA isoforms. To validate this speculation, HCC serum samples were treated with anti-SAA antibody to remove SAA by IP and then subjected to ProteinChip-SELDI TOF MS analysis. Figure 6 displays the ProteinChip spectra in the region around 12 kDa for the HCC sera before and after IP. The peak at 11,615 Da disappeared after SAA depletion with IP, indicating that this signal originates from SAA.

Western Blot Analysis

Western blotting was performed to confirm the SAA alterations in serum samples. From the three groups of serum specimens, 6 samples each were randomly selected and subjected to the immunochemical blotting with antihuman SAA sheep polyclonal antibody. Figure 7 displays the results for the immunochemical interaction. With the same amount of protein loading, strong interaction in HCC, weak signals in HBV and almost no signals in normal sera were observed. Although with some variations, these results are fully consistent with the data from SELDI profiling.

DISCUSSION

Biomarker identification by means of proteomics technology is a hot topic of research.

TABLE IV. Ten Protein Peak Profiles Screened Out by Three Step-Wise Statistical Analyses

M/Z (Da)	*P*	HCC (Mean)	HCC (SD)	HBV (Mean)	HBV (SD)	Ctrl (Mean)	Ctrl (SD)	Fold (HCC/Ctrl)
2,959	1.41E – 09	33.33	18.64	12.74	11.26	8.93	10.08	+3.7
4,574	1.30E – 09	33.94	16.21	21.05	13.76	14.35	7.76	+2.4
5,550	5.00E – 10	12.84	7.94	10.13	8.27	3.76	2.55	+3.4
5,808	4.83E – 09	39.58	21.64	13.04	17.41	6.57	4.63	+6.0
5,890	6.10E – 10	35.87	18.16	11.91	11.32	7.45	9.13	+4.8
8,337	2.90E – 09	6.32	3.78	16.82	12.96	21.00	15.58	−3.3
9,139	1.40E – 09	34.16	14.51	32.07	14.46	15.75	9.18	+2.1
11,081	3.20E – 05	4.37	1.47	2.00	2.06	1.44	0.95	+3.0
11,615	1.80E – 10	4.00	2.51	1.67	0.89	0.89	0.57	+4.5
11,724	2.00E – 08	33.94	16.21	21.05	13.76	14.35	7.76	+2.4

TABLE V. Three Most Discriminatory Protein Peaks Screened Out Through Entire Analytical Algorithm

Protein peaks	Beta-coefficients	Standard error	Pseudo R^2
M5890	−0.2613	0.0425	0.4892
M11615	−2.8525	0.6806	0.0946
M11724	0.1334	0.0456	0.0377

Identification of new markers for HCC by protein profiling has been a major research subject driven by the facts that there are still lacking sensitive and specific biomarkers for HCC surveillance and diagnosis and that protein alterations corresponding to the liver carcinogenesis should be able to be identified by proteomic profiling. Extensive efforts to discover novel HCC biomarkers have been made in these several years by profiling proteomes in liver tissues [Takashima et al., 2003, 2005; Li et al., 2004, 2005; Melle et al., 2004; Lee et al., 2005; Luk et al., 2006], cells [Tan and Chen, 2005] and sera [Le Naour et al., 2002; Poon et al., 2003; Feng et al., 2005; Kawakami et al., 2005; Paradis et al., 2005; Schwegler et al., 2005; Lee et al., 2006], and by using animal models [Block et al., 2005]. Other investigations focused on biomarker identification for HBV infection [He et al., 2003; Comunale et al., 2004] or liver fibrosis and cirrhosis that may lead to HCC [Xu et al., 2004; Zhu et al., 2004; Poon et al., 2005]. Among these studies, proteomic platforms either based on 2DE [Le Naour et al., 2002; He et al., 2003; Takashima et al., 2003, 2005; Comunale et al., 2004; Li et al., 2004, 2005; Block et al., 2005; Feng et al., 2005; Lee et al., 2005; Tan and Chen, 2005; Luk et al., 2006] or ProteinChip [Poon et al., 2003, 2005; Melle et al., 2004; Xu et al., 2004; Zhu et al., 2004; Paradis et al., 2005; Schwegler et al., 2005] have been employed individually or in combination [Lee et al., 2006]. In particular, several recent publications have reported serological profiling using SELDI-ProteinChip system to compose proteomic fingerprints or models for detecting or predicting HCC with high sensitivity and specificity [Poon et al., 2003, 2005; Paradis et al., 2005; Schwegler et al., 2005], showing the potential of building a noninvasive and high-throughput surveillance strategy by proteomics for monitoring HCC progression.

For a few years, we have been attempting to identify biomarkers for HBV-related liver diseases by using proteomics techniques. Our earlier proteomic profiling of HBV-infected serum in 2DE resulted in the identification of several proteins that changed their expression levels and patterns compared to those in normal serum [He et al., 2003]. Nevertheless, these proteins are mostly high-abundant acute-phase proteins due to the limited separation ability of 2DE for whole serum samples. In the present study, we combined both 2DE and SELDI-ProteinChip techniques to take the complementary advantages and included a HCC sample group aiming at searching for more specific serological markers to predict HBV-induced HCC incidence. Similar strategy that used SELDI-TOF MS and 2DE in combination has been successfully applied to identify a candidate

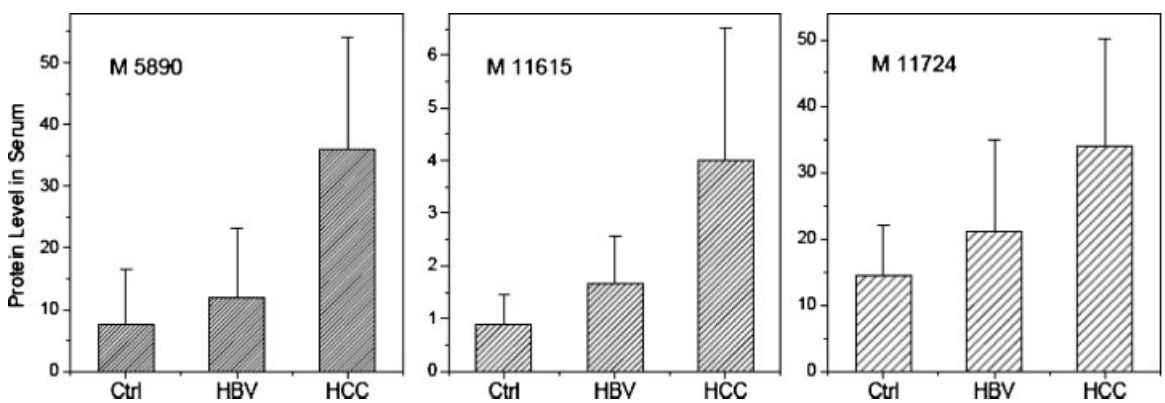


Fig. 2. Protein levels of the three most discriminatory peaks in serum of normal, HBV, and HCC subjects.

Probability of carrying HCC = 1/(1+exp(y – (- 8.5228)))

Probability of carrying HBV = 1/(1+exp(y – (- 3.1071))) - 1/(1+exp(y – (- 8.5228)))

Probability of being a control = 1 - 1/(1+exp(y – (- 3.1071)))

where

y = m5890*(-0.2613) + m11615*(-2.8525) + m11724*(0.1334)

Scheme. 2. Predictive model.

biomarker for HCV-related HCC [Lee et al., 2006].

In our experiments, we firstly profiled whole serum samples by protein chips and then screened out 10 significant protein peaks by applying bioinformatics and cluster analyses. By comparing these results with the most relevant study using similar HBV-infected serum by Poon et al. [2003], we can see data homogeneity and heterogeneity herein. In the protein/peptide peaks we found, 2 out of the 10 significant proteins with the MW 4,574 and 9,139 Da are very close to the peaks of 4,568 and 9,137 Da in their list of most significant differences (6 in total from IMAC3-Cu chip) [Poon et al., 2003], demonstrating that ProteinChip SELDI-TOF MS system can generate some consistent results. However, heterogeneity in SELDI-TOP MS analysis is also very obvious. This can also be found in the works by other groups. The highest discriminating peak at 8,900 Da (belongs to the C-terminal part of the V10 fragment of vitronectin) screened out by Paradis et al. [2005] using similar IMAC chips did not appear in the final list of both Poon's work [Poon et al., 2003, 2005] and our present results. On the other hand, Lee et al. [2006] also identified the significant peak at 8,900 Da using weak cation exchange chips from HCV-related HCC sera, but the peak is complement C3a protein. This heterogeneity demonstrates that in ProteinChip/SELDI-TOF MS profiling it is critical to use identical conditions including experimental design, sample grouping, and statistical algorithm methods.

Based on our sophisticated statistical analysis, three most discriminatory peaks were singled out from the 10 significant signals.

TABLE VI. Sensitivity and Specificity of the Prediction Model

The training set				
	Predicted status			
Actual status	Cancer	HBV	Control	Total
Cancer	46	4	0	50
HBV	3	34	8	45
Control	0	8	22	30
Total	49	46	30	125
Sensitivity	0.92	0.76		
Specificity	0.96	0.85		
The testing set				
	Predicted status			
Actual status	Cancer	HBV	Control	Total
Cancer	17	0	0	17
HBV	0	8	0	8
Control	0	4	10	14
Total	17	12	10	39
Sensitivity	1.00	1.00		
Specificity	1.00	0.87		

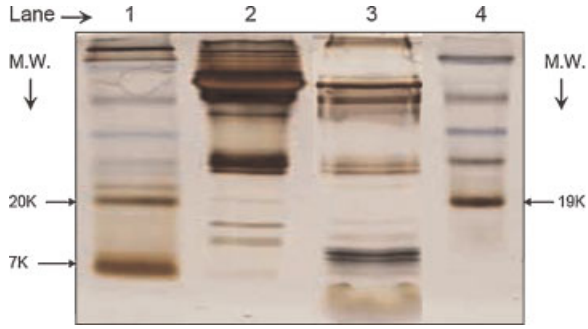


Fig. 3. SDS–PAGE images for serum samples before and after ACN precipitation. **Lane 1**: Molecular weight marker 1; (**Lane 2**) original serum; (**Lane 3**) ACN-treated serum; (**Lane 4**) molecular weight marker 2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

We were then able to use these three peaks to construct a predictive model (Scheme 2), resulting in 92% sensitivity and 96% specificity of HCC prediction for the training set and 100% of sensitivity and specificity for the test set (Table VI). This exciting result demonstrates the powerfulness of combining several biomarkers as a signature in disease detection. For the discrimination of HBV using this model, the sensitivity and specificity are reasonably at 76% and 85%, respectively. This is because that HBV infection is an early and possible stage of liver carcinogenesis, some of the serological characteristics may not be differentiated from the normal ones.

As described above and occurred in other similar studies, the deviation of up to several Daltons in peak detection is normal in SELDI-ProteinChip experiments. This imprecision makes the peak detection uncertain and confusing. The best way to solve this problem is to exactly identify the proteins or peptides corresponding to the peaks. However, it is actually the shortfall or disadvantage of SELDI-ProteinChip system to straightforwardly identify the proteins/peptides. We therefore used 2DE incorporated with MALDI-MS PMF to identify these proteins. Prior to 2DE, high-abundant and high-MW proteins in serum must be depleted and thus the low-MW proteins identified through the protein chips above may be visibly separated in 2DE. By evaluating several serum pre-treatment methods [Tirumalai et al., 2003; Chernokalskaya et al., 2004; Greenough et al., 2004; Harper et al., 2004; Merrell et al., 2004; Peng et al., 2004], we chose the optimized ACN precipitation procedure to successfully eliminate most of high-MW proteins in serum (Fig. 3) and then resumed 2DE experiment to separate the ACN-treated samples (Fig. 4). Here, 2DE exhibits its advantage in displaying the differences between HCC and normal sera

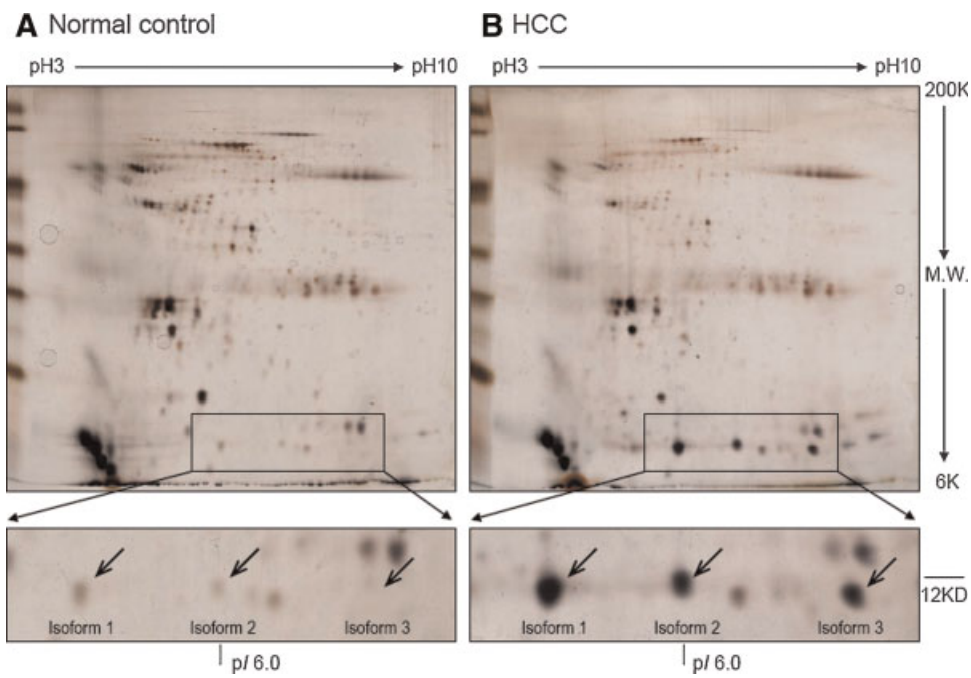


Fig. 4. 2DE images of normal control (**A**) and HCC (**B**) pooled serum samples after ACN treatment. Three significantly different spots in around 11 kDa region were analyzed by PMF to be isoforms of SAA protein. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

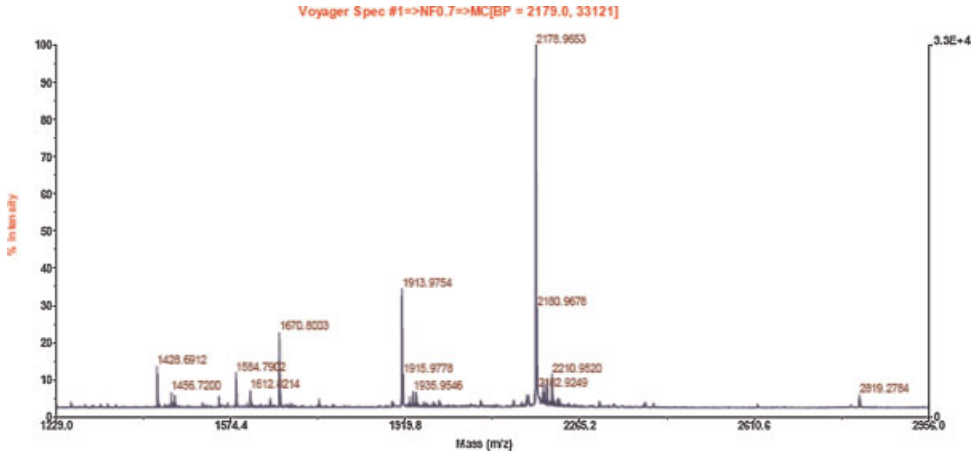


Fig. 5. MALDI-TOF MS spectrum of SAA protein. The major peaks are well matched to the SAA sequence with matched peptides coverage of 84%. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

in the region around 12 kDa MW, where the aberrant expression of three spots in HCC were obviously detected. These spots were identified through PMF to be isoforms of SAA protein and they probably correspond to one/two of the three most discriminatory peaks based on their similar MWs.

IP experiment using anti-SAA antibody followed by SELDI scanning confirmed that one of the three peaks at 11,615 Da belongs to SAA. This is not odd given the fact that the three SAA isoforms in 2DE have very close MWs (Fig. 4) and thus they may appear as a single peak in SELDI-TOF MS spectra. Our Western blot experiments further verified the gradual increase of SAA levels in serum from normal to HBV and then to HCC (Fig. 7). Unfortunately, the other two of the three most discriminatory peaks cannot be identified through this 2DE procedure. Further attempt is still being made using other strategies to identify the remained ambiguous peaks at 5,890 and 11,724 Da, respectively.

SAAs are multifunctional apolipoproteins that are secreted in the acute phase of inflammation [Uhlar and Whitehead, 1999; Zhang et al., 2005]. Most of SAAs are synthesized by the SAA1 and SAA2 genes. These proteins are found in the high-density lipoprotein fraction of the serum and are involved in transporting cholesterol to liver for secretion into the bile, induction of extracellular matrix-degrading enzymes, and chemotactic recruitment of inflammatory cells to the site of inflammation. SAAs are shown to involve in the pathogenesis of several chronic inflammatory diseases, such as amyloidosis, atherosclerosis, and rheumatoid arthritis [Uhlar and Whitehead, 1999; Zhang et al., 2005]. Although SAA cannot be regarded as a specific indicator for assessing HBV and HCC, it is interesting that SAA isoforms can be selected through the complicate statistical analysis in this study. Our present data suggest that SAAs, although belong to the acute-phase proteins, can be combined with other characteristic proteins/peptides to form a specific signature for sensitive prediction of HCC incidence.

CONCLUSION

We have integrated proteomic platforms of both SELDI-ProteinChip and 2DE-MALDI-MS technologies toward identifying serum biomarkers for HCC. Three groups of serum

TABLE VII. Results of MALDI-TOF Mass Spectra and Database Searching for Protein Identification

Spots	Protein (MW/p*I*)	Peptides matched	Sequence coverage (detected peptides)	Experimental (MW/p*I*)	Acc # (NCBI)
1	SAA1 isoform 2 (13,532 Da/6.3)	12	84% (20–122)	11.5 kDa/5.3	2144880
2	SAA1 beta (13,533 Da/5.9)	6	72% (20–122)	11.6 kDa/6.0	13937839
3	SAA2 alpha (10,167 Da/9.1)	5	73% (13–91)	11.4 kDa/8.0	36310

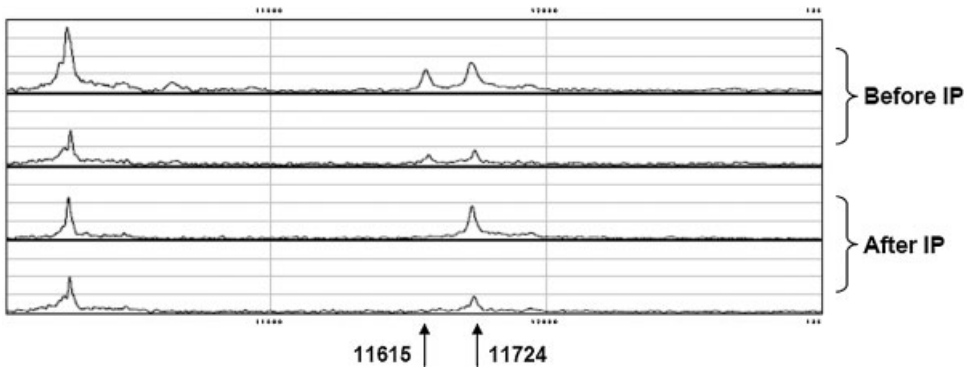


Fig. 6. IMAC3-Cu chip-SELDI TOF spectra for HCC serum samples before and after immunoprecipitation (IP) using anti-SAA antibody. The peak at 11,615 Da disappeared after IP, confirming that this signal belongs to SAA.

samples from HBV-related HCC, HBV infection and normal subjects were profiled by IMAC3 protein chip and the resulting significantly different protein signals were statistically analyzed by a sophisticated algorithm strategy. Three most discriminatory protein/peptide peaks were singled out through the stepwise algorithm analyses and a predictive model based on these three peaks was constructed and tested using the newly enrolled serum samples. One of the three peaks was identified to be SAA by using 2DE-MALDI PMF technology and validated by using IP and Western blotting. Further attempt is being made to exactly identify the other two most discriminatory peaks, with a hope that these belong to specific proteins which can be composed into the predictive model for HCC surveillance and prediction. This study demonstrated the powerfulness of combining techniques of ProteinChip, statistical algorithm, sample pre-treatment, 2DE, and PMF in biomarker identification in serum. The potential of using these biomarkers or the predictive model in HCC surveillance and prediction, although needs to be further confirmed with large sizes of samples, is indeed promising.

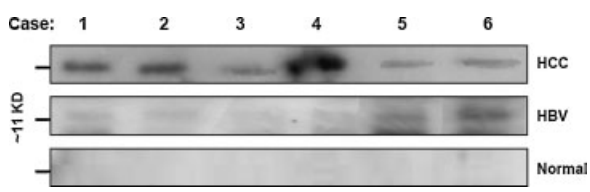


Fig. 7. Western blotting images of normal control, HBV, and HCC individual serum samples. Immunochemical interaction with SAA sheep polyclonal antibody produced strong signals in HCC, weak signals in HBV and no signals in normal sera, corresponding to the results from SELDI profiling.

REFERENCES

Block TM, Comunale MA, Lowman M, Steel LF, Romano PR, Fimmel C, Tennant BC, London WT, Evans AA, Blumberg BS, Dwek RA, Mattu TS, Mehta AS. 2005. Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. Proc Natl Acad Sci U S A 102:779–784.

Chernokalskaya E, Gutierrez S, Pitt AM, Leonard JT. 2004. Ultrafiltration for proteomic sample preparation. Electrophoresis 25:2461–2468.

Comunale MA, Mattu TS, Lowman MA, Evans AA, London WT, Semmes OJ, Ward M, Drake R, Romano PR, Steel LF, Block TM, Mehta A. 2004. Comparative proteomic analysis of de-N-glycosylated serum from hepatitis B carriers reveals polypeptides that correlate with disease status. Proteomics 4:826–838.

Feng JT, Liu YK, Song HY, Dai Z, Qin LX, Almofti MR, Fang CY, Lu HJ, Yang PY, Tang ZY. 2005. Heat-shock protein 27: A potential biomarker for hepatocellular carcinoma identified by serum proteome analysis. Proteomics 5:4581–4588.

Greenough C, Jenkins RE, Kitteringham NR, Pirmohamed M, Park BK, Pennington SR. 2004. A method for the rapid depletion of albumin and immunoglobulin from human plasma. Proteomics 4:3107–3111.

Harper RG, Workman SR, Schuetzner S, Timperman AT, Sutton JN. 2004. Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis 25:1299–1306.

He QY, Chiu JF. 2003. Proteomics in biomarker discovery and drug development. J Cell Biochem 89:868–886.

He QY, Lau GK, Zhou Y, Yuen ST, Lin MC, Kung HF, Chiu JF. 2003. Serum biomarkers of hepatitis B virus infected liver inflammation: A proteomic study. Proteomics 3:666–674.

He QY, Cheung YH, Leung SY, Yuen ST, Chu KM, Chiu JF. 2004. Diverse proteomic alterations in gastric adenocarcinoma. Proteomics 4:3276–3287.

He QY, Zhu R, Ren Y, Tam PK, Chiu JF. 2005. Serological protein profiling of neuroblastoma by ProteinChip SELDI-TOF technology. J Cell Biochem 95:165–172.

Johnson PJ. 1999. Role of alpha-fetoprotein in the diagnosis and management of hepatocellular carcinoma. J Gastroenterol Hepatol 14(Suppl):S32–S36.

Johnson PJ. 2001. The role of serum alpha-fetoprotein estimation in the diagnosis and management of hepatocellular carcinoma. Clin Liver Dis 5:145–159.

Kawakami T, Hoshida Y, Kanai F, Tanaka Y, Tateishi K, Ikenoue T, Obi S, Sato S, Teratani T, Shiina S, Kawabe T, Suzuki T, Hatano N, Taniguchi H, Omata M. 2005. Proteomic analysis of sera from hepatocellular carcinoma patients after radiofrequency ablation treatment. Proteomics 5:4287–4295.

Le Naour F, Brichory F, Misek DE, Brechot C, Hanash SM, Beretta L. 2002. A distinct repertoire of autoantibodies in hepatocellular carcinoma identified by proteomic analysis. Mol Cell Proteomics 1:197–203.

Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Yu CY, Lu FJ, Chow LP. 2005. Identification of human hepatocellular carcinoma-related biomarkers by two-dimensional difference gel electrophoresis and mass spectrometry. J Proteome Res 4:2062–2069.

Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Chen DS, Yu CY, Wen CL, Lu FJ, Chow LP. 2006. Identification of complement C3a as a candidate biomarker in human chronic hepatitis C and HCV-related hepatocellular carcinoma using a proteomics approach. Proteomics 6: 2865–2873.

Li C, Hong Y, Tan YX, Zhou H, Ai JH, Li SJ, Zhang L, Xia QC, Wu JR, Wang HY, Zeng R. 2004. Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. Mol Cell Proteomics 3:399–409.

Li C, Tan YX, Zhou H, Ding SJ, Li SJ, Ma DJ, Man XB, Hong Y, Zhang L, Li L, Xia QC, Wu JR, Wang HY, Zeng R. 2005. Proteomic analysis of hepatitis B virus-associated hepatocellular carcinoma: Identification of potential tumor markers. Proteomics 5:1125–1139.

Luk JM, Lam CT, Siu AF, Lam BY, Ng IO, Hu MY, Che CM, Fan ST. 2006. Proteomic profiling of hepatocellular carcinoma in Chinese cohort reveals heat-shock proteins (Hsp27, Hsp70, GRP78) up-regulation and their associated prognostic values. Proteomics 6:1049–1057.

Melle C, Kaufmann R, Hommann M, Bleul A, Driesch D, Ernst G, von Eggeling F. 2004. Proteomic profiling in microdissected hepatocellular carcinoma tissue using ProteinChip technology. Int J Oncol 24:885–891.

Merrell K, Southwick K, Graves SW, Esplin MS, Lewis NE, Thulin CD. 2004. Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. J Biomol Tech 15:238–248.

Paradis V, Degos F, Dargere D, Pham N, Belghiti J, Degott C, Janeau JL, Bezeaud A, Delforge D, Cubizolles M, Laurendeau I, Bedossa P. 2005. Identification of a new marker of hepatocellular carcinoma by serum protein profiling of patients with chronic liver diseases. Hepatology 41:40–47.

Peng X, Wu Y, Chen J, Wang S. 2004. Proteomic approach to identify acute phase response-related proteins with low molecular weight in loach skin following injury. Proteomics 4:3989–3997.

Petricoin EF, Liotta LA. 2003. Clinical applications of proteomics. J Nutr 133:2476S–2484S.

Poon TC, Yip TT, Chan AT, Yip C, Yip V, Mok TS, Lee CC, Leung TW, Ho SK, Johnson PJ. 2003. Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. Clin Chem 49:752–760.

Poon TC, Hui AY, Chan HL, Ang IL, Chow SM, Wong N, Sung JJ. 2005. Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: A pilot study. Clin Chem 51: 328–335.

Schwegler EE, Cazares L, Steel LF, Adam BL, Johnson DA, Semmes OJ, Block TM, Marrero JA, Drake RR. 2005. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. Hepatology 41:634–642.

Takashima M, Kuramitsu Y, Yokoyama Y, Iizuka N, Toda T, Sakaida I, Okita K, Oka M, Nakamura K. 2003. Proteomic profiling of heat shock protein 70 family members as biomarkers for hepatitis C virus-related hepatocellular carcinoma. Proteomics 3:2487–2493.

Takashima M, Kuramitsu Y, Yokoyama Y, Iizuka N, Fujimoto M, Nishisaka T, Okita K, Oka M, Nakamura K. 2005. Overexpression of alpha enolase in hepatitis C virus-related hepatocellular carcinoma: Association with tumor progression as determined by proteomic analysis. Proteomics 5:1686–1692.

Tan TL, Chen WN. 2005. A proteomics analysis of cellular proteins associated with HBV genotype-specific HBX: Potential in identification of early diagnostic markers for HCC. J Clin Virol 33:293–298.

Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. 2003. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2:1096–1103.

Uhlar CM, Whitehead AS. 1999. Serum amyloid A, the major vertebrate acute-phase reactant. Eur J Biochem 265:501–523.

Xu XQ, Leow CK, Lu X, Zhang X, Liu JS, Wong WH, Asperger A, Deininger S, Eastwood Leung HC. 2004. Molecular classification of liver cirrhosis in a rat model by proteomics and bioinformatics. Proteomics 4:3235–3245.

Zhang N, Ahsan MH, Purchio AF, West DB. 2005. Serum amyloid A-luciferase transgenic mice: Response to sepsis, acute arthritis, and contact hypersensitivity and the effects of proteasome inhibition. J Immunol 174:8125–8134.

Zhu H, Bilgin M, Snyder M. 2003. Proteomics Annu Rev Biochem 72:783–812.

Zhu XD, Zhang WH, Li CL, Xu Y, Liang WJ, Tien P. 2004. New serum biomarkers for detection of HBV-induced liver cirrhosis using SELDI protein chip technology. World J Gastroenterol 10:2327–2329.