Biomarkers, 2009; 14(5): 340–346

informa
healthcare

RESEARCH ARTICLE

New potential biomarkers in the diagnosis of esophageal squamous cell carcinoma

Shu-Yong Xu[1], Zan Liu[1], Wen-Jing Ma[1], Ilyar Sheyhidin[2], Shu-Tao Zheng[1], and Xiao-Mei Lu[1]

[1]*Medical Research Center, and* [2]*Department of Thoracic Surgery, 1st Teaching Hospital, Xinjiang Medical University, Urumqi, Xinjiang Uygur Autonomous Region, China*

Abstract

Objective: To analyse the alterations of serum proteins in cases of esophageal squamous cell carcinoma (ESCC) in order to screen and validate serum marker patterns for the diagnosis of ESCC in the high-risk populations of Xinjiang, China. *Methods*: The serum proteomic patterns of 188 cases, including 139 patients with ESCC (54 Uygur, 45 Kazakh and 40 Han subjects) and 49 sex- and age-matched healthy controls, were detected using the SELDI-TOF-MS (surface-enhanced laser desorption/ionization–time of flight–mass spectrometry) technology with the CM10 ProteinChip. Differences in protein peaks between patients with ESCC and controls were analysed using the Biomarker Pattern Software, and a primary diagnosis model of ESCC was developed and validated with SVM (support vector machines). This model was further evaluated by a large-scale blind test. *Results*: Two hundred and eighty-three protein peaks were detected within the molecular range of 0–20 kDa, among which, 140 peaks were significantly different between ESCC cases and controls ($p < 0.05$). A diagnostic pattern consisting of six protein peaks (m/z 5667, 5709, 5876, 5979, 6043 and 6102) was established with a sensitivity of 97.12% and a specificity of 83.87%. The large-scale blind test generated a sensitivity of 91.43% and a specificity of 88.89%. *Conclusions*: The differential protein peaks analysed by SELDI-TOF-MS may contain promising serum biomarkers for screening ESCC. The diagnostic model which combined only six protein peaks had a satisfactory discriminatory power. The model should be further evaluated in other populations of ESCC patients and tested against controls. The nature and function of the discriminating proteins have yet to be elucidated.

Keywords: *ESCC; SELDI-TOF-MS; bioinformatics; SVM*

Introduction

Esophageal squamous cell carcinoma (ESCC), which is the predominant histological subtype of esophageal cancer (EC), is a highly aggressive cancer, ranking fourth in China and seventh in the world (Pisani et al. 1985, Muir & McKinney 1992, Wang et al. 1997, Lightdale 1999). EC is characterized by the striking geographical variation throughout the world. For example, the so-called 'Asian esophageal cancer belt' covers the Taihang Mountain region in northern China (Lu et al. 1988) and Xinjiang is one of the regions with the highest incidence at approximately 155.9 per 100 000 population in Kazakhs, Tuoli county (Zhang 1988), which is 15-fold higher than the other high-risk regions for EC around the world. Another pronounced feature of EC is the high mortality. Most patients with EC do not survive for more than 1 year after first presenting at healthcare centres, and the 5-year survival rate for EC remains as low as 10% or even lower due to complications caused by the aberrant growth inside the esophagus, such as dysphasia, cachexia, etc. However, the 5-year survival rate for patients with EC in the early stages could be as high as 90% (Hu et al. 2001). Early detection and diagnosis of EC hold the answer for improving the curative rate of EC. To date, few specific and sensitive serum markers have been available for the diagnosis of EC. Although studies have documented many EC-associated molecules, such as oncogenes, tumour suppressor genes,

Address for Correspondence: Xiao-Mei Lu, Medical Research Center, 1st Teaching Hospital, Xinjiang Medical University, Urumqi 830054, Xinjiang Uygur Autonomous Region, China. Tel: +86-991-8189869. Fax: +86-991-4360051. E-mail: luxiaomei88@163.com

(Received 21 December 2008; revised 16 March 2009; accepted 17 March 2009)

ISSN 1354-750X print/ISSN 1366-5804 online © 2009 Informa UK Ltd
DOI: 10.1080/13547500902903055
http://www.informapharmascience.com/bmk

metastatic genes, apoptosis genes, etc., highly specific and sensitive biomarkers for the early detection of EC, and for diagnosis and evaluation of treatment efficiency have not yet been identified.

Although esophagoscopy has the potential to detect early EC, its widespread application is limited by its discomfort and high cost. Surface-enhanced laser desorption/ionization–time-of-flight–mass spectrometry (SELDI-TOF-MS) is one of the proteomics tools for screening and profiling potential serum markers in various kinds of malignant tumour. Advances in proteomics have introduced novel techniques for the screening of new cancer biomarkers and raised technology for early diagnosis of cancer to a new level (Jain 2002, Rai & Chan 2004, Graham et al. 2005, Patel et al. 2005). SELDI-TOF-MS and ProteinChip technology (Ciphergen Biosystems, San Francisco, CA, USA) offer a sensitive, high-throughput and rapid method for proteomic analysis of a complex mixture of proteins and peptides (Merchant & Weinberger 2000, Srinivas et al. 2001, Weinberger et al. 2002, Yasui et al. 2003). SELDI-TOF-MS combined with a bioinformatics approach has been successfully used to identify highly sensitive and specific potential biomarkers for the diagnosis of prostate cancer (Hellström et al. 2007, Skytt et al. 2007), ovarian cancer (Wu et al. 2006, Xia et al. 2008), brain cancer (Liu et al. 2005), colorectal cancer (Hundt et al. 2007, Ward et al. 2008), breast cancer (Goncalves 2007, Leong et al. 2007), lung cancer (Han et al. 2008), pancreatic cancer (Valkovskaya et al. 2007), liver cancer (Cui et al. 2008), etc. However, studies on EC in the region with the highest incidence in the world, Xinjiang, China, have not yet been reported.

One hundred and eighty-eight cases, including 139 serum samples of patients and 49 from controls, were analysed using SELDI-TOF-MS with weak cation exchange (CM10) ProteinChip arrays. New potential biomarkers were screened and a serum protein fingerprint model for diagnosis of ESCC was established.

Materials and methods

Samples

A total of 188 serum samples from patients and healthy individuals were obtained in the Xinjiang Medical University Hospital from 2005 to 2007, after informed consent. The sera of the patients were obtained before any therapeutic measures were implemented. A total of 139 patients with pathologically confirmed ESCC (including 54 Uygurs, 45 Kazakhs and 40 Hans) were included; the median age of the ESCC patients was 56 years (range 28–83). There were 96 men and 43 women. The controls were 49 healthy sex- and age-matched individuals (including 28 Uygurs, 10 Kazakhs and 11 Hans). All serum samples were collected in the morning before breakfast. The sera were left at room temperature for 1–2 h, centrifuged at 3000 rpm for 10 min, and then stored at -80°C.

Reagents and instruments

Sinapinic acid (SPA) was purchased from Sigma (St Louis, MO, USA). ProteinChip Biosystems (Ciphergen PBS II plus SELDI-TOF-MS) and CM10 chips were purchased from Ciphergen Biosystems. All other reagents were acquired from Sigma.

ProteinChip array analysis

All serum specimens were thawed in wet ice and then centrifuged at 10 000 rpm for 2 min. The supernatants were retained. Ten microlitres of U9 buffer (9 M urea, 2% CHAPS, 1% DTT) was added to 5 μl of each serum sample in a 96-well cell culture plate, which was then agitated on a platform shaker at 4°C for 30 min. Next, 185 μl of sodium acetate (100 mM, pH 4) was added to the U9 serum mixture and was further agitated on a platform shaker at 4°C for 2 min. CM10 chips were activated by adding 200 μl of sodium acetate and agitated for 5 min twice. Diluted samples (100 μl) were applied to each spot of the bioprocessor (Ciphergen Biosystems) that contains the ProteinChip arrays. The bioprocessor was then sealed and agitated on a platform shaker for 60 min at 4°C. The excess of the serum mixtures was discarded. The chips were then washed three times with 200 μl of sodium acetate and another two times with deionized water. Finally, the chips were removed from the bioprocessor and air-dried. Prior to the SELDI-TOF-MS analysis, 1 μl of a saturated solution of SPA in 50% ACN, 0.5% trifluoroacetic acid was applied onto each chip twice and the chips were again air-dried.

Chips were analysed by the PBS-II Plus mass spectrometer reader (Ciphergen Biosystems). Data were obtained by averaging the results from a total of 144 laser shots with an intensity of 180, a detector sensitivity of eight, a high mass of 20 000 Da and an optimized range of 2000–20 000 Da. The instrument was calibrated by the all-in-one peptide molecular mass standard (Ciphergen Biosystems).

Bioinformatics and biostatistics

To establish a diagnostic model for EC and evaluate its efficiency, 135 samples were taken for the training set (including 104 EC patients and 31 health controls). In addition, 53 new samples were taken for a test set (including 35 EC patients and 18 health controls) for a blind test on a large scale.

The data analysis was implemented by the Zhejiang University–Cancer Institute–ProteinChip Data Analysis System (ZUCI-PDAS), which was designed by Jiekai Yu on the MATLAB Web Server 1.2.4 (The MathWorks, Natick, MA, USA). The first step of our data analysis was to use the undecimated discrete wavelet transform (UDWT) method to de-noise the signals. The UDWT method is implemented by the Rice Wavelet Toolbox version 2.4. In the second step the spectra were subjected to baseline correction by fitting a monotone local minimum curve and mass calibration (adjusting the intensity scale according to three labelled peaks that appear in all the selected spectra) comments.

The proteomic peaks were detected and quantified by an algorithm that takes each local maximum of each de-noised, baseline-corrected and calibrated mass spectrum into account. The third step consisted of filtering the peaks to maintain a signal-to-noise (S/N) of more than three. The S/N of a peak is the ratio of the height of the peak above the baseline to the wavelet-defined noise. Finally, to match peaks across spectra, we pooled the detected peaks and combined them if the relative difference in their mass sizes was not more than 0.3% comments. The minimal percentage of each peak, appearing in all the spectra, is 10%. If a spectrum did not have a peak within a given cluster, the maximal height within the cluster will be assigned to its peak value comments. The normalization was performed only with the identified peak clusters. To distinguish between the different groups of data, we used a non-linear support vector machine (SVM) classifier, originally developed by Vladimir Vapnik (1982), with a radial basis function kernel, a parameter gamma of 0.6, and a cost of constraint violation of 19. The leave-one-out cross-validation approach was applied to estimate the accuracy of this classifier. This approach takes out one sample each time as the test set and keeps the remaining samples as the training set. This process was repeated until each sample had been taken once as a test sample. This SVM classifier was implemented by the shareware program OSU_SVM v.3.00 Tool box of Junshui Ma and Yi Zhao.

The capability of each peak in distinguishing different groups of data was estimated by the *p*-value of the Wilcoxon test. The top ten peaks with the smallest *p*-value were selected for further analysis. Each of the 1023 combinations of the ten peaks was analysed by the leave-one-out cross-validation SVM. Combinations with the highest accuracy in distinguishing different groups of data were selected as potential biomarkers. The SVM model with the highest Youden's index was selected as the model for detecting EC.

All these bioinformatics studies were integrated in the ZUCI-PDAS available at www.zlzx.net.

Results

Reproducibility of the experiment

Reproducibility of the protein chip was determined with quality control serum by applying 20 serum samples to each chip in a random fashion. The CV of the selected peaks, after being normalized by the intensity, was 19.7% and the CV of the selected peaks' mass was 0.05%.

Screening of relative protein and establishment of diagnosis model of EC versus healthy control

A total of 283 peaks were detected in the molecular range of 0–20 000 and 140 qualified peaks were selected after noise filtering and peak cluster identification. These peaks were ranked according to the *p*-value of the Wilcoxon rank sum test. The top ten peaks with the smallest *p*-value were selected, randomly combined, and fed into SVM. The accuracy of each combination in distinguishing ESCC from healthy controls was analysed, combinations with the highest accuracy were chosen as potential biomarkers, and the SVM model with the highest Youden's index was used as the diagnostic model. This model comprised six potential biomarkers with *m/z* of 5667, 5709, 5876, 5979, 6043, 6102 Da, respectively (Figure 1), which were all highly expressed in ESCC but weakly expressed in healthy individuals. The descriptive statistics of these six peaks are shown in Table 1. This model had a sensitivity of 97.12% and a specificity of 83.87%, as evaluated by leave-one-out cross-validation (Table 2).

Identification of 53 new serum samples was predicted using the model. Diagnosis was correctly made in 32 out of 35 patients with ESCC and in 16 out of 18 healthy individuals. The sensitivity and specificity of the blind test were 91.43% and 88.89%, respectively (Table 3).

Discussion

Although many diagnostic tools are used for EC, such as computed tomography, magnetic resonance imaging and endoscopic examination, the misdiagnosis rate of EC remains high. For a long time, researchers have looked for valuable biomarkers for EC diagnosis, such as CEA, SCC, CYFRA21-1, P53-Ab (Toshikatsu et al. 1998, Kawaguchi et al. 2000, Mao et al. 2003). However, all these biomarkers are either non-specific or have a poor positive predictive value. Therefore, there is an urgent need to search for more ideal biomarkers and seek molecular biological techniques which are less invasive and more sensitive to improve the early diagnosis of EC.

The oncogenesis of EC is a complex multistep process involving multiple genetic mutations. In light of this

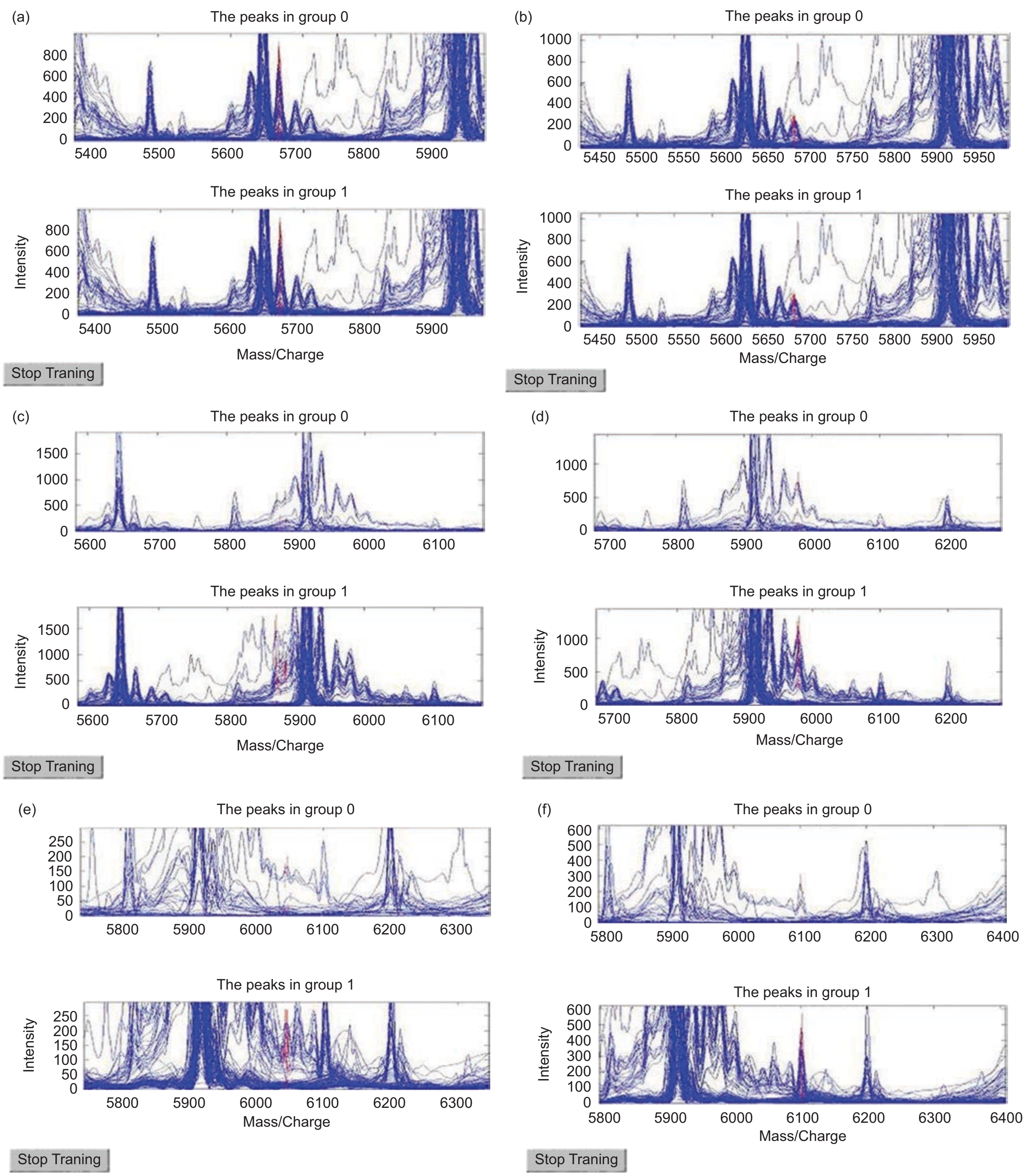

Figure 1. Representative spectra of the six selected biomarkers of esophageal squamous cell carcinoma (ESCC) and healthy controls. (a) The biomarker of 5667 *m/z*; (b) the biomarker of 5709 *m/z*; (c) the biomarker of 5876 *m/z*; (d) the biomarker of 5979 *m/z*; (e) the biomarker of 6043 *m/z*; (f) the biomarker of 6102 *m/z*. The *x*-axis is the molecular weight of the peaks and the *y*-axis is the intensity of the peaks. Group 0, normal group; group 1, patients with ESCC.

multifactorial nature of EC, it is plausible that a combination of multiple biomarkers is necessary to improve the diagnosis of EC. In order to disclose the serum protein 'fingerprints' for diagnosis of EC in the region with the highest incidence – Xinjiang, China – SELDI-TOF-MS and ProteinChip technology coupled with a

Table 1. Comparison of the six selected discrepant protein mass peaks between esophageal squamous cell carcinoma (ESCC) cases and a healthy population ($\bar{x} \pm s$).

Protein peaks (*m/z*)	Healthy control (mean ± SD)	ESCC (mean ± SD)	*p*-Value
6103	26.01 ± 27.27	165.22 ± 125.58	< 0.05%
5876	56.75 ± 55.59	221.38 ± 112.31	< 0.05%
5710	45.76 ± 79.91	201.40 ± 132.80	< 0.05%
5667	77.41 ± 103.05	329.97 ± 211.91	< 0.05%
5979	65.38 ± 91.42	266.46 ± 131.51	< 0.05%
6044	26.44 ± 38.55	122.95 ± 96.96	< 0.05%

Table 2. Leave-one-out cross-validation results of the training set by applying the diagnostic model.

Groups	ESCC	Healthy control	Total
ESCC (predicted)	101	5	106
Healthy control (predicted)	3	26	29
Total	104	31	135

ESCC, esophageal squamous cell carcinoma.

Table 3. Predicted results of the test sets in the blind test, by applying the diagnostic model.

Groups	ESCC	Healthy control	Total
ESCC (predicted)	32	2	34
Healthy control (predicted)	3	16	19
Total	35	18	53

ESCC, esophageal squamous cell carcinoma.

sophisticated bioinformatics approach was applied in the present study. Six potential biomarkers were found and a protein fingerprint pattern has been established which can distinguish EC from healthy individuals with a specificity of 88.89% and sensitivity of 91.43%. These six biomarkers were all overexpressed in EC samples compared with control samples. Although these potential biomarkers still have to be further investigated, these results suggest that this pattern can be used for detection and screening of EC and for distinguishing EC from a healthy population.

To date, five groups, all from PR China, have reported potential biomarker studies on patients with primary ESCC and subjects from control groups based on SELDI technology, Biomarker Pattern Software and different ProteinChip technologies, and they have provided several protein peak discriminating patterns (Wang et al. 2004, Wang et al. 2006, Feng et al. 2007, Liu et al 2007, Zhang et al. 2007). Wang et al. (2004) proposed a diagnostic pattern which consisted of 12 protein peaks with *m/z* of 1028, 1098, 1301, 2047, 2742, 3975, 4130, 4283, 4301, 5635, 6203, 13749 on the WCX2 ProteinChip, which generated a sensitivity of 85% (119/140) and a specificity of 84.14% (38/45). Wang et al. (2006) built a discriminating model on two proteins with *m/z* ratios of 2942 and 15953 using the WCX2 ProteinChip system and a decision-tree classification measure. Liu et al. selected six candidate protein peaks with the *m/z* values of 2545, 3371, 3746, 5009, 5021 and 15 886 to establish a predictive model on the WCX2 ProteinChip, with 77.27% (34/44) sensitivity and 75.00% (33/44) specificity. Zhang et al. identified one protein peak of *m/z* 9479 using the IMAC3 ProteinChip, able to discriminate ESCC with 95.4% (42/44) sensitivity and 95.2% (40/42) specificity. Feng et al. showed that seven tumour markers of *m/z* 9439, 6627, 2867, 4494, 7762, 6835 and 4095 could discriminate ESCC from controls with IMAC3 chips and BioMarkerWizard Software, with a sensitivity and a specificity of 88.5% and 82.0%, respectively.

From the above analysis, it is clear that all these selected markers were discordant and they usually performed less well than the model we propose in this study. A well-balanced representation of the major ethnic groups present in Xinjiang in our study may have prevented biases associated with the genetic background of the patients as well as controls and improved the diagnosis accuracy. On the other hand, the candidate ProteinChip and tools of bioinformatics were different in the various studies, and this may lead to apparent discrepancies. Because the biomarkers rely on patterns, different arithmetic produces different biomarker combinations. We can, however, note that, in the present study, the marker at *m/z* 5667 was similar to *m/z* at 5635 discovered by Wang et al. This protein certainly deserves to be explored further. In our study, the high-throughput protein mass spectrometry data analysis platform –ZUCI-PDAS – was applied. Compared with traditional software, the ZUCI-PDAS improves de-noising, calibration, normalization and reproducibility of the protein peaks. Combining the genetic arithmetic with SVM, the pattern should be better.

One of the challenges in the analysis of SELDI-TOF-MS-generated data is to reduce the false protein peaks, in which the discriminatory power is due to random variation. To solve this problem, we used a bioinformatics tool, ZUCI-PDAS, to analyse the spectral data, including de-noising with the UDWT, baseline correction, peaks detection, biomarker selection, establishment and evaluation of the SVM differential patterns. Our algorithm is likely to find most of the true, reproducible peaks. The SVM classification technique used in the ZUCI-PDAS was first described by Vapnik and is a new machine learning method based on the statistical theory. The SVM can solve problems such as the generalization of the medium and samples in pattern recognition, pattern selection and over-fitting (Brown et al. 2000, Winters-Hilt et al. 2006, Zhang et al. 2006). The SVM is specifically used for the finite samples in order to get the optimal solution with available information rather than the optimal solution with the sample number tending to infinity. In theory, the overall optimal point can be obtained to solve the local extremum problem which is unsolvable in the artificial neural network method. The algorithm transforms the actual condition into hyperplane feature space with non-linear transformation, the non-linear discriminate function in the former space is achieved by constructing the linear discriminate function in the hyperplane space. The specificity can ensure excellent generalization performance and effectively solve the dimensionality problem. Leave-one-out cross-validation is utilized to determine the accuracy of the classifier. Independent test sets were used to evaluate further the accuracy of our proposed diagnostical models. All these steps ensure that the selection of biomarkers is not influenced by systematic biases.

The sample size in the present study is still limited and a larger number of samples is needed to test our model further, which should, in addition, be tested in other populations of patients with ESCC, especially in low-risk areas such as Europe and the USA, and compared with patients with other types of tumour. Nonetheless, by establishing serum protein fingerprint models through SELDI-TOF-MS and a bioinformatic approach, a novel and highly sensitive and specific assistant method for the diagnosis of ESCC has been provided. To understand better the cell mechanisms and/or metabolic pathways in which the proteins disclosed by this screening method are involved, further steps are necessary.

The main pathways identified in other studies included cell defence mechanisms (such as heat shock proteins, which were also found in esophageal adenocarcinomas) (Fujita et al. 2008, Langer et al. 2008, Qi et al. 2008), cell motility, glycolysis, regulation of transcription, oxidative stress processes and protein folding (Du et al. 2007). Biomarkers have also been identified with prognosis value, such as transglutaminase 3 (Uemura et al. 2008). To analyse whether or not the biomarkers found in our study come directly from the tumour or are indirect markers of the presence of the ECSS, it is possible to analyse the proteome of tumours and to detect the presence or otherwise of these biomarkers. A second point is to identify these biomarkers; for this, different approaches may be used. Firstly, we plan to measure the peptide mass with a better accuracy and resolution using a matrix-assisted laser desorption/ionization (MALDI)-TOF-MS to define the mass of the monoisotopic peptides. Given the complexity of the mixture it is necessary to carry out some steps of prepurification of the samples, before identification. In order to separate the biomarkers of very large proteins, technologies such as gel exclusion or cut-off will be used. Off gel, high-performance liquid chromatography (HPLC) technologies will be applied in order to separate the peptides (biomarkers) according to their physicochemical characteristics. The candidate polypeptide peaks can now be separated by Tricine-SDS-PAGE followed by trypsin digestion and identification by HPLC-MS/MS analysis and database search (He et al. 2008). The other possibility will be to use mass spectrometers configured with electrospray ionization (ESI) source coupled with orbitrap. The ESI configuration allows the generation of multispecies-charged peptides and performance of fragmentation by the electron transfer dissociation (ETD) approach. This new ion fragmentation methodology allows a random cleavage at amide groups along the peptide backbone independent of peptide or protein size (Appella et al. 2007). The relevance of the identified peptides will then be confirmed by immunohistochemistry or Western blotting in esophageal tissues. This identification will allow us to understand better the cell mechanisms and/or metabolic pathways in which the proteins disclosed by this screening method are involved. The identification of our promising biomarkers will also allow us to set up various technologies, such as selected reaction monitoring, to increase detection sensitivity and accuracy of quantification (Lange et al. 2008).

Acknowledgements

The project was supported by the National Natural Science Foundation (no. 30760223, 30860097), Xinjiang Scientific-technique Bureau Grant (no. XJKJT200511113), and Urumqi Scientific-technique Bureau Grant (no. Y05331002). The authors thank Dr Patrick Ducoroy, from the Clinical and Innovation Proteomic Platform, University of Burgundy and University of Franche-Comte, France, and Professor Dominique Angele Vuitton, from Clinical Immunology WHO Collaborating Center CHU Besançon, University of Franche-Comté, France for their precious advice and revision of the manuscript.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Appella E, Anderson CW. (2007). New prospects for proteomics electron-capture (ECD) and electron-transfer dissociation (ETD) fragmentation techniques and combined fractional diagonal chromatography (COFRADIC). *Fed Eur Biochem Soc J* 274:6255.

Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 97:262–7.

Cui JF, Liu YK, Zhou HJ, Kang XN, Huang C, He YF, Tang ZY, Uemura T. (2008). Screening serum hepatocellular carcinoma-associated proteins by SELDI-based protein spectrum analysis. *World J Gastroenterol* 14:1257–62.

Du XL, Hu H, Lin DC, Xia SH, Shen XM, Zhang Y, Luo ML, Feng YB, Cai Y, Xu X, Han YL, Zhan QM, Wang MR. (2007). Proteomic profiling of proteins dysregulated in Chinese esophageal squamous cell carcinoma. *J Mol Med* 85:863–75.

Feng XS, Wang LD, Shan TY, Guo T, Li JL, Fan ZM, Jiao XY, Chang ZH, Gao SG, Han J, and others (2007). Proteomic analysis on esophageal cancer with lymph node metastasis. *J Zhengzhou University (Med Sci)* 42:397–9.

Fujita Y, Nakanishi T, Miyamoto Y, Hiramatsu M, Mabuchi H, Miyamoto A, Shimizu A, Takubo T, Tanigawa N. (2008). Proteomics-based identification of autoantibody against heat shock protein 70 as a diagnostic marker in esophageal squamous cell carcinoma. *Cancer Lett* 263:280–90.

Goncalves A, Bertucci F, Birnbaum D, Borg JP. (2007). Proteic profiling SELDI-TOF and breast cancer: clinical potential applications. *Med Sci (Paris)* 23:23–6.

Graham DR, Elliott ST, Van Eyk JE. (2005). Broad-based proteomic strategies: a practical guide to proteomics and functional screening. *J Physiol* 563:1–9.

Han KQ, Huang G, Gao CF, Wang XL, Ma B, Sun LQ, Wei ZJ. (2008). Identification of lung cancer patients by serum protein profiling using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. *Am J Clin Oncol* 31:133–9.

He M, Qin J, Zhai R, Wei X, Wang Q, Rong M, Jiang Z, Huang Y, Zhang Z. (2008). Detection and identification of NAP-2 as a biomarker in hepatitis B-related hepatocellular carcinoma by proteomic approach. *Proteome Sci* 6:10–20.

Hellström M, Lexander H, Franzén B, Egevad L. (2007). Proteomics in prostate cancer research. *Anal Quant Cytol Histol* 29:32–40.

Hu YC, Lam KY, Law S, Wong J, Srivastava G. (2001). Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by Cdna expression array: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. *Clin Cancer Res* 7:2213–21.

Hundt S, Haug U, Brenner H. (2007). Blood markers for early detection of colorectal cancer: a systematic review. *Cancer Epidemiol Biomarkers Prev* 16:1935–53.

Jain KK. (2002). Role of proteomics in diagnosis of cancer. *Technol Cancer Res Treat* 1:281–6.

Kawaguchi H, Ohno S, Miyazaki M, Hashimoto K, Eqashira A, Saeki H, Watanabe M, Suqimachi K. (2000). CYFRA21-1 determination in patients with esophageal squamous cell carcinoma: clinical utility for detection of recurrences. *Cancer* 89:1413–17.

Lange V, Picotti P, Domon B, Aebersold R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 4:222.

Langer R, Ott K, Specht K, Becker K, Lordick F, Burian M, Herrmann K, Schrattenholz A, Cahill MA, Schwaiger M, Hofler H, Wester HJ. (2008). Protein expression profiling in esophageal adenocarcinoma patients indicates association of heat-shock protein 27 expression and chemotherapy response. *Clin Cancer Res* 14:8279–87.

Leong S, Christopherson RI, Baxter RC. (2007). Profiling of apoptotic changes in human breast cancer cells using SELDI-TOF mass spectrometry. *Cell Physiol Biochem* 20:579–90.

Lightdale CJ. (1999). Esophageal cancer. American College of Gastroenterology. *Am J Gastroenterol* 94:20–9.

Liu CZ, Zhu PY, Shi MX, Liu JB, Wang WJ, Liao P, Xiang PP, Zhang YX. (2007). Screening of serum proteome biomarker of esophageal squamous cell carcinoma by WCX2 proteinchip. *China Oncol* 17:701–5.

Liu J, Zheng S, Yu JK, Zhang JM, Chen Z. (2005). Serum protein fingerprinting coupled with artificial neural network distinguishes glioma from healthy population or brain benign tumor. *J Zhejiang University Sci* 6:4–10.

Lu JB, Yang WX, Zu SK, Chang QL, Sun XB, Lu WQ, Quan PL, Qin YM. (1988). Cancer mortality and mortality trends in Henan, China, 1974–1985. *Cancer Detect Prev* 13:167–73.

Mao YS, Zhang DC, Zhao XH, Wang LJ, Qi J, Li XX. (2003). Significance of CEA, SCC, and Cyfra21-1 serum test in esophageal cancer. *Chinese J Oncol* 25:457–60.

Merchant M, Weinberger SR. (2000). Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21:1164–77.

Muir CS, McKinney PA. (1992). Cancer of the esophagus: a global overview. *Eur J Cancer Prev* 1:259–64.

Patel PS, Telang SD, Rawal RM, Shah MH. (2005). A review of proteomics in cancer research. *Asian Pacific J Cancer Prev* 6:113–17.

Pisani P, Parkin DM, Ferlay J. (1993). Estimates of the worldwide mortality from eighteen major cancers in 1985: implications for prevention and projections of future burden. *Int J Cancer* 55:891–903.

Qi YJ, He QY, Ma YF, Du YW, Liu GC, Li YJ, Tsao GS, Ngai SM, Chiu JF. (2008). Proteomic identification of malignant transformation-related proteins in esophageal squamous cell carcinoma. *J Cell Biochem* 104:1625–35.

Rai AJ, Chan DW. (2004). Cancer proteomics: serum diagnostics for tumor marker discovery. *Ann N Y Acad Sci* 1022:286–94.

Skytt A, Thysell E, Stattin P, Stenman UH, Antti H, Wikström P. (2007). SELDI-TOF MS versus prostate specific antigen analysis of prospective plasma samples in a nested case–control study of prostate cancer. *Int J Cancer* 121:615–20.

Srinivas PR, Srivastava S, Hanash S, Wright GL. (2001). Proteomics in early detection of cancer. *Clin Chem* 47:1901–11.

Toshikatsu, Fumio I, Yokosuka O, Saisho H, Suzuki T, Koide Y, Isono K. (1998). Expression of p53 and p21/WAFI proteins in gastric and esophageal cancer comparison with mutation of the p53 gene. *Digest Dis Sci* 43:279–83.

Uemura N, Nakanishi Y, Kato H, Saito S, Nagino M, Hirohashi S, Kondo T. (2008). Transglutaminase 3 as a prognostic biomarker in esophageal cancer revealed by proteomics. *Int J Cancer* 124:2106–15.

Valkovskaya N, Kayed H, Felix K, Hartmann D, Giese NA, Osinsky SP, Friess H, Kleeff J. (2007). ADAM8 expression is associated with increased invasiveness and reduced patient survival in pancreatic cancer. *J Cell Mol Med* 11:1162–74.

Vapnik V. (1982). *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag.

Wang LD, Wang DC, Zheng S, Fan ZM, Li JL, Feng CW, Zhang YR, Liu B, Gao SS, He X, Feng XS. (2006). Serum proteomic profiles of the subjects with esophageal precancerous and cancerous lesions from Linzhou, an area with high incidence of esophageal cancer in Henan Province, Northern China. *Chinese J Cancer* 25:549–55.

Wang LD, Zhou Q, Yang CS. (1997). Esophageal and gastric cardia epithelial cell proliferation in northern Chinese subjects living in a high incidence area. *J Cell Biochem Suppl* 28-29:159–65.

Wang Y, Deng BP, Ma LH, Xu Y, Zhang ZS, Liu F, Mao YS, Zhang JS, Zhang DC, Zhao XH. (2004). A serum pattern for the diagnosis of esophageal squamous cell carcinoma. *Chinese J Lab Med* 27:634–7.

Ward DG, Nyangoma S, Joy H, Hamilton E, Wei W, Tselepis C, Steven N, Wakelam MJ, Johnson PJ, Ismail T, Martin A. (2008). Proteomic profiling of urine for the detection of colon cancer. *Proteome Sci* 6:19.

Weinberger SR, Boschetti E, Santambien PJ. (2002). Surface-enhanced laser desorption-ionization retentate chromatography mass spectrometry (SELDI-RC-MS): a new method for rapid development of process chromatography conditions. *J Chromatogr B Anal Technol Biomed Life Sci* 782:307–16.

Winters-Hilt S, Yelundur A, McChesney C, Landry M. (2006). Support vector machine implementations for classification & clustering. *BMC Bioinform* 7 (Suppl. 2):S4.

Wu SP, Lin YW, Lai HC, Chu TY, Kuo YL, Liu HS. (2006). SELDI-TOF MS profiling of plasma proteins in ovarian cancer. *Taiwan J Obstet Gynecol* 45:26–32.

Xia T, Zheng ZG, Gao Y, Mou HZ, Xu SH, Zhang P, Zhu JQ. (2008). Application of SELDI-TOF serum proteome profiling in cervical squamous cell carcinoma. *Ai Zheng* 27:279–82.

Yasui Y, Pepe M, Thompson ML, Adam BL, Wright GL Jr, Qu Y, Potter JD, Winget M, Thornquist M, Feng Z. (2003). A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics* 4:449–63.

Zhang HL, Gao CF, Wang XL, Li DH, Hu WH. (2007). Application of SELDI-TOF-MS spectrometry in serological diagnosis for esophageal squamous cell carcinoma. *Chinese J Clin Lab (Science)* 25:280–2.

Zhang XG, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN, Iqlehart JD, Miron A, Liu JS, Wong WH. (2006). Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *BMC Bioinform* 7:197–210.

Zhang YM. (1988). Distribution of esophageal cancer in Xinjiang. *Acta Acad Med Xinjiang* 11:139–45.