Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Proteiomic patterns for endometrial cancer using SELDI-TOF-MS*

Li-rong ZHU[1], Wen-ying ZHANG[2], Li YU[1], Yan-hua ZHENG[3], Jun HU[1], Qin-ping LIAO[†‡1]

([1]*Department of Gynecology, Peking University First Hospital, Beijing 100034, China*)

([2]*Department of Gynecology, 306 Hospital of People's Liberation Army, Beijing 100101, China*)

([3]*Department of Pathology, 306 Hospital of People's Liberation Army, Beijing 100101, China*)

[†]E-mail: qinping_liao@hotmail.com

Received Dec. 11, 2007; revision accepted Jan. 28, 2008

Abstract: Serum samples from endometrial cancer (EC) patients and healthy females were analyzed using surface-enhanced laser desorption-ionization time-of-flight mass spectrometry (SELDI-TOF-MS) to discover the potential diagnostic biomarker for detection of EC. A preliminary training set of spectra derived from 40 EC patients and 30 healthy women were used to develop a proteomic model that effectively discriminated cancer patients from healthy women. The training set had a specificity of 100% and sensitivity of 92.5% in the EC detection. A blind test set, including 20 new cancer cases and 10 healthy women, was used to validate the sensitivity and specificity of this multivariate model, which had a corresponding results of 60% in specificity and 75% in sensitivity, respectively. The combination of SELDI-TOF-MS with bioinformatics tools could help find new biomarkers and establish the detection of EC with high sensitivity and specificity.

Key words: Biomarkers, Surface-enhanced laser desorption-ionization time-of-flight mass spectrometry (SELDI-TOF-MS), Endometrial cancer (EC), Proteomics

doi:10.1631/jzus.B0710589 **Document code:** A **CLC number:** R737.33

INTRODUCTION

The early detection of cancer is crucial for its ultimate control and prevention. Most of today's licensed tests for disease detection are protein-based assays. Technologies such as multidimensional separation systems directly coupled to mass spectrometry analysis represent improvement in sensitivity and throughput when compared with traditional two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) analysis for biomarker discovery. Proteomic pattern diagnostics combines proteomic pattern profiling of tissue and body fluids by mass spectrometry with sophisticated bioinformatics tools to identify patterns within the complex proteomic profile that discriminate between normal and disease states. Proteomic pattern diagnostics has been successfully applied to solve the problems of early detection for a number of different types of cancer (Wulfkuhle *et al*., 2003). Endometrial cancer (EC) is one of the most common gynecological malignancies in women. During the past two decades, the incidence of EC has been improving stably in China. Surgery alone is often curative in patients with early stage disease, while radiation and chemotherapy are reserved for patients with advanced stage or recurrent disease who posses nothing but poor prognosis. Early diagnosis of EC, therefore, is the key point for patients to survive. However, so far no good serum biomarkers with high sensitivity and specificity are available for the diagnosis and monitoring of EC, and recurrent patients are detected only following the development of symptoms or abnormalities in imaging assessments.

In this study, we aimed to search differentially expressed proteins as potential biomarkers in EC patients by surface-enhanced laser desorption-ionization time-of-flight mass spectrometry (SELDI-

‡ Corresponding author

* Project (No. 985-2-015-24) partly supported by "985" Project of Research Grants from Peking University, China

TOF-MS). We used Ciphergen weak cation exchange (WCX2) protein chip to screen potential serum biomarkers for EC detection. A total of 100 serum samples from EC patients or healthy women were collected and analyzed simultaneously. A panel of differentially expressed proteins were advocated for biomarkers of diagnosis for EC.

MATERIALS AND METHODS

Samples

Serum samples were collected from patients in Peking University First Hospital and 306 Hospital of People's Liberation Army (PLA), Beijing, China. A total of 100 serum samples were included in this study and divided into profiling group (training set) and validation group (blind test set). In the training set, the cancer group consisted of 40 serum samples from EC patients at different clinical stages: stage I (n=33), stage II (n=2), stage III (n=4) and stage IV (n=1). The healthy control group included 30 serum samples from age-matched volunteers with no evidence of disease, who participated in the EC-screening program at Peking University First Hospital. Meanwhile, the blind test set consisted of 20 EC patients with stage I (n=18) and stage II (n=2), and 10 controls were included for evaluation of this multivariate model to distinguish EC from healthy cohort. Diagnoses were pathologically confirmed, and sera were obtained before treatment. The median age of the 100 participants was 57 years (range 34~78 years). Sera were collected by a technician and analysis was performed in the same lab process. All serum samples were stored at 2~8 °C for a maximum of 48 h before freezing at −80 °C until use. The study was performed after approval by institute of Human Investigation Committee.

Protein chip array analysis

Serum samples were thawed on ice and centrifuged at 10000 r/min for 2 min at 4 °C. A total of 3 μl of each serum sample was diluted to 9 μl with U9 buffer (9 mol/L urea, 2% cholamidopropyl-dimethyl-ammonio-1-propanesulfonate (CHAPS), 1% ditheothreitol (DTT), 50 mmol/L Tris-Cl, pH 9.0). The mixture was vortex-mixed at 4 °C for 30 min and further diluted by 108 μl sodium acetate (100 mmol/L, pH 4.0). All analyses used protein chip WCX2 (Ciphergen Biosystems Inc., Fremont, CA, USA). A WCX2 chip was washed with 100 mmol/L sodium acetate (pH 4.0) twice. Then, the diluted samples (100 μl) were applied to the protein chip array and incubated for 60 min on a platform shaker at room temperature. The array was washed twice with the same sodium acetate buffer for 5 min and once with high-performance liquid chromatography H_2O, followed by a quick water rinse. After air-drying, 0.5 μl of saturated sinapinic acid was applied twice to each spot, and then the spots were air-dried. The chip was performed on Protein Biological System II (c) mass spectrometer reader (Ciphergen Biosystems Inc.). Data were collected by averaging 64 laser shots with an intensity of 190, a detector sensitivity of 8, and an optimized range of 2~50 kD. Each serum sample was spotted on all eight bait surfaces of one WCX2 array.

Bioinformatics and statistical analysis

All serum samples were analyzed by SELDI-TOF-MS technology and the data were processed with Protein Chip Software 3.1 (Ciphergen Biosystems Inc.). The spectra were generated on WCX2 chips, and protein peaks clustering and classification analyses were made using Ciphergen Biomarker Wizard and Biomarker Pattern Software, respectively. The spectra intensities of all samples were normalized to the total ion current of m/z 2000 and 100000, noise of spectra under 1000 was filtrated, and peaks were detected with an automatic peak detection pass. Peak clusters were completed to cluster the peaks in different samples that had similar masses. All protein clusters with P value less than 10^{-3} were chosen. Sensitivity was calculated as the ratio of the number of correctly classified diseased samples to the total number of diseased samples. Specificity was calculated as the ratio of the number of nondisease samples correctly classified to the total number of nondisease samples.

RESULTS

Protein peak detection of serum proteins retained on the WCX2 arrays was analyzed on a Protein Biological System II mass reader. The high mass to acquire was set to 100 kD, with an optimization range

from 2 to 50 kD. A mass accuracy of 0.1% was achieved by external calibration using the All-In-1 Protein Standard (Ciphergen Biosystems Inc.). Peak intensity was normalized to total ion current (2~50 kD). There was little variation with day-to-day sampling and instrumentation or chip variation. The results show that in two test groups of EC patients and healthy women, 13 protein biomarkers were differentially expressed, and the sensitivity using each protein biomarkers is shown in Table 1. As shown in Fig.1, when compared with the healthy control sera, the two peaks (*m*/*z* 7567 and 8605) have higher mean peak height values in EC samples, which were up-regulated proteins in EC group. In contrast, the two peaks (*m*/*z* 4797 and 6371) were down-regulated in cancer versus healthy control sera. Therefore, the differentially expressed proteins between two groups could be used for biomarker discovery.

Table 1 Descriptive data for the potential biomarkers in the detection of EC

m/*z*	Expression change	*m*/*z*	Expression change
6371	↓	7088	↓
5481	↑	15517	↑
4797	↓	5067	↑
5252	↑	5000	↓
8605	↑	15113	↑
4095	↓	7567	↑
2746	↓		

↑ and ↓ represent up-regulated and down-regulated expressions in EC, respectively

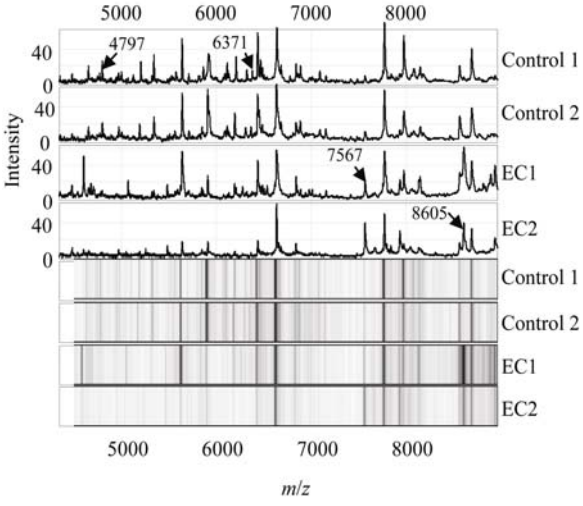

Fig.1 Representative SELIDI-TOF-MS serum spectra (2~9 kD). EC means endometrial cancer patients, and control means healthy women. Four differentially expressed proteins of relative molecular mass (M_r) 4797, 6371, 7567 and 8605 were screened up

Data analysis

A total of 13 protein peaks from relative molecular mass (M_r) 2000 to 50000 were selected to construct the classification model. Fig.2 shows the tree structure and sample distribution. Four peaks, M_r 4797, 6371, 8605 and 15517, were chosen to set up the decision tree (Kohonen, 1990; Holland, 1994; Conrads, 2004). At Node 1, samples of M_r 6371 with peak intensities lower than or equal to 7.006 went to Node 2, which had 7 control samples and 32 EC samples. Otherwise, samples entered Node 3, which had 31 samples. At Node 2, samples of M_r 4797 with peak intensities lower than or equal to 17.033 went to Node 4, which had 4 control samples and 32 EC samples. The other samples entered terminal Node 1, which had 3 control samples. At Node 4, samples of M_r 8605 with peak intensities lower than or equal to 16.696 went to terminal Node 2, which had 4 control samples and 1 EC sample. The other samples went to terminal Node 3, which had 31 EC samples. In addition, at Node 3, samples of M_r 8605 with peak intensities lower than or equal to 13.440 went to terminal Node 4, which had 4 EC samples. The other samples

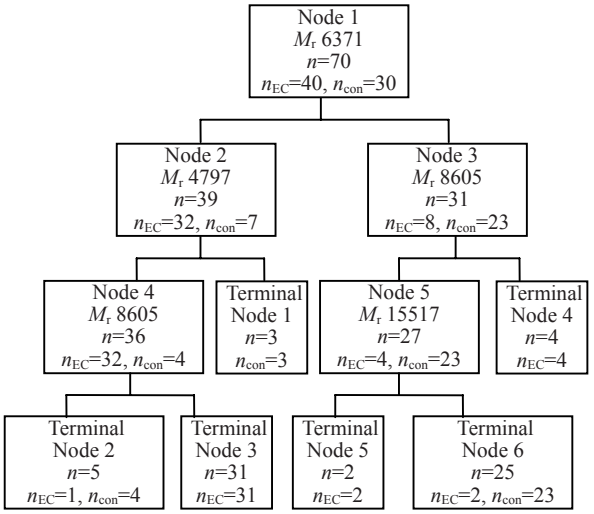

Fig.2 Tree structure and sample distribution. Peaks with M_r 6371, 4797, 8605 and 15517 were chosen to set up the decision tree

n_{EC}: The number of EC samples; n_{con}: The number of health samples

entered Node 5, which had 23 control samples and 4 EC samples. At Node 5, samples of M_r 15517 with peak intensities lower than or equal to 0.494 went to terminal Node 5, which had 2 EC samples. The other samples went to terminal Node 6, which had 23 control samples and 2 EC samples. As shown in Table 2, the model identified 70 samples of the training set, 33 in control and 37 in EC, and yielded a sensitivity of 92.5% (37/40) and specificity of 100% (30/30). While in the blind test set, 6 out of 10 true non-cancer cases were correctly classified, and 15 out of 20 EC samples were correctly classified as malignant. These results yield a sensitivity of 75% (15/20) and a specificity of 60% (6/10).

Table 2 Sensitivities and specificities of the novel protein markers for cancer detection

	Sensitivity of cancer patients (%)	Specificity of controls (%)
Training set	92.5 (n=40)	100 (n=30)
Blind test set	75 (n=20)	60 (n=10)

DISCUSSION

EC is now becoming one common malignancy, strongly threatening women's health as its incidence is growing. The high risk factors include minor labor, menopause delay, obesity, diabetes, and so on. As most cancers, the survival rates of EC mainly depend on early detection of the disease. So far, no satisfactory biomarkers exist to cope with this difficult task. New proteomic technologies have brought the hope of discovering novel early cancer-specific biomarkers in complex biologic samples and setting up new clinically relevant test systems. The development of SELDI-TOF-MS (Hutchens and Yip, 1993) offers an effective tool for protein biomarker identification. The classical approaches for discovering disease-associated proteins are 2D-PAGE and mass spectrometry. 2D-PAGE is able to resolve thousands of proteins, but labor intensive, requires large quantities of protein, and is not easily converted into a diagnostic test. Mass spectrometry is offering an alternative to 2D-PAGE. However, some limitations such as extensive sample preparation and signal-background problems resulting from inorganic and organic contaminants have hindered it from being used as a high-throughput screening tool for proteins of interest in complex biologic samples. SELDI-TOF-MS has largely overcome many of these limitations. This technology has been used successfully to discover potential serum diagnostic markers for breast (Li *et al*., 2002), lung (Yang *et al*., 2005), bladder (Liu *et al*., 2005), liver (Schwegler *et al*., 2005) and gastric (Ebert *et al*., 2004) and ovarian (Yu *et al*., 2005) cancers.

Taking advantages of the recent development in SELDI and of the protein chip technology, we were able to simultaneously analyze the protein profiles of 100 serum samples from EC patients or healthy women. We identified biomarkers specific for EC with WCX2 chip. Thirteen protein peaks appeared differentially expressed. Among them, 7 proteins (M_r 5481, 5252, 8605, 15517, 5067, 15113 and 7567) increased and 6 (M_r 6371, 4797, 4095, 2746, 7088 and 5000) decreased in EC group compared with the control group. Even though the two protein peaks of M_r 5000 and 5067 had sensitivities of 95% and 40%, respectively, they yielded specificities of 70% and 100%, respectively. By further data analysis using Biomarker Pattern Software, four protein peaks of M_r 4797, 6371, 8605 and 15517 were autoselected to form tree structure. For distinguishing women with EC from healthy women, a sensitivity of 92.5% and a specificity of 100% were obtained with the training set, and a sensitivity of 75% and a specificity of 60% were obtained with the blind test set. Therefore, better diagnostic results can be obtained when several biomarkers were combined. Two reporters (Yoshizaki *et al*., 2005; Yang *et al*., 2004) also explored the protein profiling of EC from normal endometrium using SELDI-TOF-MS, with the sample being not serum but EC tissues. Two proteins (Yoshizaki *et al*., 2005), EC1 and EC2 (the isoelectric points of EC1 and EC2 were approximately pH 5.0 and 7.0, respectively), had been found to be consistently expressed differentially. EC1 had an increased level of expression in carcinoma tissues, while EC2 was expressed at a lower level. One over expressed protein (Yang *et al*., 2004) had been further confirmed as chaperonin 10 in malignant endometrial tissues. Such results also lead the way to positively identify the candidate protein markers found in the study for a greater understanding of the mechanisms of cancer genesis and development.

CONCLUSION

Using proteomics approaches such as Ciphergen protein chip arrays and SELDI-TOF-MS in combination with bioinformatics tools could facilitate the discovery of new biomarkers. With the panel of four selected biomarkers, we could achieve high sensitivity and specificity for detection of EC. We may thus conclude that SELDI-TOF-MS is a novel, noninvasive, sensitive, highly predictive and rapid method for prediction of endometrial caner, and will be promising in future clinical screening. However, the number of samples analyzed in this study to some degree has limited the validity of the results. The difference of specificity and sensitivity between training set and blind test set may be due to the constitution of samples of both sets and the intrinsic sample-to-sample and day-to-day variabilities. Further research is needed to elucidate the sequence of the interesting peptides identified from our current study and to confirm our current findings in larger cohorts of study samples. And further analysis can be performed comparing early stage cancer with later stage cancer to discover the potential biomarker for monitoring recurrence and metastasis of cancer.

ACKNOWLEDGEMENT

We thank the investigators and the patients who contributed to this study. It is just a primary step in the huge work of screening potential serum biomarkers for detection of EC. Thanks for the comments from reviewer of Harvard University who pointed out that further research is needed to confirm our current findings in large cohorts of study samples in future.

References

Conrads, T.P., 2004. High-resolution serum proteomic features for ovarian cancer detection. *Endocr. Relat. Cancer*, **11**(2):163-178. [doi:10.1677/erc.0.0110163]

Ebert, M.P., Meuer, J., Wiemer, J.C., Schulz, H.U., Reymond, M.A., 2004. Identification of gastric cancer patients by serum protein profiling. *J. Proteome Res.*, **3**(6):1261-1266. [doi:10.1021/pr049865s]

Holland, J.H., 1994. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press, Cambridge, MA.

Hutchens, T.W., Yip, T.T., 1993. New desorption strategies for the mass spectrometric analysis of macromolecules. *Rapid Commun. Mass Spectrom.*, **7**(7):576-580. [doi:10.1002/rcm.1290070703]

Kohonen, T., 1990. The self-organizing map. *Proc. IEEE*, **78**(9):1464-1480. [doi:10.1109/5.58325]

Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y., Chan, D.W., 2002. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin. Chem.*, **48**(8):1296-1304.

Liu, W.W., Guan, M., Wu, D., Zhang, Y., Wu, Z., Xu, M., Lu, Y., 2005. Using tree analysis pattern and SELDI-TOF-MS to discriminate transitional cell carcinoma of the bladder cancer from noncancer patients. *Eur. Urol.*, **47**(4):456-462. [doi:10.1016/j.eururo.2004.10.006]

Schwegler, E.E., Cazares, L., Steel, L.F., Block, T.M., Marrero, J.A., Drake, R.R., 2005. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. *Hepatology*, **41**(3):634-642. [doi:10.1002/hep.20577]

Wulfkuhle, J.D., Liotta, L.A., Petricoin, E.F., 2003. Proteomic applications for the early detection of cancer. *Nat. Rev.*, **3**(4):267.

Yang, E.C., Guo, J., Diehl, G., DeSouza, L., Rodrigues, M.J., Romaschin, A.D., Colgan, T.J., Siu, K.W., 2004. Protein expression profiling of endometrial malignancies reveals a new tumor marker: Chaperonin 10. *J. Proteome Res.*, **3**(3):636-643. [doi:10.1021/pr049975z]

Yang, S.Y., Xiao, X.Y., Zhang, W.G., Sun, X.Z., Zhang, L.J., Zhang, W., Zhou, B., Yang, D.C., He, D.C., 2005. Application of serum SELDI proteomic patterns in diagnosis of lung cancer. *BMC Cancer*, **5**(1):83. [doi:10.1186/1471-2407-5-83]

Yoshizaki, T., Enomoto, T., Nakashima, R., Ueda, Y., Kanao, H., Yoshino, K., Fukumoto, M., Yoneda, Y., Buzard, G.S., 2005. Altered protein expression in endometrial carcinogenesis. *Cancer Lett.*, **226**(2):101-106. [doi:10.1016/j.canlet.2004.11.053]

Yu, J.K., Zheng, S., Tang, Y., Li, L., 2005. An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. *J. Zhejiang Univ. Sci. B*, **6**(4):227-231. [doi:10.1631/jzus.2005.B0227]