BIOMEDICAL AND ENVIRONMENTAL SCIENCES 16, 140-148 (2003)

A Preliminary Analysis of Non-small Cell Lung Cancer Biomarkers in Serum[1]

R73 A

XUE-YUAN XIAO[#], YING TANG[#], XIU-PING WEI[Δ], AND DA-CHENG HE[#,*]

[#]*Key Laboratory for Cell Proliferation and Regulation Biology, Ministry of Education, Beijing Normal University, Universities' of Confederated Institute for Proteomics, Beijing 100875, China;* [Δ]*Beijing Tiantan Hospital, Capital University of Medical Sciences, Beijing 100050, China*

Objective To identify potential serum biomarkers that could be used to discriminate lung cancers from normal. **Methods** Proteomic spectra of twenty-eight serum samples from patients with non-small cell lung cancer and twelve from normal individuals were generated by SELDI (Surfaced Enhanced Laser Desorption/Ionization) Mass Spectrometry. Anion-exchange columns were used to fractionate the sera into 6 designated pH groups. Two different types of protein chip arrays, IMAC-Cu and WCX2, were employed. Samples were examined in PBSII Protein Chip Reader (Ciphergen Biosystem Inc) and the discriminatory profiling between cancer and normal samples was analyzed with Biomarker Pattern software. **Results** Five distinct potential lung cancer biomarkers with higher sensitivity and specificity were found, with four common biomarkers in both IMAC-Cu and WCX2 chip; the remaining biomarker occurred only in WCX2 chip. Two biomarkers were up-regulated while three biomarkers were down-regulated in the serum samples from patients with non-small cell lung cancer. The sensitivities provided by the individual biomarkers were 75%-96.43% and specificities were 75%-100%. **Conclusions** The preliminary results suggest that serum is a capable resource for detecting specific non-small cell lung cancer biomarkers. SELDI mass spectrometry is a useful tool for the detection and identification of new potential biomarker of non-small cell lung cancer in serum.

Key words: SELDI mass spectrometry; Non-small cell lung cancer; Biomarkers

INTRODUCTION

Lung cancer is at present the number one cause of cancer death for both men and women. More than 171 000 new cases occurred in the United States in 1998[1]. Despite improvements in chemotherapy and radiotherapy over the past two decades, the overall 5-year survival rate for its patients is still below 15%[2]. Therefore, early diagnosis and treatment of lung cancer can greatly improve patient's survival, and it has become an urgent task to search and discover its specific biomarkers.

There must be quantitative and qualitative protein changes at cellular level before the occurrence of pathological changes in some cancers. Theoretically, study of the alternations of dynamic proteomics in cells will allow us to understand the tiny changes of cancers at the early stage, and specific biomarkers for particular cancers should be identified by comparing

[1]This work was supported by Science Technology Key Project of Ministry of Education (Grant No.272006) and the Major State Basic Research Project (Grant No.G1999053901).

*Correspondence should be addressed to Da-Cheng HE, E-mail: dhe@bnu.edu.cn.

Biographical note of the first author: Xue-Yuan XIAO, female, born in 1961, Associated Professor, majoring in proteomics in screening and studying disease biomarkers and functional proteins.

0895-3988/2003
CN 11-2816
Copyright © 2003 by CAPM

the protein profiling between patient samples and normal control. To this aim, some new technologies to detect biomarkers have been developed more recently. Surfaced Enhanced Laser Desorption/Ionization (SELDI) technology is one of them from Ciphergen Biosystems (Fremont, CA)[3]. It combines protein chip array with time-of-flight mass spectrometry and offers the advantages of speed, simplicity, sensitivity and suitability for a comparative study. Another advantage is that a very small amount of samples is used for the detection. SELDI technique is thus highly valuable for studying proteomics of unusual samples, such as cerebrospinal fluid and liquid in joint cavity. During the last 12 months, this technique has been successfully used to study biomarkers for breast cancer[4], bladder cancer[5] and ovarian cancer[6]. In this report, we used small size of serum samples to explore the different proteomics between patients with lung cancer and healthy individuals.

MATERIALS AND METHODS

Reagent and Protein Chips

Acetonitrile, trifluoracetic acid (TFA), CHAPS, urea, copper sulfate, Tri-base, OGP, Hepes, Aprotinin, Leupeptin and PMSF were ordered from Sigma. Sinapinic acid (SPA), anion exchange columns (Hyper Q DF), IMAC3 and WCX2 protein chips were provided by Ciphergen Biosystems Inc.

Samples

Twenty-eight serum samples from patients with non-small lung cancer (20 cases of primary brochogenic carcinoma and 8 cases of metastasis) were provided by Beijing Tiantan Hospital. The healthy individuals selected from participants in a national health survey were used as the normal control consisting of 6 males and 6 females from 40 to 70 years of age, their name list was provided by the State Sport and Physical Culture Administration.

Protein Chip Reader

It is a laser desorption/ionization time-of-flight mass spectrometer made by Ciphergen Biosystems Inc, Fremont, CA, USA. It can detect and accurately calculate the mass of compounds ranging from small molecules and peptides of less than 1 000 Da up to proteins of 500 kDa.

Protein Chip Arrays

There are two types of protein chips. One is chemical surfaces that include a series of classic chromatographic chemistries and specialized affinity capture surfaces. It can be used to detect novel proteins by obtaining its fingermap. The other is pre-activated surface arrays to investigate the specific proteins covalently immobilized on chip surfaces. In this study, two types of chemical surface protein chips were used. IMAC3 (immobilized metal affinity capture) contains nitrilotriacetic acid group on the surface that chelates cationic metal such as nickel, copper and zinc. Proteins applied to the chip surface may bind to the chelated metal ion through histidine, tryptophan, cysteine, and phosphorylated amino acids. It was used to analyze metal-binding proteins, phosphorylated proteins, and histidine tagged proteins and to eventually identify biomarkers. WCX2 (weak cation exchange) contains weak anionic carboxylated groups that interact with the positive charge on the surface of the analyte such as lysine, arginine or histidine. It was used to selectively capture the proteins

with high pI's and biomarkers that have a positive charges on the surface.

Anion-exchange Column Preparation

Resin was prepared by washing Hyper Q DF resin 3 times with 5x bed volumes of 50 mmol/L Tris-HCl pH 9. It was used to equilibrate with 50 mmol/L Tris-HCl pH 9 in a 50% suspension overnight.

Sera Samples Treatment

Twenty μL of each serum was pipetted and 30 μL 9 mol/L urea was added, and then vertexed for 20 min at 4℃. Supernatant was pipetted, 50 μL of 1 mol/L urea was added to each sample, and then they were mixed. The supernatant was loaded on the corresponding anion exchange column, vortexed for 20 min at 4℃.

Sample Fractionation

Lung cancer samples and controls were fractionated over a Hyper Q DF ion exchange resin with a gradient of washes from pH9, pH7, pH5, pH4, pH3 and organic solvent (33.3% isopropanol 16.7% acetonitrile 0.1% trifluoracetic acid). If the column was washed with pH 9 buffer, the elution was called fraction 1. Fraction 2 was the elution if the column was washed with pH 7 buffer. The same procedure was used for pH 5, pH 4, pH 3 and organic solvent washing buffer, and for all of the rest samples. Finally, 6 fractions were collected and each fraction was applied to two different kinds of protein chips (WCX2 and IAMC-3). All samples were processed in duplicate to confirm reproducibility in resolving the serum proteins.

IMAC-Cu Protocol Using a Bioprocessor

Ten uL 100 mmol/L $CuSO_4$ was applied to each spot and incubated in a humidity chamber for 15 min. Chip was rinsed with deionized water to remove the excess copper. Five μL of NaCl in PBS was applied to each spot and incubated on a shaker for 5 min. The chip was assembled in the bioprocessor and 20 μL of each eluant diluted in 80 μL binding buffer was added, and incubated with vigorous shaking for 30 min. The sample was removed from the wells and each well was washed twice, 5 min for each with 200 μL binding buffer. The chip was removed from the bioprocessor and bulk, and washed briefly with 8 mL of water in a 15 mL conical centrifuge tube. One point five μL SPA was applied after it was air dried to each spots. The SPA addition was repeated. The air dried chips were then analyzed in Protein Chip Reader (PBSII). Each fraction was duplicated on separated chips of the same type.

WCX2 Protocol Using a Bioprocessor

Each spot was pretreated for 10 min with 10 mmol/L HCl. The chip was rinsed with 10 mL water three times in a conical tube. The chip was assembled in a bioprocessor and 150 binding buffer was added to each well, the mixture was incubated for 5 min at room temperature with vigorous shaking. The buffer was removed from the wells and 100 μL sample diluted in binding buffer was immediately added and incubated for 30 min. The sample was removed from the wells and washed each well twice, 5 min for each wash with 150 μL washing buffer. The chip was removed from the bioprocessor and briefly rinsed with 8 mL of deionized water in a 15 mL conical centrifuge tube. The chip air dried and 0.5 μL SPA was applied to each spot. The SPA addition was repeated. The air dried chips were then

analyzed in Protein Chip Reader (PBSII). Each fraction was duplicated on separated chips of the same type.

Peak Detection

Peak detection was performed using PBSII Protein Chip System. Mass accuracy was assessed daily by using peptide and protein standard calibrations. A mass accuracy of 0.1% was achieved within this system. Peptide and proteins below the 200 000 mass/charge (M/Z) range were ionized with sinapinic acid as a matrix, which was most effective for the detection of proteins within this range. The mass range from 3 000 to 50 000 Da, laser intensity-230, detector sensitivity-8, and 200 000 Da were selected as optimal range for analysis. Data were collected automatically.

Biomarker Wizard and Biomarker Pattern Analysis

Ciphergen system 3.0 Biomarker Wizard was used to compare the protein profiles in the samples and biomarker pattern. 4.0 software was used to set up a classification tree and select the candidate proteins that could distinguish lung cancer from normal control samples. Peak detection was performed using biomarker wizard (Ciphergen SELDI software version 3.0). The signal/noise was 5, and minimum peak threshold was 10%. Construction of the biomarker patterns model was performed by biomarker pattern (Ciphergen SELDI software version 4.0) as follows: the level of target variable details was 2, minimum value was 0, and the tree type was classification, parent node was 2 cases and terminal node was 1, classification tree was selected as Gini using one rule at a time in the form of a peak intensity. The splitting was defined by a level higher or lower than the intensity levels of one peak. For example, if mass A had an intensity less than or equal to the "X", the data were splitted into two nodes, a left node for yes and a right node for no. This splitting process would stop if terminal nodes for further splitting had no gain.

Statistical Analysis

If the biomarkers were negatively or lowly expressed in the patient group in comparison with the normal control group, the sensitivity was defined as the ratio of the lung cancer patients without or with low expression of the biomarkers to the total number of the patients enrolled in the study. Specificity was defined as the ratio of the healthy individuals with highly expressed protein peaks to the total number of the individuals enrolled in the study. On the contrary, the sensitivity was defined as the ratio of the lung cancer patients with highly expressed biomarkers to the total number of such patients enrolled in this study. Specificity was defined as the ratio of the normal individuals without or with lowly expressed protein peaks to the total number of such individuals. The PPV (positive predictive value) for the study samples was calculated by dividing the number of true positive and false positive cases. The NPV (negative predictive value) for the study samples was calculated by dividing the number of false negative and true negative samples.

RESULTS

Distribution of Proteins on Different Protein Chips

Forty serum samples were assayed by PBSII mass spectrometry and peaks were collected automatically. All samples were processed in duplication to confirm the reproducibility of

resolving the serum proteins. Forty-two peaks in five control sera were analyzed both on IMAC-Cu and WCX2 chips to determine reproducibility of the mass spectra (data not shown). The average coefficient of variance (CV) based on five normal pool human sera for intensities of 42 peaks was lower than 20%. Except for matrix peaks, WCX2 chips appeared to bind larger number of proteins than IMAC-Cu chips. Collection of data generated 433 protein peaks in a total of 6 fractions on the IMAC-Cu chips versus 692 protein peaks on WCX2 chips. In the six corresponding fractions, an average of the protein peaks on the WCX2 chips was 59.8% higher than that on the IMAC-Cu chips. Interestingly, although IMAC-Cu chips were specific to capture the metal-binding proteins or phosphorylated proteins, the results showed that 80%-90% of protein peaks bound on the IMAC-Cu chips were also captured on the WCX2 chips. Nevertheless the relative intensity of peaks on two types of chips was obviously different.

Detection of Five Lung Cancer-associated Biomarkers

All of the protein spectra were normalized by Biomarker Wizard software at first, and then the number of peaks and their relative intensity in every fraction for both patient and control sera were analyzed by biomarker pattern software. The results showed that while the relative intensity and the number of majority of peaks in patient serum and normal control were essentially the same, some proteins showed obvious difference in their intensity between the two groups. Among them, 5 prominent proteins peaks were shown to be obviously different between the patient and normal control serum. Among them, three peaks were down-regulated and two peaks were up-regulated in the patient sera (Table 1). These three protein peaks could be captured by both WCX2 and IMAC-Cu chips, but their sensitivities and specificities for biomarkers were higher in most cases when detected using IMAC-Cu chips than using WCX2 chips (Table 1). The classification results from the decision tree algorithm for 15 848 Da peak are showed in Fig. 2. Compared with the normal sera, two proteins were up-regulated in the lung cancer sera. The average mass on SELDI was 4 463 Da and 4 669 Da (Fig. 3). The sensitivity and specificity of the classification system for 4 463 Da peak are presented in Fig. 4

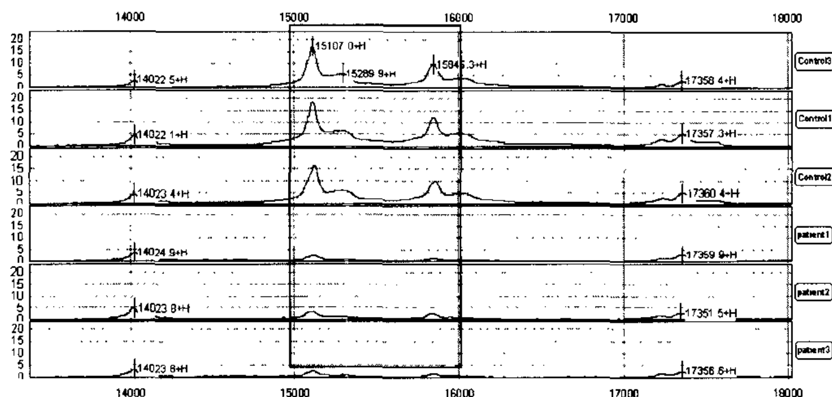

FIG. 1. Twenty-eight cases of lung cancer sera and twelve cases of normal sera were analyzed on WCX2 protein chip. In the lung cancer sera three peaks were lowly expressed in comparison with the normal. The spectrum was represented as a mass chromatogram. X-axis was molecular weight for each spectrum (M/Z values), and Y-axis was relative intensity.

TABLE 1

Summary of Lung Cancer-associated Biomarkers Data

	IMAC-Cu Protein Chip				WCX2 Protein Chips			
Marker	Sensitivity %	Specificity %	PPV %	NPV %	Sensitivity %	Specificity %	PPV %	NPV %
4463	85.71	91.67	96.00	68.75	71.43	75.00	64.52	45.00
4669					75.00	83.00	70.00	52.63
15110	92.86	91.67	96.29	78.57	89.30	75.00	86.67	71.43
15291	96.43	91.76	96.43	84.62	82.14	91.67	79.31	68.75
15848	92.86	100.00	100.00	85.71	82.14	100.00	82.14	70.59

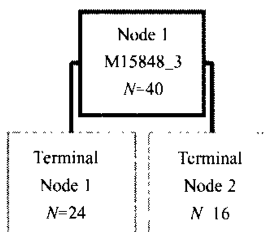

FIG. 2. Classification of the lung cancer and normal control by the decision tree alogorithm. The left terminal node is the cases of peak intensity under or equal to 4.745, while the right one is higher than 4.745. N represents the number of samples

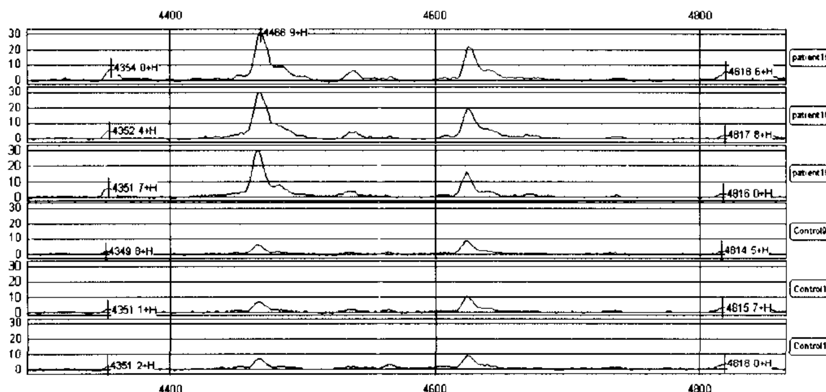

FIG. 3. Twenty-eight cases of lung cancer sera and twelve cases of normal sera were analyzed on IMAC-Cu protein chip. In the lung cancer sera one peak was highly expressed in comparison with the normal. The spectrum was represented as a mass chromatogram. X-axis was molecular weight for each spectrum (M/Z values), and Y-axis was relative intensity.

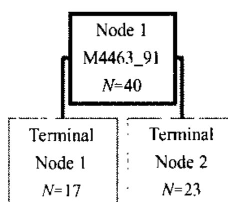

FIG. 4. Classification of the lung cancer and normal control by the decision tree alogorithm. The left terminal node is the cases of peak intensity under or equal to 28.35, while the right one is higher than 28.35. *N* represents the number of the samples.

DISCUSSION

Lung cancer is one of the most common tumors all over the world. In 2001, about 169 500 new cases of lung cancer were estimated and 157 400 lung cancers patients died in the United States[7]. Over the recent years, the morbidity and mortality of lung cancer have been steadily increased in our country. Therefore, early diagnosis and treatment can greatly improve patient's survival. The limited direct accessibility of lung cancers has led to efforts to identify cancer-associated soluble biomarkers in serum for early detection. However, so far no biomarkers have been able to detect early lung cancer based on a serum sample[8]. Therefore, the identification and simultaneous analysis of a panel of biomarkers, representative of the biological characteristics of lung cancer, have posed greater potential for improving early detection and diagnosis of lung cancer.

As major approaches toward the biomarker screening, two-dimensional protein gel electrophoresis (2-D gel) and Mass Spectrometry (MS) are widely used currently. Although the 2-DE technique has matured to a state that affords high-resolution separation of proteins, good reproducibility and sufficient sensitivity, as a labor intensive technique, it requires large quantities of starting material, and is not practical for screening cancer-related biomarkers. MS is the result of significant technological advances in protein chemistry in the last 2 decades and it is also an indispensable tool for protein study. But MS requires purified samples that limit its application in biological and clinical assay[9,10]. In the meantime, some new technologies for proteomics study are developed. Ciphergen Biosystems, Inc. has recently developed the protein chip technology coupled with SELDI- TOF-MS (surface-enhanced laser desorption/ionization time of flight mass spectrometry) to facilitate protein profiling of complex biological mixtures. Since protein chip arrays can selectively bind proteins based on their physical or chemical modification of the surface, the protein contents of the samples are thus effectively simplified, and the contaminants such as buffer salts or detergents can also be easily washed prior to MS analysis. Compared with 2D-polyacrylamide gel electrophoresis, it is much faster, with a high throughput capability, and could effectively resolve low mass proteins (2 000 to 20 000 Da), and capture hydrophobic proteins, very high or low pH proteins which sometimes are important. Particularly, this technique needs very little amount of sample to analyze biological mixtures directly or after simple pretreatment. So this technique could be directly used for clinical assay. The molecular weight (mass/charge) and pI of proteins are not only available, but some physical and chemical properties of proteins could be suggested, such as hydrophobic protein, hydrophilic protein, phosphorylated protein and metal-binding protein.

Up to the middle of 2001, there were few reports on cancer biomarkers detected with SELDI technology. However, since then, related reports have been accumulated rapidly. Eggeling[11] used this proteinchip array and SELDI to analyze the protein profiles generated from normal tissue with peripheral and central renal cell carcinoma as well as those generated from mechanical microdissected precancers and carcinoma of the cervix uteri. Differently expressed proteins were screened by comparing the protein expression patterns generated using SELDI of tumor with normal and neoplastic tissue, respectively. Paweletz[4] used proteinchip array and SELDI to analyze a set of nipple aspirate fluid samples including those from breast cancer and healthy control. The data showed that the protein signatures generated by this technique could differentiate nipple aspirate fluid samples from breast cancer patients and healthy controls, including those with an mammogram who were later proven to be biopsy normal. Vlahou[5] used protein chip technology for the detection of transitional cell carcinoma of the bladder in urine and 5 biomarkers of bladder cancer were reported. Recently, Petricoin[6] and his colleagues found five biomarkers of ovarian cancer in patient's serum by using protein chip array and SELDI. In Petricoin's article, the discriminatory pattern correctly identified all 50 ovarian cancer cases in the masked set, including all 18 cases of stage I. Of the 66 cases of non-malignant diseases, 63 were recognized as being free from cancer. His results yielded 100% sensitivity, 95% specificity, and 94% positive predictive value. SELDI technique has been applied successfully to the analysis and detection of specific cancer biomarkers in bladder, ovarian, breast and cervix uteri, though it has been developed recently.

In the present work, we also used protein chip array and SELDI to discover and identify the discrimination of protein profiles in serum sample from 28 cases of lung cancer patients and 12 cases of normal control. Before setting up a classification tree model, normalization of all the spectra is an important step due to sample-to-sample variability. This normalization process is a critical step, because nearly all of the protein alterations between lung cancer and normal control are based on the overexpression or underexpression of proteins, rather than on their presence or absence only. Our previous studies of detection of lung cancer biomarkers were conducted by biomarker wizard process and manually inspecting all spectra on naked eye. It did, however, demonstrate that the SELDI profiling can facilitate the discovery of lung cancer candidate biomarkers. Furthermore, it clearly illustrated that it is necessary to use some bioinformatics algorithm to effectively deal with the high dimensionality of the SELDI data. Therefore, in the present study the biomarker pattern software was used and the classification tree model was selected because it was easy to be interpreted and the results could be clearly presented. It was found that the multiple protein differences could be observed between lung cancer and normal control sera. But only 5 protein peaks were found with higher sensitivity and specificity, while 2 biomarkers were up-regulated and 3 biomarkers were down-regulated in serum samples from patients with lung caners. With SELDI profiling classification approach and biomarker pattern software analysis, a range of sensitivities for individual peaks was from 75% to 96.43%, specificity from 75% to 100% and PPV from 64.52% to 100%. In this study, the small size of samples was used to set up a classification tree model, and these five candidate lung cancer biomarkers need to be validated or confirmed. To complete these lung cancer specific biomarkers, we will screen a larger number of patient's sera including different sera of lung cancer and benign pulmonary diseases, such as pneumonia, pulmonary tuberculosis, asthma and chronic bronchitis. It is obvious that some types of artificial intelligence program, such as neural network and genetic algorithm will be able to make this technique much more powerful in searching early biomarkers for diseases with high sensitivity and specificity. The development of such a specific learning software and its application to an extended

scale of these candidate lung cancer biomarkers is under way in our laboratory.

REFERENCES

1. Tran, Y., Bogler, O., Gorse, K., Wieland, H., Green, M.R., and Newsham, I.F. (1999). A novel member of the NF2/ERM4.1 superfamily with growth suppressing properties in lung cancer. *Cancer Research* **59**, 35-43.
2. Wu, W., Kemp, B.L., Proctor, M.L., Gazdar, A.F., Minna, J.D., Hong, W.K., and Mao, L. (1999). Expression of DMBT1, a candidate tumor suppressor gene, is frequently lost in lung cancer. *Cancer Research* **59**,1846-1851.
3. Li, Peiwen (2001). Diagnosis and Treatment of Lung Cancer. Beijing: Traditional Chinese Medicine Press of China, Chapter 2.
4. He, D. and Xiao, X. (2002). SELDI protein chip technology and its applications in Proteomics. *Modern Instruments* **1**, 1-4.
5. Paweleta, C.P., Trock, B., Pennanen, M., Tsangaris, T., Magnant, C., Liotta, L.A., and Petricoin, E.F. (2001). Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Dis. Marker* **17**(4), 301-307.
6. Vlahou, A., Schellhammer, P.F., Mendrinos, S., Patel, K., Kondylis, F.I., Gong, L., Nasim, S., and Wright, G.L. (2001). Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. *American Journal of Pathology* **158** (4), 1491-1502.
7. Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., and Liotta, L.A. (2002). Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* **359**, 572-577.
8. Smith, R.A., Mettlin, C.S., Davis, K.J., and Eyre, H. (2000). American cancer society guidelines for the early detection of cancer. *CA. Cancer J. Clin* **50** (1), 34-49.
9. Eggeling, F.V., Gawriljuk, A., Fiedler, W., Emst, G, C;aissem, U., Klose, J., and Romer, I. (2001). Fluorescent dual colour 2D-protein gel electrophoresis for rapid detection of differences in protein pattern with standard image analysis software. *Int. J. Mo.l Med.* **8** (4), 373-377.
10. Bichsel, V.E., Liotta, L.A., and Petricoin, E.F. (2001). Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer J.* **7**(1), 69-78.
11. Eggeling, F.V., Davies, H., Lomas, L., Fiedler, W., Junker, K., Claussen, U., and Emst, G (2000). Tissue-specific microdissection coupled with Protein Chip array technologies: applications in cancer research. *Biotechniques* **29**(5), 1066-1070.

(*Received November* 5, 2002 *Accepted February* 10, 2003)