

Present Status of Clinical Proteomic Analysis for the Early Detection and Determination of Therapeutic Strategy in Lung Cancer

Takashi Hirano, MD, PhD and Harubumi Kato, MD, PhD

Although human DNA sequences have already been successfully decoded, there is still much we do not understand about the pathogenesis of human diseases. Proteomic analysis is comprehensive analysis of proteins, and there is a high probability that recent advances in this field will enable its application to individualized therapy based on thorough understanding of the pathogenesis of human diseases. Two-dimensional polyacrylamide gel electrophoresis (2-DE) and mass spectrometry (MS) are indispensable tools for proteomic analysis, and these technologies enable discovery of tiny changes of protein expression associated with pathogenesis. In this review we introduce the present status of proteomic analysis for cancer-related proteins, and especially lung cancer-related proteins in the attempt to discover new biomarkers for early detection and determine novel target-molecules for treatment. We believe that proteomic analysis will provide crucial information for diagnosis and treatment. (Ann Thorac Cardiovasc Surg 2006; 12: 4–9)

Key words: proteome, two-dimensional polyacrylamide gel electrophoresis, mass spectrometry, lung cancer

Introduction

“Proteome” and “proteomics” are relatively new words, coined by Wilkins et al. in 1996.¹⁾ Proteomic analysis means comprehensive analysis of proteins, and proteomics is the science by which proteins are comprehensively investigated with regard to their roles as functional elements. The clarification of the human genome sequence is one of the most brilliant events in life science, and its results are accelerating comprehensive analysis of human gene products, “proteins”. If the relationship between gene and protein were a one-to-one correspondence, many researchers might not recognize the importance of proteomic analysis. As we know, a gene alone is only potential information that must be put into a functional form, even

From Department of Surgery, Tokyo Medical University, Tokyo, Japan

Received September 21, 2005; accepted for publication October 17, 2005.

Address reprint requests to Takashi Hirano, MD, PhD: Department of Surgery, Tokyo Medical University, 6-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160-0023, Japan.

though the pathogenesis of a malignant neoplasm is inherently associated with genetic disorders. The DNA is transcribed into RNA, then translated into protein. During this sequence, a number of alterations or modifications occurring at transcriptional, translational and post translational levels profoundly affect function. Put another way, several proteins with different functions may be derived from one gene through either alternative splicing at the RNA level or post-translational modifications at the protein level, for instance phosphorylation. It is well-known that phosphorylation is one of the most critical phenomena in cellular functions, and that it may be intimately associated with crucial events in the pathogenesis of some human diseases, for instance malignant neoplasm. Many researchers believe that proteomics will play a dominant role in life science in the post-genome age. Though recently “post genome” as a technical term is frequently used, it would be more precise to say that at present we are in the age of the post-genome sequence, and that we now stand at the entrance to the age of functional analysis at the molecular level. The main molecules carrying out physiological as well as pathological func-

tions are, of course, proteins. Therefore, it is necessary to investigate human proteins to understand the pathogenesis of human diseases.

Human genome analysis has shown that there are approximately 35,000 human genes, and it is assumed that more than 100,000 proteins must be expressed in the human body. In addition, it is important to understand the three-dimensional structures of protein molecule to understand their functions. Even though it is very complicated to analyze protein molecules, we cannot avoid investigating proteins for the complete elucidation of the pathogenesis of any human diseases. Therefore, new technology that combines simplicity, high through-put and automatic analysis is required. In this context, recent advanced proteomic technologies have brought the hope of discovering novel biomarkers that can be used to detect the early stage of disease, to predict the effectiveness of therapy and to monitor disease progression. The precise prediction of both therapeutic effects and adverse reactions must lead us to individualized therapy. In this review we discuss the present status of clinical proteomics for cancer-related proteins, and especially its application in the field of lung cancer is reviewed. We believe that understanding the present concepts of proteomic analysis in the field of oncology is extremely valuable for thoracic surgeons who will develop new therapeutic strategies for malignant neoplasms. The final purpose of clinical proteomics is to improve diagnostic procedures including the early detection and exact evaluation of the biological characteristics of diseases, and to understand the molecular pathogenesis of diseases to permit novel therapeutic strategies.

Commencement of Clinical Proteomics—Two-dimensional Polyacrylamide Gel Electrophoresis (2-DE) Based Strategies

Though the word “proteome” as a technical term is new, the concepts of comprehensive protein analysis have been established from 1975. At that time, O’Farrell established high resolutional two-dimensional polyacrylamide gel electrophoresis (2-DE) for comprehensive protein analysis.²⁾ In this method proteins are separated using isoelectric focusing (IEF) and sequential sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE). Approximately 1,000 proteins can be evaluated on 2-DE gel showing isoelectric point and molecular weight.

We started proteomic analysis using the 2-DE method on surgically resected materials of solid cancers in 1989.

We initially used frozen surgically resected materials, but high resolutional analysis for cancer-related proteins was not possible due to contamination with serum proteins and necrotic substances as well as normal cells such as mesenchymal cells and inflammatory cells. Until we developed a new sample preparation method using fresh surgically resected materials, it was impossible to investigate cancer-related proteins with high resolution.³⁾ After the establishment of this non-enzymatic sample preparation technique we investigated the histopathological magnification of each histological type of primary lung cancer using clinical materials of primary lung cancer. More than 80% of proteins detected on 2-DE gels were not specific for any histological types. However, 2-DE patterns shown by the expression of several proteins reflected the histopathological differentiation of the primary lung cancer. During this investigation we identified 13 proteins associated with histopathological features. Most well-differentiated cases show high expression rates of proteins associated with the same histopathological differentiation and very low expression rates of proteins associated with other histopathological differentiation. On the other hand, in poorly differentiated cases we recognized a relatively wide variation of expression rates in the proteins associated with histopathological differentiation. In addition, when we evaluated primary lung adenocarcinoma, cases with the typical 2-DE pattern for lung adenocarcinoma showed a favorable outcome, and on the other hand the other cases showing the atypical 2-DE patterns showed a relatively poor outcome. The classification based upon the expression of cancer-related proteins may reflect biological characteristics of the tumor as well as histological differentiation.⁴⁾

During the investigation of histopathological differentiation-related proteins, we detected one protein with high intensity only in primary lung adenocarcinoma, TA02. This protein molecule was not expressed in either metastatic lung adenocarcinoma from the other organs or the other types of primary lung cancer except a few cases of large cell lung cancer, and in normal human tissues this molecule was distributed only in type II pneumocyte and a part of the renal tubules.^{5,6)} At present, we understand that TA02 is homologous with napsin A, a new type of aspartic proteinase, which is involved in the maturation of the biologically active form of surfactant apoprotein B (SpB). It is suggested that Napsin A cleaves the N-terminal peptide of SpB, resulting in a 25 kDa intermediate.^{7,8)} We believe that our 2-DE investigation is the first report concerning proteomic analysis using clinical samples of

primary lung cancer.

Hanash and their colleagues at the University of Michigan Medical Center constructed a database that contained protein expression data on lung cancer based upon 2-DE findings. Also, they identified histopathology-related proteins. They described the possibility of developing novel classification schemes for lung cancer and the identification of novel markers for early detection using this kind of 2-DE protein database.⁹⁾ They also investigated 93 tissue samples of lung adenocarcinoma and 10 samples of normal lung tissues, and identified 9 protein molecules with significant overexpression in lung adenocarcinoma.¹⁰⁾ Furthermore, they detected 46 survival-associated proteins by 2-DE. Sequentially, 33 out of these 46 proteins were identified using MS, and among these candidate proteins, phosphoglycerate kinase 1 was validated as a survival-associated protein based upon another investigation of both tissues and serum derived from the patients with non-small cell lung.¹¹⁾

Proteome Platforms Not Involving 2-DE

Though 2-DE is really one of the most powerful tools for proteomic analysis, it does have several shortcomings. It is too intricate to permit automatic analysis, and its reproducibility is sometimes poor. Also, highly abundant proteins interfere with identification of less common proteins on 2-DE gel when either total-cell lysates or tissue lysates are applied to 2-DE analysis. When either cell lysates or tissues lysates are analyzed, approximately 10^6 orders in the dynamic range seem to be required. Furthermore, it is difficult to investigate extremely acidic or basic proteins and hydrophobic proteins, e.g. membrane proteins, which are strongly associated with cellular functions as a growth factor receptor. Therefore, new technology with high through-put and wide dynamic range as well as with high sensitivity have been eagerly awaited.

After bioinformatics made remarkable progress due to the completion of the human genome project, high through-put proteomic technology rapidly developed in the last 10 years. Finally, Fenn and Tanaka received the Nobel Prize for Chemistry in 2002, and proteomic analysis using mass spectrometry (MS) became the center of attention. They developed ionization technology, which is essential for MS. Tanaka developed matrix-assisted laser desorption/ionization (MALDI), and Fenn also developed electro-spray ionization (ESI), which are essential elements in MS instrumentation.

The US Food and Drug Administration (FDA) and

National Cancer Institute in the US set up a Clinical Proteomics Program in 2001. The next year they reported that a specific serum-proteomic pattern of ovarian cancer was identified compared with patients and healthy donors. They used a surface-enhanced laser desorption/ionization (SELDI) MS system, which is an affinity-based MS method using a protein chip modified with a specific chromatographic surface. The SELDI MS system is a modified MALDI MS system, and it is based upon the principle that the proteins from crude mixtures are selectively attached to specific biochemical surfaces. Some proteins as potential biomarkers candidates may show a higher binding affinity to certain surfaces than common serum proteins. Serum samples from healthy donors and disease-affected individuals are processed using this protein chip. After washing steps, matrix is added to the protein spots and each proteomic pattern is acquired. Samples can be classified into a normal group, a disease-affected patients' group and additional groups, through the application of bioinformatic algorithms. SELDI MS analysis does not absolutely rely on the actual identification of the proteins to diagnose a disease. Therefore, this system has been used to detect some biomarkers (proteomic patterns) in complex protein mixtures such as cell lysates, body fluids and serum (Fig. 1).

According to initial reports concerning ovarian cancer by Petricoin et al., the discriminatory proteomic pattern correctly identified all 50 ovarian cancer cases, including 18 stage I cases. Of the 66 cases of non-malignant diseases, 63 were recognized as not being cancer. The sensitivity was 100%, specificity 95%, and positive predictive value 94%.¹²⁾ Furthermore, they investigated early detection of prostate cancer, and suggested that pathologic states within the prostate might be reflected by changes in serum proteomic patterns in relation to serum prostate specific antigen (PSA). The proteomic pattern correctly predicted 36 out of 38 patients with prostate cancer, while 177 out of 228 patients were correctly classified as having benign conditions. For men with marginally elevated PSA levels, the specificity was 71%. They concluded that serum pattern diagnostics might be of value in deciding whether to perform a biopsy on a man with an elevated PSA level.¹³⁾ This new approach is based upon the protein patterns analysis using SELDI and may provide a more effective means to diagnose some kinds of malignant neoplasm, such as ovarian, prostate, breast, and lung cancer. Over the past three years, many investigators have reported that pattern recognition algorithms based upon SELDI MS data may be successfully used to distinguish

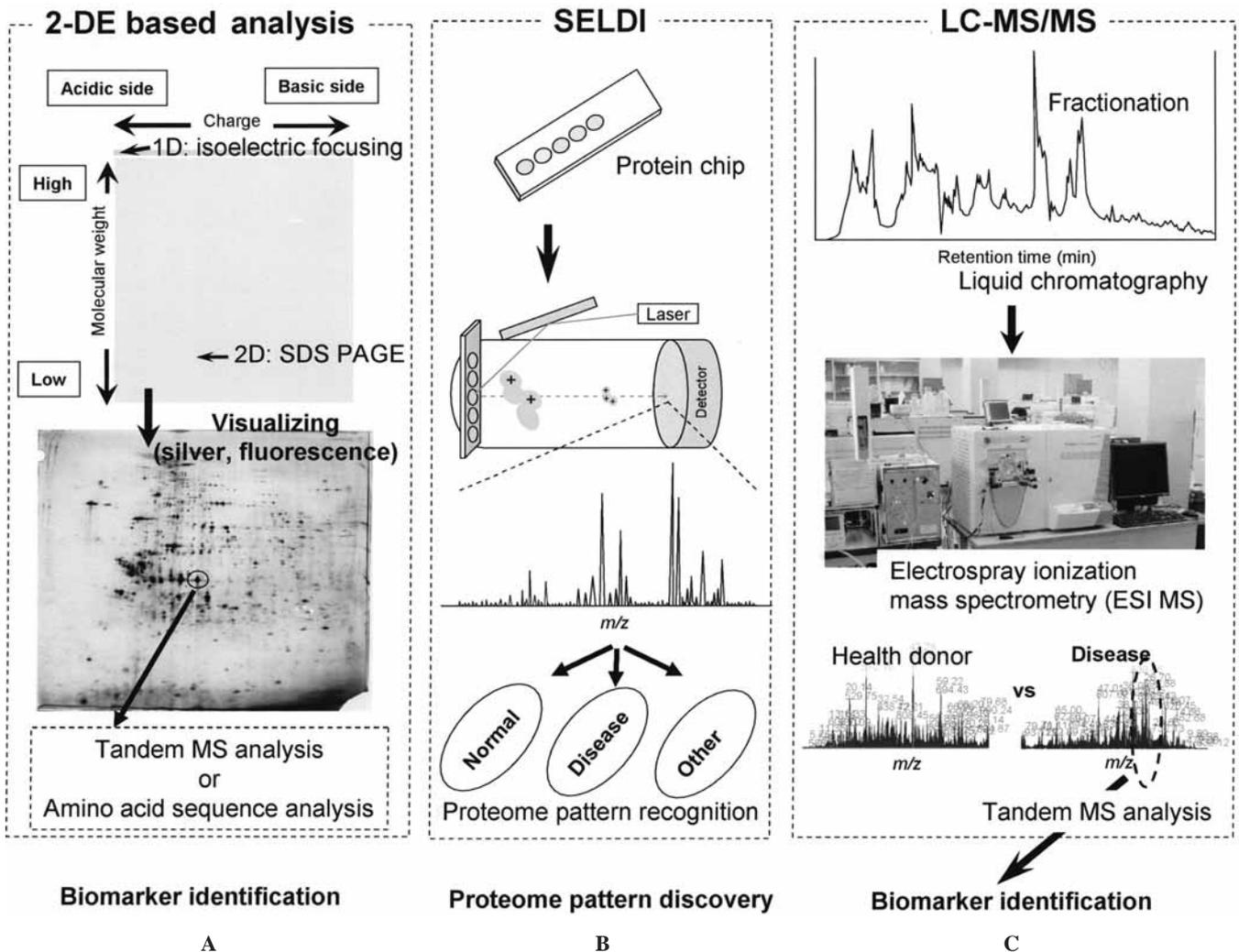


Fig. 1. Strategy for clinical proteomics.

A: Protein identification based upon two-dimensional polyacrylamide gel electrophoresis (2-DE).

A glass tube, to which the gel is cast, is used for isoelectric focusing (IEF). A sample is applied to each tube, and focused for approximately 15 hours. After the IEF gel is extruded into an equilibration buffer, sodium dodecyl sulphate (SDS) polyacrylamide gel (slab gel) is used for the second dimension. The IEF gel is sealed using agarose on the top of the slab gel, and is electrophoresed overnight. Subsequently, 2-DE proteins are visualized by either silver-staining or Comercy Blue staining.

After cutting the gel for extraction of protein, protein molecules are identified using tandem mass spectrometry (MS) or amino acid sequence analysis and bioinformatics.

B: Proteomic pattern discovery using surfaced enhanced laser desorption/ionization mass spectrometry (SELDI MS).

When the SELDI MS system is used for a diagnostic proteomics approach, samples from healthy donors and disease-affected individuals are processed using a protein chip modified with a specific chromatographic surface, and proteomic patterns are obtained as a result of the affinity to the surface of each protein chip. According to bioinformatics algorithms, the raw data are investigated to classify into either the healthy, disease, or other groups. This strategy does not rely on the actual identification of the protein molecules.

C: Protein identification using liquid chromatography-mass spectrometry (LC-MS).

This method relies upon (multidimensional) fractionation and tandem MS for protein-molecule identification. Samples derived from patients with specific diseases are compared with those derived from matched healthy donors. Attempts are made to discover unique or highly abundant proteins. Due to quantitative analysis of peptide-signal intensity and statistical analysis, it is possible to detect statistically significant differences in signal intensity, and the source protein molecule is identified using tandem MS analysis, which is performed sequentially.

between serums derived from normal donors and cancer patients.¹⁴⁻¹⁷⁾

On the other hand, there are some reports criticizing proteomic pattern analysis using SELDI MS as not being reproducible or reliable enough for practical applications.^{18,19)} Finally, they concluded that the proteomic patterns that enable successful classification are biologically implausible and the methods, properly applied, do not classify the data accurately. Though commercial laboratories planned to market a test in late 2003 or early 2004, the US FDA in the US delayed starting clinical applications, because the question has not yet been resolved. We conclude that identification of a biomarker molecule is necessary when using serum biomarkers clinically, even though combination diagnosis with several kinds of biomarkers is undertaken.

Proteomic Analysis of Lung Cancer Using MS

There are few reports concerning proteomic analysis using MS for the investigation of clinical lung cancer materials. Yanagisawa et al. investigated proteomic patterns of non-small cell lung cancer using MALDI-time of flight MS (MALDI-TOF MS). They reported that class-prediction models completely classify histology, distinguish primary tumors from metastatic lesions from other organs to the lung, and classified nodal involvement with 85% accuracy. Also, they obtained a proteomic pattern comprised of 15 distinct MS peaks that distinguish between patients with poor prognosis and good prognosis.²⁰⁾ However, they went no further than recognizing the proteome pattern. Therefore, the molecules related to specific events were not identified in this study.

Tyan investigated pleural effusion fluid derived from lung adenocarcinoma using two-dimensional liquid chromatography (LC) tandem MS, and it was reported that 124 proteins were identified. Based upon previous reports, it was concluded that 69 proteins among these proteins originated from plasma and that another 13 proteins were synthesized in the lung. Finally, a protein database concerning human pleural effusion will provide potential protein diagnostic biomarkers to be examined in further investigations.²¹⁾

We also continue to make efforts to explore biomarkers related to lung cancer using LC-MS. Recently, we established a high-throughput comprehensive protein profiling system comprising a fully automated on-line micro-flow LC/tandem MS system for clinical sample utility. In this system quantitative evaluation of signal intensity ana-

lyzes statistically significant differences between two groups, for instance groups of healthy donors and cancer patients. Furthermore, automatic operation enabled the completion of a single run of entire LC-MS/MS analysis within 11 hours. Investigation of the data extracted from the protein identification datasets of both groups could allow identification of candidate proteins of disease-specific biomarkers. We applied this high throughput micro LC-MS/MS protein profiling system to surgically resected tissues and plasma derived from patients with primary lung adenocarcinoma, and attempted to identify the specific protein-molecules showing statistically significant differences in protein-expression levels. Firstly, we applied this protein profiling system to the investigation of the proteins associated with lymph node metastasis compared with expression profiles of two groups. One group consisted of cases without lymph node involvement, and the other group consisted of cases less than 3 cm in the largest dimension with lymph node involvement. The profiles were accumulated for each group using our originally developed profile alignment program. The statistical selection was done by Student's t test using a p-value of less than 0.005 as a cut-off value. The number of selective data points was 5,889. We have to note that the number of data points is not identical to the number of proteins. Generally, one protein produces several kinds of peptide molecular ions, and each signal is expected to correspond to a single LC-MS signal. Among 5,889 significant points, 2,753 associated with protein identification information were derived from MS/MS data by the MASCOTTM protein identification software. Finally, we identified more than 500 protein molecules as potential biomarkers associated with lymph node involvement. However, a validation process is needed for clinical application of these results.

We also started a project to mine novel biomarkers for the early detection of primary lung adenocarcinoma. We believe that such protein may leak from cancerous tissue but their concentration may be less than 100 femto mol. The lower limitation of the detectable range in LC-MS system is the same order as the concentration of tissue leakage proteins. Plasma proteins consist of a large amount of classical proteins including albumin, globulin, several kinds of complement and fibrinogen, and a very small amount of tissue leakage proteins and interleukins. We used serum albumin- and Immunoglobulin-depleted samples in plasma proteome analysis, and succeeded in identifying more than 100 proteins.²²⁾ However, at present it is difficult to reliably detect a very small

amount of protein such as tissue-leakage proteins derived from malignant solid tumor. We believe that a multi-dimensional analysis system is needed before MS analysis for plasma biomarker discovery, which is the primary aim in clinical plasma proteome projects.

Conclusion

Though recent advances in proteomic analysis are conspicuous, nevertheless current research has not yet achieved the final clinical goal of producing specific biomarkers for the early detection and discovery of new molecular targets for individualized therapy. The target molecules to be detected in proteomic investigations, must be sufficiently abundant to be clinically useful. It is necessary to further improve our techniques concerning sensitivity and wide dynamic range sufficiently to analyze human plasma proteins. We believe that the remaining problems will be resolved within a few years, and that a fundamental revolution in both early detection and therapeutic strategy will occur in the near future.

References

1. Wilkins MR, Sanchez JC, Williams KL, et al. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 1996; **17**: 830–8.
2. O'Farrell PH. High resolutional two-dimensional electrophoresis of proteins. *J Biol Chem* 1975; **250**: 4007–21.
3. Franzen B, Linder S, Okuzawa K, et al. Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* 1993; **14**: 1045–53.
4. Hirano T, Franzen B, Uryu K, et al. Detection of polypeptides associated with the histopathological differentiation of primary lung carcinoma. *Br J Cancer* 1995; **72**: 840–8.
5. Hirano T, Auer G, Maeda M, et al. Human tissue distribution of TA02, which is homologous with a new type of aspartic proteinase, napsin A. *Jpn J Cancer Res* 2000; **91**: 1015–21.
6. Hirano T, Gong Y, Yoshida K, et al. Usefulness of TA02 (napsin A) to distinguish primary lung adenocarcinoma from metastatic lung adenocarcinoma. *Lung Cancer* 2003; **41**: 155–62.
7. Brasch F, Ochs M, Kahne T, et al. Involvement of napsin A in the C- and N-terminal processing of surfactant protein B in type-II pneumocytes of the human lung. *J Biol Chem* 2003; **278**: 49006–14.
8. Ueno T, Linder S, Na CL, et al. Processing of pulmonary surfactant protein B by napsin and cathepsin H. *J Biol Chem* 2004; **279**: 16178–84.
9. Oh JM, Brichory F, Puravs E, et al. A database of protein expression in lung cancer. *Proteomics* 2001; **1**: 1303–19.
10. Chen G, Gharib TG, Huang CC, et al. Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. *Clin Cancer Res* 2002; **8**: 2298–305.
11. Chen G, Gharib TG, Wang H, et al. Protein profiles associated with survival in lung adenocarcinoma. *Proc Natl Acad Sci USA* 2003; **100**: 13537–42.
12. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; **359**: 572–7.
13. Petricoin EF 3rd, Omstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002; **94**: 1576–8.
14. Zheng Y, Xu Y, Ye B, et al. Prostate carcinoma tissue proteomics for biomarker discovery. *Cancer* 2003; **98**: 2576–82.
15. Rosty C, Christa L, Kuzdzal S, et al. Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res* 2002; **62**: 1868–75.
16. Hu Y, Zhang S, Yu J, et al. SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast* 2005; **14**: 250–5.
17. Schwegler EE, Gazares L, Steel LF, et al. SELDI-TOF MS profiling of serum for detection of the progression of chronic hepatitis C to hepatocellular carcinoma. *Hepatology* 2005; **41**: 634–42.
18. Baggerly KA, Morris JS, Edmonson SR, et al. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst* 2005; **97**: 307–9.
19. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst* 2005; **97**: 315–9.
20. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003; **362**: 433–9.
21. Tyan YC, Wu HY, Lai WW, et al. Proteomic profiling of human pleural effusion using two-dimensional nano liquid chromatography tandem mass spectrometry. *J Proteome Res* 2005; **4**: 1274–86.
22. Fujii K, Nakano T, Kanazawa M, et al. Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics* 2005; **5**: 1150–9.