Mass spectrometric analysis of cerebrospinal fluid protein for glioma and its clinical application

Jian Liu[1], Jiekai Yu[2], Hong Shen[2], Jianmin Zhang[2], Weiguo Liu[2], Zhe Chen[2], Shuda He[2], Shu Zheng[2]

[1]Department of Surgery, Second Affiliated Hospital, College of Medicine, Zhejiang Chinese Medical University, Hangzhou, China
[2]Cancer Institute, Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China

Aim of the study: To establish and evaluate the fingerprint diagnostic models of cerebrospinal protein profile in glioma with surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) and bioinformatics analysis, in order to seek new tumor markers.

Material and methods: SELDI-TOF-MS was used to detect the cerebrospinal protein bond to ProteinChip H4. The cerebrospinal protein profiles were obtained and analyzed using the artificial neural network (ANN) method. Fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor, were established. The support vector machine (SVM) algorithm was used for verification of established diagnostic models. The tumor markers were screened.

Results: In a fingerprint diagnostic model of cerebrospinal protein profiles for distinguishing glioma from non-brain tumor, the sensitivity and specificity of glioma diagnosis were 100% and 91.7%, respectively. Seven candidate tumor markers were obtained. In a fingerprint diagnostic model for distinguishing glioma from benign brain tumor, the sensitivity and specificity of glioma diagnosis were 88.9% and 100%, respectively, and 8 candidate tumor markers were gained.

Conclusions: The combination of SELDI-TOF-MS and bioinformatics tools is a very effective method for screening and identifying new markers of glioma. The established diagnostic models have provided a new way for clinical diagnosis of glioma, especially for qualitative diagnosis.

Key words: glioma, cerebrospinal fluid, SELDI-TOF-MS, artificial neural network, support vector machine, diagnostic model, tumor markers.

Introduction

Glioma is a space-occupying lesion seriously endangering human health, with a high incidence in brain tumors. It is often a malignancy and has the greatest perniciousness [1]. In clinical practice, it is sometimes difficult to preoperatively distinguish glioma from other brain tumors, even if using modern imaging technologies. At present, the diagnosis of glioma lacks a tumor marker with effective clinical value. Therefore, seeking new markers of glioma and improving the clinical diagnosis level have been a hotspot in brain tumor research.

Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) is one of the most effective proteomics platforms for detecting protein profiles [2]. The basic principles of SELDI-TOF-MS are as follows: The surface-enhanced proteins are captured by a specific probe, and then are bonded by a microarray of protein biochip. Different proteins are separated according to the peak value formed by the mass and charge ratio (m/z). Each protein obtains one mass spectrum. Then these data are collected and analyzed using appropriate software. In recent years, this method has become one of the main means of finding new tumor markers in the proteomics platform [3]. The analysis software based on artificial neural network (ANN) technique has been successfully applied for analyzing and processing complex data in proteomics [4].

In our early studies, mass spectrometry and bioinformatics analysis were used for studying the serum samples of common tumors including glioma, colorectal cancer, esophageal cancer, breast cancer and ovarian cancer, and we obtained certain results. The specificity of glioma site and blood-brain barrier restrict the application of clinical blood examination. As cerebrospinal fluid directly contacts the brain tissue, the protein profiles in brain tumor can be directly reflected in cerebrospinal fluid. In this study, the SELDI-TOF-MS platform was used to detect the cerebrospinal protein profiles for glioma, and ANN was used for biological analysis of the collected data. Fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor were established. The support vector machine (SVM) algorithm was used for evaluation of established diagnostic models. The candidate tumor markers were screened. The objective was to seek new ways and methods for clinical diagnosis of glioma.

Material and methods

General data

Cerebrospinal fluid samples of glioma and benign brain tumor were collected in the period June 2009 to December 2011 in the Department of Neurosurgery,

Second Affiliated Hospital of Zhejiang University (Hangzhou, China). All samples were collected by preoperative lumbar puncture (inserting needle between the 3rd and 4th lumbar or the 4th and 5th lumbar, or between the 5th lumbar and 1st sacral vertebrae). The postoperative histopathological diagnosis was conducted on all cases. The cerebrospinal fluid was centrifuged at 5000 rpm for 5 min, then stored at –20°C for use. The sample constitutions were as follows: 22 cases (12 males and 10 females) were with glioma (grade 1 and 2, 15 cases; grade 3 and 4, 7 cases). The patients' ages were 28 to 71 years, with median age of 47.3 years. Twenty-five cases were with benign brain tumor, including 14 cases of benign meningioma (8 males and 6 females, median age 53.1 years), 6 cases of cerebral schwannoma, 4 cases of cerebral aneurysm, and 1 case of cerebral cholesteatoma.

Twenty-eight non-brain-tumor patients (17 males and 11 females) with mild traumatic brain injury (according to GCS standards) were from the Department of Neurosurgery, Shaoxing First People's Hospital (China). The cerebrospinal fluid was collected by lumbar puncture, excluding blood contamination. Their ages were 19–70 years, with median age of 48.8 years.

Instruments and analysis software

PBS-II SELDI-TOF-MS platform and ProteinChip H4 were provided by Ciphergen Biosystems, Inc. (USA). ANN in the MATLAB platform was used for biological analysis of the collected data, and SVM was used for verification.

Operation of SELDI-TOF-MS

The cerebrospinal fluid samples were unfrozen in an ice bath, followed by centrifugation at 5000 rpm for 5 min. The protein concentration of each sample was detected using the BIO-RAD DC protein assay kit. The range of protein concentration was from 0.03305 to 6.85031 mg/ml. 60 μl of sample (< 0.50000 mg/ml), 40 μl of sample (0.50000–1.00000 mg/ml), 35 μl of sample (1.00000–3.00000 mg/ml) and 30 μl of sample (> 3.00000 mg/ml) were added to each well of a 96-well plate, respectively. Then an equal amount of 0.5% CHAPS buffer (pH 7.4) was added for balance, followed by adding 20 μM HEPES buffer to adjust the total volume to 160 μl.

ProteinChip H4 was fixed to the Bioprocessor, and was previously balanced with 100 μl of 20 μM HEPES buffer (pH 7.05) 3 times, for 5 min each time. The above steps were conducted on ice (4°C). The treated sample was added to each well of the Bioprocessor, followed by centrifugation (250 rpm, 4°C) for 1 h to remove the unconjugated protein residue. Then the sample was washed with 100 μl of 20 μM HEPES 3 times (5 min each time), followed by washing with 100 μl of deionized water 2 times (for 1 min each time). The protein chip was unloaded, and naturally dried. Then 0.5 μl of CHCA was added to each well, followed by natural drying. These operations were repeated 2 times. Finally the samples were detected by SELDI-TOF-MS.

Data collection and processing

Proteinchip Software 3.0 was used to collect and process data in conditions as follows: laser intensity, 140; sensitivity, 9; optimal range of data, 2000–20 000 Da; collecting position, 20–80; 5 collections for each position, total 65 collections. A standard protein chip was used to adjust the apparatus before collecting data. 2000 to 30 000 m/z peaks were firstly filtered with signal-to-noise ratio (s/n) > 5, and then secondly filtered with S/N > 2. The screened m/z peaks existed in more than 10% samples, and the deviation of one peak value in different samples was less than 0.3%. The noise of original data was removed. ANN and SVM in the MATLAB platform were used to establish diagnostic models in the noise removed training set for distinguishing different groups.

Bioinformatics analysis and grouping

Artificial neural network based on a back propagation (BP) algorithm was used for data analysis and establishment of the diagnostic model. Comparisons between glioma and non-brain-tumor, and between glioma and benign brain tumor were conducted. 2/3 of total samples were selected as the training set to establish the diagnostic model. 1/3 of total samples were selected as the test set for the blind test. The initial screened m/z peaks were arranged from small to large, according to *P* values, and were input from small to large to train the established models. When the sensitivity and specificity no longer increased, this model was defined as the final diagnostic model. At the same time, the SVM algorithm was applied to verify the established models, using screened candidate tumor markers.

Results

Fingerprint diagnostic model of cerebrospinal protein profiles for distinguishing glioma from non-brain tumor

In order to find potential markers for distinguishing glioma from non-brain tumor, comparison was conducted between 22 protein profiles in glioma and 28 protein profiles in non-brain tumor. 65 536 m/z peaks were collected, and 103 m/z peaks were selected by clustering and peak value analysis. Then 7 m/z peaks (6089.602, 7154.886, 6055.822, 7291.292, 16021.94, 18756.25 and 7960.945) were obtained using ANN. These markers composed the optimal set and were used as the input variables of ANN and the final basis for classification (Figs. 1 and 2).

Thirty-three samples were selected as the training set to build a diagnostic model using ANN, and another 17 cases were selected for the blind test. The overall classification accuracy rate of the training set was 100% (33/33), and the accuracy rate of the test set was 94.1% (16/17). The sensitivity of the blind test was 100% (5/5), with specificity of 91.7% (11/12). The positive predictive rate was 83.3% (5/6), with a negative predictive rate of 100% (12/12). Seven markers were used to verify this model by SVM algorithm. The accuracy rate, sensitivity, specificity, positive predictive rate and negative predictive rate were 94.1% (16/17), 85.7% (6/7), 100% (10/10), 100% (7/7), and 90.9% (10/11), respectively (Tables 1 and 2).
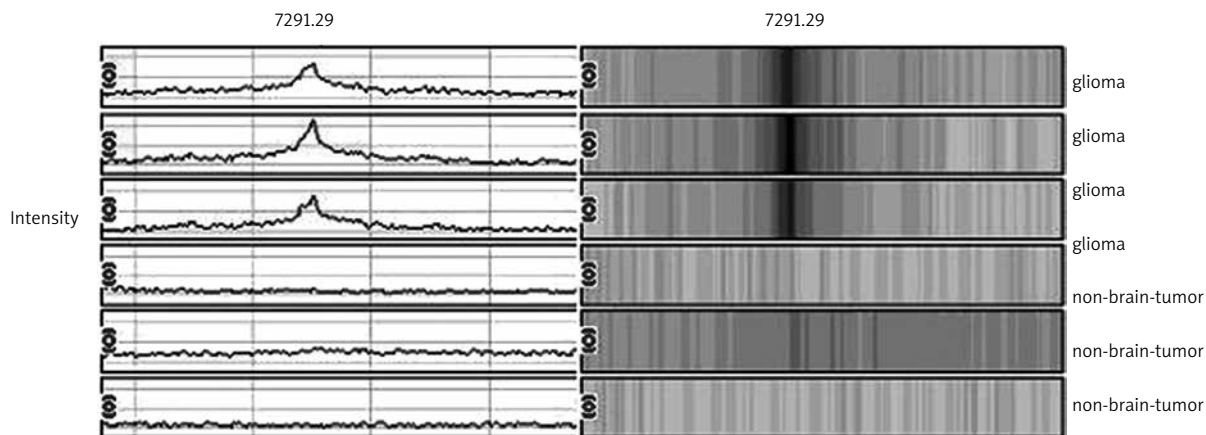
7291.29   7291.29

Intensity

glioma
glioma
glioma
glioma
non-brain-tumor
non-brain-tumor
non-brain-tumor

Fig. 1. Spectra and gel views of marker with 7291.29 m/z (left, MS; right, pseudo-gel; upper three spectra, gliomas; lower three spectra, non-brain tumors)
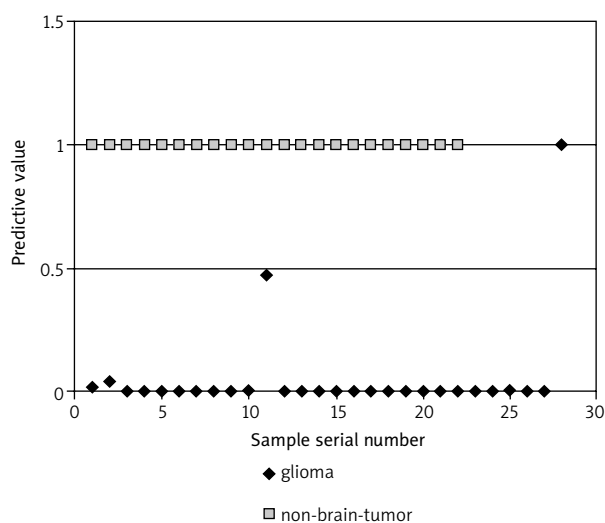

Fig. 2. Distributions of glioma and non-brain-tumor in ANN (predictive value > 0.5, glioma; predictive value ≤ 0.5, non-brain-tumor, only one case of non-brain-tumor was misjudged as glioma)

Fingerprint diagnostic model of cerebrospinal protein profiles for distinguishing glioma from benign brain tumor

Twenty-two protein profiles in glioma and 25 protein profiles in benign brain tumor were compared. Two hundred forty four preliminarily screened m/z peaks were analyzed by ANN, and a total of 47 samples were randomly divided into the training set (31 cases) and the test set (16 cases). After automatic optimization, 8 m/z peaks (3449.645, 7300.375, 16010.23, 6380.50, 8675.707, 3408.88, 17670.94 and 20238.78) were finally selected as markers. Results are

shown in Figures 3 and 4. The accuracy rate, sensitivity, specificity, positive predictive rate, and negative predictive rate were 93.8% (15/16), 88.9% (8/9), 100% (7/7), 100% (9/9) and 87.5% (7/8), respectively. Support vector machine was used to verify these 8 markers. The accuracy rate, sensitivity, specificity, positive predictive rate, and negative predictive rate were 93.8% (15/16), 100% (7/7), 88.9% (8/9), 87.5% (7/8) and 100% (9/9), respectively (Tables 3 and 4).

Discussion

Brain tumors are diseases commonly occurring in adolescents. According to the survey of NCI, the mortality of brain tumor in 2000 has the second place of tumors for adolescents in the USA, with a cure rate of around 30% [6]. Glioma is a brain tumor with the highest morbidity and greatest danger. Improving the preoperative diagnosis level for better prognosis has become one of the hotspots in current medical research. With the development of imaging technologies, the early diagnosis of glioma has made great progress. However, there is still no effective means for preoperative differential and qualitative diagnosis of glioma. The lack of specific biological markers is the main reason for this situation [7]. Tumor is a multiple-gene and multiple-step evolutionary process with interactions of internal and external factors. The tumor markers should be associated with a variety of proteins. A single protein marker could not really reflect the tumor protein expression. It is difficult for previous molecular biological technologies to complete the simultaneous detection of multiple proteins.

In recent years, the rapid development of proteomics has provided a new technical platform for seeking tumor mark-

Table 1. Sensitivity, specificity and accuracy rate for distinguishing glioma from non-brain tumor (ANN)

Tumor	Cases	Sensitivity (%)	Specificity (%)	Accuracy rate (%)
glioma	5	100 (5/5)	8.3 (1/12)	100
non-brain-tumor	12	0	91.7 (11/12)	91.7
total	17	100	100	94.1

Table 2. Sensitivity, specificity and accuracy rate for distinguishing glioma from non-brain-tumor (SVM)

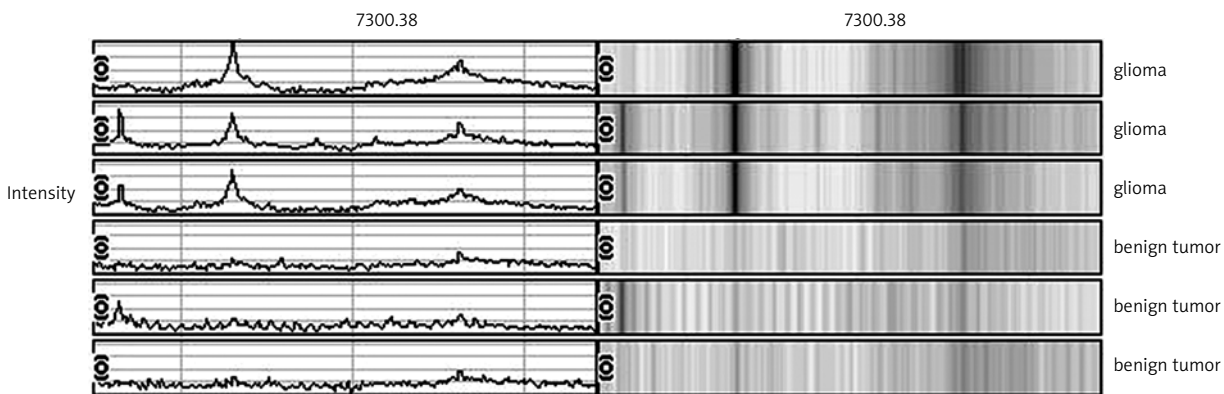| Tumor	Cases	Sensitivity (%)	Specificity (%)	Accuracy rate (%)
glioma	7	85.7 (6/7)	0	85.7
non-brain-tumor	10	14.3 (1/7)	100 (10/10)	100
total	17	100	100	94.1

Fig. 3. Spectra and gel views of marker with 7300.38 m/z (left, MS; right, pseudo-gel; upper three spectra, gliomas; lower three spectra, benign brain tumors)

ers composed of multiple proteins. With the appearance of SELDI-MS, serum tumor markers of prostate cancer, breast cancer, ovarian cancer, lung cancer, colorectal cancer and liver cancer have been found, with sensitivity and specificity higher than previous biological markers [8–13]. SELDI-MS is a protein chip technology platform invented by Ciphergen Biosystems Inc. (USA), based on studies of Hutchens and Yip [14]. This technology has provided a methodological revolution in the field of proteomics [15], and has the advantages of small amount of samples, simple operation, high sensitivity and high throughput, which previous technologies, including liquid chromatography/mass spectrometry (LC-MS), two-dimensional gel electrophoresis-mass spectrometry (2-DE-MS), enzyme-linked immunosorbent assay (ELISA) and the fluorescent labeling method, lack [16]. SELDI-MS can detect trace protein at the fmol (10^{-15} mol) level, and obtain hundreds of thousands of protein data from one sample. It overcomes many difficulties of traditional two-dimensional gel electrophoresis, including separation of membrane protein, separation of strong acidic and basic protein, and detection of low molecular weight and low abundance proteins. As a rapid, reproducible, highly sensitive, easily adoptable means of analysis and diagnostic tool, it has provided an effective technology platform for screening and identification of tumor markers [17]. In addition, this method can obtain a huge amount of data, which is often difficult for traditional data processing methods. Therefore, bioinformatics techniques are indispensable in data analysis and processing.

Artificial neural network based on the BP algorithm is a rapidly developed interdisciplinary subject composed of neuroscience, computer science, information science, and engineering science. It has the advantages of unique
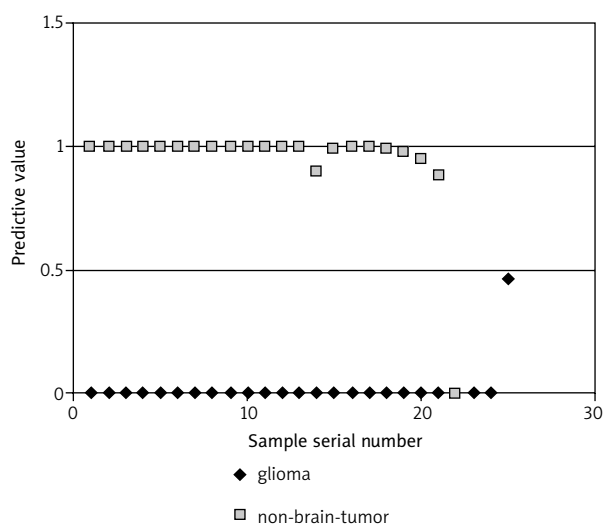

Fig. 4. Distributions of glioma and benign brain tumor in ANN (predictive value > 0.5, glioma; predictive value ≤ 0.5, brain-tumor, only one case of glioma was misjudged as benign brain tumor)

information storage way, good fault tolerance, large scale parallel processing, and strong ability of self-organization, self-learning and self-adapting, and has been used in fields of signal processing, pattern recognition, and prediction, with a wide application prospect [18]. The BP network, proposed by Rumelhart and McClelland in 1986, is a multilayer feedforward network based on back-propagation algorithm. Artificial neural network is a nonlinear dynamic system [19], of which the basic unit is the neuron. Each neuron connects with other neurons through weight value, accepts the output of other neurons, and acts with

Table 3. Sensitivity, specificity and accuracy rate for distinguishing glioma from benign brain tumor (ANN)

Tumor	Cases	Sensitivity (%)	Specificity (%)	Accuracy rate (%)
glioma	9	88.9 (8/9)	0	88.9
benign brain tumor	7	11.1 (1/9)	100 (7/7)	100
total	16	100	100	93.8 (15/16)

Table 4. Sensitivity, specificity and accuracy rate for distinguishing glioma from benign brain tumor (SVM)

Tumor	Cases	Sensitivity (%)	Specificity (%)	Accuracy rate (%)
glioma	7	100 (7/7)	0	100
benign brain tumor	9	0	88.9 (8/9)	88.9
total	16	100	100	93.8 (15/16)

other neurons by transformation of self-conversion function and threshold output. Artificial neural network is composed of several neurons with a single function by parallel distribution. In the BP algorithm, the training samples are afforded with an initial weight value. The information is input from the input layer. After processing in the hidden layer, it is transmitted to the output layer. After further treatment by output layer neurons, the results are obtained. This is as a forward process. If the desired output is not obtained, it transfers to the backward process in which the information flow is the reverse of the forward process. The interlayer connecting weight values are layer-by-layer adjusted. And then the backward process transfers to the forward process, until the error between the actual output and the expected output reaches an acceptable level. Artificial neural network is a bioinformatics algorithm most widely developed and applied in recent years. Support vector machine is a new classification technique proposed by Vapnik *et al.* in 1995 [20]. It is also a learning algorithm based on statistical learning theory, with a principle different from ANN. It provides a new algorithm for a learning machine, according to the principle of structural risk minimization. This technique can overcome large sample requirements of other algorithms, especially suitable for small samples, and can avoid the over-learning of ANN. Hence it has drawn more and more attention.

In this study, SVM was used to verify the ANN results. The principle of SVM is completely different from ANN, but the results of the two methods are very similar. This has verified the reliability of ANN results to some extent. However, the ANN results are comparatively stable. The sensitivity and specificity of most samples are more than 85%. So the ANN results are selected as the final results. In data collection and processing, the noise of original data is removed by two filtrations. A standard protein chip is used to adjust the apparatus before collecting data; this can minimize the deviation.

As cerebrospinal fluid directly contacts with the brain tumor, brain tumor markers are most likely detected in cerebrospinal fluid. In addition, the direct detection of cerebrospinal fluid also avoids the possible influence of the blood-brain barrier. Due to the difficulty of obtaining a cerebrospinal fluid sample in a normal person, nearly normal persons with mild traumatic brain injury (according to GCS standards) are selected as a control (non-brain-tumor), in which the cerebrospinal fluid is normal and non-hemorrhagic. Benign brain tumors include meningioma, neurilemmoma, hemangioma and cholesteatoma. As pituitary adenoma is a type of tumor with special secretory function, and the secreted protein might mask the detection of tumor markers, it is not included in benign brain tumors.

There are many reports about using MS and bioinformatics analysis to find a tumor marker of glioma. However, seeking brain tumor markers in cerebrospinal fluid by these methods is less often reported [21, 22]. In this study, fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor, were established. The diagnostic models employ both cross-validation and a double blind test, and the sensitivity and

specificity are over 85%. They are obviously superior to the previous single biomarker, and possess great potential application value in clinical practice. However, it is still controversial to screen potential tumor markers by SELDI-TOF-MS, due to the instable repeatability of results. The reason might be that the standards of experimental operation (sample processing, species of energy molecule, and correction of standard protein molecule) are not uniform. The established MS models are not applicable in different research groups, and the results from different analysis software are not the same. The strategies of overcoming these drawbacks include standardization of operating method, confirmation of different software and repeated verification of established models. In addition, as cerebrospinal fluid samples are not readily available, the comparisons between glioma with different grades, and between glioma and other malignant tumors, could not be conducted. Expansion of the sample size is required in subsequent research. Not all markers selected only according to *p* value are proteins with biological significance. Therefore, further separation and identification might be necessary. In the next study, one or several of the most valuable tumor markers will be screened by comparison using a network protein database, for further separation and identification. This can further clarify the biological functions of tumor markers in glioma. Related research is in progress. In addition, whether the protein profiles in cerebrospinal fluid also exist in serum, and whether there is a difference if so, will be investigated next.

In conclusion, the combination of SELDI-TOF-MS and bioinformatics analysis is a very effective method for screening and identifying new markers of glioma. The established diagnostic models have provided a new way for clinical diagnosis of glioma, especially for qualitative diagnosis, but the problems such as poor reproducibility should be solved as soon as possible.

References

1. Erick K, James H, Schwarts TM, et al. Principles of Neural Scince. 4th ed. The university of Chicago Press 2001.
2. Srinivas PR, Srivastava S, Hanash S, Wright GL Jr. Proteomics in early detection of cancer. Clin Chem 2001; 47: 1901-11.
3. Ball G, Mian S, Holding F, et al. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics 2002; 18: 395-404.
4. Burke HB. Artificial neural networks for cancer research: outcome prediction. Semin Surg Oncol 1994; 10: 73-9.
5. Decaestecker C, Camby I, Remmelink M, et al. Decision tree induction: a useful tool for assisted diagnosis and prognosis in tumor pathology. Lab Invest 2002; 76: 799-808.
6. Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. CA Cancer J Clin 2003; 53: 5-26.
7. Nambiar U, Dala Y, Rojymon A, et al. Treatment outcome in gliomas: 10-year experience. Clin Neurol Neurosurg 1997; 99: 92.

8. Qu Y, Adam BL, Yasui Y, et al. Boosted decision tree analysis of sur-face-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from non cancer patients. Clin Chem 2002; 48: 1835-43.

9. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 2002; 48: 1296-304.

10. Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Dis-covery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. Lung Cancer 2003; 40: 267-79.

11. Robert C, Bast Jr. Status of tumor markers in ovarian cancer screen-ing. J Clin Oncol 2003; 21: 200-5.

12. Liu J, Zheng S, Yu JK, Zhang JM, Chen Z. Serum protein fingerprint-ing coupled with artificial neural network distinguishes glioma from healthy population or brain benign tumor. J Zhejiang Univ Sci 2005; 6: 4-10.

13. Wang JX, Zhang B, Yu JK, Liu J, Yang MQ, Zheng S. Application of se-rum protein fingerprinting coupled with artificial neural network model in diagnosis of hepatocellular carcinoma. Chin Med J (Engl) 2005; 118: 1278-84.

14. Hutchs TW, Yip TT. New desorption strategies for the mass anal-ysis of macromolecules. Rapid Commun Mass Spectrom 1993; 7: 576-80.

15. Merchant M, Weinberger SR. Recent advancements in surface en-hanced laser desorption/ionization-time of flight-mass spectrom-etry. Electrophoresis 2000; 21: 1164-77.

16. Fung ET, Wright GL Jr, Dalmasso EA. Proteomic strategies for bio-marker identication: progress and challenges. Curr Opin Mol Ther 2002; 2: 643-50.

17. Yip TT, Van de Water J, Gershwin ME, Coppel RL, Hutchens TW. Cryptic antigenic determinants on the extracellular pyruvate de-hydrogenase complex/mimeotope found in primary bleary cirrho-sis. J Biol Chem 1996; 271: 32825-33.

18. Ball G, Mian S, Holding F, et al. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumors and rapid identification of poten-tial biomarkers. Bioinformatics 2002; 18: 395-404.

19. Therneau TM, Grambsch PM, Fleming TR. Martingale-based resid-uals for survival models. Biometrika 1990; 77: 147-53.

20. Guyon L, Weston J, Barnhill S, et al. Gene selection for cancer clas-sification using support vector machines. Machine Learning 2002; 46: 389-422.

21. Saratsis AM, Yadavilli S, Magge S, et al. Insights into pediatric dif-fuse intrinsic pontine glioma through proteomic analysis of cere-brospinal fluid. Neuro Oncol 2012; 14: 547-60.

22. Ohnishi M, Matsumoto T, Nagashio R, Kageyama T, Utsuki S, Oka H, Okayasu I, Sato Y. Proteomics of tumor-specific proteins in cere-brospinal fluid of patients with astrocytoma: usefulness of gel-solin protein. Pathol Int 2009; 59: 797-803.

Address for correspondence

Jian Liu
Department of Surgery, Second Affiliated Hospital
College of Medicine, Zhejiang Chinese Medical University
Hangzhou 310005, China
tel. 0571-85267161
fax 0571-88064725
e-mail: liujiandoc@yahoo.com.cn