Mass Spectrometric Analysis on Cerebrospinal Fluid Protein for Glioma and its Clinical Application
J Liu¹, J Yu², H Shen², J Zhang², W Liu², Z Chen², S He², S Zheng²

ABSTRACT

Objective: To establish and evaluate the fingerprint diagnostic models of cerebrospinal protein profile in glioma Using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) and bioinformatics analysis, for seeking new tumor markers.

Methods: SELDI-TOF-MS was used to detect the cerebrospinal protein bond to ProteinChip H4. The cerebrospinal protein profiles were obtained and analyzed using artificial neural network. The fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor were established, respectively. The support vector machine algorithm was used for verification of established diagnosis models. The tumor markers were screened.

Results: In 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. 7 candidate tumor markers were obtained. In 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 marker of glioma. The established diagnosis models have provided a new way for clinical diagnosis of glioma, especially for qualitative diagnosis.

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

From: ¹Department of surgery, Second Affiliated Hospital, College of Medicine, Zhejiang Chinese Medical University, Hangzhou 310005, China ²Cancer Institute, Second Affiliated Hospital, College of Medicine, Zhejiang University, HangZhou 310009, China

Correspondence: Jian Liu, Department of surgery, Second Affiliated Hospital, College of Medicine, Zhejiang Chinese Medical University, Hangzhou 310005, China. E-mail: jianliucn@126.com
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, it is sometimes difficult to preoperatively distinguish glioma from other brain tumors, even if using modern imaging technologies. At present, the diagnosis of glioma lack 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 specific probe, and then are bond by microarray of protein biochip. Different proteins are separated according to peak value formed by mass and charge ratio (m/z). Each protein obtains one mass spectrum. Then these data are collected and analyzed using corresponding software. In recent years, this method has become one of the main means of finding new tumor marker in proteomics platform (3). The analysis software based on artificial neural network (ANN) has been successfully applied to analyzing and processing complex data in proteomics (4).

In our early studies, mass spectrometry and bioinformatics analysis are used for studying the serum sample of common tumors including glioma, colorectal cancer, esophageal cancer, breast cancer and ovarian cancer, and have 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, SELDI-TOF-MS platform was used to detect the cerebrospinal protein profiles for glioma, and ANN was used for biological analysis on the collected data. The fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor were established, respectively.
The support vector machine (SVM) algorithm was used for evaluation of established diagnosis models. The candidate tumor markers were screened. The objective is to seeking new ways and methods for clinical diagnosis of glioma.

**SUBJECTS AND METHODS**

**General data**

Cerebrospinal fluid samples of glioma and benign brain tumor were collected in June 2009 to December 2011 in Department of Neurosurgery, Second Affiliated Hospital of Zhejiang University (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 centrifugated 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 28-71 years, with median age of 47.3 years. 25 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.

28 non-brain-tumor patients (17 males and 11 females) with mild traumatic brain injury (according to GCS standards) were from 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 softwares**

PBS-II SELDI-TOF-MS platform and ProteinChip H4 were provided by Ciphergen
Mass Spectrometric Analysis for Glioma

Biosystems, Inc. (USA). ANN in MATLAB platform was used for biological analysis on the collected data, and SVM was used for verification.

**Operation of SELDI-TOF-MS**

The cerebrospinal fluid samples were unfrozen in ice bath, followed by centrifugation at 5000 rpm for 5 min. The protein concentration of each sample was detected using 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 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 previously balanced with 100 μL of 20 μM HEPES buffer (pH 7.05) for 3 times, 5 min for each time. Above steps were operated on ice (4 °C). The treated sample was added to each well of 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 for 3 times (5 min each time), followed by washing with 100 μL of deionized water for 2 times (1 min for 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 for 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, 2000-20000 Da; collecting position, 20-80; 5 collection for each position, total 65 collections). Standard protein chip was used to adjust apparatus before collecting data. 2000 to 30000 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 MATLAB platform were used to establish diagnostic models in noise removed training set for distinguishing different groups.

**Bioinformatics analysis and grouping**

ANN based on Backpropagation (BP) algorithm was used for data analysis and establishment of diagnosis 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 training set to establish diagnostic model. 1/3 of total samples were selected as test set for 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 diagnosis model. At the same time, SVM algorithm was applied to verify the established models, using screened candidate tumor markers.

**RESULTS**

Fingerprint diagnosis 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. 65536 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 final basis for classification (Fig. 1 and 2).

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.

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

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

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Cases</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioma</td>
<td>5</td>
<td>100 (5/5)</td>
<td>8.3 (1/12)</td>
<td>100</td>
</tr>
<tr>
<td>Non-brain-tumor</td>
<td>12</td>
<td>0</td>
<td>91.7 (11/12)</td>
<td>91.7</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100</td>
<td>100</td>
<td>94.1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Cases</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioma</td>
<td>7</td>
<td>85.7 (6/7)</td>
<td>0</td>
<td>85.7</td>
</tr>
<tr>
<td>Non-brain-tumor</td>
<td>10</td>
<td>14.3 (1/7)</td>
<td>100 (10/10)</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>100</td>
<td>100</td>
<td>94.1</td>
</tr>
</tbody>
</table>

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

22 protein profiles in glioma and 25 protein profiles in benign brain tumor were compared. 244 preliminarily screened m/z peaks were analyzed by ANN, and a total of 47 samples were randomly divided into training set (31 cases) and 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 were shown in Fig. 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. SVM was used to verify these 8 markers. The accurate 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 (Table 3 and 4).

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

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Cases</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioma</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign brain tumor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor</td>
<td>Cases</td>
<td>Sensitivity (%)</td>
<td>Specificity (%)</td>
<td>Accuracy rate (%)</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Glioma</td>
<td>7</td>
<td>100 (7/7)</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Benign brain tumor</td>
<td>9</td>
<td>0</td>
<td>88.9 (8/9)</td>
<td>88.9</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>100</td>
<td>100</td>
<td>93.8 (15/16)</td>
</tr>
</tbody>
</table>

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

DISCUSSION

Brain tumors are the diseases commonly occurring in adolescents. According to the survey of NCI, the mortality of brain tumor in 2000 accounts for the second place of tumors for adolescents in UAS, with a cure rate around 30% (6). Glioma is a brain tumor with the highest morbidity and greatest danger. Improving preoperative diagnosis level for better prognosis has become one of hotspots in current medical research. With the development of imaging technologies, the niveau diagnosis of glioma has obtained great progress. However, there is still no effective means for preoperative differential and qualitative diagnosis of glioma. Lacking of specific biological markers is the main reason of this situation (7). Tumor is a multiple-gene and multiple-step evolution process with interactions of internal and external factors. The tumor markers should be associated with a variety of proteins. Single protein marker can 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 markers composed of multiple proteins. With the appearance of SELDI-MS, the
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, 9, 10, 11, 12, 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 proteomics field (15), and has advantages of small amount of samples, simple operation, high sensitivity and high throughput, of which previously technologies including liquid chromatography/mass spectrometry (LC-MS), Two-dimensional gel electrophoresis-mass spectrometry (2-DE-MS), enzyme-linked immunosorbent assay (ELISA) and fluorescent labeling method are lack (16). SELDI-MS can detect trace protein with fmol (10^{-15} mol) level, and obtain hundreds of thousands of protein data of 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 analysis means and diagnostic tool, it has provided an effective technology platform for screening and identification of tumor marker (17). In addition, this method can obtain huge amount of data, which is often difficult for traditional data processing methods. Therefore, the bioinformatics means are indispensable in data analysis and processing.

ANN based on BP algorithm is a rapidly developed interdisciplinary subject composing of neuroscience, computer science, information science, and engineering science. It has advantages of unique 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). BP network, proposed by Rumelhart and McClelland in 1986, is a multilayer feedforward network based on back-propagation algorithm. ANN is a nonlinear dynamic system (19), of which the basic unit is neuron. Each neuron connects with other neurons through weight value, accepts output of other
neurons, and acts with other neurons by transformation of self-conversion function and threshold output. ANN is composed of several neurons with single function by parallel distribution. In BP algorithm, the training samples are afforded with initial weight value. The information is input from the input layer. After processing in 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 backward process in which the information flow is reverse to forward process. The interlayer connecting weight values are layer-by-layer adjusted. And then the backward process transfers to forward process, until the error between actual output and the expected output reaches an acceptable level. ANN is a bioinformatics algorithm most widely developed and applied in recent years. SVM 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 learning machine, according to 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. So it has caused more and more attention.

In this study, SVM is used to verify the ANN results. The principle of SVM is completely different from ANN, but the results of 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 of samples are more than 85%. So ANN results are selected as the final results. In data collection and processing, the noise of original data is removed by two filtrations. Standard protein chip is used to adjust apparatus before collecting data, this can minimize the deviation.

As cerebrospinal fluid directly contacts with 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 blood-brain barrier. Due to the difficult of obtaining cerebrospinal
fluid sample in normal person, the nearly normal persons with mild traumatic brain injury (according to GCS standards) are selected as control (non-brain-tumor), in which the cerebrospinal fluid is normal and non-hemorrhagic. The 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 may 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 tumor marker of glioma. However, seeking brain tumor markers in cerebrospinal fluid by these methods is less reported (21, 22). In this study, the fingerprint diagnostic models of cerebrospinal protein profiles for distinguishing glioma from non-brain-tumor, and distinguishing glioma from benign brain tumor were established. The diagnosis models employ both cross validation and double blind test, and the sensitivity and specificity are over 85%. This is obviously superior to previous single biomarker, and possesses great potential application value in clinical. However, it is still controversial for screening potential tumor markers by SELDI-TOF-MS, due to the instable repeatability of results. The reasons may 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, cannot be conducted. Expansion of sample is required in subsequent researches. Not all markers selected only according to $P$ value are proteins with biological significance. Therefore, further separation and identification may be necessary. In next study, one or several most valuable tumor markers will be screened by comparison using network protein database, for further separation and identification.
This can further clarify the biological functions of tumor markers in glioma. The related researches are in progress. In addition, whether the protein profiles in cerebrospinal fluid also exist in serum, and whether there is difference if so, will be investigated next.

CONCLUSIONS

The combination of SELDI-TOF-MS and bioinformatics analysis is a very effective method for screening and identifying new marker of glioma. The established diagnosis 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.
REFERENCES


10. Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Discovery of distinct protein


20. Guyon, L , Weston, J , Barnhill S , M D, Vapnik V. Gene selection for Cancer Classification
Mass Spectrometric Analysis for Glioma


1Department of surgery, Second Affiliated Hospital, College of Medicine, Zhejiang Chinese Medical University, Hangzhou 310005, China

2Cancer Institute, Second Affiliated Hospital, College of Medicine, Zhejiang University, HangZhou 310009, China

Correspondence: Jian Liu, Department of surgery, Second Affiliated Hospital, College of Medicine, Zhejiang Chinese Medical University, Hangzhou 310005, China. E-mail: jianliucn@126.com

The work is supported by Project of Zhejiang Provincial Health Department (No. 2009A156)

**Running title:** Mass Spectrometric Analysis for Glioma

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