Identification of novel biomarker candidates by proteomic analysis of cerebrospinal fluid from patients with moyamoya disease using SELDI-TOF-MS

Yoshio Araki¹²§, Kazuhiro Yoshikawa²³§, Sho Okamoto¹§, Masaki Sumitomo¹§, Mikio Maruwaka¹§ and Toshihiko Wakabayashi¹§

¹Department of Neurosurgery, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

²Promoting Center for Clinical Research, Aichi Medical University School of Medicine, Nagakute-cho, Aichi 480-1195, Japan

³Center for Genetic and Regenerative Medicine, Nagoya University Hospital, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan

*These authors contributed equally to this work.

§Corresponding author

Email addresses:

YA: yusei.a.kanna@gmail.com
KY: yoshikaw@aichi-med-u.ac.jp

SO: soknouge@med.nagoya-u.ac.jp

MS: robmachado6925@yahoo.co.jp

MM: ciaomikio@yahoo.co.jp

TW: wakabat@med.nagoya-u.ac.jp
Abstract

Background

Moyamoya disease (MMD) is an uncommon cerebrovascular condition with unknown etiology characterized by slowly progressive stenosis or occlusion of the bilateral internal carotid arteries associated with abnormal vascular network. MMD is a major cause of stroke, specifically in the younger population. Diagnosis is based on only radiological features as no other clinical data are available. The purpose of this study was to identify novel biomarker candidate proteins differentially expressed in the cerebrospinal fluid (CSF) of patients with MMD using proteomic analysis.

Methods

For detection of biomarkers, CSF samples were obtained from 20 patients with MMD and 12 control patients. Mass spectral data were generated by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS) with an anion exchange chip in three different buffer conditions. After expression difference mapping was undertaken using the obtained protein profiles, a comparative analysis was performed.

Results

A statistically significant number of proteins (34) were recognized as single biomarker candidate proteins which were differentially detected in the CSF of patients
with MMD, compared to the control patients (p<0.05). All peak intensity profiles of
the biomarker candidates underwent classification and regression tree (CART)
analysis to produce prediction models. Two important biomarkers could successfully
classify the patients with MMD and control patients.

Conclusions

In this study, several novel biomarker candidate proteins differentially expressed in
the CSF of patients with MMD were identified by a recently developed proteomic
approach. This is a pilot study of CSF proteomics for MMD using SELDI technology.
These biomarker candidates have the potential to shed light on the underlying
pathogenesis of MMD.
Moyamoya disease (MMD) is characterized by progressive stenosis or occlusion of the bilateral internal carotid arteries associated with compensatory abnormal vascular network, so called moyamoya vessels [1]. A Japanese survey of 2075 patients with MMD found an annual incidence of 0.35, annual prevalence of 3.16 per 100,000 and a tendency of occurrence in the younger generation [2]. According to a world distribution analysis, a relatively large number of patients with MMD are present in East Asia, but rarely in Europe and the Americas [3]. Such regional and racial differences in susceptibility and familial occurrence in 10% of MMD cases [4] suggest that a genetic predisposition may be associated with the etiology and pathogenesis of this disease. Both 3p24-36 [5] and 8q23 [6] in genome-wide analyses, in addition to both 6q25 (D6S441) [7] and 17q25 [8] in chromosomal level analyses, have been identified in familial MMD and were recognized as possible sources of MMD. Pathologically, stenosis or occlusion of internal carotid arteries has been attributed to eccentric fibrocellular thickening of the intima following proliferation and necrosis of smooth muscle cells, which are associated with the thinning of the media [9-11]. These processes are reported to be regulated by the expression of several growth factors related to angiogenesis: hypoxia inducible factor-1, vascular
endothelial growth factor, basic fibroblast growth factor, transforming growth factor-β, hepatocyte growth factor and metalloproteinases [12-16]. An increasing number of reports have been focusing on not only angiogenesis related to growth factors but vasculogenesis. Vasculogenesis is considered the pathway for adult neovascularization [17-19], which induces the formation of new blood vessels from circulating bone marrow-derived endothelial progenitor cells rather than from local endothelial cells regulated by growth factors [17]. It has been hypothesized that aberrant vasculogenesis contributes to vascular abnormalities including MMD [20-22]. Despite the establishment of approaches for clarifying the disease mechanisms of MMD, the direct underlying pathogenesis remains unclear. One approach utilizing proteomics has revealed disease-associated proteins as novel biomarkers and characterized their function in pathogenesis and development of the disease. Among the many different types of approaches for CSF investigation, SELDI-TOF-MS technology [23] allows for high-throughput analysis of samples with diverse functionalization on surfaces (ProteinChip, Bio-Rad Laboratories, Hercules, CA, USA) and has been successfully used to identify protein profiles of central nervous system disorders [24-26]. The objective of this study was to identify candidate proteins in the CSF for use as biomarkers of MMD using SELDI-TOF-MS technology.
Methods

Patient Population

For proteomic analysis, CSF samples were prospectively collected from a consecutive sequence of 20 patients with MMD (11 male and 9 female; mean age, 21 years; range 1-54 years) admitted to Nagoya University Hospital, Nagoya, Japan, between February 2008 and December 2009. Diagnosis of MMD was determined by cerebral angiography or magnetic resonance imaging/angiography according to the guidelines set by the Research Committee on Moyamoya Disease (Spontaneous Occlusion of Circle of Willis) of the Ministry of Health and Welfare of Japan [27]. Type of onset of MMD includes 12 transient ischemic attacks, four cerebral infarctions, one hemorrhage and two asymptomatic cases. According to the guidelines for the diagnosis of MMD, “definite MMD” cannot have originated from any other underlying disease; the presence of an associated disease is classified as “quasi-MMD”. On the other hand, for proteomic analysis of the CSF used in this study, results are thought to be unsusceptible to the underlying disease. That is, both classifications of the disease (definite or quasi) have similar clinical characteristics and treatments. Therefore, two of the patients in this study presenting with Down syndrome were categorized as MMD (Table 1). As controls, 12 patients were
recruited from among those admitted to Nagoya University Hospital for surgical
treatment of other clinical conditions, including nine cases of cerebral aneurysm. The
control patients were matched according to gender; however, their overall age was
older than that of patients with MMD, as patients with MMD tend to be younger than
those with cerebral aneurysm (75% of control patients), which mainly affects middle-
aged individuals.

Sample Collection

All CSF samples were collected after obtaining informed written consent from the
patients following approval from the Nagoya University School of Medicine Ethical
Review Board. CSF was obtained for all subjects in the same manner during the first
phase of the operative procedure after the arachnoid membrane was dissected. Blood
contamination was avoided as much as possible by carefully performing hemostasis.
CSF sampling was performed at least 4 weeks after the stroke event associated with
MMD or other pathogenesis. Within 2 h of collection, the CSF sample was stored in
1.5-ml cryo-tubes, centrifuged at 3000 rpm for 10 min at 4°C, and the supernatants
were frozen and stored in approximately 1-ml aliquots at -80°C until analysis.

Sample preparation and analysis
CSF preparation

The CSF samples in each cryo-tube were thawed in water at room temperature, then kept on ice and well agitated before subjected to ProteinChip array. The samples were diluted by each binding/washing buffer (0.2 M citrate-phosphate, pH 5 or 7, or 0.2 M Tris-HCl, pH 9) at a final ratio of 1/5.

ProteinChip array pretreatment

Q10 (strong anion exchanger) ProteinChip array (Bio-Rad Laboratories, Hercules, CA, USA) was used for protein profile analysis. The chips were equilibrated prior to the CSF sample addition with binding/washing buffer for 2x5 min.

ProteinChip array preparation

Diluted CSF (20 µl) was applied on a spot in the chip and incubated for 40 min in a humid box at room temperature. After allowing the sample to bind, the remaining sample was removed from all spots by immediately washing the array with the corresponding binding/washing buffer 3x5 min. The arrays were rinsed with 150 µl distilled water twice and allowed to air-dry for 15 min. Then, 1 µl saturated energy absorbing molecule (EAM) solution (sinapinic acid in 50% acetonitrile and 0.5%
trifluoroacetic acid) was applied to each spot and again allowed to air-dry before analysis by the ProteinChip reader.

**SELDI-TOF-MS and data acquisition**

**Mass spectral processing**

The protein mass spectral data was generated with ProteinChip System 4000 SELDI-TOF mass spectrometer (Enterprise version; Bio-Rad Laboratories) using automated data collection protocol with Ciphergen Express version 3.0.6 software interface (Bio-Rad Laboratories). The optimal laser intensity range for spectra generation was predetermined to be 2800 to 3500 nJ by manual laser shots. For determining the maximum protein peak yield and spectral reproducibility, data were collected between 0 and 100 kDa with the ion focus mass set to 7000 Da and the matrix attenuation to 1000 Da. After two initial warming shots, 53 points equally distributed over the chip spot surface were lased 10 times for ionization of the proteins.

**Peak detection of SELDI protein profiles**

All the obtained spectra were internally mass-calibrated and normalized to the total ion current of an m/z value more than 1000 for avoiding the signal interference from EAM. Expression difference mapping (EDM) was performed automatically using
Ciphergen Express data management software version 3.0.6 under the following conditions: signal/noise ratio of 5 or higher for the first pass, 2 for the second pass, and presentation in at least 20% of spectra for identification.

**Data analysis for single biomarker candidate protein identification**

A comparative analysis of protein profiles of the MMD and control groups using univariate analysis was performed for single biomarker candidate protein identification under each pH condition. P values were calculated on the basis of the Mann-Whitney $U$ test for non-parametric data or two-tailed $t$-test for parametric data, and the area under the receiver operation characteristic (ROC) curve (AUC) of each peak cluster was determined by Ciphergen Express data management software version 3.0. A P value of less than 0.05 was considered statistically significant.

**Data mining using classification and regression trees (CART)**

Generally, stability and accuracy of the prediction model were tested to assign a training set and test set from the data set. However, as the size of the data set in this study was too small to construct an independent validation set, a tree-based model was adopted to make a prediction model for multi-biomarker candidate protein identification. This model applies non-linear regression analysis and has recently been
used broadly for data mining. CART for data mining was applied as described previously [28, 29]. A decision tree was generated using entropy and the Gini index for calculating a node [30]. Branching of the tree is created based on the values (node) calculated by these indices and pruned using the complexity parameter. In this study, the complexity parameter was set to 0.01. This analysis was performed with the R software environment for statistical computing (R Development Core Team, Vienna, Austria).

**Peak reproducibility**

Reproducibility of peak intensity and mass accuracy were evaluated using pooled CSF samples from two individuals from the control group. The coefficient of variance (CV) was calculated using randomly selected multiple protein peaks over the experiment as previously described [31, 32]. For the assessment of peak reproducibility of the SELDI profiles, both the intra-assay (spot-to-spot) and inter-assay (chip-to-chip) CV were determined.

**Results**

**Peak reproducibility**
In this study, the intra-assay (spot-to-spot) CV was 17.38% for peak intensity and 0.11% for mass accuracy. The inter-assay (chip-to-chip) CV was 25.11% for peak intensity and 0.14% for mass accuracy, indicating acceptable reproducibility of the spectra.

SELDI-TOS-MS protein profiling and data analysis

SELDI-TOF-MS analysis for CSF samples

After making the EDM cluster, the CSF samples were analyzed in a binding/washing buffer with three different pH conditions (pH 5, 7 and 9). Between 1 to 100 kDa m/z, 54, 40 and 59 m/z peak clusters were generated for the pH 5, 7 and 9 conditions, respectively; the respective minimum peaks were 1027.73, 1041.82 and 1026.39, and the respective maximum peaks were 93797.02, 94648.40 and 95278.35.

Univariate analysis using EDM data for identification of single biomarker candidate proteins

For identification of single biomarker candidates in each pH condition, univariate analysis was performed using EDM data. Table 2 shows 34 biomarker candidate proteins differentially expressed among the MMD and control groups (P < 0.05); m/z was less than 10 kDa and ranged from 1041.03 to 50989.55, for which 15 proteins
were significant (P < 0.01). Of the 34 biomarker candidates, 15, 5 and 4 were detected under pH 5, 7 and 9 conditions, respectively. The mean ROC AUC value of each cluster was 0.7723, ranging from 0.7 to 0.9. Among these protein candidates, 19 were up-regulated and 15 were down-regulated in the MMD group compared to the control group. There were 11 peak clusters thought to be originated from 5 m/z peak clusters; m/z 4473.46 at pH 5 and 4475.10 at pH 7, m/z 4589.34 at pH 9, 4588.73 at pH 5 and 4588.87 at pH 7, m/z 6941.48 at pH 9 and 6942.96 at pH 7, m/z 13882.19 at pH 9 and 13877.49 at pH 7. Within these clusters, there is a minimal difference in m/z (<5.3). Spectra of representative single biomarker candidates under each pH condition using SELDI ProteinChip analysis are shown in Fig. 1 for each CSF sample from the MMD and control groups. All proteins were up-regulated in the MMD group. A box-whisker plot of the peak intensities for representative single biomarker candidate proteins in each group is demonstrated in Fig. 2. For the MMD group, the mean peak intensity of m/z 4473 at pH 5 was 22.64±9.72 (mean±standard deviation), that of m/z 4588 at pH 7 was 56.67±28.42, and that of m/z 4746 at pH 9 was 44.94±18.76; those of the control group were 11.78±5.16 30.62±8.42, and 20.37±10.24, respectively. All peak intensities of each cluster were significantly larger in the MMD than control group (P < 0.01).
Data mining using CART analysis

CART analysis was undergone to discriminate patients with MMD from control patients using a single biomarker obtained by SELDI-TOF-MS under different pH conditions (pH 5, 7 and 9) (Fig. 3). The analysis correctly classified all 20 patients with MMD at pH 5 and 19 of them at pH 7 and 9; all control patients were classified correctly under all pH conditions. As shown in Fig. 4, CART analysis was able to classify 19 of the 20 patients with MMD and all 12 of the control patients.

Discussion

Purpose of this study

Diagnosis of MMD is currently only achievable by radiological features using cerebral angiography or magnetic resonance imaging/angiography. However, characteristics of MMD-mimicking vascular conditions are frequently encountered on radiological imaging in the clinical setting [33-35], sometimes rendering indeterminate diagnosis of MMD. In this pilot study, we utilized novel biomarker candidate proteins identified in CSF samples by SELDI-TOF-MS to aid in the definite diagnosis of MMD. Our results suggest the mechanism by which these biomarkers are associated with the pathogenesis of MMD.
Study design

Many factors have limited the advancement of basic research on MMD. Some of these include low mortality rate, difficulty of obtaining surgical specimens from the internal carotid artery or related intracranial vessels, and lack of an appropriate animal model [36]. Recent investigations of MMD have focused on its epidemiology [2-4], pathology [9-16], relationship with endothelial progenitor cells [17-19] and genetics [5-8], identifying some potential mechanisms and development pathways of this condition. However, their findings constitute only a part of the underlying cause of MMD. As the CSF contacts the extracellular space of the brain, it contains an obvious source of biomarkers which should reflect central neuro-pathologic impairments of the brain. With the development of global proteomic research for biomarker discovery [37-39], previous studies have proposed many biomarkers, especially for neuro-degenerative diseases, although mainly using two-dimensional polyacrylamide electrophoresis [40, 41]. A variety of recent proteomic techniques based on MS have also been utilized, including electrospray ionization tandem MS, matrix-assisted laser desorption/ionization MS and SELDI MS, with mechanically different methods of ion separation, mass accuracy and resolution, along with different pretreatments of low abundant proteins, such as found in the CSF [42]. Our study focused on proteomic analysis using SELDI-TOF-MS, a high-throughput, reliable technique which allows
large numbers of patient samples to be investigated with a low sample volume scale, simple maneuvering and comparatively shorter time than other methods. SELDI technology has been used for a number of clinical situations and sample types to identify biomarker protein expression patterns which discriminate patients with a certain disease from control individuals [24-26].

We tested Q10 (strong anion exchanger) ProteinChip array to capture a wide variety of proteins derived from the CSF, including physiologically active substances ionized under physiological pH (7.2 – 7.4). We also tested CM10 (weak cation exchanger) ProteinChip array, which can indiscriminately detect many kinds of biomarker proteins, although no meaningful results were obtained using our methodology (data not shown). Q10 ProteinChip array was therefore applied in this study using a binding/washing buffer at pH 5, 7, and 9, around physiological pH conditions. This is the first study to identify biomarker candidate proteins by SELDI-TOF-MS using the CSF of patients with MMD.

**Evaluation of the results in this study**

**Definite diagnostic tool for discrimination of patients with MMD and control patients**
In the course of our study using SELDI-TOF-MS, 34 candidates of single biomarker proteins within the range of 1-50 kDa were generated with Q10 ProteinChip array. SELDI technology can feasibly resolve these low molecular weight proteins which cannot be resolved by other current proteomic methods like two-dimensional electrophoresis. As a definite diagnostic tool of MMD, analysis of these proteins could adequately discriminate between the MMD and control groups (P < 0.05 and ROC AUC 0.7 - 0.9). CART analysis enabled this discrimination to provide a prediction model with relatively higher sensitivity and specificity, although the study population was too small to construct independent training and test set subjects. Therefore, a larger study population is needed for validation of these results.

**Comparison analysis using over-expressed/down-regulated proteins in the CSF of patients with MMD**

**Angiogenic factors identified in the CSF of patients with MMD**

Previous studies have suggested several growth factors that are related to angiogenesis. These include hypoxia inducible factor-1, vascular endothelial growth factor, basic fibroblast growth factor, transforming growth factor-β, hepatocyte growth factor [12-16] and metalloproteinases [17], with m/z 92670, 38200, 55960, 83133, 42005 Da, respectively. As these angiogenic factors have relatively larger
molecular weights (>10 kDa), they are more accessible to purification and exact identification using conventional methodologies. Exact identification of biomarker proteins is undoubtedly advantageous, although as in this study, purification is difficult to achieve with a lack of abundant samples from clinical subjects. Moreover, protein identification requires specialized techniques, and a highly sensitive method is needed to properly identify the target protein, especially those with a low molecular weight.

**Relationship of biomarker candidate proteins obtained in this study and angiogenic factors**

Recent evidence has suggested that the hypoxic condition attributed to arterial stenosis or occlusion in MMD induces collateralization through the formation of moyamoya vessels [43]. This process may be regulated by the expression of a variety of growth factors. On the other hand, the biomarker candidate proteins obtained in this study may be angiogenic-related substances themselves or specifically cleaved forms of these substances, although our results provide insufficient information to confirm this hypothesis. In addition, it is obscure whether these biomarkers are associated with the underlying pathology or are generated in the course of disease development.
Inferred proteins related to MMD pathophysiology

For a further understanding of the underlying pathophysiology of MMD status, the most significant m/z peaks were elucidated on the proteome-wide database level. Candidate protein (peptide) biomarkers were inferred using ExPASy protein database (http://www.expasy.org) based on the definite molecular mass; the retrieval conditions were set as pI 1.0 to 14.0 and m/z±0.5%. Six proteins (peptides) had corresponding molecular weights of two candidates, m/z 4473 and 4588, obtained by CART analysis from the 34 single biomarker candidate proteins (Table 3). These proteins were over-expressed in patients with MMD.

One of six proteins, LL-37, was identified as an antibacterial protein (peptide). Antibacterial proteins are effector molecules of the innate immune system with microbicidal and proinflammatory activities [44]. A recent report revealed an additional function of LL-37—an angiogenic role on endothelial cells that appears to differ from the other mechanism of angiogenesis. That is, it inhibits the ubiquitin-proteasome-dependent degradation of hypoxia inducible factor-1α, suggesting that vessel growth may be mediated by attracting neutrophils and monocytes which contain abundant angiogenic mediators that are released in response to cellular activation [45]. No pathological evidence has established the tissue specificity of LL-
37 in the central nervous system. Expression of LL-37 in this area could be obtained through immunohistochemical or western blotting analysis, validating its reported pathological features [9-11], as it has no underlying inflammatory involvement in MMD.

The other one of six proteins, LEAP-2, was originally purified from human blood ultrafiltrate [46], and its native form contains two disulfide bonds [47]. Although its biological function is not well known, a previous study demonstrated that chemically synthesized LEAP-2 has antimicrobial activities against several bacteria, similar to LL-37 [46-48]. Molecular cloning of its cDNA produces a 77-residue precursor which is processed to the mature LEAP-2 peptide (38-77) of 40 residues. In mammals, expression of this protein is highly conserved, with its transcripts being mainly expressed in a few tissues, including the brain (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.337588).

One of several reports suggested that the antimicrobial role of LEAP-2, with its unique primary structure, is not its main biological function. Indeed, further studies are needed to provide additional biological functions and information related to its disease pathogenesis [47].
Conclusions

Several novel biomarker candidate proteins differentially expressed in patients with MMD and control patients were identified using a recently developed proteomic approach. This is a pilot study for CSF proteomics of MMD using SELDI technology. Multi-biomarker analysis using CART made the prediction models more specific and exhibited two m/z panels as a discriminative tool for the definite diagnosis of MMD. For further understanding into the pathogenesis of MMD, a larger number of CSF samples for validation of the analysis and well-defined methodology and techniques for exact identification of biomarkers are essential.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YA and KY carried out SELDI and participated in the subject evaluations. YA, SO, SM and MM contributed to collecting CSF samples. YA, KY, SO and TW participated in study design and coordination and drafting the manuscript. All authors read and approved the final manuscript.
Acknowledgements

We are grateful to Dr. Minoru Fukayama (Aichi Medical University) for his technical assistance with SELDI data acquisition.

References


34. Sebire G: **Transient cerebral arteriopathy in childhood.** *Lancet* 2006, **368:**8-10.


   **Transient cerebral arteriopathy: a disorder recognized by serial angiograms in children with stroke.** *J Child Neurol* 1998, **13:**27-32.


40. Rohlff C, Southan C: **Proteomic approaches to central nervous system disorders.** *Curr Opin Mol Ther* 2002, **4**:251-258.


43. Achrol AS, Guzman R, Lee M, Steinberg GK: **Pathophysiology and genetic factors in moyamoya disease.** *Neurosurg Focus* 2009, **26**:E4.


Figure Legends

Figure 1. Mass spectra of representative single biomarker candidate proteins in CSF under different pH conditions.

Protein profiles of the MMD and control groups were generated using Q10 (strong anion exchanger) array. For each pH condition, the upper two spectra are protein profiles obtained between m/z 2,000 and 10,000, and the lower two spectra are expansions showing the peak intensities around m/z 4473, 4588 and 4476 for pH 5, 7 and 9, respectively. All representative peaks (red arrows) are larger for the MMD than control group under each pH condition, as determined by SELDI-TOF-MS.

Figure 2. Peak intensity of representative single biomarker candidate proteins in the MMD and control groups.

Peak intensities of representative single biomarker candidate proteins detected using SELDI-TOF-MS are significantly larger in the MMD than control group under each pH condition (m/z 4473, 4588 and 4476 for pH 5, 7 and 9, respectively). The box-whisker plots indicate the median value (thick line) and the 25th (lower line of box) and 75th (upper line of box) percentile; T bars indicate the 10th and 90th percentile. The p values between the groups were calculated using Mann-Whitney U test.

*P<0.01.
Figure 3. CART analysis using peaks obtained by SELDI-TOF-MS to discriminate between patients with MMD and control patients.

The decision tree was constructed using CSF samples from 32 patients with MMD and control patients. The classification is determined starting at the roof node, following by appropriate splitting decisions based on the peak intensity at each node. If the peak intensity is lower than the cutoff intensity value, the left node is selected. This splitting process is continued until no further classification is achieved and terminal nodes are produced. Using m/z 4473, 2406 and 6338 peaks (pH 5), m/z 4588 and 7250 peaks (pH 7), and m/z 4746 and 1044 peaks (pH 9), CART for Q10 ProteinChip was applied to identify patients with MMD and control patients. The analysis correctly classified all 20 patients with MMD under pH 5 condition and 19 of 20 under the pH 7 and 9 conditions; all 12 control patients were classified under all pH conditions.

Figure 4. CART analyses using all 34 single biomarker candidate proteins.

CART was performed for 34 single biomarker candidate proteins identified in the CSF under each pH condition (pH 5, 7 and 9) to discriminate patients with MMD from control patients. This analysis correctly classified 19 of 20 patients with MMD and all 12 control patients based on the peak intensities of the m/z 4473 peak (pH 5) and m/z 4588 peak (pH 7).
### Tables

#### Table 1. Clinical characteristics of each patient group in this study

<table>
<thead>
<tr>
<th>Variables</th>
<th>No. of patients (%)</th>
<th>MMD n=20</th>
<th>Controls n=12</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Mean±SD, years</td>
<td></td>
<td>20.5±17.4</td>
<td>61.2±9.1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>11 (55.0)</td>
<td>3 (25.0)</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9 (45.0)</td>
<td>9 (75.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of onset, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>TIA</td>
<td>12 (60.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infarction</td>
<td>4 (20.0)</td>
<td>2 (16.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>1 (5.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>2 (10.0)</td>
<td>7 (58.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>1 (5.0)</td>
<td>3 (25.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associated disease, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>None</td>
<td>18 (90.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Down syndrome</td>
<td>2 (10.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebral aneurysm</td>
<td>9 (75.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>3 (25.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suzuki’s angiographical stage (1-6)</td>
<td>Mean±SD</td>
<td>3.05±0.9</td>
<td></td>
<td>NA</td>
</tr>
</tbody>
</table>

TIA, transient ischemic attack; MMD, moyamoya disease; SD, standard deviation.
Table 2. Single biomarker candidates from univariate analysis of CSF sample comparison of MMD and control groups

<table>
<thead>
<tr>
<th>P-value*</th>
<th>m/z</th>
<th>pH of binding/washing buffer</th>
<th>ROC AUC</th>
<th>Level in MMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60E-04</td>
<td>4746.76</td>
<td>pH 9</td>
<td>0.9000</td>
<td>up</td>
</tr>
<tr>
<td>5.32E-04</td>
<td>4566.71</td>
<td>pH 5</td>
<td>0.8583</td>
<td>down</td>
</tr>
<tr>
<td>6.14E-04</td>
<td>4473.46</td>
<td>pH 5</td>
<td>0.8750</td>
<td>up</td>
</tr>
<tr>
<td>1.08E-03</td>
<td>4157.65</td>
<td>pH 9</td>
<td>0.8250</td>
<td>up</td>
</tr>
<tr>
<td>1.41E-03</td>
<td>4589.34</td>
<td>pH 9</td>
<td>0.8250</td>
<td>up</td>
</tr>
<tr>
<td>1.41E-03</td>
<td>50989.55</td>
<td>pH 9</td>
<td>0.8417</td>
<td>down</td>
</tr>
<tr>
<td>1.41E-03</td>
<td>4588.73</td>
<td>pH 5</td>
<td>0.8250</td>
<td>up</td>
</tr>
<tr>
<td>3.09E-03</td>
<td>4588.87</td>
<td>pH 7</td>
<td>0.8000</td>
<td>up</td>
</tr>
<tr>
<td>3.09E-03</td>
<td>6941.48</td>
<td>pH 9</td>
<td>0.8083</td>
<td>down</td>
</tr>
<tr>
<td>3.51E-03</td>
<td>6942.96</td>
<td>pH 7</td>
<td>0.8167</td>
<td>down</td>
</tr>
<tr>
<td>7.24E-03</td>
<td>9656.51</td>
<td>pH 5</td>
<td>0.8000</td>
<td>up</td>
</tr>
<tr>
<td>7.24E-03</td>
<td>13754.72</td>
<td>pH 5</td>
<td>0.7750</td>
<td>up</td>
</tr>
<tr>
<td>8.12E-03</td>
<td>4809.94</td>
<td>pH 5</td>
<td>0.7667</td>
<td>down</td>
</tr>
<tr>
<td>9.11E-03</td>
<td>2406.55</td>
<td>pH 5</td>
<td>0.7667</td>
<td>down</td>
</tr>
<tr>
<td>9.11E-03</td>
<td>13882.19</td>
<td>pH 9</td>
<td>0.7583</td>
<td>down</td>
</tr>
<tr>
<td>1.02E-02</td>
<td>9728.65</td>
<td>pH 5</td>
<td>0.7750</td>
<td>up</td>
</tr>
<tr>
<td>1.14E-02</td>
<td>6228.21</td>
<td>pH 5</td>
<td>0.7333</td>
<td>down</td>
</tr>
<tr>
<td>1.27E-02</td>
<td>13877.49</td>
<td>pH 7</td>
<td>0.7917</td>
<td>down</td>
</tr>
<tr>
<td>1.42E-02</td>
<td>21299.65</td>
<td>pH 5</td>
<td>0.7583</td>
<td>down</td>
</tr>
<tr>
<td>1.58E-02</td>
<td>5781.11</td>
<td>pH 5</td>
<td>0.7500</td>
<td>up</td>
</tr>
<tr>
<td>1.76E-02</td>
<td>14060.91</td>
<td>pH 9</td>
<td>0.7417</td>
<td>down</td>
</tr>
<tr>
<td>1.76E-02</td>
<td>4695.40</td>
<td>pH 5</td>
<td>0.7417</td>
<td>down</td>
</tr>
<tr>
<td>1.95E-02</td>
<td>6240.88</td>
<td>pH 9</td>
<td>0.7500</td>
<td>up</td>
</tr>
<tr>
<td>2.16E-02</td>
<td>28095.59</td>
<td>pH 9</td>
<td>0.7583</td>
<td>down</td>
</tr>
<tr>
<td>2.40E-02</td>
<td>4475.1</td>
<td>pH 7</td>
<td>0.7500</td>
<td>up</td>
</tr>
<tr>
<td>2.40E-02</td>
<td>7250.89</td>
<td>pH 7</td>
<td>0.7250</td>
<td>up</td>
</tr>
<tr>
<td>2.40E-02</td>
<td>3516.49</td>
<td>pH 9</td>
<td>0.7250</td>
<td>up</td>
</tr>
<tr>
<td>2.93E-02</td>
<td>6888.58</td>
<td>pH 5</td>
<td>0.7250</td>
<td>up</td>
</tr>
<tr>
<td>3.23E-02</td>
<td>3681.83</td>
<td>pH 9</td>
<td>0.7250</td>
<td>up</td>
</tr>
<tr>
<td>3.56E-02</td>
<td>1041.03</td>
<td>pH 9</td>
<td>0.7000</td>
<td>up</td>
</tr>
<tr>
<td>3.91E-02</td>
<td>7146.84</td>
<td>pH 5</td>
<td>0.7333</td>
<td>down</td>
</tr>
<tr>
<td>4.30E-02</td>
<td>4811.33</td>
<td>pH 9</td>
<td>0.7000</td>
<td>up</td>
</tr>
<tr>
<td>4.71E-02</td>
<td>1061.94</td>
<td>pH 9</td>
<td>0.7250</td>
<td>up</td>
</tr>
<tr>
<td>4.71E-02</td>
<td>34344.75</td>
<td>pH 5</td>
<td>0.7083</td>
<td>down</td>
</tr>
</tbody>
</table>

\* Values with the same lower-case letter indicate clusters most likely derived from the same proteins. *P<0.05, Mann-Whitney U test
Table 3. Biomarker candidate proteins (peptides) in the CSF of patients with MMD

<table>
<thead>
<tr>
<th>Theoretical MW (Da)</th>
<th>Inferred candidate protein (peptide)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4449.89</td>
<td>Oxyntomodulin (P01275)</td>
</tr>
<tr>
<td>4451.26</td>
<td>Urocortin-2 (Q96RP3)</td>
</tr>
<tr>
<td>4461.22</td>
<td>Beta-defensin 133 (Q30KQ1)</td>
</tr>
<tr>
<td>4493.32</td>
<td>Antibacterial protein LL-37 (P49913)</td>
</tr>
<tr>
<td>4585.34</td>
<td>Liver-expressed antimicrobial peptide 2 (Q969E1)</td>
</tr>
<tr>
<td>4586.6</td>
<td>Proenkephalin-A (143-183) (P01210)</td>
</tr>
</tbody>
</table>

The number in parentheses is the serial number of the protein in ExPASy database.
Figure 1

- pH 5
  - m/z 4473
- pH 7
  - m/z 4588
- pH 9
  - m/z 4746
Figure 2

- pH 9
- pH 7
- pH 5

Peak intensity vs. pH for Control and MMD m/z 4746, m/z 4588, and m/z 4473.