

SCIENTIFIC REPORTS



OPEN

A plasma proteomics method reveals links between ischemic stroke and MTHFR C677T genotype

Zhenchang Zhang¹, Qi Yan¹, Jia Guo¹, Xueping Wang¹, Wei Yuan¹, Lei Wang¹, Lixia Chen¹, Gang Su² & Manxia Wang¹

Methylene Tetrahydrofolate Reductase (MTHFR) catalyzes the conversion of methylene tetrahydrofolate to methyl tetrahydrofolate. The 677th nucleotide of the MTHFR gene is often regarded as a risk factor of cardiovascular disease. Previous studies demonstrated an elevated risk of ischemic stroke with the MTHFR677TT genotype. In this study, we employed a plasma proteomics method to investigate the connection between the polymorphism of the target nucleotide and stroke. In total, 28 protein spots were differentially expressed between the two groups, and of which, 25 protein spots were up-regulated and 3 were down-regulated. Five randomly selected spots were successfully identified as Haptoglobin (HPT) and Transferrin (TRFE). A functional analysis indicated that most of the differential expressed proteins (DEPs) were related to the inflammatory immune response. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that these DEPs were involved in the complement cascade reaction. Meanwhile, protein-protein interactions (PPIs) analysis highlighted the novel association between the C677T MTHFR genotype and Vitamin D binding protein (DBP), which was confirmed by a molecular genetic analysis. The results suggested that the phenotype of the MTHFR might be associated with multiple proteins that have a synergistic effect, which might be related to the mechanism of ischemic stroke.

Numerous polymorphisms have been identified after the completion of the sequencing of the human genome¹. However, among these polymorphisms, the phenotypic effects in cells are largely unknown, including how the polymorphisms impact physiological function, whether they are associated with diseases and so on. There is a common polymorphism in the gene encoding the catalytic domain of methylene tetrahydrofolate reductase (MTHFR), namely there is a C-T substitution at position 677 of the gene sequence, which leads to an alanine to valine switch²⁻⁴. The polymorphism of the 677th nucleotide in MTHFR gene is regarded as a risk factor for cardiovascular disease (CVD)^{5,6}. This polymorphism leads to splice site changes, decreased mRNA stability and changes of the enzymatic protein conformation, which leads to a reduction or loss of MTHFR activity, thereby increasing the level of homocysteine (HCY) in the plasma⁷. While the latter is an independent risk factor for obstructive peripheral arterial diseases, such as cardiovascular, cerebrovascular and peripheral diseases and it is also an independent risk factor for venous thrombosis⁸. Decreased enzyme activity is caused by the MTHFR genetic polymorphism, and this is a potential risk factor for the progression of hyperhomocysteinemia, but the relationships are still controversial⁷⁻⁹.

A proteomic analysis of plasma is a relatively practical snapshot tool to compare and analyze the expression atlas under different physiological and pathological genetic statuses¹⁰. This technology provides an opportunity to further understand disease pathogenesis, screen for biomarkers and early diagnosis and treatment¹¹, especially, when we focus on the study of the overall proteins' expression and the relationship between the genotype and phenotype¹².

Epidemiologic studies have revealed that a high risk of venous thrombosis is associated with the MTHFR 677TT genotype and arterial stroke¹³. However, whether this polymorphism directly participates in the progression of CVD remains unclear. In this study, we used plasma proteomics to examine whether stroke patients with the C677T polymorphisms show obviously changes in protein expression. This analysis can give some hints to explain the occurrence of stroke.

¹Department of Neurology, the Second Hospital of Lanzhou University, Lanzhou, 730030, China. ²School of Basic Medical Sciences, Lanzhou University, Lanzhou, China. Correspondence and requests for materials should be addressed to G.S. (email: sugang@lzu.edu.cn) or M.W. (email: wmx322@yeah.net)

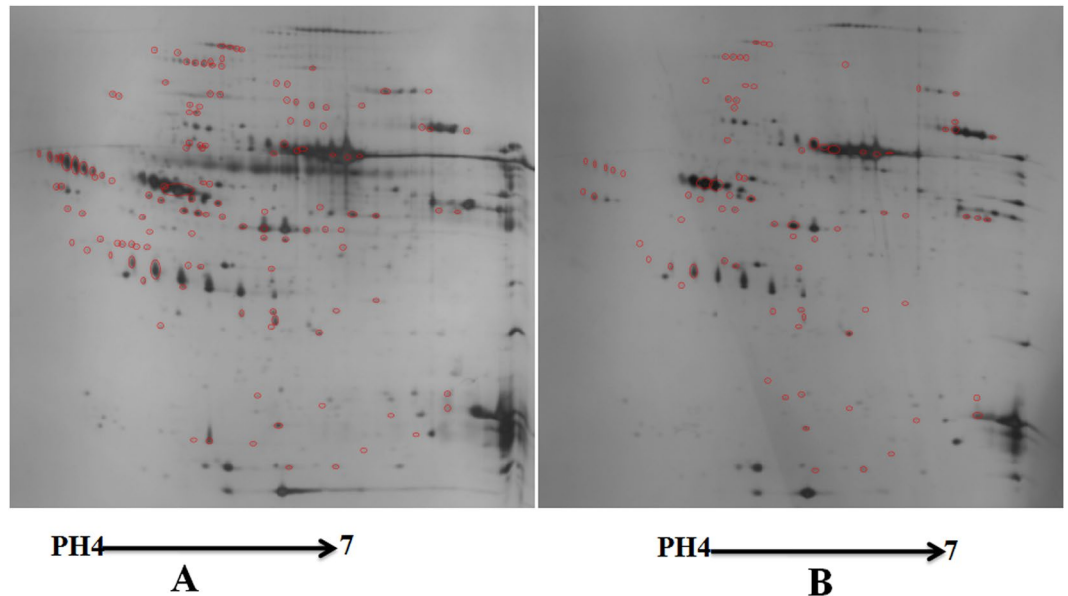


Figure 1. Results of the 2D-PAGE reveal the profiles of differential expression proteins (DEP) between the C/C genotypes and T/T genotypes. The red circles in the figure are the differential protein spots. A means the T/T genotypes and B means the C/C genotypes.

Results

Establishment of the differential protein expression spectrum. This study used 2D-gel electrophoresis to analyze plasma proteins in ischemic stroke patients with the MTHFR genotypes C/C and T/T. The results of the counts between the groups were 259 ± 12 and 195 ± 17 , respectively. The mean match ratio was 84.75%, and the correlation coefficient was 77.9% (>70%). A total of 137 different spots were found after the comprehensive comparison (Fig. 1). The correlation coefficient of the scatter diagram was 0.78 (>0.4), and the reproducibility was consistent among the groups (Supplementary Figure S1). Twenty-eight proteins were selected that had significant differences that were greater than double. Moreover, the differential expression proteins (DEPs) were the same between each of the repeats. (Supplementary Figure S2). The expression of 25 protein spots was up-regulated, while 3 protein spots were down-regulated in the mutant (TT) individuals (Table 1).

Mass spectrum identification results of the differential protein spots. Four protein spots and 3 serum albumin spots (as controls) were selected from the 28 protein spots for analysis using the SWISS-2DPAGE database to identify the spectrum using matrix assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) after extracting, digesting and performing enzymolysis on the spots. In total, 5 spots were successfully identified as two types of proteins via the Mascot Wizard software (Table 2). Using the PDquest software, the peptides of protein spot # SSP0301 were analyzed by MALDI-TOF-MS (Fig. 2A). Peptide 6 matched with the protein haptoglobin (HPT). The coverage rate of the amino acids was 14%. Moreover, the experimental relative molecular mass of peptide 6 and its theoretical relative molecular mass were exactly matched. The Mascot score was 212 (Mascot value over 67 means $p < 0.05$), which revealed that the MS identification results were significant and reliable ($P < 0.05$). Additionally, using the PDquest software, the peptides of protein spot SSP8703 were analyzed by MALDI-TOF-MS (Fig. 2B). Peptide 8 matched with the protein transferrin (TRFE). The coverage rate of the amino acids was 16%. Moreover, the experimental relative molecular mass of peptide 8 and its theoretical relative molecular mass were exactly matched. The Mascot score was 632, which revealed that the MS identification results were significant and reliable ($P < 0.05$).

Bioinformatics analysis of the differentially expressed proteins (DEPs). Figure 3 shows the 3D structure of the human MTHFR protein based on similarity to the 3apy.1.A domain (Protein Data Bank (PDB) code: 3apy.1.a), and the target amino acid is labeled and the region presented is Ays48–Thr340.

A GO (gene ontology) enrichment analysis was performed using GOEAST (<http://omicslab.genetics.ac.cn/GOEAST/index.php>). For an in-depth study, this website is convenient to use, and the data can be displayed as graphical result charts, which provide intuitive data representation. To obtain an overview of the differential expression proteins (DEPs) involved in this study, all 16 identified DEPs were used to perform a GO annotation and an enrichment analysis; the GO annotation results are listed in Supplementary Table S1. The biological processes, cellular components, and molecular functions are shown in Fig. 4. From the perspective of biological processes, homeostasis (GO: 0006879, FDR: $1.1e-21$) and transition metal ion transport (GO: 0000041, FDR: $1.74e-18$) were the top two significantly enriched terms. From a cellular component perspective, endopeptidase inhibitor activity (GO: 0004866, FDR: $1.24e-15$) was the top significantly enriched term. Ferric iron binding (GO: 0008199, FDR: $2.25e-14$) was also significantly enriched. From the molecular function perspective, the extracellular region (GO: 0005576, FDR: $2.2e-20$) was the top significantly over-represented term.

PDQuest No. (SSP)	Accession No. SWISS-PROT	Protein name	Gene name	pI	Mw (KD)	Fold change*
0515	P01011	AACT	SERPINA3	4.73	61.591	>10
0609	P01011	AACT	SERPINA3	4.70	60.696	>10
0605	P01011	AACT	SERPINA3	4.70	62.603	>10
1401	P02678	FIBG	—	5.07	51.347	8.84
1402	P02678	FIBG	—	5.13	51.207	4.31
2303	P02678	FIBG	—	5.34	49.566	2.23
4302	P02678	FIBG	—	5.56	48.107	2.01
3304	P02678	FIBG	—	5.44	48.107	3.47
0513	P02765	FETUA	AHSG	4.61	56.313	>10
3208	P02766	TTHY	TTR	5.52	35.391	3.72
0507	P01009	A1AT	SERPINA1	4.99	53.626	3.42
0505	P01009	A1AT	SERPINA1	4.95	54.065	2.36
1809	P00450	CERU	CP	5.20	123.705	4.33
2810	P99003	IGSA	—	5.23	93.722	4.13
0203	P25311	ZA2G	AZGP1	4.91	40.977	3.31
0301	P00738	HPT	HP	4.88	44.344	2.68
2205	P00738	HPT	HP	5.38	38.749	2.46
1803	P00450	CERU	CP	5.09	126.365	2.18
2705	P08697	A2AP	SERPINF2	5.17	66.097	>10
0802	P09871	C1S	C1S	4.81	88.849	>10
0803	P09871	C1S	C1S	4.83	88.173	>10
5202	P99004	NA3	—	5.77	37.824	2.68
1806	P00734	THRB	F2	5.09	80.154	2.65
0304	P27169	PON1	PON1	4.84	45.937	2.56
4101	P02743	SAMP	APCS	5.55	26.280	2.32
2501	P02774	VTDB	GC	5.24	53.918	0.43
8703	P02787	TRFE	TF	6.50	79.545	0.35
1302	P06727	APOA4	APOA4	5.16	43.627	0.20

Table 1. Differential expression proteins (DEP) between the C/C genotypes and the T/T genotype. Note: *Indicates the ratio of the protein expression level of the C/C genotypes/T/T genotypes.

PDQuest No. (SSP)	Accession No.	Protein name	pI	Mw(KD)	Score	Matches
0301	P00738	HPT	6.13	45.861	212	6(6)
8703	P02787	TRFE	6.81	79.294	632	8(8)

Table 2. Identify the differential expression proteins (DEPs) by MALDI-TOF-MS.

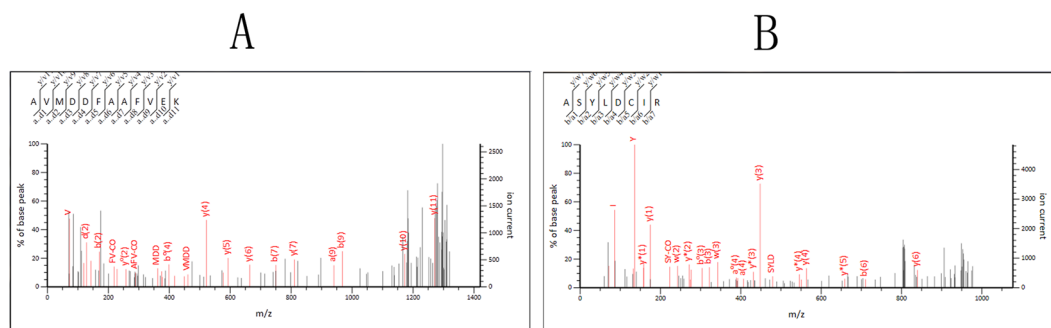


Figure 2. Identification of two randomly selected differential expression proteins (DEP) by using MALDI-TOF-MS (MS/MS) spectrum for parts of the peptides of these proteins.

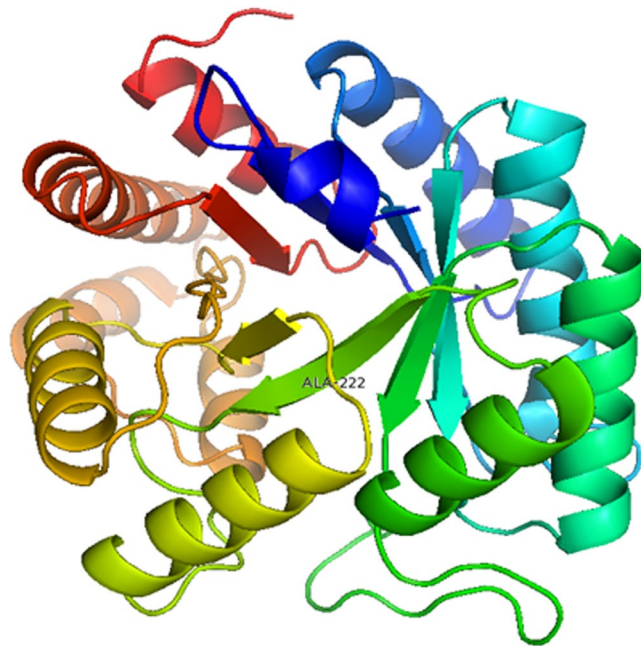


Figure 3. The ensemble of the selected structures is in stereo view. The figure was produced using the Swiss-model and pyMOL programs.

Upon further molecular function analysis, sixteen proteins were classified in 10 categories (DAVID, <https://david.ncicrf.gov/home.jsp>)¹⁴. The enrichment score was more than 6.02 (>1.3). The results showed that the biological processes were mainly involved in the acute-phase response and the inflammatory response (Supplementary Table S2). Moreover, four differential expression proteins (DEPs), including F2, SERPANI, SERPANF2, and C1S, were found to be involved in the cascade reaction of the complement system via the KEGG analysis (Fig. 5)^{15–17}. In addition, to identify the interactions between these differential expression proteins, a protein-protein interactions (PPIs) analysis was conducted by using STRING (<http://string-db.org/>). All of the 16 differential expression proteins were used to construct the interaction network (Fig. 6). To improve the coverage of the PPIs analysis, the confidence level (score) was set without a threshold. SERPINA1, TTR, and HP resided in the central part of the network and were both connected to each other, suggesting the crucial roles of these differential expression proteins (DEPs) in the whole network. Additionally, the MTHFR protein can build the connection to this network through the PON1 protein, which indicates that the MTHFR protein may be a kind of inducing factor that affects the entire cascade network. Furthermore, the C1S protein had no relationship with the other differential expression proteins (DEPs), which suggested that this protein works independently and/or with other connections with this network, which have not been deciphered.

Discussion

The C677T polymorphism in the gene MTHFR is often associated with cardiovascular diseases (CVD) because it can directly affect folic acid metabolism and also enhance the level of plasma homocysteine, which are considered independent risk factors for ischemic stroke, thrombosis, CVD, and venous thrombosis⁷. In this study, we identified different types of proteins that are expressed in the plasma of stroke patients with the C677T polymorphism.

The differentially expressed proteins mentioned above were mainly involved in the immuno-inflammatory response and the complement cascade. This finding is consistent with previous research studies¹⁸. For example, NF- κ B activation experiments in cultured cells were associated with increased MTHFR mRNA, and the co-transfection of NF- κ B and the promoter constructs demonstrated an MTHFR up-regulation by at least 2-fold through its downstream promoter in Neuro-2a cells¹⁸. Inflammation plays an important role in the formation of atherosclerosis, and numerous evidences revealed that the occurrence and development of atherosclerosis are associated with an inflammatory reaction mechanism^{19,20}. Moreover, long-term mild inflammation is also a vital risk factor for ischemic stroke and transient ischemic attack²¹. The current study provides evidence that proteins that participate in the inflammatory reaction are related to stroke and are linked to the MTHFR polymorphism. For example, we discovered that the HPT protein played an important role in the interaction with other differentially expressed proteins. Previous studies indicate that HPT is closely associated with cardiovascular and cerebrovascular diseases. For example, Kiga C., *et al.* found that HPT was a candidate marker for stroke in spontaneously hypertensive rat models²². Holme I., *et al.* examined 342,125 patients that died in Swedish hospitals over the course of 11 years and found that high levels of HPT were a significant predictor of ischemic stroke²³. Additionally, Brea D., *et al.* found that the levels of HPT and amyloid plaque were higher in patients with atherosclerotic stroke than cardiogenic stroke. The evidence previously presented indicates that HPT is an important biomarker in the diagnosis of atherosclerotic stroke²⁴.

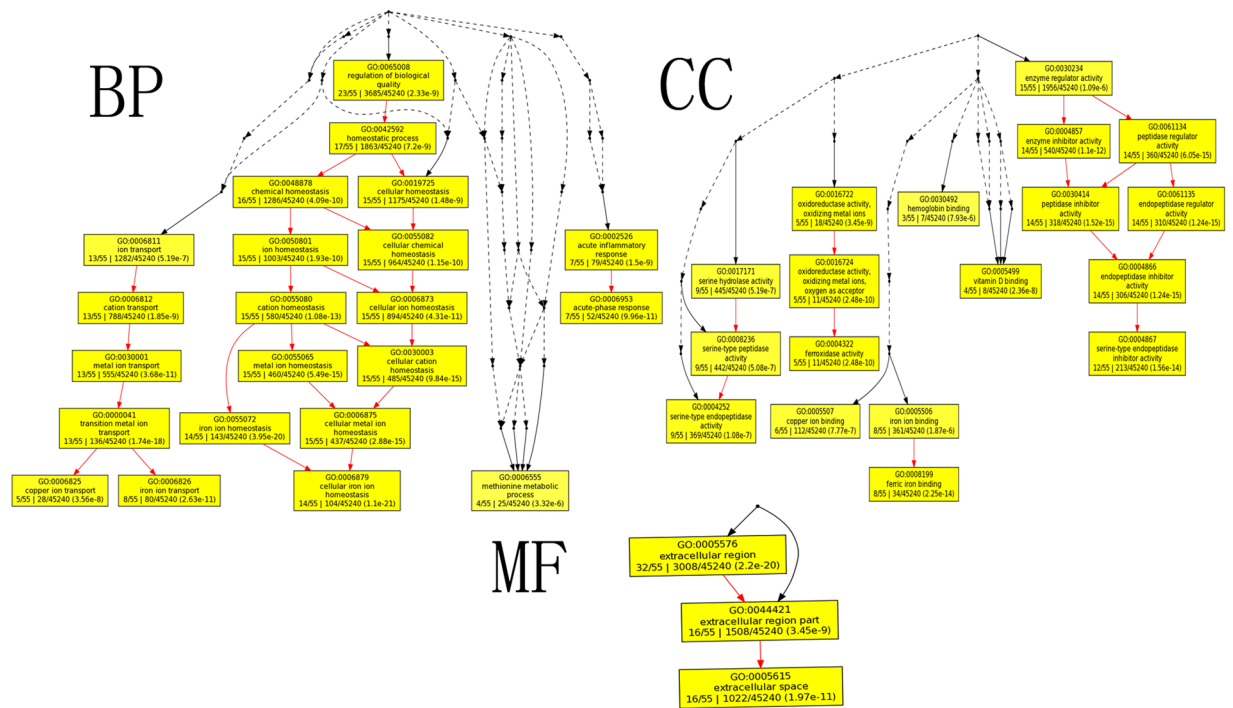


Figure 4. The graph displays enriched GOIDs of 16 seed-specific genes and their hierarchical relationships in “biological process (BP)”, “cellular component (CC)” or “molecular function (MF)” GO categories. Boxes represent GO terms, term definition, p-value and detail information. Significantly enriched GO terms are marked yellow. The degree of color saturation of each node is positively correlated with the significance of enrichment of the corresponding GO term. Non-significant GO terms within the hierarchical tree are shown as points. Branches of the GO hierarchical tree without significant enriched GO terms are not shown. Edges stand for connections between different GO terms. Red edges stand for relationship between two enriched GO terms, black solid edges stand for relationship between enriched and unenriched terms, black dashed edges stand for relationship between two unenriched GO terms.

Moreover, Chen R., *et al.* found that HPT, thyroxine carrier protein, and complement C3 significantly changed after partial cerebral ischemia/reperfusion via studying plasma proteomics in ischemic stroke models²⁵. In the protein-protein interaction analysis, HPT and thyroxine carrier protein were identified as predicted functional partners with a score of 0.751. Similarly, in the interactive network, both of these proteins were predicted to provide a basis for participating in the inflammatory and immunologic process in the occurrence of stroke.

As a vital inflammatory mediator, the complement system plays an important role in the occurrence and development of atherosclerosis and ischemic stroke²⁶. Many studies reveal that almost all of the complement inherent ingredients and the mRNA expression of various complement regulatory proteins are found in atherosclerosis lesions²⁷. Fibrinogen (Fg) plays an important role in the complement system. For example, the generation of thrombin is regarded as a critical step in the coagulation cascade, since thrombin mediates the functions that lead to the formation of blood clots by the cleavage of fibrinogen and factor XIII and the activation of platelets. Previous studies revealed that Fg is an important risk factor for coronary heart disease (CHD), stroke and peripheral angiopathy²⁸. For patients with atherosclerosis, the expression of Fg protein was significantly higher than in patients without atherosclerosis in ischemic cerebrovascular disease. In addition, the expression of Fg protein is related to the severity of the disease^{29,30}. This study revealed that Fg protein is differentially expressed in the plasma of the patients with the different genotypes.

In the KEGG analysis, four differentially expressed proteins, including SERPINF2, SERPINA1, F2, and C1S, participated in the complement response process. Individual SERPINF2 and SERPINA1 protein expression may increase with the MTHFR polymorphism, which may suppress plasminogen via the coagulation system, promoting endothelial cells to bind Fg and accelerate platelet aggregation, resulting in thrombosis. Hamzi K., *et al.* performed a meta-analysis, which revealed that prothrombin and MTHFR were risk factors in Indian patients with ischemic stroke³¹. C1S protein is important for the complement cascade and plays a role in the regulation of immune adhesion³².

Furthermore, we identified Vitamin D binding protein (DBP) in the PPIs analysis, which is a serum-rich protein that releases actin into the peripheral circulation after cellular damage. In individuals with the MTHFR polymorphism, decreased DBP lowers the affinity of vitamin D, resulting in weakening of the vascular smooth muscles. DBP alterations in MTHFR polymorphisms may have an impact on other functions of DBP³³. Plasma lipid metabolism disorder is usually associated with cerebrovascular disease. In this study, we found a quantitative reduction of Apolipoprotein A-4 in the mutant individuals. Apolipoprotein is a lipid binding protein that is involved in lipid transport that can either activate or suppress enzymes in the lipoprotein metabolic process or act

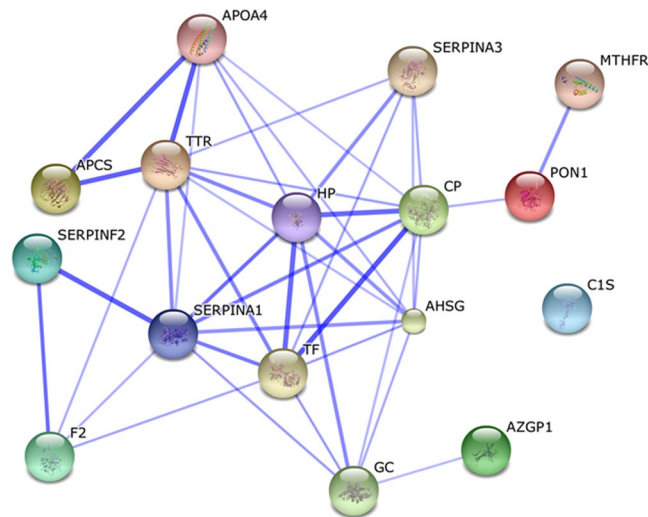


Figure 6. Protein–protein interaction network of MTHFR protein and the identified differential expression proteins (DEPs). The stronger associations are represented by the thicker lines.

MassARRAY RS1000 system using the standard protocol recommended by the manufacturer. Data were managed and analyzed using Sequenom Typer 4.0 Software. Genotype was analyzed using a Sequenom MassARRAY RS1000 system and the standard protocol recommended by the manufacturer. In total, 20 patients with the TT genotype in the C667T sites and 20 patients with the CC genotype in the C667T sites were used as experimental group and control group in this study (Table 3). The study was approved by the Ethics Committee of the Second Hospital of Lanzhou University. The Ethics Committee also approved the related screening, treatment, and data collection from these patients based on the experimental design and analysis of clinical outcome. The methods were carried out in accordance with the approved guidelines and were performed according to the Declaration of Helsinki. All of the subjects signed written informed consent forms for this study. This statement explicitly states that “informed” consent was obtained from subjects’ families.

Treatment and quantification of the samples. The plasma of 20 patients derived from experimental group and control group were mixed and prepared for protein extraction, respectively. After protein extraction, highly abundant plasma proteins were removed by using the Proteo Extract Albumin/IgG Removal Kit (Merck, Germany). The total plasma protein concentration was measured using a Bradford protein assay kit (TIANGEN, Beijing, China) according to the manufacturer’s protocol.

Two-dimensional electrophoresis of the plasma proteins. The hydration and loading buffers were combined with DTT, Bio-Lyte 4–6, 5–7, and the depleted plasma samples (70 µg for each experiment), and were then added into the focusing plates. Prefabricated IPG adhesive tapes were placed face down in the sample solution of the focusing plate. The adhesive tapes were placed in a PROTEAN IEF Cell electrophoresis system, and the isoelectric focusing procedure was set up (Table 4).

Two 10% acrylamide gels were prepared. The adhesive tapes were equilibrated and were then placed face down on a long glass board with a low melting agarose molding solution on the topside. Next, the low melting agarose molding solution was frozen and the gel was transferred into an electrophoresis tank. After finishing the electrophoresis, the gel was removed and marked by a corner cut.

Gel staining and image analysis. The gel was stained using a fast silver stain kit. Briefly, the gel was fixed, washed, and stained using the Fast Silver Stain Kit (P0017S, Beyotime, Shanghai, China) with silver staining radiation sensitivity solution (1X) and the silver solution (1X), which was adapted to the following mass spectrometry analysis. Finally, the silver staining color-developing solution was added. This step was stopped when the protein spots were visually apparent. The gel was rinsed and preserved in ultrapure water. Images of the 2D gels were captured using the Versadoc4000 images system. For experimental group and control group, three repeated experiments were performed to reduce the experimental error. Meanwhile, the protein spots were detected and optimized, and the background was reduced and additional editing was done using PDQuest8.0.1 image analysis software.

MALDI-TOF-MS analysis: identification of the differential protein points. Seven protein spots with significant differences were selected in the gel for an API 4800 cascade time-of-flight mass spectrometer MALDI-TOF/TOF (Applied Biosystems) analysis. The spots were isolated and enzymolysis by trypsin. Subsequently, peptides are extracted, vacuum dried, and ready for analysis by mass spectrometry. The analysis was done in positive ion mode and using automatic acquisition data. The PMF’s mass spectrum scanned range was 800–3500Da, and ten peaks with a maximum strength were chosen for the second phase of mass spectrometry. After the molecular weight of the peptide fractions was accurately measured, the first and second phase mass

Gender	Age	Native place	Weight	Blood Pressure(mmHg)	C677T genotype
Male	55	Lanzhou	56 kg	137/94	TT
Male	63	Lanzhou	65 kg	130/76	TT
Male	53	Lanzhou	71 kg	140/78	TT
Male	72	Lanzhou	68 kg	135/80	TT
Male	53	Lanzhou	67 kg	140/78	TT
Male	69	Lanzhou	70 kg	139/84	TT
Male	56	Lanzhou	74 kg	140/78	TT
Male	78	Lanzhou	62 kg	124/64	TT
Male	68	Lanzhou	70 kg	117/70	TT
Male	67	Lanzhou	59 kg	128/78	TT
Male	57	Lanzhou	70 kg	130/70	TT
Male	87	Lanzhou	70 kg	140/78	TT
Male	70	Lanzhou	65 kg	126/70	TT
Male	75	Lanzhou	73 kg	120/78	TT
Male	35	Lanzhou	70 kg	140/76	TT
Male	63	Lanzhou	60 kg	120/80	TT
Male	68	Lanzhou	69 kg	136/78	TT
Male	60	Lanzhou	72 kg	140/80	TT
Male	80	Lanzhou	78 kg	130/80	TT
Male	67	Lanzhou	70 kg	132/76	TT
Male	46	Lanzhou	74 kg	135/66	CC
Male	68	Lanzhou	68 kg	140/80	CC
Male	50	Lanzhou	65 kg	142/89	CC
Male	50	Lanzhou	68 kg	142/89	CC
Male	53	Lanzhou	62 kg	136/75	CC
Male	67	Lanzhou	55 kg	145/78	CC
Male	78	Lanzhou	70 kg	140/80	CC
Male	76	Lanzhou	70 kg	113/55	CC
Male	54	Lanzhou	82 kg	142/84	CC
Male	70	Lanzhou	64 kg	148/78	CC
Male	76	Lanzhou	67 kg	134/97	CC
Male	67	Lanzhou	66 kg	123/78	CC
Male	39	Lanzhou	55 kg	120/70	CC
Male	50	Lanzhou	63 kg	124/76	CC
Male	71	Lanzhou	71 kg	140/80	CC
Male	71	Lanzhou	66 kg	135/75	CC
Male	39	Lanzhou	70 kg	136/79	CC
Male	64	Lanzhou	70 kg	140/80	CC
Male	77	Lanzhou	53 kg	145/64	CC
Male	64	Lanzhou	55 kg	109/59	CC
Male	72	Lanzhou	65 kg	118/69	CC
Male	79	Lanzhou	63 kg	130/87	CC
Male	71	Lanzhou	70 kg	144/78	CC
Male	49	Lanzhou	68 kg	133/78	CC
Male	48	Lanzhou	60 kg	96/60	CC
Male	52	Lanzhou	67 kg	120/80	CC
Male	60	Lanzhou	78 kg	128/80	CC
Male	65	Lanzhou	68 kg	141/76	CC

Table 3. Basic information of the patients used in this study.

spectrometric data obtained by integration were analyzed and identified using GPS3.6 (Applied Biosystems) and Mascot2.1 (Matrix Science) software. A t-test was used for the comparison among the groups in combination with Boolean logical operators to process. A difference was considered statistically significant when $P < 0.05$ (one-way ANOVA).

Bioinformatics analysis. The standard human plasma 2D-spectrum was selected as a reference to judge the differential expression proteins (DEPs) with PDQuest8.01 software in the swiss-2DPAGE database (<http://world-2dpagexpasy.org/swiss-2dpag/>)³⁵. The protein 3D structure was predicted by Swiss-model and was viewed

Step	Voltage (V)	Patterns	Time	function
Hydration	50 V	Active hydration	12 hours	Hydration and loading
Step 1	250 V	Linear	30 min	Desalination
Step 2	1000 V	Fast	1 hour	Desalination
Step 3	10000 V	Linear	5 hours	Boosting
Step 4	10000 V	Fast	60,000 V/h	Focus
Step 5	500 V	Fast	At any time	Retention

Table 4. Setup of the PROTEAN IEF Cell Isoelectric Focusing procedure.

using PyMOL³⁶. Sixteen differential expression proteins (DEPs) were used for the GO enrichment analysis using GOEAST (<http://omicslab.genetics.ac.cn/GOEAST/index.php>). A hypergeometric distribution was employed to calculate the p-value of the GOID enrichment, and a $p < 1e-5$ cut-off value was applied³⁷. The graph size was reduced by condensing the non-significant nodes to points. The smaller the p-value is, the more significant the GO term is enriched in the dataset and the graph size was reduced by condensing the non-significant nodes to points. In addition, the DEPs were used for the preliminary signals and the metabolic pathway analysis by using the KEGG pathway mapper (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.ad.jp/kegg/pathway.html>) database^{15–17,38}. The sequences of all of the 16 DEPs were used for BLAST analysis with the National Center for Biotechnology Information (NCBI) clusters of eukaryotic Orthologous Groups (KOG) database to obtain the KOG numbers of those proteins. A data set containing all of the KOG numbers was then used for the protein–protein interactions (PPIs) analysis by using the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) database (version 9.1, <http://string-db.org>)³⁹.

References

- International Human Genome Sequencing Consortium *et al.* Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).
- Botto, L. D. & Yang, Q. 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. *Am J Epidemiol*. **151**, 862–877 (2000).
- Goyette, P. *et al.* Human methylenetetrahydrofolate reductase: isolation of cDNA mapping and mutation identification. *Nat Genet*. **7**, 551 (1994).
- Ueland, P. M., Hustad, S., Schneede, J., Refsum, H. & Vollset, S. E. Biological and clinical implications of the MTHFR C677T polymorphism. *Trends Pharmacol Sci*. **22**, 195–201 (2001).
- Boushey, C. J., Beresford, S. A., Omenn, G. S. & Motulsky, A. G. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. *Probable benefits of increasing folic acid intakes*. *JAMA*. **274**, 1049–1057 (1995).
- Graham, I. M. *et al.* Plasma homocysteine as a risk factor for vascular disease. The European Concerted Action Project. *JAMA*. **277**, 1775–1781 (1997).
- Varga, E. A., Sturm, A. C., Misita, C. P. & Moll, S. Cardiology patient pages. Homocysteine and MTHFR mutations: relation to thrombosis and coronary artery disease. *Circulation*. **111**, e289–293 (2005).
- Refsum, H., Ueland, P. M., Nygard, O. & Vollset, S. E. Homocysteine and cardiovascular disease. *Annu Rev Med*. **49**, 31–62 (1998).
- Adinolfi, L. E. *et al.* Hyperhomocysteinemia and the MTHFR C677T polymorphism promote steatosis and fibrosis in chronic hepatitis C patients. *Hepatology*. **41**, 995–1003 (2005).
- Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. **1**, 845–867 (2002).
- Pernemalm, M. & Lehtio, J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert Rev Proteomics*. **11**, 431–448 (2014).
- Gomez, Y. R. *et al.* Characterization of protein complexes using targeted proteomics. *Curr Top Med Chem*. **14**, 344–350 (2014).
- Choi BO *et al.* Homozygous C677T mutation in the MTHFR gene as an independent risk factor for multiple small-artery occlusions. (Translated from eng) *Thromb Res*. **111**, 39–44 (in eng) (2003).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. (Translated from eng) *Nat Protoc*. **4**, 44–57 (in eng) (2009).
- Kanehisa, F. M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. **45**, D353–D361 (2017).
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. **44**, D457–D462 (2016).
- Kanehisa, M. & Goro, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. **28**, 27–30 (2000).
- Pickell, L., Tran, P., Leclerc, D., Hiscott, J. & Rozen, R. Regulatory studies of murine methylenetetrahydrofolate reductase reveal two major promoters and NF-kappaB sensitivity. *Biochim Biophys Acta*. **1731**, 104–114 (2005).
- Berliner, J. A. *et al.* Atherosclerosis: basic mechanisms. Oxidation, inflammation, and genetics. *Circulation*. **91**, 2488–2496 (1995).
- Tousoulis, D., Davies, G., Stefanadis, C., Toutouzias, P. & Ambrose, J. A. Inflammatory and thrombotic mechanisms in coronary atherosclerosis. *Heart*. **89**, 993–997 (2003).
- Greisenegger, S. *et al.* Polymorphisms in inflammatory genes and the risk of ischemic stroke and transient ischemic attack: results of a multilocus genotyping assay. *Clin Chem*. **55**, 134–138 (2009).
- Kiga, C. *et al.* Proteomic identification of haptoglobin as a stroke plasma biomarker in spontaneously hypertensive stroke-prone rats. *Life Sci*. **83**, 625–631 (2008).
- Holme, I., Aastveit, A. H., Hammar, N., Jungner, I. & Walldius, G. Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure in 342,125 men and women in the Apolipoprotein MOrtality RISK study (AMORIS). *Ann Med*. **41**, 522–532 (2009).
- Brea, D. *et al.* Usefulness of haptoglobin and serum amyloid A proteins as biomarkers for atherothrombotic ischemic stroke diagnosis confirmation. *Atherosclerosis*. **205**, 561–567 (2009).
- Chen, R. *et al.* Proteomic analysis of rat plasma following transient focal cerebral ischemia. *Biomark Med*. **5**, 837–846 (2011).
- Carter, A. M. Complement activation: an emerging player in the pathogenesis of cardiovascular disease. *Scientifica (Cairo)*. **2012**, 402783 (2012).
- Bhakdi, S. Complement and atherogenesis: the unknown connection. *Ann Med*. **30**, 503–507 (1998).

28. Stec, J. J. *et al.* Association of fibrinogen with cardiovascular risk factors and cardiovascular disease in the Framingham Offspring Population. *Circulation*. **102**, 1634–1638 (2000).
29. Folsom, A. R. *et al.* Prospective study of markers of hemostatic function with risk of ischemic stroke. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Circulation*. **100**, 736–742 (1999).
30. Folsom, A. R., Wu, K. K., Rosamond, W. D., Sharrett, A. R. & Chambless, L. E. Prospective study of hemostatic factors and incidence of coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation*. **96**, 1102–1108 (1997).
31. Hamzi, K., Tazzite, A. & Nadifi, S. Large-scale meta-analysis of genetic studies in ischemic stroke: Five genes involving 152,797 individuals. *Indian J Hum Genet*. **17**, 212–217 (2011).
32. Mathern, D. R. & Heeger, P. S. Molecules Great and Small: The Complement System. *Clin J Am Soc Nephrol*. **10**, 1636–1650 (2015).
33. Otterbein, L. R., Cosio, C., Graceffa, P. & Dominguez, R. Crystal structures of the vitamin D-binding protein and its complex with actin: structural basis of the actin-scavenger system. *Proc Natl Acad Sci USA* **99**, 8003–8008 (2002).
34. Kwan, B. C., Kronenberg, F., Beddhu, S. & Cheung, A. K. Lipoprotein metabolism and lipid management in chronic kidney disease. *J Am Soc Nephrol*. **18**, 1246–1261 (2007).
35. Gasteiger, E. *et al.* The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. **31**, 3784–3788 (2003).
36. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. **42**, W252–258 (2014).
37. Zheng, Q. & Wang, X. J. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res*. **36**, W358–363 (2008).
38. Du, J. *et al.* KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst*. **10**, 2441–2447 (2014).
39. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. **39**, D561–568 (2011).

Acknowledgements

This study was supported by (1) the Gansu Provincial Natural Science Foundation of China (1506RJZA242, Study on Interaction of Multiple Locus of Susceptible Gene in Ischemic Stroke); (2) the Gansu Province 2015 Longyuan Young Talent Support Project Plan (2109901, Ischemic Stroke Early Molecular Markers Screening and Clinical Translational Research); (3) the Doctoral Research Fund Project of Second Hospital Affiliated to Lanzhou University (Differential Proteomic Study of Molecular Markers Associated with Metabolic Syndrome and Ischemic Stroke); and (4) Health Industry Planning Project of Gansu Provincial Health and Family Planning Commission (GSWSKY 2016-17, Study on the Isolation and Proteomics of PC12 Model of Ischemic Stroke).

Author Contributions

G.S. and M.X.W. designed the study and performed the study. Z.C.Z. wrote the paper. Q.Y., J.G., X.P.W., W.Y., L.W. and L.X.C. researched the data and contributed to the discussion. All of the authors approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-13542-6>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017