◇Full artical ◇

Chin J Clin Pharmacol Ther
Chinese Pharmacological Society
CN 34-1206/R, ISSN 1009-2501
E-mail: cept96@21cn.com
Oct; 12(10): 1144—1150

Metabolomics approach to the biochemical differentiation of Traditional Chinese Medicine syndrome types of hypertension

LU Yi-hong[1,3], HAO Hai-ping[1], WANG Guang-ji[1], CHEN Hu-xiao[2], ZHU Xuan-xuan[2], XIANG Bing-ren[1], HUANG Qing[1], A Ji-Y[1]

[1]*Key Laboratory of Drug Metabolism and Pharmacokinetics, China Pharmaceutical University, Nanjing 210009, Jiangsu, China*; [2]*Jinagsu Provincial Traditional Chinese Medicinal Hospital, Nanjing 210009, Jiangsu, China*; [3]*Jiangsu Institute for Drug Control, Nanjing 210008, Jiangsu, China*

ABSTRACT **AIM:** Traditional Chinese Medicine (TCM) has been practiced in China for thousands of years, providing a unique theoretical and practical approach to the treatment of diseases. In TCM theory, the notions of the "whole" and the use of "system" rather than isolation are important concepts, which well fit to systems biology theory. In the present study, we try to discover whether GC/MS-based metabolomics approaches contribute to differentiate the TCM syndrome types of hypertension. **METHODS:** The three phenotypes of constitution in patients with essential hypertension, the *hyperactivity of liver yang type*, *tan shi yong sheng type and yin xu yang kang* type, were classified by TCM approach. Serum metabolomic profiles for healthy persons and hypertension patients were acquired using GC/MS global analysis. Principal components analysis (PCA), partial least squares-discriminant analysis (PLS-DA), and Mahlanobis distance (MD) were applied to facilitate the metabolomics data differentiation and prediction. **RESULTS:** Using PCA and PLS-DA, it was capable of distinguishing normal blood pressure serum samples from those of the TCM syndrome types, while failed to discriminate the three TCM syndrome types of hypertension from each other. Further MD analysis contributed not only to a fine differentiation, but also to a clear exhibition of the progression, of the three syndrome types. **CONCLUSION:** This pilot study suggests that the metabolomics approach might be a powerful tool for exploring the scientific essence of the TCM theory.

KEY WORDS metabolomics; syndrome types; hypertension; GC/MS

With the development of modern life science, the emerging field of system biology involves the application of genomics, protenomics, metabolomics, bioinformatics *etc*. to the study of biological organisms at all levels, from the molecular, through the cellular, to the behavioral. Its aim is to understand biological processes as whole systems instead of as isolated parts[1]. Traditional Chinese Medicine (TCM) is an important part of life science, and also a mass complex system. In TCM theory, the notions of the 'whole' and the use of 'system' rather than isolate are important concepts, which well fit to systems biology theory. Therefore, it is reasonable to believe that TCM theory is intrinsically correlated with systems biology, and the TCM modernization is dependent on the development and application of systems biology.

Metabolomics, as one of the important components in systems biology, apply useful modeling tools for the classification and prediction of physiological and pathological states from metabolite profiles of biofluids such as urine[2,3] and plasma[4] and from profiles of intact tissues and tissue extracts[5]. Actually, metabolomics has now been widely applied for the research on disease diagnosis,

This work was financially supported by National natural science foundation of P. R. China (No. 30572228, 30630076), the National Eleventh Five Years Supporting Project (2006BAI08B04-05).

LU Yi-hong and HAO Hai-ping contributed equally to this work.

Dr. HAO Hai-ping, correspondence author, male.
Tel: 86-25-85391035 E-mail: hhp_770505@yahoo.com

gene function, drug discovery, toxicology and *etc*. Being aware of the intrinsic relationship between TCM theory and systems biology, some researchers began to discuss the prospective application of metabolomics approach to the scientific verification of TCM theory. However, no single report concerning metabolomics approach to TCM research has been found in the current literature. The present study was thus designed to determine whether the metabolomics strategy is useful and powerful for differentiating the syndrome types of TCM, using hypertension as a model.

According to the TCM theory, hypertension can be generally classified into four syndrome types, namely *hyperactivity of liver yang* type (A) (excessive internal heat of live), *tan shi yong sheng* type (B) (wet phlegm stasis), *yin xu yang kang* type(C) (Yin-deficiency and excessive Yang), and *yin yang liang xu* type(D) (deficiency of both *yin* and *yang*). These four syndrome types represent for different progression stages of hypertension as that, in the early stage it presents mostly with rising of liver fire and hyperacitivity of liver *yang*, in the middle stage with *yin* deficiency of liver and kidney accompanied with yang hyperactivity, in the late stage with deficiency of both *yin* and *yang* and mixed with phlegm and wet stasis[6]. Nevertheless, TCM syndrome types' theory from the generalization and summation of clinical experience and regularity for long term is helpful for the clinical diagnosis and therapeutic treatment of hypertension. However, its scientific connotation remains to be discovered and confirmed by modern scientific means.

In recent years, researchers made an endeavor trial on exploring the scientific connation of TCM syndrome types of hypertension disease, from different point of views including gene polymorphism, known biochemical indexes and *etc*[7-10]. However, these researches were nevertheless limited considering the rather isolated mythology. In the present study, we used a GC/MS based metabolomic approach for determining the biochemical profiles of different TCM syndrome types of hypertension, and for testing whether the metabolomics approach is poweful for the differentiation of TCM syndrome types.

1 MATERIALS AND METHODS

1.1 Chemicals *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide, chlorotrimethylsilane, methyl stearate, [$^2H_3$]-myristic acid (IS), and Methoxamine hydrochloride were all purchased from Sigma/Aldrich Co. The purity of all chemicals is all exceeding 98%.

1.2 Participants and blood sample collection Thirty two healthy volunteers and 33 patients with essential hypertension (EH) were included into this study. All of the clinical diagnosis and blood sample collections were conducted in the Cardiovascular Department of Jiangsu Provincial Traditional Chinese Medicinal Hospital (Nanjing, China), and all of the experiments were approved by the local ethical committee. Patients with essential hypertension were diagnosed according to the WHO/ISH guidelines for the Management of Hypertension[11], and concurrently undergone TCM syndrome types differentiation by the experienced TCM doctors. Patients, who received any drug treatment within the recent 3 days, were excluded from this study. Three milliliter blood samples, collected from each subject, were centrifuged to obtain serum which was immediately stored at −80 ℃ before analysis.

1.3 Sample preparation For GC/MS analysis, 100 μL serum from each sample was added into an Eppendorf tube which was preloaded with 400 μL methanol containing 2.5 μg [$^2H_3$]-myristic acid as internal standard, followed by vortexing for 10 min, then placed in ice bath for 1 h. Afterwards, the mixture solution was centrifuged at 22 500× *g* for 10 min and 200 μL of the supernatant was transferred to a GC vial and evaporated to dryness under N_2 at room temperature. The residues were then methoxymated in 30 μL 15 mg/mL methoxyamine hydrochloride in pyridine for 16 h, derivatized by adding 30 μL *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide (TMSFA) containing 1% Chlorotrimethylsilane for trimethylsilylation for 1 hour. The derivatized samples were reconstituted in 40 μL *n*-hexadecane with methyl stearate at 15 μg/mL for system check.

1.4 GC/MS analysis One microliter of derivatized sample was injected splitlessly into a Finigan gas chromatography (Thermofinigan, USA) coupled with mass spectrometry (TRACE DSQ). The GC is equipped with a 10 m× 0.25 mm ID, fused silica capillary column, which was chemically bonded with 0.25 μm DB1-MS stationary phase (J&W scientific, Folsom, CA, USA). The injector temperature was 270 ℃, the septum purge flow rate was 20 mL/min, and the purge was turned on after 60 s.

The gas flow rate through the column was 1 mL/min. The column initial temperature was kept at 70 ℃ for 2 min. Then temperature was increased from 70 to 320 ℃ at a rate of 20 ℃/min, and held for 2 min. Transfer line temperature was 270 ℃ and ion source temperature was 250 ℃. Ionization was achieved by a 70 eV electron beam at a current of 2.0 mA. Masses from 50 to 650 *m/z* were acquired in TIC mode when the acceleration voltage was turned on after a solvent delay of 170 s.

1.5 GC/MS data processing All data was processed by Xcalibur software (Thermofinigan, USA). Peaks of signal-to-noise (S/N) ratios below 10 were excluded. Retention time was adjusted with internal standard to minimize run-to-run errors. Retention index for each peak/compounds was calculated by comparing its retention time against those of alkane series (C8—C40, corresponding retention index values from 800 to 4000). To obtain accurate peak areas for the IS and specific peaks/compounds, one or two quant masses for each component were specified and the data were re-processed. The area of each peak was normalized against the internal standard, [$^2H_3$]-Myristic acid, before Multivariate statistical analysis. All Compounds were identified by comparing both the MS spectra and retention index with those available in libraries, i.e. NIST library 2.0 (January 31, 2001), Wiley library, and in-house mass spectra library database established by the center of development and research of drug metabolism and pharmacokinetics, China Pharmaceutical University.

1.6 Multivariate statistical analysis Multivariate statistical analysis(MVSA) was carried out using SIMCA-P 11 software (Umetrics, Umeå, Sweden). The data matrix was constructed with the observation/samples in columns and the peaks as variables in rows. It can be represented in a K-dimensional space (where K is equal to the number of variables), and then projected and reduced to a few principal components that describe the maximum variation of different groups or samples. Principal component analysis (PCA) was used to calculate a basic model and for overview of the data. Partial least squares projection to latent structures &discriminant analysis (PLS-DA) was used to calculate models differentiating groups or classes, herein, separating between healthy control and EH patients, and among three TCM syndrome types. Statistically different peaks were calculated by PLS-DA model with a confidence interval of 0.99 and significance level of 0.01. Cross-validation with seven cross-validation groups was used throughout to determine the number of components Mahlanobis distance (MD) method was applied to compute the model for classifying the three TCM syndrome types according to the shortest MD[12].

2 RESULTS AND DISCUSSION

2.1 GC/MS analysis of serum samples from EH patient and healthy control Based on the previously developed method[13], the GC/MS parameters were optimized for Finigan GC/MS system used in the present study. This system allowed the detection of a large number of peaks from the GC/MS chromatogram within 16 min analysis cycle. The typical GC/MS TIC chromatography of serum samples were shown in Fig 1 for EH patient. Sixty peaks of high intensity presenting in both EH and healthy control samples were identified and their peak areas were integrated for further MVSA.

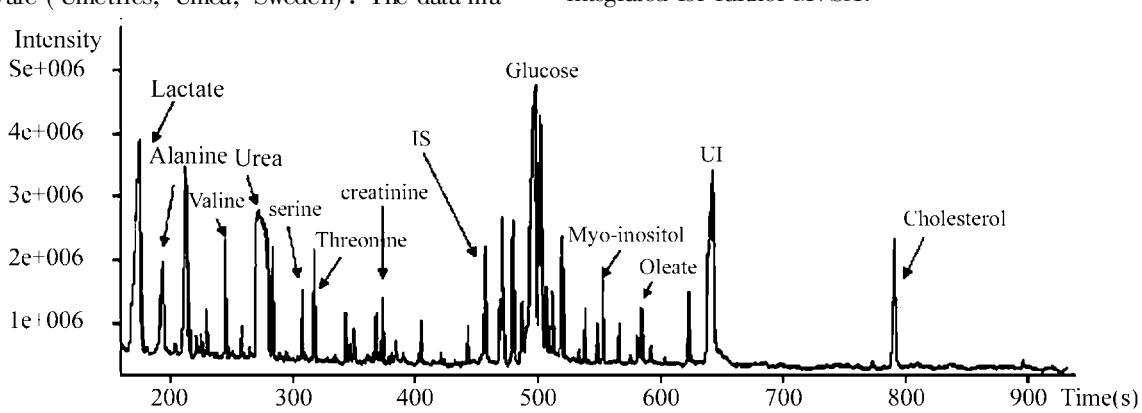

Fig 1 A typical GC/MS chromatogram (TIC) of serum samples from EH patient

2.2 Metabolomic Profiling of EH patients and the control The data matrix of peak areas obtained from

both EH and the control was subject to PCA. PCA is a routine statistical method applied in metabolomics to determine the relationship among the samples/observations. To facilitate the visualization of the data structure, PCA can reduce a great number of variables into a smaller number of uncorrelated variables, which are called principal components. The first principal component explains the greatest variability in the data, the second principal component is independent of (orthogonal to) the first component and second best explains the variability of the data and so on. There are two kinds of plots in PCA. One is "scores plots" for observations, or samples. The other is 'loading plots' for variables, or peaks in spectra. Each point on a metabolomics PCA scores plot represents all the variable data acquired in one spectrum, here 60 variables. Therefore, all of the sample points that cluster together have more similar spectra (and hence more similar biochemical makeup) than those that stand apart. However, plots should be scrutinized to include or discard especially for the plots that are far from the majority since outliers usually have negative influences on modeling. Three serious outliers of were found in PCA overviewing score plot for this dataset and therefore removed from further modeling. According to cross validation, a five-component PCA model was calculated. This model explained 68.1% (R^2) and predicted 46.9% (Q^2) of the data, with the first three components contributed 53.4% for explaining and 38.2% for prediction, respectively.

Due to diversity in gender, age and living habits, such as food, activity, samples collected from human being may have great variation on metabolome, which usually makes it difficult to calculate a very good PCA model to separate various groups and therefore find the intrinsic differences between groups. PLS-DA model was applied to the dataset to enhance this separation of EH and Healthy control. The objective with PLS-DA is to develop a model that is capable of separating the two classes of observations on the basis of their variables(X). The Y matrix encodes class membership by a set of 'dummy' variables, herein of one column with one and zero for each class. Then a PLS model is fitted between X variables and artificial Y. In this way, a discriminant plane is found in which the observations are well separated according to their class membership. A very good four-component model was computed (Fig 2), which can explain 97.8% and predict 94.8% of the data, with the first component explaining 86.1% and predicting 83.0%. Although this is a valid model to show the inner difference between EH and the control, the scores plots provide little mechanistic insight on a molecular basis, and indicate nothing about the pathological/biochemical significance of the clusters. Fortunately, the data can be investigated in more detail by examining the loadings to find out which variables are responsible for the score plots. By using this PLS-DA model, statistically significant variables were identified between the EH patients and healthy control by the first principal component, which is most reliable and covers the most significant variables. Identified compounds responsible for the model included phosphate, lactate, oxalate and some peaks not identified that are more abundant in EH, while amino isobutyric acid, cholesterol, glycerophosphate, uric acid and some amino acids, such as asparagine, serine and tryptophan that are more abundant in healthy control.

To investigate the prediction ability of the PLS-DA model, another 12 (6 samples for the patients and the control, respectively) independent observations/samples were imported and used as prediction dataset. They were calculated based on the available model for predicting which groups they belonged to in the score plot. It was shown that all of the samples can be predicted in the right region where their own groups clustered except only one control positioned somewhat between the two groups(Fig 2). The PLS-DA model was suggested to be a valid one to separate EH and the healthy control based on the GC/MS data and the multivariate statistical analysis method.

2.3 Metabolomic profiling of the three TCM syndromes PCA and PLS-DA models were calculated based on data from only EH patients of three syndrome types, type A, *hyperactivity of liver yang*; type B, *tan shi yong sheng*; type C, *yin xu yang kang*. It was found that PCA and PLS-DA failed to calculate good models presenting legible separation of the three TCM syndromes(Fig 3). Mahlanobis distance (MD) model was therefore applied to the dataset. MD is one of the commonly used distance measures of multivariate statistical analysis based on the measurement of distances between objects. The distances can be calculated in the original variable space and in the principal component (PC) space. The basic theory of MD analysis is that the shorter of the Mahalanobis distances

between the samples, the more similar of the metabolome constitution of the samples. In the original variable space, the MD takes into account the correlation in the data, since it is calculated using the inverse of the variance-covariance matrix of the dataset of interest. However, the computation of the variance-covariance matrix can cause problems. When the investigated data are measured over a large number of variables e.g., XC/MS, or NMR spectra, they can contain much redundant or correlated information. This so-called multicollinearity in the data leads to a singular or nearly singular variance-covariance matrix that cannot be inverted. The second limitation for the calculation of the variance-covariance matrix is that the number of objects in the data set has to be larger than the number of variables. For these reasons, it is clear that in many cases, feature reduction is needed. This can be done by, e.g., selecting a small number of meaningful variables. The MD can also be calculated using a smaller number of latent variables PCs obtained after PC analysis, PCA instead of the original variables. Due to the way PCA is carried out, PC1 always explains a larger amount of the total variance in the data than PC2, and than PC3. Therefore, in the 3 dimensions PC space, the scores on each PC1, PC2 and PC3 are weighted according to the amount of the variance explained by PC1, PC2 and PC3. We can, then, calculate the MD in the PC space, instead of in the original space. Equal MDs are represented by ellipses in the PC space. When the scores on each PC is divided by their weight, i.e., the amount of the total variance in the data explained, "equally weighted" or "normalized" scores is obtained. The MD computed in the normalized PC space leads to spheres around the center point, since the scores along each PC have now equal variance.

The MD model was calculated and shown in the dendrogram(Fig 4). In general, all of the samples can be divided into two principal groups by type A and B&C, where greatest differences were calculated, and then type B and C can be differentiated. The shortest distances were found amongst individuals of the same TCM syndrome, which suggested their similarity in metabolome. The only exception is that three individuals from A type were calculated and classified into type B. It can be imagined in a 3-D space that plots of type C gathered in a small spheric space, type B and 3 A individuals clustered in a middle one, while all other type A scattered in a big one. Unfortunately, statistical analysis of the variables did not suggest typical markers for the three TCM types. It was speculated that variables detected by GC/MS contribute to the model as a whole rather than several specific variables do.

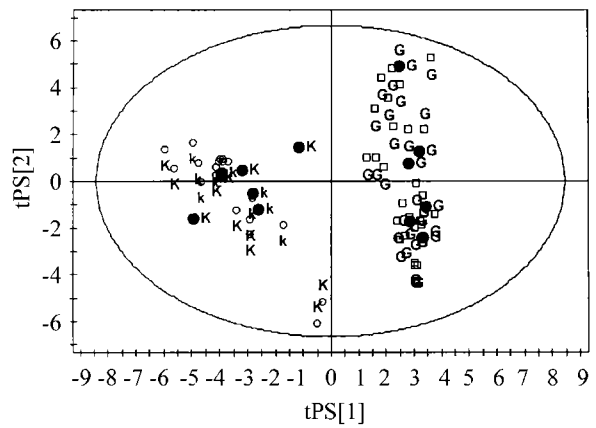

Fig 2 A very good PLS-DA model with four principal components was computed based on the data from EH patients (G□), and healthy control (K○) which can explain 97.8% and predict 94.8% of the data, with the first component explaining 86.1% and predicting 83%. The scores plot in fact showed complete separation of the two groups. Based on the calculated model, a groups of independent data from 12 samples (●, 6 for the patients and the control) was imported and predicted. All of the samples can be predicted in the right region where their own groups clustered except only one control positioned somewhat between the two groups. The PLS-DA model was suggested to be a valid one to separate EH and the healthy control based on the GC/MS data and the multivariate statistical analysis method.

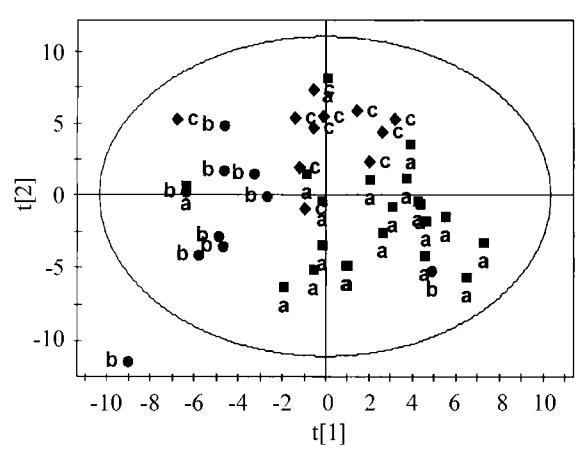

Fig 3 PLS-DA modeling of EH patients with three TCM syndrome types. It was shown that the three types of samples were poorly separated from each other in spite of a trend of separation being observed. This model gave no information of the progression relationship among the three syndrome types. a■, type A, b●, type B, c◆, type C.

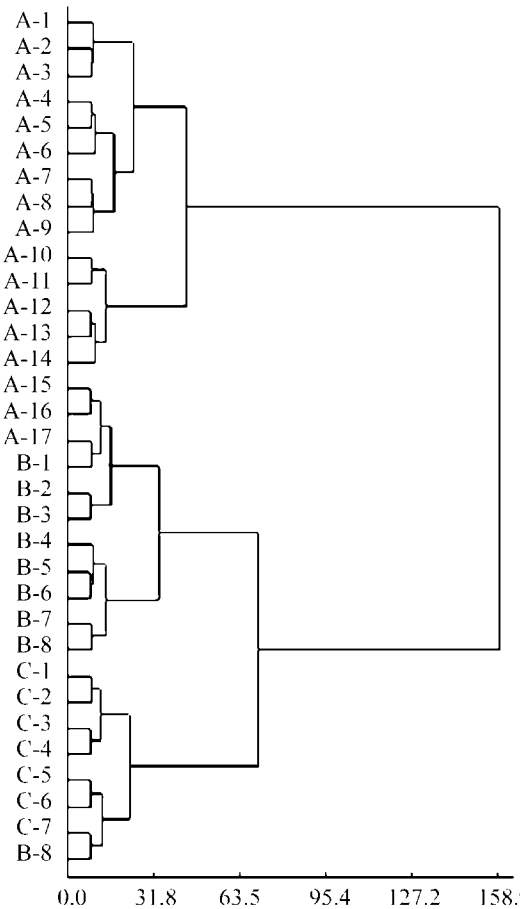

Fig 4 Hierarchical cluster analysis based on Mahlanobis distance (MD) of serum from A= type A(*n*= 17), B= type B (*n* = 8), C= type C (*n*= 8). In general, all of the samples can be divided into two principal groups by type A and B&C, where greatest differences were calculated, and then type B and C can be differentiated. The shortest distances were found amongst individuals of the same TCM syndrome, which suggested their similarity in metabolome. The only exception is that three individuals from A type (A15, 16, 17) were calculated and classified into type B.

3 CONCLUSIONS

The present study is a pioneer research of using metabolomics approach for differentiating different TCM syndrome types. Although TCM has its unique approach to the treatment of diseases, it is somewhat difficult to grasp its scientific essence. TCM syndrome differentiation still depends principally on the personal skill and experiences of making a diagnosis from the feelings, appearances, tongue fur and pulse graph of patients. For hypertension, syndrome differentiation depends on the degree of such symptoms as dizziness, headache, palpitation, insomnia, limb numbness, *etc*. In this research, chemometric analysis of GC/MS data is used to direct the identification of abnormalities in the metabolic profiles of biological matrices and thus to make the TCM syndrome types differentiation. It is found that GC/MS based metabolomics approach is successful for differentiating EH patients from healthy controls, and also for differentiating different TCM syndrome types. Potential biomarkers for differentiating EH patients from control groups are identified. This research strongly supported that the metabolomics approach would be a powerful tool for discovering the scientific connotation of TCM theory. Metabolomics approach will be prospectively powerful for bridging the TCM theory with the modern medical science.

REFERENCES

1 Kitano H. Systems Biology: a brief overview[J]. Science, 2002; 295: 1662—1664.

2 Nicholson JK, Connelly J, Lindon JC, *et al*. Metabonomics: a platform for studying drug toxicity and gene function[J]. Nat Rev Drug Discov, 2002; 1: 153—161.

3 Bollard ME, Holmes E, Lindon JC, *et al*. Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution (1) H NMR spectroscopy of urine and pattern recognition[J]. Anal Biochem, 2001; 295: 194—202.

4 Brindle JT, Antti H, Holmes E, *et al*. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabolomics[J]. Nat Med, 2002; 8: 1439—1444.

5 Waters NJ, Holmes E, Williams A, *et al*. NMR and pattern recognition studies on the time-related metabolic eVects of alpha-naphthylisothiocyanate on liver, urine, and plasma in the rat: an integrative metabonomic approach[J]. Chem Res Toxicol, 2001; 14: 1401—1412.

6 Zhang TS, Han L, Wang L. Relation of syndrome types to hypertension grades and risk stratification in patients with primary hypertension[J]. Chin J Clin Rehab, 2005; 21: 232—233.

7 Zou XG, Li YJ, Chen GT. Relationship Between TCM Syndromes and TCD Examinations in Senile Hypertension[J]. J Fujian Coll Trad Chin Med, 2001; 1: 8—11.

8 Pan MZ, Li L, Fan H. Relationship between quality of life and TCM syndrome types in patients with diabetes mellitus[J]. Chin J Integ Trad West Med, 2006; 26: 702—705.

9 Qian YS, Zhang Y, Zhang WZ. Relationship between the polymorphism of alpha-adducin gene and the two phenotypes of constitutions in patients with essential hypertension classified by TCM[J]. Chin J Integ Trad West Med, 2006; 26: 698—701.

10 Lu, QS, Lei Y, Chen KL. Relationship of the A1166C polymorphism of ATI R gene with TCM syndrome and efficacy of

Chinese hypotensor in patients with essential hypertension[J]. Chin J Integ Trad West Med, 2005; 25: 682—628.
11 Word Health Organization-International Society of Hypertension Guidelines for the Management of hypertension Guideline Subcommittee[J]. J Hypertens, 1999; 17: 158—183.
12 Maesschalck RD, Rimbaud DJ, Massart DL. The Mahalanobis distance[J]. Chemometr Intell Lab Syst, 2000; 50: 1—18.
13 A JY, Trygg J, Gullberg J, *et al*. Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome[J]. Anal Chem, 2005; 77: 8086—8094.

高血压病中医分型的代谢组学研究

摘要　目的: 将传统中医辨证方法同现代系统生物学理论相结合, 探讨原发性高血压辨证分型与基于GC/MS的血清代谢组学的关系。方法: 原发性高血压辨证分为肝火亢盛、痰湿雍盛及阴虚阳亢三型。应用GC/MS测定健康人及原发性高血压病人血清内源性代谢物, 并用主成分分析(PCA)、偏最小乘方分析(PLS-DA)和马氏距离(MD)分析他们的代谢谱。结果: PCA和PLS-DA分析的结果表明: 健康人与高血压病人血清代谢谱有明显差异, 能够被区分开, 但PCA和PLS-DA不能将中医高血压的三型完全分开。利用MD不仅可以清晰地区分上述三种类型的高血压, 同时还显示高血压的发展过程。结论: 基于GC/MS和模式识别的代谢组学在揭示传统中医理论本质上有着广泛的应用前景。

关键词　代谢组学; 中医证型; 高血压; GC/MS