Talanta 135 (2015) 108–114

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Exploring metabolic syndrome serum free fatty acid profiles based on GC–SIM–MS combined with random forests and canonical correlation analysis

Ling Dai [a], Carlos M. Vicente Gonçalves [a,b], Zhang Lin [a], Jianhua Huang [a,c], Hongmei Lu [a,*], Lunzhao Yi [a,d], Yizeng Liang [a], Dongsheng Wang [e], Dong An [a]

[a] Research Center of Modernization of Chinese Medicines, College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China
[b] Department of Chemistry, Faculty of Mathematics and Natural Sciences, University of Bergen, Bergen 5020, Norway
[c] Department of Pharmacy, Hunan University of Chinese Medicine, Changsha 410208, China
[d] Yunnan Food Safety Research Institute, Kunming University of Science and Technology, Kunming 650500, China
[e] Xiangya Hospital, Central South University, Changsha 410008, China

ARTICLE INFO

Article history:
Received 30 October 2014
Received in revised form
19 December 2014
Accepted 25 December 2014
Available online 3 January 2015

Keywords:
Metabolic syndrome
Free fatty acid
Metabolic profiling
GC–SIM–MS
Random forests
Canonical correlation analysis

ABSTRACT

Metabolic syndrome (MetS) is a cluster of metabolic abnormalities associated with an increased risk of developing cardiovascular diseases or type II diabetes. Till now, the etiology of MetS is complex and still unknown. Metabolic profiling is a powerful tool for exploring metabolic perturbations and potential biomarkers, thus may shed light on the pathophysiological mechanism of diseases. In this study, fatty acid profiling was employed to exploit the metabolic disturbances and discover potential biomarkers of MetS. Fatty acid profiles of serum samples from metabolic syndrome patients and healthy controls were first analyzed by gas chromatography–selected ion monitoring–mass spectrometry (GC–SIM–MS), a robust method for quantitation of fatty acids. Then, the supervised multivariate statistical method of random forests (RF) was used to establish a classification and prediction model for MetS, which could assist the diagnosis of MetS. Furthermore, canonical correlation analysis (CCA) was employed to investigate the relationships between free fatty acids (FFAs) and clinical parameters. As a result, several FFAs, including C16:1n-9c, C20:1n-9c and C22:4n-6c, were identified as potential biomarkers of MetS. The results also indicated that high density lipoprotein-cholesterol (HDL-C), triglycerides (TG) and fasting blood glucose (FBG) were the most important parameters which were closely correlated with FFAs disturbances of MetS, thus they should be paid more attention in clinical practice for monitoring FFAs disturbances of MetS than waist circumference (WC) and systolic blood pressure/diastolic blood pressure (SBP/DBP). The results have demonstrated that metabolic profiling by GC–SIM–MS combined with RF and CCA may be a useful tool for discovering the perturbations of serum FFAs and possible biomarkers for MetS.

1. Introduction

Metabolic syndrome (MetS), which is a constellation of metabolic abnormalities including abdominal obesity, hypertension, dyslipidemia and hyperglycemia has become one of the major public-health challenges worldwide [1,2]. The prevalence of MetS was reported to be approximately 34.6% in the United States [3], 17.8–34.0% in Europe [4,5] and 12.8–41.1% in Asia [4]. Several studies showed that the presence of MetS carries increased risk for cardiovascular disease [6,7] and type 2 diabetes [8]. The people with this syndrome possess nearly twice the risk for cardiovascular disease and 5-fold the risk for type 2 diabetes. Furthermore, MetS is associated with other diseases, including fatty liver [9] and cancer [10]. Thus, earlier identification and treatment of individuals at risk is particularly crucial.

The current method used to diagnose MetS is based on the clinical parameters in the definition, including waist circumference (WC), high density lipoprotein-cholesterol (HDL-C), triglycerides (TG), fasting blood glucose (FBG) and systolic blood pressure/diastolic blood pressure (SBP/DBP). However, these clinical parameters are not independent but rather interdependent of each other, rendering the disease rather difficult to diagnose and control. Furthermore, they might provide little additional insight regarding pathophysiologic mechanisms of MetS. Metabolomics or metabolite profiling is an emerging technology which makes it more feasible to acquire high-throughput profiles of a whole organism's metabolic status [11,12]. The strategy of metabolomics has been successfully

* Corresponding author. Tel.: +86 731 88836376; fax: +86 731 88879616.
E-mail address: Hongmeilu@csu.edu.cn (H. Lu).

http://dx.doi.org/10.1016/j.talanta.2014.12.039
0039-9140/

applied to many fields such as disease diagnosis [13], biomarker discovery [14], drug efficacy and toxicity evaluation [15], and proved to be a powerful tool for understanding the mechanism of diseases.

MetS is a complex disorder including multi-metabolic risk factors. The mechanism underlying it is complex and still unknown. More evidences demonstrated that MetS is closely associated with metabolism disorder of lipid, especially those involving free fatty acids (FFAs) [16,17]. FFAs, an important category of lipids, are active molecules, which not only provide an important energy source as nutrients, but also act as signaling molecules in various cellular processes, including insulin secretion [18]. It was found that increased blood FFA levels played an important role in the development of insulin resistance and various disturbances related to MetS [19,20]. In addition, FFAs could also contribute to oxidative stress, inflammation and endothelial dysfunction. So FFA profiling might be helpful to elucidate the pathophysiological process of MetS. To fully use the fatty acids in disease diagnosis, the entire analysis of fatty acids might be important. Traditionally, due to high separation efficiency, sensitive detection and good reproducibility, gas chromatography–mass spectrometry (GC–MS) was used as a robust metabolic tool for identification and quantification of fatty acids [21,22]. However, the analysis of fatty acid methyl esters (FAMEs) using GC–MS remains a challenge due to the coelution of many positional and geometrical isomers of fatty acids [23,24]. In order to obtain accurate and sensitive quantitation of all fatty acids, a method based on GC–SIM–MS [25] was proposed. In selected ion monitoring (SIM) mode, the most intense fragments (saturated: *m*/*z* 74, mono-unsaturated: *m*/*z* 55, di-unsaturated: *m*/*z* 67, poly-unsaturated: *m*/*z* 79) were selected as characteristic ions for quantification of fatty acids.

A cornerstone of metabolomics study is the acquisition and processing of high-quality data. The multivariate data analysis techniques and chemometrics are important to metabolomics for obtaining reliable results. Recently, many multivariate statistical analysis methods were employed to explore differences between groups of samples and discover potential biomarkers. Among these chemometric methods, random forests [26] (RF) is a new and powerful statistical classifier for exploiting differences between groups of samples, which not only possesses a high-prediction accuracy, but also can calculate the proximity between samples and visualize the data with multidimensional scaling (MDS). Meanwhile, canonical correlation analysis [27] (CCA) is a popular feature extraction method in multivariate statistical analysis, which provides an efficient way of measuring the linear relationship between two multidimensional variables. For this, CCA was utilized to discover the potential biomarkers associated with clinical parameters.

In the present study, metabolic profiling by GC–SIM–MS combined with RF and CCA was developed to explore the perturbations of serum FFAs and possible biomarkers of MetS. Serum samples from healthy controls and MetS patients were first profiled by GC–SIM–MS. In order to monitor the performance of the method and ensure the reliability of data, quality control (QC) samples were inserted throughout the run. After data preprocessing, RF was used to map the distribution of all human serum samples and classify healthy controls and MetS patients. Finally, CCA was employed to explore the relationships between FFAs and clinical parameters, so as to screen potential biomarkers that are related to clinical parameters and try to find some other useful information for dealing with metabolic abnormalities of MetS.

2. Materials and methods

2.1. Chemicals and reagents

Supelco 37 component FAME mix (No. 47885-U) used for the determination of fatty acids and cis-10-nonadecenoic acid (C19:1n-9c, $>99\%$ purity) used as internal standard were purchased from Sigma-Aldrich (St. Louis, MO, USA). The solution 5% H_2SO_4/CH_3OH was freshly prepared by diluting H_2SO_4 (98.0% purity, Zhuzhou Chemical Industry Institute, Zhuzhou, China) with chromatographic grade methanol (Tedia Company, Inc., Fairfield, USA), and 0.4 mol/L KOH/CH_3OH was freshly prepared by dissolving analytical grade KOH in methanol. Analytical grade n-hexane and anhydrous sodium sulfate were commercially obtained from Tianjin Yongda Chemical Reagent Co., Ltd. (China) and Tianjin Kermel Chemical Reagent Co., Ltd. (China), respectively.

2.2. Study population and sample collection

In our study, since the samples were recruited from Chinese people at the Xiangya Hospital of Central South University in Hunan Province, China, presence of MetS was examined according to the guidelines on prevention and treatment of dyslipidemia in Chinese adults [28] (2007). The definition for MetS in the guidelines was made on the basis of the Chinese Diabetes Society (CDS) definition [29] (2004) and the International Diabetes Federation (IDF) definition [30] (2005). It is the newest one in China. In the definition, subjects who satisfied at least three of the following risk factors were identified as MetS patients: central obesity (waist circumference >90 cm in men, >85 cm in women), increased blood pressure (SBP/DBP $\geq 130/85$ mmHg or previously diagnosed hypertension), increased fasting blood glucose (FBG ≥ 6.1 mmol/L or previously diagnosed type 2 diabetes), increased serum triglycerides (TG ≥ 1.70 mmol/L) and decreased HDL-cholesterol (HDL-C <1.04 mmol/L). Then 100 individuals collected from February 2012 to March 2013 were divided into two groups: one group containing 42 cases that had 0 risk factor was defined as healthy controls (HC), and the other group containing 58 cases that had at least three risk factors was defined as MetS patients. All clinical experiments were approved by Xiangya Institutional Human Subjects Committee and written informed consent was obtained from each participant.

2.3. Sample preparation

Serum samples were collected in the morning before breakfast and stored at -80 °C until analysis. Sample preparation procedures have been described previously in Ref. [31]. However, what we considered was only the FFAs fraction in serum, and the esterified fatty acids (EFAs) fraction in serum was left out of consideration. Moreover, internal standard (I.S.) for FFAs was replaced with C19:1n-9c since in our experiments C17:0 is detectable. In our experiments, aliquots (50 μL) of serum were spiked with internal standard (10 μL, 0.1 mg/mL C19:1n-9c/n-hexane) and 0.5 mL of 0.4 mol/L KOH/CH_3OH was added, vortexed for 30 s and placed at room temperature for 10 min. Then, 0.5 mL of n-hexane was added twice and vortexed for 30 s each time. Once the n-hexane phase containing the EFA methyl esters was removed, the residual serum phase was put into contact with anhydrous sodium sulfate to remove traces of water. Afterwards, 0.5 mL of 5% H_2SO_4/CH_3OH was added to the residuary phase of serum and the mixture was kept in a water bath at 70 °C for 30 min. Then, 1 mL of n-hexane was added twice and vortexed for 30 s each time. Prior to analysis by GC–MS, the n-hexane phases containing the methyl esters of FFAs were evaporated to dryness and subsequently reconstituted with 100 μL of n-hexane.

Quality control (QC) samples [32,33], which were employed to monitor the performance of the method and increase the credibility of downstream data analysis, were prepared by pooling 10 μL aliquots from each serum sample and vortex mixing (1 min). Sample preparation for QC samples was performed as described above.

2.4. GC–MS analysis

GC–MS analyses were performed on an Agilent 7890A gas chromatograph coupled to an Agilent 5975C quadrupole mass spectrometer. In the gas chromatographic system, a DB-23 capillary column (30 m × 0.25 mm i.d., film thickness 0.25 μm) was used. The optimized temperature program was as follows: initial temperature 100 °C, ramped to 160 °C at the rate of 20 °C/min, and followed by a second gradient of 6 °C/min to 180 °C, held at 180 °C for 3 min, finally, increased at 20 °C/min to 230 °C, held at this temperature for 5 min. The total program time was 16.833 min and the solvent delay time was 3.50 min. Helium (99.999% purity) was used as carrier gas with a flow rate of 1.0 mL/min. The temperature of the injector was 250 °C, and a sample of 2 μL was injected at the split ratio of 1:5. The mass spectrometric conditions were as follows: the temperatures of MS transfer line, quadrupole and ion source were set at 250 °C, 150 °C and 230 °C, respectively; electron impact (EI) energy was operated at 70 eV; the SIM mode was *m*/*z* 55, 67, 74, and 79; the full-scan mode was from *m*/*z* 35 to 450.

The injection order in every batch containing 12 study samples was as follows. In order to equilibrate the analytical system, 3 or 5 injections of QC samples were applied at the beginning of the batch. When the analytical system was balanced, one injection of QC sample was acquired in full scan mode. Then every fourth injection was a pooled QC sample and the final two injections of every batch were also QC samples. Except for the injection of QC-Scan, all the other injections were acquired in SIM mode.

2.5. Identification of serum free fatty acids of MetS patients and healthy controls

Since QC samples were pooled samples from MetS patients and healthy controls, they had the same composition as MetS patients and healthy controls. So QC-Scan could be used for qualitative analysis of fatty acids. The identification of FAMEs was conducted by the combination of retention time and mass spectral characteristics. The 37 component FAMEs mixture did not include all FAMEs esterified from serum free fatty acids. So, a method [34] proposed by our laboratory was adopted. In this method, the mass spectral characteristics of FAMEs were employed to automatically recognize all expected straight saturated FAMEs, whose retention times were subsequently used to calculate the equivalent chain length (ECL) values for the unsaturated FAMEs in samples. Finally, the ECL values and characteristic ions were applied to the identification of unsaturated FAMEs by comparing with those in the library.

2.6. Data processing and multivariate statistical analyses

The resulting raw data files (.D format) were converted to the NetCDF format and imported into the original software "Traditional Chinese Medicine Fingerprint Database System (TCMSys V1.0.2)", written by our group. Data preprocessing was performed here, including smoothing, baseline correction, peak shift alignment, normalization and peak area integration. As for normalization in this paper, the peak areas of corresponding metabolites were normalized to that of the internal standard on the same chromatogram to obtain the relative intensities of metabolites. Quality assurance (QA) was subsequently performed and only metabolic features with a relative standard deviation (RSD) for measured peak areas of $< 20\%$ were retained for data analysis [35]. All other metabolic features were removed from the dataset and ignored in subsequent data analysis. Then, the data matrix composed of 42 healthy controls, 58 MetS patients and 49 QC samples was imported into our custom scripts in MATLAB 7.10 (The MathWorks, Inc., USA) for subsequent multivariate statistical analyses.

2.6.1. Principal component analysis

The autoscaled data of FFAs were first analyzed by the unsupervised PCA to visualize the general trends of all the samples, including 42 healthy controls, 58 MetS patients and 49 QC samples. PCA combines the original variables linearly to produce a few new variables (called principal components, PCs) containing most of the information. The score plot of PC1, PC2 and PC3 was applied to show separation or clusters among the three groups.

2.6.2. Random forests

In this study, random forests, as a new and powerful statistical classifier, were introduced to explore the disturbances of free fatty acid profiles, so as to establish a classification and prediction model for MetS. RF is an ensemble classification algorithm. As the name suggests, RF combines many classification trees to produce more accurate classifications. The prediction accuracy for RF is estimated internally using an out-of-bag (OOB) estimation. Thus there is no need for cross-validation. Apart from the OOB estimation, RF also provides a proximity measure. The proximity between two samples is calculated as the number of times that the two samples end up in the same terminal node of a tree, divided by the number of trees in the forest. The resulting proximity matrix can be used to construct MDS plots, whose aim is to visualize the similarity or dissimilarity (calculated as 1- proximity) between samples.

More detailed theory on RF can be seen in Breiman's paper [26]. The RF algorithm for classification can be briefly described as follows:

1) Draw n_{tree} bootstrap samples from the original dataset. (n_{tree} is the number of trees in the forest. In this study, $n_{tree}=600$).
2) For each of the bootstrap samples, grow an un-pruned classification tree, and modify with the following procedure: at each node, randomly select m_{try} variables and choose the best split from among those variables. (For classification, the default value of m_{try} is equal to sqrt (m_{all}), where m_{all} is the total number of the variables in the original dataset. In this study, $m_{all}=28$, thus, $m_{try}=5$).
3) Predict new data by aggregating the predictions of the n_{tree} trees.

2.6.3. Canonical correlation analysis

Canonical correlation analysis [27,36] (CCA), which is a classic method of correlating linear relationships between two sets of multidimensional variables, was employed to explore the relationships between free fatty acids and clinical parameters, so as to discover biomarkers that are associated with clinical parameters and try to find some other useful information for dealing with metabolic abnormalities of metabolic syndrome. Considering **x** and **y** as the matrices of free fatty acids and clinical parameters, the corresponding rows and columns of **x** and **y** represent observations and variables, respectively. These two matrices were first preprocessed by centering, which can make the mean to be zero.

The total covariance matrix

$$\mathbf{C}=\begin{bmatrix}\mathbf{C_{xx}} & \mathbf{C_{xy}}\\ \mathbf{C_{yx}} & \mathbf{C_{yy}}\end{bmatrix}=\mathbf{E}\left[\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}^{\mathbf{T}}\right] \tag{1}$$

is a block matrix where $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ are the within-sets covariance matrices of **x** and **y**, respectively and $\mathbf{C_{xy}}=\mathbf{C_{yx}^{T}}$ is the between-sets covariance matrix. The canonical correlations between **x** and **y** can be found by solving the eigenvalue equations

$$\begin{cases}\mathbf{C_{xx}^{-1}C_{xy}C_{yy}^{-1}C_{yx}}\hat{\mathbf{a}}_{\mathbf{x}}=\rho^2\hat{\mathbf{a}}_{\mathbf{x}}\\ \mathbf{C_{yy}^{-1}C_{yx}C_{xx}^{-1}C_{xy}}\hat{\mathbf{b}}_{\mathbf{y}}=\rho^2\hat{\mathbf{b}}_{\mathbf{y}}\end{cases} \tag{2}$$

where the eigenvalues ρ^2 are the squared canonical correlations and the eigenvectors $\hat{\mathbf{a}}_x$ and $\hat{\mathbf{b}}_y$ are the normalized canonical correlation basis vectors. The number of non-zero solutions to these equations are limited to the smallest dimensionality of **x** and **y**. The so-called canonical variables **U** and **V** can be represented as **U**=**Xa** and **V**=**Yb**.

3. Results and discussion

3.1. Clinical parameters of MetS patients and healthy controls

The clinical parameters of MetS patients and healthy controls were measured by Xiangya Hospital according to routine physical examination methods [37]. These clinical parameters are shown in Table 1 and data was presented as mean ± SD. The significance of these clinical parameters was analyzed using *T*-test and adjusted by Bonferroni–Holm correction for multiple comparisons with the significance level set at $P<0.05$. The results reflected that several parameters such as WC, SBP/DBP, TG and FBG were significantly higher in MetS patients than that in healthy controls, in contrast, the level of HDL-C was significantly lower in MetS patients ($P<0.05$). Although these parameters may reflect the metabolic state of MetS to some extent, there are various other underlying metabolic disturbances, especially lipid metabolic disorders related to MetS. In order to elucidate the perturbations of serum fatty acids and potential biomarkers, furthermore, shed light on the pathophysiological progress of MetS, metabolic profiling was therefore performed in this study.

3.2. GC–SIM–MS profiles of serum samples from MetS patients and health controls

In order to obtain accurate quantitation of more fatty acids, GC–SIM–MS was employed to profile serum samples from MetS patients, healthy controls and QC samples. The GC–SIM–MS comparison profiles of serum free fatty acids from MetS patients and health controls (Fig. 1) and the qualitative and quantitative results of free fatty acids (Table 2) are obtained at the optimized conditions. As labeled in the solid profile, there are 28 common peaks in the TIC chromatograms of all the samples, and among the 28 peaks, 22 fatty acids were unambiguously identified and validated by the reference standards. The other fatty acids were identified based on retention times/indices and mass spectral characteristics according to the method proposed by our laboratory. Components in the solid profile could also be found in the dotted one according to their retention times and mass spectra. The comparison of fatty acid profiles of MetS patients (Fig. 1 in solid line) and health controls (Fig. 1 in dotted line) illustrated that serum FFAs of different groups of people were the same, while the concentrations were different. The concentrations of serum FFAs in MetS patients were higher than the concentrations of serum FFAs in health controls (Fig. 1 and Table 2). Moreover, most of FFAs' concentrations differed significantly ($P<0.05$) between health controls and MetS patients (Table 2). These results reflect the increasing trends of fatty acids from health controls to MetS patients. However, whether these fatty acids could completely separate MetS patients from healthy controls and which fatty acids are the most relevant to the clinical parameters are still unknown. Thus, it is necessary to further investigate the fatty acid profiles with the help of multivariate statistical analyses.

Table 1
Clinical parameters of MetS patients and healthy controls.

Clinical parameters	HC (n=42)	MetS (n=58)	P values
WC (cm)	76.76 ± 6.87	91.19 ± 4.93	1.1315e-20
SBP (mmHg)	111.33 ± 9.90	133.00 ± 14.25	9.9910e-13
DBP (mmHg)	70.95 ± 6.89	85.03 ± 11.00	2.1114e-10
TG (mmol/L)	0.89 ± 0.41	3.60 ± 2.03	9.9910e-13
HDL-C (mmol/L)	1.39 ± 0.28	1.06 ± 0.23	4.1058e-09
FBG (mmol/L)	4.89 ± 0.42	6.91 ± 2.57	2.1097e-06

A P value of <0.05 was considered statistically significant.

3.3. Results of PCA and random forests

First of all, the autoscaled free fatty acid data matrix (149 samples × 28 variables) is analyzed using PCA to explore separation or clusters of all the samples, including 42 healthy controls, 58 MetS patients and 49 QC samples. The score plot for the first three components, say PC1, PC2 and PC3, is displayed in Fig. 2. As shown in the figure, there was a trend of intergroup separation from HC, QC to MetS in the PC2 direction. However, it was still hard to completely classify the samples of these three groups, though the first three PCs have explained up to 78.25% of the total variance of the data. The results showed that the unsupervised method of PCA could not separate the three groups satisfactorily.

Therefore, the powerful supervised multivariate statistical method of RF was subsequently employed to explore the disturbances of free fatty acids of MetS, so as to establish a classification and prediction model for MetS. The RF model consists of a series of classification trees which are constructed by the bootstrap samples. After each tree

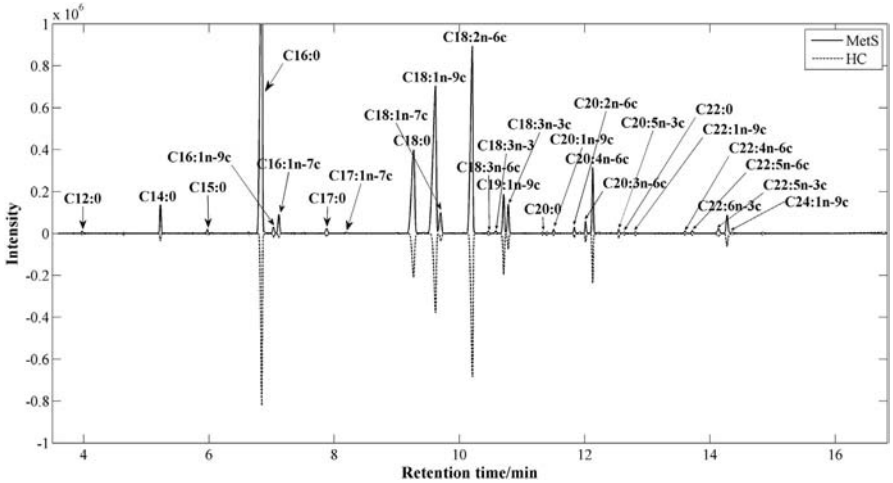

Fig. 1. Comparition of GC–SIM–MS profiles of free fatty acid methyl esters from serum samples of MetS patient (in solid line) and healthy control (in dotted line). The qualitative results were marked in the solid profile.

Table 2
Qualitative and quantitative analysis results of free fatty acids profiling of healthy controls and MetS patients.

Rt. Time (min)	FFAs	HC ($n=42$) (μg/mL)	MetS ($n=58$) (μg/mL)	RSD of QC	Bonferroni–Holm correction	
					P^*	h
3.971	**C12:0**	0.40 ± 0.23	1.02 ± 0.57	9.51%	8.9558e-07	1
5.217	**C14:0**	2.59 ± 0.94	5.94 ± 4.13	8.52%	2.7682e-05	1
5.966	**C15:0**	0.67 ± 0.14	0.96 ± 0.52	8.91%	0.0040	1
6.832	**C16:0**	89.14 ± 18.86	142.96 ± 57.47	5.45%	1.7110e-06	1
7.023	**C16:1n-9c**	1.35 ± 0.40	2.31 ± 0.73	9.44%	3.6508e-10	1
7.103	**C16:1n-7c**	2.88 ± 1.33	5.49 ± 3.03	6.36%	2.2527e-05	1
7.862	**C17:0**	1.46 ± 0.30	1.89 ± 0.58	6.20%	3.8969e-04	1
8.184	**C17:1n-7c**	0.20 ± 0.07	0.40 ± 0.30	11.49%	8.6202e-04	1
9.242	**C18:0**	37.39 ± 6.78	49.78 ± 16.99	5.09%	3.1527e-04	1
9.601	**C18:1n-9c**	56.12 ± 17.59	95.00 ± 39.14	5.14%	8.9558e-07	1
9.682	**C18:1n-7c**	4.30 ± 1.31	7.73 ± 3.56	5.33%	1.0329e-06	1
10.186	**C18:2n-6c**	72.34 ± 14.68	102.76 ± 45.08	5.54%	7.3969e-04	1
10.452	**C18:3n-6c**	0.45 ± 0.23	0.81 ± 0.46	7.29%	1.3319e-04	1
10.571	**C18:3n-3**	0.40 ± 0.19	0.69 ± 0.45	6.24%	0.0020	1
10.764	**C18:3n-3c**	4.12 ± 1.87	7.39 ± 4.97	6.63%	0.0011	1
11.315	**C20:0**	0.47 ± 0.15	0.71 ± 0.51	9.18%	0.0322	1
11.487	**C20:1n-9c**	0.74 ± 0.35	1.24 ± 0.61	5.78%	1.3243e-04	1
11.822	**C20:2n-6c**	1.33 ± 0.31	1.88 ± 0.63	4.37%	2.4112e-05	1
12.003	**C20:3n-6c**	2.83 ± 1.00	4.27 ± 1.77	6.79%	1.3905e-04	1
12.117	**C20:4n-6c**	15.08 ± 2.89	17.55 ± 5.43	6.40%	0.0462	1
12.533	**C20:5n-3c**	1.07 ± 0.72	1.49 ± 1.14	5.93%	0.0948	0
12.622	**C22:0**	0.54 ± 0.15	0.67 ± 0.37	7.76%	0.0948	0
12.801	**C22:1n-9c**	0.66 ± 0.20	0.85 ± 0.42	7.35%	0.0462	1
13.588	**C22:4n-6c**	0.49 ± 0.11	0.74 ± 0.32	5.79%	1.3905e-04	1
13.702	**C22:5n-6c**	0.67 ± 0.23	0.96 ± 0.45	5.94%	0.0027	1
14.119	**C22:5n-3c**	1.80 ± 0.41	2.38 ± 0.73	4.45%	1.5782e-04	1
14.257	**C22:6n-3c**	6.23 ± 2.19	7.26 ± 2.62	7.83%	0.0740	0
14.345	**C24:1n-9c**	0.49 ± 0.16	0.56 ± 0.16	7.53%	0.0688	0
	Total FFA	305.33 ± 61.50	464.11 ± 177.04		5.4221e-06	1

Concentration values are presented as mean ± SD.

* The statistical significance of FFAs between healthy controls and MetS patients was tested using *T*-test and adjusted by Bonferroni–Holm correction for multiple comparisons. A P value of < 0.05 was considered statistically significant and signed '1', or else signed '0'.

is built, all of the data are run down the tree, and proximities are calculated for each pair of cases. To more directly and conveniently observe the patterns in the proximity matrix, MDS is employed to map the proximity into a lower-dimensional space. The MDS plot of serum free fatty acid profiles from HC, MetS and QC groups is shown in Fig. 3. As illustrated in the figure, the samples from QC group lied between HC group and MetS group and clustered tightly in the MDS plot, confirming the reliability of the data for subsequent data analysis. Moreover, a better and clear separation was observed for HC and MetS samples. The prediction ability for RF model internally using an out-of-bag estimation was as follows: sensitivity (0.8611), specificity (0.8644) and accuracy (0.8632). These results indicated that RF algorithm could effectively reflect the distinction of free fatty acids profiles between healthy controls and MetS patients. Thus, the classification and prediction model for MetS by RF could assist the diagnosis of MetS.

3.4. Canonical correlation analysis of FFAs and clinical parameters

In clinical practice, five indicators, including WC, SBP/DBP, TG, HDL-C and FBG, are measured to diagnose MetS. MetS, as a constellation of metabolic abnormalities, especially lipid metabolic disorder, is closely associated with FFAs. However, the association between FFAs and the five components of MetS and which component should be paid more attention are still unknown. So CCA was conducted to explore the relationships between FFAs and clinical parameters, so as to discover biomarkers that are associated

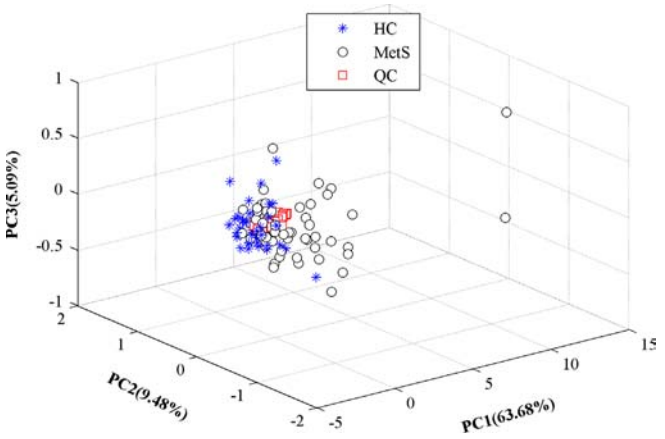

Fig. 2. PCA score plot (PC1, PC2 and PC3) of serum free fatty acid profiles from HC, MetS and QC groups.

with clinical parameters and try to find some other useful information for dealing with metabolic abnormalities of metabolic syndrome. Fig. 4 shows the scatter plot of the first couple of canonical variables for free fatty acids (U1) and clinical parameters (V1). As displayed in the figure, free fatty acids (U1) and clinical parameters (V1) had good correlation with canonical correlation coefficient of 0.9015, and it was validated by Bartlett's approximate chi-squared statistic for null hypothesis with the P value of right-tail significance level as 1.238e-13. The results indicated that free fatty acids were closely correlated with the clinical parameters.

Since FFAs can represent the underlying lipid metabolic disturbances of MetS, and were closely correlated with the clinical parameters, we intended to exploit the important clinical parameters for the correlation, so as to be considered for monitoring lipid metabolic disturbances of MetS in clinical practice. For this purpose, the coefficients of the first couple of canonical variables for the clinical parameters may be employed for representing the importance of the parameters. The plot of absolute values of each variable in the first couple of canonical variables for clinical parameters is shown in Fig. 5. As displayed in the figure, the absolute values of HDL-C, TG, and FBG, especially HDL-C, were significantly higher than those of WC and SBP/DBP. It indicates that these parameters may be more correlated with FFAs, and more attention should be paid to these parameters in clinical practice. Namely, dyslipidemia (characterized by elevation of TG and decrease of HDL-C) and hyperglycemia, especially dyslipidemia, may be more correlated with FFAs than obesity and hypertension. So, more attentions should be paid to the disorder of lipid metabolism and glucose metabolism for MetS.

As for FFAs, the absolute values of each variable in the first couple of canonical variables for FFAs may represent the importance of FFAs which are closely associated with the clinical parameters. Thus, these metabolites may be considered as the potential biomarkers of MetS. Fig. 6 shows the absolute values of the coefficients of the first couple of canonical variables for FFAs. As displayed in the figure, the absolute values of the coefficients for C16:1n-9c, C20:1n-9c, C22:0 and C22:4n-6c were obviously higher than those for other FFAs. However, as presented in Table 2, C22:0 found no significant difference between MetS patients and healthy controls by Bonferroni–Holm correction ($P < 0.05$). Therefore, only C16:1n-9c, C20:1n-9c and C22:4n-6c could be considered as the potential biomarkers of MetS.

Metabolic interpretation of the results is of great importance, although the interpretation of metabolomic research is difficult for the complexity of metabolic pathways. Palmitoleic acid (C16:1n-9c), one of the important monounsaturated fatty acids (MUFAs), can arise from the β- oxidation of oleic acid (C18:1n-9c) [38]. It has been reported that the concentration of C16:1n-9c was highly associated

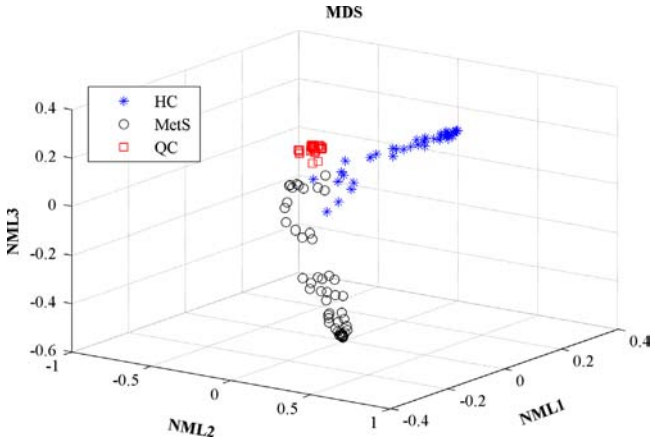

Fig. 3. MDS plot of serum free fatty acid profiles from HC, MetS and QC groups.

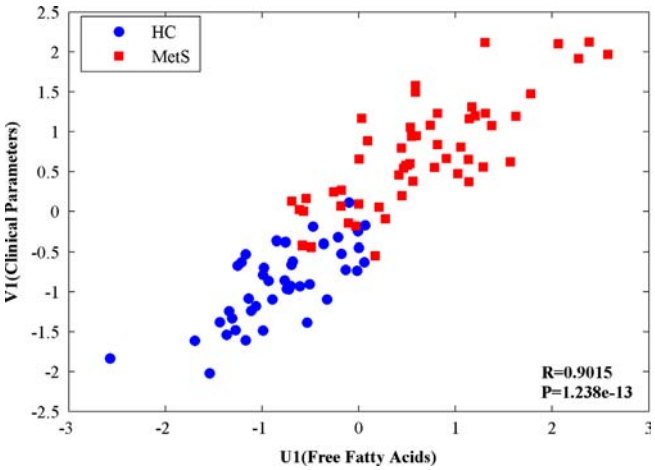

Fig. 4. Scatter plot of the first couple of canonical variables for free fatty acids (U1) and clinical parameters (V1).

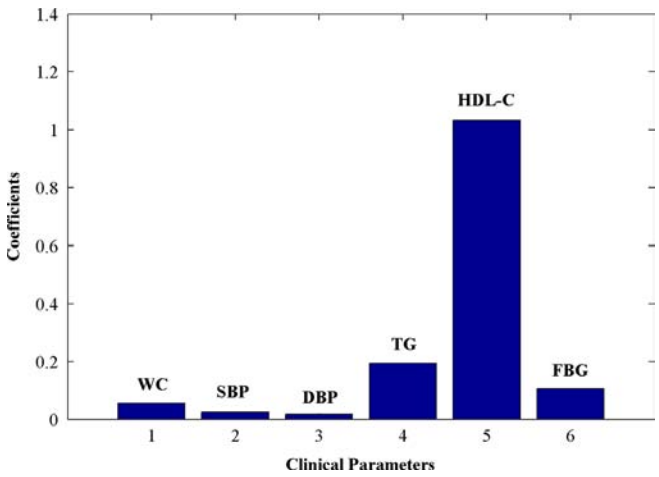

Fig. 5. Plot of the absolute values of coefficients of the first couple of canonical variables for clinical parameters.

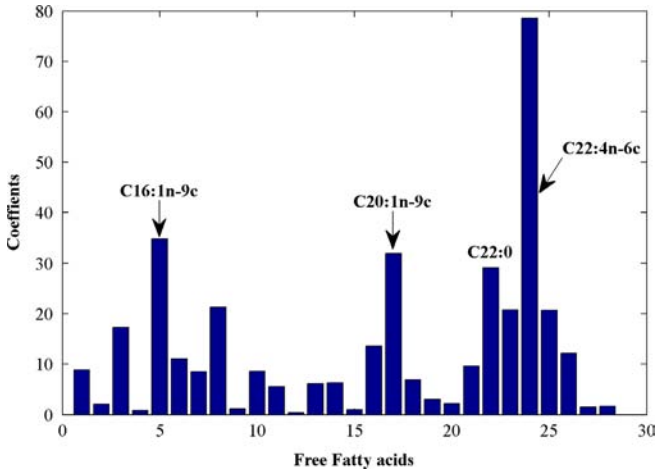

Fig. 6. Plot of the absolute values of coefficients of the first couple of canonical variables for free fatty acids.

with MetS in the study for obese prepubertal children [39]. Furthermore, in a large case-control study for predicting the metabolic syndrome status, palmitoleic acid was used as the most discriminative variable to discriminate between cases and controls since its concentration was significantly higher for MetS patients compared to controls [40]. Nervonic acid (C24:1n-9c) was reported to have preventive effects on diabetes- and obesity-related metabolic disorders [41]. 11-eicosenoic acid (C20:1n-9c), as an intermediate in the stepwise conversion of oleic acid (C18:1n-9c) to nervonic acid, was found to be significantly higher in diabetic rats than in control animals [42]. In our study, we obtained consistent results that C20:1n-9c was significantly higher in MetS patients than healthy controls and it was closely associated with the clinical parameters. Adrenic acid (AdA, C22:4n-6c), an abundant fatty acid in the vasculature, is produced by a two-carbon chain elongation of arachidonic acid (AA, C20:4n-6c) and can also be converted to AA via β-oxidation [43]. Like AA, AdA can be converted into multiple oxygenated metabolites with important roles in various physiological and pathophysiological processes [44]. As the most relevant fatty acids with clinical parameters, C22:4n-6c might be of great importance for the physiological and pathophysiological processes of MetS.

4. Conclusions

In this study, a powerful strategy of metabolic profiling by GC–SIM–MS combined with RF and CCA was developed to explore the perturbations of serum free fatty acids and possible biomarkers of MetS. The results indicated that RF models revealed characteristic and powerful advantages on the discrimination between MetS patients and healthy controls, thus the classification and prediction model for MetS by RF could be used to assist the diagnosis of MetS. Moreover, CCA showed satisfactory correlation between FFA and clinical parameters. As a result, several FFAs, including C16:1n-9c, C20:1n-9c and C22:4n-6c, were screened as the potential biomarkers associated with clinical parameters for MetS. The results of CCA also indicated that HDL-C, TG and FBG were the most important parameters closely correlated with FFAs disturbances of MetS, thus they should be paid more attention to in clinical practice for monitoring FFAs disturbances of MetS than WC and SBP/DBP. The results have demonstrated that the proposed approach may be effective for exploring metabolic perturbations and possible biomarkers for diseases, and may be able to provide useful information for clinical practice.

Acknowledgments

This work is financially supported by the National Nature Foundation Committee of PR China (Nos. 21175157, 21375151, 21105129), China Hunan Provincial Science and Technology Department (No. 2012FJ4139), and Central South University, China for special support of the basic scientific research project (No. 2010QZZD007).

References

[1] R.H. Eckel, S.M. Grundy, P.Z. Zimmet, The Lancet 365 (2005) 1415–1428.
[2] P. Zimmet, D. Magliano, Y. Matsuzawa, J.S.G. Alberti, J. Atheroscler. Thromb. 12 (2005) 295–300.
[3] E.S. Ford, Diabetes Care 28 (2005) 2745–2749.
[4] S.M. Grundy, Arterioscler. Thromb. Vasc. Biol. 28 (2008) 629–636.
[5] V.G. Athyros, E.S. Ganotakis, M. Elisaf, D.P. Mikhailidis, Curr. Med. Res. Opin. 21 (2005) 1157–1159.
[6] B. Isomaa, P. Almgren, T. Tuomi, B. Forsén, K. Lahti, M. Nissén, M.-R. Taskinen, Diabetes Care 24 (2001) 683–689.
[7] S. Mottillo, K.B. Filion, J. Genest, L. Joseph, L. Pilote, P. Poirier, S. Rinfret, E.L. Schiffrin, M.J. Eisenberg, J. Am. Coll. Cardiol. 56 (2010) 1113–1132.
[8] E.S. Ford, C. Li, N. Sattar, Diabetes Care 31 (2008) 1898–1904.
[9] G. Marchesini, E. Bugianesi, G. Forlani, F. Cerrelli, M. Lenzi, R. Manini, S. Natale, E. Vanni, N. Villanova, N. Melchionda, M. Rizzetto, Hepatology 37 (2003) 917–923.
[10] K. Esposito, P. Chiodini, A.L. ENZI, D. GIUGLIANO, Diabetes Care 35 (2012) 2402–2411.
[11] J.K. Nicholson, I.D. Wilson, Nat. Rev. Drug Discov. 2 (2003) 668–676.
[12] T.J. Wang, M.G. Larson, R.S. Vasan, S. Cheng, E.P. Rhee, E. McCabe, G.D. Lewis, C.S. Fox, P.F. Jacques, C. Fernandez, C.J. O'Donnell, S.A. Carr, V.K. Mootha, J.C. Florez, A. Souza, O. Melander, C.B. Clish, R.E. Gerszten, Nat. Med. 17 (2011) 448–453.
[13] W.M. Claudino, A. Quattrone, L. Biganzoli, M. Pestrin, I. Bertini, A. Di Leo, J. Clin. Oncol. 25 (2007) 2840–2846.
[14] M. Bogdanov, W.R. Matson, L. Wang, T. Matson, R. Saunders-Pullman, S.S. Bressman, M. Flint Beal, Brain 131 (2008) 389–396.
[15] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Rev. Drug Discov. 1 (2002) 153–161.
[16] G.D. Kolovou, K.K. Anagnostopoulou, D.V. Cokkinos, Postgrad. Med. J. 81 (2005) 358–366.
[17] S.J. Wakil, L.A. Abu-Elheiga, J. Lipid Res. 50 (2009) S138–S143.
[18] J. Girard, Med. Sci.: M/S 21 (2005) 19–25.
[19] R.N. Bergman, M. Ader, Trends Endocrinol. Metab. 11 (2000) 351–356.
[20] J.P.H. Wilding, Diabet. Med. 24 (2007) 934–945.
[21] M. Wang, R.H. Han, X. Han, Anal. Chem. 85 (2013) 9312–9320.
[22] W.J. Griffiths, Mass Spectrom. Rev. 22 (2003) 81–152.
[23] L. Akoto, R.J. Vreuls, H. Irth, R. Pel, F. Stellaard, J. Chromatogr. A 1186 (2008) 365–371.
[24] F. Destaillats, C. Cruz-Hernandez, J. Chromatogr. A 1169 (2007) 175–178.
[25] L. Zhang, P. Li, X. Sun, W. Hu, X. Wang, Q. Zhang, X. Ding, Anal. Chim. Acta 839 (2014) 44–50.
[26] L. Breiman, Mach. Learn. 45 (2001) 5–32.
[27] H. Hotelling, Biometrika 28 (1936) 321–377.
[28] Joint committee for developing Chinese guidelines on prevetion and treatment of dyslipidemia in adults, Chin. J. Cardiol. 35 (2007) 390–419.
[29] Expert Panel on Metabolic Syndrome of Chinese Diabetes Society, Chin. J. Diabetes 12 (2004) 156–161.
[30] K.G.M.M. Alberti, P. Zimmet, J. Shaw, The Lancet 366 (2005) 1059–1062.
[31] L.Z. Yi, J. He, Y.Z. Liang, D.L. Yuan, F.T. Chau, FEBS Lett. 580 (2006) 6837–6845.
[32] T. Sangster, H. Major, R. Plumb, A.J. Wilson, I.D. Wilson, Analyst 131 (2006) 1075–1078.
[33] W.B. Dunn, I.D. Wilson, A.W. Nicholls, D. Broadhurst, Bioanalysis 4 (2012) 2249–2264.
[34] L. Zhang, B. Tan, M. Zeng, H. Lu, Y. Liang, Talanta 88 (2012) 311–317.
[35] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, Nat. Protoc. 6 (2011) 1060–1083.
[36] M. Zeng, Y. Liang, H. Li, M. Wang, B. Wang, X. Chen, N. Zhou, D. Cao, J. Wu, J. Pharm. Biomed. Anal. 52 (2010) 265–272.
[37] D. Gu, K. Reynolds, X. Wu, J. Chen, X. Duan, R.F. Reynolds, P.K. Whelton, J. He, The Lancet 365 (2005) 1398–1405.
[38] J.H. Wu, R.N. Lemaitre, F. Imamura, I.B. King, X. Song, D. Spiegelman, D.S. Siscovick, D. Mozaffarian, Am. J. Clin. Nutr. 94 (2011) 431–438.
[39] M. Gil-Campos, M. del Carmen Ramirez-Tortosa, E. Larque, J. Linde, C.M. Aguilera, R. Canete, A. Gil, Lipids 43 (2008) 723–732.
[40] F. Szabo de Edelenyi, L. Goumidi, S. Bertrais, C. Phillips, R. MacManus, H. Roche, R. Planells, D. Lairon, Genes Nutr. 3 (2008) 173–176.
[41] E. Oda, K. Harada, J. Kimura, Y. Aizawa, P.V. Thanikachalam, K. Watanabe, Int. Heart J. 46 (2005) 975–985.
[42] W.J. Malaisse, L. Portois, A. Sener, Y.A. Carpentier, Int. J. Mol. Med. 22 (2008) 133–137.
[43] C.J. Mann, T.L. Kaduce, P.H. Figard, A.A. Spector, Arch. Biochem. Biophys. 244 (1986) 813–823.
[44] C. Guijas, A.M. Astudillo, L. Gil-de-Gomez, J.M. Rubio, M.A. Balboa, J Balsinde, Biochim. Biophys. Acta 2012 (1821) 1386–1393.