Talanta 110 (2013) 1–7

Contents lists available at SciVerse ScienceDirect

Talanta

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

Exploring the relationship between 5′AMP-activated protein kinase and markers related to type 2 diabetes mellitus

Jian-Hua Huang [a], Rui-Hua He [1,a], Lun-Zhao Yi [a], Hua-Lin Xie [b], Dong-sheng Cao [a], Yi-Zeng Liang [a,*]

[a] Research Center of Modernization of Traditional Chinese Medicines, Central South University, Changsha 410083, PR China
[b] School of Chemistry and Chemical Engineering, Yangtze Normal University, Fuling 408100, PR China

ARTICLE INFO

Article history:
Received 15 December 2012
Received in revised form
12 March 2013
Accepted 14 March 2013
Available online 22 March 2013

Keywords:
5′AMP-activated protein kinase (AMPK)
Metabolites
Random forest
GC/MS
Gender variation

ABSTRACT

The importance of 5′AMP-activated protein kinase (AMPK) in regulating glucose and fatty acid metabolism is increasing. Thus, it is regarded as a new pharmacological target for treatment of obesity, insulin resistance and type 2 diabetes mellitus (T2DM). In order to explore the relationships between AMPK and diabetes mellitus, urines samples from four groups of C57 mice, i.e., the normal male and female C57 mice, female C57-AMPK gene knocked-out mice, and male C57-AMPK gene knocked-out mice, were studied by coupling GC/MS with a powerful machine learning method, random forest. The experimentation has been designed as two steps: firstly, the normal male and female mice were compared with male and female C57-AMPK gene knocked-out mice, respectively; then the differences between male C57-AMPK gene knocked-out mice and female C57-AMPK gene knocked-out mice were further detected. Finally, not only the differences between the normal C57 mice and C57-AMPK gene knocked-out mice were observed, but also the gender-related metabolites differences of the C57-AMPK gene knocked-out mice were obviously visualized. The results obtained with this research demonstrate that combining GC/MS profiling with random forest is a useful approach to analyze metabolites and to screen the potential biomarkers for exploring the relationships between AMPK and diabetes mellitus.

1. Introduction

Type II diabetes mellitus (T2DM) is a worldwide health problem of the public with the rising of people's living standard. It deserves much more attentions due to the increasing prevalence of this condition in the whole world, especially western countries. Numerous human, even children, are suffering from it in the form of various acute complications, such as diabetic ketoacidosis, kidney failure, stroke amputations, and so on [1]. T2DM is considered to be a complex heterogeneous disease, a typical disease of metabolic syndrome and imbalance of energy metabolism. Previous researches have demonstrated that carbohydrate metabolic block is the most prominent characteristic of T2DM, and carbohydrate, especially the glucose has been widely used as biochemical marker for diagnosis, screening, and evaluation of glycemia control of the disease [2]. Recent data collected in several laboratories indicated that 5′AMP-activated protein kinase (AMPK) plays a key role in regulation of carbohydrate and fat metabolism, serving as a metabolic master switch in response to alterations in cellular energy charge [3–5]. Therefore, AMPK is emerging as a potentially interesting drug target for treatment of diabetes [6].

The AMPK is a multi-substrate serine/threonine protein kinase that is ubiquitously-expressed and functions as an intracellular fuel sensor activated by depletion of high energy phosphor compounds [7]. Activation of AMPK initiates a complex series of signaling events, causing an increase in uptake and oxidation of substrates for adenosine tri-phosphate (ATP) synthesis concurrent with decreasing ATP consuming biosynthetic processes such as protein, lipid, and glycogen synthesis [8]. Furthermore, activated AMPK is associated with dramatic changes in the control of glucose and fatty acid metabolism. Work by several groups during the past decade have revealed that AMPK probably serves as a key metabolic sensor in both insulin-sensitive and other tissues that is capable of responding to metabolic stresses (and in particular depletion of intracellular ATP) by shutting down the synthesis of fatty acids and cholesterol, two major energy-consuming pathways [8–12]. The dysfunction of AMPK activity would confuse many metabolic pathways and resulting in the metabolites perturbation. Therefore, monitoring the variations in metabolites could provide many evidences for exploring the relationship between AMPK and diabetes mellitus.

High-throughput metabolomics have been widely used in the biomedical sciences and have been proved to be a powerful

Abbreviations: 5′AMP-activated protein kinase, AMPK; Type II diabetes mellitus, T2DM; Adenosine tri-phosphate, ATP; Random forest, RF; Multidimensional scaling, MDS.

* Corresponding author. Tel/Fax.: +86731-88830831;
E-mail address: yizeng_liang@263.net (Y.-Z. Liang).

[1] The first two authors contributed equally to this work.

0039-9140/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.talanta.2013.03.039

approach in diagnosis of human diseases [13], physiological evaluations [14], elucidation of biomarkers [15,16], and drug toxicity [17]. In these researches, most samples were blood, serum, and urine. Compared with other matrixes, urine samples are non-invasive, easily obtained and have large sample volume, so it would be the first choice to perform such researches. To date, many analytical techniques have been applied to the metabolomic analysis of urine, including gas chromatography/mass spectrometry GC/MS [18–21], high-resolution nuclear magnetic (NMR) [22–25], ultra-performance liquid chromatography/mass spectrometry (UPLC/MS) [26]. GC/MS, as one of the most widely used techniques, can generate metabolic data and simultaneously provides quantitative, qualitative and structural information on a wide range of biological molecules. For its favorable stability, reproducibility, sensitivity and better separation of compounds [27], GC/MS is considered to be one of the most effective tools to perform the research of metabolomics.

In the present study, we constructed models by using random forest (RF) to visually discriminate C57-AMPK gene knocked-out mice from healthy controls and to investigate the metabolic differences between male and female C57-AMPK gene knocked-out mice. Firstly, urines samples from normal mice, female C57-AMPK gene knocked-out mice and male C57-AMPK gene knocked-out mice were profiled by GC/MS. Then, the qualitative and the quantitative work were carried out by using heuristic evolving latent projections (HELP) and selected ions analysis (SIA). Finally, RF was employed to visualize the differences among four groups by assembling enough classification and regression trees. Some informative metabolites or potential biomarkers have been successfully discovered by means of variable importance ranking in random forest program. Using the obtained sample proximity matrix, not only the differences between the normal C57 mice and C57-AMPK gene knocked-out mice were observed, but also the gender-related metabolites differences of the C57-AMPK gene knocked-out mice were obviously visualized.

2. Materials and methods

2.1. Chemicals, reagents and preparation of standard solutions

Adonitol, pyridine, *o*-methylhydroxylamine hydrochloride, urease and *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) containing 1% trimethylchlorosilane (TMCS) were all obtained from Sigma-Aldrich (Shanghai, China). Methanol was of analytically pure grade. Adonitol, as internal standard, was prepared in methanol at a concentration of 2 mg/mL. A 2 mg/mL urease solution was freshly prepared by dissolving urease powder in deionized water that produced by a ultra-pure water system (Molecular, Shanghai) and *o*-methylhydroxylamine hydrochloride pyridine solution with the concentration 15 mg/mL was prepared.

2.2. Acquisition of mice urine samples

Urine samples were collected from normal C57 mice (eight male mice and eight female mice) and other two groups of C57 mice, whose AMPK gene were knocked out, say C57-AMPK gene knocked-out mice of different gender (female: $n=47$ and male: $n=39$ respectively). Normal C57 mice and the C57-AMPK gene knocked-out mice were provided by Experimental Zooprophylactic Institute of Chinese Academy of Medical Sciences (CAMS) (Beijing, China). Until the urine samples were collected, the experimental mice were raised by institute of laboratory animal science in CAMS. The treated mice were raised at the temperature between 20 and 6 °C, and other ambient conditions were as follows: relative humidity was 40–70%, ventilation rate was 10–15 per hour, illuminance was 150–300 lx, and the alternation of light and shadow was 12 h per each. Food (standard mouse diet) was available. Samples were taken at the same time each morning and were stored at −80 °C prior to analysis.

2.3. Preparation of the metabolites extraction and derivatization

Aliquots of 500 μL mice urine was centrifuged at 4 °C, 16,000 rpm/min for 5 min, and 150 μL supernatant added with 100 μL, 2 mg/mL urease solution was incubated at room temperature for 30 min to remove and decompose excess urea present in it. The resulted solution was protein precipitated using 800 μL methanol, and then 2 mg/mL adonitol–methanol was added to it as the internal standard. The mixture was vigorously eddied for 30 s, followed by reaction at room temperature for 10 min. Next, the solution was centrifuged at 4 °C, 16,000 rpm ($17800\times g$) for 10 min, 150 μL supernatant was taken out and evaporated to dryness in a vacuum oven. The dried extractive was derivatized to its (MeOx-) TMS-derivatives through reaction with 40 μL of 15 mg/mL methoxyamine hydrochloride solution in pyridine at 70 °C for 30 min. After methoximation reaction, the samples were trimethylsilylated for another 30 min by adding 60 μL BSTFA-1% TMCS as catalyst. At last, the solution was vigorously eddied for 30 s again before GC–MS analysis.

2.4. GC–MS data acquisition

Following protein precipitation and derivatization, gas chromatography-mass spectrometry (Shimadzu GC/MS-QP2010, Japan) was used to analyze the metabolic profiling of all urine samples. The instrument is equipped with a 30 m × 0.25 mm ID, fused silica capillary column, which was chemically bonded with 0.25 mm DB-5ms stationary phase (Agilent, USA). The helium carrier gas flow rate was 1.0 mL/min. The column initial temperature was kept at 70 °Cfor 5 min. Then the temperature was ramped at a rate of 20 °C/min to 160 °C, 4 °C/min to 180 °C and 10 °C/min to 300 °C, and held for 1.5 min at 300 °C. 1 μL of the metabolite derivative solution with the derivatization reagent was run through the gas chromatograph-mass spectrometer with a 10:1 split throughout. The injector temperature was 280 °C, the septum purge flow rate was 3 mL/min, and the purge was turned on all the time. The total GC run time was 28 min. The interface temperature was 250 °C and ion source temperature was 200 °C. Ionization was achieved by a 70 eV electron beam. Masses were acquired in a full scan mode, over the range from m/z 35 to 800, with a scan speed of 0.2/s when the 0.9 kV of detector voltage was turned on after a solvent delay of 5 min.

2.5. Data handling and analysis

First, a chromatogram that has the greatest number of peaks was selected as the referential chromatogram. The identification of its specific metabolites and internal standard was based on the search results in NIST 05 mass chromatography library in the GC–MS Postrun Analysis software (Shimadzu), the characteristic ions according to the literature and retention time of authentic standards and those of online available data. For quantitative analysis of pure chromatographic peaks, areas were obtained directly by the GC–MS Postrun Analysis software (Shimadzu). As for some overlapped peaks, two chemometric resolution methods, say HELP [28,29] and SIA [30] were utilized to extract the pure mass spectra, and calculate the corresponding area. After the referential chromatogram was processed, those chromatograms from the other samples were compared with the referential chromatogram based on retention times. If a peak in the referential chromatogram was

not in a sample, its peak area was set to zero. If the area of a peak was zero in most of samples, this peak will be discarded. Finally, 41 peaks, which represented each urine sample, were remained. Thus, a data table, whose row and column represent a sample and a variable, respectively, was generated for statistics analysis; and subsequent statistical analysis was carried out into our custom scripts in MATLAB 7.6 (The MathWorks, Inc., USA).

2.6. Random forest

Random forest (RF) is a classifier consisting of an ensemble of tree-structured classifiers [31]. RF takes advantages of two powerful machine learning techniques: bagging and random feature selection. In bagging, each tree is trained on a bootstrap sample of the training data, and predictions are made by majority vote of trees. RF is a further development of bagging. Instead of using all features, RF randomly selects a subset of features to split at each node when growing a tree. To assess the prediction performance of the random forest algorithm, it performs a type of cross-validation in parallel with the training step by using the so-called out-of-bag (OOB) samples. Specifically, in the process of training, each tree is grown using a particular bootstrap sample. Since bootstrapping is sampling with replacement from the training data, a part of the samples will be 'left out' in the sampling, while another will be repeated in the sampling. The 'left out' data constitute the OOB sample. On average, each tree is grown using about 2/3 of the training data, leaving about 1/3 as OOB [31]. Because OOB data have not been used in the tree construction, one can use them to estimate the prediction performance. The RF algorithm was implemented by the randomForest R package. The algorithm can be stated as follows:

1. Draw n_{tree} bootstrap samples from the original data, n is the number of ensemble tree;
2. For each of the bootstrap samples, grow an un-pruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all variables, randomly sample m_{try} of the variables and choose the best split from among those variables (bagging can be thought of as the special case of random forests obtained when $m_{try}=p$, the number of variables). In general, m_{try} can be chosen to be some function of p. the default values of m_{try} ($p^{1/2}$ for classification and $p^{1/3}$ for regression) were chosen based on empirical experiments, and the performance of RF seems to change very little over a wide range of values [31].
3. Predict new data by aggregating the predictions of the n_{tree} (i.e., majority votes for classification, average for regression).

2.6.1. Variable importance

RF, as an ensemble of trees, inherits the ability to select 'important' features. A measure of how each feature contributes to the prediction performance of RF can be calculated in the course of training. The importance scores can be used to identify biomarkers. The frequently used type of the RF variable importance measure: the mean decrease in classification is based on permutation. For each tree, the classification accuracy of the OOB samples is determined both with and without random permutation of the values of the variable one by one. The prediction accuracy of after permutation is subtracted from the prediction accuracy before permutation and averaged over all trees in the forest to give the permutation importance value. In the current research, the mean decrease in classification accuracy was accepted to measure variable importance.

2.6.2. Proximity measure

Another attractive feature of RF is the proximity matrix. The proximity matrix can be used to identify structure in the data. RF not only generates variable-related information such as variable importance measures, but also calculates the proximities between samples. For proximity calculations, all samples in the original data set are classified by the forest. The proximity between two samples is calculated as the number of times the two samples end up in the same terminal node of a tree, divided by the number of trees in the forest. The resulting matrix is symmetric with diagonal element equal to 1 and off-diagonal elements ranging from 0 to 1. The proximities between similar samples are always high. Proximity scores may also be used to construct multi-dimensional scaling (MDS) plots [32]. MDS plots aim to visualize the similarity or dissimilarity (calculated as 1- proximity) between samples. Normally, a good class separation could be obtained by plotting the first two or three scaling coordinates against each other.

3. Results and discussion

3.1. Quantity and quantification analysis of urine samples

The goal of metabonomic research is to analyze as many metabolites as possible in biofluids. Therefore, our target was to develop a method that was likely to extract as many compounds as possible with high efficiency and reproducibility. Under the optimum analytical method, the metabolic profiles of metabolites from C57-AMPK gene knocked-out mice and C57 mice were obtained by GC–MS. After deconvolution, various kinds of metabolites, including carbohydrates, organic acid, fatty acid, and amine were found in the chromatograms. Then, the qualitative and the quantitative work were carried out subsequently. According to the description in "Data handling and analysis", identification of the metabolites was carried out by the standard components and NIST 05. Finally, 41 endogenous metabolites with high NIST match ($>80\%$ similarity) were selected for detailed analysis (Table 1).

Fig. 1 showed the typical total ion chromatograms (TICs) of the normal mice, male C57-AMPK gene knocked-out mice, and female C57-AMPK gene knocked-out mice. It was hard to distinguish the differences among these groups directly from chromatograms in Fig. 1. Furthermore, because of the inter-animal variation in urine matrix composition and the complexity of the total ion chromatograms (TICs), coupled with differences in the content of metabolites, visual comparison of those metabolic profiles was very difficult. Hence, some powerful statistic methods were urgently needed to overcome these problems.

3.2. Data analysis

Based on the results of qualitative and quantitative analysis, all the metabolites were used as variables for discrimination. A random forest model consists of a set of classification trees that are constructed by bootstrapping the training data. After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. The sample proximity matrix, deriving from these training of tress, can be generated to collect the samples similarity information. As similar samples may fall into the same terminal node or derive from the same parent node. Thus, the samples in the same group always have larger similarity values than compared with other group samples. Fig. 2 showed the sample proximity plot of the four group data. The samples in the same group have a deeper color than other groups.

We can use the similar matrix for finding outliers observations and discovering clusters in the data through multidimensional

Table 1
Qualitative and quantitative metabolic profile of three group mice.

Id	t_r[a] (min)	Endogenous metabolites	C57-male	C57-female	AMPK-male	AMPK-female
1	5.922	Aminoethane	0.2456 ± 0.0705	0.1782 ± 0.0352	0.1905 ± 0.0567	0.1958 ± 0.0551
2	6.593	Ethylene glycol	0.0182 ± 0.0020	0.0373 ± 0.0107	0.0530 ± 0.0428	0.0746 ± 0.0626
3	6.84	*N*,*N*-Diethylacetamide	0.0657 ± 0.0087	0.0584 ± 0.0099	0.0476 ± 0.0202	0.0557 ± 0.0107
4	7.716	Lactic acid*	0.0872 ± 0.0374	0.0765 ± 0.02934	0.0952 ± 0.0592	0.1482 ± 0.2155
5	7.934	Acetic acid	0.0856 ± 0.0333	0.0341 ± 0.0254	0.0229 ± 0.0140	0.0412 ± 0.0203
6	10.01	Phosphate	2.1278 ± 0.9173	2.7785 ± 1.4037	1.4730 ± 0.7381	1.3767 ± 0.9361
7	10.2	L-Threonine	0.0173 ± 0.0098	0.0052 ± 0.0029	0.0108 ± 0.0068	0.0096 ± 0.0065
8	10.297	Phenylacetic acid	0.0047 ± 0.0023	0.0183 ± 0.0054	0.0159 ± 0.0103	0.0147 ± 0.0097
9	10.382	Succinic acid*	0.0311 ± 0.0129	0.0117 ± 0.0084	0.0098 ± 0.0031	0.0119 ± 0.0086
10	10.447	1,2-Hydroquinone	0.0120 ± 0.0072	0.0103 ± 0.0073	0.0078 ± 0.0047	0.0067 ± 0.0039
11	10.503	Glyceric acid	0.0961 ± 0.0266	0.0250 ± 0.0120	0.0400 ± 0.0232	0.0183 ± 0.0087
12	10.723	(R^*,R^*)-2,3-Dihydroxybutanoic acid	0.0167 ± 0.0053	0.0045 ± 0.0013	0.0037 ± 0.0014	0.0053 ± 0.0029
13	11.357	2,4-Dihyoxybutanoic acid	0.0147 ± 0.0051	0.0131 ± 0.0041	0.0155 ± 0.0080	0.0166 ± 0.0047
14	11.583	(R^*,S^*)-3,4-Dihydroxybutanoic acid	0.0304 ± 0.0098	0.0161 ± 0.0053	0.0132 ± 0.0064	0.0178 ± 0.0107
15	11.797	*N*-(1-oxobutyl)- Glycine	0.0653 ± 0.0244	0.0191 ± 0.0096	0.0319 ± 0.0186	0.0274 ± 0.0151
16	12.341	Isovaleroglycine	0.0356 ± 0.0134	0.0134 ± 0.0041	0.0160 ± 0.0079	0.0107 ± 0.0073
17	12.483	D-Threitol	0.0714 ± 0.0273	0.0665 ± 0.0260	0.0290 ± 0.0130	0.0251 ± 0.0151
18	12.645	*N*-Crotonyl glycine	0.0240 ± 0.0146	0.328 ± 0.0074	0.0207 ± 0.0129	0.0148 ± 0.0099
19	12.973, 13.203	2,3,4-Trihydroxybutyrate	0.1276 ± 0.0162	0.0475 ± 0.0244	0.0631 ± 0.0343	0.0412 ± 0.0250
20	14.53	*N*-(1-oxohexyl)-glycine	0.0960 ± 0.0319	0.0112 ± 0.039	0.0421 ± 0.0273	0.0232 ± 0.0081
21	14.58	3-Hydroxyphenylacetic acid	0.0326 ± 0.0100	0.0122 ± 0.0910	0.0140 ± 0.0081	0.0134 ± 0.0088
22	14.713	D-Xylose	0.0408 ± 0.0150	0.0225 ± 0.0206	0.0182 ± 0.0044	0.0193 ± 0.0053
23	14.823, 15.057	D-Ribose	0.0926 ± 0.0370	0.0682 ± 0.0340	0.0252 ± 0.0142	0.0250 ± 0.0179
24	15.509, 15.733	Arabitol	0.0287 ± 0.0164	0.0252 ± 0.0168	0.0283 ± 0.0179	0.0278 ± 0.0215
25	16.023	6-Deoxy-D-Galactose,	0.0336 ± 0.0083	0.0370 ± 0.0064	0.0177 ± 0.0100	0.0149 ± 0.0104
26	16.087	Mannonic acid	0.0505 ± 0.0177	0.0419 ± 0.0168	0.0211 ± 0.0143	0.0168 ± 0.0138
27	16.2	*cis*-Aconitic acid*	0.0535 ± 0.0288	0.0524 ± 0.0209	0.0105 ± 0.0079	0.0168 ± 0.0147
28	16.357	Phosphoric acid	0.0414 ± 0.0202	0.0463 ± 0.0177	0.0230 ± 0.0141	0.0212 ± 0.0168
29	17.177	Isocitric acid*	0.0348 ± 0.0121	0.0410 ± 0.0157	0.0140 ± 0.0093	0.0248 ± 0.0138
30	17.563	Hippuric acid	0.0470 ± 0.0126	0.0201 ± 0.0111	0.0180 ± 0.0074	0.0156 ± 0.0096
31	17.85, 17.96	D-Fructose*	0.0512 ± 0.0286	0.0728 ± 0.0106	0.0371 ± 0.0145	0.0480 ± 0.0131
32	18.087	*N*-Phenyl glycine*	0.0596 ± 0.0214	0.0363 ± 0.0194	0.0455 ± 0.0272	0.0389 ± 0.0287
33	18.197, 18.147	D-Glucose*	0.3785 ± 0.1618	0.2254 ± 0.1328	0.1741 ± 0.0654	0.1859 ± 0.0736
34	18.507	Altronic acid	0.0302 ± 0.0069	0.0164 ± 0.0106	0.0185 ± 0.0100	0.0102 ± 0.0074
35	18.577, 18.65	D-Sorbitol*	0.0896 ± 0.0269	0.0623 ± 0.0171	0.0254 ± 0.0187	0.0300 ± 0.0275
36	18.983, 19.533	Galactonic acid	0.0613 ± 0.0282	0.0387 ± 0.0186	0.0617 ± 0.0328	0.0441 ± 0.0351
37	19.99	Palmitic acid	0.0084 ± 0.0009	0.0085 ± 0.0047	0.0067 ± 0.0017	0.0071 ± 0.0025
38	20.403	Myo-Inositol	0.0347 ± 0.0228	0.0089 ± 0.0033	0.0097 ± 0.0037	0.0134 ± 0.0129
39	25.465	D-Turanose	0.0216 ± 0.0138	0.0145 ± 0.0118	0.0197 ± 0.0090	0.0510 ± 0.0099
40	25.653, 25.783	D-(+)-lactose monohydrate*	1.0400 ± 0.3349	0.9997 ± 0.2385	0.7475 ± 0.2366	0.6559 ± 0.3286
41	25.927	Lactose*	0.0142 ± 0.0043	0.0231 ± 0.0039	0.0143 ± 0.0075	0.0190 ± 0.0163

* Identified by standard substances.
[a] Retention time.

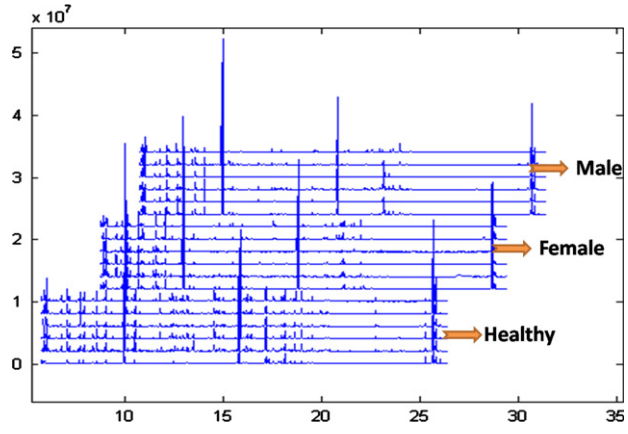

Fig. 1. The total ion chromatograms of urine samples from three groups mice.

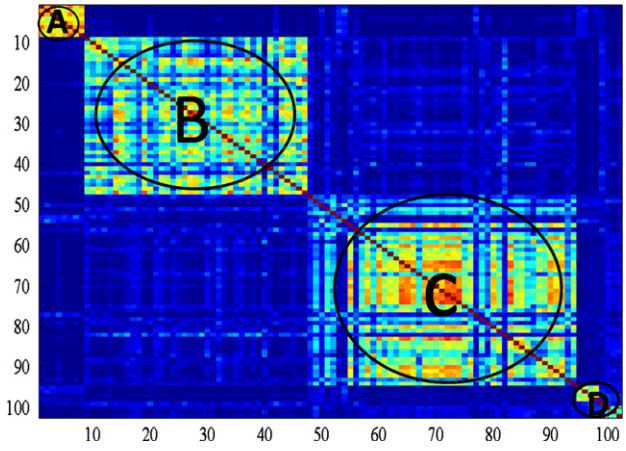

Fig. 2. The sample proximity plot of three groups mice. (A) Healthy male mice, (B) male C57-AMPK gene knocked-out mice, (C) female C57-AMPK gene knocked-out mice, (D) healthy female mice. Each group mice has a larger similar value in its group compare to that with different groups.

scaling (MDS). MDS is a set of related statistical techniques often used in information visualization for exploring similarities or dissimilarities in data. To more directly and conveniently observe the patterns in the proximity matrix, MDS was employed to map the proximity into a lower-dimensional space. From Fig. 3, a good separation for four groups could be significantly observed. Slight differences between the healthy male mice and healthy female mice could be observed, the two groups were close to each other. Comparing to the healthy male mice, the healthy female

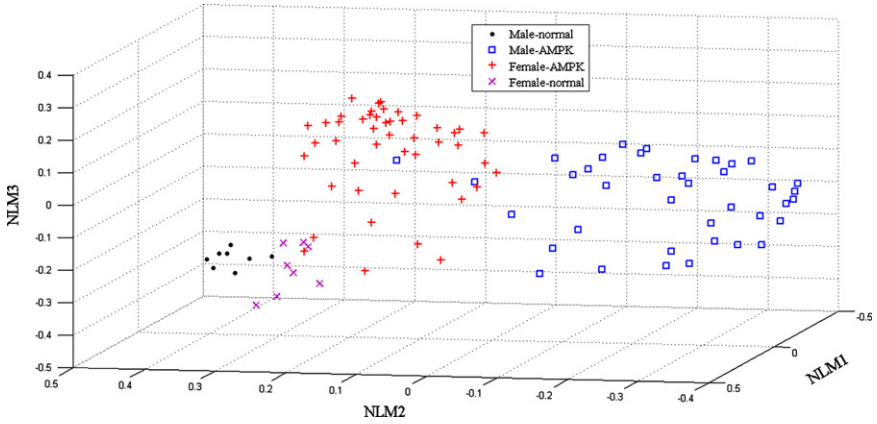

Fig. 3. The MDS plot of three group mice. A good separation can be significantly observed from this plot. Black circle: Healthy male mice, blue rectangle: Male C57-AMPK gene knocked-out mice; red plus: Female C57-AMPK gene knocked-out mice, red cross: Healthy female mice. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

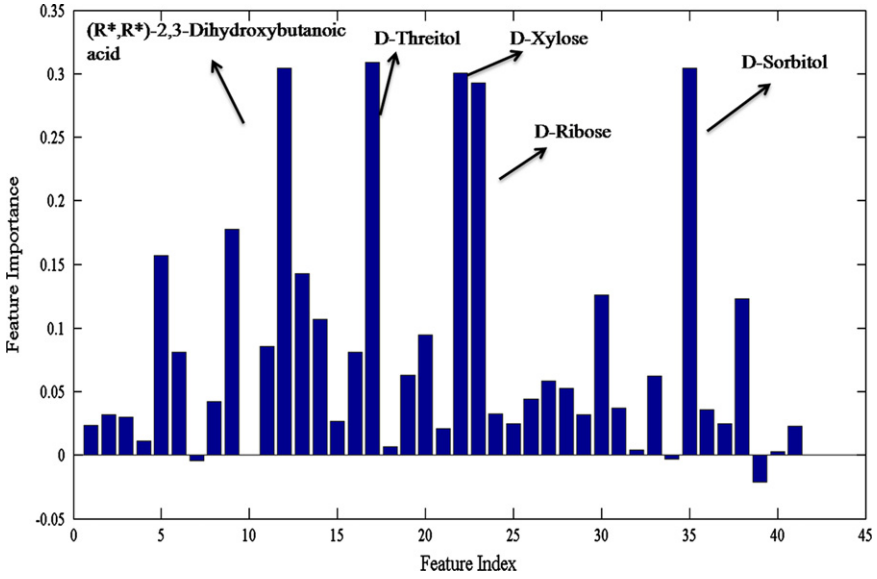

Fig. 4. The variable importance measure obtained by random forest between normal male mice and male C57-AMPK gene knocked-out mice.

mice were nearer to the female-AMPK mice. Furthermore, a good separation between male and female AMPK samples was also obtained using MDS, though two male samples were very close to female AMPK group. These sufficiently indicated that the metabolic characters between healthy and abnormal mice have an obvious distinction, and the genders of mice also have deeply influence to the metabolic. More detailed analysis for each pairs of group was done in the following sections.

3.3. Biomarker screening between normal male mice and male C57-AMPK gene knocked-out mice

Then, RF approach was applied to uncover the underlying information in data. The parameters for RF were set to $n_{tree}=8000$, $m_{try}=7$. With these settings, 8000 classification trees were independently grown and assembled for further research. In model construct, the general purpose of feature selection is to find the best combination of features, which provides the best classification result. All the 41 compounds in normal male mice and male C57-AMPK gene knocked-out mice were used as variables for discrimination analysis. The feature important for each variable was showed in Fig. 4.

Some of metabolites related to the metabolism of lipid and fatty acids, such as (R^*,R^*)-2,3-dihydroxybutanoic acid, D-threitol, D-xylose, D-ribose, and D-sorbitol, have great contribution to classification accuracy. These metabolites could be considered as potential biomarkers for diagnosing the dysfunction of AMPK in male mice. It was worth noting that the results obtained above seemed to be well consistent with the results obtained from biochemical researches.

(R^*, R^*)-2,3-Dihydroxybutanoic acid also known as 4-deoxythreonic acid is a normal organic acid present in human biofluids. 4-Deoxythreonic acid has been previously reported as potential biomarkers for diabetes using hyphenated GC–MS or HPLC-MS techniques [33–35]. Metabolic profiling of urinary organic acids from patients with diabetes mellitus has revealed significantly elevated levels of 4-deoxythreonic acids. A pathway for the likely production of this metabolite has been postulated using an animal model, in which it was suggested that this molecule was produced by L-threonine metabolism [34]. L-Threonine is a ketogenic amino acid; hence its catabolism yields products that are able to enter into energy producing metabolic pathways such as the krebs cycle [34]. L-Threonine has three major routes of degradation, two of which produce acetyl-coa that is a major source of carbon in the krebs cycle, while the third route produces a precursor for the formation of isoleucine. The identification of 4-deoxythreonic acid is expected to provide the basis

for further studies involving larger sample sets in order to establish its possible role as a disease marker.

D-Threitol is a C4-polyol, and it can be regarded as the main end product of D-xylose metabolism in man. Several studies have revealed that urinary levels of some polyols may vary in diseases associated with carbohydrate metabolism derangements such as diabetes mellitus and uremia [36]. Research by Gruzman et al. showed that D-xylose increases the rate of glucose transport in a non-insulin-dependent manner in rat and human myotubes in vitro. This effect resulted from the activation of AMP-activated protein kinase without recruiting the insulin transduction mechanism [37]. D-Ribose is a component of many so-called energy sources for exercising muscle. The research taken by Dhanoa et al. suggests the potential use of ribose in helping diabetics control their blood sugar levels [38].

D-Sorbitol is a sugar alcohol, which can be produced by the body. Only a small proportion of glucose is metabolized to sorbitol during normoglycemia, while in hyperglycemia the enzyme aldose reductase (AR) is activated, leading to an accumulation of intracellular sorbitol and fructose that increases the flux through the polyol pathway [39]. Sorbitols and other polyols accumulate intracellularly, leading to osmotic damage and swelling. AR is the first and rate-limiting enzyme of the polyol pathway, which converts monosaccharides (e.g. glucose) to their polyols or sugar alcohols. This enzyme is widely distributed throughout the body, including those tissues that are susceptible to chronic diabetic complications (e.g. retina, lens, cornea, glomerulus, nervous system and the blood vessels) [40–42]. It has been reported that "Too much sorbitol in cells can cause damage. Diabetic retinopathy and neuropathy may be related to too much sorbitol in the cells of the eyes and nerves". In fact, alterations in sorbitol and fructose metabolism are implicated as factors contributing to complications in diabetes mellitus [42]. This indicated that sorbitol metabolic disorders can possibly induce type 2 diabetes mellitus.

3.4. Biomarker screening between normal female mice and female C57-AMPK gene knocked-out mice

Just following the experimental conditions in above section, all the 41 compounds in normal female mice and female C57-AMPK gene knocked-out mice were used as variables for discrimination analysis. The feature important for each variable was showed in Fig. 5. As could be seen from Fig. 5, several metabolites were consistent with these in male mice, such as, D-threitol, and D-sorbitol. Meanwhile, some special metabolites were also found for female mice, such as *N*-crotonyl glycine, and mannonic acid. These metabolites could be considered as potential biomarkers for diagnosing the dysfunction of AMPK in female mice.

3.5. Biomarker screening between female and male C57-AMPK gene knocked-out mice

This section aims to investigate the metabolic differences between male and female C57-AMPK gene knocked-out mice. As could be seen from Fig. 3, a good separation between male and female samples was obtained using random forest and GC–MS data. In order to evaluate the predictive ability of the proposed method, RF has been employed to classify male C57-AMPK gene knocked-out mice and female C57-AMPK gene knocked-out mice. The prediction accuracy, sensitivity, and specificity for current method were 91.86%, 89.74%, and 93.62%, respectively. This indicates that good models are achieved for classifying mice gender.

A further analysis by using RF was carried out to screen the gender-specific metabolites. The variable importance measurement has been implemented. The results were showed in Fig. 6. A set of gender-specific urinary metabolites was detected. Some compounds in the urinary samples were considered to have great contributions to distinguish female and male C57-AMPK gene knocked-out mice in Fig. 6, such as (*R**,*R**)-2,3-dihydroxybutanoic acid, glyceric acid, *N*-(1-oxohexyl)-glycine, and D-turanose. The detection concentrations of (*R**,*R**)-2,3-dihydroxybutanoic acid and *N*-(1-oxohexyl)-glycine in female C57-AMPK gene knocked-out mice are double higher than these in male C57-AMPK gene knocked-out mice; while the detection concentrations of glyceric acid and D-turanose in male C57-AMPK gene knocked-out mice are about double higher than these in female C57-AMPK gene knocked-out mice. According to these results, we could say that distinctions related to gender that existed in this urinary data were reflected and differences of metabolites between male and female C57-AMPK gene knocked-out mice and C57 healthy mice were uncovered.

From all the discussion mentioned above, we could conclude that defects of the AMPK-signaling system result in many of the same metabolic perturbations found in type 2 diabetes mellitus. Some potential biomarkers have been found to have deeply contribution to discriminate the AMPK defects mice from normal ones. Further researches should be done to find out their metabolic pathways and interaction with protein or other small molecules. Because of activation of this control system either by

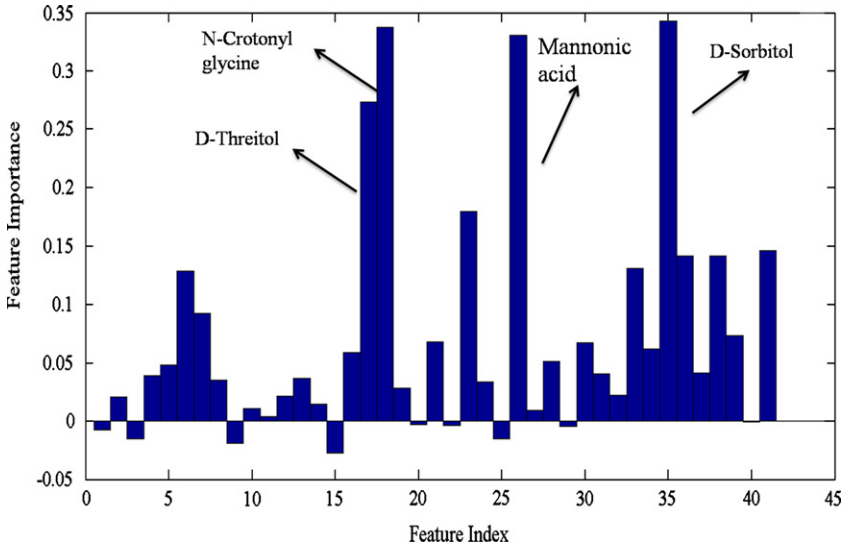

Fig. 5. The variable importance measure obtained by random forest between normal female mice and female C57-AMPK gene knocked-out mice.

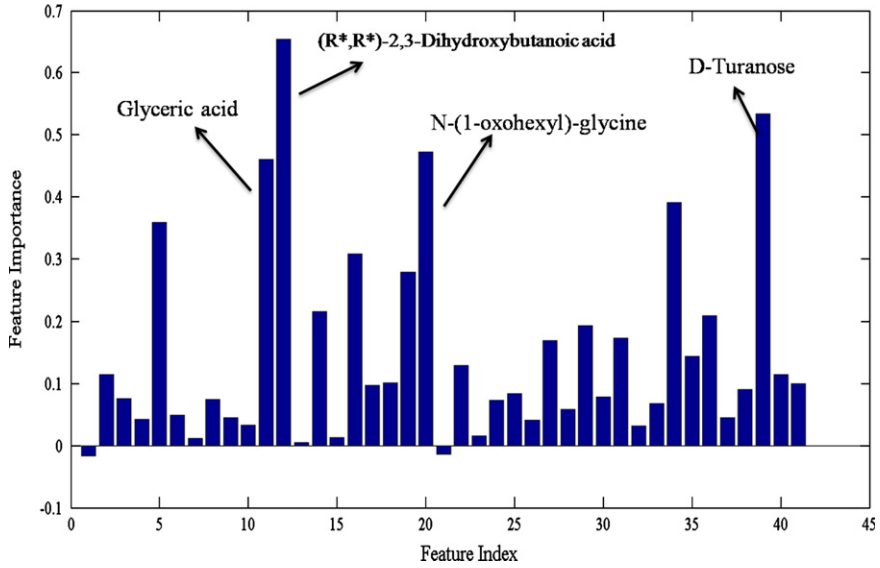

Fig. 6. The variable importance measure obtained by random forest between female and male C57-AMPK gene knocked-out mice.

exercise or pharmacological manipulation may partially correct the metabolic perturbations of the forms of type 2 diabetes resulting from defects in the insulin signaling cascade. It will be very helpful for pathogenesis research of T2DM and designing new drugs for treatment of T2DM.

4. Conclusions

This work provided that it was an efficient strategy to use GC/MS coupled with random forest to analyze metabolic fingerprints of the C57-AMPK gene knocked-out mice urine and normal C57 mice. Changes between the AMPK gene knocked-out mice and normal C57 mice on the urine metabolic profiles were revealed. Gender related differences of C57-AMPK gene knocked-out mice can be also discriminated by statistic analysis. Furthermore, the 'biomarkers' that have been screened out are expected to be helpful to evaluate the function of AMPK, and provide useful information for further investigation and study of pathogenesis of type II diabetes mellitus.

Acknowledgements

This work has been financially supported by the "National Novel Drug Development Foundation 2009ZX09303-008" and Experimental Zooprophylactic Institute of Chinese Academy of Medical Sciences. National Nature Foundation Committee of PR China (grants no. 20875104. no. 21105129.) Natural Science Foundation of Chongqing (cstc2011jjA0780). The studies meet with the approval of the university's review board.

References

[1] S.I. Taylor, Cell 97 (1999) 9–12.
[2] M.M. Engelgau, K.M.V. Narayan, W.H. Herman, Diabetes Care 23 (2000) 1563–1580.
[3] K.R. Kelly, M.J. Abbott, L.P. Turcotte, Exp. Biol. Med. 235 (2010) 514–521.
[4] M.C. Towler, D.G. Hardie, Circ. Res. 100 (2007) 328–341.
[5] W.W. Winder, D.G. Hardie, Am. J. Physiol-Endoc. M. 277 (1999) E1–E10.
[6] B.B. Zhang, G. Zhou, C. Li, Cell. Metab. 9 (2009) 407–416.
[7] D.G. Hardie, D. Carling, Eur. J. Biochem. 246 (1997) 259–273.
[8] Z.J. Luo, A.K. Saha, X.Q. Xiang, N.B. Ruderman, Trends. Pharmacol. Sci. 26 (2005) 69–76.
[9] Z.j. Luo, M. Zang, W. Guo, Future. Oncol. 6 (2010) 457–470.
[10] M. Carmen Solaz-Fuster, J. Vicente Gimeno-Alcaniz, S. Ros, M. Elena Fernandez-Sanchez, B. Garcia-Fojeda, O. Criado Garcia, D. Vilchez, J. Dominguez, M. Garcia-Rocha, M. Sanchez-Piris, C. Aguado, E. Knecht, J. Serratosa, J. Josep Guinovart, P. Sanz, S. Rodriguez de Cordoba, Hum. Mol. Genet. 17 (2008) 667–678.
[11] P. Gual, Y. Le Marchand-Brustel, J.F. Tanti, Biochimie 87 (2005) 99–109.
[12] M. Ponticos, Q.L. Lu, J.E. Morgan, D.G. Hardie, T.A. Partridge, D. Carling, EMBO J. 17 (1998) 1688–1699.
[13] C. Denkert, J. Budczies, T. Kind, W. Weichert, P. Tablack, J. Sehouli, S. Niesporek, D. Koensgen, M. Dietel, O. Fiehn, Cancer. Res. 66 (2006) 10795–10804.
[14] U. Lutz, R.W. Lutz, W.K. Lutz, Anal. Chem. 78 (2006) 4564–4571.
[15] S.J. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochhar, P.A. Guy, Anal. Chem. 81 (2009) 3285–3296.
[16] M. Oldiges, S. Luetz, S. Pflug, K. Schroer, N. Stein, C. Wiendahl, Appl. Microbiol. Biotechnol. 76 (2007) 495–511.
[17] S.P. Sawant, A.V. Dnyanmote, M.S. Mitra, J. Chilakapati, A. Warbritton, J. R. Latendresse, H.M. Mehendale, J. Pharmacol. Exp. Ther. 316 (2006) 507–519.
[18] J.H. Granger, R. Williams, E.M. Lenz, R.S. Plumb, C.L. Stumpf, I.D. Wilson, Rapid Commun. Mass Spectrom. 21 (2007) 2039–2045.
[19] Q. Zhang, G.J. Wang, Y. Du, L.L. Zhu, A. Jiye, J. Chromatogr. B 854 (2007) 20–25.
[20] K.K. Pasikanti, P.C. Ho, E.C.Y. Chan, J. Chromatogr. B 871 (2008) 202–211.
[21] H.J. Major, R. Williams, A.J. Wilson, I.D. Wilson, Rapid Commun. Mass Spectrom. 20 (2006) 3295–3302.
[22] J.C. Lindon, J.K. Nicholson, J.R. Everett, Annu. Rep. Nmr. Spectro. 38 (1999) 1–88.
[23] M.E. Bollard, E.G. Stanley, J.C. Lindon, J.K. Nicholson, E. Holmes, NMR Biomed. 18 (2005) 143–162.
[24] S. Kochhar, D.M. Jacobs, Z. Ramadan, F. Berruex, A. Fuerhoz, L.B. Fay, Anal. Biochem. 352 (2006) 274–281.
[25] E.G. Stanley, N.J.C. Bailey, M.E. Bollard, J.N. Haselden, C.J. Waterfield, E. Holmes, J.K. Nicholson, Anal. Biochem. 343 (2005) 195–202.
[26] R.S. Plumb, K.A. Johnson, P. Rainville, J.P. Shockcor, R. Williams, J.H. Granger, I. D. Wilson, Rapid Commun. Mass Spectrom. 20 (2006) 2800–2806.
[27] P. Jonsson, A.I. Johansson, J. Gullberg, J. Trygg, J.A, B. Grung, S. Marklund, M. Sjostrom, H. Antti, T. Moritz, Anal. Chem. 77 (2005) 5635–5642.
[28] L. Zhang, H. Dai, Y. Liang, O.M. Kvalheim, J. Chromatogr. A 1217 (2010) 3128–3135.
[29] C.X. Zhao, Y.Z. Liang, H.Z. Fang, X.N. Li, J. Chromatogr. A 1096 (2005) 76–85.
[30] B.B. Tan, Y.Z. Liang, L.Z. Yi, H.D. Li, Z.G. Zhou, X.Y. Ji, J.H. Deng, Metabolomics 6 (2010) 219–228.
[31] L. Breiman, Mach. Learn. 45 (2001) 5–32.
[32] W.S. Torgerson, Psychometrika 17 (1952) 401–419.
[33] O. Ballevre, J. Prugnaud, M.L. Houlier, M. Arnal, Anal. Biochem. 193 (1991) 212–219.
[34] J.A. Thompson, S.P. Markey, P.V. Fennessey, Clin. Chem. 21 (1975) 1892–1898.
[35] E. Appiah-Amponsah, N. Shanaiah, G.A. Nagana Gowda, K. Owusu-Sarfo, T. Ye, D. Raftery, J. Pharm. Biomed. Anal. 50 (2009) 878–885.
[36] E. Pitkanen, Clin. Chim. Acta 80 (1977) 49–54.
[37] A. Gruzman, O. Shamni, M. Ben Yakir, D. Sandovski, A. Elgart, E. Alpert, G. Cohen, A. Hoffman, Y. Katzhendler, E. Cerasi, S. Sasson, J. Med. Chem. 51 (2008) 8096–8108.
[38] T.S. Dhanoa, J.A. Housner, Curr. Sports. Med. Rep. 6 (2007) 254–257.
[39] G.C. Hawthorne, K. Bartlett, C.S. Hetherington, K.G. Alberti, Diabetologia 32 (1989) 163–166.
[40] M.J. Crabbe, D. Goode, Prog. Retin. Eye. Res. 17 (1998) 313–383.
[41] S. Miyamoto, Expert. Opin. Ther. Pat. 12 (2002) 621–631.
[42] D.A. Greene, S.A. Lattimer, A.A. Sima, New. Engl. J. Med. 316 (1987) 599–606.