RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Distinguish the serum metabolite profiles differences in breast cancer by gas chromatography mass spectrometry and random forest method

Jian-Hua Huang[a], Liang Fu[b], Bin Li[a], Hua-Lin Xie[b*], Xiaojuan Zhang[a], Yanjiao Chen[a], Yuhui Qin[a], Yuhong Wang[a], Shuihan Zhang[a], Huiyong Huang[a], Duanfang Liao[a], Wei Wang[a*]

a. TCM and Ethnomedicine Innovation & Development Laboratory, Sino-Luxemburg TCM Research Center, School of Pharmacy, Hunan University of Chinese Medicine, Changsha, China, 410208, P. R. China;

b. College of Chemistry and Chemical Engineering, Yangtze Normal University, Chongqing, 408100, China

[*] Correspondence to: hualinxie@163.com (H.L. Xie.); wangwei402@hotmail.com (Wei Wang)
Prof. Hua-lin Xie
Department of Chemistry and Chemical Engineering
Yangtze Normal University
Chongqing, 408100, P.R. China
E-mail: hualinxie@163.com
Prof. Wei Wang.
School of pharmacy,
Hunan University of Chinese Medicine;
Changsha 410208, P.R. China;
E-mail address: wangwei402@hotmail.com
Tel: +86731-8845 8240
Fax: +86731-8845 8227

Abbreviations: Random forests: RF; Gas chromatography-mass spectrometry: GC-MS; Breast cancer: BC; principal component analysis: PCA; partial least squares discriminant analysis: PLS-DA.

Abstract:

In this study, we proposed a Metabolomics strategy to distinguish different metabolic characters of healthy controls, breast benign (BE) patients, and breast malignant (BC) patients by using the GC-MS and random forest method (RF). In current study, the serum samples from healthy controls, BE patients, and BC patients were characterized by using GC-MS. Then, random forest (RF) models were established to visually discriminate the differences among three groups' metabolites profiles, and further investigate the progress of breast cancer from benign to malignant in patients based on these GC-MS profiles. We successfully discovered the differences between the healthy and breast cancer patients. And the metabolic changes from benign to malignant cancer were obviously visualized. The results suggested that combining GC-MS profiling with random forest method is a useful approach to analyze metabolites and to screen the potential biomarkers for exploring the serum metabolic profiles of breast cancer.

Key words: Breast cancers, Serum metabolite profiles, GC-MS, Random forest;

Introduction:

Breast cancer is the most prevalent malignant disease in women worldwide and a major cause of female deaths[1]. In clinical research, breast cancer is a prognoses disease in which larger tumor size and the presence of lymph node metastasis are associated with worse prognoses[2]. The earlier diagnosis is one of the most important strategies to reduce breast cancer morbidity rate and improve the survival rate[2, 3]. In terms of diagnosis of breast cancer, some current methods can accurately differentiate malignant from normal and benign tissue by identification of malignant tissue characteristic[4, 5]. Routine breast cancer inspection methods including periodic mammography, physical examination, and blood tests[6]. Mammography always misses small tumor that will lead to false positives, resulting in suboptimal sensitivity and specificity and unnecessary biopsies. Furthermore, these conventional BC diagnostics techniques are always expensive and time-consuming; furthermore some patients may feel discomfort during diagnosis process.

Metabolomics is an important platform for quantitative analysis of the metabolites in living systems and their dynamic responses to the changes of both endogenous and exogenous factors by using all kinds of analytical approaches, including Gas chromatography-mass spectrometry GC-MS[7-10], high-resolution nuclear magnetic (NMR)[11-14], ultra-performance liquid chromatography-mass spectrometry (UPLC-MS)[15]. Recently, the Metabolomics methods were widely used to monitor disease progression, and showed its advantages in various researches, such as diagnosis of human diseases[16], physiological evaluations[17], elucidation of

biomarkers[18, 19], and drug toxicity[20]. The transforming process from normal to malignant cells is always associated with some metabolic disturbances. Therefore, using the metabolomics method for breast cancer research is very suitable. Some previous researches have demonstrated that some volatile organic metabolites could indicate the differences between breast cancer patients and healthy controls[6, 21], and some other researchers have reported the serum concentrations of free fatty acids (FFAs) in patients with BC were significantly decreased compared with those in healthy controls[22, 23]. These researches indicated that using GC-MS metabolites profiles can help breast cancer diagnosis. Besides, GC-MS analysis method has some advantages such as, favorable stability, reproducibility, and sensitivity, and rapid analysis.

Owning to the complexity of these metabolic profiles, multivariate statistical methods are extensively used to deal with these 'Omics' data. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were the most often used method to visually represent the data information, and some other machine learning methods were applied in the researches more and more frequent [24, 25]. Random forest (RF) model, one of these machine learning methods, has its own characteristic advantages on dealing with complex metabolomics data. This algorithm has showed its advantages in dealing with these complex metabolomics data, not only distinguish different groups (patients and healthy), but also can help finding the significant changes of metabolites as a potential biomarker, as showed in our previous researches[26-28].

Based on these reasons, we established a metabolomics strategy to distinguish different metabolic characters of healthy controls, breast benign patients, and breast malignant patients by using the GC-MS and random forest (RF). The whole experiment contains several steps: Firstly, the serum samples from healthy, breast benign patients, and breast malignant patients were profiled by using GC-MS analytical technique; after being pretreated, metabolites information was processed by using RF method; Finally, RF model can calculate the sample proximity matrix, by using this sample proximity matrix, not only the differences between the healthy and breast cancer patients were observed, but also the differences between breast benign patients and breast malignant patients were obviously visualized. And some informative metabolites or potential biomarkers have been successfully discovered by means of variable importance ranking in random forest program.

2. Materials and Methods

2.1 Samples collection

23 breast benign patients and 30 breast malignant patients were collected from The Tumor Hospital of ChangDe City, Hunan Province. These patients were treated in this hospital, and were diagnosed by the standard methodologies [29]. 30 healthy controls (who were negative for breast cancer by mammography and ultrasound examination) were selected from volunteers. Total of 83 samples (30 healthy, 23 breast benign patients, and 30 breast malignant patients) were tested in current study. The protocol in this study was approved by the Ethics Committee at The Tumor Hospital of ChangDe City.

2.2 Serum collection and preparation

2 mL of venous blood samples were collected in a blank tube from each individual at 8 o'clock in the morning after overnight fasting. After obtained the serum samples, the samples were kept in -80 ℃ until analysis. Serum was thawed at 4 ℃ for 30 minutes. To the 100 μL serum samples, 350 μL methanol (including 1 mg/mL of heptadecanoic acid/methanol as internal standard) was added and vortexed for 15 s, and centrifuged for 15 min (15,000 r/min, 4 ℃). Supernatant was dried by using N_2. Then, the dried supernatant was derivatized by adding Methoxylamine/ pyridine (20 mg/mL) mixed for 15 s, and incubated for 1h (65 ℃), followed by addition of 100 μL BSTFA). All the samples were analyzed by using GC-MS at random order after being prepared by described procedure.

2.3 Equipment and reagents

Data was acquired by GC-MS using Agilent 7890A gas chromatography instrument coupled to a 5975C mass spectrometer (Agilent, Santa Clara, California, USA). Methanol (CH_3OH) was purchased from Tedia Company (Fairfield, USA). Analytical grade heptadecanoic acid (C17:0), methoxamine, pyridine and Bis(trimethylsilyl)- trifluoroacetamide (BSTFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.4 Gas chromatography–mass spectrometry conditions

GC separation was performed on Agilent DB-5MS equipped with a deactivated fused silica capillary column (0.25mm×30m×0.25μm). The oven temperature was maintained at 70 ℃ for 4 min, programmed to 300 ℃ (rate of 8 ℃/min), and then

held for 3 min. The injection volume of 1μL was used in the split ratio of 1:10. Helium was used as the carrier gas (flow rate of 1.0 mL/min). The mass spectrometery was performed using electron impact (EI) ionization source at 70 eV and a 0.90 kV detector voltage in 0.2 s/scan full scan. The mass spectrometer was operated with m/z range from 35 to 650. These analytical conditions were consistent to our previous researches [26].

2.5 Principal component analysis (PCA)

Principal component analysis (PCA) was used in current study to exhibit the cluster trend of three groups' samples. The singular value decomposition (SVD) was used to transform raw variables into a set of linearly orthogonal project variables. These project variables contain the almost useful information in the raw signals. The noise signals contain in the raw signals can be eliminated during such decomposing process. We could obtain the scores and loading by using the SVD. The scores plot can be used to present the relationships among different samples. The loading values for each variable can be used to select the informative variables. The PCA program used in this study was written by MATLAB in our group.

2.6 Random Forest

Random forest model was established by assembling enough classification and regression trees[30, 31]. The mains implements of RF are based on bagging and random feature selection strategy. The bagging method can ensemble enough tree model in the training process, and two types of datasets are established, the train dataset and "out of bag" dataset. The "out of bag" data also called OOB samples, which can be used to

estimate the model precision. This OOB estimation has been proved to be unbiased. In each tree growing process, instead of using all the features to split at each tree node, RF selects only a small subset of features, which makes each tree in the forest is different from each other. Increasing the diversities of trees is an efficient way to increase the classifying and recognition ability of RF method. The detailed RF modeling process can be found in our previous studies [26].

Here, two useful tools, the variable importance measure and proximity matrix, in the RF will be introduced, which have showed their advantages in the data interpretation and visualization. The variable importance measure can be used to estimate the importance of each metabolite in the model classification. This information can help us to find the potential biomarkers. In current study, 'the mean decrease in classification' measure was adopted. For each tree, the classification accuracy of the OOB samples is determined both with and without random permutation of the variable values one by one. The accuracy of permutation is subtracted from that before permutation, and then averaged over all trees in the forest (Calculated as Eq.1).

$$\text{Importance of j} = \text{Accuracy}_{\text{j normal}} - \text{Accuracy}_{\text{j permuted}} \qquad (1)$$

The other attractive feature in RF algorithm is the proximity matrix calculation. Proximity values can indicate the similarities among all the samples. In normal situation, samples from the same group always fall into the same or nearby tree node. (This is the principle of tree method). In tree method, distance matrix was used to calculate the similarities of samples. In RF method, the proximity between two

samples was calculated as the number of times the two samples fall into the same terminal node of a tree, and then divided by the number of trees in the forest [31]. After the proximity values are calculated, multi-dimensional scaling (MDS) plot is always used to visualize these analysis results. MDS is a set of related statistical techniques often used to visually explore similarities or dissimilarities in data[32]. We can project the first two or three scaling coordinates into low dimensions and obtain the clustering plot of all the samples.

3 Results and Discussion

3.1 Data analysis

The typical total ion chromatograms (TICs) of serum metabolic profiles for healthy control (in blue line), BE (in black line), and BC (in red line) were shown in Fig. 1. As could be seen in Fig.1, the serum metabolites profiles of three groups were similar, but the concentrations of some metabolites were differences. These results suggested that these GC-MS profiles could represent the differences among three groups.

Insert of Figure 1

After these metabolic profiles were collected, qualitative and the quantitative work were carried out, mainly metabolites, including amino acid, organic acid, fatty acid, and carbohydrates were found in the chromatograms (Detailed results were listed in Table 1). Then, these metabolites data were input to some pattern recognition algorithms for further analysis.

Insert of Table 1

Firstly, we used the principal component analysis (PCA) to present the cluster trends of these three groups' samples. PCA can project the metabolites profiles into a lower dimensional space to visually evaluate clustering trends. The first three principal components, i.e., PC1, PC2 and PC3, were used to draw the scores plot (Fig.2) which can present the samples distribution of three groups. The total contribution of these three PCs accumulated to 94.59% in the total variance of the raw data. As visually observed, the healthy controls are significant different with the BE and BC groups. But the differences of BE and BC group cannot be discriminated, some samples from two groups are overlapped.

Insert of Figure 2

Therefore, in order to further classify the BE and BC patients, random forest (RF) method was adopted to analyze these metabolites; all the metabolites were used as variables for discrimination. According to the pre-set parameters, RF models were established. During the model training process, the samples proximities are calculated for each pair of cases. As similar samples always fall into the same terminal node or derive from the same parent node. Thus, the samples in the same group always have a larger similarity value than that in other group samples.

Insert Figure 3

To more directly and conveniently observe the patterns in the proximity matrix, multidimensional scaling (MDS) was employed to map the proximity into a lower-dimensional space. From Fig.3, a good separation between the healthy controls

and breast cancer patients could be observed. Furthermore, the differences between BC and BE patients were also emerged. These results sufficiently indicated that the metabolic characters among BC patients, benign patients (BE), and healthy control are distinction. The BE patients were located in the middle of BC patients and healthy controls, and they may develop and progress to malignant tissues. More detailed analysis for each pairs of group has been done in the following sections.

3.2 Biomarkers screening between Healthy and BE

In metabolomics analysis process, the general aim is to find the best combination of metabolites which can help explain the relevant metabolic pathway. All the 41 compounds in healthy controls and BE patients were used as variables for discrimination analysis. The feature importance for each variable was showed in the Fig. 4. The prediction accuracy, sensitivity, and specificity for current method were 95.65%, 100.0%, and 96.25%, respectively.

Insert Figure 4

Some of metabolites, such as Acetic acid, (R*,R*)-2,3-Dihydroxybutanoic acid, Palmitic acid, and D-(+)-lactose monohydrate, have great contributions to classification accuracy. (R*,R*)-2,3-Dihydroxybutanoic acid is a normal organic acid in human biofluids. Palmitic acid is a saturated fatty acid, may inhibit the metabolic actions of insulin and attenuate insulin signal transduction[33]. Moreover, there is a significant direct association between palmitic acid in erythrocyte and risk of breast cancer[34]. These metabolites could be considered as potential biomarkers for diagnosing the breast benign patients.

3.3 Biomarkers screening between Healthy and BC

We further analyze the differences in metabolites between healthy controls and BC patients. The prediction accuracy, sensitivity, and specificity for healthy controls and BC patients were 100.0%, 96.67%, and 98.33%, respectively. And the feature important for each variable was showed in the Fig. 5.

Insert Figure 5

As could be seen from Fig.5, several metabolites were consistent with these in BE and healthy controls, such as, (R*,R*)-2,3-Dihydroxybutanoic acid and D-(+)-lactose monohydrate. Other metabolites such as D-Xylose and Galactonic acid were also found larger contribution for the classification. A property of many malignancies, including breast cancer, is constitutive upregulation of glycolysis with persistent glycolysis despite the present of oxygen [35]. These metabolites represented with some of changes in metabolic activity of several pathways associated with breast cancer, including amino acid metabolism, glycolysis metabolism. Galactonic acid, is a sugar acid [35] and one of the oxidized form of D-galactose. D-Xylose is a five-carbon aldose that can be catabolized or metabolized into useful product by lots of organisms [36, 37]. These means these metabolites could be considered as potential biomarkers for diagnosing the breast malignant patients.

3.4 Biomarkers screening between BE and BC

This section aims to investigate the metabolic differences between BE and BC patients. This could be seen from Fig.3, a good separation between BE and BC patients was obtained by using random forest method. In order to evaluate the

predictive ability of the proposed method, RF has been employed to classify BE and BC patients. The prediction accuracy, sensitivity, and specificity for current method were 93.33%, 86.96%, and 90.57%, respectively. The feature important for each variable was showed in the Fig. 6.

Insert Figure 6

As could be seen from Fig.6, three metabolites could be found as the potential biomarkers D-Glucose, D-(+)-lactose monohydrate, and D-Xylose. Furthermore, the D-Xylose is a special metabolite for BC patients, which is different with BE patients and healthy controls. This might be a useful potential biomarker for monitoring the transforming process and metabolic disturbances from benign to malignant cancer. These molecular biomarkers generally can provide prognostic symbols and their diagnostic detection is becoming increasingly important in early diagnosis of breast cancer.

4. Conclusion

The aim of our study was to comprehensively investigate the metabolic profiling changes of healthy control, breast benign patients, and breast malignant patients. The results provided that it was an efficient strategy to use GC-MS coupled with random forest to analyze metabolic fingerprints of the three groups. Changes between the healthy and breast cancer patients' metabolic profiles were revealed. Different metabolites of benign and malignant cancer can be also discriminated by RF analysis. What is more, rapid and reliable determination of malignancy, benign cancers could aid the current clinical approach.

Acknowledgements

This work was supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1401209), and supported by program for the Changsha Science & Technology Bureau (K1205019-31) and Graduate Student Innovation Project of Hunan Province (CX2014B369).

Reference:

1. J. R. Benson, I. Jatoi, M. Keisch, F. J. Esteva, A. Makris and V. C. Jordan, *The Lancet*, 2009, **373**, 1463-1479.
2. C. Oakman, L. Tenori, L. Biganzoli, L. Santarpia, S. Cappadona, C. Luchinat and A. Di Leo, *The International Journal of Biochemistry & Cell Biology*, 2011, **43**, 1010-1020.
3. V. M. Asiago, L. Z. Alvarado, N. Shanaiah, G. N. Gowda, K. Owusu-Sarfo, R. A. Ballas and D. Raftery, *Cancer. Res.*, 2010, **70**, 8309-8318.
4. C. Yang, A. D. Richardson, J. W. Smith and A. Osterman, 2007.
5. W. Lv and T. Yang, *Clin. Biochem.*, 2012, **45**, 127-133.
6. C. Wang, B. Sun, L. Guo, X. Wang, C. Ke, S. Liu, W. Zhao, S. Luo, Z. Guo and Y. Zhang, *Scientific reports*, 2014, **4**.
7. J. H. Granger, R. Williams, E. M. Lenz, R. S. Plumb, C. L. Stumpf and I. D. Wilson, *Rapid Communications in Mass Spectrometry*, 2007, **21**, 2039-2045.
8. Q. Zhang, G. J. Wang, Y. Du, L. L. Zhu and A. Jiye, *J. Chromatogr. B*, 2007, **854**, 20-25.
9. K. K. Pasikanti, P. C. Ho and E. C. Y. Chan, *J. Chromatogr. B*, 2008, **871**, 202-211.
10. H. J. Major, R. Williams, A. J. Wilson and I. D. Wilson, *Rapid Communications in Mass Spectrometry*, 2006, **20**, 3295-3302.
11. J. C. Lindon, J. K. Nicholson and J. R. Everett, *Annual Reports on Nmr Spectroscopy, Vol 38*, 1999, **38**, 1-88.
12. M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson and E. Holmes, *NMR Biomed.*, 2005, **18**, 143-162.
13. S. Kochhar, D. M. Jacobs, Z. Ramadan, F. Berruex, A. Fuerhoz and L. B. Fay, *Anal. Biochem.*, 2006, **352**, 274-281.

14. E. G. Stanley, N. J. C. Bailey, M. E. Bollard, J. N. Haselden, C. J. Waterfield, E. Holmes and J. K. Nicholson, *Anal. Biochem.*, 2005, **343**, 195-202.
15. R. S. Plumb, K. A. Johnson, P. Rainville, J. P. Shockcor, R. Williams, J. H. Granger and I. D. Wilson, *Rapid Communications in Mass Spectrometry*, 2006, **20**, 2800-2806.
16. C. Denkert, J. Budczies, T. Kind, W. Weichert, P. Tablack, J. Sehouli, S. Niesporek, D. Koensgen, M. Dietel and O. Fiehn, *Cancer. Res.*, 2006, **66**, 10795-10804.
17. U. Lutz, R. W. Lutz and W. K. Lutz, *Anal. Chem.*, 2006, **78**, 4564-4571.
18. S. J. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochhar and P. A. Guy, *Anal. Chem.*, 2009, **81**, 3285-3296.
19. M. Oldiges, S. Luetz, S. Pflug, K. Schroer, N. Stein and C. Wiendahl, *Appl. Microbiol. Biotechnol.*, 2007, **76**, 495-511.
20. S. P. Sawant, A. V. Dnyanmote, M. S. Mitra, J. Chilakapati, A. Warbritton, J. R. Latendresse and H. M. Mehendale, *J. Pharmacol. Exp. Ther.*, 2006, **316**, 507-519.
21. M. Phillips, J. D. Beatty, R. N. Cataneo, J. Huston, P. D. Kaplan, R. I. Lalisang, P. Lambin, M. B. Lobbes, M. Mundada and N. Pappas, *Plos One*, 2014, **9**, e90226.
22. J. Li, Y. Peng and Y. Duan, *Critical reviews in oncology/hematology*, 2013, **87**, 28-40.
23. Y. Zhang, L. Song, N. Liu, C. He and Z. Li, *Clin. Chim. Acta*, 2014, **437**, 31-37.
24. S. C. Kalhan, L. Guo, J. Edmison, S. Dasarathy, A. J. McCullough, R. W. Hanson and M. Milburn, *Metabolism*, 2011, **60**, 404-413.
25. K. Bryan, L. Brennan and P. Cunningham, *BMC Bioinformatics*, 2008, **9**, 470.
26. Z. Lin, C. M. Vicente Gonçalves, L. Dai, H.-m. Lu, J.-h. Huang, H. Ji, D.-s. Wang, L.-z. Yi and Y.-z. Liang, *Anal. Chim. Acta*, 2014, **827**, 22-27.
27. L. Dai, C. M. V. Gonçalves, Z. Lin, J. Huang, H. Lu, L. Yi, Y. Liang, D. Wang and D. An, *Talanta*, 2015, **135**, 108-114.
28. J.-H. Huang, R.-H. He, L.-Z. Yi, H.-L. Xie, D.-s. Cao and Y.-Z. Liang, *Talanta*, 2013, **110**, 1-7.
29. S. E. Singletary, C. Allred, P. Ashley, L. W. Bassett, D. Berry, K. I. Bland, P. I. Borgen, G. Clark, S. B. Edge and D. F. Hayes, *Journal of clinical oncology*, 2002, **20**, 3628-3636.
30. L. Breiman, *Mach. Learn.*, 2001, **45**, 5-32.
31. J.-H. Huang, J. Yan, Q.-H. Wu, M. Duarte Ferro, L.-Z. Yi, H.-M. Lu, Q.-S. Xu and Y.-Z. Liang, *Talanta*, 2013, **117**, 549-555.
32. H. Klock and J. M. Buhmann, *Pattern Recognit.*, 2000, **33**, 651-669.
33. M. W. Ruddock, A. Stein, E. Landaker, J. Park, R. C. Cooksey, D. McClain and M.-E. Patti, *J. Biochem.*, 2008, **144**, 599-607.
34. J. Shannon, I. B. King, R. Moshofsky, J. W. Lampe, D. L. Gao, R. M. Ray and D. B. Thomas, *The American journal of clinical nutrition*, 2007, **85**, 1090-1097.
35. Y. S. Chan and T. B. Ng, *Plos One*, 2013, **8**, e54212.
36. Y.-K. Qiu, D.-Q. Dou, L.-P. Cai, H.-P. Jiang, T.-G. Kang, B.-Y. Yang, H.-X. Kuang and M. Z. Li, *Fitoterapia*, 2009, **80**, 219-222.
37. G. Parrilli, R. V. Iaffaioli, M. Martorano, R. Cuomo, S. Tafuto, M. G. Zampino, G. Budillon and A. R. Bianco, *Cancer. Res.*, 1989, **49**, 3689-3691.

Table List:

Table 1 The metabolites quantitative results for healthy controls, breast benign (BE) patients, and breast malignant (BC) patients.

Figure Captions:

Figure 1 The typical total ion chromatograms (TICs) of healthy control (in blue line), BE (in black line), and BC (in red line).

Figure 2 The first three principal components from PCA scores plot of serum profiles for healthy, BE and BC samples.

Figure 3 The MDS plot for serum profiles for healthy, BE and BC samples.

Figure 4 The variable importance measures of healthy controls and BE samples obtained by RF models.

Figure 5 The variable importance measures of healthy controls and BC samples obtained by RF models.

Figure 6 The variable importance measures of BC and BE samples obtained by RF models.

Table 1. Qualitative and quantitative metabolic profiles of three groups' samples.

id	t_r[a] (min)	endogenous metabolites	BC group	Be group	Healthy
1	5.922	Ethylbis(trimethylsilyl)amine	0.2456±0.0705	0.1905±0.0567	0.1958±0.0551
2	6.593	Ethylene glycol	0.0182±0.0020	0.0530±0.0428	0.0746±0.0626
3	6.608	N,N-diethylacetamide	0.0120±0.0060	0.0220±0.0325	0.0227±0.0302
4	6.84	N,N-diethyl-Acetamide	0.0657±0.0087	0.0476±0.0202	0.0557±0.0107
5	7.716	Lactic acid *	0.0872±0.0374	0.0952±0.0592	0.1482±0.2155
6	7.934	Acetic acid	0.0629±0.0140	0.0856±0.0333	0.0412±0.0403
7	10.01	phosphate	3.1278±1.0173	1.4730±0.7381	1.3767±1.0361
8	10.2	l-Threonine	0.0173±0.0098	0.0108±0.0068	0.0096±0.0065
9	10.297	Acetic acid, phenyl-	0.0047±0.0023	0.0159±0.0133	0.0147±0.0117
10	10.382	Succinic acid *	0.0811±0.0429	0.0098±0.0131	0.0119±0.0086
11	10.447	[1,2-phenylenebis(oxy)]bis[trimethyl-	0.0120±0.0072	0.0078±0.0047	0.0067±0.0039
12	10.503	Glyceric acid	0.0961±0.0266	0.0400±0.0282	0.0183±0.0167
13	10.723	(R*,R*)-2,3-Dihydroxybutanoic acid	0.0267±0.0053	0.0137±0.0014	0.0053±0.0029
14	11.357	2,4-bis[(trimethylsilyl)oxy]-Butanoic acid	0.0147±0.0051	0.0055±0.0030	0.0066±0.0047
15	11.583	(R*,S*)-3,4-Dihydroxybutanoic acid	0.0304±0.0098	0.0132±0.0064	0.0178±0.0107
16	11.797	N-(1-oxobutyl)- Glycine	0.0653±0.0244	0.0319±0.0186	0.0274±0.0151
17	12.341	Isovaleroglycine	0.0356±0.0134	0.0160±0.0079	0.0107±0.0073
18	12.483	D- Threitol	0.0214±0.0073	0.0290±0.0130	0.0251±0.0151
19	12.645	N-Crotonylglycine	0.0640±0.0146	0.0207±0.0129	0.0148±0.0099
20	14.53	N-(1-oxohexyl)-glycine	0.0160±0.0072	0.0121±0.0073	0.0132±0.0081
21	14.713	d-Xylose	0.0208±0.0075	0.0082±0.0044	0.0093±0.0063
22	14.823, 15.057	d-Ribose	0.0126±0.0070	0.0152±0.0042	0.0250±0.0110
23	15.509, 15.733	Arabitol	0.0487±0.0364	0.0283±0.0179	0.0278±0.0215
24	16.023	D-Galactose, 6-deoxy-2,3,4,5-tetrakis-O-(trimethylsilyl)-	0.0336±0.0083	0.0177±0.0100	0.0149±0.0104
25	16.087	Mannonic acid	0.0505±0.0177	0.0211±0.0143	0.0168±0.0138
26	16.2	cis-Aconitic acid*	0.0435±0.0388	0.0105±0.0079	0.0168±0.0147
27	16.357	Phosphoric acid	0.0414±0.0252	0.0230±0.0141	0.0212±0.0168
28	17.177	Isocitric acid*	0.0464±0.0121	0.0340±0.0093	0.0448±0.0838
29	17.563	Hippuric acid	0.0270±0.0126	0.0180±0.0104	0.0156±0.0116
30	17.85,	D-Fructose*	0.0712±0.0586	0.0471±0.0145	0.0580±0.1031

	17.96				
31	18.087	d-Galactose*	0.0796±0.0214	0.0455±0.0272	0.0389±0.0287
32	18.197, 18.147	d-Glucose*	0.2785±0.0918	0.1741±0.7354	0.1859±0.4136
33	18.507	Altronic acid	0.0202±0.0069	0.0185±0.0100	0.0102±0.0074
34	18.577, 18.65	D-Sorbitol*	0.0259±0.0169	0.0254±0.0187	0.0300±0.0275
35	18.983, 19.533	Galactonic acid	0.1213±0.0482	0.0817±0.0328	0.0441±0.0351
36	19.99	Palmitic acid	0.0127±0.0017	0.0148±0.0029	0.0071±0.0025
37	20.403	Myo-Inositol	0.0247±0.0128	0.0197±0.0037	0.0334±0.0129
38	25.465	D-Turanose	0.0216±0.0138	0.0197±0.0190	0.0510±0.1099
39	28.125	D-（+）-lactose monohydrate*	0.8475±0.1366	1.0400±0.3349	0.6559±0.2286
40	29.927	Lactose	0.0142±0.0043	0.0143±0.0075	0.0190±0.0163
41	35.223	Cholesterol*	0.0107±0.0038	0.0101±0.0021	0.0107±0.0034

*: Identified by standard substances

[a]: Retention time;

Figure 1:

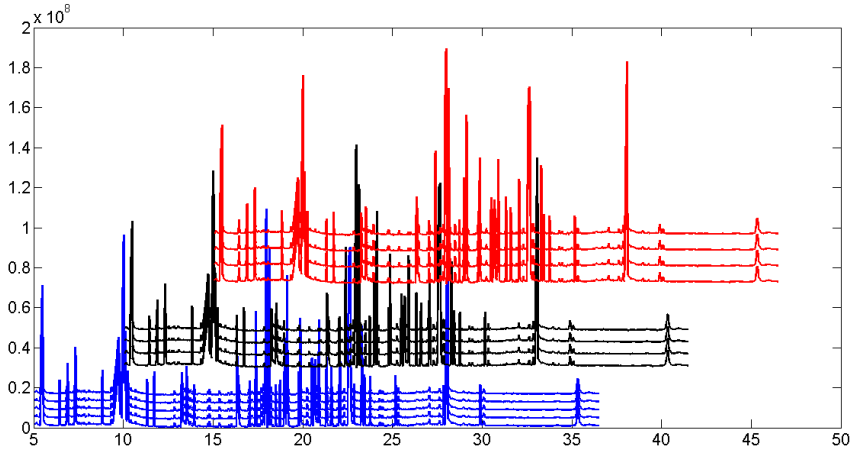

Figure 2:

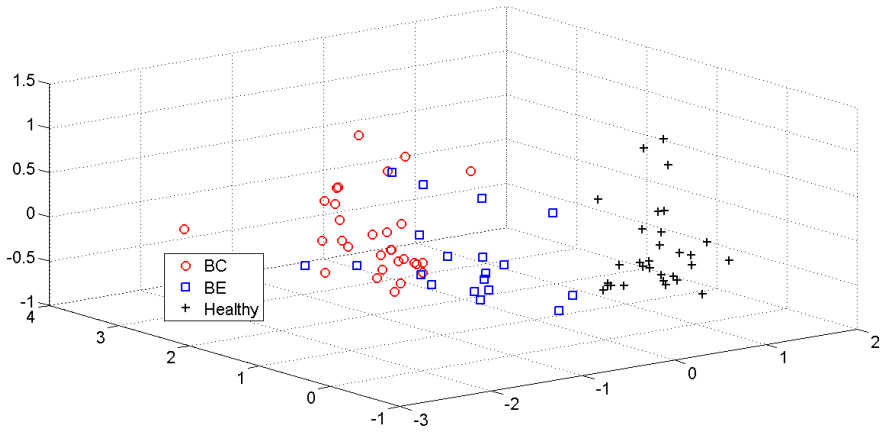

Figure 3:

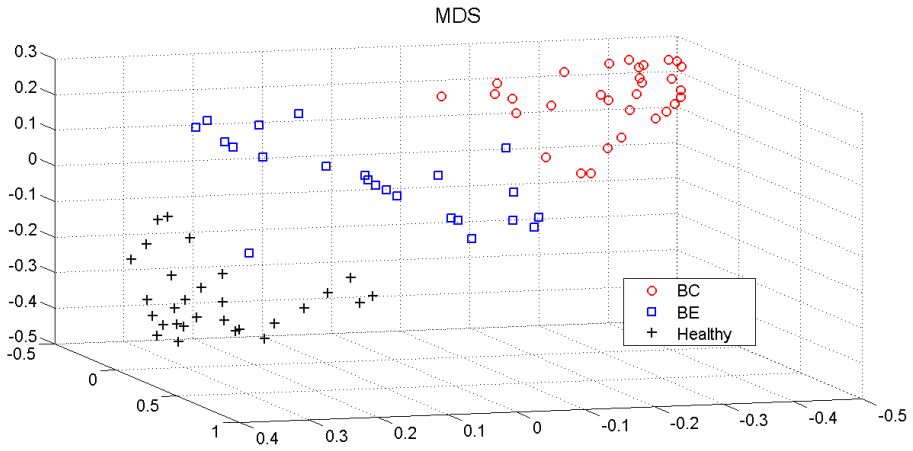

Figure 4:

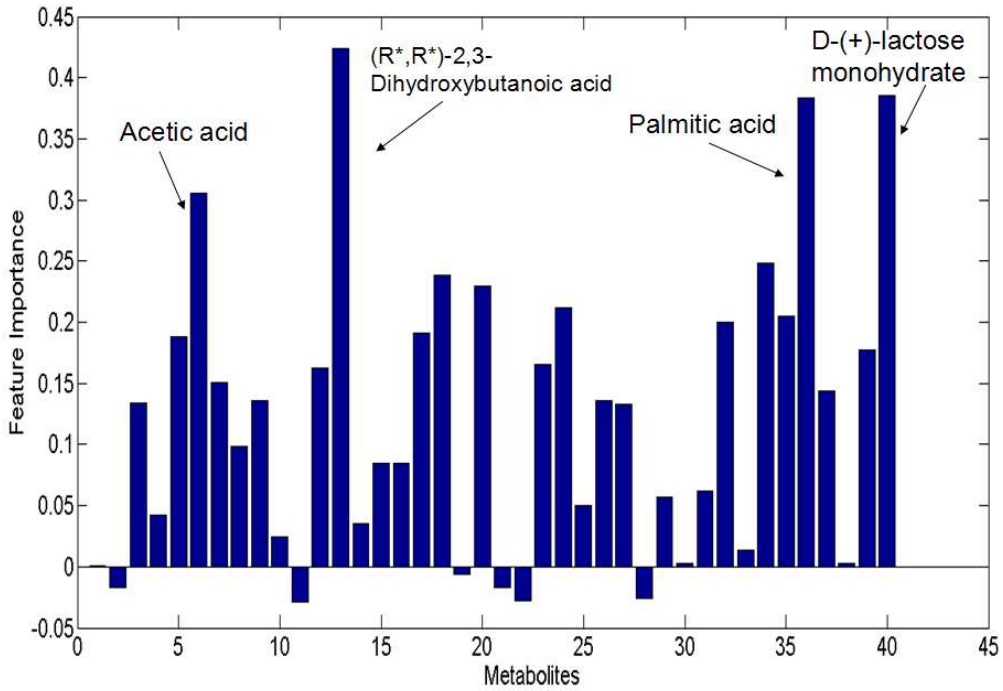

Figure 5:

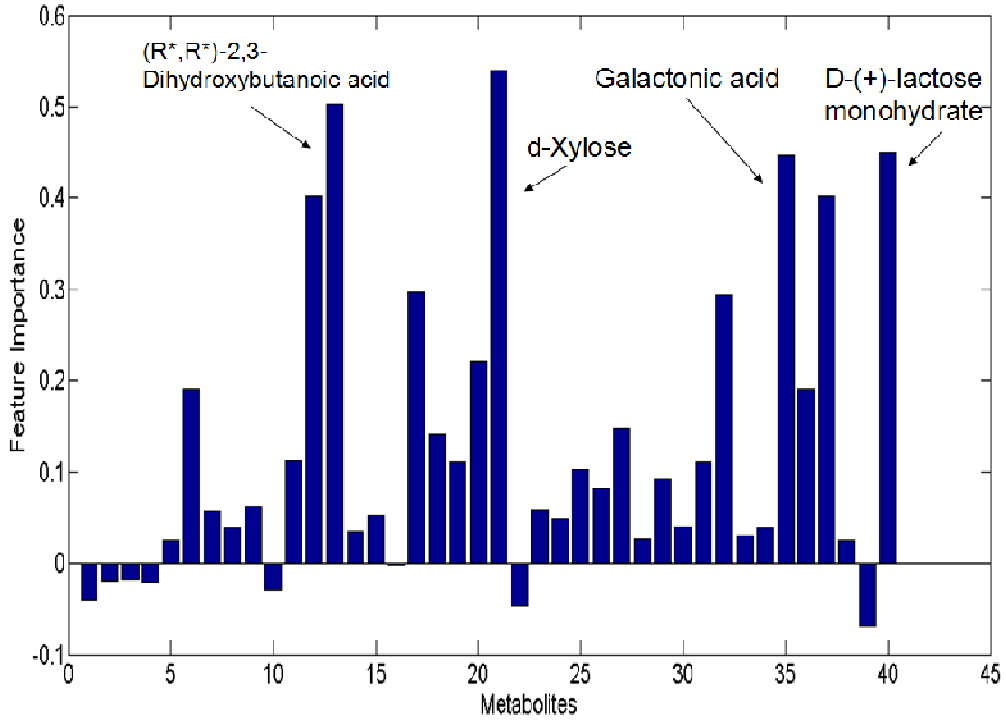

Figure 6: