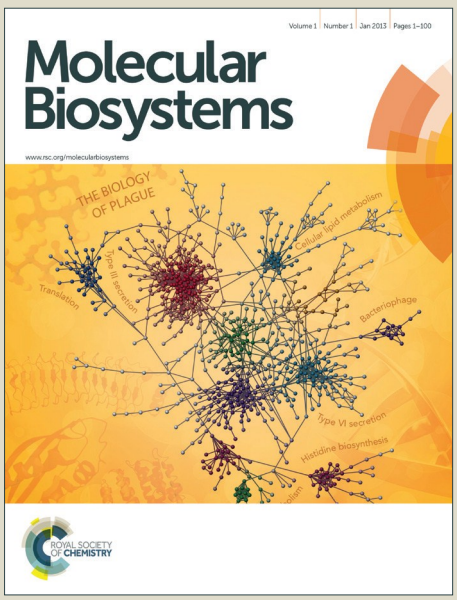
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

LC-MS based Untargeted Metabolomics Study Identified Novel Biomarkers for Coronary Heart Diseases

Ruijun Li,[†a,b] Fengyu Li,[†a,c] Qiang Feng,[c,d,e,h] Zhipeng Liu,[c] Zhuye Jie,[c] Bo Wen,[c] Xun Xu,[c] Shilong Zhong,[*f,g] Guanglei Li,[*a,c] and Kunlun He[*b]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

This study performed untargeted metabolomics for plasma sample from 40 coronary heart diseases patients and 43 healthy controls by high-performance liquid chromatography coupled with mass spectrometry technology to find a set of effective biomarkers for CHD diagnosis and prognosis. The discriminating metabolites were extracted and analyzed by univariate and multivariate analysis methods. We found five metabolites (1-Naphthol, 2-Naphthol, Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate and L-Carnitine) contributing to the separation of CHD patients from healthy controls , a subset of two metabolites in these five were identified as potential plasma biomarkers for CHD diagnosis. Major metabolic pathways associated with these potential biomarkers included nicotinate and nicotinamide metabolism, protein glycosylation, lipid metabolism and fatty acid metabolism. In addition, two potential biomarkers (GlcNAc-6-P and L-Carnitine) were found be to be associated with intestinal microflora, indicated that intestinal microflora may be related to the metabolism and progression of CHD.

1 Introduction

Coronary heart disease (CHD) is a complex disease with very high prevalence and mortality rate worldwide. The incidence of CHD has been increasing for years and it is anticipated that the growth rate of mortality caused by CHD will reach 137% for male and 120% for female from 1990 to 2020[1].The most effective way to prevent CHD from developing is diagnosing in early stage, evaluating the risk factors and taking intervention measures at proper time. So far coronary angiography is the golden standard for CHD diagnosis, but the drawback of this method is that it is invasive, not suitable for massive screening and prognosis monitoring. Therefore, a non-invasive, feasible, and effective method to diagnose CHD urgently needs to be developed. The normal function of heart highly depends on metabolism process. The major energy supplier ATPs are mainly (>90%) produced by fatty acid oxidative phosphorylation[2].The other metabolic substrates include fatty acids, glucose, ketone bodies, lactate and amino acids, which alter in line with internal or external environment change[3]. Previous studies confirmed that many cardiovascular diseases are associated with metabolic abnormality[4]. Cardiovascular diseases (CVD) such as coronary heart disease and cardiac failure undergo a "metabolic shift" as a consequence of both intrinsic and extrinsic perturbations. The fact that core defects in cardiovascular disease are lipid metabolism disorders[5]. In recent years, the progress of multiple omics techniques has allowed us to better understand various diseases at different levels. Genomics, transcriptomics, proteomics, and metabolomics reflect the patient's status at gene, RNA, protein, and metabolites level respectively. However, only changes on metabolites level illustrate the accurate physiological status of human body since they are the final physiology outcomes[4]. Therefore, metabolomics which detects the small-molecule metabolites is a vital tool for early screening, diagnosis, and prognosis of diseases. It consists of measurement of detecting small-molecule metabolites (molecular weight less than 1500 Da) in body fluids by high-throughput methods, analyzed the dynamically change of the metabolites after endogenous (such as genetic mutation) or extrinsic (environmental change) stimulation, and identified the correlation between metabolites and physiological or pathological

[a] BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China.
[b] Department of Cardiology, Chinese PLA General Hospital, Fuxing Road 28, Beijing, 100853, China.
[c] BGI-Shenzhen, Shenzhen, 518083, China.
[d] Department of Biology, University of Copenhagen, Ole MaaløesVej 5, 2200 Copenhagen, Denmark.
[e] Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal Microbiome.
[f] Medical Research Center of Guangdong General Hospital, Guangzhou, Guangdong, China.
[g] Guangdong Cardiovascular Institute, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong, China.
[h] Department of Human Microbiome, School of Stomatology, Shandong University, Shandong Provincial Key Laboratory of Oral Tissue Regeneration, Jinan, China.
[†] These authors contributed equally to this work.
[*] To whom correspondence should be addressed.

alteration[4,6]. The major metabolomics techniques include chromatography, mass spectrometry (MS), and nuclear magnetic resonance (NMR). Compared with NMR, MS-based techniques more frequently since they have higher sensitivity, wilder coverage of the metabolome, improved metabolites identification and discrimination capacity, and modularity to perform compound-class-specific analysis. In recent years, chromatography coupled with mass spectrometry, such as gas chromatography-mass spectrometry (GC-MS), capillary electrophoresis-mass spectrometry (CE-MS), and liquid chromatography mass spectrometry (LC-MS), has significantly expand the scope of metabolites analysis[7]. Since GC-MS and CE-MS are only suitable for volatile and ionic substances respectively, their applications were limited. Meanwhile, LC-MS has been widely used for metabolomics study due to its high separation capacity and the ability of detecting hundreds of compounds simultaneously. In this study, we used high performance liquid chromatography coupled with mass spectrometry (HPLC-MS) technology to detect the changes of plasma metabolites in CHD and healthy control groups. Multivariate statistical analysis methods were used to identify significantly different metabolites and potential biomarkers between CHD patients and healthy individuals, the diagnostic ability of the potential biomarkers was evaluated, and the correlation between potential biomarkers and clinical phenotypes was studied.

2 Materials and Methods

2.1 Participants

In this study, 83 individuals were recruited in Guangdong General Hospital. Among them, 40 were CHD patients diagnosed by angiography, and the other 43 were healthy controls. Blood samples were collected by Vacuette EDTA blood collection tubes and centrifuged at 14000 g for 10 min at 4°C to obtain plasma samples. The plasma samples were stored at -80°C until use. The clinical information of participants was provided in Table 1. This study was designed and performed according to the rules of the Declaration of Helsinki and approved by the Institutional Review Board of BGI-Shenzhen. Written consents were obtained from all participants before sample collecting.

2.2 Materials and Reagents

HPLC grade Formic acid was purchased from Fisher Scientific (Loughborough, UK). Water used in the experiments was obtained from a Milli-Q Ultra-pure water system (Millipore, Billerica, MA). An Agilent ZORBAX ODS C18 column (Agilent Technologies, Santa Clara, CA) (150 mm 2.1 mm, 3.5 μm) was used for all analysis.

2.3 Sample preparation

To extract low-molecular-weight (<1500 Da) metabolites in the plasma samples, these samples were prepared by using the methods mentioned in previous publication with some modifications[8]. Before experiment, all the plasma samples were thawed on ice. A "quality control"(QC) sample was made by mixing and blending equal volumes (10 μL) from each plasma samples. QC was used to represent all the analytes encountered during analysis. For plasma samples, 100 μL of samples were mixed with 200 μL methanol to precipitate protein. The mixture was then centrifuged at 14000 g for 10 min at 4°C. Supernatant was transferred into a 1.5 mL polypropylene tube for the following metabolic profiling experiment.

2.4 HPLC-MS experiments

For HPLC-MS experiment, a LTQ Orbitrap Velos instrument (Thermo Fisher Scientific, MA, USA) set at 30000 resolution was coupled with Shimadzu Prominence HPLC system (Shimadzu Scientific Instruments, Kyoto, JPN). All samples were analyzed in positive mode. Spray voltage was set to 4.5kV and the capillary temperature was set to 350°C. The mass scanning range was 50-1500 mass-to-charge (m/z). Flow rate of Nitrogen sheath gas was 30 L/min and nitrogen auxiliary was 10 L/min. The HPLC-MS system was run in binary gradient mode. Solvent A was 0.1% (v/v) formic acid/water, and solvent B was 0.1% (v/v) formic acid/methanol. The gradient was as following: 5% B at 0 min, 5% B at 5 min, 100% B at 8 min, 100% B at 9 min, 5% B at 18 min, and 5% B at 20 min. The flow rate was 0.2 mL/min. In order to ensure system equilibrium, 5 pooled QC sample were injected at the beginning of experiment. The QC samples was injected between every 5 samples to monitor system stability during the whole experiment.

2.5 Data processing and statistics analysis

The pre-treatments of HPLC-MS data include the following procedures described in previous publications: peak picking, peak grouping, retention time correction, second peak grouping, and annotation of isotopes and adducts[9]. Raw data files from LC-MS were converted into mzXML format and then processed by the XCMS[10] and CAMERA[11] toolbox implemented with the R software(v3.1.1). Intensities of each peaks were recorded and a three dimensional matrix containing arbitrarily assigned peak indices (retention time and m/z pairs), sample names (observations) and ion intensity information (variables) was generated. In order to obtain consistent results, the obtained matrix was further reduced by removing peaks with more than 80% missing values (ion intensity=0) and those with isotope ions from both CHD and healthy control groups. As a quality assurance strategy in metabolic profiling, all retained peaks were normalized to the QC sample using Robust Loess Signal Correction (R-LSC) based on the periodic analysis of a standard biological quality control sample (QC sample) together with the real plasma samples[12]. The relative standard deviation (RSD) values of metabolites in the QC samples was set at a threshold of 30% which was accepted as a standard in the assessment of repeatability in metabolomics datasets[13]. The nonparametric univariate method (Mann-Whitney-Wilcoxon test) was used to discover the significantly changed metabolites among the CHD patients and healthy control. The results was corrected by false discovery rate (FDR) to ensure that metabolite peaks were reproducibly detected[14]. The multivariate statistical analysis (PCA, PLS-DA) were performed to discriminate CHD samples from control subjects. A number of metabolites responsible for the difference in the metabolic profile scan of CHD patients

and control subjects can be obtained on the basis of variable importance in the projection (VIP) threshold of 1 from the 10-fold cross-validated PLS-DA model. The PLS-DA model was validated at a univariate level using FDR test from the R statistical toolbox with the critical p-value set to be lower than 0.05. Three dimensional PLS-DA analysis was also implemented to show the difference between CHD samples and control subjects[15]. Heatmap was used to depict the relatively disturbed and unbalanced metabolism state among CHD samples compared to control subjects. Spearman correlation analysis was implemented in those significantly changed plasma metabolites with clinical phenotype data of CHD patients and control subjects and correlations of metabolites was profiled with Cytoscape software 3.2.1[16]. In addition, receiver operating characteristic (ROC) analysis was used to evaluate diagnostic capability of identified potential biomarkers[17].

2.6 Metabolites identification and validation

The online HMDB database (http://www.hmdb.ca)[18–20] and KEGG database (http://www.genome.jp/kegg)[21,22] were used to identify the metabolites by matching the exact molecular mass data (m/z) of samples with those from database. If a mass difference between observed value and the database value was less than 10 ppm, the metabolite would be identified and the molecular formula of metabolites would further be validated by the isotopic distribution measurements. Reference standards were purchased and used to validate and confirm those significantly changed metabolites by comparing their MS/MS spectra and retention time[23].

3 Result

3.1 Clinical information of participants

An untargeted metabolomics method was used to study plasma samples from 40 CHD patients and 43 healthy controls. The participants' clinical information was listed in Table 1. A total of 13 biochemical indexes were incorporated, including age, triglyceride (TRIG), lipoprotein (LPA), body mass index (BMI), alanine transaminase (ALT), low-density lipoprotein (LDLC), cholesterol (CHOL), high-density lipoprotein (HDLC), aspartate transaminase (AST), hydroxybutyrate dehydrogenase (HBDH), albumin (ALB), total protein (TP), and lipoprotein (LPA). The average age is 59.98 $\pm$ 8.98 (mean $\pm$ SD) and 59.97 $\pm$ 7.043 for CHD patients group and control groups respectively. The average body mass index (BMI) is 24.8 $\pm$ 3.6 and 24.75 $\pm$ 5.2 for CHD and control groups respectively. There is no significant difference in age and BMI between the two groups by student's t-test.

3.2 Analysis of QC samples

Using the untargeted HPLC-MS based metabolomics technique, a total of 2588 metabolites (m/z, mass-to-charge ratio) were obtained in the plasma samples. To assess the stability and reproducibility of the current dataset, QC samples were measured during the whole experimental period. Principal component analysis (PCA) among QC samples and tested samples were performed and the two-dimensional PCA score plot was showed in Supplementary Figure 1. The result showed that QC samples formed a cluster and there is no obvious shift in tested samples, and thus confirmed that our current metabolomics data has good stability and reproducibility.

3.3 Plasma Metabolomics data analysis

To demonstrate the overall profile of the metabolomics data, a cloud plot of all 2588 metabolites were shown in Figure 1A. The results showed that the intensity of 31.22% (808) metabolites were significantly increased in CHD patients' plasma samples (fold change > 1.2) while that of 30.64% (793 m/z) samples were significantly decreased in CHD patients (fold change < 0.8) compared with healthy control group. To evaluate the discriminating power of the obtained 2588 metabolites, we performed principal component analysis for the plasma samples from 40 CHD patients and 43 healthy controls and the two-dimensional PCA score plot was shown in Supplementary Figure 2. The CHD patient group and control group formed separate clusters, indicating that they could be discriminated based on two principal component scores (PC1, PC2 as 11.09% and 6.61%). Besides, the three-dimensional partial least squares discriminant analysis (PLS-DA) scores plot (Figure 1B) also confirmed the significant difference between CHD and control group with PC1, PC2, PC3 as 42.86%, 10.72%, 7.68% respectively. The PLS-DA model was constructed by performing 10-fold cross validation and it has demonstrated good modeling and prediction using 3 components (with cumulated R2(X)=61.26%, cumulated Q2(X)=48.32). To avoid over-fitting of the model, we further validated it with a permutation multivariate analysis of variance (PERMANOVA). The R2 distribution plot of the permutation test for the PLS-DA model among plasma samples was showed in Figure 1C. The permutation model used three latent factors and the probability of this model randomly occurring was less than 0.001. From these above results, we confirmed that the metabolic profiling of CHD patients and control groups is significantly different.

3.4 Identification of metabolic biomarkers

To identify the significantly changed metabolites in CHD patients and potential biomarkers for CHD, two types of analysis were performed. First, Variable Importance for Projection (VIP) scores of all 2588 metabolites were extracted from PLS-DA model and the S-plot demonstrated putative biomarkers were showed in Figure 2A[24]. The modelled covariance (x-axis) and modelled correlation (y-axis) from the PLS-DA model were combined in a scatter plot. The red triangles in S-plot represented the significantly changed metabolites (VIP >= 1). The two-tailed Wilcoxon rank-sum tests were performed with false discovery rate correction (q-value), fold-change (FC), and average intensity of each metabolite from CHD patients normalized to the average value of the same metabolite from control group. A volcano plot combined the statistical test with the magnitude of the change was shown in Figure 2B. It enabled quick visual identification of these metabolites with large fold change values. A total of 1040 significantly different metabolites (with q-value < 0.05 and FC > 1.2 or FC < 0.8) were highlighted in volcano plot with red color. To inte-

grate the results from S-plot and volcano plot, a Venn diagram (Figure 2C) was plotted. Eventually, 83 metabolites in the overlapped area met the following criteria (q-value < 0.05, FC > 1.2 or FC < 0.8, VIP > 1). To visualize the intensity patterns of the 83 significantly changed metabolites in CHD patients and control group, a heat-map was plotted (Figure 3). Each row represents a specific metabolite's intensity and each column represents a individual (patient or healthy control). As shown in Figure 3, there is a distinct pattern between these two groups. To identify potential biomarkers, the 83 significantly changed metabolites were aligned to KEGG and HMDB databases with 10 ppm error tolerant. As a result, a total of 37 potential metabolites were identified and their detailed information is listed in Supplementary Table 1. Among these 37 metabolites, 7 (2,3-Dimethylmaleate, 2-Naphthol, Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate , 1-Naphthol, L-Carnitine, Phenylpyruvate) were verified using purchased reference standards and could be considered as candidate potential biomarkers for CHD diagnosis. The detailed information of these 7 metabolites was listed in Table 2. Among the 7 potential biomarkers, 2,3-Dimethylmaleate, 2-Naphthol, Methylitaconate, and N-Acetyl-D-glucosamine 6-phosphate were significantly increased in the CHD patients, with fold change of 4.24, 17.61, 1.40, 16.58 and q-value of 3.88 x 10-9, 0.014, 3.14 x 10-11, 2.85 x 10-10respectively. The other three metabolites (1-Naphthol, L-Carnitine, and Phenylpyruvate) were decreased in CHD patients, with fold-change value of 0.27, 0.06, 0.06 and q-value of 0.0035, 0.0079, 0.0198 respectively. An online tool MetaboAnalyst 3.0[25] was used to analyze the relevant metabolic pathway. The following 5 pathways were identified (Supplementary Table 2) : nicotinate and nicotinamide metabolism pathway, metabolism of xenobiotics by cytochrome P450 pathway, pathway for phenylalanine, tyrosine and tryptophan biosynthesis, phenylalanine metabolism pathway, amino sugar and nucleotide sugar metabolism pathways.

3.5 Correlation between potential metabolic biomarkers and clinical biochemical indexes

To evaluate the difference in 13 biochemical indexes between CHD and control groups, student's t-test was performed and the results were listed in Table 1. The results showed that LPA were significantly increased while TRIG, LDLC, CHOL, HDLC, APOB, ALB, TP were significantly decreased in CHD patients compared with control groups. Meanwhile, there is no difference in the levels of AST, ALT, and APOA between the CHD and control groups. To access the effects of the patients' covariates (age, BMI and other clinical biochemical factors) on metabolic profiles, PERMANOVA analysis was performed to evaluate the correlation between these 7 potential biomarkers and 13 clinical biochemical indexes. As shown in Figure 4A, the intensity of most potential biomarkers have a strong correlation with the levels of ALB, TP, HDLC, CHOL, LDLC, TRIG, LPA while they have no correlation with BMI, ALT, AGE, APOA, LDLC, AST. Interestingly, the levels of 3 metabolites (2,3-Dimethylmaleate, Methylitaconate, and N-Acetyl-D-glucosamine 6-phosphate) which were significantly increased in CHD patients were negatively correlated with the level of ALB, TP while they were positively correlated with the levels of TRIG and LPA. On the contrary, the levels of 3 metabolites that were decreased in CHD patients (1-Naphthol, L-Carnitine, Phenylpyruvate) have positive correlation with the levels of ALB and TP while they have negative correlation with the level of TRIG, LPA. For instance, the level of N-Acetyl-D-glucosamine 6-phosphate was increased in CHD patients and it has significantly strong negative correlation with TP (p=5.62E-05, rho=-0.442) and ALB(p=6.15E-06, rho=-0.490); the level of L-Carnitine was decreased in CHD patients and it has significantly positive correlation with TP (p=0.014, rho=0.278) and ALB(p=0.005, rho=0.315). These results from PERMANOVA analysis indicated that these clinical indexes are correlated with the plasma metabolic profile of CHD patients. In addition, to evaluate the correlation among the 7 potential biomarkers, spearman correlation analysis was performed and the results were showed in Figure 4B. Interestingly, the levels of 3 biomarkers (Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate, and 2,3-Dimethylmaleate) were increased in CHD patients and they also have significantly positive correlation among themselves. On the other hand, L-Carnitine and Phenylpyruvate were positively correlated with each other but they have no correlation with the other 5 biomarkers.

3.6 Receiver operating characteristic analysis for potential biomarkers

To evaluate the diagnosis ability of these 7 potential biomarkers, receiver operating characteristic (ROC) analysis was applied to an additional validation dataset of 102 plasma samples (59 CHD patients vs 43 healthy controls). The experiment programs and data generation strategies of the validation data set are the same with the discovery set. In the validation data, we constructed a random forest classifier and found that five potential biomarkers among the 7 identified ones (Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate, L-Carnitine, 1-Naphthol, 2-Naphthol) showed the a good discriminating ability with AUC of 89.95% and 95% confidence interval (CI) 83.29% - 96.61% (Figure 5A). The abundance of these five metabolites in CHD patient and control groups from validation dataset was showed in Figure 5B. The level of N-Acetyl-D-glucosamine 6-phosphate is significantly increased in CHD patients while 4 other metabolites are slightly decreased in CHD patients. Meanwhile, for training dataset, the AUC of this group of five potential biomarkers were showed in Supplementary Figure 3A with random forest classifier of 97.44% and 95% confidence interval (CI) of 94.69% - 100%. The ROC result of these 5 individual metabolites in training dataset was listed in Supplementary Figure 3B. The AUC were 90.99%, 89.3%, 66.63%, 68.6%, 65.17% for Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate, L-Carnitine, 1-Naphthol, 2-Naphthol respectively. And AUC of the combined classifier of the 5 metabolites by random forest is 97.44%, indicating that the discriminating power of the combined classifier is also very good in training dataset. Additionally, we compared the level of the five metabolites in two datasets. N-Acetyl-D-glucosamine 6-phosphate and 1-Naphthol showed the same trend in both datasets (Figure 5B), which indicated that

they were better potential biomarkers for CHD diagnosis. AUC of the two metabolites' combination is 91.01% and 95% confidence interval (CI) 84.69% - 97.33% (Figure 5C).

4 Discussion

Coronary heart disease is a chronic and complex disease affected by multiple human and environmental factors. It has been the top health risk for modern society, causing mortality rate higher than the sum of all types of cancers and leading to huge social and economic burden. In recent years, the progress of metabolomics technique has enabled us to measure physiological and pathogenic factors that affect the development of coronary heart disease in a high-throughput mode, to identify potential metabolic biomarkers, and to study possible mechanism for coronary heart disease progression. In this study, non-targeted HPLC-MS technique was used to measure metabolic profile of plasma from CHD patients (n=40) and healthy subjects (n=43). And a group of 5 metabolites (1-Naphthol, 2-Naphthol, Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate, L-Carnitine) has been identified as potential biomarkers in CHD patients. PERMANOVA analysis showed that there was strong correlation between these biomarkers and patients' clinical biochemical indexes such as ALB, TP, HDLC, CHOL, LDLC, TRIG, LPA. Spearman correlation analysis showed that there was significantly positive correlation among these biomarkers, indicating that there may be some unknown underlying relationships between them. The combination of these five metabolites showed an excellent separating capability ability in both training set (AUC=97.44%) and validation set (AUC=89.95%). A subset of the five metabolites consists of N-Acetyl-D-glucosamine-6-phosphate and 1-Naphthol shows a good diagnosis ability, which could potentially be used for clinical diagnosis, prognosis monitor, and early detection of CHD patients. Among these 5 potential biomarkers for CHD, N-Acetyl-D-glucosamine 6-phosphate (GlcNAc-6-P) is the phosphorylated product of N-Acetyl-D-glucosamine (GlcNAc). The process that a single GlcNAc bonds with an oxygen atom on hydroxyl of serine or threonine by O-glycosidic linkage was called protein O-GlcNAc glycosylation. Glycosylation of protein kinases, phosphatases, antioxidant enzymes, transcription factors and heat shock proteins, will change various biological function such as nuclear transport, translation, transcription, DNA repair, signal transduction and apoptosis[26]. Previous study indicates that increase of protein O-GlcNAc glycosylation can induce anti-inflammation, antioxidization, neuroprotection, reducing trauma and bleeding[27]. Especially, O-GlcNAc modification was found to be associated with atherosclerosis, myocardial reperfusion injury, arrhythmia, heart failure and other cardiovascular diseases[28–30]. Methylitaconate is metabolic product of niacin and nicotinamide. Niacin, also named vitamin B3, is an essential vitamin for human body. Niacin was transformed into nicotinamide which is an integral part of coenzyme I and II. Niacin participates in glucose glycolysis, pentose biosynthesis, metabolism of fats, amino acids, proteins and purine. Currently, niacin preparations have been widely used in the treatment of hyperlipidemia. Clinical trials showed that niacin can reduce incidence of coronary, which may reduce the overall mortality[31]. 1-Naphthol and 2-Naphthol belong to Naphthalene family. They are involved in detoxification and activation of exogenous substrates (include drugs, anti-oxidants, odorants, solvents, anesthetics) in cytochrome P450 pathway. L-Carnitine is a quaternary ammonium compound comes from methionine and lysine biosynthesis. It plays an important role in energy metabolism, especially in fatty acid catabolism. It assists transferring long-chain fatty acids through the mitochondrial membrane into the mitochondrial matrix and promoting their β-oxidation which provide energy to the myocardial cell[32,33]. It also involved in carbohydrate metabolism and glucose oxidation. Reduction of L-Carnitine will lead to elevation of blood lipids levels and abnormity in fat metabolism, and then cause acceleration of atherosclerosis and CHD progression. Recent studies also found that L-Carnitine containing trimethylamine (TMA) which can be converted to trimethylamine oxide (TMAO) in gut[34]. TMAO can accelerate atherosclerosis and further promote the progress of coronary heart disease. Previous literature indicated that over 30% of metabolites in human body originate from intestinal microbes and the interaction between microbes and host may contribute to disease progress[35]. In this study, GlcNAc-6-P was identified as potential biomarkers and it was reported to be converted to acetyl glucosamine-6-P by N-acetylglucosamine-6-phosphate deacetylase (NagA) deacetylation. NagA was originated from E. coli so it is possible that there is correlation between gut flora and CHD[36]. In addition, for another potential biomarker L-Carnitine, it contained TMA which was converted to trimethylamine oxide (TMAO) in gut by intestinal flora[37] . Thus, we speculate that gut flora disorder may be related to CHD progression in human body. In the future, microbial species and their associated metabolites might be considered as new indexes and potential targets for diagnosis and treatment of CHD. In summary, 5 potential biomarkers (2-Naphthol, Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate, 1-Naphthol, L-Carnitine) and a combination of two markers for CHD patients were identified by high-throughput non-targeted HPLC-MS method. Protein glycosylation, lipid metabolism, and fatty acid metabolism pathways are found to be associated with the development of CHD. This group of 5 potential biomarkers are strongly correlated with clinical biochemical indexes. These biomarkers could be used for early detection, clinical diagnosis and prognosis monitor of CHD patients in the future.

5 Conclusion

This project successfully applied HPLC-MS based metabolomics to investigate the differences in metabolites between coronary heart disease patients and healthy controls. An obvious separation was obtained by PLS-DA analysis on metabolites from CHD and healthy controls. The 5 metabolites identified in plasma were capable of discriminating patients from healthy subjects in both training and validation dataset, and a subset of them were identified as potential plasma biomarkers for CHD diagnosis. Protein glycosylation and energy metabolism pathways may play a role in development of coronary heart disease. Two metabolites (GlcNAc-6-P and L-Carnitine) associated with intestinal microflora were among the identified metabolites, suggesting that intestinal microflora may be related to the metabolism and pro-

gression of CHD.

Conflict of interest

The authors reported no conflict of interest. The authors alone are responsible for the content and writing of this article.

Acknowledgements

We thank Yanzhao Nong and Lige Gan for suggestions on manuscript revision. This work was supported by the Shenzhen Municipal Government of China [grant numbers JSGG20140702161403250, DRC-SZ[2015]162, KQCX20150330171652 45]

Reference

1 C. J. L. Murray and A. D. Lopez, *Geneva and Boston: World Health Organization and Harvard School of Public Health*, 1996.

2 J. R. Neely, M. J. al Rovetto and J. F. Oram, *Progress in cardiovascular diseases*, 1972, **15**, 289–329.

3 H. Taegtmeyer, L. Golfman, S. Sharma, P. Pazeghi and M. Arsdall, *Annals of the New York Academy of Sciences*, 2004, **1015**, 202–213.

4 M. Kordalewska and M. J. Markuszewski, *Journal of pharmaceutical and biomedical analysis*, 2015.

5 I. T. d. Silva, B. d. Almeida-Pititto and S. R. G. Ferreira, *Archives of endocrinology and metabolism*, 2015, **59**, 171–180.

6 S. H. Shah, W. E. Kraus and C. B. Newgard, *Circulation*, 2012, **126**, 1110–1120.

7 R. Bujak, W. Struck-Lewicka, M. J. Markuszewski and R. Kaliszan, *Journal of pharmaceutical and biomedical analysis*, 2014.

8 H. Luan, N. Meng, P. Liu, J. Fu, X. Chen, W. Rao, H. Jiang, X. Xu, Z. Cai and J. Wang, *GigaScience*, 2015, **4**, 16–19.

9 K. Lu, C. G. Knutson, J. S. Wishnok, J. G. Fox and S. R. Tannenbaum, *Journal of proteome research*, 2012, **11**, 4916–4926.

10 C. a. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, *Anal Chem*, 2006, **78**, 779–787.

11 C. Kuhl, R. Tautenhahn, C. Bottcher, T. R. Larson and S. Neumann, *Analytical chemistry*, 2011, **84**, 283–289.

12 W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall and J. N. Haselden, *Nature protocols*, 2011, **6**, 1060–1083.

13 W. B. Dunn, I. D. Wilson, A. W. Nicholls and D. Broadhurst, *Bioanalysis*, 2012, **4**, 2249–2264.

14 X. B. Ling, H. Cohen, J. Jin, I. Lau and J. Schilling, *Bioinformatics*, 2009, **25**, 1461–1462.

15 M. N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge and P. Savarin, *Molecular BioSystems*, 2015, **11**, 13–19.

16 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, 2003, **13**, 2498–2504.

17 G. Ndrepepa, S. Braun, A. Kastrati and A. Schömig, *Clinical Research in Cardiology*, 2007, **96**, 763–765.

18 D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt and S. Sawhney, *Nucleic acids research*, 2007, **35**, D521–D526.

19 D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong and S. Bouatra, *Nucleic acids research*, 2009, **37**, D603–D610.

20 D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat and E. Dong, *Nucleic acids research*, 2012, gks1065.

21 M. Kanehisa and S. Goto, *Nucleic acids research*, 2000, **28**, 27–30.

22 M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe, *Nucleic acids research*, 2014, **42**, D199–D205.

23 J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, W. Wang, X. Lu, H. U. Häring, E. D. Schleicher and R. Lehmann, *Analytical chemistry*, 2008, **80**, 1280–1289.

24 S. Wiklund, E. Johansson, L. Sjöström, E. J. Mellerowicz, U. Edlund, J. P. Shockcor, J. Gottfries, T. Moritz and J. Trygg, *Analytical Chemistry*, 2008, **80**, 115–122.

25 J. Xia, I. V. Sinelnikov, B. Han and D. S. Wishart, *Nucleic acids research*, 2015, gkv380.

26 M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz and S. Walker, *Nature*, 2011, **469**, 564–567.

27 V. V. Lima, F. R. Giachini, D. M. Hardy, R. C. Webb and R. C. Tostes, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 2011, **300**, R236–R250.

28 N. Fülöp, Z. Zhang, R. B. Marchase and J. C. Chatham, *American Journal of Physiology-Heart and Circulatory Physiology*, 2007, **292**, H2227–H2236.

29 S. P. Jones, N. E. Zachara, G. A. Ngoh, B. G. Hill, Y. Teshima, A. Bhatnagar, G. W. Hart and E. Marbán, *Circulation*, 2008, **117**, 1172–1182.

30 L. Zou, S. Yang, V. Champattanachai, S. Hu, I. H. Chaudry, R. B. Marchase and J. C. Chatham, *American Journal of Physiology-Heart and Circulatory Physiology*, 2009, **296**, H515–H523.

31 I. Jialal, W. Amess and M. Kaur, *Current diabetes reports*, 2010, **10**, 316–320.

32 J. R. Ussher and G. D. Lopaschuk, *Cardiovascular research*, 2008.

33 J. R. Ussher and G. D. Lopaschuk, *Basic research in cardiology*, 2009, **104**, 203–210.

34 R. a. Koeth, Z. Wang, B. S. Levison, J. a. Buffa, E. Org, B. T. Sheehy, E. B. Britt, X. Fu, Y. Wu, L. Li, D. Jonathan, J. a. Didonato, J. Chen, H. Li, G. D. Wu, D. James, M. Warrier, J. M. Brown, R. M. Krauss and W. H. W. Tang, 2013, **19**, 576–585.

35 E. Holmes, J. V. Li, J. R. Marchesi and J. K. Nicholson, *Cell metabolism*, 2012, **16**, 559–564.

36 R. J. White and C. A. Pasternak, *Biochem. J*, 1967, **105**, 121–125.

37 R. A. Koeth, Z. Wang, B. S. Levison, J. A. Buffa, B. T. Sheehy, E. B. Britt, X. Fu, Y. Wu, L. Li and J. D. Smith, *Nature medicine*, 2013, **19**, 576–585.

Graphics and tables

1.1 Graphics

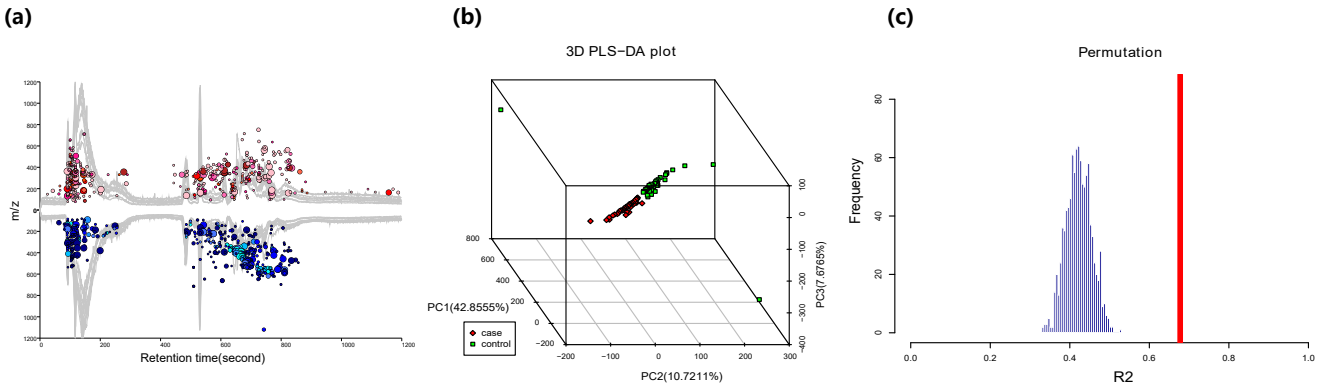


Fig. 1 (a) Cloud plot analysis of all 2588 metabolites. The grey lines in the background represent the total ion chromatograms for each sample. Each bubble on the plot represents a metabolite and the position of the bubbles means the retention time (x coordinate) and the mass-to-charge ratio (y coordinate). Bubbles on the upper half of the cloud plot shows the 808 metabolites with increased intensities in CHD patients (Fold change > 1.2) ; 793 metabolites with decreased intensities (Fold change < 0.8) in CHD patients were showed in the lower halt of the cloud plot. (b) Three-dimensional (3D) PLS-DA score plot for plasma samples. Red rectangles and green circles represent data for CHD patients and healthy control respectively. PLS-DA showed that there is obvious difference between metabolic profiles of CHD patients' plasma samples and healthy controls' samples with PC1(42.86%), PC2(10.72%), PC3(13.68%). Model was constructed by performing 10-fold cross validation with R2 of 61.26% and Q2 of 48.32%. (c) The R2 distribution plot of the permutation test for the PLS-DA model among plasma samples. The model's R2 value (red vertical bar on the right) is significantly distant from the H0 randomly classified permutation distribution (blue vertical bars on the left, n=1000). Thus, the probability that this model randomly occurs is less than 0.001.

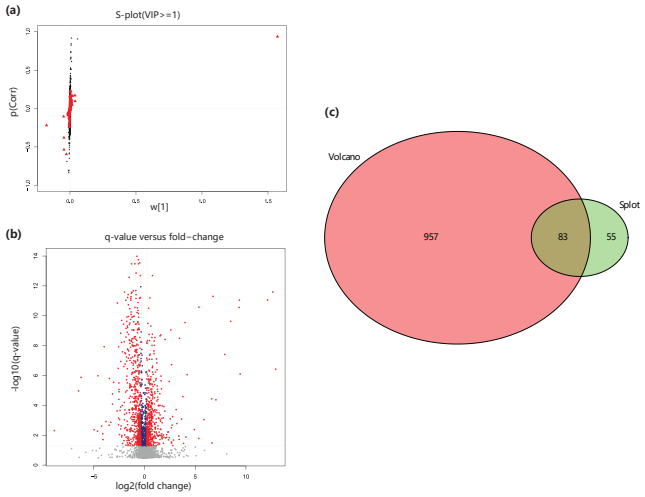


Fig. 2 (a) S-plot combined the modelled covariance (x-axis) and modelled correlation (y-axis) from the PLS-DA model in a scatter plot. The red triangles represent the metabolites with VIP > 1 and the black triangles represent the metabolites with VIP less than or equal to 1. (b) Volcano plot combines the statistical test (y-axis: -log(q-value)) and the magnitude of the change (log2(FC)) of metabolites on a scatter plot. Red points represent the metabolites with q-value < 0.05, and FC > 1.2 or FC < 0.8. Blue points represent the metabolites with q-value < 0.05 and FC between 0.8 and 1.2. Grey points represent the metabolites with q-value < 0.05. (c) Venn diagram integrating results from Volcano-plot and S-plot showed that a total number of 83 metabolites were significantly changed in CHD patients. The left red ellipse represents 1140 metabolites highlighted in volcano plot with q-value < 0.05 as well as FC > 1.2 or FC < 0.8. The right green ellipse represents 138 metabolites from S-plot with VIP > 1. The overlapped area of these two ellipses represents 83 metabolites which met these three criteria simultaneously: q-value < 0.05, FC > 1.2 or FC < 0.8, and VIP > 1.

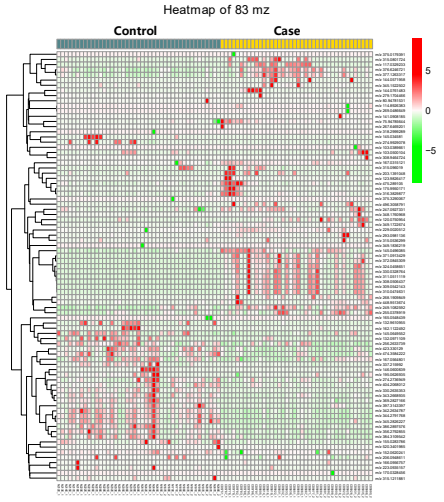


Fig. 3 Heat-map of intensity of 83 significantly different metabolites showed significantly different metabolic profiles between control samples (n=40) and CHD patients (n=43). Each row represent data for a specific metabolite and each column represents an individual (CHD patient or healthy control). Different colors correspond to the different intensity level of metabolites. Red and green colors represent increased and decreased levels of metabolites in CHD patients respectively.

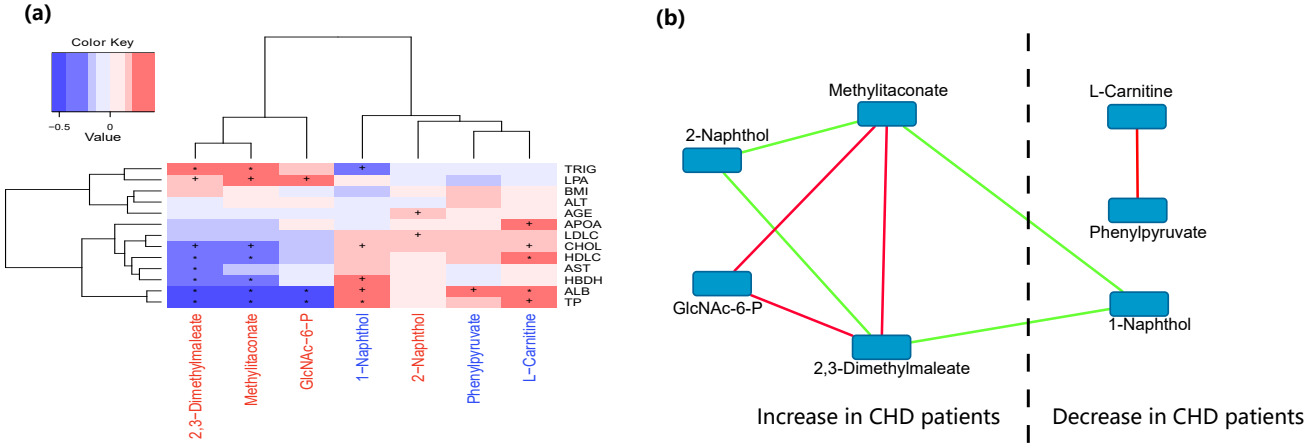


Fig. 4 (a) Spearman correlation analysis was performed for intensities of the 7 identified potential biomarkers (4,3-Dimethylmaleate, 2-Naphthol, Methylitaconate, N-Acetyl-D-glucosamine 6-phosphate , 1-Naphthol, L-Carnitine, Phenylpyruvate) and the 13 biochemical indexes. The 13 biomedical indexes include triglyceride (TRIG), lipoprotein (LPA), body mass index (BMI), alanine transaminase (ALT), low-density lipoprotein (LDLC), cholesterol (CHOL), high-density lipoprotein (HDLC), aspartate transaminase (AST), hydroxybutyrate dehydrogenase (HBDH), albumin (ALB), total protein (TP). Each row and column represents a biomedical index and a specific metabolite respectively. Red and blue colors represented positive and negative correlations respectively. Symbol + in a block means p-value of correlation is less than 0.05; Symbol * in a block means p-value of correlation is less than 0.01. Red panel indicated increased metabolites in CHD patients while blue panel suggested decreased metabolites in CHD patients. (b) Correlation among intensities of the 7 potential biomarkers were showed. Every rectangle represents a specific metabolite. The levels of metabolites on the left side of the line were increased in CHD patients while that of metabolites on the right were decreased. Lines linking the metabolites represent significant correlations (Spearman correlation adjusted p-value < 0.05). Red lines indicate positive correlation while green lines indicate negative correlation.

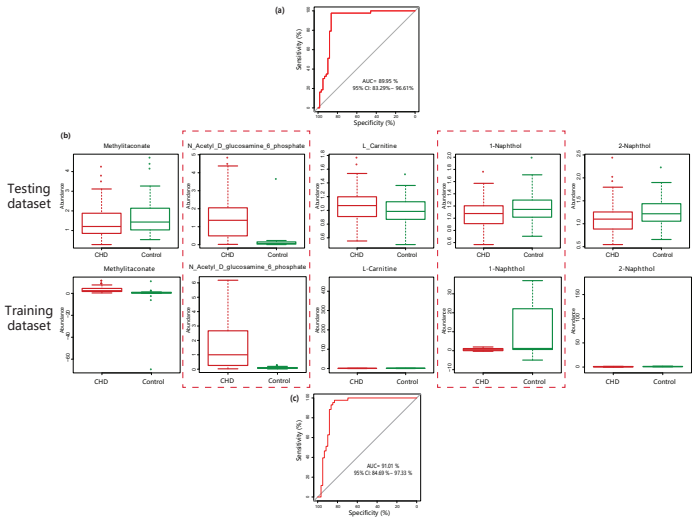


Fig. 5 (a) Receiver operating characteristic(ROC) analysis of 5 identified plasma potential biomarkers in validation datasets. (b) The boxplots from left to right represents the intensity of 5 potential biomarkers in discovery dataset and validation dataset between CHD patients and healthy controls. Metabolites with the same trend in both dataset were marked by rectangle of dash line. (c) ROC analysis of a subset of metablites (N-Acetyl-D-glucosamine 6-phosphate; 1-Naphthol) show the same trend in two dataset.

1.2 Tables

Table 1 Demographic and clinical biochemical index of participants.

	CHD case	Healthy control	p-value	No. of participents (CHD/Control)
Gender(M/F)	29/12	18/25	NA	46/37
AGE	59.98(42-77)	59.97(51-76)	9.99E-01	40/38
BMI	24.8(16.44-30.48)	24.75(17.63-50.32)	9.59E-01	38/43
TRIG	1.77(0.57-5.96)	1.11(0.45-3.49)	1.33E-03	38/38
LDLC	2.8(0.93-5.7)	3.49(2.49-5.47)	1.03E-03	38/38
CHOL	4.6(2.87-8.11)	5.62(3.86-7.88)	5.71E-05	38/38
HDLC	1.04(0.67-1.62)	1.4(0.82-2.31)	1.20E-05	38/38
APOB	0.85(0.52-1.46)	1.01(0.7-1.9)	9.29E-04	31/38
APOA	1.09(0.67-1.63)	1.23(0.67-2.01)	5.18E-02	31/38
LPA	400.1(50.1-1580.84)	149.65(11.17-933.54)	2.01E-03	32/38
ALB	35.76(25.8-42)	42.73(37-47.2)	1.54E-15	38/39
ALT	30.07(13-76.4)	26.79(2-59)	3.70E-01	35/37
TP	64.38(52.3-79.1)	74.92(67.3-86.9)	7.93E-13	38/39
AST	25.75(16-61)	25.78(17-52)	9.88E-01	38/39

Table 2 Potential biomarkers to discriminate CHD patients and controls.

	m/z	RT[2](second)	FC[3]	q-value	VIP	Formula	Adduct Ion	AUC	Pathway
1-Naphthol	145.0649552	531.467	0.268	0.004	2.166	C10H8O	H+	0.686	ko00980[4]
2-Naphthol	144.0571958	531.9175	17.607	0.014	2.712	C10H8O	NAN	0.652	ko00980[4]
2,3-Dimethylmaleate	145.0496085	1156.85	4.238	0.000	1.595	C6H8O4	H+	0.870	ko00760[5]
L-Carnitine	162.1122492	114.494	0.060	0.008	4.092	C7H15NO3	H+	0.666	ko04976[6]
Phenylpyruvate	165.0548439	183.8775	0.061	0.020	1.643	C9H8O3	H+	0.641	koxxxxx[7]
Methylitaconate	167.0315121	104.76	1.380	0.000	1.414	C6H8O4	Na+	0.910	ko00760[5]
GlcNAc-5-P[1,*]	324.0458651	541.3545	16.580	0.000	1.134	C8H16NO9P	Na+	0.893	ko0052[8]

[1] N-Acetyl-D-glucosamine 6-phosphate;
[2] Retention time;
[3] Fold change (CHD/control);
[4] ko00980, Metabolism of xenobiotics by cytochrome P450;
[5] ko00760, Nicotinate and nicotinamide metabolism;
[6] ko04976, Bile secretion;
[7] ko00360, Phenylalanine metabolism; ko00400 Phenylalanine, tyrosine and tryptophan;
[8] ko00520, Amino sugar and nucleotide sugar metabolism;
[*] Metablites that matched characteristic peaks but mistmatched retention time with purchased refernce standard.