MASS SPECTROMETRY-BASED UNTARGETED METABOLOMICS STUDIES OF SELECTED CHRONIC DISEASES: HYPERLIPIDEMIA, NON-ALCOHOLIC FATTY LIVER DISEASE AND CHRONIC KIDNEY DISEASES IN DIABETES

SHAM TUNG TING

PhD

The Hong Kong Polytechnic University

2020

The Hong Kong Polytechnic University

Department of Applied Biology and Chemical Technology

Mass Spectrometry-based Untargeted Metabolomics Studies of Selected Chronic Diseases: Hyperlipidemia, Non-Alcoholic Fatty Liver Disease and Chronic Kidney Diseases in Diabetes

Sham Tung Ting

A Thesis Submitted in Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy

August 2019

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Sham Tung Ting (Name of student)

Abstract

Chronic metabolic disorders are prevalent worldwide and become the leading health hazard and economic burden in the model world, especially three prevalent disorders commonly in modern society: hyperlipidemia, non-alcoholic fatty liver disease and diabetes. Herein, the objective is to apply untargeted metabolomics approaches for discovery of biomarkers, evaluation of potential drug therapy, and metabolic changes of these three prevalent chronic metabolic diseases in biological samples using high-resolution mass-spectrometry. Two animal studies, using piceatannol alone and an aqueous extract of a Chinese herbal medicine, Polygoni Cuspidati Rhizoma et Radix (PCRR), and a clinical observation study were conducted to demonstrate the importance of untargeted metabolomics to unveil the underlying mechanisms of diseases and treatments. The study design, analytical methods and data treatment were optimized with untargeted metabolomics workflows to fit different sample conditions.

The first research is the study is on the investigation of therapeutic effects of piceatannol on high fat diet (HFD)-induced hypercholesterolemic rats using UPLC-QTOF-MS and GC-MS. Untargeted serum metabolomics analysis of normal, hypercholesterolemic and piceatannol-treated SD rats revealed a series of bile acids and fatty acids that showed significant changes among different groups. The significant changes in the ratios of fatty acids suggested down-regulation of stearoyl-CoA desaturase and up-regulation of Δ5-desaturase activities. Quantitative analysis further validated the piceatannol's action in reducing the serum levels of primary and secondary bile acids and upregulating those of most conjugated bile acids. Secondary bile acids are gut-biotransformed metabolites from primary bile acids. The reductions of primary and secondary bile acid levels implied that the gut microbiota may play a role in the treatment of hypercholesterolemia using piceatannol. More reduction of CYP7A1 protein expression and increased levels of most conjugated bile acids in piceatannol-treated

rats compared with HFD-fed group suggested that the cholesterol-lowering effect of piceatannol supplementation did not rely on conversion of cholesterol into bile acids by CYP7A1 in the bile acid biosynthesis. The cholesterol-lowering effect might be a result of inhibition of cholesterol absorption from the intestine by the binding effect of piceatannol on the dietary cholesterols with bile acid micelles.

The second study is to evaluate the therapeutic effect of polyphenol-rich water extract from PCRR using serum metabolomics plus liver lipidomics of SD rat that were affected by HFD-induced non-alcoholic fatty liver diseases (NAFLD). Chemical analysis showed that PCRR water extract is rich in organic acids, stilbenes, flavonoids and other phenolic glucosides. Serum untargeted metabolomics revealed that bile acid synthesis is the major disturbed pathway after intake of PCRR water extract. Elevation of CYP7A1 protein expression indicated the role of PCRR in removal of cholesterols via facilitation of bile acid biosynthesis. Liver lipidomics also showed dramatic difference of lipid profile change between the HFD-fed animal group and the PCRR-treated group. Putatively identified resveratrol and/or polydatin-derived metabolites via microbial transformation were solely detected in the serum of the PCRR-treated group, indicating the action of gut microbiota had participated in the biotransformation. PCRR is likely to have action upon multiple targets to bring out hepato-protective effect.

The third study aims to identify metabolites associated with the development and risk of chronic kidney diseases (CKD), in Chinese diabetic patients. Two independent cohorts of normal and diabetic patients covering five stages of CKD were recruited at different period to cross-check the results. Over 30 identified metabolites showed a positive relationship with glomerular filtration rate (GFR) and an inverse correlation to urinary albumin to creatinine ratio (UACR). Robust correlations of serum succinyladenosine, pseudouridine and 2-(α-D-mannopyranosyl)-L-tryptophan and L,L-TMAP with MDRD GFR in type 2 diabetic patients

were observed, less dependent to sex than serum creatinine. Their associations were early observed when GFR > 60. Furthermore, prediction of MDRD GFR and CKD stages using these biomarkers alone were comparable with serum creatinine. Stepwise linear regression selected pseudourdine and L,L-TMAP as significant and independent predictors adding in the regression equation of MS-detected creatinine and gender in prediction of MDRD GFR. Combination of these biomarkers with serum creatinine improved the diagnostic ability of the serum creatinine alone in differentiation of early CKD stages. Increased production of many uremic retention solutes detected by UPLC-MS and their reduction of kidney clearance by renal degeneration accounted for the progression of CKD in diabetes. Such impairment also might be partly attributed to the increased inflammatory stimuli of diabetic redox imbalance and dysregulation of amino acid and related pathways. Our finding shows the biomarkers are potential supplement to existing clinical markers of early CKD development in diabetic patients and/or better than serum creatinine.

In combination with existing biochemical technique and other platforms, these three studies of chronic diseases demonstrated that an untargeted metabolomics approach is a useful and advanced technique for an overview of the metabolic status and discovery of the biological pathways associated to the disease pathogenesis and exploration of biological actions of drug treatments. The action of gut microbiota that interacted with diet, phytochemicals and endogenous metabolites in the three studies should be further examined. Future prospects on more organs, tissues and biofluids with integration of metabolomics and other omics tools would broaden the spectra of metabolites and give a comprehensive view of the molecular processes associated to the chronic diseases.

Research Publications

Journal Publications

1. Amy P Chiu, Barbara R Tschida, **Tung-Ting Sham**, Lilian H Lo, Branden S Moriarity, Xiao-Xiao Li, Regina CLo, David E Hinton, Dewi K Rowlands, Chi-On Chan, Daniel Kam-wah Mok, David A Largaespada, NadiaWarner, Vincent W Keng. HBx-K130M/V131I promotes liver tumorigenesis in transgenic mice viaAKT/FOXO1 signaling pathway and arachidonic acid metabolism. *Molecular Cancer Research*. 2019
2. Hui-Hui Xiao*, **Tung-Ting Sham***, Chi-On Chan, Meng-Heng Li, Xi Chen, Qing-Chang Wu, Daniel K. Mok, Xin-Sheng Yao, Man-Sau Wong. A metabolomics study on the bone protective effects of a lignan-rich fraction from Sambucus Williamsii Ramulus in aged rats. *Frontier in Pharmacology*. 2018, 22(2):703-708. ***These authors have contributed equally to this work and co-first authors.**
3. Yu-Feng Xing, Da-Qiao Zhou, Jing-Song He, Chun-Shan Wei, Wei-Hao Zhong, Zhi-Yi Han, De-Ti Peng, Mu-Min Shao, **Tung-Ting Sham**, Daniel Kam-Wah Mok, Chi-On Chan, Guang-Dong Tong. Clinical and histopathological features of chronic hepatitis B virus infected patients with high HBV-DNA viral load and normal alanine aminotransferase level: A multicentre-based study in China. *PLOS ONE*. 2018,13(9): e0203220.
4. **Tung Ting Sham,** Huan Zhang, Daniel Kam Wah Mok, Shun Wan Chan, Jianhong Wu, Songyun Tang, Chi On Chan. Chemical analysis of Astragali Complanati semen and its hypocholesterolemic effect using serum metabolomics based on Gas Chromatography-Mass Spectrometry. *Antioxidants.* 2017, *6*(3):57.
5. **Tung-Ting Sham**, Meng-Heng Li, Chi-On Chan, Huan Zhang, Shun-Wan Chan and Daniel Kam-Wah Mok. Cholesterol-lowering effects of piceatannol, a stilbene from wine, using untargeted metabolomics. *Journal of Functional Foods.* 2017;28:127-137.

6. Yam-Fung Ng, Philip Chiu-Tsun Tang, **Tung-Ting Sham**, Wing-Sum Lam, Daniel Kam-Wah Mok, Shun-Wan Chan. Semen Astragali Complanati: An ethnopharmacological, phytochemical and pharmacological review. *Journal of Ethnopharmacology*. 2014;155(1):39-53
7. **Tung-Ting Sham,** Chi-On Chan, You-Hua Wang, Jian-Mei Yang, Daniel Kam-Wah Mok, and Shun-Wan Chan. A review on the traditional Chinese Medicinal herbs and formulae with hypolipidemic effect. *BioMed Research International*. 2014 (2014), Article ID 925302, 21 pages.

Conference publications and presentations

1. **Tung-Ting Sham**, Chi-On Chan, Huan Zhang, Shun-Wan Chan, Daniel Kam-Wah Mok. Serum metabolomics in treatment of hypercholesterolemia with Polygoni Cuspidati Rhizoma et Radix extract: an *in vivo* study. Poster presentation on the 24th Symposium on Chemistry Postgraduate Research in Hong Kong at City University of Hong Kong on 6th May 2017.
2. **Tung-Ting Sham**, Meng-Heng Li, Chi-On Chan, Hui-Hui Xiao, Man-Sau Wong, Xin-Sheng Yao, Daniel Kam-Wah Mok. UPLC-QTOF-MS and GC-TOF-MS based serum metabolomics of the protective effect of Sambuci Williamsii Ramulus against ovariectomized rats. Oral presentation on Hong Kong Society of Mass Spectrometry Symposium at Hong Kong Polytechnic University, Hong Kong on 24th June, 2017. (HKSMS Conference Award)
3. Melody Yee-Man Wong, Ailsa Chui-Ying Yuen, **Tung-Ting Sham**, Huan Zhang, Chi-On Chan Shun-Wan Chan, Daniel Kam-Wah Mok. Application of lipidomics and tissue imaging mass spectrometry for the study of mechanism of action of *Scutellaria baicalensis* in lowering blood cholesterol level. Poster presented on the 7th Asia Oceania Mass Spectrometry Conference at Biopolis, Singapore on 10 - 13th December, 2017.

4. **Tung-Ting Sham**, Meng-Heng Li, Chi-On Chan, Hui-Hui Xiao, Man-Sau Wong, Xin-Sheng Yao, Daniel Kam-Wah Mok. Protective effect of ethanol fraction from *Sambucus williamsii* against ovariectomized rats using untargeted serum metabolomics. Poster presented on The Sunney and Irene Chan Lecture in Chemical Biology 2016 on 3rd October, 2016. (The second-best poster award in biological field).
5. **Tung-Ting Sham**, Chi-On Chan, Hui-Hui Xiao, Man-Sau Wong and Daniel Kam-Wah Mok. Ultra-Performance Liquid Chromatography-Quadrupole Time of Flight-Mass Spectrometry (UPLC-QTOF-MS) based serum metabolomics of the protective effect of ramulus of *Sambucus williamsii* HANCE (JieGuMu) against ovariectomized rats. Oral presentation and poster presented on the 15th Meeting of Consortium for Globalization of Chinese Medicine (CGCM) at Academia Sinica, Taipei on 22-25th August, 2016. (Travel Award)
6. **Tung-Ting Sham**, Na Ge, Chi-On Chan, Hui-Li Sun, Daniel Kam-Wah Mok. UPLC-QTOF-MS based serum metabolomic effect of diabetic kidney disease in Chinese patients. Poster presented on the12th Annual Conference of the Metabolomics Society in Dublin, Ireland on 27-30th June, 2016.
7. **Tung-Ting Sham**, Chi-On Chan, Shun-Wan Chan and Daniel Kam-Wah Mok. A study of hypolipidemic effect of piceatannol on serum metabolomics of hypercholesterolemic rats by Gas Chromatography/Mass Spectroscopy and Ultra-Performance Liquid Chromatography-Quadrupole Time of Flight/Mass Spectrometry. Poster presented on the 3rd International Conference and Exhibition on Traditional and Alternative Medicine, in Birminghan, UK on 3-5th August, 2015
8. Nai-ping Dong, Ka-Yi Man, Chi-On Chan, **Tung-Ting Sham** and Daniel Kam-Wah Mok. Componentmatcher: a package for quality control of traditional Chinese medicine. Poster presented on the XV Chemometrics in Analytical Chemistry in Changsha, China on 22-26th

June, 2015.

9. **Tung-Ting Sham**, Chi-On Chan, Shun-Wan Chan and Daniel Kam-Wah Mok. Effect of diet-induced hypercholesterolemia on serum metabolomics of rats via UPLC-QTOF/MS. Poster presented on the 22nd Symposium on Chemistry Postgraduate Research in Hong Kong on 18th April, 2015.
10. **Tung-Ting Sham**, Song-Yun Tang, Chi-On Chan, Shun-Wan Chan and Daniel Kam-Wah Mok. Effect of Total Saponins from *Panax notoginseng* on Serum Lipid Levels in Diet-induced Hypercholesterolemic Rats and Serum Fatty Acid Profiling Via Gas Chromatography-mass Spectroscopy (GC/MS). Poster presented on the 13th Meeting of Consortium for Globalization of Chinese Medicine in Beijing, China, 27 - 29th August, 2014. (Travel Award)

Patent

1. Na Ge, Huili Sun, Pengxun Han, Shunmin Li, Daniel Kam-Wah Mok, Chi-On Chan, Tung-Ting Sham, Naiping Dong. Identification of metabolites associated with chronic kidney diseases in Chinese diabetic patients: an untargeted serum metabolomics study using UPLC-Orbitrap-MS. 2020. China. Patent application number: PCT/CN2020/090927. Assignee: Shenzhen Traditional Chinese Medicine Hospital. Applicant: Shenzhen Research Institute, Hong Kong Polytechnic University. Filed on 19th May, 2020. Under consideration.

Acknowledgments

I would like to express my deep sense of gratitude to my honourable supervisor, Dr. Mok Kam-Wah Daniel, for his supervision and patient guidance of my Ph.D. study. I have received fruitful and valuable experiences and comprehensive knowledge under his inspiration and guidance for the work in analytical chemistry of traditional Chinese medicines, metabolomics and mass spectrometry. He also gave me many chances and freedom to participate different valuable on-site training workshops and overseas conferences, which are highly relevant to my researches.

I would like to specially thank Dr. Chan Chi-On for his timely, thoughtful and valuable guidance and inspiration in many ways throughout my whole study. Thanks to the help and supports from my team members, Mr. Li Meng-Heng, Ms Yuen Chui-Ying Alisa, Ms. Man Ka-Yi, Ms. Wan Siu Wei, Anna, and Mr. Timothy Tang.

I am truly fortunate and honoured to have my collaborator Dr. Chan Shun-Wan and his teammates, Dr. Zhang Huan, Dr. Ng Yam-Fung and Ms. Cheng Huan-Le for their assistance on animal studies and biological analysis, especially Dr. Zhang Huan for my Western blot experiments in Chapters 3 and 4.

My sincere and heartful thanks go to another collaborators, Dr. Hui-Li Sun and Dr. Ge Na from Shenzhen Traditional Chinese Medicine Hospital for providing clinical serum sample, clinical data and professional guides in my clinical study of Chapter 5. I would like to thank Dr. Dong Nai-ping for his assistance with programming signal correction of data matrix for my clinical data in this Chapter.

The technical supports and advice on liquid chromatography mass spectrometer from Dr. Wong Yee-Man, Melody, Dr. Pui Kin So and Ms. Xie Xiujuan are appreciated. I am also grateful to Ms Leung Hong-Man Josephine for her valuable advice in the use of English in my thesis and manuscripts.

I also acknowledge for the supports by Basic Research Project of Shenzhen Science and Technology Plans (Grant numbers: JCYJ20151030164008764, JCYJ20160330171116798, JCYJ20160229173844278, JCYJ20151030164022389), Natural Science Foundation of China (Grant No. 81303203 and 81503222), General Research Fund of the Research Grant Council of the Hong Kong Special Administrative Region (Project no. 15302718), Hong Kong Research Grants Council (CRF Grant No. C5031-14E and GRF Grant No. 5029/13P), the Large Equipment Funds and University Research Facility in Chemical and Environmental Analysis and Life Sciences of the Hong Kong Polytechnic University, the Hong Kong Chinese Materia Medica Standards Project and Studentship from Hong Kong Polytechnic University

Above all, I am heartily indebted to my parents, sister and brother for their care and continuous support. Especially, I would like to express my deepest thanks to my husband's love throughout these years. I could not have possibly finished this study without his encouragement.

Last but not the least, I thank those who offered me help directly and indirectly. Although many have not been mentioned, none of them is forgotten.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Full Form
5-LOX	5-Lipoxygenase
15-LOX	15-Lipoxygenase
ALT	Alanine transaminase
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
AUC	Area under the receiver operating characteristic curve
BP	Blood pressure
C16:0	Palmitic acid
C18:0	Stearic acid
C18:1n9c	Oleic acid
C19:1n9c	*cis*-10-nonadecenoic acid
C20:3n6c	Dihomo-γ-linolenic acid
C20:4n6c	Arachidonic acid
Cer	Ceramide
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COX-1	Cyclooxygenase-1 isoenzyme
COX-2	Cyclooxygenase-2 isoenzyme
CV	Coefficient of variation
CVD	Cardiovascular disease
CYP7A1	Cytochrome P450 7A1 (Cholesterol 7 α-hydroxylase)
DAD	Diode array detector
DG	Diacylglycerol
EC18:0	Esterified stearic acid
EC20:3n6c	Esterified dihomo-γ linoleic acid
EC20:4n6c	Esterified arachidonic acid
EFA	Esterified fatty acid
ESI	Electrospray ionization
ESRD	end-stage renal disease
FAME	Fatty acid methyl ester

FPG	Fasting plasma glucose
GC-MS	Gas Chromatography-Mass Spectrometry
GFR	Glomerular filtration rate
HCD	Higher energy collisional dissociation
HDL-C	High density lipoprotein cholesterols
HFD	High fat diet
HPLC	High performance liquid chromatography
IDO	Tryptophan 2,3-dioxygenase
IS	Internal standard
KDIGO	Kidney Disease: Improving Global Outcomes
L,L-TMAP	N, N, N-trimethyl-L-alanyl-L-proline betaine
LC	Liquid Chromatography
LC-MS	Liquid Chromatography- Mass Spectrometry
LDL-C	Low density lipoprotein cholesterols
LDLR	Low density lipoprotein receptor
LOD	Limit of detection
LOQ	Limit of quantification
LSD	Least significance difference
LysoPC	Lysophosphatidylcholine
LysoPE	Lysophosphatidylethanolamine
LysoPS	Lysophosphatidylserine
m/z	Mass-to-charge ratio
MDRD	the abbreviated Modification of Diet in Renal Disease
MDRD GFR	estimated GFR calculated by the abbreviated MDRD equation
MS	Mass spectroscopy
MS/MS	Tandem mass spectrometry
NAFLD	Non-alcoholic fatty liver disease
NAG	*N*-acetyl-β-D-glucosaminidase
NASH	nonalcoholic steatohepatitis
NEFA	Non-esterified fatty acid
OPLS-DA	Orthogonal Partial Least Squares - Discriminant Analysis
OOB	Out-of-bag
PCA	Principal component analysis

PCRR	Polygoni Cuspidati Rhizoma et Radix
PCT	Piceatannol
PG	Phosphatidylglycerol
PLS-DA	Partial Least Squares - Discriminant Analysis
QC	Quality control
RF	Random forest
ROC	Receiver Operating Characteristic
SD	Sprague-Dawley
SBP	Systolic blood pressure
SEM	Standard error of the mean
SM	Sphingomyelin
TC	Total cholesterols
TCM	Traditional Chinese medicines
TG	Triacylglycerols
UACR	Urinary albumin-to-creatinine ratio
UPLC-Orbitrap-MS	Ultra-Performance Liquid Chromatography-Orbitrap-Mass Spectrometry
UPLC-QQQ-MS/MS	Ultra-Performance Liquid Chromatography Triple Quadrupole-Tandem Mass Spectrometry
UPLC-QTOF-MS	Ultra-Performance Liquid Chromatography Quadrupole Time of Flight-Mass Spectrometry
VIP	Variable Importance in the Projection

Chapter 1 General introduction and objectives

1.1 Background

Chronic diseases are complex disorders that progress slowly with long-lasting time (Mastrangelo and Barbas 2017). They are also called non-communicable diseases as they are non-transmissible, non-infectious and always less noticeable than communicable diseases. Nowadays, chronic diseases account for the major cause of death worldwide (over 70% of all deaths globally) with an alarming rise of prevalence in developed and developing regions (World Health Organization 2018a). The rapid growth of morbidity of chronic diseases (World Health Organization 2018b) (Fig. 1.1) accounts for high proportion of mortality, economic, psychological and social impact (Mastrangelo and Barbas 2017). This will be a future threat and a huge burden to the whole population.

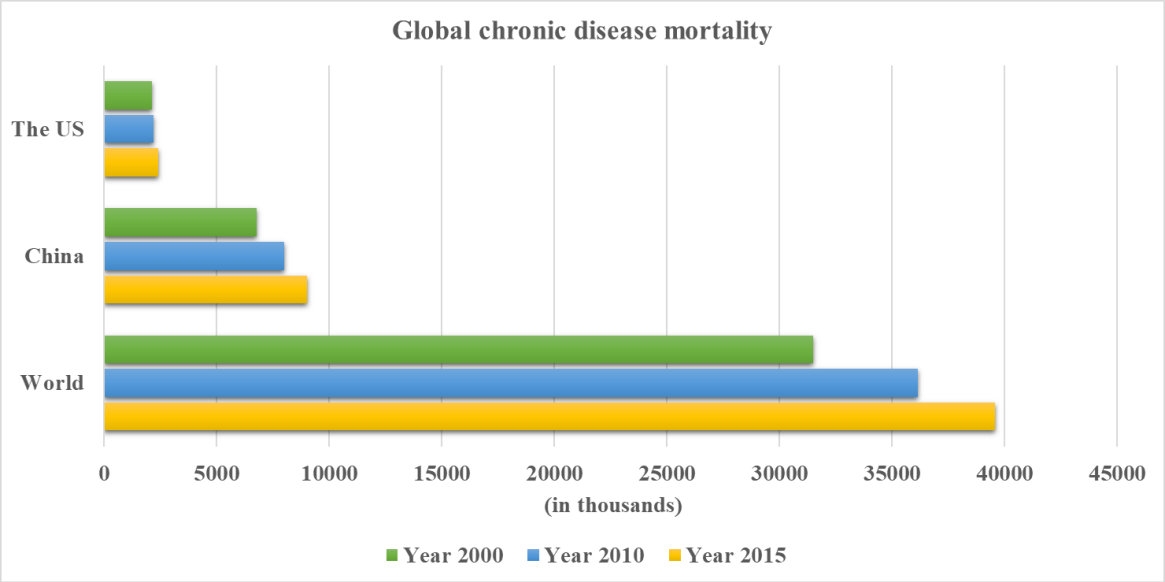


Fig. 1.1 The mortality of chronic disease in the US, China and the overall in the world (World Health Organization 2018b)

Hyperlipidemia, non-alcoholic fatty liver disease (NAFLD) and type 2 diabetes are three relatively milder and more asymptomatic chronic diseases than other major types of chronic diseases (cardiovascular diseases, cancer and respiratory chronic diseases), but they affect various parts of the body such as cardiovascular system and kidneys. They are commonly attributed to modifiable behavioural risk factors, especially over-nutrient diet and lack of exercises in the modern society and prevalent in the adult population (Younossi et al. 2016; Akbartabar Toori et al. 2018; Lecoffre et al. 2018; International Diabetes Federation 2017). Unless the patients' condition is controlled well, hyperlipidemia, NAFLD and type 2 diabetes generally progress to more severe and advanced stages such as ischaemic heart diseases (Nelson 2013; Navar-Boggan et al. 2015), hepatocellular carcinomas (Anstee, McPherson, and Day 2011; Sayiner et al. 2016) and kidney failure (Alicic, Rooney, and Tuttle 2017), respectively, especially among the aging population. Despite their prevalence in the modern society, the pathophysiology of these three chronic diseases is still not fully understood.

Metabolomics is a promising solution for the study of disease mechanism, diagnosis and assessment of treatment to chronic diseases, particularly for diabetes (Merino et al. 2018), obesity (Rauschert et al. 2014) and cardiovascular diseases (McGarrah et al. 2018). Scientific information derived from these techniques can be employed to provide accurate and clinically useful predictive or diagnostic information for the management of major chronic diseases (Kim, Ho Lee, and Sweeney 2013). In addition, the untargeted approaches have their advantage of global metabolic profiling without restriction on specific classes of metabolites or pathways. Therefore, the application of the untargeted approaches to look for new disease pathophysiology and protective mechanisms of novel drugs from natural sources can be effectively achieved with unexpected but fruitful outcomes.

1.2 Objectives of this thesis

The overall goal of this dissertation is to establish a reliable and simple metabolomics platform using untargeted metabolomics with mass spectrometry and chemometrics to provide an insight on chronic diseases. To achieve this goal, my work covered the application of appropriate methodology from study design, analytical method development, and data mining towards detection of biological and diagnostic relevant information.

The project was divided into the following parts:

1. identifying potential biomarkers and novel underlying pathways related to the intervention of piceatannol upon hypercholesterolemic rats by untargeted UPLC-MS and GC-MS serum metabolomics, validated by targeted LC-MS approaches (Chapter 3)
2. applying the method of untargeted UPLC-MS serum metabolomics in chapter 3 plus liver lipidomics to identify the hepato-protective effect of a natural product, Polygoni Cuspidati Rhizoma et Radix (PCRR) supplementation to NAFLD rats (Chapter 4)
3. improving current diagnostic methods by biomarker discovery of chronic kidney diseases (CKD) in diabetic patients using two independent cohorts with untargeted UPLC-MS serum metabolomics (Chapter 5)

1.3 Rationale and thesis outline

Chapter 2 introduced the basic workflow of mass spectrometry based untargeted metabolomics. Details about experimental conditions for reliable data acquisition, data pre-treatment and multivariate statistic technique to identify significantly changed candidates were illustrated and explained.

The first part of my research project is the study of the cholesterol-lowering effect of the administration of a single compound, piceatannol, in an animal model which is high fat-diet (HFD) induced hypercholesterolemia. Hypercholesterolemia is very prevalent in modern society and aging population. Its treatment with a phytochemical, resveratrol, has already been explored before but its solubility and oral bioavailability are low (Salehi et al. 2018). Piceatannol is also a common phytochemical found in fruit and wine and is the metabolite of resveratrol. The effect of piceatannol to high fat diet-induced hypercholesterolemia is seldom studied. With the application of GC-MS and LC-MS on serum samples, its action to the serum metabolic profiling in male rats were examined. It is well-known that gender is a biological variable and affects the metabolic status from disease pathophysiology to drug response so only male rats were used. Results of untargeted serum metabolomics that showed the change in bile acid synthesis was further validated by Western blot of the key protein expression of the enzyme in the pathway as well as quantification of bile acid concentration in serum. Most changes in the untargeted approach were successfully observed in the quantified ones. The investigation is shown in Chapter 3.

Because of the promising results of piceatannol, a similar approach was employed to examine the effect of an herbal extract in the second part of the research project. The polyphenol-rich water extract from the TCM, PCRR, is selected to treat HFD-induced male rates with NAFLD. NAFLD can be developed from hypercholesterolemia using the same type of animal models as piceatannol in the study. PCRR is rich in resveratrol and resveratrol glucoside (polydatin). It is proposed that PCRR may have similar or better protective effect than piceatannol against HFD-induced NAFLD due to its multiple bioactive compounds. Furthermore, the underlying mechanism of PCRR to NAFLD is not well discussed in the current literatures. Thus, there is a need to

systematically investigate therapeutic effects of PCRR with metabolomics. Thus, serum metabolic profiling, coupled with hepatic lipidomics was conducted to give a more comprehensive picture for elucidation of the alteration of cholesterol and lipid metabolism of NAFLD. Here, the cholesterol-lowering effect of PCRR was brought by elevation of bile acid synthesis, as evidence by increased CYP7A1 protein expression levels using Western blot. Its lipid-regulating action was also unravelled. This study was the first work of serum metabolomics and liver lipidomics of PCRR to NAFLD. Details are discussed in Chapter 4.

Finally, the untargeted metabolomics approaches have been applied to a more complicated case, human serum samples, as there are many uncontrollable confounding factors of human samples than those of monitored animal counterpart. Despite the popularity of metabolomics studies of diabetes and CKD, there is lack of comprehensive study of CKD in Chinese diabetic patients covering from a normal state to the end stage of CKD. The existing clinical markers have many drawbacks for classification of stages of CKD such as gender dependence. This part is about the improvement of classification methods and prediction using novel biomarkers across all stages with assistance of current clinical diagnostic methods. Validation through comparison of two independent cohorts is seldom found in metabolomics studies. Hence, in collaboration of a Chinese hospital, two separate cohorts of diabetic patients with CKD comprising the earliest stage to end stage of kidney diseases were recruited. Afterwards, two untargeted metabolomics of a total of 196 human samples integrated from the two cohorts were analyzed with the same platform, UPLC-Orbitrap-MS analysis at different period of time. Independence of two cohorts did enhance the robustness and diagnostic performance of CKD biomarkers. One novel biomarker, succinyladenosine, was detected in this study and showed gender independence in the

correlation with GFR and a combination of metabolites (MS-detected serum creatinine, pseudouridine and L,L-TMAP) outperformed MS-detected serum creatinine alone in the diagnosis of CKD. Chapter 5 has more details on such finding.

The last part of this project, Chapter 6 summarized and discussed some of the limitations encountered in the three chapters. Future works have to deal with such limitations to make the study design and the experiment process more reliable and robust.

Overall, untargeted metabolomics methods using mass spectrometry have been introduced in three chronic diseases including hyperlipidemia, NAFLD and CKD in type 2 diabetic patients and successfully demonstrated how this platform worked on the analysis of biological samples from rats to human samples. Using untargeted approaches could make contribution in advancing our knowledge of disease pathogenesis and actions of treatment as well as improving the clinical diagnosis and drug development.

Chapter 2 Metabolomics analysis

2.1 Introduction of metabolomics

Previously, genomics, transcriptomics and proteomics (studies of the complete set of cellular DNA, mRNAs and proteins, respectively) have been demonstrated to be powerful systematic biological tools for studying biological responses to xenobiotic exposure at various levels. However, they largely ignore the dynamic metabolic status of the whole organism for the understanding of the integrated cellular function in living systems (Nicholson, Lindon, and Holmes 1999).

Metabolomics is a method that studies the dynamic changes of metabolites or intermediates that represent the most downstream level of the flow of gene expression (Ryan and Robards 2006). Moreover, metabolomics is the nearest to biological phenotypes compared with other “omics” techniques (Fig. 2.1), with the advantage of reflecting the changes in phenotype and function. Metabolomics investigates metabolites or low-molecular-weight intermediates (50-1500 Da) in cells, tissues, and/or body fluids (Metabolomics Society 2016). Metabolites are context-dependent, and their levels vary according to the physiology, developmental or pathological state of the cell, tissue, organ or organism (Oliver 2002). Profiling the endogenous metabolites could give us an overview of the metabolic status and global biochemical events associated with the body (Metabolomics Society 2016).

Metabolomics is a newly emerging field in the advanced analytical biochemistry (Oliver et al. 1998; Courant et al. 2014). In the past two decades, the annual publication number using "metabolomics" as keyword in SciFinder (http://scifinder.cas.org) elevated exponentially up to around 5700 publications in 2019 as shown in Fig. 2.2A. Academics, regulatory authorities and commercial sectors have paid a growing

attention on metabolomics and the development of the filed (Metabolomics Society 2016). It is notable that there has been a continuous growth in literatures related to the study of diseases from 18% in 2001 to 33% in 2019 among all metabolomics publications (Fig. 2.2A). Others are related to different fields such as environmental changes, nutritional science and agricultural practiceS. This growth is promising because metabolomics is able to discover novel and potential biomarkers that are associated with pathological states among thousands of endogenous and exogenous compounds. Hence, the medical applications of metabolomics are broad (Fig. 2.2B). Potential biomarkers carry information about the sites and mechanisms of disease pathogenesis so comparison of metabolic profiles between control and study group provide important information for our understanding on diseases. These biomarkers could be used for earlier disease diagnosis, better monitoring of disease progression and establishing predictive models for disease risks (Spratlin, Serkova, and Eckhardt 2009; Hocher and Adamski 2017; Merino et al. 2018; Srivastava and Creek 2019). Metabolomics are also applied in the pharmaceutical development field for evaluation of therapeutic responses, drug discovery and drug safety (Tolstikov 2016; Tuyiringire et al. 2018).

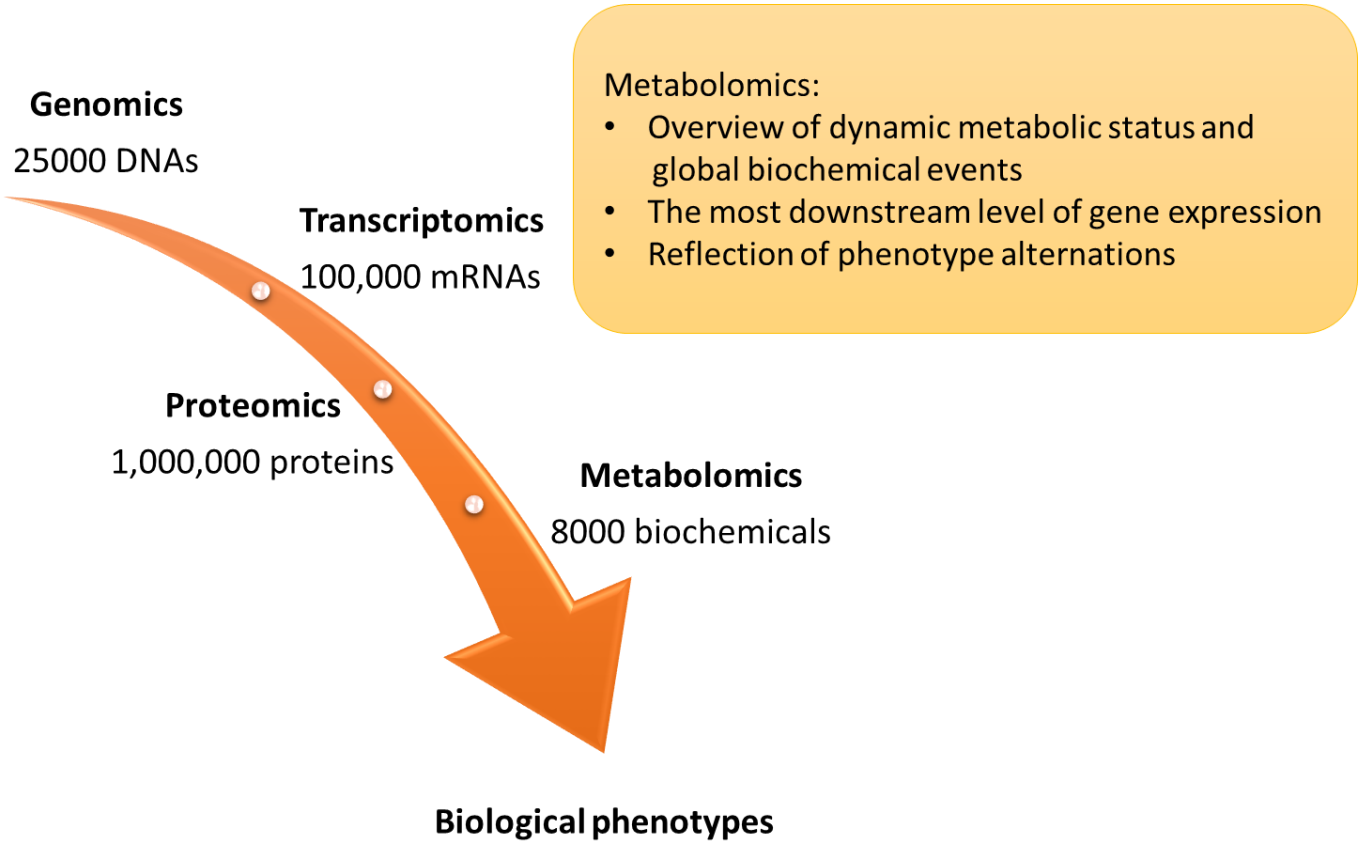
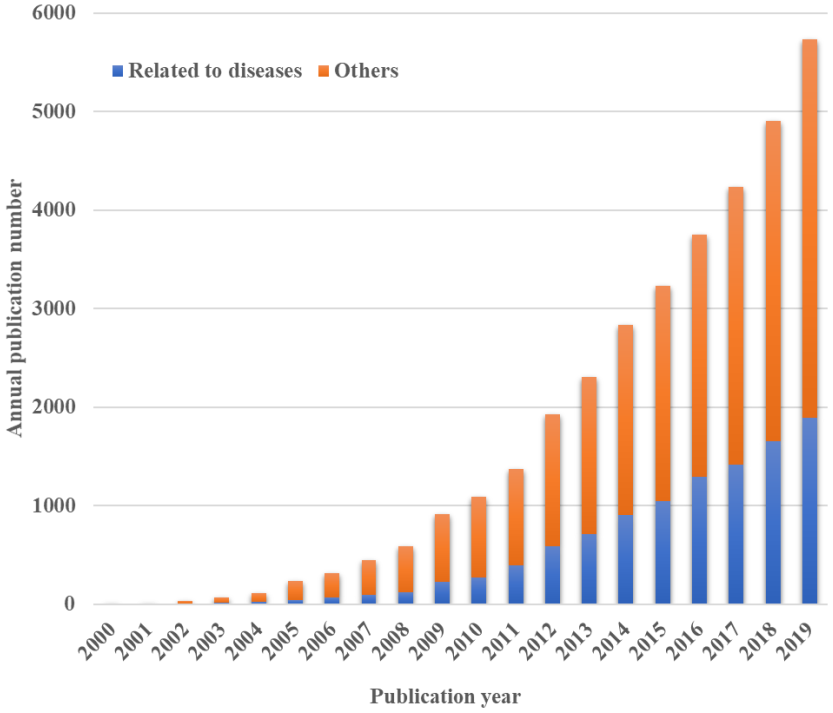


Fig. 2.1 Overview of the flow from DNA to phenotypes associated with corresponding omics from genomics to metabolomics.

This number in the figure was sourced from Aboud *et al*'s report (Aboud and Weiss 2013).

(A) Annual publication numbers related to metabolomics and diseases



(B) Application of metabolomics in diseases

Fig. 2.2 (A) Annual publication number with "metabolomics" and "metabolomics and diseases" as keywords using SciFinder (http://scifinder.cas.org) from 2000 to 2019. (B) A wide range of metabolomics applications in disease research.

2.2 Choices of metabolomics platforms

There are huge differences in chemical properties between different types of metabolites, like simple amino acids, hormones, neurotransmitters, vitamins, peptides, lipids and complex carbohydrates. Thus, no single analytical platform is capable of detecting the entire metabolome in biological samples (such as body fluids, cells and tissues). Many advanced analytical techniques are used for the analysis of these complex mixtures. The two most used platforms for metabolic profiling and quantification are mass spectrometry (MS) and nuclear magnetic resonance (NMR).

NMR spectroscopy is one of the major analytical techniques used for the analysis of multi-component mixtures in metabolomics research; it has well-known advantages including little requirement on sample preparation such as derivatization and separation, rapid, non-invasive and non-destructive sample treatment, relatively high reproducibility [coefficient of variation (CV) ~ 1-2 %] , easy identification of metabolites (Markley et al. 2017; Gowda et al. 2008; Emwas 2015). Yet, NMR spectroscopy has a key weakness – its relatively low sensitivity which means only around 100 metabolites could be detected in urine (Connor et al. 2004) and even fewer in serum and tissue samples (Rooney et al. 2003).

Compared with NMR spectroscopy, MS methods coupled with prior specific sample preparation and separation methods, such as liquid chromatography (LC) and gas chromatography (GC), can deal with complex mixtures and allow detection of more than 1000 metabolites at a time (Emwas 2015). Hence, MS methods have higher sensitivity and specificity (Dettmer, Aronov, and Hammock 2007). Tandem MS (MS/MS) methods are often used to validate the identity of unknown molecules. The introduction of Orbitrap MS allows detection of high-resolution signals with very high

mass accuracy (Barbier Saint Hilaire et al. 2018) for easy identification of metabolites. Thus, MS was chosen as the analytical platform for metabolomics study in this project.

2.3 Targeted and untargeted metabolomics

Acquisition of metabolomics data by MS generally falls into two categories: targeted metabolomics and untargeted (global) metabolomics.

Targeted metabolomics focus on quantitative analysis in one or a few classes of metabolites. The targeted compounds are usually quantified in an absolute manner using calibration curves and/or stable-isotope-labeled internal standards (Wijeyesekera et al. 2012; Fernández-Peralbo et al. 2014). Since previous understanding of intermediates and metabolites involved in the biological function or metabolic pathway is required and this becomes the major limitation of targeted methods (Roberts et al. 2012). Metabolites detected are largely restricted to only known metabolites and limited classes of compounds can be observed in single analysis. Metabolites should have its own stable and co-eluted internal standards. This largely depends on the availability of stable-isotope-labeled internal standards which are usually much more expensive than the analytes. If these standards are not available, only similar analogues can be employed, which may not have exactly the same ionization performance as the analytes. Targeted approaches could serve as a validation of results revealed from untargeted analysis, which was applied in the study of Chapter 3.

Untargeted metabolomics is most commonly used for global metabolic profiling of samples, including molecules not reported previously using full-scan mass spectrometry (Roberts et al. 2012). Thus, thousands of signals would be acquired in an analysis. Advanced chemometric techniques are needed to process and mine the large data sets to screen out unstable and insignificant signals. Given no prior assumption on

the identity of possible biomarkers, comparison between groups receiving different treatments in the study may reveal novel biomarkers or pathways that are ignored or not known in routine biochemical tests. The mechanisms of disease development and drug actions may also be unveiled. Untargeted metabolomics are always obstructed by issues such as instrumental instability, complicated data treatment and unknown peak identification. This requires experienced researchers and extra time to handle the data analysis.

2.4 Workflow of mass spectrometry-based untargeted metabolomics

Untargeted metabolomics comprises numerous steps that require careful design. There are many protocols targeted to certain classes of metabolites available for public but no single protocol would be applicable for all applications. To get fruitful and reliable results, the experimental protocol for each study needs to be cautiously designed and optimized, according to the hypothesis, sample type, metabolites of interest and analytical platform applied. An untargeted metabolomics workflow typically involves six steps (Alonso, Marsal, and Julià 2015). First is sample extraction of biological samples from the control and the experimental group. Second is the untargeted data acquisition with mass spectrometry. Third is spectral data processing. Fourth is multivariate statistical analysis to look for a list of candidates with statistically significant variations, followed by the identification of the statistically significant candidates. The final step is biological interpretation of the identified metabolites to resolve the hypothesis and provide insights for the future study. Fig. 2.3 illustrates a typical workflow of an untargeted metabolomics.

More and more comprehensive software and packages are developed by the mass spectrometry communities that integrate data picking, alignment, data filtering,

multivariate statistical analysis and even pathway analysis of metabolomics data together on the same platform to make the tedious analysis more user friendly (Table 2.1). There are command-line, web-based and integrated software packages. Matlab, Python and R (Ozgur et al. 2017) are some common command-based programming languages with many open-source packages for data processing and multivariate statistical analysis but they require fundamental knowledges of programming and statistic.

Some free user-friendly web- and graphical interface programs (Katajamaa and Oresic 2007) can also achieve multiple data treatment and statistical analysis procedures, such as XCMS online (Tautenhahn et al. 2012; Huan et al. 2017),MetaboAnalyst 4.0 (Chong et al. 2018), Automated Mass Spectral Deconvolution and Identification System (AMDIS) (Behrends, Tredwell, and Bundy 2011), MZmine2 (Pluskal et al. 2010), MetAlign (Lommen and Kools 2012). Some commercially available tools are designed by instrument vendors, such as Markerlynx Application Manager (Waters), Mass Profiler Professional Software (Agilent) and Compound Discoverer software (Thermo orbitrap). These tools already come with some standard multivariate techniques. Progenesis QI for metabolomics (Non-linear Dynamics, Waters) is another commercial software that is compatible for many high-resolution MS such as Bruker, Agilent, Waters and Thermo mass spectral data. Progenesis QI integrates general key data processing steps in one platform, including peak alignment (Fig. 2.4), peak peaking, normalization, comparison of classes and identifications of significantly changed features from multiple built-in databases. These user-friendly platforms speed up the whole metabolomics data analysis and tedious peak identification process. Table 2.1 is a summary of the mentioned graphical user-interface software tools for data-processing and/or analysis of metabolomics data.

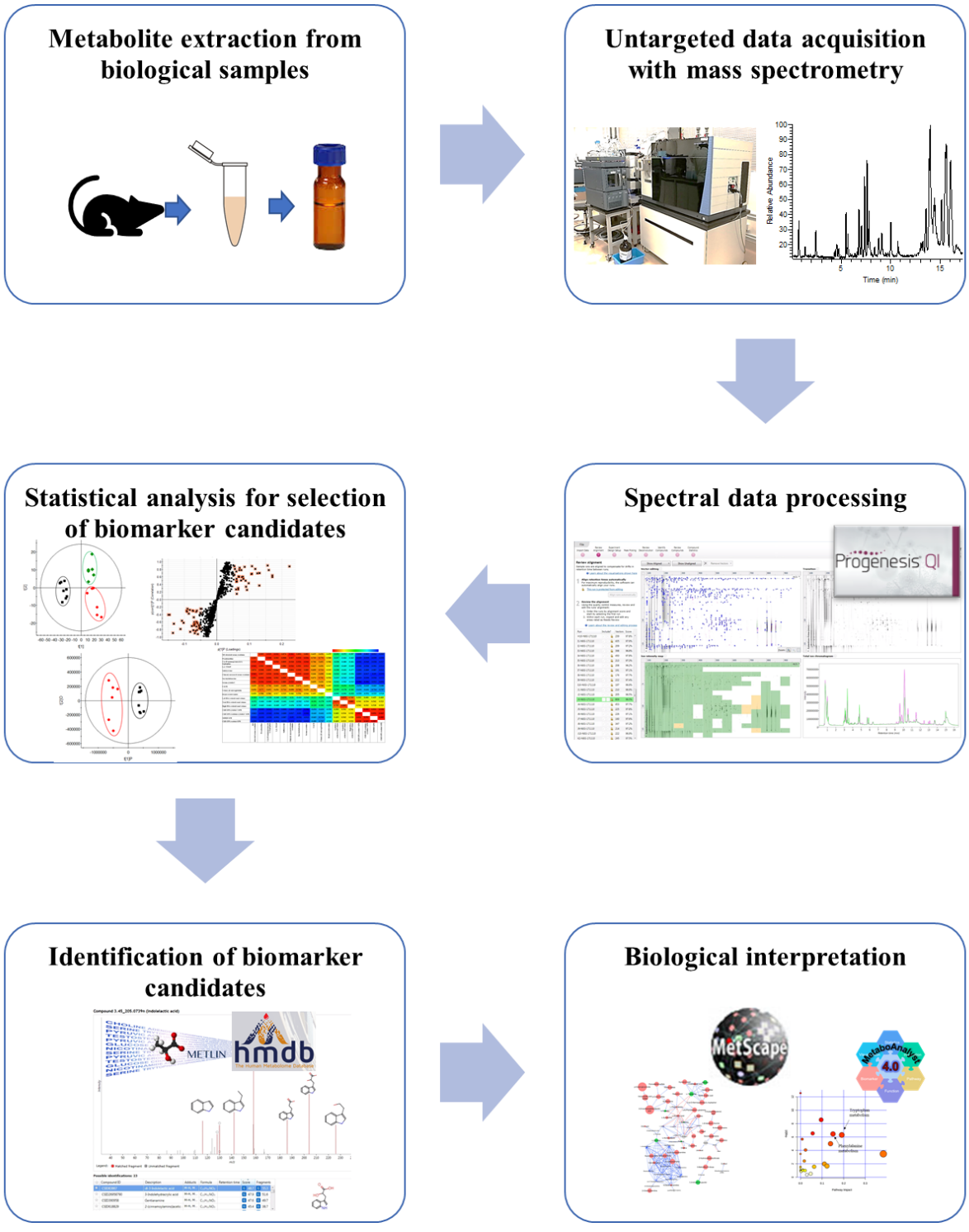


Fig. 2.3 A typical workflow of an untargeted metabolomics.

Table 2.1 Graphical user-interface software tools for data-preprocessing and/or analysis of metabolomics data

Software	Instrument data type	Interface	Webpage	Related references
AMDIS	GC–MS (preprocessing only)	Windows graphical user-interface	http://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:amdis	(Behrends, Tredwell, and Bundy 2011; Meyer, Peters, and Maurer 2010)
MZmine2	LC–MS, GC-MS (universal)	Windows graphical user-interface	http://mzmine.github.io/	(Pluskal et al. 2010)
MetAlign	LC–MS, GC–MS (universal)	Windows graphical user-interface	https://www.wur.nl/en/show/MetAlign.htm	(Lommen and Kools 2012)
XCMS Online	LC–MS, GC–MS (universal)	Web-based interface	https://xcmsonline.scripps.edu	(Tautenhahn et al. 2012; Huan et al. 2017)
MetaboAnalyst 4.0	LC–MS, GC–MS, NMR (universal)	Web-based interface	http://www.metaboanalyst.ca	(Chong et al. 2018)
Markerlynx Application Manager	LC-MS, GC-MS (Waters)	Windows graphical user-interface	Discontinued and replaced by Progenesis QI	(Stumpf and Goshawk 2004)
Mass Profiler Professional Software	LC-MS, GC-MS (Agilent)	Windows graphical user-interface	https://www.agilent.com/en/products/software-informatics/mass-spectrometry-software/data-analysis/mass-profiler-professional-software	(Technologies 2016)
Compound Discoverer software	LC–MS (Thermo orbitrap)	Windows graphical user-interface	https://www.thermofisher.com/hk/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/compound-discoverer-software.html	(Scientific 2019)
Progenesis QI for metabolomics	LC–MS (universal)	Windows graphical user-interface	http://www.nonlinear.com/progenesis/qi/	(Corporation 2015)

Fig. 2.4 An example of alignment results of UPLC-QTOF-MS data using Progenesis QI for metabolomics UPLC-QTOF-MS data extraction software.

2.5 Sample preparation

No matter what kinds of samples or classes of metabolites are being studied, the choice of sample-preparation method are always highly important. An ideal sample preparation for global metabolic profiling would be non-selective extraction with maximum coverage of metabolites, simple and fast handling, reproducible with minimum variability, simultaneous and fast quenching of all possible ongoing metabolic processes inside the samples (Dunn et al. 2011).

Many sample preparation protocols are available as guidelines for different sample types. For instance, the UCSD Metabolomics Workbench (https://www.metabolomicsworkbench.org/protocols/) and nature protocols provide general procedures, containers and chemicals for handling various biological samples such as urine (Want et al. 2010), blood (Dunn et al. 2011) and tissue (Want et al. 2013) as well as extraction of specific classes of metabolites.

A typical workflow of sample preparation for metabolomics analysis would be as follows: sample collection, homogenization (solid samples), deproteination, extract isolation, derivatization (GC-MS), lyophilization/concentration and finally reconstitution prior to injection for instrumental analysis. Before the start of sample preparation, containers, chemicals and instruments should be free of contamination and interference (Keller et al. 2008), which could be tested using blank sample such as water to go through the entire sample extraction procedures. Section 2.1.2 explained more details. This step is crucial for minimizing external contaminations.

Metabolism quenching is required after sample collection. It means a quick way to stop all cell metabolic process using cooling (cold solvent addition or freezing in liquid nitrogen) or addition of buffer (Vuckovic 2012; Pinu, Villas-Boas, and Aggio 2017).

Since the contents of metabolites in the samples may change over time very rapidly by the action of enzymes or degradation process at room temperature, the concentration detected may be far from the actual metabolic state in the body. To minimize the degradation of metabolites over time, quenching at the same time and storage of all samples at -80°C before sample preparation or injection for analysis are required. Freeze-thaw cycles should also be minimized to reduce the effect on the stability of the metabolites at different temperatures (Hirayama et al. 2015; Torell et al. 2017). This can be avoided by making smaller aliquots from the samples during sample collection. In addition, during the extract isolation and removal of solid debris by centrifugation, the Endometriosis Phenome and Biobanking Harmonisation Project SOPs recommended cooled (4℃) centrifugation as standard to avoid any effect of temperature on stability of metabolites (Rahmioglu et al. 2014).

Deproteination and removal of protein are necessary when working on all biological samples to avoid damage to the analytical column and MS capillaries. Protein precipitation by organic solvent (commonly methanol or acetonitrile), followed by centrifugation and/or ultra-filtration (10,000-14000 rpm) are common practice (Gu et al. 2010; Chen et al. 2009). Chloroform (Bird et al. 2013) and the less toxic methyl tert-butyl ether (Pizarro et al. 2013) are common solvents for lipid extraction.

Preparation of solid biological samples, such as muscle, liver and plant tissue, has a disadvantage of requiring solid homogenization for tissue and cell disruption to obtain intracellular metabolites. Freeze-drying of fresh biological samples (Kim et al. 2013) is recommended prior to homogenization for better homogeneity, repeatability and extraction capabilities (Courant et al. 2014). A bead-based homogenizer in combination with aqueous organic solvents and a simple extraction protocol (Römisch-

Margl et al. 2011) can reduce labour-intensive steps compared with manual homogenization.

GC-MS metabolomics requires a derivatization step to convert polar metabolites into volatile and thermally stable derivatives for GC-MS analysis because many primary metabolites, such as amino acids, fatty acids and steroids, are water or alcohol soluble. They usually have polar functional groups, like hydroxyl-, carboxyl- and thiol-groups which result in high boiling points. The procedures reported by Sánchez-Avila *et al* and Yi *et al* (Yi et al. 2006; Sánchez-Avila et al. 2009) suggested a two-step derivatization method for profiling analysis of fatty acids without removal of protein. It is composed of two steps: (1) base-catalyzed trans-esterification for catalyze esterified fatty acids (EFAs) such as cholesterol esters, acylcarnitines and other O-acyl esters into fatty-acid methyl esters; (2) acid-catalyzed esterification for removal of the carboxyl-groups of non-esterified fatty acids (NEFAs). Another popular derivatization protocol uses silylation to replace the reactive hydrogen atom, such as those in hydroxyl groups, carboxyl groups and enolizable ketone, with a silyl group. The most frequently used silylation agents are N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) and N,O-bis (trimethylsilyl)trifluoroacetamide (BSTFA) (Moldoveanu and David 2018) which replace reactive hydrogen atom with trimethylsilyl group (Fiehn et al. 2010; Kumari et al. 2011). Hence, the sample preparation process of GC-MS requires special care and more time (Garcia and Barbas 2011; Papadimitropoulos et al. 2018) than that of LC-MS. However, compared with LC-MS, the benefits of GC-MS in a small scale study may outweigh its drawbacks with a better separation efficiency of derivatized small molecules, higher reproducibility of generated mass spectra, broad electron impact spectral libraries, making identification of peaks easier.

2.6 Chromatography

Chromatography separates the metabolites in the samples. A mixture is dissolved in a fluid, called the mobile phase [liquid state in liquid chromatography (LC) and gas state in Gas chromatography (GC)]. The mobile phase carries the metabolites to travel through a material, called the stationary phase, which exhibits different affinity to differentiate molecules and results in different retention. GC and LC are two common chromatographic techniques used in separating the complex metabolites before MS analysis. Most biological metabolites are polar and non-volatile, therefore derivatization of metabolites into volatile and thermally stable derivates are needed before GC-MS analysis. The metabolite derivatives with lower boiling points and lower molecular sizes are eluted earlier with the carrier gases. The metabolite derivatives with characteristic and reproducible mass fragmentation patterns generally allow easier identification against libraries in GC-MS, when compared with LC-MS.

Compared to conventional high performance liquid chromatography (HPLC), ultra-high performance liquid chromatography (UPLC) uses sub-2 micron stationary phase particles and a lower mobile-phase flow rate is able to generate a chromatogram with higher resolution and sensitivity in a much shorter time and less solvent consumption (Theodoridis, Gika, and Wilson 2008; Wang 2009).

Regarding UPLC column, a variety of LC columns are available to facilitate the analysis of a range of metabolites with different polarities. The most common separation mode for LC-MS based metabolomics studies is the reversed-phase (RP) chromatographic separation within analytes coated carbon chain (especially C18) columns (Zhao et al. 2010; Ma et al. 2014). This is especially suitable for medium to weakly polar metabolites. Polar metabolites can be analyzed by hydrophilic liquid

chromatography (HILIC) such as the use of hydrophilic amide-based stationary phases (Wen et al. 2019), which is capable of improving the poor separation of polar amino acids. High strength silica (HSS) column with 100% silica-based particle, which is compatible with 100% aqueous mobile phase but not in RP column, have recently been applied in both polar and non-polar compounds in metabolomics (Liu, Peng, Jia, Zhang, et al. 2014; Guo et al. 2014; Inoue et al. 2013; Wu et al. 2014). When separating polar and non-polar compounds, HSS column can increase polar compound retention and significantly improve their separation efficiency (McDonald et al. 2006). Thus, it allows simultaneous separation of both types of metabolites in one column to get the comprehensive metabolic profiling.

2.7 Mass spectrometry

A typical mass spectrometer is composed of three basic parts: (1) an ionization source, (2) a mass analyzer, and (3) an ion detector. The ionization source and mass analyzers are the most critical in accurate mass spectral measurements.

2.7.1 Ionization source

Atmospheric pressure ionization techniques such as electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) (Mitchum and Korfmacher 1983) are the techniques most employed in LC-MS analysis (Huang et al. 2011). ESI is a soft ionization technique in which collision occurs between collision gas and molecules of analytes, resulting in gas phase ions without fragmentation. Negative ionization is particularly efficient for deprotonation (the removal of a proton) of certain classes of compounds with hydrogen bonds (e.g. fatty acids and polyphenols). ESI in positive mode can effectively protonate (add a proton) a wide range of metabolites from medium to highly polar molecules (Theodoridis et al. 2012) such as acylcarnitines, phosphatidylcholine and lysophosphatidylcholine. Neutral lipids that are not easily

ionized by ESI and they are usually present as sodium, ammonium or lithium adducts with ESI in positive mode (Dettmer, Aronov, and Hammock 2007). APCI is preferred for more non-polar metabolites such as sterols, triacylglycerols and pesticides (Christie and Han 2012; Matić et al. 2014; Glauner and Zavitsanos 2012) but it has not yet been widely employed in metabolomics investigations due to its low sensitivity and substantial in-source fragmentation (Christie and Han 2012). For GC-MS, electron impact ionization is the most typical. Energetic electrons are used to bombard the analytes in gaseous state, resulting in large characteristic and reproducible molecular fragments. This allows easier determination of their identity (Portolés et al. 2014; Li et al. 2015).

2.7.2 Mass analyser

A mass analyzer separates ions from the ion source based on their charge to mass ratios (m/z) and directs them to the detector. Different mass analysers have been utilized for ion separation, such as tandem quadrupole mass analysers (like triple quadrupole mass analysers, QQQ-MS) (Lame, Chambers, and Blatnik 2011), quadrupole-time of flight mass analysers (Q-TOF-MS) (Ponthus and Riches 2013), ion trap and Orbitrap mass analysers (Orbitrap-MS) (Makarov 2000; Eliuk and Makarov 2015). Among them, Q-TOF-MS and Orbitrap-MS mass analysers provide accurate mass measurement with much higher mass resolution than QQQ-MS and ion trap-MS. For mass fragmentation, Q-TOF-MS/MS uses the quadrupole as a scanning device to select precursor ions for fragmentation in the collision cell, producing product ions. The ions enter TOF-MS for determination of m/z by measuring their time of flight (velocity) in a known distance. The Q-TOF-MS has high-mass resolution (> 10,000) for the precursor ions to aid in metabolite characterization and identification (Stroh et al. 2007) and the dominant mass

analyser in non-targeted metabolic profiling (Wu et al. 2014; Zhang, Choi, et al. 2012; Wang, Feng, et al. 2011). Fig. 2.5 shows the schematic of Waters SYNAPT G2 HDMS mass spectrometer for Q-TOF-MS/MS (Waters Corporation, Manchester, UK) (Ponthus and Riches 2013). Recently, Orbitrap-MS analyser, an advanced ion-trap variant, is widely used in proteomics as well as metabolomics due to its high mass accuracy (2-5 ppm), much higher resolution (> 100,000 for classic analyser) (Lim et al. 2016) and large dynamic range (from 1 to 10,000 ng/mL). Orbitrap mass analyser is typically coupled with a linear ion trap to enable MS^3 determination of the fragmentation pattern of product ions of precursor ion, thereby providing more information for identification of unknown metabolites. Fig. 2.6 shows a schematic diagram of an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Eliuk and Makarov 2015).

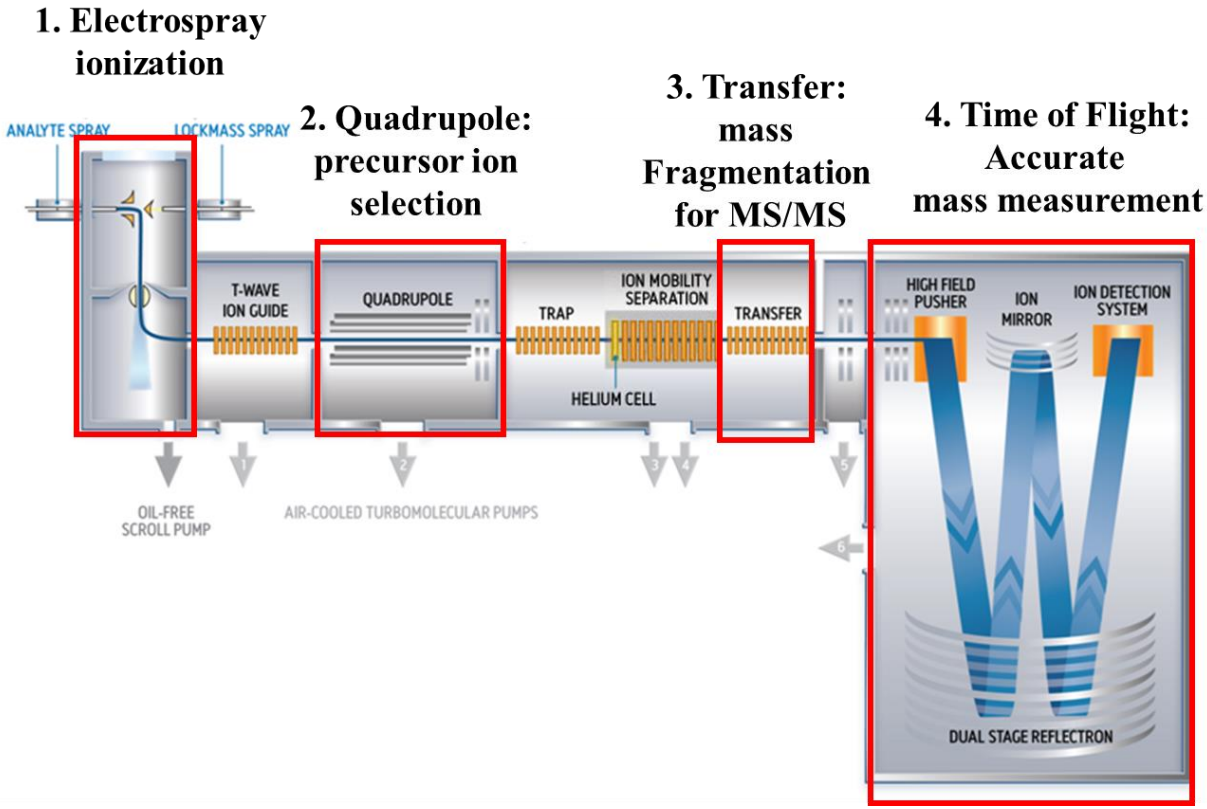


Fig. 2.5 A schematic diagram of the SYNAPT G2 HDMS mass spectrometer for Q-TOF-MS/MS modified from Ponthus *et al*'s paper (Ponthus and Riches 2013)

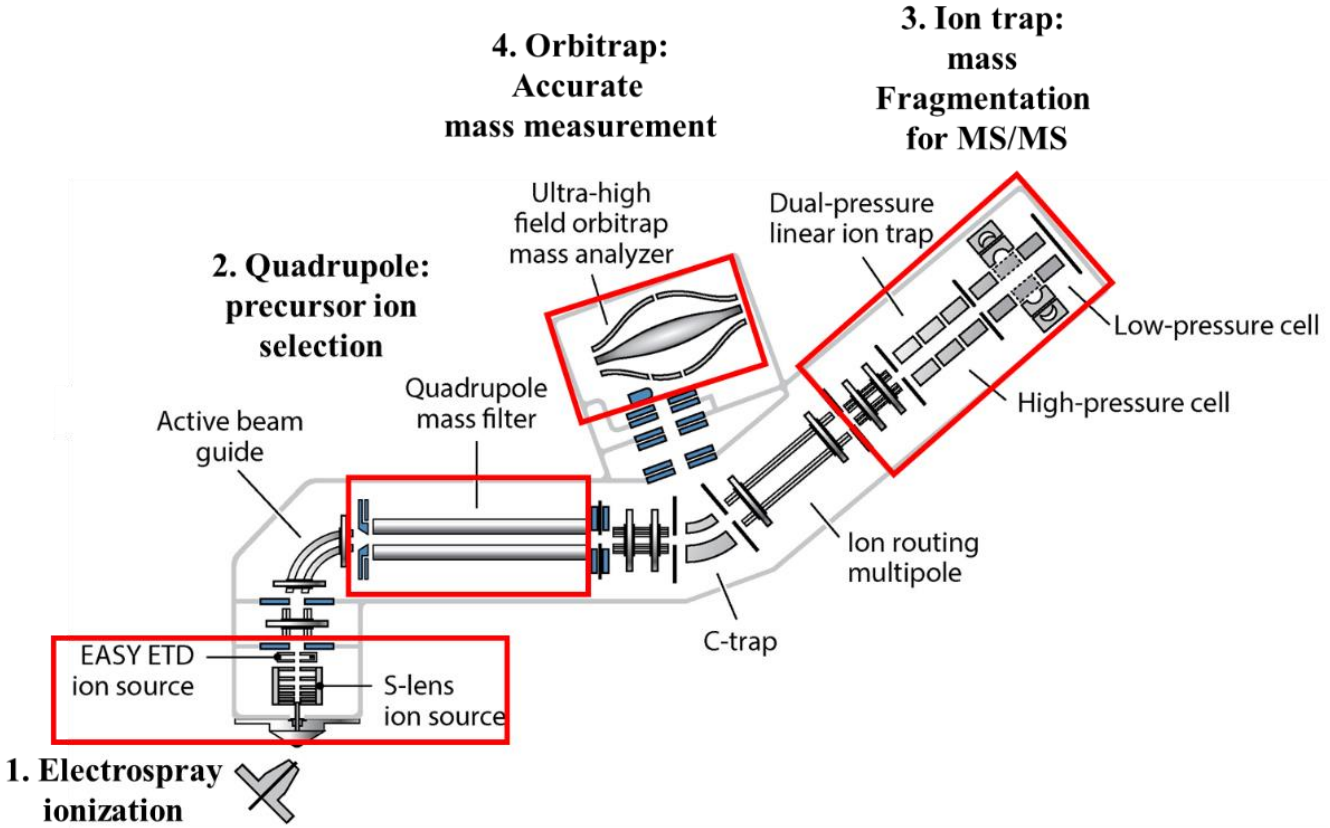


Fig. 2.6 A schematic diagram of an Orbitrap Fusion Lumos Tribrid Mass Spectrometer modified from Eliuk *et al*'s paper (Eliuk and Makarov 2015).

2.8 Multivariate statistical analyses

Univariate statistical analysis, a simple small scale data analysis, is the most widely used discrimination methods, such as Student's t-test and one-way analysis of variance (one-way ANOVA) test. Due to the comprehensive nature of the untargeted mass spectral data, a large data set (MS signals of the metabolites in this case) is usually resulted where the data matrix consists of thousands of variables in multiple classes of observations (samples). Their interpretation requires advanced multivariate statistical techniques, which have been applied in analytical chemistry application because of the development of chemometrics (Pinto 2017), instead of univariate test.

Chemometrics techniques are robust and efficient for converting complicated and correlated data matrices into more manageable, interpretable and reliable models. The chemometric models are capable of handling incomplete, noisy and collinear data structures (Trygg and Lundstedt 2007). There are two common types of chemometric techniques: linear projection-based algorithms (Worley and Powers 2013), such as Principal Components Analysis (PCA), supervised partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA), and non-linear machine learning methods (Mendez, Reinke, and Broadhurst 2019), like random forest (RF) modelling, support vector machines, genetic algorithms and artificial neural networks.

PCA is the most common chemometric method for unsupervised pattern recognition. It reduces the dimension of datasets into a few uncorrelated principal components (Ringnér 2008; Jolliffe and Cadima 2016). A data matrix for PCA contains multiple variables of interests and observations of all classes. Every variable is considered to be a coordinate axis (a dimension) and each observation is a point in the high dimensional

variable space (Fig. 2.7A). The first principal component is the axis (linear combination of all variables) that maximizes the variance of the data projected on that axis. Then, the second principal component is the axis orthogonal to the first principal component which maximizes the variance of the data projected on this axis. PCA rotates the coordinates of the high dimensional space (each axis is the content of a different metabolite) to a coordinate system according to the projected variance. Thus, the first few components are usually sufficient to capture most of the relevant variance in the data set so that we can focus on a reduced dimension and interpret the output of these components. This enhances the ease of interpretability and minimizes information loss. The primary goal of PCA in metabolomics is to explore overall interrelations and trends among all classes of observations and identify outliers (2002). Large-score potential outliers are visualized in 2D and 3D score plots of PCA. If outliers are confirmed after data inspection, they should be corrected or removed to avoid distortion of the robustness of the subsequent analysis (Bro and Smilde 2014). Hotelling's T^2 test is a common method that is a generalization of the Student's t-test to multivariate data (Bro and Smilde 2014) and always used for automatic outlier detection in the multivariate Extended Statistical tool (EZinfo Version 2.0 software, Umetrics AB). Fig. 2.7B demonstrates the distribution of metabolic profile of rat samples in the PCA score plot of UPLC-MS datasets. In the figure, one sample (bottom) is placed outside the Hotelling's T^2 range with 95% confidence level, indicating over 95% probability that the sample behaves differently from the other. Large-scale studies may include data from quality control (QC) samples in the PCA to give an overview of the variability of data in the score plot. Highly clustered QC samples shown in the PCA score plot (Fig. 2.7B) would indicate good stability and high repeatability of the analytical platforms throughout the entire experiment.

PLS-DA and OPLS-DA are two most widely used methods for discriminant analyses (Triba et al. 2015). PLS-DA (Sjöström, Wold, and Söderström 1986) is a supervised pattern recognition technique and a linear regression model, placing every observation into its own pre-assigned class, rotating PCA components into the maximum separation between classes (Barker and Rayens 2003; Eriksson et al. 2013). PLS-DA can be applied for evaluating class discrimination and selecting significant variables responsible to the separation of classes. A loading plot of PLS-DA summarizes the distribution of potential biomarkers contributing the most in the pre-assigned classification systems (Theodoridis et al. 2012). The biomarkers (X-variables) should co-locate away from the origin in the loading plot, indicating they are more correlated to each other and to the group with similar positions in the score plot (Worley and Powers 2013). In addition, candidates with the highest contribution in the separation can be assessed based on their Variable Importance in the Projection (VIP) values (the common threshold of VIP $\geq$ 1). Metabolites with larger loading and VIP values have greater importance for the differentiation among groups. The data analysed in PCA excluding QC samples and outliers are used for PLS-DA for further understanding between their differences.

OPLS-DA (Bylesjo et al. 2006) is an extension to the supervised PLS-DA. Similar to PLS-DA, OPLS-DA can be used to identify variables with the most discrimination power between groups. OPLS decomposes the PLS solution into components that are predictive (correlated) to Y response matrix (diagnostic outcome), and components that are systematic in the X descriptor matrix (signal intensity) and orthogonal (uncorrelated) to Y matrix (Trygg and Wold 2002). OPLS-DA maximizes the separation between discrete classes. Two classes are easier in statistical interpretation, such as gender, treatment group vs placebo group. The predictive value of binary

classes, Y, of an OPLS-DA model is between 0 and 1 depending on their class belonging. The variables that make the main contributions to group separation are placed in the very first component, the predictive component (Kang et al. 2008). The horizontal dimension of the score plot of OPLS-DA captures variation between classes while the vertical dimension captures variation within each class. Putative biomarkers can be identified by loading plots and VIP values. S-loading plot (S stands for sigmoid shape of the loading distribution) of OPLS-DA is dedicated in the Umetrics design softwares, such as SIMCA-P and the Extended statistic tool, and only applied in two classes. It shows the correlation arrangement and covariance between the metabolites and the predictive classes (Wiklund et al. 2008). In the S-loading plot, variables with larger absolute values of p(1) (magnitude of each variable) and p(corr) (reliability of each variable in X matrix) are the most predictive candidates because of their high contribution to the variation of the samples and correlation in the dataset while those closer to the origin have high risk for false correlation .

Due to the supervised modelling nature of PLS-DA and OPLS-DA, cross validation is more crucial than unsupervised PCA for prevention of model over-fitting (Worley and Powers 2013). Standard cross validation picks up part of samples from the data matrix for training and uses the remaining samples from the same data matrix for testing to assess the predictive accuracy of the model established using the training data only (Wu et al. 2010). There are different cross validation strategies such as Monte Carlo cross-validation (Xu and Liang 2001; Xu, Liang, and Du 2004), Leave-One-Out cross validation and permutation test. Leave-One-Out cross validation has long been incorporated in the widely used user-friendly multivariate statistical tools, SIMCA-P (Umetrics AB, Umeå Sweden) and Extended Statistical tool (Umetrics AB, Umeå Sweden) for automatic estimation of predictiveness of the models (Triba et al. 2015).

They use 7-fold Leave-One-Out cross-validation that split the dataset into seven subsets. Each subset is removed in turn. A new submodel built with the six other subsets (calibration subset) predicts the response Y values of all observations in the left-out subset (Eriksson et al. 2006). The prediction is repeated with each subset until all data are predicted. The predicted data are then compared with the actual values to get the sum of squared errors calculated for the whole dataset. The Predicted Residual Sum of Squares is converted to Q^2. The cumulative value of Q^2 from all principal components estimates the predictive accuracy between the predicted and original data. R^2X and R^2Y are the fractions of sum of squares of entire X-variables and entire Y-variables explained by the current component, respectively (Umetrics 2012). R^2X is separated into predictive and orthogonal (systematic) variation in X in the OPLS-DA. Cumulative values of vectors, R^2X_{cum} and R^2Y_{cum} are the cumulative sum of variances explained by the extracted components for estimation of the goodness of fit of the model.

A qualitative measure should have the cumulative $Q^2 > 0.5$ and the difference between Q^2 and R^2Y should not be larger than 0.3 (Eriksson et al. 2013) to avoid over-fitting. Fig. 2.8 demonstrates the 2D PLS-DA and OPLS-DA score plots with the results of cross validation generated by Extended Statistical tool.

.

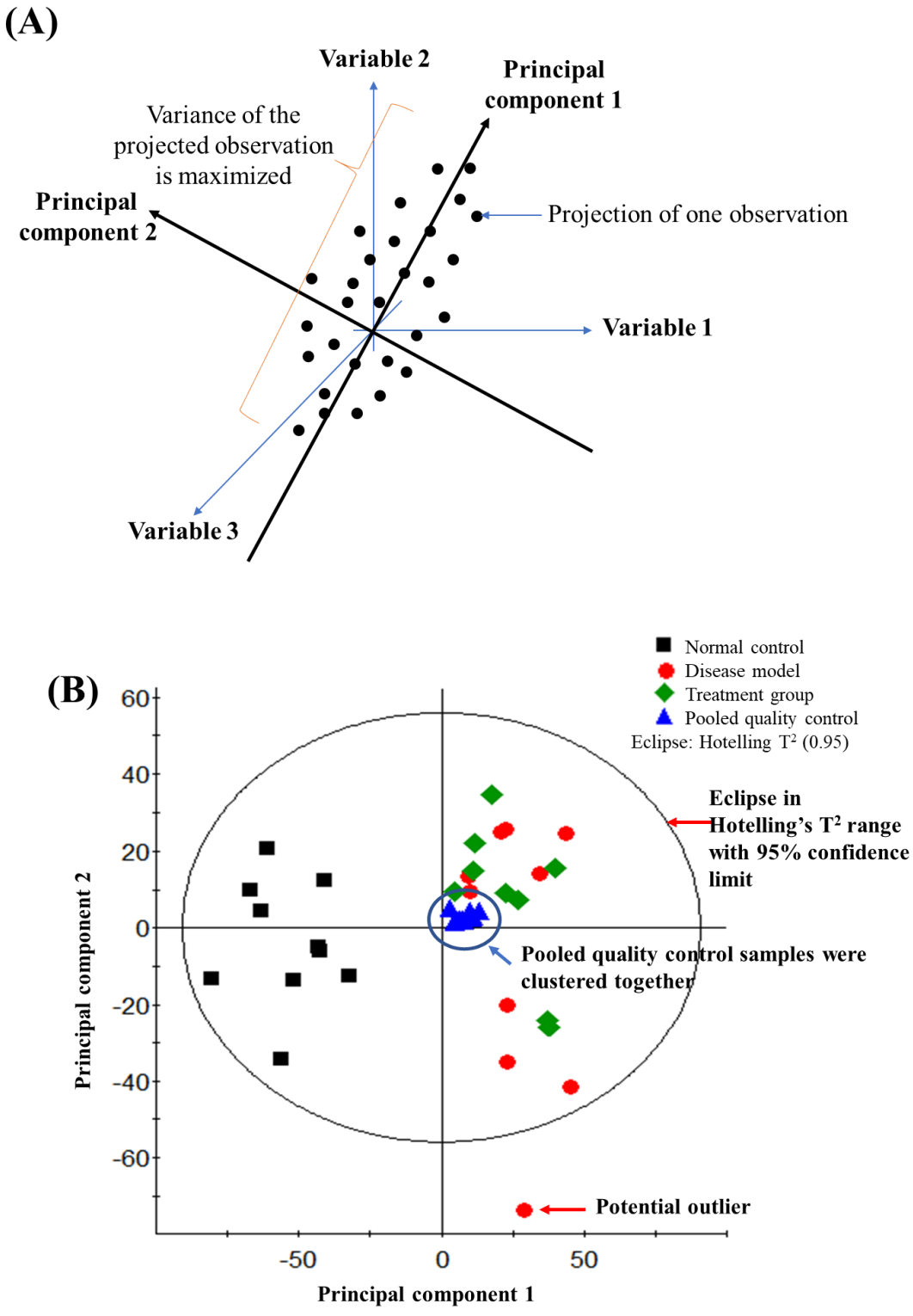


Fig. 2.7 (A) Graphical representation of the two principal components through a swarm of data points (B) 2D PCA score plot of UPLC-MS datasets from three rat groups and a pooled quality control group

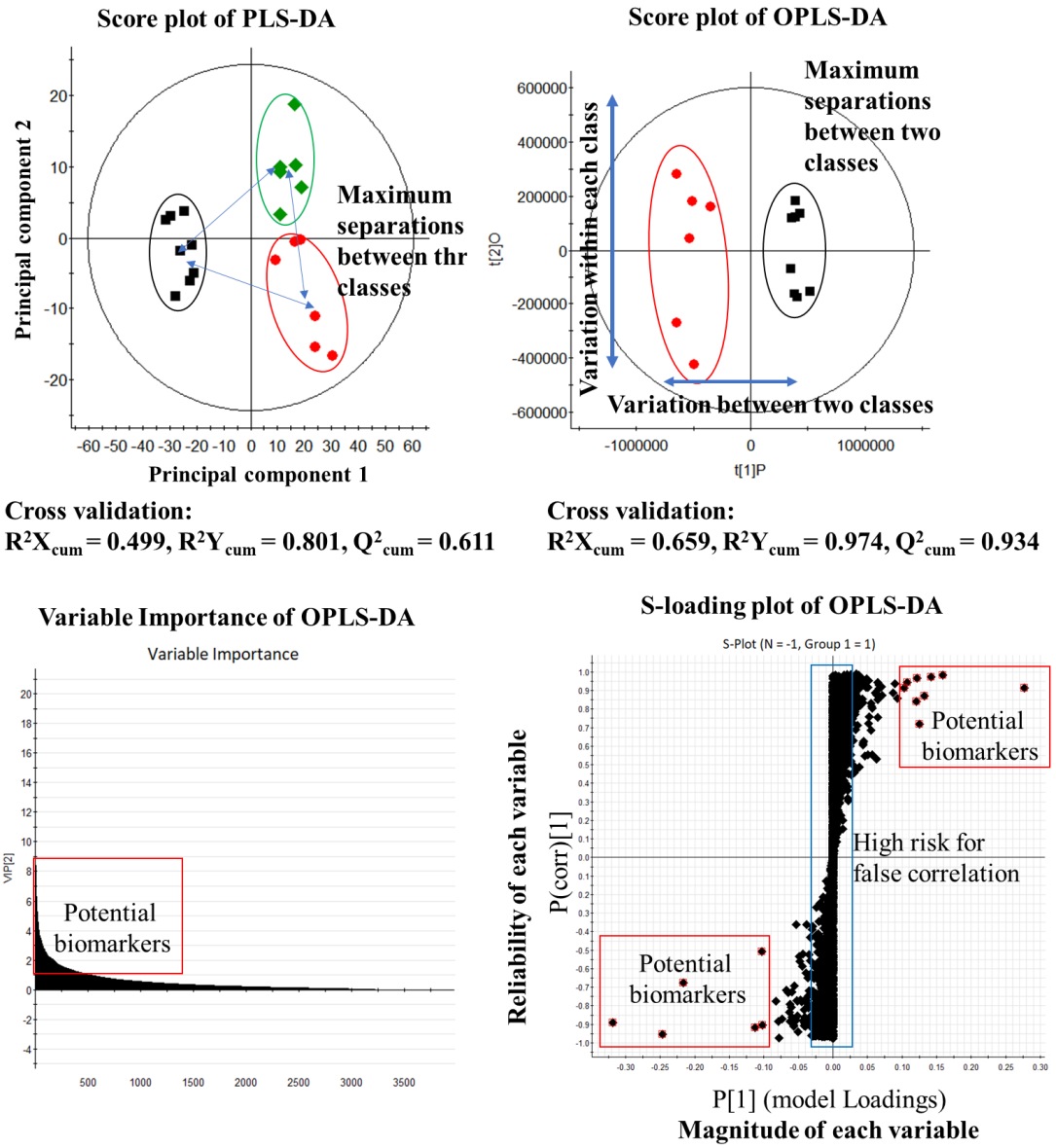


Fig. 2.8 Examples of 2D PLS-DA score plots, 2D OPLS-DA score plots, Variable Importance, S-loading plots of OPLS-DA of UPLC-MS datasets from rat groups.

Random forest (RF) (Breiman 2001) is a supervised non-parametric classification algorithm designed to identify variables that can classify different samples in a large dataset into various classes. Based on a bootstrap random sampling with replacement from original samples for the training of each single tree and about one-third of the samples being left out of the sample for testing, RF builds multiple independent decision trees (each tree is grown based on integrated rules) and merges them together to get an accurate and stable prediction (Boulesteix et al. 2012) (Fig. 2.9). For non-parametric datasets that have non-linear relationship to outcomes, RF provides a high prediction accuracy by attempting to find non-linear patterns and important features in metabolites that can explain variation in a given outcome (Chen et al. 2013). It has a relatively good tolerance for outliers (removed from the trees) and prevents over-fitting problem due to the uncorrelation of trees and combination of many single trees. Another advantage of RF also is that it does not require additional cross-validation because that is intrinsically tested. It gives an internal running out-of-bag (OOB) error estimate (Touw et al. 2012). The OOB error estimate is a method of measuring the proportion of times that the prediction result is inaccurate, averaged over all samples and is used to get estimates of variable importance. The Receiver Operating Characteristic (ROC) curve (Metz 1978) is also another measure of predictive accuracy of RF for binary classification (Calle et al. 2011). The curve is plotted with the False Positive Rate (1-specificity) on the x-axis and the True Positive Rate (sensitivity) on the y-axis. The Area Under the ROC Curve (AUC) is an overall index of diagnostic accuracy between two classes. AUC values > 0.5 means the classifier has a good measure of class separability, better than chance (Janitza, Strobl, and Boulesteix 2013)

.

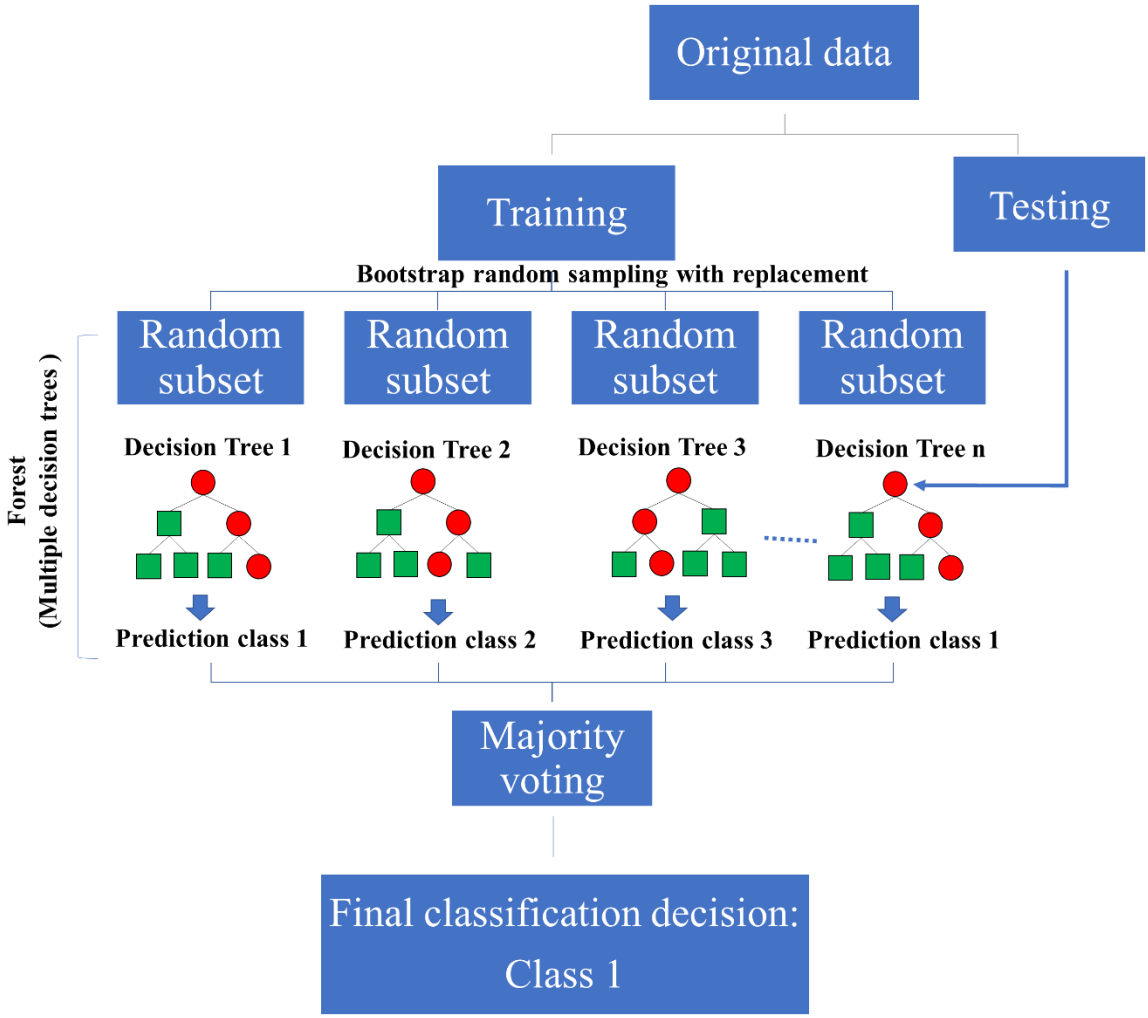


Fig. 2.9 A simplified scheme of a Random Forest Classification.

This scheme was sourced and modified from literatures (Li, Yu, et al. 2017; Harris and Grunsky 2015).

2.9 Peak identification

Peak identification is a key process in untargeted metabolomics for annotating a list of m/z ratios before processing to interpretation of biological findings. Generally, the interested peaks that are significantly altered are compound of interests for identification by matching molecular ions against the metabolite database. If acquired in high-resolution mass spectrometry (Q-TOF and Orbitrap mass analyzers), 5 ppm is the mass error tolerance between theoretical and experimental m/z of parent ions. Only those with < 5 ppm mass error would be matched with the MS/MS spectral similarity of possible candidates. Retention index is always used for GC-MS. Differences in the design of ionization source and mass fragmentation environment will bring about slightly variation in mass spectra. Common online spectral databases (Blaženović et al. 2018) are Chemspider (https://www.chemspider.com), Human Metabolome Database (HMDB, http://www.hmdb.ca/), Metlin (http://metlin.scripps.edu/index.php), Massbank (http://www.massbank.jp), mzCloud Mass Spectral Database (https://www.mzcloud.org/), Lipidmaps (http://www.lipidmaps.org) and NIST (https://www.nist.gov/). Among them, HMDB is a freely available online database that contains broad information about small molecule metabolites found in various biological sources and that is hyperlinked to other databases such as KEGG and PubMed (Wishart et al. 2013). NIST is a commercially available GC-MS library.

Currently, four levels of metabolite identifications with different levels of evidence and certainty have been proposed by Metabolomics Standards Initiative (Sumner et al. 2007) (Table 2.1). Most metabolomics studies in peer-reviewed journals did not define the level of identification (Salek et al. 2013) but defining metabolites as identified or putatively identified is highly important to provide clarity for sharing purpose.

Table 2.2 Four levels of metabolite identifications initiated by Metabolomics Standards Initiative

Levels	Annotation of identification	Criteria (Sumner et al. 2007; Viant et al. 2017; Blaženović et al. 2018)
Level 1	Identified compounds	Two or more independent and orthogonal data relative to an authentic chemical standard analysed under the same experimental conditions to validate non-novel metabolite identifications (such as retention time and MS/MS, retention index and MS/MS, retention time and NMR spectrum)
Level 2	Putatively annotated compounds	No chemical reference standards are available. A probable structure is based on physicochemical properties and/or MS/MS spectral similarity with database and/or literatures
Level 3	Tentatively characterized compound classes (isomers, substance class or substructure match)	Possible structure or class is based on characteristic physicochemical properties of a chemical class of compounds, or by similarity in parent ion to known compounds of a chemical class
Level 4	Unknown compounds of interest	Despite unidentified or unclassified status, these metabolites still can be differentiated and quantified based on reproducible spectral data.

2.10 Biological interpretation

Biological interpretation is essential to delineate the role of interested metabolites identified in the study. If only a few were identified, the interpretation of single metabolite can be achieved by HMDB, which provides related biochemical background and literatures. When a cluster of metabolites could be identified, metabolic pathway analysis could match relevant metabolites that participate in the same cellular signaling or related pathways in the metabolic network. This provides systematic insight into the underlying biological process that these altered metabolites may induce. Examples of biochemical pathway analysis for identified metabolites are KEGG database (https://www.genome.jp/kegg/), Integrated Molecular Pathway Level Analysis (IMPaLA) (http://impala.molgen.mpg.de/), Metaboanalysts (https://www.metaboanalyst.ca), Metascape plug-in of Cytoscape (https://apps.cytoscape.org/apps/metscape) and Ingenuity Pathway Analysis (IPA) (www.qiagen.com/ingenuity). Only IPA is a commercial pathway-based database and mainly for genomics while the others are free web-based databases for metabolomics. Apart from that, Metascape also has the advantage of checking the chemical structural similarity of the metabolites and allows networking metabolites with both chemical similarity or/and biological-relationship (Barupal et al. 2012).

2.11 Challenges and precautions in untargeted metabolomics studies

Metabolomics studies involve many procedures and corresponding challenges that require careful design and precautions (Fig. 2.10). Sample size and instrument stability are both important factors affecting the results of mass spectrometry based metabolomics studies

A large-scale study may improve the representation of samples in the population and statistical confidence (Dunn et al. 2011; Rizza et al. 2014) but it also makes sampling and storage of samples more challenging. Consistency in the duration and steps of sample handling minimizes day-to-day variation or degradation in samples. Sample preparation and injection in a randomized block design during sample preparation and instrumental analysis may also lower the technical bias (imbalance of sample sizes in each group in one batch). This may also reduce instrumental bias because contaminated MS source may lead to gradual changes in instrument sensitivity over time (Want et al. 2010; Dunn et al. 2012). Rigorous sample clean-up is generally not preferred for untargeted metabolite profiling due to the possible loss of some low-molecular weight compounds and reduction in precision. Thus, the development of sample preparation should aim at a simple and fast procedure that most metabolites are extractable and detectable.

Handling biological sample may have a risk of introducing external contaminations (Keller et al. 2008) into the samples. One frequent-detected external contaminant in sample tubes and reagents is polyethylene glycol (PEG). PEG is a water-soluble polyether compound and is a common surfactant used in containers (Weaver and Riley 2006) and detergents (Ahmadi and Winter 2018). It is hard to be isolated from the serum/plasma samples. This contaminant may ruin the dataset as their ions dominate in the mass spectrum and suppress positive ion signals of the target molecules. Fig. 2.11 shows the chromatograms and mass spectra of two solution (water) after filling two blank blood sample tubes. The figure shows how a series of PEG peaks (repeat signals of 44 Da) dominated in the chromatograms. Selection of containers and reagents before sample collection and preparation is important to minimize the extrinsic interference.

Instrumental instability is the major concern during the data acquisition of a mass spectrometer. During ionization, there is a direct contact of analytes to ionization sources that may lead to the gradual accumulation of contaminants and analytes on the surface of the ionization sources. These contaminants, especially lipids (Petković et al. 2001) may cause matrix effects (ion suppression and enhancement) that affect signal intensity (Mei et al. 2003; Antignac et al. 2005; Panuwet et al. 2016). The contamination may influence data collected in all subsequent analysis. There may be other sources of uncertainty, such as retention time drift due to column temperature variations and mass precision variations due to changes in room temperature and internal calibration issues.

A robust protocol should include QC samples that behave consistently throughout the analytical whole run. A typical batch injection sequence of samples consists of the consecutive analysis of one QC sample to equilibrate the analytical platform, followed by a QC sample set injected in every five to ten unknown samples (Lin et al. 2014). Intermittent monitoring of the precision and accuracy of signal from QC samples that are inserted between sample injections and exposing to the same operating conditions would be important for the evaluation of the data quality, reproducibility and instrumental stability (Sangster et al. 2006; Dunn et al. 2011; Lin et al. 2014; Guo et al. 2014). QC samples may be obtained from a representative mixture, such as aliquots pooled from each sample (Roux et al. 2011; González-Domínguez et al. 2014) or 20-50% of the studied samples prepared at the very beginning (Cui et al. 2013). There are also commercially available surrogate biological biofluids from Sigma Aldrich (Dunn et al. 2011) or NIST that could serve as QC samples for large-scale studies.

In addition, spiking multiple internal standards, each of which represents a class of metabolites into the samples before sample treatment is recommended, especially in

large scale studies that involve multiple batches of analysis. This is because each internal standard that share similar chemical and ionization properties as that class of metabolites should undergo the same extraction loss during sample preparation and face similar matrix effect in the ion source. This may evaluate the precision of the sample preparation and the analytical performance of the instrument (Raterink et al. 2014). If the internal standard and that class of metabolites ought to have similar deviation, the data of internal standard could be used to normalize them to eliminate the variation

Though there are various online databases of endogenous and exogenous compounds, traditional databases such as PubChem and Metlin have limited MS/MS spectra and may not cover novel compounds. This prolongs the exploration of the identity of unknown metabolites. In particular, the peak identification cannot be confirmed using MS/MS spectra only without commercial reference standards as the standards can provide retention time to match with the analytes in the sample (Table 2.2). Novel compounds always require additional experiments for structure elucidation. In silico fragmentation tools, such as MetFrag (Ruttkies et al. 2016) and CSI:FingerID (Dührkop et al. 2015), can be utilized to propose possible molecular structures for chemical identification (Blaženović et al. 2018). This can be achieved by matching the experimental MS/MS spectra of novel compounds against in silico fragments generalized from existing spectral database of known compounds.

In addition, no single method or platform may achieve a complete analysis of the metabolome. Multiple platforms increase large coverage of polar metabolites and lipids, such as gas chromatography coupled with liquid chromatography (Psychogios et al. 2011).

Untargeted metabolomics study requires precise consideration of many factors during the experimental designs, apart from the studying subject's inherent variations. The ultimate goal is to minimize extrinsic interference starting from the sample collections to the data acquisitions so that the comparison between groups can be achieved easier with higher reliability.

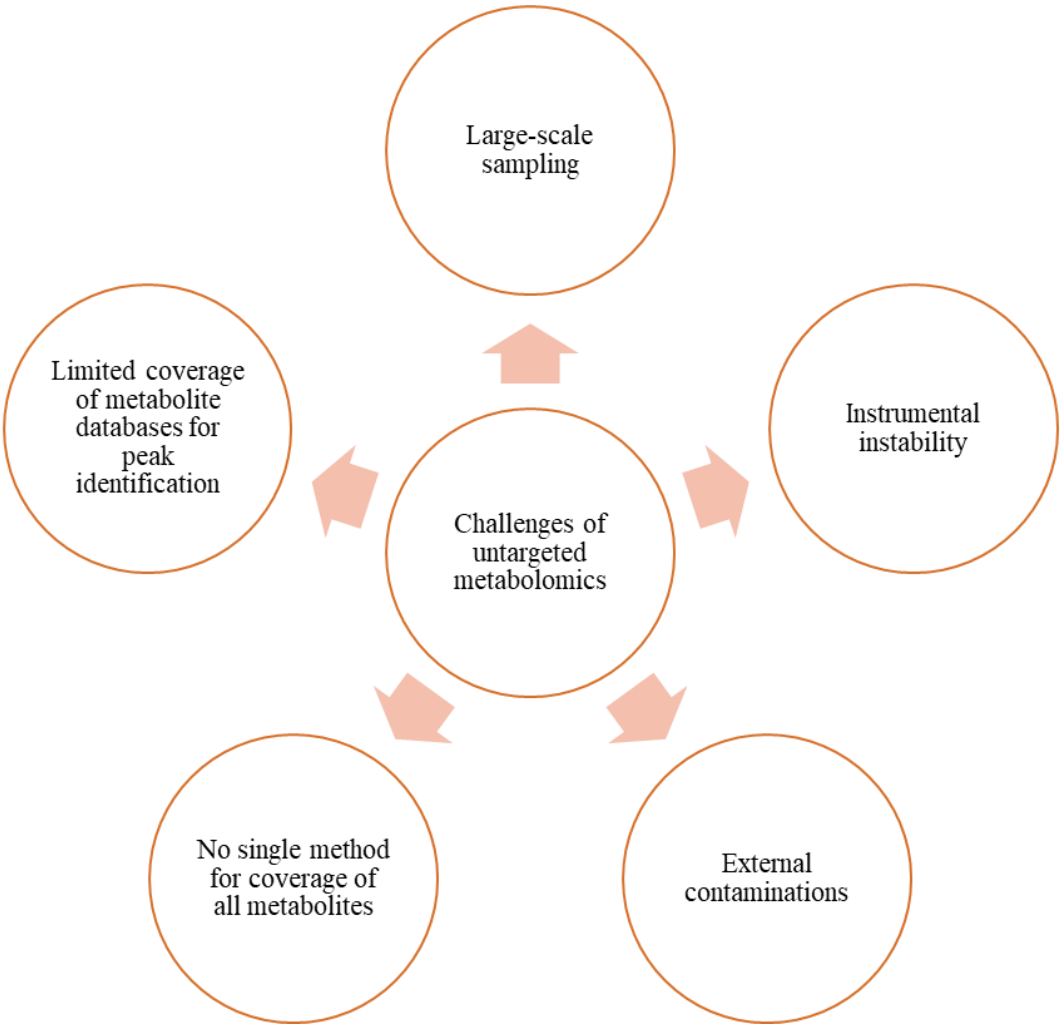
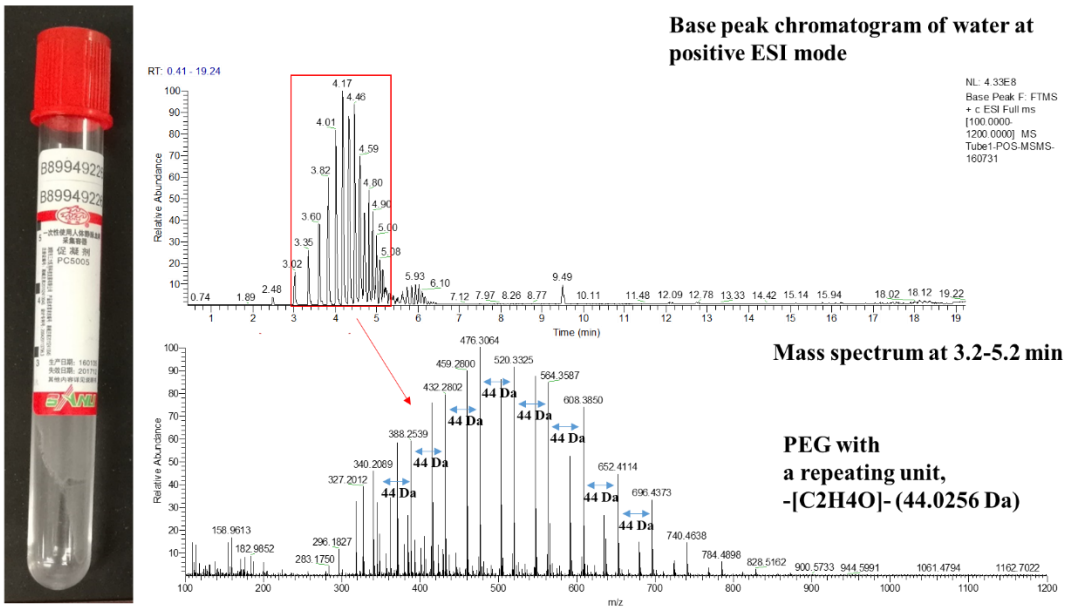


Fig. 2.10 An overview of the challenges of untargeted metabolomics

Vacuum blood collection tube 1:
SANLI tube which the inner wall was coated with clot activator



Vacuum blood collection tube 2:
BD Vacutainer® Serum Tube

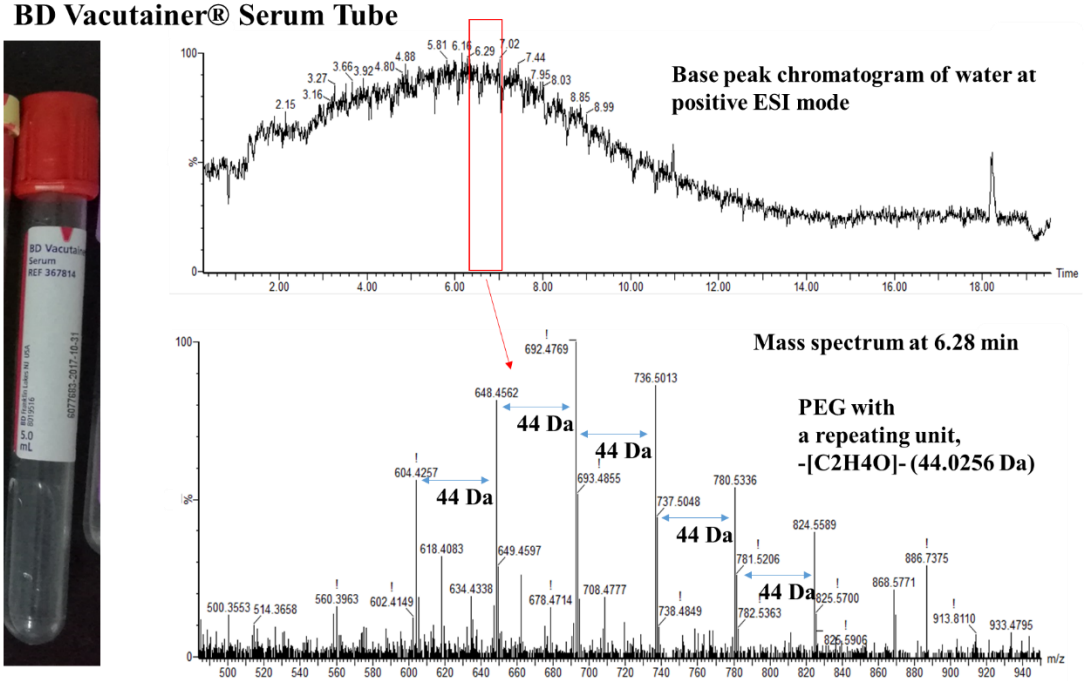


Fig. 2.11 Chromatograms and mass spectra of two solution (water) after filling two blank blood sample tubes that contained PEG.

Chapter 3 A study of cholesterol-lowering effects of piceatannol, a stilbene from wine, using untargeted metabolomics on hypercholesterolemic rats

3.1 Introduction

Hyperlipidemia is characterized by high levels of one or more lipids and/or lipoproteins including free fatty acids, small dense low density lipoprotein cholesterol (LDL-C), triacylglycerols (TG), apolipoprotein B, and/or low levels of high density lipoprotein cholesterol (HDL-C) in blood (Athyros et al. 2011; Shah et al. 2001; Durrington 2003; Sham et al. 2014b). This lipid disorder consists of primary (genetic) familial (Kusters et al. 2013; Talmud et al. 2013; Goldstein et al. 1973) and secondary causes (Fredrickson, Levy, and Lees 1967). Secondary hyperlipidemia is usually due to diet, estrogen therapy, alcohol intake and diseases such as diabetes mellitus and chronic kidney impairment (Hlaing and Park 2013; Mahamuni et al. 2012). Primary hyperlipidemia can get worse if it co-exists with secondary causes (Stone 1994). Among these causes, diets that are high in saturated fats and cholesterol are a common cause of mild hypercholesterolemia seen in modern society.

In the past several decades, the prevalence of hyperlipidemia has accelerated to cover around 10-20% of the adult population worldwide (Margaret D Carroll, Cheryl D. Fryar, and Nguyen 2017; de la Sierra et al. 2015; Lecoffre et al. 2018; Gan et al. 2018). In fact, hyperlipidemia is a major modifiable risk factor for cardiovascular diseases (CVDs) (Nelson 2013; Navar-Boggan et al. 2015). CVDs, including ischemic heart disease and stroke, are one of the major killers around the world and were the leading causes of global death from 2002 to 2016 (World Health Organization 2018a; Nelson 2013). Such prevalence has produced immense economic and health burdens

worldwide (Xi et al. 2014; Fox et al. 2015; Kim and Boye 2009; Mozaffarian et al. 2015). Thus, more attention has been paid to seek measures to relieve such burdens. It is reported that dietary intervention is effective in improving hyperlipidemia and thus CVDs, too (Mann and Piotrowski 1992; Kelly 2010; Hill, Fleming, and Kris-Etherton 2009; Toth 2010; Kim, Lim, and Shin 2019).

Stilbene is a class of compounds found in some foodstuffs of plant origin (Neveu et al. 2010). The intake of these phytochemicals has been linked to a variety of health benefits (Shahidi and Ambigaipalan 2015). For example, stilbenes in red wine like resveratrol have received a lot of interests and are regarded to contribute, in part, to the 'French paradox', an observation of relatively low rates of CVDs in France in spite of the relatively high consumption of saturated fat in diets (Catalgol et al. 2012). These phytochemicals have been reported to be safe, as evidenced upon extensive human consumption (Sham et al. 2014b) and are ideal candidates for developing protective functional foods for CVDs. Thus, understanding the health benefits of them would help improve the quality of human life, especially when we are moving towards an aging society.

Piceatannol (3,3′,4,5′-tetrahydroxy-*trans*-stilbene) is a natural hydroxylated analog of resveratrol (3,4′,5-trihydroxy-*trans*-stilbene) (Kim et al. 2009; Zhu et al. 2003; Piver et al. 2004; Potter et al. 2002); it is also a stilbene naturally available from many edible fruits and plants, such as passion fruit (Matsui et al. 2010), grapes (Flamini, De Rosso, and Bavaresco 2015), the root of peanut, traditional Chinese medicines such as rhizome of *Rheum emodi* (Chai et al. 2012) and rhizome and root of *Polygonum cuspidatum* (Lin et al. 2007) present as a free form or piceatannol glucoside. It is also commonly found in functional drinks such as tea (Viñas et al. 2011) and red wine (Viñas et al. 2009;

Tassoni, Tango, and Ferri 2014). The contents of piceatannol, *cis*-resveratrol and *trans*-resveratrol in red wine were 7.7-14 ng/mL, 4.4 -37 ng/mL and 32 - 41 ng/mL, respectively (Viñas et al. 2009). Piceatannol, like many stilbenes, exhibits health benefits such as improvement of endothelial function (Frombaum et al. 2011; Kinoshita et al. 2013) and anti-inflammatory effect (Choo et al. 2014). Its solubility and bioavailability are higher than those of resveratrol (Chen et al. 2016; Roupe et al. 2006). In terms of antioxidant activity, piceatannol is more potent than resveratrol (Rüweler et al. 2009; Ovesná et al. 2006). There is much evidence showing that piceatannol could be a promising agent for treating CVDs (Tang and Chan 2014). As long-term supplementation of *trans*-resveratrol would improve energy metabolism and metabolic profile of obese human subjects (Timmers et al. 2011), piceatannol and piceatannol-rich food may also be developed as dietary supplement. Various sources in functional foods and a wide range of biological activities have been reported but the exact mechanisms underlying the protective effects of piceatannol are still largely unknown.

In the present study, a combination of ultra-performance liquid chromatography-quadrupole time-of-flight - mass spectrometry (UPLC-QTOF-MS) and gas chromatography - mass spectrometry (GC-MS) was employed to investigate the metabolite changes in the serum samples from rats to examine the therapeutic effect of piceatannol on high fat diet (HFD)-induced hypercholesterolemic rats.

3.2 Materials and methods

3.2.1 Materials

HPLC-graded methanol, acetonitrile, analytical reagent-graded sulphuric acid (>95% purity) and potassium hydroxide were purchased from Fisher Scientific (Hampton, NH, USA) while HPLC-graded n-hexane was obtained from Duksan (ANSAN-SI, South Korea). Formic acid, cis-10-nonadecenoic acid (C19:1n9c), oleic acid and their methyl

esters, Supelco 37 Component FAME Mix, glycerol 1-oleate, L-serine, L-glutamic acid, hyodeoxycholic acid, ursodeoxycholic acid, sodium chenodeoxycholate, sodium tauroursodeoxycholate and sodium taurochenodeoxycholate were purchased from Sigma-Aldrich (St. Louis, MO, USA). Cholesterol 7α-hydroxylase (CYP7A1), α-muricholic acid, deoxycholic acid and taurodeoxycholic acid were purchased from Santa Cruz Biotechnology (Dallas, Texas, USA). Glycoursodeoxycholic acid, ursodeoxycholic acid, cholic acid-2,2,4,4-D_4, glycoursodeoxycholic acid-2,2,4,4-D_4, deoxycholic acid-2,2,4,4-D_4 and ursodeoxycholic acid-2,2,4,4-D_4 were purchased from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Sodium glycocholate hydrate was purchased from Acros Organics (Morris Plains, NJ, USA). Cholic acid, β-muricholic acid, glycodeoxycholic acid, glycohyodeoxycholic acid, taurocholic acid, sodium glycochenodeoxycholate, taurohyodeoxycholic acid, lithocholic acid, glycocholic acid-2,2,4,4-D_4 and glycochenodeoxycholic acid-2,2,4,4-D4 were obtained from Steraloids (Newport, RI, USA). Piceatannol (98% confirmed by HPLC) and simvastatin (20-mg tablets, 10% w/w, confirmed by HPLC) were obtained from Merck Sharp & Dohme (Hangzhou, China) and Nanjing Zelang Medical Technology Co. Ltd. (Nanjing, China), respectively. Water was purified in-house using a Milli-Q Advantage A10 water purification system (Millipore, Bedford, MA, USA).

3.2.2 Animal studies

Sprague-Dawley male rats of three months old with ~300g body weight were obtained from the centralized animal facilities of The Hong Kong Polytechnic University (Hong Kong, China). Rats were housed in a temperature-controlled room (25 ± 2 °C) with free access to water and rat chow and regular 12 h/12 h light/dark cycles. After acclimation for seven days in the laboratory environment, the rats were randomly assigned into four

experimental groups (n = 8 per group): (1) the normal control group, (2) HFD-fed model group, (3) the simvastatin (3mg/kg bw/day) treated HFD-fed group and (4) piceatannol (100 mg/kg bw/day)-treated HFD-fed group. The control group was fed with normal rat chow [fat (~10%), protein (~14%) and carbohydrate (~76%)]. All of the other groups were fed with HFD which was composed of addition of 2% pure cholesterol, 1% cholic acid and 5.5% peanut oil to the normal rat chow. The diets were purchased from Guangdong Medical Laboratory Animal Center (Guangzhou, China). The treatment groups were administrated with their corresponding drugs by oral gavages once every morning for four weeks while the rats in normal control group and HFD-fed model group were administrated with distilled water. The body weight of each rat was monitored daily during the whole experimental period. After four weeks, the rats were fasted overnight and then sacrificed with carbon dioxide asphyxiation, followed by collection of tissues and blood for further analysis. The experimental protocol was approved by the Animal Subjects Ethics Sub-committee of The Hong Kong Polytechnic University (ASESE no. 05/21) and the animal licence was issued by the Department of Health, the Government of the Hong Kong Special Administrative Region. All procedures complied with the Guide for the Care and Use of Laboratory Animals distributed by the US National Institutes of Health and the principles were outlined in the Declaration of Helsinki.

3.2.3 Collection of serum samples

Blood was collected with centrifuge tubes by cardiac puncture immediately after the rats were sacrificed. The blood was stood at room temperature for 30 min, and then centrifuged at 3900 ×g and 4 ℃ for 15 min. The supernatant (serum) was aliquoted and frozen at -80 ℃ before analysis. Total cholesterols, TG, LDL-C and HDL-C of serum samples were measured using the Nanjing KeyGEN Biotech's kits (Nanjing, China).

3.2.4 Untargeted metabolomics analysis

3.2.4.1 Quality control sample preparation method

Aliquots of 20 µL from all serum samples were pooled, vortexed and aliquoted to be a QC sample and stored at -80 ℃ until use. For each analytical batch, QC samples went through the same procedure as study groups for UPLC-MS and GC-MS extraction protocols as described. Prior to the initial chemical analysis, five repeated injections of a QC sample were used to equilibrate the column and check the working condition of the instruments. Afterwards, one QC sample was injected to monitor the stability of the instruments between every five study samples.

3.2.4.2 GC-MS sample preparation method

The procedures reported by Sánchez-Avila *et al* and Yi *et al* (Yi et al. 2006; Sánchez-Avila et al. 2009) were applied as a starting point for the method development here. A two-step methylation of non-esterified fatty acids (NEFAs) and esterified fatty acids (EFAs) fractions in each serum without protein removal was used. 200 µL serum aliquot was mixed with 100 µL internal standards (a mixture of 125 µM of C19:1n9c methyl ester and 125 µM C19:1n9c dissolved in methanol). In the first step of base-catalysed trans-esterification of EFAs, 2 mL of 0.4 M potassium hydroxide in methanol was added to the serum aliquot, vortexed for 30 s and stood at room temperature for 10 min. Then, 2 mL n-hexane was added, vortexed for 30 s twice and the upper layer of hexane phase extracting EFA methyl esters was isolated. The hexane partition was done twice, and the two hexane phases extracted were combined. In the second step of acid-catalysed esterification of NEFAs, 2 mL of 10% sulphuric acid in methanol was added to the remaining phase containing NEFAs and vortexed for 1 min. The mixture was incubated at 70 ℃ for 30 min. Isolation of NEFA methyl esters by hexane extraction was the same as that of EFA methyl esters. The two different hexane phases containing EFA and NEFA methyl esters respectively were dried in a rotational vacuum

concentrator (RVC 2-25, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany). The dried sample was reconstituted with 200 µL n-hexane prior to GC-MS analysis.

3.2.4.3 GC-MS condition

The GC-MS system was an Agilent 6890N GC/5975C VL MSD system equipped with an Agilent 7683 Automatic Liquid Sampler (Agilent technologies, Inc., CA, USA). The column for separation was a DB-WAX column (30 m x 0.25 mm i.d., 0.25 µm; Agilent J&W Scientific, CA, USA). The inlet temperature was 240 ℃. Helium gas (≥ 99. 999%) was used as carrier gas with a flow rate of 1.0 mL/min. 1 µL of samples was injected in splitless mode. The temperature program was optimized as follows: the initial oven temperature at 70 ℃, held for 1 min; 20 ℃/min to 170 ℃; 9 ℃/min to 190 ℃; 2 ℃/min to 220 ℃; 4 ℃/min to 230 ℃; 230 ℃ kept for 7.5 min. The mass spectrometry conditions were as follows: electron impact mode at ionization energy of 70 eV; ion source temperature at 230 ℃; full scan mode in m/z range of 35 - 550 with 0.3 s/scan velocity. The solvent delay was set at 3 min.

3.2.4.4 UPLC-QTOF-MS sample preparation method

Serum samples were thawed at 4 °C and vortexed before preparation. 100 µL serum from all samples was deproteinated with cold 300 µL methanol and vortexed for 30 s. The mixture was kept at -20 ℃ for 1 h before centrifugation at 18790 ×g for 20 min at 4 ℃. 340 µL supernatant was dried with nitrogen gas and stored at -80 ℃. Before UPLC-MS analysis, it was reconstituted in 100 µL 70% methanol, followed by centrifugation at 18790 ×g for 10 min at 4 ℃.

3.2.4.5 UPLC-QTOF-MS condition

3 µL of supernatant was injected into a Waters ACQUITY UPLC system. The separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm ×

50 mm, 1.8 μm) combined with an HSS T3 pre-column (2.1 mm × 5 mm, 1.8 μm, Waters Corporation, Milford, MA). The mobile phase consisted of mobile phase A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) in the elution gradient: 0-1.5 min, 5% B; 2 min, 35% B; 4 min, 50% B; 9 min, 55% B; 12-17 min, 95% B. A 3-min post-run time was used to fully equilibrate the column. The flow rate was set at 0.3 mL/min. The temperatures of sample chamber and UPLC column were 5℃and 35℃, respectively.

Mass spectrometry was performed on a Waters SYNAPT G2 Q-IM-TOF HDMS system (Waters, Milford, USA) with an electrospray ion source (ESI) in both negative and positive modes. Nebulization and cone gases were nitrogen. The cone gas was set at 40 L/h. The nebulization gas was set at 600 L/h and 300 ℃. The source temperature was kept at 120 ℃. The sampling and extraction cone voltages were 40 V and 4 V, respectively. The capillary voltages in negative and positive ESI modes were 2300 V and 3000 V, respectively. The data acquisition rate was 1 s plus a 0.024 s interscan delay. Signals in *m/z* 50 - 1000 were scanned in the centroid mode. A lock mass of leucine enkephalin was used for accurate mass acquisition, through a lock spray interface using 30 eV trap collision energy, 2500 V capillary energy, 40 V cone voltage at a flow rate of 4 μL/min, monitoring for negative ESI mode ($[M-H]^-$: 554.2615) and positive ESI mode ($[M+H]^+$: 556.2771) during the MS analysis. Mass fragmentation analysis of interested peaks was carried out. Argon was applied as collision gas and the collision energy was adjusted to 5 - 50 eV.

3.2.4.6 Data extraction and processing

For UPLC-MS data, the peak picking, alignment and filtering of raw data were done using Waters MarkerLynx Application Manager Version 4.1 SCN 901 (Waters, Milford, USA). The following parameters were used: m/z range of 50 -1000, mass

tolerance at 0.02 Da, intensity threshold at 500 counts, retention time tolerance at 0.2 min, apex track peak parameter calculated automatically, no smoothing and isotopic peaks excluded for processing. Quality control screening was applied by filtering out metabolites not existing in 80% of each group and with coefficient of variation > 30% (Dunn et al. 2011) to reduce the contribution of unstable peaks and minimize noise from the dataset. MarkerLynx software output a matrix of m/z, retention time and intensity pairs. Peak area was normalized to the total peak area of each individual sample and exported to the built-in Extended Statistical tool (EZinfo Version 2.0 software, Umetrics AB) for analysis.

The exported data were scaled to unit variance for PCA to give an overview of QC samples in the score plot. The study samples excluding QC samples were pareto-scaled prior to PLS-DA and OPLS-DA. Based on their contribution to the variation and correlation in the data set, potential candidates were selected from the S-plots of OPLS-DA. Markers were further identified with mass fragmentation and matched with the Human Metabolome Databases (www.hmdb.ca), the KEGG (www.kegg.com/), the METLIN (http://metlin.scripps.edu), and/or confirmed by commercially available reference standards based on their mass fragmentation pattern, retention times and mass accuracy.

For GC-MS data, peak area was obtained by integration with Agilent Chemstation (GC6890 MSD Chemstation E.02.02.1431, Agilent) and then normalized with the spiked internal standard in each individual sample. Peak identification was conducted by comparing the retention time of reference standards and matching the mass spectrum with NIST11 library. Multivariate statistical analysis with the same scaling procedures as UPLC-QTOF-MS was carried out using the EZinfo software.

3.2.5 Quantification analysis of bile acids and their conjugated bile acids in serum samples

3.2.5.1 Preparation of bile acids standard solution and calibration curve

Bile acids, conjugated bile acids (bile salts) and internal standards were dissolved individually in methanol at a concentration of 2000 ppm as stock solutions and stored at -20℃. 14 calibration standard working solutions containing the bile acids and conjugated bile acids at 7.32 ppb – 6 ppm were prepared by serial dilution (2-fold) with methanol. An internal standard mixture was prepared by mixing ursodeoxycholic acid-D_4, deoxycholic acid- D_4, cholic acid-D_4, glycoursodeoxycholic acid-D_4, glycochenodeoxycholic acid-D_4, glycocholic acid-D_4 with methanol into the final concentration of 2 ppm for each internal standard. 150 μL of each calibration standard working solution was mixed with 50 μL of the internal standard mixture. The final 14 calibration standard solutions were 0.55 ppb to 4.5 ppm and they were injected into the UPLC-QQQ-MS at the same conditions as study samples.

3.2.5.1 Sample preparation method

50 μL serum from each sample was mixed with cold 100 μL methanol and 50 μL internal standard mixture (2 ppm each of ursodeoxycholic acid-D_4, deoxycholic acid-D_4, cholic acid-D_4, glycoursodeoxycholic acid-D_4, glycochenodeoxycholic acid-D_4, glycocholic acid-D_4) and then vortexed for 30s. The mixture was cooled at an ice water bath for one hour prior to centrifugation at 18790 ×g at 4℃ for 20 min. 50 μL aliquot of supernatant was dried with nitrogen gas for storage at -80℃. It was reconstituted in 50 μL of 50% methanol in water followed by centrifugation at 18790 ×g and 4℃ for 10 min. The supernatant was obtained for subsequent ultra-performance liquid chromatography triple quadrupole-mass spectrometry (UPLC-QQQ-MS/MS) analysis.

3.2.5.2 UPLC-QQQ-MS/MS condition

This UPLC-QQQ-MS/MS condition was modified from a previous report (García-Cañaveras et al. 2012). A 4 μL aliquot was injected into Agilent 1290 Infinity Ultra High-pressure/performance liquid chromatography (Agilent Technologies, Palo Alto, CA, USA). The separation was performed on a Waters ACQUITY UPLC BEH C_{18} column (2.1mm × 100mm, 1.7μm) with BEH C_{18} guard column (2.1mm × 5mm, 1.7μm, Waters Corporation, Milford, MA). The mobile phase consisted of combinations of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 0.5 mL/min with elution gradient as follows: 0-0.5 min, 5% B; 5 min, 25% B; 10 min, 27.5% B; 18 min, 40% B; 20 min, 60% B; 22-24.5 min, 95% B. A 3-min post-run time was set to fully equilibrate the column. Column and sample chamber temperature were 65℃ and 4℃ respectively.

MS was performed on an Agilent 6460 ESI Triple Quadrupole Mass Spectrometer (Agilent Technologies, Palo Alto, CA, USA) in negative ESI mode. The ESI source operation parameters were as follows: capillary, 3500 V; nozzle voltage, 500 V; gas flow, 8 L/min; gas temperature, 350°C; nebulizer, 45 psi; sheath gas temperature, 300℃; sheath gas flow, 11 L/min. The scan mode was multiple reactions monitoring (MRM) mode.

3.2.5.3 Validation method

The quantitative analytic method of bile acids and conjugated bile acid was validated according to the guidelines given by the protocol of the bioanalytical method validation provided by the US Food and Drug Administration (FDA) with a few modifications (Administration 2018). The linearity was assessed for each bile acid and conjugated bile acid over a range of calibration concentration from 0.55 ppb to 4.5 ppm. The limit of detection (LOD) was considered with signal-to-noise ratios > 3. The lower limit of

quantification (LLOQ) was determined with signal-to-noise ratios > 10 and the upper limit of quantification (ULOQ) was the highest calibration standard concentration on the calibration curve.

Recovery was studied by comparing between study samples spiked with three concentrations (31.25 ppb, 125 ppb and 500 ppb) of bile acid mixtures before sample extraction and the theoretical concentration of the spiked standards, respectively. Blank QC serum samples spiked with internal standards only were used to determine the original endogenous bile acids and conjugated bile acids in the serum.

$$\text{Recovery} = \frac{\text{mean detected conc}_{\text{serum after spiked}} - \text{mean original conc}_{\text{blank serum}}}{\text{mean spiked conc}} \times 100\ \%$$

The precision was obtained as the CV %. Due to the limit of QC samples, each value represented the mean of four replicates.

3.2.5.4 Data processing and analysis

The peak area was extracted by Agilent MassHunter Quantitative Analysis Software (Agilent Technologies, Palo Alto, CA, USA). The peak area ratio of each standard to its internal standard was calculated from each MRM chromatogram and was used for method validation and quantitation of the samples. Calibration curves were plotted at different concentrations in the linear range and were used to calculate the diluted concentration (dilution factor = 4) in the rat serum samples.

3.2.6 Western blot immune-reactivity assay

Immunoblotting procedures were applied to quantify CYP7A1 protein. Homogenized liver tissue samples were lysed with lysis buffer and cooled in ice for 20 min. The mixture was centrifuged for 15 min at 8265 × g. The supernatant was isolated and assayed for protein contents with Bio-Rad protein assay kit (Bio-Rad, Hercules, CA,

USA). 80 µg of protein extracts were applied to 7.5% SDS-polyacrylamide gels and transferred to polyvinylidene difluoride membranes. The membranes were blotted for 1 h with 5% w/v non-fat dry milk and then incubated with the CYP7A1 antibody at 4 ℃ overnight. The membranes were washed three times in Tris-buffered saline containing 0.1% Tween 20 and were incubated with horseradish peroxidase-linked secondary antibodies for 1 h. Protein contents were detected with the chemiluminescence detection system (Amersham Biosciences, Little Chalfont, UK) and visualized with Fujifilm autoradiographic films. Densitometric analysis of optical densities was completed using AlphaEaseFC™ software (Alpha Innotech Corporation, San Leandro, CA).

3.2.7 Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 25 (Chicago, IL, USA). After removal of outliers (1.5 times of the interquartile range), statistical differences were evaluated by one-way analysis of variance (ANOVA) at a univariate level; least significance difference (LSD) *post-hoc* test was applied with assumption of equal variances. Correlation analysis was done using linear regression and the Pearson correlation coefficient (R) was determined. A p value of 0.05 was considered as statistically significant.

3.3 Results and Discussion

3.3.1 Effect on serum lipid profiles

The lipid profiles of total cholesterols, TG, HDL-C and LDL-C from different animal groups are presented in Fig. 3.1. The HFD significantly raised the total serum cholesterols (4.32 ± 0.43 mmol/L) and LDL-C (0.60 ± 0.08 mmol/L) compared with normal control (1.88 ± 0.09 mmol/L and 0.19 ± 0.01 mmol/L respectively) while TG and HDL-C showed no significant changes among all groups. The atherogenic index is

the ratio of (total serum cholesterol - HDL-C) to HDL-C and is a parameter to evaluate the risk of coronary heart disease. Administration of either piceatannol (0.89 ± 0.12 mmol/L) or simvastatin (0.83 ± 0.14 mmol/L) could significantly lower the atherogenic indexes compared with the HFD model (1.54 ± 0.25 mmol/L), suggesting that piceatannol had protective potential in atherosclerosis.

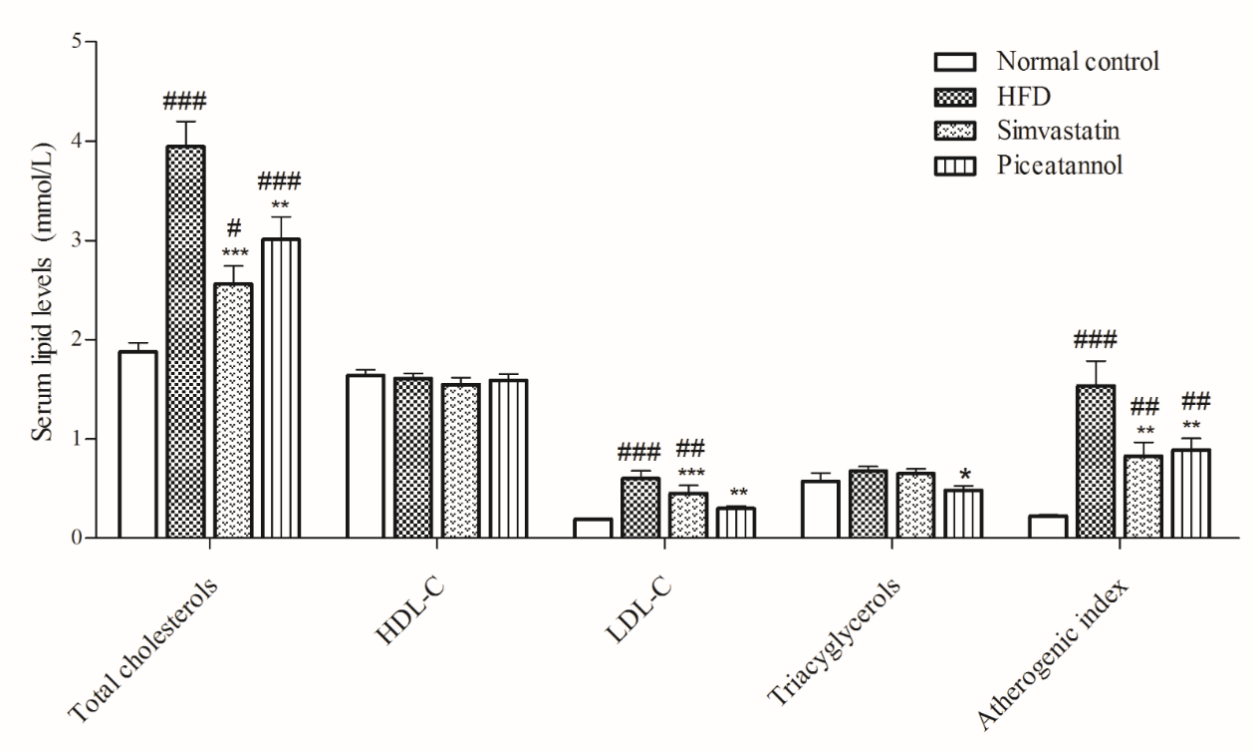


Fig. 3.1 Serum lipid levels and atherogenic indexes of rats from the normal control, HFD model, simvastatin-treated and piceatannol-treated HFD-fed groups.

Data are expressed as mean ± Standard error of the mean (SEM), n = 6 - 8. One way-ANOVA followed by LSD *post-hoc* test: #$p < 0.05$, ##$p < 0.01$ and ###$p < 0.001$ compared with the normal control. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$ compared with the HFD model group.

3.3.2 Reliability of the untargeted MS metabolomics platforms

The stabilities of GC-MS and UPLC-MS detections were assessed by inter-day measurement of a 37 FAME-standard mixture and the pooled QC sample injections, respectively. 10 metabolites including both positive and negative ESI modes from QC samples and 16 detected FAMEs from the FAME mixture were selected for comparison (Tables S3.1 and S3.2 in Appendix I). The score plot of PCA was used to evaluate the stability of the analytical instruments. Fig. 3.6A showed the QC samples were clustered together in PCA score plots, indicating the low variation between QC samples. The coefficient of variation (CV %) of 16 analytes in GC-MS was less than 7% while the CV % of ten analytes in UPLC-MS was smaller than 22%. Results showed high reproducibility was achieved across the runs. This ensured the changes among the different groups observed from the statistical analysis were mainly biologically related but not related to systematic errors in the measurements.

Additionally, the reproducibility of peak retention time was also evaluated throughout the study to calculate the CV %. The results showed that the CV % of all the retention time was less than 0.50%. Recovery of transesterification and esterification was also examined by spiking a known concentration of standards before extraction of serum samples. 50%, 100% and 200% oleic acid (a NEFA, 200, 400 and 800 μM) and glycerol 1-oleate (an EFA, 100, 200 and 400 μM) relative to the known concentration were spiked into three QC serum samples, respectively. Their recoveries were in the range at 105.45 – 113.21% for esterification of oleic acid and at 100.05-103.16 % for transesterification of glycerol 1-oleate.

High retention time and analytical reproducibility and good recovery of the metabolite detection demonstrated that the presented untargeted methodologies had the robustness and reliability required by a metabolic profiling study. The UPLC-MS and GC-MS

chromatograms of serum in normal control, HFD model and piceatannol-treated group are shown in Figs. 3.2 and 3.3.

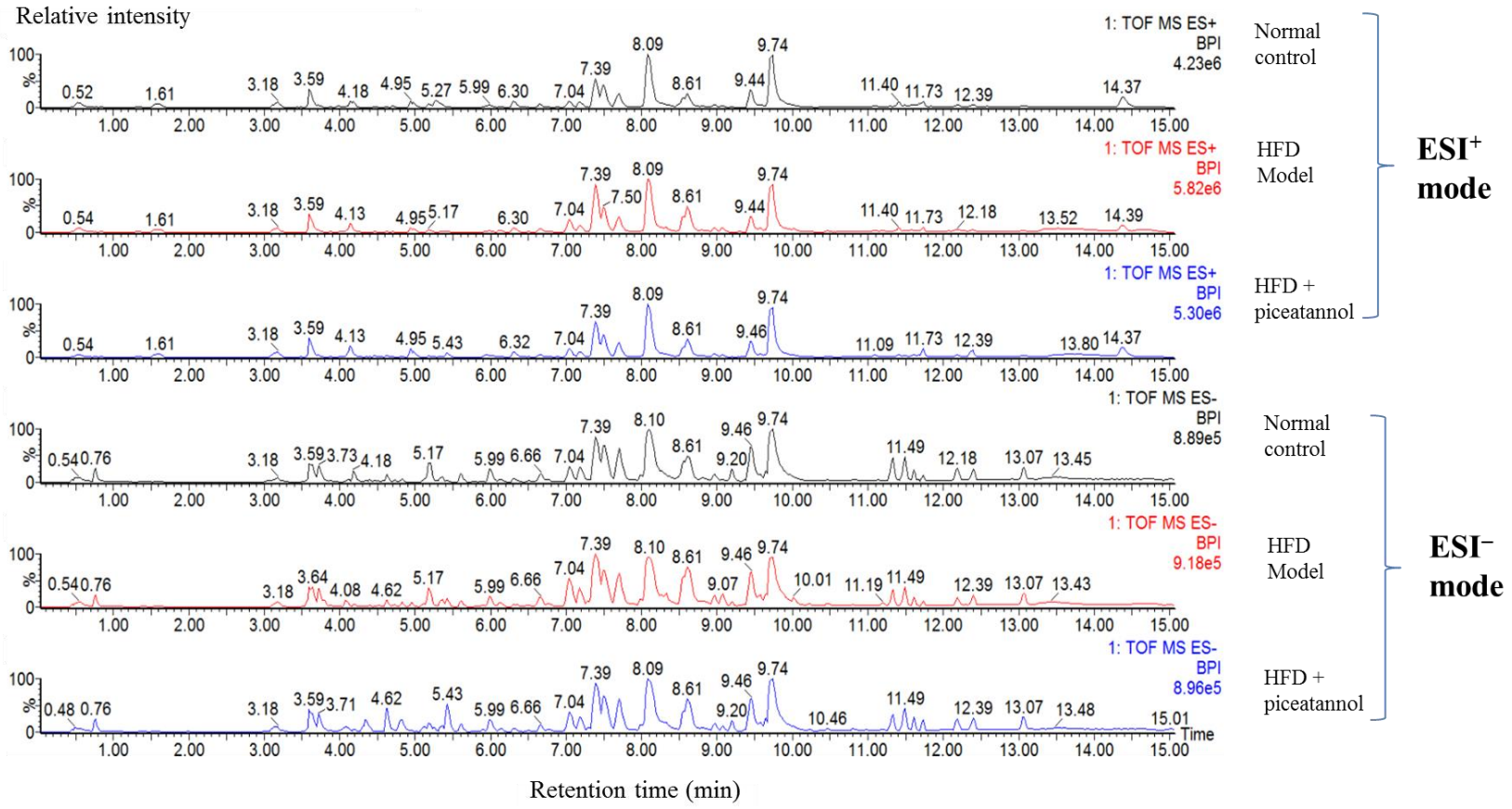


Fig. 3.2 Typical base peak intensity chromatograms of rat serums from the normal control group, HFD model and piceatannol-treated group in UPLC-QTOF-MS analysis

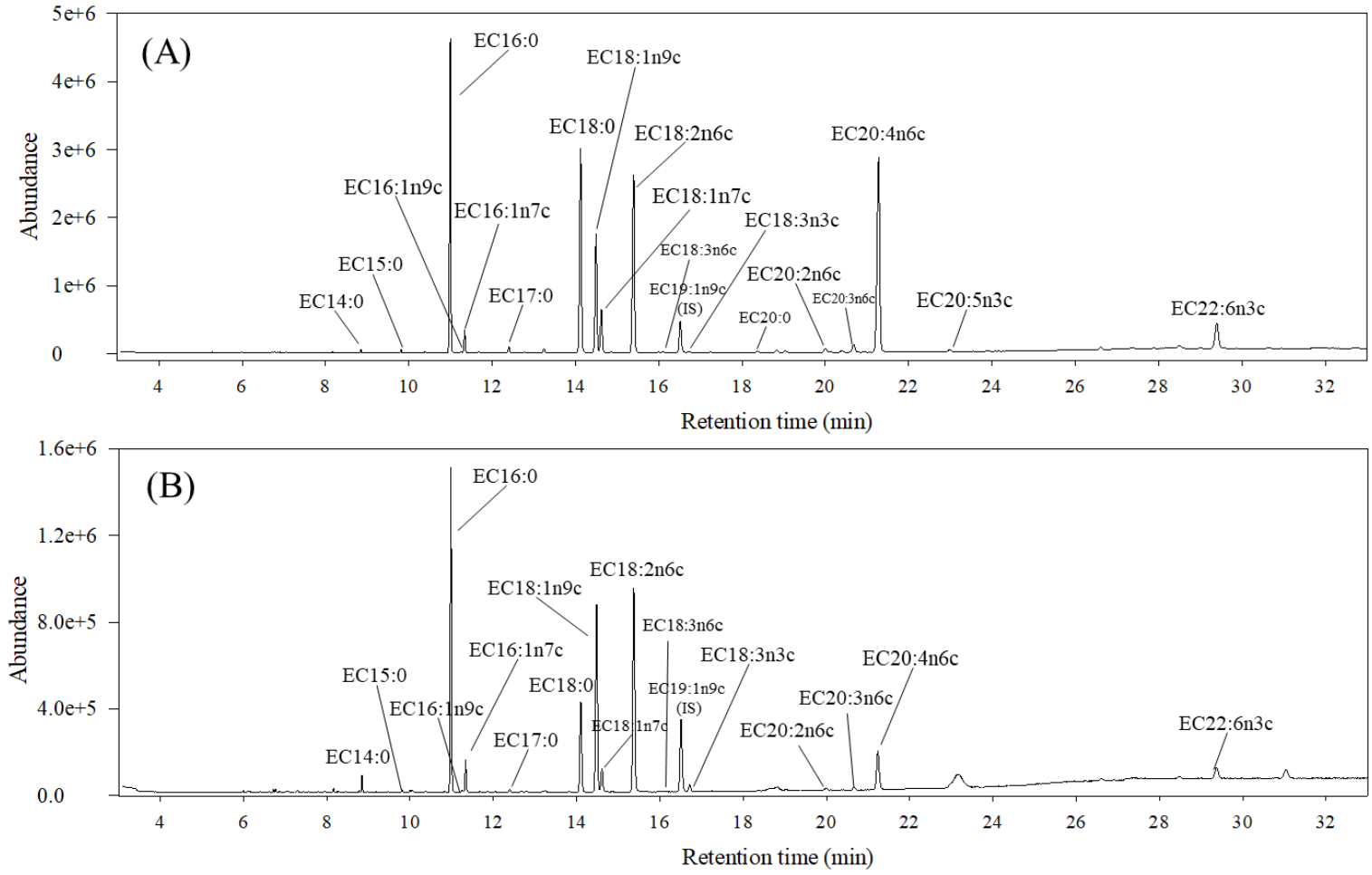


Fig. 3.3 Typical total ion chromatograms of (A) EFA and (B) NEFA of rat serums from HFD group in GC-MS Analysis.

Table 3.1 Identity of EFA and NEFA detected in the rat serum using GC-MS

Retention time (min)	Systematic name	Common name	Shorthand designation	Molecular formulae	Molecular weight (g/mol)
8.87	n-Ttetradecanoate, methyl ester	Myristic acid methyl ester	EC14:0	C15H30O2	242.40
9.83	n-Ppentadecanoate, methyl ester	N/A	EC15:0	C16H32O2	256.42
11.01	Hexadecanoic acid, methyl ester	Palmitic acid methyl ester	EC16:0	C17H34O2	270.45
11.26	cis-7-Hexadecenoic acid, methyl ester	Hypogeic acid methyl ester	EC16:1n9c	C17H32O2	268.24
11.36	cis-9-Hexadecenoic acid, methyl ester	cis-Palmitoleic acid methyl ester	EC16:1n7c	C17H32O2	268.43
12.43	Heptadecanoic acid, methyl ester	Margaric acid methyl ester	EC17:0	C18H36O2	284.48
14.14	n-Octadecanoic acid, methyl ester	Stearic acid methyl ester	EC18:0	C19H38O2	298.50
14.52	cis-9-Octadecenoic acid , methyl ester	Oleic acid methyl ester	EC18:1n9c	C19H36O2	296.49
14.65	cis-11-Octadecenoic acid methyl ester	cis-Vaccenic acid methyl ester	EC18:1n7c	C19H36O2	296.49
15.42	cis-9,12-Octadecadienoic acid	Linoleic acid methyl ester	EC18:2n6c	C19H34O2	294.47
16.06	cis-6,9,12-Octadecatrienoic acid, methyl ester	cis-γ-Linolenic acid methyl ester	EC18:3n6c	C19H32O2	292.46
16.77	cis-9,12,15-Octadecatrienoic acid, methyl ester	cis-α-Linolenic acid methyl ester	EC18:3n3c	C19H32O2	292.46
18.41	Eicosanoic acid, methyl ester	Arachidic acid methyl ester	EC20:0	C21H42O2	326.56
20.05	cis-11,14-eicosadienoic acid, metyl ester	N/A	EC20:2n6c	C21H38O2	322.53
20.73	cis-8,11,14-Eicosatrienoic acid methyl ester	Dihomo-γ-linolenic acid methyl ester	EC20:3n6c	C21H36O2	320.51
21.30	cis-5,8,11,14-Eicosatetraenoic acid, methyl ester	Arachidonic acid methyl ester	EC20:4n6c	C21H34O2	318.49
23.40	cis-5,8,11,14,17-Eicosapentaenoic acid, methyl ester	N/A	EC20:5n3c	C21H32O2	316.48
25.40	cis-13,16-Docosadienoic acid, methyl ester	N/A	EC22:2n6c	C23H42O2	350.58
16.56	cis-10-Nonadecenoic acid, methyl ester	N/A	EC19:1n9c	C20H38O2	310.51

3.3.3 Method validation of UPLC-QQQ-MS/MS analysis

MRM extracted ion chromatograms of bile acids and conjugated bile acids are shown in Fig. 3.4. Method validation of targeted quantitative analysis of bile acids and conjugated bile acids was optimized (Table 3.2 and Fig. 3.5) and suited to the analysis of the study samples. The calibration curves showed a good linearity for all standards ($R^2 \geq 0.9984$) (Table 3.3 and Fig. 3.5). Their concentrations in study and QC samples were above the LLOQ and fitted the linear range, except the concentrations of lithocholate < LOD (4.39 ppb) and that of tauroursodeoxycholate < LLOQ (LLOQ = 1.10 ppb, LOD = 0.55 ppb) in rat serum samples. Their CV % in QC serum samples (*n* = 7) injected in every four randomized rat serum samples was 2.43-15.67%. The recoveries were reproducible (77.54 – 116.56 %, CV% < 6 %) in all standards at 31.25 ppb, 125 ppb and 500 ppb concentrations that were spiked into QC serum samples (Table 3.4 and Table S3.3). These result matched the acceptance criteria of FDA (recovery = 80-120 % and variance < 15%) in term of quantification of biological samples (Administration 2018).

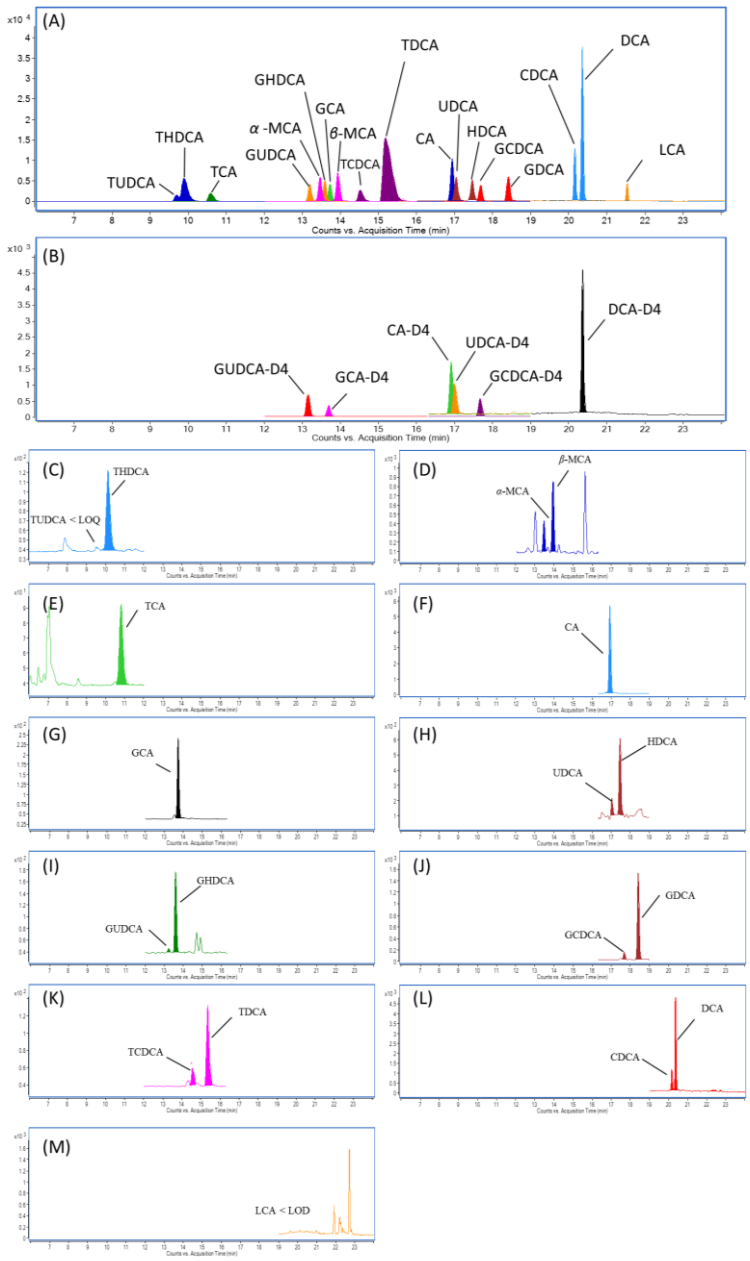


Fig. 3.4 MRM extracted ion chromatograms of bile acids and conjugated bile acids in UPLC–QQQ-MS/MS quantitative analysis.

(A) Reference standards; (B) Internal standards; (C-M) Rat QC serum sample. Abbreviations: *α*-MCA, *α*-Muricholate; *β*-MCA, *β*-muricholate; CA, cholate; CDCA, chenodeoxycholate; DCA, deoxycholate; GCA, glycocholate; GCDCA, glycochenodeoxycholate; GDCA, glycodeoxycholate; GHDCA, glycochenodeoxycholate; GUDCA, glycoursodeoxycholate; HDCA, hyodeoxycholate; LCA, lithocholate; TCA, taurocholate; TCDCA, taurochenodeoxycholate; TDCA, ursodeoxycholate; THDCA, taurohyodeoxycholate; TUDCA, tauroursodeoxycholate; UDCA, ursodeoxycholate; CA-D_4, cholate-D_4; GUDCA- D_4, glycoursodeoxycholate-D_4; DCA-D_4, deoxycholate-D_4; UDCA-D_4, ursodeoxycholate-D_4; GCA-D_4, glycocholate-D_4 ;GCDCA-D_4, glycochenodeoxycholate-D_4.

Table 3.2 Optimized multiple reaction monitoring parameters for each bile acid and each conjugated bile acids in UPLC–QQQ-MS/MS quantitative analysis

Standards	Retention time (min)	MRM transitions (*m/z*)	Fragmentor (V)	Collision energy (V)	Internal standards
Hyodeoxycholate	17.41	391 → 391	160	3	
Ursodeoxycholate	16.99	391 → 391	160	3	Ursodeoxycholate-D_4
Ursodeoxycholate-D_4	16.94	395 → 395	160	3	
Chenodeoxycholate	20.15	391 → 391	160	3	
Deoxycholate	20.35	391 → 391	160	3	Deoxycholate-D_4
Lithocholate	21.53	375 → 375	120	3	
Deoxycholate-D_4	20.33	395 → 395	160	3	
***α*-Muricholate**	13.42	407 → 407	175	3	
***β*-Muricholate**	13.89	407 → 407	175	3	Cholate-D_4
Cholate	16.88	407 → 407	175	3	
Cholate-D_4	16.86	411 → 411	175	3	
Glycoursodeoxycholate	13.14	448 → 74	120	40	
Glycohyodeoxycholate	13.54	448 → 74	120	40	Glycoursodeoxycholate-D_4
Tauroursodeoxycholate	9.60	498 → 80	140	80	
Taurohyodeoxycholate	10.10	498 → 80	140	80	
Glycoursodeoxycholate-D_4	13.11	452 → 74	120	40	
Glycochenodeoxycholate	17.64	448 → 74	120	40	
Glycodeoxycholate	18.37	448 → 74	120	40	Glycochenodeoxycholate-D_4
Taurochenodeoxycholate	14.42	498 → 80	140	80	
Taurodeoxycholate	15.02	498 → 80	140	80	
Glycochenodeoxycholate-D_4	17.62	452 → 74	120	40	
Glycocholate	13.67	464 → 74	120	40	
Taurocholate	10.49	514 → 80	140	80	Glycocholate-D_4
Glycocholate -D_4	13.66	468 → 74	120	40	

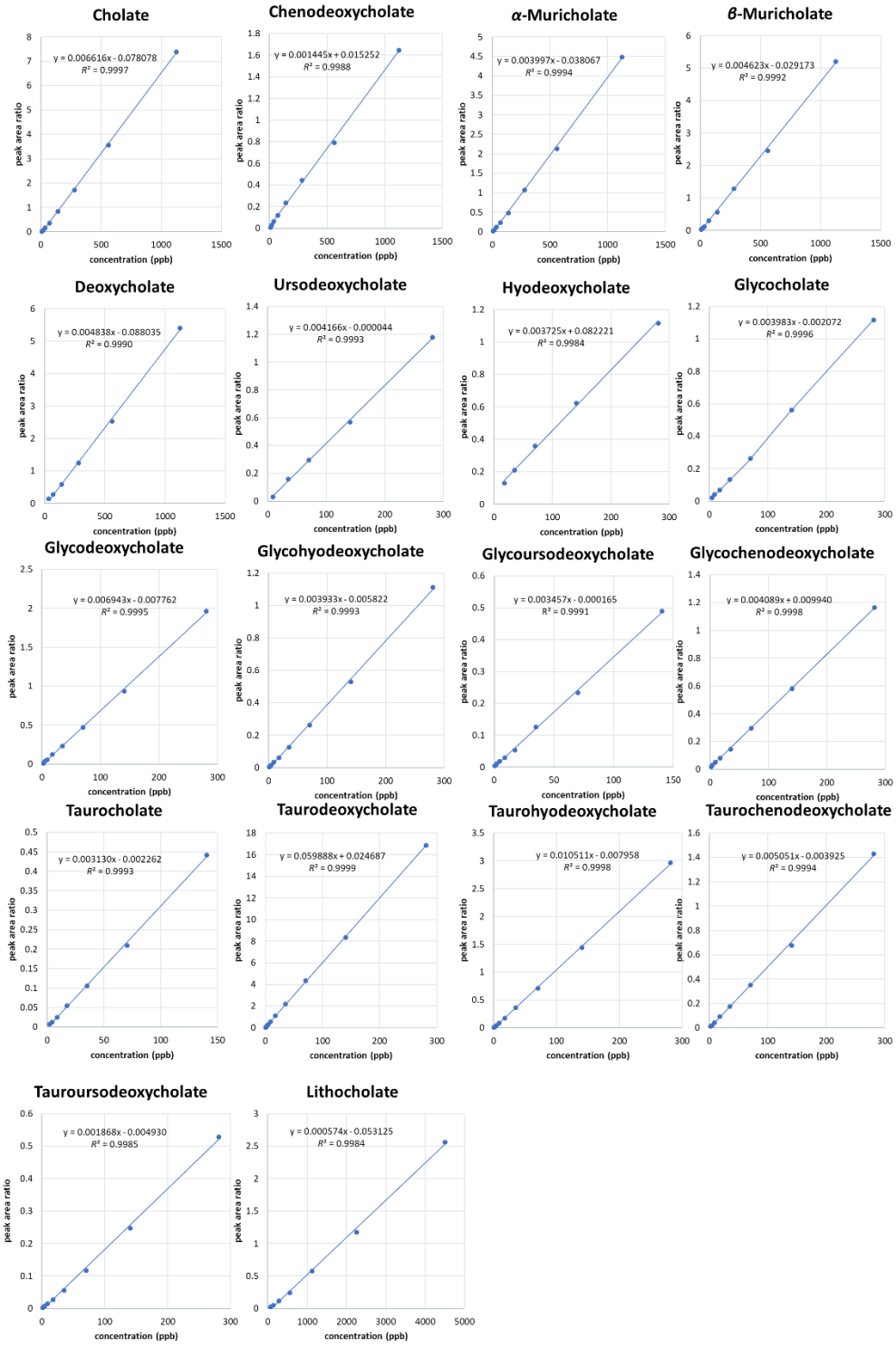


Fig. 3.5 Calibration curve of bile acids and conjugated bile acids using UPLC–QQQ-MS quantitative analysis

Table 3.3 Linear regression, linear range, LOD of each bile acid and each conjugated bile acid, and their precision in seven QC serum samples in UPLC–QQQ-MS/MS quantitative analysis

Standard	Linear regression equation	R^2	Linear range (LLOQ - ULOQ, ppb)	LOD	RSD (%) in QC samples
Ursodeoxycholate	$y = 0.004166\ x - 0.000044$	0.9993	8.79-281	4.39	6.71
Hyodeoxycholate	$y = 0.003725\ x + 0.082221$	0.9984	17.57-281	8.79	4.50
Chenodeoxycholate	$y = 0.001445\ x + 0.015252$	0.9988	8.79-281	4.39	5.14
Lithocholate	$y = 0.000574\ x - 0.053125$	0.9984	35.15-4500	4.39	ND
Deoxycholate	$y = 0.004838\ x - 0.088035$	0.9990	35.15-1125	4.39	2.42
***α*-Muricholate**	$y = 0.003997\ x - 0.038067$	0.9994	8.79-1125	4.39	5.27
***β*-Muricholate**	$y = 0.004623\ x - 0.029173$	0.9992	8.79-1125	4.39	4.35
Cholate	$y = 0.006616\ x - 0.078078$	0.9996	8.79-1125	4.39	2.55
Glycoursodeoxycholate	$y = 0.003457\ x - 0.000165$	0.9991	0.55-140.63	0.20	19.56
Glycohyodeoxycholate	$y = 0.003933\ x - 0.005822$	0.9994	1.10-281	0.55	2.73
Glycochenodeoxycholate	$y = 0.004089\ x + 0.009940$	0.9998	2.20-281	0.55	3.38
Glycodeoxycholate	$y = 0.006943\ x - 0.007762$	0.9995	2.20-281	1.10	3.65
Glycocholate	$y = 0.003983\ x - 0.002072$	0.9996	4.39-281	1.10	4.33
Tauroursodeoxycholate	$y = 0.001868\ x - 0.004930$	0.9985	1.10-281	0.55	ND
Taurohyodeoxycholate	$y = 0.010511\ x - 0.007958$	0.9998	0.55-281	0.20	5.71
Taurochenodeoxycholate	$y = 0.005051\ x - 0.003250$	0.9994	2.20-281	0.55	11.20
Taurodeoxycholate	$y = 0.059888\ x + 0.024687$	0.9999	0.55-281	0.20	6.20
Taurocholate	$y = 0.003130\ x - 0.002262$	0.9993	2.20-140.63	0.55	5.13

x, ratio of peak area of each standard to that of its respective internal standard listed in Table 3.2; *y*, standard concentration in ppb. ND, not detected in the serum samples.

Table 3.4 Extraction recovery and precision of each bile acid and each conjugated bile acid detected in QC samples with three spiked concentrations using UPLC–QQQ-MS quantitative analysis

Spiked concentration	31.25 ppb		125 ppb		500 ppb	
Standards	**Recovery extraction (%)**	**Precision (%)**	**Recovery extraction (%)**	**Precision (%)**	**Recovery extraction (%)**	**Precision (%)**
Ursodeoxycholate	102.01	2.35	81.04	2.30	-	-
Hyodeoxycholate	111.25	4.32	79.55	2.38	-	-
Chenodeoxycholate	117.30	3.12	94.59	5.52	-	-
Deoxycholate	109.56	3.61	95.82	5.91	87.21	2.79
α-Muricholate	99.52	3.19	84.59	4.31	78.86	1.95
β-Muricholate	103.70	3.83	82.91	4.11	77.54	0.74
Cholate	81.46	2.54	84.84	1.95	89.72	0.60
Glycoursodeoxycholate	82.86	1.22	81.54	2.33	-	-
Glycohyodeoxycholate	89.00	4.63	84.03	4.51	-	-
Glycochenodeoxycholate	91.07	3.22	82.88	2.07	-	-
Glycodeoxycholate	116.56	2.92	110.76	2.50	-	-
Glycocholate	89.43	1.11	83.07	2.01	-	-
Taurohyodeoxycholate	89.91	2.97	85.30	2.58	-	-
Taurochenodeoxycholate	108.63	3.08	99.84	0.89	-	-
Taurodeoxycholate	104.20	2.32	93.12	1.35	-	-
Taurocholate	104.31	1.74	99.07	2.75	-	-

Each value represents the mean of four replicates. "–" represents the concentration is out of the linear range.

3.3.4 Effect of piceatannol to HFD

After quality screening of the UPLC-QTOF-MS data acquired in negative and positive ESI modes, 878 out of 1976 peaks and 2491 out of 4904 peaks were obtained, respectively, using MarkerLynx software.

Metabolite changes in the serum collected from the animal study were investigated using OPLS-DA of the above quality-screened signals. Fig. 3.6B showed a clear separation between the normal control and the HFD model. The results obtained by VIP value ≥ 1.5 and S-loading plot (Fig. 3.6C) followed by validation with one-way ANOVA ($p < 0.05$) showed that over 20 endogenous metabolites were altered in the two groups. These metabolites belong to three major classes: lysophospholipids, fatty acids, bile acids and their conjugated products. The bile acids, conjugated bile acids, free fatty acids and fatty acid methyl esters were identified by matching with reference standards whereas the lysophosphatidylethanolamine (lysoPE) and lysophosphatidylcholines (lysoPCs) were putatively identified by accurate mass measurement, MS fragmentation, retention time matching with database search and literatures. The fold changes of these metabolites compared with those of the HFD-fed model were listed in Table 3.5 and more details could be seen in the heatmap of Fig. 3.7. The above model offered a basis for studying the role of piceatannol in reducing the blood cholesterol level in HFD-fed rats.

To study the effect on piceatannol, PLS-DA was further studied to examine differences in metabolite profiles among the normal control, HFD and piceatannol groups. As shown in Fig. 3.6D, the score plot of PLS-DA showed a clear distinction among the normal control, HFD and piceatannol groups. Significant differences based on the results of a loading plot (Fig. 3.6E) were observed in the following classes of metabolites including lysophospholipids, fatty acids and bile acids ($p < 0.05$).

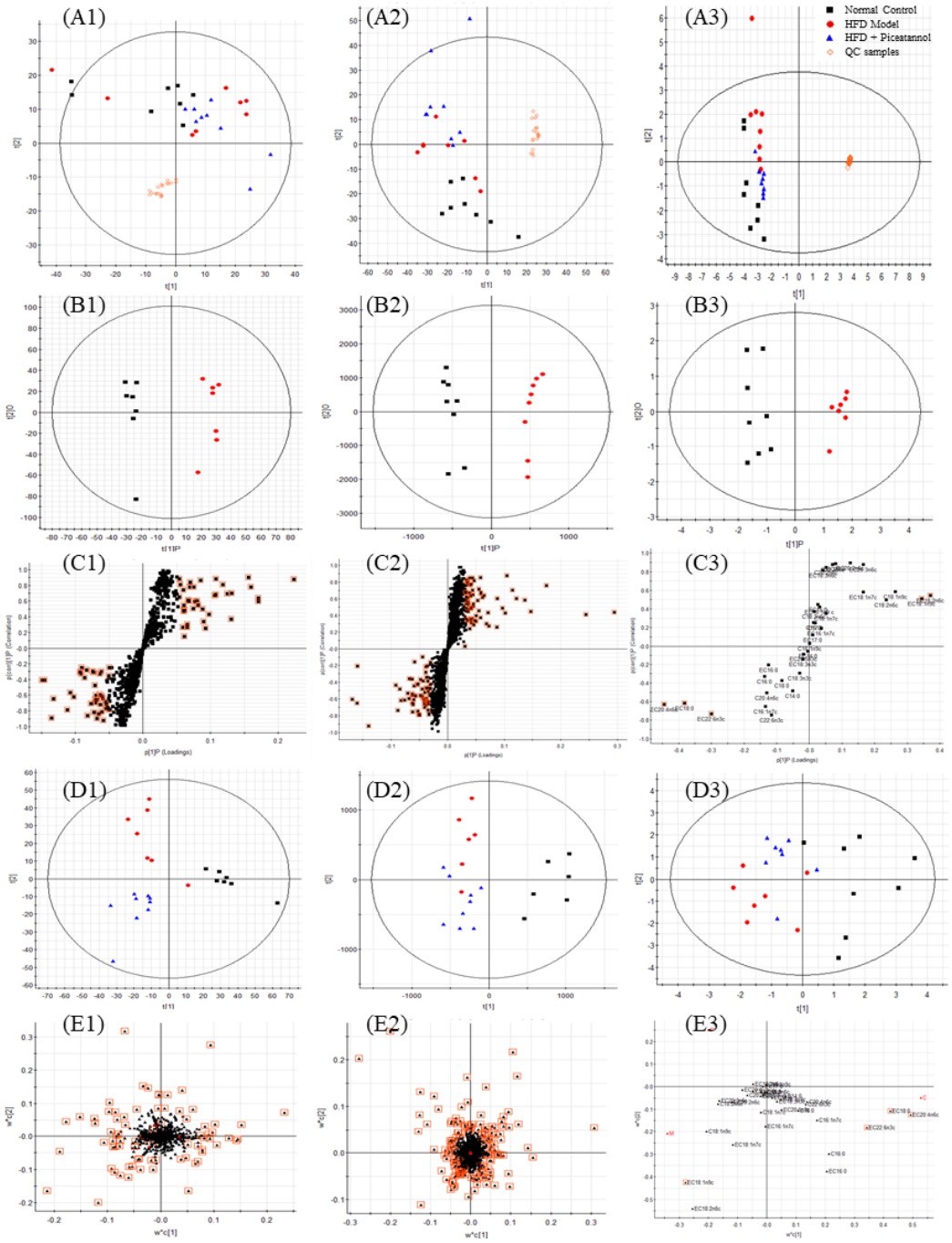


Fig. 3.6 Multivariate analyses of serum acquired using UPLC-Orbitrap-MS and GC-MS.

(A) PCA score plot, (B) OPLS-DA score plot, (C) S-loading plot of OPLS-DA, (D) PLS-DA score plot and (E) loading plot of PLS-DA of metabolites. (1) Negative and (2) positive ESI modes for UPLC-QTOF-MS; (3), GC-MS. Red squares highlighted in (C) and (E) are metabolites with VIP >1.5.

Table 3.5 Identification and fold change of metabolites compared with HFD model using UPLC-QTOF-MS and GC-MS.

Metabolites	Adducts	Retention time (min)	Theoretical m/z	Detected m/z	Mass error (ppm)	Fold change with respect to HFD model		CV (%) of QC samples
						Normal control	HFD + Piceatannol	
Taurocholate	$[M-H]^-$	4.33	514.2844	514.2834	-1.9	1.19	1.32	12.9
Taurohyodeoxycholate	$[M-H]^-$	4.31	498.2895	498.2882	-2.6	0.94	2.60^{*}	13.2
Taurohyodeoxycholate isomer	$[M-H]^-$	4.17	498.2895	498.2881	-2.8	1.10	3.79^{**}	21.2
Glycocholate	$[M+H]^+$	4.61	466.3164	466.3166	0.4	1.08	2.48^{***}	13.8
	$[M-H]^-$	4.62	464.3012	464.3006	-1.3	0.86	3.01^{***}	7.9
Glycohyodeoxycholate+	$[M+H-2H2O]^+$	4.66	414.3014	414.3004	-2.4	0.48	3.11^{*}	17.2
glycoursodeoxycholate	$[M-H]^-$	4.64	448.3068	448.3056	-2.7	0.28	3.43^{**}	7.9
Taurochenodeoxycholate	$[M-H]^-$	4.70	498.2895	498.2884	-2.2	1.33	2.74^{**}	11.4
Taurodeoxycholate	$[M-H]^-$	4.79	498.2895	498.2882	-2.6	0.26^{**}	1.90^{**}	6.6
Cholate	$[M+H-3H2O]^+$	5.19	355.2630	355.2636	1.7	1.07	0.47^{**}	17.5
	$[M-H]^-$	5.19	407.2803	407.2790	-3.2	1.29^{*}	0.63^{*}	5.6
Glycochenodeoxycholate	$[M-H]^-$	5.24	448.3068	448.3056	-2.7	0.56^{*}	1.75^{**}	8.1
Hyodeoxycholate + ursodeoxycholate	$[M+H-2H2O]^+$	5.35	357.2799	357.2791	-2.2	0.50^{**}	0.57^{**}	20.8
	$[M-H]^-$	5.34	391.2854	391.2842	-3.0	0.45^{***}	0.55^{***}	8.6
Glycodeoxycholate	$[M+H-2H2O]^+$	5.42	414.3014	414.3002	-2.9	0.16^{*}	1.86^{**}	17.7
	$[M-H]^-$	5.44	448.3068	448.3057	-2.5	0.15^{***}	1.60^{**}	7.5
Chenodeoxycholate isomer	$[M-H]^-$	5.87	391.2854	391.2841	-3.3	0.24^{***}	0.84	13.4
Chenodeoxycholate	$[M-H]^-$	6.77	391.2854	391.2842	-3.0	0.83	0.33^{**}	6.0
LysoPC(18:2)	$[M+H]^+$	7.40	520.3398	520.3401	0.57	0.39^{***}	0.76^{*}	12.5
	$[M+FA-H]^-$	7.40	564.3307	564.3300	-1.2	0.56^{***}	0.73^{*}	10.1
Deoxycholate	$[M-H]^-$	7.08	391.2854	391.2842	-3.0	0.19^{***}	0.56^{**}	16.2
LysoPC(20:3)	$[M+H]^+$	8.24	546.3554	546.3553	-0.2	0.21^{***}	0.69^{*}	13.4
	$[M+FA-H]^-$	8.26	590.3463	590.3455	-1.4	0.35^{***}	0.62^{**}	5.6
Unidentified	ESI-	8.46	/	455.3150	/	11.25^{***}	0.58	8.4
Unidentified	ESI-	8.52	/	466.3292	/	1.91^{***}	0.89	6.3

LysoPC(18:1)	$[M+H]^+$	8.61	522.3554	522.3555	0.2	0.65^{**}	0.79	13.1
	[M+FA-H]-	8.62	566.3463	566.3455	-1.5	0.64^{**}	0.69^{**}	14.9
LysoPE(20:1)	$[M-H]^-$	8.62	506.3252	506.3242	-2.0	0.55^{**}	0.59^{**}	18.6
LysoPC(20:2)	$[M+H]^+$	9.09	548.3711	548.3711	0.1	0.35^{***}	0.56^{**}	13.2
	$[M+FA-H]^-$	9.09	592.3620	592.3611	-1.5	0.22^{***}	0.59^{**}	9.6
LysoPC(20:1)	$[M+H]^+$	10.03	550.3867	550.3866	-0.2	0.51^{**}	0.75	15.3
LysoPC(20:0)	$[M+H]^+$	11.20	552.4024	552.4021	-0.5	0.31^{***}	0.56^{**}	18.1
	[M+FA-H]-	11.20	596.3933	596.3924	-1.5	0.24^{***}	0.70^{*}	17.3
4,7,10,13,16,19-Docosahexaenoic acid	$[M-H]^-$	11.33	327.2330	327.2318	-3.7	1.21^{**}	1.15	5.1
Arachidonic acid	$[M-H]^-$	11.49	303.2324	303.2320	-1.3	1.22^{*}	1.18	13.5
Palmitic acid	$[M-H]^-$	12.17	255.2330	255.2318	-4.7	1.35^{*}	1.33	5.7

EFA	**Molecular formula**	**Retention time (min)**	**Mass**	**Fold change with respect to HFD model**		**CV (%) of QC samples**
				Control	**Piceatannol +HFD**	
Esterified stearic acid (EC18:0)	$C_{19}H_{38}O_2$	14.14	298	1.22^{*}	0.98	1.51
Esterified oleic acid (EC18:1n9c)	$C_{19}H_{36}O_2$	14.52	296	0.64^{**}	0.66^{*}	1.12
Esterified linoleic acid (EC18:2n6c)	$C_{19}H_{34}O_2$	15.42	294	0.75^{**}	0.76^{*}	4.05
Esterified arachidonic acid (EC20:4n6c)	$C_{21}H_{34}O_2$	21.30	318	1.29^{**}	1.00	5.06
Esterified *cis*-4,7,10,13,16,19-docosahexaenoic acid (EC22:6n3c)	$C_{23}H_{34}O_2$	29.46	342	1.56^{***}	0.89	4.42

One-way ANOVA, followed by LSD *post-hoc* test with equal variance #p < 0.05, ##p < 0.01, ###p < 0.001; Tamhane *post-hoc* test with unequal variance *p < 0.05, **p < 0.01, ***p < 0.001.

n = 5–8.

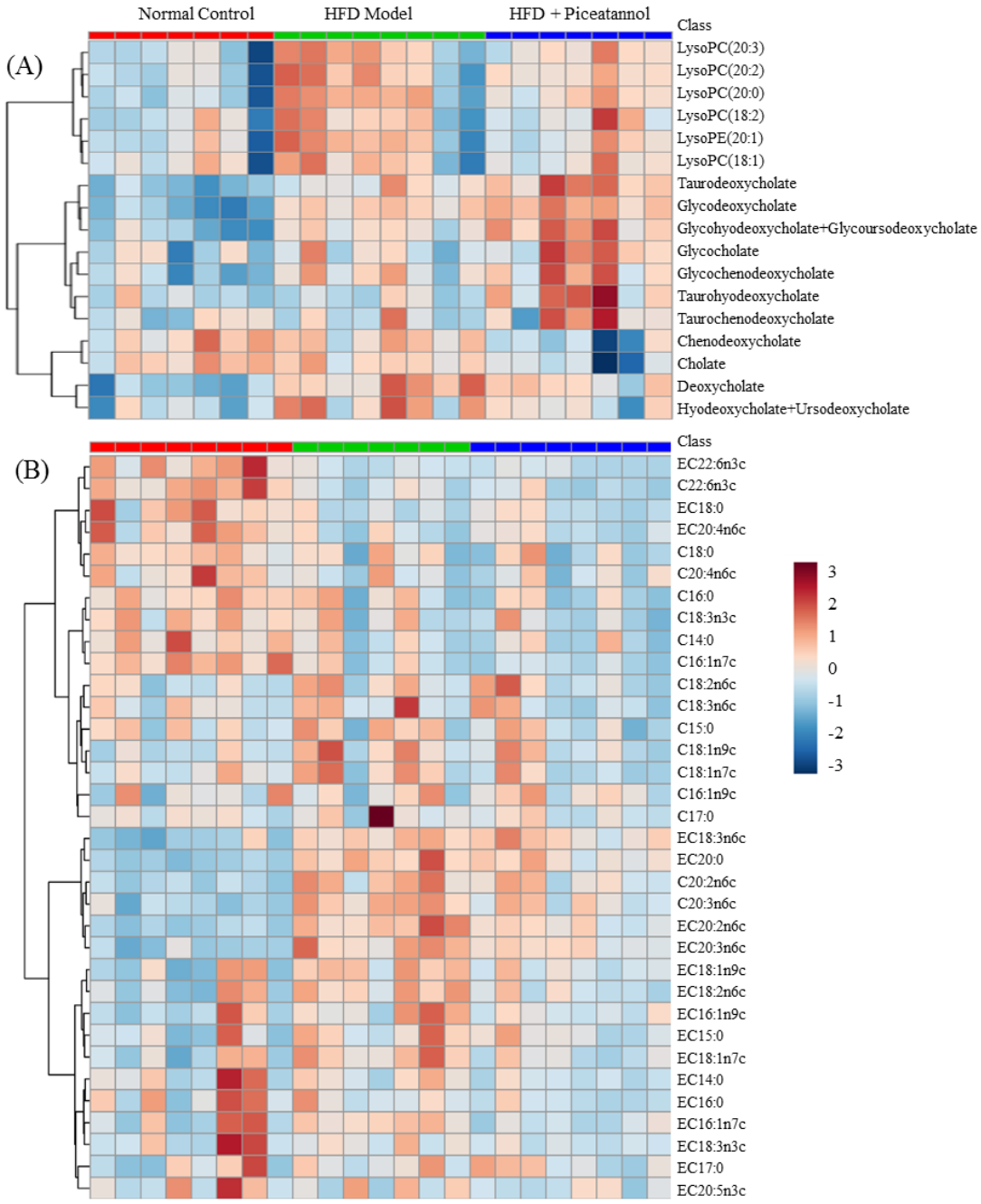


Fig. 3.7 Heatmap of (A) altered identified metabolites by piceatannol treatment acquired by untargeted serum metabolomics analysis using UPLC-QTOF-MS and (B) all GC-MS identified EFA and NEFA in normal control, HFD model and piceatannol-treated group.

3.3.5 Change of lysophospholipids in serum after piceatannol treatment

Our result showed that lysoPE (C20:1) and five lysoPCs (C18:1, C18:2, C20:0, 20:2 and 20:3) were increased in HFD models but restored in piceatannol-treated HFD models. Lysophospholipids play a structural role in the composition of cell membrane, but they are also involved in many cellular functions such as acting as messengers in signal transmission within a cell, activating cells, and synthesis of precursors of prostaglandins. In recent years, observations of the up-regulation of lysophospholipids, especially lysoPCs derived from hydrolysis of phosphatidylcholines, have been reported in hyperlipidemia-related researches, such as a study of atherosclerosis in a rabbit model (Liu, Peng, Jia, Cai, et al. 2014), HFD-fed Ossabaw pig (Hanhineva et al. 2013) and HFD-induced hyperlipidemic rats (Wu et al. 2014). Also, Frostegård's team (Frostegård 2010) pointed out that the toxicity of a high concentration of lysoPCs from the increased phosphorylcholine might cause autoimmune response and inflammation in various diseases related to hyperlipidemia.

3.3.6 Change of fatty acids in serum after piceatannol treatment

The OPLS-DA score plots from the data of LC-MS and GC-MS showed that both unsaturated and saturated NEFAs and EFAs were disturbed by HFD (Fig. 3.6B). There were lower circulating levels of free stearic acid, palmitic acid, free and esterified DHA as well as free and esterified arachidonic acid in HFD group compared with the normal control group (Table 3.5, $p < 0.05$, VIP > 1.5). Although piceatannol could not restore the above fatty acid levels, it significantly down-regulated the esterified oleic acid and linoleic acid levels and restored them to the normal level as suggested by the GC-MS results ($p < 0.05$, VIP > 1.5).

To further study the fatty acid metabolism of GC-MS results, the fatty acid product to its precursor ratio was evaluated. Such ratio commonly serves as an indicator of

endogenous fatty acid metabolism and an estimation of hepatic enzyme activities (Matthan et al. 2014). As shown in Fig. 3.8, the ratio of esterified arachidonic acid (EC20:4n6c) to esterified dihomo-γ linoleic acid (EC20:3n6c) was significantly up-regulated after the piceatannol treatment as compared with that of the HFD model ($p < 0.01$). A human study suggested that this ratio was positively associated with the activity of Δ5-desaturase (C20:4n6c / C20:3n6c) and higher activity of Δ5-desaturase was associated with lower coronary heart disease risk (Matthan et al. 2014). The present study also observed the similar trend that there was an inverse and non-linear correlation of the activity of Δ5-desaturase (EC20:4n6c / EC20:3n6c) with the atherogenic indexes of the rats ($R^2 = 0.7786$) (Fig. 3.9). This indicated that elevated activity of Δ5-desaturase promoted the conversion of dihomo-γ linoleic acid to arachidonic acid. A significant reduction of dihomo-γ linoleic acid by 20.8% was observed in the piceatannol-treated group ($p < 0.05$) compared with the HFD model while there was no significant difference in the free arachidonic acid level between these two groups ($p < 0.01$). Th observation of no significant difference may be because extra arachidonic acid generated could probably be further converted to pro-inflammatory mediators by the action of cyclooxygenase-1 and -2 isoenzymes (COX-1, COX-2), 5- and 15-lipoxygenases (5-LOX, 15-LOX) (Calder 2006) or converted to other long-chain ω6-fatty acids (Simopoulos 2008). As piceatannol is an inhibitor of COX-2 (Murias et al. 2004) and 5-LOX (Kutil et al. 2014), the arachidonic acid was probably converted to long chain fatty acids in the piceatannol-treated group. Furthermore, a non-alcoholic human fatty liver study (Kotronen et al. 2009) revealed that liver fat content was positively correlated to stearoyl-CoA desaturase 1 activity index (C18:1n9c / C18:0) and inversely correlated to hepatic elongase activity index (C18:0 / C16:0). Piceatannol treatment did not make any changes to the elongase

activity. Yet, it reduced the stearoyl-CoA desaturase activity from HFD, as evidenced by the smaller ratio of EC18:1n9c to EC18:0 than HFD group ($p < 0.01$), thereby limiting the major substrates (monounsaturated and polyunsaturated fatty acids) for synthesis of TG and other lipids (Ntambi and Miyazaki 2004). Fig. 3.10 gives a summary of the change of saturated fatty acid metabolism and ω-6 fatty acid metabolism with proposed action of piceatannol in the pathway.

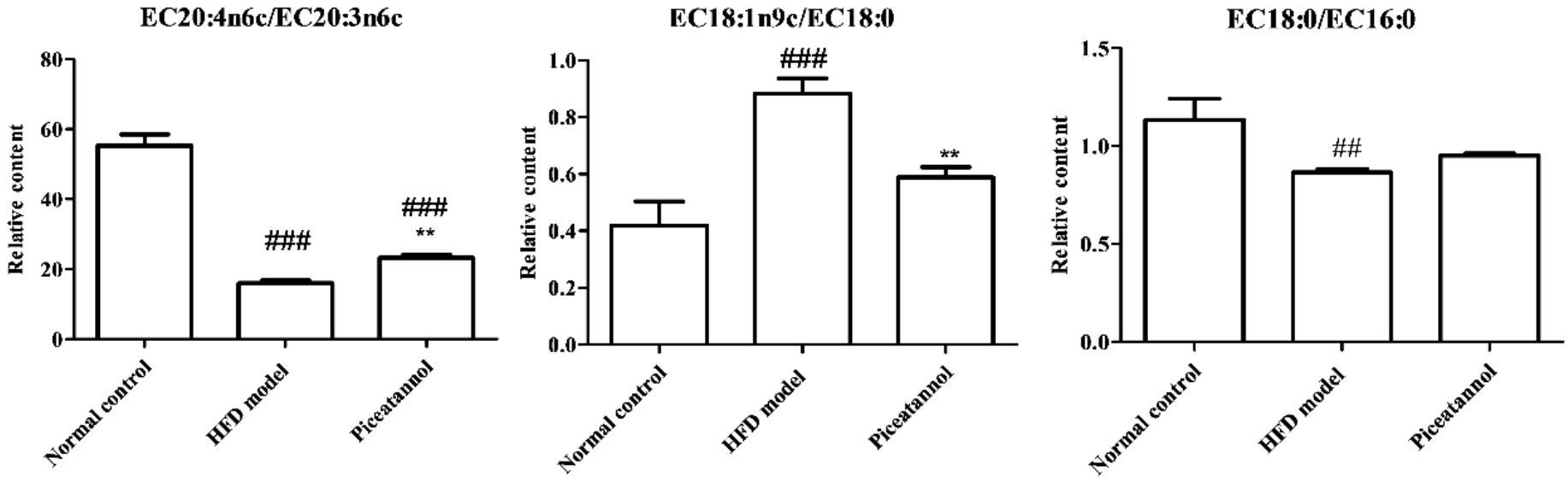


Fig. 3.8 Peak area ratios of EC20:4n6c to EC20:3n6c, EC18:In9c to EC18:0, EC18:0 to EC16:0 detected by GC-MS among the normal control, HFD model, and piceatannol-treated groups.

Data are expressed as mean ± SEM, n = 6–8.

One-way ANOVA, followed by LSD *post-hoc* test: ##p < 0.01 and ###p < 0.001 compared with the normal control group.

**p < 0.01 compared with the HFD model group.

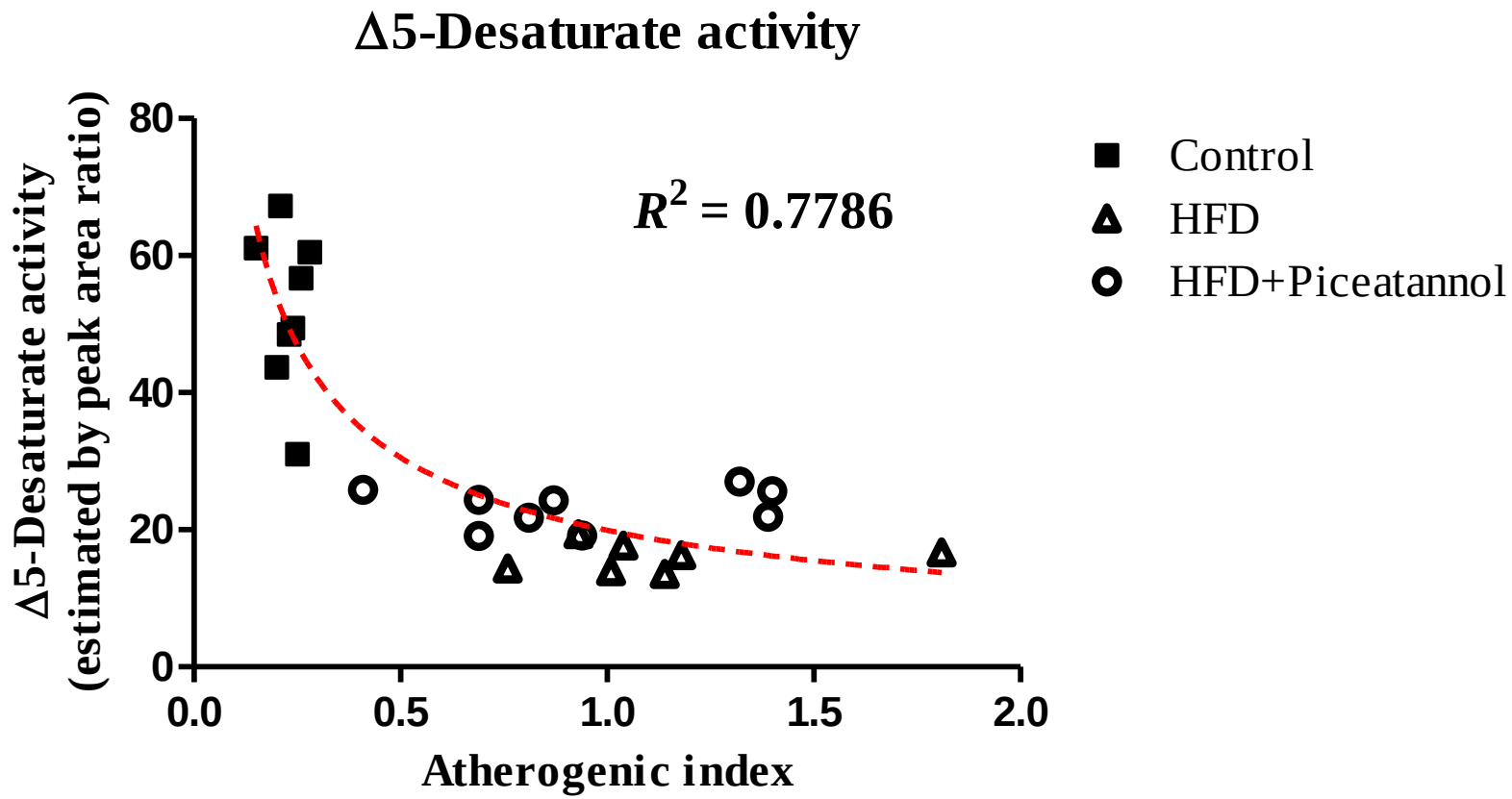


Fig. 3.9 Non-linear relationship of peak area ratios of EC20:4n6c to EC20:3n6c against atherogenic indexes of all rats from the three groups.

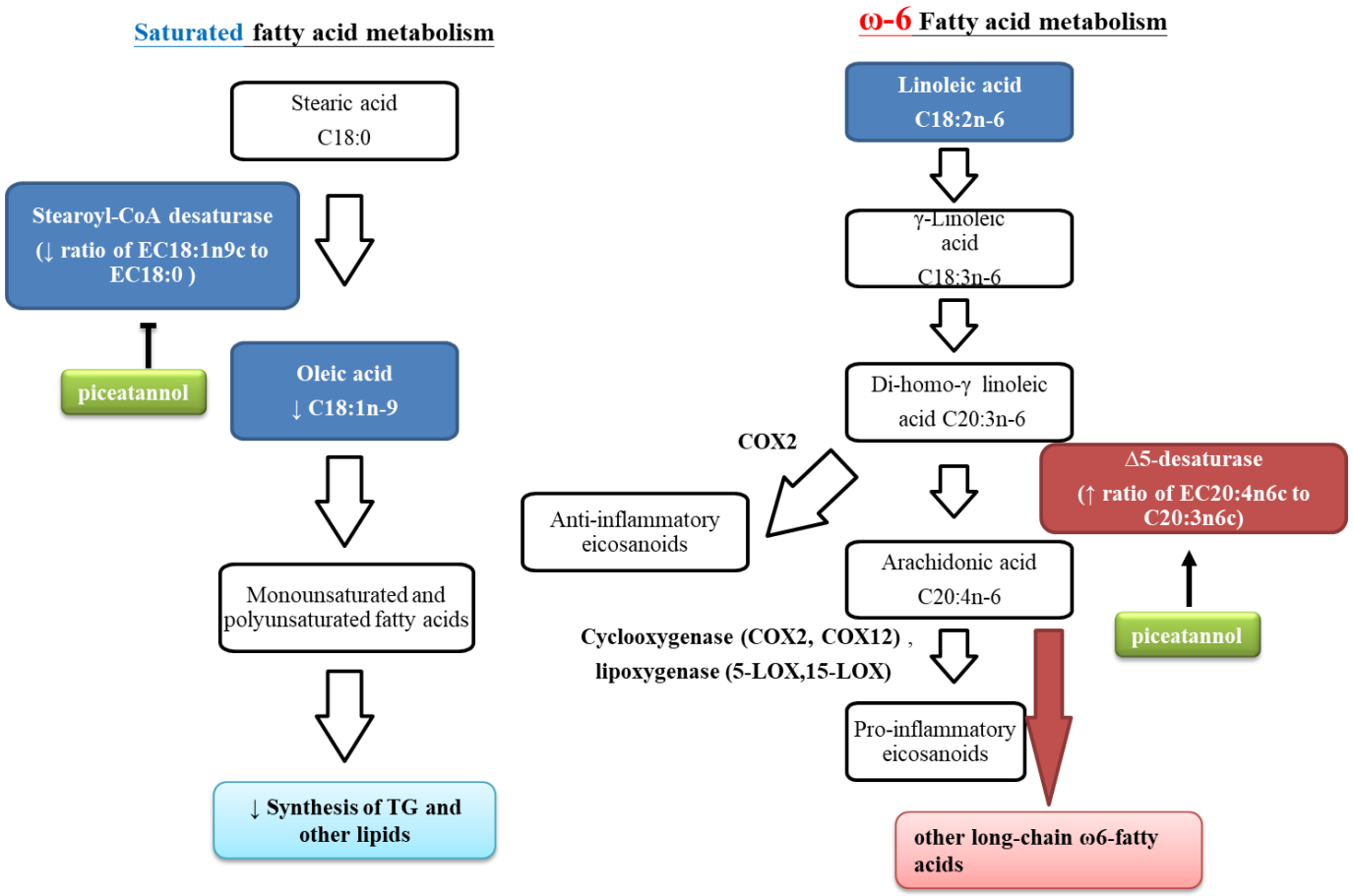


Fig. 3.10 A Simplified schematic diagram of proposed action of piceatannol in the simplified saturated fatty acid metabolism and ω-6 fatty acid metabolism

3.3.7 Change of bile acids in serum after piceatannol treatment

3.3.7.1 Piceatannol supplementation reduced primary and secondary bile acids but increased most conjugated bile acids

The results of untargeted UPLC-QTOF-MS analysis showed that bile acids and their conjugated bile acids were significantly disturbed by HFD as well as the piceatannol treatment. Their fold changes of UPLC-QTOF-MS are presented in Table 3.5.

In order to validate the results of untargeted UPLC-QTOF-MS analysis, a quantification of 16 bile acids was conducted using serum samples from the same animal study. Fig. 3.11 is a summary of a simplified pathway of bile acid biosynthesis and bar charts showing the contents of bile acids and conjugated bile acids in each group. Most results of targeted quantification method were consistent with that of untargeted metabolomics analysis. Some of the inconsistency between UPLC-QTOF-MS and UPLC-QQQ-MS/MS might be due to the signal enhancement or signal suppression effect of serum matrix during the ionization process of UPLC-QTOF-MS. This was improved by using similar internal standards as they shared similar chemical structures and co-eluted with its corresponding analytes. They could correct for the variations in ionization efficiency arising from the matrix.

The two peaks of hyodeoxycholate and ursodeoxycholate and the two peaks of glycohyodeoxycholate and glycoursodeoxycholate that were originally co-eluted in the untargeted UPLC-QTOF-MS were well separated and quantified in UPLC-QQQ-MS. Lithocholate and tauroursodeoxycholate were below the detection limit of both UPLC-QTOF-MS and UPLC-QQQ-MS. α-Muricholate and β-muricholate, which were overlapped with adjacent isomers in UPLC-QTOF-MS, were quantified in UPLC-QQQ-MS.

The UPLC-QQQ-MS study showed that primary bile acids (cholate, chenodeoxycholate, β-muricholate) did not change in the HFD model compared with normal controls but α-muricholate content was reduced. All were significantly decreased in piceatannol-treated group compared with HFD model. Untargeted and targeted study had the same trend in the secondary bile acid changes as well. The contents of secondary bile acids, deoxycholate, hyodeoxycholate and ursodeoxycholate, were significantly elevated in HFD model and reduced in concentration in the piceatannol supplementation to HFD rats.

The conjugated bile acid contents were increased more in the piceatannol treatment group than in the HFD model. The glycoursodeoxycholate, glycodeoxycholate and glycohyodeoxycholate contents were higher in both HFD models and piceatannol-treated groups compared with normal control. Yet, only glycocholate, glycochenodeoxycholate, taurodeoxycholate, taurochenodeoxycholate and taurohyodeoxycholate contents were increased in the piceatannol-treated groups but HFD models remained unchanged compared with normal control. In the contrary, the increase in glycoursodeoxycholate content was less in the piceatannol treatment group than HFD models.

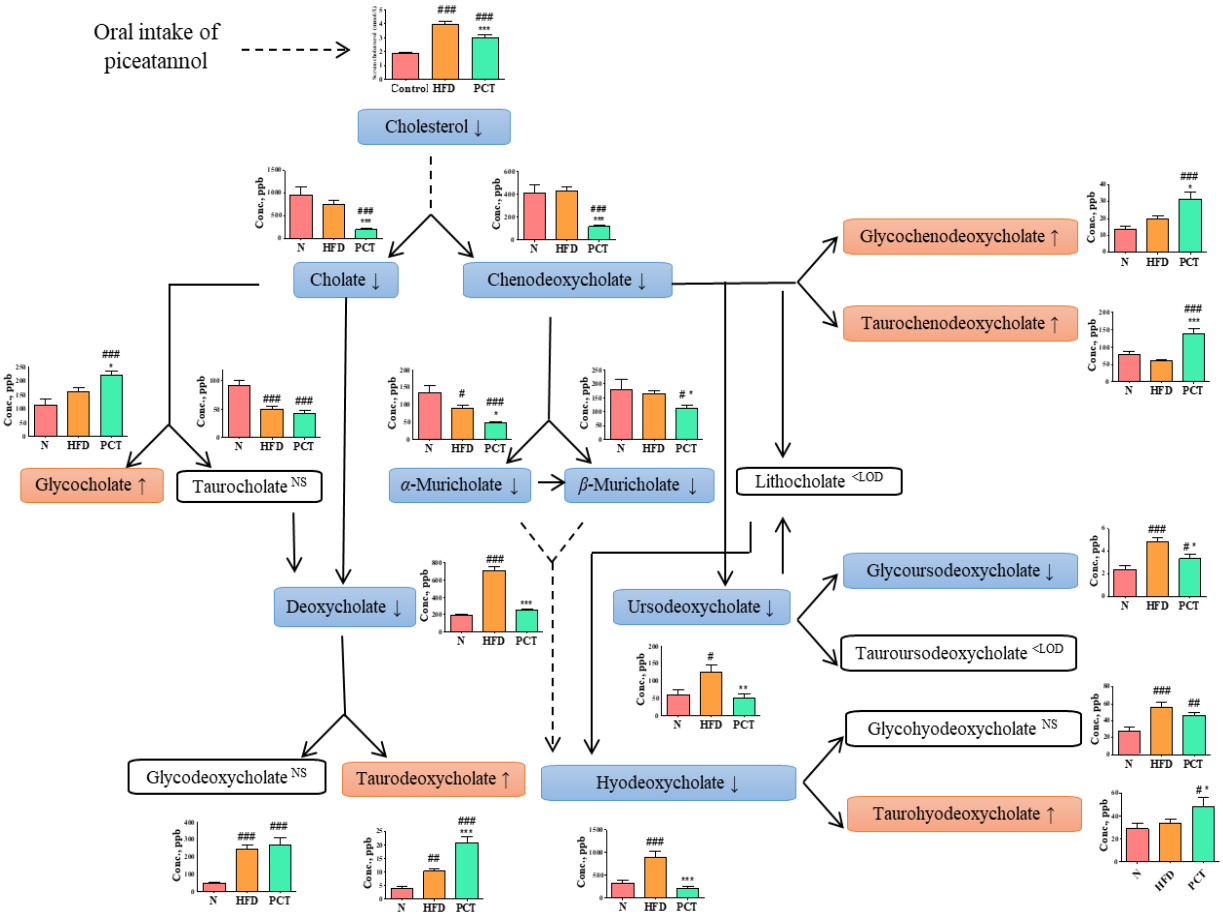


Fig. 3.11 A simplified diagram of bile acid biosynthesis altered by HFD model and piceatannol treatment, in which the quantified contents of different bile acids in serum were determined by UPLC-QQQ-MS/MS.

N, normal control; HFD, high-fat diet induced model; PCT, piceatannol-treated HFD model. One-way ANOVA, LSD *post-hoc* test: ***p<0.001, ** p< 0.01, * p <0.05 compared with the HFD model; ### p<0.001, ## p< 0.01, # p <0.05 compared with the normal control group. n=5-8. ↑: increase; ↓: decrease in piceatannol-treated group compared with HFD group. $^{<LOD}$, below limit of detection. ^{NS}, no significant change. Dotted line: skipped pathway.

3.3.7.2 Piceatannol might not rely on CYP7A1 action in elimination of cholesterol but lower the solubility of cholesterols bound on gut bile acid micelle

CYP7A1 is the rate-limiting enzyme in the bile acid biosynthetic pathway (Chiang 2009; Dawson 2016). Resveratrol has been shown to increase the expression of CYP7A1, which affected the bile acid biosynthetic pathway (Miura, Miura, and Yagasaki 2003; Zhu, Luo, and Jin 2008; Chen et al. 2012) and elevated the excretion of fecal bile acids (Zhu, Luo, and Jin 2008) to lower cholesterols in the blood circulation. It was originally hypothesized that piceatannol that has a similar chemical structure as resveratrol might also up-regulate CYP7A1, leading to higher concentration of circulating bile acids and lower cholesterols as observed here.

However, our result showed that the CYP7A1 protein expression levels in the liver of the HFD-fed model and simvastatin-treated HFD-fed group were lower than that in the liver of the normal control ($p < 0.001$) (Fig. 3.12). This was probably because the high cholesterol diet (2% cholesterols in the rat chow) of these two groups in this study might impair the bile acid biosynthetic pathway which was responsible for cholesterol elimination (Zinkhan et al. 2014), and so impaired the CYP7A1 protein expression.

The CYP7A1 protein expression of the piceatannol-treated HFD-fed group was lower than that of the HFD group ($p < 0.01$), suggesting that piceatannol might not rely on CYP7A1 action in elimination of cholesterol. The cholesterol-lowering effect of piceatannol might be contributed by the inhibition of dietary cholesterol absorption from the small intestine with the assistance of binding effects of piceatannol to conjugated bile acids (Montilla et al. 2004). Conjugated bile acids are amphipathic molecules with non-polar (steroid) and polar regions (glycine or taurine) and behave as biosurfactants (Hofmann 1963). Within the gut lumen, they form

micelles with cholesterols and other lipids to solubilize them and facilitate their absorption in the intestine. However, there was a region-specific interaction of conjugated bile acids with polyphenols found in grapes (Ngamukote et al. 2011) and tea (Ogawa et al. 2016) to lower the solubility of cholesterols bound in the bile acid micelle. The interaction might be hydrogen bonds formed between keto groups of bile acids and resveratrol's hydroxyl groups (Atanacković et al. 2009). These binding effects in the micelle probably resulted in the delay of cholesterol absorption. As piceatannol is a polyphenol and has one more hydroxyl group than resveratrol, it might also interact with the conjugated bile acids to lower the solubility of bile acid micelles that was bound to cholesterols. It was found that polyphenols found in red wine probably inhibited the absorption of lipids and exogenous cholesterols in the small intestine, resulting in a rise in the cholesterol sulphate content in faeces (Jimeez-Giro et al. 2014). Lower dietary cholesterol absorption was likely to reduce the hepatic cholesterol levels and so the stimulation to CYP7A1 receptor, resulting in a lower protein expression of CYP7A1 in the piceatannol-treated group compared with the HFD model, resulting smaller contents of most primary and secondary bile acids in the piceatannol-treated group.

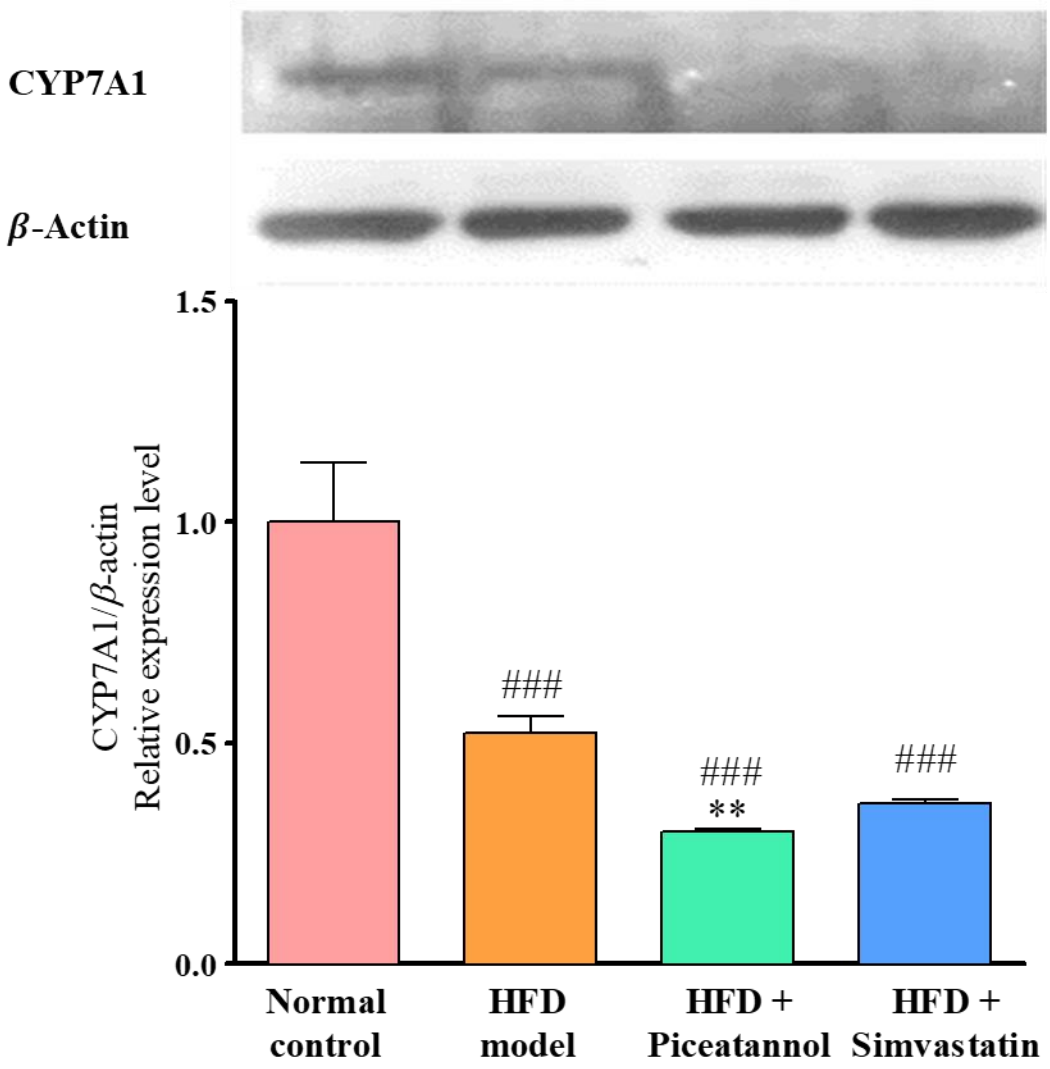


Fig. 3.12 Representative Western blots (upper figure) and the graph (bottom figure) demonstrating the quantitative comparisons of protein expressions of CYP7A1 in the hepatocytes of the normal control, HFD model, piceatannol-treated and simvastatin-treated groups.

The expression level of each protein was normalized to that of the *β*-actin protein in each sample. Data are expressed as mean ± SEM, $n = 8$. One way-ANOVA, followed by LSD *post-hoc* test: ### $p < 0.001$ compared with the normal control group. ** $p < 0.01$ compared with the HFD group.

3.3.7.3 Piceatannol might interfere the biotransformation of secondary bile acids by modification of the microbial bile acid metabolism

Primary bile acids are produced from cholesterols in the hepatic cells and metabolized in the intestine by the gut microbiota to generate secondary bile acids (Qi et al. 2015). For instance, cholic acid (a primary bile acid) is metabolized into deoxycholic acids (a secondary bile acid) by the 7α-dehydroxylation reaction with the intestinal microbiota (Wang et al. 2019).

In this study, the contents of secondary bile acids (deoxycholate, hyodeoxycholate and ursodeoxycholate) were in excess after consumption of HFD and largely reduced after the piceatannol supplementation to HFD rats. A research showed that dietary supplementation of polyphenols, such as curcumin and caffeic acid, lowered faecal deoxycholic acid which was increased in the HFD-fed rats (Han et al. 2009).

These changes were consistent in the HFD model and the piceatannol-treated group shown in the current study. It was likely that piceatannol, which is also a polyphenol, might interfere the gut microflora for biotransformation of secondary bile acids and deconjugation of bile salts in the gut. Hijona *et al.* reported that in a genetically obese Zucker rat model that took normal diet and piceatannol together, piceatannol supplementation did not change the gut microbial composition but it changed the amounts of several species of *Firmicutes* and *Bacteroidetes* phyla (Hijona et al. 2016). Another study demonstrated that resveratrol improved the HFD-induced intestinal microbial imbalance (Qiao et al. 2014). This study proposed that piceatannol supplementation to HFD-fed rats altered the gut microbial contents and reduced the biotransformation of primary bile acids into secondary bile acids and deconjugation of bile acids. Modification of the microbial bile acid metabolism induced by piceatannol or other stilbenes supplementation may play a vital role in the intestinal health. The

present study offered some evidence of this and further research is certainly warranted to provide a better picture on the role of intestinal microbiota in improving the health of the host via the consumptions of phytochemicals.

3.4 Conclusion

The results acquired in this study supported the cardioprotective and lipid-lowering effects of piceatannol. Untargeted metabolomic approach has be employed to characterize the serum metabolic profile of HFD-fed rats that received piceatannol supplementation to provide more insights into the underlying mechanisms of piceatannol in regulating the atherogenic index. GC-MS indicated that piceatannol restored the circulating esterified fatty acid profiles whereas over 20 biomarkers identified from UPLC-MS exhibited that piceatannol interrupted bile acid metabolism and glycerophospholipid metabolism. The observed changes of fatty acids and their ratios suggested the regulation of fatty acid pathways was to down-regulate stearoyl-CoA desaturase and up-regulate Δ5-desaturase activities. Quantitative analysis further validated the piceatannol's action in reducing primary and secondary bile acids and upregulating most conjugated bile acids. Reduction of CYP7A1 protein expression and reduction in primary and secondary bile acids in the piceatannol-treated group compared with that of HFD-fed model suggested that piceatannol did not rely on the CYP7A1 activation for removal of extra dietary cholesterols. But it might be because of the reduction in dietary cholesterol absorption in the intestine, which was likely due to the binding effect of piceatannol to the bile acids and cholesterols with bile acid micelles. The observation of lower serum levels of secondary bile acids and higher conjugated bile acids in the piceatannol-treated HFD-fed group gave us hints that the action of piceatannol may be related to the modification of gut microbial contents,

thereby reducing biotransformation of primary bile acids into secondary bile acids and deconjugation of bile salts. To the best of my knowledge, this is the first study showing piceatannol supplementation would reduce the serum cholesterol levels from HFD by regulating the bile acid biosynthesis and has been published (Sham et al. 2017). This knowledge would contribute to the development of new functional foods for CVDs and offer support of consumption of piceatannol-rich food. These conclusions may also be applied to other plant polyphenols and propose a new direction in understanding the health benefits of this kind of phytochemicals. More studies on the fecal bile acid compositions and the gut microbial population and content are warranted to support the proposed action of piceatannol on microbial bile acid metabolisms.

Chapter 4 The protective effect of water extract of Polygoni Cuspidati Rhizoma et Radix on NAFLD of rats induced by HFD using serum metabolomics and liver lipidomics

4.1 Background

NAFLD is defined by the accumulation of fatty acid content greater than 5% of liver weight in the absence for secondary causes of hepatic fat accumulation (Nassir et al. 2015). The common primary causes of NAFLD are related to metabolic disorders such as type 2 diabetes, hyperlipidemia, obesity and insulin resistance. The secondary causes of hepatic fat accumulation include excess alcohol consumption (a threshold of >20 g/day for women and >30 g/day for men is typically adopted) (Anstee, McPherson, and Day 2011), viral infection, use of steatogenic drugs and hereditary diseases (Chalasani et al. 2012; Nassir et al. 2015). NAFLD is the mildest form of a histological spectrum of hepatic diseases, starting from simple steatosis (fatty liver) to more progressive non-alcoholic steatohepatitis, cirrhosis and in some cases hepatocellular carcinoma (Anstee, McPherson, and Day 2011; Sayiner et al. 2016). NAFLD is a growing health problem in the world. In 2016, a meta-analysis of 86 studies involving over 8.5 million persons from 22 countries indicated that the global prevalence of NAFLD was 25.24% (Younossi et al. 2016). It showed that the highest metabolic comorbidity associated with NAFLD is hyperlipidemia (around 70%), followed by obesity (over 50%). NAFLD is becoming a public health concern due to the rising obesity epidemics and increasing popularity of metabolic syndrome (Wree et al. 2013).

In spite of the incomplete understanding of the pathogenesis of NAFLD, it is believed that one of the major causes is an imbalance between lipid acquisition (uptake of fatty acids and de novo

lipogenesis) and removal (mitochondrial fatty acid oxidation and export of triglycerides as a component of very-low-density lipoprotein particles) in the liver which leads to lipid accumulation (Kawano and Cohen 2013).

The classical theory of NAFLD pathogenesis is the multiple "hits" theory (Day and James 1998). The first "hit" is lipid accumulation leading to inflammation in liver. The second "hit" is a growth in oxidative stress and lipid peroxidation. The first "hit" is linked with insulin resistance and raises the susceptibility to the second "hit", resulting in progression of NAFLD to non-alcoholic steatohepatitis (NASH) (de Castro and Calder 2018). The early stages of the diseases (the first and second "hits") are reversible in some patients after appropriate treatments (de Castro and Calder 2018).

There are limited approaches for treating NAFLD (Day 2011). The common potential ways targeting NAFLD are medication treatments without strong evidence but for hepatic protection (Dajani and AbuHammour 2016), such as statin, vitamin E and pioglitazone (Ipsen, Lykkesfeldt, and Tveden-Nyborg 2018). Yet, these drugs are associated with possible adverse effects, which limit their treatment potential. For instance, vitamin E is a potent antioxidant to reduce oxidative stress in NAFLD, but its long-term use may increase the incidence of stroke (Schürks et al. 2010) and prostate cancer (Klein et al. 2011). Prolong use of pioglitazone is correlated to reduced bone density (Pop et al. 2017) and bladder cancer (Tang et al. 2018). Lifestyle modifications, such as dietary intervention, more physical activity and weight loss, are strongly recommended for relieving the symptoms with NAFLD (Ueno et al. 1997; Cave et al. 2007). However, lifestyle interventions are very difficult to comply (Middleton, Anton, and Perri 2013), suggesting that pharmacological therapy may be more effective for some patients.

Hundreds of Chinese herbal medicines were reported to be effective for treatment of hyperlipidemia, especially the cases related to HFD in the past few decades (Xie, Zhao, and Du 2012). Chinese herbs and classic traditional Chinese medicine (TCM) formulae could improve hyperlipidemic conditions with multiple mechanisms of actions similar to conventional Western drugs (Sham et al. 2014a). TCM is a possible answer for the alternative treatment for NAFLD. As NAFLD is caused by multiple factors, the multi-targeted nature of TCM may be more effective in regulating the conditions. The use of TCM to prevent and treat NAFLD has also received increasing attention due to its low risk of side effects and high efficacy (Yao et al. 2016).

Four TCM formulae were reported to be effective for the treatment of NAFLD (Fan 2004; Ji et al. 2008; Wang Lu-wen et al. 2005; Feng et al. 2013; Zhang et al. 2008). They all contained the same Chinese herb, Polygoni Cuspidati Rhizoma et Radix (PCRR) as one of the components. Also, previous studies found that resveratrol and piceatannol had strong hypolipidemic effect (Miura, Miura, and Yagasaki 2003; Zhu, Luo, and Jin 2008; Sham et al. 2017; Simental-Mendía and Guerrero-Romero 2019). PCRR is rich in resveratrol and resveratrol glucoside (polydatin). Piceatannol and resveratrol shared similar chemical structure. Piceatannol is in fact a natural analogue and metabolite of resveratrol and is found in the PCRR water extract used in this study. PCRR might play a role in these formulae in treating NAFLD.

Polygoni Cuspidati Rhizoma et Radix (PCRR) is the dried rhizome and root of *Polygonum cuspidatum* Sieb. et Zucc. and is called “Hu Zhang” (“虎杖”) in Chinese. It is listed in the Pharmacopoeia of the People’s Republic of China 2015 (Chinese Pharmacopoeia Commission 2015) and Hong Kong Chinese Materia Medica Vol. 4 (Hong Kong Chinese Materia Medica Standards Office 2012). PCRR is also used as folk medicine in Japan and Korea (Zhang et al.

2013). The action of PCRR according to traditional Chinese medicine (TCM) theory is to drain dampness jaundice and clear heat and remove toxin. Its medical uses by TCM practitioners are very diverse, from treatment of cough, phlegm, burns, jaundice to hyperlipidemia (Chinese Pharmacopoeia Commission 2015) . Recent clinical and pharmacological studies had indicated that PCRR has antimicrobial, antiviral, anti-inflammatory, cardioprotective and neuroprotective functions (Zhang et al. 2013).

Metabolomics, which aims at studying the whole metabolome, aligns with the holistic concept of TCM (Li and Yang 2008; Li, Yang, and Gong 2009). It provides another perspective in understanding the effects of the TCM inside the body (Liu et al. 2012; Dai et al. 2011; Lv et al. 2010; Liu, Peng, Jia, Cai, et al. 2014; Guo et al. 2014). Untargeted serum metabolomics and hepatic lipidomic profiling were employed to understand the protective effects of PCRR supplementation on lipid metabolism on HFD-induced NAFLD rats in the present study.

4.2 Materials and method

4.2.1 Materials

Isopropanol, chloroform, acetonitrile and methanol (HPLC grade) were purchased from Duksan (ANSAN-SI, South Korea). Distilled water was purified using a Milli-Q water purification system (Millipore, Bedford, MA, USA). Internal standards, succinic acid-2,2,3,3-D_4 and C19:1n9c were from Sigma Aldrich (St. Louis, MO, USA). D-glucose (U-$^{13}C_6$), cholic acid-2,2,4,4-D_4 and L-tryptophan-(indole-D_5) were from Cambridge Isotope Laboratories (Tewksbury, MA, USA).

Ammonium formate, formic acid, hyodeoxycholic acid, ursodeoxycholic acid, sodium chenodeoxycholate, sodium tauroursodeoxycholate and sodium taurochenodeoxycholate were purchased from Sigma-Aldrich (St. Louis, MO, USA). α-Muricholic acid, deoxycholic acid

and taurodeoxycholic acid were purchased from Santa Cruz Biotechnology (Dallas, Texas, USA). Glycoursodeoxycholic acid and ursodeoxycholic acid were purchased from Cambridge Isotope Laboratories (Tewksbury, MA, USA). Sodium glycocholate hydrate was purchased from Acros Organics (Morris Plains, NJ, USA). Cholic acid, *β*-muricholic acid, glycodeoxycholic acid, glycohyodeoxycholic acid, taurocholic acid, sodium glycochenodeoxycholate, taurohyodeoxycholic acid and lithocholic acid were obtained from Steraloids (Newport, RI, USA).

Piceatannol (98% confirmed by HPLC) and simvastatin (20-mg tablets, 10% w/w, confirmed by HPLC) were obtained from Merck Sharp & Dohme (Hangzhou, China) and Nanjing Zelang Medical Technology Co. Ltd. (Nanjing, China), respectively. Polydatin was purchased from The Testing Laboratory for Chinese Medicine (Hong Kong University of Science and Technology, Hong Kong, China). Emodin, resveratrol, citric acid, gallic acid, catechin, epicatechin and physcion were purchased from Sigma-Aldrich (St. Louis, MO, USA). Emodin 8-*O*-*β*-D-glucoside, torachrysone 8-*O*-*β*-D-glucoside, physcion 8-*O*-*β*-D-glucoside were supplied by Chengdu Must Biotechnology (Chengdu, China). All the above had purity ≥ 97%. Water was purified using a Milli-Q Advantage A10 water purification system (Millipore, MA, USA) before use in this study.

4.2.2 Source and preparation of PCRR water extract

PCRR was obtained from Eu Yan Sang Company Limited (Hong Kong, China). 1 kg of dried PCRR decoction pieces was extracted with 15L boiling water for 2 h, mimicking domestic consumption of TCM. The water extract was centrifuged at 4000 rpm for 10 min. The supernatant was freeze-dried (Labconco, Freezone 6) to collect dried water extract.

Fig. 4.1 Decoction pieces of Polygoni Cuspidati Rhizoma et Radix.

4.2.3 Animal studies

Sprague-Dawley male rats of three months old with 160 g ± 10 g body weight were obtained from the Guangdong Medical Laboratory Animal Center (Guangzhou, China). Rats were housed in a temperature-controlled room (20 ± 1 ℃) with 60% ± 10% humidity, free access to water and rat chow and a regular 12 h/12 h light/dark cycle. After acclimation for seven days in the laboratory environment, 48 rats were randomly assigned into six experimental groups (*n* = 10 per group): (1) the normal control group, (2) HFD-fed model group, (3 and 4) low- and high-dose simvastatin (3 and 10 mg/kg bw/day) treated HFD-fed groups, (5 and 6) PCRR water extract in low and high doses (150 and 450 mg/kg bw/day) treated HFD-fed group. The control group was fed with normal rat chow [fat (~10%), protein (~14%) and carbohydrate (~76%)]. All of the other groups were fed with HFD which was composed of additional 2% pure cholesterol, 1% cholic acid and 5.5% peanut oil to the normal rat chow. The diets were purchased from Guangdong Medical Laboratory Animal Center (Guangzhou, China). The treatment groups were administrated with their corresponding drugs by oral gavages once per day for four weeks while the rats in the normal control group and HFD-fed model were administrated with distilled water. The body weight of each rat was monitored daily during the whole experimental period. After four weeks, the rats were fasted overnight and then sacrificed with carbon dioxide asphyxiation, followed by collection of tissues and blood for further analysis.

In animal experiments, the two doses of PCRR water extract (150 and 450 mg/kg bw/day) were determined upon the recommended human dose of PCRR decoction pieces (10 - 30 g) (Chinese Pharmacopoeia Commission 2015), the effective doses of polydatin used in rat and mice studies of diabetes, NAFLD and hyperlipidemia (Arichi et al. 1982; Zhang, Tan, et al. 2012; Hao et al. 2014; Wang et al. 2016) and the doses used in two rat studies that used crude PCRR decoction pieces (Jiang et al. 2009; Yiyong 2016) In this animal study, the doses of (150 and

450 mg/kg bw) were converted to the weight of crude PCRR decoction pieces were 2 g (150mg PCRR water extract /7.5% yield of water extract), and 6 g (450mg PCRR water extract /7.5% yield).

The experimental protocol was conducted under approval of the Animal Subjects Ethics Sub-committee of The Hong Kong Polytechnic University (ASESE no. 05/21) and the animal license issued by the Department of Health, the Government of the Hong Kong Special Administrative Region. All procedures were applied in compliance with the Guide for the Care and Use of Laboratory Animals distributed by the US National Institutes of Health and the principles were outlined in the Declaration of Helsinki.

4.2.4 Quantitative analysis of PCRR water extract using High Performance Liquid Chromatography-Diode Array Detector analysis (HPLC-DAD)

100 mg of dried PCRR water extract was accurately weighed and ultrasonicated with 10 mL methanol for 30 min. The extract methanol solution was filtrated with a syringe filter (0.45 μm). A 10 μL aliquot was injected into the HPLC-DAD. The HPLC separation was conducted on an Alltima HP C_{18} column (250 mm × 4.6 mm, 5 μm; Grace, USA) using an Agilent 1100 liquid chromatography system equipped with a quaternary solvent delivery system, an autosampler and a DAD detector. A detection wavelength of 290 nm was selected. The mobile phase consisted of mobile phase A (water) and B (0.05% methanol in acetonitrile, v/v) in the elution gradient: 0-15 min, 20% B; 25 min, 50 % B; 30 min, 70% B; 40-45 min, 100 % B. A 10-min post-run time was used to fully equilibrate the column. The flow rate was set at 0.8 mL/min. The temperatures of sample chamber and HPLC column were at room temperature.

4.2.5 Qualitative chemical analysis of PCRR water extract with UPLC-Orbitrap-MS

PCRR water extract was further analyzed using a more advanced platform, high-resolution UPLC-Orbitrap-MS for discovery of novel compounds. 0.5 g dried PCRR water extract was

ultrasonicated with 50 mL 70% ethanol for 30 min. Then, 2 mL extracted solution was centrifuged at 18700 × g for 15 min. An aliquot of 3 μL supernatant was injected into a Waters ACQUITY UPLC system. The separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm) with HSS T3 pre-column (2.1 mm × 5 mm, 1.8 μm, Waters Corporation, Milford, MA). The mobile phase consisted of combinations of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 0.3 mL/min with elution gradient as follows: 0 min, 5% B; 10 min, 20% B; 15 min, 30% B; 20 min, 50% B; 22 min, 95% B; 22.1-25 min, 5% B. Column and sample chamber temperature were 40 ℃ and 6 ℃ respectively.

Mass spectrometry analysis was conducted using a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a heated electrospray ionization (H-ESI) interface (Thermo Fisher, Waltham, MA, USA). The mass-spectrometric conditions were optimized as follows: spray voltage, 3500 V in positive H-ESI mode and 2300 V in negative H-ESI mode; ion transfer tube and vaporizer temperature, 300 ℃. Nitrogen gas was used as the sheath gas and the aux gas at a flow rate of 30 and 10 L/min, respectively. The instrument was operated in a data-dependent acquisition mode, with full MS scans over a mass range of m/z 90–1000 with detection in the Orbitrap (120,000 resolution) and with auto gain control (AGC) set to 80000 and a maximum injection time at 100 ms. In each cycle (0.6 s) of data-dependent acquisition analysis, the most intense ions with intensity threshold above 20000 were selected for fragmentation at normalized collision energy of 25±5% higher energy collisional dissociation (HCD). The number of selected precursor ions for fragmentation was determined by the “Top Speed” acquisition algorithm. Fragment ion spectra were acquired in the Orbitrap (15,000 resolution) with an AGC of 20000 and a maximum injection time of 35 ms for Orbitrap MS^2 detection. All the data analysis was carried out using the Thermo Xcalibur Qual Browser software (Thermo Fisher Scientific).

4.2.6 Serum metabolomics using UPLC-QTOF-MS

4.2.6.1 Serum sample preparation method

Serum samples were thawed at 4 °C and vortexed before preparation. 100 µL serum from all samples were deproteinated with cold 300 µL methanol containing internal standards [100 ppm C19:1n9c and 40 ppm L-tryptophan-(indole-D_5), succinic acid-2,2,3,3-D_4 and D-glucose (U-$^{13}C_6$)] and vortexed for 30 s. The mixture was kept at -20 ℃ overnight before centrifugation at 18790 ×*g* for 20 min at 4 ℃. 340 µL supernatant was dried with nitrogen gas and stored at -80 ℃. Before UPLC-MS analysis, it was reconstituted in 100 µL 5% acetonitrile in water as initial gradient of UPLC, vortexed for 30s followed by centrifugation at 18790 ×*g* for 20 min at 4 ℃.

4.2.6.2 Quality control sample preparation method

An aliquot of 20 µL of each serum sample was pooled, vortexed and aliquoted to provide QC samples and kept at -80°C until use. For each analytical batch, QC samples went through extraction protocols as described below similar to all other samples. Before the start of chemical analysis, five repeated injections of the same QC sample were used to verify the working condition of the instruments. Afterwards, a QC sample was injected to monitor the stability of the instrument after every five-sample run.

4.2.6.3 UPLC-QTOF-MS condition

A 3 µL aliquot was injected into a Waters ACQUITY UPLC system. The separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 µm) with HSS T3 pre-column (2.1 mm × 5 mm, 1.8 µm, Waters Corporation, Milford, MA). The mobile phase consisted of combinations of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 0.3 mL/min with elution gradient as follows: 0-1 min, 5% B; 5 min, 35% B; 10 min, 50% B; 18 min, 65% B; 19-23 min, 95% B. A 3-min post-run time

was set to fully equilibrate the column between injections. Column and sample chamber temperature were set at 40°C and 6°C, respectively.

Mass spectrometry (MS) was performed on a Waters SYNAPT G2 Q-IM-TOF HDMS system (Waters, Milford, USA) operating in an electrospray ion source (ESI) in both positive and negative modes. Nitrogen and argon were used as cone and collision gases. The desolvation gas flow was set to 600 L/h at a desolvation temperature of 400°C, and the cone gas was set to 40 L/h. The source temperature was set at 120°C. The capillary voltages in positive and negative ion modes were 3 kV and 2.2 kV, respectively. The sampling and extraction cone voltages were 40 V and 4 V, respectively. The scan time was 0.5 s with a 0.024 s interscan delay. Data scan ranging from m/z 50 to 1000 was recorded in the centroid data format. For accurate mass acquisition, a lock-mass of leucine enkephalin was used to monitor for positive ion mode ($[M+H]^+$: m/z 556.2771) and negative ion mode ($[M-H]^-$: m/z 554.2615) to ensure accuracy during MS analysis. MS/MS analysis was carried out to study the structure of potential biomarkers. In this section, the collision energy was set between 5 and 50 eV according to the extent of a metabolite's fragmentation.

4.2.7 Liver lipidomics

4.2.7.1 Liver sample preparation method

For liver lipid extraction, this procedure had been optimized according to previous methods (Bird et al. 2013). 10 mg lyophilized rat liver samples were homogenized with 800 μL chloroform / methanol mixture (1: 1, v / v) containing internal standard mixture (1 ppm succinic acid-2,2,3,3-D_4, 2ppm L-tryptophan-(indole-D_5), 2ppm cholic acid-D_4 and 100ppm C19:1n9c) at 4°C with Precellys Evolution homogenizer (Bertin Technologies). After addition of 800 μL of water, vortexing and centrifugation at 18700 g for 15 min, 180 μL of lower organic phase was isolated. The mixture was dried under nitrogen prior to storage at -80 °C. The dried sample

was reconstituted in 160 μL acetonitrile / isopropanol / water (65: 30: 5, v / v / v) and centrifuged at 18700 g for 15 min, then subjected to LC-MS analysis.

4.2.7.2 Quality control sample preparation method

An aliquot of 20 μL of each supernatant in the lower organic phase of liver sample was pooled, vortexed and aliquoted to provide QC samples. They were dried by gentle nitrogen gas and kept at -80°C until use. For each analytical batch, QC samples went through extraction protocols as described in the serum QC sample preparation steps.

4.2.7.3 UPLC-Orbitrap-MS analysis

For UPLC-MS analysis, a Thermo Orbitrap Fusion Lumos Tribrid Mass Spectrometry (Thermo Fisher Scientific, Waltham, MA, USA) was connected to Waters Acquity UPLC System (Waters Corp., Milford, USA) equipped with a heated electrospray ionization (H-ESI) source. UPLC separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8μm) with HSS T3 pre-column (2.1 mm × 5 mm, 1.8μm) at 40 ℃. A gradient elution of solvent A [water/acetonitrile (6:4, v/v) containing 10mM ammonium formate and 0.1% formic acid] and solvent B [isopropanol/acetonitrile (9:1, v/v) containing 10mM ammonium formate and 0.1% formic acid] was applied with the modified gradient program: 0-1.5 min, 30% B; 4 min, 45% B; 8 min, 58% B; 11 min, 66% B; 14 min, 70% B; 18 min, 75% B; 21-25 min, 97% B; 25.1 - 28 min, 30% B. The flow rate was 0.3 mL/min, and the injection volume was 3 μL. The column and sample chamber temperature were 40 ℃ and 4 ℃ respectively. The H-ESI-MS spectra were acquired in both positive and negative modes. The H-ESI parameters were as follows: spray voltage, 3600V for positive ESI mode and 3000V for negative ESI mode; sheath gas, 35 units; nebulizer auxiliary gas, 20 units. General instrumental parameters were set as follows: ion transfer tube temperature, 350 ℃; vaporizer temperature, 200 ℃. The full scan centroid mode was used for acquiring the ESI-MS. Mass range was set at 100 - 1200 m/z with 120,000 resolutions. Automatic gain control (AGC) target was set as 2

$\times 10^5$ with a maximum injection time of 100 ms. For MS/MS scan, HCD fragmentation method was used for fragmenting the precursor ions selected by Quadrupole isolation.

4.2.8 Data extraction and processing

Chromatographic peak picking and alignment were performed using Progenesis QI 2.3 (Nonlinear Dynamics, Newcastle upon Tyne, United Kingdom). The data type was centroid. Sensitivity and chromatographic peak width were optimized accordingly. Ion abundance of each signal was normalized with internal standards for serum metabolomics. On the other hand, the normalization of ion abundance in the dataset of liver lipidomics was done by using total ion abundance of each individual sample. Quality screening was conducted by filtering out those unstable metabolites and applying a cut-off on coefficient of variation > 30% in all QC samples. A data matrix of normalized ion abundance of all samples was exported to the Extended Statistical tool (EZinfo Version 2.0 software, Umetrics AB) for data pre-processing.

The exported data were scaled to unit variance for PCA, giving an overview of QC samples in the score plot. The study samples excluding QC samples were pareto-scaled before PLS-DA and OPLS-DA. Based on their contribution to the variation and correlation in the data set, potential candidates were selected from the S-plots of OPLS-DA. The markers were further identified with mass fragmentation and matched with the Human Metabolome Databases (www.hmdb.ca), the KEGG (www.kegg.com/), the METLIN (http://metlin.scripps.edu), and/or confirmed by commercially available reference standards based on their mass fragmentation pattern, retention times and mass accuracy.

4.2.9 Measurement of serum and liver biochemical markers

Blood samples were collected and stood at room temperature for 1 h, and then centrifuged at 1500 rpm for 10 minutes to obtain serum samples. Aliquots of the serum samples were stored

at -80°C prior to analysis. Serum TC, TG, LDL-C and HDL-C were measured with the ALYCON® systems using Roche Reagents.

About 2 g liver tissue was homogenized with chloroform-methanol solution (2:1, v/v) to a final dilution of 1:20 (w/v) using Ultra-turrax T-25 homogenizer. After filtration, 10 mL of filtrate was mixed with 2 mL water and centrifuged at 900 g for 20 min. The lower phase of the two layers was collected and dried. The dried extract was weighed as liver lipid content and expressed as the weight of lipid per gram of liver. The content of TG and TC were measured with a colorimetric enzymatic kit (Nanjing Jiancheng Bioengineering Institute, Nanjing, China).

4.2.10 Western blot immune-reactivity assay

Immunoblotting procedures were conducted to quantify the protein expression of cholesterol 7α-hydroxylase (CYP7A1) by my teammate. The methodology was described for reference. Homogenized liver tissue samples were lysed with lysis buffer and cooled in ice for 20 min. The mixture was centrifuged for 15 min at 8265 × *g*. The supernatant was isolated and assayed for protein contents with Bio-Rad protein assay kit (Bio-Rad, Hercules, CA, USA). Protein extracts were applied to 7.5%, 10% or 15% SDS-polyacrylamide gels and transferred to polyvinylidene fluoride membranes. The membranes were incubated with primary antibodies: rabbit anti-CYP7A1 (Santa Cruz Biotechnology) and rabbit anti-β-actin (Cell Signaling Technology). Then signals were obtained by binding a secondary antibody. Protein contents were visualized with a Clarity™ Enhanced Chemiluminescence Western blotting substrate on Azure™ Biosystems C600 (Bio-Rad, Hercules, USA). The band intensities of proteins were quantified with ImageJ software. The expression level of each protein was normalized to that of the *β*-actin protein in each sample.

4.2.11 Statistical analysis

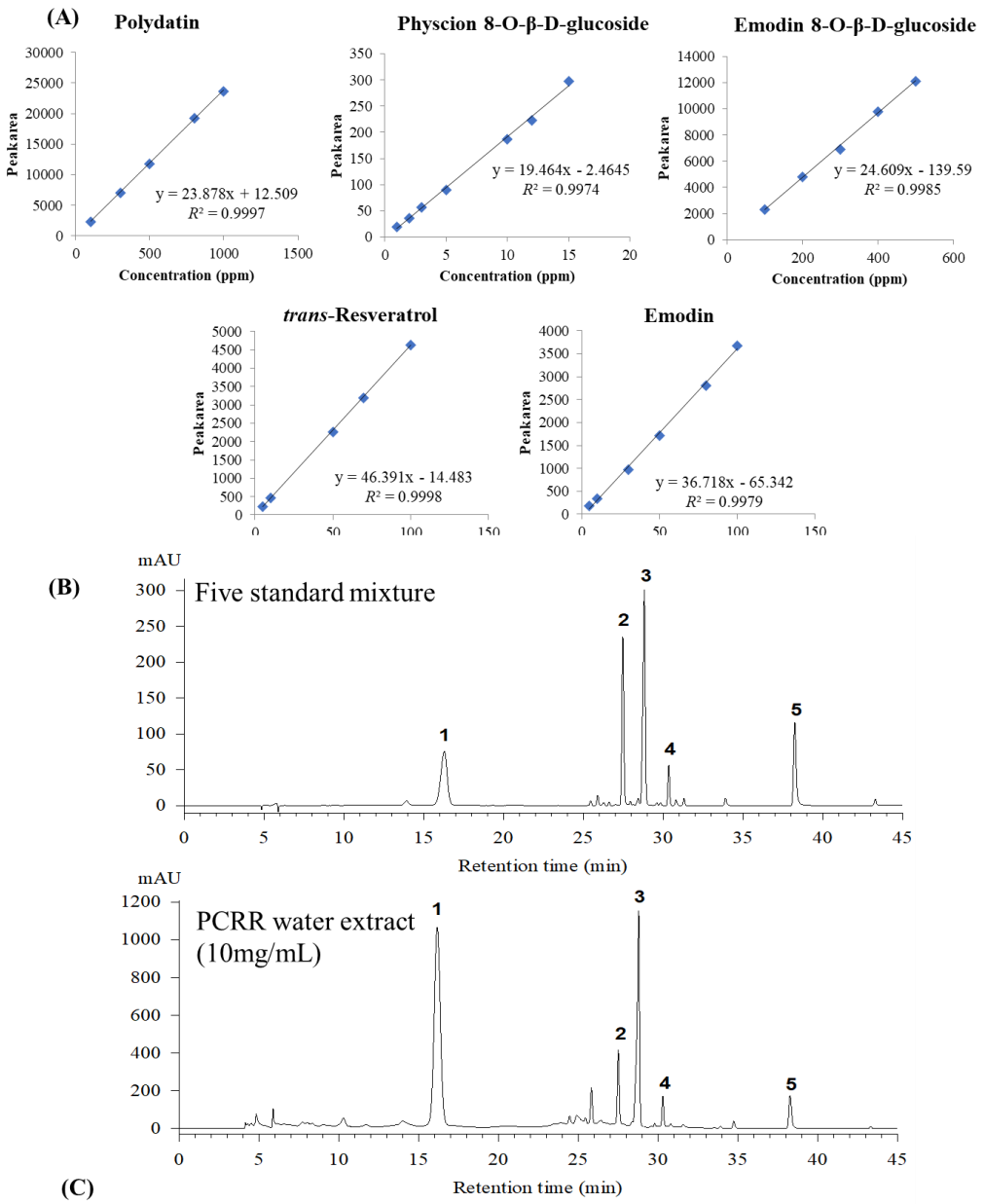
Statistical analyses were performed using IBM SPSS Statistics version 25 (Chicago, IL, USA). After $\log_{10}$ transformation and removal of outliers (1.5 times of the interquartile range), statistical differences were evaluated by one-way ANOVA at a univariate level, followed by Tukey HSD *post-hoc* test (if there were no significant differences, LSD would be replaced) with assumption of equal variances. A p value of 0.05 was considered statistically significant.

4.3 Results and discussion

4.3.1 Chemical analysis of PCRR-water extract

4.3.1.1 Quantitative analysis of major chemicals in PCRR-water extract using HPLC-DAD

The major peaks detected in PCRR water extract were identified with reference standards and quantified with HPLC-DAD (Fig. 4.2). A representative HPLC chromatogram of PCRR water extract with annotation of the five identified peaks is shown in Fig. 4.2B. According to the Pharmacopoeia of the People's Republic of China 2015 (Chinese Pharmacopoeia Commission 2015), the dried PCRR crude drug should contain not less than 0.60 % of emodin (emodin and aglycon isolated after acid hydrolysis of emodin glycosides) and 0.15% of polydatin. As the yield of the dried water extract of PCRR from its raw decoction pieces after freeze-drying was 7.5%, their contents with respect to the dried crude herb weight (%) listed in a descending order were: polydatin (1.058 ± 0.022 %), emodin 8-*O*-*β*-D-glucoside (0.477 ± 0.003 %), resveratrol (0.066 ± 0.001 %), emodin (0.058 ± 0.001 %) and physcion 8-*O*-*β*-D-glucoside (0.053 ± 0.005 %) (Fig. 4.2C). The chemical quality of the PCRR sample used in this experiment matched the requirement and was good for medicinal purpose.



(C)

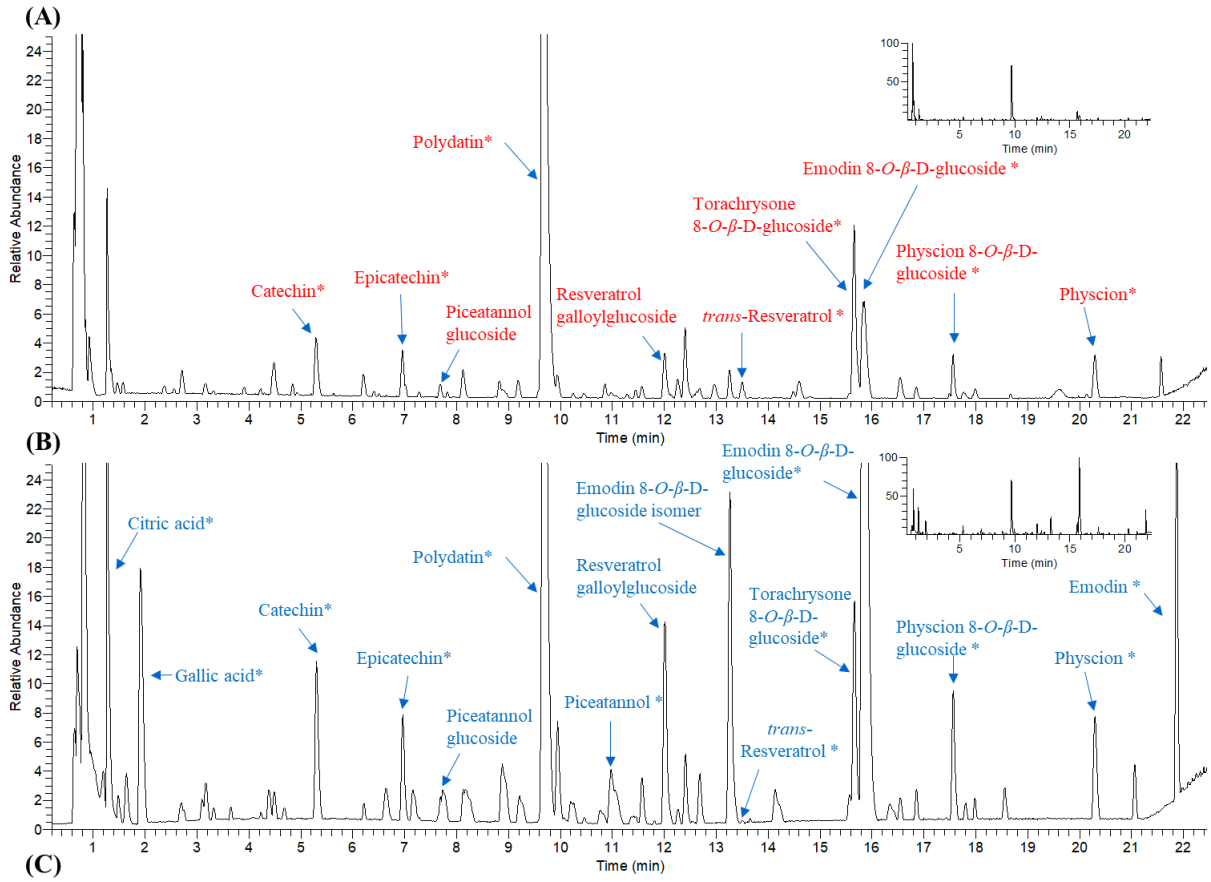
Peak	Retention time (min)	Major chemicals	Average contents (%, w/w)*	
			dried water extract	dried crude herb ^
1	16.31	Polydatin	7.939 ± 0.177	1.058 ± 0.022
2	27.51	*trans*-Resveratrol	0.498 ± 0.006	0.066 ± 0.001
3	28.76	Emodin 8-*O*-*β*-D-glucoside	3.580 ± 0.025	0.477 ± 0.003
4	30.30	Physcion 8-*O*-*β*-D-glucoside	0.398 ± 0.036	0.053 ± 0.005
5	38.20	Emodin	0.432 ± 0.004	0.058 ± 0.001

Fig. 4.2 (A) Calibration curves and linearity of five major chemicals of PCRR water extract; (B) HPLC-DAD chromatogram; (C) their average contents with respect to the dried weight of water extract and crude herb.

*Data were expressed as average ± SD calculated by three replicates. ^ Data were calculated with 7.5 % yield of water extraction from dried crude herb.

4.3.1.2 Qualitative analysis of PCRR-water extract using UPLC-Orbitrap-MS

In order to discover more unknown chemicals, the PCRR water extract was further analysed using a more advanced platform, high-resolution UPLC-Orbitrap-MS. 15 compounds, namely organic acids (citric acid and gallic acid), stilbene and its glycosides (trans-resveratrol, piceatannol, polydatin, piceatannol glucoside and resveratrol galloylglucoside), flavonoids and its glycosides (catechin, epicatechin, emodin, physcion, physcion 8-*O*-*β*-D-glucoside, emodin 8-*O*-*β*-D-glucoside and emodin 8-*O*-*β*-D-glucoside isomer) as well as another phenolic (naphthol) glucoside (torachrysone 8-*O*-*β*-D-glucoside), were identified with reference standards or putatively identified by UPLC-Orbitrap-MS after comparison of the retention time, accurate m/z and mass fragmentation pattern with literatures and online database. The identified peaks shown in the base peak chromatograms of the PCRR water extract and their identification information are depicted in Fig. 4.3. Among them, piceatannol was first reported to be identified in the herb.



(C)

Identity	Molecular formulae	Adducts	Retention time (min)	Theoretical m/z	Detected m/z	Mass error (ppm)
Citric acid*	$C_6H_8O_7$	[M-H]-	1.26	191.0197	191.0200	1.5
Gallic acid*	$C_7H_6O_5$	[M-H]-	1.91	169.0142	169.0144	1.2
Catechin*	$C_{15}H_{14}O_6$	[M+H]+	5.29	291.0863	291.0853	-3.4
		[M-H]-	5.30	289.0718	289.0723	1.9
Epicatechin*	$C_{15}H_{14}O_6$	[M+H]+	6.96	291.0863	291.0854	-3.1
		[M-H]-	6.96	289.0718	289.0723	1.9
Piceatannol glucoside	$C_{20}H_{22}O_9$	[M+H]+	7.68	407.1336	407.1325	-2.7
		[M-H]-	7.69	405.1191	405.1197	1.48
Polydatin*	$C_{20}H_{22}O_8$	[M+H]+	9.67	391.1387	391.1375	-3.2
		[M-H]-	9.69	389.1242	389.1248	1.6
Piceatannol *	$C_{14}H_{12}O_4$	[M-H]-	10.94	243.0663	243.0665	0.9
Resveratrol galloylglucoside	$C_{27}H_{26}O_{12}$	[M+H]+	12.01	543.1497	543.1483	-2.6
		[M-H]-	12.01	541.1351	541.1363	2.1
Emodin 8-O-β-D-glucoside isomer	$C_{21}H_{20}O_{10}$	[M-H]-	13.26	431.0984	431.0992	1.9
***trans*-Resveratrol ***	$C_{14}H_{12}O_3$	[M+H]+	13.51	229.0859	229.0852	-3.1
		[M-H]-	13.51	227.0714	227.0715	0.4
Torachrysone 8-O-β-D-glucoside*	$C_{20}H_{24}O_9$	[M+H]+	15.66	409.1493	409.1480	-3.2
		[M-H]-	15.66	407.1348	407.1355	1.7
Emodin 8-O-β-D-glucoside*	$C_{21}H_{20}O_{10}$	[M+H]+	15.83	433.1129	433.1116	-3.1
		[M-H]-	15.84	431.0984	431.0991	1.7
Physcion 8-O-β-D-glucoside*	$C_{22}H_{22}O_{10}$	[M+Na]+	17.56	469.1105	469.1091	-3.0
		[M-H]-	17.57	445.1140	445.1148	1.8
Physcion *	$C_{16}H_{12}O_5$	[M+H]+	20.30	285.0757	285.0748	-3.3
		[M-H]-	20.29	283.0612	283.0618	2.1
Emodin *	$C_{15}H_{10}O_5$	[M-H]-	21.87	269.0455	269.0461	2.1

Fig. 4.3 Base peak chromatogram of identified peaks from PCRR water extract (10mg/mL in 70% methanol) using UPLC-Orbitrap-MS. (A) ESI$^+$ mode; (B) ESI$^-$ mode.

*Peaks labelled with asterisk were identified with reference standards. Peaks labelled without asterisk were putatively identified with literatures and online database.

4.3.2 Effect on serum and liver lipid profiles

The HFD-induced fatty liver rat model is a good model for the study of NAFLD (Lieber et al. 2004; Lau, Zhang, and Yu 2016). After four-week HFD feeding, measurement of liver and serum lipid levels showed that the weight of liver, the content of TC in both of serum and liver, liver lipid content, serum LDL-C of rats ($p < 0.001$) and liver TG ($p < 0.05$) were markedly increased in the HFD model. Table 4.1 is a summary of the serum and liver lipid levels in all groups. Serum glucose, ALT and AST elevations usually reflect the presence of hepatocellular injury, but this did not happen in HFD model in this study. These observations showed that in this study, oral intake of HFD for 4 weeks just induced simple fatty livers (hepatic steatosis) which represent the early stage of NAFLD, and potential metabolic dysfunction accompanying development of NAFLD.

PCRR treatments in both low and high doses led to significant reduction of liver TC ($p < 0.05$), TG ($p < 0.05$), and lipid content in liver ($p < 0.001$), similar to that of the positive control group with simvastatin at high dose ($p < 0.05$). Only high-dose PCRR treatment reduced the increased ratio of liver weight to body weight by HFD ($p < 0.01$). Increased serum TC, and serum LDL-C levels by HFD were reduced by treatment of PCRR in both low and high doses. The levels of serum glucose, alanine aminotransferase, aspartate aminotransferase and HDL-C in the HFD model, PCRR treatment group and positive control group were not changed. These changes indicated that PCRR supplementation ameliorated the enlargement of liver and fatty liver, upregulation of cholesterol contents in the blood induced by HFD in this study.

Table 4.1 Serum and liver lipid levels of rats from the normal control, HFD model, simvastatin-treated and PCRR-treated HFD-fed groups.

	Normal control	HFD model	HFD+Low dose of simvastatin	HFD+High dose of simvastatin	HFD+Low dose of PCRR water extract	HFD+High dose of PCRR water extract
Body weight gain (g)	187.8 ± 9.12	207.1 ± 13.65	179.6 ± 6.94	181.5 ± 5.58	199.6 ± 14.99	202.6 ± 10.13
			Serum			
TC (mmol/L)	1.36 ± 0.09***	2.93 ± 0.29	2.11 ± 0.27	1.76 ± 0.15**	1.98 ± 0.10*	1.93 ± 0.19*
TG (mmol/L)	0.57 ± 0.03	0.54 ± 0.11	0.36 ± 0.02	0.48 ± 0.05	0.47 ± 0.04	0.51 ± 0.06
LDL-C (mmol/L)	0.21 ± 0.02***	0.56 ± 0.07	0.35 ± 0.08	0.22 ± 0.02**	0.28 ± 0.03**	0.30 ± 0.04*
HDL-C (mmol/L)	1.10 ± 0.04	1.42 ± 0.05	1.24 ± 0.08	1.45 ± 0.06	1.40 ± 0.09	1.35 ± 0.13
ALT (U/L)	42.1 ± 2.78	42.4 ± 1.28	44.1 ± 1.51	42.6 ± 2.81	39.4 ± 2.08	40.3 ± 2.58
AST (U/L)	148 ± 6.02	156 ± 3.36	148 ± 8.49	150 ± 7.63	124 ±4.21	126 ± 10.92
Glucose (mmol/L)	5.96 ± 0.33	6.42 ± 0.30	5.91 ± 0.40	6.05 ± 0.35	6.07 ± 0.35	5.85 ± 0.27
			Liver			
Ratio of liver to body weight (g/kg)	26.6 ± 0.45***	39.4 ± 1.83	35.4 ± 0.64	32.8 ± 1.97**	35.1 ± 0.63	33.4 ± 0.99**
Lipid content (g/g of liver)	0.14 ± 0.01***	0.34 ± 0.03	0.24 ± 0.01*	0.14 ± 0.04***	0.18 ± 0.02***	0.21 ± 0.02***
TC (% of control)	100 ± 5.78***	321 ± 24.7	258 ± 10.1	243 ± 10.7*	238 ± 15.9*	232 ± 17.4**
TG (% of control)	100 ± 7.77*	154 ± 16.0	115 ± 11.1	88.3 ± 8.91**	110 ± 9.62*	106 ± 4.6*

Data are expressed as means ± SEM. n = 6-9. Bold numbers represent the significant difference. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$ compared with the HFD model. ALT, alanine aminotransferase; AST, aspartate aminotransferase; HDL-C, high-density lipoprotein cholesterols; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterols; TG, triaglycerides.

4.3.3 Stability of the serum metabolomics and liver lipidomics measurements

The stability of serum metabolomics and liver lipidomics of UPLC-MS detections was assessed by the internal standards and pooled QC sample injections. The CV% of internal standards was < 20%. The bile acids listed in Table 4.2 had < 5% CV, expect glycocholate that had < 30% CV. Lipids that are listed in Table 4.3 had < 20%, except phosphatidylglycerols that had < 30% CV. The score plot of PCA was used to evaluate the stability of the overall signals. Fig. 4.6A showed the QC samples were clustered together in the PCA score plots, indicating high stability throughout the run. The UPLC-MS base peak chromatograms of serum and liver in normal control, HFD model and PCRR-treated group by serum metabolomics analysis and liver lipidomics are shown in Fig. 4.4 and Fig. 4.5, respectively.

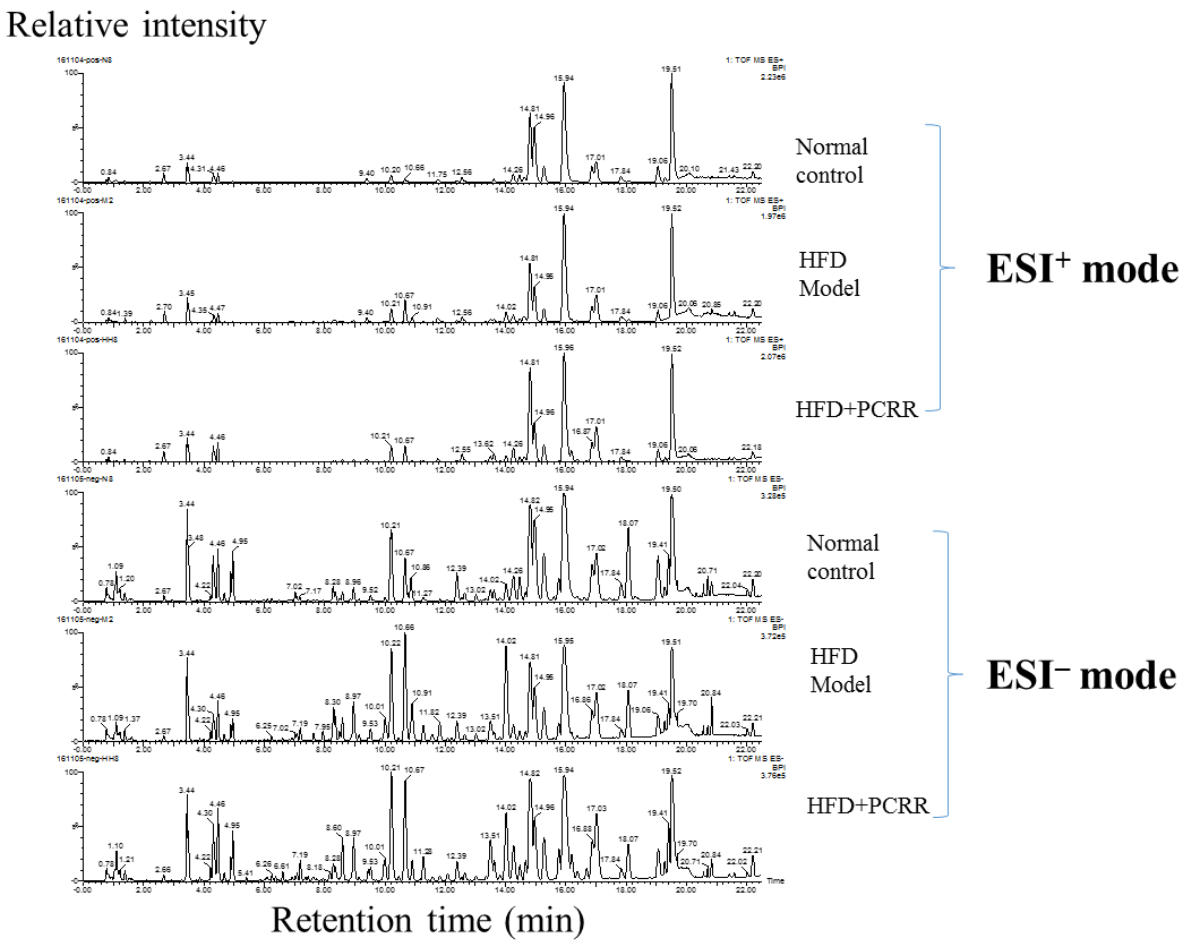


Fig. 4.4 Representative base peak chromatograms of UPLC-QTOF-MS acquired in two ESI modes of rat serum metabolomics from different groups.

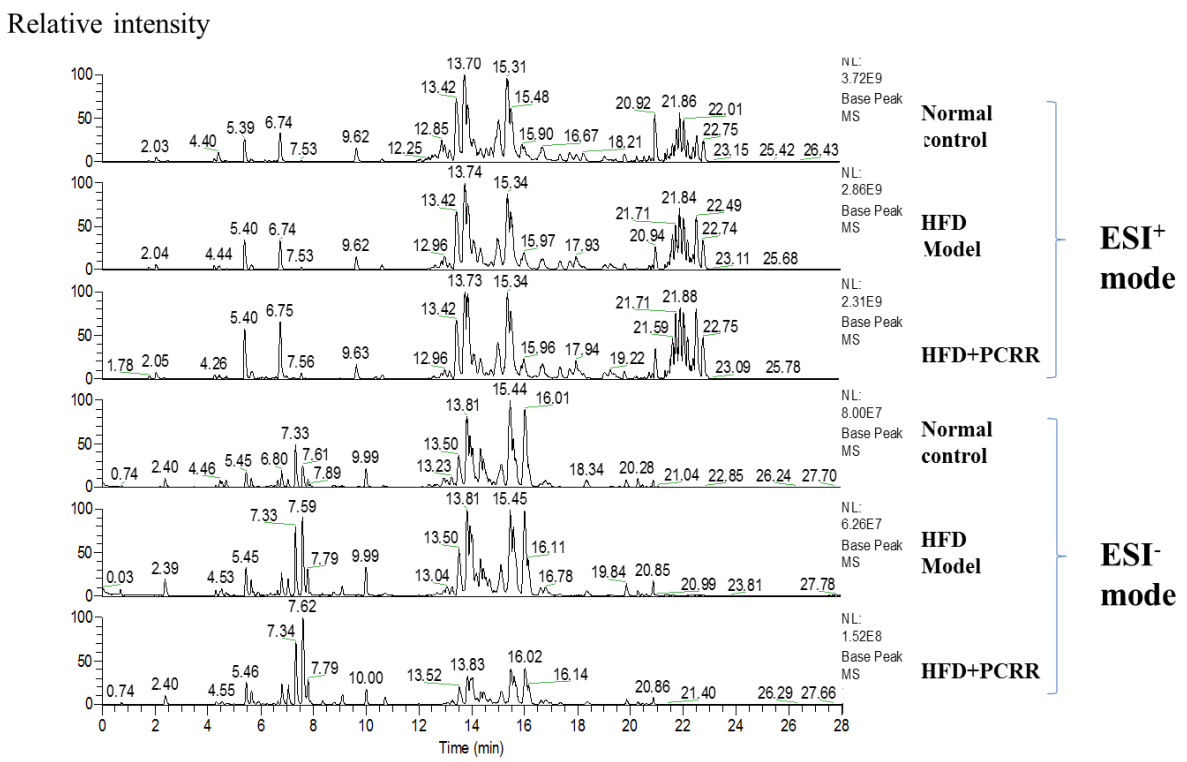


Fig. 4.5 Representative base peak chromatograms of UPLC-Orbitrap-MS acquired in two ESI modes of rat liver lipidomics from different groups.

4.3.4 Multivariate statistics of serum metabolomics and liver lipidomics

The multivariate statistical analysis of serum metabolomics data showed that the goodness of fit for PLS-DA and OPLS-DA were acceptable in the view of the internal cross validation results (PLS-DA: $R^2X_{cum} = 0.366$, $R^2Y_{cum} = 0.789$, $Q^2_{cum} = 0.531$ in positive ESI mode; $R^2X_{cum} = 0.468$, $R^2Y_{cum} = 0.752$, $Q^2_{cum} = 0.536$ in negative ESI mode. OPLS-DA: $R^2X_{cum} = 0.451$, $R^2Y_{cum} = 0.987$, $Q^2_{cum} = 0.788$ in positive ESI mode; $R^2X_{cum} = 0.498$, $R^2Y_{cum} = 0.940$, $Q^2_{cum} = 0.641$ in negative ESI mode) (Fig. 4.6a). The result of OPLS-DA showed that there were 118 peaks with VIP >1 in negative ESI mode and 295 peaks with VIP > 1 in positive ESI mode (Fig. 4.6Da).

The separation of the liver lipidomics data in PLS-DA and OPLS-DA was better than that of serum metabolomics (PLS-DA: $R^2X_{cum} = 0.499$, $R^2Y_{cum} = 0.801$, $Q^2_{cum} = 0.611$ in positive ESI mode; $R^2X_{cum} = 0.502$ $R^2Y_{cum} = 0.746$, $Q^2_{cum} = 0.660$ in negative ESI mode. OPLS-DA: $R^2X_{cum} = 0.659$, $R^2Y_{cum} = 0.974$, $Q^2_{cum} = 0.934$ in positive ESI mode; $R^2X_{cum} = 0.750$, $R^2Y_{cum} = 0.900$, $Q^2_{cum} = 0.840$ in negative ESI mode) (Fig. 4.6b). The result of OPLS-DA showed that there were 121 peaks with VIP >1 in negative ESI mode and 463 peaks with VIP >1 in positive ESI mode (Fig. 4.6Db).

The cluster of samples of the PCRR-treated group was well separated from those of the normal control group and HFD-fed group in the score plot of PLS-DA (Fig. 4.6B). This indicated that the serum metabolite and liver lipid profiles of the PCRR-treated group were very different from the other two groups

After statistical analysis and identification of peaks listed in OPLS-DA, amino acids, bile acids and their conjugated products as well as lipids were highlighted. The fold changes and their identities of bile acids, their conjugated products and lipids compared with the HFD-fed model were listed in Table 4.2 and Table 4.3, respectively.

The above OPLS-DA model offered the foundation for studying the role of PCRR water extract in reducing blood cholesterol level and TG in the livers of HFD-fed rats, which would be further discussed in the following sections. Fig. 4.7 shows a heatmap of bile acids and liver lipids that were significantly changed in PCRR water extract treated group compared with the HFD model from the results of serum metabolomics and liver lipidomics.

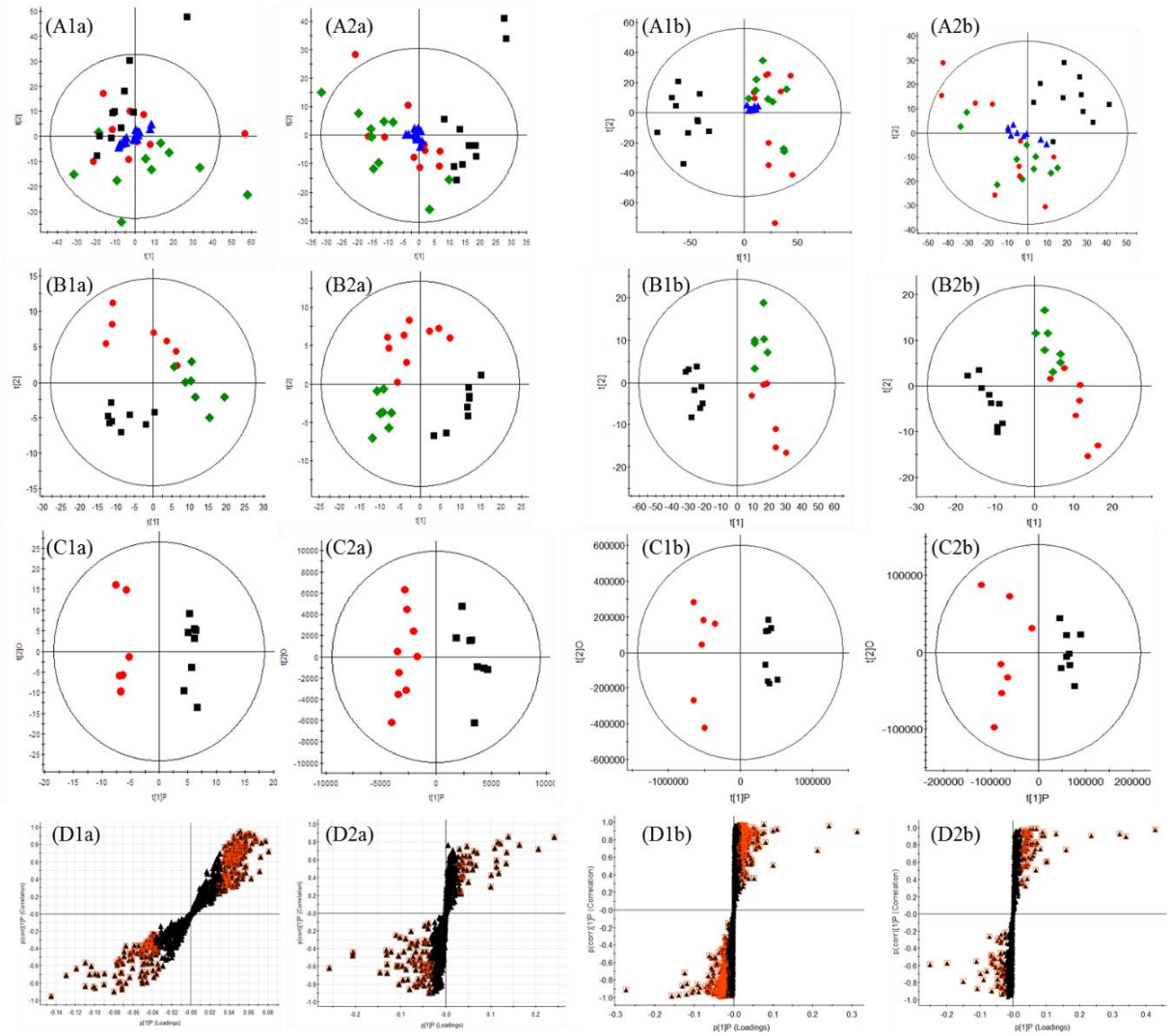


Fig. 4.6 Multivariate analyses of serum metabolomics and liver lipidomics samples acquired using UPLC- MS.

(A) Score plots of PCA; (B) Score plots of PLS-DA; (C) Score plots of OPLS-DA; (D) Loading plots of OPLS-DA: red squares highlighted are variables with VIP >1; serum ESI-, 118 peaks; serum ESI+, 295 peaks; liver ESI-, 121 peaks; liver ESI+, 463 peaks.

(1) ESI+ mode, (2) ESI- mode.

(a) serum metabolomics data; (b) liver lipidomics data.

Normal control (■); HFD-fed model (●); high-dosed PCRR water extract-treated HFD-fed model (◆) and pooled quality control group (▲).

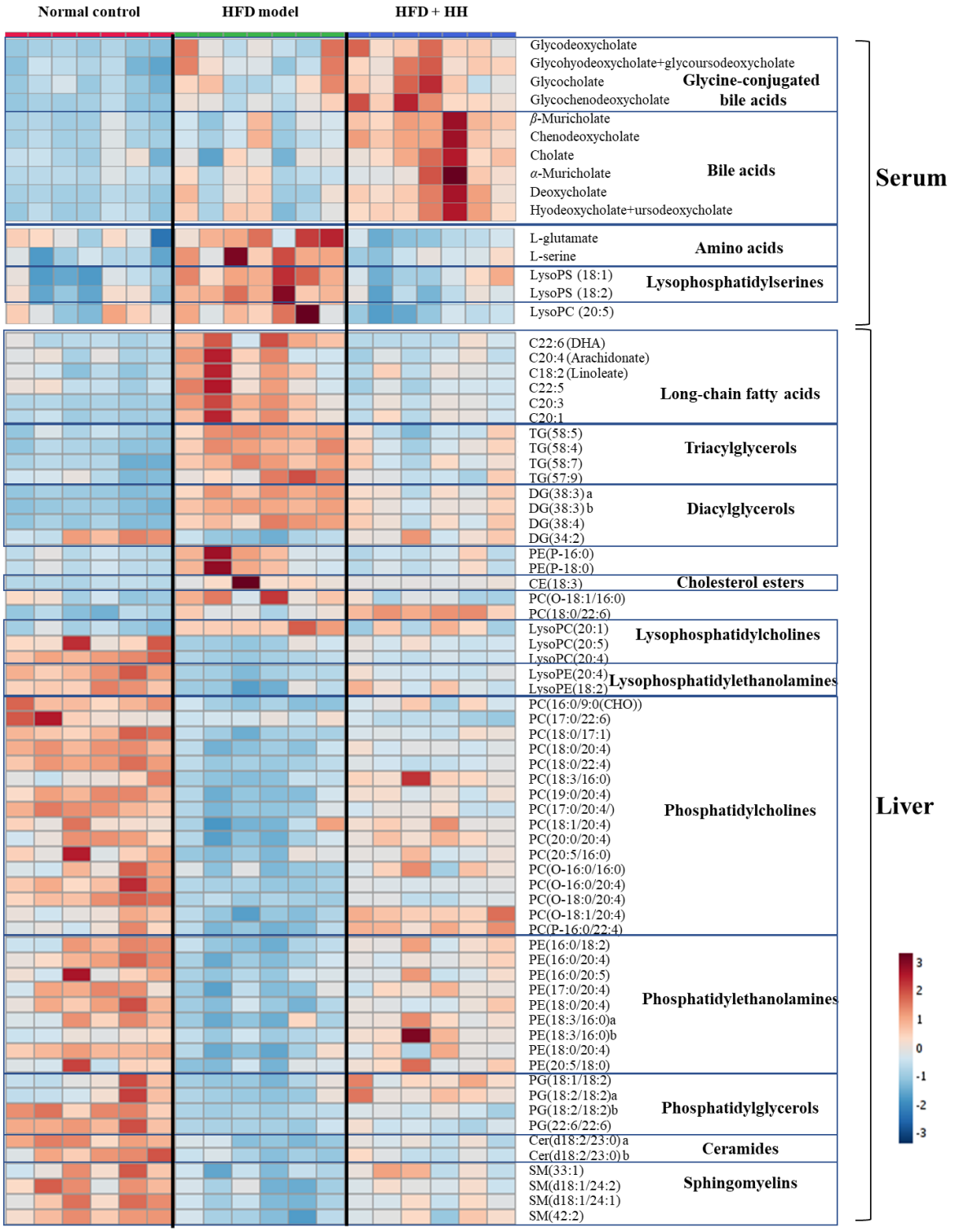


Fig. 4.7 Heatmap of bile acids and liver lipids that were significantly changed in PCRR water extract treated group compared with the HFD model from the results of serum metabolomics and liver lipidomics.

4.3.5 Putative identification of lipids using mass fragmentations

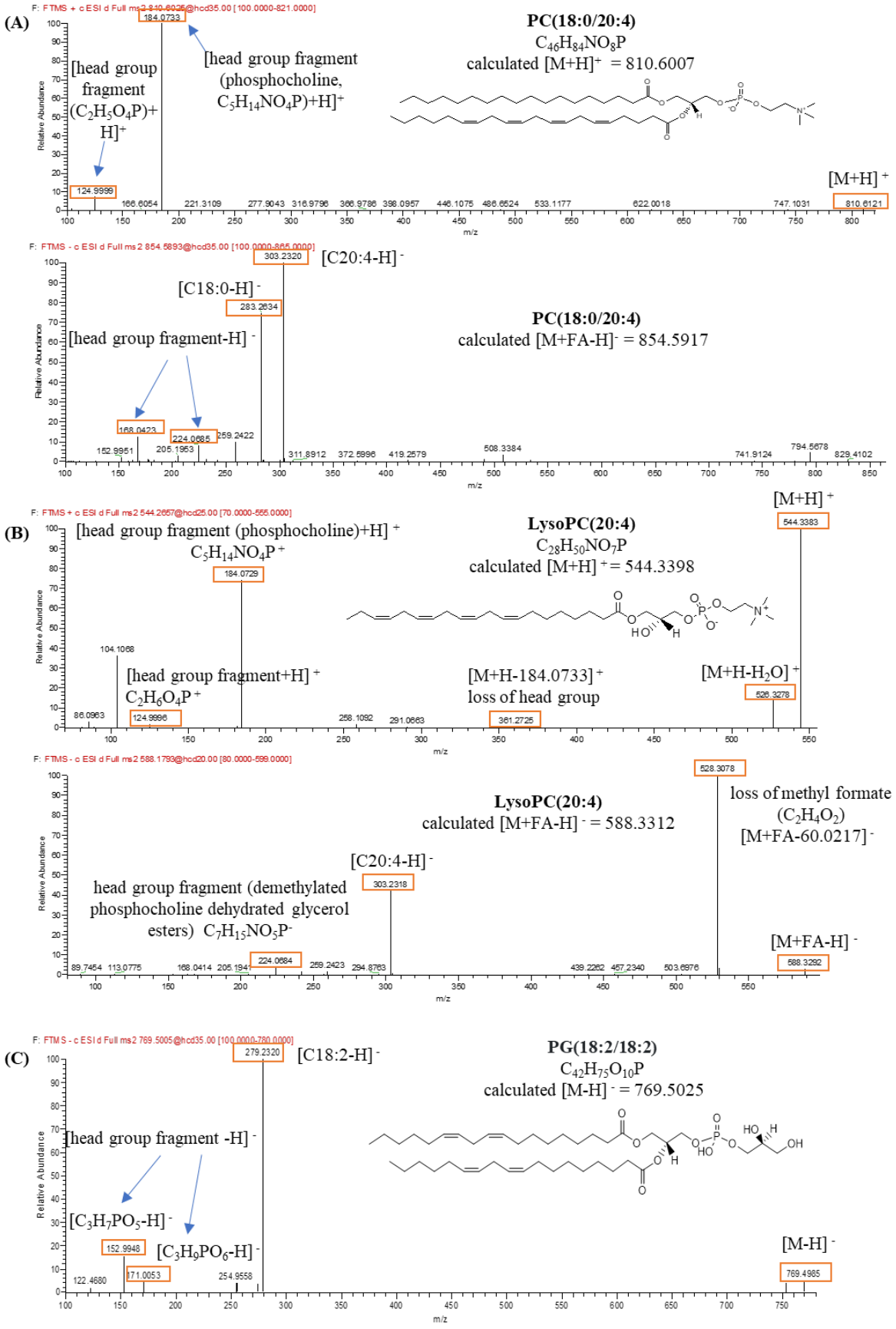
Various lipids derived from lipidomics and serum metabolomics were putatively identified according to their classes' characteristic mass fragments and matching with databases, such as lipidmaps and HMDB, and literatures, except acylcarnitine, fatty acids and amino acids which were further confirmed with reference standards when applicable. Fig. 4.8 demonstrates the mass fragmentation pattern of parent ions of representatives from different classes, including PC(18:0/20:4), lysoPC(20:4), PG(18:2/18:2), PE(18:0/20:4), lysoPE (20:4), lysoPS (18:1), TG(58:4), DG (34:2), CE (22:6), Cer (d18:2/23:0), SM (42:3), arachidonic acid, L-serine and cholic acid.

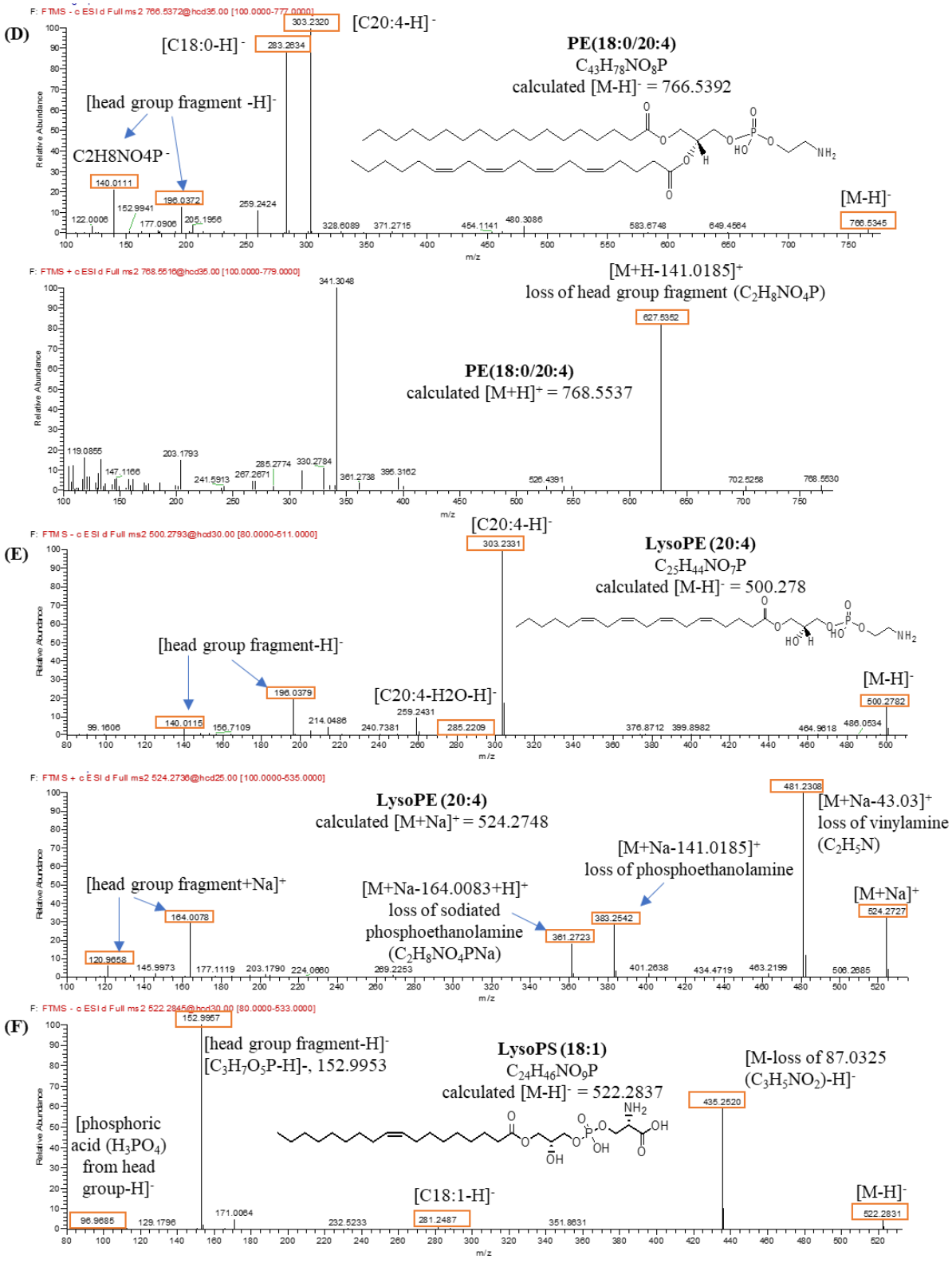
For instance, MS/MS of parent ion, $[M+H]^+$ (*m/z* 810.6000) at 15.47 min showed a strong characteristic head group fragment ion, a protonated phosphocholine $[C_5H_{14}NO_4P+H]^+$ with *m/z* 184. Under negative ESI mode, it was associated with a strong negative adduct ion, formate, denoted as $[M+FA-H]^-$ (*m/z* 854.5893) and yielded two prominent fatty acid carboxylate anions, $[C18:0-H]^-$ (*m/z* 208) and $[C20:4-H]^-$ (*m/z* 303) derived from its *sn*-1 and *sn*-2 acyl substituents, respectively. Combination of the detection of phosphocholine head group, two fatty acid fragments and the mass error of the two parent ions within 5 ppm gave the putative identity of the peak at 15.47 min as PC(18:0/20:4) (Frega, Pacetti, and Boselli 2012). Another peak at 4.41 min that also yielded the same head group fragment ion *m/z* 184 in $[M+H]^+$ (*m/z* 544.3396) and fatty acid carboxylate ion $[C20:4-H]^-$ (*m/z* 303) from $[M+FA-H]^-$ = *m/z* 588.3295 was proposed as lysoPC (20:4) because lysophosphatidylcholines were eluted faster than phosphatidylcholines

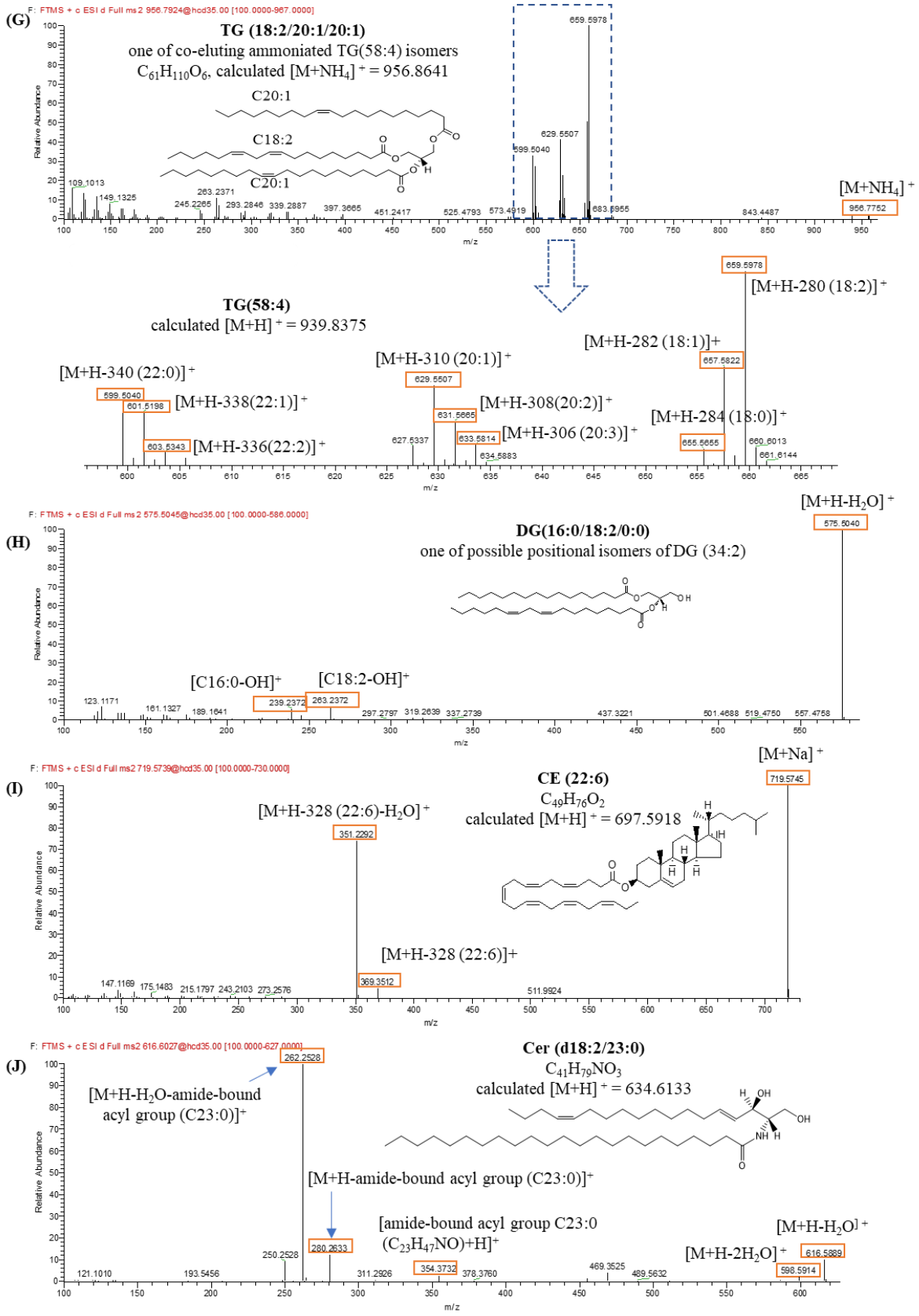
SM (d42:3) was proposed from its head group fragment (demethylated phosphocholine), m/z 168 from [M+FA-H]- (m/z 855.6575) at 16.98 min and did not yield any strong fatty acid ions so its substitutes were not known. However, its $[M+H]^+$ (*m/z* 811.6683) eluted at the similar time gave a protonated phosphocholine, and more importantly, the diagnostic fragment ion

$[C_{18}H_{33}N+H]^+$ which was a sphingosine backbone ($C_{18}H_{37}NO_2$) with loss of two H_2O, resulting from the bond cleavage at the branching point of sphingomyelin backbone (Baba et al. 2016). Thus, the putative identity of SM (d42:3) at 16.98 min was SM(d18:1/24:2).

TG structural isomers, that shared the same *m/z* and had strong ammoniated adduct ions $[M+NH_4]^+$, had the tendency to be co-eluted in the single chromatographic peaks (Jiao et al. 2015). This resulted in the presence of many DG fragment ions that showed the loss of a fatty acid substitute in the same MS/MS spectrum and was hard to figure out its substitutes. For instance, TG (58:4) at 22.34 min had the combination of ammoniated fragment ions losing C18:0, C18:1, C18:2, C20:1, C20:2, C20:3, C22:0, C22:1 and C22:2, which might be come from TG (18:2/20:1/20:1), TG (18:1/20:2/20:1), TG (18:0/20:2/20:2) or other TG species. Thus, they were presented with the identities as TG (58:4).


(A)
PC(18:0/20:4)
C46H84NO8P
calculated [M+H]+ = 810.6007
[head group fragment (C2H5O4P)+H]+
[head group fragment (phosphocholine, C5H14NO4P)+H]+
[M+H]+
[C20:4-H]-
[C18:0-H]-
[head group fragment-H]-
PC(18:0/20:4)
calculated [M+FA-H]- = 854.5917
(B)
LysoPC(20:4)
C28H50NO7P
calculated [M+H]+ = 544.3398
[head group fragment (phosphocholine)+H]+
C5H14NO4P+
[head group fragment+H]+
C2H6O4P+
[M+H-184.0733]+
loss of head group
[M+H-H2O]+
LysoPC(20:4)
calculated [M+FA-H]- = 588.3312
loss of methyl formate (C2H4O2)
[M+FA-60.0217]-
[C20:4-H]-
head group fragment (demethylated phosphocholine dehydrated glycerol esters) C7H15NO5P-
[M+FA-H]-
(C)
PG(18:2/18:2)
C42H75O10P
calculated [M-H]- = 769.5025
[C18:2-H]-
[head group fragment -H]-
[C3H7PO5-H]-
[C3H9PO6-H]-
[M-H]-


(D)
[C20:4-H]-
[C18:0-H]-
PE(18:0/20:4)
C43H78NO8P
calculated [M-H]- = 766.5392
[head group fragment -H]-
C2H8NO4P-
[M-H]-
[M+H-141.0185]+
loss of head group fragment (C2H8NO4P)
PE(18:0/20:4)
calculated [M+H]+ = 768.5537
(E)
[C20:4-H]-
LysoPE (20:4)
C25H44NO7P
calculated [M-H]- = 500.278
[head group fragment-H]-
[C20:4-H2O-H]-
[M-H]-
LysoPE (20:4)
calculated [M+Na]+ = 524.2748
[M+Na-43.03]+
loss of vinylamine
(C2H5N)
[M+Na-141.0185]+
loss of phosphoethanolamine
[head group fragment+Na]+
[M+Na-164.0083+H]+
loss of sodiated
phosphoethanolamine
(C2H8NO4PNa)
[M+Na]+
(F)
[head group fragment-H]-
[C3H7O5P-H]-, 152.9953
LysoPS (18:1)
C24H46NO9P
calculated [M-H]- = 522.2837
[M-loss of 87.0325
(C3H5NO2)-H]-
[phosphoric
acid (H3PO4)
from head
group-H]-
[C18:1-H]-
[M-H]-


(G)
TG (18:2/20:1/20:1)
one of co-eluting ammoniated TG(58:4) isomers
$C_{61}H_{110}O_6$, calculated $[M+NH_4]^+$ = 956.8641
C20:1
C18:2
C20:1
$[M+NH_4]^+$
TG(58:4)
calculated $[M+H]^+$ = 939.8375
$[M+H-280\ (18:2)]^+$
$[M+H-282\ (18:1)]+$
$[M+H-310\ (20:1)]^+$
$[M+H-340\ (22:0)]^+$
$[M+H-338(22:1)]^+$
$[M+H-308(20:2)]^+$
$[M+H-284\ (18:0)]^+$
$[M+H-306\ (20:3)]^+$
$[M+H-336(22:2)]^+$
(H)
DG(16:0/18:2/0:0)
one of possible positional isomers of DG (34:2)
$[M+H-H_2O]^+$
$[C16:0\text{-}OH]^+$
$[C18:2\text{-}OH]^+$
(I)
CE (22:6)
$C_{49}H_{76}O_2$
calculated $[M+H]^+$ = 697.5918
$[M+Na]^+$
$[M+H-328\ (22:6)\text{-}H_2O]^+$
$[M+H-328\ (22:6)]+$
(J)
Cer (d18:2/23:0)
$C_{41}H_{79}NO_3$
calculated $[M+H]^+$ = 634.6133
$[M+H\text{-}H_2O\text{-amide-bound acyl group (C23:0)}]^+$
$[M+H\text{-amide-bound acyl group (C23:0)}]^+$
[amide-bound acyl group C23:0 $(C_{23}H_{47}NO)+H]^+$
$[M+H-H_2O]^+$
$[M+H-2H_2O]^+$
Relative Abundance
m/z

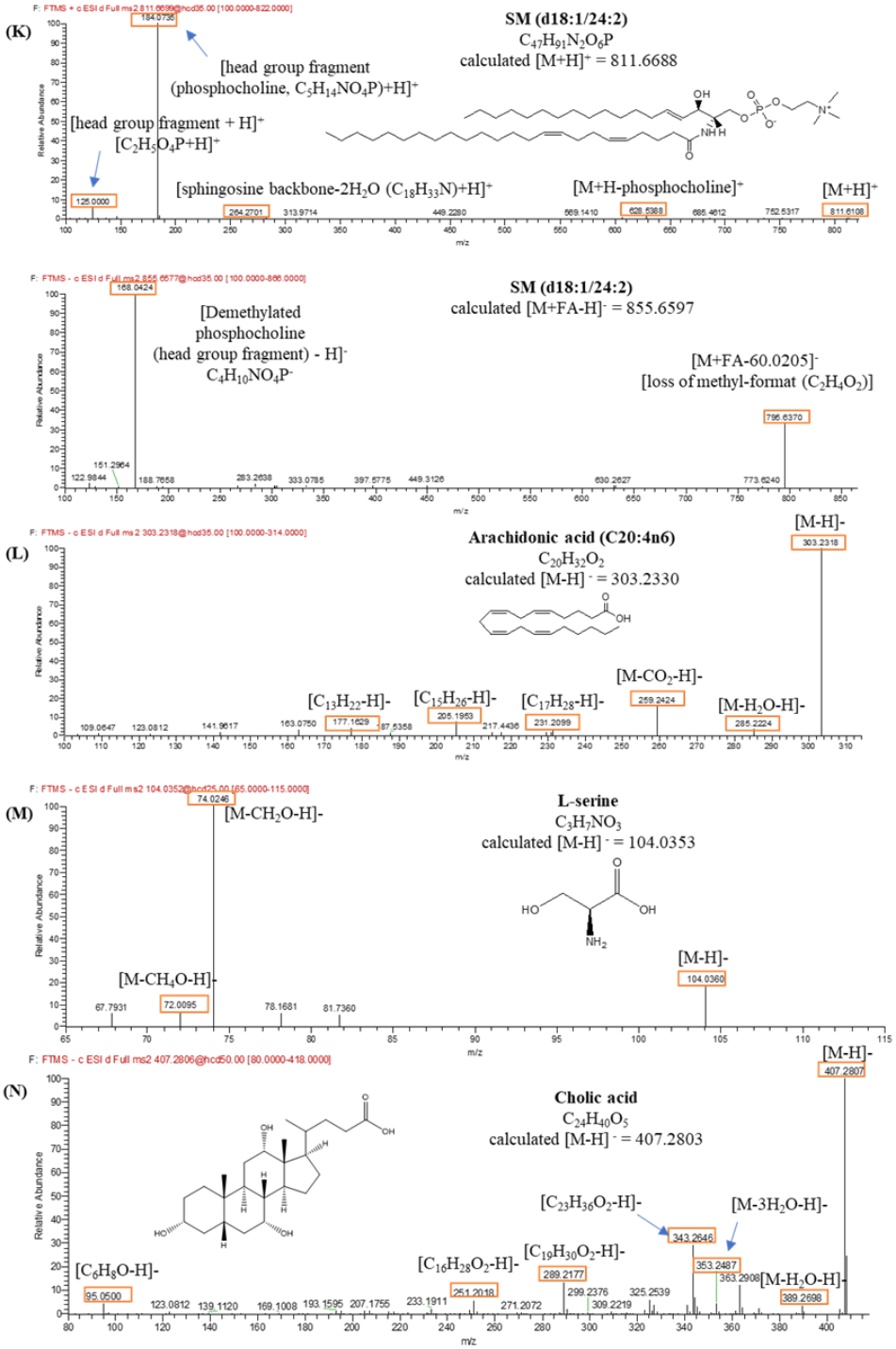


Fig. 4.8 Representative mass spectra of putatively identified altered lipids and amino acids by UPLC-Orbitrap-MS. (A) PC(18:0/20:4); (B) lysoPC(20:4); (C) PG(18:2/18:2); (D) PE(18:0/20:4); (E) lysoPE (20:4); (F) lysoPS (18:1); (G) TG(58:4); (H) DG (34:2); (I) CE (22:6); (J) Cer (d18:2/23:0); (K) SM (d18:1/24:2); (L) arachidonic acid; (M) L-serine; (N) Cholic acid.

4.3.6 Effect of PCRR water extract on NAFLD as shown in the results of serum metabolomics

Two amino acids (L-serine and L-glutamic acid), one lysophosphatidylcholine [LysoPC (20:5)], two lysophosphatidylserines [LysoPS (18:2) and LysoPS (18:1)], as well as a large increase in the four serum primary bile acids (cholate and chenodeoxycholate, α-muricholate and β-muricholate), three secondary bile acids (deoxycholate, sum of hyodeoxycholate and ursodeoxycholate), five glycine-conjugated bile acids (glycocholate, glycochenodeoxycholate and glycodeoxycholate, sum of glycohyodeoxycholate and glycoursodeoxycholate) and two taurine-conjugated bile acids (sum of taurochenodeoxycholate and taurodeoxycholate) were elevated in the HFD group.

PCRR treatment reduced the levels of the amino acids, the lysophosphatidylcholine and the lysophosphatidylserines, close to the normal levels. Furthermore, the treatment caused significant rise in the serum levels of most of them: α-muricholate, β-muricholate, cholate, chenodeoxycholate, deoxycholate, sum of hyodeoxycholate and ursodeoxycholate, glycocholate, glycochenodeoxycholate, glycodeoxycholate, glycohyodeoxycholate sum of glycohyodeoxycholate and glycoursodeoxycholate, except taurochenodeoxycholate and taurodeoxycholate. Among them, fold changes of α-muricholate, chenodeoxycholate, deoxycholate, glycochenodeoxycholate and glycodeoxycholate in PCRR treatment group with respect to HFD model were about double. The proposed pathway of bile acid biosynthesis and level changes of the bile acids and conjugated bile acids by HFD and PCRR treatment were shown in Fig. 4.10.

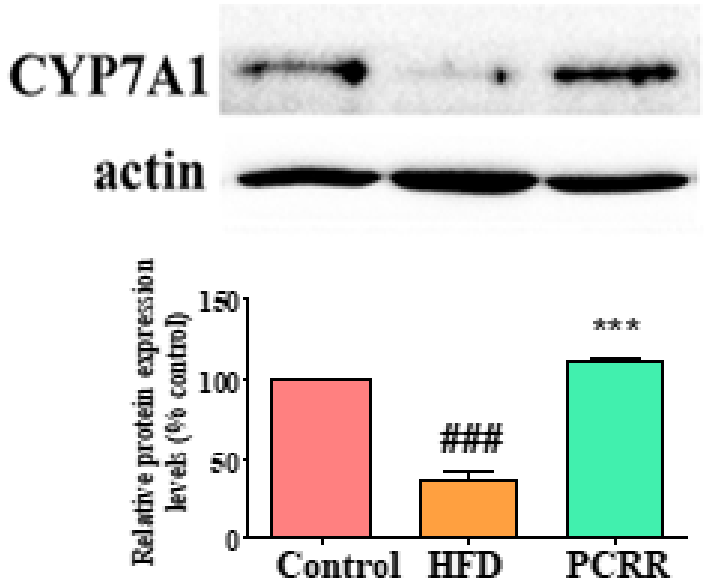
Since a large amount of bile acids were altered, the enzymes participating in the bile acid biosynthesis were also investigated. CYP7A1 is the main rate-limiting enzyme in the classical pathway of bile acid biosynthesis to remove cholesterols (Chiang 2009; Dawson 2016). In this study, the protein expression levels of CYP7A1 was determined in the mechanism of bile acid

disturbance. The liver of the HFD model with NAFLD showed lower CYP7A1 protein expression than the normal control ($p < 0.001$) while PCRR supplement had higher CYP7A1 protein expression than that of the HFD model ($p < 0.001$) (Fig. 4.9). This indicated that HFD suppressed the expression of CYP7A1 protein but PCRR restored its expression back to a normal level. This further explained that PCRR might facilitate the removal of cholesterols via the activation of CYP7A1 expression. There were many reports related to the elevation of this enzyme expression by resveratrol which affects the bile acid biosynthesis (Miura, Miura, and Yagasaki 2003; Zhu, Luo, and Jin 2008; Chen et al. 2012) and increases the excretion of fecal bile acids (Zhu, Luo, and Jin 2008).

Table 4.2 Fold change of serum metabolites in normal controls and PCRR-treated groups compared with the HFD model from the results of serum metabolomics.

Metabolites	Adducts	Molecular formulae	Retention time (min)	Theoretical m/z	Detected m/z	Mass error (ppm)	Fold change with respect to HFD model (mean)		CV (%) in QC samples
							Normal control	HFD+ PCRR	
Bile acids									
α-Muricholate	$[M-H]^-$	C24H40O5	8.31	407.2803	407.2796	-1.72	0.24**	2.49*	4.06
β-Muricholate	$[M-H]^-$	C24H40O5	8.59	407.2803	407.2796	-1.72	0.24**	1.85*	3.15
Cholate	$[M-H]^-$	C24H40O5	10.21	407.2803	407.2796	-1.72	0.63*	1.53*	3.41
Sum of hyodeoxycholate and ursodeoxycholate	$[M-H]^-$	C24H40O4	10.67	391.2854	391.2848	-1.53	0.49*	1.48*	3.41
Chenodeoxycholate	$[M-H]^-$	C24H40O4	13.51	391.2854	391.2848	-1.53	0.51**	2.65***	3.97
Deoxycholate	$[M-H]^-$	C24H40O4	14.02	391.2854	391.2848	-1.53	0.36*	2.28**	3.73
Conjugated bile acids									
Glycocholate	$[M-H]^-$	C26H43NO6	8.28	464.3018	464.3013	-1.08	0.63*	1.92**	29.3
Sum of glycohyodeoxycholate and glycoursodeoxycholate	$[M-H]^-$	C26H43NO5	8.36	448.3068	448.3062	-1.34	0.63*	1.68*	3.82
Glycochenodeoxycholate	$[M-H]^-$	C26H43NO5	10.46	448.3068	448.3063	-1.34	0.22***	2.33*	4.48
Glycodeoxycholate	$[M-H]^-$	C26H43NO5	10.91	448.3068	448.3063	-1.12	0.27***	2.08***	4.34
Taurohyodeoxycholate	$[M-H]^-$	C26H45NO6S	7.02	498.2895	498.2880	-3.01	1.14	0.93	3.49
Taurocholate	$[M-H]^-$	C26H45NO7S	7.18	514.2844	514.2839	-0.97	0.85	0.93	4.25
Sum of taurochenodeoxycholate and taurodeoxycholate	$[M-H]^-$	C26H45NO6S	9.14	498.2895	498.2891	-0.80	0.40**	1.19	4.48
Amino acids									
L-serine	$[M-H]^-$	C3H7NO3	0.81	104.0353	104.0350	-2.88	0.78**	0.71***	10.87
L-glutamate	$[M-H]^-$	C5H9NO4	0.85	146.0459	146.0458	-0.68	0.60***	0.69***	17.84
Lysophosphatidylcholines and Lysophosphatidylserines									
LysoPC(20:5)	$[M+H]+$	C28H48NO7P	10.74	542.3241	542.3243	0.37	0.66*	0.38***	13.12
LysoPS(18:2)	$[M-H]^-$	C24H44NO9P	14.77	520.2681	520.2676	-0.96	0.50***	0.52***	6.79
LysoPS(18:1)	$[M-H]^-$	C24H46NO9P	17.13	522.2837	522.2831	-1.15	0.40***	0.57***	7.45

*One-way ANOVA, followed by LSD *post-hoc* test compared with HFD-fed model. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. n = 6-7. CE, cholesterol esters; LysoPC, lysophosphatidylcholine; LysoPE, lysophosphatidylethanolamine; LysoPS, lysophosphatidylserine.



Protein	Relative protein expression levels to normal control		Trends of protein expression		Related pathway
	HFD	HFD+PCRR	HFD model vs Control	HFD+PCRR vs HFD model	
CYP7A1	0.364 ± 0.054	1.107 ± 0.030	↓###	↑***	Bile acid synthesis (conversion of cholesterols to bile acids)

Fig. 4.9 Quantitative comparison of protein expression in the hepatocytes of the normal control, HFD model and PCRR -treated HFD-fed group using Western blots

The expression level of each protein was normalized to that of the *β*-actin protein in each sample. Data are expressed as means ± SEM from three independent experiments. One way-ANOVA followed by LSD *post-hoc* test: ### $p < 0.001$ compared with normal control group and ***$p < 0.001$ compared with the HFD model. CYP7A1, cholesterol 7α-hydroxylase.

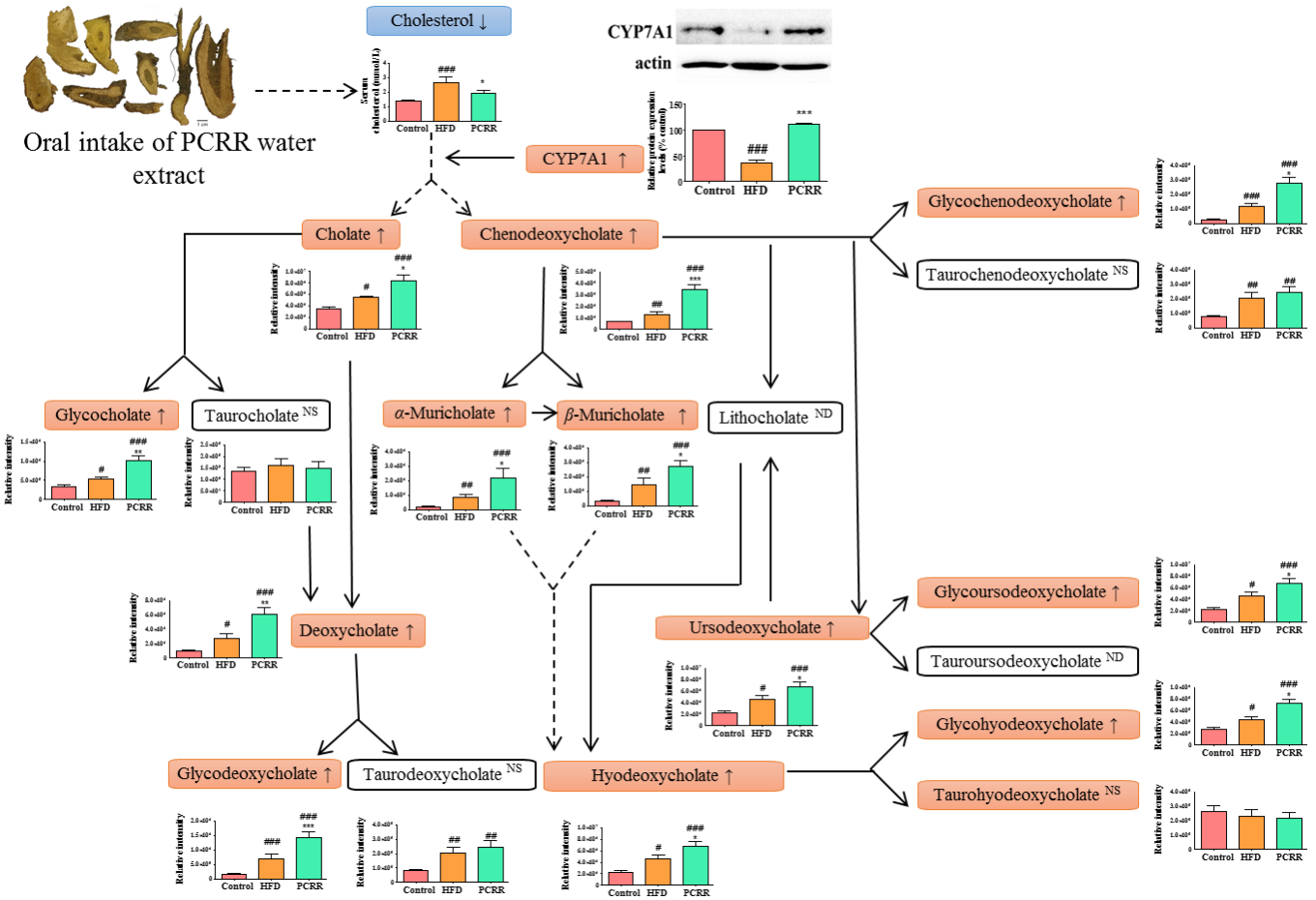


Fig. 4.10 A simplified diagram of bile acid biosynthesis altered by HFD feeding and PCRR treatment.

Data expressed as mean ± SEM. One-way ANOVA, LSD *post-Hoc* test compared with HFD-fed model. **p* < 0.05, ***p* < 0.01, ****p* < 0.001 compared with HFD group; # *p* < 0.05, ## *p* < 0.01, ### *p* < 0.001 compared with normal control group. *n* = 6-7. Control, normal control; HFD, high-fat diet induced model; PCRR, HFD model treated with high dose of PCRR water extract. Orange box with ↑, increased serum level and blue box with ↓, decreased serum level in PCRR-treated group compared with the HFD model.

4.3.7 Effects of PCRR water extract to NAFLD from liver lipidomics results

Compared with the normal control rats, the levels of liver polyunsaturated long-chained fatty acids, four TGs and three diacylglycerols (DGs) in the HFD model were obviously upregulated, which might be mainly the results of the rat chow, which led to fatty acid accumulation. Other lipid contents in the HFD model were mostly lower than those in the normal control. The changed lipids in the PCRR-treated rats were found to be mostly restored. In order to observe the overall lipid class changes, the average percentages of the sum of lipids in each class occupied in the total sum of all classes were determined and shown in Fig. 4.10. Compared with the HFD model, PCRR water extract supplementation led to reduction of the total relative contents of polyunsaturated long-chain fatty acids, cholesterol esters, DGs and TGs as well as elevation of the total relative contents of phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), lysophosphatidylcholines (LysoPC) and lysophosphatidylethanolamines (LysoPE). More details of changes in lipid profile are shown in Fig. 4.7 and Table 4.3.

The above findings were most in line with the results reported by Buechler *et al.* (Buechler 2014). There was an increase in the liver fatty acid in HFD-induced rat with NAFLD compared with the normal controls (Liu et al. 2016). Liu et al suggested that free fatty acids or their metabolites played a key role in liver injury through increased oxidative stress. Liver free fatty acids reflected the influx of fatty acids into the liver. Their high accumulation often resulted in TG synthesis along with an increased concentration of LDL-C in the circulation (Karjalainen et al. 1998). Their results were line with the observation of this animal study that the levels of liver TG and serum LDL-C in HFD model were elevated (Table 4.1).

The two kinds of lipids that were down-regulated the most in the liver with NAFLD were PCs and PEs. PCs and PEs occupy the most in phospholipids of all mammalian subcellular organelles and cell types (van der Veen et al. 2017). Total PCs tended to be lower in steatotic

and NASH liver (Buechler 2014). It was reported that the liver was a recipient of PCs from plasma lipoproteins (HDL and LDL) and donated these PCs during the assembly and secretion of lipoproteins, of which half were converted into TGs (Vance 2008). More PCs were probably converted into TGs than that received from plasma lipoproteins in the HFD-fed rats. PCRR hindered such conversion to an extent that most in the PCRR-treated group had higher PC contents than the HFD model (Table 4.3 and Fig. 4.11).

Table 4.3 Fold change of liver lipids in normal controls and PCRR-treated groups compared with the HFD model from the results of liver lipidomics

Metabolites	Adducts	Molecular formulae	Retention time (min)	Theoretical m/z	Detected m/z	Mass error (ppm)	Fold change with respect to HFD model (mean)		CV (%) in QC samples
							Normal control	HFD+ PCRR	
Phosphatidylcholines									
PC(16:0/9:0(CHO))	[M+H]+	C33H64NO9P	7.21	650.4391	650.4391	<0.1	4.15***	2.39***	8.6
PC(O-16:0/20:4)	[M+FA-H]-	C44H82NO7P	14.87	812.5811	812.5788	-2.8	3.12***	1.77***	14.3
PC(O-18:0/20:4)	[M+H]+	C46H86NO7P	14.46	796.6215	796.6211	-0.5	3.01***	1.51***	8.5
PC(20:0/20:4)	[M+FA-H]-	C48H88NO8P	16.95	882.6230	882.6205	-2.8	2.98***	2.65***	11.8
PC(18:0/17:1)	[M+Na]+	C43H84NO8P	15.71	796.5827	796.5842	1.9	2.81***	1.69**	8.4
PC(18:0/22:4)	[M+FA-H]-	C48H88NO8P	16.63	882.6230	882.6206	-2.7	2.69***	1.41***	9.7
PC(17:0/20:4)	[M+FA-H]-	C45H82NO8P	14.65	840.5760	840.5736	-2.9	2.69***	1.60**	9.5
PC(19:0/20:4)	[M+FA-H]-	C47H86NO8P	16.33	868.6073	868.6049	-2.8	2.41***	1.71***	11.9
PC(20:5/16:0)	[M+FA-H]-	C44H78NO8P	12.89	824.5447	824.5430	-2.1	2.35***	1.72**	10.9
PC(18:0/20:4)	[M+FA-H]-	C46H84NO8P	15.47	854.5917	854.5893	-2.8	1.96***	1.37***	7.4
PC(18:3/16:0)	[M+H]+	C42H78NO8P	13.13	756.5538	756.5534	-0.5	1.86***	2.17***	6.3
PC(P-16:0/22:4)	[M+FA-H]-	C46H84NO7P	15.04	838.5967	838.5942	-3.0	1.79***	2.16***	12.8
PC(17:0/22:6)	[M-CH3]-	C47H82NO8P	14.31	804.5549	804.5510	-4.8	1.74**	0.70*	7.5
PC(O-18:1/20:4)	[M+H]+	C46H84NO7P	14.92	794.6058	794.6052	-0.8	1.61***	1.89***	3.1
PC(18:1/20:4)	[M+FA-H]-	C46H82NO8P	14.06	852.5760	852.5732	-3.3	1.61**	1.56**	9.0
PC(O-16:0/16:0)	[M+H]+	C40H82NO7P	15.82	720.5902	720.5900	-0.3	1.47**	1.37*	6.3
PC(O-18:1/16:0)	[M+FA-H]-	C42H84NO7P	16.14	790.5967	790.5944	-2.9	0.80***	1.26***	9.9
PC(18:0/22:6)	[M+Na]+	C48H84NO8P	15.12	856.5827	856.5818	-1.1	0.59**	0.60**	17.3
Phosphatidylethanolamines									
PE(16:0/20:5)	[M-H]-	C41H72NO8P	13.35	736.4923	736.4910	-1.7	2.56**	1.89*	7.7
PE(18:3/16:0)a	[M-H]-	C39H72NO8P	13.42	712.4923	712.4908	-2.1	2.41***	2.16**	7.8
PE(18:0/20:4)	[M+FA-H]-	C43H78NO8P	13.07	812.5447	812.5434	-1.6	2.38***	1.74**	12.5
PE(16:0/20:4)	[M-H]-	C41H74NO8P	14.33	738.5079	738.5065	-1.9	2.02***	1.53***	4.0
PE(16:0/18:2)	[M+H]+	C39H74NO8P	14.34	716.5225	716.5223	-0.3	1.76***	1.57**	3.3
PE(17:0/20:4)	[M-H]-	C42H76NO8P	15.16	752.5236	752.5215	-2.8	1.75***	1.48**	6.4
PE(18:3/16:0)b	[M-H]-	C39H72NO8P	13.69	712.4923	712.4906	-2.4	1.74**	2.68***	8.4
PE(18:0/20:4)	[M-H]-	C43H78NO8P	16.01	766.5392	766.5378	-1.9	1.58***	1.28**	3.6

PE(20:5/18:0)	[M-H]-	C43H76NO8P	14.97	764.5236	764.5219	-2.2	1.49*	1.53**	6.4
PE(P-16:0)	[M-H]-	C21H44NO6P	6.22	436.2833	436.2819	-3.3	0.18**	0.27**	12.0
PE(P-18:0)	[M-H]-	C23H48NO6P	7.76	464.3146	464.3140	-1.4	0.15**	0.24*	13.9
Phosphatidylglycerols									
PG(22:6/22:6)	[M-H]-	C50H75O10P	10.66	865.5025	865.5012	-1.5	4.13***	1.69**	23.5
PG(18:2/18:2)a	[M-H]-	C42H75O10P	11.31	769.5025	769.5005	-2.6	2.44**	2.73***	28.5
PG(18:2/18:2)b	[M-H]-	C42H75O10P	12.06	769.5025	769.5006	-2.5	4.13***	1.69**	19.0
PG(18:1/18:2)	[M-H]-	C42H77O10P	12.23	771.5182	771.5163	-2.4	2.22**	2.51***	24.0
Sphingomyelins									
SM(41:2)	[M+H]+	C46H92N2O6P	17.40	799.6693	799.6685	-1.0	4.13***	2.39**	9.9
SM(d18:1/24:2)	[M+FA-H]-	C47H91N2O6P	16.98	855.6597	855.6575	-2.5	1.92***	1.39*	12.9
SM(d18:1/24:1)	[M+Na]+	C47H93N2O6P	18.22	835.6663	835.6658	-0.7	1.82***	1.41**	5.3
SM(33:1)	[M+H]+	C38H77N2O6P	12.40	689.5592	689.5589	-0.4	1.65**	1.46*	10.1
Ceramides									
Cer(d18:2/23:0) a	[M+H-H2O]+	C41H79NO3	19.85	616.6027	616.6024	-0.5	2.45***	1.47**	13.3
Cer(d18:2/23:0) b	[M+H-H2O]+	C41H79NO3	20.05	616.6024	616.6024	0.0	2.43***	1.42*	10.4
Cer(d18:1/18:0)	[M+H-H2O]+	C36H71NO3	17.42	548.5401	548.5399	-0.4	1.87***	1.16	8.1
Cer(d18:1/24:1)	[M+H-H2O]+	C42H81NO3	20.23	630.6183	630.6181	-0.3	1.74***	1.04	5.9
Triacylglycerols									
TG(58:5)	[M+NH4]+	C61H108O6	22.22	954.8484	954.8473	-1.2	0.53***	0.64***	8.4
TG(57:9)	[M+NH4]+	C60H98O6	21.58	932.7702	932.7698	-0.4	0.53**	0.65*	7.7
TG(58:7)	[M+H]+	C61H104O6	22.30	933.7906	933.7876	-3.2	0.48***	0.72**	12.4
TG(58:4)	[M+NH4]+	C61H110O6	22.34	956.8641	956.8610	-3.2	0.37***	0.64**	11.9
Diacylglycerols									
DG(34:2)	[M+H-H2O]+	C37H68O5	14.34	575.5034	575.5030	-0.7	1.76***	1.58**	5.0
DG(38:4)	[M+H-H2O]+	C41H72O5	21.93	627.5347	627.5348	0.2	0.43***	0.77**	8.0
DG(38:3) a	[M+H-H2O]+	C41H74O5	22.05	629.5503	629.5504	0.1	0.29***	0.69**	6.7
DG(38:3) b	[M+H-H2O]+	C41H74O5	22.19	629.5503	629.5499	-0.7	0.36***	0.73***	7.8
Cholesterol esters									
CE(22:6)	[M+Na]+	C49H76O2	22.28	719.5737	719.5733	-0.6	0.37***	0.80	8.7
CE(18:3)	[M+NH4]+	C45H74O2	22.37	664.6027	664.6024	-0.5	0.25***	0.57*	10.5
Lysophosphatidylcholines and lysophosphatidylethanolamines									
LysoPE(20: 4)	[M+Na]+	C25H44NO7P	4.63	524.2748	524.2745	-0.6	5.30***	2.37**	9.8
LysoPC(20:4)	[M+H]+	C28H50NO7P	4.41	544.3398	544.3396	-0.4	4.89***	1.86***	8.5
LysoPC(20:5)	[M+H]+	C28H48NO7P	3.52	542.3241	542.3240	-0.2	3.84***	1.87*	8.6
LysoPE(18:2)	[M+H]+	C23H44NO7P	4.67	478.2928	478.2926	-0.4	2.23***	1.72**	7.8
LysoPC(20:1)	[M+FA-H]-	C28H56NO7P	6.95	594.3776	594.3763	-2.3	0.36***	0.62*	10.3
Polyunsaturated long chain fatty acids									

C22:6 (DHA)	[M+2Na-H]+	C22H32O2	7.29	373.2119	373.2110	-2.5	0.67***	0.70***	11.1
C22:5	[M-H]-	C22H34O2	7.87	329.2486	329.2478	-2.4	0.45*	0.39**	16.2
C18:2 (Linoleate)	[M-H]-	C18H32O2	7.78	279.2330	279.2326	-1.3	0.41***	0.61*	17.9
C20:4 (Arachidonate)	[M-H]-	C20H32O2	7.60	303.2330	303.2328	-0.5	0.38*	0.39*	19.5
C20:1	[M-H]-	C20H38O2	10.88	309.2799	309.2788	-3.6	0.18***	0.45**	15.0
C20:3	[M-H]-	C20H34O2	8.35	305.2486	305.2476	-3.3	0.09***	0.37***	15.2

*One-way ANOVA, followed by LSD *post-hoc* test compared with HFD-fed model. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. n = 6-7. CE, cholesterol esters; LysoPC, lysophosphatidylcholine; LysoPE, lysophosphatidylethanolamine; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; Cer, ceramide; SM, sphingomyelin; DG, diacylglycerol; TG, triacylglycerol; ‘P-’ in bracket, presence of an alkenyl ether (plasmalogen) substituent; ‘O-’ in bracket, presence of an alkyl ether substituent. ‘a’ and ‘b’ labelled after the lipids’ name, isomers sharing similar mass fragmentation patterns.

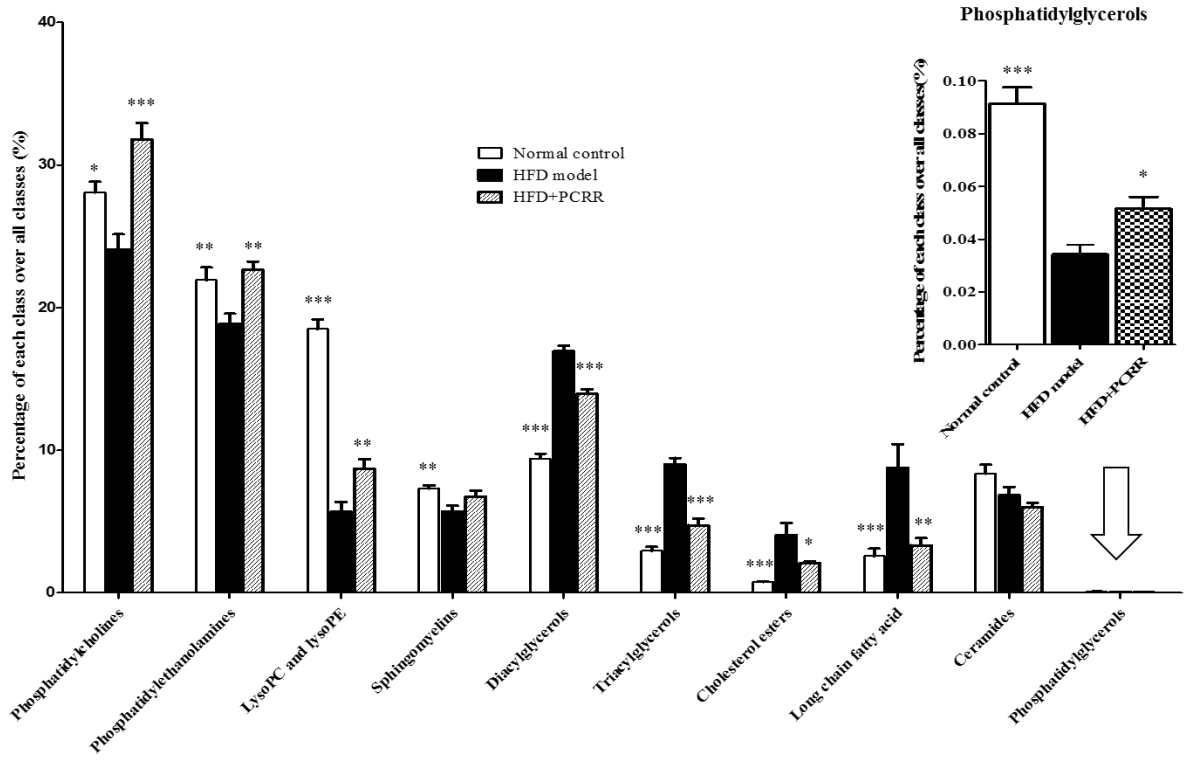


Fig. 4.11 Average percentages of each lipid class over the total sum of all classes of lipids in each group from the results of liver lipidomics.

Data expressed as mean ± SEM. One-way ANOVA, LSD *post-Hoc* test compared with HFD-fed model. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. $n = 6$.

4.3.8 Identification of unique PCRR metabolites in the PCRR-treated group

Since PCRR had high content of polydatin (resveratrol 3-O-β-D-glucoside) and resveratrol, it is believed that they are the main bioactive compounds of PCRR for the facilitation of bile acid biosynthesis from the cholesterols to primary bile acids which eventually became conjugated bile acids to be removed through the liver.

Polyphenols and their glycosides are the main bioactive chemical class from PCRR. The presence of metabolites from their metabolism after 4-week usage of PCRR was further investigated compared with HFD model. Interestingly, three of resveratrol and/ polydatin-transformed metabolites via microbial transformation were solely identified in the serum of the PCRR-treated group, matched with the observation from literatures (Bode et al. 2013; Etxeberria et al. 2015; Andres-Lacueva et al. 2012). These metabolites were putatively identified as 3,4'-dihydrobibenzyl glucuronide (lunularin glucuronide) ([M-H]-, *m/z* 389.1242), dihydroresveratrol glucuronide ([M-H]-, *m/z* 405.1194) and 3,4'-dihydrobibenzyl 3-glucuronide 4'-sulfate (lunularin 3-glucuronide 4'-sulfate) ([M-H]-, *m/z* 469.0810). Their extracted ion chromatograms in PCRR-treated serum were shown in Fig. 4.12. Their proposed mass fragmentation pathways were depicted as well (Fig. 4.13-4.15). None of them was detected in the serum of HFD model or in the water extract of PCRR. One possible explanation of their absence could be the biotransformation of polyphenols and their glycoside via gut microbiota before absorption into the circulations (Fig. 4.16). Gut microbiota could cause the reduction of the carbon-carbon double bond of resveratrol to form dihydroresveratrol and 3,4'-dihydrobibenzyl (lunularin) (Bode et al. 2013). Phase 2 transformation increased solubility of dihydroresveratrol and 3,4'-dihydrobibenzyl for circulation by glucuronide and sulfate conjugation.

It is likely that treatment of PCRR water extract might be associated with the action of gut microbiota during intervention to facilitate the dietary cholesterol-lowering effect. Further

study of gut bacteria incubation with PCRR is certainly necessary to provide a better picture on the role of gut microbiota in improving NAFLD after PCRR treatment. Future confirmation of the identity of these metabolites would be required as there is a lack of commercially available reference standards.

More studies of these resveratrol and polydatin metabolites are required in the terms of gut absorption, metabolism and their protection in humans. Also, there may be synergistic effect of resveratrol and polydatin with other PCRR polyphenols to enhance in the hypolipidemic effect in NALFD rats. Since the animal experiment in this project did not apply single polydatin or resveratrol as a positive control group to compare with PCRR, this should be carried out in future related researches.

4.3.9 Toxicity of long-term use of PCRR water extract

Although there was no observation of liver toxicity on rats after the four-week administration of PCRR water extract in this study, many plants that belong to *Polygonaceae* family have reported to have hepatotoxicity in high doses after long term use (Li, Wang, et al. 2017; Xia, Yuan, and Liu 2017; Wang, Zhao, et al. 2011). Partial hepatotoxicity was observed on human hepatocytes LO2 cells after long term exposure to physcion, emodin, emodin 8-*O*-*β*-D-glucoside and physcion-8-*O*-*β*-D-glucoside (Lv et al. 2015), which could also be found in PCRR water extract. Ethanol extract might contain higher contents and bring out stronger hepatotoxicity. Conventional water extraction of PCRR is recommended but the adverse effect for prolonged single use is still not clear and needs more *in vitro* and *in vivo* toxicity studies.

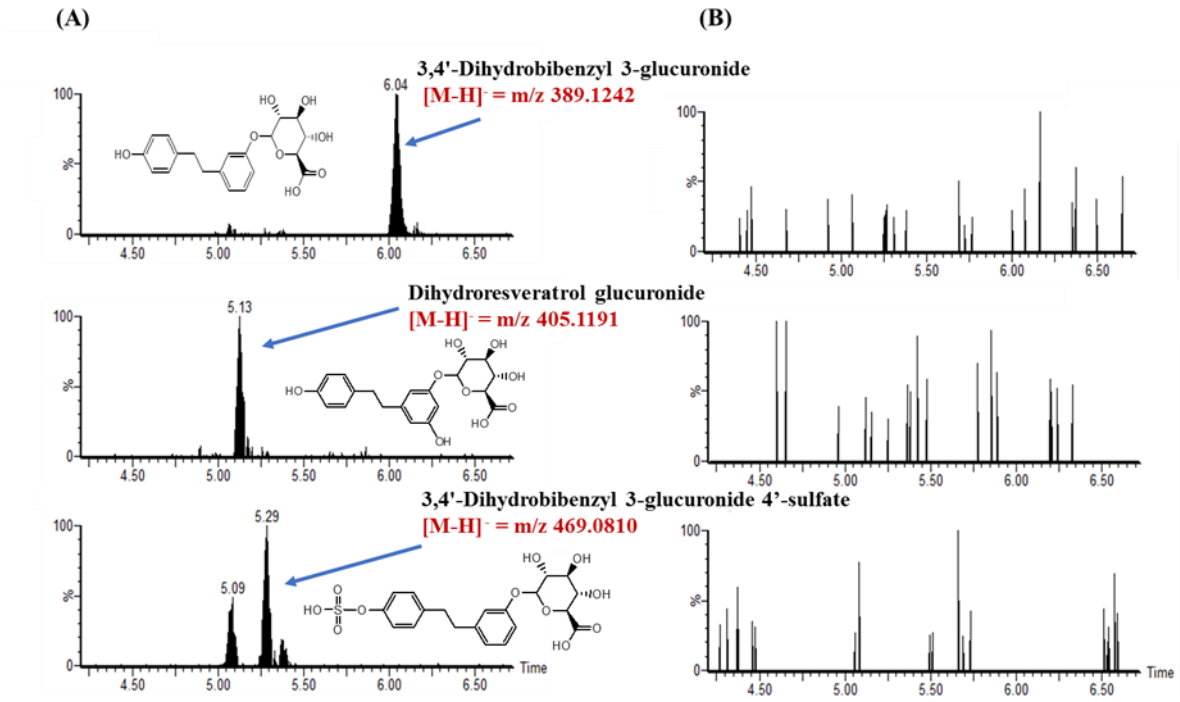


Fig. 4.12 Extracted ion chromatograms of unique and putatively identified PCRR metabolites in the PCRR-treated serum samples, acquired in UPLC-QTOF-MS at negative ESI mode.

(A) Serum of PCRR-treated HFD group. (B) Serum of HFD model with NAFLD.

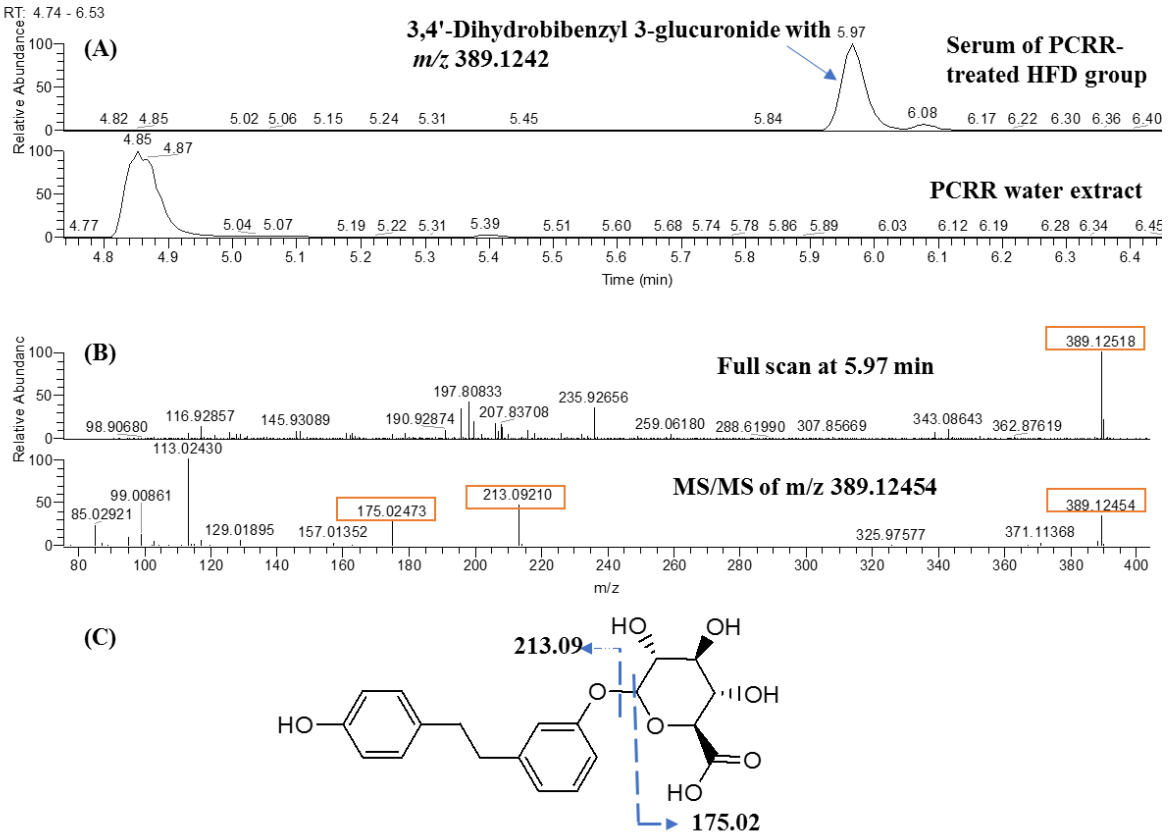


Fig. 4.13 UPLC-Orbitrap-MS results of putatively identified 3,4'-dihydrobibenzyl 3-glucuronide.

(A) Extracted ion chromatograms with [M-H]- equal to *m/z* 389.1242 at 5.97 min of serum of PCRR-treated HFD group and PCRR water extract that were at the same UPLC-Orbitrap-MS gradient. (B) Mass spectrum of serum of PCRR-treated HFD group at 5.97 min. (C) Proposed mass fragmentation pathway of the molecular structure.

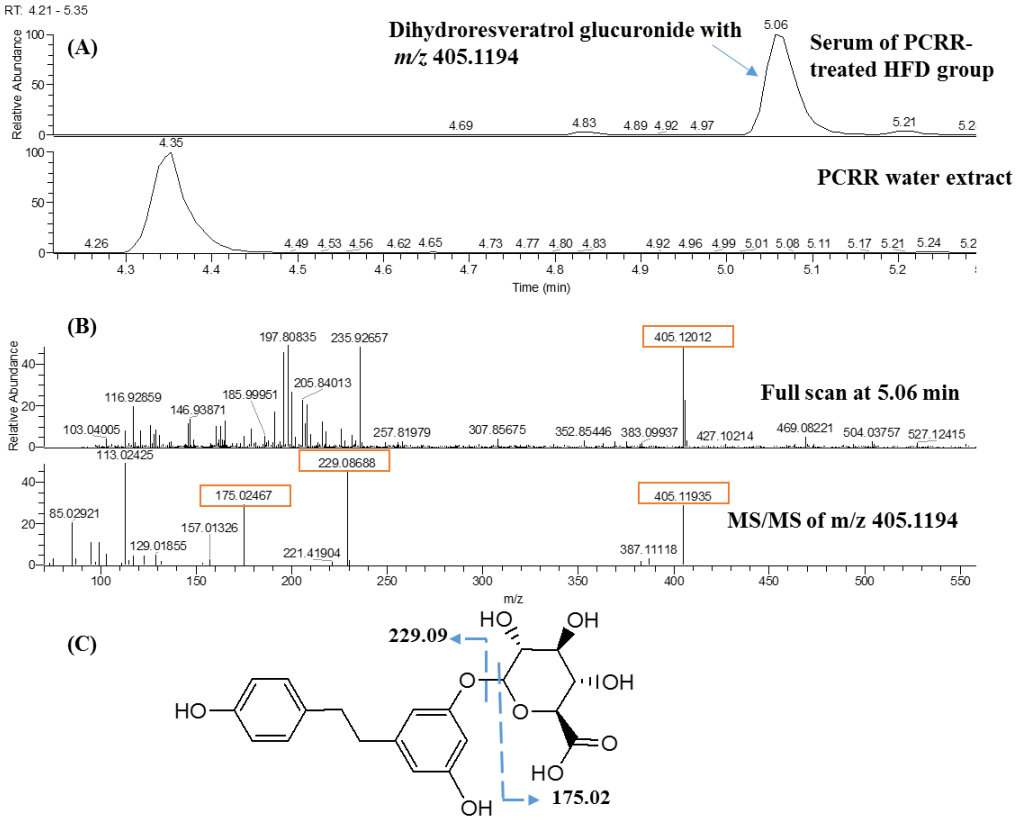


Fig. 4.14 UPLC-Orbitrap-MS results of putatively identified dihydroresveratrol glucuronide.

(A) Extracted ion chromatograms with [M-H]- equal to *m/z* 405.1191 at 5.06 min of serum of PCRR-treated HFD group and PCRR water extract that were at the same UPLC-Orbitrap-MS gradient. (B) Mass spectrum of serum of PCRR-treated HFD group at 5.06 min. (C) Proposed mass fragmentation pathway of the molecular structure.

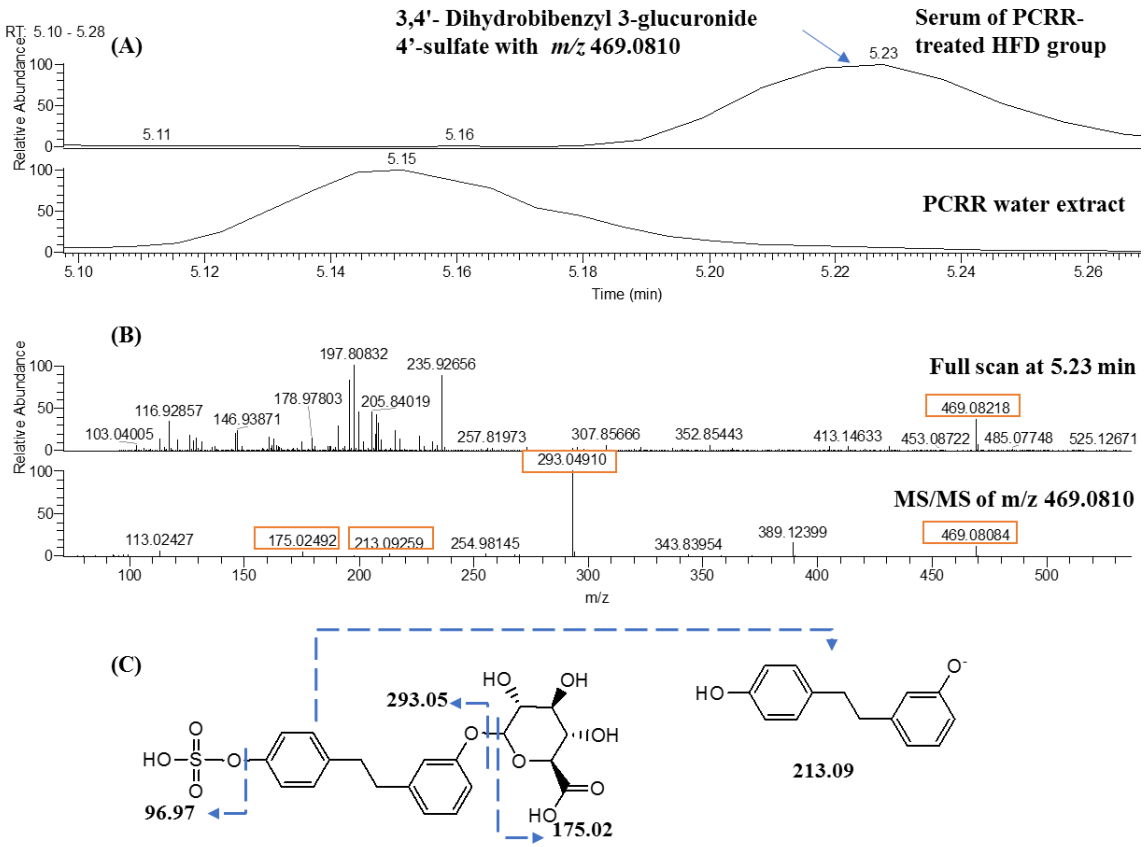


Fig. 4.15 UPLC-Orbitrap-MS results of putatively identified 3,4'-dihydrobibenzyl 3-glucuronide 4'-sulfate

(A) Extracted ion chromatograms with [M-H]- equal to *m/z* 469.0810 at 5.23 min of serum of PCRR-treated HFD group and PCRR water extract that were at the same UPLC-Orbitrap-MS gradient. (B) Mass spectrum of serum of PCRR-treated HFD group at 5.23 min. (C) Proposed mass fragmentation pathway of the molecular structure.

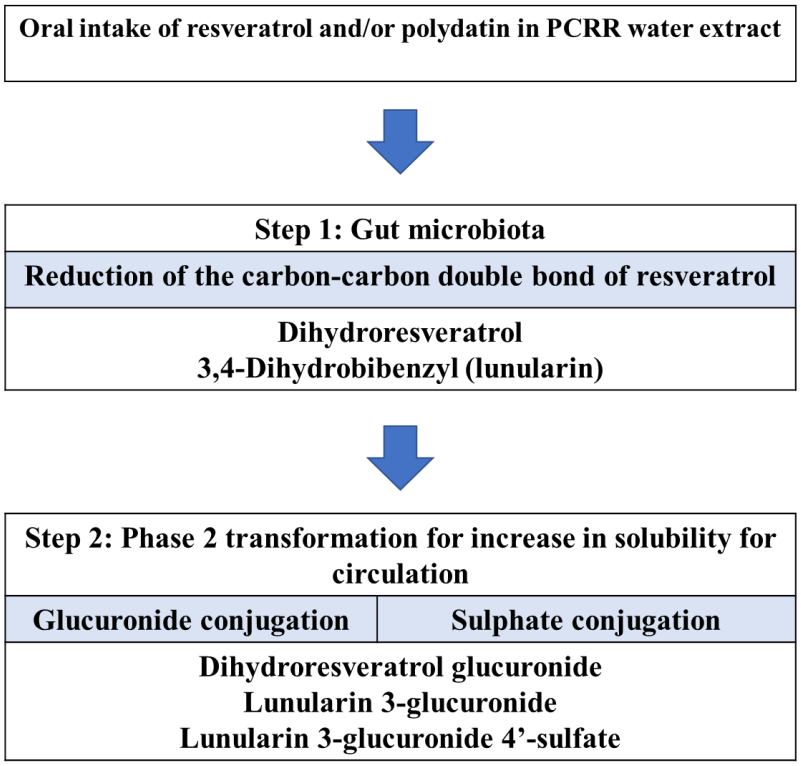


Fig. 4.16 Proposed biotransformation of orally administrated resveratrol and/or polydatin (resveratrol 3-glucoside) in PCRR water extract.

4.4 Conclusion

Chemical analysis has showed that PCRR water extract contained abundant polyphenols. Animal study showed that it lowered the serum and hepatic total cholesterol compared with the HFD model with NAFLD. The analysis of serum metabolomics showed that bile acids and their conjugated bile acids were upregulated. Lipidomics profiling of UPLC-MS revealed that it significantly restored levels of circulating free fatty acids and seven other lipid classes of liver tissues compared with the HFD model. The Western blot of the animals' liver tissue also validated that PCRR water extract upregulated the protein expressions of CYP7A1, which were related to bile acid synthesis. Resveratrol and/ polydatin-derived metabolites via microbial transformation were solely detected in the serum of the PCRR-treated group. Taken together, the treatment of PCRR water extract might be contributed by up-regulation of the bile acid biosynthesis and lipid metabolism with the action of gut microbiota, resulting in the protective effects against NAFLD. Although no adverse effect was observed in the animals after oral administration of PCRR water extract, further studies are warranted to investigate the hepatotoxicity and other adverse effects of its long-term use.

Chapter 5 Identification of metabolites associated with chronic kidney diseases in Chinese diabetic patients: an untargeted serum metabolomics study using UPLC-Orbitrap-MS

5.1 Background

Chronic kidney disease (CKD) is one of the most frequently occurred complications of diabetes. The prevalence of diabetes mellitus in adult grew significantly throughout the world from 2000 to 2017 (International Diabetes Federation 2017; International Diabetes Federation 2000) and incidence of diabetes in China was more severe than the global average rate (10.9 %) (International Diabetes Federation 2017). The American Diabetes Association indicated that 20–40% of diabetic patients would suffer from CKD (American Diabetes Association 2018). In a study of hospitalized patients in China, an increasing trend of CKD was observed from 2010 to 2015 (19.5% vs 24.3%) (Huang et al. 2018). CKD gives rise to significant health care and economic burdens to the patients for medication, dialysis and kidney transplantation therapy, especially for moderate CKD patients at Stages 3 - 4 (McQueen et al. 2017; Vupputuri et al. 2014). It is also one of the most common causes for cardiovascular events and premature death in both the developed and developing countries (Tuttle et al. 2014). The mortality risk of individuals with coexistence of type 2 diabetes and kidney disease is more than twice of those with either one only (Afkarian et al. 2013).

CKD in diabetic patients ordinarily develops from glomerular hyperfiltration, progressive albuminuria, deteriorating glomerular filtration rate (GFR), and finally to the most severe end-stage renal diseases (ESRD) (Alicic, Rooney, and Tuttle 2017). The earliest stage of CKD,

glomerular hyperfiltration, which is an aberrant increase in GFR, is an independent risk factor for progressive renal dysfunction (Ruggenenti et al. 2012). Glomerular hyperfiltration is mostly due to abnormal high blood sugar in diabetic patients (Palatini 2012). It could occur in more than 40% of type 2 diabetic patients (Jerums et al. 2010) but part of them are asymptomatic. Among patients with hyperfiltration, 49% had hyperfiltration without albuminuria (Haymann et al. 2010). Persistent glomerular hyperfiltration would cause faster renal function decline in type 2 diabetic patients (Ruggenenti et al. 2012) and raised progression risk to diabetic nephropathy (Magee et al. 2009). When kidney diseases proceed into stage 3, they are irreversible, and treatment can only slow down its progression into ESRD (Kidney disease: improving global outcomes 2013; Mora-Fernández et al. 2014; Gross et al. 2005). To prevent the decline in kidney function and extra medical burden caused by ESRD, early intervention and accurate diagnosis are extremely important. Also, accurate diagnosis can prevent unnecessary drug intake of those at low risk, minimize possible adverse effects and avert delay in proper intervention, which may lead to irreversible disease outcomes.

The main diagnostic criteria for CKD are a persistent rise in the urinary albumin excretion rate and abnormal glomerular filtration rate, indicating an impaired excretion function in the kidney. The major four renal excretion processes include glomerular filtration, active tubular secretion, active reabsorption and passive reabsorption (Felmlee, Dave, and Morris 2013), and different parts of the kidney are involved. Traditionally, GFR and UACR are common diagnostic markers but they do not reflect all of the above four excretion processes.

Estimation of GFR using serum creatinine is the most common practice but serum creatinine only reflects the glomerular filtration effect but does not cover efficiency of renal tubular reabsorption of compounds or the entire renal injury. Serum creatinine concentration also varies between male and female. It may also be increased with older age, high muscle mass

(Baxmann et al. 2008), protein-rich diet (Lew and Bosch 1991), medication (Andreev, Koopman, and Arisz 1999) and other factors. Cystatin C (Shlipak, Mattes, and Peralta 2013) and isotopic GFR measurement (Seegmiller et al. 2016; Kwong et al. 2010) are recent alternatives to measure GFR. Yet, they are not better than serum creatinine for routine use. The concentration of cystatin C could be affected by the thyroid and liver diseases (Ismail et al. 2012). Also, cystatin C alone does not have great improvement in accuracy compared with creatinine (Eriksen et al. 2012). Exogenous isotopic markers like ^{125}I-iothalamate and iohexol are costly, inconvenient to use, unsuitable in iodine allergy and may cause kidney toxicity at high dosages (Levey and Inker 2016). These clinical markers may cause false positive CKD diagnosis or underestimation, leading to delay of treatment. New, reliable, cost-effective, less invasive and convenient measurement methods and biomarkers are needed for better prediction of incidence and progression of CKD.

UACR is another marker of glomerular disease and often the first clinical indicator of the presence of CKD (2013). But there are considerable limitations about UACR. First of all, not everyone with diabetic CKD and reduced GFR has increased UACR. In the UKPDS Study, 51% of those who developed GFR < 60 mL/min/1.73 m^2 were not tested positive for albuminuria (abnormal presence of albumin in urine) (Retnakaran et al. 2006). The DEMAND Study showed 23% of type 2 diabetic patients with reduced GFRs had normal albuminuria (Thomas et al. 2009; Koye et al. 2018). Secondly, albuminuria may be affected by different factors, which influenced various body conditions, such as high-protein diet, exercise, fever, infection, hyperglycemia, high blood pressure and congestive heart failure (American Diabetes Association 2004; Bachmann et al. 2014). Thirdly, measurement of albuminuria lacked standardization and demonstrates significant imprecision (Seegmiller, Miller, and Bachmann 2017). Lastly, discordance between albuminuria changes and renal disease events has also been

observed in the ACCORD Study (Ismail-Beigi et al. 2010) and the UKPDS Study (Bilous 2008).

These insensitive and inaccurate clinical markers may cause false positive or false negative CKD diagnosis or underestimation, leading to delay of treatment. More new, reliable, cost-effective, less invasive and convenient methods are needed for better prediction of incidence and progression of CKD.

So far, many metabolomics studies have investigated for new risk markers of kidney diseases but they mainly focus on the prediction of incidence of kidney damages at the later stages 3-5 of CKD (Zhang, Zhang, and Wang 2015; Davies 2018; Canadas-Garre et al. 2019; Chen et al. 2019). Trimethyaminde N-oxide, kynurenine and critulline are common potential biomarkers in CKD. Acylcarnitines and phospholipids from CKD patients have been the most popular research targets owing to their targeted metabolomics method (Canadas-Garre et al. 2019). N, N, N-trimethyl-L-alanyl-L-proline betaine (L,L-TMAP) (Velenosi et al. 2019), a newly identified biomarkers for patients with ESRD from normal subjects, was found in a recent CKD study.

This study aimed at increasing the accuracy in differentiating various CKD stages in diabetic patients, especially its earliest stages, via exploring novel CKD stage-related metabolites and comparing them with existing clinical markers. Herein, we used the non-targeted metabolomics profiling to look for more sensitive and specific potential biomarkers that could predict renal disease progression across four stages of CKD, including CKD patients with glomerular hyperfiltration. The biomarkers were then validated independently with another cohort, a step that was usually absent in previous studies (Colhoun and Marcovecchio 2018).

5.2 Research design and methods

5.2.1 Chemicals

Acetonitrile and methanol (HPLC grade) were purchased from Duksan (ANSAN-SI, South Korea). Distilled water was purified using a Milli-Q water purification system (Millipore, Bedford, MA, USA). Formic acid, *cis*-10-nonadecenoic acid (C19:1n9c) was obtained from Sigma Aldrich (St. Louis, MO, USA). Cholic acid-2,2,4,4-D_4 and L-tryptophan-(indole-D_5) was purchased from Cambridge Isotope Laboratories (Tewksbury, MA, USA). The suppliers of reference standards used for metabolite identification are listed in Table 5.7.

5.2.2 Study participants

The serum samples used in this study were collected from two independent cohorts recruited at two different periods of time; they were referred as the discovery set and validation set, respectively.

For the discovery set, 108 eligible inpatients with type 2 diabetes and 20 normal subjects aged 40-70 were recruited from the Shenzhen Traditional Chinese Medicine Hospital between October, 2016 and October, 2017. At the subject enrollment stage, each participant was clinically examined and interviewed by our trained recruiters. The exclusion criteria were acute renal failure, rapidly increasing proteinuria or nephrotic syndrome, refractory hypertension, serious infections, signs or symptoms of other systemic disease, known renal tubular acidosis, pregnancy, type 1 diabetes, gestational diabetes, chronic liver disease, serious cardiovascular diseases, alcoholics or malignant tumour. For healthy control, they had not received any treatment like antibiotics, probiotics and hormone therapy in the past two months, did not have proteinuria or history of kidney disease, their oral glucose tolerance test and other related clinical test are in normal levels. The diagnosis of diabetes, kidney disease and classification were based on the criteria of the Kidney Disease: Improving Global Outcomes (KDIGO) 2012 Clinical Practice Guideline (Kidney Disease: Improving Global Outcomes (KDIGO) CKD

Work Group 2013). Patients were grouped by CKD severity according to their estimated GFR and UACR.

Their GFRs were used to classify 108 patients in the discovery set: stage 1 (GFR ≥ 90 ml/min/1.73 m^2); Stage 2, 60-89 mL/min/1.73 m^2; Stage 3, 30-59 mL/min/1.73 m^2; Stage 4, < 30 mL/min/1.73 m^2 . Stage 1 consisted of Stage 1a and 1b. The GFR range of Stage 1b was defined as normal (GFR = 90-120 mL/min/1.73 m^2) and those over the normal upper limit (> 120 mL/min/1.73 m^2) were assigned as stage 1a of CKD in this study.). In this study, early stages was indicated as Stages 1a-2 (GFR ≥ 60 ml/min/1.73 m^2) while later stages was indicated as Stages 3-4 (GFR < 60 ml/min/1.73 m^2). In the normal group at Stage 0, 20 subjects with GFR = 90-120 mL/min/1.73m^2 were recruited during routine body check. Their age range and gender ratio matched with those of the inpatients in discovery set.

UACR (mg/g Cr.) is the ratio of urinary albumin concentration (mg) to urinary creatinine concentration (g). Persistent increased protein in the urine (two positive tests over three months or more) is considered as a sign of kidney damage. There were three persistent albuminuria categories according to KDIGO guideline (Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group 2013): normoalbuminuria, (normal to mildly increased group, UACR < 30 mg/g Cr.), microalbuminuria (moderately increased group, UACR 30-300 mg/g Cr.) and macroalbuminuria (severely increased group, UACR > 300 mg/g Cr.). The stages of CKD and albumuria categories of the present study are shown in Table 5.1.

An external validation set, that was another separate cohort, was recruited in October, 2017-April, 2018 and it contained 56 type 2 diabetic inpatients and 10 normal subjects, who were enrolled independently and was classified under the same condition as described above as those in the discovery set. Body surface-area related renal volume and serum cystatin C of subjects

in the validation set were also measured. Their baseline characteristics are also shown in Table 5.3.

Table 5.1 Stages of CKD and albumuria categories that were applied in the present study

Stages of CKD*		**GFR (mL/min/1.73 m^2)**
Stage 1	**Stage 1a (glomerular hyperfiltration)**	>120
	Stage 1b (normal kidney function)	90-120
Stage 2 (mild loss of kidney function)		60-89
Stage 3 (mild to moderate loss of kidney function)		30-59
Stage 4 (severe loss of kidney function)		< 30
Albumuria categories*		**UACR (mg/g** Cr.**)**
Normoalbuminuria (normal to mildly increased)		<30
Microalbuminuria (moderately increased)		30-300
Macroalbuminuria (severely increased)		> 300

*The diagnosis of CKD and albumuria classification were modified from the criteria of the KDIGO 2012 Clinical Practice Guideline (Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group 2013).

5.2.3 Ethical approval

The use of protocol for this study and informed consent forms was approved by the Research Ethics Board at the Shenzhen Hospital of Guangzhou University of Chinese Medicine [Shenzhen Traditional Chinese Medicine Ethics (Research) (2016) No.8]. Written informed consent was obtained from all subjects prior to this study.

5.2.4 Specimen collection

Serum and urine samples were collected in patients in the hospital. No additional treatments or drug were given. Normal subjects and diabetic patients were instructed to have simple meals and stop taking unnecessary drugs the day before blood collection. The subjects were requested to fast 8 hours. Serum and urine were collected at 8:30 - 9:30 am. About 3 mL blood was drawn by venipuncture of antecubital area and clotted for serum collection. About 3 mL of the first morning midstream urine was collected into a clean container. Clotted blood and freshly collected urine were centrifuged at 4 ℃ at 3000 rpm for 15 min. Supernatants were aliquoted and stored at -80 ℃ prior to testing.

5.2.5 Selected equations for comparisons with MDRD GFR

This study applied one widely used equation for GFR estimation, the abbreviated Modification of Diet in Renal Disease (MDRD) Study equation (Levey et al. 2003) to calculate MDRD GFR in mL/min/1.73 m^2. The relationship of three other published equations, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) $_{creatinine}$ equation (Levey et al. 2009), CKD-EPI $_{cystatin\ C}$ equation (Inker et al. 2012) and CKD-EPI $_{creatinine–cystatin\ C}$ equation (Inker et al. 2012) with MDRD GFR by correlation were assessed and their equation details are listed in Table 5.2.

Table 5.2 MDRD study equation and other existing GFR equations.

GFR calculation methods	Equations
MDRD Study equation **(Levey et al. 2003)**	Estimated GFR = 186 × (serum creatinine)$^{-1.154}$ × (age in years)$^{-0.203}$ × 0.742 (if female) × 1.210 (if African American)
CKD-EPI $_{\text{creatinine}}$ equation **(Levey et al. 2009)**	Estimated GFR = 141 × min (serum creatinine /κ, 1)α × max (serum creatinine /κ, 1)$^{-1.209}$ × 0.993$^{\text{Age in year}}$ × 1.018 (if female) × 1.159 (if black), where κ is 0.7 for females and 0.9 for males, α is -0.329 for females and -0.411 for males. min indicates the minimum of ratio of serum creatinine to κ or 1, and max indicates the maximum of ratio of serum creatinine to κ or 1.
CKD-EPI $_{\text{cystatin C}}$ equation **(Inker et al. 2012)**	Estimated GFR = 133 × min(serum cystatin C /0.8, 1)$^{-0.499}$ × max(serum cystatin C /0.8, 1)$^{-1.328}$ × 0.996$^{\text{Age}}$ × 0.932 (if female), where min indicates the minimum of ratio of serum cystatin C to 0.8 or 1, and max indicates the maximum of ratio of serum cystatin C to 0.8 or 1.
CKD-EPI $_{\text{creatinine–cystatin C}}$ equation **(Inker et al. 2012)**	Estimated GFR = 135 × min (serum creatinine /κ, 1)$^{\alpha}$ × max (serum creatinine /κ, 1)$^{-0.601}$ × min (serum cystatin C/0.8, 1)$^{-0.375}$ × max(serum cystatin C /0.8, 1) $^{-0.711}$ × 0.995$^{\text{Age}}$ × 0.969 (if female) × 1.08 (if black), where κ is 0.7 for females and 0.9 for males, and α is -0.207 for males and -0.248 for females.

GFR, serum creatinine and serum cystatin C are expressed in mL/min/1.73 m^2 of body surface area, mg/dL and mg/L respectively.

5.2.6 Calculation of body-surface-area (BSA)-related renal volume and renal resistive index (RRI)

BSA (m^2) was calculated using the D. Du Bois and E. F. Du Bois's formula: weight (kg) $^{0.425}$ × height (cm) $^{0.725}$ × 0.007184 (Du Bois and Du Bois 1989; Scholbach and Weitzel 2012). Renal volume (mL) was calculated based on the formula using kidney dimension measured by ultrasound imaging: length × width × depth × 0.523 (Scholbach and Weitzel 2012). BSA-related renal volume (mL /m^2) (Hricak and Lieto 1983) was calculated as a ratio of renal volume to body surface area. RRI was calculated as (peak systolic velocity - end diastolic velocity) / peak systolic velocity derived from the kidney doppler ultrasonography.

5.2.7 Non-targeted UPLC-Orbitrap-MS metabolites profiling analysis

5.2.7.1 Serum preparation

60 µL serum was deproteinated with 240 µL cold methanol containing 0.5 ppm L-tryptophan (indole-D_5) and 0.5 ppm cholic acid-2,2,4,4-D_4 and 50 ppm C19:1n9c. They were vortexed for 1 min and stood at -20 ^{o}C overnight for complete deproteination. Then, they were centrifuged at 18700 × g for 20 min. 250 µL supernatant was collected and dried under nitrogen gas and stored at -80 ^{o}C. The dried supernatant was reconstituted with initial UPLC gradient (5% acetonitrile in water), vortexed for 30 s and was centrifuged at 18700 × g for 20 min. The supernatant was transferred to a glass insert in an amber HPLC vial prior to UPLC-Orbitrap-MS analysis.

5.2.7.2 UPLC condition

3 µL aliquot was injected into a Waters ACQUITY UPLC system. UPLC separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm x 100 mm, 1.8 µm) with HSS T3 guard column (2.1mm x 5 mm, 1.8 µm, Waters Corporation, Milford, MA). The mobile phase consisted of combinations of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 0.3 mL/min with elution gradient as follows: 0-1.5 min, 5% B; 2 min, 35% B; 4 min, 50% B; 8 min, 55% B; 11-14 min, 95% B. A 3-min

post-run time was set to fully equilibrate the column. Column and sample chamber temperature were 40 ^{o}C and 4 ^{o}C respectively.

5.2.7.3 Mass spectrometry condition

Mass spectrometry analysis was conducted by a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with a heated electrospray ionization (H-ESI) interface (Thermo Fisher, Waltham, MA, USA). The mass-spectrometric parameters were set as follows: spray voltage, 2300 V and 3500 V in ESI negative and positive ionization modes respectively; ion transfer tube and vaporizer temperature, 300℃. Nitrogen gas was used as the sheath gas and the aux gas with a flow rate of 25 and 10 L/min, respectively. The analyser was operated in a data-dependent acquisition mode, with full MS scans of mass range at 90–1000 m/z with detection in the Orbitrap (120000 resolution) and with auto gain control targeted at 20000 count and a maximum injection time at 100 ms.

20 µL aliquots from each sample of all groups were mixed and aliquoted as QC samples. QC samples were injected between every six-sample injections to monitor the stability of the instruments throughout the UPLC-MS signal acquisition. The order of injection for all samples was randomized.

5.2.7.4 Data extraction and pre-processing

Chromatographic peak picking and alignment were performed using Progenesis QI 2.3 (Nonlinear Dynamics, Newcastle upon Tyne, United Kingdom). Sensitivity and chromatographic peak width were optimized accordingly. Peaks at retention time 0.3 min – 14 min were selected for study. Matrix of all raw ion abundance without any normalization were exported to Matlab (MathWorks, Natick, MA, USA) for data pre-processing (missing value imputation and signal correction).

Variables, whose values were missed in over 40% of samples in any group, were excluded from the subsequent analysis because the results of these variables were unreliable (Armitage et al. 2015; Wei et al. 2018). Then, imputation was performed on the remaining variables that contained missing values with equation (1) in each group separately. The missed value was the median in each group plus random errors based on the variations of raw ion abundance.

$$\boldsymbol{x_{g.i}^{mis} = median(x_{g,i}^{raw}) + e_{g,i}} \quad (1)$$

where $\boldsymbol{i}$: variable in *m/z* ratio;

$\boldsymbol{g}$: group;

$\boldsymbol{x_{g,i}^{mis}}$: missing value in the variable $\boldsymbol{i}$ in the group $\boldsymbol{g}$ for imputation;

$\boldsymbol{x_{g,i}^{raw}}$: vector containing all the raw ion abundances of the variable $\boldsymbol{i}$ in the group $\boldsymbol{g}$;

$\boldsymbol{e_{g,i}}$: a random error related to the standard deviation of $\boldsymbol{x_{g,i}^{raw}}$, which was a simulation of the variation of ion abundance in the same variable in the group between different experimental runs.

Signal correction was then performed by smoothing through QC samples in the injection sequence that was served as baseline correction using cubic spline interpolation. Cubic splines were very flexible smoothers that could catch the variations of ion abundances caused by systematic bias in the instrumental responses, with a very wide range of curve shapes (e.g., linear and nonlinear curve) (van der Kloet et al. 2009). The optimal smoothing parameter was set at 0.01 and fit well to the variations. Then, the relative ion abundance $\boldsymbol{r_{n,i}}$ was corrected with the instrumental variation and would be used for subsequent analysis. It was a ratio of raw ion abundance with respect to the predicted ion abundance of the variable $\boldsymbol{i}$ of the QC sample at the $\boldsymbol{n}^{\text{th}}$ injection after signal correction and was calculated by the equation (2) below.

$$\boldsymbol{r_{n,i} = \frac{x_{n,i}}{q_{n,i}}} \quad (2)$$

where n: order number of the QC and study samples at the n^{th} injection in the injection sequence;

$r_{n,i}$: relative ion abundance r with respect to the predicted ion abundance of the variable i of the QC sample at the n^{th} injection;

$x_{n,i}$: raw ion abundance x of the variable i at the n^{th} injection;

$q_{n,i}$: predicted ion abundance q of the variable i at the n^{th} injection that was generated by using QC samples that fitted into the interpolating cubic splines.

In the equation (2), $q_{n,i}$ was a new predicted ion abundance of the variable i at the n^{th} injection that was generated by a smoothing curve. The smoothing curve was created by fitting the raw ion abundances of all the QC samples into the interpolating cubic spline. In an ideal case that all signals were perfectly fitted into the curve, all $q_{n,i}$ were equal to their corresponding raw ion abundances, $x_{n,i}^{QC}$, in each QC sample, and the curve constructed by cubic splines should go through every raw ion abundance of QC samples. If there was no significant deviation between the serum levels of study samples and QC samples, the value of $x_{n,i}^{sample}$ should be similar to that of $q_{n,i}$. Hence, $r_{n,i}$ should be close to 1. Fig. 5.1 gave a demonstration of the comparison of raw and relative ion abundances of L-tyrosine and L-phenylalanine in the study samples and QC samples before and after signal correction. The figure showed that after signal correction, QC samples clearly followed a straighter horizontal line with less variations.

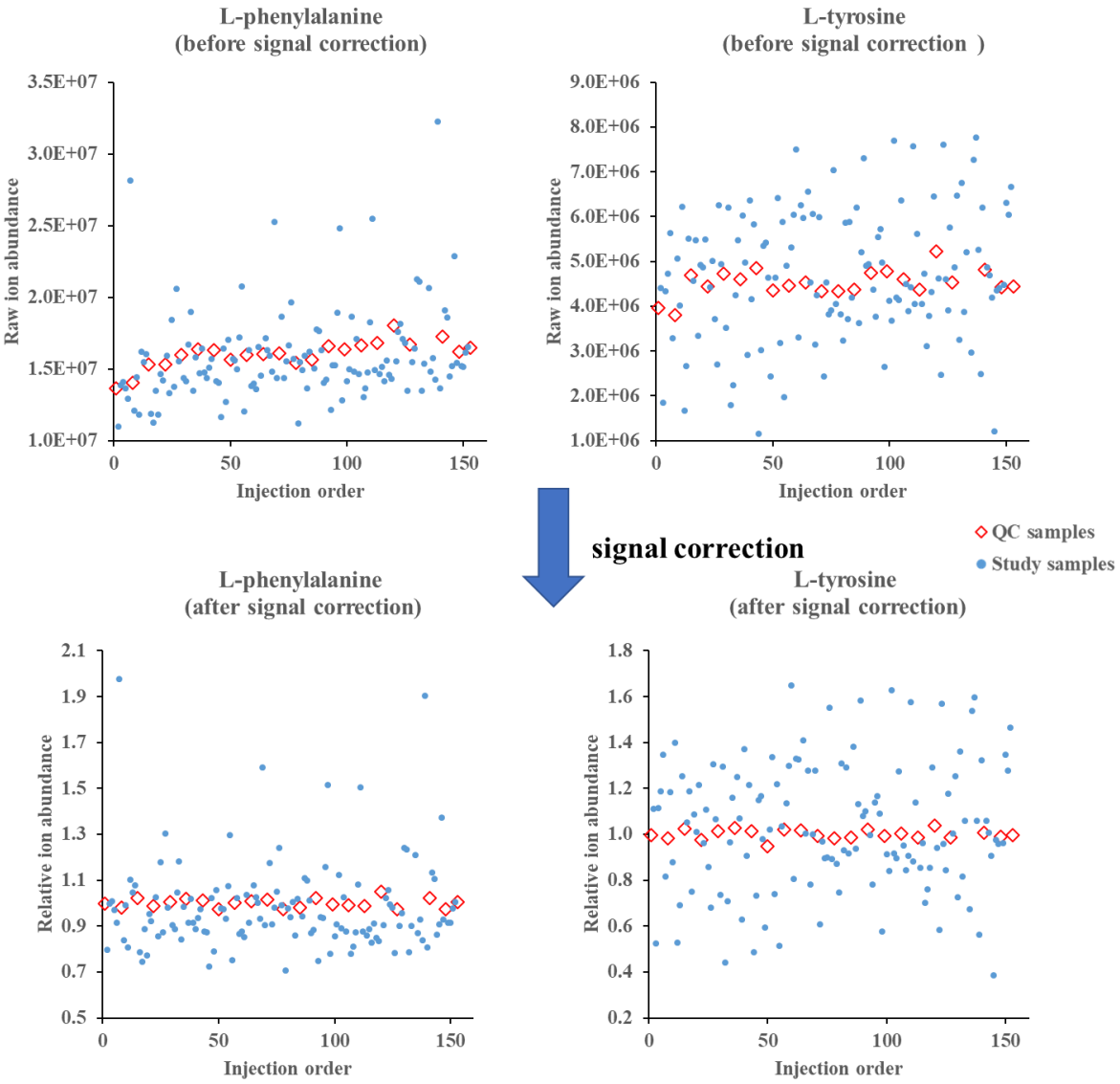


Fig. 5.1 Comparison of raw and relative ion abundances of L-tyrosine and L-phenylalanine in the study samples and QC samples before and after signal correction of data acquired by UPLC-Orbitrap-MS.

5.2.8 Initial variable selection

First, quality screening was done by filtering out those unstable metabolites by applying cut-off CV > 30% in all QC samples. Then, variable selection of Spearman rank correlation of LC-MS data with GFR and UACR was evaluated in the discovery set separately ($p \leq 0.05$). After Student's *t*-test and one-way ANOVA test with comparison between each stage and normal group after $\log_{10}$ transformation, metabolites with $p \leq 0.05$ were selected for peak identification.

5.2.9 Peak identification

Metabolites were first putatively annotated based upon their mass to charge ratio (*m/z*) and mass fragmentation patterns with the built-in database of Progenesis QI 2.3 (Chemspider, Lipidmaps), online metabolite databases such as Human Metabolome Database (http://www.hmdb.ca/), Metlin (http://metlin.scripps.edu), MassBank and literatures. Most of their identities were further confirmed with commercial reference standards under the same UPLC-MS conditions. Those markers that shared different retention time but had similar mass fragmentation pattern with reference standards were given the name "isomer" following the reference standard's name as in "arabinose isomer".

In order to compensate for the insufficient information of online identification database, those identified metabolites were placed into two online pathway analyses [MetScape that was built in Cytoscape 3.7.1 (Boston, MA, USA) and Ingenuity Pathway Analysis (QIAGEN Inc, USA)] for pathway discovery to check any unexamined metabolites that were present in the pathway. The presence of these unexamined metabolites in the serum sample was checked again.

5.2.10 Validation of discovery set

Those identified peaks in the discovery set were further confirmed in the independent validation dataset. They were examined through the same procedure of the variable selection and peak identification. Metabolites in the validation set that had the same mass fragmentation

pattern, retention time and that shared the same trend with significant statistical differences with those in the discovery sets were further investigated.

5.2.11 Metabolic network

Networks of altered metabolites (only those with Pubmed ID listed in Table 5.8) were generated through MetaMapp based on their metabolic pathway and chemical similarity. Their relationship output and fold change were processed with CytoScape to create visual diagrams. Their metabolic pathway output were based on their KEGG reaction pair database while their chemical and structural relationship depended on their Tanimoto similarity (Barupal et al. 2012). The pathway enrichment analysis from Metaboanalyst 4.0 (http://www.metaboanalyst.ca) was conducted using the above metabolites with HMDB ID to further identify their metabolic networks.

5.2.12 Comparison method and statistical analysis

5.2.12.1 Statistical analysis for selection of potential biomarkers

Results were expressed as means (standard deviation, SD) or medians (interquartile range), counts (*n*), as appropriate. All non-normally distributed variables were log10 transformed. Statistical analyses were performed using IBM SPSS Statistics version 25 (Chicago, IL, USA). The significant features between groups in continuous variables were analyzed using unpaired Student's *t* - test, one-way ANOVA followed by LSD *post-hoc* test. All tests were two-tailed with $p \leq 0.05$ considered as statistically significant. Screening of metabolites were further tested with false discovery rate (FDR, < 0.1 using Metaboanalyst) adjusted for multiple comparisons. Fold change was calculated using median to minimize the effect of outliers. Spearman rank correction could be used for non-parametric data to look for biomarker candidates with high accuracy of prediction. Thus, Spearman rank correlation of identified metabolites with fasting blood glucose (FBG), hemoglobin, hemoglobin A1c, BSA-related

renal volume, renal resistive index, GFR, serum creatinine, cystatin C, UACR and urinary β2-microglobulin were calculated. Selected metabolites that reached significant associations with correlation coefficient (Spearman R) $\geq$ 0.4 were listed.

5.2.12.2 Area under the receiver operating characteristic curve (AUC)

To evaluate the performance of selected metabolites on differentiating CKD stages, AUC was calculated by RF and logistic regression using single, two- and three-metabolite models. RF is a powerful supervised classification technique that works by building a large number of decision tree models on a sub-dataset obtained by bootstrap sampling of input data matrix and merging all predictions from these trees to get an accurate and stable prediction (Boulesteix et al. 2012). Samples were randomly split into a training set (50%) and a testing set (50%) for model training and prediction respectively, to avoid overfitting. This procedure was repeated 100 times to calculate the average AUC of the testing set. The statistical analyses were scripted in house using Python (https://www.python.org/). Logistic regression and RF classification were implemented by scikit-learn (https://scikit-learn.org) (Pedregosa et al. 2011).

5.2.12.3 Multivariate linear regression analyses

Multivariate linear regressions by IBM SPSS Statistics were applied to determine the best models with biomarkers that could jointly predict MDRD GFR better than MS-detected creatinine. The relative intensities of potential biomarkers, values of MDRD GFR and UACR were log10-transformed in favour of normal distribution requirement for the analysis. The multivariate linear regression used ENTER method (all predictors were forced into the model simultaneously) to assess the performance of each variable in one model and stepwise method (all redundant predictors were removed step by step) to select the best model using discovery set as training samples (108 diabetic patients) with common confounding factors [age, gender, systolic blood pressure (SBP), BMI and log [UACR]]. The unstandardized regression coefficients (β) of the best fitted model were applied to obtain the predicted log [MDRD GFR]

using the independent validation sets (56 diabetic patients). Only variables with significant $p < 0.05$ that contributed to the linear regression model fit were selected in the best model while other predictors in the models that had no effect on the response were removed in favour of a simpler model. The predicted log [MDRD GFR] and predicted MDRD GFR (after conversion from log10 transformation for easy interpretation) were compared with the actual values in the separate validation sets by linear regression again to assess the performance of the prediction equation. Using two separate datasets for construct a model and test for prediction performance not only could remove systematic bias of samplings from identical datasets and overfitting but increase the robustness of the final model.

5.3 Results

5.3.1 Clinical characteristics of diabetes and CKD in different stages

Both cohorts were comparable in terms of baseline characteristics (Table 5.3). Among 194 participants in both cohorts, those with diabetes [164 subjects; mean age, 55.3 (7.1) years] and those without diabetes [30 subjects, mean ages, 49.9 (6.3) years] were different in GFR and UACR. The median (interquartile range) duration of diabetes was 8 (4-14) years. The levels of FBG of patients at Stage 1a were higher than those of the normal group ($p < 0.05$). GFR was reduced and the UACR, SBP, the contents of serum creatinine, homocysteine, fasting C-peptide, urea, uric acid, cystatin C, urinary microalbumin, and urinary protein to creatinine ratio, urinary β2-microglobulin were increased with the stages of CKD, which were more obvious in Stage 4. All these changes showed the loss of kidney clearance function, leading to accumulation of uremic toxin.

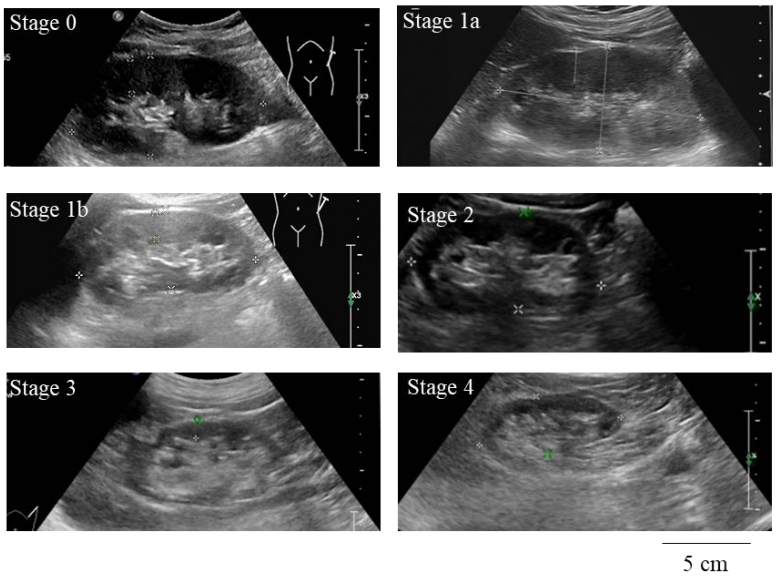
The renal ultrasound images were assessed in the validation sets and kidney size was enlarged at Stage 1a and gradually diminished in the present study as shown by renal ultrasound images and BSA-related renal volumes (Fig. 5.2). Patients at Stage 1a had the largest left [median (interquartile), 92 (84-125) mL/m^2], right [96 (88-108) mL/m^2] and total BSA-related renal volumes [189 (177-220) mL/m^2] while those at Stage 4 had the smallest left [median (interquartile), 55 (41-72) mL/m^2], right [66 (47-75) mL/m^2] and total BSA-related renal volumes [124 (95-136) mL/m^2] (Table 1). Patients at Stage 4 also had the highest RRI [mean (SD) = 0.75 (11)] and systolic blood pressure (149-152 mmHg) (Table 5.3).

Table 5.3 Clinical characteristics of participants in the discovery and validation sets according to stages of diabetic CKD.

Clinical parameters	Stages (MDRD GFR, Discovery set)						Stages (MDRD GFR, Validation set)					
	0	1a	1b	2	3	4	0	1a	1b	2	3	4
n	20	20	19	30	19	20	10	10	10	10	15	11
Age (years)	50(7)	51(5)	54(5) [a]	55(7) [a]	60(9) [abcd]	61(5)[abcd]	50(5.6)	50(2)	50(3)	54(7)	58.9(6) [abc]	55.7(9)
Gender (Male, n)	10	10	10	15	9	12	5	5	5	5	9	7
Duration of diabetes (years)	-	7(6)	5(6)	10(6) [c]	10(8) [c]	12(6)[bc]	-	5.9(4.7)	8.7(3.3)	9.3(7.4)	8.1 (4.7)	15.4(10.2) [be]
BMI (kg/m^2)	21.8(1.5)	23.3(4.1)	25.7(3.5)[a]	25.2 (3.1) [a]	27.0(3.0)[ab]	27.0(3.3) [a]	20.8(1.2)	23.4(2.8) [a]	24.9(1.5) [af]	25.9(3.5) [af]	25.0 (3.0) [af]	22.5(2.8)
SBP (mmHg)	117(13)	128(20) [a]	136(23) [a]	132(20) [a]	135(20) [a]	149(26)[abd]	111(5)	122(17)	126(2.3)	147 (19) [abc]	139 (20) [ab]	152(22)[abc]
DBP (mmHg)	74(11)	80(11)	86(14) [abf]	78(12)	79(9)	77(12)	77(7)	82(11)	87(21)	91 (9) [a]	86 (12)	85(11)
Total cholesterols (mmol/L)	4.7(0.5)	5.3(1.1) [ac]	4.5(1.2)	4.9(1.2)	5.0(1.4)	4.8(1.8)	4.5(0.5)	5.0(0.7)	5.2(1.1)	4.7 (0.9)	4.9 (1.1)	4.8(1.5)
LDL-C(mmol/L)	3.1(0.5)	3.5(1.1)	2.9(1.1)	3.1(1.1)	3.4(1.4)	3.1(1.6)	2.7(0.4)	3.4(0.8) [a]	3.5(0.9) [a]	3.1 (0.6)	3.2 (1.0)	2.7(1.1)
HDL-C(mmol/L)	1.45(0.2)[bc def]	1.1(0.3)	1.1(0.5)	1.0(0.3)	1.1(0.3)	1.1(0.3)	1.5(0.4)[def]	1.3(0.3)	1.2(0.4)	1.2 (0.3)	1.2 (0.3)	1.0(0.3)
Triglycerides (mmol/L)	1.2(0.5)	2.7(3.4)	1.9(0.7) [a]	2.5(2.3) [a]	2.0(1.0) [a]	1.5(0.8)	1.1(0.7)	1.5(0.6)	1.8(0.8) [a]	1.7 (1.0)	2.0 (0.9) [a]	3.6(5.7)
Serum creatinine (umol/L)	65(62-75)	49(45-62) [a]	73(55-77) [b]	83(72-91) [abc]	121(107-139) [abcd]	335(232-627)[abcde]	62(51-63)	48(43-60) [a]	65(56-73)[b]	86 (72-106) [abc]	119 (101-125) [abcd]	709(289-858) [abcde]
MDRD GFR (ml/min/1.73 m^2)	97(93-104)[def]	135(124-151) [acdef]	102(96-105) [def]	78(72-84) [ef]	52 (41-56)[f]	15(6.4-25)	110(98-120) [def]	137(121-149) [acdef]	104(99-110)[def]	75 (70-78) [ef]	55 (43-58)[f]	7.0(5.4-21)
CKD-EPI GFR (ml/min/1.73 m^2)	101(97-105)[def]	111(109-116) [acdef]	100(96-103) [def]	82(74-89) [ef]	53 (39-55)[f]	14(5.8-24)	106(102-107)[def]	111(109-118) [acdef]	103(102-106) [def]	78 (73-84) [ef]	53 (43-58)[f]	6.3(5.0-20)
Serum urea (mmol/L)	4.3(1.1)	4.3(0.9)	4.9(0.9) [b]	4.9(1.3)	8.2(2.3)[abcd]	20.4(8.9)[ab cde]	5.1(1.4)	5.0(1.0)	4.4(1.3)	6.7 (1.8) [abc]	7.7 (3.1) [abc]	20.4(7.6)[a bcde]
Serum uric acid (μmol/L)	312(57)	305(81)	340 (89)	324 (75)	416(73) [abcd]	424(89)[abcd]	256(48)	297(107)	336(113)	368(115) [a]	377(121) [a]	429.3(157) [ab]
FBG (mmol/L)	5.1(0.3)	9.9(2.7) [acdef]	7.6(2.2) [af]	8.0(2.7) [af]	7.18(2.6)[af]	4.2(1.2) [a]	4.2(0.5)	10.8(4.3)[ab]	8.3(2.3) [a]	6.7(3.8)	7.8(3.1) [a]	9.8(6.4)[a]
Hemoglobin A1c (%)	-	9.9(2.1)[cf]	7.8(1.5) [b]	9.06 (1.7)[f]	8.3(2.2) [b]	6.6(1.3) [bcde]	5.4(0.3)	9.79(3.2)[ade]	8.1(1.4) [a]	7.0(1.5) [a]	7.71.5) [a]	8.2(2.6) [a]
Serum hemoglobin (g/L)	142(23)	139(20)	141(17)	135(22)	128(19) [ac]	87(20) [abcde]	136(11)	144(14)	143(13)	146(19)	135(16)	109(30)[abc de]
Fasting C-peptide (nmol/L)	-	0.4(0.3)	0.6(0.3) [b]	0.5(0.4)	0.7(0.5)	0.3(0.3)	0.4(0.1)	0.5(0.3)	0.6(0.2) [a]	0.6(0.3)	1.1(0.4) [abcd]	2.0(1.4) [abcde]
UACR (mg/g Cr.)	2.7(2.33-3.59)	27(3.3-59) [a]	12(2.8-34) [a]	103(17-329) [abc]	357(191-1325)[abcd]	2337(1556-3814)[abcde]	2.7(1.8-4.3)	17(8.6-65) [a]	4.5(2.0-39)	42(7.0-159)	73(10-426)	2518(1784-4374) [abcde]
Urinary creatinine (μmol/L)	11793(7909-16647) [cf]	9369(6552-14750)[f]	6900(4795-14550) [f]	8846(5144-9828) [a]	6560(4595-9769) [ab]	4285(3462-5807) [abcde]	8014(3877-13799)	12211(5264-13947)	9478(8607-12632)	10270 (8543-14183)	8498(6812-13085)	8287(5891-10347)
Urinary β2-MG (mg/L)	0.14(0.06-0.20)	0.09(0.01-0.17)[a]	0.12(0.03-0.18)	0.23(0.08-0.60)	0.42(0.20-3.74)	17.58(2.06-27.27) [abcde]	0.09(0.04-0.16)	0.07(0.05-0.21)	0.09 (0.06-0.13)	0.13(0.11-0.34)	1.32(0.42-5.07)	32.79(15.21-72.81) [abcde]
Urinary NAG (ng/mL)	13.1(11.1-15.5)	14.8(12.8-18.3)	13.4(10.5-17.2)	16.5(13.5-18.9)	16.9(10.7-20.2)	14.0(11.2-18.2)	8.0(5.8-10.3)	18.9(11.6-27.1) [ac]	10.9(7.1-15.4)	15.4(12.5-20.0) [ac]	15.3(11.2-18.3) [ac]	18.2(15.5-20.7) [ac]
Urinary microalbumin(mg/L)	-	30(6.3-51)	12 (0.9-23)	82 (4.7-216) [bc]	305(120-944)[bc]	827(99-1680)[bcde]	3.9(0.5-5.0)	20(11-37)	5.0(3.3-64)	41(9.9-296) [a]	163(15-971) [a]	2027(825-3687) [abcde]

Antihypertensive treatment (%)	-	20	16	53	45	90	-	10	40	70	73	73
Antidyslipidemia treatment (%)	-	25	26	30	45	45	-	0	10	10	27	0
Hypertension (%)	-	10	37	50	41	60	-	-	-	-	-	-
BSA-related left renal volume (mL/m^2)	-	-	-	-	-	-	72(64-77)	92(84-125)^{acdef}	79(78-87)	76(62-97)	62(61-78)	55(41-72)^{abcd}
BSA-related right renal volume (mL/m^2)	-	-	-	-	-	-	78(67-87)	96(88-107)^{a}	82(79-108)	80(71-92)	68(62-75)^{bc}	66(47-75)^{abcd}
BSA-related total renal volume (mL/m^2)	-	-	-	-	-	-	150(126-176)	189(177-220)^{adef}	175(156-176)	152(132-173)	138(120-156)	124(95-136)^{abcd}
Renal resistive index	-	-	-	-	-	-	-	0.60(0.07)	0.60(0.06)	0.64(0.05)	0.64(0.05)	0.75(0.11)^{bcde}
Serum cystatin C (mg/L)	-	-	-	-	-	-	0.8(0.7-0.8)	0.7(0.6-0.8)	0.8(0.7-0.9)	1.2 (1.0-1.4)^{abc}	1.8 (1.3-2.1)^{abcd}	6.47(4.3-8.0)^{abcde}
Serum homocysteine (µmol/L)							12.3(10.1-19.0)	12.6(10.1-13.9)	12.4(11.6-15.3)	15.3(11.9-23.9)^{b}	17.9(17.1-24.9)^{abc}	21.3(17.0-26.8)^{abc}

Data are expressed as median (interquartile range), mean (standard deviation) or count (*n*). Significance level for superscript case letters (^{a,b,c,d,e,f} represent comparison with Stages 0, 1a, 1b, 2, 3 and 4, respectively): Student's *t*-tests, $p \leq 0.05$. SBP, systolic blood pressure; DBP, diastolic blood pressure; FBG, fasting blood glucose; GFR, estimated glomerular filtration rate; UACR, urinary albumin to creatinine ratio; β2-MG, urinary β2-microglobulin; NAG, N-acetyl-β-D-glucosaminidase. –, not available. Renal resistive index (RRI) is defined as (peak systolic velocity - end diastolic velocity) / peak systolic velocity derived from the kidney doppler ultrasonography.



Median of BSA-related renal volume (interquartile range, mL/m^2)	Stages (MDRD GFR, Validation set)					
	0	**1a**	**1b**	**2**	**3**	**4**
Left kidney	72(64-77)	92 (84-125) ^{acdef}	79 (78-87)	76 (62-97)	62 (61-78)	55 (41-72) ^{abcd}
Right kidney	78 (67-87)	96(88-107) ^{a}	82 (79-108)	80 (71-92)	68 (62-75) ^{bc}	66 (47-75) ^{abcd}
Left + right kidneys	150 (126-176)	189 (177-220) ^{adef}	175 (156-176)	152 (132-173)	138 (120-156)	124 (95-136) ^{abcd}
Average renal resistive index (SD)	NA	0.60(0.07)	0.60(0.06)	0.64(0.05)	0.64(0.05)	0.75(0.11) ^{bcde}

Fig. 5.2 Images of ultrasonic scanning of the left kidneys in normal subject and diabetic patients in different stages and the comparison of BSA-related renal volumes and renal resistive index between stages.

Significance level for superscript case letters (^{a,b,c,d,e,f} represent comparison with Stages 0, 1a, 1b, 2, 3 and 4, respectively): Student's *t*-test, $p \leq 0.05$.

5.3.1.1 High correlations between different GFRs calculated by different equations

Comparison between routine MDRD GFR and several newly reported GFRs (CKD-EPI $_{creatinine}$ GFR, CKD-EPI $_{cystatin\ C}$ GFR and CKD-EPI $_{creatinine\text{-}cystatin\ C}$ GFR) using creatinine and cystatin C in this study, they had a high similarity, using correlation analysis (Pearson $R > 0.94$). Table 5.4 shows their Pearson correlation coefficients.

Table 5.4 Pearson correlation and Spearman rank correlation of MDRD GFR with three another GFRs that were calculated by the three newly reported equations in all samples of discovery set and/or validation set

Correlation to MDRD GFR	Dataset and size	Pearson correlation
CKD-EPI $_{creatinine}$ GFR	Discovery (n=128)	0.9523
	Validation (n=66)	0.9729
CKD-EPI $_{cystatin\ C}$ GFR	Validation (n=58)	0.9468
CKD-EPI $_{creatinine\text{-}cystatin\ C}$ GFR	Validation (n=58)	0.9681

5.3.1.2 Correlations of clinical parameters to GFR and UACR in diabetic patients

The correlation of clinical parameters to GFR and UACR with absolute Spearman $R > 0.40$ is listed in Table 5.5. In diabetic patients, BSA-related renal volumes of the left, right and the sum of two kidneys were positively associated to MDRD GFR (Spearman $R = 0.5838$, 0.6514, 0.6783, $n = 46$). Serum hemoglobin, urinary microalbumin, β2-microglobulin, urinary total protein, urinary protein-to-creatinine ratio, renal resistive index, serum homocysteine, serum urea, serum creatinine and serum cystatin C were inversely correlated to MDRD GFR (absolute Spearman $R > 0.400$ in Stages 1-4). Urinary protein-to-creatinine ratio, urinary total protein, urinary microalbumin, urinary β2-microglobulin, renal resistive index, serum homocysteine, serum creatinine, serum cystatin C, and serum urea increased with the rise in UACR (Spearman $R > 0.400$) while serum hemoglobin decreased with the increase in UACR (Spearman $R = -0.4743$ and -0.4040).

Table 5.5 Spearman rank correlation of clinical parameters to MDRD GFR and UACR in discovery set and validation set (Absolute Spearman *R* > 0.40).

Clinical parameters	Stages 0-4				Stages 1-4			
	Discovery set (n=128)		Validation set (n=66)		Discovery set (n =108)		Validation set (n =56)	
	Spearman *R*	*p* value	Spearman *R*	*p* value	Correlation	*p* value	Spearman *R*	*p* value
MDRD GFR								
Serum hemoglobin (g/L)	0.4507	1.18E-07	0.3340	0.0061	0.4902	9.61E-08	0.4430	0.0006
Serum uric acid (μmol/L）	-0.4392	2.15E-07	-0.4438	0.0002	-0.4579	6.29E-07	-0.3358	0.0114
Age (years)	-0.4757	1.39E-08	-0.3857	0.0014	-0.4985	4.00E-08	-0.3919	0.0028
Urinary microalbumin(mg/L)	-0.6141	8.56E-12	-0.5841	1.85E-06	-0.6141	8.56E-12	-0.6332	1.78E-06
Urinary *β*2-microglobulin (mg/L)	-0.6285	1.99E-15	-0.7140	2.43E-11	-0.6714	1.83E-15	-0.7571	2.26E-11
Urinary total protein (mg/L)	-0.6851	4.71E-19	-0.6480	5.38E-09	-0.7229	1.02E-18	-0.6752	1.56E-08
Serum urea (mmol/L)	-0.6974	5.93E-20	-0.7014	5.39E-11	-0.7632	8.00E-22	-0.7481	3.44E-11
Urinary protein-to-creatinine ratio (mg/g)	-0.7324	8.95E-23	-0.6469	5.84E-09	-0.7686	2.73E-22	-0.6718	1.97E-08
UACR (mg/g Cr.)	-0.7326	8.62E-23	-0.6634	1.28E-09	-0.7786	3.50E-23	-0.6911	3.73E-09
Serum creatinine (mg/dl)	-0.9308	6.25E-57	-0.9500	4.66E-34	-0.9546	1.44E-57	-0.9592	2.68E-31
BSA-related total renal volume (mL/m^2)			0.5835	2.37E-06			0.6783	2.22E-07
BSA-related right renal volume (mL/m^2)			0.5607	6.98E-06			0.6514	9.44E-07
BSA-related left renal volume (mL/m^2)			0.4747	0.0002			0.5838	2.05E-05
Renal resistive index			-0.5253	0.0001			-0.5253	0.0001
Serum homocysteine (μmol/L)			-0.6371	1.30E-07			-0.6768	2.42E-07
Serum cystatin C (mg/L)			-0.9126 (n = 58)	2.08E-23			-0.9400	4.10E-23 (n = 44)
UACR								
Urinary protein-to-creatinine ratio (mg/g)	0.9459	1.81E-63	0.8780	7.86E-22	0.9660	4.04E-64	0.9157	1.25E-22
Urinary total protein (mg/L)	0.8703	1.42E-40	0.8392	2.57E-18	0.8939	9.89E-39	0.8056	1.20E-13
Urinary microalbumin(mg/L)	0.8390	6.44E-28	0.9767	1.81E-38	0.8390	6.44E-28	0.9735	1.79E-30
Serum creatinine (mg/dl)	0.6880	2.92E-19	0.6464	4.58E-09	0.7367	1.02E-19	0.6668	2.02E-08
Urinary *β*2-microglobulin (mg/L)	0.6215	4.93E-15	0.7000	8.65E-11	0.6473	3.75E-14	0.6938	4.30E-09
Serum urea (mmol/L)	0.5795	7.64E-13	0.4915	2.78E-05	0.6011	6.08E-12	0.5098	5.98E-05
Serum hemoglobin (g/L)	-0.4511	1.16E-07	-0.2896	0.0183	-0.4743	2.80E-07	-0.4040	0.0020
MDRD GFR (ml/min/1.73 m2)	-0.7326	8.62E-23	-0.6634	1.28E-09	-0.7786	3.50E-23	-0.6911	3.73E-09
Serum cystatin C (mg/L)			0.6549	2.44E-08			0.6758	1.36E-07
Renal resistive index			0.4949	0.0003			0.4949	0.0003
Serum homocysteine (μmol/L)			0.4695	0.0003			0.5107	0.0003

5.3.2 Stability and accuracy of the metabolomics platforms

5.3.2.1 Low relative standard deviation among the signals of metabolites and internal standards

The base peak chromatograms of representative pooled QC samples acquired using UPLC-Orbitrap-MS in negative and positive ESI modes from the discovery set and the validation set are shown in Fig. 5.3.

After signal correction and signal quality screening in the discovery set, LC–MS analysis in the negative ESI mode displayed 4316 (82.46 %) out of 5234 variables with coefficient variation (CV %) ≤ 30 %, and 3596 (68.70 %) variables had RSD ≤ 20 %. LC–MS analysis in the positive ESI mode displayed 3164 (67.38 %) out of 4701 variables with CV ≤ 30 %, and 2368 (50.37 %) variables had CV ≤ 20 %. In the validation set, LC–MS analysis in the negative ESI displayed 7249 (87.32 %) out of 8302 variables with CV ≤ 30 %, and 6255 (75.34 %) variables had CV ≤ 20 %. LC–MS analysis in the positive ESI mode displayed 10279 (80.29 %) out of 12802 variables with CV ≤ 30 %, and 8125 (63.47 %) variables had CV ≤ 20 %. The internal standards in QC samples of both the discovery and validation sets had CV % below 5% in positive and negative modes except C19:1n9c which was only detected in the negative mode and its CV% in the discovery and validation sets was 15.27 % and 11.37 %, respectively. Details could be seen in Table 5.6.

PCA score plots (CV ≤ 30% in QC samples) showed that signal correction minimized the variation caused by systemic fluctuations from instrumental response. Study samples and QC samples were clustered together after smoothing of the ion abundance with QC samples (Fig. 5.4).

These illustrated the stabilities of most metabolites and internal standards in the QC samples were generally good across the runs in both data sets. This ensured the observed changes between groups were really due to biological reasons rather than the instrumental bias.

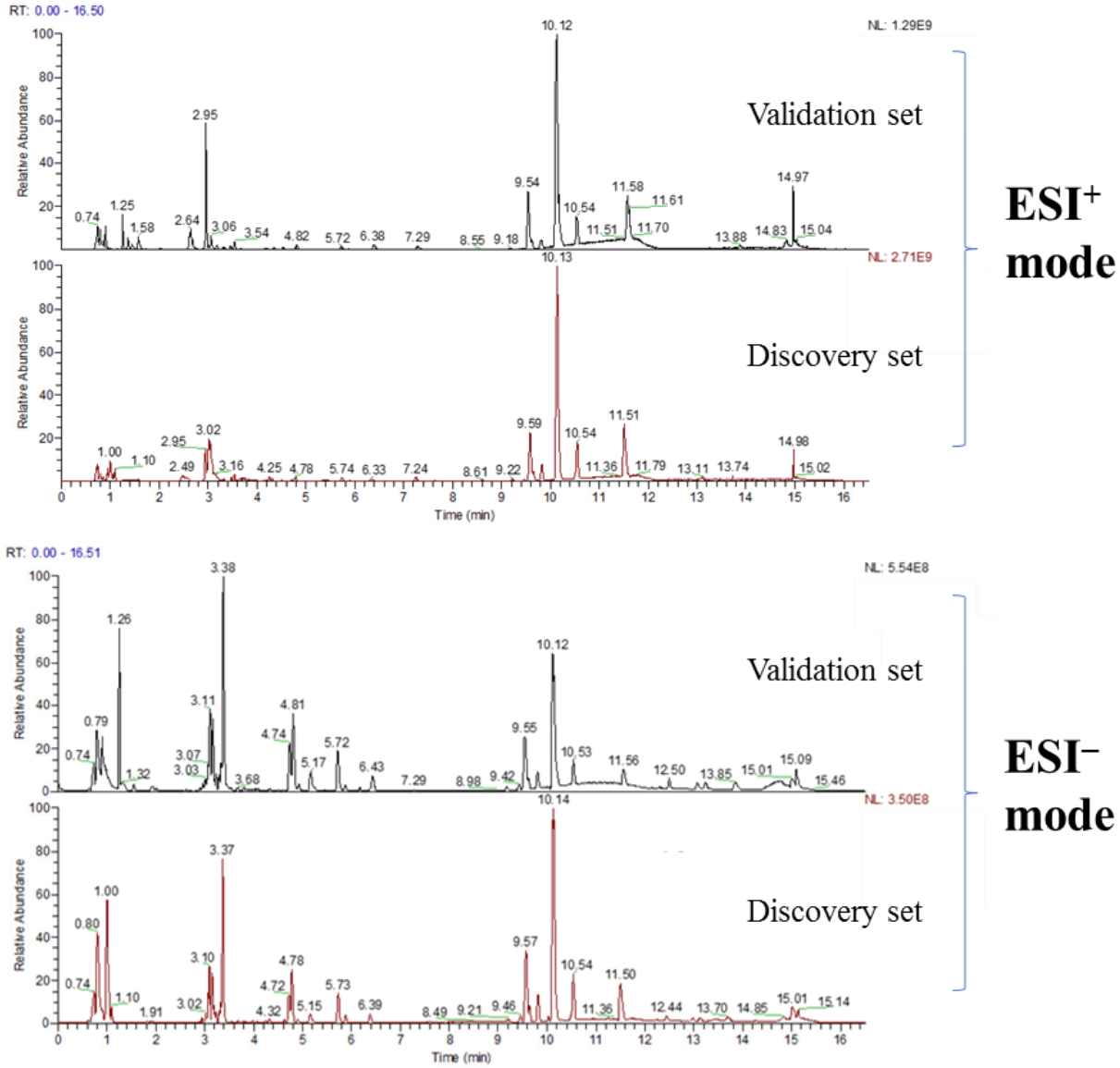


Fig. 5.3 Base peak chromatograms of representative pooled QC samples acquired using UPLC-Orbitrap-MS.

Table 5.6 The stability of relative ion abundance of internal standards in serum of subject samples and QC samples and overall features in QC samples

Internal standards in serum of subject samples and QC samples							
Internal standard	**Dataset**	**Adduct**	**Retention time (min)**	**Detected m/z**	**Theoretical m/z**	**CV (%) of subject samples**	**CV (%) of Quality control samples**
L-Tryptophan-(indole-D_5)	Discovery set	[M-H]-	2.95	208.1137	208.1140	18.04	2.16
	Validation set		2.96	208.1142	208.1140	27.08	1.61
	Discovery set	[M+H]+	2.95	210.1285	210.1285	4.24	2.43
	Validation set		2.96	210.1279	210.1285	12.14	4.28
Cholic acid D_4	Discovery set	[M-H]-	5.73	411.3054	411.3054	3.31	2.74
		[M+H-3H_2O]+	5.74	359.2878	359.2882	4.10	1.99
	Validation set	[M-H]-	5.72	411.3059	411.3054	3.44	1.31
		[M+H-3H_2O]+	5.73	359.2870	359.2882	9.82	2.84
C19:1n9c	Discovery set	[M-H]-	13.68	295.2642	295.2643	29.61	15.28
	Validation set	[M-H]-	13.85	295.2647	295.2643	25.67	11.37

Overall features in QC samples							
Dataset	**ESI mode**	**QC samples (*n*)**	**Total features**	**Number of features with CV ≤ 30% in QC samples**	**Number of features with CV ≤ 20% in QC samples**	**Percentage in all features (CV ≤ 30%)**	**Percentage in all features (CV ≤ 20%)**
Discovery set	-	22	5234	4316	3596	82.46	68.70
	+	25	4701	3164	2368	67.30	50.37
Validation set	-	12	8302	7249	6255	87.32	75.34
	+	12	12802	10279	8125	80.29	63.47

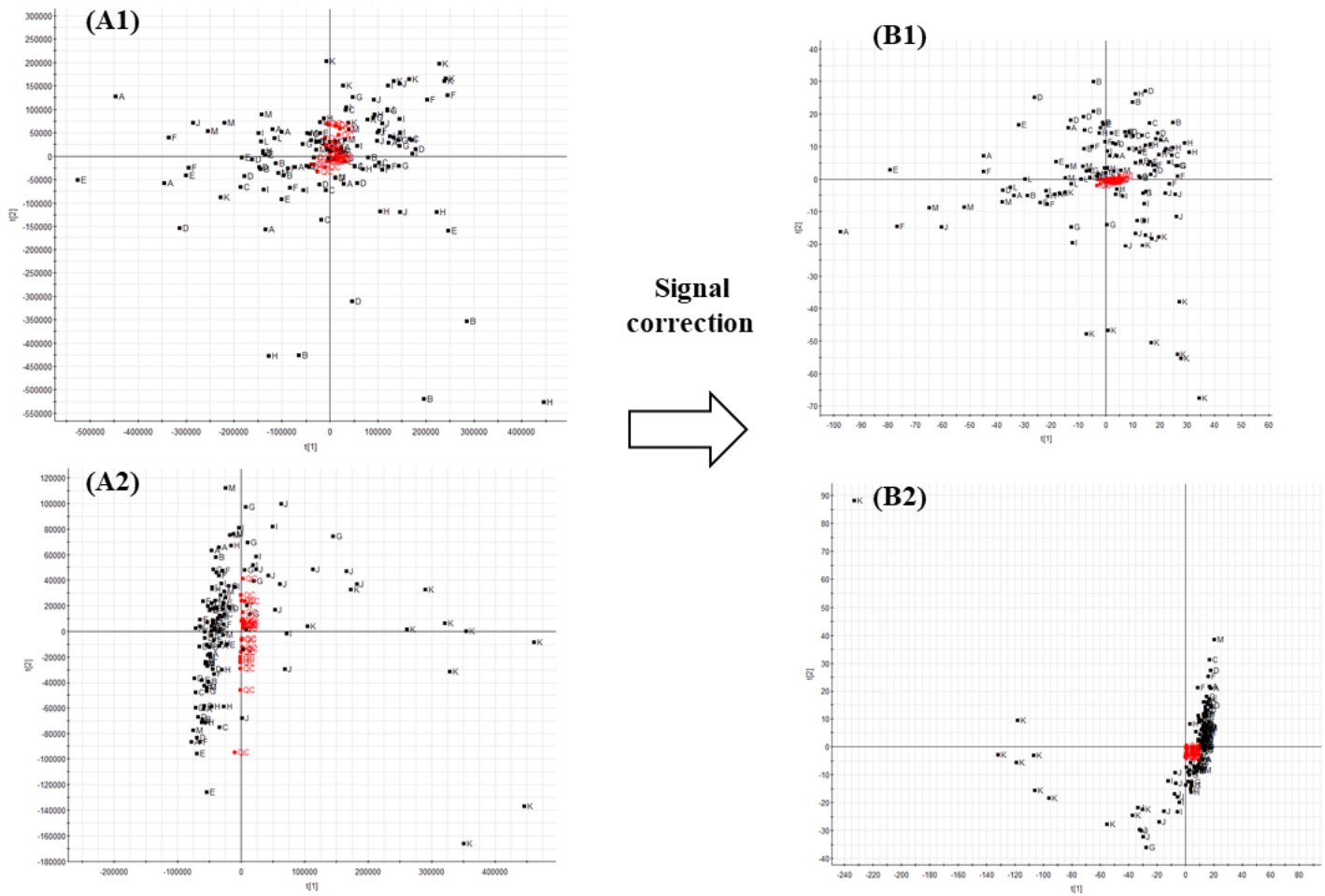


Fig. 5.4 Score plots of PCA of all human study samples (black box) and QC samples (red box) (A) before and (B) after signal correction acquired in (1) ESI+ mode and (2) ESI- mode.

5.3.2.2 High linear relationship between clinical serum parameters and their corresponding values detected by MS

The UPLC MS-detected signal corrected relative intensities, D-glucose (discovery set: R^2 = 0.7727, validation set: R^2 = 0.9414; both R^2 = 0.8043), uric acid (discovery set: R^2 = 0.8424, validation set : R^2 = 0.8487, both: R^2 = 0.8448) and serum creatinine (discovery set: R^2 = 0.9409, validation set : R^2 = 0.9851; both R^2 = 0.9602) showed a very good linear relationship to their corresponding values measured clinically (uric acid, serum creatinine and FBG) (Fig. 5.5). Such high linear relationship demonstrated that the measurements were reliable.

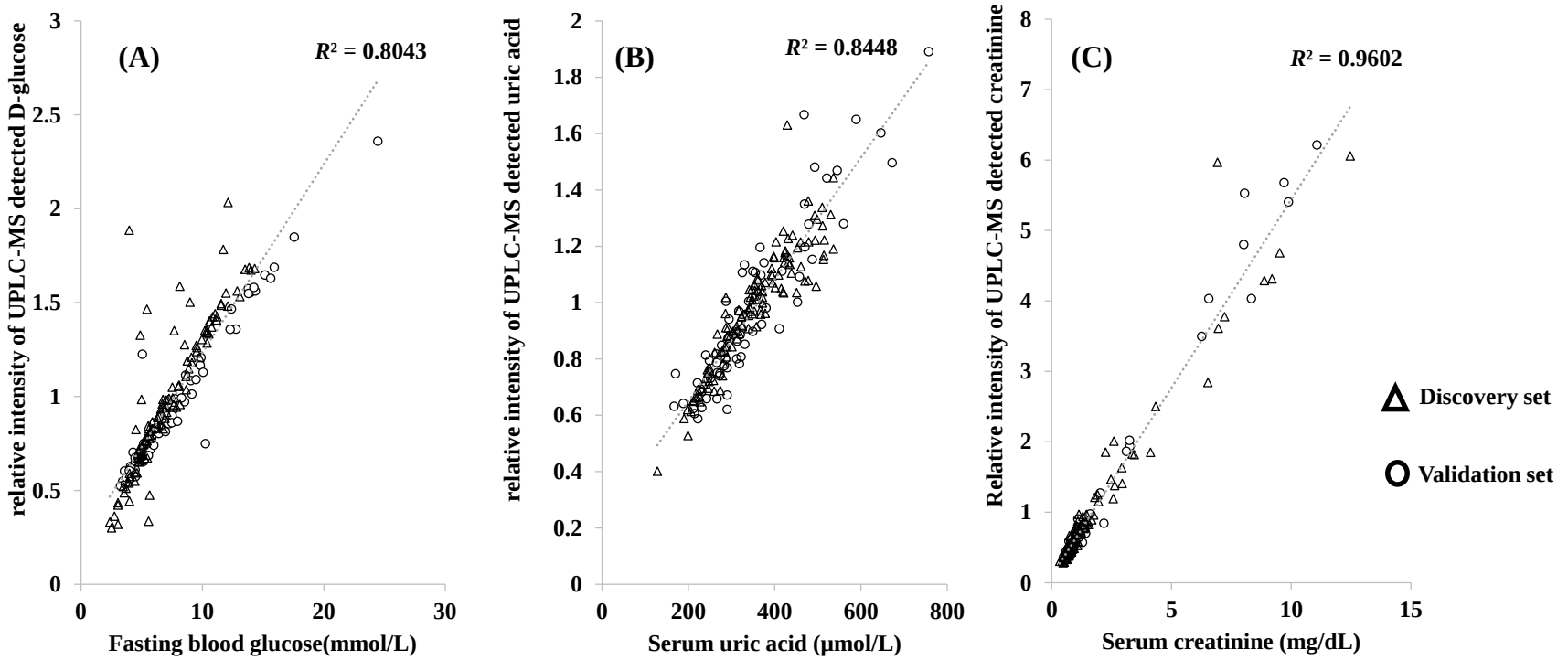


Fig. 5.5 Linear relationship of UPLC-MS detected (A) serum D-glucose, (B) serum uric acid and (C) MS-detected serum creatinine against clinically measured FBG, serum uric acid, serum creatinine, respectively.

Trendlines were formed with both datasets.

5.3.3 Selection and identification of metabolites

After univariate comparison between stages in the discovery set, global metabolomics profiling resulted in detection of 80 candidates (72 metabolites and 8 pairs of metabolite ratios) in both sets of data. The mean CV % of each metabolite in the pooled QC samples was 5.27 % and 4.85% in the discovery set and validation set, respectively. Their identification information and CV % could be found in Table 5.7 and their ID listed in online database are shown in Table 5.8. Fig. 5.6 was an example of the peak identification process, which showed the extracted ion chromatograms and mass fragmentation pattern of succinyladenosine against its reference standard.

Metabolites were ranked according to their fold changes (Table 5.9) with the normal group and their later stages (Student's *t*-test, $p < 0.05$) (Table 5.10) as well as their correlations to MDRD GFR (Table 5.11), clinically measured serum creatinine, UACR, urinary β-microglobulin, cystatin C, FPG, total BSA-related renal volume and renal resistive index (Spearman $R > 0.4$ among diabetic patients) (Table S5.1-S5.6, Appendix II). Those lacking identification were not considered for subsequent analyses. The relationship of metabolites with serum creatinine was similar to that with serum cystatin C. The correlation of metabolites with MDRD GFR was also close to that with CKD-EPI $_{\text{creatinine–cystatin C}}$ GFR and CKD-EPI $_{\text{cystatin C}}$ GFR (Table S5.4)

Annotated metabolites were further checked by multivariate analysis with one-way ANOVA with LSD *post-hoc* test ($p < 0.05$) and false discovery rate (FDR < 0.1 using Metaboanalyst) adjusted for multiple comparisons, 60 candidates (53 metabolites and 7 ratios) had significant changes compared with normal samples or samples in subsequent stages.

Statistical consistency was checked in both cohorts. 55 candidates (49 metabolites and 6 ratios) in the validation set that shared the same trends in fold changes with respect to normal group as those

in the discovery set eventually remained (Table 5.9) and/or between two consecutive stages of CKD (Table 5.10).

Fig. 5.7 is a summary of the selection process and contains two heatmaps of the overview of the metabolite changes among Stages 0 - 4. 38 metabolites were identified with reference standards and 11 were putatively/tentatively identified with online databases or literatures. The heatmaps were drawn by using MetaboAnalyst (https://www.metaboanalyst.ca/) and it also contained clinical markers and clinical indexes for easy comparison.

Table 5.7 Identification details and stability of metabolites in QC samples using UPLC-Orbitrap-MS

Identified metabolites	Retention time (min)	Detected m/z	Theoretical m/z	Mass error (ppm)	Adduct	Molecular formula	Confirmation / Supplier*	CV of QC (%)	
								Discovery set	Validation set
L-Arginine	0.67	173.1042	173.1044	0.6	$[M-H]^-$	$C_6H_{14}N_4O_2$	Sigma Aldrich	10.17	8.58
L-Ornithine	0.67	131.0823	131.0826	0	$[M-H]^-$	$C_5H_{12}N_2O_2$	Sigma Aldrich	11.08	5.23
Choline	0.78	104.1071	104.1070	-2.9	$[M]^+$	$C_5H_{14}NO$	Acros Organics	4.67	5.48
L-Glutamine	0.79	145.0616	145.0619	0.0	$[M-H]^-$	$C_5H_{10}N_2O_3$	Sigma Aldrich	1.08	1.25
L-Citrulline	0.80	174.0882	174.0884	0.6	$[M-H]^-$	$C_6H_{13}N_3O_3$	Sigma Aldrich	2.26	3.29
D-Glucose	0.80	215.0328	215.0328	1.4	$[M+Cl]^-$	$C_6H_{12}O_6$	International laboratory	1.14	0.68
L-Carnitine	0.80	162.1125	162.1125	-3.1	$[M+H]^+$	$C_7H_{15}NO_3$	CIL	1.50	3.63
L-Glutamic acid	0.80	148.0604	148.0604	-2.7	$[M+H]^+$	$C_5H_9NO_4$	Sigma Aldrich	3.80	4.09
L-Threonine	0.80	120.0655	120.0655	-2.5	$[M+H]^+$	$C_4H_9NO_3$	Sigma Aldrich	6.47	5.57
Arabinose isomer	0.80	195.0513	195.0510	1.5	[M+FA-H]-	$C_5H_{10}O_5$	Santa Cruz	1.46	1.88
Betaine	0.81	118.0863	118.0863	-3.4	$[M+H]^+$	$C_5H_{11}NO_2$	Sigma Aldrich	1.70	3.06
Creatinine	0.81	114.0662	114.0662	-2.6	$[M+H]^+$	$C_4H_7N_3O$	Acros Organics	2.24	3.93
2-Hydroxyethanesulfonate	0.82	124.9911	124.9914	0.0	$[M-H]^-$	$C_2H_6O_4S$	Sigma Aldrich	3.78	8.46
***γ*-Butyrobetaine**	0.83	146.1175	146.1176	-3.4	$[M+H]^+$	$C_7H_{15}NO_2$	TRC	3.76	2.37
L-Proline	0.83	116.0707	116.0706	-2.6	$[M+H]^+$	$C_5H_9NO_2$	Sigma Aldrich	1.56	3.69
1,5-Anhydro-D-glucitol	0.85	199.0376	199.0379	1.0	[M+Cl]-	$C_6H_{12}O_5$	TRC	2.09	1.82
N-Acetylcarnosine	0.89	269.1243	269.1243	-2.6	$[M+H]^+$	$C_{11}H_{16}N_4O_4$	Santa Cruz	5.91	7.34
5-Methylthio-D-ribose	0.93	181.0529	181.0529	-2.8	$[M+H]^+$	$C_6H_{12}O_4S$	Online database (HMDB)	6.12	8.15
Pseudouridine	0.93	243.0622	243.0623	1.2	$[M-H]^-$	$C_9H_{12}N_2O_6$	Supelco	1.78	3.65
L-Valine	0.94	118.0863	118.0863	-3.4	$[M+H]^+$	$C_5H_{11}NO_2$	Sigma Aldrich	1.89	4.29
L-Acetylcarnitine	0.95	204.1230	204.1230	-2.9	$[M+H]^+$	$C_9H_{17}NO_4$	CIL	2.35	6.67
L,L-TMAP isomer	0.96	229.1546	229.1547	-3.5	$[M+H]^+$	$C_{11}H_{20}N_2O_3$	ChemPartner and (Zhang, Ford, et al. 2017)	4.05	6.07
Uric acid	1.00	169.0356	169.0356	-3.0	$[M+H]^+$	$C_5H_4N_4O_3$	Sigma Aldrich	1.78	4.69
L,L-TMAP	1.06	229.1546	229.1547	-3.5	$[M+H]^+$	$C_{11}H_{20}N_2O_3$	ChemPartner and (Zhang, Ford, et al. 2017)	2.31	5.44
L-Methionine	1.07	148.0435	148.0438	1.8	[M-H]-	$C_5H_{11}NO_2S$	Sigma Aldrich	5.03	7.31
Citric acid	1.09	191.0195	191.0197	1.0	$[M-H]^-$	$C_6H_8O_7$	Sigma Aldrich	3.55	3.64
Hydroxybutyrylcarnitine	1.09	248.1491	248.1492	-3.6	$[M+H]^+$	$C_{11}H_{21}NO_5$	Online databases	11.84	8.58
Succinylcarnitine	1.09	262.1284	262.1285	-3.8	$[M+H]^+$	$C_{11}H_{19}NO_6$	Supelco	5.29	11.02
Uracil	1.09	113.0346	113.0346	-2.7	$[M+H]^+$	$C_4H_4N_2O_2$	Wako	5.11	6.35
Uridine	1.09	243.0622	243.0623	1.2	$[M-H]^-$	$C_9H_{12}N_2O_6$	Wako	1.69	1.14
L-Tyrosine	1.33	180.0664	180.0666	0.6	$[M-H]^-$	$C_9H_{11}NO_3$	Sigma Aldrich	2.11	1.06
Sulfotyrosine	1.46	260.0234	260.0234	1.9	$[M-H]^-$	$C_9H_{11}NO_6S$	(Turyan, Frenkel, and Sosic 2018)	1.79	0.88
Inosine	1.50	267.0735	267.0735	1.9	$[M-H]^-$	$C_{10}H_{12}N_4O_5$	Acros Organics	2.30	1.17
L-Leucine	1.56	132.1019	132.1019	-3.0	$[M+H]^+$	$C_6H_{13}NO_2$	Sigma Aldrich	1.81	6.27
4-Acetamidobutanoic acid	1.59	144.0664	144.0666	0.7	$[M-H]^-$	$C_6H_{11}NO_3$	Matrix Scientific	1.62	1.44
Propionylcarnitine	1.67	218.1387	218.1387	-3.2	$[M+H]^+$	$C_{10}H_{19}NO_4$	CIL	2.61	6.39
2-Hydroxybutyric acid	1.90	103.0398	103.0401	-1.0	$[M-H]^-$	$C_4H_8O_3$	Sigma Aldrich	1.85	1.35
2-(*α*-D-Mannopyranosyl)-L-tryptophan	2.21	367.1497	367.1500	-3.8	$[M+H]^+$	$C_{17}H_{22}N_2O_7$	TRC	2.16	6.02
L-Kynurenine	2.42	209.0921	209.0921	-3.3	$[M+H]^+$	$C_{10}H_{12}N_2O_3$	Sigma Aldrich	3.00	8.09
L-Phenylalanine	2.48	164.0714	164.0717	0.6	$[M-H]^-$	$C_9H_{11}NO_2$	Sigma Aldrich	2.02	2.36
Succinyladenosine	2.89	382.1005	382.1004	1.6	$[M-H]^-$	$C_{14}H_{17}N_5O_8$	TRC	3.06	3.07
***O*-Adipoylcarnitine**	2.90	290.1597	290.1598	-3.8	$[M+H]^+$	$C_{13}H_{23}NO_6$	Supelco	3.39	5.11
Butyrylcarnitine	2.92	232.1543	232.1543	-3.4	$[M+H]^+$	$C_{11}H_{21}NO_4$	CIL	2.62	5.65
L-*β*-aspartyl-L-leucine	2.93	247.1287	247.1288	-3.2	$[M+H]^+$	$C_{10}H_{18}N_2O_5$	Online database (HMDB)	6.71	4.89
L-Tryptophan	2.95	203.0824	203.0826	1.0	$[M-H]^-$	$C_{11}H_{12}N_2O_2$	Sigma Aldrich	2.21	1.26
Homovanillic acid sulfate	2.97	261.0073	261.0074	1.9	$[M-H]^-$	$C_9H_{10}O_7S$	Cayman Chemical	12.40	2.42

Kynurenic acid	2.99	190.0499	190.0499	-3.2	$[M+H]^+$	C10H7NO3	Sigma Aldrich	14.71	4.95
2-[3-(sulfooxy)phenyl]acetic acid	3.00	230.9967	230.9969	1.3	$[M-H]^-$	C8H8O6S	Online database (HMDB)	2.64	4.81
Valerylcarnitine	3.00	246.1699	246.1700	-3.7	$[M+H]^+$	C12H23NO4	Cayman Chemical	5.96	7.49
Pyrocatechol sulfate	3.07	188.9865	188.9863	1.1	[M-H]-	C6H6O5S	Online database (HMDB)	2.82	1.57
***α*-*N*-Phenylacetyl-L-glutamine**	3.07	263.1037	263.1037	1.9	$[M-H]^-$	C13H16N2O4	Santa Cruz	1.87	1.41
Phenol sulfate	3.10	172.9912	172.9914	1.2	$[M-H]^-$	C6H6O4S	Online database (HMDB)	2.32	0.99
Hexanoylcarnitine	3.11	260.1855	260.1856	-3.5	$[M+H]^+$	C13H25NO4	Santa Cruz	28.21	8.88
Hippuric acid	3.14	178.0508	178.0510	0.6	$[M-H]^-$	C9H9NO3	Acros Organics	2.03	0.91
Indoxyl sulfate	3.15	212.0022	212.0023	1.4	$[M-H]^-$	C8H7NO4S	Sigma Aldrich	2.82	1.43
***p*-Cresol glucuronide**	3.17	283.0823	283.0823	1.8	$[M-H]^-$	C13H16O7	TRC	2.09	1.42
2-Octenoylcarnitine	3.32	286.2011	286.2013	-3.8	$[M+H]^+$	C15H27NO4	Online databases	3.15	6.83
***p*-Cresol sulfate**	3.36	187.0070	187.0071	1.1	$[M-H]^-$	C7H8O4S	CIL	1.82	1.00
Indole-3-lactic acid	3.44	204.0664	204.0666	1.0	$[M-H]^-$	C11H11NO3	Santa Cruz	2.25	1.28
3-Hydroxydecanoyl carnitine	3.62	332.2429	332.2431	-3.6	$[M+H]^+$	C17H33NO5	Online databases	5.28	5.99
L-Octanoylcarnitine	3.65	288.2167	288.2169	-3.5	$[M+H]^+$	C15H29NO4	CIL	2.63	5.72
3-Indoleacetic acid	3.78	176.0706	176.0706	-2.8	$[M+H]^+$	C10H9NO2	Sigma Aldrich	3.54	5.14
Cortisol	4.03	363.2163	363.2166	-3.6	$[M+H]^+$	C21H30O5	Sigma Aldrich	3.00	3.57
9-Decenoylcarnitine	4.13	314.2324	314.2326	-3.8	$[M+H]^+$	C17H31NO4	Online databases	2.18	6.51
Bilirubin	4.30	585.2706	585.2708	-2.7	$[M+H]^+$	C33H36N4O6	Acros Organics	4.95	5.13
Decanoylcarnitine	4.51	316.2480	316.2482	-3.8	$[M+H]^+$	C17H33NO4	Sigma Aldrich	2.80	2.82
Dehydroepiandrosterone sulfate	4.73	367.1584	367.1585	1.1	$[M-H]^-$	C19H28O5S	Cayman Chemical	3.63	1.64
3,5-Tetradecadiencarnitine	5.49	368.2793	368.2795	-3.5	$[M+H]^+$	C21H37NO4	Online databases	5.88	6.73
Dodecanoylcarnitine	5.57	344.2793	344.2795	-3.5	$[M+H]^+$	C19H37NO4	CIL	9.44	4.34
***cis*-5-Tetradecenoylcarnitine**	6.39	370.2949	370.2952	-3.5	$[M+H]^+$	C21H39NO4	Online databases	12.23	9.59
LysoPE(18:1(11Z)/0:0)	10.46	480.3083	480.3085	-2.9	$[M+H]^+$	C23H46NO7P	Online databases	14.54	9.97
Stearoylcarnitine	10.52	428.3732	428.3734	-3.3	$[M+H]^+$	C25H49NO4	CIL	16.91	8.10

*Acros Organics, New Jersey, US. Cayman Chemical, Michigan, US. CIL, Cambridge Isotope Laboratories, MA, US. ChemPartner, Shanghai ChemPartner Co., Ltd., China. International laboratory, CA, USA. Sgima Aldrich and Supelco, MO, US. Matrix Scientific, SC, US. Santa Cruz, Texas, US. TRC, Toronto Research Chemicals, Canada. Wako, Wako Pure Chemical Industries, Japan. N, N, N-trimethyl-L-alanyl-Lproline betaine (L,L-TMAP).

Table 5.8 List of ID in online database for networking

Metabolites	PubChem ID	HMDB ID	KEGG ID	SMILES code
1,5-Anhydro-D-glucitol	64960	HMDB0002712	C07326	C1[C@@H]([C@H]([C@@H]([C@H](O1)CO)O)O)O
2-[3-(sulfooxy)phenyl] acetic acid	131831698	HMDB0125163	N/A	C1=CC(=CC(=C1)OS(=O)(=O)O)CC(=O)O
2-(*α*-D-Mannopyranosyl)-L-tryptophan	10981970	HMDB0240296	N/A	N[C@@H](CC1=C(NC2=C1C=CC=C2)[C@H]1O[C@H](CO)[C@@H](O)[C@H](O)[C@@H]1O)C(O)=O
2-Hydroxybutyric acid	11266	HMDB0000008	C05984	CCC(C(=O)O)O
2-Hydroxyethanesulfonate	7866	HMDB0003903	C05123	OCCS(O)(=O)=O
2-Octenoylcarnitine	53481667	HMDB0013324	N/A	CCCCC/C=C/C(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
3,5-Tetradecadiencarnitine	53481681	HMDB0013331	N/A	CCCCCCCC/C=C/C=C/CC(=O)O[C@@H](CCC(=O)[O-])[N+](C)(C)C
3-Hydroxydecanoyl carnitine	121454166	HMDB0061636	N/A	CCCCCCCC(CC(=O)OC(CC(=O)[O-])C[N+](C)(C)C)O
3-Indoleacetic acid	802	HMDB0000197	C00954	C1=CC=C2C(=C1)C(=CN2)CC(=O)O
4-Acetamidobutanoic acid	18189	HMDB0003681	C02946	CC(=O)NCCCC(=O)O
5-Methylthio-D-ribose	439904	HMDB0001087	C03089	CSC[C@@H]1[C@H]([C@H]([C@@H](O1)O)O)O
9-Decenoylcarnitine	53481651	HMDB0013205	N/A	C[N+](C)(C)CC(CC(=O)[O-])OC(=O)CCCCCCCC=C
Betaine	247	HMDB0000043	C00719	C[N+](C)(C)CC(=O)[O-]
Bilirubin	5280352	HMDB0000054	C00486	CC1=C(C=C)\C(NC1=O)=C\C1=C(C)C(CCC(O)=O)=C(CC2=C(CCC(O)=O)C(C)=C(N2)\C=C2/NC(=O)C(C=C)=C2C)N1
Butyrylcarnitine	213144	HMDB0002013	C02862	CCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
Choline	305	HMDB0000097	C00114	C[N+](C)(C)CCO
***cis*-5-Tetradecenoylcarnitine**	22833575	HMDB0002014	N/A	CCCCCCCC/C=C\CCCC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
Citric acid	311	HMDB0000094	C00158	OC(=O)CC(O)(CC(O)=O)C(O)=O
Cortisol	5754	HMDB0000063	C00735	C[C@]12CCC(=O)C=C1CC[C@@H]3[C@@H]2[C@H](C[C@]4([C@H]3CC[C@@]4(C(=O)CO)O)C)O
Creatinine	588	HMDB0000562	C00791	CN1CC(=O)N=C1N
Decanoylcarnitine	10245190	HMDB0000651	N/A	CCCCCCCCCC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
Dehydroepiandrosterone sulfate	12594	HMDB0001032	C04555	C[C@]12CC[C@H]3[C@H]([C@@H]1CCC2=O)CC=C4[C@@]3(CC[C@@H](C4)OS(=O)(=O)O)C
D-Glucose	5793	HMDB0000122	C00031	C([C@@H]1[C@H]([C@@H]([C@H](C(O1)O)O)O)O)O
Dodecanoylcarnitine	168381	HMDB0002250	N/A	CCCCCCCCCCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
Hexanoylcarnitine	6426853	HMDB0000705	N/A	CCCCCC(=O)OC(CC([O-])=O)C[N+](C)(C)C
Hippuric acid	464	HMDB0000714	C01586	C1=CC=C(C=C1)C(=O)NCC(=O)O
Homovanillic acid sulfate	29981063	HMDB0011719	N/A	COC1=C(C=CC(=C1)CC(=O)O)OS(=O)(=O)O
Hydroxybutyrylcarnitine	53481617	HMDB0013127	N/A	CC(CC(=O)O[C@@H](CC(=O)[O-])C[N+](C)(C)C)O
Indole-3-lactic acid	92904	HMDB0000671	C02043	C1=CC=C2C(=C1)C(=CN2)CC(C(=O)O)O
Indoxyl sulfate	10258	HMDB0000682	N/A	C1=CC=C2C(=C1)C(=CN2)OS(=O)(=O)O
Inosine	6021	HMDB0000195	C00294	C1=NC(=O)C2=C(N1)N(C=N2)[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O
Kynurenic acid	3845	HMDB0000715	C01717	OC(=O)C1=CC(=O)C2=CC=CC=C2N1
L,L-TMAP	134218393	HMDB0240365	N/A	C[C@@H](C(=O)N1CCC[C@H]1C([O-])=O)[N+](C)(C)C

L-Acetylcarnitine	7045767	HMDB0000201	C02571	CC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
L-Arginine	6322	HMDB0000517	C00062	C(C[C@@H](C(=O)O)N)CN=C(N)N
L-Carnitine	10917	HMDB0000062	C00318	C[N+](C)(C)C[C@@H](CC(=O)[O-])O
L-Citrulline	9750	HMDB0000904	C00327	C(C[C@@H](C(=O)O)N)CNC(=O)N
L-Glutamic acid	33032	HMDB0000148	C00025	C(CC(=O)O)[C@@H](C(=O)O)N
L-Glutamine	5961	HMDB0000641	C00064	C(CC(=O)N)[C@@H](C(=O)O)N
L-Homocysteine	91552	HMDB0000742	C00155	N[C@@H](CCS)C(O)=O
L-Kynurenine	161166	HMDB0000684	C00328	N[C@@H](CC(=O)C1=CC=CC=C1N)C(O)=O
L-Leucine	6106	HMDB0000687	C00123	CC(C)C[C@H](N)C(O)=O
L-Methionine	6137	HMDB0000696	C00073	CSCC[C@@H](C(=O)O)N
L-Octanoylcarnitine	11953814	HMDB0000791	C02838	CCCCCCCC(=O)O[C@H](CC(=O)[O-])C[N+](C)(C)C
L-Ornithine	6262	HMDB0000214	C00077	C(C[C@@H](C(=O)O)N)CN
L-Phenylalanine	6140	HMDB0000159	C00079	C1=CC=C(C=C1)C[C@@H](C(=O)O)N
L-Proline	145742	HMDB0000162	C00148	C1C[C@H](NC1)C(=O)O
L-Threonine	6288	HMDB0000167	C00188	C[C@H]([C@@H](C(=O)O)N)O
L-Tryptophan	6305	HMDB0000929	C00078	C1=CC=C2C(=C1)C(=CN2)C[C@@H](C(=O)O)N
L-Tyrosine	6057	HMDB0000158	C00082	C1=CC(=CC=C1C[C@@H](C(=O)O)N)O
L-Valine	6287	HMDB0000883	C00183	CC(C)[C@@H](C(=O)O)N
LysoPE(18:1(11Z)/0:0)	53480949	HMDB0011505	N/A	[H][C@@](O)(COC(=O)CCCCCCCCC\C=C/CCCCCC)COP(O)(=O)OCCN
L-*β*-aspartyl-L-leucine	3549397	HMDB0011166	N/A	CC(C)CC(NC(=O)CC(N)C(O)=O)C(O)=O
***N*-Acetylcarnosine**	10221026	HMDB0012881	N/A	CC(=O)NCCC(=O)N[C@H](CC1=CN=CN1)C(=O)O
***O*-Adipoylcarnitine**	71296139	HMDB0061677	N/A	C[N+](C)(C)C[C@@H](CC([O-])=O)OC(=O)CCCCC(O)=O
***p*-Cresol glucuronide**	154035	HMDB0011686	N/A	CC1=CC=C(C=C1)O[C@H]2[C@@H]([C@H]([C@@H]([C@H](O2)C(=O)O)O)O)O
***p*-Cresol sulfate**	4615423	HMDB0011635	N/A	CC1=CC=C(C=C1)OS(=O)(=O)O
Phenol sulfate	74426	HMDB0060015	C02180	C1=CC=C(C=C1)OS(=O)(=O)O
Propionylcarnitine	107738	HMDB0000824	C03017	CCC(=O)OC(CC(=O)[O-])C[N+](C)(C)C
Pseudouridine	15047	HMDB0000767	C02067	C1=C(C(=O)NC(=O)N1)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O
Pyrocatechol sulfate	3083879	HMDB0059724	N/A	OC1=C(OS(O)(=O)=O)C=CC=C1
Stearoylcarnitine	3006797	HMDB0000848	N/A	CCCCCCCCCCCCCCCCCC(=O)O[C@H](CC(=O)O)C[N+](C)(C)C
Succinyladenosine	20849086	HMDB0000912	N/A	C1=NC2=C(C(=N1)N[C@@H](CC(=O)O)C(=O)O)N=CN2[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O
Succinylcarnitine	71464481	HMDB0061717	N/A	C[N+](C)(C)C[C@@H](CC([O-])=O)OC(=O)CCC(O)=O
Sulfotyrosine	514186	N/A	N/A	C1=CC(=CC=C1C[C@@H](C(=O)O)N)OS(=O)(=O)O
Uracil	1174	HMDB0000300	N/A	O=C1NC=CC(=O)N1
Urea	1176	HMDB0000294	C00086	C(=O)(N)N
Uric acid	1175	HMDB0000289	C00366	C12=C(NC(=O)N1)NC(=O)NC2=O
Uridine	6029	HMDB0000296	C00299	C1=CN(C(=O)NC1=O)[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O
Valerylcarnitine	53481619	HMDB0013128	N/A	CCCCC(=O)O[C@@H](CC(=O)[O-])C[N+](C)(C)C
***α*-*N*-Phenylacetyl-L-glutamine**	92258	HMDB0006344	C04148	C1=CC=C(C=C1)CC(=O)N[C@@H](CCC(=O)N)C(=O)O
***γ*-Butyrobetaine**	725	HMDB0006831	C01181	C[N+](C)(C)CCCC(=O)[O-]

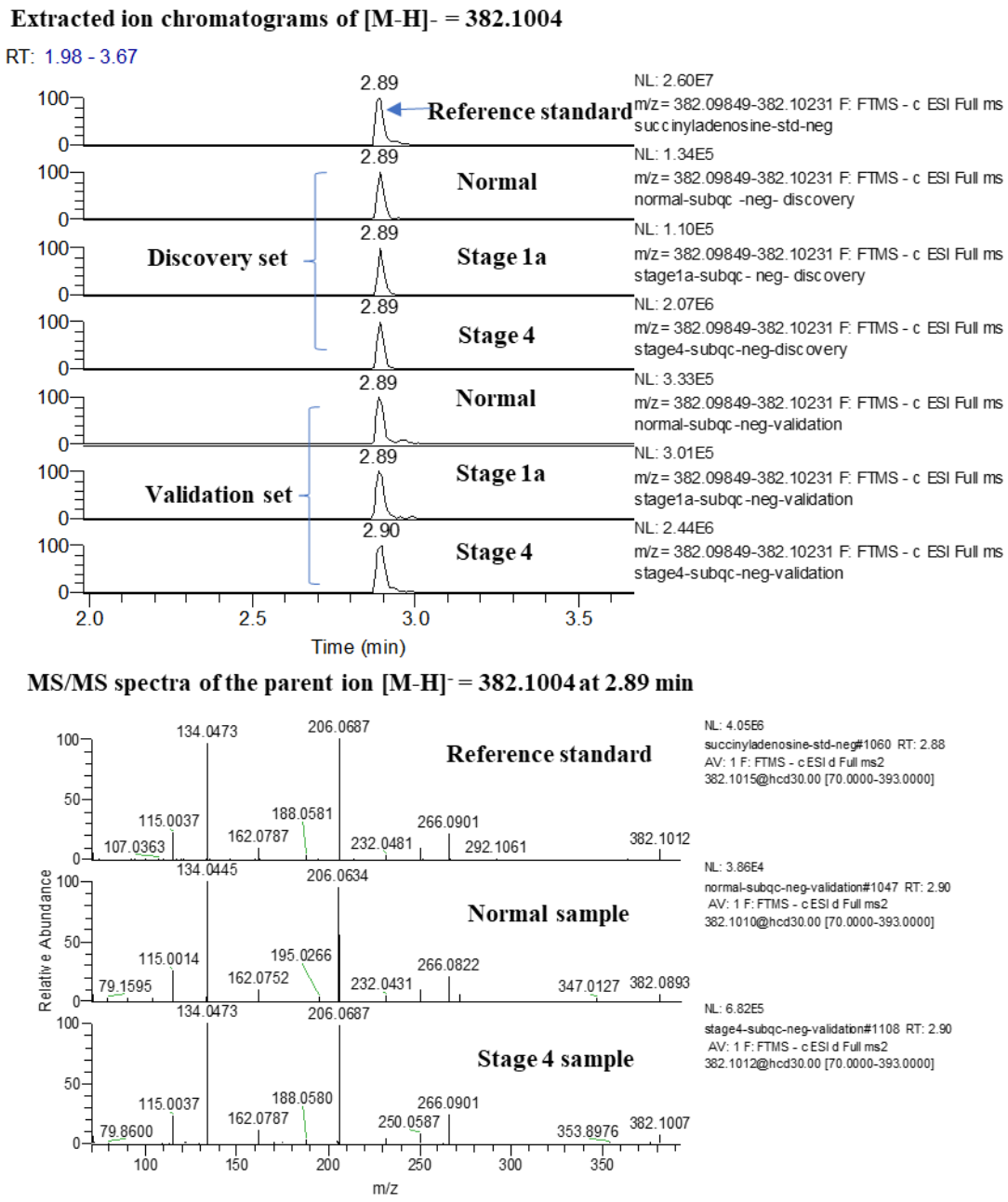


Fig. 5.6 Demonstration of peak identification of the adduct [M-H]- of succinyladenosine compared with its reference standard.

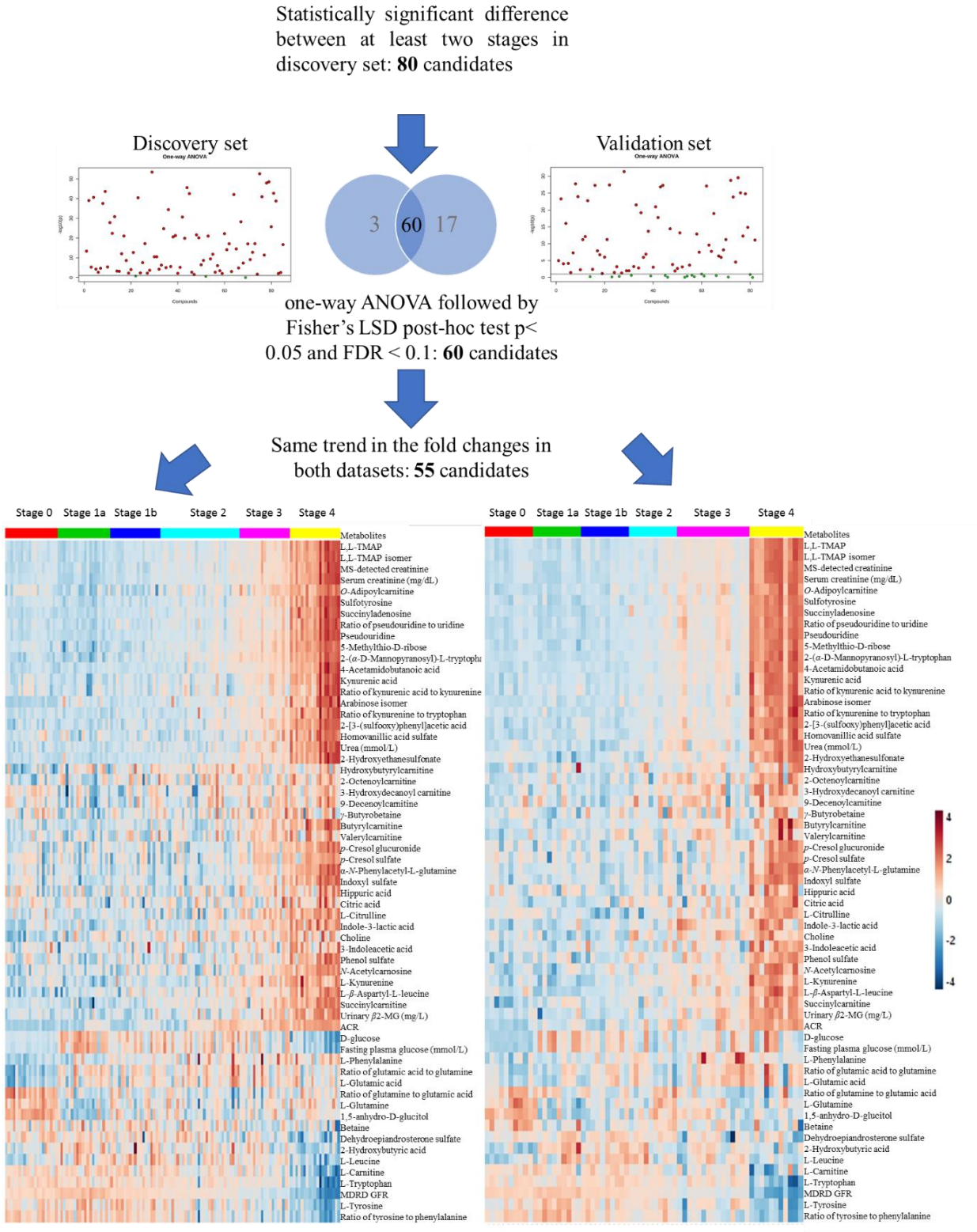


Fig. 5.7 Validation process of potential candidates in all stages of CKD in discovery set using validation set

Table 5.9 Metabolites that had significant fold changes in different stages of MDRD GFR with respect to the normal groups in both discovery and validation sets.

	Fold change		*P* value	
Stage 1a vs normal control	**Discovery set**	**Validation set**	**Discovery set**	**Validation set**
Hydroxybutyrylcarnitine	4.66	2.33	0.0095	0.0067
D-Glucose	1.82	2.19	1.63E-08	0.0001
Stearoylcarnitine	1.44	1.66	0.0064	0.0099
2-Hydroxybutyric acid	1.38	2.20	0.0014	3.72E-07
Cortisol	1.35	1.79	0.0189	0.0003
L-Leucine	1.26	1.37	0.0039	0.0291
***γ*-Butyrobetaine**	-1.28	-1.20	0.0015	0.0403
L-Glutamine	-1.56	-1.35	1.06E-10	0.0025
1,5-Anhydro-D-glucitol	-26.5	-17.75	1.68E-11	0.0005
Stage 1b vs normal control				
Hydroxybutyrylcarnitine	3.91	1.62	0.0015	0.0032
Ratio of glutamic acid to glutamine	2.28	1.43	7.80E-07	0.0294
L-Leucine	1.39	1.54	2.11E-05	0.0002
D-Glucose	1.32	1.65	2.73E-05	0.0001
2-Hydroxybutyric acid	1.17	1.87	0.0371	5.83E-06
L-Glutamine	-1.25	-1.19	0.0001	0.0031
Ratio of glutamine to glutamic acid	-2.28	-1.43	7.80E-07	0.0294
1,5-Anhydro-D-glucitol	-7.88	-7.18	4.93E-06	8.40E-06
Stage 2 vs normal control				
Hydroxybutyrylcarnitine	5.29	1.78	4.46E-05	0.0028
Arabinose isomer	2.40	1.70	8.83E-11	0.0006
2-(*α*-D-Mannopyranosyl)-L-tryptophan	2.08	1.43	1.46E-07	0.0001
2-[3-(sulfooxy)phenyl]acetic acid	1.86	2.39	0.0047	0.0011
4-Acetamidobutanoic acid	1.61	1.58	0.00010	0.0010
Succinyladenosine	1.55	1.35	2.29E-07	0.0107
D-Glucose	1.41	1.24	9.69E-07	0.0325
LysoPE(18:1(11Z)/0:0)	1.38	1.43	0.0004	0.0116
L,L-TMAP isomer	1.36	1.52	2.42E-05	0.0033
Succinylcarnitine	1.35	1.73	4.48E-05	0.0118
Creatinine	1.33	1.33	2.59E-06	0.0161
5-Methylthio-D-ribose	1.32	1.66	0.0010	0.0019
L-Leucine	1.28	1.37	0.0005	0.0053
L,L-TMAP	1.27	1.40	0.0003	0.0029
Pseudouridine	1.23	1.52	3.02E-07	0.0001
Ratio of pseudouridine to uridine	1.18	1.38	0.001	0.0063
Sulfotyrosine	1.15	1.28	0.0206	0.0189
Ratio of kynurenine to tryptophan	1.15	1.36	0.0214	0.0081
1,5-Anhydro-D-glucitol	-8.97	-2.26	1.25E-11	0.0197
Stage 3 vs normal control				
2-[3-(sulfooxy)phenyl]acetic acid	7.86	6.15	9.99E-10	1.71E-05
Hydroxybutyrylcarnitine	5.79	2.50	0.0002	1.43E-06
2-(*α*-D-Mannopyranosyl)-L-tryptophan	4.23	1.83	4.46E-14	1.96E-06
***O*-Adipoylcarnitine**	4.22	5.53	2.39E-06	9.12E-07

Arabinose isomer	3.8	2.45	7.10E-12	2.54E-06
N-Acetylcarnosine	3.11	1.93	0.0006	0.0039
4-Acetamidobutanoic acid	2.82	2.32	6.23E-14	3.99E-07
Kynurenic acid	2.60	3.05	2.83E-07	0.0001
5-Methylthio-D-ribose	2.48	2.22	2.97E-10	1.40E-06
Ratio of kynurenine to tryptophan	2.23	1.62	8.39E-11	0.0001
Succinyladenosine	2.20	1.95	1.85E-12	4.07E-07
Ratio of glutamic acid to glutamine	2.18	1.97	4.77E-07	0.0005
Homovanillic acid sulfate	2.09	3.72	0.0055	1.52E-06
L,L-TMAP	2.01	1.91	7.94E-10	1.25E-06
Ratio of kynurenic acid to kynurenine	1.97	1.91	0.0026	0.0029
2-Octenoylcarnitine	1.95	2.88	0.0019	0.0051
Sulfotyrosine	1.93	2.00	1.76E-09	7.44E-06
Urea (mmol/L)	1.91	1.61	5.01E-09	0.0124
Creatinine	1.88	1.78	4.34E-10	1.28E-05
L,L-TMAP isomer	1.88	2.09	5.13E-09	1.17E-07
Indole-3-lactic acid	1.87	1.63	8.09E-07	0.0002
Pseudouridine	1.84	2.24	4.94E-14	2.77E-07
Succinylcarnitine	1.79	2.52	2.25E-05	0.0002
L-Glutamic acid	1.73	1.78	2.31E-06	0.0016
Ratio of pseudouridine to uridine	1.72	2.15	1.14E-08	2.85E-06
Valerylcarnitine	1.66	1.46	6.70E-06	0.0261
Butyrylcarnitine	1.59	1.41	8.46E-06	0.0440
L-Kynurenine	1.54	1.42	1.98E-07	0.0006
Cortisol	1.38	1.42	0.02880	0.0206
Choline	1.37	1.26	5.59E-07	0.0138
D-Glucose	1.22	1.43	0.0310	4.68E-05
Uric acid	1.21	1.25	0.0011	0.0177
L-Leucine	1.20	1.30	0.0007	0.0192
L-Phenylalanine	1.18	1.28	0.0001	0.0013
L-Glutamine	-1.24	-1.16	1.27E-05	0.0035
Betaine	-1.28	-1.54	0.0048	0.0015
Ratio of tyrosine to phenylalanine	-1.49	-1.34	6.26E-06	0.0004
Ratio of glutamine to glutamic acid	-2.18	-2.05	4.77E-07	0.0005
1,5-Anhydro-D-glucitol	-5.47	-2.88	3.22E-07	0.0001

Stage 4 vs normal control				
2-[3-(sulfooxy)phenyl]acetic acid	139.84	267.82	1.96E-16	8.60E-12
***p*-Cresol glucuronide**	115.02	80.02	6.88E-12	0.0001
Homovanillic acid sulfate	36.66	73.94	1.50E-11	1.04E-08
***O*-Adipoylcarnitine**	21.46	43.80	2.62E-13	1.29E-08
Arabinose isomer	20.08	24.29	7.28E-11	3.51E-08
L-*β*-aspartyl-L-leucine	14.67	8.44	2.42E-13	3.26E-06
Kynurenic acid	14.29	57.82	2.90E-11	6.14E-08
2-(*α*-D-Mannopyranosyl)-L-tryptophan	14.06	10.87	2.47E-16	5.27E-08
***N*-Acetylcarnosine**	11.41	10.69	1.14E-11	3.14E-06
2-Hydroxyethanesulfonate	11.33	12.88	8.50E-12	1.00E-07
***α*-N-Phenylacetyl-L-glutamine**	11.05	31.16	1.92E-11	2.83E-07
5-Methylthio-D-ribose	9.94	12.06	8.00E-18	3.92E-10
4-Acetamidobutanoic acid	9.88	36.26	5.32E-11	1.76E-09
Indoxyl sulfate	9.48	21.24	2.71E-10	3.22E-08
Succinyladenosine	9.13	10.67	9.46E-14	1.85E-09
Hydroxybutyrylcarnitine	9.07	4.50	0.0001	0.0002
***p*-Cresol sulfate**	8.61	11.55	2.92E-06	3.41E-05

Sulfotyrosine	7.60	8.34	7.50E-12	1.02E-11
Ratio of kynurenic acid to kynurenine	7.31	20.32	2.25E-08	1.55E-07
L,L-TMAP	7.29	12.58	2.44E-12	1.92E-09
Ratio of kynurenine to tryptophan	6.67	8.65	1.84E-11	2.37E-06
Ratio of pseudouridine to uridine	6.64	12.07	7.51E-13	7.23E-12
Phenol sulfate	6.50	10.97	1.15E-07	1.07E-07
Pyrocatechol sulfate	6.32	4.90	5.31E-10	0.0207
Pseudouridine	5.17	12.36	2.11E-13	1.96E-11
L,L-TMAP isomer	5.16	11.91	3.41E-12	1.77E-09
Creatinine	4.39	9.01	1.03E-11	7.46E-10
Urea	4.05	3.44	2.73E-14	7.18E-09
Hippuric acid	3.38	6.08	0.0002	0.0118
Succinylcarnitine	3.22	5.52	7.75E-15	3.71E-07
Butyrylcarnitine	3.06	3.59	1.48E-10	0.0005
Indole-3-lactic acid	2.89	2.95	1.04E-09	0.0001
2-Octenoylcarnitine	2.49	4.44	1.21E-06	0.0003
L-Citrulline	2.30	2.43	1.25E-08	1.28E-06
Ratio of L-citrulline to arginine	2.12	2.35	0.0001	0.0002
L-Kynurenine	1.83	2.26	3.47E-10	8.51E-07
Valerylcarnitine	1.81	2.26	0.0002	0.0022
3-Indoleacetic acid	1.79	2.64	0.0003	9.34E-06
Ratio of glutamic acid to glutamine	1.70	1.98	5.21E-06	0.0182
L-Glutamic acid	1.63	1.62	1.59E-05	0.0438
Choline	1.57	1.31	2.89E-08	0.0006
9-Decenoylcarnitine	1.47	2.04	0.0059	0.0021
Citric acid	1.30	1.86	0.0027	1.22E-05
L-Phenylalanine	1.23	1.10	1.73E-06	0.0460
Uric acid	1.18	1.37	0.0364	0.0087
Cortisol	1.13	2.00	0.0330	4.35E-05
L-Glutamine	-1.17	-1.17	0.0146	0.0054
Betaine	-1.32	-1.78	0.0122	0.0063
L-Tyrosine	-1.59	-1.58	1.86E-09	0.0002
Ratio of glutamine to glutamic acid	-1.7	-2.05	5.21E-06	0.0182
L-Carnitine	-2.2	-1.66	2.69E-08	0.0068
Ratio of tyrosine to phenylalanine	-2.29	-1.97	2.38E-12	7.91E-06
L-Tryptophan	-2.93	-5.26	4.66E-08	0.0001
Bilirubin	-3.49	-2.42	1.95E-09	0.0235
1,5-Anhydro-D-glucitol	-4.88	-7.15	7.64E-08	1.40E-06

Fold change is the ratio of median at different stages over those at Stage 0. Negative sign indicates downward trend. P value is the Student's t-test of log10-transformed data to minimize the variance.

Table 5.10 Metabolites that had significant fold changes between consecutive stages of MDRD GFR in discovery and validation sets.

	Fold change		*p* value	
Stage 1b vs Stage 1a	**Discovery set**	**Validation set**	**Discovery set**	**Validation set**
L,L-TMAP	1.28	1.36	0.0005	0.0462
Creatinine	1.27	1.24	0.0349	0.0261
***γ*-Butyrobetaine**	1.25	1.17	0.0012	0.0303
L,L-TMAP isomer	1.21	1.23	0.0030	0.0413
Sulfotyrosine	1.21	1.43	0.0096	0.0138
Succinyladenosine	1.17	1.14	0.0095	0.0066
Stage 2 vs Stage 1b				
2-(*α*-D-Mannopyranosyl)-L-tryptophan	1.70	1.50	0.0001	0.0003
5-Methylthio-D-ribose	1.49	1.52	0.0003	0.0019
4-Acetamidobutanoic acid	1.42	1.59	0.0021	0.0008
Succinyladenosine	1.34	1.43	0.0089	0.0047
Creatinine	1.27	1.31	0.0005	0.0294
Sulfotyrosine	1.21	1.28	0.0111	0.0308
L,L-TMAP	1.21	1.31	0.0005	0.0150
Ratio of pseudouridine to uridine	1.21	1.23	0.0384	0.0140
Pseudouridine	1.19	1.35	4.33E-05	0.0002
Stage 3 vs Stage 2				
***O*-Adipoylcarnitine**	3.62	3.33	1.59E-05	0.0124
2-(*α*-D-Mannopyranosyl)-L-tryptophan	2.04	1.28	1.28E-08	0.0194
5-Methylthio-D-ribose	1.88	1.34	2.34E-07	0.0144
4-Acetamidobutanoic acid	1.75	1.47	7.44E-07	0.0025
L-Kynurenine	1.68	1.11	4.61E-07	0.0475
Indole-3-lactic acid	1.68	1.27	2.43E-07	0.0407
Sulfotyrosine	1.67	1.57	5.04E-07	0.0019
Arabinose isomer	1.59	1.44	0.0151	0.0393
L,L-TMAP	1.58	1.36	1.29E-06	0.0062
Pseudouridine	1.49	1.47	2.99E-09	0.0027
Ratio of pseudouridine to uridine	1.46	1.56	0.0002	0.0157
Creatinine	1.42	1.34	3.60E-05	0.0122
Succinyladenosine	1.42	1.45	4.72E-06	0.0063
L,L-TMAP isomer	1.38	1.38	0.0001	0.0047
Dehydroepiandrosterone sulfate	-1.27	-1.38	0.0401	0.0333
Stage 4 vs Stage 3				
2-[3-(sulfooxy)phenyl]acetic acid	17.78	43.54	1.19E-09	1.69E-09
Homovanillic acid sulfate	17.54	19.86	1.74E-07	2.05E-06
p-Cresol glucuronide	6.43	53.75	1.39E-05	0.0003
Indoxyl sulfate	5.62	15.23	2.69E-07	2.28E-08
Kynurenic acid	5.49	18.97	3.35E-07	4.84E-06
Arabinose isomer	5.28	9.91	9.51E-07	2.10E-07
***O*-Adipoylcarnitine**	5.08	7.93	3.04E-06	4.49E-06
Phenol sulfate	4.69	4.22	1.10E-06	1.16E-05
2-Hydroxyethanesulfonate	4.42	7.68	4.55E-08	2.54E-07
Succinyladenosine	4.15	5.46	2.97E-10	8.56E-12

5-Methylthio-D-ribose	4.01	5.43	2.30E-11	8.06E-08
***α*-N-Phenylacetyl-L-glutamine**	3.97	11.58	1.44E-08	2.97E-06
Sulfotyrosine	3.94	4.17	1.40E-08	3.22E-09
N-Acetylcarnosine	3.67	5.53	1.00E-06	0.0002
L,L-TMAP	3.63	6.59	1.21E-08	2.78E-07
L-*β*-aspartyl-L-leucine	3.52	5.28	3.91E-08	2.52E-05
4-Acetamidobutanoic acid	3.51	15.61	3.50E-07	7.30E-08
2-(*α*-D-Mannopyranosyl)-L-tryptophan	3.32	5.94	2.12E-08	2.97E-07
Pseudouridine	2.81	5.53	1.98E-09	1.84E-09
L,L-TMAP isomer	2.75	5.71	2.50E-08	2.91E-07
Pyrocatechol sulfate	2.42	4.26	0.0067	0.0135
MS-detected creatinine	2.33	5.06	4.73E-08	1.27E-07
***p*-Cresol sulfate**	2.33	13.03	0.0100	1.49E-05
Urea	2.12	2.13	1.71E-08	1.21E-06
Butyrylcarnitine	1.92	2.54	2.39E-05	0.0039
Hippuric acid	1.91	4.83	0.0184	0.0067
Succinylcarnitine	1.80	2.19	7.46E-06	0.0002
3-Hydroxydecanoyl carnitine	1.65	1.95	0.0393	0.0037
3-Indoleacetic acid	1.38	2.40	0.0051	0.0001
L-Citrulline	1.25	2.19	0.0038	4.88E-05
Citric acid	1.21	1.56	0.0365	0.0009
L-Kynurenine	1.19	1.60	0.0128	0.0004
Uracil	-1.14	-1.27	0.0484	0.0423
L-Leucine	-1.25	-1.31	2.97E-06	0.0172
L-Tyrosine	-1.31	-1.84	4.60E-05	0.0008
L-Valine	-1.47	-1.08	1.51E-07	0.0392
L-Carnitine	-1.98	-1.49	1.04E-06	0.0290
Bilirubin	-1.99	-2.83	0.0003	0.0376
L-Tryptophan	-2.20	-4.43	2.43E-06	0.0001
Ratio of pseudouridine to uridine	3.86	5.61	1.86E-09	4.68E-09
Ratio of kynurenic acid to kynurenine	3.70	10.62	3.70E-06	3.61E-06
Ratio of kynurenine to tryptophan	2.99	5.34	1.80E-07	1.74E-05
Ratio of tyrosine to phenylalanine	-1.54	-1.47	0.0004	0.0178

Fold change is the ratio of median at the later stage over the previous stage. Negative sign indicates downward trend. *p* value is the Student's *t*-test of log10-transformed data to minimize the variance.

Table 5.11 Spearman rank correlation of metabolites with MDRD GFR at different ranges with discovery and validation sets (absolute Spearman $R > 0.40$).

Metabolites associated to MDRD GFR	Stage 0-4		Stage 1-4		Stage 1-2 (GFR > = 60)	
	Discovery (n=128)	Validation (n=66)	Discovery (n=108)	Validation (n=56)	Discovery (n=69)	Validation (n=30)
Ratio of tyrosine to phenylalanine	0.6981	0.7673	0.7243	0.7849	0.4603	0.4198
L-Tryptophan	0.5866	0.4874	0.6230	0.5493	0.1238	-0.0625
L-Tyrosine	0.5503	0.5576	0.6020	0.6455	0.2420	0.2943
L-Carnitine	0.5132	0.4390	0.5393	0.4403	0.1801	0.2472
Bilirubin	0.4928	0.5255	0.5562	0.6417	-0.0285	0.4652
***γ*-Butyrobetaine**	-0.3502	-0.4051	-0.4073	-0.4200	-0.4149	-0.3789
3-Indoleacetic acid	-0.3886	-0.4846	-0.4303	-0.5126	0.0665	-0.1693
3-Hydroxydecanoyl carnitine	-0.3907	-0.5761	-0.4998	-0.5889	-0.3837	-0.2254
Valerylcarnitine	-0.4040	-0.4385	-0.4265	-0.4062	-0.1181	-0.1733
9-Decenoylcarnitine	-0.4191	-0.4195	-0.4965	-0.4292	-0.2803	-0.1168
***p*-Cresol sulfate**	-0.4741	-0.4007	-0.5342	-0.4733	0.1000	0.0278
Phenol sulfate	-0.5065	-0.6050	-0.5619	-0.6265	-0.2028	-0.2774
Homovanillic acid sulfate	-0.5241	-0.8027	-0.5923	-0.8144	0.1074	-0.4300
***p*-Cresol glucuronide**	-0.5253	-0.5068	-0.5708	-0.5697	0.1694	-0.0968
L-Citrulline	-0.5498	-0.5760	-0.6201	-0.6898	-0.1090	-0.0407
Ratio of kynurenic acid to kynurenine	-0.5650	-0.6602	-0.6263	-0.7037	-0.1099	-0.1631
Indoxyl sulfate	-0.5656	-0.6264	-0.6418	-0.6804	-0.0882	-0.4376
***α*-N-Phenylacetyl-L-glutamine**	-0.5675	-0.5627	-0.6374	-0.6221	0.0031	-0.0407
N-Acetylcarnosine	-0.5839	-0.6347	-0.6304	-0.6384	-0.1987	-0.3540
Urinary *β*2-microglobulin (mg/L)	-0.6285	-0.7140	-0.6714	-0.7571	-0.2854	-0.1552
Butyrylcarnitine	-0.6353	-0.6309	-0.6996	-0.6592	-0.2030	-0.3117
L-*β*-Aspartyl-L-leucine	-0.6773	-0.7862	-0.7055	-0.8608	-0.3486	-0.7095
Arabinose isomer	-0.6815	-0.7660	-0.7161	-0.7932	-0.3068	-0.3090
Indole-3-lactic acid	-0.6894	-0.6941	-0.7438	-0.6713	-0.3363	-0.4478
Serum urea (mmol/L)	-0.6974	-0.7014	-0.7632	-0.7481	-0.2001	-0.3596
L-Kynurenine	-0.7096	-0.8325	-0.7580	-0.8486	-0.3547	-0.6583
Kynurenic acid	-0.7208	-0.8394	-0.7852	-0.8756	-0.3101	-0.6329
2-[3-(sulfooxy)phenyl]acetic acid	-0.7579	-0.8080	-0.8211	-0.8225	-0.3793	-0.4149
***O*-Adipoylcarnitine**	-0.7663	-0.8150	-0.8233	-0.8513	-0.4774	-0.4643
Ratio of kynurenine to tryptophan	-0.7711	-0.8004	-0.8236	-0.8342	-0.4089	-0.5186
4-Acetamidobutanoic acid	-0.7810	-0.8614	-0.8299	-0.8986	-0.4113	-0.5350
Succinylcarnitine	-0.8077	-0.6536	-0.8369	-0.6622	-0.6178	0.0581
2-Hydroxyethanesulfonate	-0.8136	-0.7400	-0.8710	-0.8170	-0.6043	-0.5310
Ratio of pseudouridine to uridine	-0.8137	-0.8836	-0.8601	-0.9063	-0.6251	-0.5591
Sulfotyrosine	-0.8200	-0.8874	-0.8719	-0.9226	-0.6038	-0.7152
Succinyladenosine	-0.8546	-0.9262	-0.8986	-0.9321	-0.7060	-0.8176
L,L-TMAP isomer	-0.8637	-0.8695	-0.9118	-0.9113	-0.7314	-0.6392
L,L-TMAP	-0.8659	-0.9013	-0.9206	-0.9222	-0.7341	-0.7393
Creatinine	-0.8692	-0.9124	-0.9004	-0.9161	-0.7124	-0.7393
5-Methylthio-D-ribose	-0.8731	-0.8979	-0.9110	-0.9171	-0.6975	-0.6574
2-(*α*-D-Mannopyranosyl)-L-tryptophan	-0.8753	-0.9276	-0.9180	-0.9353	-0.7326	-0.8469
Pseudouridine	-0.8962	-0.9389	-0.9394	-0.9530	-0.7889	-0.7918
Serum creatinine (mg/dL)	-0.9308	-0.9500	-0.9546	-0.9592	-0.8446	-0.8546
Serum homocysteine (μmol/L)		-0.6371		-0.6768		-0.3877
Serum cystatin C (mg/L)	-	-0.9126	-	-0.9400	-	-0.8016

5.3.4 Network of metabolites

In this study, metabolomics network diagrams were constructed using 72 metabolites (those had Pubmed ID listed in Table 5.8) based on their metabolic pathway and chemical similarity. They showed a gradual elevation of up-regulated metabolite numbers from 6 metabolites at Stage 1a up to 40 metabolites at Stage 4 with increasing fold changes with the severity of kidney function (Fig. 5.8). The serum levels of metabolites in the later stages [Stages 3 (31 metabolites) and 4 (47 metabolites)] were more changed than those in the early Stages 1a (9 metabolites), 1b (6 metabolites) and 2 (15 metabolites) when all were compared with the normal group. The metabolites that were structurally similar or KEGG-reaction pairs were clustered together and could be categorized into six classes: sulfate metabolites, amino acids and organic acids, acylcarnitine, purine derivatives, monosaccharides and their derivatives, and steroids. Among the altered metabolites, 1,5-anhydro-D-glucitol was markedly downregulated and hydroxybutrylcarnitine was upregulated at each stage of CKD compared with the normal group.

The pathway enrichment analysis using Metaboanalyst with HMDB ID identified six amino-acid related metabolic pathways, namely tryptophan, phenylalanine, nitrogen, arginine and proline metabolisms, aminoacyl-tRNA biosynthesis and tryptophan, phenylalanine and tyrosine biosynthesis (FDR < 0.05) (Fig. 5.9). Among them, eight significantly changed metabolites matched the best with tryptophan metabolism (tryptophan, 3-indoleacetic acid, kynurenine and kynurenic acid) and phenylalanine metabolism (phenylalanine, tyrosine, hippuric acid and *α-N*-phenylacetyl-L-glutamine) (pathway impact values from the pathway topology analysis > 0.1) (Fig. 5.9A-Fig. 5.9B). In comparison with the normal subjects, the kynurenine to tryptophan ratio [estimation of activity of indoleamine 2,3-dioxygenase (IDO)] (Debnath et al. 2017)] and kynurenic acid to kynurenine ratio [(estimation of activity of kynurenine aminotransferase (KAT)

(Myint et al. 2007; Birner et al. 2017)] were also increased in Stage 4, which might be related to the increase in the above metabolites in the tryptophan metabolism (Fig. 5.9C). The reduction of tyrosine to phenylalanine ratio (estimation of activity of phenylalanine hydroylase) supported the elevation in hippuric acid and *N*-phenylacetyl-L-glutamine sourced from phenylalanine in the phenylalanine metabolism (Fig. 5.9D). These altered metabolites, excluding phenylalanine and hippuric acid and the three ratios were also correlated with GFR (Absolute Spearman R = 0.43-0.88 among diabetic patients) (Table 5.11), indicating the alteration of these pathways occurred during the development of CKD.

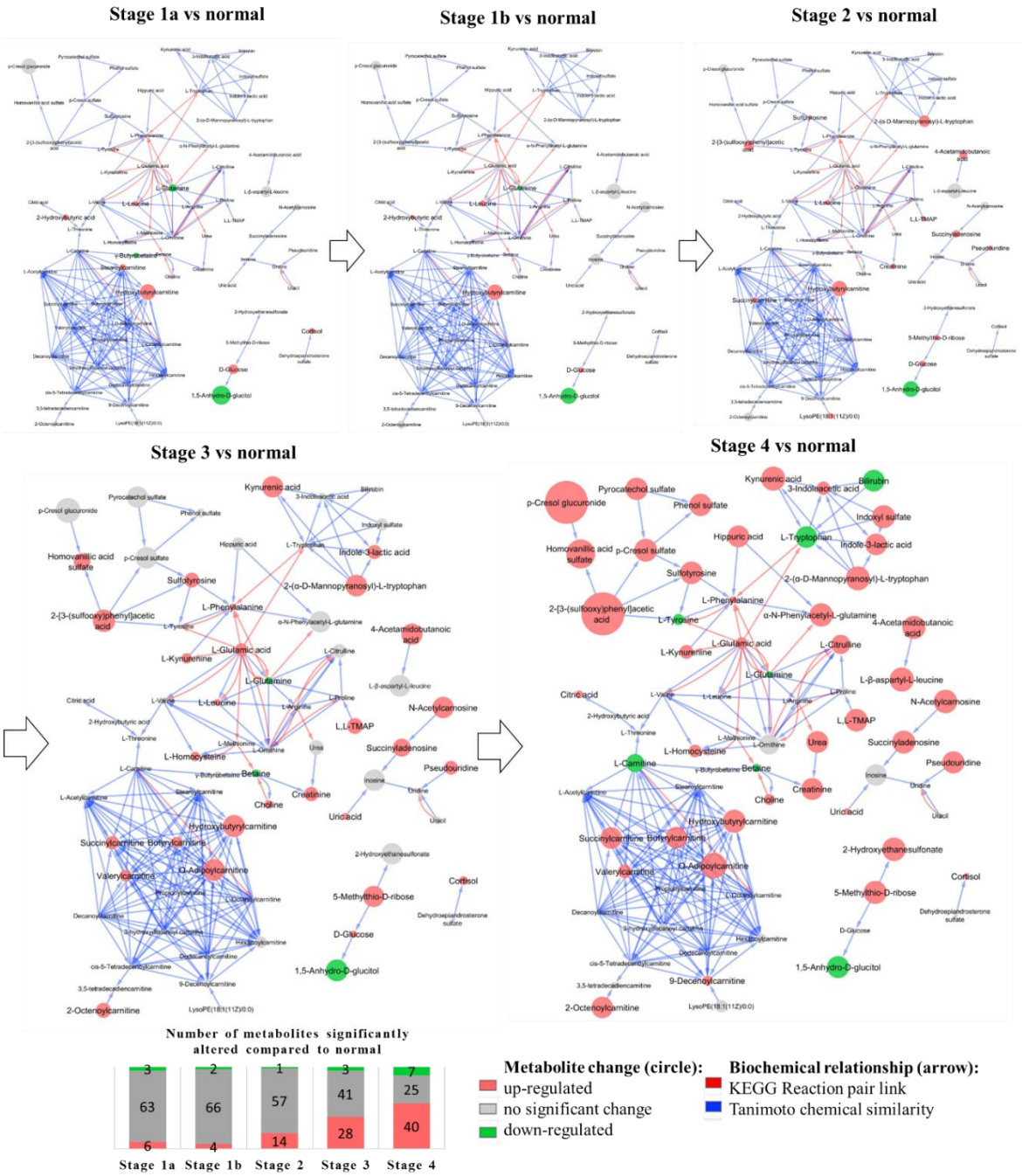
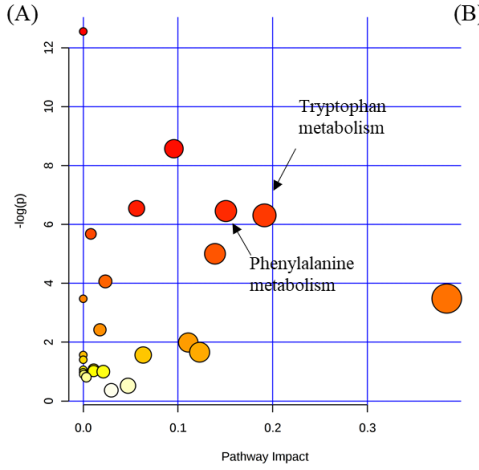


Fig. 5.8 Metabolomics network diagrams of 72 metabolites for visualization of their significant change in each stage of CKD compared with normal subjects.

Red circle: upregulated; green circle: metabolites downregulated; grey circle denotes metabolites without insignificant change or inconsistent in both datasets. Circle sizes were proportional to the absolute value of fold change with respect to the normal group.

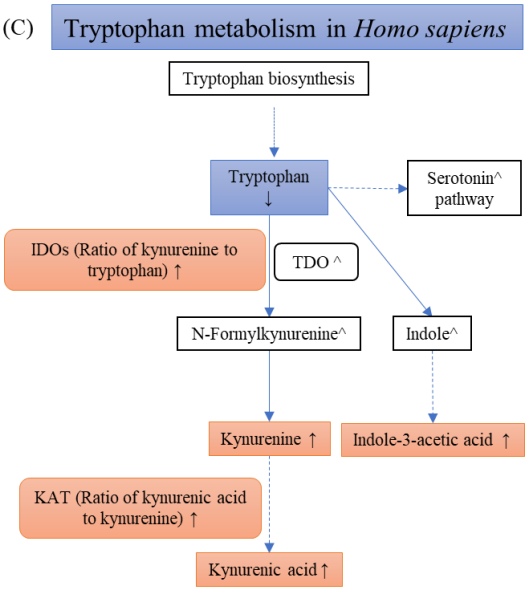
(A)



(B)

Pathway with KEGG database with Homo sapiens	Total	Hits	*p* value	-log(p)	FDR	Impact
Tryptophan metabolism	**79**	**4**	**0.0018**	**6.3081**	**0.0291**	**0.191**
Phenylalanine metabolism	**45**	**4**	**0.0016**	**6.4543**	**0.0291**	**0.151**
Alanine, aspartate and glutamate metabolism	24	2	0.0308	3.4815	0.2479	0.384
D-Glutamine and D-glutamate metabolism	11	2	0.0067	5.0014	0.0769	0.139
Cysteine and methionine metabolism	56	2	0.1369	1.9882	0.9129	0.111
Arginine and proline metabolism	77	6	0.0002	8.5717	0.0076	0.096
Aminoacyl-tRNA biosynthesis	75	5	0.0014	6.5426	0.0291	0.056
Glycine, serine and threonine metabolism	48	3	0.0172	4.0635	0.1719	0.023
Purine metabolism	92	3	0.0889	2.4198	0.6468	0.018
Porphyrin and chlorophyll metabolism	104	2	0.3429	1.0703	1	0.011
Phenylalanine, tyrosine and tryptophan biosynthesis	27	3	0.0034	5.6775	0.0456	0.008
Nitrogen metabolism	39	6	3.53E-06	12.555	0.0003	0
Pyrimidine metabolism	60	3	0.0310	3.474	0.2479	0

(C)



(D)

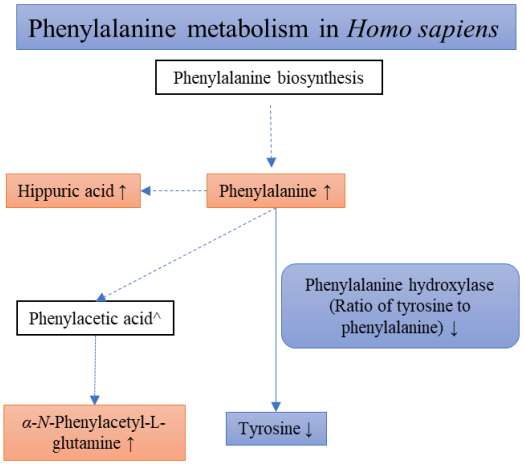


Fig. 5.9 Results of pathway analysis of all significantly altered metabolites with HMDB ID.

(A) Metabolome view. (B) Table of the matched pathway with *p* values from pathway enrichment analysis and pathway impact values from the pathway topology analysis using MetaboAnalyst and KEGG database (Hits ≥ 2). (C-D) Simplified pathways of tryptophan metabolism and phenylalanine metabolism with the change trends of metabolites and their ratios at Stage 4 compared with the normal group. A Dashed arrow indicates the skipped pathway. ^, not detected.

5.3.5 Significantly changed metabolites at early stages: 1,5-anhydro-D-glucitol.

Compared with the normal group, diabetic patients at Stage 1a had reduced levels of 1,5-anhydro-D-glucitol, γ-butyrobetaine and L-glutamine, and increased levels of hydroxybutyrylcarnitine, D-glucose, stearoylcarnitine, 2-hydroxybutyric acid, cortisol and L-leucine.

Only 1,5-anhydro-D-glucitol were markedly reduced in each stage (fold change = -26.5 to -2.60 against the normal individuals, Table 5.9). The contents of two clinical glycemic markers, FBG and serum hemoglobin A1c, were highly and inversely correlated with the serum 1,5-anhydro-D-glucitol level within each stage (Stages 1-3) (Pearson *R* with hemoglobin A1c = -0.95 to -0.64; Pearson *R* with FBG = -0.87 to -0.42) (Table 5.12). This was more noticeable in hemoglobin A1c (Linear regression with R^2 = 0.63-0.71 among Stage 1-3) (Fig. 5.10)**.** The positive linear relationship of serum 1,5-anhydro-D-glucitol level with FBG was the strongest in Stage 1a (Pearson *R* with FBG = -0.87 to -0.72), followed by Stage 1b (Pearson *R* with FBG = -0.77 to -0.65). On the contrary, there were generally no association between the two clinical glycemic markers and MDRD GFR ($p > 0.05$, Table 5.12).

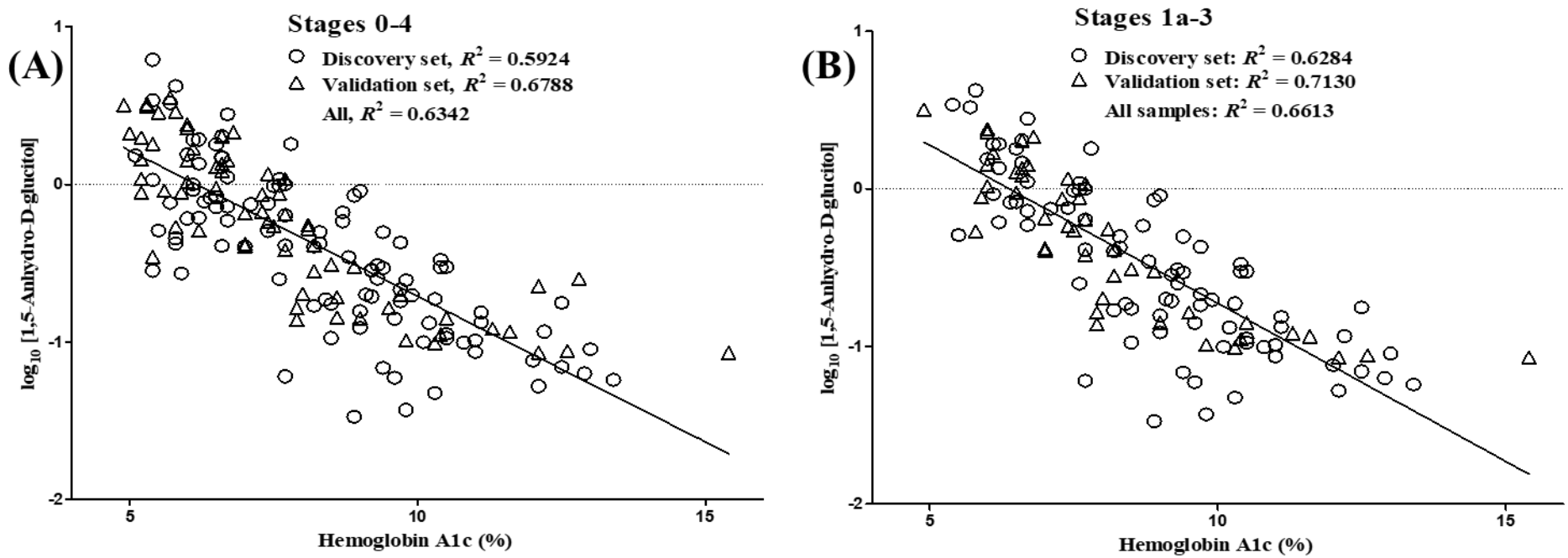


Fig. 5.10 Linear regression of UPLC-MS detected log[1,5-anhydro-D-glucitol] against hemoglobin A1c in (A) Stages 0-4 and (B) Stages 1-3 of CKD.

Table 5.12 Correlation between UPLC-MS detected 1,5-anhydro-D-glucitol, hemoglobin A1c, FBG, MS-detected D-glucose and MDRD GFR in each stage.

Correlation	Normal		Stage 1a		Stage 1b		Stage 2		Stage 3		Stage 4	
	Discovery	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation
	n = 20	n =10	HbA1c: n = 18 FBG: n = 20	n = 10	HbA1c: n = 17 FBG: n = 19	n = 10	HbA1c: n = 29 FBG: n = 30	n = 10	HbA1c: n = 17 FBG: n = 19	n = 15	HbA1c: n = 19 FBG: n = 20	n = 11
log_{10} (1,5-Anhydro-D-glucitol)												
FBG^	0.184 ns	0.311 ns	-0.723	-0.865	-0.652	-0.766	-0.422	-0.650	-0.721	-0.585	0.225 ns	-0.603
Hemoglobin A1c^	-	0.534 ns	-0.821	-0.831	-0.874	-0.835	-0.639	-0.954	-0.839	-0.827	-0.445 ns	-0.669
MDRD GFR^	0.141 ns	0.312 ns	-0.235 ns	-0.443 ns	0.420 ns	-0.448 ns	-0.043 ns	0.075 ns	-0.242 ns	0.052 ns	0.217 ns	0.587 ns
Hemoglobin A1c												
MDRD GFR*	-	-0.308 ns	0.366 ns	0.663	-0.367 ns	0.673	-0.014 ns	-0.172 ns	0.185 ns	0.117 ns	0.181 ns	-0.132 ns
FBG (mmol/L)												
MDRD GFR*	-0.154 ns	0.219 ns	-0.036 ns	0.588 ns	-0.326 ns	0.527 ns	-0.056 ns	-0.273 ns	0.160 ns	0.236 ns	0.627	< 0.001 ns

ns , no statistical significance ($p > 0.05$).
^ Pearson correlation.
*Spearman rank correlation
HbA1c**,** hemoglobin A1c.

5.3.6 Potential uremic retention solutes that were highly correlated to MDRD GFR

Over 30 metabolites and 4 ratios of metabolites had high correlation with MRDR GFR range in Stages 0-4 (including diabetic patients and normal subjects) and Stages 1-4 (only diabetic patients: Spearman $R > 0.4$, Table 5.11**).** Among them, 28 MS-detected metabolites were uremic retention solutes in the serum samples and two were peptide uremic solutes that were elevated with the stages of CKD. According to their protein-binding affinity, molecular size and publication history (Vanholder, Smet, and Lameire 2004; Neirynck et al. 2012), they were classified into five classes: (1) five low-molecular-weight (≤ 300 Da), water-soluble and protein-unbound uremic solutes, (2) nine low-molecular-weight and protein-bound uremic solutes, (3) two middle-molecular-weight (300-15k Da) peptides, (4) nine uremic solutes that were reported but not classified, and (5) five uremic solutes that were not reported and could not be identified with standards. Most of their serum levels were much higher at Stage 4 than those at other stages **(**Tables 5.9-5.10**),** indicating a great loss of kidney clearance function. Their classification and correlation to MDRD GFR were summarized in Table 5.13.

Using a Spearman rank coefficient cutoff of 0.8 for diabetic patients and 0.7 for those at early stages to select biomarkers altered along with MDRD GFR, four serum metabolites by UPLC-Orbitrap-MS were significantly associated with the severity of kidney functions. The four metabolites were succinyladenosine ($[M-H]^-$ = 382.1005 at 2.89 min), pseudouridne ($[M-H]^-$ = 243.0622 at 0.93 min), 2-(α-D-Mannopyranosyl)-L-tryptophan ($[M-H]^-$ = 367.1497 at 2.21 min) and L,L-TMAP ($[M+H]^+$ = 229.1546 at 1.06 min). They were highly correlated with MDRD GFR (Spearman R = -0.95 to - 0.90) and were also correlated to each other (Spearman R =0.86 - 0.96) among Stages 1a - 4 (Fig. 5.11). Furthermore, similar high associations were observed in partial correlation with MDRD GFR (partial correlation R = -0.95 to -0.84) after controlling gender, age, SBP and UACR using log10-transformed data (Table 5.14). Their

correlations remained high in early Stages 1-2 (Spearman R = -0.85 to -0.71, partial correlation R = -0.76 to -0.53).

Correlation analyses with other clinical factors showed that the four selected metabolites were also associated with other kidney function related markers (Fig. 5.11 and Table 5.15). They were increased with the rise in UACR (Spearman R = 0.68-0.80), urinary β2-microglobulin (Spearman R = 0.67-0.79), renal resistive index (Spearman R = 0.52 to 0.56) and the decline in the total BSA-related renal volume (Spearman R = -0.62 to -0.60) among diabetic patients**.** Importantly, most of their correlation performances were better than that of serum creatinine and only succinyladenosine had all of these correlation coefficients higher than that of serum cystatin C.

Table 5.13 Classification of 28 uremic solutes, 4 ratios and non-uremic solutes found in this studies and Spearman rank correlation to MDRD GFR at all, diabetic and early stages of CKD.

Uremic solutes	Metabolites associated to MDRD GFR	Stage 0-4		Stage 1-4		Stage 1- 2	
		Discovery (n=128)	Validation (n=66)	Discovery (n=108)	Validation (n=56)	Discovery (n=69)	Validation (n=30)
Middle-molecular-weight peptides	Serum cystatin C (mg/L)	-	-0.913	-	-0.940	-	-0.802
	Urinary *β*2-microglobulin (mg/L)	-0.629	-0.714	-0.671	-0.757	-0.285	-0.155
Low-molecular-weight protein-unbound molecules	Clinically measured serum creatinine (mg/dL)	-0.931	-0.950	-0.955	-0.959	-0.845	-0.855
	Pseudouridine	-0.896	-0.939	-0.939	-0.953	-0.789	-0.792
	MS-detected creatinine	-0.869	-0.912	-0.900	-0.916	-0.712	-0.739
	Serum urea (mmol/L)	-0.697	-0.701	-0.763	-0.748	-0.200	-0.360
	α-*N*-Phenylacetyl-L-glutamine	-0.568	-0.563	-0.637	-0.622	0.003	-0.041
Low-molecular-weight and protein-bound molecules	Kynurenic acid	-0.721	-0.839	-0.785	-0.876	-0.310	-0.633
	L-Kynurenine	-0.710	-0.833	-0.758	-0.849	-0.355	-0.658
	Indole-3-lactic acid	-0.689	-0.694	-0.744	-0.671	-0.336	-0.448
	Serum homocysteine (µmol/L)	-	-0.637	-	-0.677	-	-0.388
	Indoxyl sulfate	-0.566	-0.626	-0.642	-0.680	-0.088	-0.438
	p-Cresol glucuronide	-0.525	-0.507	-0.571	-0.570	0.169	-0.097
	Phenol sulfate	-0.507	-0.605	-0.562	-0.627	-0.203	-0.277
	p-Cresol sulfate	-0.474	-0.401	-0.534	-0.473	0.100	0.028
	3-Indoleacetic acid	-0.389	-0.485	-0.430	-0.513	0.067	-0.169
Other molecules that were reported but not classified	2-(*α*-D-Mannopyranosyl)-L-tryptophan	-0.875	-0.928	-0.918	-0.935	-0.733	-0.847
	L,L-TMAP	-0.866	-0.901	-0.921	-0.922	-0.734	-0.739
	Succinyladenosine	-0.855	-0.926	-0.899	-0.932	-0.706	-0.818
	Sulfotyrosine	-0.820	-0.887	-0.872	-0.923	-0.604	-0.715
	2-Hydroxyethanesulfonate	-0.814	-0.740	-0.871	-0.817	-0.604	-0.531
	Succinylcarnitine	-0.808	-0.654	-0.837	-0.662	-0.618	0.058
	4-Acetamidobutanoic acid	-0.781	-0.861	-0.830	-0.899	-0.411	-0.535
	O-Adipoylcarnitine	-0.766	-0.815	-0.823	-0.851	-0.477	-0.464
	Butyrylcarnitine	-0.635	-0.631	-0.700	-0.659	-0.203	-0.312
	N-Acetylcarnosine	-0.584	-0.635	-0.630	-0.638	-0.199	-0.354
	L-Citrulline	-0.550	-0.576	-0.620	-0.690	-0.109	-0.041
	Homovanillic acid sulfate	-0.524	-0.803	-0.592	-0.814	0.107	-0.430

Other molecules that were not reported and putatively identified	**5-Methylthio-D-ribose**	-0.873	-0.898	-0.911	-0.917	-0.698	-0.657
	L,L-TMAP isomer	-0.864	-0.870	-0.912	-0.911	-0.731	-0.639
	2-[3-(sulfooxy)phenyl]acetic acid	-0.758	-0.808	-0.821	-0.823	-0.379	-0.415
	Arabinose isomer	-0.682	-0.766	-0.716	-0.793	-0.307	-0.309
	L-*β*-Aspartyl-L-leucine	-0.677	-0.786	-0.706	-0.861	-0.349	-0.710
Non-uremic solutes and ratios of metabolites	**Ratio of pseudouridine to uridine**	-0.814	-0.884	-0.860	-0.906	-0.625	-0.559
	Ratio of kynurenine to tryptophan	-0.771	-0.800	-0.824	-0.834	-0.409	-0.519
	Ratio of kynurenic acid to kynurenine	-0.565	-0.660	-0.626	-0.704	-0.110	-0.163
	L-Tyrosine	0.550	0.558	0.602	0.646	0.242	0.294
	L-Tryptophan	0.587	0.487	0.623	0.549	0.124	-0.063
	Ratio of tyrosine to phenylalanine	0.698	0.767	0.724	0.785	0.460	0.420

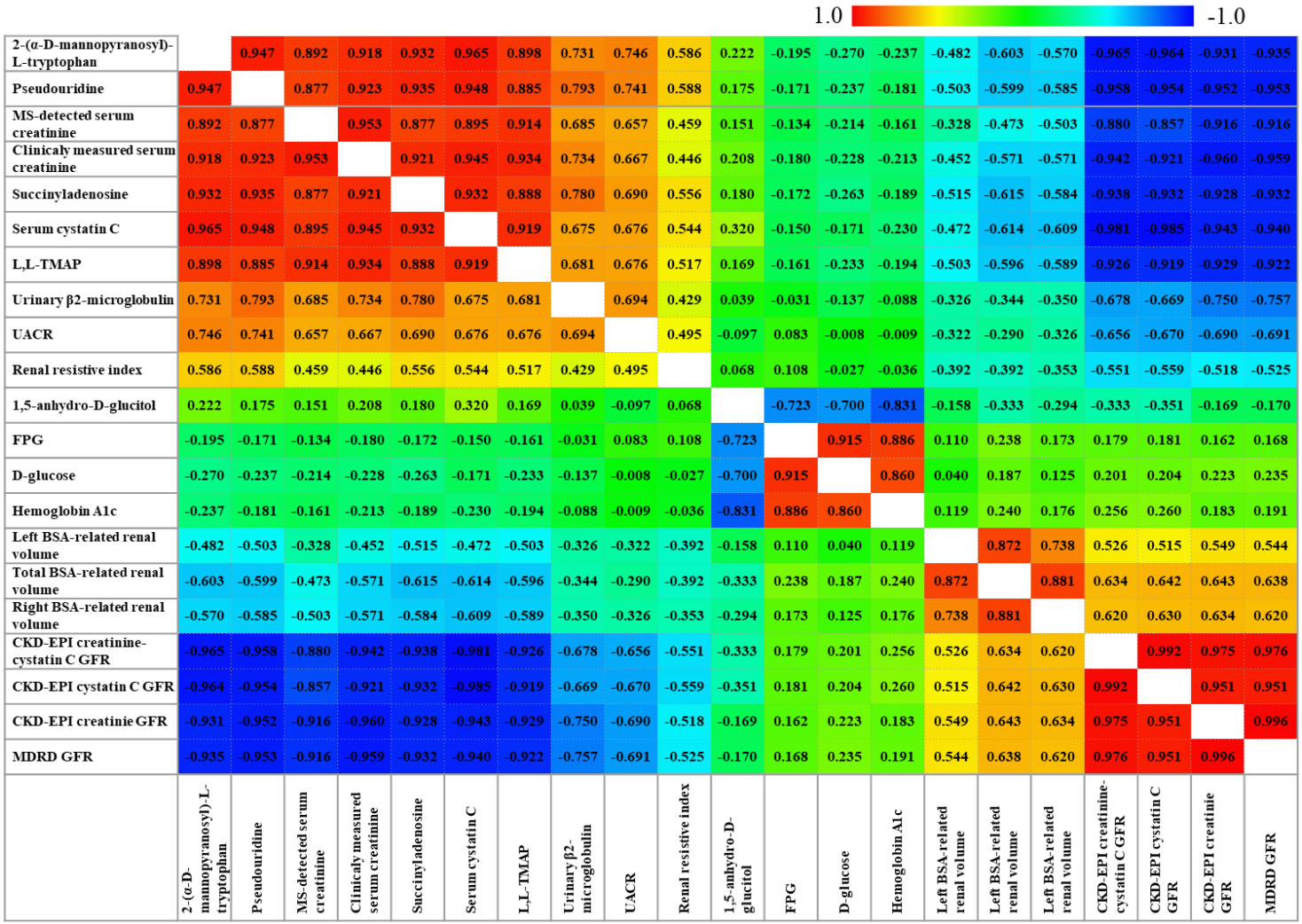


Fig. 5.11 Correlogram of Spearman rank correlation coefficient of biomarkers with clinical markers and GFRs calculated by different equations in validation sets at Stages 1-4.

Table 5.14 Correlation of clinical markers and biomarkers with MDRD GFR at different ranges of stages with and without controlling confounding factors.

Metabolites	Stage 0-4		Stage 1-4		Stage 1-2	
	Discovery (n = 128)	Validation (n = 66)	Discovery (n = 108)	Validation (n = 56)	Discovery (n = 69)	Validation (n = 30)
Spearman Rank Correlation with MDRD GFR						
Succinyladenosine	-0.855	-0.926	-0.899	-0.932	-0.706	-0.818
L,L-TMAP	-0.866	-0.901	-0.921	-0.922	-0.734	-0.739
MS-detected creatinine	-0.869	-0.912	-0.900	-0.916	-0.712	-0.739
2-(*α*-D-Mannopyranosyl)-L-tryptophan	-0.875	-0.928	-0.918	-0.935	-0.733	-0.847
Pseudouridine	-0.896	-0.939	-0.939	-0.953	-0.789	-0.792
Clinically measured serum creatinine	-0.931	-0.950	-0.955	-0.959	-0.845	-0.855
Serum cystatin C	-	-0.913	-	-0.940	-	-0.802
Pearson Correlation with log [MDRD GFR]						
log [2-(α-D-Mannopyranosyl)-L-tryptophan]	-0.921	-0.976	-0.932	-0.975	-0.770	-0.819
log [Succinyladenosine]	-0.940	-0.957	-0.944	-0.954	-0.639	-0.697
log [MS-detected creatinine]	-0.955	-0.975	-0.957	-0.977	-0.715	-0.759
log [L,L-TMAP]	-0.959	-0.973	-0.965	-0.975	-0.739	-0.787
log [Pseudouridine]	-0.972	-0.976	-0.976	-0.976	-0.796	-0.793
log [Serum cystatin C]	N/A	-0.972	N/A	-0.973	N/A	-0.851
log [Clinically measured serum creatinine]	-0.983	-0.988	-0.984	-0.989	-0.869	-0.879
Partial correlation of metabolites with log [MDRD GFR] after controlling gender, age, SBP and log [UACR]						
log [2-(α-D-Mannopyranosyl)-L-tryptophan]	-0.823	-0.956	-0.844	-0.958	-0.701	-0.755
log [Succinyladenosine]	-0.867	-0.898	-0.874	-0.898	-0.533	-0.571
log [L,L-TMAP]	-0.924	-0.950	-0.930	-0.953	-0.685	-0.698
log [Pseudouridine]	-0.941	-0.949	-0.947	-0.952	-0.739	-0.709
log [MS-detected creatinine]	-0.957	-0.975	-0.962	-0.976	-0.850	-0.847
log [Serum cystatin C]	N/A	-0.949	N/A	-0.953	N/A	-0.831
log [Clinically measured serum creatinine]	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000

Table 5.15 Spearman rank correlation of selected four metabolites to UACR, urinary *β*2-microglobulin, total BSA-related renal volume and renal resistive index among diabetic patients at Stage 1-4 in the discovery and validation sets.

CKD risk factors	UACR		Urinary *β*2-microglobulin		Total BSA-related renal volume	Renal resistive index
	Discovery (n=108)	Validation (n=55)	Discovery (n=108)	Validation (n=55)	Validation (n=48)	Validation (n=48)
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.801	0.746	0.664	0.731	-0.604	0.586
Succinyladenosine	0.795	0.690	0.635	0.780	-0.615	0.556
Pseudouridine	0.794	0.741	0.690	0.793	-0.599	0.588
L,L-TMAP	0.743	0.676	0.670	0.681	-0.596	0.517
Serum creatinine (mg/dL)	0.737	0.667	0.658	0.734	-0.571	0.446
MS-detected creatinine	0.681	0.657	0.667	0.685	-0.473	0.459
Serum cystatin C (mg/L)	-	0.676	-	0.675	-0.614	0.544

5.3.7 Gender differences and linear regression of potential biomarkers along with MDRD GFR

The four biomarkers that had the highest correlation with MDRD GFR: succinyladenosine, 2-(α-D-mannopyranosyl)-L-tryptophan, pseudouridine and L,L-TMAP had no gender difference after comparison between male and female among all diabetic patients (Mann-Whitney U, $p > 0.07$) (Fig. 5.12A). In the opposite, there were significant gender discrepancies in serum creatinine and cystatin C ($p < 0.05$) (Fig. 5.12A). Furthermore, univariate linear regression of female and male samples against MDRDGFR showed that the gender discrepancy of clinically measured and MS-detected creatinine was more obvious than that of pseudouridine, succinyladenosine and 2-(α-D-mannopyranosyl)-L-tryptophan and L,L-TMAP in the sample distribution and trendlines (Fig. 5.12B**)**. The four biomarkers' relative intensity changes along with MDRD GFR (slope of the trendlines in linear regression = -1.155 to -0.8328) were more sensitive than that of MS-detected creatinine (slopes = -0.7926) (Fig. 5.12B).

Multivariate linear regression included gender as a covariate to test for the effect of gender in the relationship of four biomarkers to log [MDRD GFR]. Only L,L-TMAP showed gender significantly contributed to the linear fit ($p < 0.05$) but its model had a smaller unstandardized coefficient ($\beta = 0.0627$) than serum creatinine ($\beta = 0.1584$). Gender had no effect in the linear relationship of log [pseudouridine], log [succinyladenosine] and log [2-(α-D-mannopyranosyl) -L-tryptophan] with log [MDRD GFR].

These results suggested that the relationships of pseudouridine, succinyladenosine and 2-(α-D-mannopyranosyl)-L-tryptophan with MDRD GFR are gender-independent. Gender had less effect on L,L-TMAP than serum creatinine.

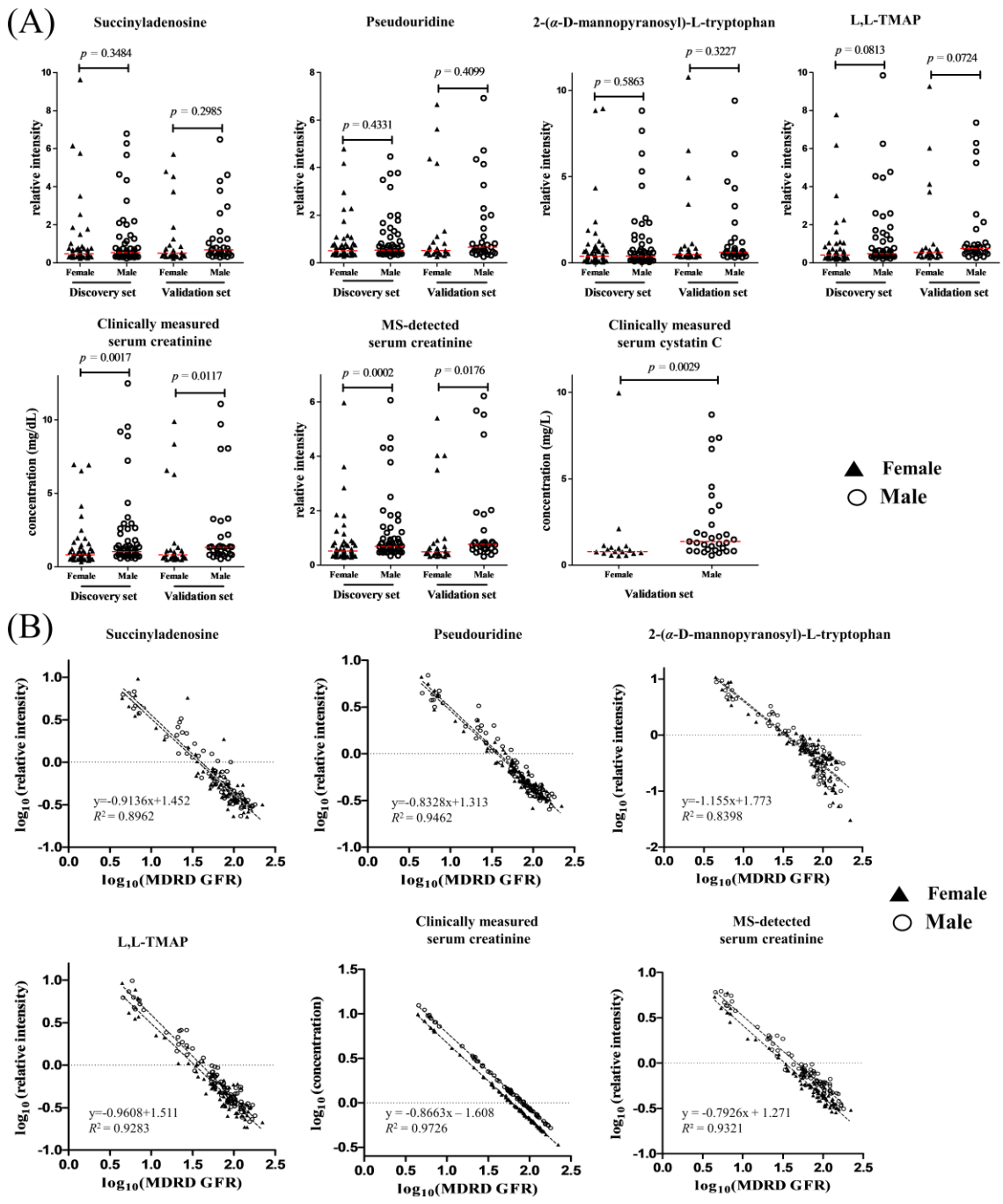


Fig. 5.12 (A) Scatter plots in gender difference. (B) Relationship of clinically measured markers and MS-detected biomarkers with MDRD GFR.

p value was calculated by Mann-Whitney U due to non-normal distributed data. Regression equations and R^2 were determined by the trendline of all diabetic males and females from two cohorts.

Table 5.16 Univariate and multivariate linear regression analyses of biomarkers with log [MDRD GFR] trained with discovery set and tested with validation set among diabetic patients

log [MDRD GFR]	Discovery set					Validation set
	Unstandardized Coefficients (β)	p value	R^2 *	Adjusted R^2	RMSE	R^2 †
			Univariate linear regression			
log [MS-detected creatinine]	-1.1804	$p < 0.001$	0.9160	-	0.1082	0.9541
log [pseudouridine]	-1.2182	$p < 0.001$	0.9527	-	0.0812	0.9519
log [L,L-TMAP]	-0.9410	$p < 0.001$	0.9305	-	0.0985	0.9500
log [succinyladenosine]	-0.9431	$p < 0.001$	0.8905	-	0.1236	0.9104
log [2-(α-D-mannopyranosyl)-L-tryptophan]	-0.6691	$p < 0.001$	0.8688	-	0.1353	0.9496
			Multivariate linear regression			
Model L1			0.9583	0.9577	0.0765	0.9709
constant	1.3591	$p < 0.001$				
log [MS-detected creatinine]	-1.2513	$p < 0.001$				
gender	0.1584	9.76E-18				
Model L2			0.9531	0.9523	0.0812	0.95283
constant	1.5353	$p < 0.001$				
log [pseudouridine]	-1.2203	$p < 0.001$				
gender	0.0152	0.3352				
Model L3			0.9375	0.9363	0.0938	0.9569
constant	1.4681	$p < 0.001$				
log [L,L-TMAP]	-0.9531	$p < 0.001$				
gender	0.0627	0.0009				
Model L4			0.8908	0.8888	0.1240	0.9116
constant	1.5852	$p < 0.001$				
log [succinyladenosine]	-0.9446	$p < 0.001$				
gender	0.0138	0.5662				
Model L5			0.8690	0.8665	0.1358	0.9499
constant	1.5295	$p < 0.001$				
log [2-(α-d-mannopyranosyl)-l-tryptophan]	-0.6699	$p < 0.001$				
gender	0.0107	0.6834				
Model L6 (enter method)			0.9580	0.9519	0.0480	0.9872
Constant	1.4315	$p < 0.001$				
log [MS-detected creatinine]	-0.6727	$p < 0.001$				
log [pseudouridine]	-0.3135	0.0095				
log [L,L-TMAP]	-0.1575	0.0227				
log [succinyladenosine]	-0.0226	0.6510				

log [2-(α-D-mannopyranosyl)-L-tryptophan]	-0.0426	0.2328				
gender	0.0967	$p < 0.001$				
age	-0.0019	0.0705				
SBP	0.0003	0.2942				
BMI	0.0018	0.3006				
log [UACR]	0.0033	0.6613				
Model L7 (the best model by stepwise method of model 2)			0.9788	0.9781	0.0550	0.9879
Constant	1.4315	$p < 0.001$				
log [MS-detected creatinine]	-0.6083	$p < 0.001$				
log [L,L-TMAP]	-0.1403	0.0214				
log [pseudouridine]	-0.4803	$p < 0.001$				
gender	0.0934	$p < 0.001$				

Gender, female = 1 and male = 2.
Discovery set, n = 106; validation set, n =56.
***** R^2 was based on the predicted log [MDRD GFR] against actual log [MDRD GFR] using the equation of the model and data of discovery set.

† R^2 was based on the predicted log [MDRD GFR] against actual log [MDRD GFR] using the equation of the model of discovery set and data of validation set.

.

5.3.8 Multivariate linear regression for prediction of MDRD GFR using potential MS-detected biomarkers

Potential biomarkers had a strong linear relationship with log [MDRD GFR] with high R^2 in univariate linear regression in the descending order; log [pseudouridine] > log [L,L-TMAP] > log [MS-detected creatinine] > log [succinyladenosine] > log [2-(α-D-mannopyranosyl)-L-tryptophan] (Table 5.16).

Stepwise multivariate linear regression analysis is a method that automatically do multiple regressions, remove variables with weak statistical contribution and retain variables that explain the distribution best. Only the patient's gender, log [MS-detected creatinine], log [pseudourdine] and log [L,L-TMAP] remained significant ($p < 0.05$) as independent predictors for log [MDRD GFR] in the regression equation (Table 5.16, Model L7) while age, BMI, SBP, UACR, log [succinyladensine] and log [2-(α-D-mannopyranosyl)-L-tryptophan] were insignificant and excluded. The selected multiple regression equation of Model L7 is here:

$$\begin{aligned}\textbf{predicted log}\,[\textbf{MDRD GFR}] &= -\,\mathbf{0.6083}\,\textbf{log}\,[\textbf{MS}-\textbf{detected creatinine}]\;\mathbf{0.4803}\,\textbf{log}\,[\textbf{pseudouridine}] \\ &-\,\mathbf{0.1403}\,\textbf{log}\,[\textbf{L,L}-\textbf{TMAP}]\;+\;\mathbf{0.0934}\,\textbf{gender}\,(\textbf{female},\mathbf{1};\;\textbf{male},\mathbf{2}) \\ &+\;\mathbf{1.4315}\end{aligned}$$

Model L7 had a high adjusted R^2 value of 0.978 accounting for 97.8% variance of overall prediction. Root Mean Square Error (RMSE) is the square root of the residual mean square. It estimated the mean squared error in prediction and equaled to 0.055, indicating a high reliability of the variables' contribution to the model.

External validation was applied using the separate validation set to avoid overfitting. The R^2 of the linear regression of predicted log [MDRD GFR] against actual values using validation set as testing samples was 0.988, indicating a good quality of curve fit was obtained. The two linear curves of the training and testing datasets were overlapped in Fig. 5.13 A2-B2 after conversion of log values to MDRD GFR. Model L7 (Fig. 5.13 B2) showed smaller variance in the distribution of the samples compared with log creatinine and gender alone (R^2 = 0.915, Fig. 5.13 A2). These findings suggest that multiple regression equation using Model L7 could be used to predict MDRD GFR according to the regression equation.

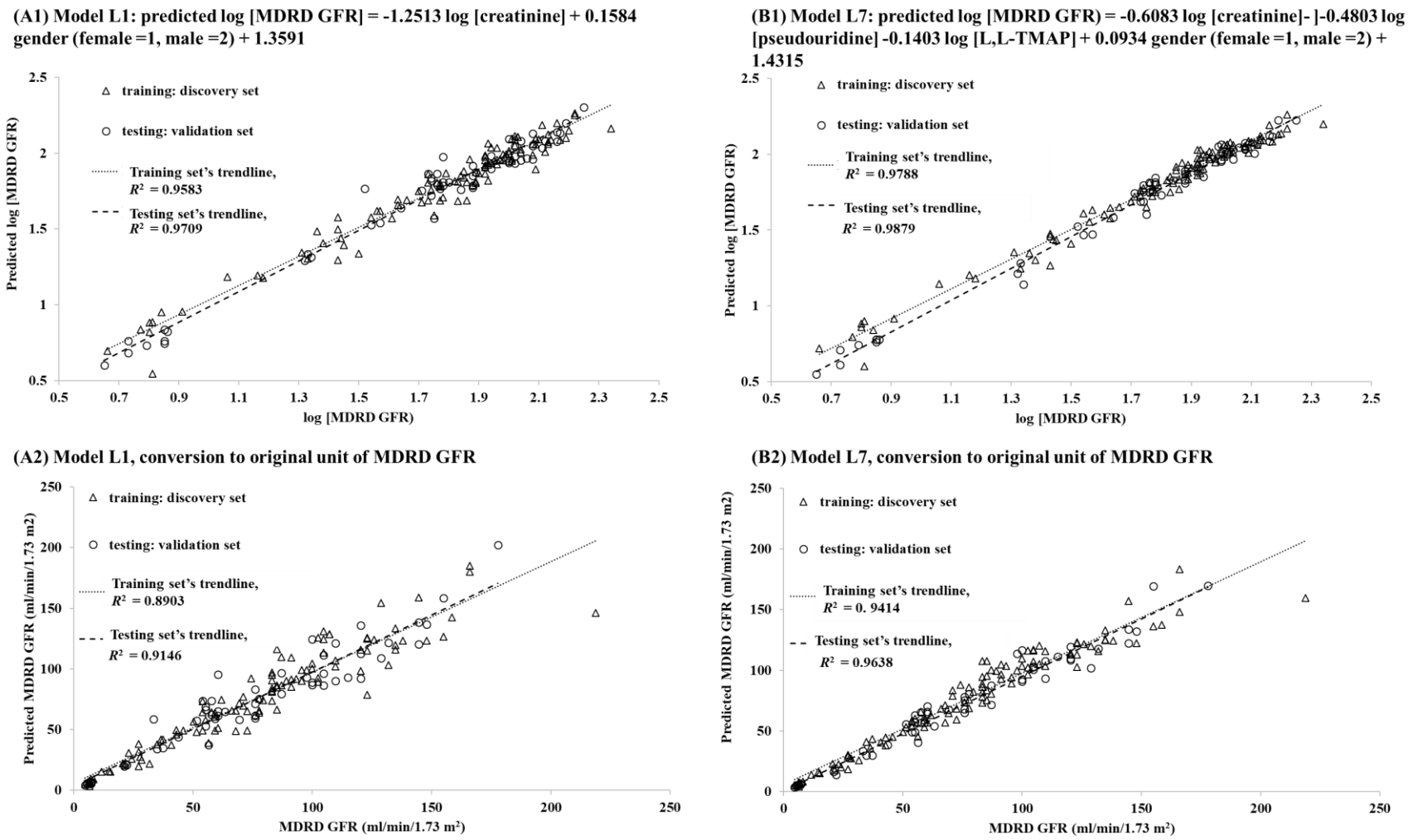


Fig. 5.13 Multivariate linear regressions of (A) Models L1 and (B) Model L7 with (1) log [MDRD GFR] and (2) conversion of log [MDRD GFR] trained with discovery set and tested with validation set.

5.3.9 Prediction of accuracy for differentiating CKD stages using random forest classification.

The use of a combinations of multiple biomarkers might be more sensitive and specific in classification of the kidney function of diabetic patients than the simple measurement of serum creatinine method. Herein, we built a non-parametric method, univariate and multivariate random forest classification model with equal-size sampling in two classes for training and testing models to distinguish different GFR thresholds. Single metabolite model candidates that were used for differentiation of diabetic patients at Stage 1a (GFR >120) from patients at Stage 1b-2 (GFR= 60-120), Stage 1a (GFR > 120) from Stages 1b-4 (GFR ≤ 120), Stage 1 (GFR ≥ 90) from Stages 2-4 (GFR < 90), Stages 3-4 (GFR < 60) from Stage 1-2 (GFR ≥ 60), Stage 4 (GFR < 30) to Stage 1-3 (GFR ≥ 30) are listed in Tables S5.7 - S5.10 (single metabolite models). The two-metabolite models were first built on using the following four biomarkers: 2-(*α*-D-Mannopyranosyl)-L-tryptophan, pseudouridine, succinyladenosine and L,L-TMAP with each other and in combination with MS-detected creatinine. Three-metabolite models were constructed using the four biomarkers with MS-detected creatinine. Their average AUC, sensitivity and specificity after 100 sampling tests using different GFR thresholds were listed in Tables S5.11 - S5.16. The tables also showed the results of commonly used logistic regression method for comparison, but random forest classification generally gained a better sensitivity result in ROC than logistic regression did. This suggested that the relative intensity of these biological metabolites did not have normal distribution and did not show a linear relationship with MDRD GFR.

Random forest classification results revealed that the four metabolites accurately differentiated patients at early Stage 1a from Stage 1b-4, Stage 1a from Stages 1b-2, Stage 1b from Stage 2, Stages 1 from Stages 2-4 of diabetic patients (average AUC > 0.700, Tables 5.17)**.** Their performances were comparable with that of MS-detected creatinine. Seven novel multiple-

metabolite models were more predictive of CKD stages than MS-detected creatinine alone (Table 5.18). Among them, Model 1 (MS-detected serum creatinine plus pseudouridine), Model 1 plus either L,L-TMAP (Model 5), succinyladenosine (Model 6) or 2-(α-D-Mannopyranosyl)-L-tryptophan (Model 7) predicted more accurately (Model 1, Model 5-7; average AUC > 0.800) than MS-detected creatinine alone when differentiating patients at Stage 1a (hyperfiltration) from Stages 1b-2 (MS-detected creatine : AUC =0.754 - 0.763) and when differentiating patients at Stage 1b (normal GFR range) from Stage 2 (mild decreased GFR) (MS-detected creatinine: average AUC = 0.600 - 0.680) (Table 5.18**).** Model 1, 5-7 could be detected in a single UPLC-MS platform, so they were selected as better candidates for prediction of kidney function.

Table 5.17 Average AUC of single metabolite models for differentiation between CKD stages in diabetic patients using random forest classification

Classification	Data set	MS-detected serum creatinine	Pseudouridine	L,L-TMAP	succinyladenosine	2-(α-D-mannopyranosyl)-L-tryptophan
Stage 1a vs Stages 1b-4	**Discovery**	0.849 ± 0.051	0.867 ± 0.072	0.849± 0.057	0.816 ± 0.065	0.811 ± 0.074
(GFR > 120 vs GFR ≤ 120)	**Validation**	0.879 ± 0.078	0.757 ± 0.078	0.774 ± 0.115	0.858 ± 0.066	0.780 ± 0.098
Stage 1a vs Stages 1b-2	**Discovery**	0.754 ± 0.080	0.867 ± 0.072	0.849 ± 0.057	0.816 ± 0.065	0.811 ± 0.074
(GFR > 120 vs GFR = 60-120)	**Validation**	0.763 ± 0.109	0.757 ± 0.078	0.774 ± 0.115	0.858 ± 0.066	0.781 ± 0.100
Stage 1b vs Stages 2	**Discovery**	0.681 ± 0.094	0.741 ± 0.079	0.680 ± 0.078	0.777 ± 0.069	0.707± 0.097
(GFR = 90-120 vs GFR = 60-89)	**Validation**	0.600 ± 0.170	0.982 ± 0.031	0.668 ± 0.157	0.889 ± 0.078	0.870 ± 0.092
Stages 1 vs Stages 2-4	**Discovery**	0.918 ± 0.028	0.935± 0.035	0.906 ± 0.038	0.899 ± 0.038	0.935 ± 0.026
(GFR ≥ 90 vs GFR < 90)	**Validation**	0.927 ± 0.048	0.994 ± 0.011	0.928 ± 0.054	0.972 ± 0.031	0.954 ± 0.034
Stages 1-2 vs Stages 3-4	**Discovery**	0.950 ± 0.025	0.985 ± 0.017	0.969 ± 0.020	0.958 ± 0.028	0.972 ± 0.025
(GFR ≥ 60 vs GFR < 60)	**Validation**	0.931 ± 0.039	0.949 ± 0.036	0.915 ± 0.046	0.912 ± 0.044	0.938 ± 0.035
Stages 1-3 vs Stages 4	**Discovery**	0.989 ± 0.009	0.993 ± 0.020	0.983 ± 0.034	0.993 ± 0.018	0.988 ± 0.024
(GFR ≥ 30 vs GFR < 30)	**Validation**	0.976 ± 0.081	0.968 ± 0.061	0.984 ± 0.054	0.955 ± 0.051	0.953± 0.056

Table 5.18 Average AUC of MS-detected serum creatinine and multiple-metabolite models for differentiation between CKD stages in diabetic patients using random forest classification in discovery and validation sets.

Classification	Data set	MS-detected serum creatinine	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
			MS-detected creatinine + pseudouridine	MS-detected creatinine + succinyladenosine	Pseudouridine + L,L-TMAP	succinyladenosine + pseudouridine	MS-detected creatinine+ pseudouridine + L,L-TMAP	MS-detected creatinine+ pseudouridine + succinyladenosine	MS-detected creatinine + Pseudouridine + 2-(α-D-Mannopyranosyl) -L-tryptophan
Stage 1a vs Stages 1b-4	**Discovery**	0.849 ± 0.051	0.930 ± 0.025	0.912 ± 0.029	0.942 ± 0.022	0.921 ± 0.049	0.945 ± 0.021	0.937 ± 0.023	0.933 ± 0.024
(GFR > 120 vs GFR ≤ 120)	**Validation**	0.879 ± 0.078	0.932 ± 0.034	0.940 ± 0.041	0.917 ± 0.032	0.935 ± 0.028	0.930 ± 0.035	0.937 ± 0.030	0.922 ± 0.034
Stage 1a vs Stages 1b-2	**Discovery**	0.754 ± 0.080	0.874 ± 0.047	0.840 ± 0.057	0.905 ± 0.037	0.877 ± 0.068	0.909 ± 0.035	0.888 ± 0.045	0.886 ± 0.041
(GFR > 120 vs GFR = 60-120)	**Validation**	0.763 ± 0.109	0.849 ± 0.063	0.869 ± 0.072	0.834 ± 0.091	0.870 ± 0.066	0.892 ± 0.061	0.850 ± 0.067	0.811 ± 0.089
Stage 1b vs Stages 2	**Discovery**	0.681 ± 0.094	0.810 ± 0.064	0.787 ± 0.070	0.776 ± 0.065	0.803 ± 0.065	0.807 ± 0.059	0.821 ± 0.059	0.831 ± 0.053
(GFR = 90-120 vs GFR = 60-89)	**Validation**	0.600 ± 0.170	0.967 ± 0.058	0.856 ± 0.080	0.966 ± 0.056	0.986 ± 0.030	0.951 ± 0.067	0.982 ± 0.034	0.962 ± 0.052
Stages 1 vs Stages 2-4	**Discovery**	0.918 ± 0.028	0.955 ± 0.016	0.932 ± 0.024	0.937 ± 0.023	0.942 ± 0.020	0.948 ± 0.021	0.957 ± 0.016	0.960 ± 0.016
(GFR ≥ 90 vs GFR < 90)	**Validation**	0.927 ± 0.048	0.995 ± 0.009	0.979 ± 0.018	0.999 ± 0.003	0.972 ± 0.022	0.993 ± 0.012	0.998 ± 0.004	0.997 ± 0.005
Stages 1-2 vs Stages 3-4	**Discovery**	0.950 ± 0.025	0.990 ± 0.006	0.973 ± 0.012	0.988 ± 0.008	0.982 ± 0.010	0.987 ± 0.008	0.984 ± 0.008	0.990 ± 0.007
(GFR ≥ 60 vs GFR < 60)	**Validation**	0.931 ± 0.039	0.955 ± 0.026	0.955 ± 0.029	0.959 ± 0.026	0.941 ± 0.032	0.968 ± 0.019	0.959 ± 0.027	0.962 ± 0.026
Stages 1-3 vs Stages 4	**Discovery**	0.989 ± 0.009	0.995 ± 0.014	0.999 ± 0.002	0.996 ± 0.012	0.999 ± 0.001	0.997 ± 0.006	0.999 ± 0.001	0.996 ± 0.009
(GFR ≥ 30 vs GFR < 30)	**Validation**	0.976 ± 0.081	0.992 ± 0.039	0.996 ± 0.020	0.990 ± 0.041	0.985 ± 0.034	0.980 ± 0.060	0.990 ± 0.030	0.988 ± 0.038

5.3.10 Correlations of endogenous metabolites to urinary markers: UACR and urinary β2-microglobulin

Correlations of 4-acetamidobutanoic acid, succinyladenosine, 2-(α-D-mannopyranosyl)-L-tryptophan, pseudouridine, 2-[3-(sulfooxy)phenyl]acetic acid, 5-methylthio-D-ribose, 2-hydroxyethanesulfonate, L,L-TMAP isomer, L,L-TMAP, ratio of kynurenine to tryptophan, sulfotyrosine, O-adipoylcarnitine, to UACR (Spearman R = 0.71-0.81 in the discovery set and 0.66-0.75 in the validation set) were better than those of MS-detected creatinine (Spearman R = 0.68 in discovery set and 0.65 in the validation set) in diabetic patients at Stages 1-4 **(**Table 5.19). The first five metabolites were better than serum cystatin C (Spearman R = 0.68 in validation set). However, the above metabolites did not show statistical consistency in the correlations between two cohorts among diabetic patients with UACR < 300 mg/g Cr. (p > 0.05).

Additionally, there were also significant positive correlations between proximal tubule injury indicator, urinary β2-microglobulin, and 13 MS-detected metabolites (including pseudouridine, L,L-TMAP, 2-(α-D-mannopyranosyl)-L-tryptophan, 4-acetamidobutanoic acid and succinyladenosine), two ratios of metabolites, clinically measured serum creatinine, serum cystatin C and UACR (Spearman R = 0.60-0.79) (Table 5.19**).** Pseudouridine (Spearman R = 0.74-0.79) and 4-acetamidobutanoic acid (Spearman R = 0.71-0.81) outperformed clinically measured serum creatinine (Spearman R = 0.67-0.74) in the Spearman rank correlation. For diabetic patients with UACR < 300 mg/g Cr., their relationship to urinary β2-microglobulin was weak or insignificant (absolute Spearman R < 0.40).

Details of the correlations of other metabolites with UACR and urinary β2-microglobulin could be seen in in Table S5.2 and Table S5.3, respectively.

Table 5.19 Spearman rank correlation of metabolites to UACR and urinary *β*2-microglobulin at different ranges with combination of discovery and validation sets. (Absolute Spearman *R* > 0.60).

Metabolites	Stage 0-4		Stage 1-4		Stage 1-3 (UACR < 300)	
	Discovery (n=128)	Validation (n=65)	Discovery (n=108)	Validation (n=55)	Discovery (n=63)	Validation (n=38)
Metabolites associated to UACR						
4-Acetamidobutanoic acid	0.7903	0.7244	0.8119	0.7076	0.4984	0.2306
Succinyladenosine	0.7799	0.6706	0.7947	0.6897	0.5159	0.1752
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.7643	0.7062	0.8009	0.7459	0.4483	0.2765
Pseudouridine	0.7514	0.7330	0.7936	0.7411	0.4494	0.2734
2-[3-(sulfooxy)phenyl] acetic acid	0.7311	0.6968	0.7540	0.6789	0.3972	0.1896
5-Methylthio-D-ribose	0.7124	0.6551	0.7626	0.6633	0.3946	0.1012
2-Hydroxyethanesulfonate	0.6971	0.6040	0.7137	0.7109	0.3094	0.2520
L,L-TMAP isomer	0.6951	0.6704	0.7308	0.6695	0.2957	0.1649
L,L-TMAP	0.6940	0.6628	0.7426	0.6756	0.3237	0.1566
Serum creatinine (mg/dL)	0.6880	0.6464	0.7367	0.6668	0.3251	0.2334
Ratio of kynurenine to tryptophan	0.6837	0.6625	0.7302	0.7114	0.3828	0.2518
Sulfotyrosine	0.6466	0.6986	0.7160	0.7321	0.3260	0.3785
MS-detected creatinine	0.6434	0.6349	0.6805	0.6569	0.2244	0.1737
Serum cystatin C (mg/L)		0.6549	-	0.6758	-	0.1923
Metabolites associated to urinary *β*2-microglobulin						
4-Acetamidobutanoic acid	0.6056	0.7871	0.6457	0.8056	0.2009	0.5240
5-Methylthio-D-ribose	0.6569	0.6962	0.7100	0.7152	0.3875	0.3131
L,L-TMAP isomer	0.6278	0.6460	0.6962	0.6814	0.2256	0.2824
Pseudouridine	0.6270	0.7696	0.6900	0.7933	0.3189	0.5088
UACR	0.6215	0.7000	0.6473	0.6938	0.1689	0.2709
MS-detected creatinine	0.6183	0.6565	0.6670	0.6853	0.203	0.2908
2-Hydroxyethanesulfonate	0.6179	0.5959	0.6684	0.6468	0.3413	0.1986
Serum Creatinine (mg/dL)	0.6102	0.6981	0.6583	0.7336	0.2493	0.3919
L,L-TMAP	0.6081	0.6557	0.6697	0.6809	0.1815	0.2534
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.6063	0.6925	0.6638	0.7314	0.2206	0.3510
O-Adipoylcarnitine	0.5962	0.7536	0.6181	0.7719	0.3647	0.4841
Ratio of kynurenine to tryptophan	0.5891	0.6671	0.6759	0.6979	0.2470	0.3280
Succinyladenosine	0.5884	0.7412	0.6353	0.7802	0.1537	0.4771
2-[3-(sulfooxy)phenyl] acetic acid	0.5676	0.6673	0.5882	0.6990	0.2838	0.2898
Ratio of pseudouridine to uridine	0.5661	0.7365	0.6380	0.7657	0.2235	0.441
Sulfotyrosine	0.5492	0.7136	0.6049	0.7680	0.2104	0.4638
Serum cystatin C (mg/L)	-	0.6365	-	0.6749	-	0.2779

5.3.11 Prediction of accuracy for distinguishing the different ranges of UACR using random forest classification

CKD albuminuria can be divided into three clinical stages: normoalbuminuria (UACR < 30 mg/g Cr.), microalbuminuria (UACR = 30-300 mg/g Cr.) and macroalbuminuria (UACR > 300 mg/g Cr.). Pseudouridine, 2-(α-D-mannopyranosyl)-L-tryptophan, succinyladenosine, 4-acetamidobutanoic acid, and sulfotyrosine (AUC = 0.711-0.837) are better markers to differentiate patients with normoalbuminuria (UACR < 30 mg/g Cr.) from later clinical stages of albuminuria than serum creatinine (AUC = 0.677-0.774) but L,L-TMAP and its isomer were not in the validation set (AUC = 0.667-0.678) (Table 5.20 and S5.17). For differentiating patients with UACR = 30-300 mg/g Cr. from those with UACR < 30 mg/g Cr., the five metabolites (AUC = 0.404- 0.701) did not have improved AUC compared with MS-detected creatinine (AUC =0.444 - 0.665) (Table S5.18). Four two-metabolite models that used sulfotyrosine plus either MS-detected creatinine, pseudouridine, 2-(α-D-mannopyranosyl)-L-tryptophan, succinyladenosine (AUC =0.668-0.706 in discovery set, 0.559-0.619 in validation set) had slightly improved AUC for differentiating patients with UACR = 30-300 mg/g Cr. from those with UACR < 30 mg/g Cr compared with MS-detected creatinine alone (AUC = 0.665 in discovery set, 0.444 in validation set) while the performances of the models that used sulfotyrosine plus L,L-TMAP (AUC = 0.653 in discovery set, 0.597 in validation set) and sulfotyrosine plus or 4-acetamidobutanoic acid (AUC = 0.660 in discovery set, 0.606 in validation set) were comparable with MS-detected creatinine alone (Table 5.21 and S5.19). The three-metabolite model that used these metabolites in combination did not have any better results (data not shown). Single biomarkers, pseudouridine, 2-(α-D-mannopyranosyl)-L-tryptophan, succinyladenosine, 4-acetamidobutanoic acid, L,L-TMAP and L,L-TMAP isomer (AUC = 0.893 - 0.957) slightly outperformed clinically measured serum creatinine (AUC = 0.884 - 0.887) and

MS-detected serum creatinine (AUC = 0.857 - 0.889) in differentiating patients at macroalbuminuria (UACR > 300 mg/g Cr.) from the rest (Tables 5.20, S5.20 - S5.21).

Table 5.20 AUC of significant metabolites for distinguishing CKD albuminuria categories at different UACR ranges using random forest classification.

UACR range (mg/g Cr.)	Dataset	AUC	Clinically measured serum creatinine	MS-detected creatinine	4-Acetamidobutanoic acid	Pseudouridine	2-(α-D-Mannopyranosyl) -L-tryptophan	L,L-TMAP	L,L-TMAP isomer	Succinyl-adenosine	Sulfotyrosine
≥ 30 vs <30	Discovery set	AUC	0.774 ± 0.046	0.773 ± 0.050	0.811 ± 0.047	0.837 ± 0.038	0.829 ± 0.036	0.784 ± 0.055	0.753 ± 0.054	0.823 ± 0.051	0.776 ± 0.051
		Sens	0.587 ± 0.078	0.670 ± 0.068	0.693 ± 0.069	0.713 ± 0.064	0.726 ± 0.054	0.604 ± 0.107	0.581 ± 0.118	0.655 ± 0.081	0.576 ± 0.097
		Spec	0.900 ± 0.106	0.826 ± 0.134	0.826 ± 0.166	0.932 ± 0.080	0.848 ± 0.138	0.865 ± 0.126	0.799 ± 0.137	0.895 ± 0.100	0.896 ± 0.112
	Validation set	AUC	0.705 ± 0.082	0.677 ± 0.077	0.738 ± 0.070	0.734 ± 0.057	0.711 ± 0.072	0.667 ± 0.080	0.678 ± 0.087	0.742 ± 0.076	0.823 ± 0.064
		Sens	0.437 ± 0.138	0.502 ± 0.118	0.563 ± 0.116	0.594 ± 0.098	0.624 ± 0.129	0.577 ± 0.181	0.4890 ± 0.171	0.551 ± 0.102	0.765 ± 0.115
		Spec	0.886 ± 0.168	0.874 ± 0.164	0.851 ± 0.167	0.848 ± 0.136	0.717 ± 0.220	0.660 ± 0.201	0.762 ± 0.176	0.886 ± 0.146	0.723 ± 0.174
30-300 vs < 30	Discovery set	AUC	0.633 ± 0.063	0.665 ± 0.077	0.676 ± 0.072	0.701 ± 0.075	0.685 ± 0.067	0.633 ± 0.074	0.565 ± 0.078	0.686 ± 0.076	0.630 ± 0.063
		Sens	0.916 ± 0.086	0.777 ± 0.157	0.907 ± 0.119	0.919 ± 0.116	0.881± 0.137	0.766 ± 0.135	0.553 ± 0.154	0.829 ± 0.137	0.793 ± 0.160
		Spec	0.304 ± 0.126	0.543 ± 0.135	0.397 ± 0.086	0.486 ± 0.106	0.472 ± 0.106	0.479 ± 0.133	0.575 ± 0.148	0.470 ± 0.130	0.418 ± 0.129
	Validation set	AUC	0.505 ± 0.119	0.444 ± 0.105	0.421 ± 0.102	0.428 ± 0.100	0.404 ± 0.100	0.452 ± 0.101	0.472 ± 0.112	0.464 ± 0.111	0.632 ± 0.125
		Sens	0.851 ± 0.177	0.434 ± 0.198	0.627 ± 0.182	0.448 ± 0.202	0.793 ± 0.183	0.854 ± 0.173	0.813 ± 0.180	0.377 ± 0.182	0.536 ± 0.186
		Spec	0.194 ± 0.150	0.479 ± 0.176	0.336 ± 0.151	0.426 ± 0.177	0.158 ± 0.127	0.090 ± 0.119	0.112 ± 0.145	0.580 ± 0.208	0.664 ± 0.206
> 300 vs 30-300	Discovery set	AUC	0.833 ± 0.059	0.820 ± 0.055	0.876 ± 0.054	0.855 ± 0.052	0.848 ± 0.056	0.845 ± 0.054	0.841 ± 0.049	0.840 ± 0.055	0.820 ± 0.044
		Sens	0.704 ± 0.107	0.753 ± 0.097	0.741 ± 0.142	0.747 ± 0.140	0.825 ± 0.148	0.702 ± 0.118	0.828 ± 0.097	0.753 ± 0.151	0.612 ± 0.101
		Spec	0.854 ± 0.102	0.854 ± 0.085	0.839 ± 0.109	0.777 ± 0.128	0.700 ± 0.118	0.852 ± 0.097	0.811 ± 0.074	0.739 ± 0.118	0.893 ± 0.099
	Validation set	AUC	0.860 ± 0.077	0.880 ± 0.081	0.915 ± 0.059	0.926 ± 0.056	0.948 ± 0.041	0.939 ± 0.043	0.892 ± 0.080	0.913 ± 0.057	0.921 ± 0.048
		Sens	0.701 ± 0.150	0.708 ± 0.182	0.833 ± 0.135	0.813 ± 0.156	0.873 ± 0.141	0.769 ± 0.139	0.761 ± 0.201	0.862 ± 0.137	0.804 ± 0.120
		Spec	0.899 ± 0.156	0.843 ± 0.160	0.901 ± 0.121	0.853 ± 0.138	0.844 ± 0.112	0.929 ± 0.135	0.804 ± 0.153	0.850 ± 0.097	0.870 ± 0.142
> 300 vs ≤ 300	Discovery set	AUC	0.887 ± 0.044	0.857 ± 0.059	0.922 ± 0.041	0.916 ± 0.038	0.915 ± 0.038	0.908 ± 0.034	0.895 ± 0.045	0.893 ± 0.035	0.873 ± 0.041
		Sens	0.773 ± 0.097	0.789 ± 0.091	0.909 ± 0.071	0.944 ± 0.057	0.945 ± 0.064	0.805 ± 0.117	0.825 ± 0.105	0.879 ± 0.080	0.833 ± 0.112
		Spec	0.878 ± 0.058	0.867 ± 0.082	0.792 ± 0.090	0.770 ± 0.062	0.808 ± 0.054	0.862 ± 0.068	0.858 ± 0.072	0.778 ± 0.078	0.750 ± 0.088
	Validation set	AUC	0.884 ± 0.068	0.889 ± 0.060	0.930 ± 0.051	0.946 ± 0.048	0.957 ± 0.045	0.941 ± 0.046	0.908 ± 0.052	0.939 ± 0.041	0.940 ± 0.043
		Sens	0.724 ± 0.136	0.689 ± 0.137	0.851 ± 0.122	0.849 ± 0.157	0.922 ± 0.106	0.821 ± 0.155	0.832 ± 0.170	0.871 ± 0.121	0.838 ± 0.121
		Spec	0.926 ± 0.135	0.920 ± 0.098	0.963 ± 0.040	0.908 ± 0.081	0.890 ± 0.062	0.851 ± 0.116	0.812 ± 0.094	0.947 ± 0.041	0.932 ± 0.071

Table 5.21 AUC of two-metabolite model using sulfotyrosine and other significant metabolites for distinguishing between UACR =30-300 and UACR <30 using random forest classification.

Two-metabolite models			Sulfotyrosine					
			MS-detected creatinine	Succinyladenosine	Pseudouridine	2-(α-D-Mannopyranosyl)-L-tryptophan	L,L-TMAP	4-Acetamidobutanoic acid
30-300 vs <30	Discovery set	AUC	0.671 ± 0.065	0.706 ± 0.076	0.691 ± 0.064	0.668 ± 0.063	0.653 ± 0.060	0.660 ± 0.068
		Sens	0.599 ± 0.193	0.663 ± 0.154	0.765 ± 0.165	0.744 ± 0.155	0.811 ± 0.131	0.852 ± 0.110
		Spec	0.653 ± 0.111	0.636 ± 0.142	0.549 ± 0.122	0.515 ± 0.129	0.450 ± 0.123	0.438 ± 0.103
	Validation set	AUC	0.619 ± 0.124	0.616 ± 0.108	0.578 ± 0.101	0.559 ± 0.101	0.597 ± 0.099	0.606 ± 0.100
		Sens	0.529± 0.169	0.490 ± 0.158	0.517 ± 0.157	0.472 ± 0.154	0.519 ± 0.141	0.499 ± 0.171
		Spec	0.664 ± 0.186	0.693 ± 0.162	0.648 ± 0.188	0.665 ± 0.169	0.678 ± 0.157	0.659 ± 0.180

5.4 Discussion

To our knowledge, this is the first clinical study of untargeted serum metabolomics in association with all stages of CKD among type 2 diabetic Chinese patients. Patients showed a clear progression of metabolic changes with the severity of kidney function from the early Stages 1-2 to later Stages 3-4 (Fig. 5.8).

5.4.1 Abnormal kidney size in different stages

This study examine early stages which many related studies (Sekula et al. 2016; Niewczas et al. 2017; Velenosi et al. 2019; Luo et al. 2019) excluded. In this study, the observation of enlargement of kidney size, followed by gradual diminishment in the present study **(**Fig.1.2**)** supported that diabetic patients starting from Stage 1a were most likely to have abnormal kidney hypertrophy and enlargement, followed with gradual renal atrophy and volume depletion from Stages 2 to 4 (Rigalleau et al. 2010; Jovanović et al. 2013). Experimental studies suggested that the renal hypertrophy might be due to the inhibition of kidney AMP-activated protein kinase by high glucose level, and that it is followed by glomerular hyperfiltration, and tubulo-glomerular feedback mechanism (Rigalleau et al. 2010).

The degeneration in the later stage might lead to increases in renal vascular resistance and blood pressure as shown by the highest RRI and SBP in Stage 4. The elevation of RRI in this study supported that the increasing kidney arterial damage and renal arterial resistance with the rise in severity of CKD in diabetic patients (Afsar and Elsurer 2017). Research found that strict insulin treatment could reduce kidney size to normal size in diabetic patients (Mogensen and Andersen 1975). Thus, early intervention in Stage 1a patients is of high importance.

5.4.2 1,5-Anhydro-D-glucitol level monitoring as a marker of type 2 diabetes at early Stage 1

Correlation analysis showed that the FBG and serum hemoglobin A1c levels, two clinical glycemic markers, did not show significant changes along with MDRD GFR but were highly correlated to the serum level of 1,5-anhydro-D-glucitol among diabetic patients within each stage of Stages 1-3 (Table 5.12), especially for hemoglobin A1c (Fig. 5.10). Patients with a high level of serum hemoglobin A1c might have a high level of serum 1,5-anhydro-D-glucitol.

Previous studies had also indicated similar findings that serum 1,5- anhydro-D-glucitol levels were little affected by GFR (Yamanouchi et al. 1988) and were not altered in diabetic patients at Stage 4 (Kim et al. 2012). 1,5-anhydro-D-glucitol is a dietary 1-deoxyglucose (Yamanouchi et al. 1992) that glucose competes with it in the binding sites of transporters in the renal proximal tubules and blocks tubular reabsorption of 1,5-anhydro-D-glucitol when blood glucose is higher than the renal threshold (Yamanouchi et al. 1996; Yamanouchi and Akanuma 1994; Kametani et al. 1987; Dworacka et al. 2002), leading to a rise in urinary loss and a reduction of blood concentration of 1,5-anhydro-D-glucitol. This was more obvious during the early stages of CKD probably because of tubular hyper-reabsorption of glucose (Anders et al. 2018). The serum level of 1,5-anhydro-D-glucitol would be a more sensitive clinical glycemic markers at early Stage 1 and might be mainly affected by the function of renal proximal tubules.

5.4.3 Potential biomarkers for prediction of MDRD GFR and UACR

It has been well reported that serum creatinine is an inaccurate indicator of renal function. After investigating the correlation between metabolites and GFR, using ROC

for diagnosis of patients at different stages of CKD and prediction for MDRD GFR in this study, succinyladenosine, pseudouridine, L,L-TMAP and 2-(α-D-mannopyranosyl)-L-tryptophan, that were identified with commercially available standards, had a high correlation with GFR and improved the accuracy in prediction.

Using ROC for diagnosis of patients at different CKD albuminuria, using a single metabolite model or two-metabolite models with the above four biomarkers, 4-acetamidobutanoic acid and sulfotyrosine did not have obvious improvement on AUC compared with MS-detected serum creatinine, thus 4-acetamidobutanoic acid and sulfotyrosine were not discussed here (Table 5.20).

Pseudouridine is a C-glycosidic derivative of uridine, a modified nucleoside found in RNA. The blood level of pseudouridine was elevated in the patients with diabetic kidney injuries and pseudouridine was suggested as a filtration marker for kidney function (Niewczas et al. 2017; Dzúrik et al. 1992; Titan et al. 2019; Sekula et al. 2016).

L,L-TMAP, a recently identified metabolite that was newly identified by Velenosi's team (Velenosi et al. 2019). L,L-TMAP might be produced from the degradation of myosin light chain proteins (Henry et al. 1985). L,L-TMAP was reported to outperform serum creatinine in accurately identifying ESRD patients with MDRD GFR < 30 mL/min per 1.73 m^2 (Velenosi et al. 2019).

Succinyladenosine is a succinylpurine found in urine (Monostori et al. 2019), serum and the liver (Han et al. 2019). Succinyladenosine is traditionally used as a diagnostic target of patients with adenylosuccinate lyase deficiency (Hartmann et al. 2006; Jurecka et al. 2015; Monostori et al. 2019). Succinyladenosine has been reported in adenine-induced rats with kidney failure (Kobayashi et al. 2014) and mentioned in one cohort

studying the relationship with the end stage-renal disease in patients with proteinuria. It has not been mentioned as a biomarker to differentiate CKD patients in different stages, especially in the early stages, and there was a lack of researches related to its correlation to other clinical markers like UACR and β2-microglobulin.

2-(α-D-mannopyranosyl)-L-tryptophan is formed by *C*-mannosylation of an α-mannose to the indole C2 carbon atom of a tryptophan residue. 2-(α-D-Mannopyranosyl)-L-tryptophan is found in human serum, urine, and cerebrospinal fluid as well as in various food products (Gutsche et al. 1999).

They were remarkably associated with severity of kidney function in diabetic patients. The Spearman rank correlations remained strong after controlling common confounding factors (Table 5.14). Compared with serum creatinine and cystatin C, their associations with MDRD GFR were comparably high and more sensitive in changes with the advantage of being independent or less dependent on gender (Fig. 5.12 and Table 5.16). Future application of the first three metabolites in the estimation of GFR might simplify calculation without concerning a subject's gender.

This study had concordant results with four other researches in the association of pseudouridine, 2-(α-D-mannopyranosyl)-L-tryptophan (Sekula et al. 2016; Niewczas et al. 2017), L,L-TMAP (Velenosi et al. 2019), succinyladenosine (Luo et al. 2019) with renal function decline in patients at later stage (GFR < 60). The results in this study suggested that these biomarkers were also useful to identify early stages of renal functions (GFR > 60), which were not investigated in these previous studies. Only one study identified pseudouridine and 2-(α-D-mannopyranosyl)-L-tryptophan as predictive markers in differentiating individuals in GFR < 60 from GFR > 60 (Solini et al. 2016). Succinyladenosine has not been reported in diabetic patients with chronic

kidney diseases at the early stages (Stages 1-2). Our study extended these findings to show that the association of the four metabolites with GFR also occurred at early Stages 1-2 (GFR > 60) of diabetic patients (Table 5.14**).** The four biomarkers comparably had accurate differentiate patients at early Stages as MS-detected creatinine did **(**Table 5.17**).** Pseudouridine, 2-(*α*-D-mannopyranosyl)-L-tryptophan and succinyladenosine could differentiate between patients with UACR > 30 and patients with UACR ≤ 30 better than clinically measured serum creatinine (Table 5.20).

Prediction using combination of these metabolites with MS-detected serum creatinine revealed four novel multiple-metabolite models, Model 1 (MS-detected serum creatinine plus pseudouridine), Model 1 plus either L,L-TMAP, succinyladenosine or 2-(α-D-Mannopyranosyl)-L-tryptophan as being more predictive of CKD stages than MS-detected creatinine alone (Table 5.18) which might improve the accuracy of serum creatine alone. Stepwise multivariate linear regression analyses selected Model L7 (log [MS-detected serum creatinine], log [pseudouridine], log [L,L-TMAP] plus gender) as the best regression equation to predict log [MDRD GFR] (Table 5.16).

5.4.4 Correlations of biomarkers with UACR and urinary *β*2-microglobulin to renal tubular markers

Interestingly, the levels of pseudouridine, L,L-TMAP, 2-(*α*-D-mannopyranosyl)-L-tryptophan and succinyladenosine were not only increased significantly with the decline in GFR but were also positively correlated with UACR, urinary *β*2-microglobulin and were negatively correlated with BSA-related renal volume and renal resistive index **(**Fig. 5.11**).** Importantly, such correlation performances with these kidney-function related markers were generally better than that of serum creatinine and cystatin C.

The current results expanded their association with GFR to other clinical renal markers that might reflect other kidney physiological damages. This is because urinary β2-microglobulin has a low molecular weight and about 95% is filtered through normal glomerulus freely and then almost completely reabsorbed in proximal tubules (Schardijn and Statius van Eps 1987). When the renal tubules become damaged or diseased, the concentrations of renal tubular reabsorption markers in urine would increase due to the decreased ability to reabsorb these proteins. 55% of normalbuminuric and 57% of microalbuminuric type 2 diabetic patients were found to have elevated β2-microglobulin as an indicator of diabetic tubulopathy in the early stages of CKD (Tanaka et al. 1989).

5.4.5 Biomarkers might be sourced from posttranslational modifications of nucleosides or proteins

The levels of pseudouridine, L,L-TMAP, 2-(α-D-mannopyranosyl)-L-tryptophan, succinyladenosine and sulfotyrosine were also found to be highly correlated with each other and might have similarity in chemical structures or sources. This study suggested that they might be sourced from posttranslational modifications of nucleosides or proteins. A previous study proposed that the elevation in serum pseudouridine and 2-(α-D-mannopyranosyl)-L-tryptophan might be atttributed to the increased posttranslational modifications in tissues of diabetic patients (Niewczas et al. 2017). Pseudouridine and 2-(α-D-mannopyranosyl)-L-tryptophan were *C*-glycosylated uracil and tryptophan, respectively (Sekula et al. 2016). Additionally, sulfotyrosine, that have a high correlation with UACR, is *O*-sulfation of protein tyrosine residues, a post-translational modification facilitated by Golgi tyrosyl-protein sulfotransferases in the secreted and transmembrane proteins (Ouyang, Lane, and Moore 1998; Bundgaard and Rehfeld 2013). Noticeably, the other two metabolites, succinyladenosine and L,L-

TMAP were derived from succinylation of adenosine and methylation of alanylproline, respectively. These modifications might arise from chronic oxidative stress and/or carbon stress induced by cellular diabetic redox imbalance (Lyons and Jenkins 1997; Zheng et al. 2016). Persistent hyperglycemia would cause redox imbalance between NADH and NAD^+, resulting overproduction of reactive oxygen species to attack proteins and DNA and induce protein and nucleoside modifications (Yan 2014) .

5.4.6 Uremic retention solutes quicken CKD progression

Most of the altered metabolites that were inversely correlated to GFR in this study, including the selected biomarkers have previously been reported as uremic retention solutes (Mair et al. 2018; Velenosi et al. 2019). Uremic retention solutes were accumulated in the circulation and tissues during the development of CKD, directly or indirectly owing to a lack of renal clearance by kidney degradation (Vanholder and De Smet 1999). Uremic retention solutes have been shown to quicken CKD progression by causing oxidative stress, renal fibrosis, functional impairment on kidney and cardiovascular system and other organs (Mair, Sirich, and Meyer 2018; Lisowska-Myjak 2014). Indoxyl sulfate and *p*-cresol sulfate are two well-known protein-bound uremic solutes that depend on tubular secretion for elimination. Their accumulation induced the progression of cardiovascular and renal diseases and damaged bone metabolism in CKD patients (Liu, Tomino, and Lu 2018). At early stages of diabetes, uremic solute accumulation in the circulation would be a consequence of oxidative stress induced by diabetic redox imbalance but also played a role in quickening CKD progression simultaneously.

5.4.7 Pathway analysis suggested dysregulation of amino acid and related pathways

Evaluating pathway changes provides a systematic biology approach that can be used to understand the metabolic effects of CKD in diabetic pathobiology. In the present study, alternations were identified in the pathways that suggested dysregulation of amino acid and related pathways, especially tryptophan and phenylalanine metabolisms **(**Fig. 5.9). In the tryptophan metabolism, the rise in IDO (indoleamine 2,3-dioxygenase) activity as estimated by the kynurenine to tryptophan ratio was in line with the result of the Family Investigation of Nephropathy in Diabetes study (Debnath et al. 2017). The drop in the tyrosine to phenylalanine ratio in the phenylalanine metabolism was in line with a research that indicated an impairment in the conversion of phenylalanine to tyrosine in chronic kidney failure (Kopple 2007). In particular, the observation of increased IDO activity in the tryptophan metabolisms was in line with several CKD studies (Karu et al. 2016; Debnath et al. 2017; Zhang, Ruan, et al. 2017). Noticeably, the increased IDO activity was reported to be activated by the inflammatory signals to increase the biosynthesis of NAD^+ for promoting immune tolerance and preventing inflammatory responses (Sorgdrager et al. 2019). The positive correlation of tryptophan with GFR and negative correlation of estimated IDO activity (the kynurenine to tryptophan ratio) implied that tryptophan catabolism got more active at later stages of CKD.

5.4.8 Strength of this study

One major strength of this study is that the metabolite alternations in one cohort were reproducible in another independent cohort conducted at different time, a step that was usually absent in other studies (Colhoun and Marcovecchio 2018). Another strength

was that this study covered normal individuals and diabetic patient at all CKD stages, especially those at early Stages with GFR up to 60 - 150, suffering from glomerular hyperfiltration, mild and moderate renal damage, which other studies always lacked. More importantly, to our knowledge, succinyladenosine was the only biomarker that was first reported to have association with GFR > 60 in diabetic patients in this study. The correlation of succinyladenosine with GFR was first found to have no independence on gender whereas gender had smaller effect on L,L-TMAP than serum creatinine in relationship with with GFR.

5.4.9 Limitation

There were limitations to this study. First, this study only focused on type 2 diabetic Chinese patients, so it might not be generalizable to the other racial populations and type 1 diabetic patients. Second, samples size in each group may not be enough to reflect entire population. Third, the relative intensities of the UPLC-MS detected biomarkers were used in the analysis of this study rather than their actual concentrations, but their stabilities monitored by QC samples were high. Thus, the relative intensity would be compared among samples. Fourth, this was a cross-sectional study and the serum samples were only collected at a single time point, so this was unable to monitor the trends of metabolite change over time in the same individual. A longitudinal study of the disease progression with repeated measures over prolonged period of time would certainly further validate the findings reported here.

5.5 Conclusion

This study of two independent cohorts demonstrated the robust correlation of serum succinyladenosine, pseudouridine and 2-(α-D-mannopyranosyl)-L-tryptophan and L,L-TMAP with MDRD GFR in type 2 diabetic patients, which were observed early at

MDRD GFR ≥ 60. Gender as a covariate had no effect in the linear regression using either succinyladenosine, pseudouridine or 2-(α-D-mannopyranosyl)-L-tryptophan. Furthermore, prediction of MDRD GFR and CKD stages using these four biomarkers alone were comparable with serum creatinine. Combination of these biomarkers with serum creatinine improved the diagnostic ability of the serum creatinine alone in differentiation of early CKD stages. Stepwise multivariate linear regression selected pseudourdine and L,L-TMAP as significant and independent predictors, adding in the regression equation of MS-detected creatinine and gender in prediction of MDRD GFR. A longitudinal study and targeted analysis of these biomarker are warranted to replicate the respective changes along with the CKD progression for the same diabetic individuals and validate the associations reported in this study.

Chapter 6 Overall conclusion and prospects

6.1 Overall conclusion

Untargeted metabolomics has emerged as a promising approach to address important research questions. Untargeted metabolomics with MS plays key roles in applications of underlying pathologic mechanism, biomarker and drug discovery. Unexpected changes in metabolite levels can be exploited to uncover novel mechanisms related to health and disease.

Metabolomics has been used to examine the changes of metabolites in three cases with increasing complexity. The first study is a classic serum untargeted metabolic profiling for investigation of cholesterol-lowering effects of piceatannol. Validation of the results were done with a quantitative analysis and subsequent protein expression of the associated enzymes discovered from the pathways of biomarkers found.

In the second study, more complex phytochemicals sourced from a Chinese herbal medicine that also contained piceatannol and its glucoside was studied in a more serious metabolic disorder, NAFLD. The herbal extract contains more complex chemicals and it may act on multiple sites and targets in the body to bring out the protective effect. In order to cover more metabolites, this study includes serum metabolomics and liver lipidomics. The two different sample preparations and analytical methods allow profiling of polar small molecules and less polar lipids to maximize the coverage of metabolites. In this study, bio-transformed resveratrol and/or polydatin derived metabolites via microbial transformation were unveiled with the help of multivariate statistics without any previous hypothesis.

The last cross-sectional study is a large-scale serum metabolomics of about 200 human serum samples with five stages of CKD in diabetes. This study involved plenty of data treatments as five groups were compared, and two independent cohorts were collected to enhance the robustness of the biomarker discovery. Clinical biochemical data were incorporated into MS-detected small molecule candidates for the differentiation between CKD stages and improvement of GFR estimation. The linear regression results showed there was a high relationship between MDRD GFR and MS-detected succinyladenosine, pseudouridine or 2-(α-D-mannopyranosyl)-L-tryptophan with the advantage of being independent on gender compared with the routine marker, serum creatinine. The finding proved the combination of pseudouridine and L,L-TMAP with serum creatinine and gender improved the accuracy in differentiation at early stages of CKD.

Untargeted metabolomics in this study provided new research directions in drug discovery and gave hints of drug target sites for drug development. Resveratrol bio-transformed metabolites in PCRR-retreated rat serum explained the reason of low bioavailability of resveratrol via oral administration and they might be candidates modified as new medicines. Moreover, piceatannol treatment on HFD-fed rats showed that Δ5-desaturase activity in fatty acid synthesis were upregulated in response to the influx of fat from HFD intake. The CYP7A1 activity was elevated after treatment with PCRR water extract to remove hepatic cholesterols. Chronic inflammation in diabetic patients activated kynurenine pathways in response to the inflammatory signals. However, the activated pathways might also generate excess products into toxic levels in the biological system such as kynurenine whose metabolic products (3-hydroxykynurenine and quinolinic acid) in high levels may cause neurotoxicity (Parrott et al. 2016). Since the three pathways comprised multiple functional and important

enzymes and receptors for regulation, additional studies on the changes of enzyme protein expression and receptor genomic expression are warranted to figure out the possible drug targets for development of new medicines. Meanwhile, their possible adverse effects of drug actions should also be examined.

The action of gut microbiota was commonly observed after investigation of the altered metabolites in the three chronic disorders and treatments. This could be explained by the imbalance of gut microbiota in association with chronic diseases (Jackson et al. 2018; Durack and Lynch 2019). Their actions were observed in our metabolomics results: they bio-transformed primary bile acids into secondary bile acids in the gut; they deconjugated glycine and taurine-conjugated bile acids; they bio-transformed resveratrol and/or polydatin from PCRR extract into dihydroresveratrol and 3,4-dihydrobibenzyl by reduction of the carbon-carbon double bond of resveratrol; they converted dietary tryptophan and tyrosine into indole and cresol respectively, and then their derivatives such as 3-indoleacetic acid, indoxyl sulfate and *p*-cresol sulfate. There were plenty of direct interactions between gut microbiota and diets, medications and circulating metabolites. Thus, comparison of the relative abundance and genera of intestinal bacteria from the feces between groups could be achieved by the extraction of fecal microbial DNA. This would reveal new pathways on pathogenesis of chronic diseases. Treatment could be achieved by balancing targeted intestinal microflora's abundance or composition.

It is quite clear that more sophisticated data processing and multivariate techniques ae important for complex samples. The first two experiments analysed serum and tissue samples from rats. These animals were kept in well controlled facilities with consistent diets to reduce confounding factors. Thus, PLS-DA and OPLS-DA are already good enough to provide clear results. In the third experiment, when human subjects were

involved, there were more factors that could not be controlled. More sophisticated data processing methods were used to uncover and validate the differences observed.

6.2 Future Prospects

These experiments demonstrated an untargeted metabolomics approach is a powerful technique for an overview of the metabolic status, discovery of the biological pathways associated to disease pathogenesis and identifying the biological actions of treatment. Yet, there are still some future works for confirmation and validation of the findings.

Outliers were observed in the animal studies and clinical cohorts but identification and removal of the outlier in each group caused a lot of works and reduced the sample size. A larger sample size would certainly improve the robustness of each study and weaken the effect of outliers in each group. The samples in each experiment conducted were only collected at a single time point. A longitudinal study may improve our understanding of the causes and effects of the diseases and therapy, such as collection of samples before and after treatments as well as a long term follow-up of the same subjects for disease progression.

Apart from the bile acids in the first study, only relative quantification of metabolites was performed rather than actual concentration. The high stability in the data acquisition was reflected by QC samples, suggesting that the relative intensities of the metabolites could be used as dependable comparison among samples. To get the absolute content for future clinical application, the main challenges was the unavailability of each biomarker's isotopic-labelled internal standard. This requires more explorative experiments to look for alternative internal standards that may have similar chemical properties as the biomarkers.

Another difficulty is the identification of novel compounds, like the transformed metabolites discovered in PCRR-treated rat serum in Chapter 4. Their identifies could only be confirmed by mass fragmentation due to the lack of commercially available reference standards. More experiments are needed to confirm its chemical identification.

In the CKD study, protein or nucleoside modifications by the attack of overproduced reactive oxygen species in diabetic patients was proposed but their underlying mechanism was not validated in the current CKD clinical study. More investigation on the linkage of the reactive oxygen species with the modified metabolites was still warranted.

In addition, studies of various samples types from the subjects, like gut microbiota, urine and faeces, would broaden the spectra of metabolites and discover inter-organ crosstalk. In particular, the changes of composition and contents of gut microbiota in the faeces due to the three chronic disorders and drug treatments might help explain the difference of circulating metabolic changes between controls, disease models and treatment groups. Integration of metabolomics and the other omics tools such as proteomics and genomics may improve the interpretation of the discover at different levels of omics data and generate new insights. These combination of sample types and omics tools may give a comprehensive view of the changes of each single molecular process.

Appendix I

Table S3.1 The variation of intensity of the QC samples (A) negative ESI mode in UPLC-MS; (B) positive ESI mode in UPLC-QTOF-MS

Table S3.1A

	L-Tryptophan	Taurocholate	Taurodeoxy-cholate	Cholate	Chenodeoxy-cholate	LysoPC (18:2)	LysoPC (18:1)	LysoPC (20:2)	LysoPC (20:0)	Stearate
Retention time (min	3.16	4.06	4.79	5.19	6.77	7.4	8.62	9.09	.2	13.7
Measured m/z	203.0812	514.2832	498.2882	407.279	391.2842	564.33	566.3455	592.3611	596.3924	283.263
QC1	12250	10771	31848	30274	9775	83436	59307	5878	4276	2186
QC2	12488	11166	33649	30699	9943	83488	71877	5590	4114	2335
QC3	12267	10147	31918	29817	9474	101152	79956	5061	3528	2215
QC4	11447	9880	31940	28503	9000	73454	61078	4638	3211	2069
QC5	11262	9555	31301	28486	8735	73007	50262	5199	2848	1988
QC6	12647	7735	29932	27103	8583	74342	58353	6440	5160	1750
QC7	11913	8072	29814	28179	8717	73407	55552	5877	3874	1829
QC8	12007	8621	32474	28755	9079	75912	55791	5255	3474	1860
QC9	11591	8343	32264	28400	8759	72313	52709	5705	4074	1806
QC10	12002	8311	32416	27396	8744	71854	55293	5163	3806	1787
Mean of peak area	11987	9260	31756	28761	9081	78237	60018	5481	3836	1983
Standard deviation	449	1205	1163	1173	483	9127	9137	518	639	208
CV (%)	3.75	13.01	3.66	4.08	5.32	11.67	15.22	9.45	16.66	10.49

Table S3.1B

	L-Tryptophan	Phytosphingo-sine	Cholic acid	Chenodeoxy-cholic acid	LysoPC (18:2)	LysoPC (20:4)	LysoPC (18:1)	LysoPC (18:0)	Palmitic amide	13-Docosen-amide
Retention time (min)	3.16	4.99	5.19	6.76	7.4	7.51	8.61	9.75	11.09	14.39
Measured m/z	205.0983	318.3007	355.2636	357.2793	520.3401	544.34	522.3555	524.3712	256.2643	338.3422
QC1	2318	27808	29718	12278	244547	180388	138260	391141	20072	131534
QC2	2542	28289	29016	12520	229777	170500	126757	357984	14587	125434
QC3	2457	25610	26869	11824	204855	152653	113120	320687	11332	121769
QC4	2360	23909	24515	10809	185431	134544	105649	287219	16834	116217
QC5	2260	22968	24200	10757	182025	133367	98453	275012	16515	114107
QC6	2944	27927	26489	11661	245094	192316	143677	397220	18411	159475
QC7	2933	26670	25759	11422	233686	183302	136330	364942	17892	148929
QC8	3005	26658	24845	10956	218304	170244	124738	342377	18886	146047
QC9	2506	22593	21218	8900	183155	144834	107954	283546	16064	120597
QC10	2121	21405	19527	7853	172423	136774	101015	254091	14131	113248
Mean of peak area	2545	25384	25216	10898	209930	159892	119595	327422	16472	129736
Standard deviation	312	2489	3151	1474	27872	22108	16500	50781	2589	16284
CV (%)	12.27	9.8	12.5	13.52	13.28	13.83	13.8	15.51	15.72	12.55

Table S3.2 The variation of peak area ratio of EFA of the QC serum samples injected between study samples in GC-MS

EFA	EC14:0	EC15:0	EC16:0	EC16:1n7c	EC17:0	EC18:0	EC18:1n9c	EC18:2n6c	EC18:3n6c	EC18:3n3c	EC20:0	EC20:2n6c	EC20:3n6c	EC20:4n6c	EC20:5n3c	EC22:6n3c
Q1	0.4041	0.2074	0.6872	0.1863	0.2163	0.4630	0.6632	0.1771	0.1722	0.1835	0.4633	0.1773	0.1700	0.1378	0.1693	0.1428
Q2	0.3926	0.2056	0.6669	0.1870	0.2171	0.4579	0.6618	0.1875	0.1794	0.1874	0.4754	0.1871	0.1798	0.1486	0.1821	0.1599
Q3	0.4035	0.2092	0.6782	0.1896	0.2159	0.4657	0.6630	0.1848	0.1823	0.1889	0.4702	0.1870	0.1782	0.1533	0.1803	0.1661
Q4	0.3929	0.2017	0.6637	0.1805	0.2109	0.4554	0.6526	0.1746	0.1696	0.1785	0.4541	0.1714	0.1646	0.1425	0.1666	0.1437
Q5	0.3909	0.2021	0.6641	0.1839	0.2144	0.4615	0.6477	0.1820	0.1781	0.1830	0.4634	0.1834	0.1748	0.1519	0.1716	0.1513
Q6	0.3847	0.2005	0.6670	0.1848	0.2149	0.4584	0.6532	0.1836	0.1834	0.1863	0.4659	0.1872	0.1851	0.1587	0.1798	0.1653
Q7	0.3826	0.2017	0.6633	0.1841	0.2118	0.4593	0.6546	0.1947	0.1768	0.1843	0.4665	0.1872	0.1768	0.1624	0.1809	0.1555
Q8	0.3872	0.2005	0.6726	0.1803	0.2111	0.4792	0.6601	0.2015	0.1763	0.1809	0.4629	0.1856	0.1765	0.1706	0.1753	0.1625
Q9	0.3863	0.1993	0.6631	0.1799	0.2139	0.4565	0.6491	0.1873	0.1711	0.1812	0.4629	0.1823	0.1750	0.1507	0.1772	0.1571
Q10	0.3857	0.2022	0.6630	0.1808	0.2106	0.4523	0.6519	0.1827	0.1756	0.1856	0.4630	0.1816	0.1729	0.1517	0.1783	0.1609
Q11	0.3965	0.2044	0.6707	0.1858	0.2140	0.4549	0.6474	0.1826	0.1746	0.1833	0.4641	0.1840	0.1753	0.1477	0.1757	0.1564
Q12	0.3867	0.2028	0.6627	0.1829	0.2107	0.4619	0.6549	0.1874	0.1756	0.1832	0.4684	0.1874	0.1790	0.1543	0.1786	0.1575
Q13	0.3870	0.2015	0.6815	0.1862	0.2153	0.4755	0.6675	0.2015	0.1774	0.1886	0.4711	0.1906	0.1817	0.1674	0.1772	0.1657
Q14	0.4024	0.2054	0.6818	0.1873	0.2169	0.4608	0.6643	0.1807	0.1818	0.1899	0.4785	0.1848	0.1757	0.1526	0.1813	0.1612
Q15	0.3942	0.2059	0.6712	0.1848	0.2158	0.4608	0.6611	0.1799	0.1784	0.1848	0.4755	0.1845	0.1781	0.1479	0.1707	0.1612
Q16	0.3956	0.2043	0.6840	0.1874	0.2171	0.4734	0.6733	0.1940	0.1774	0.1871	0.4683	0.1834	0.1769	0.1622	0.1756	0.1670
Q17	0.4032	0.2099	0.6791	0.1885	0.2172	0.4672	0.6661	0.1761	0.1734	0.1849	0.4735	0.1820	0.1746	0.1506	0.1695	0.1587
Q18	0.3951	0.2064	0.6728	0.1855	0.2165	0.4583	0.6621	0.1830	0.1803	0.1889	0.4697	0.1905	0.1791	0.1532	0.1785	0.1647
Q19	0.3960	0.2070	0.6781	0.1853	0.2155	0.4631	0.6664	0.1812	0.1788	0.1899	0.4708	0.1883	0.1803	0.1542	0.1830	0.1634
Q20	0.3974	0.2053	0.6705	0.1864	0.2156	0.4612	0.6526	0.1860	0.1808	0.1872	0.4672	0.1933	0.1841	0.1522	0.1777	0.1500
Mean of peak area	0.3932	0.2041	0.6721	0.1849	0.2146	0.4623	0.6586	0.1854	0.1772	0.1854	0.4677	0.1849	0.1769	0.1535	0.1764	0.1586
Standard deviation	0.007	0.003	0.008	0.003	0.002	0.007	0.007	0.008	0.004	0.003	0.006	0.005	0.005	0.008	0.005	0.007
CV (%)	1.72	1.46	1.18	1.50	1.09	1.51	1.12	4.05	2.10	1.72	1.21	2.61	2.61	5.06	2.65	4.42

Table S3.3 The variation of concentration of the QC serum samples injected between study samples in UPLC-QQQ-MS/MS

Standards	QC1	QC2	QC3	QC4	QC5	QC6	QC7	CV (%)
Ursodeoxycholate	95.27	91.48	108.48	97.72	89.28	96.48	92.52	6.56
Hyodeoxycholate	678.71	746.84	743.44	729.68	669.20	745.24	723.08	4.50
Chenodeoxycholate	362.69	352.16	371.84	394.96	359.52	373.44	404.92	5.13
Lithocholate	< LOD	< LOD	< LOD	< LOD	< LOD	< LOD	< LOD	N/A
Deoxycholate	623.11	612.80	640.64	628.40	626.36	652.52	653.44	2.43
***α*-Muricholate**	121.03	120.04	118.00	120.44	128.48	132.72	133.88	5.28
***β*-Muricholate**	220.57	238.16	235.96	242.84	231.96	223.84	249.52	4.35
Cholate	956.44	967.84	989.88	944.56	1000.64	994.64	1012.56	2.55
Glycoursodeoxycholate	6.07	6.13	5.86	4.95	5.74	7.75	7.45	15.67
Glycohyodeoxycholate	117.91	121.20	113.20	116.16	114.80	122.04	117.20	2.74
Glycochenodeoxycholate	40.28	42.60	40.36	40.28	38.52	42.24	40.36	3.38
Glycodeoxycholate	382.55	376.24	369.04	357.48	345.76	369.56	352.24	3.65
Glycocholate	327.19	318.68	339.00	332.60	302.72	316.24	303.60	4.34
Tauroursodeoxycholate	< LOD	< LOD	< LOD	< LOD	< LOD	< LOD	< LOD	N/A
Taurohyodeoxycholate	99.68624	96.66	95.05	111.32	97.67	105.64	106.94	6.02
Taurochenodeoxycholate	193.86	184.72	171.40	158.04	143.28	175.04	148.00	11.20
Taurodeoxycholate	32.68	28.72	27.40	28.96	28.52	29.88	27.48	6.19
Taurocholate	164.39	169.32	165.28	167.92	166.08	185.20	183.32	5.12

Appendix II

Table S5.1 Spearman rank correlation of metabolites with clinically measured serum creatinine at different ranges with discovery and validation sets (absolute Spearman *R* > 0.40)

Metabolites associated to clinically measured serum creatinine	Stages 0-4		Stages 1-4		Stage 1-2 (GFR ≥ 60)	
	Discovery (n=128)	Validation (n=66)	Discovery (n=108)	Validation (n=56)	Discovery (n=69)	Validation (n=30)
MS-detected serum creatinine	0.9569	0.9611	0.9648	0.9534	0.9115	0.8943
L,L-TMAP	0.8848	0.9258	0.9181	0.9342	0.7267	0.7104
L,L-TMAP isomer	0.8683	0.9014	0.9045	0.9261	0.7071	0.6732
Pseudouridine	0.8654	0.9073	0.9052	0.9232	0.6776	0.6823
Succinyladenosine	0.8301	0.9187	0.8765	0.9212	0.6343	0.7674
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8234	0.9176	0.8720	0.9184	0.5849	0.7718
5-Methylthio-D-ribose	0.8028	0.9287	0.8555	0.9395	0.5053	0.7885
Succinylcarnitine	0.7851	0.7069	0.8152	0.6929	0.5454	0.1383
Sulfotyrosine	0.7839	0.8944	0.8269	0.9099	0.4597	0.7115
4-Acetamidobutanoic acid	0.7705	0.8815	0.8157	0.9045	0.3640	0.5572
Ratio of pseudouridine to uridine	0.7652	0.9001	0.8074	0.9078	0.4771	0.5975
2-[3-(sulfooxy)phenyl]acetic acid	0.7617	0.8006	0.8137	0.8247	0.3819	0.4165
Indole-3-lactic acid	0.7576	0.7355	0.7763	0.7017	0.4384	0.5376
2-Hydroxyethanesulfonate	0.7428	0.7514	0.8076	0.8298	0.3909	0.5465
***N*-Acetylcarnosine**	0.7377	0.7324	0.7592	0.7106	0.5395	0.5468
***O*-Adipoylcarnitine**	0.7294	0.8577	0.7920	0.8929	0.3833	0.6567
Ratio of kynurenine to tryptophan	0.7076	0.7567	0.7593	0.8055	0.2187	0.3925
L-Kynurenine	0.7071	0.8375	0.7294	0.8450	0.2809	0.6285
Kynurenic acid	0.6972	0.8217	0.7618	0.8552	0.2480	0.5069
Arabinose isomer	0.6913	0.7409	0.7300	0.7595	0.3595	0.2801
L-*β*-Aspartyl-L-leucine	0.6640	0.7789	0.6971	0.8462	0.3200	0.5846
Butyrylcarnitine	0.6470	0.6518	0.7093	0.6580	0.2215	0.2500
Indoxyl sulfate	0.5682	0.7004	0.6681	0.7397	0.1713	0.5519
Homovanillic acid sulfate	0.5661	0.7584	0.6254	0.7844	0.0205	0.2847
***α*-N-Phenylacetyl-L-glutamine**	0.5590	0.5878	0.6468	0.6365	0.0047	0.0677
Ratio of kynurenic acid to kynurenine	0.5383	0.6293	0.6112	0.6725	0.0800	0.0363
L-Citrulline	0.5163	0.5932	0.5708	0.6802	-0.0098	0.0888
***p*-Cresol glucuronide**	0.5129	0.5016	0.5841	0.5602	-0.1374	0.1757
Valerylcarnitine	0.4974	0.5439	0.4945	0.5182	0.2781	0.3618
Phenol sulfate	0.4908	0.5900	0.5601	0.6252	0.1925	0.1567
p-Cresol sulfate	0.4845	0.4110	0.5747	0.4776	0.0138	-0.0443
3-Hydroxydecanoyl carnitine	0.4776	0.6505	0.5806	0.6507	0.5575	0.2674
9-Decenoylcarnitine	0.4726	0.5253	0.5553	0.5239	0.4369	0.2785
γ-Butyrobetaine	0.4269	0.4627	0.4619	0.4591	0.5170	0.5755
3-Indoleacetic acid	0.4056	0.5727	0.4345	0.5803	-0.0373	0.2643
L-Tryptophan	-0.4774	-0.4083	-0.5430	-0.5071	0.0792	0.2164
L-Carnitine	-0.5005	-0.4330	-0.5309	-0.4659	-0.1464	-0.1060
L-Tyrosine	-0.5126	-0.5369	-0.5773	-0.6564	-0.1645	-0.2531
Ratio of tyrosine to phenylalanine	-0.6663	-0.7429	-0.6955	-0.7613	-0.3634	-0.3680
Serum homocysteine (μmol/L)	-	0.7384	-	0.7501	-	0.6594
Serum cystatin C (mg/L)	-	0.9288	-	0.9450	-	0.8361

Table S5.2 Spearman rank correlation of metabolites with UACR at different ranges with discovery and validation sets. (Absolute Spearman *R* > 0.40)

Metabolites	Stages 0-4		Stages 1-4		Stages 1-3 (UACR<300)		Stages 1-2 (UACR<300)	
	Discovery (n= 128)	Validation (n=66)	Discovery (n=108)	Validation (n=56)	Discovery (n=69)	Validation (n=38)	Discovery (n=69)	Validation (n=30)
4-Acetamidobutanoic acid	0.7903	0.7244	0.8119	0.7076	0.5128	0.2306	0.5065	0.1849
Succinyladenosine	0.7799	0.6706	0.7947	0.6897	0.5327	0.1752	0.4900	0.1791
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.7643	0.7062	0.8009	0.7459	0.4686	0.2765	0.4739	0.3068
Pseudouridine	0.7514	0.7330	0.7936	0.7411	0.4685	0.2734	0.4467	0.2494
2-[3-(sulfooxy)phenyl] acetic acid	0.7311	0.6968	0.7540	0.6789	0.4164	0.1896	0.3369	0.1697
5-Methylthio-D-ribose	0.7124	0.6551	0.7626	0.6633	0.4176	0.1012	0.3665	0.0754
L-β-aspartyl-L-leucine	0.7095	0.5806	0.6736	0.6724	0.3563	0.1485	0.3807	0.1902
Arabinose isomer	0.7089	0.6716	0.6420	0.6591	0.2954	0.1095	0.2465	0.0171
2-Hydroxyethanesulfonate	0.6971	0.6040	0.7137	0.7109	0.3328	0.2520	0.2812	0.3086
L,L-TMAP isomer	0.6951	0.6704	0.7308	0.6695	0.3200	0.1649	0.2805	0.1248
L,L-TMAP	0.6940	0.6628	0.7426	0.6756	0.3471	0.1566	0.2907	0.1132
Serum creatinine (mg/dL)	0.6880	0.6464	0.7367	0.6668	0.3486	0.2334	0.2627	0.1734
Ratio of kynurenine to tryptophan	0.6837	0.6625	0.7302	0.7114	0.4041	0.2518	0.3247	0.3330
Sulfotyrosine	0.6466	0.6986	0.7160	0.7321	0.3497	0.3785	0.2515	0.2659
MS-detected creatinine	0.6434	0.6349	0.6805	0.6569	0.2506	0.1737	0.1389	0.0843
***O*-Adipoylcarnitine**	0.6311	0.7017	0.7089	0.6785	0.3586	0.2993	0.2846	0.2436
Ratio of pseudouridine to uridine	0.6212	0.7404	0.6672	0.7718	0.1955	0.3645	0.1498	0.3379
Succinylcarnitine	0.6152	0.6863	0.6446	0.6102	0.1968	0.1688	0.1394	0.1800
Indolelactic acid	0.5997	0.6167	0.6372	0.5987	0.2592	0.3951	0.2280	0.4047
L-Kynurenine	0.5886	0.6263	0.6384	0.6223	0.3312	0.2960	0.2369	0.2747
Urea (mmol/L)	0.5795	0.4845	0.6011	0.5098	0.1697	-0.0623	-0.0643	0.0574
***α*-N-Phenylacetyl-L-glutamine**	0.5616	0.4664	0.6036	0.4714	0.2370	0.0233	0.1102	-0.0278
Kynurenic acid	0.5578	0.7553	0.6187	0.7460	0.3206	0.3667	0.0581	0.4670
Butyrylcarnitine	0.5355	0.4758	0.5603	0.4732	0.0866	0.0795	-0.0010	0.1288
***p*-Cresol glucuronide**	0.5324	0.4299	0.5283	0.4478	0.1847	-0.0349	-0.0029	0.1831
N-Acetylcarnosine	0.5315	0.6329	0.5504	0.6179	0.0275	0.2465	0.0621	0.3321
L-Citrulline	0.5219	0.4984	0.5822	0.6252	0.1946	0.3623	0.1328	0.1835
Homovanillic acid sulfate	0.5172	0.7267	0.5944	0.6837	0.2009	0.2397	0.1270	0.2476
Choline	0.4907	0.5668	0.3795	0.5949	0.2332	0.1572	0.2669	0.1795
Indoxyl sulfate	0.4592	0.4917	0.5338	0.5281	0.1205	0.1529	-0.0106	0.1457
Ratio of kynurenic acid to kynurenine	0.4466	0.7075	0.4927	0.7032	0.2141	0.2638	-0.0547	0.3081
Phenol sulfate	0.4312	0.5684	0.5028	0.5444	0.0956	0.2095	0.0955	0.2147
Bilirubin	-0.5010	-0.3373	-0.5507	-0.4443	-0.2458	-0.1008	-0.1470	-0.1003
L-Tyrosine	-0.5461	-0.4178	-0.6023	-0.5382	-0.3021	0.0614	-0.3089	0.1546
L-Tryptophan	-0.5934	-0.5703	-0.5917	-0.6472	-0.3116	-0.1084	-0.1961	-0.2494
Ratio of tyrosine to phenylalanine	-0.6742	-0.5629	-0.6760	-0.5526	-0.3970	-0.0675	-0.4015	-0.0113
MDRD GFR	-0.7325	-0.6911	-0.7786	-0.6634	-0.4146	-0.2244	-0.3927	-0.1889
Serum cystatin C (mg/L)		0.6549		0.6758		0.1923		0.3427

Table S5.3 Spearman rank correlation of metabolites with urinary *β*2-microglobulin at different ranges with discovery and validation sets (absolute Spearman *R* > 0.40)

Metabolites	Stage 0-4		Stage 1-4		Stages 1-3 (UACR<300)		Stages 1-2 (UACR<300)	
	Discovery (n=128)	Validation (n=65)	Discovery (n=108)	Validation (n=55)	Discovery (n=63)	Validation (n=38)	Discovery (n=55)	Validation (n=28)
5-Methylthio-D-ribose	0.6569	0.6962	0.7100	0.7152	0.3875	0.3131	0.3243	-0.0373
L,L-TMAP isomer	0.6278	0.6460	0.6962	0.6814	0.2256	0.2824	0.1247	-0.0351
Pseudouridine	0.6270	0.7696	0.6900	0.7933	0.3189	0.5088	0.2296	0.2950
UACR	0.6215	0.7000	0.6473	0.6938	0.1689	0.2709	0.0436	0.2966
MS-detected creatinine	0.6183	0.6565	0.6670	0.6853	0.2030	0.2908	0.0914	0.0027
2-Hydroxyethanesulfonate	0.6179	0.5959	0.6684	0.6468	0.3413	0.1986	0.2561	0.1468
Serum Creatinine (mg/dL)	0.6102	0.6981	0.6583	0.7336	0.2493	0.3919	0.1490	0.1006
L,L-TMAP	0.6081	0.6557	0.6697	0.6809	0.1815	0.2534	0.0409	-0.0381
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.6063	0.6925	0.6638	0.7314	0.2206	0.3510	0.0881	0.0985
4-Acetamidobutanoic acid	0.6056	0.7871	0.6457	0.8056	0.2009	0.5240	0.0449	0.3155
***O*-Adipoylcarnitine**	0.5962	0.7536	0.6181	0.7719	0.3647	0.4841	0.2653	0.3883
Ratio of kynurenine to tryptophan	0.5891	0.6671	0.6759	0.6979	0.2470	0.3280	0.1344	0.0225
Succinyladenosine	0.5884	0.7412	0.6353	0.7802	0.1537	0.4771	0.0225	0.2011
Succinylcarnitine	0.5781	0.6436	0.6332	0.6151	0.2566	0.3277	0.1547	0.1147
2-[3-(sulfooxy)phenyl] acetic acid	0.5676	0.6673	0.5882	0.6990	0.2838	0.2898	0.1620	0.0195
Ratio of pseudouridine to uridine	0.5661	0.7365	0.6380	0.7657	0.2235	0.4410	0.1539	0.1926
Sulfotyrosine	0.5492	0.7136	0.6049	0.7680	0.2104	0.4638	0.1109	0.1976
Kynurenic acid	0.5382	0.7501	0.5848	0.7538	0.2076	0.4278	0.0916	0.2244
Arabinose isomer	0.5205	0.5946	0.5371	0.6354	0.0313	0.1557	-0.0963	-0.0966
***α*-N-Phenylacetyl-L-glutamine**	0.5147	0.5675	0.5628	0.6177	0.1606	0.3378	0.0413	-0.0601
L-Kynurenine	0.5125	0.6294	0.5864	0.6352	0.2289	0.3371	0.1068	0.1023
***p*-Cresol glucuronide**	0.4832	0.5133	0.5337	0.5160	0.0916	0.2349	-0.0307	-0.0209
***N*-Acetylcarnosine**	0.4767	0.5208	0.5208	0.5100	-0.1054	0.1027	-0.2168	0.1136
Indolelactic acid	0.4628	0.6052	0.4914	0.6101	0.1275	0.4050	-0.0194	0.3446
Ratio of kynurenic acid to kynurenine	0.4597	0.6744	0.4888	0.6760	0.1562	0.2601	0.0680	-0.0060
Butyrylcarnitine	0.4594	0.4768	0.5073	0.5014	0.1768	0.1621	0.0459	-0.1043
Indoxyl sulfate	0.4543	0.4795	0.5176	0.5467	0.1589	0.1708	0.0886	-0.0628
Homovanillic acid sulfate	0.4254	0.6172	0.4636	0.6284	-0.1413	0.1796	-0.1859	-0.1561
L-*β*-aspartyl-L-leucine	0.4194	0.5703	0.4462	0.6591	-0.0253	0.2073	-0.0974	-0.0996
Urea (mmol/L)	0.4159	0.5956	0.4526	0.6375	-0.0189	0.3580	-0.1780	0.0578

***p*-Cresol sulfate**	0.4044	0.4221	0.4468	0.4853	0.1235	0.2826	-0.0038	0.1432
L-Citrulline	0.3746	0.5191	0.4037	0.5951	-0.0207	0.2944	-0.0010	0.2939
Phenol sulfate	0.3376	0.5120	0.4198	0.5291	-0.0060	0.2225	-0.0155	0.0252
Bilirubin	-0.3860	-0.3606	-0.4272	-0.4709	-0.0724	-0.1488	-0.0234	-0.0313
L-Tryptophan	-0.5280	-0.4947	-0.5652	-0.5620	-0.0672	-0.0449	0.0273	0.1649
L-Tyrosine	-0.5377	-0.4873	-0.5757	-0.6038	-0.1692	-0.1175	-0.1504	-0.0903
Ratio of tyrosine to phenylalanine	-0.5831	-0.6767	-0.6141	-0.7038	-0.1606	-0.3807	-0.1135	-0.2297
MDRD GFR	-0.6283	-0.7571	-0.6714	-0.7571	-0.2710	-0.4184	-0.1830	-0.0826
Serum homocysteine (µmol/L)	-	0.4174	-	0.4328	-	0.0119	-	-0.0175
Serum cystatin C (mg/L)	-	0.6365	-	0.6749	-	0.2779	-	0.1290

Table S5.4 Spearman rank correlation of metabolites with clinically measured serum cystatin C, serum creatinine, CKD-EPI $_{creatinine}$ GFR, CKD-EPI $_{creatinine–cystatin\ C}$ GFR, CKD-EPI $_{cystatin\ C}$ GFR at different ranges in validation sets according to MDRD-GFR stages (absolute Spearman $R > 0.40$)

Metabolites	CKD-EPI $_{creatinine}$ GFR (ml/min/1.73 m2)			CKD-EPI $_{creatinine–cystatin\ C}$ GFR (ml/min/1.73 m2)			CKD-EPI $_{cystatin\ C}$ GFR (ml/min/1.73 m2)			Serum cystatin C			Serum creatinine (mg/dL)		
Stages	**0-4 (n=66)**	**1-4 (n=56)**	**1-2 (n=31)**	**0-4 (n=58)**	**1-4 (n=48)**	**1-2 (n=28)**	**0-4 (n=58)**	**1-4 (n=48)**	**1-2 (n=27)**	**0-4 (n=58)**	**1-4 (n=48)**	**1-2 (n=30)**	**0-4 (n=66)**	**1-4 (n=56)**	**1-2 (n=30)**
MDRD GFR	0.9850	0.9963	0.9851	0.9512	0.9756	0.9141	0.9212	0.9506	0.7825	-0.9500	-0.9592	-0.8546	-0.9126	-0.9400	-0.8016
Ratio of tyrosine to phenylalanine	0.7710	0.7707	0.5730	0.7472	0.7550	0.7441	0.7235	0.7359	0.4579	-0.7086	-0.7237	-0.3153	-0.7429	-0.7613	-0.3680
L-Tyrosine	0.5428	0.6302	0.2617	0.5050	0.6448	0.6548	0.4915	0.6337	0.2372	-0.4718	-0.6247	-0.1885	-0.5369	-0.6564	-0.2531
Bilirubin	0.5360	0.6349	0.3625	0.5496	0.6698	0.6994	0.5514	0.6712	0.4308	-0.5256	-0.6593	-0.4885	-0.5061	-0.6384	-0.4744
L-Tryptophan	0.4723	0.5561	-0.1367	0.4608	0.5638	0.5107	0.4918	0.5990	0.1939	-0.4478	-0.5573	-0.1090	-0.4083	-0.5071	0.2164
L-Carnitine	0.4183	0.4396	0.2560	0.5356	0.5620	0.4718	0.5275	0.5517	-0.1612	-0.5072	-0.5453	-0.0779	-0.4330	-0.4659	-0.1060
2-Hydroxybutyric acid	0.3315	0.5278	0.3911	0.2936	0.5280	0.5053	0.3026	0.5349	0.2442	-0.2639	-0.4925	-0.2991	-0.3031	-0.4918	-0.3495
Betaine	0.1461	0.0623	-0.1711	0.0374	-0.0852	-0.0633	0.0322	-0.0822	-0.3081	-0.0017	0.1155	0.5735	-0.0568	0.0324	0.5001
L-Threonine	0.0959	0.0519	0.0633	0.0283	-0.0213	-0.0069	0.0172	-0.0307	-0.0119	0.0239	0.0653	0.2139	-0.0006	0.0233	0.2974
D-Glucose	0.0492	0.2229	0.4694	0.0182	0.2006	0.2263	0.0214	0.2039	0.5563	0.0103	-0.1711	-0.4420	-0.0467	-0.2277	-0.4181
1,5-Anhydro-D-glucitol	0.0260	-0.1692	-0.4851	-0.0709	-0.3333	-0.4100	-0.0928	-0.3510	-0.5709	0.0575	0.3204	0.5483	-0.0078	0.2075	0.4323
L-Glutamine	-0.0312	-0.1656	-0.3573	-0.1323	-0.2841	-0.2837	-0.1394	-0.2879	-0.4296	0.1399	0.2845	0.5319	0.0988	0.2398	0.4989
Propionylcarnitine	-0.1538	-0.0781	-0.2157	-0.1698	-0.1254	-0.1575	-0.1612	-0.1212	-0.4170	0.1884	0.1357	0.3935	0.1749	0.1089	0.2654
Decanoylcarnitine	-0.2092	-0.2067	-0.0004	-0.3299	-0.3835	-0.4350	-0.3272	-0.3784	-0.2513	0.3786	0.4189	0.2515	0.3116	0.2991	0.2366
Cortisol	-0.2192	-0.1374	0.1500	-0.1177	-0.0014	-0.0819	-0.1440	-0.0262	0.0089	0.1567	0.0318	-0.0245	0.2292	0.1373	-0.1167
Hexanoylcarnitine	-0.2711	-0.2667	-0.0048	-0.2369	-0.2566	-0.3310	-0.2342	-0.2511	-0.4799	0.2789	0.2882	0.3739	0.3254	0.3048	0.2956
L-Octanoylcarnitine	-0.2723	-0.3018	0.0016	-0.3183	-0.3817	-0.3786	-0.3162	-0.3769	-0.3486	0.3626	0.4117	0.2877	0.3569	0.3668	0.2171
Uric acid	-0.2971	-0.1811	-0.3177	-0.3924	-0.2644	-0.3843	-0.4099	-0.2865	-0.5074	0.4430	0.3186	0.6242	0.3886	0.2776	0.5343
L-Proline	-0.3022	-0.2905	-0.0032	-0.2465	-0.2012	-0.2062	-0.2323	-0.1886	-0.0116	0.3136	0.2527	0.2455	0.4107	0.3454	0.3653
3,5-Tetradecadiencarnitine	-0.3051	-0.3231	-0.0153	-0.3546	-0.4031	-0.3318	-0.3714	-0.4179	-0.2442	0.4004	0.4364	0.2984	0.3785	0.3747	0.1162
γ-Butyrobetaine	-0.3522	-0.3679	-0.3520	-0.2597	-0.2800	-0.3624	-0.2250	-0.2399	-0.3233	0.2755	0.2852	0.4099	0.4059	0.3949	0.5755
Hydroxybutyrylcarnitine	-0.3870	-0.3822	-0.0992	-0.3306	-0.3340	-0.2216	-0.3409	-0.3503	-0.1984	0.3651	0.3705	0.2671	0.4186	0.4240	0.1503
L-Phenylalanine	-0.3958	-0.3039	-0.4883	-0.3946	-0.2778	-0.3104	-0.3645	-0.2446	-0.3685	0.3781	0.2531	0.2693	0.3826	0.2713	0.3217
***p*-Cresol sulfate**	-0.4105	-0.4841	-0.0722	-0.3498	-0.4280	-0.3954	-0.3590	-0.4346	0.0827	0.3440	0.4170	0.0162	0.4110	0.4776	-0.0443
9-Decenoylcarnitine	-0.4319	-0.4506	-0.0266	-0.4818	-0.5343	-0.5498	-0.4955	-0.5455	-0.3679	0.5417	0.5786	0.4150	0.5253	0.5239	0.2785
L-Glutamic acid	-0.4437	-0.4053	-0.3113	-0.3471	-0.2896	-0.3387	-0.3326	-0.2770	-0.2638	0.3508	0.2979	0.1271	0.4605	0.4229	0.2696
Valerylcarnitine	-0.4482	-0.4157	-0.0306	-0.5332	-0.4904	-0.4579	-0.5230	-0.4753	-0.2946	0.5560	0.5032	0.3737	0.5438	0.5182	0.3618
Ratio of L-citrulline to arginine	-0.4753	-0.5818	0.1020	-0.3585	-0.4734	-0.4399	-0.3733	-0.4807	-0.0018	0.3530	0.4766	0.0113	0.4312	0.5563	-0.0494
***p*-Cresol glucuronide**	-0.5067	-0.5768	-0.1476	-0.4094	-0.5021	-0.4342	-0.4157	-0.5015	-0.0363	0.4206	0.5099	0.1998	0.5016	0.5602	0.1756
Indoleacetic acid	-0.5156	-0.5157	-0.1544	-0.4923	-0.5103	-0.4183	-0.4897	-0.5151	0.1011	0.5113	0.5266	0.1591	0.5727	0.5803	0.2642
2-Octenoylcarnitine	-0.5300	-0.5492	-0.2194	-0.5570	-0.5936	-0.5256	-0.5848	-0.6216	-0.2888	0.6028	0.6317	0.4912	0.5483	0.5587	0.3103

α-N-Phenylacetyl-L-glutamine	-0.5861	-0.6291	-0.1734	-0.5141	-0.5710	-0.4520	-0.5320	-0.5897	0.1966	0.5205	0.5738	0.0372	0.5878	0.6365	0.0677
L-Citrulline	-0.5954	-0.7010	0.0593	-0.4990	-0.6193	-0.5537	-0.5184	-0.6324	-0.0940	0.5206	0.6346	0.1653	0.5932	0.6802	0.0888
Citric acid	-0.5959	-0.6349	-0.2085	-0.5212	-0.5618	-0.6384	-0.4977	-0.5352	-0.3377	0.4932	0.5393	0.3287	0.5365	0.5878	0.1995
3-hydroxydecanoyl carnitine	-0.6048	-0.6068	-0.2448	-0.6825	-0.7103	-0.6420	-0.6948	-0.7197	-0.3474	0.7207	0.7364	0.4952	0.6505	0.6507	0.2674
Phenol sulfate	-0.6058	-0.6314	-0.3359	-0.6287	-0.6418	-0.6710	-0.6424	-0.6548	-0.5169	0.6251	0.6379	0.3805	0.5900	0.6252	0.1567
Ratio of kynurenic acid to kynurenine	-0.6442	-0.6962	-0.0218	-0.5997	-0.6657	-0.5413	-0.6025	-0.6690	0.0519	0.5975	0.6604	0.1260	0.6293	0.6725	0.0363
Indoxyl sulfate	-0.6500	-0.6837	-0.4254	-0.6251	-0.6892	-0.6960	-0.6381	-0.7046	-0.3920	0.6430	0.6993	0.4960	0.7004	0.7397	0.5519
Butyrylcarnitine	-0.6511	-0.6603	-0.1298	-0.6202	-0.6620	-0.6124	-0.6060	-0.6506	-0.2729	0.5988	0.6271	0.2884	0.6518	0.6580	0.2500
***N*-Acetylcarnosine**	-0.6588	-0.6304	-0.3552	-0.6554	-0.6499	-0.6389	-0.6432	-0.6391	-0.1945	0.6974	0.7006	0.4883	0.7324	0.7106	0.5467
Choline	-0.6707	-0.6855	-0.2919	-0.6243	-0.6657	-0.5761	-0.6096	-0.6483	-0.1731	0.5991	0.6421	0.2192	0.6277	0.6411	0.2317
Succinylcarnitine	-0.6793	-0.6699	0.0024	-0.6267	-0.5973	-0.4085	-0.6233	-0.5941	0.1230	0.6706	0.6437	0.1177	0.7069	0.6929	0.1382
UACR	-0.6800	-0.6899	-0.0931	-0.6430	-0.6562	-0.1111	-0.6539	-0.6700	-0.2693	0.6464	0.6668	0.1734	0.6549	0.6758	0.3427
Indolelactic acid	-0.6874	-0.6630	-0.3298	-0.7540	-0.7584	-0.6752	-0.7518	-0.7592	-0.3395	0.7771	0.7801	0.5121	0.7355	0.7017	0.5376
Urea (mmol/L)	-0.7029	-0.7610	-0.3107	-0.6830	-0.7559	-0.6480	-0.6817	-0.7547	-0.1898	0.6936	0.7645	0.4504	0.6997	0.7643	0.3428
2-Hydroxyethanesulfonate	-0.7169	-0.8126	-0.4395	-0.7169	-0.8342	-0.7426	-0.7143	-0.8292	-0.4021	0.7278	0.8445	0.5844	0.7514	0.8298	0.5465
Arabinose isomer	-0.7654	-0.7932	-0.3440	-0.6825	-0.7252	-0.6039	-0.6781	-0.7253	-0.1762	0.6981	0.7430	0.3229	0.7409	0.7595	0.2801
Homovanillic acid sulfate	-0.7967	-0.8090	-0.3585	-0.7535	-0.7776	-0.6708	-0.7425	-0.7691	-0.1734	0.7285	0.7530	0.3027	0.7584	0.7844	0.2847
Ratio of kynurenine to tryptophan	-0.7973	-0.8373	-0.4565	-0.8335	-0.8804	-0.8378	-0.8549	-0.8982	-0.7199	0.8292	0.8809	0.7259	0.7567	0.8055	0.3925
L-*β*-aspartyl-L-leucine	-0.7993	-0.8609	-0.6270	-0.7489	-0.8716	-0.8512	-0.7316	-0.8590	-0.5911	0.7257	0.8544	0.6525	0.7789	0.8462	0.5846
2-[3-(sulfooxy)phenyl]acetic acid	-0.8176	-0.8300	-0.4044	-0.7769	-0.7980	-0.6505	-0.7654	-0.7894	-0.1417	0.7760	0.7988	0.3866	0.8006	0.8246	0.4165
Kynurenic acid	-0.8335	-0.8710	-0.4685	-0.8538	-0.9023	-0.8281	-0.8514	-0.8984	-0.5706	0.8511	0.8992	0.6803	0.8217	0.8552	0.5069
***O*-Adipoylcarnitine**	-0.8383	-0.8568	-0.5085	-0.8320	-0.8531	-0.7586	-0.8213	-0.8381	-0.4708	0.8415	0.8612	0.6231	0.8577	0.8929	0.6567
L-Kynurenine	-0.8386	-0.8485	-0.6440	-0.8575	-0.8721	-0.8942	-0.8646	-0.8740	-0.7794	0.8660	0.8790	0.8494	0.8375	0.8450	0.6285
4-Acetamidobutanoic acid	-0.8755	-0.8946	-0.4871	-0.8671	-0.8829	-0.7266	-0.8676	-0.8836	-0.3630	0.8818	0.8951	0.5837	0.8815	0.9045	0.5572
Ratio of pseudouridine to uridine	-0.8839	-0.9054	-0.5851	-0.8739	-0.9018	-0.7897	-0.8760	-0.9037	-0.6201	0.8989	0.9125	0.7041	0.9001	0.9078	0.5975
L,L-TMAP isomer	-0.8986	-0.9219	-0.6831	-0.8634	-0.8992	-0.7964	-0.8575	-0.8919	-0.5147	0.8743	0.9021	0.6776	0.9014	0.9261	0.6732
5-Methylthio-D-ribose	-0.8993	-0.9191	-0.6952	-0.8876	-0.9145	-0.8162	-0.8867	-0.9113	-0.5535	0.9171	0.9404	0.7876	0.9287	0.9395	0.7885
MS-detected creatinine	-0.9092	-0.9159	-0.7621	-0.8618	-0.8796	-0.8216	-0.8378	-0.8566	-0.4824	0.8818	0.8951	0.7206	0.9611	0.9534	0.8943
Sulfotyrosine	-0.9121	-0.9240	-0.7552	-0.8891	-0.9081	-0.8646	-0.8898	-0.9081	-0.7297	0.8891	0.9028	0.7509	0.8944	0.9099	0.7115
L,L-TMAP	-0.9145	-0.9288	-0.7387	-0.9062	-0.9261	-0.8571	-0.8998	-0.9186	-0.6494	0.9116	0.9194	0.7475	0.9258	0.9342	0.7104
Succinyladenosine	-0.9244	-0.9281	-0.8129	-0.9301	-0.9378	-0.9102	-0.9187	-0.9317	-0.7864	0.9221	0.9321	0.8390	0.9187	0.9212	0.7674
2-(*α*-D-Mannopyranosyl)-L-tryptophan	-0.9268	-0.9306	-0.8069	-0.9426	-0.9649	-0.9601	-0.9409	-0.9643	-0.9015	0.9437	0.9647	0.9209	0.9176	0.9184	0.7718
Pseudouridine	-0.9302	-0.9517	-0.7831	-0.9393	-0.9581	-0.9060	-0.9348	-0.9544	-0.7935	0.9356	0.9483	0.8174	0.9073	0.9232	0.6823
Serum cystatin C (mg/L)	-0.9307	-0.9426	-0.8210	-0.9814	-0.9811	-0.9582	-0.9859	-0.9852	-0.9181	-	-	-	0.9288	0.9450	0.8361
Serum creatinine (mg/dL)	-0.9527	-0.9603	-0.8594	-0.9173	-0.9416	-0.8841	-0.8947	-0.9206	-0.6328	0.9288	0.9450	0.8361	-	-	-

Table S5.5 Spearman rank correlation of metabolites with clinical glucose markers at different ranges with discovery and validation sets (absolute Spearman *R* > 0.40)

Metabolites	Hemoglobin A1c (%)					Fasting plasma glucose (mmol/L)					
	Stage 0-4	Stage 1-4		Stage 1-2		Stage 0-4		Stage 1-4		Stage 1-2	
	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation	Discovery	Validation
Hemoglobin A1c (%)	-	-	-	-	-	0.6013	0.9145	0.6013	0.8860	0.4857	0.8961
Fasting plasma glucose (mmol/L)	0.9145	0.6013	0.8860	0.4857	0.8961	-	-	-	-	-	-
D-Glucose	0.9020	0.6046	0.8603	0.4358	0.8908	0.8960	0.9363	0.8877	0.9150	0.8089	0.8621
2-Hydroxybutyric acid	0.3883	0.4041	0.2418	0.2233	0.5240	0.5288	0.4133	0.5851	0.2733	0.5581	0.5172
Stearoylcarnitine	0.3092	0.2490	0.2473	0.2809	0.4270	0.2808	0.2709	0.3063	0.2044	0.4300	0.4492
MS-detected creatinine	0.0545	-0.2923	-0.1610	0.0758	-0.3457	-0.1682	0.0418	-0.3258	-0.1343	0.2565	-0.3099
Serum cystatin C (mg/L)	0.0257	-	-0.2304	-	-0.4957	-	0.0180	-	-0.1501	-	-0.4248
L-Glutamine	-0.6750	-0.5939	-0.6290	-0.5494	-0.6920	-0.6461	-0.6469	-0.6412	-0.5920	-0.6039	-0.6538
1,5-anhydro-D-glucitol	-0.8530	-0.7777	-0.8309	-0.7809	-0.9298	-0.5963	-0.7690	-0.6022	-0.7228	-0.6409	-0.8291

Table S5.6 Spearman rank correlation of metabolites with total BSA-related renal volume and renal resistive index in validation set (absolute Spearman *R* > 0.40)

Metabolites	Total BSA-related renal volume		Renal resistive index
	Validation (Stage 0-4) (n=58)	Validation (Stage 1-4) (n=48)	Validation (Stage 1-4) (n=48)
MDRD GFR	0.5423	0.6376	-0.5253
Bilirubin	0.5388	0.5533	-0.4234
Stearoylcarnitine	0.4303	0.4694	-0.1478
Ratio of tyrosine to phenylalanine	0.3868	0.4253	-0.4676
L-Tryptophan	0.3190	0.3752	-0.5688
L-Tyrosine	0.3086	0.3259	-0.4337
SBP(mmHg)	-0.0823	-0.1285	0.5961
***N*-Acetylcarnosine**	-0.1373	-0.2279	0.4343
Succinylcarnitine	-0.2057	-0.3588	0.4011
Ratio of kynurenic acid to kynurenine	-0.2335	-0.3091	0.4990
Citric acid	-0.2454	-0.3223	0.4839
Phenol sulfate	-0.2479	-0.3182	0.4510
Arabinose isomer	-0.2643	-0.4015	0.6066
Indolelactic acid	-0.3073	-0.4087	0.4482
3-hydroxydecanoyl carnitine	-0.3101	-0.4894	0.3840
Indoleacetic acid	-0.3199	-0.4440	0.1961
2-Octenoylcarnitine	-0.3418	-0.4858	0.2388
MS-detected creatinine	-0.3521	-0.4729	0.4593
Ratio of pseudouridine to uridine	-0.3693	-0.5276	0.5435
L,L-TMAP isomer	-0.3756	-0.5602	0.4952
***O*-Adipoylcarnitine**	-0.3898	-0.5784	0.3563
Urea (mmol/L)	-0.3912	-0.5352	0.5543
Renal resistive index	-	-0.3916	-
2-[3-(sulfooxy)phenyl]acetic acid	-0.4008	-0.5164	0.4072
2-Hydroxyethanesulfonate	-0.4291	-0.4807	0.5026
Kynurenic acid	-0.4322	-0.5372	0.5670
Homovanillic acid sulfate	-0.4365	-0.5474	0.4551
4-Acetamidobutanoic acid	-0.4397	-0.5573	0.4941
Sulfotyrosine	-0.4414	-0.5723	0.5347
Serum Creatinine (mg/dL)	-0.4434	-0.5705	0.4456

L,L-TMAP	-0.4492	-0.5959	0.5173
Butyrylcarnitine	-0.4572	-0.6169	0.3690
Pseudouridine	-0.4656	-0.5988	0.5884
5-Methylthio-D-ribose	-0.4745	-0.6183	0.4655
2-(*α*-D-Mannopyranosyl)-L-tryptophan	-0.4801	-0.6035	0.5863
L-*β*-aspartyl-L-leucine	-0.4865	-0.5892	0.5783
Serum cystatin C (mg/L)	-0.5032	-0.6136	0.5442
Succinyladenosine	-0.5426	-0.6145	0.5561
L-Kynurenine	-0.5622	-0.6833	0.4724
Ratio of kynurenine to tryptophan	-0.5630	-0.6403	0.6128

Table S5.7 AUC of metabolites for distinguishing Stage 1a patients from Stage 1b-4 patients using random forest classification and logistic regression.

Metabolites	Stage 1a vs Stages 1b-4 (GFR > 120 vs GFR ≤ 120)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9626+/- 0.0154	0.8807+/- 0.0402	0.9560+/- 0.1424	0.9648+/- 0.0144	0.8909+/- 0.0779	0.8460+/- 0.1694	0.9399+/- 0.0700	0.8561+/- 0.0664	0.8800+/- 0.2078	0.9597+/- 0.0226	0.9587+/- 0.0921	0.3580+/- 0.3547
Pseudouridine	0.9148 ± 0.0543	0.9334 ± 0.0534	0.6880 ± 0.1485	0.9432 ± 0.0239	0.8134 ± 0.1081	0.8570 ± 0.1344	0.9148 ± 0.0543	0.9334 ± 0.0534	0.6880 ± 0.1485	0.9432 ± 0.0239	0.8134 ± 0.1081	0.8570 ± 0.1344
L,L-TMAP	0.8984 ± 0.0461	0.8102 ± 0.0661	0.9020 ± 0.1319	0.9394 ± 0.0219	0.9305 ± 0.0832	0.8984 ± 0.0461	0.8102 ± 0.0661	0.9020 ± 0.1319	0.9394 ± 0.0219	0.9305 ± 0.0832	0.8984 ± 0.0461	0.8102 ± 0.0661
L,L-TMAP isomer	0.8915 ± 0.0492	0.7639 ± 0.0670	0.8820 ± 0.1431	0.9280 ± 0.0237	0.8995 ± 0.0901	0.5910 ± 0.2940	0.8696 ± 0.0684	0.8204 ± 0.0656	0.7700 ± 0.2689	0.9086 ± 0.0389	0.3743 ± 0.1503	1.0000
2-(α-D-Mannopyranosyl)-L-tryptophan	0.8878 ± 0.0659	0.8320 ± 0.1096	0.7530 ± 0.1565	0.9212 ± 0.0315	0.8932 ± 0.0922	0.6680 ± 0.2395	0.8930 ± 0.0648	0.8522 ± 0.0740	0.8060 ± 0.2521	0.9398 ± 0.0303	0.3861 ± 0.1378	1.0000
Succinyladenosine	0.8877 ± 0.0584	0.8318 ± 0.0591	0.7910 ± 0.1582	0.9147 ± 0.0306	0.8973 ± 0.1112	0.5720 ± 0.2895	0.9228 ± 0.0499	0.8657 ± 0.0674	0.8240 ± 0.1795	0.9513 ± 0.0291	0.4691 ± 0.1296	1.0000
MS-detected serum creatinine	0.8491 ± 0.0508	0.7207 ± 0.1175	0.8760 ± 0.1201	0.8966 ± 0.0337	0.8470 ± 0.0774	0.7080 ± 0.1677	0.8787 ± 0.0777	0.9196 ± 0.0996	0.6740 ± 0.1825	0.9298 ± 0.0414	0.4352 ± 0.1679	0.9940 ± 0.0341
Sulfotyrosine	0.8513 ± 0.0529	0.7945 ± 0.0780	0.7970 ± 0.1513	0.8965 ± 0.0338	0.4707 ± 0.0849	0.9990 ± 0.0099	0.8333 ± 0.1104	0.7974 ± 0.1225	0.7740 ± 0.2361	0.9003 ± 0.0498	0.4583 ± 0.1252	0.9880 ± 0.0475
5-Methylthio-D-ribose	0.8446 ± 0.0606	0.7396 ± 0.0774	0.8420 ± 0.2471	0.8871 ± 0.0469	0.4317 ± 0.1245	1.0000	0.8446 ± 0.0606	0.7396 ± 0.0774	0.8420 ± 0.2471	0.8871 ± 0.0469	0.4317 ± 0.1245	1.0000

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.8 AUC of metabolites for distinguishing Stage 2-4 patients from Stage 1 patients using random forest classification and logistic regression.

Metabolites	Stage 2-4 vs Stages 1 (GFR < 90 vs GFR ≥ 90)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9529+/-0.0242	0.7721+/-0.0619	0.9568+/-0.0897	0.9594+/-0.0149	0.8565+/-0.0932	0.8600+/-0.1112	0.9568+/-0.0334	0.8678+/-0.0675	0.8880+/-0.1699	0.9737+/-0.0175	0.6917+/-0.2010	0.9930+/-0.0453
Pseudouridine	0.9348 ± 0.0349	0.8741 ± 0.0579	0.9047 ± 0.0828	0.9526 ± 0.0184	0.7574 ± 0.1130	0.9605 ± 0.0455	0.9942 ± 0.0105	0.9661 ± 0.0359	0.9760 ± 0.0634	0.9992 ± 0.0020	0.5911 ± 0.1728	0.9980 ± 0.0199
2-(α-D-Mannopyranosyl)-L-tryptophan	0.9348 ± 0.0259	0.7779 ± 0.0639	0.9195 ± 0.0977	0.9383 ± 0.0197	0.8279 ± 0.0947	0.8316 ± 0.1112	0.9544 ± 0.0337	0.9394 ± 0.0940	0.8570 ± 0.0828	0.9769 ± 0.0179	0.6117 ± 0.2081	0.9640 ± 0.1171
5-Methylthio-D-ribose	0.9294 ± 0.0298	0.6891 ± 0.0914	0.9474 ± 0.0640	0.9424 ± 0.0208	0.7747 ± 0.1142	0.8979 ± 0.0905	0.9579 ± 0.0284	0.8967 ± 0.0657	0.9410 ± 0.0708	0.9686 ± 0.0220	0.8361 ± 0.1704	0.8080 ± 0.2942
L,L-TMAP isomer	0.9214 ± 0.0314	0.8094 ± 0.0786	0.9389 ± 0.0665	0.9489 ± 0.0190	0.8494 ± 0.1119	0.8442 ± 0.1598	0.8953 ± 0.0518	0.7817 ± 0.1098	0.8730 ± 0.1441	0.9277 ± 0.0343	0.6250 ± 0.2138	0.9560 ± 0.1116
L,L-TMAP	0.9064 ± 0.0380	0.7809 ± 0.0571	0.9395 ± 0.0810	0.9333 ± 0.0234	0.8144 ± 0.1209	0.8321 ± 0.1899	0.9281 ± 0.0536	0.7844 ± 0.0851	0.9290 ± 0.1409	0.9509 ± 0.0252	0.6350 ± 0.2076	0.9600 ± 0.1058
MS-detected serum creatinine	0.9183 ± 0.0281	0.7244 ± 0.0708	0.9068 ± 0.0953	0.9197 ± 0.0229	0.6665 ± 0.0818	0.9642 ± 0.0464	0.9268 ± 0.0480	0.8111 ± 0.0813	0.9100 ± 0.1453	0.9518 ± 0.0280	0.6589 ± 0.2239	0.9590 ± 0.0960
Succinyladenosine	0.8990 ± 0.0378	0.7874 ± 0.0594	0.9016 ± 0.0917	0.9213 ± 0.0278	0.8097 ± 0.1092	0.8468 ± 0.1739	0.9715 ± 0.0314	0.9306 ± 0.0787	0.9250 ± 0.1062	0.9901 ± 0.0088	0.8250 ± 0.1479	0.9370 ± 0.1324
Sulfotyrosine	0.8682 ± 0.0382	0.7206 ± 0.0755	0.9458 ± 0.0739	0.9034 ± 0.0292	0.7965 ± 0.1135	0.8142 ± 0.1587	0.9118 ± 0.0597	0.8872 ± 0.0928	0.8250 ± 0.1486	0.9404 ± 0.0313	0.6272 ± 0.1493	0.9640 ± 0.0625
O-Adipoylcarnitine	0.8687 ± 0.0445	0.7247 ± 0.0769	0.8816 ± 0.0778	0.9048 ± 0.0270	0.6100 ± 0.1153	0.9626 ± 0.0472	0.8556 ± 0.0612	0.7344 ± 0.1295	0.7600 ± 0.1530	0.8789 ± 0.0483	0.6261 ± 0.1897	0.8540 ± 0.2651
Ratio of pseudouridine to uridine	0.8679 ± 0.0436	0.7997 ± 0.0805	0.8505 ± 0.1009	0.9013 ± 0.0336	0.7918 ± 0.1059	0.8568 ± 0.1087	0.9320 ± 0.0366	0.8056 ± 0.0776	0.9280 ± 0.0906	0.9486 ± 0.0275	0.7583 ± 0.2007	0.8640 ± 0.2052
2-Hydroxyethanesulfonate	0.8650 ± 0.0449	0.7218 ± 0.0709	0.9205 ± 0.0827	0.9075 ± 0.0327	0.5891 ± 0.1401	0.9695 ± 0.0511	0.8120 ± 0.0657	0.7150 ± 0.1068	0.8250 ± 0.1615	0.8761 ± 0.0484	0.6261 ± 0.1767	0.9110 ± 0.1348
4-Acetamidobutanoic acid	0.8541 ± 0.0409	0.7679 ± 0.0722	0.9158 ± 0.0973	0.8850 ± 0.0328	0.7029 ± 0.1476	0.9100 ± 0.1921	0.9444 ± 0.0296	0.8567 ± 0.0734	0.9030 ± 0.1307	0.9590 ± 0.0200	0.5406 ± 0.2425	0.9480 ± 0.1775

Ratio of kynurenine to tryptophan	0.8322 ± 0.0425	0.6268 ± 0.0814	0.9084 ± 0.0779	0.8583 ± 0.0366	0.5091 ± 0.1134	0.9842 ± 0.0273	0.8930 ± 0.0669	0.7572 ± 0.0999	0.8740 ± 0.1481	0.9171 ± 0.0352	0.7083 ± 0.2181	0.8110 ± 0.2912
Kynurenic acid	0.8319 ± 0.0417	0.7150 ± 0.0940	0.8174 ± 0.0969	0.8526 ± 0.0394	0.6441 ± 0.1717	0.8274 ± 0.2677	0.9357 ± 0.0365	0.8228 ± 0.1026	0.8810 ± 0.1146	0.9499 ± 0.0270	0.6950 ± 0.2875	0.7010 ± 0.4241
2-[3-(sulfooxy)phenyl]acetic acid	0.8250 ± 0.0504	0.5982 ± 0.1090	0.9311 ± 0.0758	0.8587 ± 0.0366	0.2794 ± 0.0695	1.0000 ± 0.0000	0.8854 ± 0.0456	0.6922 ± 0.1366	0.8620 ± 0.1690	0.9039 ± 0.0376	0.6100 ± 0.3210	0.6720 ± 0.4407
Ratio of tyrosine to phenylalanine	0.8226 ± 0.0469	0.6865 ± 0.0734	0.8763 ± 0.0840	0.8595 ± 0.0384	0.7194 ± 0.0787	0.8063 ± 0.1466	0.8648 ± 0.0553	0.8589 ± 0.1404	0.7030 ± 0.1153	0.8884 ± 0.0378	0.5900 ± 0.1229	0.9310 ± 0.1083

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.9 AUC of metabolites for distinguishing Stage 3-4 patients from Stage 1-2 patients using random forest classification and logistic regression.

Metabolites	Stage 3-4 vs Stages 1-2 (GFR < 60 vs GFR ≥ 60)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9715+/-0.0306	0.9542+/-0.0754	0.8818+/-0.0757	0.9901+/-0.0060	0.9347+/-0.0752	0.9003+/-0.0979	0.9665+/-0.0340	0.9192+/-0.1073	0.8600+/-0.0978	0.9769+/-0.0163	0.7262+/-0.2228	0.9447+/-0.0725
Pseudouridine	0.9820 ± 0.0213	0.9042 ± 0.0757	0.9491 ± 0.0444	0.9912 ± 0.0056	0.9195 ± 0.0753	0.9335 ± 0.0535	0.9515 ± 0.0341	0.9054 ± 0.0960	0.8827 ± 0.0905	0.9702 ± 0.0201	0.7931 ± 0.1993	0.8727 ± 0.1744
2-(α-D-Mannopyranosyl)-L-tryptophan	0.9743 ± 0.0212	0.9232 ± 0.0646	0.8979 ± 0.0549	0.9825 ± 0.0095	0.9705 ± 0.0449	0.8226 ± 0.1001	0.9395 ± 0.0393	0.7862 ± 0.1335	0.8633 ± 0.1358	0.9387 ± 0.0274	0.6800 ± 0.2069	0.8987 ± 0.1607
L,L-TMAP	0.9691 ± 0.0199	0.9495 ± 0.0566	0.8512 ± 0.0737	0.9833 ± 0.0089	0.9411 ± 0.0806	0.8418 ± 0.1369	0.9151 ± 0.0459	0.8446 ± 0.1109	0.8580 ± 0.0878	0.9438 ± 0.0268	0.6962 ± 0.2467	0.8980 ± 0.1224
5-Methylthio-D-ribose	0.9687 ± 0.0243	0.9100 ± 0.0679	0.9529 ± 0.0458	0.9854 ± 0.0079	0.8684 ± 0.0882	0.9629 ± 0.0457	0.9424 ± 0.0386	0.9346 ± 0.0900	0.7753 ± 0.0903	0.9528 ± 0.0257	0.8223 ± 0.1819	0.7773 ± 0.2239
2-Hydroxyethanesulfonate	0.9685 ± 0.0179	0.9800 ± 0.0420	0.8503 ± 0.0422	0.9692 ± 0.0128	0.8032 ± 0.1458	0.9026 ± 0.1390	0.8551 ± 0.0692	0.7785 ± 0.1436	0.7540 ± 0.1243	0.8793 ± 0.0485	0.6808 ± 0.1785	0.8033 ± 0.1898
4-Acetamidobutanoic acid	0.9666 ± 0.0196	0.9611 ± 0.0570	0.8447 ± 0.0440	0.9693 ± 0.0130	0.9432 ± 0.1174	0.6774 ± 0.2850	0.9408 ± 0.0396	0.9131 ± 0.0719	0.9080 ± 0.0842	0.9548 ± 0.0267	0.6531 ± 0.2508	0.8660 ± 0.2767
2-[3-(sulfooxy)phenyl]acetic acid	0.9663 ± 0.0197	0.9558 ± 0.0599	0.8982 ± 0.0431	0.9742 ± 0.0131	0.5426 ± 0.1386	0.9776 ± 0.1403	0.8775 ± 0.0533	0.7977 ± 0.1134	0.7627 ± 0.1829	0.9073 ± 0.0396	0.4600 ± 0.1804	0.9607 ± 0.1749
L,L-TMAP isomer	0.9560 ± 0.0217	0.9337 ± 0.0643	0.8094 ± 0.0890	0.9702 ± 0.0140	0.8026 ± 0.1149	0.9388 ± 0.0612	0.9188 ± 0.0441	0.8777 ± 0.0949	0.8040 ± 0.1131	0.9510 ± 0.0250	0.7031 ± 0.2393	0.8947 ± 0.1427
Succinyladenosine	0.9581 ± 0.0276	0.9116 ± 0.0755	0.8412 ± 0.0774	0.9648 ± 0.0141	0.9163 ± 0.1114	0.7968 ± 0.1588	0.9118 ± 0.0437	0.8992 ± 0.0861	0.8427 ± 0.1005	0.9522 ± 0.0292	0.8346 ± 0.1792	0.8380 ± 0.1494
Ratio of kynurenine to tryptophan	0.9582 ± 0.0260	0.8853 ± 0.0856	0.9229 ± 0.0477	0.9705 ± 0.0154	0.8468 ± 0.1153	0.9318 ± 0.0759	0.8814 ± 0.0644	0.7531 ± 0.1137	0.8787 ± 0.1205	0.8968 ± 0.0427	0.5531 ± 0.2242	0.9400 ± 0.1102
MS-detected serum creatinine	0.9500 ± 0.0247	0.9189 ± 0.0789	0.8803 ± 0.0800	0.9658 ± 0.0173	0.8758 ± 0.1072	0.9147 ± 0.0477	0.9306 ± 0.0385	0.9323 ± 0.0655	0.8173 ± 0.1090	0.9502 ± 0.0253	0.6854 ± 0.2396	0.9133 ± 0.0877
Sulfotyrosine	0.9553 ± 0.0325	0.9305 ± 0.0640	0.9009 ± 0.0547	0.9763 ± 0.0159	0.8537 ± 0.1174	0.9588 ± 0.0319	0.9314 ± 0.0420	0.8300 ± 0.0892	0.9287 ± 0.0906	0.9562 ± 0.0230	0.8323 ± 0.1611	0.8593 ± 0.1417

***O*-Adipoylcarnitine**	0.9463 ± 0.0308	0.9026 ± 0.0638	0.8800 ± 0.0651	0.9635 ± 0.0210	0.8200 ± 0.1302	0.9388 ± 0.0697	0.9195 ± 0.0477	0.9169 ± 0.0989	0.7767 ± 0.0949	0.9245 ± 0.0341	0.5623 ± 0.1996	0.9707 ± 0.0392
Kynurenic acid	0.9398 ± 0.0324	0.8453 ± 0.0825	0.9253 ± 0.0475	0.9643 ± 0.0153	0.6100 ± 0.1330	0.9809 ± 0.0218	0.9258 ± 0.0423	0.9200 ± 0.0910	0.7533 ± 0.0950	0.9203 ± 0.0408	0.4985 ± 0.2345	0.9573 ± 0.1585
Ratio of pseudouridine to uridine	0.9323 ± 0.0242	0.9416 ± 0.0566	0.7697 ± 0.0600	0.9414 ± 0.0199	0.7153 ± 0.1248	0.9144 ± 0.0731	0.9407 ± 0.0378	0.9262 ± 0.1002	0.8407 ± 0.0666	0.9605 ± 0.0256	0.7600 ± 0.2090	0.8560 ± 0.1786
Succinylcarnitine	0.9318 ± 0.0267	0.9053 ± 0.0666	0.8176 ± 0.0663	0.9375 ± 0.0290	0.9258 ± 0.0494	0.7991 ± 0.0610	0.8318 ± 0.0882	0.8138 ± 0.1183	0.7800 ± 0.1172	0.8780 ± 0.0498	0.8062 ± 0.1209	0.8013 ± 0.1139
L-Kynurenine	0.9303 ± 0.0277	0.9100 ± 0.0810	0.8847 ± 0.0421	0.9475 ± 0.0231	0.9021 ± 0.0674	0.8903 ± 0.0448	0.9087 ± 0.0478	0.8162 ± 0.1181	0.7787 ± 0.1388	0.9187 ± 0.0354	0.7823 ± 0.1292	0.8387 ± 0.1021
Indole-3-lactic acid	0.9272 ± 0.0327	0.8258 ± 0.0989	0.8482 ± 0.0787	0.9372 ± 0.0219	0.8784 ± 0.0743	0.8321 ± 0.0597	0.8240 ± 0.0599	0.8015 ± 0.1492	0.7613 ± 0.1258	0.8583 ± 0.0493	0.8738 ± 0.1018	0.7100 ± 0.1231
Butyrylcarnitine	0.9159 ± 0.0423	0.8268 ± 0.0939	0.8365 ± 0.0768	0.9191 ± 0.0259	0.8047 ± 0.0843	0.8709 ± 0.0575	0.7186 ± 0.0743	0.5969 ± 0.1480	0.7660 ± 0.1462	0.8027 ± 0.0623	0.6500 ± 0.1481	0.8093 ± 0.1120
Indoxyl sulfate	0.9144 ± 0.0398	0.8047 ± 0.0888	0.8771 ± 0.1088	0.8904 ± 0.0394	0.7558 ± 0.1396	0.9103 ± 0.0876	0.7888 ± 0.0745	0.6800 ± 0.1514	0.7027 ± 0.2165	0.7966 ± 0.0617	0.4977 ± 0.1812	0.9780 ± 0.0517
***α*-N-Phenylacetyl-L-glutamine**	0.9133 ± 0.0419	0.7842 ± 0.0827	0.9412 ± 0.0566	0.9106 ± 0.0308	0.7837 ± 0.0911	0.8862 ± 0.0978	0.8578 ± 0.0501	0.7669 ± 0.1006	0.8193 ± 0.1097	0.8220 ± 0.0647	0.4423 ± 0.2113	0.9700 ± 0.0791
***p*-Cresol glucuronide**	0.9082 ± 0.0419	0.8495 ± 0.0819	0.8497 ± 0.0781	0.9081 ± 0.0370	0.6184 ± 0.1795	0.9141 ± 0.1573	0.7330 ± 0.1003	0.7185 ± 0.1596	0.5993 ± 0.1664	0.7917 ± 0.0569	0.4438 ± 0.2323	0.8807 ± 0.2468
***p*-Cresol sulfate**	0.9081 ± 0.0475	0.7447 ± 0.0948	0.9353 ± 0.0561	0.8676 ± 0.0483	0.7568 ± 0.0811	0.9385 ± 0.0285	0.6847 ± 0.0882	0.5046 ± 0.1386	0.7760 ± 0.1504	0.7284 ± 0.0670	0.4869 ± 0.1595	0.8893 ± 0.1186
Homovanillic acid sulfate	0.8958 ± 0.0336	0.8305 ± 0.1085	0.7779 ± 0.0858	0.9019 ± 0.0313	0.6679 ± 0.1486	0.8926 ± 0.1065	0.8705 ± 0.0605	0.8877 ± 0.0853	0.7280 ± 0.1169	0.8984 ± 0.0418	0.4754 ± 0.2209	0.9300 ± 0.1903
Arabinose isomer	0.8852 ± 0.0409	0.8089 ± 0.1078	0.7974 ± 0.0853	0.9038 ± 0.0260	0.8532 ± 0.1562	0.6138 ± 0.3202	0.8547 ± 0.0686	0.7646 ± 0.1235	0.8387 ± 0.1337	0.9090 ± 0.0432	0.5046 ± 0.1748	0.9740 ± 0.0603
Ratio of kynurenic acid to kynurenine	0.8622 ± 0.0390	0.7221 ± 0.0836	0.8915 ± 0.0573	0.8801 ± 0.0287	0.6016 ± 0.1436	0.9297 ± 0.0539	0.8190 ± 0.0634	0.6938 ± 0.1250	0.8440 ± 0.1311	0.8547 ± 0.0555	0.5062 ± 0.1869	0.9660 ± 0.0792
L-Citrulline	0.8591 ± 0.0479	0.7663 ± 0.1196	0.8509 ± 0.0703	0.8717 ± 0.0414	0.8095 ± 0.0899	0.8159 ± 0.0556	0.8109 ± 0.0612	0.7015 ± 0.1034	0.8467 ± 0.1507	0.8743 ± 0.0500	0.6846 ± 0.1501	0.9300 ± 0.0911
Pyrocatechol sulfate	0.8535 ± 0.0462	0.7558 ± 0.1035	0.8291 ± 0.0768	0.8763 ± 0.0390	0.8563 ± 0.1198	0.6388 ± 0.2827	0.4680 ± 0.1102	0.9062 ± 0.1353	0.0607 ± 0.1059	0.6316 ± 0.0816	0.4231 ± 0.1586	0.7827 ± 0.1199
Bilirubin	0.8455 ± 0.0499	0.7932 ± 0.0894	0.7826 ± 0.0810	0.8622 ± 0.0343	0.8232 ± 0.0949	0.7841 ± 0.0943	0.6323 ± 0.0816	0.7415 ± 0.1763	0.4200 ± 0.1653	0.7246 ± 0.0748	0.6154 ± 0.1849	0.6213 ± 0.1185
L-*β*-aspartyl-L-leucine	0.8426 ± 0.0382	0.7784 ± 0.1102	0.8359 ± 0.0639	0.8671 ± 0.0344	0.8868 ± 0.0830	0.6294 ± 0.1657	0.8481 ± 0.0653	0.7546 ± 0.1458	0.7680 ± 0.1516	0.8857 ± 0.0462	0.5869 ± 0.1568	0.9420 ± 0.0728
Ratio of tyrosine to phenylalanine	0.8424 ± 0.0519	0.8121 ± 0.0991	0.7282 ± 0.1396	0.8637 ± 0.0334	0.8189 ± 0.0633	0.7785 ± 0.0652	0.8965 ± 0.0528	0.8092 ± 0.0958	0.8400 ± 0.1024	0.9233 ± 0.0334	0.7592 ± 0.1118	0.9173 ± 0.0755

L-Tryptophan	0.8367 ± 0.0492	0.8463 ± 0.0851	0.6153 ± 0.1489	0.8640 ± 0.0361	0.8258 ± 0.0877	0.7171 ± 0.1002	0.6763 ± 0.0621	0.5377 ± 0.1299	0.7393 ± 0.1617	0.7516 ± 0.0675	0.6031 ± 0.1442	0.7907 ± 0.1135
N-Acetylcarnosine	0.8137 ± 0.0561	0.7753 ± 0.1039	0.7068 ± 0.1200	0.8554 ± 0.0440	0.7647 ± 0.1190	0.7853 ± 0.0796	0.7471 ± 0.0821	0.7654 ± 0.1732	0.6167 ± 0.1468	0.8030 ± 0.0573	0.7015 ± 0.2416	0.6493 ± 0.1861

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.10 AUC of metabolites for distinguishing Stage 4 patients from Stage 1-3 patients using random forest classification and logistic regression.

Metabolites	Stage 4 vs Stages 1-3 (GFR < 30 vs GFR ≥ 30)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9817+/-0.0308	0.9330+/-0.0849	0.9814+/-0.0194	0.9989+/-0.0016	0.9640+/-0.0671	0.9755+/-0.0254	0.9760+/-0.0814	0.9520+/-0.1628	1.0000+/-0.0000	1.0000+/-0.0000	0.9580+/-0.1531	0.9964+/-0.0153
Pseudouridine	0.9925 ± 0.0197	0.9140 ± 0.0959	0.9734 ± 0.0267	0.9977 ± 0.0025	0.9980 ± 0.0199	0.9245 ± 0.0479	0.9679 ± 0.0606	0.8820 ± 0.1676	0.9832 ± 0.0219	0.9975 ± 0.0040	0.9740 ± 0.0673	0.9727 ± 0.0280
Succinyladenosine	0.9933 ± 0.0176	0.9420 ± 0.0929	0.9834 ± 0.0154	0.9994 ± 0.0010	0.9690 ± 0.0643	0.9850 ± 0.0148	0.9550 ± 0.0508	0.8860 ± 0.1241	0.9873 ± 0.0386	0.9956 ± 0.0066	0.8420 ± 0.1531	1.0000 ± 0.0000
L,L-TMAP isomer	0.9893 ± 0.0130	0.9570 ± 0.0840	0.9645 ± 0.0243	0.9955 ± 0.0042	0.9850 ± 0.0654	0.9530 ± 0.0241	0.9850 ± 0.0572	0.9580 ± 0.1394	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9840 ± 0.0784	0.9977 ± 0.0099
MS-detected serum creatinine	0.9894 ± 0.0093	0.9670 ± 0.0736	0.9748 ± 0.0147	0.9930 ± 0.0059	0.9780 ± 0.0642	0.9655 ± 0.0285	0.9760 ± 0.0814	0.9520 ± 0.1628	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9760 ± 0.1176	0.9968 ± 0.0116
Sulfotyrosine	0.9882 ± 0.0347	0.9390 ± 0.1130	0.9598 ± 0.0271	0.9939 ± 0.0047	0.8650 ± 0.1472	0.9684 ± 0.0300	0.9630 ± 0.0483	0.9140 ± 0.0990	0.9914 ± 0.0220	1.0000 ± 0.0000	0.9680 ± 0.0733	0.9727 ± 0.0380
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.9879 ± 0.0239	0.9260 ± 0.0976	0.9730 ± 0.0276	0.9972 ± 0.0027	0.9530 ± 0.0806	0.9618 ± 0.0328	0.9529 ± 0.0555	0.9060 ± 0.1112	0.9955 ± 0.0198	1.0000 ± 0.0000	0.8720 ± 0.1372	1.0000 ± 0.0000
Ratio of pseudouridine to uridine	0.9859 ± 0.0269	0.9090 ± 0.1158	0.9634 ± 0.0234	0.9958 ± 0.0040	0.9870 ± 0.0439	0.8825 ± 0.0944	0.9564 ± 0.0803	0.9080 ± 0.1610	0.9718 ± 0.0307	0.9959 ± 0.0058	0.9500 ± 0.1245	0.9586 ± 0.0347
5-Methylthio-D-ribose	0.9852 ± 0.0272	0.9210 ± 0.1089	0.9684 ± 0.0271	0.9974 ± 0.0028	0.9980 ± 0.0140	0.9491 ± 0.0300	0.9534 ± 0.0541	0.8780 ± 0.1323	0.9877 ± 0.0294	0.9973 ± 0.0042	0.8540 ± 0.1381	0.9991 ± 0.0064
L,L-TMAP	0.9825 ± 0.0340	0.9340 ± 0.1142	0.9505 ± 0.0285	0.9944 ± 0.0046	0.9340 ± 0.0874	0.9561 ± 0.0288	0.9840 ± 0.0543	0.9380 ± 0.1347	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9580 ± 0.0992	1.0000 ± 0.0000
2-Hydroxyethanesulfonate	0.9845 ± 0.0202	0.9200 ± 0.1000	0.9720 ± 0.0195	0.9952 ± 0.0045	0.9990 ± 0.0099	0.3800 ± 0.3754	0.9486 ± 0.0605	0.8900 ± 0.1338	0.9873 ± 0.0232	1.0000 ± 0.0000	0.8840 ± 0.1528	0.9700 ± 0.0597
4-Acetamidobutanoic acid	0.9814 ± 0.0292	0.9200 ± 0.0906	0.9868 ± 0.0133	0.9991 ± 0.0014	0.9640 ± 0.1044	0.8057 ± 0.2874	0.9477 ± 0.0683	0.8920 ± 0.1426	0.9918 ± 0.0260	1.0000 ± 0.0000	0.8800 ± 0.1523	0.9959 ± 0.0145
α-*N*-Phenylacetyl-L-glutamine	0.9715 ± 0.0259	0.9150 ± 0.0887	0.9705 ± 0.0380	0.9962 ± 0.0040	0.9480 ± 0.0781	0.9448 ± 0.0554	0.9335 ± 0.0755	0.7900 ± 0.1480	0.9614 ± 0.0734	0.9767 ± 0.0226	0.8240 ± 0.1504	1.0000 ± 0.0000

Ratio of kynurenine to tryptophan	0.9693 ± 0.0219	0.9450 ± 0.0517	0.9648 ± 0.0194	0.9867 ± 0.0094	0.9560 ± 0.0864	0.9336 ± 0.0655	0.9727 ± 0.0596	0.8940 ± 0.1279	0.9895 ± 0.0191	1.0000 ± 0.0000	0.8200 ± 0.1822	0.9968 ± 0.0173
2-[3-(sulfooxy)phenyl]acetic acid	0.9674 ± 0.0306	0.8960 ± 0.1019	0.9748 ± 0.0280	0.9971 ± 0.0032	0.9210 ± 0.1251	0.6755 ± 0.4440	0.9740 ± 0.0550	0.9000 ± 0.1637	0.9836 ± 0.0218	0.9979 ± 0.0038	0.7600 ± 0.1789	1.0000 ± 0.0000
***O*-Adipoylcarnitine**	0.9638 ± 0.0396	0.9160 ± 0.1074	0.8793 ± 0.0554	0.9799 ± 0.0122	0.8420 ± 0.1408	0.9173 ± 0.0824	0.9493 ± 0.0509	0.8140 ± 0.1860	0.9732 ± 0.0397	0.9861 ± 0.0152	0.8540 ± 0.1646	0.9768 ± 0.0284
Succinylcarnitine	0.9539 ± 0.0444	0.8980 ± 0.1149	0.9182 ± 0.0340	0.9843 ± 0.0089	0.9340 ± 0.0851	0.9055 ± 0.0517	0.8957 ± 0.0936	0.6940 ± 0.1969	0.9564 ± 0.0718	0.9578 ± 0.0345	0.7280 ± 0.1588	0.9836 ± 0.0269
L-*β*-aspartyl-L-leucine	0.9514 ± 0.0311	0.9070 ± 0.1051	0.8677 ± 0.0753	0.9728 ± 0.0139	0.8950 ± 0.1381	0.9016 ± 0.0492	0.9624 ± 0.0757	0.8240 ± 0.2421	0.9400 ± 0.0476	0.9825 ± 0.0139	0.9460 ± 0.1493	0.7541 ± 0.2792
Kynurenic acid	0.9475 ± 0.0499	0.8490 ± 0.1179	0.9473 ± 0.0432	0.9879 ± 0.0093	0.9250 ± 0.0973	0.7411 ± 0.3489	0.9870 ± 0.0688	0.9640 ± 0.1480	1.0000 ± 0.0000	1.0000 ± 0.0000	0.7140 ± 0.2388	0.9877 ± 0.1010
Arabinose isomer	0.9454 ± 0.0440	0.8280 ± 0.1217	0.9525 ± 0.0284	0.9766 ± 0.0134	0.9260 ± 0.1246	0.7507 ± 0.2860	0.9855 ± 0.0565	0.9440 ± 0.1551	0.9868 ± 0.0206	1.0000 ± 0.0000	0.9260 ± 0.1566	0.9141 ± 0.1946
N-Acetylcarnosine	0.9431 ± 0.0399	0.8760 ± 0.1087	0.8868 ± 0.0749	0.9702 ± 0.0220	0.9100 ± 0.0806	0.8916 ± 0.0522	0.8760 ± 0.0971	0.7320 ± 0.2195	0.9195 ± 0.1098	0.9418 ± 0.0539	0.7600 ± 0.2154	0.9486 ± 0.0673
Indoxyl sulfate	0.9429 ± 0.0497	0.9020 ± 0.1166	0.8893 ± 0.0638	0.9825 ± 0.0104	0.8630 ± 0.1055	0.9468 ± 0.0403	0.9800 ± 0.0600	0.8860 ± 0.1477	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9020 ± 0.1732	0.9873 ± 0.0335
***p*-Cresol glucuronide**	0.9424 ± 0.0504	0.8650 ± 0.1472	0.8455 ± 0.0719	0.9690 ± 0.0159	0.7580 ± 0.1234	0.9720 ± 0.0502	0.8286 ± 0.1005	0.6800 ± 0.1897	0.9955 ± 0.0151	0.9301 ± 0.0679	0.7260 ± 0.2166	0.9450 ± 0.1954
Homovanillic acid sulfate	0.9401 ± 0.0381	0.8820 ± 0.0829	0.8941 ± 0.0948	0.9756 ± 0.0165	0.8570 ± 0.1290	0.9466 ± 0.0582	0.9620 ± 0.0772	0.8780 ± 0.1741	0.9959 ± 0.0171	1.0000 ± 0.0000	0.9240 ± 0.1569	0.6623 ± 0.4233
L-Tryptophan	0.9324 ± 0.0437	0.8470 ± 0.1253	0.9586 ± 0.0285	0.9770 ± 0.0129	0.9470 ± 0.0591	0.8823 ± 0.0663	0.9488 ± 0.0520	0.8800 ± 0.1327	0.9214 ± 0.1096	0.9854 ± 0.0195	0.8880 ± 0.1451	0.9927 ± 0.0167
Ratio of kynurenic acid to kynurenine	0.9311 ± 0.0564	0.8710 ± 0.1275	0.8791 ± 0.0713	0.9667 ± 0.0219	0.9050 ± 0.1126	0.7264 ± 0.2806	0.9673 ± 0.0750	0.8700 ± 0.1752	0.9759 ± 0.0276	0.9983 ± 0.0036	0.8140 ± 0.2470	0.8891 ± 0.1770
Butyrylcarnitine	0.9219 ± 0.0548	0.8690 ± 0.1247	0.8634 ± 0.0727	0.9524 ± 0.0267	0.8610 ± 0.0835	0.9223 ± 0.0482	0.8547 ± 0.0913	0.7040 ± 0.1865	0.9732 ± 0.0511	0.9150 ± 0.0656	0.7440 ± 0.1835	0.9664 ± 0.0415
L-Valine	0.9181 ± 0.0538	0.8170 ± 0.1209	0.9259 ± 0.0394	0.9585 ± 0.0215	0.8850 ± 0.0865	0.8936 ± 0.0466	0.6429 ± 0.1260	0.4060 ± 0.2473	0.7523 ± 0.1554	0.7281 ± 0.0829	0.2340 ± 0.1651	0.9655 ± 0.0538
L-Carnitine	0.9076 ± 0.0571	0.6810 ± 0.1197	0.9948 ± 0.0215	0.9464 ± 0.0306	0.8890 ± 0.1224	0.7611 ± 0.0894	0.7720 ± 0.1050	0.5180 ± 0.2317	0.8941 ± 0.0811	0.8105 ± 0.1024	0.7140 ± 0.1606	0.8223 ± 0.0818
Ratio of tyrosine to phenylalanine	0.9019 ± 0.0484	0.8210 ± 0.1306	0.8714 ± 0.0499	0.9293 ± 0.0292	0.9270 ± 0.0947	0.8402 ± 0.0570	0.8518 ± 0.0926	0.7440 ± 0.2401	0.8818 ± 0.0564	0.8905 ± 0.0644	0.9000 ± 0.1077	0.8555 ± 0.0650
D-Glucose	0.9006 ± 0.0672	0.7790 ± 0.1802	0.8607 ± 0.0617	0.9364 ± 0.0283	0.9030 ± 0.0932	0.8164 ± 0.0728	0.5169 ± 0.1040	0.2300 ± 0.1682	0.8995 ± 0.0834	0.4265 ± 0.1026	0.2260 ± 0.1460	0.8050 ± 0.1415

Phenol sulfate	0.8992 ± 0.0533	0.8150 ± 0.1459	0.9064 ± 0.0357	0.9292 ± 0.0403	0.9040 ± 0.0799	0.8284 ± 0.0823	0.9347 ± 0.0570	0.7740 ± 0.2013	0.9523 ± 0.0465	0.9714 ± 0.0230	0.8460 ± 0.1852	0.9436 ± 0.0531
Pyrocatechol sulfate	0.8970 ± 0.0474	0.8970 ± 0.1330	0.7793 ± 0.1098	0.9330 ± 0.0256	0.7410 ± 0.1632	0.9055 ± 0.0458	0.8291 ± 0.1119	0.5840 ± 0.2310	0.9495 ± 0.0514	0.7875 ± 0.1221	0.7240 ± 0.1644	0.8909 ± 0.0884
Indole-3-lactic acid	0.8952 ± 0.0801	0.8320 ± 0.1555	0.8436 ± 0.0715	0.9403 ± 0.0319	0.9430 ± 0.0738	0.7691 ± 0.0719	0.7320 ± 0.1188	0.6660 ± 0.2546	0.7309 ± 0.1293	0.8455 ± 0.0796	0.6180 ± 0.2099	0.8814 ± 0.0567
***p*-Cresol sulfate**	0.8942 ± 0.0465	0.7620 ± 0.1522	0.9118 ± 0.0645	0.9137 ± 0.0514	0.8900 ± 0.0964	0.8155 ± 0.0732	0.9099 ± 0.0718	0.7780 ± 0.1467	0.9932 ± 0.0371	0.9492 ± 0.0467	0.8220 ± 0.1622	0.9923 ± 0.0204
Bilirubin	0.8869 ± 0.0439	0.8510 ± 0.1446	0.8173 ± 0.0666	0.9228 ± 0.0279	0.9450 ± 0.0555	0.7870 ± 0.0531	0.8309 ± 0.1032	0.5780 ± 0.1741	0.9173 ± 0.0817	0.8963 ± 0.0566	0.7280 ± 0.1638	0.8814 ± 0.0665
L-Tyrosine	0.8811 ± 0.0577	0.7830 ± 0.1436	0.8984 ± 0.0625	0.9282 ± 0.0336	0.8560 ± 0.0875	0.8707 ± 0.0605	0.8984 ± 0.0829	0.7340 ± 0.2118	0.9555 ± 0.0450	0.9214 ± 0.0577	0.8640 ± 0.1439	0.8995 ± 0.0693
L-Citrulline	0.8740 ± 0.0611	0.8500 ± 0.1360	0.7702 ± 0.0680	0.9141 ± 0.0343	0.9070 ± 0.1022	0.7302 ± 0.0752	0.9276 ± 0.0647	0.8640 ± 0.1670	0.8968 ± 0.0721	0.9704 ± 0.0233	0.8840 ± 0.1419	0.8986 ± 0.0449
L-Kynurenine	0.8739 ± 0.0514	0.8360 ± 0.1473	0.8052 ± 0.0716	0.9233 ± 0.0260	0.8420 ± 0.0951	0.8541 ± 0.0469	0.9110 ± 0.0822	0.7640 ± 0.1609	0.9518 ± 0.0609	0.9652 ± 0.0228	0.8000 ± 0.1497	0.9400 ± 0.0683
L-Leucine	0.8732 ± 0.0618	0.6240 ± 0.1443	0.9245 ± 0.0429	0.9047 ± 0.0366	0.8120 ± 0.1125	0.8216 ± 0.0615	0.7799 ± 0.0738	0.7060 ± 0.2374	0.7009 ± 0.1373	0.8104 ± 0.0571	0.8860 ± 0.1175	0.6850 ± 0.1047

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.11 AUC of selected single and multiple metabolites for distinguishing Stage 1a patients from Stages 1b-4 patients using random forest classification and logistic regression.

Metabolites	Stage 1a vs Stages 1b-4 (GFR > 120 vs GFR ≤ 120)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9626+/-0.0154	0.8807+/-0.0402	0.9560+/-0.1424	0.9648+/-0.0144	0.8909+/-0.0779	0.8460+/-0.1694	0.9399+/-0.0700	0.8561+/-0.0664	0.8800+/-0.2078	0.9597+/-0.0226	0.9587+/-0.0921	0.3580+/-0.3547
MS-detected creatinine	0.8491 ± 0.0508	0.7207 ± 0.1175	0.8760 ± 0.1201	0.8966 ± 0.0337	0.8470 ± 0.0774	0.7080 ± 0.1677	0.8787 ± 0.0777	0.9196 ± 0.0996	0.6740 ± 0.1825	0.9298 ± 0.0414	0.4352 ± 0.1679	0.9940 ± 0.0341
2-(α-D-Mannopyranosyl)-L-tryptophan	0.8878 ± 0.0659	0.8320 ± 0.1096	0.7530 ± 0.1565	0.9212 ± 0.0315	0.8932 ± 0.0922	0.6680 ± 0.2395	0.8930 ± 0.0648	0.8522 ± 0.0740	0.8060 ± 0.2521	0.9398 ± 0.0303	0.3861 ± 0.1378	1.0000 ± 0.0000
Pseudouridine	0.9148 ± 0.0543	0.9334 ± 0.0534	0.6880 ± 0.1485	0.9432 ± 0.0239	0.8134 ± 0.1081	0.8570 ± 0.1344	0.8742 ± 0.0555	0.8461 ± 0.0697	0.7280 ± 0.2458	0.9297 ± 0.0329	0.4096 ± 0.1309	1.0000 ± 0.0000
Succinyladenosine	0.8877 ± 0.0584	0.8318 ± 0.0591	0.7910 ± 0.1582	0.9147 ± 0.0306	0.8973 ± 0.1112	0.5720 ± 0.2895	0.9228 ± 0.0499	0.8657 ± 0.0674	0.8240 ± 0.1795	0.9513 ± 0.0291	0.4691 ± 0.1296	1.0000 ± 0.0000
L,L-TMAP	0.8984 ± 0.0461	0.8102 ± 0.0661	0.9020 ± 0.1319	0.9394 ± 0.0219	0.9305 ± 0.0832	0.5070 ± 0.2779	0.8686 ± 0.0887	0.8322 ± 0.1316	0.6720 ± 0.2616	0.9105 ± 0.0467	0.3961 ± 0.1607	0.9960 ± 0.0398
MS-detected creatinine reatinine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9229 ± 0.0282	0.8711 ± 0.0666	0.7220 ± 0.1467	0.9287 ± 0.0260	0.7966 ± 0.0980	0.8740 ± 0.1254	0.9181 ± 0.0426	0.9283 ± 0.0568	0.6240 ± 0.1882	0.9465 ± 0.0309	0.9043 ± 0.1120	0.5820 ± 0.3465
MS-detected creatinine reatinine + Pseudouridine	0.9296 ± 0.0249	0.8482 ± 0.0574	0.8090 ± 0.1504	0.9451 ± 0.0223	0.7632 ± 0.1081	0.9590 ± 0.0680	0.9317 ± 0.0344	0.8909 ± 0.0610	0.6740 ± 0.2239	0.9561 ± 0.0288	0.8474 ± 0.1369	0.7720 ± 0.2815
MS-detected creatinine reatinine + Succinyladenosine	0.9119 ± 0.0288	0.8198 ± 0.0541	0.7990 ± 0.1284	0.9235 ± 0.0265	0.7732 ± 0.0890	0.9260 ± 0.0986	0.9396 ± 0.0407	0.9361 ± 0.0646	0.7400 ± 0.1685	0.9526 ± 0.0275	0.8343 ± 0.1330	0.8520 ± 0.1889
MS-detected creatinine reatinine + L,L-TMAP	0.9100 ± 0.0429	0.8441 ± 0.0744	0.8770 ± 0.1441	0.9226 ± 0.0280	0.7816 ± 0.0911	0.8940 ± 0.0988	0.9168 ± 0.0517	0.8887 ± 0.0881	0.7220 ± 0.1895	0.9371 ± 0.0355	0.9017 ± 0.1158	0.6480 ± 0.2759
Succinyladenosine + Pseudouridine	0.9206 ± 0.0492	0.7950 ± 0.0929	0.8920 ± 0.1146	0.9482 ± 0.0234	0.7309 ± 0.1130	0.9270 ± 0.0798	0.9347 ± 0.0278	0.8926 ± 0.0572	0.7520 ± 0.2138	0.9501 ± 0.0309	0.7991 ± 0.1465	0.8980 ± 0.2107
Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9203 ± 0.0422	0.8239 ± 0.0883	0.8120 ± 0.1373	0.9353 ± 0.0271	0.7389 ± 0.1050	0.8920 ± 0.0891	0.8988 ± 0.0367	0.8630 ± 0.0505	0.7100 ± 0.2439	0.9346 ± 0.0386	0.8630 ± 0.1371	0.6340 ± 0.3991
Pseudouridine + L,L-TMAP	0.9419 ± 0.0222	0.8673 ± 0.0467	0.8320 ± 0.1580	0.9546 ± 0.0175	0.7348 ± 0.1107	0.9670 ± 0.0679	0.9167 ± 0.0323	0.8787 ± 0.0607	0.7160 ± 0.2564	0.9441 ± 0.0305	0.8548 ± 0.1518	0.6740 ± 0.3402
Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9159 ± 0.0387	0.8607 ± 0.0827	0.7820 ± 0.1374	0.9258 ± 0.0312	0.5914 ± 0.0741	0.9950 ± 0.0218	0.9270 ± 0.0451	0.8900 ± 0.0541	0.7920 ± 0.1853	0.9465 ± 0.0320	0.8478 ± 0.1308	0.7620 ± 0.3196
Succinyl-adenosine + L,L-TMAP	0.9432 ± 0.0217	0.8684 ± 0.0579	0.8300 ± 0.1480	0.9489 ± 0.0199	0.7602 ± 0.1314	0.9030 ± 0.1559	0.9073 ± 0.0597	0.8700 ± 0.0737	0.8120 ± 0.1894	0.9457 ± 0.0335	0.8339 ± 0.1384	0.7820 ± 0.2385
MS-detected creatinine reatinine + Succinyladenosine + Pseudouridine	0.9374 ± 0.0248	0.8952 ± 0.0520	0.7440 ± 0.1485	0.9518 ± 0.0190	0.7120 ± 0.1089	0.9860 ± 0.0583	0.9371 ± 0.0299	0.9296 ± 0.0641	0.6400 ± 0.1855	0.9561 ± 0.0291	0.7930 ± 0.1346	0.9460 ± 0.1260

MS-detected creatinine reatinine + Pseudouridine + L,L-TMAP	0.9452 ± 0.0212	0.8970 ± 0.0471	0.8150 ± 0.1590	0.9519 ± 0.0192	0.7177 ± 0.1105	0.9770 ± 0.0581	0.9297 ± 0.0352	0.8830 ± 0.0740	0.6600 ± 0.2236	0.9572 ± 0.0300	0.5709 ± 0.1526	1.0000 ± 0.0000
MS-detected creatinine reatinine + L,L-TMAP + Succinyladenosine	0.9375 ± 0.0240	0.8723 ± 0.0572	0.8430 ± 0.1557	0.9395 ± 0.0237	0.7230 ± 0.1075	0.9520 ± 0.0700	0.9293 ± 0.0467	0.9426 ± 0.0659	0.6760 ± 0.1975	0.9628 ± 0.0259	0.5970 ± 0.1340	1.0000 ± 0.0000
MS-detected creatinine + Succinyladenosine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.9248 ± 0.0291	0.8775 ± 0.0593	0.7280 ± 0.1422	0.9388 ± 0.0229	0.7266 ± 0.1041	0.9570 ± 0.0863	0.9362 ± 0.0436	0.9130 ± 0.0735	0.6840 ± 0.2063	0.9534 ± 0.0299	0.5783 ± 0.1474	1.0000 ± 0.0000
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.9375 ± 0.0240	0.8723 ± 0.0572	0.8430 ± 0.1557	0.9395 ± 0.0237	0.7230 ± 0.1075	0.9520 ± 0.0700	0.9293 ± 0.0467	0.9426 ± 0.0659	0.6760 ± 0.1975	0.9628 ± 0.0259	0.5970 ± 0.1340	1.0000 ± 0.0000
MS-detected creatinine reatinine + L,L-TMAP + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.9306 ± 0.0251	0.8511 ± 0.0731	0.8260 ± 0.1572	0.9397 ± 0.0218	0.7305 ± 0.1089	0.9500 ± 0.0854	0.9201 ± 0.0522	0.8791 ± 0.0812	0.6980 ± 0.2182	0.9512 ± 0.0343	0.5613 ± 0.1614	1.0000 ± 0.0000
MS-detected creatinine + Pseudouridine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.9331 ± 0.0244	0.8705 ± 0.0538	0.7750 ± 0.1329	0.9522 ± 0.0193	0.7152 ± 0.1087	0.9730 ± 0.0563	0.9215 ± 0.0337	0.8817 ± 0.0674	0.6340 ± 0.2228	0.9558 ± 0.0293	0.5504 ± 0.1639	1.0000 ± 0.0000

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.12 AUC of selected single and multiple metabolites for distinguishing Stage 1a patients from Stage 1b-2 patients using random forest classification and logistic regression.

Metabolites	Stage 1a vs Stages 1b-2 (GFR > 120 VS GFR= 60-120)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.8569+/-0.0749	0.9933+/-0.0663	0.5090+/-0.1167	0.8416+/-0.0698	0.9744+/-0.1121	0.4930+/-0.1329	0.9053+/-0.0712	0.8936+/-0.1461	0.6260+/-0.1847	0.9169+/-0.0544	0.6436+/-0.1503	0.9780+/-0.0795
MS-detected serum creatinine	0.7543 ± 0.0800	0.6433 ± 0.1007	0.8050 ± 0.1615	0.8077 ± 0.0612	0.5721 ± 0.1124	0.8800 ± 0.1105	0.7633 ± 0.1085	0.9091 ± 0.1150	0.6360 ± 0.2086	0.8596 ± 0.0805	0.9300 ± 0.1112	0.6560 ± 0.1768
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8111 ± 0.0737	0.8400 ± 0.0931	0.7040 ± 0.1455	0.8576 ± 0.0501	0.8604 ± 0.0840	0.6990 ± 0.1300	0.7805 ± 0.0996	0.6718 ± 0.1408	0.7940 ± 0.2340	0.8747 ± 0.0687	0.7627 ± 0.1219	0.7960 ± 0.1523
Pseudouridine	0.8672 ± 0.0721	0.8812 ± 0.0889	0.6960 ± 0.1385	0.8973 ± 0.0393	0.9312 ± 0.0505	0.6490 ± 0.1323	0.7566 ± 0.0776	0.6509 ± 0.1418	0.7600 ± 0.2433	0.8469 ± 0.0669	0.7391 ± 0.1065	0.8240 ± 0.1607
Succinyladenosine	0.8157 ± 0.0652	0.6883 ± 0.1224	0.8070 ± 0.1675	0.8472 ± 0.0528	0.6687 ± 0.1186	0.8540 ± 0.1187	0.8575 ± 0.0656	0.8191 ± 0.1434	0.7300 ± 0.1729	0.9051 ± 0.0646	0.6836 ± 0.1388	0.9260 ± 0.0966
L,L-TMAP	0.8485 ± 0.0567	0.7508 ± 0.0779	0.8120 ± 0.1762	0.8806 ± 0.0379	0.7171 ± 0.0917	0.8720 ± 0.1217	0.7736 ± 0.1154	0.8636 ± 0.1209	0.4940 ± 0.2087	0.8325 ± 0.0877	0.8455 ± 0.1203	0.6740 ± 0.1591
MS-detected Creatinine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8390 ± 0.0530	0.6833 ± 0.1090	0.8040 ± 0.1489	0.8754 ± 0.0424	0.8775 ± 0.0642	0.6660 ± 0.1751	0.8260 ± 0.0826	0.8491 ± 0.1268	0.5920 ± 0.1831	0.8929 ± 0.0524	0.8318 ± 0.1249	0.7260 ± 0.1803
MS-detected Creatinine + Pseudouridine	0.8743 ± 0.0466	0.7079 ± 0.1253	0.8130 ± 0.1376	0.9106 ± 0.0341	0.8462 ± 0.0606	0.7300 ± 0.1367	0.8494 ± 0.0629	0.8355 ± 0.1340	0.5880 ± 0.1807	0.8962 ± 0.0525	0.6682 ± 0.1571	0.8900 ± 0.1819
MS-detected Creatinine + Succinyladenosine	0.8398 ± 0.0568	0.6758 ± 0.1154	0.8010 ± 0.1439	0.8757 ± 0.0419	0.7504 ± 0.0940	0.7900 ± 0.1493	0.8687 ± 0.0717	0.8736 ± 0.1198	0.6980 ± 0.1822	0.9055 ± 0.0548	0.9473 ± 0.0805	0.6760 ± 0.1955
MS-detected Creatinine + L,L-TMAP	0.8684 ± 0.0557	0.7908 ± 0.0903	0.8180 ± 0.1615	0.8875 ± 0.0442	0.8646 ± 0.0701	0.7330 ± 0.1349	0.8340 ± 0.0793	0.9036 ± 0.0870	0.5980 ± 0.1822	0.8769 ± 0.0655	0.8291 ± 0.1300	0.7320 ± 0.1702
Succinyladenosine + Pseudouridine	0.8767 ± 0.0676	0.8337 ± 0.1128	0.7920 ± 0.1560	0.9070 ± 0.0394	0.7988 ± 0.0882	0.8240 ± 0.1150	0.8695 ± 0.0659	0.8264 ± 0.1207	0.7460 ± 0.1621	0.8942 ± 0.0650	0.7664 ± 0.1362	0.8840 ± 0.1362
Pseudouridine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8598 ± 0.0572	0.8842 ± 0.0839	0.6520 ± 0.1466	0.8891 ± 0.0476	0.8800 ± 0.0861	0.7150 ± 0.1252	0.7981 ± 0.0813	0.7091 ± 0.1091	0.7560 ± 0.2128	0.8604 ± 0.0661	0.6700 ± 0.1155	0.9680 ± 0.0786
Pseudouridine + L,L-TMAP	0.9054 ± 0.0374	0.7917 ± 0.0750	0.8060 ± 0.1593	0.9240 ± 0.0311	0.8604 ± 0.0701	0.7490 ± 0.1453	0.8342 ± 0.0908	0.7336 ± 0.1416	0.7560 ± 0.1888	0.8822 ± 0.0617	0.7036 ± 0.1713	0.8540 ± 0.1763
Succinyladenosine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8374 ± 0.0628	0.8938 ± 0.0749	0.6830 ± 0.1497	0.8663 ± 0.0522	0.8858 ± 0.0696	0.7220 ± 0.1110	0.8700 ± 0.0823	0.7645 ± 0.1299	0.8360 ± 0.1706	0.8878 ± 0.0618	0.6155 ± 0.1537	0.9320 ± 0.1174
Succinyl-adenosine + L,L-TMAP	0.9126 ± 0.0439	0.8908 ± 0.0738	0.6730 ± 0.1642	0.9116 ± 0.0362	0.7913 ± 0.1076	0.8500 ± 0.1367	0.8615 ± 0.0810	0.8036 ± 0.1371	0.7600 ± 0.1766	0.8862 ± 0.0668	0.6345 ± 0.1641	0.9060 ± 0.1482
MS-detected creatinine + Succinyladenosine + Pseudouridine	0.8882 ± 0.0449	0.8262 ± 0.0856	0.7530 ± 0.1315	0.9051 ± 0.0365	0.7871 ± 0.0799	0.8350 ± 0.1178	0.8496 ± 0.0674	0.7636 ± 0.1483	0.7240 ± 0.2311	0.8947 ± 0.0539	0.7491 ± 0.1312	0.8420 ± 0.1656

Creatinine + Pseudouridine + L,L-TMAP	0.9091 ± 0.0355		0.8379 ± 0.0721		0.7810 ± 0.1586		0.9210 ± 0.0288		0.6967 ± 0.1080		0.9570 ± 0.0738		0.8915 ± 0.0614		0.8082 ± 0.1226		0.7120 ± 0.2405		0.8844 ± 0.0536		0.6582 ± 0.1349		0.8740 ± 0.1460	
MS-detected Creatinine + L,L-TMAP + Succinyladenosine	0.9081 ± 0.0399		0.8688 ± 0.0581		0.7090 ± 0.1443		0.9019 ± 0.0354		0.7754 ± 0.0803		0.8700 ± 0.1145		0.8633 ± 0.0905		0.8945 ± 0.1245		0.5840 ± 0.2292		0.8891 ± 0.0621		0.8845 ± 0.1127		0.6680 ± 0.1503	
MS-detected Creatinine + Succinyladenosine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8849 ± 0.0485		0.9017 ± 0.0772		0.6330 ± 0.1569		0.8790 ± 0.0463		0.8250 ± 0.0784		0.7370 ± 0.1197		0.8538 ± 0.0849		0.8964 ± 0.1136		0.6700 ± 0.1841		0.8995 ± 0.0561		0.9191 ± 0.0935		0.6040 ± 0.1720	
MS-detected Creatinine + L,L-TMAP + Succinyladenosine	0.9081 ± 0.0399		0.8688 ± 0.0581		0.7090 ± 0.1443		0.9019 ± 0.0354		0.7754 ± 0.0803		0.8700 ± 0.1145		0.8633 ± 0.0905		0.8945 ± 0.1245		0.5840 ± 0.2292		0.8891 ± 0.0621		0.8845 ± 0.1127		0.6680 ± 0.1503	
MS-detected Creatinine + L,L-TMAP + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8995 ± 0.0366		0.8604 ± 0.0715		0.7380 ± 0.1660		0.9009 ± 0.0387		0.8304 ± 0.0757		0.7840 ± 0.1206		0.8413 ± 0.0830		0.8973 ± 0.1018		0.6640 ± 0.1895		0.8733 ± 0.0644		0.8936 ± 0.1106		0.6560 ± 0.1551	
MS-detected Creatinine + Pseudouridine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8858 ± 0.0412		0.8308 ± 0.0835		0.7170 ± 0.1607		0.9037 ± 0.0398		0.7879 ± 0.0894		0.8390 ± 0.1094		0.8109 ± 0.0890		0.8518 ± 0.1352		0.6280 ± 0.2079		0.8887 ± 0.0552		0.6673 ± 0.1337		0.9440 ± 0.1098	

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.13 AUC of selected single and multiple metabolites for distinguishing Stage 1b patients from Stage 2 patients using random forest classification and logistic regression.

Metabolite and model	Stage 1b vs Stages 2 (GFR = 90-120 vs GFR = 60-89)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.7955+/-0.0647	0.4660+/-0.1235	0.9389+/-0.1518	0.8171+/-0.0655	0.4433+/-0.1393	0.9578+/-0.1284	0.7678+/-0.1288	0.5360+/-0.1937	0.8520+/-0.2377	0.8100+/-0.0912	0.4940+/-0.2029	0.9460+/-0.1545
MS-detected serum creatinine	0.6805 ± 0.0938	0.7847 ± 0.1766	0.4467 ± 0.1872	0.7547 ± 0.0719	0.8227 ± 0.1276	0.4900 ± 0.1268	0.6002 ± 0.1698	0.4800 ± 0.2383	0.6840 ± 0.2656	0.7336 ± 0.1457	0.3220 ± 0.1764	0.9840 ± 0.0784
2-(α-D-Mannopyranosyl)-L-tryptophan	0.7073 ± 0.0974	0.6873 ± 0.1796	0.5978 ± 0.1904	0.8176 ± 0.0629	0.7687 ± 0.1109	0.7122 ± 0.1353	0.8698 ± 0.0918	0.8420 ± 0.2566	0.7920 ± 0.1521	0.9296 ± 0.0623	0.7360 ± 0.1830	0.8380 ± 0.1434
Pseudouridine	0.7410 ± 0.0791	0.6787 ± 0.1085	0.7722 ± 0.1527	0.8205 ± 0.0667	0.7413 ± 0.1148	0.7922 ± 0.1615	0.9822 ± 0.0307	0.9000 ± 0.1000	0.9160 ± 0.1501	0.9900 ± 0.0173	0.8700 ± 0.1634	0.9680 ± 0.0786
Succinyladenosine	0.7766 ± 0.0693	0.5560 ± 0.1212	0.8156 ± 0.1701	0.7785 ± 0.0667	0.7020 ± 0.1822	0.6733 ± 0.2282	0.8890 ± 0.0784	0.7660 ± 0.2210	0.8020 ± 0.1887	0.9424 ± 0.0535	0.5420 ± 0.2861	0.9400 ± 0.1149
L,L-TMAP	0.6801 ± 0.0780	0.5180 ± 0.1268	0.7600 ± 0.1851	0.7575 ± 0.0824	0.4387 ± 0.1016	0.9544 ± 0.0787	0.6684 ± 0.1565	0.4780 ± 0.2448	0.7780 ± 0.2575	0.8076 ± 0.0946	0.6440 ± 0.2201	0.7300 ± 0.1634
MS-detected creatinine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.7813 ± 0.0734	0.5907 ± 0.1054	0.8844 ± 0.1594	0.8627 ± 0.0494	0.6920 ± 0.1107	0.8589 ± 0.1413	0.8620 ± 0.0954	0.7620 ± 0.2293	0.8120 ± 0.1351	0.9312 ± 0.0554	0.7000 ± 0.1755	0.8840 ± 0.1239
MS-detected creatinine + Pseudouridine	0.8096 ± 0.0644	0.6507 ± 0.1383	0.8311 ± 0.1383	0.8541 ± 0.0528	0.7613 ± 0.1124	0.7778 ± 0.1523	0.9666 ± 0.0579	0.8980 ± 0.1183	0.9140 ± 0.1679	0.9440 ± 0.0625	0.6480 ± 0.2334	0.9840 ± 0.0674
MS-detected creatinine + Succinyladenosine	0.7868 ± 0.0700	0.5820 ± 0.1511	0.8800 ± 0.1437	0.7698 ± 0.0671	0.4553 ± 0.1455	0.9078 ± 0.1043	0.8562 ± 0.0803	0.7580 ± 0.1904	0.7700 ± 0.2007	0.8944 ± 0.0864	0.6740 ± 0.2556	0.8720 ± 0.1638
MS-detected creatinine + L,L-TMAP	0.7303 ± 0.0796	0.5747 ± 0.1001	0.8933 ± 0.1414	0.7933 ± 0.0578	0.5140 ± 0.1064	0.9289 ± 0.0778	0.6918 ± 0.1321	0.5560 ± 0.2070	0.7080 ± 0.2675	0.8220 ± 0.0987	0.5100 ± 0.1884	0.8980 ± 0.1483
Succinyladenosine + Pseudouridine	0.8034 ± 0.0654	0.7213 ± 0.1105	0.7400 ± 0.1581	0.7918 ± 0.0686	0.6347 ± 0.1368	0.8656 ± 0.0972	0.9860 ± 0.0305	0.8860 ± 0.1449	0.9060 ± 0.1660	0.9976 ± 0.0111	0.8160 ± 0.2239	0.9860 ± 0.0510
Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.8040 ± 0.0531	0.6893 ± 0.0910	0.8144 ± 0.1432	0.8359 ± 0.0597	0.6140 ± 0.1234	0.8567 ± 0.1147	0.9882 ± 0.0313	0.9260 ± 0.1154	0.8920 ± 0.1560	0.9636 ± 0.0463	0.7560 ± 0.2376	0.8920 ± 0.1181
Pseudouridine + L,L-TMAP	0.7763 ± 0.0653	0.7093 ± 0.0892	0.7533 ± 0.1584	0.8086 ± 0.0543	0.5667 ± 0.1305	0.9167 ± 0.0935	0.9660 ± 0.0560	0.8580 ± 0.1632	0.9240 ± 0.1544	0.9896 ± 0.0352	0.6560 ± 0.2099	0.9980 ± 0.0199
Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.7918 ± 0.0746	0.7227 ± 0.1382	0.7178 ± 0.1808	0.7781 ± 0.0676	0.4213 ± 0.1087	0.9400 ± 0.0656	0.9346 ± 0.0592	0.8660 ± 0.1498	0.8340 ± 0.1298	0.9568 ± 0.0451	0.7900 ± 0.1729	0.8880 ± 0.1107
Succinyl-adenosine + L,L-TMAP	0.7473 ± 0.0719	0.6267 ± 0.1178	0.7189 ± 0.1795	0.7424 ± 0.0667	0.5280 ± 0.1235	0.9011 ± 0.0830	0.8348 ± 0.1039	0.7200 ± 0.2383	0.7840 ± 0.1869	0.8716 ± 0.0927	0.6980 ± 0.1865	0.8800 ± 0.1356

MS-detected creatinine + Succinyladenosine + Pseudouridine	0.8215 ± 0.0589	0.7300 ± 0.0995	0.7911 ± 0.1468	0.8238 ± 0.0616	0.6940 ± 0.1222	0.7956 ± 0.1062	0.9818 ± 0.0342	0.8320 ± 0.1406	0.9500 ± 0.1034	0.9644 ± 0.0614	0.7420 ± 0.2232	0.9900 ± 0.0520
MS-detected creatinine + Pseudouridine + L,L-TMAP	0.8067 ± 0.0593	0.7187 ± 0.0955	0.7789 ± 0.1637	0.8436 ± 0.0483	0.7147 ± 0.1219	0.7967 ± 0.1187	0.9514 ± 0.0672	0.8320 ± 0.1434	0.9140 ± 0.1703	0.9492 ± 0.0662	0.6760 ± 0.2112	0.9880 ± 0.0621
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.7762 ± 0.0641	0.6473 ± 0.0954	0.8211 ± 0.1523	0.7797 ± 0.0626	0.6873 ± 0.1023	0.7567 ± 0.1537	0.8428 ± 0.0966	0.6840 ± 0.2353	0.8300 ± 0.2086	0.9024 ± 0.0962	0.6960 ± 0.2088	0.9000 ± 0.1755
MS-detected creatinine + Succinyladenosine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.7904 ± 0.0546	0.7407 ± 0.1062	0.6411 ± 0.1968	0.8347 ± 0.0635	0.7033 ± 0.1175	0.7844 ± 0.1350	0.9200 ± 0.0682	0.8640 ± 0.1546	0.8360 ± 0.1584	0.9480 ± 0.0498	0.8400 ± 0.1549	0.8860 ± 0.1030
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.7762 ± 0.0641	0.6473 ± 0.0954	0.8211 ± 0.1523	0.7797 ± 0.0626	0.6873 ± 0.1023	0.7567 ± 0.1537	0.8428 ± 0.0966	0.6840 ± 0.2353	0.8300 ± 0.2086	0.9024 ± 0.0962	0.6960 ± 0.2088	0.9000 ± 0.1755
MS-detected creatinine + L,L-TMAP + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.7726 ± 0.0630	0.6180 ± 0.0874	0.8900 ± 0.1356	0.8483 ± 0.0521	0.6833 ± 0.1186	0.8122 ± 0.1429	0.8698 ± 0.0879	0.7180 ± 0.2351	0.8000 ± 0.1855	0.9208 ± 0.0601	0.4680 ± 0.1726	0.9940 ± 0.0341
MS-detected creatinine + Pseudouridine + 2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.8313 ± 0.0533	0.7220 ± 0.1011	0.7989 ± 0.1488	0.8581 ± 0.0484	0.7280 ± 0.1243	0.8211 ± 0.1369	0.9622 ± 0.0524	0.8680 ± 0.1420	0.8800 ± 0.1549	0.9520 ± 0.0496	0.7560 ± 0.1992	0.8720 ± 0.1281

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.14 AUC of selected single and multiple metabolites for distinguishing Stages 1 patients from Stages 2-4 patients using random forest classification and logistic regression.

Metabolite	Stages 1 vs Stages 2-4 (GFR ≥ 90 vs GFR < 90)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9529+ /-0.0242	0.7721+ /-0.0619	0.9568+ /-0.0897	0.9594+ /-0.0149	0.8565+ /-0.0932	0.8600+ /-0.1112	0.9568+ /-0.0334	0.8678+ /-0.0675	0.8880+ /-0.1699	0.9737+ /-0.0175	0.6917+ /-0.2010	0.9930+ /-0.0453
MS-detected serum creatinine	0.9183 ± 0.0281	0.7244 ± 0.0708	0.9068 ± 0.0953	0.9197 ± 0.0229	0.6665 ± 0.0818	0.9642 ± 0.0464	0.9268 ± 0.0480	0.8111 ± 0.0813	0.9100 ± 0.1453	0.9518 ± 0.0280	0.6589 ± 0.2239	0.9590 ± 0.0960
2-(α-D-Mannopyranosyl)-L-tryptophan	0.9348 ± 0.0259	0.7779 ± 0.0639	0.9195 ± 0.0977	0.9383 ± 0.0197	0.8279 ± 0.0947	0.8316 ± 0.1112	0.9544 ± 0.0337	0.9394 ± 0.0940	0.8570 ± 0.0828	0.9769 ± 0.0179	0.6117 ± 0.2081	0.9640 ± 0.1171
Pseudouridine	0.9348 ± 0.0349	0.8741 ± 0.0579	0.9047 ± 0.0828	0.9526 ± 0.0184	0.7574 ± 0.1130	0.9605 ± 0.0455	0.9942 ± 0.0105	0.9661 ± 0.0359	0.9760 ± 0.0634	0.9992 ± 0.0020	0.5911 ± 0.1728	0.9980 ± 0.0199
Succinyladenosine	0.8990 ± 0.0378	0.7874 ± 0.0594	0.9016 ± 0.0917	0.9213 ± 0.0278	0.8097 ± 0.1092	0.8468 ± 0.1739	0.9715 ± 0.0314	0.9306 ± 0.0787	0.9250 ± 0.1062	0.9901 ± 0.0088	0.8250 ± 0.1479	0.9370 ± 0.1324
L,L-TMAP	0.9064 ± 0.0380	0.7809 ± 0.0571	0.9395 ± 0.0810	0.9333 ± 0.0234	0.8144 ± 0.1209	0.8321 ± 0.1899	0.9281 ± 0.0536	0.7844 ± 0.0851	0.9290 ± 0.1409	0.9509 ± 0.0252	0.6350 ± 0.2076	0.9600 ± 0.1058
MS-detected creatinine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9488 ± 0.0169	0.8224 ± 0.0561	0.9042 ± 0.1110	0.9575 ± 0.0152	0.8053 ± 0.0751	0.9621 ± 0.0610	0.9811 ± 0.0181	0.9300 ± 0.0710	0.8730 ± 0.1130	0.9738 ± 0.0191	0.8817 ± 0.1323	0.8750 ± 0.1602
MS-detected creatinine + Pseudouridine	0.9549 ± 0.0161	0.8909 ± 0.0417	0.8605 ± 0.0977	0.9596 ± 0.0135	0.8865 ± 0.0837	0.8389 ± 0.1096	0.9954 ± 0.0094	0.9711 ± 0.0328	0.9530 ± 0.1081	0.9914 ± 0.0087	0.8133 ± 0.1502	0.9710 ± 0.0941
MS-detected creatinine + Succinyladenosine	0.9317 ± 0.0238	0.8341 ± 0.0626	0.8742 ± 0.0835	0.9398 ± 0.0198	0.7932 ± 0.0876	0.9242 ± 0.0734	0.9789 ± 0.0184	0.9161 ± 0.0752	0.9120 ± 0.1211	0.9887 ± 0.0119	0.8150 ± 0.1350	0.9880 ± 0.0515
MS-detected creatinine + L,L-TMAP	0.9238 ± 0.0281	0.8297 ± 0.0525	0.8705 ± 0.1011	0.9425 ± 0.0177	0.7974 ± 0.0866	0.9016 ± 0.0949	0.9588 ± 0.0267	0.8800 ± 0.0697	0.9360 ± 0.1221	0.9706 ± 0.0211	0.6522 ± 0.2030	0.9970 ± 0.0222
Succinyladenosine + Pseudouridine	0.9422 ± 0.0197	0.8718 ± 0.0479	0.8621 ± 0.0911	0.9456 ± 0.0171	0.8824 ± 0.0808	0.8084 ± 0.1493	0.9723 ± 0.0220	0.9072 ± 0.0694	0.9230 ± 0.1103	0.9992 ± 0.0020	0.7617 ± 0.1458	0.9940 ± 0.0443
Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9490 ± 0.0199	0.8521 ± 0.0496	0.9016 ± 0.0904	0.9976 ± 0.0034	0.7944 ± 0.1726	0.9600 ± 0.1421	0.9988 ± 0.0027	0.9811 ± 0.0307	0.9480 ± 0.0889	0.9659 ± 0.0207	0.7869 ± 0.1839	0.8933 ± 0.1300
Pseudouridine + L,L-TMAP	0.9374 ± 0.0227	0.8679 ± 0.0475	0.8747 ± 0.0867	0.9529 ± 0.0161	0.8826 ± 0.0796	0.8179 ± 0.1379	0.9987 ± 0.0034	0.9794 ± 0.0357	0.9640 ± 0.0641	0.9996 ± 0.0022	0.8094 ± 0.1510	0.9790 ± 0.1061
Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9909 ± 0.0106	0.9622 ± 0.0502	0.8780 ± 0.0955	0.9940 ± 0.0070	0.8006 ± 0.1423	0.9850 ± 0.0572	0.9300 ± 0.0263	0.8432 ± 0.0677	0.8474 ± 0.1011	0.9459 ± 0.0269	0.8038 ± 0.1745	0.8613 ± 0.1349

Succinyl-adenosine + L,L-TMAP	0.9331 ± 0.0243	0.8376 ± 0.0530	0.8711 ± 0.0963	0.9356 ± 0.0198	0.7703 ± 0.0962	0.9611 ± 0.0480	0.9967 ± 0.0097	0.9689 ± 0.0579	0.9640 ± 0.0889	0.9844 ± 0.0125	0.8256 ± 0.1237	0.9730 ± 0.0676
MS-detected creatinine + Succinyladenosine + Pseudouridine	0.9573 ± 0.0165	0.8703 ± 0.0520	0.9084 ± 0.0589	0.9591 ± 0.0176	0.8591 ± 0.0762	0.8900 ± 0.0859	0.9982 ± 0.0038	0.9567 ± 0.0475	0.9840 ± 0.0463	0.9973 ± 0.0038	0.9128 ± 0.1123	0.9030 ± 0.1700
MS-detected creatinine + Pseudouridine + L,L-TMAP	0.9477 ± 0.0208	0.8635 ± 0.0548	0.8963 ± 0.0796	0.9612 ± 0.0164	0.8612 ± 0.0807	0.8926 ± 0.0947	0.9930 ± 0.0116	0.9411 ± 0.0597	0.9940 ± 0.0420	0.9898 ± 0.0114	0.7817 ± 0.1914	0.9770 ± 0.0847
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.9368 ± 0.0224	0.8124 ± 0.0524	0.9368 ± 0.0729	0.9459 ± 0.0197	0.7874 ± 0.0784	0.9558 ± 0.0530	0.9766 ± 0.0184	0.9022 ± 0.0699	0.9400 ± 0.0735	0.9888 ± 0.0123	0.8072 ± 0.1736	0.9860 ± 0.0633
MS-detected creatinine + Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9484 ± 0.0194	0.8676 ± 0.0576	0.8363 ± 0.1036	0.9589 ± 0.0171	0.8782 ± 0.0681	0.8684 ± 0.1132	0.9888 ± 0.0095	0.9494 ± 0.0504	0.9230 ± 0.0870	0.9899 ± 0.0113	0.8017 ± 0.1845	0.9790 ± 0.0637
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.9368 ± 0.0224	0.8124 ± 0.0524	0.9368 ± 0.0729	0.9459 ± 0.0197	0.7874 ± 0.0784	0.9558 ± 0.0530	0.9766 ± 0.0184	0.9022 ± 0.0699	0.9400 ± 0.0735	0.9888 ± 0.0123	0.8072 ± 0.1736	0.9860 ± 0.0633
MS-detected creatinine + L,L-TMAP + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9451 ± 0.0188	0.8253 ± 0.0478	0.9605 ± 0.0588	0.9589 ± 0.0167	0.8741 ± 0.0737	0.8521 ± 0.1238	0.9794 ± 0.0149	0.9150 ± 0.0705	0.8970 ± 0.1063	0.9803 ± 0.0145	0.8067 ± 0.1959	0.9300 ± 0.1473
MS-detected creatinine + Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9600 ± 0.0160	0.8512 ± 0.0562	0.9368 ± 0.0686	0.9688 ± 0.0151	0.8629 ± 0.0781	0.9184 ± 0.0810	0.9970 ± 0.0050	0.9683 ± 0.0498	0.9480 ± 0.0741	0.9913 ± 0.0092	0.7689 ± 0.1989	0.9710 ± 0.0840

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.15 AUC of selected single and multiple metabolites for distinguishing Stages 1-2 patients from Stages 3-4 patients using random forest classification and logistic regression.

Metabolite	Stages 1-2 vs Stages 3-4 (GFR ≥ 60 vs GFR < 60)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9715 ± 0.0306	0.9542 ± 0.0754	0.8818 ± 0.0757	0.9901 ± 0.0060	0.9347 ± 0.0752	0.9003 ± 0.0979	0.9665 ± 0.0340	0.9192 ± 0.1073	0.8600 ± 0.0978	0.9769 ± 0.0163	0.7262 ± 0.2228	0.9447 ± 0.0725
MS-detected serum creatinine	0.9500 ± 0.0247	0.9189 ± 0.0789	0.8803 ± 0.0800	0.9658 ± 0.0173	0.8758 ± 0.1072	0.9147 ± 0.0477	0.9306 ± 0.0385	0.9323 ± 0.0655	0.8173 ± 0.1090	0.9502 ± 0.0253	0.6854 ± 0.2396	0.9133 ± 0.0877
2-(α-D-Mannopyranosyl)-L-tryptophan	0.9715 ± 0.0254	0.8921 ± 0.0760	0.9259 ± 0.0546	0.9841 ± 0.0092	0.9779 ± 0.0466	0.8053 ± 0.1143	0.9378 ± 0.0347	0.8492 ± 0.1569	0.7853 ± 0.1274	0.9402 ± 0.0315	0.6854 ± 0.2157	0.9000 ± 0.1669
Pseudouridine	0.9850 ± 0.0168	0.9242 ± 0.0696	0.9321 ± 0.0509	0.9921 ± 0.0057	0.9963 ± 0.0226	0.8121 ± 0.1213	0.9492 ± 0.0361	0.8923 ± 0.0973	0.8947 ± 0.0817	0.9724 ± 0.0195	0.8069 ± 0.1935	0.8767 ± 0.1425
Succinyladenosine	0.9581 ± 0.0276	0.9116 ± 0.0755	0.8412 ± 0.0774	0.9648 ± 0.0141	0.9163 ± 0.1114	0.7968 ± 0.1588	0.9118 ± 0.0437	0.8992 ± 0.0861	0.8427 ± 0.1005	0.9522 ± 0.0292	0.8346 ± 0.1792	0.8380 ± 0.1494
L,L-TMAP	0.9691 ± 0.0199	0.9495 ± 0.0566	0.8512 ± 0.0737	0.9833 ± 0.0089	0.9411 ± 0.0806	0.8418 ± 0.1369	0.9151 ± 0.0459	0.8446 ± 0.1109	0.8580 ± 0.0878	0.9438 ± 0.0268	0.6962 ± 0.2467	0.8980 ± 0.1224
MS-detected creatinine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9814 ± 0.0094	0.8911 ± 0.0753	0.9282 ± 0.0466	0.9860 ± 0.0090	0.9205 ± 0.0972	0.9159 ± 0.0659	0.9493 ± 0.0271	0.8669 ± 0.1225	0.8720 ± 0.0770	0.9668 ± 0.0198	0.8569 ± 0.1949	0.8033 ± 0.1737
MS-detected creatinine + Pseudouridine	0.9898 ± 0.0062	0.9179 ± 0.0597	0.9347 ± 0.0400	0.9905 ± 0.0060	0.9774 ± 0.0628	0.8800 ± 0.0769	0.9550 ± 0.0264	0.8977 ± 0.1061	0.8987 ± 0.0797	0.9741 ± 0.0161	0.8115 ± 0.1870	0.8967 ± 0.1156
MS-detected creatinine + Succinyladenosine	0.9726 ± 0.0124	0.8553 ± 0.0830	0.9235 ± 0.0489	0.9754 ± 0.0117	0.8795 ± 0.1309	0.9050 ± 0.0664	0.9553 ± 0.0294	0.9008 ± 0.0787	0.8840 ± 0.0798	0.9583 ± 0.0252	0.8346 ± 0.1831	0.8693 ± 0.1123
MS-detected creatinine + L,L-TMAP	0.9795 ± 0.0110	0.9121 ± 0.0614	0.9044 ± 0.0479	0.9804 ± 0.0103	0.8747 ± 0.1131	0.9182 ± 0.0573	0.9426 ± 0.0317	0.8769 ± 0.0991	0.8727 ± 0.0766	0.9730 ± 0.0182	0.8669 ± 0.1988	0.8047 ± 0.1794
Succinyladenosine + Pseudouridine	0.9819 ± 0.0098	0.9053 ± 0.0845	0.9115 ± 0.0477	0.9885 ± 0.0068	0.9553 ± 0.0872	0.8835 ± 0.0770	0.9414 ± 0.0317	0.8423 ± 0.0919	0.8900 ± 0.0824	0.9597 ± 0.0233	0.7746 ± 0.1785	0.9047 ± 0.1008
Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9883 ± 0.0072	0.9226 ± 0.0611	0.9118 ± 0.0452	0.9900 ± 0.0056	0.9705 ± 0.0615	0.8703 ± 0.0732	0.9492 ± 0.0259	0.8746 ± 0.1001	0.8813 ± 0.0820	0.9659 ± 0.0207	0.7869 ± 0.1839	0.8933 ± 0.1300
Pseudouridine + L,L-TMAP	0.9879 ± 0.0082	0.9063 ± 0.0814	0.9397 ± 0.0398	0.9932 ± 0.0048	0.9668 ± 0.0769	0.8918 ± 0.0758	0.9594 ± 0.0261	0.8762 ± 0.0978	0.9000 ± 0.0745	0.9706 ± 0.0191	0.8177 ± 0.1936	0.8987 ± 0.1157
Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9747 ± 0.0123	0.8679 ± 0.0724	0.9121 ± 0.0546	0.9807 ± 0.0099	0.8742 ± 0.1372	0.8962 ± 0.0830	0.9559 ± 0.0234	0.8662 ± 0.0927	0.8560 ± 0.1014	0.9459 ± 0.0269	0.8038 ± 0.1745	0.8613 ± 0.1349

Succinyl-adenosine + L,L-TMAP	0.9766 ± 0.0104	0.8721 ± 0.0820	0.9076 ± 0.0539	0.9799 ± 0.0093	0.8400 ± 0.1494	0.9206 ± 0.0689	0.9553 ± 0.0287	0.8662 ± 0.0914	0.8847 ± 0.0711	0.9504 ± 0.0269	0.8331 ± 0.1846	0.8480 ± 0.1337
MS-detected creatinine + Succinyladenosine + Pseudouridine	0.9837 ± 0.0082	0.8832 ± 0.0867	0.9271 ± 0.0439	0.9875 ± 0.0070	0.9653 ± 0.0788	0.8985 ± 0.0530	0.9586 ± 0.0275	0.8969 ± 0.0732	0.9060 ± 0.0865	0.9612 ± 0.0208	0.9015 ± 0.1389	0.8020 ± 0.1437
MS-detected creatinine + Pseudouridine + L,L-TMAP	0.9874 ± 0.0075	0.8989 ± 0.0914	0.9450 ± 0.0403	0.9921 ± 0.0056	0.9716 ± 0.0696	0.9062 ± 0.0548	0.9677 ± 0.0189	0.8923 ± 0.0857	0.8787 ± 0.0780	0.9722 ± 0.0189	0.9300 ± 0.1157	0.7700 ± 0.1797
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.9805 ± 0.0096	0.9174 ± 0.0687	0.8941 ± 0.0487	0.9824 ± 0.0089	0.8721 ± 0.1255	0.9353 ± 0.0508	0.9539 ± 0.0271	0.8915 ± 0.0800	0.8993 ± 0.0751	0.9526 ± 0.0231	0.9338 ± 0.0948	0.7353 ± 0.1688
MS-detected creatinine + Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9802 ± 0.0077	0.8805 ± 0.0815	0.9088 ± 0.0527	0.9848 ± 0.0083	0.8963 ± 0.1123	0.9282 ± 0.0513	0.9603 ± 0.0293	0.9015 ± 0.0785	0.8853 ± 0.0885	0.9503 ± 0.0242	0.9200 ± 0.1151	0.7213 ± 0.1844
MS-detected creatinine + L,L-TMAP + Succinyladenosine	0.9805 ± 0.0096	0.9174 ± 0.0687	0.8941 ± 0.0487	0.9824 ± 0.0089	0.8721 ± 0.1255	0.9353 ± 0.0508	0.9539 ± 0.0271	0.8915 ± 0.0800	0.8993 ± 0.0751	0.9526 ± 0.0231	0.9338 ± 0.0948	0.7353 ± 0.1688
MS-detected creatinine + L,L-TMAP + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9862 ± 0.0077	0.9042 ± 0.0796	0.9282 ± 0.0349	0.9890 ± 0.0073	0.9074 ± 0.1045	0.9462 ± 0.0370	0.9573 ± 0.0238	0.8715 ± 0.1003	0.8780 ± 0.0778	0.9690 ± 0.0182	0.8362 ± 0.1839	0.8673 ± 0.1220
MS-detected creatinine + Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9896 ± 0.0066	0.9137 ± 0.0786	0.9309 ± 0.0458	0.9923 ± 0.0052	0.9832 ± 0.0470	0.8974 ± 0.0510	0.9624 ± 0.0258	0.9046 ± 0.1136	0.8653 ± 0.1028	0.9735 ± 0.0188	0.9162 ± 0.1337	0.7513 ± 0.1957

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.16 AUC of selected single and multiple metabolites for distinguishing Stages 1-3 patients from Stage 4 patients using random forest classification and logistic regression

Top metabolite	Stages 1-3 vs Stages 4 (GFR ≥ 30 vs GFR < 30)											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Clinically measured serum creatinine	0.9817+/-0.0308	0.9330+/-0.0849	0.9814+/-0.0194	0.9989+/-0.0016	0.9640+/-0.0671	0.9755+/-0.0254	0.9760+/-0.0814	0.9520+/-0.1628	1.0000+/-0.0000	1.0000+/-0.0000	0.9580+/-0.1531	0.9964+/-0.0153
MS-detected serum creatinine	0.9894 ± 0.0093	0.9670 ± 0.0736	0.9748 ± 0.0147	0.9930 ± 0.0059	0.9780 ± 0.0642	0.9655 ± 0.0285	0.9760 ± 0.0814	0.9520 ± 0.1628	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9760 ± 0.1176	0.9968 ± 0.0116
2-(α-D-Mannopyranosyl)-L-tryptophan	0.9879 ± 0.0239	0.9260 ± 0.0976	0.9730 ± 0.0276	0.9972 ± 0.0027	0.9530 ± 0.0806	0.9618 ± 0.0328	0.9529 ± 0.0555	0.9060 ± 0.1112	0.9955 ± 0.0198	1.0000 ± 0.0000	0.8720 ± 0.1372	1.0000 ± 0.0000
Pseudouridine	0.9925 ± 0.0197	0.9140 ± 0.0959	0.9734 ± 0.0267	0.9977 ± 0.0025	0.9980 ± 0.0199	0.9245 ± 0.0479	0.9679 ± 0.0606	0.8820 ± 0.1676	0.9832 ± 0.0219	0.9975 ± 0.0040	0.9740 ± 0.0673	0.9727 ± 0.0280
Succinyladenosine	0.9933 ± 0.0176	0.9420 ± 0.0929	0.9834 ± 0.0154	0.9994 ± 0.0010	0.9690 ± 0.0643	0.9850 ± 0.0148	0.9550 ± 0.0508	0.8860 ± 0.1241	0.9873 ± 0.0386	0.9956 ± 0.0066	0.8420 ± 0.1531	1.0000 ± 0.0000
L,L-TMAP	0.9825 ± 0.0340	0.9340 ± 0.1142	0.9505 ± 0.0285	0.9944 ± 0.0046	0.9340 ± 0.0874	0.9561 ± 0.0288	0.9840 ± 0.0543	0.9380 ± 0.1347	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9580 ± 0.0992	1.0000 ± 0.0000
MS-detected creatinine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9953 ± 0.0164	0.9100 ± 0.1034	0.9875 ± 0.0141	0.9970 ± 0.0029	0.9620 ± 0.0892	0.9741 ± 0.0196	0.9940 ± 0.0310	0.9120 ± 0.1275	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9780 ± 0.0795	0.9823 ± 0.0287
MS-detected creatinine + pseudouridine	0.9950 ± 0.0141	0.9180 ± 0.1004	0.9875 ± 0.0145	0.9969 ± 0.0030	0.9600 ± 0.0721	0.9743 ± 0.0184	0.9920 ± 0.0392	0.9340 ± 0.1498	0.9977 ± 0.0099	1.0000 ± 0.0000	0.9840 ± 0.0612	0.9732 ± 0.0328
MS-detected creatinine + succinyladenosine	0.9990 ± 0.0024	0.9340 ± 0.1051	0.9886 ± 0.0146	0.9977 ± 0.0024	0.9980 ± 0.0199	0.9680 ± 0.0236	0.9960 ± 0.0196	0.9020 ± 0.1400	0.9995 ± 0.0045	1.0000 ± 0.0000	1.0000 ± 0.0000	0.6345 ± 0.2829
MS-detected creatinine + L,L-TMAP	0.9955 ± 0.0049	0.9080 ± 0.1294	0.9802 ± 0.0160	0.9952 ± 0.0043	0.9990 ± 0.0099	0.9355 ± 0.0465	0.9920 ± 0.0392	0.9580 ± 0.1274	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9740 ± 0.1045	0.9932 ± 0.0175
Succinyladenosine + pseudouridine	0.9990 ± 0.0015	0.9090 ± 0.1150	0.9898 ± 0.0141	0.9999 ± 0.0004	0.9980 ± 0.0199	0.9616 ± 0.0270	0.9851 ± 0.0344	0.8880 ± 0.1336	0.9868 ± 0.0225	0.9978 ± 0.0039	1.0000 ± 0.0000	0.7018 ± 0.2406
Pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9921 ± 0.0226	0.8760 ± 0.1106	0.9880 ± 0.0177	0.9989 ± 0.0016	0.8890 ± 0.1067	0.9859 ± 0.0195	0.9832 ± 0.0415	0.8960 ± 0.1341	0.9932 ± 0.0175	1.0000 ± 0.0000	0.9560 ± 0.0828	0.9636 ± 0.0407
Pseudouridine + L,L-TMAP	0.9956 ± 0.0124	0.8770 ± 0.1190	0.9811 ± 0.0206	0.9987 ± 0.0016	0.9650 ± 0.0669	0.9702 ± 0.0220	0.9900 ± 0.0412	0.9340 ± 0.1328	0.9986 ± 0.0078	1.0000 ± 0.0000	0.9820 ± 0.0638	0.9750 ± 0.0393
Succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9988 ± 0.0017	0.8920 ± 0.1093	0.9907 ± 0.0129	0.9994 ± 0.0010	0.9240 ± 0.1040	0.9893 ± 0.0126	0.9576 ± 0.0499	0.9080 ± 0.1146	0.9977 ± 0.0099	0.9965 ± 0.0056	0.9940 ± 0.0341	0.5018 ± 0.3906
Succinyl-adenosine + L,L-TMAP	0.9977 ± 0.0054	0.9040 ± 0.1113	0.9857 ± 0.0169	0.9979 ± 0.0022	0.9930 ± 0.0406	0.9573 ± 0.0299	0.9920 ± 0.0271	0.9080 ± 0.1146	1.0000 ± 0.0000	0.9998 ± 0.0013	0.9960 ± 0.0280	0.5432 ± 0.3547

MS-detected creatinine + succinyladenosine + pseudouridine	0.9993 ± 0.0014	0.9390 ± 0.1019	0.9850 ± 0.0158	0.9986 ± 0.0019	1.0000 ± 0.0000	0.9718 ± 0.0176	0.9900 ± 0.0300	0.8720 ± 0.1537	0.9982 ± 0.0089	1.0000 ± 0.0000	0.9640 ± 0.0768	0.9700 ± 0.0296
MS-detected creatinine + pseudouridine + L,L-TMAP	0.9969 ± 0.0057	0.9110 ± 0.1057	0.9809 ± 0.0205	0.9974 ± 0.0027	1.0000 ± 0.0000	0.9564 ± 0.0220	0.9800 ± 0.0600	0.8960 ± 0.1907	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9860 ± 0.0707	0.9900 ± 0.0219
MS-detected creatinine + L,L-TMAP + succinyladenosine	0.9988 ± 0.0023	0.9560 ± 0.0791	0.9807 ± 0.0194	0.9975 ± 0.0026	1.0000 ± 0.0000	0.9675 ± 0.0219	0.9890 ± 0.0343	0.8980 ± 0.1709	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9580 ± 0.0815	0.9973 ± 0.0126
MS-detected creatinine + succinyladenosine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9992 ± 0.0016	0.9410 ± 0.1078	0.9843 ± 0.0160	0.9981 ± 0.0022	1.0000 ± 0.0000	0.9730 ± 0.0187	0.9898 ± 0.0300	0.8760 ± 0.1379	0.9995 ± 0.0045	1.0000 ± 0.0000	0.9440 ± 0.0898	0.9823 ± 0.0294
MS-detected creatinine + L,L-TMAP + succinyladenosine	0.9988 ± 0.0023	0.9560 ± 0.0791	0.9807 ± 0.0194	0.9975 ± 0.0026	1.0000 ± 0.0000	0.9675 ± 0.0219	0.9890 ± 0.0343	0.8980 ± 0.1709	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9580 ± 0.0815	0.9973 ± 0.0126
MS-detected creatinine + L,L-TMAP + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9975 ± 0.0030	0.9160 ± 0.1093	0.9798 ± 0.0195	0.9966 ± 0.0033	0.9980 ± 0.0140	0.9536 ± 0.0240	0.9880 ± 0.0382	0.8840 ± 0.1654	1.0000 ± 0.0000	1.0000 ± 0.0000	0.9720 ± 0.0849	0.9968 ± 0.0116
MS-detected creatinine + pseudouridine + 2-(α-D-Mannopyranosyl)-L-tryptophan	0.9963 ± 0.0089	0.8970 ± 0.1044	0.9814 ± 0.0207	0.9974 ± 0.0026	0.9960 ± 0.0242	0.9525 ± 0.0286	0.9880 ± 0.0382	0.8780 ± 0.1467	0.9995 ± 0.0045	1.0000 ± 0.0000	0.9760 ± 0.0709	0.9800 ± 0.0304

AUC, area under curve of ROC; Sens: Sensitivity; Spec: Specificity.

Table S5.17 AUC of metabolites for distinguishing UACR ≥ 30 from UACR < 30 of diabetic patients using logistic regression and random forest classification (AUC > 0.70).

Top metabolite	UACR ≥ 30 VS UACR <30											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Pseudouridine	0.8375 ± 0.0382	0.7126 ± 0.0644	0.9320 ± 0.0805	0.8511 ± 0.0352	0.5544 ± 0.0887	1.0000 ± 0.0000	0.7341 ± 0.0566	0.5939 ± 0.0977	0.8480 ± 0.1360	0.8013 ± 0.0567	0.5717 ± 0.1308	0.9370 ± 0.0868
2-(α-D-Mannopyranosyl)-L-tryptophan	0.8293 ± 0.0358	0.7262 ± 0.0545	0.8480 ± 0.1376	0.8595 ± 0.0339	0.5667 ± 0.1071	0.9987 ± 0.0133	0.7106 ± 0.0720	0.6239 ± 0.1291	0.7170 ± 0.2205	0.7991 ± 0.0577	0.5772 ± 0.1529	0.8820 ± 0.1519
Succinyladenosine	0.8227 ± 0.0513	0.6549 ± 0.0813	0.8953 ± 0.0999	0.8454 ± 0.0376	0.4705 ± 0.1115	0.9940 ± 0.0191	0.7416 ± 0.0765	0.5511 ± 0.1020	0.8860 ± 0.1463	0.7749 ± 0.0605	0.5083 ± 0.0954	0.9800 ± 0.0616
4-Acetamidobutanoic acid	0.8109 ± 0.0474	0.6928 ± 0.0690	0.8260 ± 0.1665	0.8419 ± 0.0389	0.4510 ± 0.1469	0.9967 ± 0.0197	0.7380 ± 0.0700	0.5633 ± 0.1160	0.8510 ± 0.1670	0.7804 ± 0.0604	0.3461 ± 0.1204	0.9990 ± 0.0099
L,L-TMAP	0.7843 ± 0.0554	0.6041 ± 0.1072	0.8647 ± 0.1263	0.8233 ± 0.0406	0.7762 ± 0.1330	0.6373 ± 0.2764	0.6673 ± 0.0796	0.5772 ± 0.1814	0.6600 ± 0.2010	0.7589 ± 0.0653	0.5594 ± 0.1915	0.7940 ± 0.1495
Sulfotyrosine	0.7763 ± 0.0508	0.5759 ± 0.0967	0.8960 ± 0.1125	0.7872 ± 0.0414	0.4744 ± 0.0798	0.9973 ± 0.0161	0.8231 ± 0.0641	0.7650 ± 0.1152	0.7230 ± 0.1737	0.8228 ± 0.0554	0.6006 ± 0.1188	0.9180 ± 0.0953
Clinically measured serum_creatinine (mg/dL)	0.7738 ± 0.0459	0.5867 ± 0.0781	0.9000 ± 0.1065	0.8036 ± 0.0367	0.4844 ± 0.0927	0.9873 ± 0.0308	0.7054 ± 0.0822	0.4372 ± 0.1385	0.8860 ± 0.1679	0.7521 ± 0.0686	0.5672 ± 0.1524	0.8000 ± 0.1364
MS-detected creatinine	0.7726 ± 0.0499	0.6703 ± 0.0676	0.8260 ± 0.1343	0.7646 ± 0.0384	0.5559 ± 0.0820	0.9547 ± 0.0639	0.6770 ± 0.0775	0.5017 ± 0.1180	0.8740 ± 0.1641	0.7329 ± 0.0664	0.3178 ± 0.0846	1.0000 ± 0.0000
L,L-TMAP isomer	0.7535 ± 0.0539	0.5810 ± 0.1176	0.7993 ± 0.1371	0.8083 ± 0.0402	0.7885 ± 0.1194	0.5960 ± 0.2508	0.6783 ± 0.0866	0.4889 ± 0.1714	0.7620 ± 0.1765	0.7535 ± 0.0680	0.2867 ± 0.0877	0.9990 ± 0.0099
2-Hydroxyethanesulfonate	0.7248 ± 0.0561	0.5785 ± 0.0943	0.8200 ± 0.1129	0.7682 ± 0.0428	0.4000 ± 0.1217	0.9500 ± 0.0502	0.6796 ± 0.0830	0.6978 ± 0.1289	0.5990 ± 0.1622	0.7919 ± 0.0595	0.3694 ± 0.1386	0.9420 ± 0.0737
1,5-Anhydro-D-glucitol	0.4544 ± 0.0668	0.8915 ± 0.0746	0.0940 ± 0.0828	0.4518 ± 0.0539	0.9756 ± 0.0416	0.0153 ± 0.0387	0.5109 ± 0.0790	0.8300 ± 0.1283	0.2140 ± 0.1364	0.5461 ± 0.1071	0.9656 ± 0.0532	0.1410 ± 0.1087

Table S5.18 AUC of significant metabolites for distinguishing UACR = 30-300 from UACR < 30 of diabetic patients using random forest classification and logistic regression (AUC > 0.70).

Top metabolite	UACR = 30-300 vs UACR < 30											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Pseudouridine	0.7011 ± 0.0755	0.9193 ± 0.1156	0.4858 ± 0.1065	0.7061 ± 0.0636	0.8107 ± 0.1203	0.5337 ± 0.1323	0.4279 ± 0.0996	0.4478 ± 0.2021	0.4260 ± 0.1770	0.5699 ± 0.1115	0.9956 ± 0.0218	0.0010 ± 0.0099
Succinyladenosine	0.6857 ± 0.0761	0.8293 ± 0.1367	0.4700 ± 0.1300	0.7124 ± 0.0580	0.6320 ± 0.1851	0.6579 ± 0.1388	0.4643 ± 0.1113	0.3767 ± 0.1825	0.5800 ± 0.2078	0.5164 ± 0.1101	0.6400 ± 0.1906	0.3910 ± 0.2210
2-(α-D-Mannopyranosyl)-L-tryptophan	0.6854 ± 0.0673	0.8813 ± 0.1371	0.4716 ± 0.1063	0.7135 ± 0.0652	0.6840 ± 0.0996	0.6532 ± 0.1018	0.4045 ± 0.0996	0.7933 ± 0.1833	0.1580 ± 0.1266	0.5537 ± 0.1142	0.9733 ± 0.0687	0.0290 ± 0.0816
4-Acetamidobutanoic acid	0.6762 ± 0.0716	0.9073 ± 0.1188	0.3968 ± 0.0860	0.6888 ± 0.0582	0.8833 ± 0.0845	0.4516 ± 0.1104	0.4212 ± 0.1018	0.6267 ± 0.1816	0.3360 ± 0.1513	0.5232 ± 0.1157	0.9956 ± 0.0218	0.0000 ± 0.0000
MS-detected creatinine	0.6649 ± 0.0767	0.7773 ± 0.1570	0.5426 ± 0.1347	0.5782 ± 0.0760	0.8360 ± 0.1450	0.3574 ± 0.1166	0.4443 ± 0.1052	0.4344 ± 0.1982	0.4790 ± 0.1762	0.4299 ± 0.0832	0.9567 ± 0.1029	0.0150 ± 0.0517
Serum_creatinine (mg/dL)	0.6329 ± 0.0634	0.9160 ± 0.0862	0.3037 ± 0.1256	0.6445 ± 0.0575	0.8347 ± 0.1216	0.4358 ± 0.1200	0.5053 ± 0.1192	0.8511 ± 0.1772	0.1940 ± 0.1502	0.4907 ± 0.1190	0.5389 ± 0.2145	0.4940 ± 0.2541
L,L-TMAP	0.6328 ± 0.0744	0.7660 ± 0.1352	0.4795 ± 0.1331	0.6679 ± 0.0660	0.8020 ± 0.1308	0.4842 ± 0.1104	0.4517 ± 0.1010	0.8544 ± 0.1733	0.0900 ± 0.1187	0.4393 ± 0.1000	0.5889 ± 0.1882	0.3790 ± 0.1935
Sulfotyrosine	0.6302 ± 0.0632	0.7933 ± 0.1604	0.4184 ± 0.1288	0.6209 ± 0.0656	0.8753 ± 0.1153	0.3932 ± 0.1145	0.6316 ± 0.1255	0.5356 ± 0.1862	0.6640 ± 0.2057	0.6740 ± 0.1125	0.4789 ± 0.1762	0.8370 ± 0.1128
L,L-TMAP isomer	0.5648 ± 0.0785	0.5533 ± 0.1545	0.5753 ± 0.1480	0.6387 ± 0.0656	0.7387 ± 0.1138	0.5174 ± 0.1203	0.4719 ± 0.1120	0.8133 ± 0.1805	0.1120 ± 0.1451	0.4519 ± 0.1066	0.5833 ± 0.2105	0.3950 ± 0.2174
2-Hydroxyethanesulfonate	0.5331 ± 0.0885	0.7307 ± 0.1630	0.3468 ± 0.1281	0.5843 ± 0.0773	0.9140 ± 0.0999	0.2032 ± 0.1285	0.4554 ± 0.1112	0.6444 ± 0.1963	0.3290 ± 0.1669	0.4847 ± 0.1370	0.8011 ± 0.2190	0.1940 ± 0.2249

Table S5.19 AUC of two metabolite models using significant metabolites for distinguishing UACR = 30-300 from UACR < 30 of diabetic patients using random forest classification and logistic regression

Top metabolite models		UACR = 30-300 vs UACR <30											
		Discovery set						Validation set					
		Random Forest			Logistic Regression			Random Forest			Logistic Regression		
		AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Sulfotyrosine	**Succinyladenosine**	0.7065 ± 0.0763	0.6633 ± 0.1539	0.6358 ± 0.1421	0.6727 ± 0.0695	0.8413 ± 0.1173	0.5163 ± 0.0984	0.6159 ± 0.1085	0.4900 ± 0.1580	0.6930 ± 0.1620	0.6528 ± 0.0795	0.4344 ± 0.1626	0.8230 ± 0.1295
Sulfotyrosine	**Pseudouridine**	0.6908 ± 0.0636	0.7653 ± 0.1648	0.5489 ± 0.1221	0.6899 ± 0.0745	0.9040 ± 0.0886	0.4626 ± 0.0994	0.5776 ± 0.1010	0.5167 ± 0.1566	0.6480 ± 0.1879	0.6464 ± 0.0970	0.4722 ± 0.1688	0.8060 ± 0.1362
Sulfotyrosine	**MS-detected creatinine**	0.6708 ± 0.0654	0.5993 ± 0.1926	0.6526 ± 0.1114	0.6110 ± 0.0690	0.9567 ± 0.0621	0.2679 ± 0.0991	0.6187 ± 0.1237	0.5289 ± 0.1686	0.6640 ± 0.1858	0.6573 ± 0.0964	0.4767 ± 0.1628	0.7880 ± 0.1437
Sulfotyrosine	**2-(α-D-Mannopyranosyl)-L-tryptophan**	0.6684 ± 0.0630	0.7440 ± 0.1548	0.5147 ± 0.1290	0.6881 ± 0.0760	0.9440 ± 0.0691	0.3953 ± 0.0994	0.5594 ± 0.1011	0.4722 ± 0.1535	0.6650 ± 0.1693	0.6412 ± 0.0996	0.6111 ± 0.1544	0.6670 ± 0.1844
Sulfotyrosine	**4-Acetamidobutanoic acid**	0.6605 ± 0.0681	0.8520 ± 0.1098	0.4379 ± 0.1031	0.6771 ± 0.0679	0.9407 ± 0.0816	0.4116 ± 0.0840	0.6063 ± 0.0996	0.4989 ± 0.1710	0.6590 ± 0.1801	0.6420 ± 0.0951	0.5356 ± 0.1574	0.7370 ± 0.1641
Sulfotyrosine	**L,L-TMAP**	0.6526 ± 0.0597	0.8113 ± 0.1307	0.4505 ± 0.1226	0.6475 ± 0.0740	0.8920 ± 0.0770	0.4011 ± 0.0926	0.5967 ± 0.0993	0.5189 ± 0.1441	0.6780 ± 0.1572	0.6519 ± 0.0924	0.5011 ± 0.1461	0.7940 ± 0.1399
Sulfotyrosine	**L,L-TMAP isomer**	0.6188 ± 0.0637	0.7913 ± 0.1499	0.4179 ± 0.1176	0.6242 ± 0.0677	0.9133 ± 0.0775	0.3279 ± 0.0896	0.5724 ± 0.0944	0.5200 ± 0.1448	0.6030 ± 0.1775	0.6574 ± 0.0948	0.4889 ± 0.1625	0.7700 ± 0.1552
Sulfotyrosine	**2-Hydroxyethanesulfonate**	0.5927 ± 0.0565	0.6107 ± 0.1602	0.5121 ± 0.1302	0.5987 ± 0.0767	0.9260 ± 0.0938	0.3111 ± 0.0942	0.5714 ± 0.1074	0.5078 ± 0.1657	0.6520 ± 0.1723	0.6177 ± 0.0900	0.4878 ± 0.1498	0.7370 ± 0.1623
2-Hydroxyethanesulfonate	**4-Acetamidobutanoic acid**	0.6418 ± 0.0747	0.7213 ± 0.1559	0.4989 ± 0.1425	0.6729 ± 0.0694	0.8600 ± 0.1141	0.4468 ± 0.1006	0.4259 ± 0.1135	0.5822 ± 0.1827	0.3270 ± 0.1714	0.4723 ± 0.1297	0.2944 ± 0.1590	0.6790 ± 0.1745
MS-detected creatinine	**2-Hydroxyethanesulfonate**	0.6572 ± 0.0772	0.7753 ± 0.1332	0.4558 ± 0.1187	0.5686 ± 0.0821	0.8640 ± 0.1107	0.3358 ± 0.1088	0.3951 ± 0.1059	0.3756 ± 0.1654	0.4650 ± 0.1819	0.4161 ± 0.1132	1.0000 ± 0.0000	0.0000 ± 0.0000
MS-detected creatinine	**4-Acetamidobutanoic acid**	0.7446 ± 0.0724	0.8400 ± 0.1139	0.5784 ± 0.1161	0.6718 ± 0.0681	0.9667 ± 0.0529	0.3495 ± 0.0850	0.4068 ± 0.1030	0.4878 ± 0.1839	0.3810 ± 0.1547	0.4757 ± 0.1169	1.0000 ± 0.0000	0.0000 ± 0.0000
MS-detected creatinine	**L,L-TMAP**	0.6914 ± 0.0771	0.5687 ± 0.1696	0.7042 ± 0.1321	0.6444 ± 0.0726	0.8847 ± 0.0884	0.3811 ± 0.1079	0.3931 ± 0.1108	0.4511 ± 0.1720	0.4020 ± 0.1667	0.4078 ± 0.0926	1.0000 ± 0.0000	0.0000 ± 0.0000
MS-detected creatinine	**L,L-TMAP isomer**	0.6642 ± 0.0703	0.6293 ± 0.1544	0.6289 ± 0.1334	0.6126 ± 0.0700	0.8527 ± 0.1038	0.3684 ± 0.1182	0.4074 ± 0.1045	0.3689 ± 0.1609	0.5060 ± 0.1917	0.4068 ± 0.1022	1.0000 ± 0.0000	0.0000 ± 0.0000
MS-detected creatinine	**2-(α-D-Mannopyranosyl)-L-tryptophan**	0.7616 ± 0.0555	0.9073 ± 0.0997	0.5184 ± 0.1122	0.6869 ± 0.0725	0.8887 ± 0.0966	0.4584 ± 0.1043	0.4203 ± 0.0999	0.6544 ± 0.1859	0.2520 ± 0.1389	0.4748 ± 0.1101	0.6578 ± 0.2181	0.3100 ± 0.1921
MS-detected creatinine	**Pseudouridine**	0.7584 ± 0.0625	0.9173 ± 0.0865	0.5274 ± 0.1126	0.6873 ± 0.0752	0.9887 ± 0.0284	0.3284 ± 0.0993	0.4194 ± 0.1085	0.7889 ± 0.1856	0.1330 ± 0.1312	0.5347 ± 0.1083	0.4567 ± 0.1865	0.5660 ± 0.1893
MS-detected creatinine	**Succinyladenosine**	0.7256 ± 0.0727	0.7000 ± 0.1701	0.6368 ± 0.1353	0.6709 ± 0.0793	0.7967 ± 0.0954	0.4879 ± 0.1096	0.4530 ± 0.1129	0.4156 ± 0.1857	0.5360 ± 0.2003	0.4763 ± 0.1098	1.0000 ± 0.0000	0.0000 ± 0.0000

L,L-TMAP	**2-Hydroxyethanesulfonate**	0.6262 ± 0.0680	0.8153 ± 0.1302	0.3474 ± 0.1191	0.6349 ± 0.0718	0.8560 ± 0.1010	0.3426 ± 0.1110	0.4156 ± 0.1222	0.5133 ± 0.2047	0.3870 ± 0.1592	0.4207 ± 0.1311	0.8178 ± 0.1739	0.1580 ± 0.1662
L,L-TMAP	**4-Acetamidobutanoic acid**	0.6809 ± 0.0781	0.8913 ± 0.1094	0.3921 ± 0.1002	0.6939 ± 0.0628	0.9060 ± 0.0989	0.4416 ± 0.0964	0.4244 ± 0.1206	0.4433 ± 0.1928	0.4590 ± 0.1408	0.4748 ± 0.1205	0.3022 ± 0.1548	0.7090 ± 0.1855
L,L-TMAP	**2-(α-D-Mannopyranosyl)-L-tryptophan**	0.6888 ± 0.0624	0.8027 ± 0.1346	0.5042 ± 0.1156	0.7039 ± 0.0656	0.8780 ± 0.0905	0.4600 ± 0.1128	0.4336 ± 0.1140	0.8067 ± 0.1883	0.1400 ± 0.1407	0.4867 ± 0.1226	0.2822 ± 0.1486	0.6780 ± 0.2033
L,L-TMAP	**Succinyladenosine**	0.6879 ± 0.0681	0.6840 ± 0.1478	0.5932 ± 0.1220	0.7094 ± 0.0632	0.8007 ± 0.0996	0.4942 ± 0.1058	0.4435 ± 0.1109	0.5033 ± 0.1795	0.4140 ± 0.1643	0.4738 ± 0.1041	0.9578 ± 0.0979	0.0180 ± 0.0456
L,L-TMAP isomer	**2-Hydroxyethanesulfonate**	0.5938 ± 0.0736	0.8147 ± 0.1378	0.2979 ± 0.1264	0.6034 ± 0.0747	0.8207 ± 0.1098	0.3184 ± 0.1251	0.4373 ± 0.1186	0.6133 ± 0.1760	0.3050 ± 0.1819	0.4262 ± 0.1272	1.0000 ± 0.0000	0.0000 ± 0.0000
L,L-TMAP isomer	**4-Acetamidobutanoic acid**	0.6604 ± 0.0721	0.8827 ± 0.1189	0.3968 ± 0.1008	0.6862 ± 0.0663	0.8833 ± 0.1068	0.4405 ± 0.1028	0.4401 ± 0.1019	0.6433 ± 0.1814	0.2810 ± 0.1501	0.4774 ± 0.1131	0.3033 ± 0.1594	0.6790 ± 0.1996
L,L-TMAP isomer	**L,L-TMAP**	0.6080 ± 0.0687	0.7120 ± 0.1511	0.4742 ± 0.1235	0.6711 ± 0.0717	0.7733 ± 0.1353	0.5174 ± 0.1097	0.4349 ± 0.1088	0.4544 ± 0.1860	0.4510 ± 0.1769	0.4257 ± 0.1051	1.0000 ± 0.0000	0.0000 ± 0.0000
L,L-TMAP isomer	**2-(α-D-Mannopyranosyl)-L-tryptophan**	0.6757 ± 0.0669	0.8340 ± 0.1384	0.4595 ± 0.1143	0.6949 ± 0.0680	0.9633 ± 0.0592	0.3300 ± 0.0926	0.4397 ± 0.1048	0.6711 ± 0.1993	0.2560 ± 0.1675	0.4908 ± 0.1172	0.8244 ± 0.1983	0.1650 ± 0.1711
L,L-TMAP isomer	**Succinyladenosine**	0.6771 ± 0.0666	0.7213 ± 0.1446	0.5284 ± 0.1348	0.6967 ± 0.0612	0.7980 ± 0.0921	0.4816 ± 0.1067	0.4479 ± 0.1055	0.4467 ± 0.1721	0.5200 ± 0.1844	0.4846 ± 0.1131	1.0000 ± 0.0000	0.0000 ± 0.0000
2-(α-D-Mannopyranosyl)-L-tryptophan	**2-Hydroxyethanesulfonate**	0.6812 ± 0.0635	0.8733 ± 0.1168	0.4211 ± 0.0968	0.6859 ± 0.0729	0.9687 ± 0.0592	0.3274 ± 0.0981	0.4297 ± 0.1076	0.7978 ± 0.1759	0.1550 ± 0.1596	0.4680 ± 0.1223	0.2900 ± 0.1721	0.7100 ± 0.1847
2-(α-D-Mannopyranosyl)-L-tryptophan	**4-Acetamidobutanoic acid**	0.6991 ± 0.0688	0.8673 ± 0.1149	0.5026 ± 0.0989	0.7188 ± 0.0660	0.9633 ± 0.0606	0.4368 ± 0.0952	0.4313 ± 0.0948	0.8156 ± 0.1709	0.1390 ± 0.1413	0.5010 ± 0.1308	0.2933 ± 0.1659	0.6900 ± 0.1814
2-(α-D-Mannopyranosyl)-L-tryptophan	**Succinyladenosine**	0.7004 ± 0.0678	0.8387 ± 0.1324	0.4779 ± 0.0932	0.7150 ± 0.0673	0.8780 ± 0.0864	0.4679 ± 0.1104	0.4328 ± 0.1052	0.8156 ± 0.1694	0.1310 ± 0.1332	0.5070 ± 0.1187	0.4633 ± 0.1618	0.5740 ± 0.1942
Pseudouridine	**2-Hydroxyethanesulfonate**	0.6659 ± 0.0698	0.8900 ± 0.1021	0.4100 ± 0.1040	0.6837 ± 0.0756	0.9820 ± 0.0409	0.3258 ± 0.0908	0.4077 ± 0.1112	0.5633 ± 0.1881	0.3030 ± 0.1473	0.4989 ± 0.1215	0.3367 ± 0.1651	0.7030 ± 0.1857
Pseudouridine	**4-Acetamidobutanoic acid**	0.7070 ± 0.0714	0.9053 ± 0.0905	0.4842 ± 0.1039	0.7178 ± 0.0614	0.9687 ± 0.0503	0.4237 ± 0.0928	0.4273 ± 0.1145	0.8211 ± 0.1699	0.1040 ± 0.1318	0.5409 ± 0.1281	0.3111 ± 0.1474	0.7350 ± 0.1780
Pseudouridine	**L,L-TMAP**	0.6921 ± 0.0685	0.8640 ± 0.1233	0.4737 ± 0.1153	0.7072 ± 0.0672	0.8973 ± 0.0819	0.4458 ± 0.1105	0.4401 ± 0.1091	0.8344 ± 0.1774	0.0910 ± 0.1242	0.5068 ± 0.1254	1.0000 ± 0.0000	0
Pseudouridine	**L,L-TMAP isomer**	0.6875 ± 0.0709	0.8460 ± 0.1250	0.4589 ± 0.1119	0.6988 ± 0.0704	0.8940 ± 0.0948	0.4347 ± 0.1075	0.4311 ± 0.0964	0.7600 ± 0.1897	0.1510 ± 0.1526	0.5184 ± 0.1349	1.0000 ± 0.0000	0
Pseudouridine	**2-(α-D-Mannopyranosyl)-L-tryptophan**	0.6872 ± 0.0658	0.8993 ± 0.1125	0.4953 ± 0.1010	0.7189 ± 0.0678	0.9073 ± 0.0952	0.4947 ± 0.1045	0.3888 ± 0.0981	0.7167 ± 0.2002	0.1950 ± 0.1431	0.5438 ± 0.1426	1.0000 ± 0.0000	0
Pseudouridine	**Succinyladenosine**	0.6963 ± 0.0715	0.7580 ± 0.1637	0.5616 ± 0.1179	0.7163 ± 0.0630	0.8360 ± 0.0892	0.5453 ± 0.0922	0.4193 ± 0.1064	0.8478 ± 0.1830	0.0870 ± 0.1197	0.5301 ± 0.1263	0.4256 ± 0.1626	0.6050 ± 0.1878
Succinyladenosine	**2-Hydroxyethanesulfonate**	0.6361 ± 0.0647	0.7747 ± 0.1410	0.4579 ± 0.1003	0.6788 ± 0.0675	0.7773 ± 0.1008	0.4832 ± 0.1054	0.4287 ± 0.1086	0.4278 ± 0.1535	0.4790 ± 0.1633	0.4780 ± 0.1203	0.4422 ± 0.1610	0.5570 ± 0.2178

Succinyladenosine	**4-Acetamidobutanoic acid**	0.6843 ± 0.0693	0.8413 ± 0.1312	0.5000 ± 0.0989	0.7168 ± 0.0604	0.8893 ± 0.1076	0.4895 ± 0.0958	0.4385 ± 0.0944	0.4100 ± 0.1864	0.4930 ± 0.1444	0.5093 ± 0.1242	0.4222 ± 0.1507	0.6290 ± 0.1941

Table S5.20 AUC of significant metabolites for distinguishing UACR > 300 from UACR = 30-300 of diabetic patients using random forest classification and logistic regression (AUC > 0.70).

Top metabolite	UACR > 300 vs UACR = 30-300											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
4-Acetamidobutanoic acid	0.8763 ± 0.0541	0.7411 ± 0.1425	0.8389 ± 0.1088	0.9021 ± 0.0343	0.4363 ± 0.1279	0.9753 ± 0.0302	0.9146 ± 0.0586	0.8333 ± 0.1347	0.9011 ± 0.1206	0.9484 ± 0.0418	0.5511 ± 0.1523	1.0000 ± 0.0000
Pseudouridine	0.8548 ± 0.0518	0.7468 ± 0.1400	0.7774 ± 0.1280	0.8801 ± 0.0405	0.5516 ± 0.0937	0.9200 ± 0.0502	0.9261 ± 0.0562	0.8133 ± 0.1563	0.8533 ± 0.1378	0.9553 ± 0.0279	0.6667 ± 0.1414	0.9911 ± 0.0301
2-(α-D-Mannopyranosyl)-L-tryptophan	0.8483 ± 0.0564	0.8247 ± 0.1483	0.7000 ± 0.1176	0.8711 ± 0.0421	0.4911 ± 0.1063	0.9400 ± 0.0542	0.9485 ± 0.0410	0.8733 ± 0.1406	0.8444 ± 0.1122	0.9711 ± 0.0184	0.5711 ± 0.1449	0.9989 ± 0.0111
L,L-TMAP	0.8450 ± 0.0536	0.7021 ± 0.1177	0.8521 ± 0.0975	0.8900 ± 0.0414	0.5589 ± 0.1195	0.9458 ± 0.0360	0.9393 ± 0.0430	0.7689 ± 0.1394	0.9289 ± 0.1347	0.9590 ± 0.0296	0.5822 ± 0.1483	1.0000 ± 0.0000
L,L-TMAP isomer	0.8412 ± 0.0494	0.8284 ± 0.0975	0.8111 ± 0.0745	0.8880 ± 0.0415	0.5684 ± 0.0933	0.9437 ± 0.0409	0.8925 ± 0.0796	0.7611 ± 0.2015	0.8044 ± 0.1533	0.9199 ± 0.0454	0.5933 ± 0.1476	1.0000 ± 0.0000
Succinyladenosine	0.8397 ± 0.0549	0.7526 ± 0.1508	0.7395 ± 0.1182	0.8656 ± 0.0426	0.5126 ± 0.0844	0.9442 ± 0.0481	0.9127 ± 0.0566	0.8622 ± 0.1371	0.8500 ± 0.0973	0.9390 ± 0.0388	0.6100 ± 0.1544	0.9389 ± 0.0636
Serum creatinine (mg/dL)	0.8327 ± 0.0590	0.7037 ± 0.1075	0.8537 ± 0.1025	0.8779 ± 0.0421	0.5547 ± 0.1233	0.9426 ± 0.0415	0.8599 ± 0.0775	0.7011 ± 0.1505	0.8989 ± 0.1564	0.9166 ± 0.0481	0.6100 ± 0.1674	1.0000 ± 0.0000
2-Hydroxyethanesulfonate	0.8262 ± 0.0583	0.7647 ± 0.1403	0.6921 ± 0.1421	0.8588 ± 0.0450	0.4037 ± 0.1300	0.9468 ± 0.0390	0.9377 ± 0.0433	0.8044 ± 0.1441	0.8944 ± 0.1211	0.9683 ± 0.0217	0.5433 ± 0.1539	1.0000 ± 0.0000
MS-detected creatinine	0.8205 ± 0.0549	0.7526 ± 0.0969	0.8537 ± 0.0854	0.8782 ± 0.0455	0.5705 ± 0.1304	0.9432 ± 0.0399	0.8805 ± 0.0806	0.7078 ± 0.1817	0.8433 ± 0.1603	0.9210 ± 0.0448	0.6089 ± 0.1543	1.0000 ± 0.0000
Sulfotyrosine	0.8200 ± 0.0439	0.6116 ± 0.1015	0.8932 ± 0.0986	0.8515 ± 0.0463	0.5379 ± 0.1249	0.9595 ± 0.0439	0.9206 ± 0.0485	0.8044 ± 0.1198	0.8700 ± 0.1423	0.9307 ± 0.0543	0.6567 ± 0.1736	1.0000 ± 0.0000

Table S5.21 AUC of significant metabolites for distinguishing UACR > 300 from UACR ≤ 300 of diabetic patients using random forest classification and logistic regression (AUC > 0.70).

Top metabolite	UACR > 300 vs UACR ≤ 300											
	Discovery set						Validation set					
	Random Forest			Logistic Regression			Random Forest			Logistic Regression		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
4-Acetamidobutanoic acid	0.9217 ± 0.0408	0.9089 ± 0.0706	0.7921 ± 0.0897	0.9361 ± 0.0231	0.9200 ± 0.1190	0.5394 ± 0.3445	0.9305 ± 0.0513	0.8511 ± 0.1219	0.9632 ± 0.0404	0.9607 ± 0.0337	0.7978 ± 0.2295	0.7342 ± 0.3783
Pseudouridine	0.9160 ± 0.0385	0.9442 ± 0.0566	0.7697 ± 0.0617	0.9235 ± 0.0259	0.7363 ± 0.1320	0.8856 ± 0.0621	0.9459 ± 0.0476	0.8489 ± 0.1568	0.9084 ± 0.0811	0.9737 ± 0.0189	0.7889 ± 0.1629	0.9468 ± 0.0616
2-(*α*-D-Mannopyranosyl)-L-tryptophan	0.9151 ± 0.0381	0.9447 ± 0.0638	0.8085 ± 0.0542	0.9259 ± 0.0233	0.9600 ± 0.0582	0.6450 ± 0.1948	0.9573 ± 0.0448	0.9222 ± 0.1060	0.8895 ± 0.0625	0.9811 ± 0.0143	0.8278 ± 0.2151	0.8305 ± 0.2424
L,L-TMAP	0.9076 ± 0.0339	0.8053 ± 0.1173	0.8618 ± 0.0682	0.9183 ± 0.0307	0.9153 ± 0.0902	0.6276 ± 0.2500	0.9415 ± 0.0461	0.8211 ± 0.1547	0.8511 ± 0.1156	0.9513 ± 0.0275	0.7778 ± 0.2250	0.8053 ± 0.2155
L,L-TMAP isomer	0.8949 ± 0.0453	0.8247 ± 0.1053	0.8576 ± 0.0718	0.9141 ± 0.0296	0.7663 ± 0.1487	0.8688 ± 0.0942	0.9085 ± 0.0519	0.8322 ± 0.1696	0.8121 ± 0.0936	0.9218 ± 0.0375	0.5711 ± 0.1499	0.9968 ± 0.0125
Succinyladenosine	0.8931 ± 0.0355	0.8795 ± 0.0796	0.7782 ± 0.0785	0.9094 ± 0.0282	0.9042 ± 0.1180	0.6012 ± 0.2688	0.9392 ± 0.0407	0.8711 ± 0.1214	0.9468 ± 0.0411	0.9607 ± 0.0270	0.8000 ± 0.2031	0.9105 ± 0.0879
Serum creatinine (mg/dL)	0.8869 ± 0.0440	0.7732 ± 0.0972	0.8785 ± 0.0576	0.9054 ± 0.0326	0.8700 ± 0.1049	0.6935 ± 0.1820	0.8840 ± 0.0681	0.7244 ± 0.1356	0.9258 ± 0.1348	0.9202 ± 0.0412	0.6856 ± 0.1572	0.9705 ± 0.0757
2-Hydroxyethanesulfonate	0.8817 ± 0.0381	0.8079 ± 0.1144	0.7674 ± 0.0839	0.8913 ± 0.0296	0.8068 ± 0.1718	0.6738 ± 0.2742	0.9318 ± 0.0414	0.8489 ± 0.1436	0.8589 ± 0.0838	0.9647 ± 0.0209	0.6667 ± 0.2172	0.9479 ± 0.0739
Sulfotyrosine	0.8735 ± 0.0415	0.8326 ± 0.1120	0.7503 ± 0.0884	0.8867 ± 0.0338	0.8611 ± 0.1199	0.6091 ± 0.2372	0.9404 ± 0.0434	0.8378 ± 0.1211	0.9321 ± 0.0715	0.9450 ± 0.0450	0.8522 ± 0.1441	0.9289 ± 0.0715
MS-detected creatinine	0.8569 ± 0.0594	0.7895 ± 0.0906	0.8674 ± 0.0816	0.8953 ± 0.0377	0.7989 ± 0.1056	0.8500 ± 0.0809	0.8895 ± 0.0602	0.6889 ± 0.1370	0.9205 ± 0.0978	0.9320 ± 0.0363	0.6600 ± 0.1593	0.9742 ± 0.0638
L-*β*-aspartyl-L-leucine	0.8513 ± 0.0436	0.7838 ± 0.0999	0.7379 ± 0.1131	0.8579 ± 0.0329	0.7626 ± 0.1145	0.7426 ± 0.1372	0.9200 ± 0.0441	0.8184 ± 0.0749	0.9044 ± 0.1187	0.9492 ± 0.0231	0.8900 ± 0.1319	0.6878 ± 0.1653

Reference

Aboud, Omran Abu, and Robert H Weiss. 2013. 'New Opportunities from the Cancer Metabolome', *Clinical Chemistry*, 59: 138-46.

Administration, U.S. Food and Drug. 2018. "Bioanalytical Method Validation. Guidance for Industry." In.

Afkarian, Maryam, Michael C. Sachs, Bryan Kestenbaum, Irl B. Hirsch, Katherine R. Tuttle, Jonathan Himmelfarb, and Ian H. de Boer. 2013. 'Kidney Disease and Increased Mortality Risk in Type 2 Diabetes', *Journal of the American Society of Nephrology*, 24: 302-08.

Afsar, Baris, and Rengin Elsurer. 2017. 'Increased renal resistive index in type 2 diabetes: Clinical relevance, mechanisms and future directions', *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 11: 291-96.

Ahmadi, Shiva, and Dominic Winter. 2018. 'Identification of Poly(ethylene glycol) and Poly(ethylene glycol)-Based Detergents Using Peptide Search Engines', *Analytical Chemistry*, 90: 6594-600.

Akbartabar Toori, Mehdi PhD, Faezeh MSc Kiani, Fatemeh PhD Sayehmiri, Kourosh PhD Sayehmiri, Yousof M. D. Mohsenzadeh, Rahim PhD Ostovar, Parvin MSc Angha, and Yazdan MSc Mohsenzadeh. 2018. 'Prevalence of Hypercholesterolemia, High LDL, and Low HDL in Iran: A Systematic Review and Meta-Analysis', *Iranian Journal of Medical Sciences*, 43: 449-65.

Alicic, Radica Z., Michele T. Rooney, and Katherine R. Tuttle. 2017. 'Diabetic Kidney Disease', *Challenges, Progress, and Possibilities*, 12: 2032-45.

Alonso, Arnald, Sara Marsal, and Antonio Julià. 2015. 'Analytical Methods in Untargeted Metabolomics: State of the Art in 2015', *Frontiers in Bioengineering and Biotechnology*, 3: 23.

American Diabetes Association 2004. 'Nephropathy in Diabetes', *Diabetes Care*, 27: s79-s83.

American Diabetes Association. 2018. '10. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes-2018', *Diabetes Care*, 41: S105-S18.

Anders, Hans-Joachim, Tobias B. Huber, Berend Isermann, and Mario Schiffer. 2018. 'CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease', *Nature Reviews Nephrology*, 14: 361-77.

Andreev, E., M. Koopman, and L. Arisz. 1999. 'A rise in plasma creatinine that is not a sign of renal failure: which drugs can be responsible?', *Journal of Internal Medicine*, 246: 247-52.

Andres-Lacueva, Cristina, M. Teresa Macarulla, Maria Rotches-Ribalta, María Boto-Ordóñez, Mireia Urpi-Sarda, Víctor M. Rodríguez, and María P. Portillo. 2012. 'Distribution of Resveratrol Metabolites in Liver, Adipose Tissue, and Skeletal Muscle in Rats Fed Different Doses of This Polyphenol', *Journal of Agricultural and Food Chemistry*, 60: 4833-40.

Anstee, Quentin M, Stuart McPherson, and Christopher P Day. 2011. 'How big a problem is non-alcoholic fatty liver disease?', *BMJ*, 343: d3897.

Antignac, Jean-Philippe, Katia de Wasch, Fabrice Monteau, Hubert De Brabander, François Andre, and Bruno Le Bizec. 2005. 'The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis', *Analytica Chimica Acta*, 529: 129-36.

Arichi, Hideko, Yoshiyuki Kimura, Hiromichi Okuda, Kimiye Baba, Mitsugi Kozawa, and Shigeru Arichi. 1982. 'Effects of Stilbene Components of the Roots of Polygonum

cuspidatum SIEB. et ZUCC. on Lipid Metabolism', *Chemical and Pharmaceutical Bulletin*, 30: 1766-70.

Armitage, Emily Grace, Joanna Godzien, Vanesa Alonso-Herranz, Ángeles López-Gonzálvez, and Coral Barbas. 2015. 'Missing value imputation strategies for metabolomics data', *Electrophoresis*, 36: 3050-60.

Atanacković, Milica, Mihalj Posa, Helmut Heinle, Ljiljana Gojković-Bukarica, and Jelena Cvejić. 2009. 'Solubilization of resveratrol in micellar solutions of different bile acids', *Colloids and Surfaces B: Biointerfaces*, 72: 148-54.

Athyros, V. G., K. Tziomalos, A. Karagiannis, and D. P. Mikhailidis. 2011. 'Dyslipidaemia of obesity, metabolic syndrome and type 2 diabetes mellitus: the case for residual risk reduction after statin treatment', *The Open Cardiovascular Medicine Journal*, 5: 24-34.

Baba, Takashi, J. Larry Campbell, J. C. Yves Le Blanc, and Paul R. S. Baker. 2016. 'In-depth sphingomyelin characterization using electron impact excitation of ions from organics and mass spectrometry', *Journal of Lipid Research*, 57: 858-67.

Bachmann, Lorin M., Goran Nilsson, David E. Bruns, Matthew J. McQueen, John C. Lieske, Jack J. Zakowski, and W. Greg Miller. 2014. 'State of the Art for Measurement of Urine Albumin: Comparison of Routine Measurement Procedures to Isotope Dilution Tandem Mass Spectrometry', *Clinical Chemistry*, 60: 471.

Barbier Saint Hilaire, Pierre, Ulli M. Hohenester, Benoit Colsch, Jean-Claude Tabet, Christophe Junot, and François Fenaille. 2018. 'Evaluation of the High-Field Orbitrap Fusion for Compound Annotation in Metabolomics', *Analytical Chemistry*, 90: 3030-35.

Barker, Matthew, and William Rayens. 2003. 'Partial least squares for discrimination', *Journal of Chemometrics*, 17: 166-73.

Barupal, D. K., P. K. Haldiya, G. Wohlgemuth, T. Kind, S. L. Kothari, K. E. Pinkerton, and O. Fiehn. 2012. 'MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity', *BMC Bioinformatics*, 13: 99.

Baxmann, A. C., M. S. Ahmed, N. C. Marques, V. B. Menon, A. B. Pereira, G. M. Kirsztajn, and I. P. Heilberg. 2008. 'Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C', *Clinical Journal of the American Society of Nephrology*, 3: 348-54.

Behrends, Volker, Gregory D. Tredwell, and Jacob G. Bundy. 2011. 'A software complement to AMDIS for processing GC-MS metabolomic data', *Analytical Biochemistry*, 415: 206-08.

Bilous, R. 2008. 'Microvascular disease: what does the UKPDS tell us about diabetic nephropathy?', *Diabetic Medicine*, 25: 25-29.

Bird, S. S., V. R. Marur, I. G. Stavrovskaya, and B. S. Kristal. 2013. 'Qualitative Characterization of the Rat Liver Mitochondrial Lipidome using LC-MS Profiling and High Energy Collisional Dissociation (HCD) All Ion Fragmentation', *Metabolomics*, 9: 67-83.

Birner, Armin, Martina Platzer, Susanne Astrid Bengesser, Nina Dalkner, Frederike T. Fellendorf, Robert Queissner, Rene Pilz, Philipp Rauch, Alexander Maget, Carlo Hamm, Simone Herzog-Eberhard, Harald Mangge, Dietmar Fuchs, Natalie Moll, Sieglinde Zelzer, Gregor Schütze, Markus Schwarz, Bernd Reininghaus, Hans-Peter Kapfhammer, and Eva Z. Reininghaus. 2017. 'Increased breakdown of kynurenine towards its neurotoxic branch in bipolar disorder', *PLOS ONE*, 12: e0172699.

Blaženović, Ivana, Tobias Kind, Jian Ji, and Oliver Fiehn. 2018. 'Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics', *Metabolites*, 8: 31.

Bode, Lisa M, Diana Bunzel, Melanie Huch, Gyu-Sung Cho, Denise Ruhland, Mirko Bunzel, Achim Bub, Charles MAP Franz, and Sabine E Kulling. 2013. 'In vivo and in vitro metabolism of trans-resveratrol by human gut microbiota', *The American Journal of Clinical Nutrition*, 97: 295-309.

Boulesteix, Anne-Laure, Silke Janitza, Jochen Kruppa, and Inke R. König. 2012. 'Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2: 493-507.

Breiman, Leo. 2001. 'Random Forests', *Machine Learning*, 45: 5-32.

Bro, Rasmus, and Age K. Smilde. 2014. 'Principal component analysis', *Analytical Methods*, 6: 2812-31.

Buechler, Christa. 2014. 'Regard Lipidomics in human nonalcoholic fatty liver disease', *Clinical Lipidology*, 9: 279-82.

Bundgaard, Jens R., and Jens F. Rehfeld. 2013. 'Chapter 250 - Tyrosylprotein sulfotransferases.' in Abba J. Kastin (ed.), *Handbook of Biologically Active Peptides (Second Edition)* (Academic Press: Boston).

Bylesjo, Max, Mattias Rantalainen, Olivier Cloarec, Jeremy Nicholson, Elaine Holmes, and Johan Trygg. 2006. 'OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification', *Journal of Chemometrics*, 20: 341-51.

Calder, Philip C. 2006. 'Long-chain polyunsaturated fatty acids and inflammation', *Food & Nutrition Research*, 50: 54-61.

Calle, M. L., V. Urrea, A. L. Boulesteix, and N. Malats. 2011. 'AUC-RF: A New Strategy for Genomic Profiling with Random Forest', *Human Heredity*, 72: 121-32.

Canadas-Garre, M., K. Anderson, J. McGoldrick, A. P. Maxwell, and A. J. McKnight. 2019. 'Proteomic and metabolomic approaches in the search for biomarkers in chronic kidney disease', *Journal of Proteomics*, 193: 93-122.

Catalgol, B., S. Batirel, Y. Taga, and N. K. Ozer. 2012. 'Resveratrol: French paradox revisited', *Frontiers in Pharmacology*, 3: 1-18.

Cave, Matthew, Ion Deaciuc, Christian Mendez, Zhenyuan Song, Swati Joshi-Barve, Shirish Barve, and Craig McClain. 2007. 'Nonalcoholic fatty liver disease: predisposing factors and the role of nutrition', *The Journal of Nutritional Biochemistry*, 18: 184-95.

Chai, Yuan Yuan, Fang Wang, Yan Li Li, Ke Liu, and Hui Xu. 2012. 'Antioxidant activities of stilbenoids from Rheum emodi Wall', *Evidence-Based Complementary and Alternative Medicine*, 2012: 1-7.

Chalasani, N., Z. Younossi, J. E. Lavine, A. M. Diehl, E. M. Brunt, K. Cusi, M. Charlton, and A. J. Sanyal. 2012. 'The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association', *Hepatology*, 55: 2005-23.

Chen, Dan-Qian, Gang Cao, Hua Chen, Christos P. Argyopoulos, Hui Yu, Wei Su, Lin Chen, David C. Samuels, Shougang Zhuang, George P. Bayliss, Shilin Zhao, Xiao-Yong Yu, Nosratola D. Vaziri, Ming Wang, Dan Liu, Jia-Rong Mao, Shi-Xing Ma, Jin Zhao, Yuan Zhang, You-Quan Shang, Huining Kang, Fei Ye, Xiao-Hong Cheng, Xiang-Ri Li, Li Zhang, Mei-Xia Meng, Yan Guo, and Ying-Yong Zhao. 2019. 'Identification of serum metabolites associating with chronic kidney disease progression and anti-fibrotic effect of 5-methoxytryptophan', *Nature Communications*, 10: 1476.

Chen, Jing, Wenzhao Wang, Shen Lv, Peiyuan Yin, Xinjie Zhao, Xin Lu, Fengxia Zhang, and Guowang Xu. 2009. 'Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations', *Analytica Chimica Acta*, 650: 3-9.

Chen, Qiong, Er Mao Wang, Li Ping Ma, and Pei Zhai. 2012. 'Dietary resveratrol increases the expression of hepatic 7α-hydroxylase and ameliorates hypercholesterolemia in high-fat fed C57BL/6J mice', *Lipids in Health and Disease*, 11: 1-8.

Chen, Tianlu, Yu Cao, Yinan Zhang, Jiajian Liu, Yuqian Bao, Congrong Wang, Weiping Jia, and Aihua Zhao. 2013. 'Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection', *Evidence-Based Complementary and Alternative Medicine*, 2013: 11.

Chen, Wan, Samuel Chao Ming Yeo, Mai Gamal Ahmed Ahmed Elhennawy, and Hai-Shu Lin. 2016. 'Oxyresveratrol: A bioavailable dietary polyphenol', *Journal of Functional Foods*, 22: 122-31.

Chiang, John Y. L. 2009. 'Bile acids: regulation of synthesis', *Journal of Lipid Research*, 50: 1955-66.

Chinese Pharmacopoeia Commission. 2015. *Pharmacopoeia of the People's Republic of China, 2015ed (Chinese version)* (Chinese Medical Science Press: Beijing).

Chong, Jasmine, Othman Soufan, Carin Li, Iurie Caraus, Shuzhao Li, Guillaume Bourque, David S. Wishart, and Jianguo Xia. 2018. 'MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis', *Nucleic Acids Research*, 46: W486-W94.

Choo, Qiu Yi, Samuel Chao Ming Yeo, Paul C. Ho, Yoshiya Tanaka, and Hai Shu Lin. 2014. 'Pterostilbene surpassed resveratrol for anti-inflammatory application: Potency consideration and pharmacokinetics perspective', *Journal of Functional Foods*, 11: 352-62.

Christie, William W., and Xianlin Han. 2012. 'Chapter 13 - Introduction to mass spectrometric analysis of lipids in lipidomics.' in William W. Christie and Xianlin Han (eds.), *Lipid Analysis (Fourth Edition)* (Woodhead Publishing).

Colhoun, Helen M., and M. Loredana Marcovecchio. 2018. 'Biomarkers of diabetic kidney disease', *Diabetologia*, 61: 996-1011.

Connor, S. C., W. Wu, B. C. Sweatman, J. Manini, J. N. Haselden, D. J. Crowther, and C. J. Waterfield. 2004. 'Effects of feeding and body weight loss on the 1H-NMR-based urine metabolic profiles of male Wistar Han rats: implications for biomarker discovery', *Biomarkers*, 9: 156-79.

Corporation, Waters. 2015. 'Progenesis QI v2.4 product solution '.

Courant, Frédérique, Jean Philippe Antignac, Gaud Dervilly-Pinel, and Bruno Le Bizec. 2014. 'Basics of mass spectrometry based metabolomics', *Proteomics*, 14: 2369-88.

Cui, Liang, Yie Hou Lee, Yadunanda Kumar, Fengguo Xu, Kun Lu, Eng Eong Ooi, Steven R. Tannenbaum, and Choon Nam Ong. 2013. 'Serum metabolome and lipidome changes in adult patients with primary dengue infection', *PLOS Neglected Tropical Diseases*, 7: e2373-e73.

Dai, Wei Dong, Cong Wei, Hong Wei Kong, Zhen Hua Jia, Jian Ke Han, Feng Xia Zhang, Ze Ming Wu, Yan Gu, Shi Li Chen, Qun Gu, Xin Lu, Yi Ling Wu, and Guo Wang Xu. 2011. 'Effect of the traditional Chinese medicine tongxinluo on endothelial dysfunction rats studied by using urinary metabonomics based on liquid chromatography-mass spectrometry', *Journal of Pharmaceutical and Biomedical Analysis*, 56: 86-92.

Dajani, Asad, and Adnan AbuHammour. 2016. 'Treatment of nonalcoholic fatty liver disease: Where do we stand? an overview', *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association*, 22: 91-105.

Davies, Robert. 2018. 'The metabolomic quest for a biomarker in chronic kidney disease', *Clinical Kidney Journal*, 11: 694-703.

Dawson, Paul A. 2016. 'Chapter 12 - Bile Acid Metabolism.' in Roger S. McLeod (ed.), *Biochemistry of Lipids, Lipoproteins and Membranes (Sixth Edition)* (Elsevier: Boston).

Day, C. P., and O. F. W. James. 1998. 'Steatohepatitis: A tale of two 'Hits'?', *Gastroenterology*, 114: 842-45.

Day, Christopher P. 2011. 'Non-alcoholic fatty liver disease: a massive problem', *Clinical Medicine Insights. Pathology*, 11: 176-78.

de Castro, Gabriela S., and Philip C. Calder. 2018. 'Non-alcoholic fatty liver disease and its treatment with n-3 polyunsaturated fatty acids', *Clinical Nutrition*, 37: 37-55.

de la Sierra, Alex, Xavier Pintó, Carlos Guijarro, José López Miranda, Daniel Callejo, Jesús Cuervo, Rudi Subirà, and Marta Rubio. 2015. 'Prevalence, Treatment, and Control of Hypercholesterolemia in High Cardiovascular Risk Patients: Evidences from a Systematic Literature Review in Spain', *Advances in Therapy*, 32: 944-61.

Debnath, Subrata, Chakradhar Velagapudi, Laney Redus, Farook Thameem, Balakuntalam Kasinath, Claudia E Hura, Carlos Lorenzo, Hanna E Abboud, and Jason C O'Connor. 2017. 'Tryptophan Metabolism in Patients With Chronic Kidney Disease Secondary to Type 2 Diabetes: Relationship to Inflammatory Markers', *International Journal of Tryptophan Research*, 10: 1178646917694600.

Dettmer, Katja, Pavel A. Aronov, and Bruce D. Hammock. 2007. 'Mass spectrometry-based metabolomics', *Mass Spectrometry Reviews*, 26: 51-78.

Du Bois, D., and E. F. Du Bois. 1989. 'A formula to estimate the approximate surface area if height and weight be known. 1916', *Nutrition*, 5: 303-11; discussion 12-3.

Dührkop, Kai, Huibin Shen, Marvin Meusel, Juho Rousu, and Sebastian Böcker. 2015. 'Searching molecular structure databases with tandem mass spectra using CSI:FingerID', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 12580-85.

Dunn, Warwick B., David Broadhurst, Paul Begley, Eva Zelena, Sue Francis-McIntyre, Nadine Anderson, Marie Brown, Joshau D. Knowles, Antony Halsall, John N. Haselden, Andrew W. Nicholls, Ian D. Wilson, Douglas B. Kell, and Royston Goodacre. 2011. 'Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry', *Nature Protocols*, 6: 1060-83.

Dunn, Warwick, Ian Wilson, Andrew Nicholls, and David Broadhurst. 2012. 'The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans', *Bioanalysis*, 4: 2249-64.

Durack, Juliana, and Susan V. Lynch. 2019. 'The gut microbiome: Relationships with disease and opportunities for therapy', *The Journal of experimental medicine*, 216: 20-40.

Durrington, Paul. 2003. 'Dyslipidaemia', *Lancet*, 362: 717-31.

Dworacka, M., H. Winiarska, M. Szymanska, S. Kuczynski, K. Szczawinska, and B. Wierusz-Wysocka. 2002. '1,5-anhydro-D-glucitol: a novel marker of glucose excursions', *International Journal of Clinical Practice. Supplement*: 40-4.

Dzúrik, R., I. Lajdová, V. Spustová, and K. Opatrný Jr. 1992. 'Pseudouridine Excretion in Healthy Subjects and Its Accumulation in Renal Failure', *Nephron*, 61: 64-67.

Eliuk, Shannon, and Alexander Makarov. 2015. 'Evolution of Orbitrap Mass Spectrometry Instrumentation', *Annual Review of Analytical Chemistry*, 8: 61-80.

Emwas, Abdul-Hamid M. 2015. 'The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research.' in Jacob T. Bjerrum (ed.), *Metabonomics: Methods and Protocols* (Springer New York: New York, NY).

Eriksen, Bjørn Odvar, Ulla Dorte Mathisen, Toralf Melsom, Ole Christian Ingebretsen, Trond Geir Jenssen, Inger Njølstad, Marit Dahl Solbu, and Ingrid Toft. 2012. 'The Role of Cystatin C in Improving GFR Estimation in the General Population', *American Journal of Kidney Diseases*, 59: 32-40.

Eriksson, L., T. Byrne, E. Johansson, J. Trygg, and C. Vikström. 2013. *Multi- and megavariate data analysis : basic principles and applications* (MKS Umetrics AB: Malmö, Sweden).

Eriksson, L., E. Johansson, N. Kettaneh-Wold, Johan Trygg, C. Wikstr, and Svante Wold. 2006. 'Multi- and Megavariate Data Analysis. Part I Basic Principles and Applications. Second revised and enlarged edition', *Ume Sweden: MKS Umetrics AB*: 1-103.

Etxeberria, Usune, Noemi Arias, Noemí Boqué, Ana Romo-Hualde, M. Teresa Macarulla, María P. Portillo, Fermín I. Milagro, and J. Alfredo Martínez. 2015. 'Metabolic faecal fingerprinting of trans-resveratrol and quercetin following a high-fat sucrose dietary model using liquid chromatography coupled to high-resolution mass spectrometry', *Food & Function*, 6: 2758-67.

Fan, J. G. 2004. 'Evaluating the efficacy and safety of Danning Pian in the short-term treatment of patients with non-alcoholic fatty liver disease: a multicenter clinical trial', *Hepatobiliary & Pancreatic Diseases International* 3: 375-80.

Felmlee, Melanie A., Rutwij A. Dave, and Marilyn E. Morris. 2013. 'Mechanistic models describing active renal reabsorption and secretion: a simulation-based study', *The AAPS journal*, 15: 278-87.

Feng, Qin, Xiao-Jun Gou, Sheng-Xi Meng, Cheng Huang, Yu-Quan Zhang, Ya-Jun Tang, Wen-Jing Wang, Lin Xu, Jinghua Peng, and Yi-Yang Hu. 2013. 'Qushi Huayu Decoction Inhibits Hepatic Lipid Accumulation by Activating AMP-Activated Protein Kinase In Vivo and In Vitro', *Evidence-Based Complementary and Alternative Medicine*, 2013: 184358.

Fernández-Peralbo, M. A., C. Ferreiro Vera, F. Priego-Capote, and M. D. Luque de Castro. 2014. 'Stable isotopic internal standard correction for quantitative analysis of hydroxyeicosatetraenoic acids (HETEs) in serum by on-line SPE–LC–MS/MS in selected reaction monitoring mode', *Talanta*, 126: 170-76.

Fiehn, Oliver, W. Timothy Garvey, John W. Newman, Kerry H. Lok, Charles L. Hoppel, and Sean H. Adams. 2010. 'Plasma metabolomic profiles reflective of glucose homeostasis in non-diabetic and type 2 diabetic obese African-American women', *PLOS ONE*, 5: e15234-e34.

Flamini, Riccardo, Mirko De Rosso, and Luigi Bavaresco. 2015. 'Study of grape polyphenols by liquid chromatography-high-resolution mass spectrometry (UHPLC/QTOF) and suspect screening analysis', *Journal of Analytical Methods in Chemistry*, 2015: 350259.

Fox, K. M., L. Wang, S. R. Gandra, R. G. W. Quek, L. Li, and O. Baser. 2015. 'Long-term economic burden associated with cardiovascular events among high-risk patients with hyperlipidemia', *Value in Health*, 18: A139-A40.

Fredrickson, Donald S., Robert I. Levy, and Robert S. Lees. 1967. 'Fat Transport in Lipoproteins — An Integrated Approach to Mechanisms and Disorders', *New England Journal of Medicine*, 276: 34-44.

Frega, Natale G., Deborah Pacetti, and Emanuele Boselli. 2012. *Characterization of Phospholipid Molecular Species by Means of HPLC-Tandem Mass Spectrometry, Tandem Mass Spectrometry - Applications and Principles* (IntechOpen).

Frombaum, Matthieu, Patrice Therond, Raja Djelidi, Jean Louis Beaudeux, Dominique Bonnefont-Rousselot, and Didier Borderie. 2011. 'Piceatannol is more effective than resveratrol in restoring endothelial cell dimethylarginine dimethylaminohydrolase expression and activity after high-glucose oxidative stress', *Free Radical Research*, 45: 293-302.

Frostegård, Johan. 2010. 'Low level natural antibodies against phosphorylcholine: A novel risk marker and potential mechanism in atherosclerosis and cardiovascular disease', *Clinical Immunology*, 134: 47-54.

Gan, Wei, Ying Liu, Kai-Hong Luo, Shan-Shan Liang, Hui Wang, Meng Li, Yan-Xing Zhang, and Heng-Jian Huang. 2018. 'The prevalence change of hyperlipidemia and hyperglycemia and the effectiveness of yearly physical examinations: an eight-year study in Southwest China', *Lipids in Health and Disease*, 17: 70.

García-Cañaveras, Juan C., M. Teresa Donato, José V. Castell, and Agustín Lahoz. 2012. 'Targeted profiling of circulating and hepatic bile acids in human, mouse, and rat using a UPLC-MRM-MS-validated method', *Journal of lipid research*, 53: 2231-41.

Garcia, A., and C. Barbas. 2011. 'Gas chromatography-mass spectrometry (GC-MS)-based metabolomics', *Methods in Molecular Biology* 708: 191-204.

Glauner, Thomas, and A. Paul Zavitsanos. 2012. 'Chapter 6 - Electrospray Operational Parameters in TOF-MS.' in Amadeo R. Fernandez-Alba (ed.), *Comprehensive Analytical Chemistry* (Elsevier).

Goldstein, J L, H G Schrott, W R Hazzard, E L Bierman, and A G Motulsky. 1973. 'Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia', *The Journal of Clinical Investigation*, 52: 1544-68.

González-Domínguez, Raúl, Rocío Castilla-Quintero, Tamara García-Barrera, and José Luis Gómez-Ariza. 2014. 'Development of a metabolomic approach based on urine samples and direct infusion mass spectrometry', *Analytical Biochemistry*, 465: 20-27.

Gowda, G. A. Nagana, Shucha Zhang, Haiwei Gu, Vincent Asiago, Narasimhamurthy Shanaiah, and Daniel Raftery. 2008. 'Metabolomics-based methods for early disease diagnostics', *Expert Review of Molecular Diagnostics*, 8: 617-33.

Gross, Jorge L., Mirela J. de Azevedo, Sandra P. Silveiro, Luís Henrique Canani, Maria Luiza Caramori, and Themis Zelmanovitz. 2005. 'Diabetic Nephropathy: Diagnosis, Prevention, and Treatment', *Diabetes Care*, 28: 164-76.

Gu, Y., Y. Zhang, X. Shi, X. Li, J. Hong, J. Chen, W. Gu, X. Lu, G. Xu, and G. Ning. 2010. 'Effect of traditional Chinese medicine berberine on type 2 diabetes based on comprehensive metabonomics', *Talanta*, 81: 766-72.

Guo, Na, Da Wei Yang, Xiao Xing Wang, Jin Gang Dai, Mei Wang, and Yan Lei. 2014. 'Metabonomic study of chronic heart failure and effects of Chinese herbal decoction in rats', *Journal of Chromatography A*, 1362: 89-101.

Gutsche, Birgit, Christoph Grun, Dieter Scheutzow, and Markus Herderich. 1999. 'Tryptophan glycoconjugates in food and human urine', *Biochemical Journal*, 343: 11.

Han, Jun, Wen-xing Qin, Zhen-li Li, Ai-jing Xu, Hao Xing, Han Wu, Han Zhang, Ming-da Wang, Chao Li, Lei Liang, Bing Quan, Wen-tao Yan, Feng Shen, Meng-chao Wu, and Tian Yang. 2019. 'Tissue and serum metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma', *Clinica Chimica Acta*, 488: 68-75.

Han, Yunkyung, Tomoaki Haraguchi, Sumie Iwanaga, Hiroyuki Tomotake, Yukako Okazaki, Shigeru Mineo, Akiho Moriyama, Junji Inoue, and Norihisa Kato. 2009. 'Consumption of some polyphenols reduces fecal deoxycholic acid and lithocholic acid, the secondary bile acids of risk factors of colon cancer', *Journal of Agricultural and Food Chemistry*, 57: 8587-90.

Hanhineva, Kati, Thaer Barri, Marjukka Kolehmainen, Jenna Pekkinen, Jussi Pihlajamäki, Arto Vesterbacka, Gloria Solano-Aguilar, Hannu Mykkänen, Lars Ove Dragsted, Joseph F. Urban, and Kaisa Poutanen. 2013. 'Comparative nontargeted profiling of metabolic changes in tissues and biofluids in high-fat diet-fed Ossabaw pig', *Journal of Proteome Research*, 12: 3980-92.

Hao, Jie, Cheng Chen, Kaipeng Huang, Junying Huang, Jie Li, Peiqing Liu, and Heqing Huang. 2014. 'Polydatin improves glucose and lipid metabolism in experimental diabetes

through activating the Akt signaling pathway', *European Journal of Pharmacology*, 745: 152-65.

Harris, J. R., and E. C. Grunsky. 2015. 'Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data', *Computers & Geosciences*, 80: 9-25.

Hartmann, Susen, Jürgen G. Okun, Christiane Schmidt, Claus-Dieter Langhans, Sven F. Garbade, Peter Burgard, Dorothea Haas, Jörn Oliver Sass, William L. Nyhan, and Georg F. Hoffmann. 2006. 'Comprehensive Detection of Disorders of Purine and Pyrimidine Metabolism by HPLC with Electrospray Ionization Tandem Mass Spectrometry', *Clinical Chemistry*, 52: 1127-37.

Haymann, J. P., K. Stankovic, P. Levy, V. Avellino, P. L. Tharaux, E. Letavernier, G. Grateau, L. Baud, R. Girot, and F. Lionnet. 2010. 'Glomerular hyperfiltration in adult sickle cell anemia: a frequent hemolysis associated feature', *Clinical journal of the American Society of Nephrology*, 5: 756-61.

Henry, Gillian D., Ian P. Trayer, Susan Brewer, and Barry A. Levine. 1985. 'The widespread distribution of α-N-trimethylalanine as the N-terminal amino acid of light chains from vertebrate striated muscle myosins', *European Journal of Biochemistry*, 148: 75-82.

Hijona, E., L. Aguirre, P. Pérez-Matute, M. J. Villanueva-Millán, A. Mosqueda-Solis, M. Hasnaoui, F. Nepveu, J. M. Senard, L. Bujanda, L. Aldámiz-Echevarría, M. Llarena, F. Andrade, P. Perio, F. Leboulanger, L. Hijona, J. M. Arbones-Mainar, M. P. Portillo, and C. Carpéné. 2016. 'Limited beneficial effects of piceatannol supplementation on obesity complications in the obese Zucker rat: gut microbiota, metabolic, endocrine, and cardiac aspects', *Journal of Physiology and Biochemistry*, 72: 567-82.

Hill, A. M., J. A. Fleming, and P. M. Kris-Etherton. 2009. 'The role of diet and nutritional supplements in preventing and treating cardiovascular disease', *Current opinion in cardiology*, 24: 433-41.

Hirayama, Akiyoshi, Masahiro Sugimoto, Asako Suzuki, Yoko Hatakeyama, Ayame Enomoto, Sei Harada, Tomoyoshi Soga, Masaru Tomita, and Toru Takebayashi. 2015. 'Effects of processing and storage conditions on charged metabolomic profiles in blood', *ELECTROPHORESIS*, 36: 2148-55.

Hlaing, Thinn Thinn, and Adrian Park. 2013. 'Hyperlipidaemia', *Medicine*, 41: 607-09.

Hocher, Berthold, and Jerzy Adamski. 2017. 'Metabolomics for clinical use and research in chronic kidney disease', *Nature Reviews Nephrology*, 13: 269-84.

Hofmann, A. F. 1963. 'THE FUNCTION OF BILE SALTS IN FAT ABSORPTION. THE SOLVENT PROPERTIES OF DILUTE MICELLAR SOLUTIONS OF CONJUGATED BILE SALTS', *The Biochemical journal*, 89: 57-68.

Hong Kong Chinese Materia Medica Standards Office. 2012. "Hong Kong Chinese Materia Medica Standards - Polygoni Cuspidati Rhizoma et Radix." In, 236-45. Chinese Medicine Division, Department of Health, HKSAR Goverment.

Hricak, H, and R P Lieto. 1983. 'Sonographic determination of renal volume', *Radiology*, 148: 311-12.

Huan, Tao, Erica M. Forsberg, Duane Rinehart, Caroline H. Johnson, Julijana Ivanisevic, H. Paul Benton, Mingliang Fang, Aries Aisporna, Brian Hilmers, Farris L. Poole, Michael P. Thorgersen, Michael W. W. Adams, Gregory Krantz, Matthew W. Fields, Paul D. Robbins, Laura J. Niedernhofer, Trey Ideker, Erica L. Majumder, Judy D. Wall, Nicholas J. W. Rattray, Royston Goodacre, Luke L. Lairson, and Gary Siuzdak. 2017. 'Systems biology guided by XCMS Online metabolomics', *Nature Methods*, 14: 461-62.

Huang, Min-Zong, Sy-Chi Cheng, Yi-Tzu Cho, and Jentaie Shiea. 2011. 'Ambient ionization mass spectrometry: a tutorial', *Analytica Chimica Acta*, 702: 1-15.

Huang, Y. M., D. Xu, J. Long, Y. Shi, L. Zhang, H. Wang, A. Levin, and M. H. Zhao. 2018. 'Spectrum of chronic kidney disease in China: A national study based on hospitalized patients from 2010 to 2015', *Nephrology (Carlton)*: 1-12.

Inker, Lesley A., Christopher H. Schmid, Hocine Tighiouart, John H. Eckfeldt, Harold I. Feldman, Tom Greene, John W. Kusek, Jane Manzi, Frederick Van Lente, Yaping Lucy Zhang, Josef Coresh, Andrew S. Levey, and Ckd-Epi Investigators. 2012. 'Estimating glomerular filtration rate from serum creatinine and cystatin C', *The New England Journal of Medicine*, 367: 20-29.

Inoue, Koichi, Haruhito Tsutsui, Hiroyasu Akatsu, Yoshio Hashizume, Noriyuki Matsukawa, Takayuki Yamamoto, and Toshimasa Toyo'oka. 2013. 'Metabolic profiling of Alzheimer's disease brains', *Scientific Reports*, 3.

International Diabetes Federation. 2000. "IDF Diabetes Atlas 2000." In, 306.

International Diabetes Federation 2017. "IDF Diabetes Atlas 2017." In, 110.

Ipsen, David Højland, Jens Lykkesfeldt, and Pernille Tveden-Nyborg. 2018. 'Molecular mechanisms of hepatic lipid accumulation in non-alcoholic fatty liver disease', *Cellular and Molecular Life Sciences*, 75: 3313-27.

Ismail-Beigi, Faramarz, Timothy Craven, Mary Ann Banerji, Jan Basile, Jorge Calles, Robert M. Cohen, Robert Cuddihy, William C. Cushman, Saul Genuth, Richard H. Grimm, Bruce P. Hamilton, Byron Hoogwerf, Diane Karl, Lois Katz, Armand Krikorian, Patrick O'Connor, Rodica Pop-Busui, Ulrich Schubart, Debra Simmons, Harris Taylor, Abraham Thomas, Daniel Weiss, and Irene Hramiak. 2010. 'Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: an analysis of the ACCORD randomised trial', *The Lancet*, 376: 419-30.

Ismail, Nagwa, Naglaa Omar, Sonia Habib, Nagwa abd el-ghaffar, Mona el kafoury, and Ahmed Talaat. 2012. 'Impact of Splenectomy and Chelating Agents on Serum Cystatin C Levels in Egyptian Children With Beta-Thalassemia 1', *Australian Journal of Basic and Applied Sciences*, , 6: 85-89.

Jackson, Matthew A., Serena Verdi, Maria-Emanuela Maxan, Cheol Min Shin, Jonas Zierer, Ruth C. E. Bowyer, Tiphaine Martin, Frances M. K. Williams, Cristina Menni, Jordana T. Bell, Tim D. Spector, and Claire J. Steves. 2018. 'Gut microbiota associations with common diseases and prescription medications in a population-based cohort', *Nature Communications*, 9: 2655.

Janitza, Silke, Carolin Strobl, and Anne-Laure Boulesteix. 2013. 'An AUC-based permutation variable importance measure for random forests', *BMC bioinformatics*, 14: 119-19.

Jerums, G., E. Premaratne, S. Panagiotopoulos, and R. J. MacIsaac. 2010. 'The clinical significance of hyperfiltration in diabetes', *Diabetologia*, 53: 2093-104.

Ji, G., J. G. Fan, J. J. Chen, L. G. Lu, L. J. Xing, P. Y. Zheng, H. G. Gu, H. F. Wei, S. F. You, and P. T. Zhu. 2008. 'Effectiveness of Danning Tablet in patients with non-alcoholic fatty liver of damp-heat syndrome type: a multicenter randomized controlled trial', *Journal of Chinese Integrative Medicine*, 6: 128-33.

Jiang, Qing-lan, Jun Ma, Jin-yao Pan, Yu-yuan Li, and Jia-yan Lian. 2009. 'Gene Expressions in the Adipose Tissue of NAFLD Rats Intervened with the Extracts of Polygonum Caspidatum', *Journal of Medical Research*, 38: 54-57.

Jiao, Guangling, Joseph P. M. Hui, Ian W. Burton, Marie-Hélène Thibault, Claude Pelletier, Josée Boudreau, Nadia Tchoukanova, Balaji Subramanian, Yahia Djaoued, Stephen Ewart, Jacques Gagnon, Kathryn Vanya Ewart, and Junzeng Zhang. 2015. 'Characterization of Shrimp Oil from Pandalus borealis by High Performance Liquid Chromatography and High Resolution Mass Spectrometry', *Marine drugs*, 13: 3849-76.

Jimeez-Giro, Ana, Clara Ibã Ez, Alejandro Cifuentes, Carolina Simo, Irene Muñ Oz-Gonzaíez, Pedro J. Martín-Lvarez, Begoñ A. Bartolome, and M. Victoria Moreno-Arribas. 2014.

'Faecal Metabolomic Fingerprint after Moderate Consumption of Red Wine by Healthy Subjects', *Journal of Proteome Research*, 14: 897-905.

Jolliffe, I. T. 2002. 'Outlier Detection, Influential Observations, Stability, Sensitivity, and Robust Estimation of Principal Components.' in I. T. Jolliffe (ed.), *Principal Component Analysis* (Springer New York: New York, NY).

Jolliffe, Ian T., and Jorge Cadima. 2016. 'Principal component analysis: a review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374: 20150202.

Jovanović, Dijana, Branislav Gasic, Stevan Pavlovic, and Radomir Naumovic. 2013. 'Correlation of kidney size with kidney function and anthropometric parameters in healthy subjects and patients with chronic kidney diseases', *Renal Failure*, 35: 896-900.

Jurecka, Agnieszka, Marie Zikanova, Stanislav Kmoch, and Anna Tylki-Szymańska. 2015. 'Adenylosuccinate lyase deficiency', *Journal of Inherited Metabolic Disease*, 38: 231-42.

Kametani, Shunichi, Yutaka Hashimoto, Toshikazu Yamanouchi, Yasuo Akanuma, and Hiroshi Akanuma. 1987. 'Reduced Renal Reabsorption of 1, 5-Anhydro-D-Glucitol in Diabetic Rats and Mice', *The Journal of Biochemistry*, 102: 1599-607.

Kang, Jinho, Moon-Young Choi, Sunmi Kang, Hyuk Nam Kwon, He Wen, Chang Hoon Lee, Minseok Park, Susanne Wiklund, Hyo Jin Kim, Sung Won Kwon, and Sunghyouk Park. 2008. 'Application of a 1H Nuclear Magnetic Resonance (NMR) Metabolomics Approach Combined with Orthogonal Projections to Latent Structure-Discriminant Analysis as an Efficient Tool for Discriminating between Korean and Chinese Herbal Medicines', *Journal of Agricultural and Food Chemistry*, 56: 11589-95.

Karjalainen, Leena, Jussi Pihlajamäki, Pauli Karhapää, and Markku Laakso. 1998. 'Impaired Insulin-Stimulated Glucose Oxidation and Free Fatty Acid Suppression in Patients with Familial Combined Hyperlipidemia', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 18: 1548-53.

Karu, Naama, Charlotte McKercher, David S. Nichols, Noel Davies, Robert A. Shellie, Emily F. Hilder, and Matthew D. Jose. 2016. 'Tryptophan metabolism, its relation to inflammation and stress markers and association with psychological and cognitive functioning: Tasmanian Chronic Kidney Disease pilot study', *BMC Nephrology*, 17: 171-71.

Katajamaa, Mikko, and Matej Oresic. 2007. 'Data processing for mass spectrometry-based metabolomics', *Journal of Chromatography A*, 1158: 318-28.

Kawano, Yuki, and David E. Cohen. 2013. 'Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease', *Journal of Gastroenterology*, 48: 434-41.

Keller, Bernd O., Jie Sui, Alex B. Young, and Randy M. Whittal. 2008. 'Interferences and contaminants encountered in modern mass spectrometry', *Analytica Chimica Acta*, 627: 71-81.

Kelly, R. B. 2010. 'Diet and exercise in the management of hyperlipidemia', *Am Fam Physician*, 81: 1097-102.

Kidney disease: improving global outcomes. 2013. 'Chapter 1: Definition and classification of CKD', *Kidney International Supplements*, 3: 19-62.

Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. 2013. 'KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease', *Kidney International Supplements*, 3: 1-150.

Kim, Dong Hyun, Taeho Ahn, Heung Chae Jung, Jae Gu Pan, and Chul Ho Yun. 2009. 'Generation of the human metabolite piceatannol from the anticancer-preventive agent resveratrol by bacterial cytochrome P450 BM3', *Drug Metabolism and Disposition*, 37: 932-36.

Kim, H. J., C. W. Cho, J. T. Hwang, N. Son, J. H. Choi, G. S. Shim, and C. K. Han. 2013. 'LC-MS-based metabolomic analysis of serum and livers from red ginseng-fed rats', *J Ginseng Res*, 37: 371-8.

Kim, Oh, Jong Ho Lee, and Gary Sweeney. 2013. 'Metabolomic profiling as a useful tool for diagnosis and treatment of chronic disease: Focus on obesity, diabetes and cardiovascular diseases', *Expert Review of Cardiovascular Therapy*, 11: 61-8.

Kim, Seong-Ah, Kyungjoon Lim, and Sangah Shin. 2019. 'Associations between Low-Carbohydrate Diets from Animal and Plant Sources and Dyslipidemia among Korean adults', *Journal of the Academy of Nutrition and Dietetics*.

Kim, Sunny, and Kristina S. Boye. 2009. 'Excessive hospitalizations and its associated economic burden among people with diabetes in the United States', *Value in Health*, 12: 267-72.

Kim, Won Jun, Cheol-Young Park, Kyu-Beck Lee, Se Eun Park, Eun Jung Rhee, Won Young Lee, Ki Won Oh, and Sung Woo Park. 2012. 'Serum 1,5-Anhydroglucitol Concentrations Are a Reliable Index of Glycemic Control in Type 2 Diabetes With Mild or Moderate Renal Dysfunction', *Diabetes Care*, 35: 281-86.

Kinoshita, Yosuke, Shinpei Kawakami, Koji Yanae, Shoko Sano, Hiroko Uchida, Hiroyuki Inagaki, and Tatsuhiko Ito. 2013. 'Effect of long-term piceatannol treatment on eNOS levels in cultured endothelial cells', *Biochemical and Biophysical Research Communications*, 430: 1164-68.

Klein, E. A., I. M. Thompson, Jr., C. M. Tangen, J. J. Crowley, M. S. Lucia, P. J. Goodman, L. M. Minasian, L. G. Ford, H. L. Parnes, J. M. Gaziano, D. D. Karp, M. M. Lieber, P. J. Walther, L. Klotz, J. K. Parsons, J. L. Chin, A. K. Darke, S. M. Lippman, G. E. Goodman, F. L. Meyskens, Jr., and L. H. Baker. 2011. 'Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT)', *Journal of the American Medical Association*, 306: 1549-56.

Kobayashi, Toshihiro, Yuriko Matsumura, Toshihiko Ozawa, Hiroyuki Yanai, Atsuo Iwasawa, Toshiaki Kamachi, Kouichi Fujiwara, Noriaki Tanaka, and Masahiro Kohno. 2014. 'Exploration of novel predictive markers in rat plasma of the early stages of chronic renal failure', *Analytical and Bioanalytical Chemistry*, 406: 1365-76.

Kopple, Joel D. 2007. 'Phenylalanine and Tyrosine Metabolism in Chronic Kidney Failure', *The Journal of Nutrition*, 137: 1586S-90S.

Kotronen, Anna, Tuulikki Seppänen-Laakso, Jukka Westerbacka, Tuula Kiviluoto, Johanna Arola, Anna-Liisa Ruskeepää, Matej Oresic, and Hannele Yki-Järvinen. 2009. 'Hepatic stearoyl-CoA desaturase (SCD)-1 activity and diacylglycerol but not ceramide concentrations are increased in the nonalcoholic human fatty liver', *Diabetes*, 58: 203-08.

Koye, Digsu N., Dianna J. Magliano, Christopher M. Reid, Christopher Jepson, Harold I. Feldman, William H. Herman, and Jonathan E. Shaw. 2018. 'Risk of Progression of Nonalbuminuric CKD to End-Stage Kidney Disease in People With Diabetes: The CRIC (Chronic Renal Insufficiency Cohort) Study', *American Journal of Kidney Diseases*, 72: 653-61.

Kumari, Sangeeta, Doug Stevens, Tobias Kind, Carsten Denkert, and Oliver Fiehn. 2011. 'Applying in-silico retention index and mass spectra matching for identification of unknown metabolites in accurate mass GC-TOF mass spectrometry', *Analytical chemistry*, 83: 5895-902.

Kusters, D. M., H. J. Avis, M. J. Braamskamp, R. Huijgen, F. A. Wijburg, J. J. Kastelein, A. Wiegman, and B. A. Hutten. 2013. 'Inheritance pattern of familial hypercholesterolemia and markers of cardiovascular risk', *Journal of Lipid Research*, 54: 2543-49.

Kutil, Zsofia, Veronika Temml, David Maghradze, Marie Pribylova, Marcela Dvorakova, Daniela Schuster, Tomas Vanek, and Premysl Landa. 2014. 'Impact of wines and wine constituents on cyclooxygenase-1, cyclooxygenase-2, and 5-lipoxygenase catalytic activity', *Mediators of Inflammation*, 2014: 1-8.

Kwong, Yuen-Ting, Lesley A. Stevens, Elizabeth Selvin, Yaping Zhang, Tom Greene, Frederick Van Lente, Andrew S. Levey, and Josef Coresh. 2010. 'Imprecision of Urinary Iothalamate Clearance as a Gold-Standard Measure of GFR Decreases the Diagnostic Accuracy of Kidney Function Estimating Equations', *American Journal of Kidney Diseases*, 56: 39-49.

Lame, Mary E., Erin E. Chambers, and Matthew Blatnik. 2011. 'Quantitation of amyloid beta peptides Aβ(1-38), Aβ(1-40), and Aβ(1-42) in human cerebrospinal fluid by ultra-performance liquid chromatography-tandem mass spectrometry', *Analytical biochemistry*, 419: 133-9.

Lau, Jennie Ka Ching, Xiang Zhang, and Jun Yu. 2016. 'Animal models of non-alcoholic fatty liver disease: current perspectives and recent advances', *The Journal of Pathology*.

Lecoffre, C., A.L. Perrine, J. Blacher, and V. Olié. 2018. 'National Prevalence Of Hypercholesterolemia, Treatment And Control, In France In 2015 And Temporal Trends Since 2006', *Journal of Hypertension*, 36: e37-e38.

Levey, A. S., L. A. Stevens, C. H. Schmid, Y. L. Zhang, A. F. Castro, 3rd, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, and J. Coresh. 2009. 'A new equation to estimate glomerular filtration rate', *Annals of Internal medicine*, 150: 604-12.

Levey, Andrew S., Josef Coresh, Ethan Balk, Annamaria T. Kausz, Adeera Levin, Michael W. Steffes, Ronald J. Hogg, Ronald D. Perrone, Joseph Lau, and Garabed Eknoyan. 2003. 'National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification', *Annals of Internal Medicine*, 139: 137-47.

Levey, Andrew S., and Lesley A. Inker. 2016. 'GFR as the "Gold Standard": Estimated, Measured, and True', *American Journal of Kidney Diseases*, 67: 9-12.

Lew, S W, and J P Bosch. 1991. 'Effect of diet on creatinine clearance and excretion in young and elderly healthy subjects and in patients with renal disease', *Journal of the American Society of Nephrology*, 2: 856-65.

Li, Du-Xin, Lin Gan, Amela Bronja, and Oliver J. Schmitz. 2015. 'Gas chromatography coupled to atmospheric pressure ionization mass spectrometry (GC-API-MS): Review', *Analytica Chimica Acta*, 891: 43-61.

Li, H., X. Wang, Y. Liu, D. Pan, Y. Wang, N. Yang, L. Xiang, X. Cai, and Y. Feng. 2017. 'Hepatoprotection and hepatotoxicity of Heshouwu, a Chinese medicinal herb: Context of the paradoxical effect', *Food Chem Toxicol*, 108: 407-18.

Li, Ke, Nan Yu, Pengfei Li, Shimin Song, wu Lei, Yang Li, and Meng Liu. 2017. 'Multi-label spacecraft electrical signal classification method based on DBN and random forest', *PLOS ONE*, 12.

Li, P., and L. P. Yang. 2008. '[Application of systems biology method in the research of traditional Chinese medicine]', *Journal of Chinese Integrative Medicine*, 6: 454-7.

Li, Ping, Li Ping Yang, and Yue Wen Gong. 2009. 'Application of Systems Biology Technology in Research of Traditional Chinese Medicine', *Journal of Traditional Chinese Medicine*, 29: 153-57.

Lieber, Charles S, Maria A Leo, Ki M Mak, Youqing Xu, Qi Cao, Chaoling Ren, Anatoly Ponomarenko, and Leonore M DeCarli. 2004. 'Model of nonalcoholic steatohepatitis', *The American Journal of Clinical Nutrition*, 79: 502-09.

Lim, Lucy, Fangzhi Yan, Stephen Bach, Katianna Pihakari, and David Klein. 2016. 'Fourier Transform Mass Spectrometry: The Transformation of Modern Environmental Analyses', *International journal of molecular sciences*, 17: 104.

Lin, Li Lian, Ching Yi Lien, Ya Chin Cheng, and Kuo Lung Ku. 2007. 'An effective sample preparation approach for screening the anticancer compound piceatannol using HPLC coupled with UV and fluorescence detection', *Journal of Chromatography B*, 853: 175-82.

Lin, Zhang, Carlos M. Vicente Gonçalves, Ling Dai, Hong-mei Lu, Jian-hua Huang, Hongchao Ji, Dong-sheng Wang, Lun-zhao Yi, and Yi-zeng Liang. 2014. 'Exploring metabolic syndrome serum profiling based on gas chromatography mass spectrometry and random forest models', *Analytica Chimica Acta*, 827: 22-7.

Lisowska-Myjak, B. 2014. 'Uremic Toxins and Their Effects on Multiple Organ Systems', *Nephron Clinical Practice*, 128: 303-11.

Liu, Jiali, Lina Han, Leilei Zhu, and Yerong Yu. 2016. 'Free fatty acids, not triglycerides, are associated with non-alcoholic liver injury progression in high fat diet induced obese rats', *Lipids in Health and Disease*, 15: 27.

Liu, Ping, Songlin Liu, Daizhi Tian, and Ping Wang. 2012. 'The applications and obstacles of metabonomics in traditional chinese medicine', *Evidence-Based Complementary and Alternative Medicine*, 2012: 945824-24.

Liu, Wen-Chih, Yasuhiko Tomino, and Kuo-Cheng Lu. 2018. 'Impacts of Indoxyl Sulfate and p-Cresol Sulfate on Chronic Kidney Disease and Mitigating Effects of AST-120', *Toxins (Basel)*. 10: 367.

Liu, Yue-Tao, Jing-Bo Peng, Hong-Mei Jia, Hong-Wu Zhang, Gang Ding, and Zhong-Mei Zou. 2014. 'Urinary metabonomic evaluation of the therapeutic effect of traditional Chinese medicine Xin-Ke-Shu against atherosclerosis rabbits using UPLC–Q/TOF MS', *Chemometrics and Intelligent Laboratory Systems*, 136: 104-14.

Liu, Yue Tao, Jing Bo Peng, Hong Mei Jia, Da Yong Cai, Hong Wu Zhang, Chang Yuan Yu, and Zhong Mei Zou. 2014. 'UPLC-Q/TOF MS standardized Chinese formula Xin-Ke-Shu for the treatment of atherosclerosis in a rabbit model', *Phytomedicine*, 21: 1364-72.

Lommen, Arjen, and Harrie J. Kools. 2012. 'MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware', *Metabolomics : Official journal of the Metabolomic Society*, 8: 719-26.

Luo, Shengyuan, Josef Coresh, Adrienne Tin, Casey M. Rebholz, Lawrence J. Appel, Jingsha Chen, Ramachandran S. Vasan, Amanda H. Anderson, Harold I. Feldman, Paul L. Kimmel, Sushrut S. Waikar, Anna Köttgen, Anne M. Evans, Andrew S. Levey, Lesley A. Inker, Mark J. Sarnak, and Morgan Erika Grams. 2019. 'Serum Metabolomic Alterations Associated with Proteinuria in CKD', *Clinical Journal of the American Society of Nephrology*, 14: 342-53.

Lv, G. P., L. Z. Meng, D. Q. Han, H. Y. Li, J. Zhao, and S. P. Li. 2015. 'Effect of sample preparation on components and liver toxicity of Polygonum multiflorum', *Journal of pharmaceutical and biomedical analysis*, 109: 105-11.

Lv, Yong Hai, Xin Ru Liu, Shi Kai Yan, Xu Liang, Yun Yang, Wei Xing Dai, and Wei Dong Zhang. 2010. 'Metabolomic study of myocardial ischemia and intervention effects of Compound Danshen Tablets in rats using ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry', *Journal of Pharmaceutical and Biomedical Analysis*, 52: 129-35.

Lyons, Timothy J., and Alicia J. Jenkins. 1997. 'Glycation, oxidation, and lipoxidation in the development of the complications of diabetes: a carbonyl stress hypothesis', *Diabetes reviews (Alexandria, Va.)*, 5: 365-91.

Ma, Jing, Jiong Yu, Xiaoru Su, Chengxing Zhu, Xiao Yang, Huawang Sun, Deying Chen, Ying Wang, Hongcui Cao, and Jianxin Lu. 2014. 'UPLC-MS-based serum metabonomics for identifying acute liver injury biomarkers in Chinese miniature pigs', *Toxicology letters*, 225: 358-66.

Magee, G. M., R. W. Bilous, C. R. Cardwell, S. J. Hunter, F. Kee, and D. G. Fogarty. 2009. 'Is hyperfiltration associated with the future risk of developing diabetic nephropathy? A meta-analysis', *Diabetologia*, 52: 691.

Mahamuni, Sagar P., Rekha D. Khose, Farid Menaa, and Sachin L. Badole. 2012. 'Therapeutic approaches to drug targets in hyperlipidemia', *BioMedicine*, 2: 137-46.

Mair, Robert D., Tammy L. Sirich, and Timothy W. Meyer. 2018. 'Uremic Toxin Clearance and Cardiovascular Toxicities', *Toxins*, 10: 226.

Mair, Robert D., Tammy L. Sirich, Natalie S. Plummer, and Timothy W. Meyer. 2018. 'Characteristics of Colon-Derived Uremic Solutes', *Clinical Journal of the American Society of Nephrology*: CJN.03150318.

Makarov, Alexander. 2000. 'Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis', *Analytical Chemistry*, 72: 1156-62.

Mann, H. D., and P. Piotrowski. 1992. 'Diet Modification for Hyperlipidemia: Individual approach to diet planning and education', *Canadian family physician Medecin de famille canadien*, 38: 1483-89.

Margaret D Carroll, Cheryl D. Fryar, and Duong T. Nguyen. 2017. "Total and High-density Lipoprotein Cholesterol in Adults: United States, 2015–2016." In *NCHS Data Brief No. 209*. National Center for Health Statistics.

Markley, John L., Rafael Brüschweiler, Arthur S. Edison, Hamid R. Eghbalnia, Robert Powers, Daniel Raftery, and David S. Wishart. 2017. 'The future of NMR-based metabolomics', *Current Opinion in Biotechnology*, 43: 34-40.

Mastrangelo, Annalaura, and Coral Barbas. 2017. 'Chronic Diseases and Lifestyle Biomarkers Identification by Metabolomics.' in Alessandra Sussulini (ed.), *Metabolomics: From Fundamentals to Clinical Applications* (Springer International Publishing: Cham).

Matić, Ivana, Svetlana Grujić, Zorica Jauković, and Mila Laušević. 2014. 'Trace analysis of selected hormones and sterols in river sediments by liquid chromatography-atmospheric pressure chemical ionization-tandem mass spectrometry', *Journal of Chromatography A*, 1364: 117-27.

Matsui, Yuko, Kenkichi Sugiyama, Masanori Kamei, Toshio Takahashi, Tamio Suzuki, Yohtaro Katagata, and Tatsuhiko Ito. 2010. 'Extract of passion fruit (Passiflora edulis) seed containing high amounts of piceatannol inhibits melanogenesis and promotes collagen synthesis', *Journal of Agricultural and Food Chemistry*, 58: 11112-18.

Matthan, Nirupa R., Esther M. Ooi, Linda Van Horn, Marian L. Neuhouser, Richard Woodman, and Alice H. Lichtenstein. 2014. 'Plasma phospholipid fatty acid biomarkers of dietary fat quality and endogenous metabolism predict coronary heart disease risk: a nested case-control study within the women's health initiative observational study', *Journal of the American Heart Association*, 3: e000764.

McDonald, Patrick D., Douglas McCabe, Bonnie A. Alden, Nicole Lawrence, Daniel P. Walsh, Pamela C. Iraneta, Eric Grumbach, Fang Xia, and Paula Hong. 2006. 'Topics in Liquid Chromatography: Part 1. Designing a Reversed-Phase Column for Polar Compound Retention', *Waters Whitepaper*, September: 8.

McGarrah, Robert W., Scott B. Crown, Guo-Fang Zhang, Svati H. Shah, and Christopher B. Newgard. 2018. 'Cardiovascular Metabolomics', *Circulation Research*, 122: 1238-58.

McQueen, R. B., S. Farahbakhshian, K. F. Bell, K. V. Nair, and J. J. Saseen. 2017. 'Economic burden of comorbid chronic kidney disease and diabetes', *Journal of Medical Economics*, 20: 585-91.

Mei, Hong, Yunsheng Hsieh, Cymbylene Nardo, Xiaoying Xu, Shiyong Wang, Kwokei Ng, and Walter A. Korfmacher. 2003. 'Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: application to drug discovery', *Rapid Communications in Mass Spectrometry*, 17: 97-103.

Mendez, Kevin M., Stacey N. Reinke, and David I. Broadhurst. 2019. 'A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification', *Metabolomics*, 15: 150.

Merino, J., A. Leong, C. T. Liu, B. Porneala, G. A. Walford, M. von Grotthuss, T. J. Wang, J. Flannick, J. Dupuis, D. Levy, R. E. Gerszten, J. C. Florez, and J. B. Meigs. 2018. 'Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose', *Diabetologia*, 61: 1315-24.

Metabolomics Society. 2016. 'Metabolomics', Accessed 24 July, .2016. http://metabolomicssociety.org/index.php/metabolomics.

Metz, Charles E. 1978. 'Basic principles of ROC analysis', *Seminars in Nuclear Medicine*, 8: 283-98.

Meyer, M. R., F. T. Peters, and H. H. Maurer. 2010. 'Automated mass spectral deconvolution and identification system for GC-MS screening for drugs, poisons, and metabolites in urine', *Clin Chem*, 56: 575-84.

Middleton, Kathryn R., Stephen D. Anton, and Michal G. Perri. 2013. 'Long-Term Adherence to Health Behavior Change', *American Journal of Lifestyle Medicine*, 7: 395-404.

Mitchum, Ronald K., and Walter A. Korfmacher. 1983. 'Atmospheric Pressure Ionization Mass Spectrometry', *Analytical Chemistry*, 55: 1485A-99A.

Miura, Daiki, Yutaka Miura, and Kazumi Yagasaki. 2003. 'Hypolipidemic action of dietary resveratrol, a phytoalexin in grapes and red wine, in hepatoma-bearing rats', *Life Sciences*, 73: 1393-400.

Mogensen, C. E., and M. J. Andersen. 1975. 'Increased kidney size and glomerular filtration rate in untreated juvenile diabetes: normalization by insulin-treatment', *Diabetologia*, 11: 221-4.

Moldoveanu, Serban, and Victor David. 2018. 'Derivatization Methods in GC and GC/MS, Gas Chromatography - Derivatization, Sample Preparation, Application.' in Peter Kusch (ed.) (IntechOpen).

Monostori, Péter, Glynis Klinke, Jana Hauke, Sylvia Richter, Jörgen Bierau, Sven F. Garbade, Georg F. Hoffmann, Claus-Dieter Langhans, Dorothea Haas, and Jürgen G. Okun. 2019. 'Extended diagnosis of purine and pyrimidine disorders from urine: LC MS/MS assay development and clinical validation', *PLOS ONE*, 14: e0212458.

Montilla, P., M. C. Mu, I. Bujalance, J. R. Mu, and I. Tepartment. 2004. 'Effect of red wine on oxidative stress and hyper- cholesterolemia induced by feeding a high-cholesterol diet in rat', *Journal of Physiology and Biochemistry*, 60: 259-64.

Mora-Fernández, Carmen, Virginia Domínguez-Pimentel, Mercedes Muros de Fuentes, José L. Górriz, Alberto Martínez-Castelao, and Juan F. Navarro-González. 2014. 'Diabetic kidney disease: from physiology to therapeutics', *Journal of Physiology*, 592: 3997-4012.

Mozaffarian, Dariush, Emelia J. Benjamin, Alan S. Go, Donna K. Arnett, Michael J. Blaha, Mary Cushman, Sarah de Ferranti, Jean-Pierre Després, Heather J. Fullerton, Virginia J. Howard, Mark D. Huffman, Suzanne E. Judd, Brett M. Kissela, Daniel T. Lackland, Judith H. Lichtman, Lynda D. Lisabeth, Simin Liu, Rachel H. Mackey, David B. Matchar, Darren K. McGuire, Emile R. Mohler, Claudia S. Moy, Paul Muntner, Michael E. Mussolino, Khurram Nasir, Robert W. Neumar, Graham Nichol, Latha Palaniappan, Dilip K. Pandey, Mathew J. Reeves, Carlos J. Rodriguez, Paul D. Sorlie,

Joel Stein, Amytis Towfighi, Tanya N. Turan, Salim S. Virani, Joshua Z. Willey, Daniel Woo, Robert W. Yeh, and Melanie B. Turner. 2015. 'Heart disease and stroke statistics—2015 update: A report from the American Heart Association', *Circulation*, 131: e29-e322.

Murias, Marek, Norbert Handler, Thomas Erker, Karin Pleban, Gerhard Ecker, Philipp Saiko, Thomas Szekeres, and Walter Jäger. 2004. 'Resveratrol analogues as selective cyclooxygenase-2 inhibitors: synthesis and structure-activity relationship', *Bioorganic & Medicinal Chemistry*, 12: 5571-78.

Myint, Aye-Mu, Yong Ku Kim, Robert Verkerk, Simon Scharpé, Harry Steinbusch, and Brian Leonard. 2007. 'Kynurenine pathway in major depression: Evidence of impaired neuroprotection', *Journal of Affective Disorders*, 98: 143-51.

Nassir, Fatiha, R. Scott Rector, Ghassan M. Hammoud, and Jamal A. Ibdah. 2015. 'Pathogenesis and Prevention of Hepatic Steatosis', *Gastroenterology & hepatology*, 11: 167-75.

Navar-Boggan, Ann Marie, Eric D. Peterson, Ralph B. D'Agostino, Benjamin Neely, Allan D. Sniderman, and Michael J. Pencina. 2015. 'Hyperlipidemia in early adulthood increases long-term risk of coronary heart disease', *Circulation*, 131: 451-8.

Neirynck, Nathalie, Rita de Smet, Eva Schepers, Raymond Vanholder, and Griet Glorieux. 2012. 'Classification and a List of Uremic Toxins.' in Toshimitsu Niwa (ed.), *Uremic Toxins*.

Nelson, Robert H. 2013. 'Hyperlipidemia as a risk factor for cardiovascular disease', *Primary Care*, 40: 195-211.

Neveu, V., J. Perez-Jimenez, F. Vos, V. Crespy, L. du Chaffaut, L. Mennen, C. Knox, R. Eisner, J. Cruz, D. Wishart, and A. Scalbert. 2010. 'Phenol-Explorer: an online comprehensive database on polyphenol contents in foods', *Database*, 2010: bap024.

Ngamukote, Sathaporn, Kittana Mäkynen, Thavaree Thilawech, and Sirichai Adisakwattana. 2011. 'Cholesterol-lowering activity of the major polyphenols in grape seed', *Molecules*, 16: 5054-61.

Nicholson, J. K., J. C. Lindon, and E. Holmes. 1999. "Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data', *Xenobiotica*, 29: 1181-89.

Niewczas, Monika A., Anna V. Mathew, Stephanie Croall, Jaeman Byun, Melissa Major, Venkatta S. Sabisetti, Adam Smiles, Joseph V. Bonventre, Subramaniam Pennathur, and Andrzej S. Krolewski. 2017. 'Circulating Modified Metabolites and a Risk of ESRD in Patients With Type 1 Diabetes and Chronic Kidney Disease', *Diabetes Care*, 40: 383-90.

Ntambi, James M., and Makoto Miyazaki. 2004. 'Regulation of stearoyl-CoA desaturases and role in metabolism', *Progress in Lipid Research*, 43: 91-104.

Ogawa, Kazuki, Sayumi Hirose, Satoshi Nagaoka, and Emiko Yanase. 2016. 'Interaction between tea polyphenols and bile acid inhibits micellar cholesterol solubility', *Journal of Agricultural and Food Chemistry*, 64: 204-09.

Oliver, Stephen G. 2002. 'Functional genomics: lessons from yeast', *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 357: 17-23.

Oliver, Stephen G., Michael K. Winson, Douglas B. Kell, and Frank Baganz. 1998. 'Systematic functional analysis of the yeast genome', *Trends in Biotechnology*, 16: 373-78.

Ouyang, Ying-bin, William S. Lane, and Kevin L. Moore. 1998. 'Tyrosylprotein sulfotransferase: Purification and molecular cloning of an enzyme that catalyzes tyrosine O-sulfation, a common posttranslational modification of eukaryotic proteins', *Proceedings of the National Academy of Sciences of the United States of America.*, 95: 2896-901.

Ovesná, Zdenka, Katarína Kozics, Yvonne Bader, Philipp Saiko, Norbert Handler, Thomas Erker, and Thomas Szekeres. 2006. 'Antioxidant activity of resveratrol, piceatannol and 3,3',4,4',5,5'-hexahydroxy-trans-stilbene in three leukemia cell lines', *Oncology Reports*, 16: 617-24.

Ozgur, Ceyhun, Taylor Colliau, Grace Rogers, Zachariah Hughes, and Elyse Bennie. 2017. 'MatLab vs. Python vs. R', *Journal of data science: JDS*, 15: 355-72.

Palatini, Paolo. 2012. 'Glomerular hyperfiltration: a marker of early renal damage in pre-diabetes and pre-hypertension', *Nephrology Dialysis Transplantation*, 27: 1708-14.

Panuwet, Parinya, Ronald E. Hunter, Jr., Priya E. D'Souza, Xianyu Chen, Samantha A. Radford, Jordan R. Cohen, M. Elizabeth Marder, Kostya Kartavenka, P. Barry Ryan, and Dana Boyd Barr. 2016. 'Biological Matrix Effects in Quantitative Tandem Mass Spectrometry-Based Analytical Methods: Advancing Biomonitoring', *Critical reviews in analytical chemistry*, 46: 93-105.

Papadimitropoulos, M. P., C. G. Vasilopoulou, C. Maga-Nteve, and M. I. Klapa. 2018. 'Untargeted GC-MS Metabolomics', *Methods in Molecular Biology*, 1738: 133-47.

Parrott, J. M., L. Redus, D. Santana-Coelho, J. Morales, X. Gao, and J. C. O'Connor. 2016. 'Neurotoxic kynurenine metabolism is increased in the dorsal hippocampus and drives distinct depressive behaviors during inflammation', *Translational psychiatry*, 6: e918-e18.

Pedregosa, F.; , G.; Varoquaux, A.; Gramfort, V.; Michel, B.; Thirion, O.; Grisel, M.; Blondel, P.; Prettenhofer, R.; Dubourg Weiss, V.; , J.; Vanderplas, A.; Passos, D.; Cournapeau, M.; Brucher, M.; Perrot, and E. Duchesnay. 2011. 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12: 2825-30.

Petković, Marijana, Jürgen Schiller, Matthias Müller, Stefan Benard, Sabine Reichl, Klaus Arnold, and Jürgen Arnhold. 2001. 'Detection of Individual Phospholipids in Lipid Mixtures by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry: Phosphatidylcholine Prevents the Detection of Further Species', *Analytical Biochemistry*, 289: 202-16.

Pinto, Rui Climaco. 2017. 'Chemometrics Methods and Strategies in Metabolomics.' in Alessandra Sussulini (ed.), *Metabolomics: From Fundamentals to Clinical Applications* (Springer International Publishing: Cham).

Pinu, R. Farhana, G. Silas Villas-Boas, and Raphael Aggio. 2017. 'Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols', *Metabolites*, 7.

Piver, Bertrand, Maude Fer, Xavier Vitrac, Jean-Michel Merillon, Yvonne Dreano, François Berthou, and Danièle Lucas. 2004. 'Involvement of cytochrome P450 1A2 in the biotransformation of trans-resveratrol in human liver microsomes', *Biochemical Pharmacology*, 68: 773-82.

Pizarro, C., I. Arenzana-Ramila, N. Perez-del-Notario, P. Perez-Matute, and J. M. Gonzalez-Saiz. 2013. 'Plasma lipidomic profiling method based on ultrasound extraction and liquid chromatography mass spectrometry', *Analytical Chemistry*, 85: 12085-92.

Pluskal, Tomás, Sandra Castillo, Alejandro Villar-Briones, and Matej Oresic. 2010. 'MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data', *BMC bioinformatics*, 11: 395-95.

Ponthus, Jérémie, and Eleanor Riches. 2013. 'Evaluating the multiple benefits offered by ion mobility-mass spectrometry in oil and petroleum analysis', *International Journal for Ion Mobility Spectrometry*, 16: 95-103.

Pop, L. M., I. Lingvay, Q. Yuan, X. Li, B. Adams-Huet, and N. M. %J Osteoporosis International Maalouf. 2017. 'Impact of pioglitazone on bone mineral density and bone marrow fat content', *Osteoporosis International*, 28: 3261-69.

Portolés, Tania, Johannes G. J. Mol, Juan V. Sancho, and Félix Hernández. 2014. 'Use of electron ionization and atmospheric pressure chemical ionization in gas chromatography coupled to time-of-flight mass spectrometry for screening and identification of organic pollutants in waters', *Journal of Chromatography A*, 1339: 145-53.

Potter, G. A., L. H. Patterson, E. Wanogho, P. J. Perry, P. C. Butler, T. Ijaz, K. C. Ruparelia, J. H. Lamb, P. B. Farmer, L. A. Stanley, and M. D. Burke. 2002. 'The cancer preventative agent resveratrol is converted to the anticancer agent piceatannol by the cytochrome P450 enzyme CYP1B1', *British Journal of Cancer*, 86: 774-78.

Psychogios, Nikolaos, David D. Hau, Jun Peng, An Chi Guo, Rupasri Mandal, Souhaila Bouatra, Igor Sinelnikov, Ramanarayan Krishnamurthy, Roman Eisner, Bijaya Gautam, Nelson Young, Jianguo Xia, Craig Knox, Edison Dong, Paul Huang, Zsuzsanna Hollander, Theresa L. Pedersen, Steven R. Smith, Fiona Bamforth, Russ Greiner, Bruce McManus, John W. Newman, Theodore Goodfriend, and David S. Wishart. 2011. 'The Human Serum Metabolome', *PLOS ONE*, 6: e16957-e57.

Qi, Yun Peng, Chang Tao Jiang, Jie Cheng, Kristopher W. Krausz, Tiangang Li, Jessica M. Ferrell, Frank J. Gonzalez, and John Y. L. Chiang. 2015. 'Bile acid signaling in lipid metabolism: metabolomic and lipidomic analysis of lipid and bile acid markers linked to anti-obesity and anti-diabetes in mice', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851: 19-29.

Qiao, Yi, Jin Sun, Shufang Xia, Xue Tang, Yonghui Shi, and Guowei Le. 2014. 'Effects of resveratrol on gut microbiota and fat storage in a mouse model with high-fat-induced obesity', *Food & Function*, 5: 1241-49.

Rahmioglu, Nilufer, Amelie Fassbender, Allison F. Vitonis, Shelley S. Tworoger, Lone Hummelshoj, Thomas M. D'Hooghe, G. David Adamson, Linda C. Giudice, Christian M. Becker, Krina T. Zondervan, and Stacey A. Missmer. 2014. 'World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: III. Fluid biospecimen collection, processing, and storage in endometriosis research', *Fertility and Sterility*, 102: 1233-43.

Raterink, Robert Jan, Peter W. Lindenburg, Rob J. Vreeken, Rawi Ramautar, and Thomas Hankemeier. 2014. 'Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics', *Trends in Analytical Chemistry*, 61: 157-67.

Rauschert, Sebastian, Olaf Uhl, Berthold Koletzko, and Christian Hellmuth. 2014. 'Metabolomic Biomarkers for Obesity in Humans: A Short Review', *Annals of Nutrition & Metabolism*, 64: 314-24.

Retnakaran, Ravi, Carole A. Cull, Kerensa I. Thorne, Amanda I. Adler, and Rury R. Holman. 2006. 'Risk Factors for Renal Dysfunction in Type 2 Diabetes', *Diabetes*, 55: 1832.

Rigalleau, Vincent, Magalie Garcia, Catherine Lasseur, François Laurent, Michel Montaudon, Christelle Raffaitin, Nicole Barthe, Marie-Christine Beauvieux, Benoit Vendrely, Philippe Chauveau, Christian Combe, and Henri %J BMC Nephrology Gin. 2010. 'Large kidneys predict poor renal outcome in subjects with diabetes and chronic kidney disease', *BMC Nephrology*, 11: 3.

Ringnér, Markus. 2008. 'What is principal component analysis?', *Nature Biotechnology*, 26: 303-04.

Rizza, S., M. Copetti, C. Rossi, M. a Cianfarani, M. Zucchelli, a Luzi, C. Pecchioli, O. Porzio, G. Di Cola, a Urbani, F. Pellegrini, and M. Federici. 2014. 'Metabolomics signature improves the prediction of cardiovascular events in elderly subjects', *Atherosclerosis*, 232: 260-4.

Roberts, Lee D., Amanda L. Souza, Robert E. Gerszten, and Clary B. Clish. 2012. 'Targeted metabolomics', *Current protocols in molecular biology*, Chapter 30: Unit30.2-30.2.24.

Römisch-Margl, Werner, Cornelia Prehn, Ralf Bogumil, Cornelia Röhring, Karsten Suhre, and Jerzy Adamski. 2011. 'Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics', *Metabolomics*, 8: 133-42.

Rooney, O.M., J. Troke, J.K. Nicholson, and J.L. Griffin. 2003. 'High-resolution diffusion and relaxation-edited magic angle spinning 1H NMR spectroscopy of intact liver tissue', *Magnetic Resonance in Medicine*, 50: 925-30.

Roupe, Kathryn A., Jaime A. Yanez, Xiao Wei Teng, and Neal M. Davies. 2006. 'Pharmacokinetics of selected stilbenes: rhapontigenin, piceatannol and pinosylvin in rats', *Journal of Pharmacy and Pharmacology*, 58: 1443-50.

Roux, Aurélie, Dominique Lison, Christophe Junot, and Jean-François Heilier. 2011. 'Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review', *Clinical Biochemistry*, 44: 119-35.

Ruggenenti, Piero, Esteban L. Porrini, Flavio Gaspari, Nicola Motterlini, Antonio Cannata, Fabiola Carrara, Claudia Cella, Silvia Ferrari, Nadia Stucchi, Aneliya Parvanova, Ilian Iliev, Alessandro Roberto Dodesini, Roberto Trevisan, Antonio Bossi, Jelka Zaletel, and Giuseppe Remuzzi. 2012. 'Glomerular Hyperfiltration and Renal Disease Progression in Type 2 Diabetes', *Diabetes Care*, 35: 2061-68.

Ruttkies, Christoph, Emma L. Schymanski, Sebastian Wolf, Juliane Hollender, and Steffen Neumann. 2016. 'MetFrag relaunched: incorporating strategies beyond in silico fragmentation', *Journal of Cheminformatics*, 8: 3.

Rüweler, Milena, Michael Gülden, Edmund Maser, Marek Murias, and Hasso Seibert. 2009. 'Cytotoxic, cytoprotective and antioxidant activities of resveratrol and analogues in C6 astroglioma cells in vitro', *Chemico-Biological Interactions*, 182: 128-35.

Ryan, D., and K. Robards. 2006. 'Metabolomics: The greatest omics of them all?', *Analytical Chemistry*, 78: 7954-58.

Salehi, Bahare, Abhay Prakash Mishra, Manisha Nigam, Bilge Sener, Mehtap Kilic, Mehdi Sharifi-Rad, Patrick Valere Tsouh Fokou, Natália Martins, and Javad Sharifi-Rad. 2018. 'Resveratrol: A Double-Edged Sword in Health Benefits', *Biomedicines*, 6: 91.

Salek, Reza M., Christoph Steinbeck, Mark R. Viant, Royston Goodacre, and Warwick B. Dunn. 2013. 'The role of reporting standards for metabolite annotation and identification in metabolomic studies', *GigaScience*, 2: 13-13.

Sánchez-Avila, N., J. M. Mata-Granados, J. Ruiz-Jiménez, and M. D. Luque de Castro. 2009. 'Fast, sensitive and highly discriminant gas chromatography-mass spectrometry method for profiling analysis of fatty acids in serum', *Journal of Chromatography A*, 1216: 6864-72.

Sangster, Timothy, Hilary Major, Robert Plumb, Amy J. Wilson, and Ian D. Wilson. 2006. 'A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis', *Analyst*, 131: 1075-78.

Sayiner, Mehmet, Aaron Koenig, Linda Henry, and Zobair M. Younossi. 2016. 'Epidemiology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis in the United States and the Rest of the World', *Clinics in Liver Disease*, 20: 205-14.

Schardijn, G. H., and L. W. Statius van Eps. 1987. 'β2-microglobulin: its significance in the evaluation of renal function', *Kidney international*, 32: 635-41.

Scholbach, Th., and D. Weitzel. 2012. 'Body-Surface-Area Related Renal Volume: A Common Normal Range from Birth to Adulthood', *Scientifica*, 2012: 4.

Schürks, Markus, Robert J. Glynn, Pamela M. Rist, Christophe Tzourio, and Tobias Kurth. 2010. 'Effects of vitamin E on stroke subtypes: meta-analysis of randomised controlled trials', *BMJ*, 341: c5702-c02.

Scientific, Thermo Fisher. 2019. 'Manual: Compound Discoverer 3.1 – Metabolomics Tutorial (Revision A)'.

Seegmiller, Jesse C., Bradley E. Burns, Carrie A. Schinstock, John C. Lieske, and Timothy S. Larson. 2016. 'Discordance Between Iothalamate and Iohexol Urinary Clearances', *American Journal of Kidney Diseases*, 67: 49-55.

Seegmiller, Jesse C., W. Greg Miller, and Lorin M. Bachmann. 2017. 'Moving Toward Standardization of Urine Albumin Measurements', *Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*, 28: 258-67.

Sekula, Peggy, Oemer-Necmi Goek, Lydia Quaye, Clara Barrios, Andrew S. Levey, Werner Römisch-Margl, Cristina Menni, Idil Yet, Christian Gieger, Lesley A. Inker, Jerzy Adamski, Wolfram Gronwald, Thomas Illig, Katja Dettmer, Jan Krumsiek, Peter J. Oefner, Ana M. Valdes, Christa Meisinger, Josef Coresh, Tim D. Spector, Robert P. Mohney, Karsten Suhre, Gabi Kastenmüller, and Anna Köttgen. 2016. 'A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population', *Journals of the American Society of Nephrology*, 27: 1175-88.

Shah, P. K., S. Kaul, J. Nilsson, and B. Cercek. 2001. 'Exploiting the vascular protective effects of high-density lipoprotein and its apolipoproteins: an idea whose time for testing is coming, part I', *Circulation*, 104: 2376-83.

Shahidi, Fereidoon, and Priyatharini Ambigaipalan. 2015. 'Phenolics and polyphenolics in foods, beverages and spices: Antioxidant activity and health effects – A review', *Journal of Functional Foods*, 18: 820-97.

Sham, Tung-Ting, Chi-On Chan, You-Hua Wang, Jian-Mei Yang, Daniel Kam-Wah Mok, and Shun-Wan Chan. 2014a. 'A Review on the Traditional Chinese Medicinal Herbs and Formulae with Hypolipidemic Effect', *BioMed Research International*, 2014: 21.

Sham, Tung-Ting, Meng-Heng Li, Chi-On Chan, Huan Zhang, Shun-Wan Chan, and Daniel Kam-Wah Mok. 2017. 'Cholesterol-lowering effects of piceatannol, a stilbene from wine, using untargeted metabolomics', *Journal of Functional Foods*, 28: 127-37.

Sham, Tung Ting, Chi On Chan, You Hua Wang, Jian Mei Yang, Daniel Kam wah Mok, and Shun Wan Chan. 2014b. 'A review on the traditional Chinese medicinal herbs and formulae with hypolipidemic effect', *BioMed Research International*, 2014: 925302.

Shlipak, Michael G., Monica D. Mattes, and Carmen A. Peralta. 2013. 'Update on Cystatin C: Incorporation Into Clinical Practice', *American Journal of Kidney Diseases*, 62: 595-603.

Simental-Mendía, Luis E., and Fernando Guerrero-Romero. 2019. 'Effect of resveratrol supplementation on lipid profile in subjects with dyslipidemia: A randomized double-blind, placebo-controlled trial', *Nutrition*, 58: 7-10.

Simopoulos, Artemis P. 2008. 'The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases', *Experimental Biology and Medicine*, 233: 674-88.

Sjöström, M., S. Wold, and M. Söderström. 1986. 'PLS discriminant plots.' in E. S. Gelsema and L. N. Kanal (eds.), *Pattern recognition in practice II* (Elsevier Science: North-Holland).

Solini, Anna, Maria Laura Manca, Giuseppe Penno, Giuseppe Pugliese, Jeff E. Cobb, and Ele Ferrannini. 2016. 'Prediction of Declining Renal Function and Albuminuria in Patients With Type 2 Diabetes by Metabolomics', *The Journal of Clinical Endocrinology & Metabolism*, 101: 696-704.

Sorgdrager, Freek J. H., Petrus J. W. Naudé, Ido P. Kema, Ellen A. Nollen, and Peter P. De Deyn. 2019. 'Tryptophan Metabolism in Inflammaging: From Biomarker to Therapeutic Target', *Frontiers in Immunology* 10: 2565.

Spratlin, Jennifer L., Natalie J. Serkova, and S. Gail Eckhardt. 2009. 'Clinical Applications of Metabolomics in Oncology: A Review', *Clinical Cancer Research*, 15: 431.

Srivastava, Anubhav, and Darren John Creek. 2019. 'Discovery and Validation of Clinical Biomarkers of Cancer: A Review Combining Metabolomics and Proteomics', *Proteomics*, 19: 1700448.

Stone, Neil J. 1994. 'Secondary causes of hyperlipidemia', *Medical Clinics of North America*, 78: 117-41.

Stroh, Justin G., Christopher J. Petucci, Scott J. Brecker, Nelson Huang, and James M. Lau. 2007. 'Automated Sub-ppm Mass Accuracy on an ESI-TOF for Use with Drug Discovery Compound Libraries', *Journal of the American Society for Mass Spectrometry*, 18: 1612-16.

Stumpf, Chris L., and Jeff Goshawk. 2004. 'The Markerlynx Application Manager: Informatics for Mass Spectrometric Metabonomic Discovery', *Waters Application note*.

Sumner, Lloyd W., Alexander Amberg, Dave Barrett, Michael H. Beale, Richard Beger, Clare A. Daykin, Teresa W. M. Fan, Oliver Fiehn, Royston Goodacre, Julian L. Griffin, Thomas Hankemeier, Nigel Hardy, James Harnly, Richard Higashi, Joachim Kopka, Andrew N. Lane, John C. Lindon, Philip Marriott, Andrew W. Nicholls, Michael D. Reily, John J. Thaden, and Mark R. Viant. 2007. 'Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)', *Metabolomics*, 3: 211-21.

Talmud, P. J., S. Shah, R. Whittall, M. Futema, P. Howard, J. A. Cooper, S. C. Harrison, K. Li, F. Drenos, F. Karpe, H. A. Neil, O. S. Descamps, C. Langenberg, N. Lench, M. Kivimaki, J. Whittaker, A. D. Hingorani, M. Kumari, and S. E. Humphries. 2013. 'Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study', *Lancet*, 381: 1293-301.

Tanaka, A., K. Shima, M. Fukuda, Y. Tahara, Y. Yamamoto, and Y. Kumahara. 1989. 'Tubular dysfunction in the early stage of diabetic nephropathy', *Medical Journal of Osaka University*, 38: 57-63.

Tang, Huilin, Weilong Shi, Shuangshuang Fu, Tiansheng Wang, Suodi Zhai, Yiqing Song, and Jiali Han. 2018. 'Pioglitazone and bladder cancer risk: a systematic review and meta-analysis', *Cancer medicine*, 7: 1070-80.

Tang, Yee Ling, and Shun Wan Chan. 2014. 'A review of the pharmacological effects of piceatannol on cardiovascular diseases', *Phytotherapy Research*, 28: 1581-88.

Tassoni, Annalisa, Nunzio Tango, and Maura Ferri. 2014. 'Polyphenol and biogenic amine profiles of Albana and Lambrusco grape berries and wines obtained following different agricultural and oenological practices', *Food and Nutrition Sciences*, 5: 8-16.

Tautenhahn, Ralf, Gary J. Patti, Duane Rinehart, and Gary Siuzdak. 2012. 'XCMS Online: a web-based platform to process untargeted metabolomic data', *Analytical chemistry*, 84: 5035-39.

Technologies, Agilent. 2016. 'MassHunter Mass Profiler Software Quick Start Guide'.

Theodoridis, Georgios a, Helen G. Gika, Elizabeth J. Want, and Ian D. Wilson. 2012. 'Liquid chromatography-mass spectrometry based global metabolite profiling: a review', *Analytica Chimica Acta*, 711: 7-16.

Theodoridis, Georgios, Helen G. Gika, and Ian D. Wilson. 2008. 'LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics', *Trends in Analytical Chemistry*, 27: 251-60.

Thomas, Merlin C., Richard J. MacIsaac, George Jerums, Andrew Weekes, John Moran, Jonathan E. Shaw, and Robert C. Atkins. 2009. 'Nonalbuminuric Renal Impairment in Type 2 Diabetic Patients and in the General Population (National Evaluation of the

Frequency of Renal Impairment cO-existing with NIDDM [NEFRON] 11)', *Diabetes Care*, 32: 1497-502.

Timmers, Silvie, Ellen Konings, Lena Bilet, Riekelt H Houtkooper, Tineke van de Weijer, Gijs H Goossens, Joris Hoeks, Sophie van der Krieken, Dongryeol Ryu, Sander Kersten, Esther Moonen-Kornips, Matthijs K C. Hesselink, Iris Kunz, Vera B Schrauwen-Hinderling, Ellen E. Blaak, Johan Auwerx, and Patrick Schrauwen. 2011. 'Calorie Restriction-like Effects of 30 Days of Resveratrol Supplementation on Energy Metabolism and Metabolic Profile in Obese Humans', *Cell Metabolism*, 14: 612-22.

Titan, S. M., G. Venturini, K. Padilha, G. Tavares, R. Zatz, I. Bensenor, P. A. Lotufo, E. P. Rhee, R. I. Thadhani, and A. C. Pereira. 2019. 'Metabolites related to eGFR: Evaluation of candidate molecules for GFR estimation using untargeted metabolomics', *Clinica Chimica Acta*, 489: 242-48.

Tolstikov, Vladimir. 2016. 'Metabolomics: Bridging the Gap between Pharmaceutical Development and Population Health', *Metabolites*, 6: 1-13.

Torell, Frida, Kate Bennett, Stefan Rännar, Katrin Lundstedt-Enkel, Torbjörn Lundstedt, and Johan Trygg. 2017. 'The effects of thawing on the plasma metabolome: evaluating differences between thawed plasma and multi-organ samples', *Metabolomics : Official journal of the Metabolomic Society*, 13: 66-66.

Toth, PeterP. 2010. 'Drug Treatment of Hyperlipidaemia: a guide to the rational use of lipid-lowering drugs.', *Drugs*, 70: 1363-79.

Touw, Wouter G., Jumamurat R. Bayjanov, Lex Overmars, Lennart Backus, Jos Boekhorst, Michiel Wels, and Sacha A. F. T. van Hijum. 2012. 'Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle?', *Briefings in Bioinformatics*, 14: 315-26.

Triba, Mohamed N., Laurence Le Moyec, Roland Amathieu, Corentine Goossens, Nadia Bouchemal, Pierre Nahon, Douglas N. Rutledge, and Philippe Savarin. 2015. 'PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters', *Molecular BioSystems*, 11: 13-19.

Trygg, Johan, and Torbjörn Lundstedt. 2007. 'Chapter 6 - Chemometrics Techniques for Metabonomics.' in John C. Lindon, Jeremy K. Nicholson and Elaine Holmes (eds.), *The Handbook of Metabonomics and Metabolomics* (Elsevier Science B.V.: Amsterdam).

Trygg, Johan, and Svante Wold. 2002. 'Orthogonal projections to latent structures (O-PLS)', *Journal of Chemometrics*, 16: 119-28.

Turyan, Iva, Ruth Frenkel, and Zoran Sosic. 2018. 'Rapid quantification of tyrosine sulfation in therapeutic proteins', *Analytical Biochemistry*, 549: 96-98.

Tuttle, Katherine R., George L. Bakris, Rudolf W. Bilous, Jane L. Chiang, Ian H. de Boer, Jordi Goldstein-Fuchs, Irl B. Hirsch, Kamyar Kalantar-Zadeh, Andrew S. Narva, Sankar D. Navaneethan, Joshua J. Neumiller, Uptal D. Patel, Robert E. Ratner, Adam T. Whaley-Connell, and Mark E. Molitch. 2014. 'Diabetic Kidney Disease: A Report From an ADA Consensus Conference', *American Journal of Kidney Diseases*, 64: 510-33.

Tuyiringire, Naasson, Deusdedit Tusubira, Jean-Pierre Munyampundu, Casim Umba Tolo, Claude M. Muvunyi, and Patrick Engeu Ogwang. 2018. 'Application of metabolomics to drug discovery and understanding the mechanisms of action of medicinal plants with anti-tuberculosis activity', *Clinical and translational medicine*, 7: 29-29.

Ueno, Takato, Hiroshi Sugawara, Koodo Sujaku, Osamu Hashimoto, Riko Tsuji, Seishu Tamaki, Takuji Torimura, Sadataka Inuzuka, Michio Sata, and Kyuichi Tanikawa. 1997. 'Therapeutic effects of restricted diet and exercise in obese patients with fatty liver', *Journal of Hepatology*, 27: 103-07.

Umetrics, MKS. 2012. "User Guide to SIMCA." In. Sweden.

van der Kloet, Frans M., Ivana Bobeldijk, Elwin R. Verheij, and Renger H. Jellema. 2009. 'Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping', *Journal of Proteome Research*, 8: 5132-41.

van der Veen, Jelske N., John P. Kennelly, Sereana Wan, Jean E. Vance, Dennis E. Vance, and René L. Jacobs. 2017. 'The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease', *Biochimica et Biophysica Acta - Biomembranes*, 1859: 1558-72.

Vance, D. E. 2008. 'Role of phosphatidylcholine biosynthesis in the regulation of lipoprotein homeostasis', *Current opinion in lipidology*, 19: 229-34.

Vanholder, Raymond Camille, Rita De Smet, and Norbert Hendrik Lameire. 2004. "Uremic toxicity." In *Replacement of renal function by dialysis*, edited by Walter H Hörl, 15-55. Dordrecht ; Boston: Kluwer Academic Publishers.

Vanholder, Raymond, and Rita De Smet. 1999. 'Pathophysiologic Effects of Uremic Retention Solutes', *Journal of the American Society of Nephrology*, 10: 1815-23.

Velenosi, Thomas J., Benjamin K. A. Thomson, Nicholas C. Tonial, Adrien A. E. RaoPeters, Megan A. Mio, Gilles A. Lajoie, Amit X. Garg, Andrew A. House, and Bradley L. Urquhart. 2019. 'Untargeted metabolomics reveals N, N, N-trimethyl-L-alanyl-L-proline betaine (TMAP) as a novel biomarker of kidney function', *Scientific Reports*, 9: 6831.

Viant, Mark R., Irwin J. Kurland, Martin R. Jones, and Warwick B. Dunn. 2017. 'How close are we to complete annotation of metabolomes?', *Current Opinion in Chemical Biology*, 36: 64-69.

Viñas, Pilar, Natalia Campillo, Nelson Martínez-Castillo, and Manuel Hernández-Córdoba. 2009. 'Solid-phase microextraction on-fiber derivatization for the analysis of some polyphenols in wine and grapes using gas chromatography–mass spectrometry', *Journal of Chromatography A*, 1216: 1279-84.

Viñas, Pilar, Nelson Martínez-Castillo, Natalia Campillo, and Manuel Hernández-Córdoba. 2011. 'Directly suspended droplet microextraction with in injection-port derivatization coupled to gas chromatography–mass spectrometry for the analysis of polyphenols in herbal infusions, fruits and functional foods', *Journal of Chromatography A*, 1218: 639-46.

Vuckovic, Dajana. 2012. 'Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry', *Analytical and Bioanalytical Chemistry*, 403: 1523-48.

Vupputuri, Suma, Teresa M. Kimes, Michael O. Calloway, Jennifer B. Christian, David Bruhn, Alan A. Martin, and Gregory A. Nichols. 2014. 'The economic burden of progressive chronic kidney disease among patients with type 2 diabetes', *Journal of Diabetes and its Complications*, 28: 10-16.

Wang, C., R. Feng, D. Sun, Y. Li, X. Bi, and C. Sun. 2011. 'Metabolic profiling of urine in young obese men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC/Q-TOF MS)', *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 879: 2871-76.

Wang, Jia-bo, Hai-ping Zhao, Yan-ling Zhao, Cheng Jin, Dao-jian Liu, Wei-jun Kong, Fang Fang, Lin Zhang, Hong-juan Wang, and Xiao-he Xiao. 2011. 'Hepatotoxicity or Hepatoprotection? Pattern Recognition for the Paradoxical Effect of the Chinese Herb Rheum palmatum L. in Treating Rat Liver Injury', *PLOS ONE*, 6: e24498.

Wang, Jian. 2009. 'Analysis of macrolide antibiotics, using liquid chromatography-mass spectrometry, in food, biological and environmental matrices', *Mass Spectrometry Reviews*, 28: 50-92.

Wang Lu-wen, Xiang Long-kui, Yan Shao-nan, Deng Bin, and Gong Zuo-jiong. 2005. 'Therapeutic effects of Huganning pill on patients with nonalcoholic steatohepatitis', *Chinese Journal of Integrated Traditional and Western Medicine On Gastro-spleen*, 13: 373-75.

Wang, Sinan, Wenxiao Dong, Li Liu, Mengque Xu, Yu Wang, Tianyu Liu, Yujie Zhang, Bangmao Wang, and Hailong Cao. 2019. 'Interplay between bile acids and the gut microbiota promotes intestinal carcinogenesis', *Molecular carcinogenesis*, 58: 1155-67.

Wang, Yu, Jiantao Ye, Jie Li, Cheng Chen, Junying Huang, Peiqing Liu, and Heqing Huang. 2016. 'Polydatin ameliorates lipid and glucose metabolism in type 2 diabetes mellitus by downregulating proprotein convertase subtilisin/kexin type 9 (PCSK9)', *Cardiovascular diabetology*, 15: 19-19.

Want, Elizabeth J., Perrine Masson, Filippos Michopoulos, Ian D. Wilson, Georgios Theodoridis, Robert S. Plumb, John Shockcor, Neil Loftus, Elaine Holmes, and Jeremy K. Nicholson. 2013. 'Global metabolic profiling of animal and human tissues via UPLC-MS', *Nature Protocols*, 8: 17-32.

Want, Elizabeth J., Ian D. Wilson, Helen Gika, Georgios Theodoridis, Robert S. Plumb, John Shockcor, Elaine Holmes, and Jeremy K. Nicholson. 2010. 'Global metabolic profiling procedures for urine using UPLC-MS', *Nature protocols*, 5: 1005-18.

Weaver, Richard, and Rob J. Riley. 2006. 'Identification and reduction of ion suppression effects on pharmacokinetic parameters by polyethylene glycol 400', *Rapid Communications in Mass Spectrometry*, 20: 2559-64.

Wei, Runmin, Jingye Wang, Mingming Su, Erik Jia, Shaoqiu Chen, Tianlu Chen, and Yan Ni. 2018. 'Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data', *Scientific Reports*, 8: 663.

Wen, Yongqing, Xuemei Yuan, Feng Qin, Longshan Zhao, and Zhili Xiong. 2019. 'Development and validation of a hydrophilic interaction ultra-high-performance liquid chromatography–tandem mass spectrometry method for rapid simultaneous determination of 19 free amino acids in rat plasma and urine', *Biomedical Chromatography*, 33: e4387.

Wijeyesekera, A., P. A. Clarke, M. Bictash, I. J. Brown, M. Fidock, T. Ryckmans, I. K. Yap, Q. Chan, J. Stamler, P. Elliott, E. Holmes, and J. K. Nicholson. 2012. 'Quantitative UPLC-MS/MS analysis of the gut microbial co-metabolites phenylacetylglutamine, 4-cresyl sulphate and hippurate in human urine: INTERMAP Study', *Analytical Methods*, 4: 65-72.

Wiklund, Susanne, Erik Johansson, Lina Sjöström, Ewa J. Mellerowicz, Ulf Edlund, John P. Shockcor, Johan Gottfries, Thomas Moritz, and Johan Trygg. 2008. 'Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models', *Analytical Chemistry*, 80: 115-22.

Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, and A. Scalbert. 2013. 'HMDB 3.0--The Human Metabolome Database in 2013', *Nucleic Acids Research*, 41: D801-07.

World Health Organization. 2018a. 'The top 10 causes of death', World Health Organization, Accessed 1 August, 2019. http://www.who.int/mediacentre/factsheets/fs310/en/.

———. 2018b. 'Total NCD Mortality'. http://apps.who.int/gho/data/node.main.A860?lang=en.

Worley, Bradley, and Robert Powers. 2013. 'Multivariate Analysis in Metabolomics', *Current Metabolomics*, 1: 92-107.

Wree, Alexander, Lori Broderick, Ali Canbay, Hal M. Hoffman, and Ariel E. Feldstein. 2013. 'From NAFLD to NASH to cirrhosis—new insights into disease mechanisms', *Nature Reviews Gastroenterology & Hepatology*, 10: 627.

Wu, Qiong, Hai Zhang, Xin Dong, Xiao-Fei Chen, Zhen-Yu Zhu, Zhan-Ying Hong, and Yi-Feng Chai. 2014. 'UPLC-Q-TOF/MS based metabolomic profiling of serum and urine of hyperlipidemic rats induced by high fat diet', *Journal of Pharmaceutical Analysis*: 1-8.

Wu, Zaibin, Dapeng Li, Jie Meng, and Huiwen Wang. 2010. 'Introduction to SIMCA-P and Its Application.' in Vincenzo Esposito Vinzi, Wynne W. Chin, Jörg Henseler and Huiwen Wang (eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (Springer Berlin Heidelberg: Berlin, Heidelberg).

Xi, Bo, Fangchao Liu, Yongchen Hao, Hongbo Dong, and Jie Mi. 2014. 'The growing burden of cardiovascular diseases in China', *International Journal of Cardiology*, 174: 736-37.

Xia, X. H., Y. Y. Yuan, and M. Liu. 2017. 'The assessment of the chronic hepatotoxicity induced by Polygoni Multiflori Radix in rats: A pilot study by using untargeted metabolomics method', *Journal of Ethnopharmacology*, 203: 182-90.

Xie, W., Y. Zhao, and L. Du. 2012. 'Emerging approaches of traditional Chinese medicine formulas for the treatment of hyperlipidemia', *Journal of Ethnopharmacology*, 140: 345-67.

Xu, Qing-Song, and Yi-Zeng Liang. 2001. 'Monte Carlo cross validation', *Chemometrics and Intelligent Laboratory Systems*, 56: 1-11.

Xu, Qing-Song, Yi-Zeng Liang, and Yi-Ping Du. 2004. 'Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration', *Journal of Chemometrics*, 18: 112-20.

Yamanouchi, T., H. Akanuma, T. Nakamura, I. Akaoka, and Y. Akanuma. 1988. 'Reduction of plasma 1,5-anhydroglucitol (1-deoxyglucose) concentration in diabetic patients', *Diabetologia*, 31: 41-5.

Yamanouchi, T., Y. Tachibana, H. Akanuma, S. Minoda, T. Shinohara, H. Moromizato, H. Miyashita, and I. Akaoka. 1992. 'Origin and disposal of 1,5-anhydroglucitol, a major polyol in the human body', *The American Journal of Physiology*, 263: E268-73.

Yamanouchi, Toshikazu, and Yasuo Akanuma. 1994. 'Serum 1,5-anhydroglucitol (1,5 AG): New clinical marker for glycemic control', *Diabetes Research and Clinical Practice*, 24: S261-S68.

Yamanouchi, Toshikazu, Takaomi Shinohara, Nobuyuki Ogata, Yumi Tachibana, Ieo Akaoka, and Hideo Miyashita. 1996. 'Common reabsorption system of 1,5-anhydro-d-glucitol, fructose, and mannose in rat renal tubule', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1291: 89-95.

Yan, Liang-Jun. 2014. 'Positive oxidative stress in aging and aging-related disease tolerance', *Redox biology*, 2: 165-69.

Yao, Hong, Yu-Jie Qiao, Ya-Li Zhao, Xu-Feng Tao, Li-Na Xu, Lian-Hong Yin, Yan Qi, and Jin-Yong Peng. 2016. 'Herbal medicines and nonalcoholic fatty liver disease', *World Journal of Gastroenterology*, 22: 6890-905.

Yi, Lun Zhao, Jun He, Yi Zeng Liang, Da Lin Yuan, and Foo Tim Chau. 2006. 'Plasma fatty acid metabolic profiling and biomarkers of type 2 diabetes mellitus based on GC/MS and PLS-LDA', *FEBS Letters*, 580: 6837-45.

Yiyong, Song. 2016. 'Experimental study on effects of Giant knotweed on lipid peroxidation and oxidative stress in nonalcoholic fatty liver disease', *Chinese Journal of Clinicians (Electronic Edition)*, 10: 1580-83.

Younossi, Zobair M., Aaron B. Koenig, Dinan Abdelatif, Yousef Fazel, Linda Henry, and Mark Wymer. 2016. 'Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes', *Hepatology*, 64: 73-84.

Zhang, Huan, Chang Li, Sin-Tung Kwok, Qing-Wen Zhang, and Shun-Wan Chan. 2013. 'A Review of the Pharmacological Effects of the Dried Root of Polygonum cuspidatum (Hu Zhang) and Its Constituents', *Evidence-Based Complementary and Alternative Medicine*, 2013: 208349-49.

Zhang, J., Y. Tan, F. Yao, and Q. Zhang. 2012. 'Polydatin alleviates non-alcoholic fatty liver disease in rats by inhibiting the expression of TNF-alpha and SREBP-1c', *Molecular Medicine Reports*, 6: 815-20.

Zhang, Qibo, Lisa A. Ford, Anne M. Evans, and Douglas R. Toal. 2017. 'Structure elucidation of metabolite x17299 by interpretation of mass spectrometric data', *Metabolomics*, 13: 92.

Zhang, Shi-Jun, Ze-Xiong Chen, Kai-Ping Jiang, Yong-Hua Cheng, and Yan-Li Gu. 2008. 'The effect of QuYuHuaTanTongLuo Decoction on the non-alcoholic steatohepatitis', *Complementary Therapies in Medicine*, 16: 192-98.

Zhang, Xiaojun, Franky F. K. Choi, Yan Zhou, Feung P. Leung, Shun Tan, Shuhai Lin, Hongxi Xu, Wei Jia, Joseph J. Y. Sung, Zongwei Cai, and Zhaoxiang Bian. 2012. 'Metabolite profiling of plasma and urine from rats with TNBS-induced acute colitis using UPLC-ESI-QTOF-MS-based metabonomics - A pilot study', *FEBS Journal*, 279: 2322-38.

Zhang, Yue, Yizhe Ruan, Ping Zhang, and Li Wang. 2017. 'Increased indoleamine 2,3-dioxygenase activity in type 2 diabetic nephropathy', *Journal of Diabetes and its Complications*, 31: 223-27.

Zhang, Yumin, Siwen Zhang, and Guixia Wang. 2015. 'Metabolomic biomarkers in diabetic kidney diseases—A systematic review', *Journal of Diabetes and its Complications*, 29: 1345-51.

Zhao, Xinjie, Jens Fritsche, Jiangshan Wang, Jing Chen, Kilian Rittig, Philippe Schmitt-Kopplin, Andreas Fritsche, Hans-Ulrich Häring, Erwin D. Schleicher, Guowang Xu, and Rainer Lehmann. 2010. 'Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits', *Metabolomics*, 6: 362-74.

Zheng, Hong, Jinzi Wu, Zhen Jin, and Liang-Jun Yan. 2016. 'Protein Modifications as Manifestations of Hyperglycemic Glucotoxicity in Diabetes and Its Complications', *Biochemistry Insights*, 9: BCI.S36141.

Zhu, Lixian, Xin Luo, and Zhengyu Jin. 2008. 'Effect of resveratrol on serum and liver lipid profile and antioxidant activity in hyperlipidemia rats', *Asian-Australasian Journal of Animal Science*, 21: 890-95.

Zhu, Yongxin, Hwa Chiang, Jean Zhou, Peter T. Kissinger, Bioanalytical Systems, West Lafayette, Beckman Coulter, and San Jose. 2003. 'In vitro netabolism study of resveratrol and identification and determination of its metabolite piceatannol by LC/EC and LC/MSMS', *Discovery*, 3: 93-96.

Zinkhan, Erin K., Jeanette R. Chin, Jennifer M. Zalla, Baifeng Yu, Ben Numpang, Xing Yu, Chengshe Jiang, Christopher W. Callaway, Robert A. McKnight, Lisa Joss-Moore, and Robert H. Lane. 2014. 'Combination of intrauterine growth restriction and a high fat diet impairs cholesterol elimination in rats', *Pediatric Research*, 76: 432-40.