

University of Alberta

**Development of Isotope Labeling Liquid Chromatography Mass
Spectrometry for Biofluid Metabolomics and Applications in
Disease Studies**

by

Jun Peng

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

©Jun Peng
Spring 2014
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

Metabolomics is a research field focusing on global study of all the metabolites present in a biological system (i.e., the metabolome). Metabolome analysis involves identifying and quantifying as many small molecule metabolites as possible in a biological sample. Metabolomics has attracted much attention in recent years and holds promise in application areas such as disease biomarker discovery. The objective of this thesis is to develop isotope labeling LC-MS metabolomics methods to improve the metabolome coverage and quantification precision in biofluids including human urine, mouse urine, and bronchoalveolar lavage fluids, and apply the developed metabolomics methods to study animal model of diseases including Alzheimer's disease and asthma. We developed an offline two-dimensional LC separation strategy based on ion pairing reversed phase LC, coupled with dansylation labeling LC-MS for comprehensive metabolomic profiling of human urine. We also developed a liquid-liquid extraction method coupled with dimethylphenacyl labeling LC-MS for improved metabolomic profiling of organic acids in human urine. We developed a metabolomics workflow for comparative study of mouse model of Alzheimer's disease and wild type, and it is demonstrated that mouse urine metabolomics could be a useful approach to study Alzheimer's disease. We also developed an isotope labeling LC-MS method for bronchoalveolar lavage fluids and applied it to study a rat model of allergic inflammation. Metabolic pathway analysis implicates that arginine-proline metabolic pathway may be associated with

allergic inflammation. Through these research activities, my thesis work has contributed to the development of isotope labeling LC-MS for metabolomics applications in disease studies and biomarker discovery.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Liang Li, for giving me the opportunity to pursue my PhD study in his world-class lab. I really appreciate his great guidance, direction, financial support and patience throughout my studies. He is a terrific mentor who is inspiring and influencing me very much.

I would like to thank the other members of my supervisory committee, Professor Mark McDermott and Professor Derrick Clive, the members of my examining committee, Professor James Harynuk and Professor Dean Befus, and the external thesis examiner, Professor David Muddiman, North Carolina State University, for their active participation during my oral examination, reviews and comments on my thesis, and their invaluable advice on my research.

I want to express my gratitude to many people who I collaborated with during my thesis research. Especially, I thank Dr. Kevin Kun Guo for his training and collaborative work on the ion pairing reversed phase LC offline coupled to dansylation labeling LC-MS for urine metabolome (Chapter 2). I also would like to thank Roukun Zhou for his help on isotope labeling reagent synthesis (Chapter 3). I am very grateful to have worked with Jeff Xia and Ben Zhou from Professor David Wishart's lab at Department of Computer Science and would like to thank them for their contributions in the development of the software for automated data preprocessing (Chapter 4). I would also thank Jing Yang, and Professor David Westaway, and Professor David Wishart on sample collection and advice on

mouse model of Alzheimer's disease study. I am also very grateful to Professor Dean Befus and Chris St. Laurent for collecting BALF samples from rats model of allergic inflammation and their great contributions to the manuscript (Chapter 5).

I would like to acknowledge Dr. Randy Whittal, from the Mass Spectrometry Facility in the Department of Chemistry at the University of Alberta for their maintenance of FTICR and suggestions.

I would also like to thank the departmental staff in the general office, purchasing office, mailroom, supply room, and the electronics and machine shops for their kind assistance. I would like to thank all the members in Professor Li's research group for their helpful discussions and assistance throughout these years.

Finally, I especially thank my parents, Mr. Benxiang Peng, Mrs. Xinglan Han, and my wife Jing Qu for their love, encouragement, and support over the years, and my older daughter, Eileen and my younger daughter, Caitlynn, for the endless happiness they have brought to me.

List of Tables

Table 1.1 Comparison between GC-MS and LC-MS based metabolomics.....	10
Table 2.1. List of metabolites identified by 2D-RPLC-FTICR-MS from a human urine sample along with those identified by 1D-RPLC-FTICR-MS in bold and underlined.....	70
Table 3.1 Summary of the analyte recovery rate in the sample extraction, drying and redissolution.....	94
Table 3.2 Partial list of organic acids putatively identified based on accurate mass Matching HMDB.....	104
Table 4.1. Information on the mouse urine samples used in this study.....	141
Table 4.2 List of top 15 important metabolites found in the mouse urine samples that differentiated the APP mutant group and the wild type group.....	142
Supplemental Table T 4.1. Summary of HMDB and EML database matches for 8 individual QC injections and the combined results.....	144
Table 5.1 List of identified metabolites in normal rat bronchoalveolar lavage fluid.....	190
Table 5.2 Absolute quantification of metabolites in normal rat BALF.....	191
Table 5.3 Metabolites dysregulated in a rat model of allergic airways inflammation.....	192

Table 5.4 Pearson correlation coefficients comparing metabolite levels and leucocyte numbers in bronchoalveolar lavage fluids.....	193
Table 5.5 Pearson correlation coefficients comparing metabolite levels and the percentages of different leucocytes in bronchoalveolar lavage fluids.....	194
Supplemental Table T 5.1. Comparison of discriminant metabolites found in OPLS-DA and PLS-DA.....	196
Supplemental Table T 5.2 Quantification results of two metabolites using the internal standard method and validated using the standard addition method.....	197
Supplemental Table T 5.3 Total cell and differential cell count numbers in the rat BALF.....	198

List of Figures

Figure 1.1 Metabolomics: a branch of systems biology	2
Figure 1.2 General workflows of untargeted profiling and targeted analysis.....	7
Figure 1.3 Schematic drawing of processes of electrospray ionization	13
Figure 1.4 Schematic drawing of ion evaporation mechanism model a) and charge residue mechanism b). Reproduction with permission.....	14
Figure 1.5 Schematic drawing of ion cyclotron motion in the magnet.....	20
Figure 1.6 Schematic drawing of processes of ion excitation and detection, and Four transform mass spectrum.....	21
Figure 1.7 Schematic drawing of Triple quadrupole and MRM.....	23
Figure 1.8 Comparison between MRM and scheduled MRM.....	23
Figure 1.9 Schematic drawing of dansylation labeling.....	27
Figure 1.10 Schematic drawing of dimethylamine phenacyl bromide labling.....	28
Figure 1.11 Schematic drawing of PCA decomposition into scores plot.....	31

Figure 1.12 Schematic drawing of PLS-DA.....	32
Figure 2.1 Schematic of the 2D-RPLC-FTICR-MS workflow for metabolome analysis.....	56
Figure 2.2 Superimposed ion pairing RPLC-UV chromatograms from four consecutive injections of a human urine sample. Several fractions were collected for RPLC-FTICR-MS analysis.....	60
Figure 2.3 Base-peak ion chromatograms of the dansylation-labeled fraction #1 (top trace) and the unlabeled fraction #1 (bottom trace). (B) A representative mass spectrum showing a pair of ion peaks obtained at a particular retention time. Accurate mass match against the Human Metabolome Database indicates that the ion pairs were likely from homovanillic acid or other isomers.....	63
Figure 2.4 Distribution of the number of ion pairs detected by RPLC-FTICR-MS from the 7 fractions collected by IP RPLC of a human urine sample; the numbers of the unique and common ion pairs detected in neighboring fraction(s) are shown. (B) Comparison of the number of ion pairs detected by 2D-RPLC-FTICR-MS and 1D-RPLC-FTICR-MS.....	66
Figure 3.1 Workflow of an existing protocol involving direct DmPA labeling of a sample (A). Working flows of a new protocol using liquid-liquid extraction followed by DmPA labeling with a starting material of (B) 90 μ L of human urine, (C) 20 μ L of human urine and (D) 10 μ L of human urine.....	86
Figure 3.2 Extracted ion chromatograms of a mixture of DmPA labeled ethanolamine (m/z 384.2233) and DmPA labeled hippuric acid (341.1148) (the concentration was 2 mM for each standard and the injection volume was 2 μ L) obtained by (A) the old protocol with direct DmPA labeling and (B) the new	

protocol with LLE and DmPA labeling. Extracted ion chromatograms of asparagine labeled with 2 DmPA tags and asparagine labeled with 3 DmPA tags (the concentration was 2 mM and the injection volume was 2 μ L) obtained by (C) the old protocol and (D) the new protocol.....91

Figure 3.3 Base-peak ion chromatograms of DmPA labeled urine sample prepared using (A) the direct DmPA labeling method and (B) the new LLE DmPA method. (C) the overlaid ion chromatograms of (A) and (B) for comparison.....96

Figure 3.4 Number of peak pairs detected as a function of the volume of the starting101

Figure 4.1 Experimental workflow of isotope labeling LC-MS for comparative metabolomics of mouse urine.....118

Figure 4.2 (A) Inter-day repeatability of overlaid base peak chromatograms of 8 injections of quality control (QC) samples during 4 days LC-MS running; B) the frequency histogram of RSD% relative intensity of 840 metabolites in the QC samples.....122

Figure 4.3 (A) PCA score plot of all the mice urine samples. Green dots were the 8 injections of the quality control sample. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. (B) PLS-DA score plot of all the mice urine samples. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. $R^2Y(\text{cum})$ was 0.719, $Q^2(\text{cum})$ was 0.516. (C) OPLS-DA score plot of all the mice urine samples. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. $R^2Y(\text{cum})$ was 0.854, $Q^2(\text{cum})$ was 0.665.....125

Figure 4.4	Score plots of OPLS-DA at (A) the age of 15-17 weeks, (B) the age of 25-28 weeks, and (C) the age of 30-31 weeks.....	128
Figure 4.5	MS/MS spectra of (A) ¹² C-dansylated methionine standard, (B) ¹² C-dansylated methionine in the mouse urine sample, (C) ¹³ C-dansylated methionine standard, and (D) ¹³ C-dansylated methionine in the mouse urine sample.....	131
Figure 4.6	Plots showing the changed relative concentration level of candidate metabolite biomarkers in the APP mutant (app) and wild type (wt) groups at three different ages: 1 corresponds to 15-17 weeks, 2 corresponds to 25-28 weeks, and 3 corresponds to 30-31 weeks. The horizontal dark line is the median value. The p-value of the difference between the mutant APP and wild type groups at each of three ages was calculated to be <0.001 for these metabolites.....	132
Supplemental Figure S 4.1	Overlaid base-peak ion chromatograms of three experimental replicates of a mixture of ¹² C-dansylated mouse urine and ¹³ C-dansylated mouse urine (1:1, v/v).....	139
Supplemental Figure S 4.2	Score plot of PCA showing the separation between the male mice and the female mice.....	140
Supplemental Figure S 4.3	Validation of the PLS-DA model using 20-permutation test built in the SIMCA-P ⁺ program.....	142
Figure 5.1	Experimental workflow of isotope labeling LC-MS for comparative metabolomics between 18 control rats (naive [n=9] pooled with sensitized but saline challenged, n=9) and 18 sensitized rats challenged with aerosolized ovalbumin (inflamed). The pooled BALF sample was generated from aliquots of	

36 rat BALF samples and used as a global internal standard.....	158
Figure 5.2 Base-peak ion chromatograms obtained (A) without SPE and analyte concentration and (B) with SPE and analyte concentration.....	165
Figure 5.3 (A) Score plot of PCA showing some separation between 9 naïve rats (red triangles), 9 rats sensitized but saline challenged (blue diamonds), and 18 inflamed rats (green squares); black dots were from 8 injections of the quality control sample. (B) Score plot of OPLS-DA showing distinct separation between 18 control rats (red triangles) and 18 inflamed rats (green squares).....	168
Figure 5.4 Heat map generated by the hierarchical clustering method. The rows are individual rat samples and the columns are the top 25 important metabolites. All 18 control rats are clustered together at the bottom half labeled with red and all 18 inflamed rats are clustered together at the top half labeled with green.....	170
Figure 5.5 Arginine-proline metabolic pathway with five metabolites, arginine, proline, ammonia, hydroxyproline, and 4-aminobutanal, found to be dysregulated by metabolomic analysis (highlighted in red) in the rat model of asthma.....	172
Figure 5.6 Cell numbers in bronchoalveolar lavage fluids from control rats (n=18, naïve [n=9] pooled with sensitized but saline challenged, n=9) and sensitized rats challenged with aerosolized ovalbumin (n=18). The sensitized, ovalbumin challenged rats are separated into two subgroups, low and high inflammation, based on metabolomic profiles (see text). (A) total leucocyte numbers, (B) neutrophils, (C) eosinophils, and (D) macrophages; the P values between “low inflammation” and “high inflammation” subgroups are indicated, and also P values between control groups and “low inflammation” or “high	

inflammation”	subgroups	are	
indicated.....			176
Supplemental Figure S 5.1	Effect of sample pH on SPE: (A) pH 3, (B) pH 7 and		
(C) pH 10.....			185
Supplemental Figure S 5.2	Ion chromatogram from normal rat BALF obtained by		
using scheduled MRM for quantification of 25			
metabolites.....			186
Supplemental Figure S 5.3	Cross validation of the OPLS-DA model using 100		
permutation test built in SIMCA-P+			
software.....			187
Supplemental Figure S 5.4	PCA score plot using only 9 naïve rats (labeled as		
red squares), 9 rats sensitized, challenged with saline (labeled as blue diamond)			
and quality control samples (labeled as purple			
dots).....			188
Supplemental Figure S 5.5	Heat map using only 9 naïve rats (labeled as red), 9		
rats sensitized, challenged with saline(labeled as blue) and quality control			
samples			189

List of Abbreviations

ACN	Acetonitrile
ATP	Adenosine triphosphate
BALF	Bronchoalveolar lavage fluid
BPC	Base peak chromatogram
CID	Collision-induced dissociation
CSF	Cerebrospinal fluid
DmPA	p-dimethylaminophenacyl bromide
DnCl	Dansyl chloride
EI	Electron impact ionization
ESI	Electrospray ionization
nM	Nano molarity
nm	Nano meter
FTICR	Fourier-transform ion cyclotron resonance
FWHM	Full width at half maximum
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
h	Hour
HFBA	heptafluorobutyric acid
HILIC	Hydrophilic interaction chromatography
HMDB	Human metabolome database
HPLC	High performance liquid chromatography
IP	Ion pairing
LC-MS	Liquid chromatography mass spectrometry
LC	Liquid chromatography
LLE	Liquid liquid extraction
m	milli- (10^{-3})
m/z	mass to charge
MeOH	Methanol

min	Minute
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
NIST	National Institute of Standards and Technology
PLS-DA	Partial least square-discriminant analysis
	Orthogonal projections to latent structures-
OPLS-DA	discriminant analysis
PCA	Principal component analysis
NMR	Nuclear magnetic resonance
PBr	Phenacyl bromide
ppm	part(s) per million
QC	Quality control
QTOF	Quadrupole time-of-flight
RF	Radio frequency
RP	Reversed-phase
RSD	Relative standard derivation
S/N	Signal to noise ratio
SPE	Solid phase extraction
SIL	stable-isotope-labeled
TEA	Triethylamine
TCA	Tricarboxylic acid
TLC	Thin layer chromatography
TOF	Time-of-flight
UV	Ultra- violet
VIP	Variable importance on the projection
μM	Micro molarity

Contents

Chapter 1	1
Overview of Mass Spectrometry-Based Metabolomics	1
1.1 Introduction of metabolism, metabolome, and metabolomics	1
1.2 Metabolomics in comparison with traditional clinical chemistry	3
1.3 Workflow of untargeted metabolomics and targeted metabolomics	5
1.4 Mass spectrometry techniques	8
1.4.1 LC-MS in comparison with GC-MS	8
1.4.2 Electrospray ionization (ESI)	11
1.4.3 Ion suppression in Electrospray	15
1.4.4 FTICR-MS	15
1.4.5 Triple quadrupole and scheduled MRM	21
1.5 Analytical Challenges in LC-MS Metabolomics	24
1.6 Isotope labeling by chemical derivatization in LC-MS	25
1.7 Data Analysis	28
1.7.1 PCA	28
1.7.2 PLS-DA	31
1.7.3 OPLS-DA	32
1.7.4 Model validation	33
1.7.5 Metabolic Pathway analysis	35
1.8 Overview of the thesis	36
References	38
Chapter 2	51
Ion-Pairing Reversed-Phase Liquid Chromatography Fractionation in Combination with Isotope Labeling Reversed-Phase Liquid Chromatography Mass Spectrometry for Comprehensive Metabolome Profiling	51
2.1 Introduction	51
2.2 Experimental	53
2.2.1 Chemicals and reagents	53
2.2.2 Labeling reaction	53
2.2.3 Ion-pairing RPLC	54
2.2.4 LC-FTICR-MS	55

2.3 Results and discussion.....	57
2.4 Conclusions	68
References	72
Chapter 3	77
Liquid-Liquid Extraction Combined with Differential Isotope Dimethylaminophenacyl Labeling for Improved Metabolomic Profiling of Organic Acids	77
3.1 Introduction	77
3.2 Experimental	80
3.2.1 Chemicals and Reagents	80
3.2.2 Urine and standard solution preparation.....	80
3.2.3 DmPA labeling using the existing protocol.....	80
3.2.4 Liquid-liquid extraction combined with DmPA labeling	81
3.2.5 Measurement of extraction recovery rate	82
3.2.6 Comparison of the overall extraction and labeling efficiency.....	82
3.2.7 UPLC-UV	83
3.2.8 LC-MS	83
3.2.9 Data processing and metabolite database search	84
3.3 Results and Discussion.....	87
3.3.1 Workflow for liquid-liquid extraction and DmPA labeling	87
3.3.2 Method comparison for analyzing standard metabolites	88
3.3.3 Method comparison for analyzing urine samples	94
3.3.4 Low volume sample analysis.....	97
3.3.5 Putative metabolite identification	99
3.4 Conclusions	101
References	106
Chapter 4	109
Development of Isotope Labeling Liquid Chromatography Mass Spectrometry for Mouse Urine Metabolomics: Quantitative Metabolomic Study of Transgenic Mice Related to Alzheimer's Disease	109
4.1 Introduction	109
4.2 Experimental Section	111
4.2.1 Chemicals and reagents.	111

4.2.2 Mouse model and sample collection.	112
4.2.3 Sample preparation.	112
4.2.4 LC-MS.	113
4.2.5 Data analysis.	114
4.2.6 Metabolite identification.	115
4.3 Results and Discussion.	116
Development of automated data preprocessing for isotope labeling LC-MS.	116
Mouse model.	117
Metabolome profiling method.	119
Comparative metabolome analysis in mouse model of AD.	123
Identification of candidate metabolite biomarkers.	129
Significance of candidate metabolite biomarkers.	133
4.4 Conclusions.	137
References.	145
Chapter 5	153
Metabolomic Profiling of Bronchoalveolar Lavage Fluids by Isotope Labeling Liquid Chromatography Mass Spectrometry: a Promising Approach to Studying Experimental Asthma	153
5.1 Introduction	153
5.2 Experimental Section	155
5.2.1 Chemicals and reagents.	155
5.2.2 Rat model and sample collection.	156
5.2.3 BALF sample preparation.	159
5.2.4 Dansylation labeling.	160
5.2.5 Sample preparation for method development.	160
5.2.6 LC-MS.	161
5.2.7 Data analysis.	161
5.2.8 Metabolite identification.	163
5.2.9 Metabolic pathway analysis.	163
5.3 Results	163
5.3.1 Method development.	164
5.3.2 Metabolite identification and absolute concentrations.	166
5.3.3 Comparative metabolome analysis between inflamed rats and control rats.	167
5.3.4 Metabolite biomarkers and pathway analysis.	171

5.3.5 Cell count analysis and its potential relevance to metabolomic profiling.	173
5.4 Discussion	178
5.5 Conclusions	183
Supporting Information Available	184
References	199
Chapter 6	205
Summary and future work	205
6.1 Summary	205
6.2 Future work	209

Chapter 1

Overview of Mass Spectrometry-Based Metabolomics

1.1 Introduction of metabolism, metabolome, and metabolomics

Metabolism refers to the total biochemical reactions taking place within living cells.¹ There are two types of metabolism: one is catabolism and the other is anabolism. Catabolism involves the breaking down of macromolecules, such as proteins and polysaccharides, into small molecules, such as amino acids and monosaccharides, and releasing of energy. Anabolism involves making the macromolecules from small molecules and requiring of energy.² Small molecule metabolites are the intermediates and products of cellular process, including glucose, amino acids, organic acids, ATP, etc.³ The metabolome is the complete collection of metabolites in a cell, organ, tissue, or an organism.^{4, 5} The metabolome mainly includes endogenous metabolites, which are naturally occurring biochemicals in a biosystem, such as human body. It could also include some exogenous compounds, such as commonly used drug and drug metabolites, food components, and some environmental chemicals in the body.^{6, 7} Unlike the genome, metabolome size is not well defined. In 2006 it was estimated there are about 2500 metabolites in the human body, but recently it was estimated to be around 40,000 metabolites.⁶⁻⁸ The number of metabolites might be increased with the advancement of analytical technologies. Metabolomics is the study of metabolome and commonly defined as the global identification and quantification of all small molecules (typically molecular weight of less than 1500 Da) in a biological sample.^{9, 10} Along with genomics, transcriptomics, and proteomics,

metabolomics is an important branch of systems biology. Metabolites are the very end products not only of the genetic setup of the human body, but also of the environmental influence.^{11, 12} (See Figure 1.1) Therefore, the metabolome is closer to the phenotype. Measurement of the metabolome will provide a functional readout of the physiological and pathophysiological state of the human body.^{13, 14} Metabolomics has great potential in many biological applications, such as diseases biomarker discovery,¹⁵⁻²² pharmaceutical development,²³⁻²⁵ toxicology,²⁶⁻²⁸ and plant biology.²⁹⁻³¹

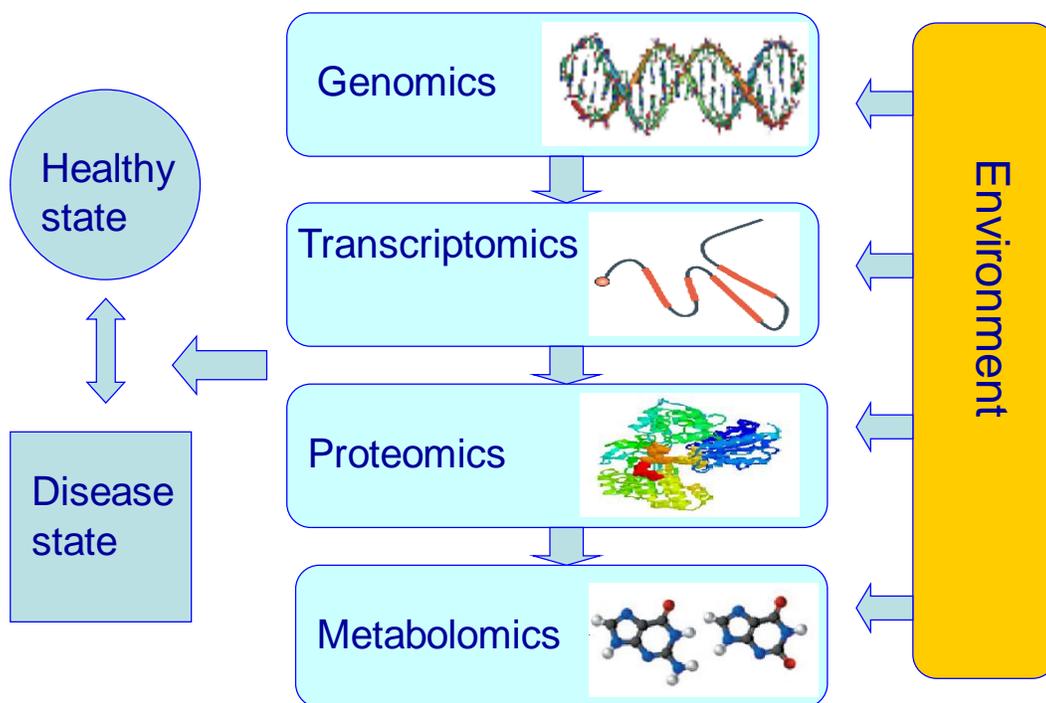


Figure 1.1 Metabolomics: a branch of systems biology. Adapted from Kaddurah-Daouk et al. (Reference 12).

1.2 Metabolomics in comparison with traditional clinical chemistry

Metabolite analysis plays an important role in traditional clinical laboratory tests. About 80% of clinical assays routinely performed in clinical settings are for small molecule metabolites.³² For example, blood glucose test is used for diagnosis of diabetics. Blood lipid analysis can be used to assess the risk of heart diseases. Blood creatinine test can be used to evaluate the function of the kidney. Urine organic acids test can be used to diagnose in-born errors of metabolism, such as methylmalonic acidemia. The analytical techniques used for these assays are mainly based on colorimetry methods and traditional chromatography methods.³³ The majority of these assays involve analyzing a single metabolite or a few metabolites. However, analyzing a few metabolites may not allow us to monitor all the relevant metabolic pathways related to disease diagnosis. Within the concept of systems biology or medicine, it is increasingly important to employ a holistic approach to do the clinical diagnosis and treatment.³⁴ The major difference between traditional clinical chemistry and metabolomics is that metabolomics is attempting to globally identify and quantify as many metabolites as possible in clinical specimens and have the ability to discover novel metabolite biomarkers using the untargeted approach. It may require traditional clinical chemistry approach to translate metabolomics biomarker discovery into clinical assays in clinical laboratories.^{11, 34, 35} The emerging of metabolomics may arise from 1) advances of modern analytical technologies; 2) advances in sophisticated statistical analysis tools; and 3) development of comprehensive human metabolome databases. Firstly, the analytical technologies continue to be

advancing, which includes better separation tools, such as ultra-high performance liquid chromatography (UPLC),³⁶ two-dimensional gas chromatography (2D GC),³⁷ and high resolution and high mass accuracy mass spectrometry (MS).³⁸ The high resolution MS includes Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS), Orbitrap, and new generation quadrupole-time of flight mass spectrometer (Q-TOF).³⁸⁻⁴⁰ These modern techniques allow us to detect many more metabolites that was not possible using traditional methods. Because traditional clinical methods used low-performance analytical techniques such as HPLC with UV detector, the sensitivity and specificity –is not as good as the modern techniques such as LC coupled with high resolution MS. I will discuss the FT-ICR-MS in more detail later. Secondly, advanced statistical analysis tools allow us to do multivariate data analysis, which include principal component analysis (PCA), hierarchical clustering analysis, partial least square regression-discriminant analysis (PLS-DA), and orthogonal projections to latent structures-discriminant analysis (OPLS-DA).⁴¹⁻⁴⁴ Traditional clinical chemistry only monitors single metabolite, therefore, it most often used univariate statistical analysis, such as *t*-test, in two classes. I will discuss these multivariate statistical tools in more details later in this chapter. Thirdly, the development of the metabolome database is also critical, which facilitates the identification of metabolites. For example, the Human Metabolome Database (HMDB) is the by far the most comprehensive and largest human metabolome database^{7, 8} and it contains both NMR spectra and MS/MS spectra of many metabolite standards.

Other databases such as METLIN⁴⁵ and Massbank⁴⁶ also contain many useful MS/MS spectra of metabolite standards.

1.3 Workflow of untargeted metabolomics and targeted metabolomics

There are two approaches to perform metabolomic studies to address biological problems. One approach is untargeted metabolomic profiling and the other approach is targeted analysis.^{47, 45, 48, 49} Figure 1.2 shows the general workflows of untargeted profiling and targeted analysis. As this thesis is based on LC-MS analytical technology, I will use LC-MS technology as examples to illustrate the difference between the two approaches. Both untargeted and targeted metabolomics have advantages and disadvantages. Untargeted profiling starts without *a priori* knowledge on specific metabolites and metabolic pathways⁵⁰ can lead to some hypothesis of biological systems to be studied. In contrast, targeted metabolomics starts with some knowledge on specific metabolites and metabolic pathways and it is used to test a hypothesis.⁴⁷ The sample preparation for untargeted profiling is trying to extract a wide range of metabolites when extraction is necessary. In contrast, extensive sample preparation methods are used for targeted analysis to extract specific metabolites to improve the specificity of the analytical method.⁵¹ The extensive sample preparation may include liquid-liquid extraction (LLE) or solid phase extraction (SPE).^{52, 53} Data acquisition methods are different for untargeted and targeted metabolomics. In untargeted profiling experiment, we typically run LC-MS in full scan mode to acquire the data and the mass analyzer should provide high resolution and mass accuracy, such as FTICR, orbitrap or Q-TOF. In targeted analysis, we typically use LC-

MS/MS in multiple reaction monitoring (MRM) mode to acquire the data, and the mass analyzer is usually a triple quadrupole instrument.⁵⁴⁻⁵⁶ In untargeted profiling, the raw data may contain much noise and data reduction is necessary. In contrast, the raw data of targeted analysis is much cleaner. The data interpretation for untargeted profiling may not be direct and it may not be easy to do specific pathway analysis.⁵⁷ In contrast, the data interpretation for targeted analysis is easier and metabolic pathway analysis is more straightforward.⁵⁸ The choice of untargeted or targeted approach may be dependent on the specific biological problem and the availability of analytical instrument. It might be ideal to combine two approaches to provide complementary data.⁵⁹

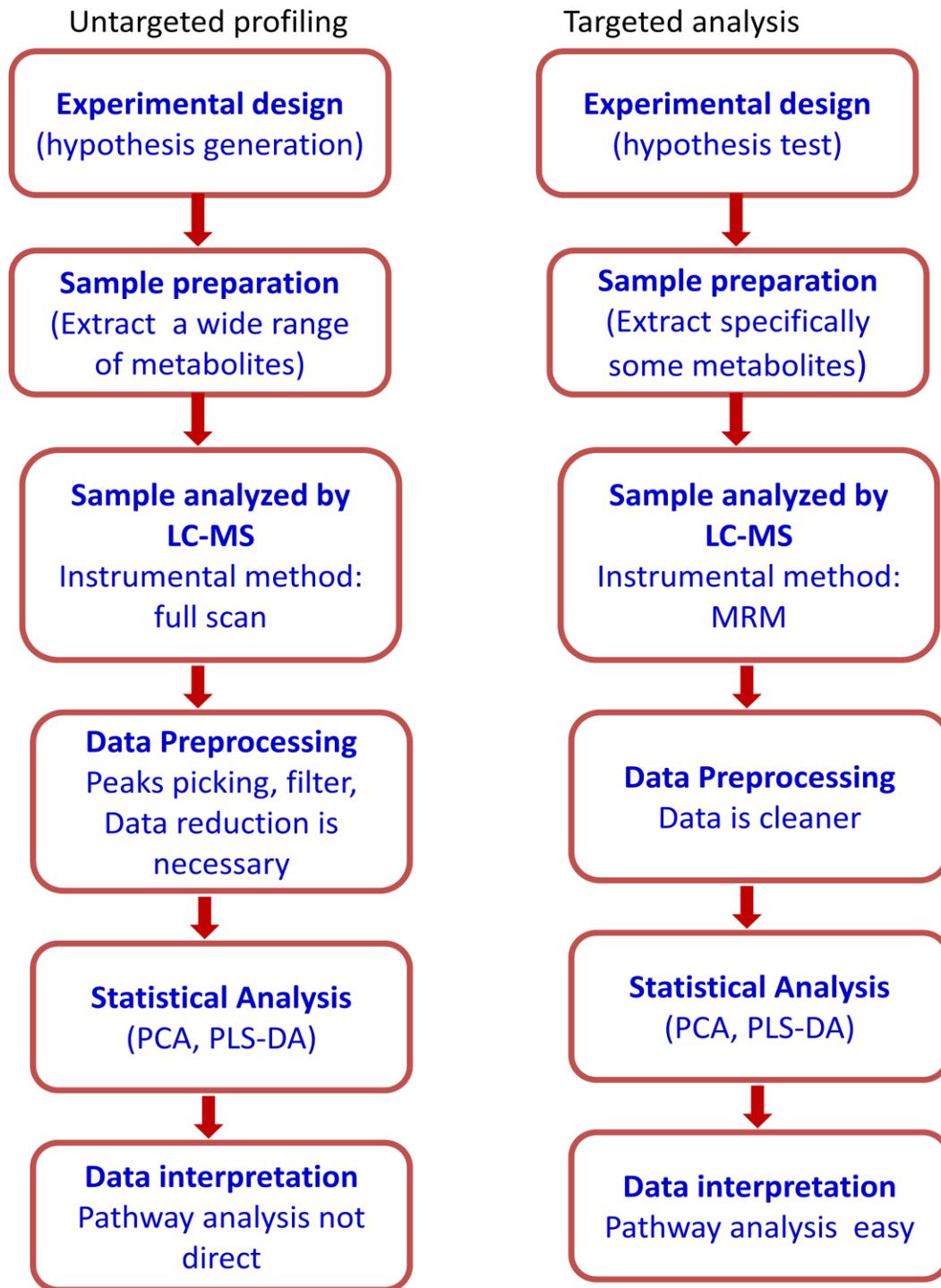


Figure 1.2 General workflows of untargeted profiling and targeted analysis

1.4 Mass spectrometry techniques

There are three most commonly used analytical technologies for metabolomic studies. They are nuclear magnetic resonance spectroscopy (NMR), gas chromatography-mass spectrometry (GC-MS), and liquid chromatography-mass spectrometry (LC-MS). Generally speaking, NMR has an advantage of excellent quantitative reproducibility, but it is less sensitive than GC-MS or LC-MS.⁶⁰ The commonly used 600 MHz NMR typically has a limit of detection around 1-10 μM for most metabolites, therefore, the metabolome coverage is limited for NMR.⁶¹ GC-MS traditionally has been widely used in clinical laboratories for metabolite analysis, and LC-MS has been increasingly employed for metabolomics in recent years.⁶²

1.4.1 LC-MS in comparison with GC-MS

Although GC-MS and LC-MS are mass spectrometry based analytical technologies, they are quite different in terms of separation mechanism, ionization mechanism, etc. Both GC-MS and LC-MS have advantages and disadvantages, therefore, it is better to combine them to provide complementary data to each other.^{63, 64} Here, I just give a brief comparison to introduce some features related to each technology (See table 1). It is not an easy task to compare them since there are different types of GC-MS from different vendors, and also many types of LC-MS from different vendors. Furthermore, metabolites have diverse chemical structures and each metabolite has a different signal response. The following table shows the comparison between the commonly used GC-MS and LC-MS

platforms in most metabolomics laboratories. It needs to be noted that it can only be applicable in most common situations and cannot be generalized for all situations. GC-MS has some strengths which LC-MS does not have. For example, electron impact ionization (EI) is the most commonly used ionization method in GC-MS.⁶⁵ EI spectra are very reproducible; EI spectra library is comprehensive (e.g., NIST has more than 200,000 spectra of compounds).⁶⁶⁻⁶⁸ Compound identification is easier for GC-MS due to the reproducible EI spectra and the larger EI library.^{66, 69} In contrast, electrospray ionization (ESI) MS/MS spectra vary from different mass analyzers and MS/MS library (around 2000 metabolites having MS/MS spectra in HMDB) is not as comprehensive as EI spectra library. GC-MS can readily detect glucose and other sugars, after derivatization, which are not easily detected by LC-MS method.⁶⁷ The advantages of LC-MS include that LC-MS could detect a wider range of metabolites and it could detect some compounds easily which GC-MS could not, such as thermally unstable compounds and large molecules, such as peptides.⁷⁰ LC-MS can easily measure the molecular weight of metabolites accurately and can putatively identify much more metabolites than GC-MS. Nevertheless, they are complementary techniques and should be combined for more comprehensive analysis of the metabolome.^{61, 71}

Table 1.1 Comparison between GC-MS and LC-MS based metabolomics

Techniques	GC-MS	LC-MS
The most commonly used platform	GC-Quadrupole GC-TOF	LC-high resolution QTOF LC-Orbitrap, LC-FTICR
Limit of Detection (specific examples)	Less sensitive For example: around 0.5 μ M for alanine using GC-Quadrupole (Agilent)	More sensitive For example: around 0.5 nM for dansylated phenylalanine using LC-FTICR (Bruker)
Number of metabolite detected	100-500 depending on sample	200-2000 depending on sample
Sample preparation	Chemical derivatization required to make volatile, tedious and long process	directly inject into LC-MS; chemical derivatization is able to improve the performance and quantification by introducing isotope labeling tags
Sample volume	30-100 μ L	10 μ L for direct LC-MS; 20-50 μ L for derivatization
The range of metabolites	Nonvolatile must be derivatized, limited to thermostable, relatively small molecules. Commonly-detected compounds: glucose, fatty acids, organic acids.	A wide range of metabolites including nonvolatile, polar, thermally unstable, large molecules, etc. Commonly-detected compounds: amino acids, hydrophobic amines or acids, phospholipids, peptides.
Commonly used Ionization method	Electron impact ionization (EI)	Electrospray ionization (ESI)
MS spectra	EI spectra are more reproducible in generating fragment ion patterns using different mass spectrometers, compound identification by	Having accurate mass information on molecule weight, MS/MS spectra may vary from different mass spectrometers, MS/MS library not

	matching fragment ion patterns to library, EI spectra library comprehensive, EI has less ion suppression and quantitative precision is better	comprehensive, compound identification by matching accurate mass and MS/MS spectra to library, ESI has severe ion suppression issues and quantitative precision not as good as EI
Cost	Inexpensive \$150 K-250 K	Expensive \$400K-1M

1.4.2 Electrospray ionization (ESI)

The idea of electrospray as an ionization source was originally explored by Dole et al.⁷² Fenn and his colleagues were the first to develop electrospray ionization mass spectrometry and make it possible for analyzing large biomolecules.^{73,74} Nowadays, electrospray is the most commonly used ionization source for LC-MS.^{73,74} In contrast to EI, ESI is a “soft” ionization method, which is able to transfer the ions in solution phase to gas phase without localized heating. Since there is very little extra internal energy imparted to the molecular ions, the molecular ions are often stable and usually do not dissociate into fragment ions.⁷⁵

There are mainly three steps involved in the ESI process.⁷⁶

1) **Generation of charged droplets at the tip of ESI spray capillary.** As is shown in Figure 1.3, the LC effluent will flow through a spray needle, where a high voltage (2-5KV) is applied between the needle and the counterplate or orifice. There is ion separation at the tip of the spray needle and the building up of the charges at the liquid surface causes a Taylor cone to form. At a high enough

voltage, the cone becomes unstable and charged droplets will form and evaporate.⁷⁷

2) **Shrinkage of charged droplets and repeated droplet disintegrations.** As the solvent continues to evaporate from the charged droplet, the droplet will get smaller and smaller. At a certain droplet radius, the repulsion between the charges on the surface will increase to overcome the cohesive force by surface tension. The parent droplet can release a jet of small charged progeny droplets, and the so called Columbic fission can happen. The following Rayleigh equation was used to describe the Columbic fission.⁷⁶

$$Q_{Ry}=8\pi(\epsilon_0\gamma R^3)^{1/2} \quad (\text{Eq. 1})$$

Q_{Ry} is the charge on the droplet. γ is the surface tension of the solvent. R is the radius of the droplet. ϵ_0 is the electrical permittivity.

3) **Formation of gas phase ions.** There are two mechanisms proposed to explain the formation of gas phase ions: Ion evaporation model (IEM) and charged residue mode (CRM).⁷⁸ It is generally agreed that IEM is more applicable for small molecules and CRM is more applicable for large molecules, such as proteins.

In the IEM mechanism, it is assumed that an ion can leave a charged droplet in a hydrated form. As the droplet shrinks and reaches a small droplet size at around 20 nm in diameter, the electric field on the surface is sufficiently high so that the solvated ion can be expelled from the droplet.⁷⁸ (see Figure 1.4a) The transition

state theory is used to describe the ion evaporation process. The reaction rate kinetic is dependent on chemical properties of the ion.

For the CRM, I take an example of electrospray ionization of a protein analyte to explain the mechanism. The spraying process produces droplets that ultimately contain only one protein ion. As the solvent evaporated completely, the charge on the droplet will transfer to the protein molecule. Theoretical calculation showed that at a concentration of $1 \mu\text{M}$, the droplet with a diameter of 200 nm contains on average less than one protein molecule.⁷⁸ (See Figure 1.4b)

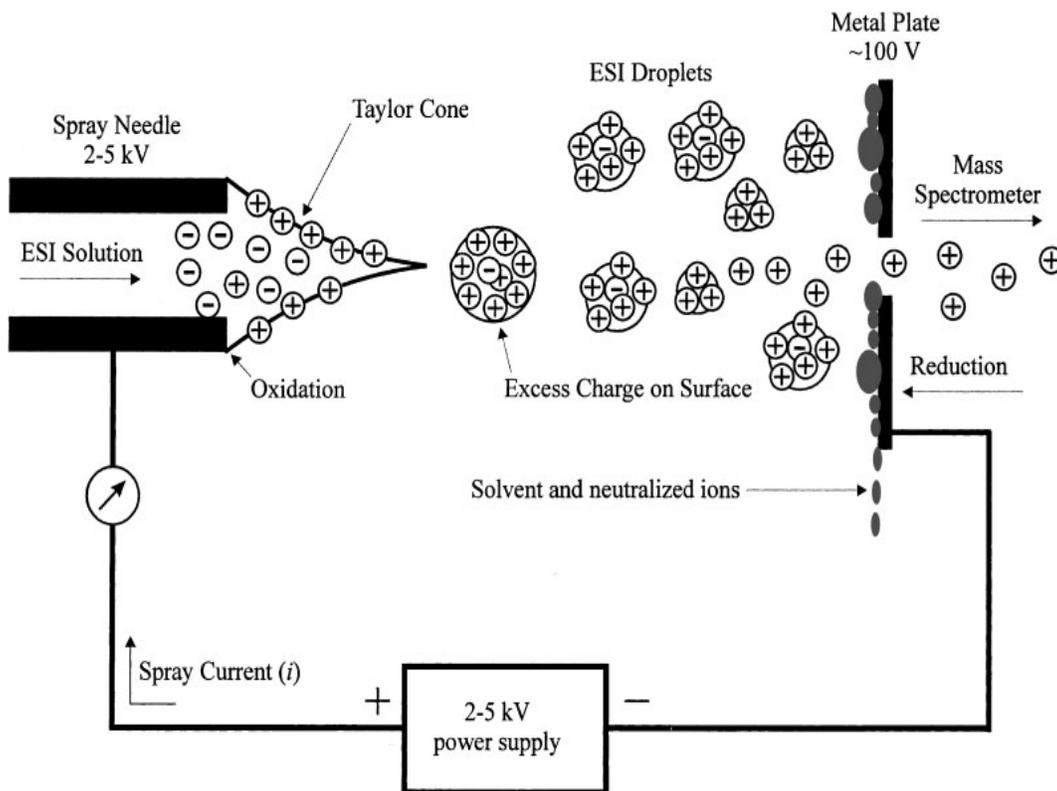
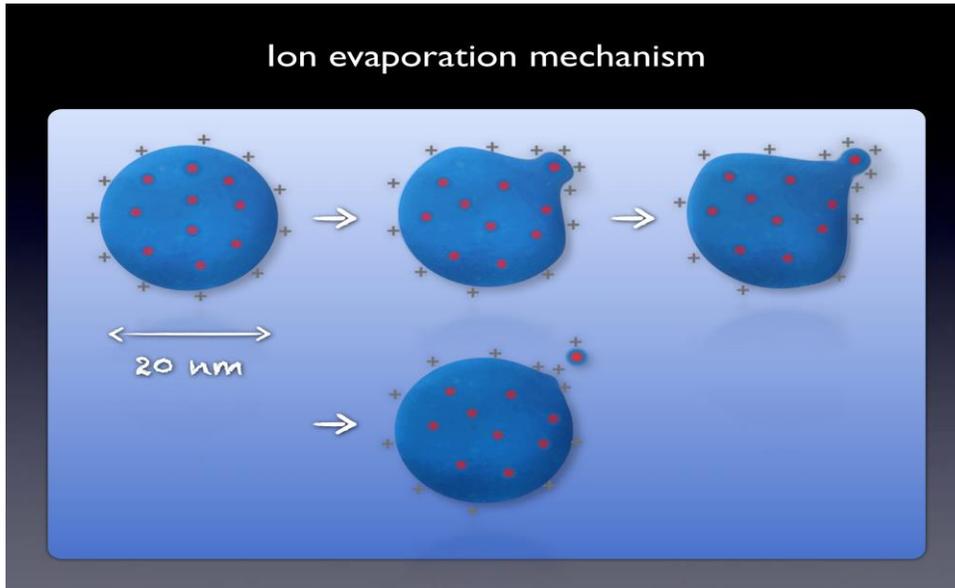


Figure 1.3 Schematic drawing of processes of electrospray ionization.

Reproduction with permission from Cech et al. (Reference 77).

a)



b)

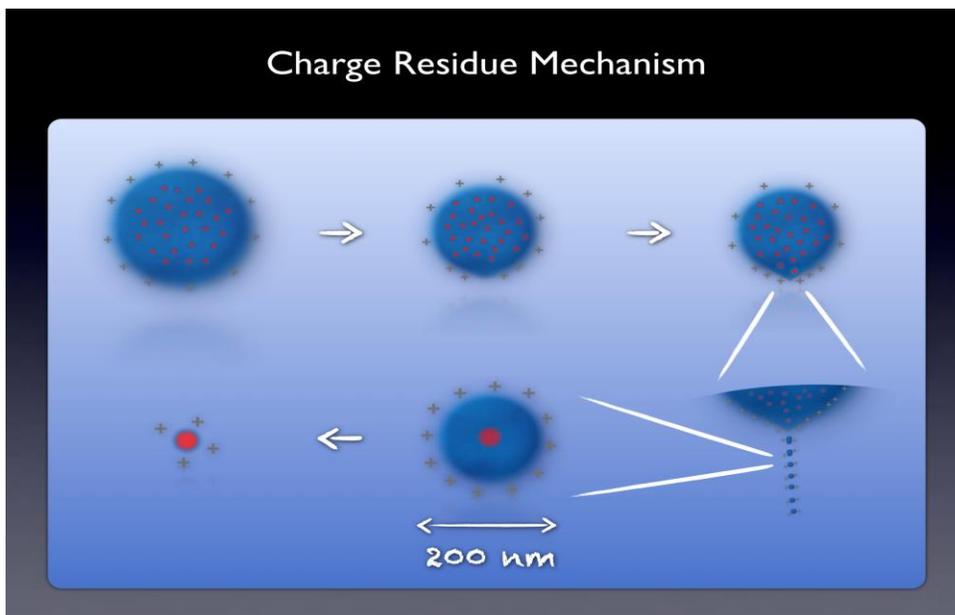


Figure 1.4 Schematic drawing of ion evaporation mechanism model a); and charge residue mechanism b). Reproduction with permission from Wilm (Reference 78).

1.4.3 Ion suppression in Electrospray

LC-ESI-MS is a powerful method for quantitative analysis; however, it can still encounter one major issue, namely ion suppression.^{79, 80} Ion suppression may affect the precision and accuracy of quantification results, therefore, we need to be aware of it and have a deep understanding of its origin and mechanism. Ion suppression is one form of matrix effect, arising from the coeluting endogenous compounds in the matrix, and also exogenous compounds contaminated.⁸¹ One of the mechanisms of ion suppression is that high concentrations of nonvolatile compounds, such as sulfate and phosphate, will inhibit the release of analytes into gas phase.⁸² Nonvolatile material may prevent droplets from reaching their critical radius required for the gas phase ions to be emitted or may cause analyte precipitation.⁸¹ Another proposed mechanism for ion suppression is that there is competition due to limited excess charge and space on the surface of droplet. The coeluting compounds in the matrix that have higher basicities or surface activities will out-compete and cause the suppression of analyte ions.

1.4.4 FTICR-MS

The resolving power of a mass spectrometer is defined as the $m/\Delta m$. m is the measured mass of an isolated peak and Δm is the full width at the half maximum height of the isolated peak. High resolving power of a mass spectrometer means the better ability to distinguish or separate the slight different peaks. The use of high resolution mass spectrometers is desirable to detect a complex sample, such as the metabolome in a biological sample.⁴⁰ High resolving

power can lead to excellent mass accuracy, which is required to deduce the elemental composition of a metabolite. Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) is able to achieve the highest resolution among all different types of mass spectrometers.⁸³ Apex-Qe 9.4 Tesla FTICR from Bruker was used in my thesis research. In FTICR, the ions were generated from the electrospray ionization source. Ion funnels were used to help transport the ions from the atmosphere to low vacuum, therefore improving the ion transmission efficiency.⁸⁴ The ions transported through ion paths, including hexapoles and quadrupole, and then were injected into a cell in the superconducting magnet. A cylindrical cell is placed inside the magnet (see Figures 1.5 and 1.6). The wall of cylindrical cell is divided by four equal plates, two excitation plate, and two detection plates. Two ends of the cell are trapping plates. Ions were injected into the cell and were trapped into the cell. In a strong magnet field, the charged particle (ions) with q and velocity v , will experience Lorentz force. The Lorentz force inward is counterbalanced by the centrifugal force outward.⁸⁵ Ions move circularly in a plane perpendicular to the direction of magnet. The frequency of ion cyclotron is related to magnet field and the mass-to-charge ratio. This is only applicable in the homogeneous magnetic field without applying any electric field. However, it is a necessity to apply the trapping potential to trap ions in the axial direction, therefore the ions motion actually is a complex three dimensional movement in the analyzer cell. In additional to cyclotron motion, there is trapping motion, which is harmonic oscillation in the electric field between trapping plates. There is also magnetron motion, which arises from the combination of magnet

field and electric field.⁸⁶ The frequency of the magnetron motion ranges from 1 to 100 Hz, which is much smaller than the frequency of the cyclotron motion ranged from 5 KHz to 5MHz.⁸⁷ The so called “effective” cyclotron frequency is the cyclotron frequency minus the magnetron frequency. When ultra high resolution is required, it is necessary to reduce the effect of magnetron frequency as much as possible.

Ion cyclotron rotates very close to center of the cell and the orbital radii is at the level of sub millimeter at the room temperature; therefore, it is not easy to detect them. In order to detect those trapped ions, we need to excite them. RF voltage is applied to the excitation plate of the cylindrical cell. The so called Chirp frequency can be swept at the same time, which means a range of frequencies was applied simultaneously.⁸⁵ When the RF frequency is the same as the frequency of ion cyclotron, the ions will be resonant and excited from the center to the larger orbital. The RF field also makes all the ions coherent. (spatial coherence to form the ion packets) When the ions pass by the detection plates, an image current is induced. The image current signal decays with time and is measured as a function of time. Then we can use Fourier transform to convert the time domain to frequency domain. Finally we can obtain a mass spectrum after converting from frequency to mass/charge ratio. (See Figure 1.6)

Since the resolving power is the most important parameter, it is essential to have a good understanding what factors affect the resolution of FTICR. The resolving power at low pressure can be defined as the following equations.³⁹

$$\frac{m}{\Delta m} = \frac{1.274 \times 10^7 \text{ z } B_0 T_{\text{aqn}}}{m} \quad (\text{Eq. 2})$$

Where m/z is mass-to-charge ratio, B_0 is the magnetic field, and T_{aqn} is the acquisition time.

We can deduce from the above equation that the higher magnetic field, the better resolution we can achieve. The longer acquisition during or transient, the better resolution we get. As the magnetic field increases, the resolution and scan speed will increase linearly. The low m/z ion has a higher resolution than high m/z ions under the same conditions. Longer transient will be beneficial to improve the resolution, however, in a LC-MS experiment, the transient time needs to be compatible with a LC time scale. Some other factors may affect the resolution; for example, space charge⁸⁸ can affect the resolution. As is mentioned earlier, the magnetron frequency due to the trapping electric field can also affect the resolution.

FTICR technique continues to evolve and advance. Two recent advances are the dynamic harmonization cell⁸⁹ and the absorption mode.^{90, 91} Both can improve the resolution significantly. These techniques have been implemented in the latest model of Bruker FTICR instrument.

The FTICR-MS needs to be calibrated properly to achieve the excellent mass accuracy. The external calibration method is used in our experiments and typically 2 ppm mass accuracy is readily achieved within 4 days. A mixture of four authentic standards: dansylated glycine, dansylated phenylalanine, dansylated threonine, and dansylated glycine-tryptophan (each having 0.5 μM concentration)

was directly infused into MS. The following equation is applied for the calibration method we used in Bruker FTICR-MS.

$$f = \frac{a}{m} + b \quad (\text{Eq. 3})$$

f is frequency; a and b are determined experimentally; m is the mass of the standards.

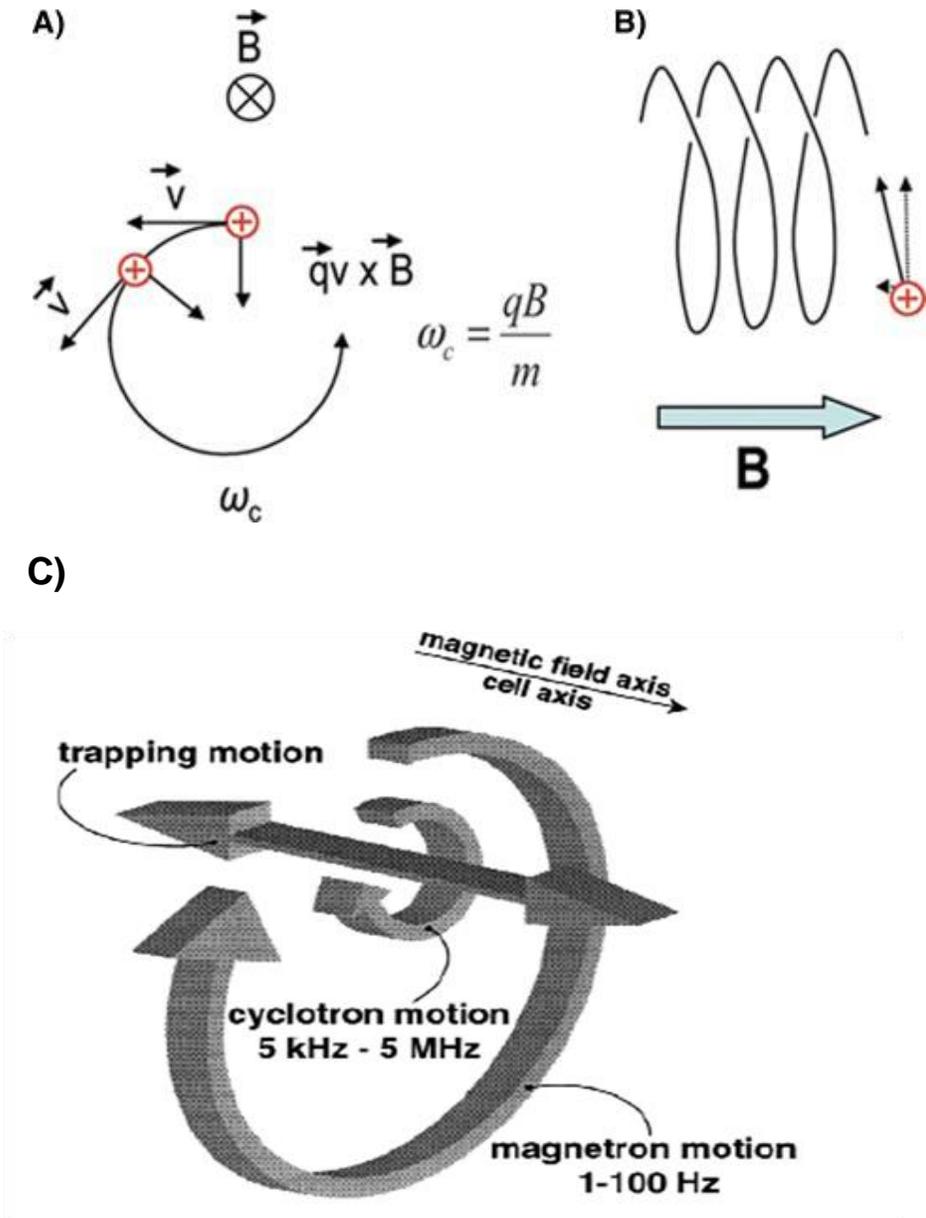


Figure 1.5 Schematic drawing of ion cyclotron motion in the magnet. Circular trajectory in homogeneous magnet A); and B) magnetic field trapping ion radially; and C) complex cyclotron motion with trapping motion, magnetron motion. Reproduction with permission from Scigelova et al. (Reference 39) and Schmid et al. (Reference 87).

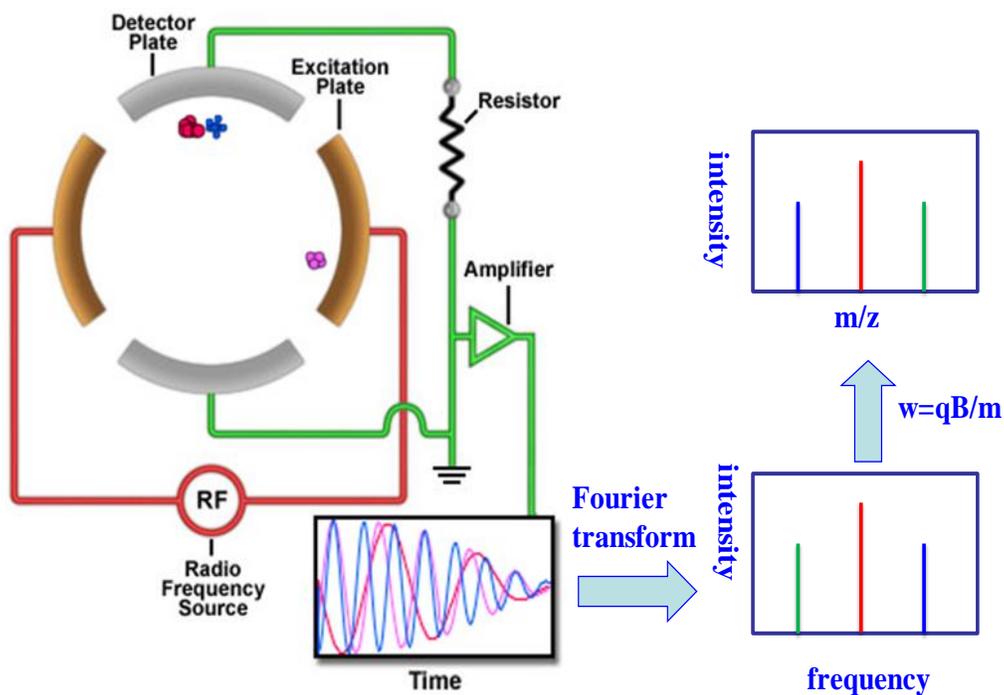


Figure 1.6 Schematic drawing of processes of ion excitation and detection, and Fourier transform mass spectrum.

Adapted from

<http://www.magnet.fsu.edu/education/tutorials/magnetacademy/fticr/page7.html>

1.4.5 Triple quadrupole and scheduled MRM

Quadrupole MS is also called quadrupole mass filter.⁹² It has four parallel cylindrical rods. Two opposite rods are labeled as positive, and two opposite rods are labeled as negative. Both the DC voltage and AC voltage applied between positive rods and negative rods. These applied voltages affect the trajectory of

ions. For given DC and AC voltages, only ions of a specific mass to charge ratio can go through the quadrupole filter.⁹³

In triple quadrupole, there are three quadrupoles Q1, Q2, Q3 in tandem physically (see Figure 1.7).⁹⁴ Multiple reaction monitoring (MRM) is a specific MS/MS acquisition mode for triple quadrupole.⁹⁵ Q1 is used to select the precursor ion. Only RF is applied to Q2, and Q2 does not function as mass filter. Actually, it is used for collision cell. Q3 is used to select specific fragment ions, for example one fragment ion as quantitative, one fragment ion as qualitative. By reducing the chemical noise, MRM can improve sensitivity and specificity dramatically.

Each MRM transition needs a dwell time. The typical dwell time can be set to 50 ms in the QTRAP 2000 from AB Sciex. If we want to monitor many MRM transitions at one time, for example 40 transitions, then dwell time of 40 MRM transitions add up to have a long cycle time (2 s). The longer cycle time can affect the number of data points across the chromatography peak, therefore, it can affect the precision of quantification. In the scheduled MRM, the algorithm intelligently uses the retention time of each metabolite. Each MRM transition will be monitored at a shorter retention time window.⁹⁶ Therefore, the software can reduce the number of concurrent MRM transitions and allow the optimal cycle time and dwell time.⁹⁷ The scheduled MRM can improve the peak shape of chromatogram and improve the quantification precision (see Figure 1.8).

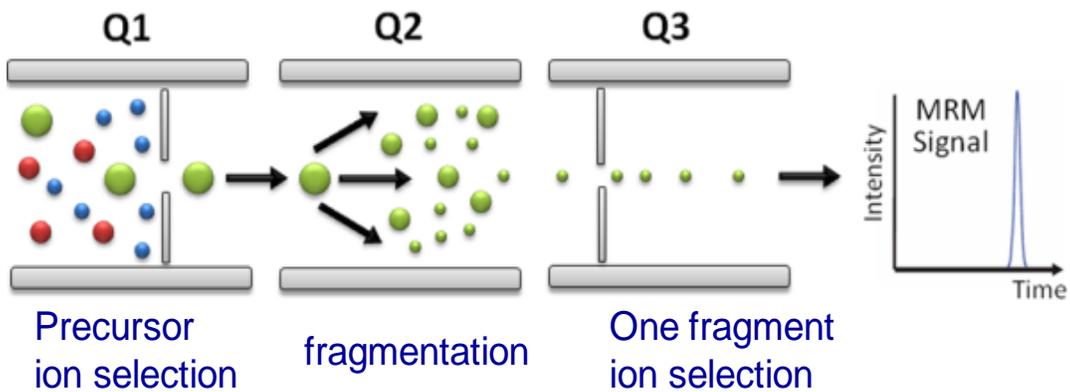


Figure 1.7 Schematic drawing of triple quadrupole MS and MRM.

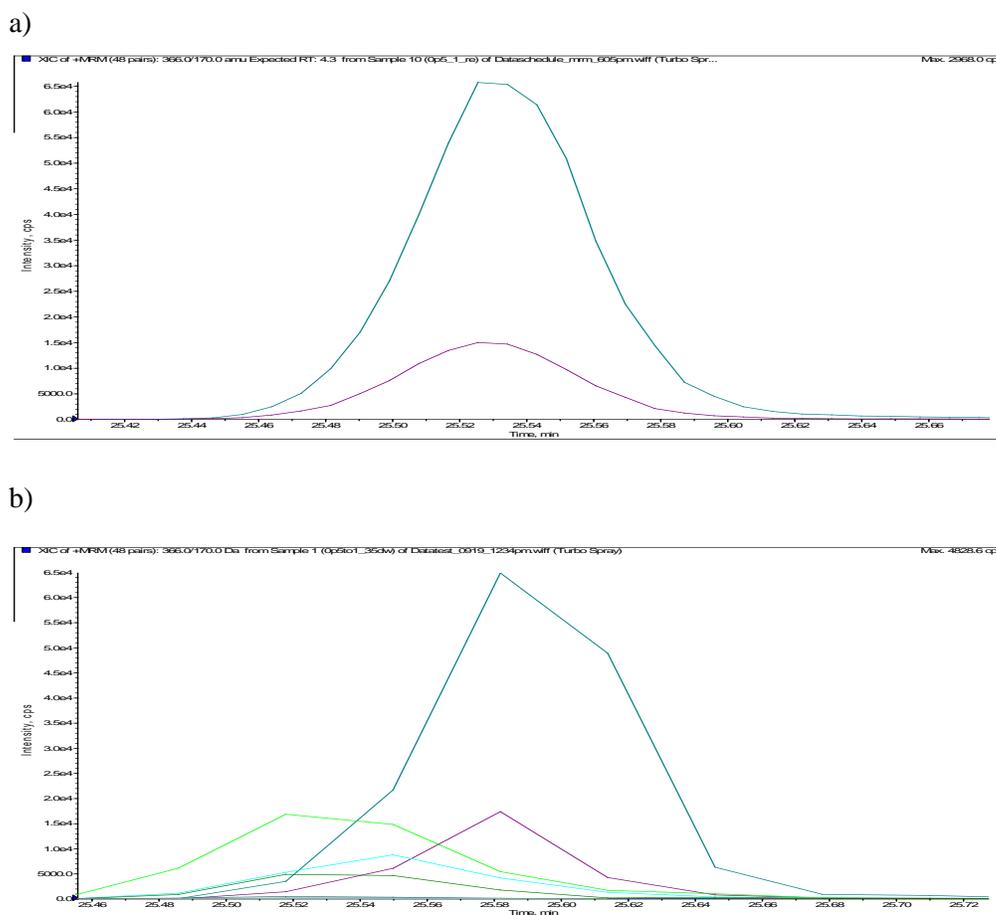


Figure 1.8 Comparison between (a) scheduled MRM; and (b) MRM.

1.5 Analytical Challenges in LC-MS Metabolomics

Metabolomics technologies advance significantly in recent years; however, there are still quite a number of major technical challenges that need to be addressed from the analytical chemistry perspective.

The 1st major analytical challenge is metabolite identification. Metabolome consists of different varieties of small chemicals. Metabolites contain many diverse functional groups, such as amine, carboxylic acid, hydroxyl, ester, polyol, aldehyde, ketone, etc. The diverse chemical structures present a daunting task to identify all the metabolites in the biological sample. It is practically impossible to obtain all authentic standards of the entire metabolome. Often time we just putatively identify most metabolites by searching metabolome databases. The establishment of comprehensive metabolome databases on accurate mass and tandem MS is essentially helpful to facilitate the metabolite identification. However, there are still many unknown metabolites existing and they are not included in any database.

The 2nd major challenge is that metabolome coverage in current metabolome profiling studies is limited, not comprehensive and there is no single analytical technique which is able to detect and quantify all the metabolites in any sample. The metabolite concentration dynamic range is extremely high. For example, in the blood plasma, it is estimated to be 9 orders of magnitude. One dimensional LC-MS at most is able to detect metabolites from high concentration 1 mM to low concentration 1 nM (6 order of magnitude), which is not sufficient

to cover the wider range of concentration differences in a biofluid sample. The physical and chemical diversity of metabolites requires a combination of different analytical techniques to cover as many metabolites as possible. The conventional single LC-MS approach only detects and quantifies less than 1000 metabolites, which is only a small fraction of the whole metabolome.

The 3rd major challenge is that the reproducibility is not satisfactory, especially in the case of LC-MS. It is mainly due to the ion suppression issue discussed earlier. Different samples may have different matrix effect on the metabolite signal on LC-MS, which may cause relatively large technical variations. So it is desirable to have a better reproducible and more accurate LC-MS method to minimize the technical variation, which is particularly important for reliable biomarker discovery.

The 4th major challenge is a need to improve data preprocessing software. The peak picking algorithm can detect many features; however, many of the peaks are from noise and background.⁹⁸ It is a difficult task to pick all the biological relevant peaks confidently at a lower false positive rate and a lower false negative rate.

1.6 Isotope labeling by chemical derivatization in LC-MS

To address these challenges in LC-MS metabolomics, especially the challenges related to metabolome coverage and quantitative precision, our lab has

developed isotope labeling LC-MS using chemical derivatization. As I discussed earlier, the metabolome is quite complex and metabolites are diverse in chemical property. Therefore, there is no single analytical technique that can handle the complex metabolome. We undertake the “divide and conquer” strategy as is applied in proteomics.^{99, 100} We use a classical chemical derivatization reaction, dansylation reaction, to target the primary, secondary amine and phenol compounds (See Figure 1.9). We introduce two ¹³C isotope into the dimethylamino group in the dansyl chloride reagent.¹⁰¹⁻¹⁰³ The naphthalene in the dansyl reagent can improve hydrophobicity and retention of metabolites in the reversed-phase (RP) LC. And more strikingly, the dansyl group can enhance the ESI response by 1-3 orders of magnitude due to the dimethyl amine basicity and surface activity of naphthalene.^{77, 104} Dansyl labeling can improve the detectability, therefore, leading to better metabolome coverage. In our relative quantitative work, a pooled sample is ¹³C dansyl labeled, and is spiked into the individual sample, which is ¹²C dansyl labeled. Therefore the ¹³C labeled pooled sample is used as a global internal standard, and this isotope dilution strategy^{105, 106} can improve the precision of relative quantitative for all detected metabolites. We also can do the absolute quantitative analysis by spiking authentic standards which are ¹³C dansyl labeled into individual sample which is ¹²C dansyl labeled.

Since dansylation reaction mainly targets primary, secondary amine groups and phenol groups, our lab has also developed an isotope labeling reagent to deal with metabolites containing carboxylic acid groups, which are also an important class of metabolites. (See Figure 1.10) The derivatization was based on classical

phenacyl ester reaction, but dimethylamine group was introduced to improve ESI signal intensity.¹⁰⁷ As in the case of dansyl labeling, ¹²C/¹³C differential isotope dilution strategy used for acid labeling can improve the reproducibility of LC-MS metabolomics.

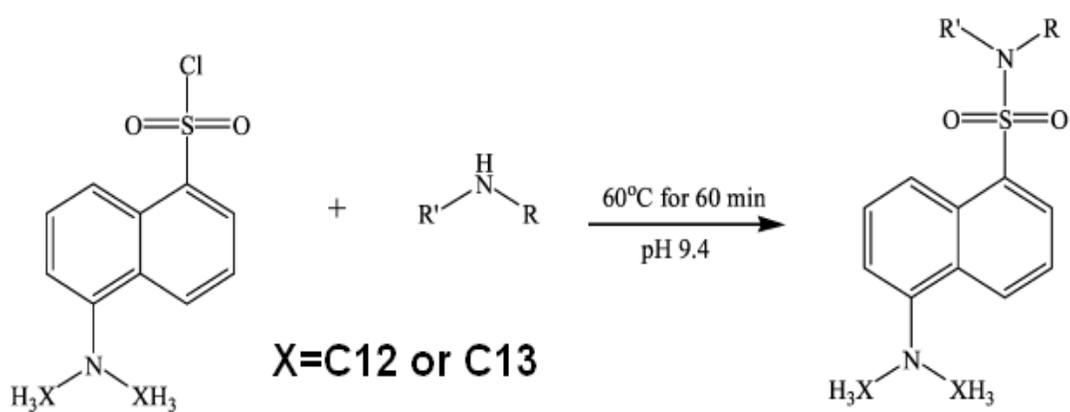


Figure 1.9 Schematic drawing of dansylation labeling.

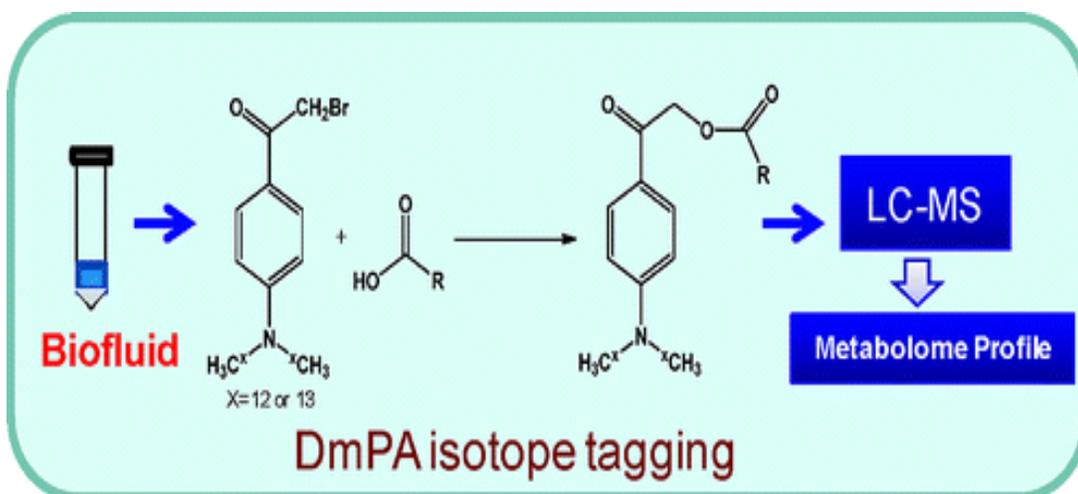


Figure 1.10 Schematic drawing of *p*-dimethylaminophenacyl bromide labeling.

1.7 Data Analysis

As is mentioned earlier in traditional biochemistry, univariate statistical analysis, such as *t*-test, was commonly used. However, metabolomics data matrix is quite complicated, univariate analysis is not sufficient to extract rich information from the data. Metabolomics data has many variables, i.e., up to several thousand variables (metabolites itself or metabolite features), and more importantly, there are some correlation between these metabolites. Therefore, it is required to use more advanced statistical tools, such as multivariate statistical analysis methods including PCA and OPLS-DA. These statistical methods not only can reduce the multidimensional matrix to lower dimension plane, but also take into account the correlation between the metabolites.

1.7.1 PCA

Principal component analysis (PCA) is one of the most commonly used chemometrics tools. PCA is a tool to reduce the multidimensional data matrix into lower dimensional plane.¹⁰⁸ For example, by projection, the three dimensional data can be transformed into a new coordinate system. The greatest variance of the data will lie in the first coordinate, also called first principal component.⁴² The first component captures the most variance in the table. (See Figure 1.11) PCA is an unsupervised visualization method. The class information about the groups is not used and we do not label the samples in advance.

To understand PCA better, it is necessary to have a brief introduction of the mathematical principles behind PCA. A data matrix X with m observations or samples, and n variables or metabolites, can be decomposed using the following equation:¹⁰⁹

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E} \quad (\text{Eq. 4})$$

Graphically, the equation can be expressed as the following:¹⁰⁹

Here the \mathbf{t}_i vectors are called scores, which contains the information on how the samples related to each other. The \mathbf{p}_i vectors are called loadings, which contains the information on how the variables related to each other.¹⁰⁹ Usually the decomposition is truncated at some point k leaving a small amount of variance in a residual matrix E .

Mathematically, PCA actually involves an eigenvector decomposition of the covariance matrix of the variables. Covariance matrix is expressed as the following equation:

$$\mathbf{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{m - 1} \quad (\text{Eq. 5})$$

$$\mathbf{cov}(\mathbf{X})\mathbf{p}_i = \lambda_i \mathbf{p}_i \quad (\text{Eq. 6})$$

In the PCA decomposition, the \mathbf{p}_i vectors are eigenvectors of the covariance matrix. The λ_i is the eigenvalue, which is a measure of the amount of variance.

PCA can generate scores plots and loadings plots. Each dot in scores plot represents an individual sample, and the relation between samples, such as class separation, can be visualized. A loadings plot shows how much each of the variables (metabolites) contributed to the different principal components. Variables at the extreme corners contribute most to the scores plot separation.

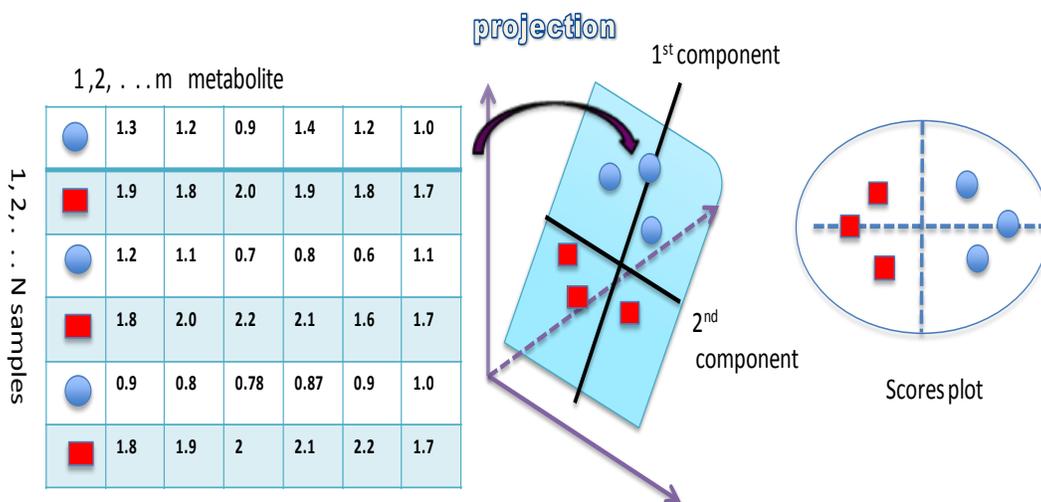


Figure 1.11 Schematic drawing of PCA decomposition into scores plot.

1.7.2 PLS-DA

PCA is a good tool to give an overview of the data, allowing us to find the general trends and outliers. However, PCA may not be good to predict the unknown samples and it is not easy to find most significant variables related to the class separation. PLS-DA stands for partial least square discriminant analysis. It is a supervised classification method. We label the samples in advance and tell the sample belonging to which classes. We can label one class 0 and the other class 1. Then we can construct an additional dummy matrix Y matrix using the class membership information. PLS-DA uses linear regression method to find the correlation between X matrix (data table) and Y matrix, which is a maximum of covariance between X matrix and Y matrix. (See Figure 1.12) In contrast, PCA is to find the maximum of variance of the X matrix. PLS-DA can be expressed in a simplified equation⁴²

Model of X : $X = TP^T + E$ (Eq. 7)

Model of Y : $Y = TC^T + F$ (Eq. 8)

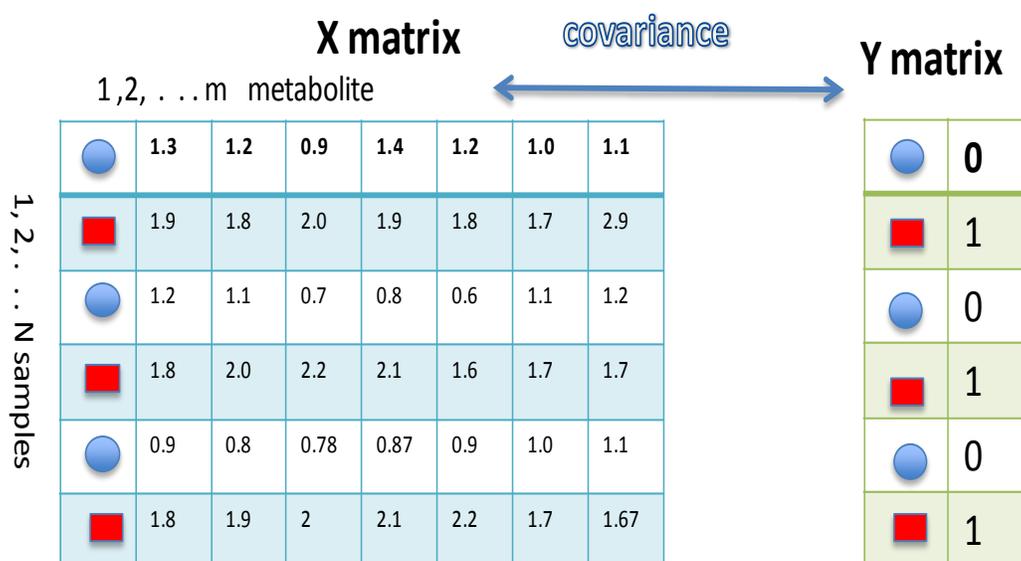


Figure 1.12 Schematic drawing of PLS-DA.

1.7.3 OPLS-DA

PLS-DA model may be affected by the unrelated variation in the data. Therefore, OPLS-DA is a modified version of PLS-DA and is trying to deal with

unwanted variation. OPLS-DA stands for orthogonal projections to latent structures-discriminant analysis (OPLS-DA). This model is a unique method in the commercial software SIMCA-P+, which is most widely used in metabolomics. OPLS-DA is more often used compared with PLS-DA. In OPLS-DA, the data X matrix will be divided into two parts, one is linearly related to Y, and the other is unrelated (orthogonal) to Y. The OPLS-DA can be expressed as following:⁴²

$$\text{Model of X: } \mathbf{X} = T_p P_p^T + T_o P_o^T + \mathbf{E} \quad (\text{Eq. 9})$$

$$\text{Model of Y: } \mathbf{Y} = T_p C_p^T + \mathbf{F} \quad (\text{Eq. 10})$$

$T_p P_p$ is the predictive variations, and $T_o P_o$ is the orthogonal variations. Only $T_p P_p$ is used to model the relationship to the Y matrix. Compared with PLS-DA, the OPLS-DA greatly facilitates the data interpretation, but it does not necessarily improve the predictability of the model.

VIP (variable importance on the projection) is a parameter which summarizes the importance of variables (metabolites) both for X and Y matrix.¹⁰⁹ PLS-DA or OPLS-DA can generate a VIP list and the importance of metabolites will be ranked according to their VIP value. Generally speaking, the metabolite with VIP larger than 1 would be considered as important and need close examination.

1.7.4 Model validation

Either PLS-DA or OPLS-DA is susceptible to overfitting, an artifact of the model fitting. Due to random chance, the overfitted model may have a clear class separation between samples which have no difference in fact. In many metabolomics studies, the number of samples is much less than the number of variables. This is one of main sources causing overfitting. Therefore, both PLS-DA and OPLS-DA need to be validated. The SIMCA-P+ software has two built-in functions to validate the quality of PLS-DA or OPLS-DA. One is Q^2 and the other is permutation test. Q^2 is defined as the ability of the model to predict the unknown samples (goodness of prediction), while R^2 is the goodness of fitting. The original data are divided into 7 parts and each 1/7th in turn is removed. A cross validation model is built on the 6/7 th data the 1/7th data are predicted from the new model.¹⁰⁹ The predicted data are then compared with the original data and the sum of squared errors calculated for the whole dataset. This is then called the Predicted Residual Sum of Squares (PRESS).

$Q^2 = 1 - \text{PRESS}/SS$ or more mathematically expressed as¹¹⁰

$$Q^2 = 1 - \frac{\sum_i^n (y_i - \hat{y})^2}{\sum_i^n (y_i - \bar{y})^2} \quad (\text{Eq. 11})$$

Here SS is the sum of squares of the elements of \mathbf{Y} after pre-processing.

As a general guide, if the $Q^2 > 0.5$ for PLS-DA or OPLS-DA, then the model is considered as good . If the $Q^2 > 0.9$, then the model is considered as excellent.¹⁰⁹

Permutation test should be also performed to check model validity. In the permutation test, the class label, such as control and disease class, will be

permutated, and the class information will be randomly assigned to each individual sample.¹¹¹ Then the model will be rebuilt using this random “wrong” class information. The assumption here is that if the model is valid, then the R^2 and Q^2 values after permutation test should be significantly different from the R^2 and Q^2 values using the original model, which means all the Q^2 in the permuted data set are lower than the Q^2 in the original data set.¹¹⁰ The model is not valid or overfitting if the R^2 and Q^2 values after permutation test is not different from the R^2 and Q^2 using the original model.¹¹²

1.7.5 Metabolic Pathway analysis

Metabolic pathways are the networks of metabolites through a series of biochemical reactions using many enzymes. Metabolic pathway analysis is important because it can be helpful to interpretate the biological meaning of the results. Each metabolite in the network can be considered as node, and one reaction can be considered as edges. There are several metabolic pathway database, such as KEGG;¹¹³ Nevertheless, it is not straightforward to pinpoint the most significant pathway. MetPA¹¹⁴ is a web-based pathway analysis software with visualization function and it is based on pathway topological analysis by measuring the centrality of the compounds in a network.¹¹⁵ By measuring the centrality, the position of a metabolite or node in the network will tell us the relative importance of the metabolite in a pathway. Finally the

significant metabolic pathway will be picked by calculating the overall impact of the pathway.

1.8 Overview of the thesis

To continue improving LC-MS metabolomics techniques, we aim to develop more comprehensive and reproducible metabolomic methods and workflows based on isotope labeling LC-MS for human and animal biofluids. We also want to apply our optimized metabolomics techniques to study animal models of diseases, such as Alzheimer's disease and asthma, in search of potential metabolite biomarkers in disease diagnosis. Specifically, there are four research projects in the thesis, which correspond to the following chapters (Chapters 2 to 5).

Chapter 2 describes development of an off-line two dimensional LC separation method for comprehensive urine metabolomics profiling. In this approach, an ion pairing reversed phase LC was used as a first dimensional separation, then seven fractions were differentially $^{12}\text{C}/\text{-}^{13}\text{C}$ dansyl labeled, and second dimensional reversed phase LC-MS was used.

Chapter 3 combines liquid-liquid extraction with $^{12}\text{C}/\text{-}^{13}\text{C}$ dimethylaminophenacyl labeling LC-MS for the improved metabolomics profiling of organic acids in urine.

Chapter 4 describes development of a relative quantitative metabolomics workflow based on $^{12}\text{C}/^{13}\text{C}$ differential dansyl labeling LC-FTICR-MS for mouse urine metabolomic profiling and its application to urine metabolomics in a mouse model of Alzheimer's disease.

Chapter 5 describes development of a more sensitive metabolomics method for bronchoalveolar lavage fluid of rats and its application to investigate a rat model of airway allergic inflammation (experimental asthma).

Chapter 6 is a summary of what we have done, and gives some concluding remarks and also briefly discusses what we need to do in future studies.

References

1. Salway, J. G., Metabolism at a glance. *Wiley, John & Sons, Incorporated* **2004**, 3rd ed.
2. <http://www.nature.com/scitable/topicpage/cell-metabolism-14026182>.
3. DeBerardinis, R. J.; Thompson, C. B., Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **2012**, 148, (6), 1132-1144.
4. Nicholson, J. K.; Wilson, I. D., Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discovery* **2003**, 2, (8), 668-676.
5. Fiehn, O., Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* **2001**, 2, (3), 155-168.
6. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the human metabolome database. *Nucleic Acids Res.* **2007**, 35, D521-D526.
7. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, 41, (D1), D801-D807.
8. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.;

Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J.; Forsythe, I., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37*, D603-D610.

9. Fiehn, O., Metabolomics - the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, (1-2), 155-171.

10. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discovery* **2002**, *1*, (2), 153-161.

11. Weckwerth, W., Metabolomics in systems biology. *Annu. Rev. Plant Biol.* **2003**, *54*, 669-689.

12. Kaddurah-Daouk, R.; Krishnan, K. R. R., Metabolomics: A global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* **2009**, *34*, (1), 173-186.

13. Patti, G. J.; Yanes, O.; Siuzdak, G., Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, (4), 263-269.

14. Holmes, E.; Wilson, I. D.; Nicholson, J. K., Metabolic phenotyping in health and disease. *Cell* **2008**, *134*, (5), 714-717.

15. Vinayavekhin, N.; Homan, E. A.; Saghatelian, A., Exploring Disease through Metabolomics. *ACS Chem. Biol.* **2010**, *5*, (1), 91-103.

16. Nordstrom, A.; Lewensohn, R., Metabolomics: moving to the clinic. *J. Neuroimmune Pharmacol.* **2010**, *5*, (1), 4-17.

17. Abu Aboud, O.; Weiss, R. H., New opportunities from the Cancer Metabolome. *Clin. Chem.* **2013**, *59*, (1), 138-146.

18. Bodi, V.; Marrachelli, V. G.; Husser, O.; Chorro, F. J.; Vina, J. R.; Monleon, D., Metabolomics in the diagnosis of acute myocardial ischemia. *J. Cardiovasc. Transl. Res.* **2013**, *6*, (5), 808-815.

19. Zhang, A. H.; Sun, H.; Wang, X. J., Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Anal. Bioanal. Chem.* **2012**, 404, (4), 1239-1245.
20. Mishur, R. J.; Rea, S. L., Applications of mass spectrometry to metabolomics and metabonomics: Detection of biomarkers of aging and of age-related diseases. *Mass Spectrom. Rev.* **2012**, 31, (1), 70-95.
21. Yin, P. Y.; Xu, G. W., Metabolomics for tumor marker discovery and identification based on chromatography-mass spectrometry. *Expert Rev. Mol. Diagn.* **2013**, 13, (4), 339-348.
22. Weiss, R. H.; Kim, K. M., Metabolomics in the study of kidney diseases. *Nat. Rev. Nephrol.* **2012**, 8, (1), 22-33.
23. Wei, R., Metabolomics and Its Practical Value in Pharmaceutical Industry. *Curr. Drug Metab.* **2011**, 12, (4), 345-358.
24. Xu, E. Y.; Schaefer, W. H.; Xu, Q. W., Metabolomics in pharmaceutical research and development: Metabolites, mechanisms and pathways. *Curr. Opin. Drug Discovery Dev.* **2009**, 12, (1), 40-52.
25. Wilcoxon, K. M.; Uehara, T.; Myint, K. T.; Sato, Y.; Oda, Y., Practical metabolomics in drug discovery. *Expert Opin. Drug Discov.* **2010**, 5, (3), 249-263.
26. Bouhifd, M.; Hartung, T.; Hogberg, H. T.; Kleensang, A.; Zhao, L., Review: Toxicometabolomics. *J. Appl. Toxicol.* **2013**, 33, (12), 1365-1383.
27. Griffin, J. L.; Bollard, M. E., Metabonomics: Its potential as a tool in toxicology for safety assessment and data integration. *Curr. Drug Metab.* **2004**, 5, (5), 389-398.
28. Robertson, D. G., Metabonomics in toxicology: A review. *Toxicol. Sci.* **2005**, 85, (2), 809-822.

29. Satheeshkumar, N.; Nisha, N.; Sonali, N.; Nirmal, J.; Jain, G. K.; Spandana, V., Analytical profiling of bioactive constituents from herbal products, using metabolomics - a review. *Nat. Prod. Commun.* **2012**, 7, (8), 1111-1115.
30. Sumner, L. W.; Mendes, P.; Dixon, R. A., Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **2003**, 62, (6), 817-836.
31. Kueger, S.; Steinhauser, D.; Willmitzer, L.; Giavalisco, P., High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. *Plant J.* **2012**, 70, (1), 39-50.
32. Kadplan, L. A. P., A.J. Kazmierczak, S.C., *Clinical chemistry: theory, analysis, correlation Mosby Inc.* **2003**.
33. Wishart, D. S., Metabolomics: The principles and potential applications to transplantation. *Am. J. Transplant.* **2005**, 5, (12), 2814-2820.
34. Snyder, M.; Li, X. Y., Metabolomics as a robust tool in systems biology and personalized medicine: an open letter to the metabolomics community. *Metabolomics* **2013**, 9, (3), 532-534.
35. German, J. B.; Hammock, B. D.; Watkins, S. M., Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **2005**, 1, (1), 3-9.
36. Denoroy, L.; Zimmer, L.; Renaud, B.; Parrot, S., Ultra high performance liquid chromatography as a tool for the discovery and the analysis of biomarkers of diseases: A review. *J. Chromatogr. B* **2013**, 927, 37-53.
37. Almstetter, M. F.; Oefner, P. J.; Dettmer, K., Comprehensive two-dimensional gas chromatography in metabolomics. *Anal. Bioanal. Chem.* **2012**, 402, (6), 1993-2013.
38. Xian, F.; Hendrickson, C. L.; Marshall, A. G., High resolution mass spectrometry. *Anal. Chem.* **2012**, 84, (2), 708-719.

39. Scigelova, M.; Hornshaw, M.; Giannakopoulos, A.; Makarov, A., Fourier transform mass spectrometry. *Mol. Cell. Proteomics* **2011**, 10, (7), M111.009431.
40. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P., Liquid chromatography-mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *J. Chromatogr. A* **2013**, 1292, 51-65.
41. Madsen, R.; Lundstedt, T.; Trygg, J., Chemometrics in metabolomics-A review in human disease diagnosis. *Anal. Chim. Acta* **2009**, 659, (1-2), 23-33.
42. Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J. Proteome Res.* **2007**, 6, (2), 469-479.
43. Xia, J. G.; Psychogios, N.; Young, N.; Wishart, D. S., MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, 37, W652-W660.
44. Xia, J. G.; Wishart, D. S., Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* **2011**, 6, (6), 743-760.
45. Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z. J.; Patti, G. J.; Siuzdak, G., An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat. Biotechnol.* **2012**, 30, (9), 826-828.
46. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, 45, (7), 703-714.

47. Lv, H. T., Mass Spectrometry based metabolomics: towards understanding of gene functions with a diversity of biological contexts. *Mass Spectrom. Rev.* **2013**, 32, (2), 118-128.
48. Griffiths, W. J.; Koal, T.; Wang, Y. Q.; Kohl, M.; Enot, D. P.; Deigner, H. P., Targeted metabolomics for biomarker discovery. *Angew. Chem. Int. Ed.* **2010**, 49, (32), 5426-5445.
49. Ganti, S.; Weiss, R. H., Urine metabolomics for kidney cancer detection and biomarker discovery. *Urol. Oncol. Semin. Ori.* **2011**, 29, (5), 551-557.
50. Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L., Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **2011**, 40, (1), 387-426.
51. Urban, M.; Enot, D. P.; Dallmann, G.; Korner, L.; Forcher, V.; Enoh, P.; Koal, T.; Keller, M.; Deigner, H. P., Complexity and pitfalls of mass spectrometry-based targeted metabolomics in brain research. *Anal. Biochem.* **2010**, 406, (2), 124-131.
52. Fernandez-Peralbo, M. A.; de Castro, M. D. L., Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. *Trac-Trends Anal. Chem* **2012**, 41, 75-85.
53. Ferreiro-Vera, C.; Ribeiro, J. P. N.; Mata-Granados, J. M.; Priego-Capote, F.; de Castro, M. D. L., Standard operation protocol for analysis of lipid hydroperoxides in human serum using a fully automated method based on solid-phase extraction and liquid chromatography-mass spectrometry in selected reaction monitoring. *J. Chromatogr. A* **2011**, 1218, (38), 6720-6726.
54. Theodoridis, G.; Gika, H. G.; Wilson, I. D., Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. *Mass Spectrom. Rev.* **2011**, 30, (5), 884-906.

55. Kitteringham, N. R.; Jenkins, R. E.; Lane, C. S.; Elliott, V. L.; Park, B. K., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B* **2009**, 877, (13), 1229-1239.
56. Guo, B.; Chen, B.; Liu, A. M.; Zhu, W. T.; Yao, S. Z., Liquid chromatography-mass spectrometric multiple reaction monitoring-based strategies for expanding targeted profiling towards quantitative metabolomics. *Curr. Drug Metab.* **2012**, 13, (9), 1226-1243.
57. Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D., Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Anal. Chim. Acta* **2012**, 711, 7-16.
58. Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B., Targeted metabolomics. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2012**, Chapter 30, Unit 30.2.1-24.
59. Brunelli, L.; Ristagno, G.; Bagnati, R.; Fumagalli, F.; Latini, R.; Fanelli, R.; Pastorelli, R., A combination of untargeted and targeted metabolomics approaches unveils changes in the kynurenine pathway following cardiopulmonary resuscitation. *Metabolomics* **2013**, 9, (4), 839-852.
60. Wishart, D. S., Quantitative metabolomics using NMR. *Trac-Trends Anal. Chem.* **2008**, 27, (3), 228-237.
61. Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J. G.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S., The human serum metabolome. *PLoS ONE* **2011**, 6, (2), 23.
62. Roux, A.; Lison, D.; Junot, C.; Heilier, J. F., Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clin. Biochem.* **2011**, 44, (1), 119-135.

63. Issaq, H. J.; Van, Q. N.; Waybright, T. J.; Muschik, G. M.; Veenstra, T. D., Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **2009**, 32, (13), 2183-2199.
64. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, 26, (1), 51-78.
65. Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y., Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J. Chromatogr. B* **2008**, 871, (2), 202-211.
66. Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T., Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **2011**, 7, (3), 307-328.
67. Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **2011**, 6, (10), 1483-1499.
68. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2005**, 56, (410), 219-243.
69. Zhang, Q.; Wang, G. J.; Du, Y.; Zhu, L. L.; Jiye, A., GC/MS analysis of the rat urine for metabonomic research. *J. Chromatogr. B* **2007**, 854, (1-2), 20-25.
70. Lei, Z. T.; Huhman, D. V.; Sumner, L. W., Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **2011**, 286, (29), 25435-25442.
71. Doerfler, H.; Lyon, D.; Nagele, T.; Sun, X. L.; Fragner, L.; Hadacek, F.; Egelhofer, V.; Weckwerth, W., Granger causality in integrated GC-MS and LC-MS metabolomics data reveals the interface of primary and secondary metabolism. *Metabolomics* **2013**, 9, (3), 564-574.
72. Dole, M.; Mack, L. L.; Hines, R. L., Molecular beams of macroions. *J. Chem. Phys.* **1968**, 49, (5), 2240-2249.

73. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, 246, (4926), 64-71.
74. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B., Electrospray interface for liquid chromatography and mass spectrometry. *Anal. Chem.* **1985**, 57, (3), 675-679.
75. Kebarle, P.; Tang, L., From ions in solution to ions in the gas-phase-the mechanism of electrospray mass spectrometry. *Anal. Chem.* **1993**, 65, (22), A972-A986.
76. Kebarle, P.; Verkerk, U. H., Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **2009**, 28, (6), 898-917.
77. Cech, N. B.; Enke, C. G., Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **2001**, 20, (6), 362-387.
78. Wilm, M., Principles of electrospray ionization. *Mol. Cell. Proteomics* **2011**, 10, (7), 8.
79. Annesley, T. M., Ion suppression in mass spectrometry. *Clin. Chem.* **2003**, 49, (7), 1041-1044.
80. Taylor, P. J., Matrix effects: The Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clin. Biochem.* **2005**, 38, (4), 328-334.
81. Jessome, L. L. V., D.A., Ion suppression: a major concern in mass spectrometry. *LCGC North Am.* **2006**, 24, (5), 498-510.
82. King, R.; Bonfiglio, R.; Fernandez-Metzler, C.; Miller-Stein, C.; Olah, T., Mechanistic investigation of ionization suppression in electrospray ionization. *J. Am. Soc. Mass Spectrom.* **2000**, 11, (11), 942-950.

83. Marshall, A. G.; Guan, S. H., Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, 10, (14), 1819-1823.
84. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K. Q.; Smith, R. D., The ion funnel: theory, implementations, and applications. *Mass Spectrom. Rev.* **2010**, 29, (2), 294-312.
85. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* **1998**, 17, (1), 1-35.
86. Amster, I. J., Fourier transform mass spectrometry. *J. Mass Spectrom.* **1996**, 31, (12), 1325-1337.
87. Schmid, D. G.; Grosche, P.; Bandel, H.; Jung, G., FTICR-Mass spectrometry for high-resolution analysis in combinatorial chemistry. *Biotechnol. Bioeng.* **2000**, 71, (2), 149-161.
88. Taylor, P. K.; Amster, I. J., Space charge effects on mass accuracy for multiply charged ions in ESI-FTICR. *Int. J. Mass Spectrom.* **2003**, 222, (1-3), 351-361.
89. Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G., Initial experimental characterization of a new ultra-high resolution FTICR cell with dynamic harmonization. *J. Am. Soc. Mass Spectrom.* **2011**, 22, (7), 1125-1133.
90. Qi, Y. L.; Li, H. L.; Wills, R. H.; Perez-Hurtado, P.; Yu, X.; Kilgour, D. P. A.; Barrow, M. P.; Lin, C.; O'Connor, P. B., Absorption-mode Fourier transform mass spectrometry: The effects of apodization and phasing on modified protein spectra. *J. Am. Soc. Mass Spectrom.* **2013**, 24, (6), 828-834.
91. Qi, Y. L.; Barrow, M. P.; Li, H. L.; Meier, J. E.; Van Orden, S. L.; Thompson, C. J.; O'Connor, P. B., Absorption-Mode: The next generation of Fourier transform mass spectra. *Anal. Chem.* **2012**, 84, (6), 2923-2929.

92. Brubaker, W. M.; Tuul, J., Performance study of quadrupole mass filter. *Rev. Sci. Instrum.* **1964**, 35, (8), 1007-1010.
93. Miller, P. E.; Denton, M. B., The quadrupole mass filter - basic operating concepts. *J. Chem. Educ.* **1986**, 63, (7), 617-622.
94. Yost, R. A.; Enke, C. G., Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **1979**, 51, (12), 1251-1270A.
95. Kondrat, R. W.; McClusky, G. A.; Cooks, R. G., Multiple reaction monitoring mass spectrometry for direct analysis of complex mixtures. *Anal. Chem.* **1978**, 50, (14), 2017-2021.
96. Dziadosz, M.; Weller, J. P.; Klintschar, M.; Teske, J., Scheduled multiple reaction monitoring algorithm as a way to analyse new designer drugs combined with synthetic cannabinoids in human serum with liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B* **2013**, 929, 84-89.
97. Liu, Y.; Uboh, C. E.; Soma, L. R.; Li, X. Q.; Guan, F. Y.; You, Y. W.; Chen, J. W., Efficient Use of Retention Time for the Analysis of 302 Drugs in Equine Plasma by Liquid Chromatography-MS/MS with Scheduled Multiple Reaction Monitoring and Instant Library Searching for Doping Control. *Anal. Chem.* **2011**, 83, (17), 6834-6841.
98. Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R., Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets. *Metabolomics* **2012**, 8, (1), S29-S36.
99. Liu, T.; Qian, W. J.; Gritsenko, M. A.; Xiao, W. Z.; Moldawer, L. L.; Kaushal, A.; Monroe, M. E.; Varnum, S. M.; Moore, R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G.; Smith, R. D.; Inflammation host response, I., high dynamic range characterization of the trauma patient plasma proteome. *Mol. Cell. Proteomics* **2006**, 5, (10), 1899-1913.

100. Teufelhofer, O.; Parzefall, W.; Elbling, L.; Kainzbauer, E.; Grasl-Kraupp, B.; Zielinski, C.; Schulte-Hermann, R.; Gerner, C., Divide and conquer: Rat liver tissue proteomics based on the analysis of purified constituents. *Electrophoresis* **2006**, *27*, (20), 4112-4120.
101. Guo, K.; Li, L., Differential C-12/C-13-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal. Chem.* **2009**, *81*, (10), 3919-3932.
102. Guo, K.; Bamforth, F.; Li, L. A., Qualitative metabolome analysis of human cerebrospinal fluid by C-13-/C-12-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, (2), 339-347.
103. Bruheim, P.; Kvitvang, H. F. N.; Villas-Boas, S. G., Stable isotope coded derivatizing reagents as internal standards in metabolite profiling. *J. Chromatogr. A* **2013**, *1296*, 196-203.
104. Walker, S. H.; Papas, B. N.; Comins, D. L.; Muddiman, D. C., Interplay of permanent charge and hydrophobicity in the electrospray ionization of glycans. *Anal. Chem.* **2010**, *82*, (15), 6636-6642.
105. Metz, T. O.; Zhang, Q. B.; Page, J. S.; Shen, Y. F.; Callister, S. J.; Jacobs, J. M.; Smith, R. D., Future of liquid chromatography-mass spectrometry in metabolic profiling and metabolomic studies for biomarker discovery. *Biomark. Med.* **2007**, *1*, (1), 159-185.
106. Ciccimaro, E.; Blair, I. A., Stable-isotope dilution LC-MS for quantitative biomarker analysis. *Bioanalysis* **2010**, *2*, (2), 311-341.
107. Guo, K.; Li, L., High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry. *Anal. Chem.* **2010**, *82*, (21), 8789-8793.

108. Liland, K. H., Multivariate methods in metabolomics - from pre-processing to dimension reduction and statistical analysis. *Trac-Trends Anal. Chem.* **2011**, 30, (6), 827-841.
109. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Trygg, J.; Wikstrom, C.; Wold, S., multi- and megavariate data analysis part I basic principles and applications. *Umetrics AB Sweden* **2006**.
110. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A., Assessment of PLSDA cross validation. *Metabolomics* **2008**, 4, (1), 81-89.
111. Rubingh, C. M.; Bijlsma, S.; Derks, E.; Bobeldijk, I.; Verheij, E. R.; Kochhar, S.; Smilde, A. K., Assessing the performance of statistical validation tools for megavariate metabolomics data. *Metabolomics* **2006**, 2, (2), 53-61.
112. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **2010**, 5, (6), 1005-1018.
113. Kanehisa, M.; Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, 28, (1), 27-30.
114. Xia, J. G.; Wishart, D. S., MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics* **2010**, 26, (18), 2342-2344.
115. Aittokallio, T.; Schwikowski, B., Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.* **2006**, 7, (3), 243-255.

Chapter 2

Ion-Pairing Reversed-Phase Liquid Chromatography Fractionation in Combination with Isotope Labeling Reversed- Phase Liquid Chromatography Mass Spectrometry for Comprehensive Metabolome Profiling

2.1 Introduction

Liquid chromatography combined with mass spectrometry (LC-MS) has become an important technique for metabolome profiling due to its high sensitivity and specificity. However, one of the major analytical challenges in current LC-MS-based metabolome analysis is the difficulty to generate a very comprehensive profile of a metabolome sample due to great diversity in chemical and physical properties of metabolites and a wide range of concentrations of different metabolites. Due to matrix and ion suppression effects often encountered in MS detection, highly efficient metabolite separation becomes critical for improving the overall detectability of LC-MS in metabolome analysis.

Multi-dimensional separation of metabolites provides a means of reducing sample complexity for MS analysis¹⁻³. A variety of combinations of LC separation mechanisms including size-exclusion, ion exchange or reversed phase (RP) have been used in two-dimensional (2D) LC¹⁻³. In an ideal situation, all the

A form of this chapter was published as: Kevin Guo, Jun Peng, Ruokun Zou, Liang Li. "Ion-pairing reversed-phase liquid chromatography fractionation in combination with isotope labeling reversed-phase liquid chromatography-mass spectrometry for comprehensive metabolome profiling *J. Chromatog. A.* **2011**, 1218(23):3689-94. I worked with Ruokun Zou to carry out the experiments under the supervision of Kevin Guo.

analytes in a complex sample will be differentially retained by two orthogonal mechanisms in 2D-LC, and each mechanism is considered an independent separation dimension^{2,4}.

One constraint in selecting the separation modes for LC-MS is related to the compatibility of the LC separation with MS detection. RPLC is commonly used for online LC-MS due to its high resolving power and readiness to combine with MS. For metabolome analysis, ion exchange-reversed phase (IEC-RP)⁵ and size exclusion-reversed phase (SEC-RP)⁶ LC-MS have been reported. Hydrophilic interaction liquid chromatography (HILIC) has also been used for online LC-MS⁷⁻¹⁰. Although the resolving power of HILIC is not as good as RPLC, it provides the opportunity of combining with other separation mode, such as strong cation exchange (SCX), for 2D-LC-MS⁷. The combination of HILIC and RPLC for metabolome analysis has also been reported¹¹⁻¹⁴.

Because of relatively low resolving power of IEC, SEC or HILIC, the peak capacity of 2D-LC involving one of these modes would be lower than a RPLC - RPLC, assuming the RPLC - RPLC mode is capable of offering true orthogonality. However, only limited differences in selectivity of the RP stationary phase are achievable using different chemistries or mobile phases. For example, peptides can be separated and fractionated at low pH mobile phases in the 1st dimension RPLC and a certain degree of orthogonal separation can be done by the 2nd dimension RPLC at a relatively higher pH¹⁵. There are also reports of using diagonal chromatography¹⁶ to create some orthogonality between two RPLC dimensions for peptide or protein separation; modifications of peptides or proteins after the 1st dimension separation (e.g., oxidation of certain amino acid side chains) can result in changes in retention properties, thereby allowing altered separation in the 2nd dimension¹⁷.

In this work, we report a 2D-LC strategy in which the orthogonality of the separation is achieved by altering the hydrophobicity of the analytes prior to the 2nd dimension of RPLC separation. Thus, stationary phases of differing chemistries are not used for each dimension. The hydrophobicity of targeted analytes is altered by a chemical derivatization, dansylation, thus the derivatized metabolites can be analyzed by MS with much enhanced sensitivity, compared to the un-derivatized counterparts.

2.2 Experimental

2.2.1 Chemicals and reagents

All chemicals and reagents were purchased from Sigma-Aldrich Canada (Markham, ON, Canada) except those otherwise noted. The isotopic compound, ¹³C₂-dimethyl sulfate, used to synthesize the isotope tagged dansylation reagent (¹³C-dansyl chloride) was purchased from Cambridge Isotope Laboratories (Cambridge, MA, US). LC-MS grade water, methanol and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada). Urine samples were collected from a healthy individual and processed by adding 50% (v/v) LC-MS grade acetonitrile, then stored at -20 °C. The samples were centrifuged for 10 min at 13,800 g before injecting into LC.

2.2.2 Labeling reaction

The synthesis of ¹³C-dansyl chloride as the labeling reagent and dansylation labeling reaction have been described elsewhere ¹⁸. Briefly, two

aliquots were taken from an individual fraction collected from the 1st dimension separation for isotope labeling. Each aliquot was mixed with an equal volume of sodium carbonate/sodium bicarbonate buffer (0.5 mol/L, pH 9.4) in reaction vials. About a 4-fold molar ratio excess ¹²C-dansyl chloride solution in acetonitrile (20 mg/mL) (for light labeling) or ¹³C-dansyl chloride in acetonitrile (20 mg/mL) (for heavy labeling) was then added, and the reaction was allowed to proceed for 60 min at 60 °C with shaking at 150 rpm. After 60 min, methylamine (0.5 mol/L) was added to the reaction mixture to consume the excess dansyl chloride and quench the dansylation reaction. After an additional 30 min of 60 °C incubation, the ¹³C-labeled mixture was combined with its ¹²C-labeled counterpart for LC-FTICR-MS analysis. Note that the 4-fold molar excess of the reagent used was estimated based on the peak intensity of methylamine. The remaining methylamine in the labeled sample was the difference of the amount added and the amount consumed to react with the excess amount of the labeling reagent.

2.2.3 Ion-pairing RPLC

Figure 2.1 shows the workflow of the 2D-LC-FTICR-MS experiment. An Agilent 1100 series quaternary HPLC system (Agilent, Palo Alto, CA), and an Agilent Zorbax Rx-C18 column (9.4 × 250 mm, 5 μm particle size) were used in the 1st dimension of separation. Mobile phase A was 12 mM heptafluorobutyric acid (HFBA) and mobile phase B was 100 % acetonitrile. 254 nm was chosen as the UV detector wavelength. The gradient elution profile was as follows: t = 0, 0% B; t = 8 min, 10% B; t = 20 min, 30% B; t = 23 min, 95% B; t = 23.5 min, 0%

B; t = 50 min, 0% B. The flow rate was 5 mL/min, and sample injection volumes were 800 μ L.

2.2.4 LC-FTICR-MS

An Agilent 1100 series binary system (Agilent, Palo Alto, CA) and an Agilent reversed-phase Eclipse plus C₁₈ column (2.1 \times 100 mm, 1.8 μ m particle size, 95 Å pore size) were used for online LC-MS. LC solvent A was 0.1% (v/v) LC-MS grade formic acid in 5% (v/v) LC-MS grade ACN, and solvent B was 0.1% (v/v) LC-MS grade formic acid in LC-MS grade acetonitrile. The gradient elution profile was as follows: t = 0 min, 20% B; t = 3.0 min, 35% B; t = 16 min, 65% B; t = 18.6 min, 95% B; t = 21 min, 95% B; t = 21.3 min, 98% B; t = 23.0 min, 98% B; t = 24.0 min, 20% B. The flow rate was 150 μ L/min. The flow from RPLC was split 1:3 and a 50 μ L/min flow was loaded to the electrospray ionization (ESI) source of a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA, USA), while the rest of the flow was delivered to waste. All MS spectra were obtained in the positive ion mode.

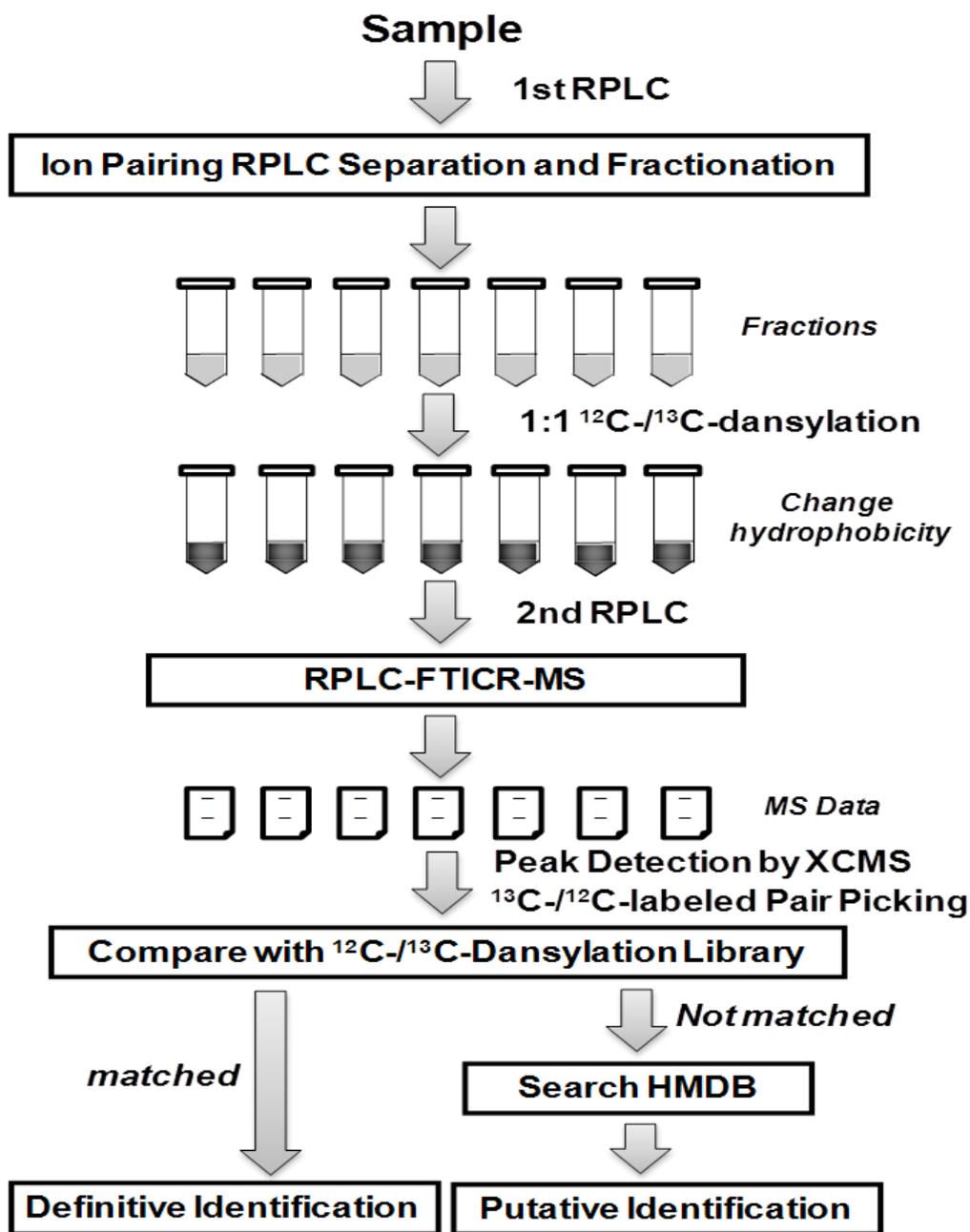


Figure 2.1 Schematic of the 2D-RPLC-FTICR-MS workflow for metabolome analysis.

2.3 Results and discussion

A large proportion of human metabolites in biofluids are highly polar compounds, and conventional RPLC often lacks the capability to adequately retain and separate ionic, polar metabolites. Recently HILIC has been used in bioanalytical applications, but the chromatographic efficiency is typically not as good as the RP separation¹⁰. An alternative approach to improve the retention of polar metabolites and maintain high separation efficiency is to use ion-pairing (IP) RPLC in which cationic species, such as amines or amino acids at low pH, are reported to be successfully separated using long chain perfluorocarboxylic acids, such as heptafluorobutyric acid (HFBA) as the volatile ion pairing reagent¹⁹⁻²⁷. In IP RPLC, the cationic species form an ion-pair with the negatively charged ion-pairing reagent in the mobile phase to become electrically neutral. Using HFBA the resulting hydrophobic character of the ion-pair leads to a greater affinity for the stationary reversed-phase and thus, results in greater retention of cationic polar compounds in the RPLC column. As a result, ion pairing RPLC can be used to generate selective retention of cationic polar metabolites and elute the anionic polar metabolites at or near void volume.

We observed that a long column equilibration is necessary to maintain reproducible chromatographic separation in IP RPLC. Systems containing 12 mM heptafluorobutyric acid required at least 25 column volumes to be well equilibrated. Figure 2.2 shows the superimposed RPLC-UV chromatograms of a human urine sample from four consecutive injections. All the major chromatographic peaks are perfectly superimposed in both retention times and

intensities. This high reproducibility of IP RPLC ensures the integrity of the fraction collections for multiple LC injections, if a large quantity of analytes are required for the 2nd dimension separation and analysis. This is also important for quantitative metabolome profiling where replicate runs are required ¹⁸.

It is well known that a high ionic strength of the mobile phase, and particularly when using many ion-pairing reagents, causes discharges and severe ion suppression in ESI ²⁸. It has been reported that, with commonly used volatile ion-pairing reagents, such as perfluoroheptanoic acid, ESI signal intensity decreased about 30 – 80%, compared with the use of formic acid ²¹. This problem of ion suppression on ESI from high concentrations of ion pairing reagents did not present a problem in this work, because heptafluorobutyric acid was eluted in the RPLC void volume of our 2nd dimension LC-MS analysis. However, it is still preferable to use volatile ion-pairing reagents, such as HFBA, as a non-volatile ion-pairing reagent may form crystals and pose a problem by contaminating the MS interface.

In contrast to most current 2D-LC separations that employ two columns with two different stationary phases with orthogonal retention mechanisms, in this work, we generate the orthogonality of RPLC-RPLC by altering the hydrophobicity of analytes through a chemical derivatization prior to the 2nd dimension separation on a RPLC column (see Figure 2.1). The major components of cationic metabolites, such as amines and amino acids, and polar phenolic compounds in a given sample are separated by IP RPLC in the 1st dimension separation. Their hydrophobicity can be significantly altered by a simple and

robust dansylation derivatization procedure. Dansylation is a well-studied derivatization chemistry that targets primary amines, secondary amines and phenolic hydroxyls^{18, 29-34}. Tertiary amines and alkyl hydroxyls cannot be dansylated. The large hydrophobic dansylation tag attached to the metabolites changes their hydrophobicity, and thus, their retention in the 2nd dimension RPLC. The degree to which the hydrophobicity changes is mainly dependent on the number of dansylation tag(s) added on, whilst the structures of the polar cationic compounds have less impact on their hydrophobicity and their resulting retention on the 2nd dimensional RPLC. In general, these dansylated compounds elute according to the number of incorporated tags with more tags resulting in longer retention. We note that one limitation of the current approach of using IP RPLC is that the anionic metabolites will not be efficiently separated in the first dimension separation. However, we expect that not many amine-containing metabolites are present in anionic forms in a urine sample. Some anionic phenolic compounds may not be separated in IP RPLC.

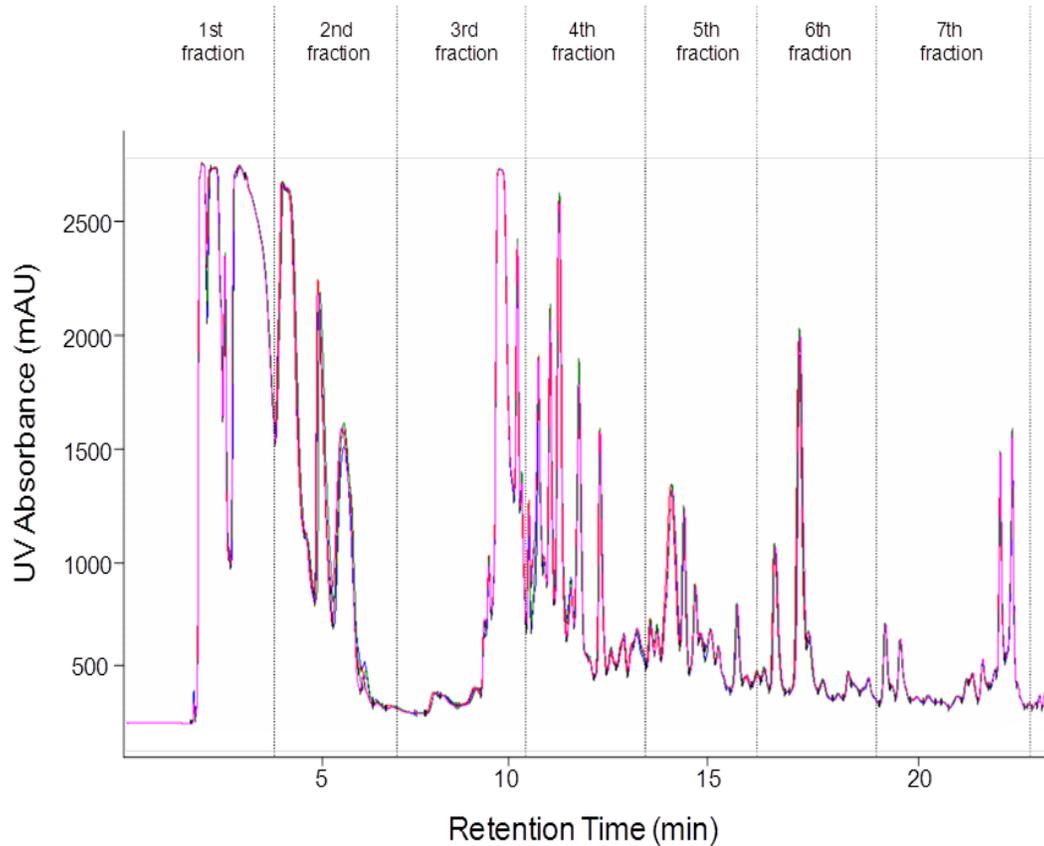


Figure 2.2 Superimposed ion pairing RPLC-UV chromatograms from four consecutive injections of a human urine sample. Seven fractions were collected for RPLC-FTICR-MS analysis.

Seven fractions were collected from the 1st dimensional IP RPLC as shown in Figure 2.2. Aliquots of individual IP RPLC fractions were differentially dansylated by ¹³C-dansyl chloride and ¹²C-dansyl chloride, and then combined in 1:1 equal molar ratio. The use of differential isotope labeling allows easy picking of the metabolite peaks in the mass spectra (see below). The top trace of Figure 2.3A shows a representative base-peak ion chromatogram obtained from a labeled fraction (Fraction #1 in Figure 2.2). For comparison, the lower trace of Figure 2.3A shows the ion chromatogram of the unlabeled fraction. A representative mass spectrum showing a pair of peaks is shown in Figure 2.3B. As Figure 2.3A shows, many chromatographic peaks were observed over the entire gradient elution window after labeling. Compared to the 1st dimension separation, a significant increase in separation space was achieved in the 2nd dimension, indicating that some orthogonality has been attained between the two RPLCs. In addition, much stronger signals were obtained from the dansylated fraction, which is consistent with the notion that dansylation generally improves the detection sensitivity by 1 to 3 orders of magnitude for the reasons already discussed in our previous report ¹⁸.

Based on accurate masses and retention time information of the ion pairs detected, both definitive and putative (or tentative) metabolite identification were carried out from RPLC-FTICR-MS analysis of the seven IP RPLC fractions. Using the differential isotope labeling method, metabolite peaks can be readily detected by examining the masses and intensities of the peak pairs. Error in mass difference was the mass error between the measured mass difference and

theoretical mass difference for $^{13}\text{C}/^{12}\text{C}$ -dansylation ion pairs. The theoretical mass difference for one dansylation tag is 2.00671. Two parts per million (2 ppm) for error in mass difference was used as a key criterion to assign the $^{13}\text{C}/^{12}\text{C}$ -dansylation ion pairs. High mass accuracy FTICR-MS measurement ensures confident assignment of the true $^{13}\text{C}/^{12}\text{C}$ -ion pairs. Note that non-reactive metabolites, background ions, instrumental and electronic noises will not have characteristic mass differences as $^{13}\text{C}/^{12}\text{C}$ -ion pairs. In addition, $^{13}\text{C}/^{12}\text{C}$ -dansylation ion pairs do not show isotopic chromatographic separation in RPLC, i.e. $^{13}\text{C}/^{12}\text{C}$ -ion pairs will be shown in the same spectrum.

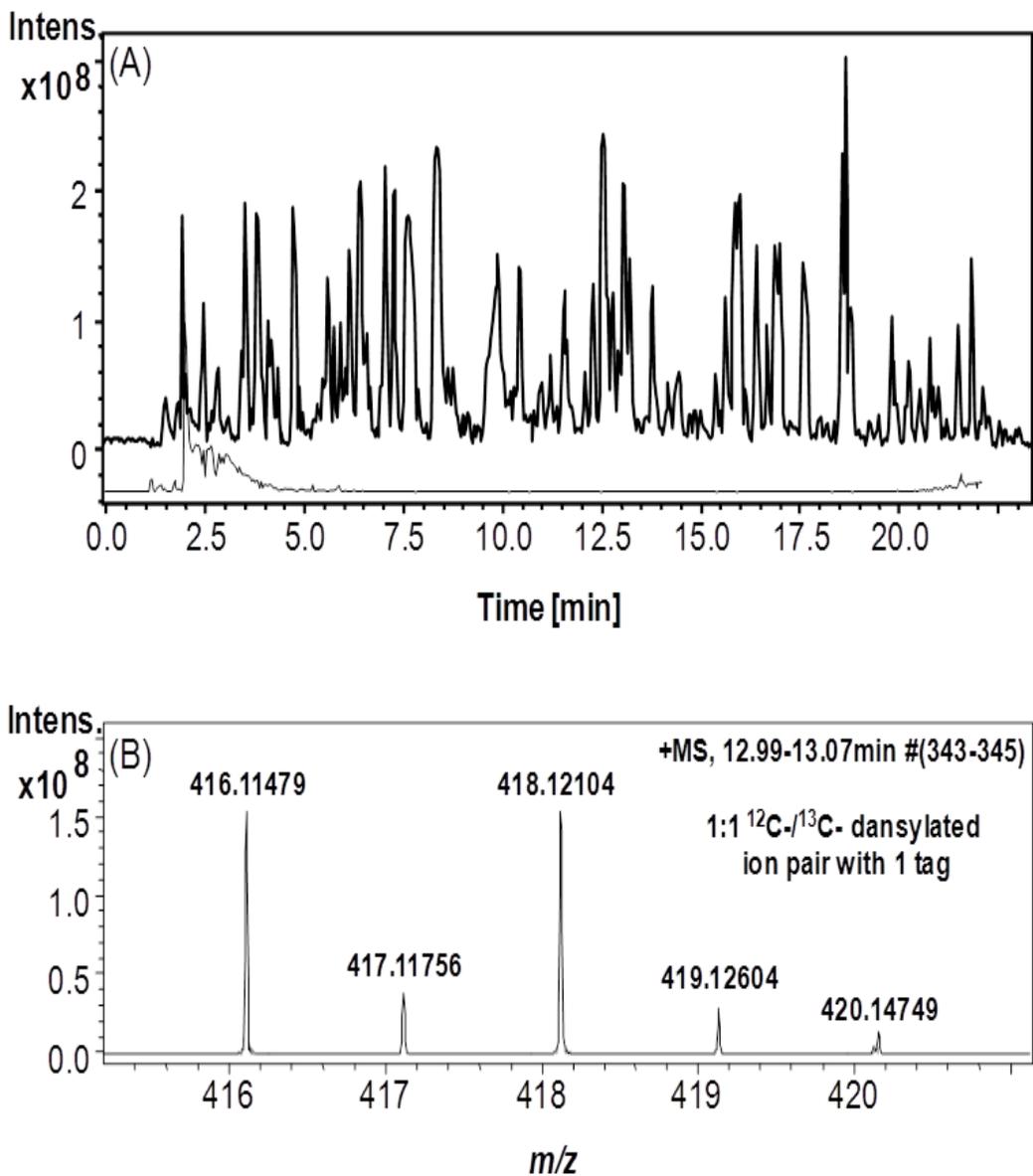


Figure 2.3 Base-peak ion chromatograms of the dansylation-labeled fraction #1 (top trace) and the unlabeled fraction #1 (bottom trace). (B) A representative mass spectrum showing a pair of ion peaks obtained at a particular retention time. Accurate mass match against the Human Metabolome Database indicates that the ion pairs were likely from homovanillic acid or other isomers.

In our workflow shown in Figure 2.1, a software program based on a peak analysis script, XCMS³⁵, was written to pick up the ¹³C-/¹²C-ion pairs. The program eliminated isotopic peaks, common adduct ions, multiply charged ions and multimers. Only the protonated ion pairs were exported to an Excel table for further analysis. If the ion pairs were detected as multiply charged ions, not the usual singly charged ions, only one form of the multiply charged ions was retained and exported to Excel.

Figure 2.4A shows the distribution of the number of ion pairs or metabolites detected from individual IP RPLC fractions. The number of ion pairs detected was 1042, 1111, 902, 877, 523, 366, and 211 from fractions 1 to 7, respectively. The number of common ion pairs found in adjacent fractions is also indicated in Figure 2.4A. As Figure 2.4A shows, there are many unique ion pairs detected in each fraction. In total, 3564 ion pairs were detected, which is almost three-fold of ion pairs detected from 1D-RPLC-FTICR-MS of the same urine sample after dansylation (i.e., 1218 ion pairs). Figure 2.4B shows the comparison of the number of ion pairs detected from off-line 2D-RPLC-FTICR-MS and 1D-RPLC-FTICR-MS. A major portion of the ion pairs found in the 1D experiment were detected in 2D-LC-MS. A small number of ion pairs uniquely found in the 1D experiment reflects the complexity of the urine metabolome. These unique metabolite were preferentially ionized in 1D-LC-MS while they might be suppressed during the 2D experiment; the extent of ion suppression is dependent on the chemical composition of the elute at a given MS detection window. Reduced complexity results in overall reduction of ion suppression. However,

this does not mean ion suppression is reduced evenly for all the ions. In fact, some ions may be more suppressed due to the increase in signals from other otherwise suppressed ions. It is also plausible that some of the missing metabolites in 2D-LC-MS were lost during the fraction collection and chemical labeling processes. Further optimization of these processes, such as the use of different solvents to rinse the collection vial more thoroughly during fraction transfer, is needed to avoid sample loss.

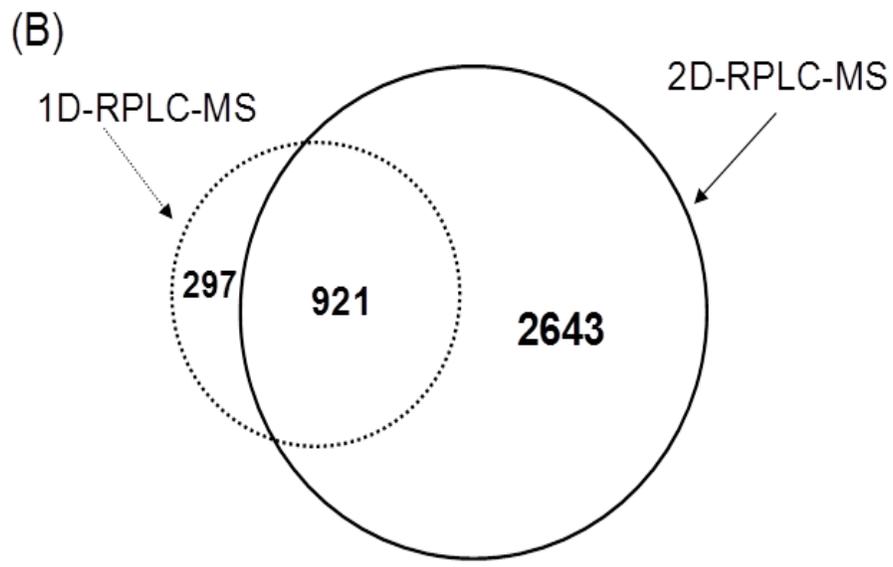
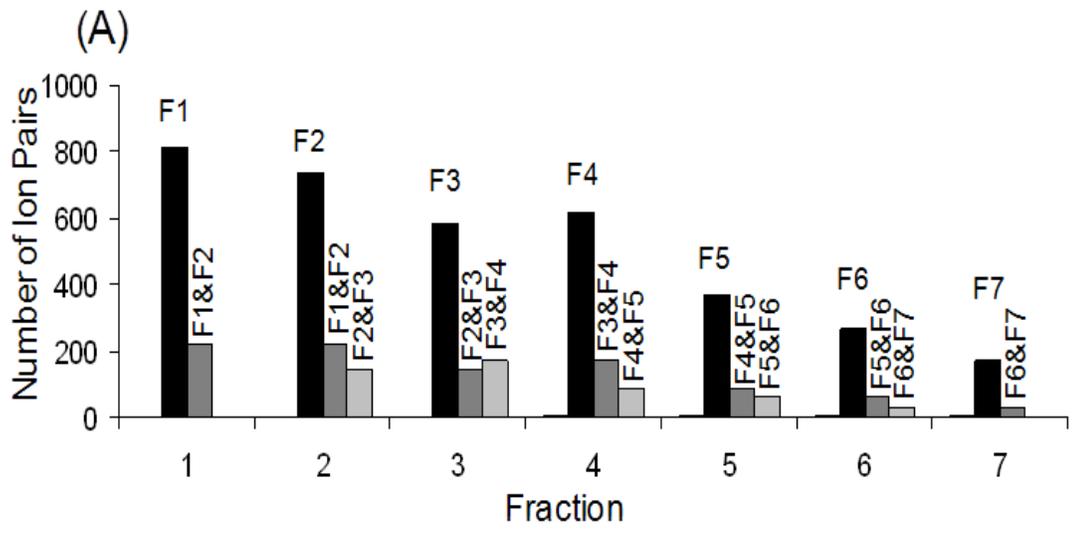


Figure 2.4 Distribution of the number of ion pairs detected by RPLC-FTICR-MS from the 7 fractions collected by IP RPLC of a human urine sample; the numbers of the unique and common ion pairs detected in neighboring fraction(s) are shown. (B) Comparison of the number of ion pairs detected by 2D-RPLC-FTICR-MS and 1D-RPLC-FTICR-MS.

Nevertheless, the 2D-LC-MS experiment clearly increases the metabolome coverage significantly, compared to 1D-LC-MS. We note that 2D-LC separation should benefit the detection of low abundance ions in FTICR-MS. Since the number of ions trapped in the FTICR cell is limited, reducing the number of co-eluting metabolites by using 2D-LC should facilitate the trapping of the low abundance ions from the mixture, thereby increasing the detectability of these ions.

To provide definitive identification of the metabolites in the urine sample, we have constructed a dansylation compound library consisting of 220 authentic standards of amine- or phenol-containing metabolites. Definitive identification was done by matching the accurate mass and retention time of the $^{13}\text{C}/^{12}\text{C}$ -dansylation ion pairs detected in the urine sample to those of the $^{13}\text{C}/^{12}\text{C}$ -dansylated standards. In total, 167 ion pairs were found to match the accurate mass and retention time of the authentic $^{13}\text{C}/^{12}\text{C}$ -dansylated standards. Table 2.1 lists the compound names of the 167 metabolites definitively identified from the human urine sample from this work, according to their order of elution in RPLC-MS. The number of definitively identified metabolites from the 2D-LC-MS experiment is about 2.5 times the number of metabolites definitively identified by 1D-LC-MS analysis of the same urine sample after $^{13}\text{C}/^{12}\text{C}$ -dansylation (i.e., 68 metabolites as listed in bold and underlined in Table 2.1).

All the ion pairs that were not matched to the $^{13}\text{C}/^{12}\text{C}$ -labeled authentic standard library were used to search against the Human Metabolome Database

(HMDB) for putative identification ³⁶. The HMDB database search was based on matching the accurate mass of the measured ions (¹²C-dansyl ion ions minus the dansyl group) against the accurate masses of 10364 human metabolites listed in the database. 53% of the ion pairs do not match with any metabolites in the HMDB database and 47% of pairs (1586 pairs) match with one or more putative metabolite (18% of pairs match to one putative metabolite). Many of the matches are likely coincidental, as an accurate mass can generate a number of possible molecular formula and, even with one formula, many possible chemical structures can be postulated. Clearly definitive identification of many ion pairs detected from the 2D-LC-MS experiment remains to be an analytical challenge. One strategy of averting this problem is to carry out relative quantification of the metabolomes of a number of comparative samples (e.g., diseased vs. controlled) first, followed by statistics analysis, such as principal component analysis (PCA), to discover one or a few putative biomarkers. After the discovery of the putative biomarkers, major efforts are then devoted to the identification of these metabolites using techniques such as tandem MS, NMR and synthesis of standard compounds. Definitive identification of the biomarkers is important for both studying the disease biology and obtaining regulatory approval for clinical applications.

2.4 Conclusions

We have developed a new off-line 2D-LC strategy combined with FTICR-MS for analyzing a large number of amine- and phenol-containing metabolites in

a complex biological sample. Cationic and polar species in a biological sample were successfully separated and fractionated by ion-pairing RPLC. Many of these species contained amine moiety which could be easily derivatized by dansyl chloride. The RP chromatographic retention behavior of the polar amines, amino acids, phenols, etc., were altered to an extent after dansylation derivatization such that they could be well retained and separated with high efficiency in the 2nd RPLC interfaced to FTICR-MS. We have shown that this 2D-RPLC-MS technique was able to detect 3564 metabolites in a human urine sample, compared to 1218 metabolites detected in 1D-RPLC-MS. While this strategy was demonstrated for the analysis of amine- and phenol-containing metabolites using dansylation derivatization, it should be generally applicable to other derivatization chemistries where derivatized metabolites have vastly different hydrophobicity than their un-derivatized counterparts (e.g., p-dimethylaminophenacyl labeling for carboxylic acid-containing metabolites³⁷).

Table 2.1. List of metabolites identified by 2D-RPLC-FTICR-MS from a human urine sample along with those identified by 1D-RPLC-FTICR-MS in bold and underlined.

Compound Name	Compound Name	Compound Name
Phospho-tyrosine	5-hydroxymethyluracil	<u>Lysine</u>
Hydroxylamine	5-aminopentanoic acid	3-hydroxybenzoic acid
Hydrochlorothiazide	2-aminoisobutyric acid	Vanillic acid
Phospho-serine	2-aminobutyric acid	Isoferulic acid
<u>Phosphoethanolamine</u>	<u>Sarcosine</u>	<u>4-hydroxybenzoic acid</u>
Glucosamine	Pyridoxine	Aniline
<u>Taurine</u>	<u>Proline</u>	<u>Histidine</u>
Saccharopine	<u>Methylamine</u>	Desaminotyrosine
Phospho-threonine	Aminocaproic acid	3-hydroxyanthranilic acid
<u>3-methylhistidine</u>	<u>Valine</u>	Benzylamine
<u>1-methylhistidine</u>	Salicylic acid	Tryptamine
Carnosine	<u>Methionine</u>	m-coumaric acid
Hypotaurine	<u>3-hydroxyl-picolinic acid</u>	trans-ferulic acid
<u>Arginine</u>	Gly-Trp	Ephedrine
Guanidine	3-nitrotyrosine	2-aminooctanoic acid
<u>Asparagine</u>	<u>Tryptophan</u>	Pyridoxamine
<u>Homoarginine</u>	Kynurenine	5-hydroxytryptophan
Histamine	Phenylephrine	<u>1,3-diaminopropane</u>
<u>Glutamine</u>	<u>2-phenylglycine</u>	<u>Tyrosinamide</u>
<u>Citrulline</u>	3-aminobenzoic acid	1,2-diaminopropane
1-methylhistamine	3-aminosalicylic acid	<u>1,4-diaminobutane</u>
3-methylhistamine	Ethylamine	o-tyrosine
3-sn-	Diaminopimelic acid	Thyroxine
phosphatidylethanolamine	Vanillylmandelic acid	<u>Cadaverine</u>
Aspartic acid amide	<u>Pipecolic acid</u>	<u>Tyrosine</u>
<u>Methylguanidine</u>	<u>Phenylalanine</u>	<u>Metoprolol</u>
<u>Homoserine</u>	Hydroxyphenylacetyl glycine	<u>Phenol</u>
Adenosine	Acetyl-tyrosine	4-nitrophenol
<u>Methionine sulfoxide</u>	Leu-Pro	<u>Cysteamine</u>
Homocitrulline	<u>3-hydroxymandelic acid</u>	16b-hydroxyestradiol
<u>Serine</u>	<u>Isoleucine</u>	4,9-dioxo-1,12-
<u>Glutamic acid</u>	L-cystathionine	dodecanediamine
<u>Aspartic acid</u>	<u>Leucine</u>	<u>Octopamine</u>
Diglycine	5-hydroxylysine	p-Cresol
<u>4-hydroxy-proline</u>	<u>Cystine</u>	Protocatechuic acid
<u>Amino adipic acid</u>	4-hydroxy-3-methoxyphenyllactic acid	Gentisic acid
<u>Threonine</u>	<u>Phenylethanolamine</u>	o-Cresol
<u>Folic acid</u>	<u>Hydroxyphenyllactic acid</u>	Serotonin
Dopamine	<u>5-HIAA</u>	Caffeic acid
Iminodiacetic acid		Metanephrine
<u>Diethanolamine</u>		Piperazine
<u>Ethanolamine</u>		

Epinephrine	<u>Dimethylamine</u>	Thyronine
<u>Glycine</u>	<u>2,4-Diaminobutyric acid</u>	Phenylephrine or Synephrine
Glycylproline	<u>Homocystine</u>	<u>Tyramine</u>
Beta-alanine	Salicylic acid	<u>Spermidine</u>
<u>Tyrosine methylester</u>	<u>Ornithine</u>	Xanthurenic acid
<u>Alanine</u>	Methyl-phenylalanine	Estradiol
<u>γ-aminobutyric acid</u>	<u>5-Methoxysalicylic acid</u>	3-isopropylphenol
Aminolevulinic acid	3- or 4-hydroxyphenylacetic acid	Pyrocatechol
Procaine	<u>Homovanillic acid</u>	Estrone
Pantothenic acid	5-Methoxytryptamine	Norepinephrine
<u>p-aminohippuric acid</u>	Syringic acid	<u>Thymol</u>
Salbutamol	<u>Homocarnosine</u>	Hydroquinone
Hypoxanthine	<u>3-Cresotinic acid</u>	Deoxyepinephrine
Isoguanine	<u>Carnosine</u>	Desipramine
3-aminoisobutyric acid	<u>Gentisic acid</u>	

References

1. Dixon, S. P.; Pitfield, I. D.; Perrett, D., Comprehensive multi-dimensional liquid chromatographic separation in biomedical and pharmaceutical analysis: a review. *Biomed. Chromatogr.* **2006**, 20, (6-7), 508-529.
2. Stoll, D. R.; Wang, X.; Carr, P. W., Comparison of the practical resolving power of one- and two-dimensional high-performance liquid chromatography analysis of metabolomic samples. *Anal. Chem.* **2007**, 80, (1), 268-278.
3. Stoll, D. R.; Li, X. P.; Wang, X. O.; Carr, P. W.; Porter, S. E. G.; Rutan, S. C., Fast, comprehensive two-dimensional liquid chromatography. *J. Chromatogr. A* **2007**, 1168, (1-2), 3-43.
4. Horvath, K.; Fairchild, J. N.; Guiochon, G., Generation and limitations of peak capacity in online two-dimensional liquid chromatography. *Anal. Chem.* **2009**, 81, (10), 3879-3888.
5. Edwards, J. L.; Edwards, R. L.; Reid, K. R.; Kennedy, R. T., Effect of decreasing column inner diameter and use of off-line two-dimensional chromatography on metabolite detection in complex mixtures. *J. Chromatogr. A* **2007**, 1172, (2), 127-134.
6. de Souza, L. M.; Cipriani, T. R.; Sant'Ana, C. F.; Iacomini, M.; Gorin, P. A. J.; Sasaki, G. L., Heart-cutting two-dimensional (size exclusion x reversed phase) liquid chromatography-mass spectrometry analysis of flavonol glycosides from leaves of *Maytenus ilicifolia*. *J. Chromatogr. A* **2009**, 1216, (1), 99-105.
7. Fairchild, J. N.; Horvath, K.; Gooding, J. R.; Campagna, S. R.; Guiochon, G., Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. *J. Chromatogr. A* **2010**, 1217, (52), 8161-6.
8. Bajad, S. U.; Lu, W.; Kimball, E. H.; Yuan, J.; Peterson, C.; Rabinowitz, J. D., Separation and quantitation of water soluble cellular metabolites by

hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **2006**, 1125, (1), 76-88.

9. Spagou, K.; Tsoukali, H.; Raikos, N.; Gika, H.; Wilson, I. D.; Theodoridis, G., Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *J. Sep. Sci.* **2010**, 33, (6-7), 716-727.

10. Guo, K.; Ji, C.; Li, L., Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. *Anal. Chem.* **2007**, 79, (22), 8631-8638.

11. Yang, Q.; Shi, X. Z.; Wang, Y. A.; Wang, W. Z.; He, H. B.; Lu, X.; Xu, G. W., Urinary metabonomic study of lung cancer by a fully automatic hyphenated hydrophilic interaction/RPLC-MS system. *J. Sep. Sci.* **2010**, 33, (10), 1495-1503.

12. Sato, Y.; Nakamura, T.; Aoshima, K.; Oda, Y., Quantitative and wide-ranging profiling of phospholipids in human plasma by two-dimensional liquid chromatography/mass spectrometry. *Anal. Chem.* **2010**, 82, (23), 9858-9864.

13. Wang, Y.; Lehmann, R.; Lu, X.; Zhao, X. J.; Xu, G. W., Novel, fully automatic hydrophilic interaction/reversed-phase column-switching high-performance liquid chromatographic system for the complementary analysis of polar and apolar compounds in complex samples. *J. Chromatogr. A* **2008**, 1204, (1), 28-34.

14. Nunez, O.; Nakanishi, K.; Tanaka, N., Preparation of monolithic silica columns for high-performance liquid chromatography. *J. Chromatogr. A* **2008**, 1191, (1-2), 231-252.

15. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* **2005**, 28, (14), 1694-1703.

16. Hais, I. M., Diagonal techniques : Introductory remarks. *J. Chromatogr. A* **1970**, 48, 200-206.

17. Gevaert, K.; Van Damme, P.; Martens, L.; Vandekerckhove, J., Diagonal reverse-phase chromatography applications in peptide-centric proteomics: Ahead of catalogue-omics? *Anal. Biochem.* **2005**, 345, (1), 18-29.
18. Guo, K.; Li, L., Differential ¹²C-/¹³C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal. Chem.* **2009**, 81, (10), 3919-3932.
19. Chaimbault, P.; Petritis, K.; Elfakir, C.; Dreux, M., Ion-pair chromatography on a porous graphitic carbon stationary phase for the analysis of twenty underivatized protein amino acids. *J. Chromatogr. A* **2000**, 870, (1-2), 245-254.
20. Eckstein, J. A.; Ammerman, G. M.; Reveles, J. M.; Ackermann, B. L., Analysis of glutamine, glutamate, pyroglutamate, and GABA in cerebrospinal fluid using ion pairing HPLC with positive electrospray LC/MS/MS. *J. Neurosci. Methods* **2008**, 171, (2), 190-196.
21. Gustavsson, S. A.; Samskog, J.; Markides, K. E.; Langstrom, B., Studies of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents. *J. Chromatogr. A* **2001**, 937, (1-2), 41-47.
22. Gao, S. M.; Bhoopathy, S.; Zang, Z. P.; Wright, D. S.; Jenkins, R.; Karnes, H. T., Evaluation of volatile ion-pair reagents for the liquid chromatography-mass spectrometry analysis of polar compounds and its application to the determination of methadone in human plasma. *J. Pharm. Biomed. Anal.* **2006**, 40, (3), 679-688.
23. Kaspar, H.; Dettmer, K.; Gronwald, W.; Oefner, P. J., Advances in amino acid analysis. *Anal. Bioanal. Chem.* **2009**, 393, (2), 445-452.
24. Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S., Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **2009**, 5, (4), 435-458.

25. Armstrong, M.; Jonscher, K.; Reisdorph, N. A., Analysis of 25 underivatized amino acids in human plasma using ion-pairing reversed-phase liquid chromatography/time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, 21, (16), 2717-2726.
26. Bartha, A.; Vigh, G.; Vargapuchony, Z., Basics of the rational selection of the hydrophobicity and concentration of the ion pairing reagent in reversed-phase HPLC. *J. Chromatogr.* **1990**, 499, 423-434.
27. Sanchez-Lopez, J.; Camanes, G.; Flors, V.; Vicent, C.; Pastor, V.; Vicedo, B.; Cerezo, M.; Garcia-Agustin, P., Underivatized polyamine analysis in plant samples by ion pair LC coupled with electrospray tandem mass spectrometry. *Plant Physiol. Biochem.* **2009**, 47, (7), 592-598.
28. Kebarle, P.; Verkerk, U. H., Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **2009**, 28, (6), 898-917.
29. Seiler, N.; Deckardt, K., Determination of amines and amino acids in sugar-containing samples by dansylation. *J. Chromatogr.* **1975**, 107, (1), 227-229.
30. Seiler, N.; Knodgen, B.; Eisenbeiss, F., Determination of di- and polyamines by high-performance liquid-chromatographic separation of their 5-dimethylaminonaphthalene-1-sulfonyl derivatives. *J. Chromatogr.* **1978**, 145, (1), 29-39.
31. Barrett, D. A.; Shaw, P. N.; Davis, S. S., Determination of morphine and 6-acetylmorphine in plasma by high-performance liquid chromatography with fluorescence detection. *J. Chromatogr.* **1991**, 566, (1), 135-145.
32. Loukou, Z.; Zotou, A., Determination of biogenic amines as dansyl derivatives in alcoholic beverages by high-performance liquid chromatography with fluorimetric detection and characterization of the dansylated amines by liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *J. Chromatogr. A* **2003**, 996, (1-2), 103-113.

33. Minocha, R.; Long, S., Simultaneous separation and quantitation of amino acids and polyamines of forest tree tissues and cell cultures within a single high-performance liquid chromatography run using dansyl derivatization. *J. Chromatogr. A* **2004**, 1035, (1), 63-73.
34. Zezza, F.; Kerner, J.; Pascale, M. R.; Giannini, R.; Martelli, E. A., Rapid determination of amino acids by high-performance liquid chromatography: release of amino acids by perfused rat liver. *J. Chromatogr.* **1992**, 593, (1-2), 99-101.
35. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, 78, (3), 779-787.
36. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, 37, D603-D610.
37. Guo, K.; Li, L., High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry. *Anal. Chem.* **2010**, 82, (21), 8789-8793.

Chapter 3

Liquid-Liquid Extraction Combined with Differential Isotope Dimethylaminophenacyl Labeling for Improved Metabolomic Profiling of Organic Acids

3.1 Introduction

Comprehensive and quantitative profiling of the metabolome of a biofluid or a biological sample is an analytical challenge, due to great diversity of chemical and physical properties of metabolites and a large difference of metabolite concentrations in a complex mixture. One strategy to address this challenge is to divide the metabolome into sub-metabolomes according to the presence of a specific functional group in a metabolite and then develop optimal analytical tools tailored to the analysis of the sub-metabolomes. For example, we have recently reported the use of dansylation chemistry to label all the amine- and phenol-containing metabolites, followed by liquid chromatography mass spectrometry (LC-MS) analysis of the labeled metabolites ¹. With enhanced chromatographic separation and improved ionization efficiency, this method can be used to generate a much more comprehensive profile of the amine- and phenol-containing metabolites in various biological samples ¹⁻⁴, compared to

A form of this chapter was published as Jun Peng, Liang Li. "Liquid-Liquid Extraction Combined with Differential Isotope Dimethylaminophenacyl Labeling for Improved Metabolomic Profiling of Organic Acids" *Anal. Chim. Acta.* **2013**, 803, 97-105.

conventional LC-MS methods. Several other isotope reagents offering enhanced analytical performance for sub-metabolome profiling have also been reported⁵⁻¹¹.

The success of this divide-and-conquer metabolome profiling strategy is very much dependent on the availability of robust chemical labeling methods that target different groups of metabolites. An ideal labeling method should provide high labeling efficiency and alter the metabolite properties to enable better chromatographic separation and enhance ionization efficiency. In addition to dansyl labeling, we have developed an isotope labeling reagent *p*-dimethylaminophenacyl bromide (DmPA) for profiling the organic acid sub-metabolome¹⁰. This labeling reagent can be used to detect a large number of organic acids with signal enhancement of 10 to 1000 fold over the unlabeled counterparts. However, there are two major drawbacks of the current DmPA labeling method. One drawback is that the presence of water in a sample, such as a urine sample, could affect the labeling reaction efficiency of the acid metabolites. For example, a urine sample prepared in 80% acetonitrile (volume/volume) can maintain high solubility of metabolites present in urine. However, 20% water is still present in the urine sample. Water is nucleophilic under the alkaline condition used for labeling and, therefore, can react with DmPA to form the hydrolyzed DmPA compound. Water can compete with organic acid metabolites, particularly the low abundance metabolites, resulting in no or reduced labeling of some acid metabolites present in urine. Previous studies suggested that even 5% water in the sample could affect the phenacyl reaction kinetics and product yield for the tested compounds¹². The other drawback of the

existing labeling method is that the amine metabolites, specially the high abundance amine metabolites, can suppress the signals of the acid metabolites. Even though the reaction condition used for DmPA labeling was optimized for labeling the acid metabolites, some amines, particularly the high abundance ones, could still be labeled.

To circumvent the above drawbacks, we have developed an improved acid labeling method based on the use of liquid-liquid extraction (LLE) for enriching the acid metabolites prior to carrying out the DmPA labeling. LLE is simple and rapid. It has been widely used for gas chromatography (GC) MS analysis of organic acids in clinical laboratories^{13,14}. In GC-MS analysis of organic acids, it typically uses a starting material of 0.5 to 2 mL of urine¹⁵. A recent report indicated the possibility of analyzing organic acids by GC-MS using 200 μ L of urine as the starting material¹⁶. However, for metabolomic profiling work, it is desirable to use a smaller volume of a biofluid due to a limited sample available for many metabolomics studies. For example, in profiling the mouse urine metabolome, only tens of microlitres of urine are available. Fortunately, the DmPA labeling reaction offers much improved detectability for profiling the acid metabolome¹⁰. Thus, in this work, we report the novel combination of liquid-liquid extraction and DmPA labeling for handling small volumes of samples (tens of microlitres). Using human urine as an example, we demonstrate that this method offers much improved performance for profiling the sub-metabolome of organic acids.

3.2 Experimental

3.2.1 Chemicals and Reagents

All chemicals and reagents, except those specifically noted, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). LC-MS grade water and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada). Synthesis of ^{12}C - and ^{13}C - *p*-dimethylaminophenacyl bromide was performed according to the published procedure ¹⁰.

3.2.2 Urine and standard solution preparation

Urine samples were collected from healthy individuals. An informed consent was obtained from individual volunteers and ethics approval was obtained from the University of Alberta in compliance with the University of Alberta Health Information policy. The collected urine samples were stored at -80°C until further use. Urine was thawed on ice and centrifuged for 10 min at 13000 rpm. A standard solution of organic acid was prepared by dissolving each metabolite standard in water to give a concentration of 2 mM.

3.2.3 DmPA labeling using the existing protocol

For DmPA labeling, an existing protocol reported earlier (see Figure 3.1A for the workflow) ^{10, 17, 18} was adapted for labeling the urine and metabolite standards for comparison with the new protocol (see below). An aliquot of 50 μL of urine or a metabolite standard prepared in 80% acetonitrile solution (volume/volume) was added into a vial with a screw cap. Then 50 μL of triethylamine (TEA) dissolved in fresh acetonitrile (20 mg/mL) were added. The vial was vortexed and spun down. Then a solution of either ^{12}C -DmPA or ^{13}C -

DmPA in acetonitrile (20 mg/mL) was added into the vial. The vial was again vortexed and spun down. The vial was placed in an oven at 85°C for 55 min. After combining the ¹²C-labeled sample with the ¹³C-labeled sample, 2 µL of the mixture were injected into LC-MS for analysis.

3.2.4 Liquid-liquid extraction combined with DmPA labeling

For the new method of combining liquid-liquid extraction with DmPA labeling, Figure 3.1B shows the workflow using the 90 µL of urine as the starting material. In this case, 90 µL of urine or standard solution was aliquoted into a centrifuging plastic vial. Ten µL of 6M HCl solution was added, followed by adding 10 µL of saturated NaCl solution, and the vial was vortexed and spun down. The sample was extracted using 300 µL of ethyl acetate. Each vial was vortexed for 30 s and was then centrifuged at 8000 rpm for 5 min. The organic phase was transferred into another plastic vial with a screw cap. The pH of the organic phase was adjusted to 8 by adding 20 µL of TEA solution (180 mg/mL in acetonitrile). The sample was dried down using SpeedVac. Then 60 µL of TEA solution in acetonitrile (20 mg/mL) were added. The vial was vortexed and spun down. A solution of either ¹²C-DmPA or ¹³C-DmPA in acetonitrile (20 mg/mL) was added into the vial. The vial was again vortexed and spun down. The vial was placed in an oven at 85 °C for 55 min. After combining the ¹²C-labeled sample with ¹³C-labeled sample, 2 µL of the mixture were injected into LC-MS. Injection of this small volume of sample in pure acetonitrile did not affect the chromatographic peak shape, as the flow rate used was quite high, 0.18 mL/min,

with a 2.1-mm \times 100 column (see below). Similar peak shapes could be obtained from injection of a sample with reduced organic content (i.e., 20% acetonitrile).

The experimental workflows used for handling 20 μ L and 10 μ L of urine samples are shown in Figures 3.1C and 3.1D, respectively.

3.2.5 Measurement of extraction recovery rate

Three different concentrations of standard solutions (25 μ M, 125 μ M, and 250 μ M) of four organic acids, phenylacetic acid, 3-hydroxybenzoic acid, hydrocinnamic acid, and phenylglyoxylic acid, were prepared. For each sample, 90 μ L solution was extracted according to the LLE protocol described in Figure 3.1. After adding TEA to adjust pH to 8 and drying down by SpeedVac, the samples were redissolved in 50 μ L of acetonitrile. The standard solutions before extraction and after extraction were injected into UPLC-UV for measuring the UV absorbance at 210 nm to determine the recovery rate. The recovery rate was obtained by comparing the peak area of a standard solution after extraction with the peak area of the same standard solution before extraction. Three experimental replicates were performed for each extraction.

3.2.6 Comparison of the overall extraction and labeling efficiency

To compare the overall extraction and labeling efficiencies of the original protocol and the new protocol, acid standards, lactic acid (250 μ M), hydrocinnamic acid (125 μ M), 3-hydroxybenzoic acid (125 μ M) and phenylacetic acid (125 μ M), were spiked into a urine sample individually. The urine sample with a spiked standard was processed and labeled by the original protocol and the new protocol. The labeled urine sample spiked with one standard was injected

into UPLC-UV for absorbance measurement at 338 nm. The chromatograms of the labeled standard obtained using the two protocols were overlaid, and the peak area change was calculated to determine the overall efficiency difference of the two methods.

3.2.7 UPLC-UV

A Waters ACQUITY UPLC system with binary solvent manager and a photodiode array (PDA) detector was used for the detection and quantification of organic acid standards before and after extraction or after labeling. A Waters ACQUITY BEH C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 130 Å pore size) was used. The flow rate was 0.42 mL/min, and the injection volume was 4 μL. Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient was set as follows: 0 min, 15%B; 1 min, 15%; 8.5min, 50%B; 14.7min, 95%B; 15.2 min, 98%B; 16min, 15%B; 20min, 15%B.

3.2.8 LC-MS

An Agilent capillary 1100 HPLC system (Agilent, Palo Alto, CA) coupled with a Bruker Apex-Qe 9.4-T Fourier-transform Ion Cyclotron Resonance (FT-ICR) MS (Bruker, Billerica, MA, U.S.A.) with electrospray ionization (ESI) was used to analyze the labeled samples. The column used was a reversed-phase Eclipse C18 column (2.1 mm × 100 mm, 1.8 μm particle size, 95Å pore size). For the LC-MS work, solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile (ACN), and solvent B was 0.1% (v/v) formic acid in ACN. The flow rate was 180 μL/min. The gradient elution profile was as follows: t=0 min, B%=20; t=9 min,

B%=50; t=22 min, B%=65; t=26 min, B%=80; t=28 min, B%=98; t=30 min, B%=98; t=30.5 min, B%=20. Injection of 1 μ L of ACN was done during the column re-equilibrium time from 31 to 45 min at 20% B. The injection sequence was setup for each sample injection to wash the injection needle through the flushing port in order to minimize the sample carryover from the needle. The elute from the LC column was split at a ratio of 3:1, and about 60 μ L/min was introduced into MS. All MS spectra were obtained in the positive ion mode. The MS conditions used for FTICR-MS were as follows: nitrogen nebulizer gas: 2.3 L/min, dry gas flow: 7.0 L/min, dry temperature: 190°C, capillary voltage: 4200 V, spray shield: 3700 V, acquisition size: 256 k, scan range: 200-1100, and ion accumulation time: 1 sec. Due to the slow scan speed of FTICR-MS, we optimized the 1.8- μ m particle UPLC column separation conditions to achieve reasonably high efficiency while maintaining a sufficiently wide of peak. The typical peak width was about 18 s for relatively high abundant analytes and only one or a few spectra could be detected from the low abundant analytes.

3.2.9 Data processing and metabolite database search

All LC-FTICR-MS data were processed using an in-house developed software called IsoMS. First, the mass spectral peaks were picked using the Bruker DataAnalysis software 4.0. The $^{12}\text{C}/^{13}\text{C}$ -ion pairs were found by their accurate mass difference of 2.0067 Da within the 2 ppm window. Redundant peaks of the same metabolite including natural isotope peaks, adducts peaks such as sodium adducts, dimer peaks, and multiply charged peaks were removed by the IsoMS software.

For putative metabolite identification based on accurate mass matches with the metabolites in a database, the Human Metabolome Database (HMDB)¹⁹ was used. The mass accuracy window used for database search was set at 20 ppm. We note that, for targeted metabolite or compound analysis where the analyte signal intensity is optimized at the optimal detection range of FTICR-MS (i.e., there is no signal saturation at the high end and the signal intensity is much higher than the detection limit at the low end), 2 ppm mass accuracy can be readily achieved. However, for metabolome profiling work where the analyte concentration differences in a complex sample can be quite large, not all analyte signals will fall within the optimal detection range of FTICR-MS. The high abundant ions may cause space charge effect that reduces the mass accuracy. For the low abundant ions, there may not be a sufficient number of ions to produce a perfect peak shape for peak detection. Mass errors of up to 20 ppm can be observed in these cases. Thus, in our work, we used 20 ppm mass window for database search.



Figure 3.1 Workflow of an existing protocol involving direct DmPA labeling of a sample (A). Working flows of a new protocol using liquid-liquid extraction followed by DmPA labeling with a starting material of (B) 90 µL of human urine, (C) 20 µL of human urine and (D) 10 µL of human urine.

3.3 Results and Discussion

3.3.1 Workflow for liquid-liquid extraction and DmPA labeling

Figure 3.1 shows the general workflows of the existing protocol and the new protocol using liquid-liquid extraction. Instead of preparing the urine sample in 80% acetonitrile in the old protocol, a new protocol was developed to separate the amine metabolites into the aqueous phase as well as remove the water in the sample by using an organic solvent to extract the acid metabolites. The starting volume of the urine sample was chosen to be 90 μ L for the method development (Figure 3.1B). This volume of human urine is available for many metabolomics studies using samples typically collected for bio-banking. For the liquid-liquid extraction of organic acids from urine, the most commonly used organic solvent is ethyl acetate and urine is typically extracted under a very acidic condition (pH around 1) ²⁰. Under the very acidic condition, amine metabolites are protonated and dissolved readily in the aqueous phase. The organic acids are not ionized under this acidic condition and remain to be relatively hydrophobic and, thus, they are more readily partitioned into the organic phase. A saturated NaCl solution is used to separate the organic phase from the aqueous phase better, since ethyl acetate is less soluble in a high ionic solution such as the saturated NaCl ²¹.

From our working experience in dealing with various metabolomic samples, metabolite sample loss can sometime be encountered during the sample drying down by SpeedVac ³. Thus, in the liquid-liquid extraction protocol, before drying down the organic phase, TEA is added to adjust the pH to 8 in order to

keep the organic acids more ionized to reduce volatility so that the loss of the acid metabolites can be minimized during solvent evaporation by SpeedVac. For the extraction, 300 μ L of ethyl acetate is used to obtain an about 3:1 volume ratio of the organic phase to the aqueous phase, which is commonly used in organic acid extraction^{15, 20}.

For the DmPA isotope labeling reaction, the procedure used in the current protocol is the same as the one used previously¹⁰. We note that, after DmPA labeling, the labeled metabolites are stable and storage of the labeled samples over a period of several months (e.g., up to 6 months) did not alter the quantitative results. However, we have not investigated the stability of the carboxylic acids in stored urine. We will investigate this issue in detail in the future, particularly within the context of studying the effect of sample storage on biomarker discovery.

3.3.2 Method comparison for analyzing standard metabolites

To demonstrate the performance of the new protocol in comparison to the existing protocol, a mixture of two metabolite standards were analyzed. One was a common amine metabolite in urine, ethanolamine, and the other one was a common acid metabolite in urine, hippuric acid. Panels A and B of Figure 3.2 show the extracted ion chromatograms of ethanolamine and hippuric acid obtained by LC-FTICR-MS using the old and new protocols, respectively. Figure 3.2A shows that an intense peak from DmPA labeled ethanolamine is detected in the old protocol without liquid-liquid extraction. Figure 3.2B shows that the new

protocol with liquid-liquid extraction effectively removes the ethanolamine peak. Moreover, the peak intensity of DmPA labeled hippuric acid is increased by about 2-fold. The improved labeling efficiency offered by the new protocol can be very useful for detecting low abundance metabolites (see below). Several other organic acid standards including citric acid, succinic acid and other acids involved in the tricarboxylic acid (TCA) cycle of metabolism were tested using the new protocol and it was generally found that the new protocol improved the detectability of the organic acids compared to the old protocol. For examples, the signal of fumaric acid was found to be enhanced by 2.3-fold and the signal of oxoglutaric acid was enhanced by 5.9-fold. For these two acids each containing two COOH groups, two DmPA tags were attached to each molecule. In general, for a metabolite containing multiple carboxylic groups, the multiply labeled metabolite is the dominant product. The IsoMS program selects the peak pair of differentially labeled metabolites and calculates the peak ratio for relative quantification; other low abundance peaks from less than full labeling are not considered for quantification.

We also tested the new protocol for the standard amino acids that are usually present in high abundance in biofluids, such as urine. They can be readily analyzed using the dansylation labeling chemistry and thus removing these high abundance amino acids during the acid analysis process can be beneficial in detecting other organic acids. One example is shown in Figure 3.2C where the extracted ion chromatogram of asparagine labeled with 2 DmPA tags and asparagine labeled with 3 DmPA tags obtained by using the old protocol is

displayed. However, by applying liquid-liquid extraction, asparagine was removed, prior to DmPA labeling, and thus it was not detected in LC-MS (see Figure 3.2D).

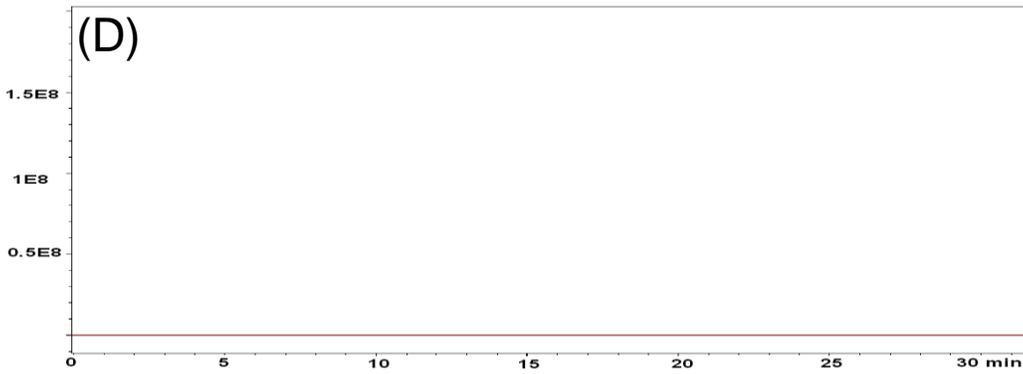
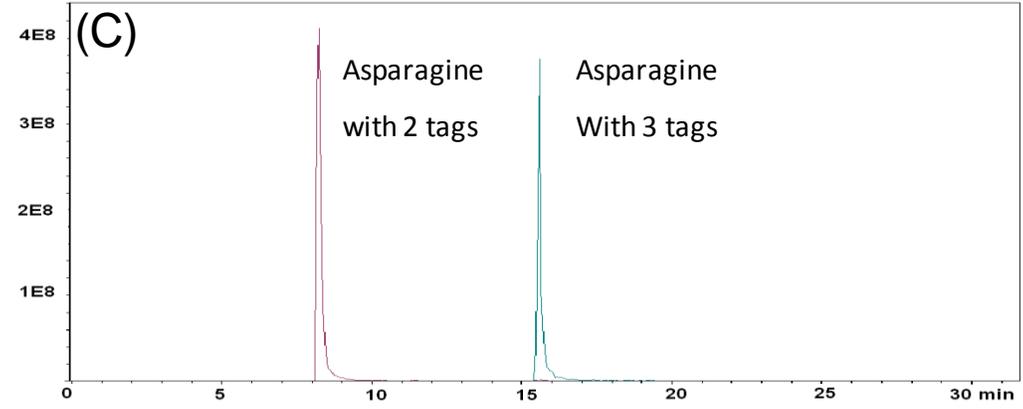
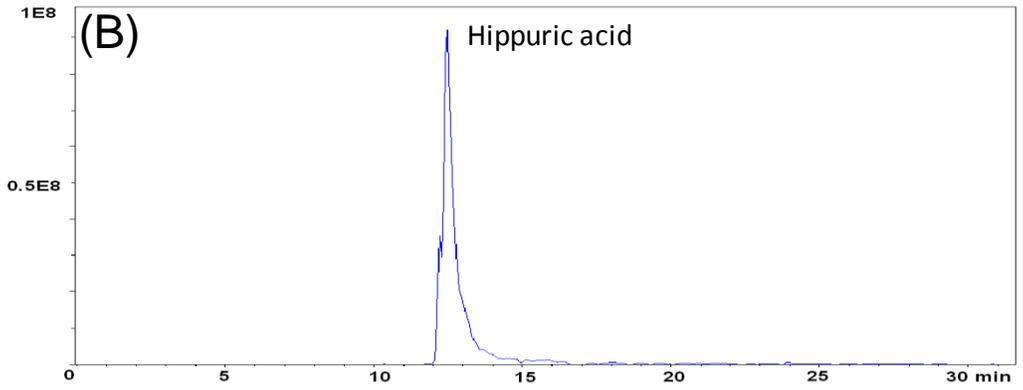
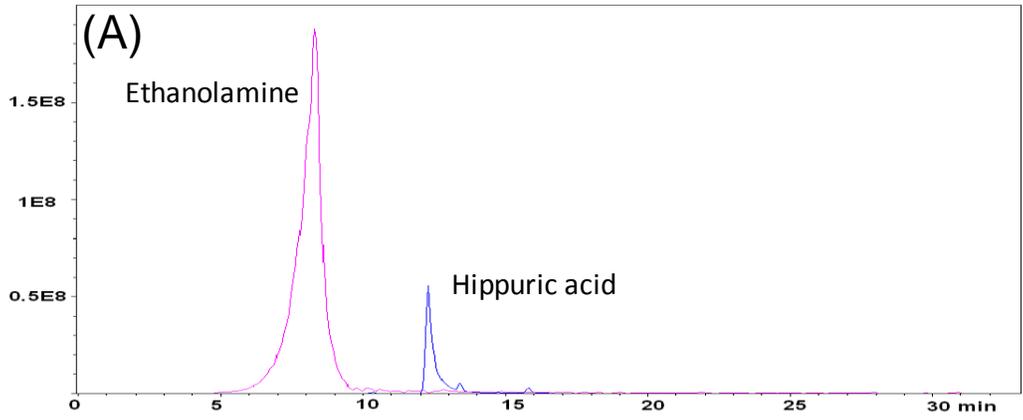


Figure 3.2 Extracted ion chromatograms of a mixture of DmPA labeled ethanolamine (m/z 384.2233) and DmPA labeled hippuric acid (341.1148) (the concentration was 2 mM for each standard and the injection volume was 2 μ L) obtained by (A) the old protocol with direct DmPA labeling and (B) the new protocol with LLE and DmPA labeling. Extracted ion chromatograms of asparagine labeled with 2 DmPA tags and asparagine labeled with 3 DmPA tags (the concentration was 2 mM and the injection volume was 2 μ L) obtained by (C) the old protocol and (D) the new protocol.

One potential concern for the new protocol is related to the analyte recovery rate during the sample extraction, drying and re-dissolution process. We have investigated this issue by determining the recovery rates for four standard acids, phenylacetic acid, 3-hydroxybenzoic acid, hydrocinnamic acid and phenylglyoxylic acid, prepared in three different concentrations (25 μ M, 125 μ M, and 250 μ M). UPLC-UV measurements of the analyte before and after subjecting the standard solution to the LLE process were performed. Table 3.1 summarizes the results obtained. The average recovery rate was found to be about 84% ranging from 73% to 95%. The recovery rate is more consistent for a given acid at different concentrations, which suggests that, for relative quantification of a metabolite, the LLE process should not introduce a significant error.

We have compared the overall sampling processing and labeling efficiency of the original and new protocols for four standards, lactic acid, hydrocinnamic acid, 3-hydroxybenzoic acid and phenylacetic acid. The individual standard solution was spiked into a human urine sample where the endogenous concentration of the metabolite was much lower than that of the spiked solution (250 μM for lactic acid and 125 μM for other three acids). After processing and labeling, UPLC-UV measurement of the labeled acid was carried out to determine the extent of improvement of the new protocol versus the old protocol. It was found that the improvement was 26-fold for lactic acid, 28-fold for hydrocinnamic acid, 5.6-fold for 3-hydroxybenzoic acid, and 4.4-fold for phenylacetic acid.

The above examples and results indicate that liquid-liquid extraction can be used to separate the amine metabolites from the organic acids and remove the water content, which can lead to improvement in the detection of organic acid metabolites by isotope labeling LC-MS.

Table 3.1 Summary of the analyte recovery rate in the sample extraction, drying and re-dissolution procedure.

Organic acid	Concentration	Recovery rate (n=3)
phenylglyoxylic acid	25 μ M	0.77 \pm 0.02
	125 μ M	0.91 \pm 0.01
	250 μ M	0.880 \pm 0.009
hydrocinnamic acid	25 μ M	0.88 \pm 0.04
	125 μ M	0.73 \pm 0.02
	250 μ M	0.75 \pm 0.02
phenylacetic acid	25 μ M	0.88 \pm 0.02
	125 μ M	0.82 \pm 0.02
	250 μ M	0.86 \pm 0.03
3-hydroxylbenzoic acid	25 μ M	0.95 \pm 0.02
	125 μ M	0.85 \pm 0.04
	250 μ M	0.84 \pm 0.05

3.3.3 Method comparison for analyzing urine samples

The new protocol was applied to analyze the metabolome of human urine. In this case, a urine sample was divided into two aliquots. One aliquot was subjected to liquid-liquid extraction and ^{12}C -DmPA labeling and another one was processed in the same way except it was labeled with ^{13}C -DmPA. The $^{12}\text{C}/^{13}\text{C}$ -differentially labeled aliquots were mixed in 1:1 for LC-MS analysis. All the mass spectral peaks of the labeled metabolites would be shown as peak pairs in the mass spectra where background peaks would appear as singlet's. Using the IsoMS program, all the peak pairs can be picked up and then filtered by removing

redundant peaks. The intensity ratios of the final list of peak pairs are calculated. Each peak pair with unique mass and retention time should represent a metabolite.

Panels A and B of Figure 3.3 show the ion chromatograms of human urine obtained using the old and new protocols, respectively. Figure 3.3C shows the overlaid chromatograms for direct comparison. There were 2491 peak pairs detected using the new protocol, while 874 peak pairs were found using the old protocol. The detection improvement is very significant, reflecting the ability of the new protocol to remove the amines and water that can interfere with the labeling and detection of the organic acids. As Figure 3.3C illustrates, amine metabolites such as ethanolamine, ethylamine and glycine in the urine sample were mostly removed, while at the same time the peak intensity was significantly increased for the organic acids such as citric acid, hippuric acid and lactic acid. In addition, the peak intensity from the hydrolyzed DmPA reagent was much reduced (i.e., 3.5% of the original peak intensity).

The above example demonstrates the new DmPA labeling protocol with liquid-liquid extraction can be used to improve the metabolome coverage of the organic acids greatly and also improve the specificity of the acid sub-metabolome analysis (see below; most matched peak pairs are belonging to the organic acids). It should be noted that, in a typical GC-MS experiment, less than 200 organic acids can be detected^{22, 23}. As it is shown in this work, the isotope labeling LC-MS method can detect ~2500 organic acids in urine. Thus, this method can be used to profile the organic acid sub-metabolome in a much more comprehensive manner.

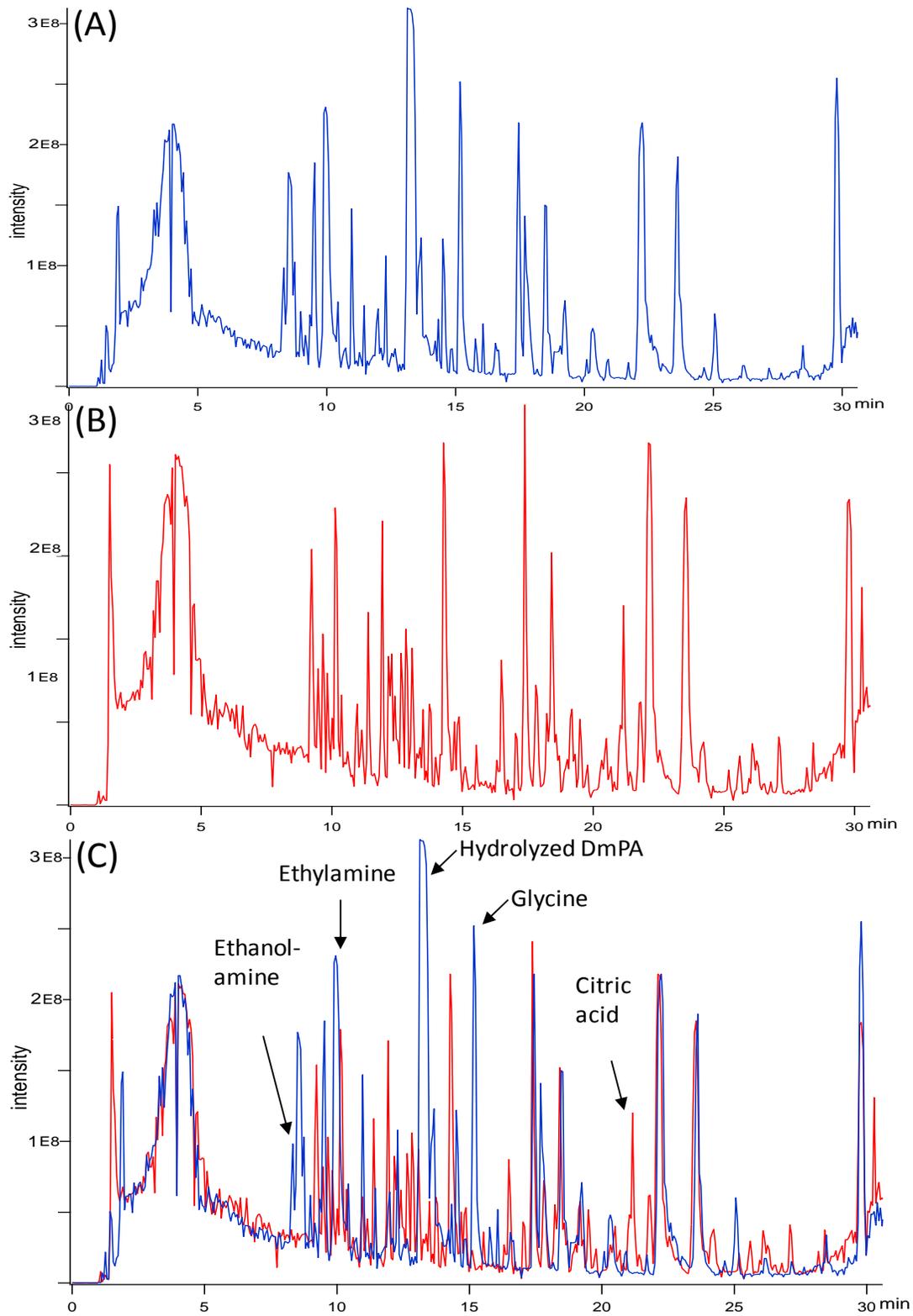


Figure 3.3 Base-peak ion chromatograms of DmPA labeled urine sample prepared using (A) the direct DmPA labeling method and (B) the new LLE DmPA method. (C) the overlaid ion chromatograms of (A) and (B) for comparison.

3.3.4 Low volume sample analysis

While the above liquid-liquid extraction protocol developed for handling a 90 μL of urine as a starting material is adequate for metabolomics applications where relatively larger volumes of samples are available, it is often desirable and, sometimes, necessary to analyze a smaller volume of sample. For example, in some metabolomics work, the sample volume is limited, particularly in animal studies such as mice. In other cases, a valuable biofluid may be divided into small volume aliquots for multiple analyses including proteomics and genomics analysis, in addition to metabolome profiling. And even for metabolomics studies, a sample may be further divided into several aliquots for different isotopic labeling methods in order to increase the metabolome coverage. Therefore, the volume of an analysis sample can be 20 μL or even lower. We have explored the feasibility of scaling down the new protocol to handle 20 and 10 μL of starting materials.

One straightforward approach for scaling down to handle 10 μL of sample volume is to reduce the liquid-liquid extraction volume from 90 μL to 10 μL , and also the chemical reaction volume from 60 μL to 10 μL proportionally. However, this approach did not produce the same results as those with the 90 μL of starting

materials; both the number of peak pairs detected and reproducibility were significantly lower. On the other hand, it was found that scaling down from 90 μL to 45 μL in extraction and from 60 μL to 30 μL in reaction did not sacrifice the analytical performance. When using this way of scaling down, after combining the ^{12}C -labeled sample with the ^{13}C -labeled sample, the final volume would be as large as 120 μL , from which only 2 μL was injected into LC-MS. In order to offset the limitation of low volume starting material and hence overall low concentration of metabolites, the labeled mixture was concentrated using SpeedVac to a volume with an overall concentration equivalent to that produced using the 90 μL of starting material. Panels B, C and D of Figure 3.1 show the workflow for handling 90 μL , 20 μL and 10 μL of urine starting material and the difference in the three experiments are highlighted.

Figure 3.4 shows the comparison of the number of peak pairs detected from three volumes of starting material. For the 90 μL starting material, 2561 ± 27 ($n=3$) peak pairs were detected. The number of peak pairs obtained from the 20 μL starting material was 2548 ± 118 ($n=3$). The number of peak pairs from the 10 μL starting material was 2301 ± 249 ($n=3$). These results show that it is still possible to detect similar number of metabolites using low volume of samples even with 10 μL of urine. However, the number reproducibility from the 10 μL sample is not as good as those from the 90 or 20 μL samples. This may be due to sample loss encountered during the drying down by SpeedVac. The effect of sample loss is expected to be more pronounced for a small volume sample (10 μL), compared to large volume samples (20 or 90 μL). In other words, if the total

amount of sample loss is the same for all samples (e.g., via the adsorption to the container walls), the percentage of sample loss is much higher for a small volume sample compared to a large volume samples.

Another important criterion to gauge the performance of the new protocol for handling various volumes of samples is the relative standard deviation (RSD) of the intensity ratio for all the detected metabolite peak pairs. It was found that the average RSD was 8.8% for the 20 μL sample and 12% for the 10 μL sample. Thus, the precision of the method should be sufficient for comparative metabolomics where the relative changes of the metabolomes from different cohorts of samples are analyzed. The application of this method for comparative metabolomics is currently underway and will be reported in the future.

The above results indicate that the liquid-liquid extraction DmPA labeling method can be used to handle low volumes of urine samples. While the number of peak pairs detected and the precision of the peak intensity ratio for all the peak pairs are slightly reduced when 10 μL of starting volume was used, compared to 90 μL of volume, the use of 20 μL of urine does not affect the performance. We feel that for routine metabolomics applications 20 μL of urine can be conveniently used for acid metabolome profiling. However, if the sample volume is very limited, 10 μL of urine can still generate good results.

3.3.5 Putative metabolite identification

FTICR-MS can provide high mass measurement accuracy for measuring the masses of metabolite ions and the mass error is typically less than 2 ppm,

although signal saturation from very high abundant ions or less than ideal peak shapes from low abundant ions can increase the mass error up to 20 ppm. The measured mass of an organic acid can be calculated from the detected m/z of the peak pair by subtracting the isotope tag. By matching the accurate mass with those of the metabolites in HMDB, a list of putatively identified organic acids in urine can be generated. Table 3.2 shows a partial list of putatively identified organic acids in a urine sample. In the urine sample, there were 1004 matched metabolites out of 2527 detected peak pairs. About 95% of the 1004 matched metabolites have a carboxylic acid group; there are about 5% of them matched with acylcarnitines, diols, metabolites with hydroxyl group(s) only (e.g., sucrose), and other no-carboxylic-acid metabolites. The matched acids include many common organic acids analyzed by GC-MS in clinical laboratories. Although these matched metabolites with HMDB need to be confirmed for positive identification by authentic standards, this list should still be useful to clinicians and biochemists to explore some unusual organic acids in some metabolism disorders that are detectable in urine ²⁴. To further broaden the metabolome coverage and increase the number of putatively identified organic acids, we are planning to use two-dimensional LC separation, combined with MS ²⁵, to further expand the concentration dynamic range and detect more organic acids.

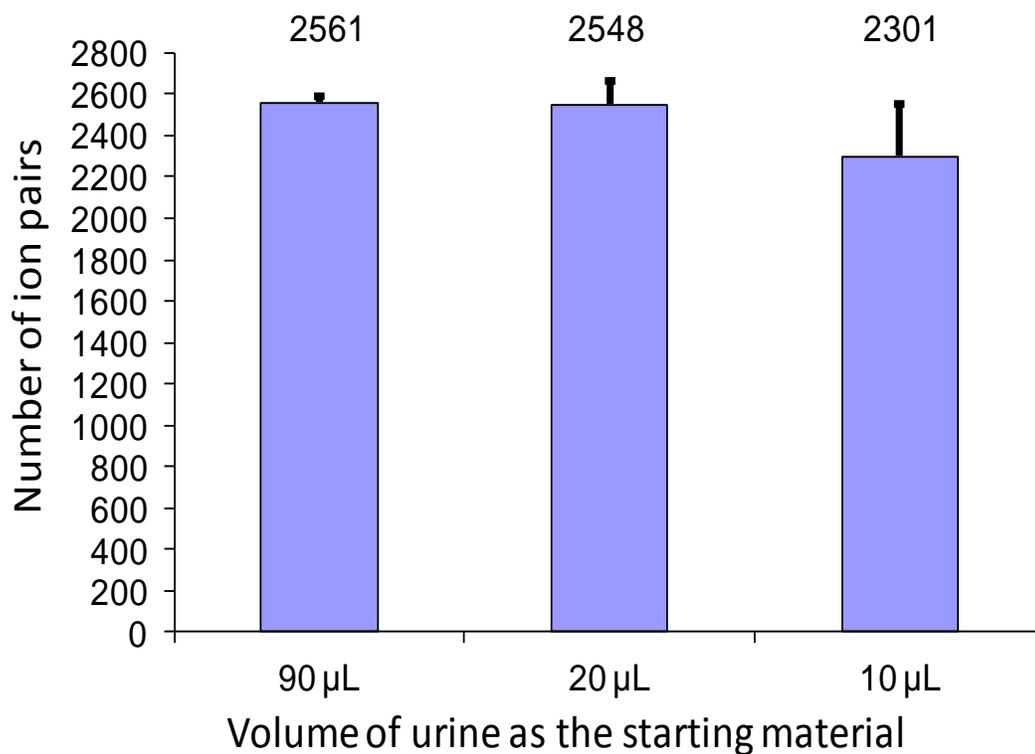


Figure 3.4 Number of peak pairs detected as a function of the volume of the starting material.

3.4 Conclusions

We have developed an improved isotope labeling method based on liquid-liquid extraction and *p*-dimethylaminophenacyl reaction for LC-MS metabolomic profiling of organic acid metabolites. LLE under acidic condition can effectively separate amine metabolites from organic acid metabolites, thereby increasing the detectability of organic acids using DmPA labeling LC-MS. LLE can also remove

the water from a sample, which eliminates the water interference in DmPA labeling reaction and greatly reduces the water-induced hydrolysis product of the DmPA reagent that can interfere with the LC-MS detection of labeled organic acids. This method can be used to handle low volumes of urine samples; 2561 ± 27 (n=3), 2548 ± 118 (n=3), and 2301 ± 249 (n=3) peak pairs could be detected in human urine with a starting material of 90, 20 and 10 μL , respectively. These numbers are almost three times more than that detected using the original method without LLE. We envisage the routine use of this method for metabolomic profiling of organic acids for comparative metabolomics applications. In addition, we will combine this method with the isotope dansyl labeling amine- and phenol-metabolome profiling workflow to extend the metabolome coverage. The applications of this method for handling other types of biofluids and biological samples will also be reported in the future.

Table 3.2 Partial list of organic acids putatively identified based on accurate mass match with the HMDB metabolites.

Peak pair series number	Retention time (min)	Intensity	Accurate Mass	Name of putative metabolite	HMDB ID
1	10.5	1.12E+06	46.0047	Formic acid	HMDB00142
2	16.7	6.20E+05	60.0203	Acetic acid	HMDB00042
				Glycolaldehyde	HMDB03344
3	24.8	3.48E+05	60.0204	Acetic acid	HMDB00042
				Glycolaldehyde	HMDB03344
4	20.3	4.06E+05	132.0413	Dimethylmalonic acid	HMDB02001
				(S)-2-Acetolactate	HMDB06855
				2-Acetolactate	HMDB06833
				Ethylmalonic acid	HMDB00062
				Glutaric acid	HMDB00066
				Methylsuccinic acid	HMDB01844
				Monoethyl malonic acid	HMDB00057
5	21.9	9.54E+04	141.0385	2-Aminomuconic acid semialdehyde	HMDB01280
6	21.4	7.01E+05	143.0540	Vinylacetylglycine	HMDB00089
7	20.2	1.41E+05	153.0371	3-Aminosalicyclic acid	HMDB01972
				3-Hydroxyanthranilic acid	HMDB01476
8	25.6	1.90E+05	160.0691	Pimelic acid	HMDB00085
				3,3-Dimethylglutaric acid	HMDB02441
				3-Methyladipic acid	HMDB00055
9	18.0	8.52E+05	161.0623	Amino adipic acid	HMDB00051
10	15.2	3.30E+05	162.0399	1,2-Dihydroxy-3-keto-5-methylthiopentene	HMDB12134
				3 Hydroxycoumarin	HMDB02149

11	18.5	2.98E+0 5	166.0321	Terephthalic acid	HMDB0242 8
				Benzoquinoneacetic acid	HMDB0233 4
				Phthalic acid	HMDB0210 7
12	22.8	1.80E+0 5	168.0374	Homogentisic acid	HMDB0013 0
				3,4-Dihydroxybenzeneacetic acid	HMDB0133 6
				3,4-Dihydroxymandelaldehyde	HMDB0624 2
				3-Hydroxymandelic acid	HMDB0075 0
				5-Methoxysalicylic acid	HMDB0186 8
				p-Hydroxymandelic acid	HMDB0082 2
				Uric acid	HMDB0028 9
				Vanillic acid	HMDB0048 4
13	16.0	1.81E+0 5	169.0293	L-2,3-Dihydrodipicolinate	HMDB1224 7
				2-Furoylglycine	HMDB0043 9
14	15.5	1.67E+0 5	171.0588	Tetrahydrodipicolinate	HMDB1228 9
15	27.7	1.22E+0 5	172.1534	Capric acid	HMDB0051 1
16	26.1	6.50E+0 5	173.1001	Isovalerylalanine	HMDB0074 7
				Hexanoylglycine	HMDB0070 1
				Isovalerylsarcosine	HMDB0208 7
				N-Acetylleucine	HMDB1175 6
17	25.7	1.25E+0 5	174.0854	Suberic acid	HMDB0089 3
				Demethylated antipyrine	HMDB0624 0
				Ethyladipic acid	HMDB0202 3
18	18.4	7.25E+0 4	176.0669	3-Isopropylmalate	HMDB1215 6
				2,3-Dimethyl-3-hydroxyglutaric acid	HMDB0202 5
				2-Isopropylmalic acid	HMDB0040 2
19	11.9	1.00E+0 6	90.0296	Dihydroxyacetone	HMDB0188 2
				D-Lactic acid	HMDB0131 1

				Glyceraldehyde	HMDB01051
				Hydroxypropionic acid	HMDB00700
				L-Lactic acid	HMDB00190
20	8.4	9.18E+05	90.0299	Dihydroxyacetone	HMDB01882
				D-Lactic acid	HMDB01311
				Glyceraldehyde	HMDB01051
				Hydroxypropionic acid	HMDB00700
				L-Lactic acid	HMDB00190
21	9.2	8.04E+06	90.0301	Dihydroxyacetone	HMDB01882
				D-Lactic acid	HMDB01311
				Glyceraldehyde	HMDB01051
				Hydroxypropionic acid	HMDB00700
				L-Lactic acid	HMDB00190

References

1. Guo, K.; Li, L., Differential C-12/C-13-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal. Chem.* **2009**, 81, (10), 3919-3932.
2. Guo, K.; Bamforth, F.; Li, L., Qualitative metabolome analysis of human cerebrospinal fluid by C-13-/C-12-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011**, 22, (2), 339-347.
3. Zheng, J. M.; Dixon, R. A.; Li, L., Development of isotope labeling LC-MS for human salivary metabolomics and application to profiling metabolome changes associated with mild cognitive impairment. *Anal. Chem.* **2012**, 84, (24), 10802-10811.
4. Wu, Y.; Li, L., Development of isotope labeling liquid chromatography mass spectrometry for metabolic profiling of bacterial cells and its application for bacterial differentiation. *Anal. Chem.* **2013**, in press. DOI: 10.1021/ac400330z.
5. Dai, W. D.; Huang, Q.; Yin, P. Y.; Li, J.; Zhou, J.; Kong, H. W.; Zhao, C. X.; Lu, X.; Xu, G. W., Comprehensive and highly sensitive urinary steroid hormone profiling method based on stable isotope-labeling liquid chromatography mass spectrometry. *Anal. Chem.* **2012**, 84, (23), 10245-10251.
6. Zhang, S. J.; You, J. M.; Ning, S. J.; Song, C. H.; Suo, Y. R., Analysis of estrogenic compounds in environmental and biological samples by liquid chromatography-tandem mass spectrometry with stable isotope-coded ionization-enhancing reagent. *J. Chromatogr. A* **2013**, 1280, 84-91.
7. Mazzotti, F.; Benabdelkamel, H.; Di Donna, L.; Athanassopoulos, C. M.; Napoli, A.; Sindona, G., Light and heavy dansyl reporter groups in food chemistry: amino acid assay in beverages. *J. Mass Spectrom.* **2012**, 47, (7), 932-939.
8. Wang, H.; Wang, H. Y.; Zhang, L.; Zhang, J.; Guo, Y. L., N-Alkylpyridinium isotope quaternization for matrix-assisted laser desorption/ionization Fourier transform mass spectrometric analysis of cholesterol and fatty alcohols in human hair. *Anal. Chim. Acta* **2011**, 690, (1), 1-9.
9. Huang, Y. Q.; Liu, J. Q.; Gong, H. Y.; Yang, J.; Li, Y. S.; Feng, Y. Q., Use of isotope mass probes for metabolic analysis of the jasmonate biosynthetic pathway. *Analyst* **2011**, 136, (7), 1515-1522.

10. Guo, K.; Li, L., High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry. *Anal. Chem.* **2010**, 82, (21), 8789-8793.
11. Armenta, J. M.; Cortes, D. F.; Pisciotta, J. M.; Shuman, J. L.; Blakeslee, K.; Rasoloson, D.; Ogunbiyi, O.; Sullivan, D. J.; Shulaev, V., Sensitive and rapid method for amino acid quantitation in Malaria biological samples using AccQ.Tag ultra performance liquid chromatography-electrospray ionization-MS/MS with multiple reaction monitoring. *Anal. Chem.* **2010**, 82, (2), 548-558.
12. Nallu, M.; Subramanian, M.; Vembu, N.; Hussain, A. A., Effect of binary aqueous-organic solvents on the reaction of phenacyl bromide with nitrobenzoic acid(s) in the presence of triethylamine. *Int. J. Chem.l Kinet.* **2004**, 36, (7), 401-409.
13. Liebich, H. M., Sample preparation for organic acids in biological fluids. *Anal. Chim. Acta* **1990**, 236, (1), 121-130.
14. Jones, P. M.; Bennett, M. J., Urine Organic Acid Analysis for Inherited Metabolic Disease Using Gas Chromatography-Mass Spectrometry. In *Clinical Applications of Mass Spectrometry: Methods and Protocols*, Humana Press Inc: Totowa, 2010; Vol. 603, pp 423-431.
15. Duez, P.; Kumps, A.; Mardens, Y., GC-MS profiling of urinary organic acids evaluated as a quantitative method. *Clin. Chem.* **1996**, 42, (10), 1609-1615.
16. Nakagawa, K.; Kawana, S.; Hasegawa, Y.; Yamaguchi, S., Simplified method for the chemical diagnosis of organic aciduria using GC/MS. *J. Chromatogr. B* **2010**, 878, (13-14), 942-948.
17. Borch, R. F., Separation of long-chain fatty-acids as phenacyl esters by high-pressure liquid-chromatography. *Anal. Chem.* **1975**, 47, (14), 2437-2439.
18. Lewis-Stanislaus, A. E.; Li, L., A Method for comprehensive analysis of urinary acylglycines by using ultra-performance liquid chromatography quadrupole linear ion trap mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, 21, (12), 2105-2116.
19. Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L., HMDB: the human metabolome database. *Nucleic Acids Res.* **2007**, 35, D521-D526.

20. Tanaka, K.; Hine, D. G.; Westdull, A.; Lynn, T. B., Gas-chromatographic method of analysis for urinary organic-acids 1. retention indexes of 155 metabolically important compounds. *Clin. Chem.* **1980**, 26, (13), 1839-1846.
21. Suh, J. W.; Lee, S. H.; Chung, B. C., GC-MS determination of organic acids with solvent extraction after cation-exchange chromatography. *Clin. Chem.* **1997**, 43, (12), 2256-2261.
22. Kaluzna-Czaplinska, J., Current applications of gas chromatography/mass spectrometry in the study of organic acids in urine. *Crit. Rev. Anal. Chem.* **2011**, 41, (2), 114-123.
23. Wasant, P.; Liammongkolkul, S.; Kuptanon, C.; Vatanavicharn, N.; Sathienkijakanchai, A.; Shinka, T., Organic acid disorders detected by urine organic acid analysis: Twelve cases in Thailand over three-year experience. *Clin. Chim. Acta* **2008**, 392, (1-2), 63-68.
24. Lehotay, D. C.; Clarke, J. T. R., Organic acidurias and related abnormal. *Crit. Rev. Clin. Lab. Sci.* **1995**, 32, (4), 377-429.
25. Guo, K.; Peng, J.; Zhou, R. K.; Li, L., Ion-pairing reversed-phase liquid chromatography fractionation in combination with isotope labeling reversed-phase liquid chromatography-mass spectrometry for comprehensive metabolome profiling. *J. Chromatogr. A* **2011**, 1218, (23), 3689-3694.

Chapter 4

Development of Isotope Labeling Liquid Chromatography Mass Spectrometry for Mouse Urine Metabolomics: Quantitative Metabolomic Study of Transgenic Mice Related to Alzheimer's Disease

4.1 Introduction

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders in the elderly.¹ Currently there is no effective treatment for AD. Furthermore, there are no definitive biomarkers for the reliable clinical diagnosis of AD at the early stage of development, although late-stage AD can be diagnosed using MRI brain scans and protein biomarkers in cerebrospinal fluid.²⁻⁴ The difficulty in diagnosing AD may arise from the complex perturbations occurring in the genome, transcriptome, proteome, and metabolome of the affected individuals. While most AD biomarker studies have focused on transcriptomic or proteomic assays, metabolomic assays may prove to be sensitive enough to distinguish AD phenotypes from normal.⁵⁻⁷ Traditional biochemical studies have shown that neurodegenerative disorders are often linked to disturbances in metabolic pathways related to neurotransmitter synthesis,^{8,9} oxidative stress¹⁰ and mitochondrial function.¹¹ Metabolome analysis with a broad metabolite coverage

A form of this chapter was submitted to Journal of Proteome Research as Jun Peng, Kevin Guo, Jianguo Xia, Jianjun Zhou, Jing Yang, David Westaway, David Wishart and Liang Li "Development of Isotope Labeling Liquid Chromatography Mass Spectrometry for Mouse Urine Metabolomics: Quantitative Metabolomic Study of Transgenic Mice Related to Alzheimer's Disease". I did the metabolome profiling experiments and analyzed the data.

may allow us to monitor metabolite perturbations in many biochemical pathways and to discover a panel of metabolite biomarkers specific to AD.

Over the past few years, a significant number of metabolomics studies on AD have been reported. These include animal models as well as human studies. In human studies, most investigated CSF samples,¹²⁻¹⁵ plasma samples,¹⁵⁻¹⁸ or brain samples.¹⁹ In animal models, brain tissues of mouse models were most commonly used.²⁰⁻²² Two mouse model studies used both brain and plasma.^{23, 24} All these studies indicated the metabolomic changes associated with AD and showed the initial promise of metabolomics techniques in the investigation of AD disease. The analytical techniques employed include LC-electrochemical array (ECA),^{12, 13} GC-MS,²³ LC-MS,^{16, 18, 19, 25} capillary electrophoresis (CE)-MS,²⁶ and NMR.^{20, 24} However, each of these techniques could only detect a relatively small number of metabolites. There is a need to develop more sensitive and quantitative metabolomic approaches to search for biomarkers of AD. We have recently reported a quantitative metabolomic technique based on ¹³C-/¹²C-isotope dansylation labeling combined with LC Fourier-transform ion cyclotron resonance (FTICR) MS.^{27,28} Dansylation is a chemical derivatization method that greatly enhances the electrospray ionization (ESI) signal response and improves reversed-phase (RP) LC separation, hence offers much more comprehensive metabolome coverage. Furthermore, the introduction of ¹³C-/¹²C isotope tags enables relative (and absolute) quantification by isotope dilution with much improved measurement precision. This method has been successfully applied for

analyzing biological samples, including human CSF,²⁸ human urine,²⁹ human saliva³⁰ and bacterial cells^{31, 32}.

Urine metabolomics in a transgenic mouse model of AD has not been extensively studied. Only one urine metabolomics study of mouse model of AD using NMR was reported very recently.³³ Because of a limited volume of urine available from a mouse, it is technically challenging to perform urine metabolomics with high metabolome coverage. However, due to the noninvasive nature of sample collection, urine is an excellent source for metabolomics, especially if one wishes to find a set of non-invasive early-stage diagnostic biomarkers, which could be complementary to biomarkers in CSF, blood, or brain tissue. Here we report a quantitative metabolomics approach based on differential ¹³C-/¹²C-isotope labeling combined with LC-FTICR-MS for urinary biomarker discovery in a transgenic mouse model of AD.

4.2 Experimental Section

4.2.1 Chemicals and reagents.

All the chemicals and reagents, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). For dansylation labeling reactions, the ¹²C-labeling reagents were from Sigma-Aldrich and the ¹³C-labeling reagents were synthesized in our lab using the procedures published previously²⁷. LC-MS grade water, methanol, and acetonitrile (ACN) were purchased from ThermoFisher Scientific (Nepean, ON, Canada).

4.2.2 Mouse model and sample collection.

The TgCRND8 transgenic mice (see section 4.3 for more information) were used in this study. The strains of all the mice were 129Svev and the mice were fed with 4% mice chow. To avoid possible contamination with bedding material and animal waste the urine samples were collected by lifting the mouse from the home cage and immediately placing into a brand new disposable plastic cage (Innovive Inc. San Diego, CA). Urine was pipetted from the floor of the cage into a 1.5 mL Eppendorf tube and snap-frozen on dry-ice. The urine samples were collected at three different ages, namely 15-17 weeks, 25-28 weeks, and 30-31 weeks. Table 4.1 shows the sample information about the mouse urine samples.

4.2.3 Sample preparation.

Figure 1 shows the experimental workflow for sample preparation and LC-MS analysis. A pooled sample was created by aliquoting 10 μL of each of the 75 mice urine samples and mixing them well. The method for dansylation labeling of urine was adapted from our previous report.²⁷ Briefly, 10 μL volume of individual mouse urine sample or 10 μL of the pooled sample were diluted to 50 μL by adding 40 μL water (LC-MS grade), then mixed with 50 μL sodium carbonate/sodium bicarbonate buffer (0.5 mol/L, pH 9.5) in reaction vials. Fifty μL of freshly prepared ^{12}C -dansyl chloride solution (20 mg/mL) was added to each of the individual samples for light labeling. Fifty μL of ^{13}C -dansyl chloride solution (20 mg/mL) were added to each of the pooled sample for heavy labeling. The dansylation reaction was performed in an Innova-4000 bench top incubator

shaker at 60 °C for 60 min. Ten μL of the 250 mM sodium hydroxide solution was added to quench the reaction. After additional 10 min incubation, the ^{12}C -dansylated individual sample was combined with the ^{13}C -dansylated pooled sample. The pH of the combined mixture was adjusted to pH 3 by adding formic acid. The mixture was then centrifuged for 10 min at 13,800 g and was ready for LC-MS injection. A mixture of ^{12}C -dansylated pooled sample and ^{13}C -dansylated pooled sample in 1:1 (v:v) was used as a quality control (QC) sample.

4.2.4 LC-MS.

The HPLC system was an Agilent capillary 1100 binary system (Agilent, Palo Alto, CA). A reversed-phase Acquity BEH C18 column (1.0 \times 100 mm, 1.7 μm particle size) was purchased from Waters (Milford, MA, U.S.A). LC Solvent A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and Solvent B was 0.1% (v/v) formic acid in acetonitrile. The gradient elution profile was as follows: t=0 min, 20% B; t=3.5 min, 35% B; t=18 min, 65% B; t=21 min, 95% B; t=21.5 min, 95% B; t=23 min, 98% B; t=24 min, 98% B; t=25.5 min, 20% B; t=36.6 min, 20% B. The flow rate was 35 $\mu\text{L}/\text{min}$. The sample injection volume was 1.0 μL . To minimize the carryover and reduce the background noise peaks, a fast gradient elution of 22 min was run after each sample injection. The LC column was directly connected to a Bruker 9.4 Tesla Apex-Qe FTICR mass spectrometer (Bruker, Billerica, MA, USA) without splitting. The electrospray ionization (ESI) mass spectra were collected in the positive ion mode. The MS conditions used for FTICR-MS were as follows: nitrogen nebulizer gas: 2.3 L/min, dry gas flow: 7.0 L/min, dry temperature: 195 °C, capillary voltage: 4200 V, spray shield: 3700 V,

acquisition size: 256 k, mass scan range: m/z 200-1000, ion accumulation time: 1 s, TOF (AQS): 0.007 s, DC Extract Bias: 0.7 V. All the urine samples were put into random order for injections. Eight injections of the QC sample (one QC injection after 9 sample injections) were used to monitor the performance of LC-MS running during the whole experiment.

For metabolite identification, an AB Sciex QTRAP 2000 hybrid triple-quadrupole with a linear ion trap mass spectrometer (Toronto, ON, Canada) was used to collect the MS/MS spectra of the interesting metabolites.

4.2.5 Data analysis.

The LC-MS raw data were first converted to netcdf format using Bruker software DataAnalysis 4.0. The publicly available software XCMS³⁴ was used for peak picking. The XCMS parameters were set as fwhm=3 and step=0.005, sn=1. An in-house written R program was used to find the ¹²C-/¹³C isotopic pairs based on the mass difference of 2.00671 Da of isotopic pairs and the mass accuracy tolerance of 2 ppm. The relative ion intensity of ¹²C- labeled/¹³C-labeled pairs was calculated. The redundant peaks such as natural isotopic peaks, sodium adduct peaks, potassium adduct peaks, ammonia adduct peaks, dimer peaks, doubly-charged peaks, and triply charged peaks were automatically removed by the program. We applied a threshold filter to remove noise peaks according to the background mass peak intensities. An in-house written Perl program was used to align or group the LC-MS data across the different urine samples, which was based on the mass accuracy tolerance of 5 ppm and retention time shift tolerance of 15 s. The table resulted from the Perl program contained the rows which were

the individual samples, columns which were the unique metabolites and their relatively ion intensity between the individual sample and the pooled sample (i.e., peak ratios of ^{12}C -labeled/ ^{13}C -labeled pairs). The peak ratios of an individual sample were normalized to the total relative ion intensity to account for the urine volume dilution difference.

Multivariate statistical analysis of the LC-MS data was carried out using SIMCA-P+ 11.5 (Umetrics AB, Umea, Sweden). Principal components analysis (PCA), partial least squares discriminant analysis (PLS-DA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were used to analyze the data. Twenty permutation test and CV-ANOVA, built in SIMCA-P, were used to conduct cross validation for the PLS-DA and OPLS-DA models. A list of interesting metabolites that contributed mostly to the model was generated from the VIP list. The Metaboanalyst³⁵ software was used to do the t-test for the individual metabolite.

4.2.6 Metabolite identification.

Accurate mass of an underivatized metabolite was obtained by subtracting the measured mass of the dansylation labeled metabolite to the mass of the dansyl group. Based on the accurate mass information, the Human Metabolome Database (HMDB)³⁶ and the Evidence-based Metabolome Library (EML)³⁷ were searched using MyCompoundID,³⁷ with a mass accuracy tolerance of 5 ppm, to generate a list of mass-matched putative metabolites from the ion pairs detected. For positive metabolite identification, MS/MS spectra of the significant metabolites found to be differentially expressed in two groups were generated and manually

interpreted. A metabolite was deemed to be definitively identified if the MS/MS spectrum and retention time matched with those of an authentic standard. A metabolite was considered to be putatively identified if there was no authentic standard available.

4.3 Results and Discussion

Development of automated data preprocessing for isotope labeling LC-MS.

The workflow of our relative quantitative metabolomics approach for urine metabolomic biomarker discovery in a transgenic mouse model was showed in Figure 4.1. This approach was based on a differential ^{13}C -/ ^{12}C isotope dansylation labeling LC-MS method, which was previously developed in our group. The method has great potential for untargeted metabolomics applications. However, there was no robust data preprocessing software freely available which could meet our needs for isotope labeling quantitative metabolomics. It is essential to develop automated data preprocessing method for our isotope labeling LC-FTICR-MS data. XCMS is one of the most widely used data preprocessing software for metabolomics. However, it was mainly used for label free LC-MS metabolomics and it also did not address the issue of many redundant ion peaks, which added difficulty in data interpretation. In order to apply the differential ^{13}C -/ ^{12}C isotope dansylation labeling method to large scale metabolomics studies, we developed an automated data preprocessing method, which integrated the XCMS with our in-house written R program and Perl program. The method was able to find the ^{13}C -/ ^{12}C isotope peak pairs, calculate the relative ion intensity, remove the redundant

peaks, and finally align the ion pairs across different samples. Thus, the LC-MS data from ^{13}C -/ ^{12}C isotope dansylation labeling was compatible for downstream chemometric analysis.

Mouse model.

Metabolomics has been applied to biomarker discovery in both human studies and animal studies. However, some confounding factors such as diet, environment and genetics can cause significant individual variability in human studies which complicate the biomarker discovery process.³⁸ On the other hand, animal models enable a much more controlled study and minimize the effects of confounding factors.³⁹ For studies of AD, an animal model would be expected to be particularly useful for identifying or generating a set of candidate markers that could be extended to human studies. Many transgenic mouse models of AD have been constructed based on the amyloid cascade hypothesis.^{40, 41} Amyloid beta ($\text{A}\beta$) plaque is one of the major pathophysiological features of AD. $\text{A}\beta$ is the peptide fragment cleaved from an amyloid precursor protein (APP) by secretases. It aggregates extra-cellularly to form $\text{A}\beta$ plaques, leading to neurodegeneration through a number of ill-defined neurotoxic mechanisms. Several transgenic mouse models expressing human mutant APP have been shown to develop amyloid

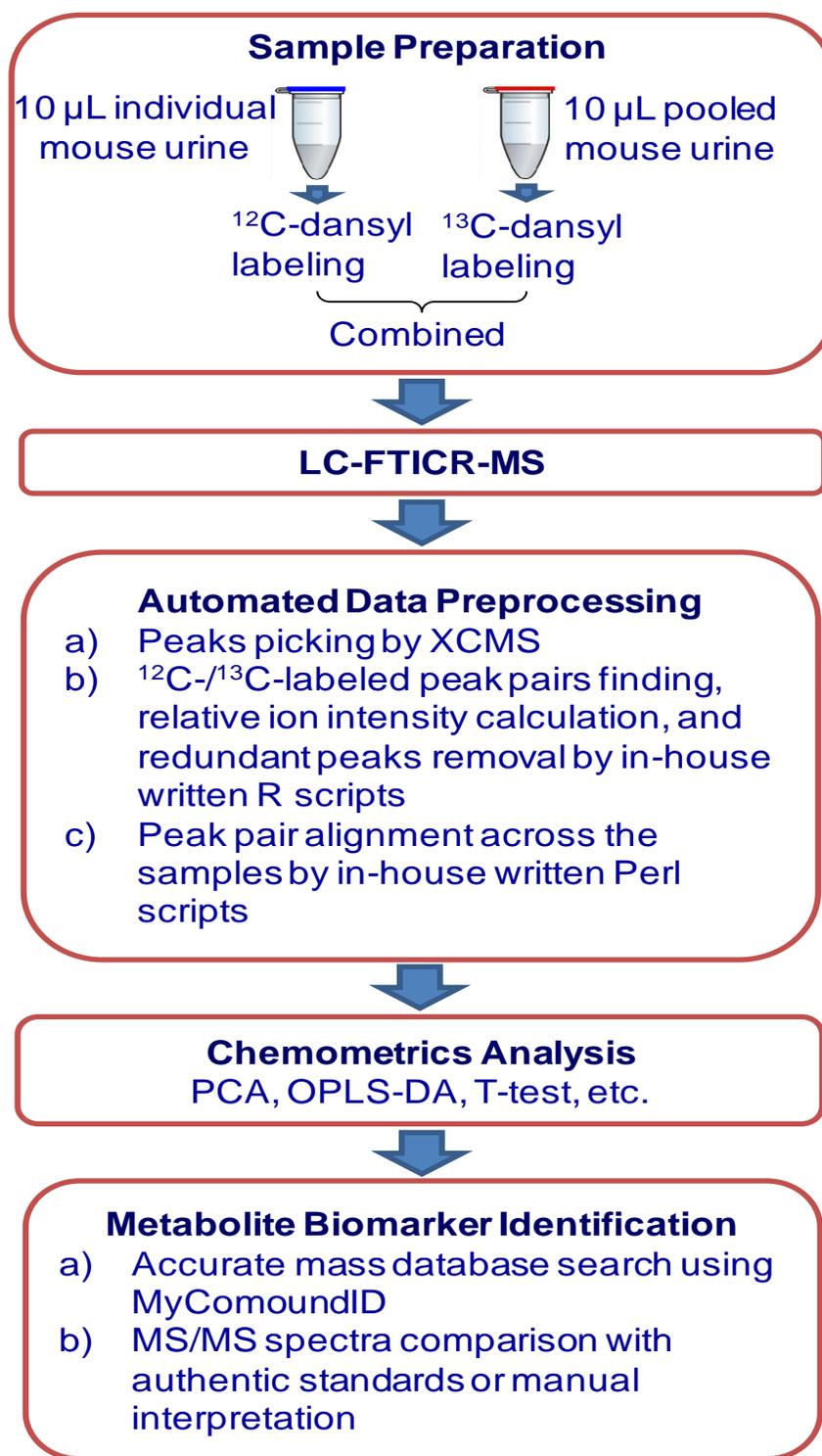


Figure 4.1 Experimental workflow of isotope labeling LC-MS for comparative metabolomics of mouse urine.

plaque pathology and dense core plaques.⁴² The TgCRND8 transgenic mouse model used in this study expresses a double mutant form of human APP 695 isoform and develops amyloid deposits as early as 2–3 months.⁴³ Therefore, the TgCRND8 transgenic mouse model allows one to perform AD biomarker discovery in a relatively short time period.

Metabolome profiling method.

The differential ^{13}C -/ ^{12}C dansylation labeling method was able to detect and quantify over 600 metabolites in human urine in a single LC-MS run.²⁷ However, unlike human urine samples, the volume of mouse urine that can be conveniently collected without significant contamination from the environment is often limited. Collection of 20 to 50 μL per animal at a given time point is possible by removing mice from their home cage and placing them into a new disposable cage. This method does not cause any contact of urine on any surface inside the animal cage and thus prevents feces, food, and other species from contaminating the sample. The freshly collected urine sample can also be frozen immediately for storage, reducing the risk of bacterial growth and degradation of metabolites.

To handle the small volume of mouse urine (e.g., 10 μL), it was necessary to optimize the LC-FTICR-MS settings in order to achieve better detection of metabolites, as well as the dansylation labeling protocol. In this work, a 1 \times 100 mm LC column, instead of a 2.1 \times 100 mm column, was used, which could theoretically improve the detection sensitivity by four-fold. In addition, the ion accumulation time for data collection was optimized to be 1 s, instead of shorter time (e.g., <0.5 s). The limit of detection for the optimized LC-FTICR-MS

method was found to be 0.24 nM for the dansylated phenylalanine. In the reported protocol,²⁷ dansylation labeling reaction requires a relatively large volume of sample, i.e., 50 μ L. It was found that merely scaling down the reaction volume from 50 to 10 μ L could cause great variations in the mass spectrometric results, likely due to inhomogeneity of the reaction solution when 10 μ L urine was used. Instead, we diluted the mouse urine from 10 μ L to 50 μ L and kept all the reagent solution volumes the same as those used in the reported protocol. The detection of metabolome coverage was not compromised even when we started from 10 μ L mouse urine sample, since we have improved the limit of detection of the LC-MS method.

In our workflow, a 20 μ L of mouse urine was initially divided into two halves, one half labeled by ¹²C-dansylation and other half combined with others to form a pooled sample that was labeled by ¹³C-dansylation. Generally speaking, we could detect more than 950 unique ion pairs or putative metabolites in the mixture of the differentially labeled mice urine sample. Many of them could match with the metabolites in HMDB and EML, based on accurate mass search using 5 ppm mass error tolerance. For example, for the 8 QC injections from the pooled sample, the average number of peak pairs found was 1082 \pm 148 (the peak pair number was lower for the 1st injection, like due to insufficient conditioning of the column). There were a total of 1454 peak pairs detected from the combined results. About 92% of them could match with the metabolites in the two libraries. It is clear that our method can provide much better metabolome coverage than a conventional LC-MS method.

Besides metabolite detectability, analytical variation of a global metabolomic profiling method should be minimized for biomarker discovery work, as a large analytical variation can mask the true biological variation between the disease state and the control state. We have evaluated the inter-day repeatability of our method using the pooled QC sample. Figure 4.2A shows the overlaid base-peak chromatograms of 8 injection of the QC sample within 4 days. These chromatograms overlay very well, and the maximum retention time shift within 4 days was less than 15 seconds. We calculated the relative standard derivation (RSD) of the peak ratios individually for the 840 metabolites that were commonly detected in the 8 injections, and the mean RSD was found to be 6.1% and the median RSD was 5.0%. Figure 4.2B shows the histogram of the frequency distribution of RSD% for the 840 metabolites. More than 96% metabolites have RSD% less than 15%, which was much better than the previously reported data using label-free LC-MS method.^{44, 45} The better inter-day repeatability of our method is mainly due to the fact that we used ¹³C-labeled pooled sample as a global internal standard, which could minimize the variations for the injection volume, electrospray response, instrumental settings, etc. We also investigated the experimental repeatability by running 3 replicates of dansylation reaction of a mouse urine sample (see Supplemental Figure S 4.1 for the overlaid base-peak ion chromatograms). The RSD% of peak ratios for all commonly detected metabolites was

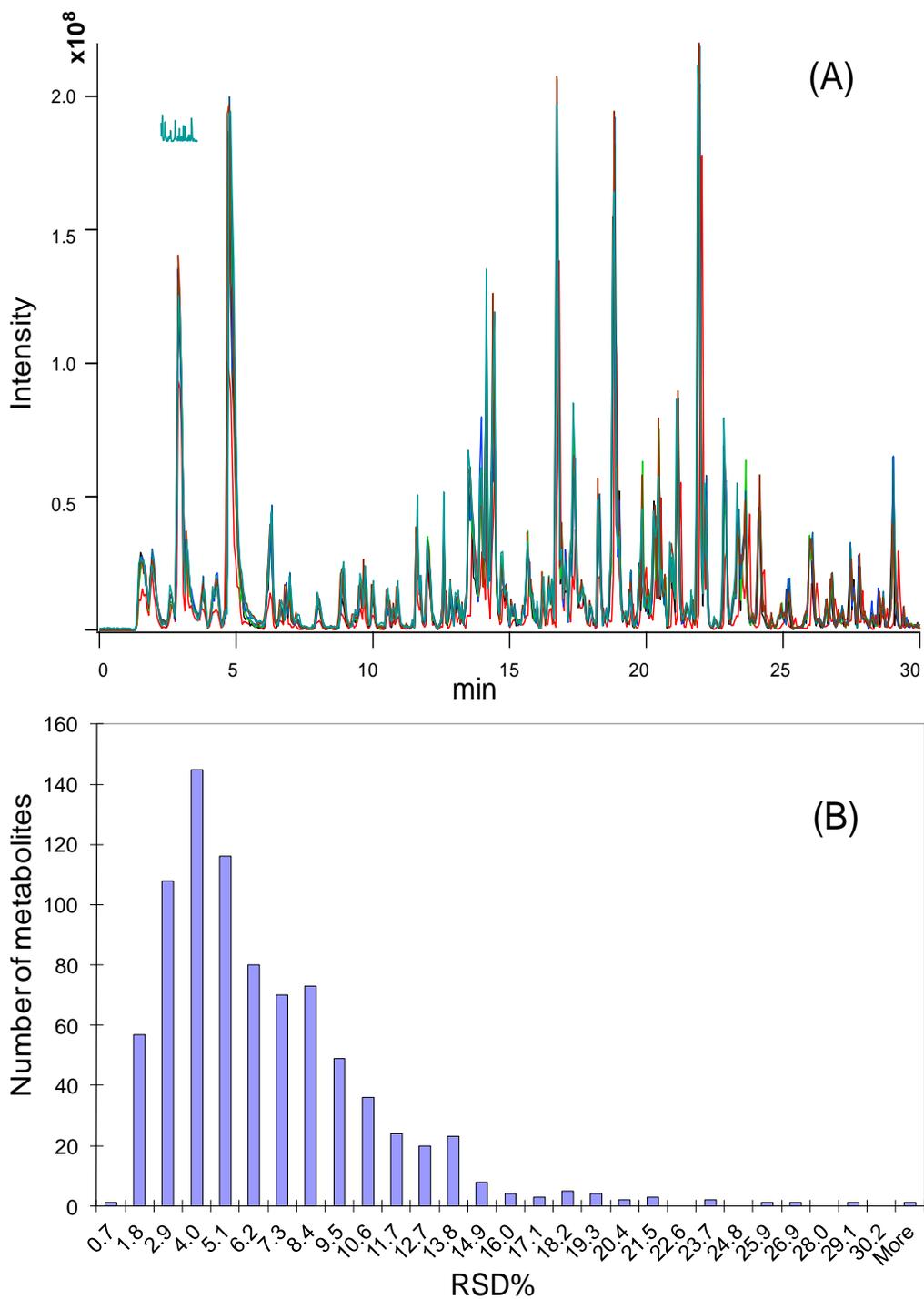


Figure 4.2 (A) Inter-day repeatability of overlaid base peak chromatograms of 8 injections of quality control (QC) samples during 4 days LC-MS running; B) the frequency histogram of RSD% relative intensity of 840 metabolites in the QC samples.

calculated. The mean RSD was 10.7%, and the median RSD was 7.4%. These results indicate that, despite the use of a small volume of urine sample (10 μ L for labeling), good reproducibility could be obtained by the isotope labeling LC-MS method.

Comparative metabolome analysis in mouse model of AD.

PCA, an unsupervised chemometric method, was used to obtain an overall picture of the whole data sets, and to see if there was any clustering, trends, or outliers. Figure 4.3A shows the PCA score plot of all the urine samples labeled as QC samples (in green), APP mutant group (in red), and wild type group (in blue). X-axis is t1 component one; Y-axis is t2 component 2. Component 1 has 20 percent of variance explained, and component 2 has 13 percent of variance explained. The 8 QC injections cluster together closely, indicating that our method was sufficiently robust that the quality of the whole LC-MS data for this study should be satisfactory.⁴⁶ Figure 4.3A shows that there is a likely trend to separation between APP mutant and wild type groups, but the separation is not obvious. If we label the samples as male and female (see Supplemental Figure S 4.2), there is a clear separation between male and female mouse urine samples, which is consistent with the previous report.⁴⁷

In order to maximum the separation between the APP mutant and wild type groups, we need to input *a priori* information about the two classes and used the supervised models, such as PLS-DA and OPLS-DA.^{48, 49} Figure 4.3B shows the PLS-DA score plot of urine samples between APP mutant and wild type groups.

There is a clear separation between APP mutant and wild type groups, although there are a few overlapping data points between the two groups. OPLS-DA removes the unrelated variation in the data sets, such as gender in this case. Figure 4.3C shows that the OPLS-DA model could discriminate the APP mutant group from the wild type group. R^2

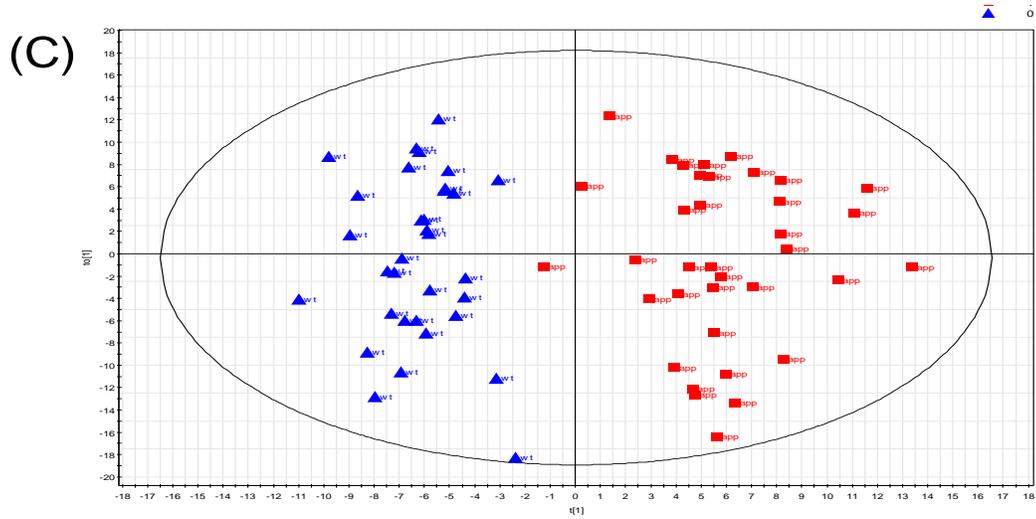
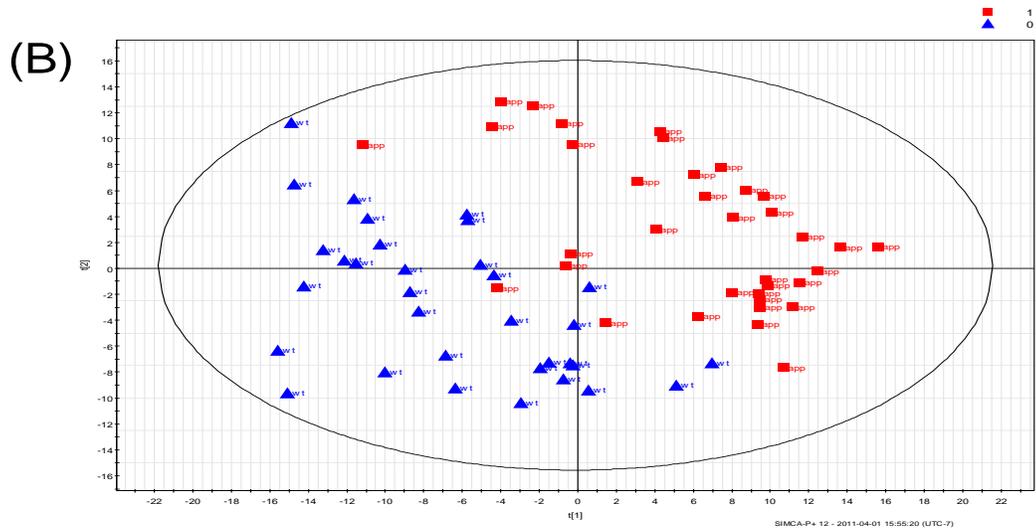
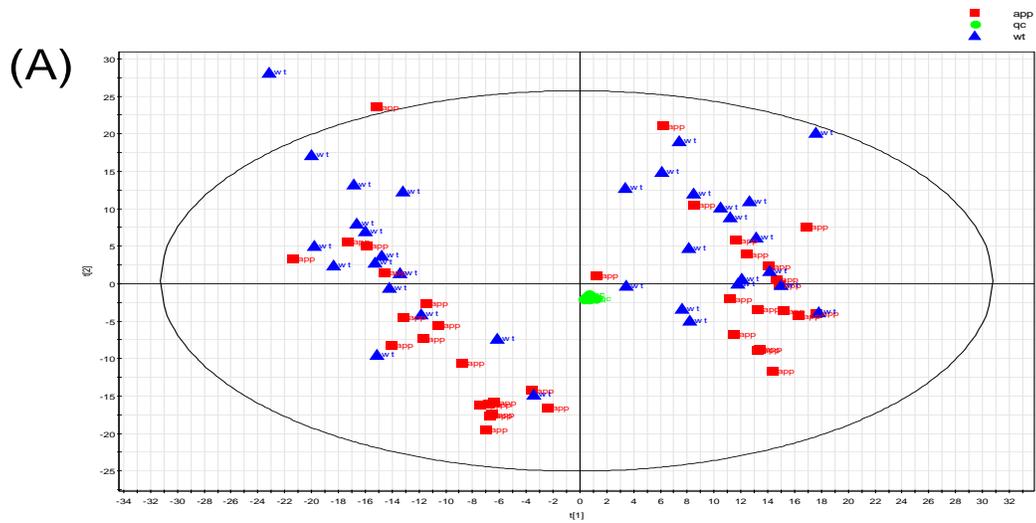


Figure 4.3 (A) PCA score plot of all the mice urine samples. Green dots were the 8 injections of the quality control sample. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. (B) PLS-DA score plot of all the mice urine samples. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. $R^2Y(\text{cum})$ was 0.719, $Q^2(\text{cum})$ was 0.516. (C) OPLS-DA score plot of all the mice urine samples. Red boxes were labeled as mutant APP mice and blue triangles were labeled as wild type mice. $R^2Y(\text{cum})$ was 0.854, $Q^2(\text{cum})$ was 0.665.

is used to evaluate the goodness of fitting, while Q^2 is used to evaluate the goodness of prediction. The OPLS-DA had R^2Y and Q^2Y values larger than 0.5, suggesting that it was a robust and predictable model ($R^2X_{cum}=0.513$, $R^2Y_{cum}=0.854$, $Q^2Y_{cum}=0.665$ for one predictive component and 3 orthogonal components). We did the cross validation for the PLS-DA and OPLS-DA models. Supplemental Figure S 4.3 shows the 20 permutation test for PLS-DA model. The slopes of both R and Q are positive and permutation data in the left are lower than the original point on the right top, which suggests that the model was valid. The CV-ANOVA, a built-in cross validation method in SIMCA-P, was used to do the cross validation of the OPLS-DA model, and it showed the model was robust (data not shown).

Since the mouse urine samples were collected at three different time points, we also divided the data into three individual time point and performed the OPLS-DA analysis. Figure 4.4 shows that the OPLS-DA can separate the APP mutant group and the wild type group in each of three time points. The separations between the APP mutant and the wild type group at 25-28 weeks and at 30-31 weeks are similar, and they both are more pronounced than that at the 15-17 weeks. This observation can be explained by the phenotype change of amyloid plaques in the APP mutant mouse brain with time. The amyloid plaques appear as early as 2-3 months and become more dense at 6 months.⁴³

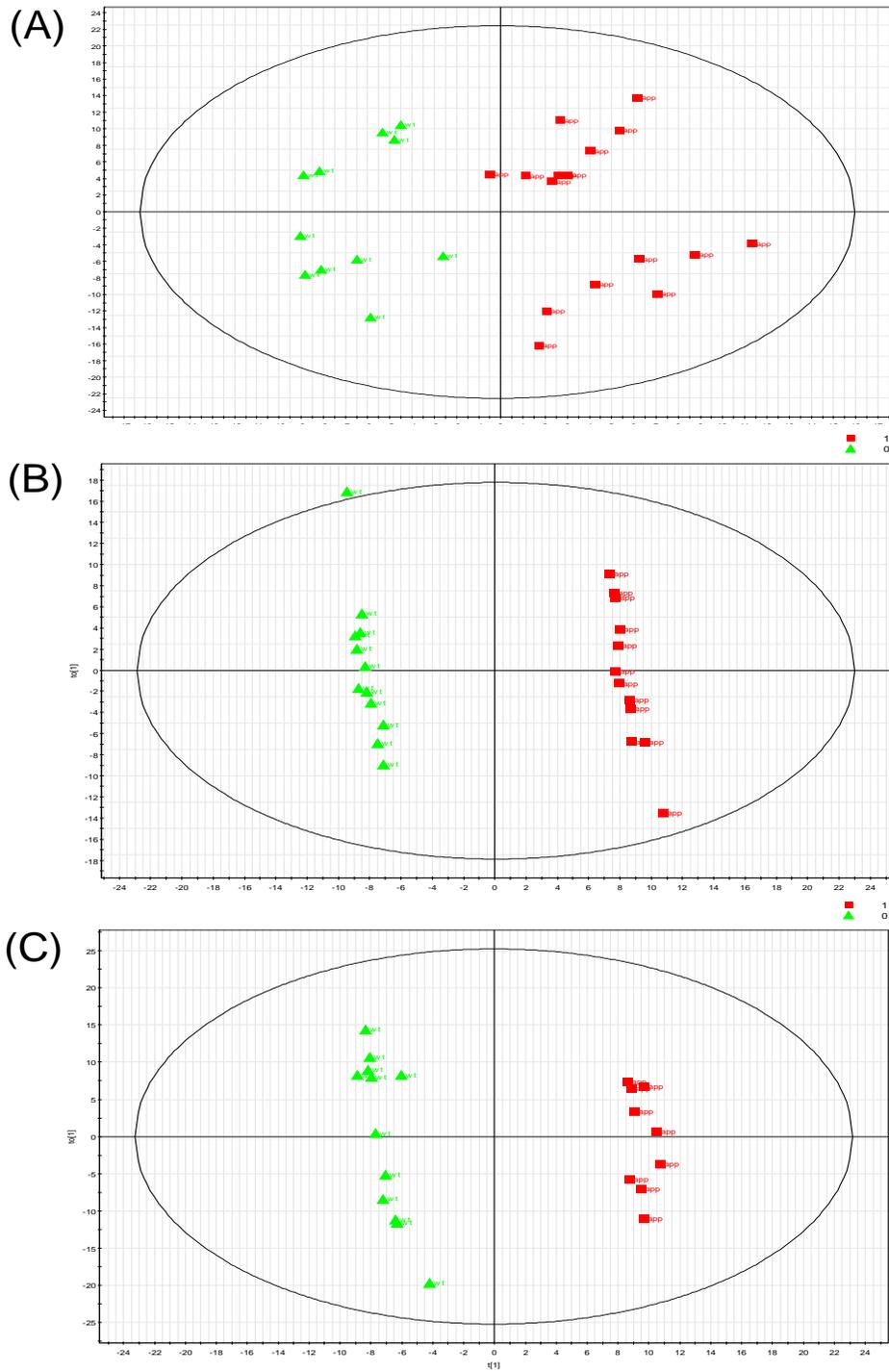


Figure 4.4 Score plots of OPLS-DA at (A) the age of 15-17 weeks, (B) the age of 25-28 weeks, and C) the age of 30-31 weeks.

Identification of candidate metabolite biomarkers.

Metabolite identification is a challenging and time-consuming process.³⁸ We focused on a relatively short list of candidate metabolites for identification. The top 15 most important metabolites were selected from the loading plot based on their variable importance in the project value and they were also validated by the student's t-test. Table 4.2 shows a list of these top 15 metabolites with their retention time, accurate mass of dansylated metabolite, accurate mass of underivatized metabolite, mass accuracy error, fold change and P-value. The use of high resolution and high mass accuracy FTICR MS permitted the use of small mass error tolerance window, 5 ppm, to perform mass search in a metabolome database and gave us a relatively short list of possible hits. Furthermore, MS/MS spectrum was used to facilitate the identification of the potential metabolites. As an example, Figure 4.5 shows the MS/MS spectra of ¹²C-dansylated and ¹³C-dansylated methionine in the mouse sample matched with those of the reference standard.

In this work, methionine, desaminotyrosine, taurine, N1-acetylspermidine, 5-hydroxyindoleacetic acid and choline were definitely identified. Dihydrouracil, lysine-valine, thiocysteine and ureidopropionic acid were putatively identified. The remaining significant metabolites could not be identified. Figure 4.6 shows the change of relative concentration level of some metabolites in the APP mutant group and the wild type group at three different ages of 15-17 weeks, 25-28 weeks, and 30-31 weeks. We can see that methionine and desaminotyrosine are upregulated and N1-acetylspermidine and 5-hydroxyindoleacetic acid are

downregulated in the APP mutant group. They show some differences among the three different ages within the

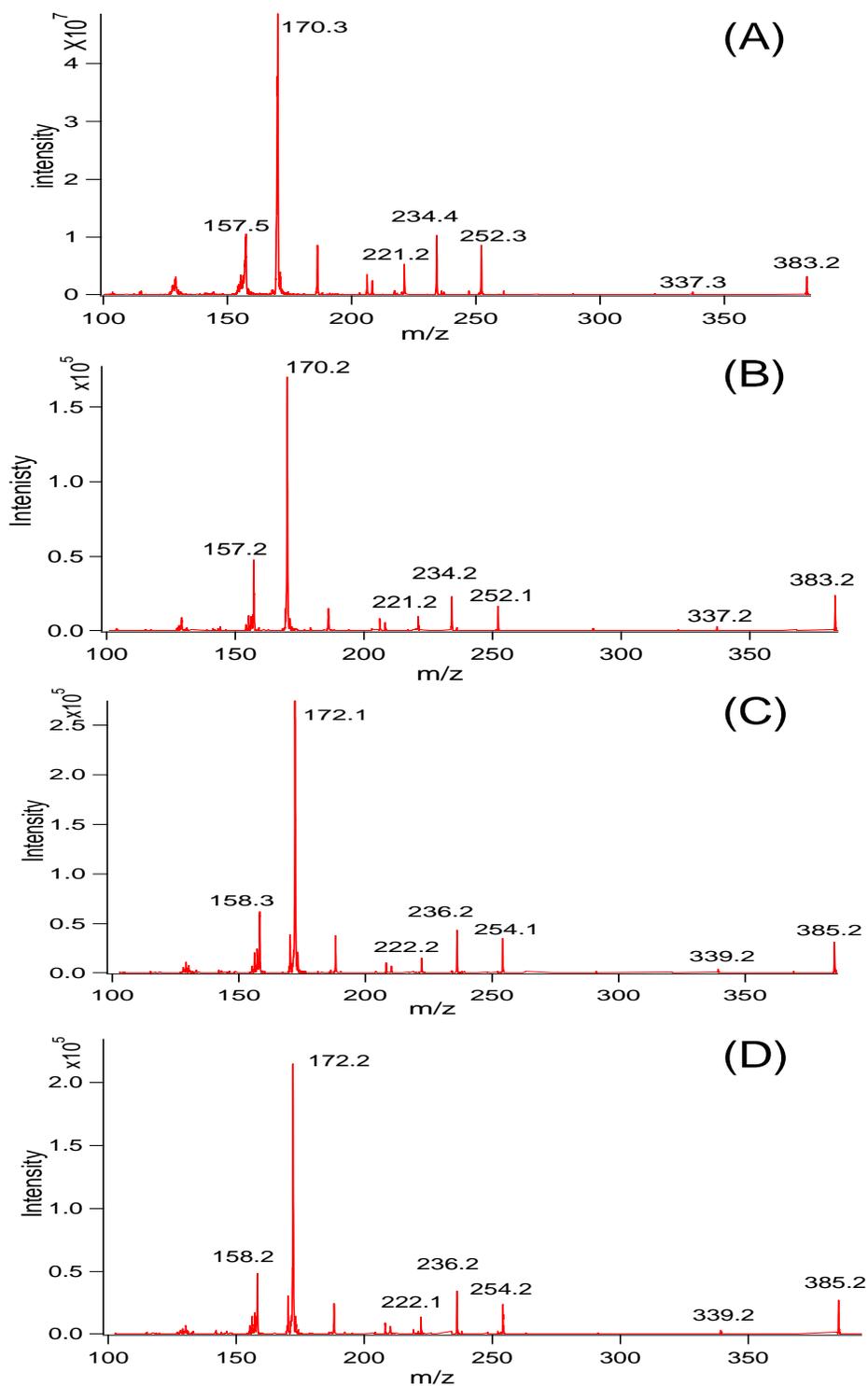


Figure 4.5 MS/MS spectra of (A) ^{12}C -dansylated methionine standard, (B) ^{12}C -dansylated methionine in the mouse urine sample, (C) ^{13}C -dansylated methionine standard, and (D) ^{13}C -dansylated methionine in the mouse urine sample.

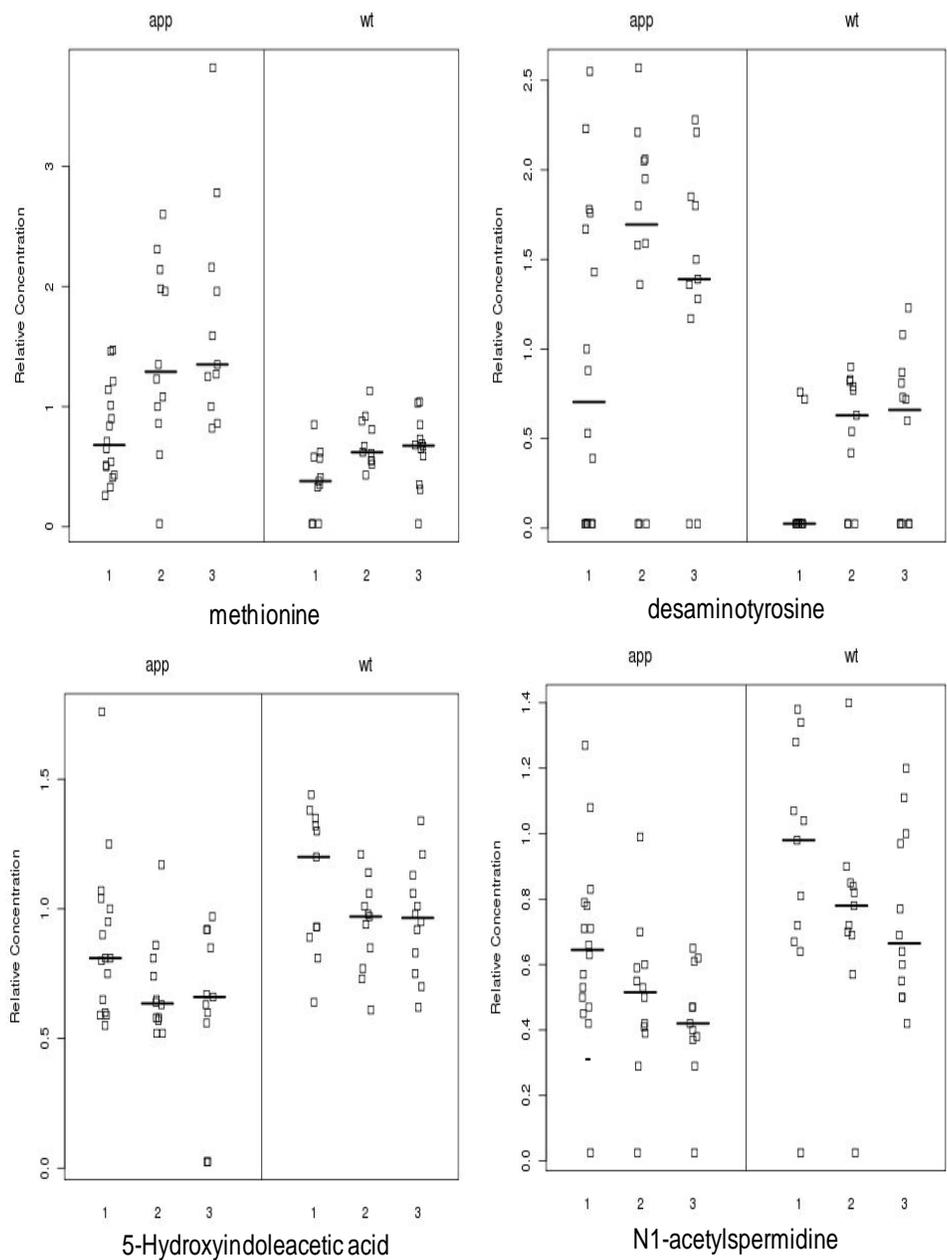


Figure 4.6 Plots showing the changed relative concentration level of candidate metabolite biomarkers in the APP mutant (app) and wild type (wt) groups at three different ages: 1 corresponds to 15-17 weeks, 2 corresponds to 25-28 weeks, and 3 corresponds to 30-31 weeks. The horizontal dark line is the median value. The p-value of the difference between the mutant APP and wild type groups at each of three ages was calculated to be <0.001 for these metabolites.

wild type group, but these metabolites show more significant differences from the APP mutant group. The differences between the APP mutant and the wild type group for all these metabolites at each three different ages are statistically significant and the P-values are all <0.001.

Significance of candidate metabolite biomarkers.

The metabolomic profile changes as shown in Figure 4.4 and the relative concentration changes of individual metabolites as shown in Figure 4.6 are significant for understanding the AD disease development. The metabolomic changes between the APP mutant group and the wild type group may be used in the future to examine the efficacy of a treatment at any time point of the disease development - a reversal of metabolomic changes to the normal control would indicate that a treatment could reverse the disease development process. In addition, if some of the metabolite biomarkers are transferable to human samples, they could potentially be served as biomarkers for disease diagnosis or monitoring therapeutic efficacy in human. The biological significances of the potential metabolite biomarkers found in this mouse model study are briefly discussed below.

Methionine is an essential amino acid containing sulfur, which mice could not synthesize. Homocysteine also containing sulfur is an intermediate product in the methionine metabolism pathway. The elevated level of blood plasma homocysteine was reported to be associated with high risk of AD.^{50, 51} A methionine-rich diet could cause high level of plasma homocysteine, which

resulted in the acceleration of brain AB accumulation in the APP mutant mouse.⁵² In methionine metabolism pathway, methionine can be converted to S-adenosylmethionine (SAM), which serves as ubiquitous methyl group donor and is necessary for the synthesis of neurotransmitters, neuronal membrane stability, and DNA methylation. Decreased level of SAM in CSF was reported to be associated with the AD patients.⁵³ Methionine is also required for synthesis of cysteine and the sulfur atom from methionine is transferred to cysteine. We also observed that a putatively identified metabolite thiocysteine were up-regulated in the mutant APP mice. Thiocysteine was an intermediate product in the cysteine metabolism pathway. In this study we observed that the methionine and cysteine metabolism was disturbed, which provided evidence that methionine and cysteine metabolism pathway warrant further investigation for AD transgenic mice model. Interestingly, one recent metabolomics study in human CSF also showed that methionine level was significantly increased in AD patients.¹⁷ These results suggest that the methionine related pathway may play an important role in the pathophysiology of AD.

Taurine is also a sulfur amino acid, which can be synthesized by human body from cysteine. Taurine has many diverse biological functions serving as a neurotransmitter in the brain, a stabilizer of cell membranes and a facilitator in the transport of ions. Recently it was reported that a relative decrease in the concentration of taurine in the frontal cortex and midbrain of the TgCRND8 APP 695 transgenic mouse, compared to the control.²⁰ Another recent study also reported taurine level changes in the Tg2576 AD in a mouse model.³³ In a study

of human saliva metabolomics, it was found that taurine level was reduced in patients with mild cognitive impairment (MCI), a disease related to AD.³⁰ Consistent with previous studies, in our study we observed a decreased level of taurine in the mutant APP mouse urine samples.

The relative concentration of desaminotyrosine was increased in mutant APP mice. It is one of the phenolic acid metabolites of tyrosine by tyrosine aminotransferase.⁵⁴ There is very little literature information on the biological function of this metabolite. However, the tyrosine metabolism pathway is important and some well-known neurotransmitters, such as L-Dopa, dopamine, are derivatized from tyrosine. Further investigation of desaminotyrosine is needed to find out what role it plays in the metabolite pathway related to AD disease.

N1-acetylspermidine is one of aliphatic polyamines occurring ubiquitously in organisms. Polyamines have important functions in the stabilization of cell membranes, biosynthesis of informing molecules, cell growth and differentiation. Reduced concentration of N1-acetylspermidine was previously reported in the urine of AD patients by traditional biochemical study method.⁵⁵ In our work, a reduced level of N1-acetylspermidine in the transgenic mouse model of AD was observed.

5-hydroxyindoleacetic acid is a metabolite of serotonin which is one of the important neurotransmitters in the brain. Neurotransmitters react with receptors in the postsynaptic membrane, thereby activating the neuron and transmitting the signal. The concentration changes of several neurotransmitters including serotonin were implicated in AD.⁵⁶ Although there was no change observed in the human

brain tissue, decreased level of 5-hydroxyindoleacetic acid in human CSF samples was reported to be associated with AD by traditional biochemistry study.^{57, 58} One recent metabolomics study also showed that 5-hydroxyindoleacetic acid was dysregulated in human CSF samples of AD patients.¹³ In this work, we observed that 5-hydroxyindoleacetic acid level was changed in the APP mutant urine samples, which provides further evidence that serotonin metabolic pathway is likely to be associate with the pathology of AD.

Choline is an important metabolite since it serves as a precursor of acetylcholine, as a methyl donor in various metabolic processes. We found that the level change of choline may be involved in the AD. Our result was also in agreement with another recent metabolomics study in human CSF samples, in which they identified choline as a potential disease progress biomarker.²⁶

The findings of these potentially biological relevant metabolites, as discussed above, demonstrated that our metabolomic profiling method could be useful in untargeted metabolomic biomarker discovery. In addition, some candidate metabolites, such as methionine, 5-hydroxyindoleacetic acid, choline, and taurine, identified in this mouse model study are in agreement with previous metabolomics studies in human CSF and blood samples. Although urine is not in close proximity to brain tissue, it might reflect the systems body response to AD. Urine metabolomics may have potential to be a non-invasive diagnostic tool for AD. Furthermore, urine metabolomics may also be used to assess the therapeutic intervention for the management and treatment of AD, particularly at the early stage of the disease development.

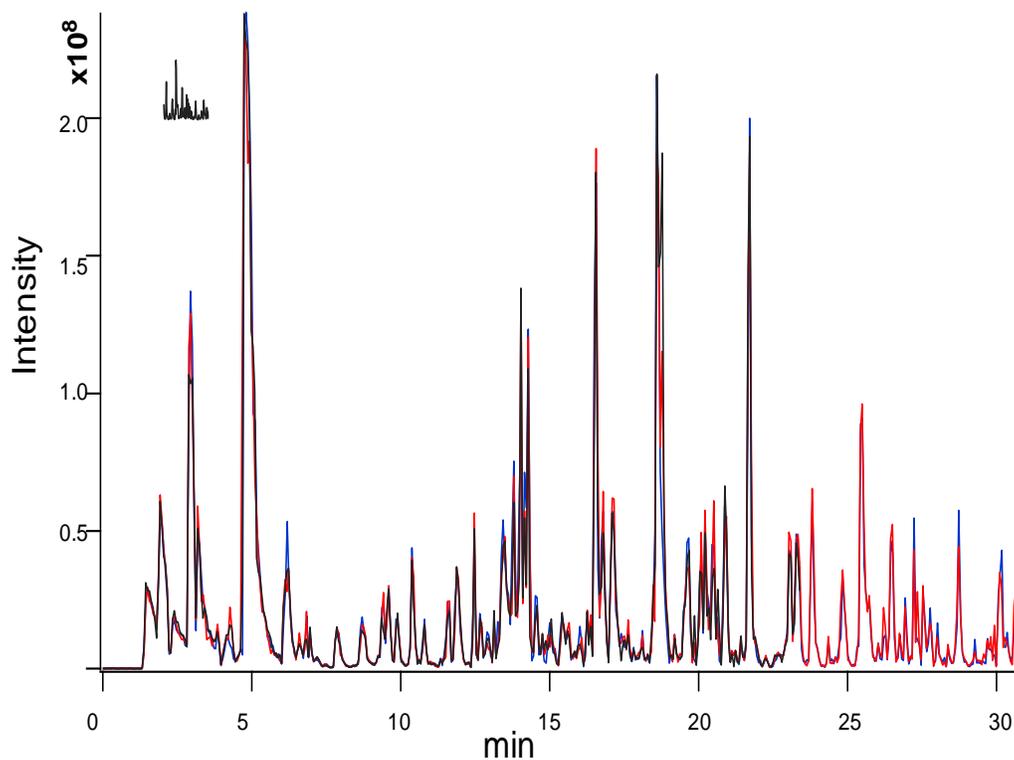
4.4 Conclusions

We have developed a new LC-MS method for analyzing the mouse urine metabolome with much improved metabolome coverage. This method allowed us to detect and quantify more than 950 metabolites in mouse urine samples using a starting material of 20 μ L. From 8 LC-MS runs of the quality control sample prepared from a pooled urine sample, a total of 1454 peak pairs could be detected and 92% of them could match with the metabolites in either the HMDB database or the EML library based on accurate mass search. Urinary metabolomic study on a transgenic mice model of AD showed that metabolic profiles could differentiate the mutant APP from the wild type mice. Several metabolite candidate biomarkers identified in this work were consistent with previous metabolomics study in human CSF. Our study indicated that complementary to CSF, brain, and plasma, urine has the potential to be used to search for non-invasive, inexpensive, sensitive metabolic biomarkers for AD disease diagnosis and also the therapeutic effect monitoring.

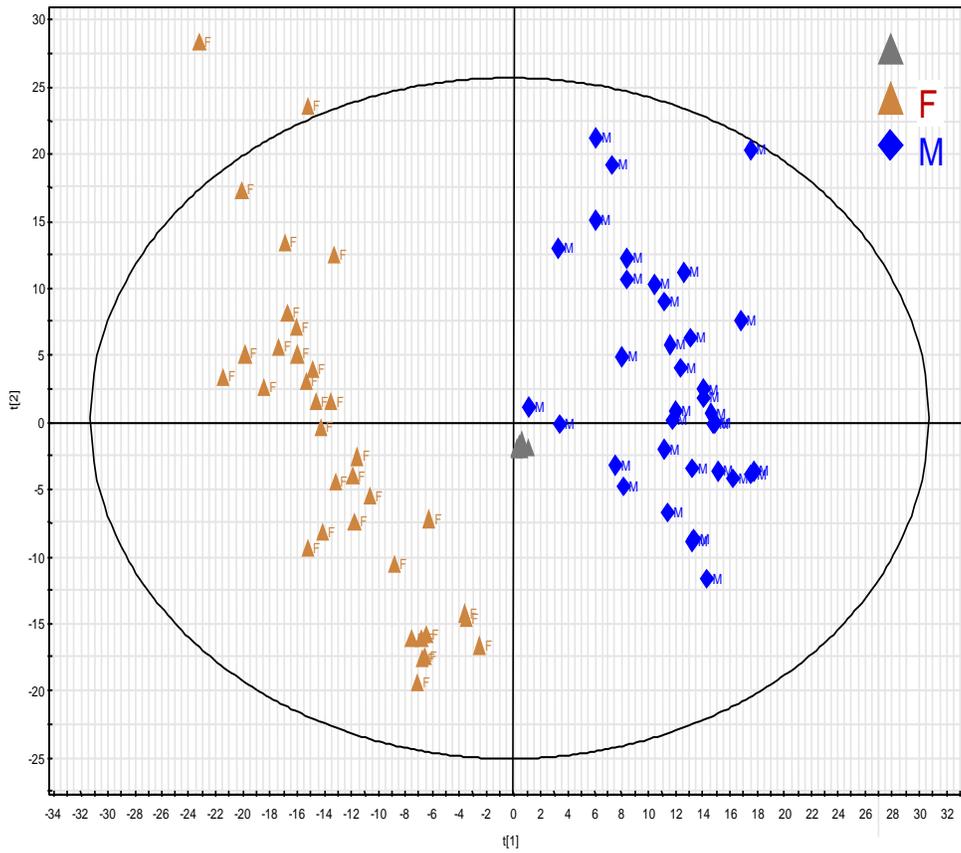
One of limitations in this study is that dansylation labeling mainly enhanced the detection of the compounds containing primary, secondary amines and phenol groups, but it could not detect some other class of compounds like glucose, fatty acids, which might be interesting to AD. This limitation can be overcome using other isotopic labeling reagents. For example, isotope-coded *p*-dimethylaminophenacyl bromide can be used to detect compounds containing carboxylic groups.⁵⁹ We plan to apply these newly developed labeling methods to

comprehensively study the mouse urine metabolome to obtain the enhanced picture of metabolic perturbations in the transgenic mouse of AD for better understanding of the disease and searching for more specific and sensitive biomarkers for early diagnosis of AD.

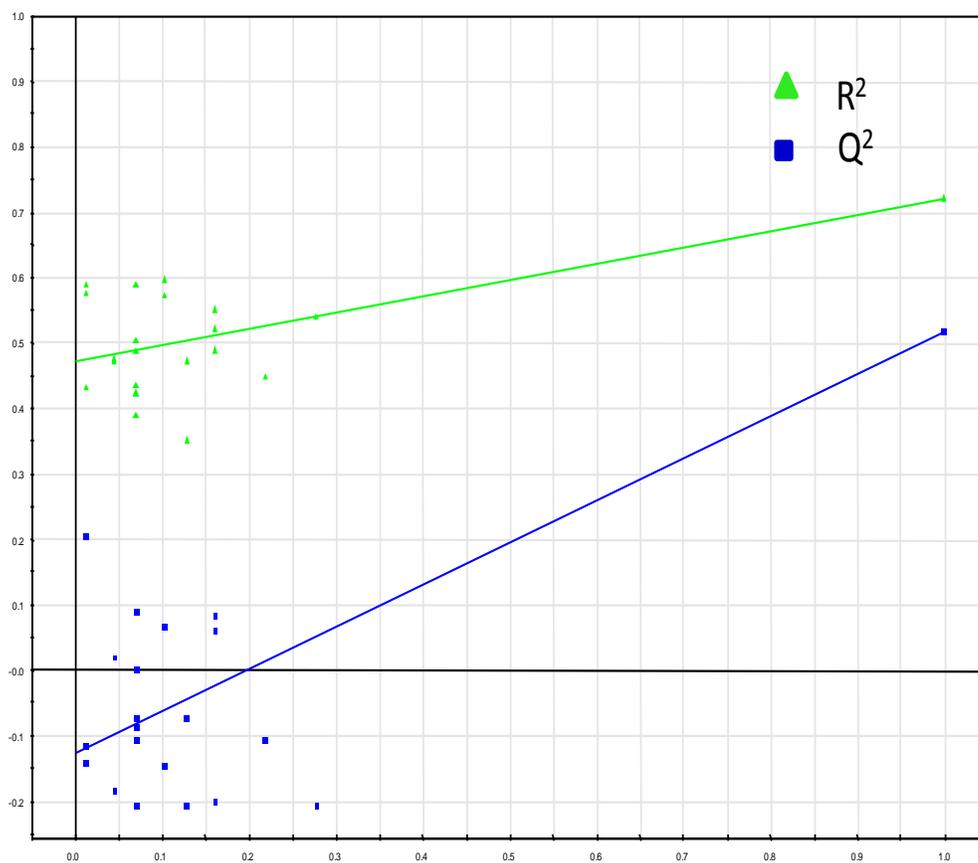
- On average, 950 pairs were detected
- Repeatability (RSD %): 10.7% mean; 7.4% median



Supplemental Figure S 4.1 Overlaid base-peak ion chromatograms of three experimental replicates of a mixture of ¹²C-dansylated mouse urine and ¹³C-dansylated mouse urine (1:1, v/v).



Supplemental Figure S 4.2 Score plot of PCA showing the separation between the male mice and the female mice.



Supplemental Figure S 4.3 Validation of the PLS-DA model using 20-permutation test built in the SIMCA-P⁺ program.

Table 4.1. Information on the mouse urine samples used in this study.

Total number of mice	24		
number of mice (male and female)		12 male	12 female
number of mice (APP mutant and wild type)		12 APP	12 WT
Total number of samples (APP mutant and WT)	75	39 APP	36 WT
at the ages of 15-17 weeks	28	15 APP	13 WT
at the ages of 25-28 weeks	23	12 APP	11 WT
at the ages of 30-31 weeks	24	12 APP	12 WT

Table 4.2 List of top 15 important metabolites found in mouse urine that differentiated the APP mutant group and the wild type group.

Ret. Time (min)	Mass of dansylated metabolite	Mass of metabolite	Metabolite	Mass accuracy (ppm)	ID *	Fold change	T-test (p-value)
21.4						2.20	10e-5.6
8	400.12156	166.06323	Desaminotyrosine	1.4	D		
12.4						-2.00	0.00033
7	337.15813	103.09980	Choline	0.9	D		
14.3						1.81	0.0028
3	383.10911	149.05078	L-Methionine	1.8	D		
18.7			5-Hydroxyindoleacetic acid			-1.80	0.0010
1	425.11673	191.05840		0.8	D		
2.91	359.07268	125.01435	Taurine	2.5	D	-1.70	0.0090
20.9						-1.50	0.0014
8	327.64260	187.16853	N1-Acetylspermidine	0.4	D		
27.9	330.59553	193.07439	Phenylacetyl glycine or methylhippuric acid	2.5	P	2.20	0.00045
0							
16.8						2.19	10e-6.0
7	348.10153	114.04320	Dihydrouracil	2.3	P		
12.2						1.94	10e-5.5
2	366.11202	132.05369	Ureidopropionic acid	1.4	P		
15.6						1.82	0.0050
7	401.06614	167.00781	Thiocysteine	1.9	P		
26.7	317.58747	167.05828	Hydroxyphenylglycine or pyridoxal	0.2	P	-1.60	0.00027
1							
6.66	479.23249	245.17416	Lysine-valine	1.1	P	-1.55	0.00034
28.1	341.06009	214.00353	Unknown			2.35	0.0014
4							
26.4	309.09803	75.03970	Unknown			2.07	0.00012
7							
9.42	362.11708	128.05875	Unknown			2.00	0.00015

*D=definitely identified; P=putatively identified.

Supplemental Table T 4.1. Summary of HMDB and EML database matches for 8 individual QC injections and the combined results.

sample number	QC1	QC2	QC3	QC4	QC5	QC6	QC7	QC8	mean	std dev	CV%
number of ion pairs	751	1027	1085	1079	1131	1196	1175	1209	1082	148	14.4%
HMDB matches	438	592	617	610	643	683	659	673	614	78	12.7%
EML matches	242	348	379	383	407	431	432	446	384	66	17.2%
total matches	680	940	996	993	1050	1114	1091	1119	998	143	14.4%
percentage of total matches	90.5%	91.5%	91.8%	92.0%	92.8%	93.1%	92.8%	92.6%	92.1%	0.86%	0.9%

References

1. Bishop, N. A.; Lu, T.; Yankner, B. A., Neural mechanisms of ageing and cognitive decline. *Nature* **2010**, 464, (7288), 529-535.
2. Lovestone, S., Searching for biomarkers in neurodegeneration. *Nat. Med.* **2010**, 16, (12), 1371-1372.
3. Gupta, V. B.; Sundaram, R.; Martins, R. N., Multiplex biomarkers in blood. *Alzheimers Res. Ther.* **2013**, 5, (3), 6.
4. Grossman, M., Multimodal Comparative Studies of Neurodegenerative Diseases. *J. Alzheimers Dis.* **2013**, 33, S379-S383.
5. Kaddurah-Daouk, R.; Krishnan, K. R. R., Metabolomics: A global biochemical approach to the study of central nervous system diseases. *Neuropsychopharmacology* **2009**, 34, (1), 173-186.
6. Quinones, M. P.; Kaddurah-Daouk, R., Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. *Neurobiol. Dis.* **2009**, 35, (2), 165-176.
7. Mishur, R. J.; Rea, S. L., Applications of mass spectrometry to metabolomics and metabonomics: Detection of biomarkers of aging and of age-related diseases. *Mass Spectrom. Rev.* **2012**, 31, (1), 70-95.
8. Bierer, L. M.; Haroutunian, V.; Gabriel, S.; Knott, P. J.; Carlin, L. S.; Purohit, D. P.; Perl, D. P.; Schmeidler, J.; Kanof, P.; Davis, K. L., Neurochemical correlates of dementia severity in Alzheimers disease - relative importance of the cholinergic deficits. *J. Neurochem.* **1995**, 64, (2), 749-760.
9. Kapogiannis, D.; Mattson, M. P., Disrupted energy metabolism and neuronal circuit dysfunction in cognitive impairment and Alzheimer's disease. *Lancet Neurol.* **2011**, 10, (2), 187-198.

10. Butterfield, D. A., beta-Amyloid-associated free radical oxidative stress and neurotoxicity: Implications for Alzheimer's disease. *Chem. Res. Toxicol.* **1997**, 10, (5), 495-506.
11. Manczak, M.; Anekonda, T. S.; Henson, E.; Park, B. S.; Quinn, J.; Reddy, P. H., Mitochondria are a direct site of A beta accumulation in Alzheimer's disease neurons: implications for free radical generation and oxidative damage in disease progression. *Hum. Mol. Genet.* **2006**, 15, (9), 1437-1449.
12. Kaddurah-Daouk, R.; Rozena, S.; Matson, W.; Han, X.; Hulette, C. M.; Burke, J. R.; Doraiswamy, P. M.; Welsh-Bohmer, K. A., Metabolomic changes in autopsy-confirmed Alzheimer's disease. *Alzheimers Dement.* **2010**, 1-9.
13. Kaddurah-Daouk, R.; Zhu, H.; Sharma, S.; Bogdanov, M.; Rozen, S. G.; Matson, W.; Oki, N. O.; Motsinger-Reif, A. A.; Churchill, E.; Lei, Z.; Appleby, D.; Kling, M. A.; Trojanowski, J. Q.; Doraiswamy, P. M.; Arnold, S. E.; Pharmacometabolomics Research Network, Alterations in metabolic pathways and networks in Alzheimer's disease. *Transl. Psychiatry* **2013**, 3, 8.
14. Ibanez, C.; Simo, C.; Barupal, D. K.; Fiehn, O.; Kivipelto, M.; Cedazo-Minguez, A.; Cifuentes, A., A new metabolomic workflow for early detection of Alzheimer's disease. *J. Chromatogr. A* **2013**, 1302, 65-71.
15. Trushina, E.; Dutta, T.; Persson, X. M. T.; Mielke, M. M.; Petersen, R. C., Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's Disease using metabolomics. *PLoS One* **2013**, 8, (6), 13.
16. Greenberg, N.; Grassano, A.; Thambisetty, M.; Lovestone, S.; Legido-Quigley, C., A proposed metabolic strategy for monitoring disease progression in Alzheimer's disease. *Electrophoresis* **2009**, 30, (7), 1235-1239.
17. Oresic, M.; Hyotylainen, T.; Herukka, S. K.; Sysi-Aho, M.; Mattila, I.; Seppanan-Laakso, T.; Julkunen, V.; Gopalacharyulu, P. V.; Hallikainen, M.; Koikkalainen, J.; Kivipelto, M.; Helisalmi, S.; Lotjonen, J.; Soininen, H.,

Metabolome in progression to Alzheimer's disease. *Transl. Psychiatry* **2011**, 1, 9.

18. Li, N. J.; Liu, W. T.; Li, W.; Li, S. Q.; Chen, X. H.; Bi, K. S.; He, P., Plasma metabolic profiling of Alzheimer's disease by liquid chromatography/mass spectrometry. *Clin. Biochem.* **2010**, 43, (12), 992-997.

19. Graham, S. F.; Chevallier, O. P.; Roberts, D.; Holscher, C.; Elliott, C. T.; Green, B. D., Investigation of the human brain metabolome to identify potential markers for early diagnosis and therapeutic targets of Alzheimer's disease. *Anal. Chem.* **2013**, 85, (3), 1803-1811.

20. Salek, R. M.; Xia, J.; Innes, A.; Sweatman, B. C.; Adalbert, R.; Randle, S.; McGowan, E.; Emson, P. C.; Griffin, J. L., A metabolomic study of the CRND8 transgenic mouse model of Alzheimer's disease. *Neurochem. Int.* **2010**, 56, (8), 937-947.

21. Lin, S. H.; Liu, H. D.; Kanawati, B.; Liu, L. F.; Dong, J. Y.; Li, M.; Huang, J. D.; Schmitt-Kopplin, P.; Cai, Z. W., Hippocampal metabolomics using ultrahigh-resolution mass spectrometry reveals neuroinflammation from Alzheimer's disease in CRND8 mice. *Anal. Bioanal. Chem.* **2013**, 405, (15), 5105-5117.

22. Hurko, O.; Boudonck, K.; Gonzales, C.; Hughes, Z. A.; Jacobsen, J. S.; Reinhart, P. H.; Crowther, D., Ablation of the locus coeruleus increases oxidative stress in tg-2576 transgenic but not wild-type mice. *Int. J. Alzheimer's Dis.* **2010**, 2010, 864625.

23. Hu, L. P.; Browne, E. R.; Liu, T.; Angel, T. E.; Ho, P. C.; Chan, E. C. Y., Metabonomic profiling of TASTPM transgenic Alzheimer's disease mouse model. *J. Proteome Res.* **2012**, 11, (12), 5903-5913.

24. Graham, S. F.; Holscher, C.; McClean, P.; Elliott, C. T.; Green, B. D., H-1 NMR metabolomics investigation of an Alzheimer's disease (AD) mouse model pinpoints important biochemical disturbances in brain and plasma. *Metabolomics* **2013**, 9, (5), 974-983.

25. Hurko, O.; Boudonck, K.; Gonzales, C.; Hughes, Z. A.; Jacobsen, J. S.; Reinhart, P. H.; Crowther, D., Ablation of the locus coeruleus increases oxidative stress in tg-2576 transgenic but not wild-type mice. *Int. J. Alzheimer's Dis.* **2011**, 2010, 864625.
26. Ibanez, C.; Simo, C.; Martin-Alvarez, P. J.; Kivipelto, M.; Winblad, B.; Cedazo-Minguez, A.; Cifuentes, A., Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry metabolomics. *Anal. Chem.* **2012**, 84, (20), 8532-8540.
27. Guo, K.; Li, L., Differential C-12/C-13-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal. Chem.* **2009**, 81, (10), 3919-3932.
28. Guo, K.; Bamforth, F.; Li, L. A., Qualitative metabolome analysis of human cerebrospinal fluid by C-13-/C-12-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011**, 22, (2), 339-347.
29. Wu, Y.; Li, L., Determination of total concentration of chemically labeled metabolites as a means of metabolome sample normalization and sample loading optimization in mass spectrometry-based metabolomics. *Anal. Chem.* **2012**, 84, (24), 10723-10731.
30. Zheng, J. M.; Dixon, R. A.; Li, L., Development of isotope labeling LC-MS for human salivary metabolomics and application to profiling metabolome changes associated with mild cognitive impairment. *Anal. Chem.* **2012**, 84, (24), 10802-10811.
31. Wu, Y.; Li, L., Development of isotope labeling liquid chromatography-mass spectrometry for metabolic profiling of bacterial cells and its application for bacterial differentiation. *Anal. Chem.* **2013**, 85, (12), 5755-63.
32. Fu, F.; Cheng, V. W. T.; Wu, Y.; Tang, Y.; Weiner, J. H.; Li, L., Comparative proteomic and metabolomic analysis of staphylococcus warneri

SG1 cultured in the presence and absence of butanol. *J. Proteome Res.* **2013**, 12, (10), 4478-89.

33. Fukuhara, K.; Ohno, A.; Ota, Y.; Senoo, Y.; Maekawa, K.; Okuda, H.; Kurihara, M.; Okuno, A.; Niida, S.; Saito, Y.; Takikawa, O., NMR-based metabolomics of urine in a mouse model of Alzheimer's disease: identification of oxidative stress biomarkers. *J. Clin. Biochem. Nutr.* **2013**, 52, (2), 133-8.

34. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, 78, (3), 779-787.

35. Xia, J. G.; Psychogios, N.; Young, N.; Wishart, D. S., MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, 37, W652-W660.

36. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazzyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, 37, D603-D610.

37. Li, L.; Li, R. H.; Zhou, J. J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y. M.; Huan, T.; Zheng, J. M.; Shi, Y.; Wishart, D. S.; Lin, G. H., MyCompoundID: using an evidence-based metabolome library for metabolite identification. *Anal. Chem.* **2013**, 85, (6), 3401-3408.

38. Baker, M., Metabolomics: from small molecules to big ideas. *Nat. Methods* **2011**, 8, (2), 117-121.

39. Barba, I.; Fernandez-Montesinos, R.; Garcia-Dorado, D.; Pozo, D., Alzheimer's disease beyond the genomic era: nuclear magnetic resonance

(NMR) spectroscopy-based metabolomics. *J. Cell. Mol. Med.* **2008**, 12, (5A), 1477-1485.

40. Janus, C.; Chishti, M. A.; Westaway, D., Transgenic mouse models of Alzheimer's disease. *BBA-Mol. Basis Dis.* **2000**, 1502, (1), 63-75.

41. Kokjohn, T. A.; Roher, A. E., Amyloid precursor protein transgenic mouse models and Alzheimer's disease: Understanding the paradigms, limitations, and contributions. *Alzheimers Dement.* **2009**, 5, (4), 340-347.

42. Phinney, A. L.; Horne, P.; Yang, J.; Janus, C.; Bergeron, C.; Westaway, D., Mouse models of Alzheimer's disease: The long and filamentous road. *Neurol. Res.* **2003**, 25, (6), 590-600.

43. Chishti, M. A.; Yang, D. S.; Janus, C.; Phinney, A. L.; Horne, P.; Pearson, J.; Strome, R.; Zuker, N.; Loukides, J.; French, J.; Turner, S.; Lozza, G.; Grilli, M.; Kunicki, S.; Morissette, C.; Paquette, J.; Gervais, F.; Bergeron, C.; Fraser, P. E.; Carlson, G. A.; St George-Hyslop, P.; Westaway, D., Early-onset amyloid deposition and cognitive deficits in transgenic mice expressing a double mutant form of amyloid precursor protein 695. *J. Biol. Chem.* **2001**, 276, (24), 21562-21570.

44. Crews, B.; Wikoff, W. R.; Patti, G. J.; Woo, H. K.; Kalisiak, E.; Heideker, J.; Siuzdak, G., Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. *Anal. Chem.* **2009**, 81, (20), 8538-8544.

45. Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D., Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2008**, 871, (2), 299-305.

46. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **2010**, 5, (6), 1005-1018.

47. Plumb, R.; Granger, J.; Stumpf, C.; Wilson, I. D.; Evans, J. A.; Lenz, E. M., Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* **2003**, 128, (7), 819-823.
48. Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J. Proteome Res.* **2007**, 6, (2), 469-479.
49. Madsen, R.; Lundstedt, T.; Trygg, J., Chemometrics in metabolomics-A review in human disease diagnosis. *Anal. Chim. Acta* **2010**, 659, (1-2), 23-33.
50. Popp, J.; Lewczuk, P.; Linnebank, M.; Cvetanovska, G.; Smulders, Y.; Kolsch, H.; Frommann, I.; Kornhuber, J.; Maier, W.; Jessen, F., Homocysteine metabolism and cerebrospinal fluid markers for Alzheimer's disease. *J. Alzheimers Dis.* **2009**, 18, (4), 819-828.
51. McCampbell, A.; Wessner, K.; Marlatt, M. W.; Wolffe, C.; Toolan, D.; Podtelezhnikov, A.; Yeh, S.; Zhang, R.; Szczerba, P.; Tanis, K. Q.; Majercak, J.; Ray, W. J.; Savage, M., Induction of Alzheimer's-like changes in brain of mice expressing mutant APP fed excess methionine. *J. Neurochem.* **2011**, 116, (1), 82-92.
52. Zhuo, J. M.; Pratico, D., Normalization of hyperhomocysteinemia improves cognitive deficits and ameliorates brain amyloidosis of a transgenic mouse model of Alzheimer's disease. *FASEB J.* **2010**, 24, (10), 3895-3902.
53. Linnebank, M.; Popp, J.; Smulders, Y.; Smith, D.; Semmler, A.; Farkas, M.; Kulic, L.; Cvetanovska, G.; Blom, H.; Stoffel-Wagner, B.; Kolsch, H.; Weller, M.; Jessen, F., S-Adenosylmethionine Is Decreased in the Cerebrospinal Fluid of Patients with Alzheimer's Disease. *Neurodegener. Dis.* **2009**, 7, (6), 373-378.
54. Booth, A. N.; Masri, M. S.; Robbins, D. J.; Emerson, O. H.; Jones, F. T.; Deeds, F., Urinary phenolic acid metabolites of tyrosine. *J. Biol. Chem.* **1960**, 235, (9), 2649-2652.

55. Paik, M. J.; Lee, S.; Cho, K. H.; Kim, K. R., Urinary polyamines and N-acetylated polyamines in four patients with Alzheimer's disease as their N-ethoxycarbonyl-N-pentafluoropropionyl derivatives by gas chromatography-mass spectrometry in selected ion monitoring mode. *Anal. Chim. Acta* **2006**, 576, (1), 55-60.
56. Keverne, J.; Ray, M., Neurochemistry of Alzheimer's disease. *Psychiatry* **2005**, 4, (1), 40-42.
57. Lopez, O. L.; Kaufer, D.; Reiter, C. T.; Carra, J.; DeKosky, S. T.; Palmer, A. M., Relationship between CSF neurotransmitter metabolites and aggressive behavior in Alzheimer's disease. *Eur. J. Neurol.* **1996**, 3, (2), 153-155.
58. Sjogren, M.; Minthon, L.; Passant, U.; Blennow, K.; Wallin, A., Decreased monoamine metabolites in frontotemporal dementia and Alzheimer's disease. *Neurobiol. Aging* **1998**, 19, (5), 379-384.
59. Guo, K.; Li, L., High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry. *Anal. Chem.* **2011**, 82, (21), 8789-8793.

Chapter 5

Metabolomic Profiling of Bronchoalveolar Lavage Fluids by Isotope Labeling Liquid Chromatography Mass Spectrometry: a Promising Approach to Studying Experimental Asthma

5.1 Introduction

Asthma is a global public health problem that in 2010 affected 300 million people¹. It is a heterogeneous syndrome with many clinical classifications. Allergic asthma is the most common form, and allergic inflammation is an important pathophysiological feature. Allergic inflammation may involve the interplay of several cell types including: epithelial cells, smooth muscle cells, mast cells, dendritic cells, eosinophils, neutrophils, macrophages and T and B lymphocytes². However, the pathogenesis is incompletely understood and there is no objective diagnostic biomarker of asthma in a typical clinical setting³.

Bronchoalveolar lavage (BALF) is a procedure for sampling the epithelial lining fluid of the respiratory tract using a bronchoscope^{4,5}. A saline solution is instilled and then recovered to wash cells and soluble constituents from the alveolar and bronchial airspaces⁶. BALF contains different cell types and a variety of soluble substances, such as protein and small molecule

A form of this chapter was submitted to Journal of Proteome Research as Jun Peng, Chris D. St. Laurent, A. Dean Befus, Ruokun Zhou and Liang Li “Metabolomic Profiling of Bronchoalveolar Lavage Fluids by Isotope Labeling Liquid Chromatography Mass Spectrometry: a Promising Approach to Studying Experimental Asthma”. I did the metabolome profiling experiments and analyzed the metabolome data.

metabolites⁷. Chemical analysis of BALF may aid in diagnosis and in definition of pathophysiological mechanisms of pulmonary disease, as BALF is in close proximity to lung tissue, and thus likely to be a better source of biomarker discovery than other biofluids, such as urine and blood. Cytological analysis of BALF can be useful in diagnosis of some pulmonary diseases, however, this analysis lacks sensitivity and specificity⁸. Accordingly, there is both interest and clinical need in the development of novel assays of BALF for pulmonary diseases.

Metabolomics seeks to study the complete set of small molecule metabolites in a biological sample using a holistic and unbiased approach. The metabolome measurement can provide a functional readout of the physiological and pathophysiological state and has potential to identify novel biomarkers for pulmonary diseases and help understand molecular mechanisms of the diseases such as asthma^{9, 10}. Metabolomic profiling of urine and breath samples have been used to investigate asthma^{11, 12, 13}. However, until recently metabolomic profiling of BALF has not been widely used, because the metabolite level in BALF can be relatively low compared with blood or urine, presenting an analytical challenge^{7, 14}. A few metabolomics studies of BALF have been reported, including in humans¹⁵, mice¹⁶, and rats¹⁷ and most studies used NMR spectroscopy, a relatively insensitive method with limited metabolome coverage. Gas chromatography MS (GC-MS) was applied to analyze BALF from human infants¹⁸ and more recently, GC-MS along with LC-MS was used to analyze BALF in experimental allergic asthma in mice and revealed novel changes in lung metabolic pathways as compared with normal controls¹⁹. This work also illustrated the challenge in using

conventional LC-MS for detecting BALF metabolites; only about 300 features were detectable using both positive and negative ion modes, compared to thousands of features detected in urine or serum/plasma samples. Moreover, <20% of detected features are actually from unique metabolites in LC-MS analysis of biofluids²⁰. Thus, the actual number of metabolites detected in BALF was likely small. Clearly, there is a need to develop a more sensitive LC-MS technique for metabolome profiling of BALF samples with greater metabolome coverage.

We recently developed a differential ¹²C-/¹³C-dansylation isotope labeling LC-FTICR-MS method for profiling metabolites containing primary, secondary amine or phenol groups²¹. This method can enhance sensitivity significantly (1-3 orders of magnitude) and improve quantification precision using ¹³C isotopic global internal standards. This method has been used in metabolic profiling of a variety of biological samples to discover potential metabolite biomarkers of diseases and cellular metabolomics for biological studies^{22,23}. Here we report a method based on dansyl-labeling LC-MS for metabolomic profiling of rat BALF samples and apply this to investigate experimental asthma in rats.

5.2 Experimental Section

5.2.1 Chemicals and reagents.

All the chemicals and reagents, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). For dansylation labeling reactions, the ¹²C-labeling reagents were from Sigma-Aldrich and the ¹³C-labeling reagents were synthesized in our lab using the procedures published previously²¹.

LC-MS grade water, methanol, and acetonitrile (ACN) were purchased from ThermoFisher Scientific (Nepean, ON, Canada).

5.2.2 Rat model and sample collection.

Male 10-week old Brown Norway rats (Harlan Sprague Dawley, Indianapolis, IN) were housed in the Health Sciences Laboratory Animal Services (University of Alberta, Edmonton, Alberta, Canada), on a 12-h light-dark cycle, with access to food and water *ad libitum*. The animals were rested for a minimum of 1 wk after arrival before experimentation. Rats were sensitized to ovalbumin (OA) with an intraperitoneal injection of 10 μ g OA, 50 ng pertussis toxin, and 150 mg aluminum hydroxide. On d 21 rats were placed in plastic chambers with lids, and challenged for 5 min with a 5% OA solution or saline using a Micro Mist nebulizer (Hudson RCI, Durham, NC). This work was approved by the University of Alberta Health Sciences Animal Policy and Welfare Committee in accordance with guidelines of the Canadian Council on Animal Care.

To minimize day-to-day and animal batch variation, all animal experiments were conducted by a single investigator using rigorously standardized methods, and on each day of animal sacrifice, rats from all experimental groups were done. Twenty-four h following OA challenge, rats were euthanized, and the trachea of each rat was cannulated with polyethylene tubing attached to an 18-gauge needle, and 5 mL of ice-cold phosphate buffered saline (PBS) was instilled into the airways and the lungs were gently massaged. The PBS was aspirated into ice-cold polypropylene tubes, and the 5 mL instillation and removal was repeated six

times. For the first 5 mL instillation, 2 to 2.5 mL was recovered, kept separate and following cell collection by centrifugation (see below), the fluid was stored in a -80 °C freezer to be used for metabolomic analysis. The remaining BALF collected (on average 23 mL of the remaining 25 mL instilled) were centrifuged for 5 min at 300 x g to pellet the cells. Cell pellets from the initial 2 mL and remaining BALF were combined and resuspended in PBS, yielding a cell viability of >95% as determined by trypan blue exclusion. The isolated cells were counted and cell smears were prepared using a Cytospin (Thermo Fisher Scientific). Cell differentials (minimum of 300 cells assessed) were determined using cytopins of the BALF cells following staining with PROTOCOL Hema 3 staining system (Thermo Fisher Scientific).

There were 18 inflamed rats sensitized with OA and challenged with aerosolized OA. There were 18 control rats, which included 9 naïve rats and 9 rats sensitized with OA but challenged with saline, not OA. Cell counts and differential analysis established that naïve rats and rats sensitized with OA but challenged with saline had similar cell profiles in BALF. As shown in the Results, principal components analysis (PCA) and hierarchical clustering analysis (HCA) of metabolites also showed that there was no obvious separation between naïve rats and rats sensitized with OA but challenged with saline. This data also established that the use of pertussis toxin and aluminum hydroxide as adjuvants for OA immunization did not significantly alter the base metabolome profile in BALF.

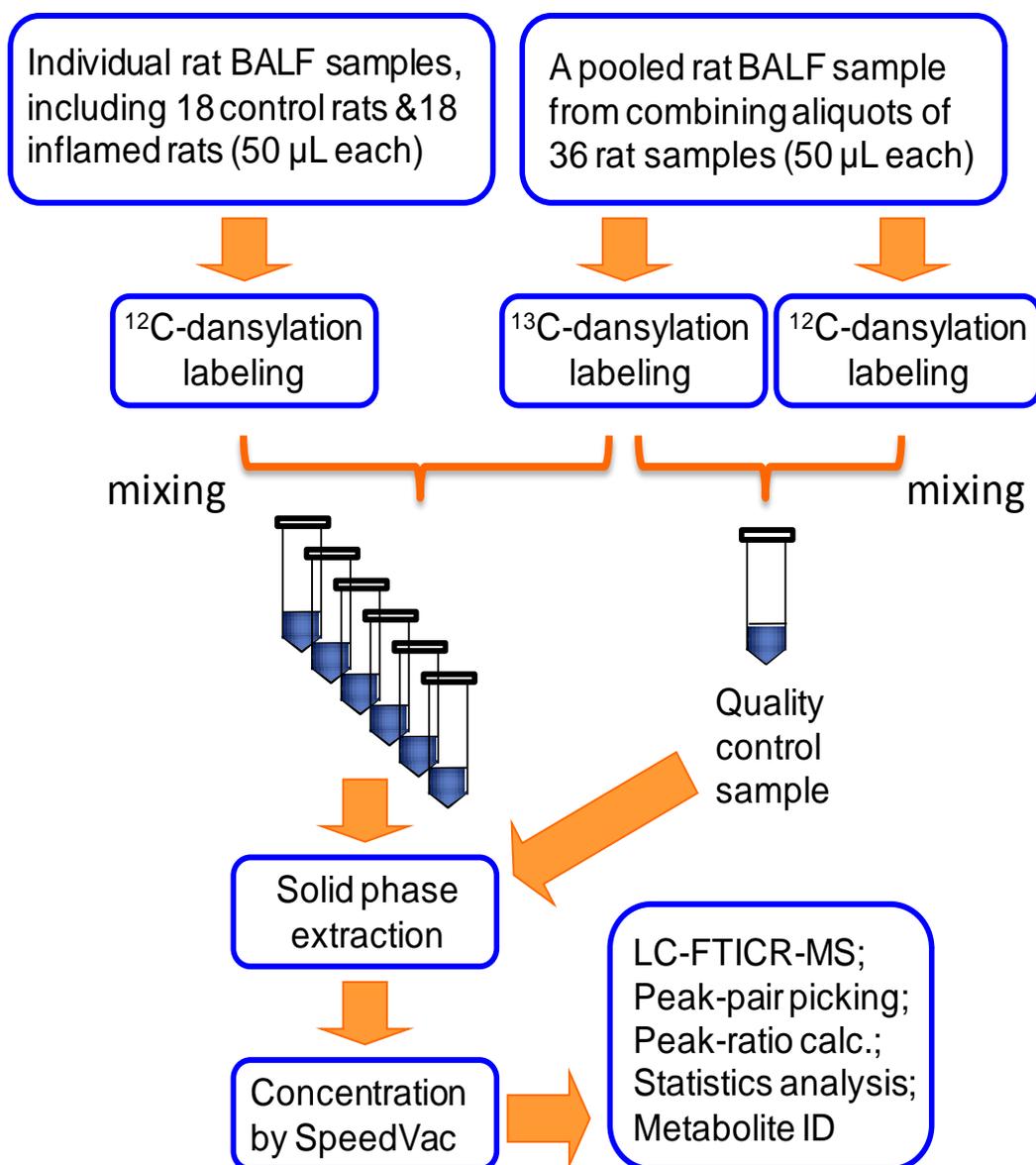


Figure 5.1 Experimental workflow of isotope labeling LC-MS for comparative metabolomics between 18 control rats (naive [n=9] pooled with sensitized but saline challenged, n=9) and 18 sensitized rats challenged with aerosolized ovalbumin (inflamed). The pooled BALF sample was generated from aliquots of 36 rat BALF samples and used as a global internal standard.

5.2.3 BALF sample preparation.

Figure 5.1 shows the workflow for sample preparation and LC-MS analysis. A pooled sample was generated from aliquots of 36 rat BALF samples (from the 36 rats listed above, both control and inflamed rats). Fifty μL of each of 36 rat samples was ^{12}C -dansylation labeled (see next section). Fifty μL of the pooled sample was ^{13}C -dansylation labeled. After they were combined, the mixture was run through solid phase extraction (SPE Oasis HLB cartridge, Waters Limited, Mississauga, Ontario, Canada). The SPE cartridge was first conditioned with methanol, then equilibrated with water and after the sample was loaded, the cartridge was washed with water to remove the salts. The sample was eluted with acetonitrile and the eluate was concentrated 10.5 times using a SpeedVac and was injected into LC-MS.

A 50 μL aliquot of the pooled sample that was ^{12}C -dansylation labeled was mixed with a 50 μL aliquot of the pooled sample that was ^{13}C -dansylation labeled. This mixture was used as a quality control (QC) sample. It was run through SPE, followed by analyte enrichment, just as in the case of a sample mixture. There were four QC samples prepared in this way. Each QC sample was injected twice. Thus, there were a total of 8 QC injections. Each QC injection was followed by 5 sample injections except the last batch with 6 sample injections for a total of 36 samples.

5.2.4 Dansylation labeling.

All the dansylation labeling reactions for rat BALF samples were adapted from our previous report²¹. Briefly, 50 μL of the sample was mixed with 50 μL sodium carbonate/sodium bicarbonate buffer (0.5 mol/L, pH 9.5) in reaction vials. Fifty μL of freshly prepared ^{12}C dansyl chloride solution (20 mg/mL) was added to each of individual sample for light labeling. Fifty μL of ^{13}C -dansyl chloride solution (20 mg/mL) was then added to each pooled sample for heavy labeling. The dansylation reaction was performed in an Innova-4000 bench top incubator shaker (New Brunswick Scientific, Enfield, CT, USA) at 60 $^{\circ}\text{C}$ for 60 min. Finally, 10 μL of a 250 mM sodium hydroxide solution was added to quench the reaction.

5.2.5 Sample preparation for method development.

Pooled normal rat BALF was split into two identical aliquots of 100 μL . One aliquot was ^{12}C -dansylation labeled and the other aliquot was ^{13}C -dansylation labeled. They were combined into one vial. After adjusting pH to 3, the mixture was run through SPE. For the absolute quantification, one aliquot of pooled normal rat BALF sample was ^{12}C -dansylation labeled and then spiked with a fixed amount of authentic standard which was ^{13}C -dansylation labeled. Calibration was prepared by using a series of concentrations of a mixture of authentic standards, which were ^{12}C -dansylation labeled. They were spiked into the fixed amount of authentic standard, which was ^{13}C -dansylation labeled.

5.2.6 LC-MS.

An Agilent 1100 series binary system (Agilent, Palo Alto, CA) and an Agilent reversed-phase Eclipse plus C18 column (2.1 mm×100 mm, 1.8 μm particle size, 95 Å pore size) were used for LC-MS. LC solvent A was 0.1% (v/v) LC-MS grade formic acid in 5% (v/v) LC-MS grade ACN, and solvent B was 0.1% (v/v) LC-MS grade formic acid in LC-MS grade acetonitrile. The gradient elution profile was as follows: t = 0 min, 20% B; t = 3.0 min, 35% B; t = 16 min, 65% B; t = 18.6 min, 95% B; t = 21 min, 95% B; t = 21.3 min, 98% B; t = 23.0 min, 98% B; t = 24.0 min, 20% B. The flow rate was 150 μL/min. The flow from RPLC was split 1:2 and a 50 μL/min flow was loaded to the electrospray ionization (ESI) source of a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA, USA), while the rest of the flow was delivered to waste. All MS spectra were obtained in the positive ion mode.

The same LC equipment with the same LC running conditions was connected with QTRAP 2000 quadrupole linear trap mass spectrometer (AB Sciex, Concord, ON, Canada) to do the absolute quantification work.

5.2.7 Data analysis.

An integrated approach was used to do the automated data preprocessing. The XCMS software²⁴ was used for peak picking from the LC-MS data. An in-house written R program was used to find ¹²C- /¹³C-dansyl labeled peak pairs based on the mass difference of 2.00671 Da of isotopic pairs and the mass

accuracy tolerance of 2 ppm. The relative ion intensity of ^{12}C -labeled/ ^{13}C -labeled pairs was calculated. The redundant peaks of each metabolite, such as natural isotopic peaks, sodium/potassium/ammonia adduct peaks, doubly or triply charged peaks, and dimer peaks, were automatically removed by the program. An in-house written Perl program was used to align the peak pairs across the different BALF samples, which was based on the mass accuracy tolerance of 5 ppm and retention time shift tolerance of 25 sec. The table resulting from the Perl program contained information on relative ion intensity of each unique peak pair in each BALF sample.

Multivariate statistical analysis was carried out using SIMCA-P+ 11.5 (Umetrics AB, Umea, Sweden). PCA, partial least squares discriminant analysis (PLS-DA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were used to analyze the data. One hundred permutation tests, built in SIMCA-P, were used to conduct cross validation for OPLS-DA model. A list of top important metabolites which contributed most to the model was generated from the variable importance of the projection (VIP) plot. There were 48 metabolites with VIP score of greater than 1.0. Among them, 44 metabolites were also found in the PLS-DA model (see Supplemental Table T1) A web-based software Metaboanalyst²⁵ was used to construct the heat map using hierarchical clustering method, and to perform the t-test, calculate the P-value and fold change for individual metabolites. Correlations between the important metabolites and inflammatory cells were calculated using Pearson's correlation analysis built in Metaboanalyst 2.0.

5.2.8 Metabolite identification.

The potential metabolite biomarkers were selected based on their rank in VIP value in the OPLS-DA model, P-value, and fold changes. Each metabolite had its retention time and accurate mass information. Mass of the un-derivatized metabolites was obtained by subtracting the measured mass of the dansylation labeled metabolite from the dansyl group. The MycompoundID program²⁶ was used to search the accurate mass within the human metabolome database (HMDB)²⁷ with a mass accuracy tolerance of 5 ppm. The metabolite markers were definitively identified if they were matched with the retention time and accurate mass of authentic standards under the same experimental conditions. The matched metabolites were deemed to be putatively identified if their authentic standards were not available.

5.2.9 Metabolic pathway analysis.

The construction and analysis of metabolic pathways of potential biomarkers was performed using a web-based tool, MetPA²⁸ that was associated with Metaboanalyst 2.0 software (<http://metpa.metabolomics.ca/MetPA/>). MetPA combines advanced pathway enrichment analysis along with analysis of pathway topological characteristics to identify and visualize the most relevant metabolic pathways.

5.3 Results

5.3.1 Method development.

After dansylation the hydrophobic aromatic structure of the dansyl group improves separation of labeled metabolites on reversed phase liquid chromatography (RPLC). This allows the analysis of polar and ionic metabolites using RPLC, rendering the possibility of analyzing the amine and phenol sub-metabolome with diverse physiochemical properties using only one LC-MS condition. Additionally, the dimethylamino moiety can be readily protonated in ESI. The detection sensitivity of dansyl labeled metabolites is usually enhanced by 10 to 1000-fold, compared to unlabeled counterparts²¹.

Initially we applied dansylation labeling to the rat BALF samples directly, followed by injection into LC-MS, as we would do for other biofluids, such as urine. However, few peaks were detected in the ion chromatogram (see Figure 5.2A). The earlier elution peaks were from the quenched labeling reagent, dansyl hydroxide, and the intense peak at around 19 min was from an unknown impurity with m/z 266.0833; the mass spectra of this peak only shows a single peak, not peak pair. Low metabolite detectability is due to low concentrations of metabolites in BALF collected using a washing procedure. Because of high concentrations of sodium phosphate buffer and salts, it was not feasible to use solvent evaporation to concentrate the metabolites. To overcome this challenge, we developed an analytical strategy using reversed phase SPE to remove high concentrations of buffer, salts and other impurities after dansylation labeling of the metabolites. The SPE eluate was concentrated before injecting into LC-MS. A

10.5 fold increase in analyte concentration was achieved while removing salt interference in LC-MS.

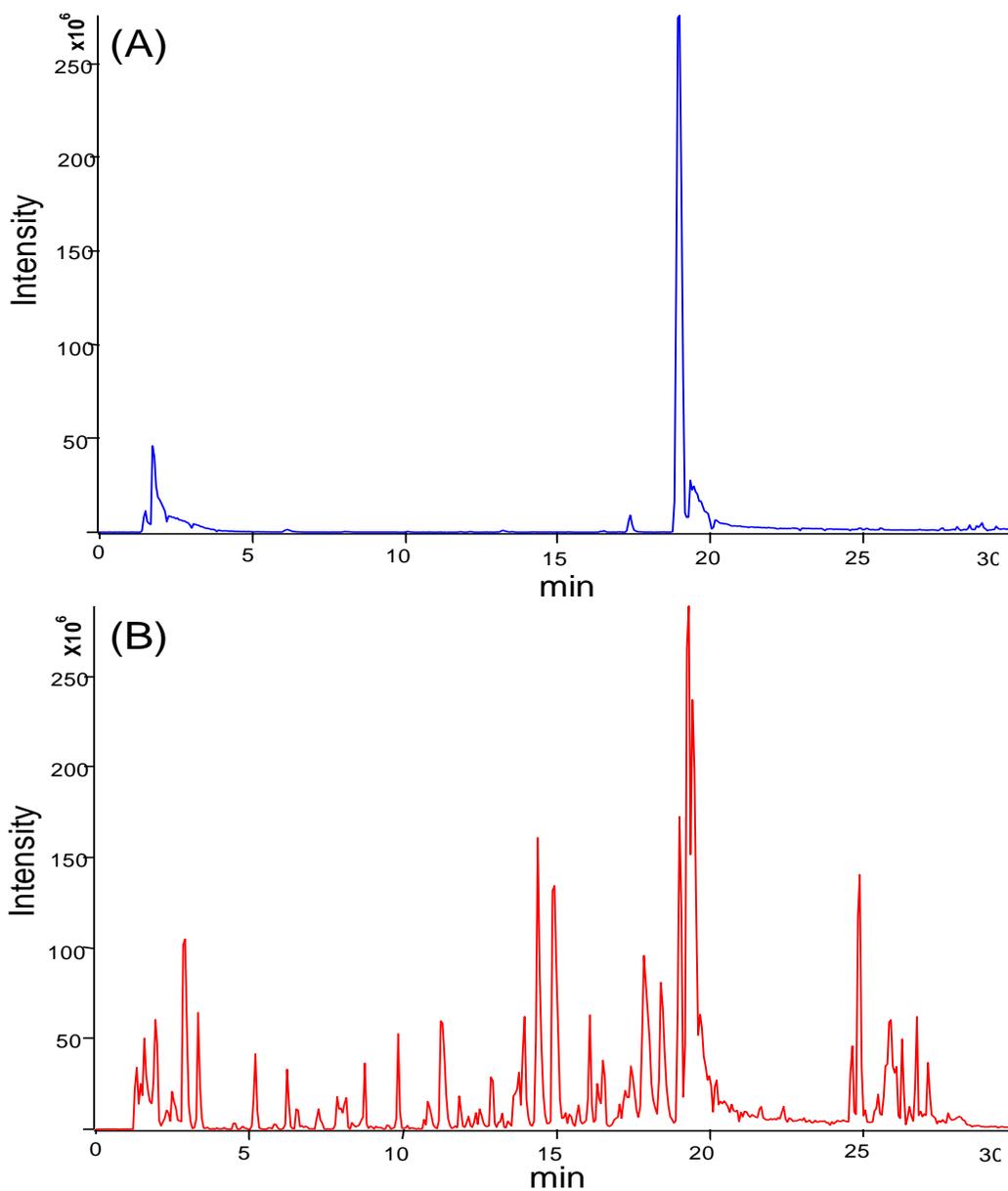


Figure 5.2 Base-peak ion chromatograms obtained (A) without SPE and analyte concentration and (B) with SPE and analyte concentration.

Figure 5.2B shows the ion chromatogram of normal rat BALF obtained using dansyl labeling, SPE and LC-MS. About 250 unique peak pairs were detected, most probably corresponding to 250 different metabolites containing primary, secondary amine or phenol groups. While applying SPE for analyte concentration is straightforward, the pH of the samples should be controlled to generate reproducible results. We studied the effect of different pH of the samples before loading into the SPE cartridge, and found that pH 3 was the best condition compared with pH 7 and pH 10 (Supplemental Figure S 5.1).

5.3.2 Metabolite identification and absolute concentrations.

Among the 250 peak pairs detected, we identified 36 metabolites based on matching mass and retention time with their authentic standards; our current standard library contains 296 metabolites. Table 5.1 shows the list of metabolites positively identified. The majority of them were not identified by previous NMR study of rat BALF samples¹⁷ and mouse BALF samples using LC-MS¹⁹.

There are no reports on the absolute quantification of metabolites in rat BALF samples. To fill this knowledge gap, we developed a scheduled multiple reaction monitoring (MRM) method using QTRAP mass spectrometry to quantify some of the metabolites in normal rat BALF samples. Supplemental Figure S 5.2 shows the ion chromatogram for quantification of 24 metabolites in rat BALF using the MRM method. These 24 metabolites were selected after considering the 36 metabolites identified in BALF and others arbitrarily chosen from our

compound library. We validated the internal standard method by comparing the results obtained with the standard addition method for two metabolites, glycine and alanine. Supplemental Table T2 shows results obtained by the two methods; quantitative results were essentially the same. This is not surprising, as we used SPE in our LC-MS quantitative analysis that would minimize the matrix effect in analyte quantification. Table 5.2 lists the information on the absolute concentrations of 24 metabolites in normal rat BALF. The concentrations of 15 metabolites range from 6.7 nM to 3.93 μ M, whereas concentrations of the other 9 metabolites are below the limit of quantification (50 nM except 5 nM for cresol). These results indicate that metabolite concentrations in BALF are very low, presenting a major detection challenge to conventional metabolome profiling techniques such as NMR and LC-MS. However, dansyl labeling LC-MS can provide a means of detecting many amine- and phenol-containing metabolites in BALF. Note that, using the QTRAP2000 mass spectrometer which is not very sensitive, compared to the latest model, the 9 metabolites could not be detected using the MRM method. However, they were detectable in the full scan mode of the more sensitive FT-ICR-MS instrument.

5.3.3 Comparative metabolome analysis between inflamed and control rats.

We have developed a workflow (Figure 5.1) that incorporated SPE in sample handling for differential isotope labeling of amine- and phenol-containing metabolites of BALF samples. A total of 36 rat samples including 18 control rats (9 naïve and 9 sensitized but challenged with aerosolized saline rather than OA) and 18 inflamed rats were analyzed.

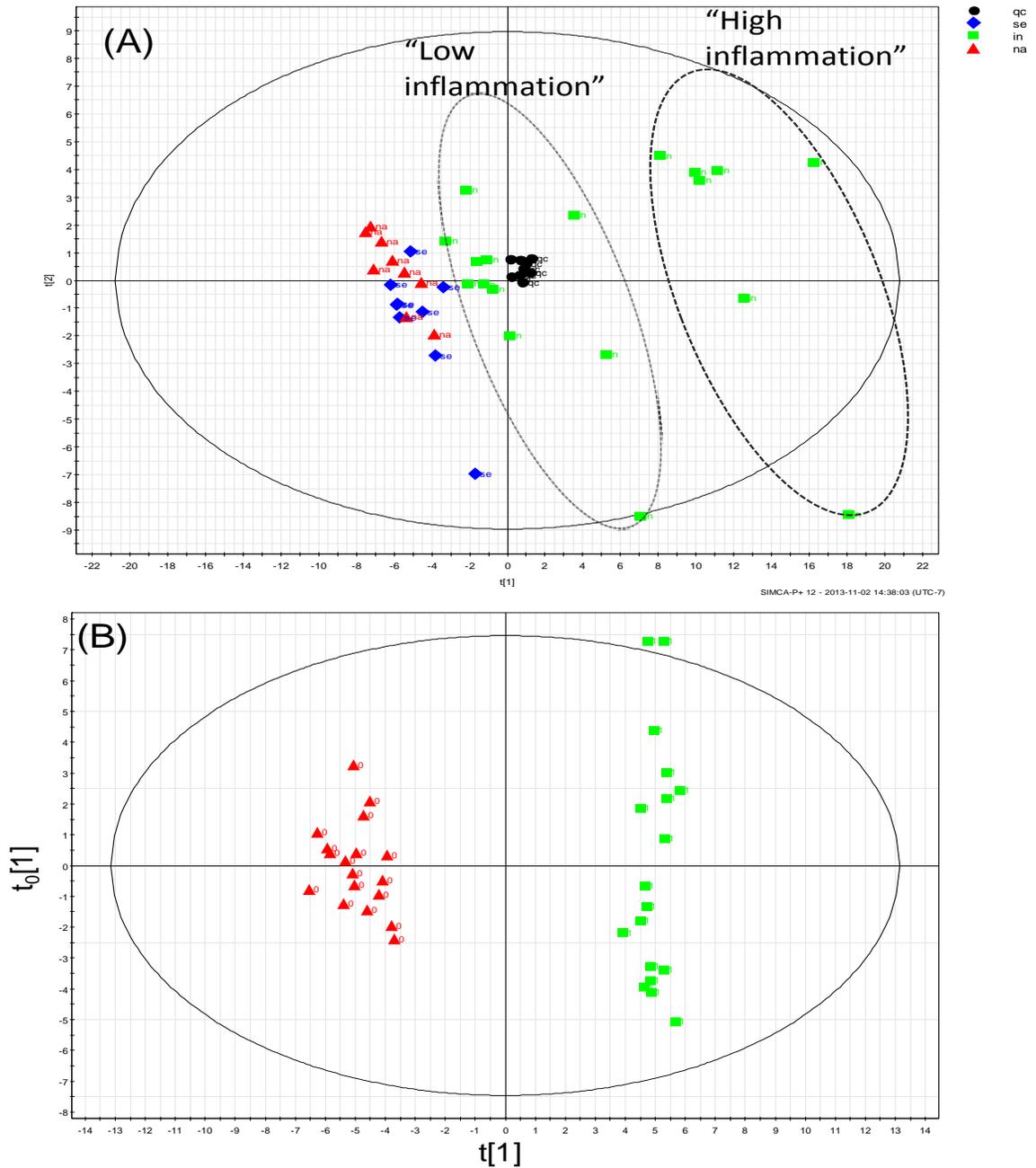


Figure 5.3 (A) Score plot of PCA showing some separation between 9 naïve rats (red triangles), 9 rats sensitized but saline challenged (blue diamonds), and 18 inflamed rats (green squares); black dots were from 8 injections of the quality control sample. (B) Score plot of OPLS-DA showing distinct separation between 18 control rats (red triangles) and 18 inflamed rats (green squares).

During the data collection, a quality control (QC) sample (see Figure 1) was used to confirm if the analytical performance was robust. Figure 5.3A shows the PCA score plot of the individual rat samples as well as the QC samples. X-axis is t1 component one, Y-axis is t2 component 2. Component 1 has 19 percent of variance explained, and component 2 has 3.5 percent of variance explained. All QC sample injections (8 injections in total from four QC samples) were tightly clustered (labeled as black dots in Figure 5.3A). This indicates that overall technical variation was small. PCA is an unsupervised model, which can provide an overview of the dataset revealing general trends, clustering and outliers. Figure 5.3A shows that there was no obvious separation between 9 naïve rats (labeled as red triangles) and 9 rats sensitized with OA, challenged with saline (labeled as blue diamonds). In contrast, there is clear separation between 18 inflamed rats (all labeled as green squares) and 18 control rats. In this PCA model, $R^2X(\text{cum})=0.49$, $Q^2(\text{cum})=0.39$.

OPLS-DA, a supervised model, was used to maximize the separation and enhance model visualization and interpretation. Figure 5.3B is the score plot of OPLS-DA showing that there was distinct separation between 18 control rats (red triangles) and 18 inflamed rats (green squares). In this OPLS-DA model, $R^2X(\text{cum})=0.98$, $Q^2(\text{cum})=0.91$. Since OPLS-DA is prone to overfitting, we did cross validation using the 100 permutation test. The validation plot shown in Supplemental Figure S3 indicates that the original model was valid, as the Q2

regression line has a negative intercept and all permuted R2 and Q2 values to the left are lower than the original point to the right.

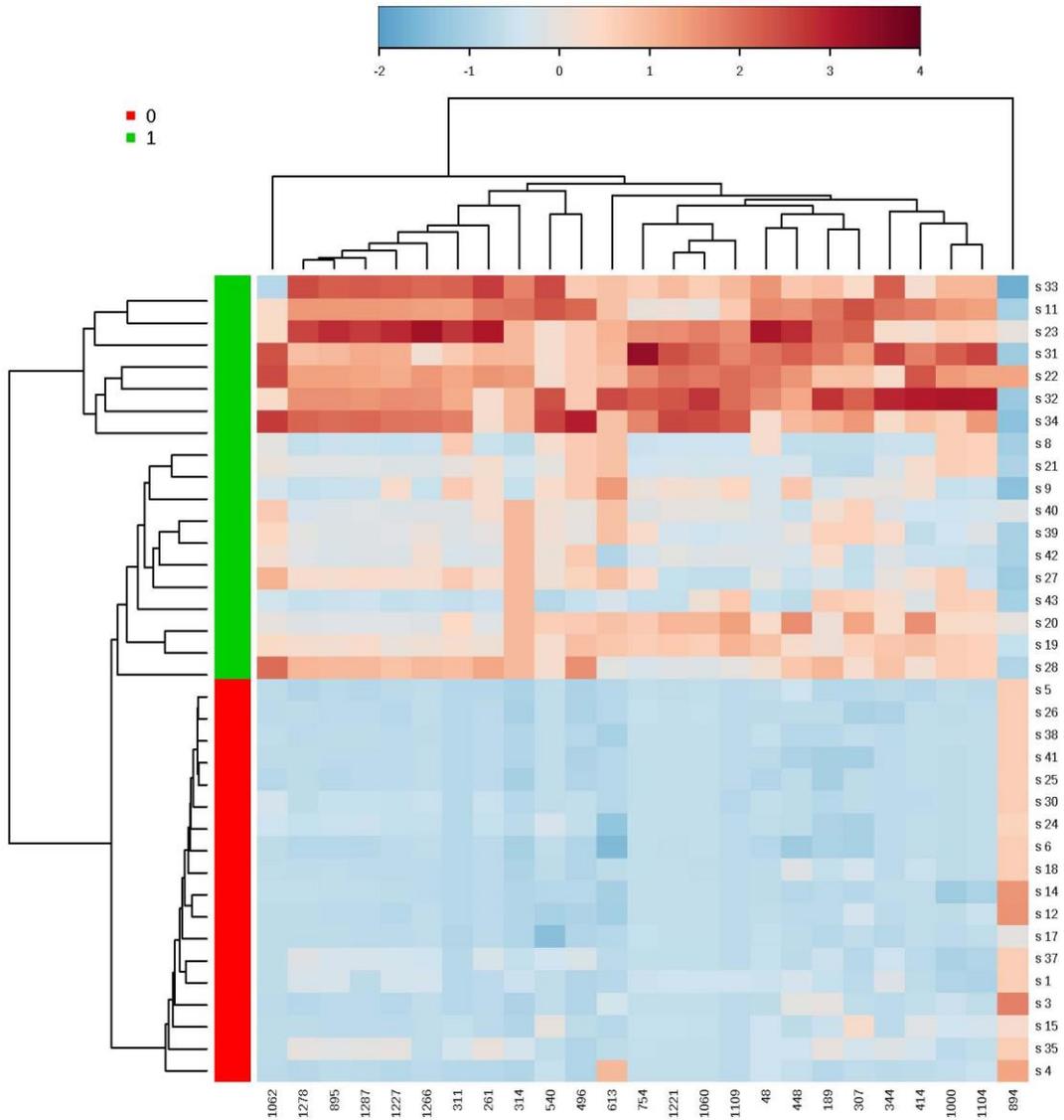


Figure 5.4 Heat map generated by the hierarchical clustering method. The rows are individual rat samples and the columns are the top 25 important metabolites. All 18 control rats are clustered together at the bottom half labeled with red and all 18 inflamed rats are clustered together at the top half labeled with green

HCA is another unsupervised chemometrics method. A dendrogram was generated to see the hierarchical structure of all samples and a heat map was constructed to visualize the intensities of individual metabolites. Figure 5.4 shows the heat map of metabolomic profiles between 18 control rats and 18 inflamed rats, using intensity information of the top 25 metabolites. All 18 control rats cluster at the bottom and all 18 inflamed rats cluster at the top of the heat map, consistent with PCA and OPLS-DA results.

We also did the PCA and HCA analysis only using 9 naïve rats and 9 rats sensitized, challenged with saline (see Supplemental Figures S 5.4 and S 5.5). There is no clear metabolomic difference between two subgroups.

5.3.4 Metabolite biomarkers and pathway analysis.

Among the most significant metabolites, we definitively identified 11 metabolites based on matching retention time and accurate mass with their authentic standards under the same experimental conditions. We also putatively identified 2 metabolites based on accurate mass matches alone, which were not validated since their authentic standards were not commercially available. The list of the top important metabolites identified includes: arginine, proline, hydroxyproline, ammonia, 4-hydroxybenzaldehyde, asparagine, valine, alanine, threonine, 2-phenylglycine, xanthine, formamide, and 4-aminobutanal (Table 5.3). Most of these metabolites were up-regulated in inflamed rats and only one metabolite, xanthine was down-regulated in inflamed rats. T-tests for these 13 metabolites showed that their concentrations were significantly different between

the groups. The fold changes of these metabolites are also distinct, ranging from 1.8 to 5.3.

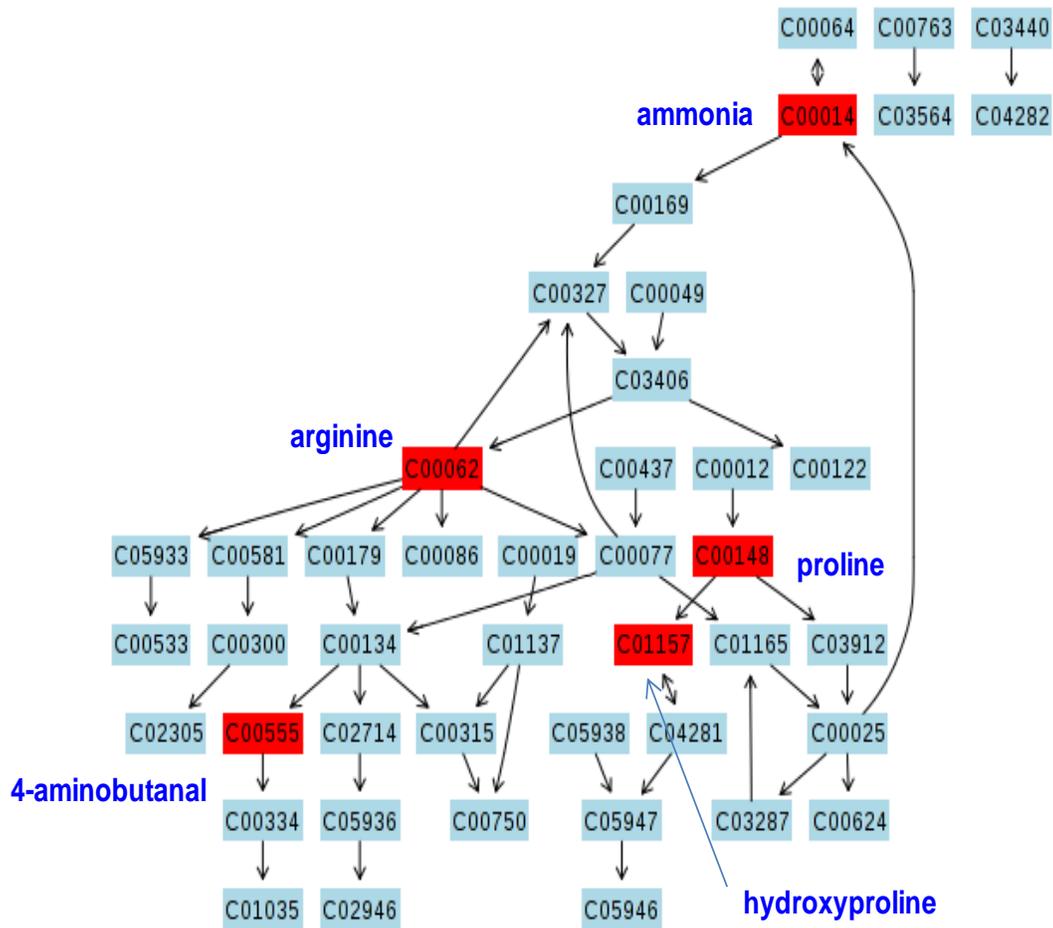


Figure 5.5 Arginine-proline metabolic pathway with five metabolites, arginine, proline, ammonia, hydroxyproline, and 4-aminobutanal, found to be dysregulated by metabolomic analysis (highlighted in red) in the rat model of asthma.

Metabolic pathway analysis was used to determine if these metabolites are related to specific metabolic pathways. The web-based tool, MetPA, combines powerful pathway enrichment analysis with pathway topology analysis to help identify the most relevant pathways. Pathway topology particularly takes into account that centered positions of a network will have a bigger impact in that pathway than marginal or relatively isolated positions. Applying MetPA to our data indicates that it is highly probable that the arginine-proline pathway was dysregulated in allergic inflammation in the lung. Figure 5 shows the perturbed arginine-proline pathway, where five dysregulated metabolites including arginine, proline, hydroxyproline, ammonia, and 4-aminobutanal are highlighted in red.

5.3.5 Cell count analysis and its potential relevance to metabolomic profiling.

There was variability among the inflamed rats in the PCA model and subgroup separation among the inflamed rats (Figure 5.3A). The heat map (Figure 4) also identified two subgroup clusters in the inflamed rats. The top 7 samples clustered into one subgroup and the bottom 11 samples from inflamed rats clustered into another subgroup. The separation was consistent with the subgroup separation in PCA in the inflamed rats. Accordingly, we subdivided the 18 inflamed rats into two subgroups using their metabotype implicated by PCA and the heat map. One subgroup has 7 samples, which is tentatively labeled as “high inflammation”, while the other subgroup has 11 samples, tentatively labeled as “low inflammation” in the PCA score plot (Figure 5.3A). We tested if there was correlation between metabolomic profiling and inflammatory cell numbers in

BALF samples, total cell counts and differential cell analysis including eosinophils, neutrophils, and macrophages (Supplemental Table T3).

Figure 5.6 shows comparisons among the total leucocyte numbers, neutrophils, eosinophils and macrophages in BALF from control rats and OA challenged, sensitized rats in the low and high inflammation subgroups. For all panels (A-D) the cell numbers in both the low and high inflammation groups are statistically different than the numbers in the control group ($p < 0.001$). For the total leucocyte numbers (A) and the neutrophils (B) the differences between the low and high inflammation groups are statistically significant ($p < 0.001$), whereas for eosinophils and macrophages, there is no significant difference. Thus the metabolomic profile of BALF in this rat model of allergic inflammation is associated with the influx of neutrophils rather than eosinophils or macrophages. Pearson correlation analysis was used to identify if there is correlation between individual metabolites and the numbers of infiltrating leucocytes. Table 5.4 shows that among the 13 most interesting metabolites there are 12 metabolites (excluding xanthine with a moderate negative correlation) with a strong positive correlation to neutrophils numbers (correlation coefficients from 0.546 to 0.911). There is very weak correlation between individual metabolites and macrophage numbers. Table 5.5 shows the correlation coefficients of levels of individual metabolites compared to the percentages of leucocytes. Twelve of the metabolites have a positive correlation (one a negative correlation) with the percentage of neutrophils. There are 8 metabolites with a positive correlation with the

percentage of eosinophils and 13 metabolites with a negative correlation with the percentage of macrophages.

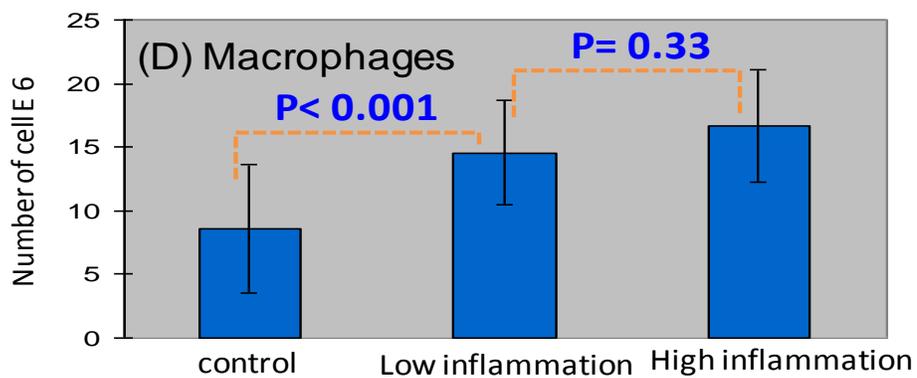
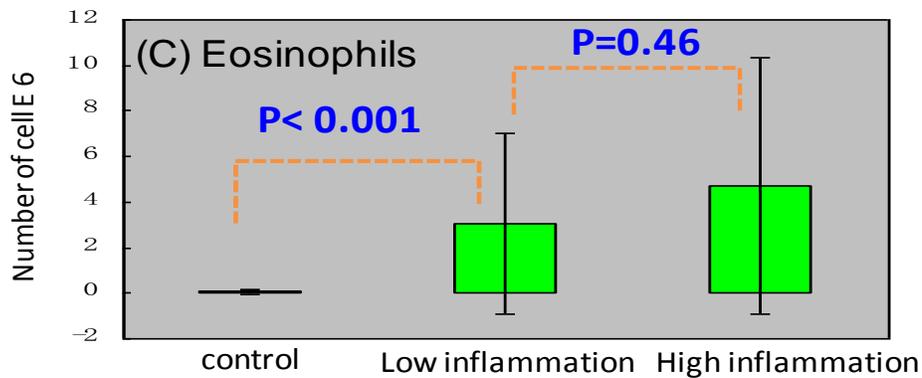
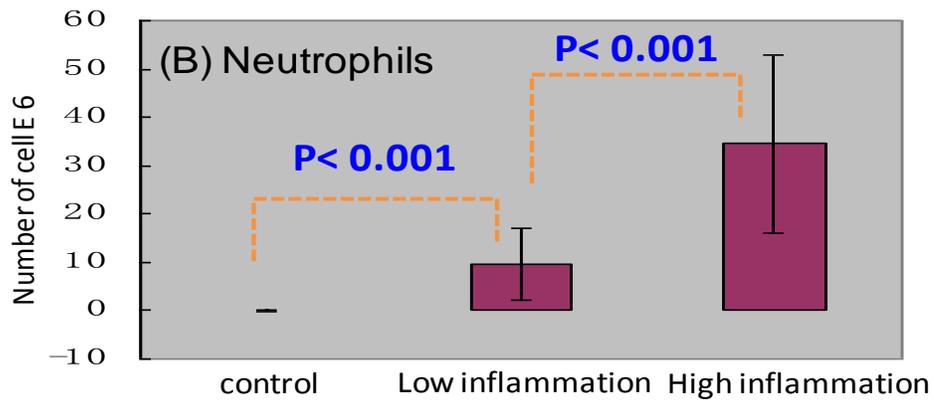
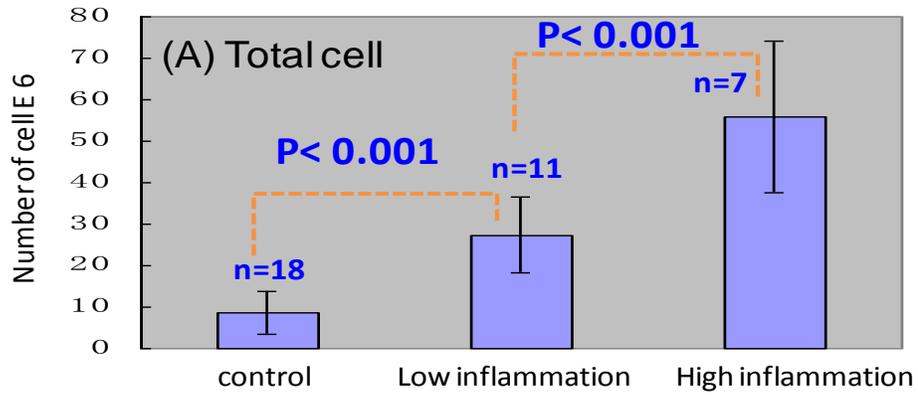


Figure 5.6 Cell numbers in bronchoalveolar lavage fluids from control rats (n=18, naive [n=9] pooled with sensitized but saline challenged, n=9) and sensitized rats challenged with aerosolized ovalbumin (n=18). The sensitized, ovalbumin challenged rats are separated into two subgroups, low and high inflammation, based on metabolomic profiles (see text). (A) total leucocyte numbers, (B) neutrophils, (C) eosinophils, and (D) macrophages; the P values between “low inflammation” and “high inflammation” subgroups are indicated, and also P values between control groups and “low inflammation” or “high inflammation” subgroups are indicated.

5.4 Discussion

Our first objective was to develop a metabolomic profiling method for rat BALF samples using dansylation isotopic labeling LC-MS that would allow the detection of many amine- and phenol-containing metabolites. Metabolites in this sub-metabolome are key components of several metabolic pathways. The dansyl labeling LC-MS method has been successfully applied for comprehensive metabolic profiling of urine, cerebral spinal fluid, saliva, and cell extracts^{21-23, 29, 30}. However, the typical dansyl labeling LC-MS protocol encountered difficulties with rat BALF samples containing phosphate saline buffer that caused ion suppression and metabolite dilution. Accordingly, we used SPE for sample cleanup and analyte enrichment prior to injection of the labeled samples into LC-MS.

There are several advantages of this workflow. Firstly, SPE after dansylation labeling can minimize sample loss, as the labeled metabolites have a higher hydrophobicity than their unlabeled counterparts and are more readily retained on the reversed phase SPE cartridge. Secondly, quantification accuracy and precision are improved by spiking ¹³C-dansyl standards or internal control (i.e., pooled sample) into the ¹²C-dansyl labeled sample before running SPE and concentrating the eluate. Thirdly, dansyl labeling improves sensitivity, allowing the detection of metabolites in nM concentrations, thus increasing the metabolome coverage. A previous NMR study of rat BALF showed small numbers of peaks detected in the NMR spectra with about 20 metabolites identified. In a recent study of BALF from mice 300 peaks were detected by LC-MS using both positive

and negative ion modes. It has been estimated that >80% of the peaks detected in LC-MS analysis of biofluids such as serum and urine are not related to the signals of metabolites and <20 percent of those peaks could be assigned to unique metabolites²⁰. Thus, conventional LC-MS is not sufficiently sensitive to detect many metabolites in BALF. We detected 250 peak pairs or putative metabolites in rat BALF and identified 36 metabolites using authentic standards. We determined that the absolute concentrations of 15 metabolites ranged from 6.7 nM to 3.93 μ M; another 9 metabolites were below the limit of quantification using the QTrap instrument (50 nM except 5 nM for cresol).

In addition to improved detectability, differential isotope labeling allows quantitative metabolomic profiling to reveal metabolic changes with high accuracy and precision. We used a pooled BALF sample labeled with ¹³C-dansyl chloride as a control to compare the metabolite concentration changes among all the individual samples that were labeled separately with ¹²C-dansyl chloride. The peak intensity ratio of the ¹²C-/¹³C-peak pair from a metabolite reflects the concentration of the metabolite in a sample relative to the control or pooled sample. As the same control or pooled sample was used, the peak intensity ratios of a given peak pair determined in different samples can be used to gauge the relative concentration differences among these samples. These ratios from individual samples were analyzed by PCA and other statistics or chemometrics tools to determine the metabolomic variations between 18 control rats and 18 rats with airways inflammation.

Urine samples from human adults with asthma¹², children with asthma³¹, and a guinea pig model of asthma¹¹ have been used to conduct metabolomic profiling and search for metabolite biomarkers of asthma. Urine metabolomics showed great potential for developing a noninvasive tool for diagnosis of asthma. However, urine might not directly reflect the pathophysiological status of lung tissue, the location of allergic inflammation. By contrast, BALF is a washing from the lumen of the lung, and studying the BALF metabolome is likely to better reflect the metabolomic changes associated with the pathophysiology of asthma. Indeed, using BALF from a strongly eosinophilic model of experimental asthma in the mouse, Ho et al. identified alterations in energy, amino acid and lipid metabolism in the lung¹⁹. Previous urine studies identified some potential biomarkers associated with asthma, however, they did not report any dysregulated metabolic pathways. We established that the arginine-proline metabolic pathway was likely dysregulated in our rat model of allergic airways inflammation. None of those metabolites in arginine-proline pathway were reported in previous urine studies, and the mouse BALF study reported only 2-oxoarginine, one of metabolites in arginine-proline pathway was affected by dexamethasone treatment used in human asthma. Our method has limitations as it targets metabolites containing primary, secondary, or phenol groups. It does not detect metabolites such as carbohydrate, lipid, sterol, organic acids, which were detected by GC-MS in Ho et al. study¹⁹. Our lab is currently developing other isotope labeling LC methods to target those metabolites.

Rat and mouse models of experimental asthma differ in the immunization protocols for OA, and perhaps most importantly in the challenge protocols with aerosolized OA (d 21, 5 min, 5% exposure in the rat; d 22, 23 and 24, 30 min, 1% exposures in the mouse³². In the rat model there is neutrophilia in BALF (~19%), as well as a moderate eosinophilia (~3.7%) and an increased number of macrophages, whereas in the mouse model there is a prominent eosinophilia (>50%) and only a modest neutrophilia (<5%)¹⁹. These distinct inflammatory cell infiltrates, as well as the duration of OA challenges, likely contribute to differences in lung metabolism seen in the models.

Arginine is an important metabolite of the urea cycle that can be metabolized into urea and ornithine by the enzyme arginase³³. Ornithine can serve as a substrate for ornithine decarboxylase, leading to downstream proline. Hydroxyproline is a product of proline hydrolyzed by 4-hydroxylase. Both proline and hydroxyproline are required for collagen synthesis, which is associated with pulmonary fibrosis^{34, 35}. Ornithine can also be a substrate for ornithine aminotransferase, which synthesizes polyamines such as putrescine. 4-aminobutanal is a metabolite of putrescine using diamine oxidase. Polyamines and their metabolites may be involved in promotion of cell proliferation and differentiation³⁶. Therefore, our metabolomics data suggest that the arginine-proline metabolic pathway might play an important role in the pathogenesis of allergic asthma. Indeed, there is considerable evidence that dysregulated arginase and altered arginine metabolism are associated with asthma^{33, 36, 37, 38}. One of the earliest studies using global microarray analysis found that arginase isoforms I

and II were dysregulated in lung samples from a mouse model of asthma^{33, 37}. Arginase expression was induced in the mouse model, as well as in humans with asthma. Arginase is mainly responsible for regulating arginine metabolism into urea and ornithine as well as downstream metabolites, such as putrescine, proline and hydroxyproline. Therefore our untargeted metabolomics data helps validate previous studies suggesting the importance of the arginine-proline metabolic pathway in the pathogenesis of asthma. Further studies are warranted to investigate the role of enzymes other than arginase in this metabolic pathway.

We identified that xanthine, an intermediate in the degradation of adenosine monophosphate to uric acid, was downregulated in allergic asthma. Although it is unclear whether it is associated with pathogenesis of asthma, it has been reported that xanthine could be used as a bronchodilator to treat asthma³⁹.

Our metabolomic profiling of the rat model of allergic inflammation is associated with the influx of inflammatory cells into the airways, especially neutrophils and to a lesser extent eosinophils. Although eosinophilic airway inflammation is one pathological hallmark in allergic inflammation⁴⁰, it cannot explain the diversity of phenotypes of asthma and increased numbers of neutrophils are associated with some severe asthma⁴¹⁻⁴⁴. Although the exact role of neutrophils in the asthma has not been elucidated, neutrophils release mediators such as elastase, which can promote activation of matrix metalloproteinases and the degradation of type III collagen². Given that proline and hydroxyproline dysregulation is associated with collagen deposition, the perturbed arginine-proline metabolic pathway we identified may be associated with neutrophil

activation and function. However, several other cell types can be associated with collagen metabolism and the arginine-proline pathway and further study is needed to elucidate the relationship between the metabolomic profile of BALF and inflammatory cell numbers and their activities.

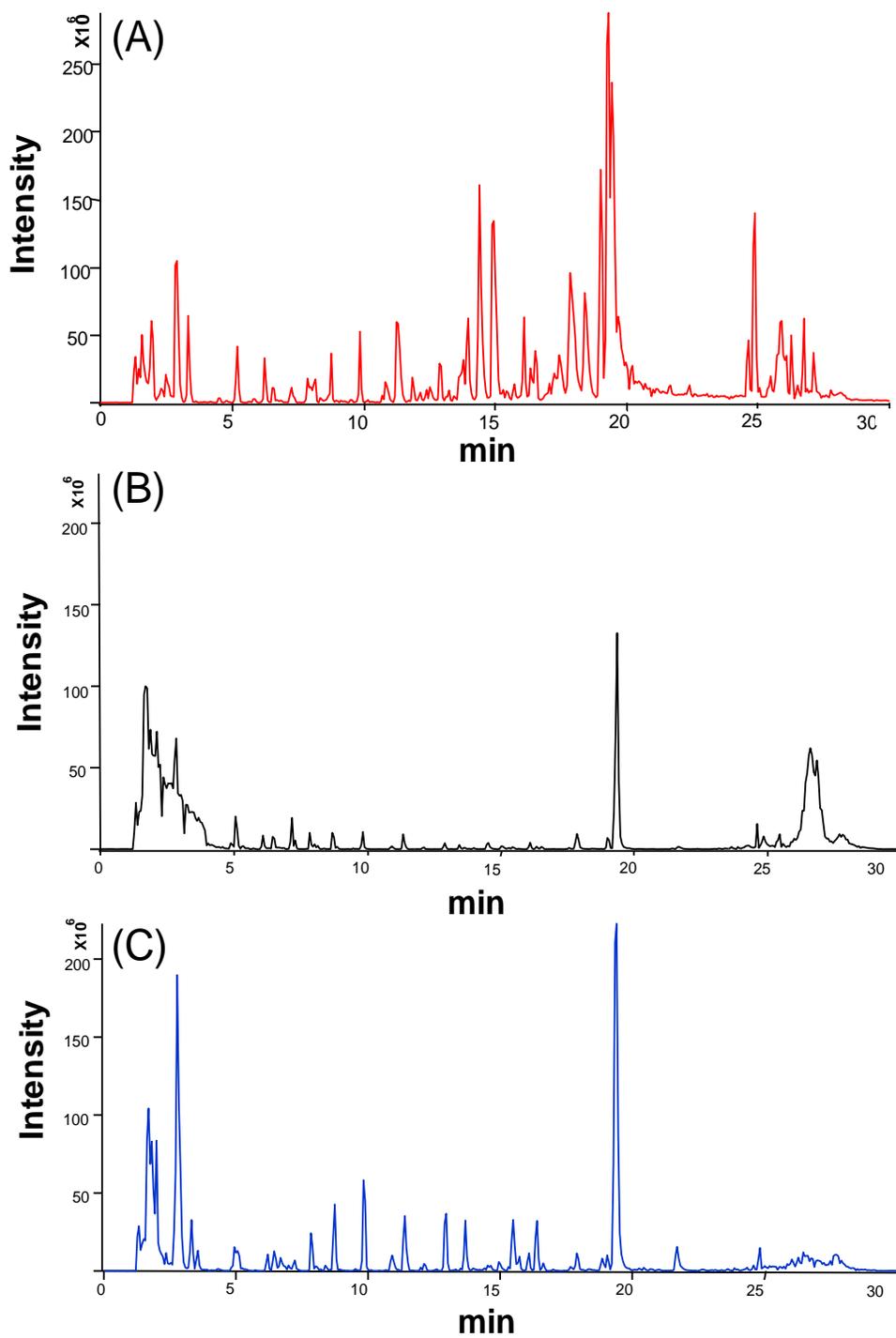
5.5 Conclusions

We have developed a new LC-MS method for analyzing the metabolome of rat BALF with much improved sensitivity. This method is based on differential $^{12}\text{C}/^{13}\text{C}$ dansyl labeling of BALF metabolites, SPE for analyte concentration, and LC-FTICR-MS for analysis. Dansylation selectively labels the amine- and phenol-containing metabolites and improves their separation in RPLC and detection in ESI-MS. We detected 250 peak pairs or putative metabolites in rat BALF, among which we could identify 36 metabolites. We applied this method to analyze the BALF metabolome of an experimental model of asthma. Comparative metabolomics using chemometrics methods including PCA, OPLS-DA and hierarchical clustering identified distinct separation between 18 control rats and 18 inflamed rats. Metabolic pathway analysis showed that the arginine-proline metabolic pathway was dysregulated in the rat model of allergic inflammation. Thus isotope labeling LC-MS can be used to improve metabolomic profiling of BALF. Future work in applying and developing other labeling chemistries targeting other functional groups will increase the overall metabolome coverage. Metabolomic profiling of BALF is a promising approach to studying asthma and has potential as a novel diagnostic tool in clinical settings. With improved

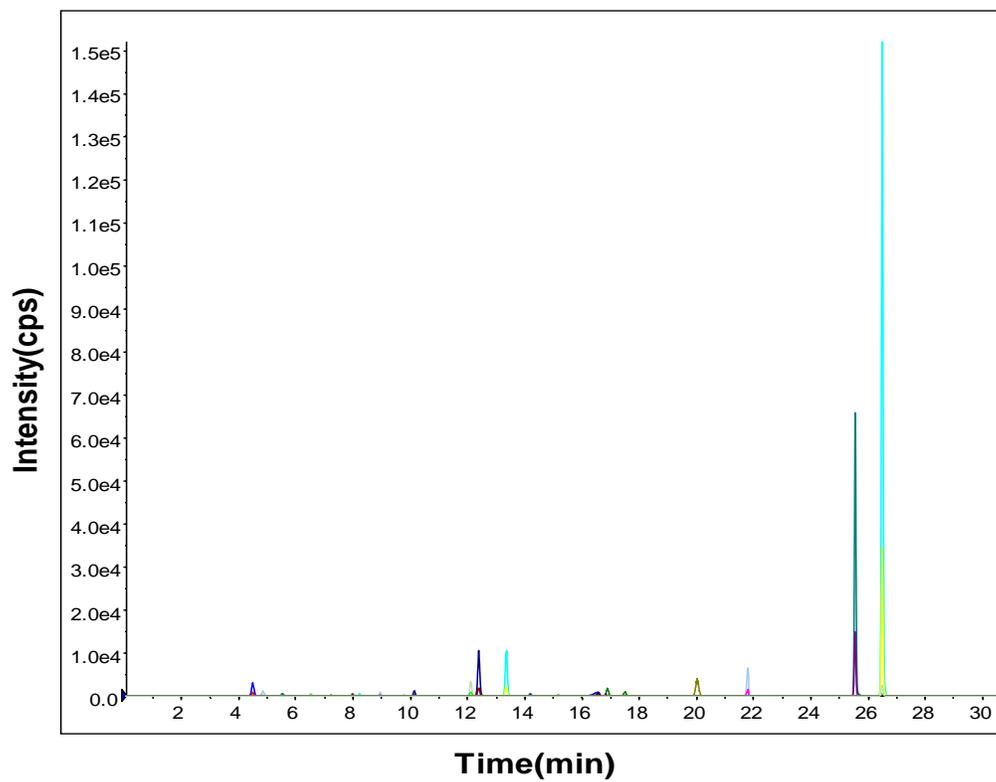
coverage, metabolome profiles of BALF may also serve as an indicator for monitoring the effect of various therapeutics on the treatment of asthma^{11, 19}.

Supporting Information Available

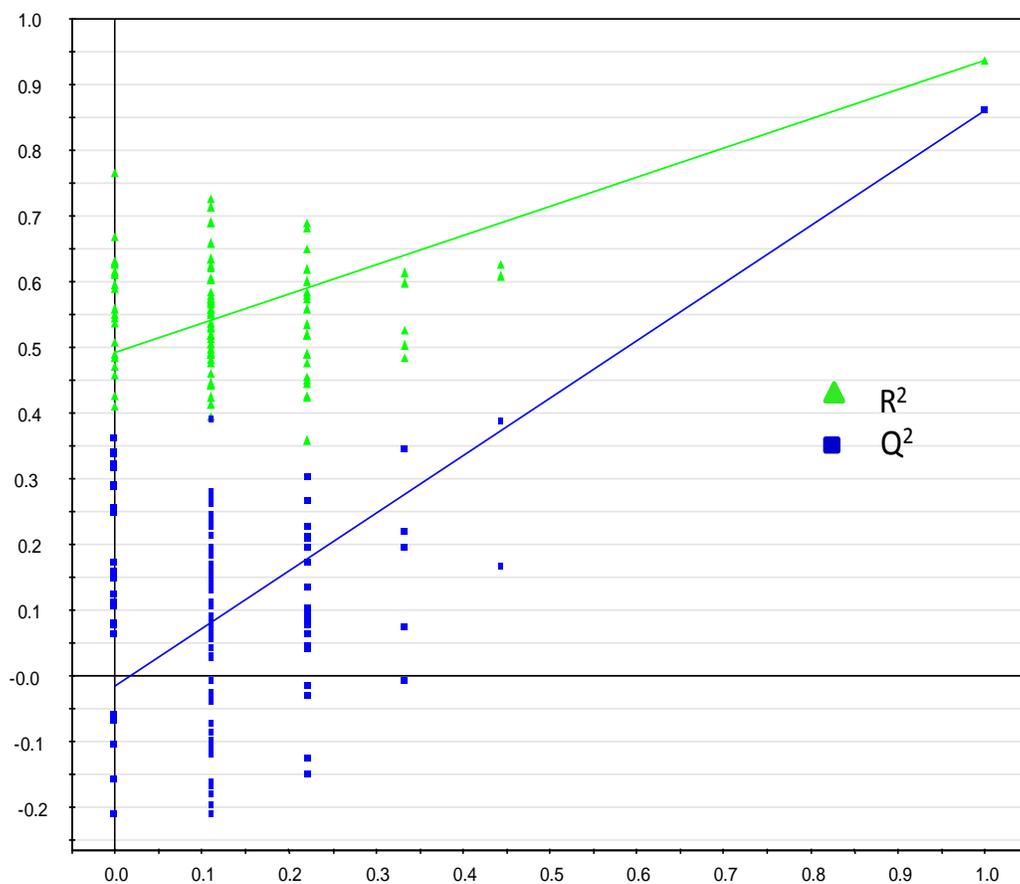
Effect of sample pH on SPE (Supplemental Figure S5.1); ion chromatogram for absolute quantification of 25 metabolites (Supplemental Figure S 5.2); validation of OPLS-DA model (Supplemental Figure S 5.3); PCA score plot between 9 naïve rats and rats sensitized, challenged with saline (Supplemental Figure S 5.4); Heat map between 9 naïve rats and rats sensitized, challenged with saline (Supplemental Figure S 5.5); comparison of discriminant metabolites found in OPLS-DA and PLS-DA models (Supplemental Table T 5.1); quantification method validation (Supplemental Table T 5.2); and leucocyte counting in rat BALF (Supplemental Table T 5.3) are provided.



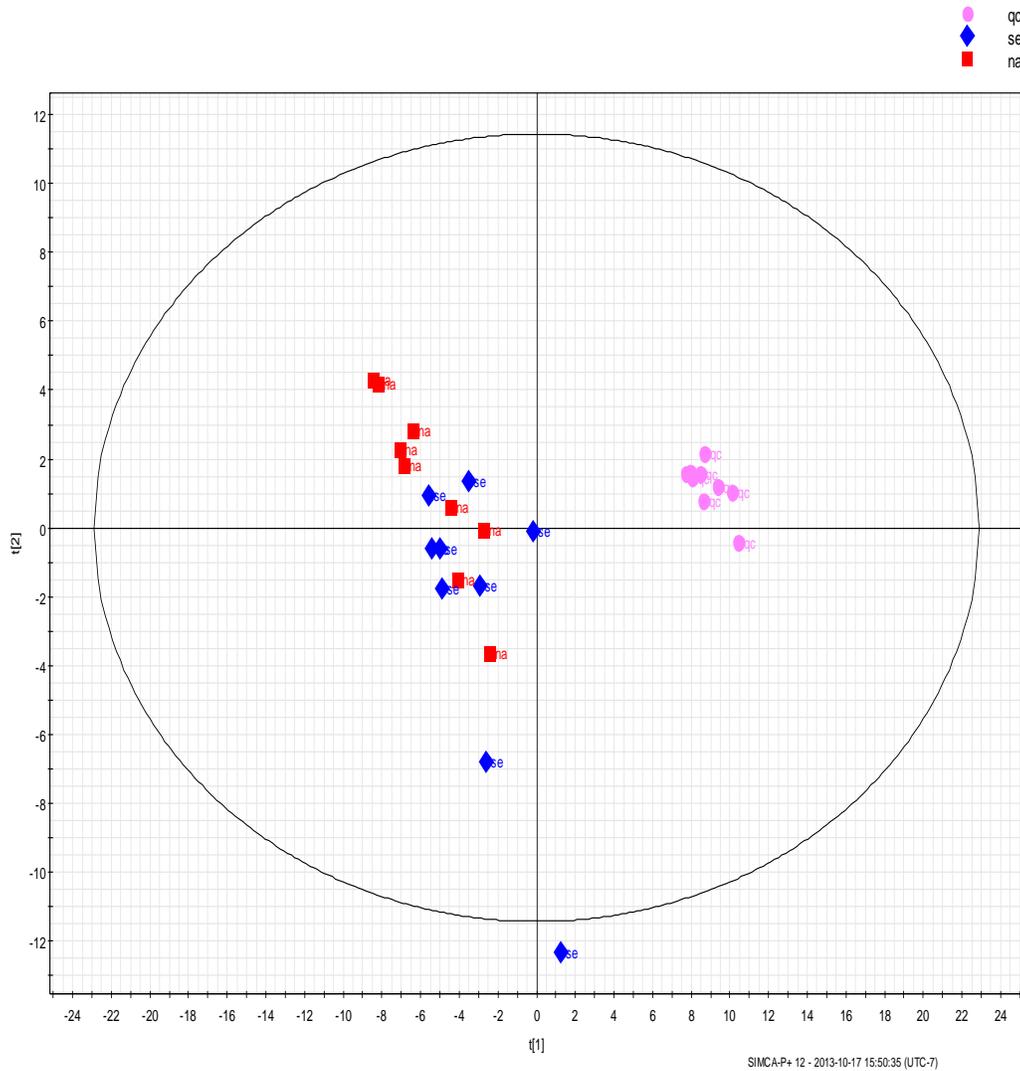
Supplemental Figure S 5.1 Effect of sample pH on SPE: (A) pH 3, (B) pH 7 and (C) pH 10.



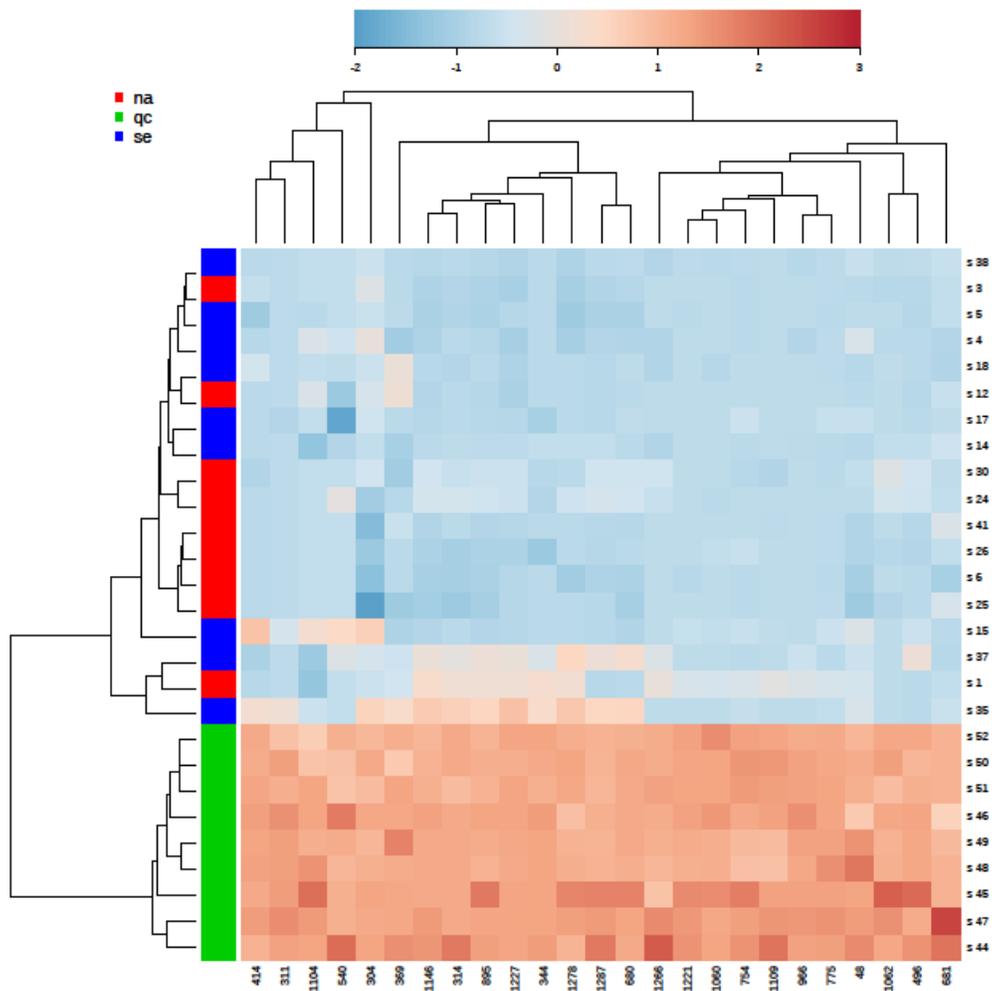
Supplemental Figure S 5.2 Ion chromatogram from normal rat BALF obtained by using scheduled MRM for quantification of 25 metabolites.



Supplemental Figure S 5.3 Cross validation of the OPLS-DA model using 100 permutation test built in SIMCA-P+ software.



Supplemental Figure S 5.4 PCA score plot using only 9 naïve rats (labeled as red squares), 9 rats sensitized, challenged with saline (labeled as blue diamond) and quality control samples (labeled as purple dots)



Supplemental Figure S 5.5 Heat map using only 9 naïve rats (labeled as red), 9 rats sensitized, challenged with saline(labeled as blue) and quality control samples (labeled as green)

Table 5.1 List of identified metabolites in normal rat bronchoalveolar lavage fluid.

Retention time (minutes)	Mass dansylated	Mass underivatized	Metabolite identified
3.41	359.07285	125.01452	taurine
3.83	408.16983	174.11150	arginine*
4.61	366.11175	132.05343	asparagine*
5.09	359.15317	125.09485	1-methyl-histamine*
5.33	380.12728	146.06896	glutamine
5.39	359.15323	125.09490	3-methyl-histamine*
5.75	409.15396	175.09563	citrulline*
6.65	339.10073	105.04240	serine
7.31	381.11133	147.05300	glutamic acid
7.43	367.09576	133.03743	aspartic acid*
7.49	422.18568	188.12735	homoarginine*
8.03	353.11646	119.05814	threonine
8.81	309.09022	75.03190	glycine
9.95	323.10577	89.04744	alanine
10.91	395.12723	161.06891	aminoadipic acid*
13.01	349.12151	115.06319	proline
13.25	323.10584	89.04751	sarcosine*
13.79	383.10944	149.05112	methionine*
15.11	337.12155	103.06323	2-aminobutyric acid*
16.01	462.24235	228.18402	leu-pro*
16.06	402.10071	168.04238	3-hydroxymandelic acid*
16.30	385.12157	151.06324	2-phenylglycine*
16.42	365.15291	131.09458	leucine
17.14	354.07017	240.02369	cystine*
19.18	363.13733	129.07900	isoleucine
20.14	365.15300	131.09468	norleucine*
20.50	307.11091	73.05259	lysine
24.28	400.12146	166.06313	desaminotyrosine*
24.34	368.08575	268.05486	homocystine*
24.88	328.10047	94.04215	phenol*
25.72	342.11597	108.05764	cresol*
28.24	384.16295	150.10462	thymol*

*metabolite not previously reported to be present in BALF.

Table 5.2 Absolute quantification of metabolites in normal rat BALF

Metabolite name	Regression equation	R ²	LOQ (nM)	Linear range (μM)	Conc. (μM)	Std dev	CV % (n=3)
proline	y = 6.298x + 0.669	0.996	5	0.005-5	3.93	0.26	6.6
glycine	y = 4.622x + 1.078	0.994	5	0.005-5	3.64	0.56	15
glutamic acid	y = 5.275x + 1.107	0.993	50	0.05-5	3.39	0.28	8.3
serine	y = 3.800x + 0.494	0.998	5	0.005-5	1.46	0.26	18
alanine	y = 5.63x + 0.641	0.995	5	0.005-5	1.37	0.08	5.9
threonine	y = 4.399x + 0.925	0.993	50	0.05-5	0.751	0.073	9.7
aspartic acid	y = 5.437x + 0.646	0.994	50	0.05-5	0.537	0.098	18
leucine	y = 6.019x + 0.497	0.997	5	0.005-5	0.531	0.025	4.7
isoleucine	y = 5.790x + 0.443	0.999	50	0.05-5	0.262	0.039	15
asparagine	y = 6.656x + 0.411	0.996	5	0.005-5	0.212	0.012	5.6
methionine	y = 6.314x + 0.573	0.999	50	0.05-5	0.1027	0.007	7.6
sarcosine	y = 9.730x + 0.025	0.999	5	0.005-5	0.0100	0.001	16
desaminotyrosine	y = 9.078x + 0.016	0.999	5	0.005-5	0.008	0.001	18
citrulline	y = 7.034x + 0.177	0.996	50	0.05-0.5	0.083	0.015	18
phenol	y = 8.776x + 0.057	0.999	5	0.005-0.5	0.0067	0.000	9.2
homoarginine	y = 5.740x + 0.689	0.997	50	0.05-5	<0.05		
methylhisamine	y = 13.331x + 1.966	0.995	50	0.05-5	<0.05		
aminoadipic acid	y = 6.821x + 0.810	0.999	50	0.05-5	<0.05		
2-aminobutyric acid	y = 6.213x + 0.603	0.999	50	0.05-5	<0.05		
2-phenylglycine	y = 5.715x + 1.334	0.993	50	0.05-5	<0.05		
norleucine	y = 6.272x + 0.609	0.997	50	0.05-5	<0.05		
thymol	y = 6.692x + 0.725	0.998	50	0.05-5	<0.05		
leu-pro	y = 8.497x + 0.110	0.997	50	0.05-0.5	<0.05		
cresol	y = 6.275x + 0.494	0.998	5	0.005-5	5	<0.00	

Table 5.3 Metabolites dysregulated in a rat model of allergic airways inflammation.

Metabolites	RT (min s)	Mass dansylated	Mass underivatized	Mass accuracy (ppm)	T test (P value)	Fold change	VIP score	Remarks
valine	11.93	351.1376 2	117.07929	2.30	1.50E-02	3.4	1.70	definitive
4-aminobutanal	13.22	321.1269 3	87.06861	2.20	1.68E-06	3.2	1.40	putative
4-hydroxybenzaldehyde	24.63	356.0956 6	122.03734	4.55	1.50E-03	4.4	1.35	definitive
2-phenylglycine	16.92	385.1220 7	151.06374	2.73	5.00E-03	3.3	1.34	definitive
formamide	11.43	279.0798 4	45.02151	1.06	1.90E-04	5.3	1.30	putative
asparagine	4.51	366.1120 6	132.05373	1.80	3.20E-04	4.1	1.29	definitive
proline	13.59	349.1218 3	115.06351	1.56	3.50E-05	2.7	1.23	definitive
hydroxyproline	7.49	365.1167 5	131.05842	1.35	6.80E-03	2.4	1.19	definitive
alanine	10.37	323.1061 6	89.04783	1.72	8.30E-05	2.6	1.19	definitive
threonine	8.14	353.1167 6	119.05843	1.56	1.40E-05	2.4	1.17	definitive
xanthine	11.99	386.0919 8	152.03365	1.50	1.60E-03	-3.9	1.11	definitive
arginine	3.61	408.1703 3	174.11200	1.86	4.00E-03	2.3	1.10	definitive
ammonia	8.22	251.0848 2	17.02650	3.00	1.80E-02	1.8	1.10	definitive

Table 5.4 Pearson correlation coefficients comparing metabolite levels and leucocyte numbers in bronchoalveolar lavage fluids.

Metabolite	Total cell number	Eosinophils	Neutrophils	Macrophages
formamide	0.88	NS	0.911	0.479
proline	0.85	0.334	0.883	0.424
ammonia	0.799	0.626	0.79	NS
threonine	0.793	NS	0.84	NS
4-aminobutanol	0.751	0.482	0.73	0.398
valine	0.719	0.448	0.64	0.535
alanine	0.711	0.446	0.716	NS
2-phenylglycine	0.676	0.528	0.583	0.484
arginine	0.641	0.485	0.6	0.343
hydroxyproline	0.638	0.57	0.607	NS
asparagine	0.637	0.44	0.613	NS
4-hydroxybenzaldehyde	0.608	0.366	0.546	0.446
xanthine	-0.423	-0.343	-0.386	NS

*NS no significant correlation.

*Moderate correlations are values of -0.5 to -0.3 (negative correlations) or 0.3 to 0.5; strong correlations are values of -1.0 to -0.5 (negative correlations) or strong correlations 0.5 to 1.

Table 5.5 Pearson correlation coefficients comparing metabolite levels and the percentages of different leucocytes in bronchoalveolar lavage fluids.

metabolite name	Eosinophils %	Neutrophils %	Macrophages %
formamide	NS	0.882	-0.815
proline	NS	0.832	-0.78
threonine	NS	0.816	-0.765
valine	0.409	0.779	-0.82
4-aminobutanal	0.377	0.775	-0.8
asparagine	0.331	0.767	-0.779
alanine	0.36	0.738	-0.764
2-phenylglycine	0.346	0.712	-0.74
4-hydroxybenzaldehyde	0.317	0.707	-0.724
ammonia	0.463	0.702	-0.771
arginine	NS	0.653	-0.664
hydroxyproline	0.497	0.652	-0.743
xanthine	NS	-0.438	-0.48

*NS no significant correlation.

*Moderate correlations are values of -0.5 to -0.3 (negative correlations) or 0.3 to 0.5; strong correlations are values of -1.0 to -0.5 (negative correlations) or strong correlations 0.5 to 1.

Supplemental Table T 5.1. Comparison of discriminant metabolites found in OPLS-DA and PLS-DA.

OPLS-DA				PLS-DA			
metabolite ID	VIP score	identified metabolite	remark	metabolite ID	VIP score	identified metabolite	remark
314	1.70	<i>valine</i>		894	1.86		
496	1.63			78	1.64		*
894	1.61			613	1.59		
613	1.58			314	1.52	<i>valine</i>	
311	1.54			371	1.51		*
1000	1.45			496	1.43		
1109	1.45			311	1.34		
189	1.40	<i>4-aminobutanal</i>		739	1.29		
1104	1.39			1000	1.29		
1062	1.35			1109	1.21		
307	1.35			1104	1.20		
344	1.35	<i>4-hydroxybenzaldehyde</i>		189	1.19	<i>4-aminobutanal</i>	
540	1.34	<i>2-phenylglycine</i>		1287	1.17		
1227	1.33			1227	1.17		
1287	1.33			304	1.17	<i>proline</i>	
261	1.32			344	1.16	<i>4-hydroxybenzaldehyde</i>	
48	1.30	<i>formamide</i>		193	1.16	<i>alanine</i>	
448	1.30			48	1.16	<i>formamide</i>	
414	1.29	<i>asparagine</i>		261	1.16		
895	1.28			448	1.15		
1266	1.28			414	1.15	<i>asparagine</i>	
1278	1.27			895	1.15		
1221	1.26			342	1.15		
1060	1.25			966	1.15		
754	1.25		*	1216	1.15		*
342	1.24			1062	1.15		
304	1.23	<i>proline</i>		288	1.14		
230	1.24			307	1.14		
1146	1.23			1278	1.13		
680	1.22			540	1.13	<i>2-phenylglycine</i>	
966	1.20			87	1.13		
288	1.20			1266	1.12		
400	1.19	<i>hydroxyproline</i>		322	1.12	<i>threonine</i>	
193	1.19	<i>alanine</i>		546	1.12	<i>xanthine</i>	
516	1.18			230	1.11		
322	1.17	<i>threonine</i>		60	1.11		*
87	1.15			313	1.10		*
130	1.15			618	1.09		
83	1.12			400	1.09	<i>hydroxyproline</i>	
546	1.11	<i>xanthine</i>		369	1.09		
649	1.10	<i>arginine</i>		649	1.08	<i>arginine</i>	
2	1.10		*	1221	1.08		
5	1.10	<i>ammonia</i>		130	1.08		
369	1.09			680	1.08		
739	1.09			1060	1.08		
405	1.06		*	83	1.08		
1168	1.01		*	516	1.08		
618	1.00			1146	1.06		
				257	1.06		*
				120	1.05		*
				5	1.05	<i>ammonia</i>	

There are 48 metabolites with VIP score larger than 1.0.

44 out of 48 are shared by OPLS-DA and PLS-DA models.

Common metabolites in both OPLS-DA and PLS-DA are highlighted in yellow and the unique metabolites are indicated by *

Supplemental Table T 5.2 Quantification results of two metabolites using the internal standard method and validated using the standard addition method.

Metabolite	Internal standard method (calibration curve)			Standard addition method		
	BALF pooled sample (μM)	Standard deviation	CV% (n=3)	BALF pooled sample (μM)	Standard deviation	CV% (n=3)
glycine	3.64	0.56	15	3.30	0.26	7.8
alanine	1.37	0.08	5.9	1.13	0.11	9.6

Supplemental Table T 5.3 Total cell and differential cell count numbers in the rat BALF samples.

groups	Sample number	Total Cells	Macrophages	Neutrophils	Eosinophils
normal	1	5.6	5.1	0.2	0.2
	24	5.3	5.2	0.0	0.0
	30	10.5	10.2	0.2	0.1
	35	10.1	9.7	0.1	0.4
	37	8.8	8.8	0.0	0.0
	38	8.5	8.5	0.0	0.0
	4	4.0	4.0	0.0	0.0
	5	7.0	7.0	0.0	0.0
	6	7.5	7.3	0.1	0.0
	12	10.0	10.0	0.0	0.0
	14	23.5	23.3	0.0	0.1
	15	10.0	9.6	0.0	0.4
	17	9.0	9.0	0.0	0.0
	18	4.5	4.5	0.0	0.0
	25	1.0	0.9	0.1	0.0
26	6.5	6.5	0.0	0.0	
41	16.4	15.9	0.3	0.2	
low inflammation subgroup	43	19.5	16.9	2.2	0.3
	42	20.9	15.4	3.2	2.2
	27	27.0	13.5	5.0	8.4
	28	24.0	6.3	5.1	12.6
	40	21.3	11.9	8.2	1.2
	20	21.5	11.2	8.7	1.6
	39	24.4	14.5	9.0	0.9
	9	25.5	15.3	9.5	0.2
	21	29.5	14.1	11.1	4.3
	19	51.0	21.5	28.5	1.1
8	36.0	19.7	15.8	0.5	
high inflammation subgroup	33	44.5	25.8	18.4	0.3
	11	55.5	15.8	39.3	0.4
	23	92.0	17.4	72.3	2.3
	31	41.0	14.2	23.5	3.3
	22	54.0	14.1	36.9	3.0
	32	40.0	12.2	20.3	7.5
	34	64.5	17.3	30.9	16.3

References

1. Locksley, R. M., Asthma and allergic inflammation. *Cell* **2010**, 140, (6), 777-783.
2. Galli, S. J.; Tsai, M.; Piliponsky, A. M., The development of allergic inflammation. *Nature* **2008**, 454, (7203), 445-454.
3. Adamko, D. J.; Sykes, B. D.; Rowe, B. H., The metabolomics of asthma: novel diagnostic potential. *Chest* **2012**, 141, (5), 1295-302.
4. Connett, G. J., Bronchoalveolar lavage. *Paediatr. Respir. Rev.* **2000**, 1, (1), 52-56.
5. Reynolds, H. Y., Use of bronchoalveolar lavage in humans - Past necessity and future imperative. *Lung* **2000**, 178, (5), 271-293.
6. Meyer, K. C., The role of bronchoalveolar lavage in interstitial lung disease. *Clinics in Chest Medicine* **2004**, 25, (4), 637-640.
7. WALTERS, E.; WARD, C.; Li, X., Bronchoalveolar lavage in asthma research. *Respirology* **1996**, 1, (4), 233-245.
8. Szeffler, S. J.; Wenzel, S.; Brown, R.; Erzurum, S. C.; Fahy, J. V.; Hamilton, R. G.; Hunt, J. F.; Kita, H.; Liu, A. H.; Panettieri, R. A.; Schleimer, R. P.; Minnicozzi, M., Asthma outcomes: Biomarkers. *J. Allergy Clin. Immunol.* **2012**, 129, (3), S9-S23.
9. Atzei, A.; Atzori, L.; Moretti, C.; Barberini, L.; Noto, A.; Ottonello, G.; Pusceddu, E.; Fanos, V., Metabolomics in paediatric respiratory diseases and bronchiolitis. *J. Matern.-Fetal Neonatal Med.* **2011**, 24, 60-63.
10. Kominsky, D. J.; Campbell, E. L.; Colgan, S. P., Metabolic shifts in immunity and inflammation. *J. Immunol.* **2010**, 184, (8), 4062-4068.
11. Saude, E. J.; Obiefuna, I. P.; Somorjai, R. L.; Ajamian, F.; Skappak, C.; Ahmad, T.; Dolenko, B. K.; Sykes, B. D.; Moqbel, R.; Adamko, D. J.,

Metabolomic biomarkers in a model of asthma exacerbation urine nuclear magnetic resonance. *Am. J. Respir. Crit. Care Med.* **2009**, 179, (1), 25-34.

12. Saude, E. J.; Skappak, C. D.; Regush, S.; Cook, K.; Ben-Zvi, A.; Becker, A.; Moqbel, R.; Sykes, B. D.; Rowe, B. H.; Adamko, D. J., Metabolomic profiling of asthma: Diagnostic utility of urine nuclear magnetic resonance spectroscopy. *J. Allergy Clin. Immunol.* **2011**, 127, (3), 757-764

13. Carraro, S.; Rezzi, S.; Reniero, F.; Heberger, K.; Giordano, G.; Zanconato, S.; Guillou, C.; Baraldi, E., Metabolomics applied to exhaled breath condensate in childhood asthma. *Am. J. Respir. Crit. Care Med.* **2007**, 175, (10), 986-990.

14. Serkova, N. J.; Van Rheen, Z.; Tobias, M.; Pitzer, J. E.; Wilkinson, J. E.; Stringer, K. A., Utility of magnetic resonance imaging and nuclear magnetic resonance-based metabolomics for quantification of inflammatory lung injury. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **2008**, 295, (1), L152-L161.

15. Wolak, J. E.; Esther, C. R.; O'Connell, T. M., Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients. *Biomarkers* **2009**, 14, (1), 55-60.

16. Hu, J. Z.; Rommereim, D. N.; Minard, K. R.; Woodstock, A.; Harrer, B. J.; Wind, R. A.; Phipps, R. P.; Sime, P. J., Metabolomics in lung inflammation: A high-resolution H-1 NMR study of mice exposed to silica dust. *Toxicol. Mech. Methods* **2008**, 18, (5), 385-398.

17. Azmi, J.; Connelly, J.; Holmes, E.; Nicholson, J. K.; Shore, R. F.; Griffin, J. L., Characterization of the biochemical effects of 1-nitronaphthalene in rats using global metabolic profiling by NMR spectroscopy and pattern recognition. *Biomarkers* **2005**, 10, (6), 401-416.

18. Fabiano, A.; Gazzolo, D.; Zimmermann, L. J. I.; Gavilanes, A. W. D.; Paolillo, P.; Fanos, V.; Caboni, P.; Barberini, L.; Noto, A.; Atzori, L., Metabolomic analysis of bronchoalveolar lavage fluid in preterm infants complicated by respiratory distress syndrome: preliminary results. *J. Matern.-Fetal Neonatal Med.* **2012**, 24, ((S2)), 56-59.

19. Ho, W. E.; Xu, Y. J.; Xu, F. G.; Cheng, C.; Peh, H. Y.; Tannenbaum, S. R.; Wong, W. S. F.; Ong, C. N., Metabolomics reveals altered metabolic pathways in experimental asthma. *Am. J. Respir. Cell Mol. Biol.* **2013**, 48, (2), 204-211.
20. Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R., Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets. *Metabolomics* **2012**, 8, (1), S29-S36.
21. Guo, K.; Li, L., Differential C12-/C13-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal. Chem.* **2009**, 81, (10), 3919-3932.
22. Zheng, J. M.; Dixon, R. A.; Li, L., Development of isotope labeling LC-MS for human salivary metabolomics and application to profiling metabolome changes associated with mild cognitive impairment. *Anal. Chem.* **2012**, 84, (24), 10802-10811.
23. Wu, Y.; Li, L., Development of isotope labeling liquid chromatography-mass spectrometry for metabolic profiling of bacterial cells and its application for bacterial differentiation. *Anal. Chem.* **2013**, 85, (12), 5755-63.
24. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, 78, (3), 779-787.
25. Xia, J. G.; Psychogios, N.; Young, N.; Wishart, D. S., MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, 37, W652-W660.
26. Li, L.; Li, R. H.; Zhou, J. J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y. M.; Huan, T.; Zheng, J. M.; Shi, Y.; Wishart, D. S.; Lin, G. H., MyCompoundID: using an evidence-based metabolome library for metabolite identification. *Anal. Chem.* **2013**, 85, (6), 3401-3408.

27. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J.; Forsythe, I., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37*, D603-D610.
28. Xia, J. G.; Wishart, D. S., MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics* **2010**, *26*, (18), 2342-2344.
29. Guo, K.; Bamforth, F.; Li, L. A., Qualitative metabolome analysis of human cerebrospinal fluid by C-13-/C-12-isotope dansylation labeling combined with liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, (2), 339-347.
30. Wu, M. Y., S.; Zurek, G.; Barsch, A., Metabolomics studies on yeast arginine synthesis pathway mutants using maXis impact with ionBooster. *Application Note #ET-33 Bruker Daltonik GmbH, Bremen, Germany.* **2013**.
31. Mattarucchi, E.; Baraldi, E.; Guillou, C., Metabolomics applied to urine samples in childhood asthma; differentiation between asthma phenotypes and identification of relevant metabolites. *Biomed. Chromatogr.* **2012**, *26*, (1), 89-94.
32. Cheng, C.; Ho, W. E.; Goh, F. Y.; Guan, S. P.; Kong, L. R.; Lai, W. Q.; Leung, B. P.; Wong, W. S. F., Anti-Malarial drug attenuates experimental allergic asthma via inhibition of the phosphoinositide 3-Kinase/Akt pathway. *PLoS One* **2011**, *6*, (6), 9.
33. King, N. E.; Rothenberg, M. E.; Zimmermann, N., Arginine in asthma and lung inflammation. *J. Nutr.* **2004**, *134*, (10), 2830S-2836S.
34. Ou, X. M.; Feng, Y. L.; Wen, F. Q.; Huang, X. Y.; Xiao, J.; Wang, K.; Wang, T., Simvastatin attenuates bleomycin-induced pulmonary fibrosis in mice. *Chin. Med. J.* **2008**, *121*, (18), 1821-1829.

35. Zeki, A. A.; Bratt, J. M.; Rabowsky, M.; Last, J. A.; Kenyon, N. J., Simvastatin inhibits goblet cell hyperplasia and lung arginase in a mouse model of allergic asthma: a novel treatment for airway remodeling? *Transl. Res.* **2010**, 156, (6), 335-349.
36. Benson, R. C.; Hardy, K. A.; Morris, C. R., Arginase and arginine dysregulation in asthma. *J. Allergy* **2011**, 2011, (2011), 736319.
37. Zimmermann, N.; King, N. E.; Laporte, J.; Yang, M.; Mishra, A.; Pope, S. M.; Muntel, E. E.; Witte, D. P.; Pegg, A. A.; Foster, P. S.; Hamid, Q.; Rothenberg, M. E., Dissection of experimental asthma with DNA microarray analysis identifies arginase in asthma pathogenesis. *J. Clin. Invest.* **2003**, 111, (12), 1863-1874.
38. Lara, A.; Khatri, S. B.; Wang, Z.; Comhair, S. A. A.; Xu, W.; Dweik, R. A.; Bodine, M.; Levison, B. S.; Hammel, J.; Bleecker, E.; Busse, W.; Calhoun, W. J.; Castro, M.; Chung, K. F.; Curran-Everett, D.; Gaston, B.; Israel, E.; Jarjour, N.; Moore, W.; Peters, S. P.; Teague, W. G.; Wenzel, S.; Hazen, S. L.; Erzurum, S. C., Alterations of the arginine metabolome in asthma. *Am. J. Respir. Crit. Care Med.* **2008**, 178, (7), 673-681.
39. Seddon, P.; Bara, A.; Ducharme, F. M.; Lasserson, T. J., Oral xanthines as maintenance treatment for asthma in children. *Cochrane Database Syst Rev.* **2006**, (1), 131.
40. Kamath, A. V.; Pavord, I. D.; Ruparelia, P. R.; Chilvers, E. R., Is the neutrophil the key effector cell in severe asthma? *Thorax* **2005**, 60, (7), 529-530.
41. Jatakanon, A.; Uasuf, C.; Maziak, W.; Lim, S.; Chung, K. F.; Barnes, P. J., Neutrophilic inflammation in severe persistent asthma. *Am. J. Respir. Crit. Care Med.* **1999**, 160, (5), 1532-1539.
42. Monteseirin, J., Neutrophils and asthma. *J. Invest. Allergol. Clin. Immunol.* **2009**, 19, (5), 340-354.

43. Fahy, J. V., Eosinophilic and neutrophilic inflammation in asthma: insights from clinical studies. *Proc. Am. Thor. Soc.* **2009**, 6, (3), 256-9.
44. Lacoste, J. Y.; Bousquet, J.; Chanez, P.; Vanvyve, T.; Simonylafontaine, J.; Lequeu, N.; Vic, P.; Enander, I.; Godard, P.; Michel, F. B., Eosinophilic and neutrophilic inflammation in asthma, chronic-bronchitics, and chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* **1993**, 92, (4), 537-548.

Chapter 6

Summary and future work

6.1 Summary

Before I give a brief summary of each chapter, I would like to describe the contributions of my thesis to the mass spectrometry based metabolomics research in a big picture. When I started my PhD study, our lab has successfully developed a differential ^{13}C -/ ^{12}C -dansylation labeling LC-MS method for targeting amines and phenol compounds, and a ^{13}C -/ ^{12}C - dimethylaminophenacyl labeling method for carboxylic acids. Those developed methods laid an excellent foundation for my thesis. In order for those methods to be applicable in real metabolomics applications, automated data preprocessing is essential. I worked together with computer scientists to develop in-house written software. My thesis firstly contributed to the development of an automated data preprocessing method in combination with chemometrics tools for isotope labeling LC-MS metabolomics. This automated data preprocessing method was demonstrated to be useful in my research projects and also research projects of other group members. In Chapter 2, I contributed to the method development of ion pairing reversed phase LC-UV as the first dimensional fractionation including the choice of an ion pairing reagent and optimization of the concentration of ion pairing agent, flow rate, and injection amount. The automated data preprocessing software, which I developed, was used to find the ion pairs in this project. In chapter 3, the ^{13}C -/ ^{12}C -dimethylaminophenacyl labeling reagent was developed in our lab; however, the labeling method has some drawbacks which need to be overcome. The specificity

of the labeling method is not good because the amines and water interfere with the organic acids signals. The method had not been applied to any real metabolomics project yet. My contribution is that I developed a new labeling method by combining LLE with this DmPA labeling, which greatly improved the performance of the method. The new method was successfully used in metabolomics application project by other group members. In Chapter 4, I contributed to the development of isotope labeling LC-MS for low volume mouse urine metabolomics. This is the first application of isotope dansylation labeling LC-MS in a real metabolomics project. I demonstrated that urine metabolomics was useful to study mice model of Alzheimer's disease and I found some interesting metabolite biomarkers in our animal study which were also discovered in previous human studies. In Chapter 5, I contributed to the development of isotope labeling LC-MS method for rat BALF and this method allows us to perform more comprehensive metabolomics profiling (250 putative metabolites compared with previous studies of 30 metabolites using NMR). The data showed that there was clear metabolome change between control rats and inflamed rats and arginine-proline metabolic pathway was dysregulated in rats model of allergic inflammation. The metabolomics study generated interesting results, which will be beneficial to study experimental asthma.

In Chapter 1, I gave the overview of the basic concepts of metabolomics, LC-MS techniques, isotope labeling methods, and the data analysis in metabolomics study.

In Chapter 2 we developed a novel two-dimensional (2D) separation strategy which was based on the use of ion-pairing (IP) reversed-phase LC as the first dimension separation to fractionate the metabolites, followed by isotope labeling of individual fractions using dansylation chemistry to alter the physiochemical properties of the metabolites. The labeled metabolites having different hydrophobicity from their unlabeled counterparts are then separated and analyzed by on-line LC-FTICR-MS. This off-line 2D-LC-MS strategy offers significant improvement over the one-dimensional LC-MS technique in terms of the number of detectable metabolites. As an example, in the analysis of a human urine sample, 3564 ^{13}C -/ ^{12}C -dansylated ion pairs or metabolites were detected from seven IP RPLC fractions, compared to 1218 metabolites found in 1D-RPLC-MS. Using a library of 220 amine- and phenol-containing metabolite standards, 167 metabolites were positively identified based on retention time and accurate mass matches, which was about 2.5 times the number metabolites identified by 1D-RPLC-MS analysis of the same urine sample.

In Chapter 3, we developed an improved method for detecting organic acid metabolites. This method is based on the use of liquid-liquid extraction (LLE) to selectively extract the organic acids, followed by using differential isotope *p*-dimethylaminophenacyl (DmPA) labeling of the acid metabolites. The ^{12}C -/ ^{13}C -labeled samples are analyzed by liquid chromatography Fourier-transform ion cyclotron resonance mass spectrometry (LC-FTICR-MS). It is shown that this LLE DmPA labeling method offers superior performance over the method of direct DmPA labeling of biofluids such as human urine. LLE of organic acids

reduces the interference of amine-containing metabolites that may also react with DmPA. It can also remove the interference of water that is present in a biofluid. Using human urine as an example, it is demonstrated that about 2500 peak pairs or putative metabolites could be detected in a 30-min gradient LC-MS run, which is about 3 times more than that detected in a sample prepared using direct DmPA labeling. Most of the 1000 or so matched metabolites to the Human Metabolome Database are organic acids.

In Chapter 4, a differential isotope labeling LC-MS method was developed for mouse urine metabolomics, which allowed relative quantification of over 950 putative metabolites using 20 μ L of mouse urine as the starting material. This method was applied to a metabolomic biomarker discovery study using urine samples obtained from the TgCRND8 mouse model of early-onset familial Alzheimer's disease throughout the course of their pathological deposition of beta amyloid ($A\beta$). A number of metabolites were found to be differentially expressed and could be used to separate the mutant and the control groups.

In Chapter 5, an isotope labeling LC-MS method was developed for metabolomic profiling of bronchoalveolar lavage fluid (BALF) with improved metabolite detectability. This method was applied to investigate metabolomic changes in BALF samples between 18 control rats and 18 rats with allergic inflammation. Statistical analysis of the resultant data showed that there was distinct separation between control and inflamed rats. Metabolic pathway analysis implicated that the arginine-proline metabolic pathway was probably dysregulated in rats with asthma.

6.2 Future work

There has been increased research interest in the past few years during my Ph.D study and this increase will continue in the coming years. Metabolomics holds potential promise in many applications areas, especially in biomarker discovery for disease diagnosis and monitoring of toxicity in drug discovery. However, metabolomics is still a young field compared with genomics or proteomics and the technology itself still needs to advance. There is no single analytical technique which is capable of doing comprehensive metabolome profiling. To further improve the metabolome coverage, there is a need for combined techniques or methods. One of the future directions is that we can combine the dansylation labeling for targeting amine or phenol compounds with dimethylphenacyl labeling for targeting organic acids in one study. For example, we are able to routinely detect and quantify 2000 putative metabolites in one human urine sample by combining two isotope labeling methods.

Both LC-MS untargeted method and targeted method have advantages and disadvantages. The combined two methods in one study can be explored in the future to synergize both strengths. They could be performed in two different approaches in study design. One approach is that we could perform the untargeted analysis using LC-high resolution MS full scan, at the same time, we perform the targeted method using LC-MRM, in which 100-200 common metabolites can be quantified in a single LC-MRM run. Another approach is that we firstly do the untargeted method using LC-high resolution MS, identify the most interesting metabolite biomarkers, and then we perform the targeted analysis using LC-MRM

for those significant metabolites. The combined untargeted method and targeted method will improve the data quality and reduce the false positive metabolite biomarkers.

Metabolite identification is the biggest technical bottleneck. Comprehensive database on accurate MS and MS/MS of authentic metabolites is needed. The usefulness of retention time information obtained from LC may be explored further. Retention index of metabolites in GC is useful to identify metabolites; likewise, it will be helpful to use the retention time information to facilitate identifying metabolites in LC-MS. For example, we could establish the relationship between retention time of dansylated metabolite in commonly used LC column and its chemical structure using some mathematical model. Then we could predict the potential chemical structure using the retention time information. Although this prediction may not be necessarily accurate, it will be useful especially when accurate mass and MS/MS database searching have several matches.