

Development and Application of Chemical Isotope Labeling Methods and Metabolite
Identification Solution for Liquid Chromatography-Mass Spectrometry-Based
Metabolomics

by

Shuang Zhao

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

© Shuang Zhao, 2018

Abstract

Metabolomics, the comprehensive analysis of small molecules in biological specimens, has become an emerging and indispensable tool for systems biology and clinical research. Liquid chromatography-mass spectrometry (LC-MS) is a dominant analytical platform for metabolomics, featuring high metabolite detectability, accurate quantification ability and good versatility. However, quantitative metabolomic analysis with very high coverage is still a challenge due to great diversity of metabolites and their wide concentration ranges. Traditional approach of increasing coverage is to combine lists of metabolites detected by several complementary analytical methods (e.g., combined use of reversed phase LC and hydrophilic interaction LC). Alternatively, chemical isotope labeling (CIL) LC-MS method has been developed to improve the overall analytical performance for metabolomics.

My research focuses on 1) improving CIL LC-MS analysis power to eventually realize high-performance and comprehensive profiling of the entire metabolome and 2) enhancing high-confidence metabolite identification for metabolomics.

In the first part of my thesis, three novel CIL LC-MS methods for comparative metabolomics study were developed targeting hydroxyl submetabolome (Chapter 2), carbonyl submetabolome (Chapter 3) and carboxyl submetabolome (Chapter 4), respectively. These labeling methods significantly increased metabolite detection sensitivity and improved LC separation efficiency. Using a pair of isotope reagents, very accurate and precise relative quantification could be achieved. The application of these methods was validated in the analysis of urinary submetabolomes. To perform positive metabolite identification, labeled standard libraries were constructed for each reaction.

Each method, used alone, should enable in-depth analysis of the corresponding submetabolome. In combination with other CIL LC-MS methods, high-coverage metabolome profiling could be carried out.

In the second part, the idea that using multiple CIL LC-MS methods to perform high-coverage near-complete metabolome profiling was validated (Chapter 5) and applied for biomarker discovery (Chapter 6). The newly developed methods were integrated with previously reported CIL LC-MS methods to form a multichannel labeling technique. In Chapter 5, human plasma metabolome was analyzed to demonstrate the high performance of this technique. In Chapter 6, the technique was applied in differential biomarker discovery of Alzheimer's disease and cerebral amyloid angiopathy. The method successfully differentiated disease groups from healthy controls. A panel of metabolites was selected as biomarker candidates for each disease and between diseases, showing good discriminative ability. It demonstrated that multichannel CIL LC-MS approach is a powerful tool for metabolome profiling and biomarker discovery.

Lastly, to improve the metabolite identification in metabolomics, in Chapter 7, a high-resolution MS/MS-retention time (RT) library was constructed using 825 human endogenous metabolites. Based on the library, a convenient metabolite identification solution was developed for real world sample analysis. The performance and portability were validated by analyzing various biological samples in different laboratories. The approach was proved to be a useful and powerful tool for endogenous metabolite identification with high confidence.

Preface

A version of Chapter 2 was published as: Shuang Zhao, Xian Luo and Liang Li, 2016, “Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics”, *Anal. Chem.*, 88, 10617-10623. I was responsible for the experimental design, data collection and analysis, as well as the manuscript writing. Xian Luo contributed towards experimental design and data collection. Professor Liang Li supervised the project and edited the manuscript.

A version of Chapter 3 was published as: Shuang Zhao, Margot Dawe, Kevin Guo and Liang Li, 2017, “Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome”, *Anal. Chem.*, 89, 6758-6765. I was responsible for the experimental design, data collection and analysis, as well as the manuscript preparation. Margot Dawe and Kevin Guo were involved with concept discussion and experimental design. Professor Liang Li supervised the project and edited the manuscript.

A version of Chapter 4 was submitted to *Analytical Chemistry* as: Dansylhydrazine Isotope Labeling LC-MS for Comprehensive Carboxylic Acid Submetabolome Profiling. I was responsible for the experimental design, data collection and analysis, as well as the manuscript writing. Professor Liang Li supervised the project and edited the manuscript.

Chapter 5 was finished by me and Hao Li. I was responsible for experimental design, sample analysis and data processing. Hao Li carried out the chemical group diversity analysis.

Chapter 6 was a collaborative project with Dr. Eric Smith at the University of Calgary and Dr. Roger Dixon in Department of Psychology, University of Alberta. I was responsible for experimental design, sample analysis and data processing. Dr. Eric Smith collected samples and contributed in data analysis.

Chapter 7 was an international collaboration with Dr. Aiko Barsch and Ulrike Schweiger Hufnagel from Bruker Daltonics (Bremen, Germany). Professor Liang Li, Dr. Aiko Barsch, Dr. Mingguo Xu, Dr. Zhendong Li, Dr. Nan Wang, Xian Luo and I were responsible for the project design. Dr. Zhendong Li was in charge of library construction in positive ion mode acquisition. I and Xian Luo were in charge of library construction in negative ion mode acquisition. Dr. Mingguo Xu, Dr. Zhendong Li, I, Xian Luo, Dr. Nan Wang, Jaspaul Tatlay, Tran Tran Ngoc, Dr. Tao Huan, Dr. Yiman Wu, Dr. Ruokun Zhou, Dr. Chiao-Li Tseng, Dr. Wei Han, Yunong Li, Kevin Hooton, Dorothea Mung, Xiaohang Wang, and Adriana Zardini Buzatto contributed towards the MS/MS spectra collection. I, Xian Luo, Wan Chan, and Dr. Zhendong Li contributed towards the retention time collection. Dr. Aiko Barsch and Ulrike Schweiger Hufnagel provided technique support and carried out portability test in Germany. I finalized the library, developed the solution and analyzed biological samples. The project was presented as an oral talk at the 66th ASMS Conference, Jun 6th, 2018, San Diego, US. The spectral-RT database, standard operating procedures of handling real world samples such as human urine and LC-MS setups are a major part of the T-Rex Metabolomics Solution introduced by Bruker as a commercial product at the 2018 ASMS conference.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Liang Li, who has supported me throughout my study with his patient guidance and immense knowledge. It has been a great pleasure and an honour to engage in research on mass spectrometry and metabolomics under his direction. I enjoyed each time we discussed research project, solved a challenging problem, and imagined the future plan. His passion and enthusiasm for research not only encouraged me for past and current thesis study, but also will inspire my future life.

I would also like to thank my supervisory committee, Professor Mark McDermott and Professor Sarah Styler, and other members of my examining committee, Professor Arthur Mar and Professor Jeff Smith, for their time attending my oral examination and reviewing my thesis. I am grateful for their valuable comments and suggestions for my thesis work. Thanks to Professor Bob Jordan for agreeing to be the non-examining chair of my oral examination. I would also like to thank Professor John Vederas for attending my candidacy examination and providing insightful advice on my research projects.

I am also deeply thankful to all the collaborators. Specifically, I thank Dr. Aiko Barsch and Ulrike Schweiger Hufnagel from Bruker Daltonics for their professional and timely technical supports. Thanks for answering every question with patience and details. The progress of MS/MS-RT library project that has been made would have been impossible without their efforts. I would like to thank Professor Eric Smith for providing clinical samples and knowledgeable support for biomarker discovery.

I really appreciate all the help and support from my labmates in Li group. It has been a very enjoyable and enriching experience to work with them. Special thanks for their great contributions in MS/MS-RT library construction. In no particular order, thanks to: Xian Luo, Dr. Nan Wang, Dr. Chiao-Li, Tseng, Dorothea Mung, Dr. Zhendong Li, Wan Chan, Hao Li, Yunong Li, Dr. Yiman Wu, Dr. Liyan Liu, Dr. Wei Han, Dr. Tao Huan, Kevin Hooton, Jaspaul Tatlay, Xiaohang Wang, Adriana Zardini Buzatto, Xinyun Gu and Minglei Zhu. Special thanks to Monica Li for valuable comments of my thesis.

My sincere thanks also go to Professor David Wishart, Professor John Vederas, Professor Todd Lowary, and Professor Derrick Clive, for providing many of the compound standards used in my research. Thanks to Dr. Randy Whittal, Jing Zheng, and Béla Reiz in mass spectrometry lab for their kind and knowledgeable assistance in instrument maintenance and repair. I also acknowledge the electronics shop, Kim Do and Andrew Hillier, for guiding and helping me to build and fix many electronics. Thanks to machine shop, Dirk Kelm, Paul Crothers, and Dieter Starke, for all completed projects. Special thanks to Ryan Lewis for all the help with sample shipping and receiving.

I would like to thank Alberta Advanced Education and Alberta Innovates for providing Alberta Innovates Graduate Student Scholarship to support me to pursue my study here. Also thanks to Mrs. Margaret Harris and Mr. David Jones, for their generous scholarship support.

Last but not the least, I would like to express my heartiest thanks to my parents: Mr. Yingjun Zhao and Mrs. Wenjing Gan for their unconditional love and sacrifice. It was their support and encouragement that made this thesis possible.

Table of Contents

Chapter 1 Introduction.....	1
1.1. Metabolomics.....	1
1.2. LC-MS based Metabolomics: Conventional Approach and Chemical Isotope Labeling LC-MS Approach	2
1.3. Rationales and Technical Considerations of CIL LC-MS Approach	6
1.3.1. From Quantification Perspective	6
1.3.2. From Separation Perspective	7
1.3.3. From Detection Perspective.....	10
1.3.4. From Stability Perspective.....	11
1.3.5. From Metabolite Picking Perspective.....	12
1.4. Derivatization Reagents for CIL LC-MS Metabolomics.....	13
1.4.1. Reagents for amine-containing metabolites derivatization.....	13
1.4.2. Reagents for carboxyl-containing metabolites derivatization.....	18
1.4.3. Reagents for carbonyl-containing metabolites derivatization.....	20
1.4.4. Reagents for hydroxyl-containing metabolites derivatization	21
1.4.5. Reagents for thiol-containing metabolites derivatization	23
1.5. Limitations and Future Directions of CIL LC-MS approach.....	28
1.6. Overview of Thesis	30
Chapter 2 Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics.....	32
2.1. Introduction.....	32
2.2. Experimental Section	33
2.2.1. Workflow	33
2.2.2. Chemicals and Reagents	35
2.2.3. Standard Solution Preparation	36
2.2.4. Urine Sample Collection and Preparation.....	36
2.2.5. Dansylation Labeling Reaction.....	36
2.2.6. LC-UV Quantification	37
2.2.7. LC-MS Analysis	37

2.2.8.	Data Processing.....	38
2.3.	Results and Discussion	39
2.3.1.	Reaction Condition Optimization with Standard.....	39
2.3.2.	Reaction Condition Optimization with Urine Sample	41
2.3.3.	Stability Test of Labeled Samples	43
2.3.4.	Sample Quantification and Normalization.....	44
2.3.5.	Hydroxyl Submetabolome Profiling	47
2.3.6.	Hydroxyl Metabolite Identification.	51
2.4.	Conclusions.....	59
Chapter 3 Development of High-Performance Chemical Isotope Labeling LC-MS		
for Profiling the Carbonyl Submetabolome		
60		
3.1.	Introduction.....	60
3.2.	Experimental Section	62
3.2.1.	Principle and Workflow	62
3.2.2.	Chemicals and Reagents	64
3.2.3.	Synthesis of Dansylhydrazine.....	65
3.2.4.	Solution Preparation.....	68
3.2.5.	Urine Sample Collection and Preparation.....	68
3.2.6.	Dansylhydrazine Labeling Reaction.	68
3.2.7.	LC-MS and MS/MS	69
3.2.8.	Data Processing and Analysis.....	70
3.3.	Results and Discussion	70
3.3.1.	Standard-Mixture Labeling.....	70
3.3.2.	Urine Sample Labeling	75
3.3.3.	Accuracy and Precision.....	78
3.3.4.	Product Stability.....	80
3.3.5.	Dansylhydrazine-labeled Standard Library	81
3.3.6.	Urine Carbonyl Submetabolome Profiling	85
3.4.	Conclusions.....	90
Chapter 4 Dansylhydrazine Isotope Labeling LC-MS for Comprehensive		
Carboxylic Acid Submetabolome Profiling.....		
91		

4.1.	Introduction.....	91
4.2.	Experimental Section.....	93
4.2.1.	Chemicals and Urine Sample.....	93
4.2.2.	Preparation of Solution.....	94
4.2.3.	Chemical Isotope Labeling.....	94
4.2.4.	LC-MS Analysis.....	95
4.2.5.	Data Processing.....	96
4.3.	Results and Discussion.....	97
4.3.1.	Principle of DnsHz Labeling LC-MS Method.....	97
4.3.2.	Optimization of Labeling Reaction.....	99
4.3.3.	Validation of Labeling Conditions with Urine Sample.....	104
4.3.4.	Metabolite Detectability.....	107
4.3.5.	Quantification.....	108
4.3.6.	Stability of the Labeled Samples.....	111
4.3.7.	Metabolite Identification.....	112
4.4.	Conclusions.....	122
Chapter 5 Multichannel Chemical Isotope Labeling Mass Spectrometry for High-Coverage Quantitative Metabolomics.....		123
5.1.	Introduction.....	123
5.2.	Experimental Section.....	125
5.2.1.	Chemical Group Diversity Analysis.....	125
5.2.2.	Principle and Workflow of CIL LC-MS Methods.....	126
5.2.3.	Amine/Phenol Submetabolome Profiling.....	129
5.2.4.	Carboxyl Submetabolome Profiling.....	129
5.2.5.	Hydroxyl Submetabolome Profiling.....	130
5.2.6.	Carbonyl Submetabolome Profiling.....	130
5.2.7.	LC-UV Quantification and Normalization.....	131
5.2.8.	LC-MS Analysis.....	131
5.2.9.	Data Processing.....	132
5.3.	Results and Discussion.....	133
5.3.1.	Chemical Group Diversity Analysis.....	133

5.3.2.	Benefits of high-performance CIL LC-MS methods for metabolomics.	135
5.3.3.	Metabolite detection.....	136
5.3.4.	Metabolite identification.....	140
5.3.5.	Metabolite quantification.....	141
5.4.	Conclusions.....	153
Chapter 6 Comprehensive Profiling of Cerebral Amyloid Angiopathy and Alzheimer’s Disease Using In-depth Metabolomic Analysis for Biomarkers		
Discovery.....		154
6.1.	Introduction.....	154
6.2.	Experimental Section.....	156
6.2.1.	Sample Collection and Metabolite Extraction.....	156
6.2.2.	Multichannel Chemical Isotope Labeling.....	157
6.2.3.	LC-MS Analysis.....	159
6.2.4.	Data Processing.....	160
6.2.5.	Statistical Analysis.....	161
6.3.	Results and Discussion.....	161
6.3.1.	Comprehensive Metabolomic Analysis of Plasma.....	161
6.3.2.	Metabolomic Comparisons of All Groups.....	166
6.3.3.	Comparative Metabolome Analysis of AD and Control.....	168
6.3.4.	Comparative Metabolome Analysis of CAA and Control.....	172
6.3.5.	Comparative Metabolome Analysis of CAA and AD.....	175
6.3.6.	Biological Pathway Analysis.....	178
6.4.	Conclusion.....	181
Chapter 7 Construction and Application of a High-resolution MS/MS-RT Library for Rapid Identification of Endogenous Metabolites in Metabolomics.....		
		183
7.1.	Introduction.....	183
7.2.	Experimental Section.....	185
7.2.1.	Chemicals and Reagents.....	185
7.2.2.	Standard Solution Preparation.....	185
7.2.3.	Quality Control and Retention Time Calibrants.....	186
7.2.4.	MS/MS Spectra Acquisition.....	187

7.2.5.	Retention Time Acquisition	187
7.2.6.	Biological Sample Preparation and Analysis.....	189
7.2.7.	Data Processing.....	190
7.3.	Results and Discussion	191
7.3.1.	Construction of MS/MS Library.....	191
7.3.2.	Construction of RT Library	196
7.3.3.	Metabolite Identification Solution	199
7.3.4.	Analysis of Biological Samples	202
7.3.5.	Portability Test.....	209
7.3.6.	Proof-of-concept Study.....	210
7.4.	Conclusions.....	212
Chapter 8 Conclusion and Future Work		214
8.1.	Thesis Summary.....	214
8.2.	Future Work	218
Reference		221
Appendix.....		238

List of Figures

- Figure 1.1 The overview scheme of (A) conventional LC-MS based metabolomics approach and (B) chemical isotope labeling LC-MS based metabolomics approach. 4
- Figure 1.2 The general workflow of using chemical isotope labeling (CIL) combined with LC-MS for relative or absolute quantification in metabolomics..... 6
- Figure 2.1 Workflow of the base-activated dansylation isotope labeling LC-MS technique for in-depth relative quantification of the hydroxyl submetabolome for metabolomics..... 35
- Figure 2.2 (A) LC-UV chromatogram of dansyl labeled 1-propanol. (B) Comparison of labeling efficiencies under different labeling reaction conditions: from left to right: incubation solvent, reaction time, reaction temperature, DnsCl/DMAP mole ratio, and reaction quenching method. Data were presented as mean \pm S.D. from three independent experiments (n=3). 40
- Figure 2.3 Peak pair numbers detected from ^{12}C -/ ^{13}C -dansyl labeled urine samples prepared under different labeling reaction conditions: (A) effect of reaction time, (B) effect of reaction temperature, (C) effect of DnsCl to DMAP mole ratio, and (D) effect of quenching method. Data are presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6). * p<0.05; ** p< 0.01.43
- Figure 2.4 (A) Sample storage stability. Labeled urine samples were stored under different temperatures for different periods of time. (B) Water stability. Labeled urine samples were dissolved with or without water. Data were presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6). ** p<0.01. 44
- Figure 2.5 (A) Absorption spectra of 10 dansyl labeled alcohols. (B) UV chromatogram of labeled urine sample. (C) Calibration curve of labeled urine sample. (D) Calibration curve of a mixture of 17-dansyl labeled amino acids..... 46
- Figure 2.6 (A) Peak pair number detected as a function of injection amount of a ^{12}C -/ ^{13}C -labeled human urine (n=3). (B) Ion chromatogram (IC) of labeled urine. (C) IC

of a mixture of 22 labeled standards containing amine or hydroxyl group. *	
p<0.05.....	48
Figure 2.7 (A) Venn diagram of the peak pair numbers detected in triplicate analysis of labeled urine. Number distribution of peak pairs as a function of (B) averaged peak ratio and (C) RSD (n=9). (D) Classes of labeled metabolites.	51
Figure 2.8 Examples of entries in the labeled hydroxyl standards library: (A) IC and (B) MS/MS spectrum of dansyl labeled 1-phenyl-1-propanol, (C) IC and (D) MS/MS spectrum of dansyl labeled 3-phenyl-1-propanol. Each entry contains accurate mass, retention time, ion chromatogram, and MS/MS spectrum obtained using RPLC-QTOF-MS.....	54
Figure 3.1 Workflow of ¹² C-/ ¹³ C-dansylhydrazine labeling for relative quantification of the carbonyl submetabolome differences in comparative samples. The dansylhydrazine labeling reaction scheme is also shown.	64
Figure 3.2 ¹ H NMR spectrum of (A) ¹² C-dansylhydrazine standard and (B) synthesized ¹³ C-dansylhydrazine.	66
Figure 3.3 LC-MS analysis of ¹² C-dansylhydrazine standard and synthesized ¹³ C-dansylhydrazine: (A) LC ion chromatogram of ¹³ C-dansylhydrazine, (B) LC ion chromatogram of ¹² C-dansylhydrazine, (C) MS spectrum of ¹³ C-dansylhydrazine, and (D) MS spectrum of ¹² C-dansylhydrazine.....	67
Figure 3.4 (A) Scheme of ¹² C- or ¹³ C-dansylhydrazine synthesis. (B) Structures of six carbonyl-containing metabolites used to prepare a standard-mixture for method development and evaluation.....	71
Figure 3.5 Comparison of efficiency and reproducibility for labeling a standard-mixture under different reaction conditions: (A) effect of solvent type, (B) effect of acid concentration, (C) effect of acid type, (D) effect of reaction temperature, and (E) effect of reaction time. Data are presented as the mean ± S.D. of six replicates (n=6).....	72
Figure 3.6 Peak pair numbers detected from ¹² C-/ ¹³ C-dansylhydrazine labeled urine samples prepared under different labeling conditions: (A) effect of sample preparation, (B) effect of acid concentration, (C) effect of acid type, (D) effect of temperature, (E) effect of time, and (F) effect of dansylhydrazine amount.	

Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).	77
Figure 3.7 Venn diagrams of the peak pair numbers detected in triplicate labeled urine from (A) the intra-day experiments and (D) the inter-day experiments. Distributions of peak pair numbers as a function of averaged peak ratio of (B) intra-day results and (E) inter-day results and RSD of (C) intra-day results and (F) inter-day results. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).	79
Figure 3.8 Storage condition test.	81
Figure 3.9 (A) LC-MS ion chromatogram of RT calibrants consisting of 17 labeled standards. (B) LC-MS ion chromatogram of ^{12}C -/ ^{13}C -dansylhydrazine labeled urine. (C) Peak pair number detected as a function of injection amount of ^{12}C -/ ^{13}C -dansylhydrazine labeled urine. Data are presented as mean \pm S.D. (n=3).	86
Figure 4.1 Workflow of differential CIL LC-MS method using ^{12}C -/ ^{13}C -DnsHz for carboxyl submetabolome analysis.	98
Figure 4.2 Structures of seven carboxyl-containing metabolites used to prepare a standards mixture for method development and evaluation.	100
Figure 4.3 Comparison of efficiency for labeling a standards mixture under different reaction conditions: (A) effect of buffer solution and pH, (B) effect of EDC concentration, (C) effect of additive type, (D) effect of HOAT concentration, (E) effect of reaction temperature, and (F) effect of reaction time. Data are presented as the mean \pm S.D. of six replicates (n=6).	102
Figure 4.4 (A) LC-UV chromatograms of DnsHz and CuCl_2 mixture for different incubation times at 40 °C. (B) LC-MS chromatograms of DnsHz incubated with or without CuCl_2	104
Figure 4.5 Peak pair numbers detected from ^{12}C -/ ^{13}C -DnsHz labeled urine samples prepared under different conditions: (A) effect of buffer solution and pH, (B) effect of EDC concentration, (C) effect of additives type, (D) effect of HOAT concentration, (E) effect of temperature and (F) effect of time. Data are	

presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).	106
Figure 4.6 (A) LC-MS ion chromatogram of ^{12}C -/ ^{13}C -DnsHz labeled urine. (B) Peak pair number detected as a function of injection amount of ^{12}C -/ ^{13}C -DnsHz labeled urine. Data are presented as mean \pm S.D. (n=3). (C) Venn diagrams of the peak pair numbers detected in triplicate labeled urine. (D) Peak pair number detected as a function of peak intensity (black: all peak pairs; red: peak pairs with peak ratio >1.25 or <0.8).	108
Figure 4.7 Distributions of peak pair numbers as a function of (A) averaged peak ratio and (B) RSD of peak ratio. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6). (C) Peak ratios of all pairs as a function of their absolute signal intensities.	110
Figure 4.8 (A) Short term stability test of labeled urine samples. (B) Long term stability test of labeled urine samples. Data are presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6).	111
Figure 5.1 Workflow of chemical group diversity analysis	126
Figure 5.2 Workflow of 4-channel CIL LC-MS for in-depth relative quantification of the plasma metabolome.	127
Figure 5.3 Reaction schemes of (A) dansylation labeling for amine/phenol-containing metabolites; (B) DmPA bromide labeling for carboxylic acid-containing metabolites; (C) base-activated dansylation labeling for hydroxyl-containing metabolites; (D) dansylhydrazine labeling for carbonyl-containing metabolites.	128
Figure 5.4 Classification of chemical groups of (A) MyCompoundID library, (C)ECMDB, (E)YMDB and (G) HMDB; Percent distributions of metabolites belonging to the four groups that are analyzed using the 4-channel CIL-LC-MS approach in (B) MyCompoundID library, (D) ECMDB, (F) YMDB and (H) HMDB.	134
Figure 5.5 Base peak ion chromatograms of CIL LC-MS method profiling (A) Amine/phenol submetabolome; (B) Carboxylic acid submetabolome; (C) Carbonyl submetabolome; (D) Hydroxyl submetabolome.	137

Figure 5.6 (A) Ion chromatograms of 22 dansyl labeled metabolites with different structures. (B) List of 22 dansyl labeled metabolites.	138
Figure 5.7 Peak pair number detected as a function of injection amount of ^{12}C -/ ^{13}C -labeled plasma of (A) Amine/phenol submetabolome; (B) Carboxylic acid submetabolome; (C) Carbonyl submetabolome; (D) Hydroxyl submetabolome. Data are presented as mean \pm S.D. (n=3).	139
Figure 5.8 (A) Percentage of peak pair detected in four submetabolome profiling methods as a function of peak intensity. (B) Venn diagram of the numbers of peak pairs or metabolites detected in four methods.....	140
Figure 5.9 Distributions of peak pair numbers as a function of (A) averaged peak ratio and (B) RSD. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=3).....	142
Figure 6.1 (A) PCA, (B) 2D PLS-DA and (C) 3D PLS-DA scores plots of metabolomics dataset from three groups, including Alzheimer's disease (AD, in red). cerebral amyloid angiopathy (CAA, in green) and control group (in blue). QC samples were shown in PCA as cyan dots. R^2 and Q^2 values given by cross-validation for PLS-DA are: 0.998 and 0.902, respectively. (D) Permutation test of PLS-DA model for the dataset. (E) Clustered heatmap of relative intensity comparison of important peak pairs from three groups.	167
Figure 6.2 (A) PLS-DA scores plot of dataset from AD (in red) and control group (in green). $R^2 = 0.990$; $Q^2 = 0.897$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between AD and control group with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in AD and control group.....	169
Figure 6.3 (A) ROC curve generated by the random forest model to differentiate AD and control using 5 metabolite biomarker candidates: methionine sulfoxide, arginine, desaminotyrosine, 3-hydroxybutyric acid and glutaric acid. (B) Permutation test result for the ROC curve of AD vs. control.....	172

- Figure 6.4 (A) PLS-DA scores plot of dataset from CAA (in red) and control group (in green). $R^2 = 0.978$; $Q^2 = 0.760$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between CAA and control group with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in CAA and control group..... 173
- Figure 6.5 (A) ROC curve generated by the random forest model to differentiate CAA and control using 5 metabolite biomarker candidates: methionine sulfoxide, arginine, asparagine, aspartic acid and glutamic acid. (B) Permutation test result for the ROC curve of CAA vs. control. 174
- Figure 6.6 (A) PLS-DA scores plot of dataset from CAA (in green) and AD (in red). $R^2 = 0.981$; $Q^2 = 0.802$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between CAA and AD with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in CAA and AD group..... 175
- Figure 6.7 (A) ROC curve generated by the random forest model to differentiate CAA and AD using five positively identified candidates: benzaldehyde, hypoxanthine, glyceraldehyde, 4-hydroxybenzoic acid and glutaric acid. (B) Permutation test result for the ROC curve of CAA vs. AD using positively identified metabolites. (C) A series of ROC curves generated with different numbers of putative identified metabolites in the list on the right. (D) ROC curve generated using top three putatively identified metabolites and (E) its permutation test results..... 177
- Figure 6.8 Overview of pathway analysis in (A) AD vs. control comparison; (B) CAA vs. control comparison; (C) CAA vs. AD comparison. Each circle represents a matched pathway. The color of circle is based on its p value and the radius is determined by its pathway impact values. Four significant pathways are labeled. (D) Box plots of six significantly altered amino acids. AD: in red; CAA: in green; control group: in blue..... 180

Figure 6.9 (A) Schematic view of butanoate metabolism. Compounds were represented by KEGG numbers. Matched metabolites were shown as nodes with varied heat map colors based on p-values (red indicates lower p-value). The information of matched metabolites was listed in the table below. (B) Box plots of five significant metabolites in butanoate metabolism. AD: in red; CAA: in green.	181
Figure 7.1 Effect of different additives on signal intensity of (A) citric acid, (B) traumatic acid, (C) adenosine 3',5'-diphosphate, (D) diacetyl and (E) maltose in negative ion mode detection.	193
Figure 7.2 MS/MS spectra of citric acid using different additives in mobile phase.	194
Figure 7.3 A typical LC-MS/MS data of one metabolite for MS/MS library construction.	195
Figure 7.4 LC ion chromatogram and list of RT-CalMix containing 28 components.	198
Figure 7.5 (A) Principle of retention time normalization approach. (B) Distribution of RTs and their standard deviations of metabolites in the library.	199
Figure 7.6 The workflow of metabolite identification solution.	200
Figure 7.7 Metabolite identification using constructed MS/MS-RT library in Metaboscape of (A) creatine in positive ion mode and (B) citric acid in negative ion mode.	201
Figure 7.8 LC-MS/MS chromatograms and identification results of (A) human urine sample, (B) human plasma sample and (C) yeast cell extracts in both positive and negative ion mode. “Filtered” features are the ones after removing redundant spectral features (adduct ions, multimers, etc.).	203
Figure 7.9 Identification of xanthosine in human urine using MS/MS and RT matches.	204
Figure 7.10 Inter-lab portability test of the metabolite identification solution.	209
Figure 7.11 Application of the developed solution in a proof-of-concept experiment. The study compared the metabolome differences of urine samples before and after coffee intake. (A) PCA scores plot of LC-MS/MS data in positive ion mode. (B) PCA scores plot of LC-MS/MS data in negative ion mode. (C) Volcano plot showing significant metabolites in two groups. (D) Box plots of three significant metabolites in two groups as example.	211

List of Tables

Table 1.1 Summary of common reagents for CIL LC-MS based metabolomics method.	25
Table 2.1 List of 10 labeled alcohols analyzed by LC-UV.	45
Table 2.2 List of RT calibrants consisting of 20 labeled standards.	49
Table 2.3 List of 85 compounds in the current dansyl hydroxyl standards library.	52
Table 2.4 List of hydroxyl metabolites identified based on accurate mass and retention time matches to the dansyl hydroxyl standards library.	57
Table 3.1 Signal comparison between unlabeled metabolites and labeled metabolites.	74
Table 3.2 List of 90 entries from 78 compounds for the DnsHz-labeled standard library.	82
Table 3.3 List of RT calibrants consisting of 17 labeled standards.	85
Table 3.4 List of metabolites identified based on accurate mass and retention time matches to the DnsHz-labeled standard library.	88
Table 4.1 List of 193 compounds in the current labeled DnsHz-carboxyl standard library.	113
Table 4.2 List of carboxylic acid metabolites identified based on accurate mass (10 ppm) and retention time (30 s) matches to the DnsHz labeled carboxyl standard library.	118
Table 5.1 List of metabolites identified based on accurate mass and retention time matches to the labeled standard library.	143
Table 6.1 List of positively identified metabolites based on accurate mass and retention time match to the labeled standard library.	162
Table 6.2 List of positively identified significant metabolites in comparison of AD and control.	170
Table 6.3 List of positively identified significant metabolites in comparison of CAA and control.	173
Table 6.4 List of positively identified significant metabolites in comparison of CAA and AD.	176
Table 7.1 List of 177 metabolites that were identified in human urine	204

List of Abbreviations

ACN	Acetonitrile
AD	Alzheimer's Disease
ANOVA	Analysis of Variance
AUC	Area-under-the-curve
A β	beta-Amyloid Protein
BPC	Base Peak Chromatogram
CAA	Cerebral Amyloid Angiopathy
CE	Capillary Electrophoresis
CIC	Chemical Isotope Coding
CID	Collision Induced Dissociation
CIL	Chemical Isotope Labeling
CSF	Cerebrospinal Fluid
Da	Dalton
DCC	N,N'-Dicyclohexylcarbodiimide
DMAP	4-Dimethylaminopyridine
DmPA	p-Dimethylaminophenacyl
DnsCl	Dansyl Chloride
DnsHz	Dansylhydrazine
ECMDB	E. coli Metabolome Database
EDC	1-Ethyl-3-dimethylaminopropylcarbodiimide
ESI	Electrospray Ionization
FA	Formic Acid
FC	Fold Change
GC	Gas Chromatography
h	Hour
HILIC	Hydrophilic Interaction Liquid Chromatography
HMDB	Human Metabolome Database
HOAT	1-Hydroxy-7-azabenzotriazole
IC	Ion Chromatogram
iTRAQ	The IsobaricTag for Relative and Absolute Quantification
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LLE	Liquid-Liquid Extraction
m/z	Mass to Charge
MALDI	Matrix-Assisted Laser Desorption/Ionization
MCID	MyCompoundID
mCIL	Multichannel Chemical Isotope Labeling

MeOH	Methanol
MES	2-(N-morpholino)ethanesulfonic acid
min	Minute
MP	Mobile Phase
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffered Saline
PCA	Principle Component Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
ppm	part(s) per million
QC	Quantity Control
Q-TOF	Quadrupole Time-of-Flight
ROC	Receiver Operating Characteristic
RPLC	Reversed Phase Liquid Chromatography
RSD	Relative Standard Deviation
RT	Retention Time
RT-CalMix	Retention Time Calibration Mixture
s	Second
S/N	Signal to Noise Ratio
SD	Standard Deviation
SIL	Stable Isotope Labeled
SRM	Selected Reaction Monitoring
TEA	Triethylamine
TFA	Trifluoroacetic Acid
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
VIP	Variable Importance on the Projection
YMDB	Yeast Metabolome Database

Chapter 1

Introduction

1.1. Metabolomics

The concept of “metabolome”, which was introduced twenty years ago, refers to a complete set of small chemical species (50-1000 Da) within a biological sample.¹ However, the attempt to use analytical approaches to perform more or less comprehensive metabolic profiling of a biological system has lasted for long time, which is the initial idea of metabolomics.²⁻⁴

As the development of systems biology, “omics” studies become more and more important to understand biological processes and to discover the mysteries behind disease and life. Metabolome is the final downstream product of the “omics” cascade including genome, transcriptome and proteome. It also directly reflects the effect of environmental exposures on organism, such as diet and contaminants.⁵ Its close link to phenotype makes metabolomics a powerful tool to discover the molecular mechanism of biological process, to monitor the metabolic status of a biological system and to assist in the diagnosis of disease. Nowadays, metabolomics study has become an incredible growing area in both the development and application of novel methods. It has been widely used in various fields, such as biomarker discovery,⁶⁻⁷ drug development,⁸⁻⁹ medical research,¹⁰ etc. Several other related “omics” areas are also rapidly developing, including foodomics¹¹ and exposomics.¹² However in the meantime, it is also more complex than other “omics” studies due to the tremendously large variety of metabolites with wide concentration ranges.

The most common analytical detection platforms for metabolomics analysis are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). NMR provides detailed structural information, relatively simple sample preparation and highly robust and reproducible results due to low instrument drift, though suffering from less sensitivity, large sample amount requirement and convoluted metabolite signals.¹³ In contrast, MS is much more sensitive than NMR. However interpretation of complicated spectral signatures is sometimes hard and ambiguous.^{7, 14} Separation techniques are usually incorporated in MS-based metabolomics study to reduce the complexity of data and improve the detection ability, such as gas chromatography (GC), liquid chromatography (LC) and capillary electrophoresis (CE). Among all the platforms, the usage of liquid chromatography-mass spectrometry (LC-MS) to perform metabolomics study expanded rapidly due to several unique advantages, including high metabolite detectability, good compatibility with majority of samples types and accurate quantification ability.¹³

1.2. LC-MS based Metabolomics: Conventional Approach and Chemical Isotope Labeling LC-MS Approach

The ultimate goal of metabolomics is to provide qualitative and quantitative information for as many metabolites as possible. Method to achieve it requires high metabolite coverage, sensitive detection, accurate quantification and confident unknown metabolite identification ability. However, realizing this goal is hindered by the extreme complexity of biological metabolome and several practical considerations. First, metabolites with various physicochemical properties coexist in the biological samples,

which may not show acceptable responses simultaneously in one analytical instrument. Therefore combinations of instrumental platforms and/or methods are required for comprehensively profiling metabolome. Second, wide concentration ranges of metabolites (from pM to mM) worsen the detection and quantification of all metabolites at the same time. Even for a given metabolite, its concentration in one type of biological sample may vary greatly among individuals. For example, the concentrations of some metabolites are very dependent on diet intake.¹⁵ The detection and quantification of low abundant metabolites would be very hard, especially if the the metabolites have poor response to one detection instrument. Third, biological samples are usually very precious, whatever they are collected from animal models or clinical volunteers. For some biological samples, the sample amount would be very limited. Cerebrospinal fluid (CSF) is one of the best samples as reflection of brain activities. However collection of CSF is very difficult. The volume of CSF that collected from one single mouse is about 10 microliter.¹⁶ Very sensitive methods are required to analyze these samples. Finally, although many metabolites databases have been constructed during last two decades,¹⁷⁻²¹ the identification of unknown metabolites in metabolomics is still difficult due to the extreme complex matrices of the sample.

In conventional LC-MS based metabolomics approach, increasing metabolite coverage is achieved by combing different instrumental platforms to analyze metabolites according to their physical and chemical properties. A typical procedure for conventional LC-MS based comprehensive metabolomics is shown in Figure 1.1A. In this example, after sample collection and pretreatment (e.g. protein removal, metabolite extraction, etc.), reversed phase liquid chromatography (RPLC) is used for separating hydrophobic

metabolites and hydrophilic interaction liquid chromatography (HILIC) is used for separating more polar compounds in both positive ion mode and negative ion mode. In this way four LC-MS runs using different instrumental setups are required to analyze one sample. The introduction of extra cost in terms of both instrument investment and analysis time is a limitation of this approach. In addition, it is still difficult to detect many metabolites in this approach for several reasons, such as extremely low amount and poor ionization of some metabolites.

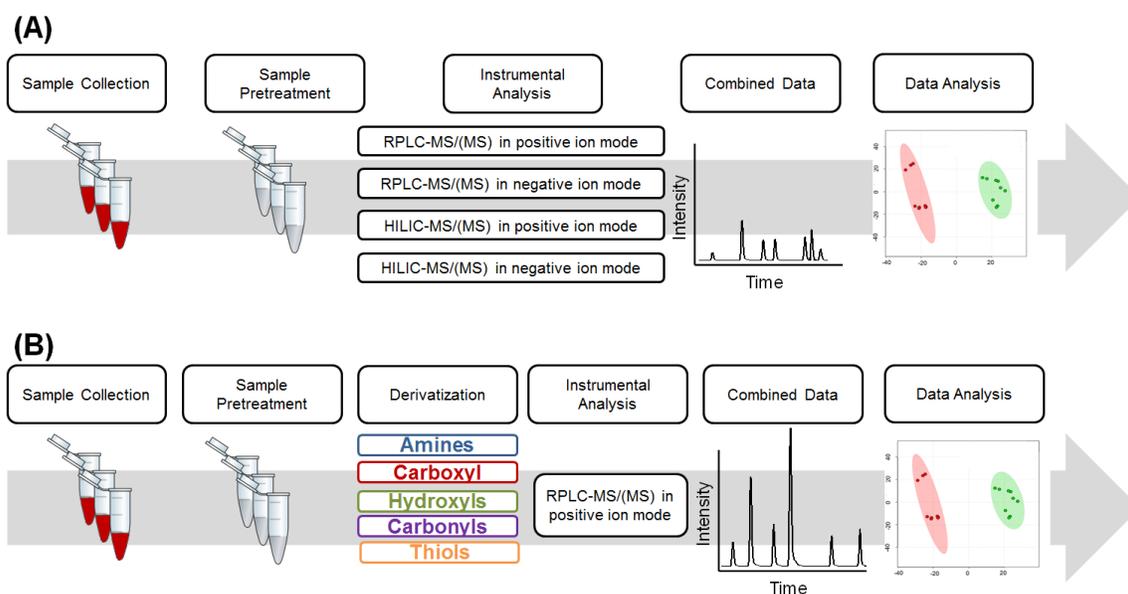


Figure 1.1 The overview scheme of (A) conventional LC-MS based metabolomics approach and (B) chemical isotope labeling LC-MS based metabolomics approach.

The quantification in conventional LC-MS analysis relies on internal standards to correct the sample loss, instrument drift and matrix effect. Compared with chemical structural analogue, stable isotope labeled (SIL) internal standard is a better choice because of its nearly identical physical and chemical properties to the interested analyte. However usually SIL internal standard is very expensive and the accessibility is limited.

Particularly in metabolomics area, it is impossible to purchase or synthesize the isotopic internal standards for all the metabolites.

To address these issues, chemical derivatization combined with LC-MS approach has been developed as a relatively new LC-MS based metabolomics approach. In this approach, the entire metabolome of samples is divided into several chemical-group-based submetabolomes. Prior to LC-MS analysis, biological samples are “in vitro” derivitized using a chemical reagent targeting specific chemical functional group. The metabolites in the samples containing certain chemical moiety react with the reagent forming new derivitized metabolites, followed by LC-MS analysis. Then the combination of submetabolome information produces comprehensive metabolome profiling (Figure 1.1B). This “divide-and-conquer” strategy has been proved to be successful in metabolome profiling when using combination of reagents that target different groups of metabolites.²²⁻²⁴

In addition, many reagents can incorporate isotopic atom (e.g. H/D, $^{12}\text{C}/^{13}\text{C}$, $^{14}\text{N}/^{15}\text{N}$, etc.) and introduce the isotopic moiety into the labeled metabolites during derivatization. Then metabolites derivitized by heavy reagents can serve as the internal standard for light reagent labeled metabolites. This chemical isotope labeling (CIL), or chemical isotope coding (CIC) strategy, has been widely use in relative quantification for untargeted metabolomics and absolute quantification for targeted metabolomics.²⁵⁻²⁶

1.3. Rationales and Technical Considerations of CIL LC-MS Approach

1.3.1. From Quantification Perspective

As mentioned above, one of the most prominent features of CIL LC-MS technique is the introduction of CIL internal standard for each labeled metabolites, which is illustrated in Figure 1.2, showing the general workflow using CIL internal standard for quantification.

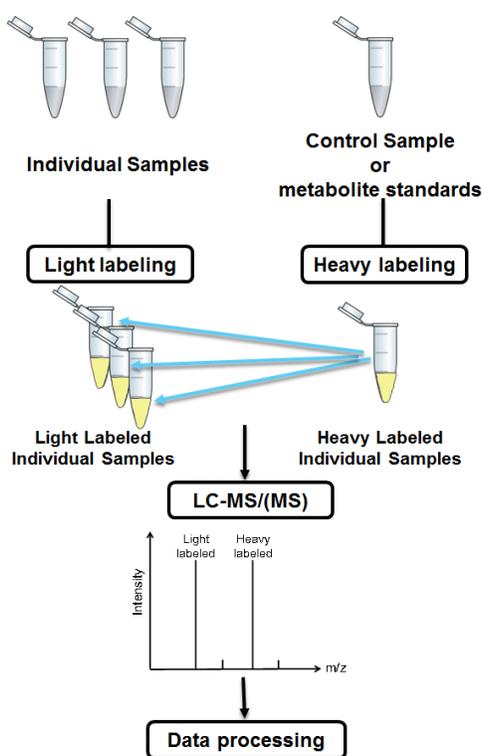


Figure 1.2 The general workflow of using chemical isotope labeling (CIL) combined with LC-MS for relative or absolute quantification in metabolomics.

In this workflow, for metabolome profiling or relative quantification in two comparative groups of samples, one group of samples (e.g. healthy or pooled control) is labeled by heavy reagent, while samples from the other group (e.g. disease or experimental group) is derivatized by heavy reagent. Then the two derivatized samples

are mixed and then injected into LC-MS for analysis. Since for certain metabolite, light labeled derivative and heavy labeled derivative have nearly identical properties, they elute out at the same time. In mass spectra, the two derivatives are shown as a peak pair. The relative amount of the metabolite can be concluded from the comparison of peak areas of two derivatives. Combined with database identification and statistical analysis, this approach has been successfully used for biomarker discovery.²⁷ For absolute quantification, similar as before, samples are derivatized with light labeling reagent and interested analyte standards of known concentration are labeled with heavy labeling. The resulting heavy derivative will be used as CIL internal standard for performing quantification with tandem mass spectrometer. By using selected reaction monitoring (SRM) or multiple reaction monitoring (MRM), the sensitivity for targeted metabolites would further increase.

1.3.2. From Separation Perspective

Ion suppression is one obstacle in detection of complex mixtures using mass spectrometry, especially in the electrospray ionization (ESI)-MS. Chemical species competes with each other to reach to the surface of droplets and get ionized. This may adversely alter the ionization efficiency of many metabolites in the ESI interface²⁸. Analytical separation techniques are usually employed in metabolomics study to reduce the ion suppression effect. Liquid chromatography is one of the most commonly used platforms. In conventional approach, reversed-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) are either solely or combined used (Figure 1.1A). However this approach suffers from poor retention of polar

compounds on RPLC, relatively poor reproducibility of HILIC separation²⁹ and multiple system requisition.

Derivatization strategy has been proved to be effective to overcome these issues and improve the separation ability of a complex biological sample. Usually derivatization introduces another chemical moiety into the original chemical species, resulting in a larger molecule. If the introduced moiety contains aromatic ring (e.g. dansyl chloride for amine/phenol/hydroxyl group derivatization³⁰⁻³¹), alkyl chain (e.g. N,N'-dimethylethylenediamine for carboxyl group derivatization³²) or parts that can block polar groups (e.g. acetone for ribonucleosides derivatization³³), the hydrophobicity of the metabolites would increase, resulting in better retention and peak shapes on RPLC column. In this way, polar metabolites within a biological sample can be retained and separated effectively even only using RPLC, avoiding the drawback of co-elution without derivatization. Many derivatization reagents for LC-MS have this benefit. Thus in CIL LC-MS based metabolomics method, only RPLC column can be used to separate the whole metabolome (Figure 1.1B).

The improvement of HILIC separation after derivatization was also reported. Bawazeer et al. found that if using d₅-aniline to tag the sugars, better separation and peak shape of monosaccharide with ZIC-HILIC column was observed.³⁴ With this tagging reagent, they can perform separation of fructose, glucose, galactose and mannose with baseline resolution and better peak shape, which is much improved compared with separation of underivatized sugar. The approach was validated to profile sugars in urine and brain samples.

Determination of chiral metabolites is another challenge in metabolome analysis, such as identification and quantification of D/L amino acids. The separation of enantiomers using liquid chromatography mainly relies on chiral chromatography, in which a specific chiral HPLC column is required.³⁵ The second common way is to use chiral derivatization reagent which contains another chiral center to react with enantiomers by generating diastereomers. Unlike enantiomers, diastereomers have different chemical properties and can be easily separated on an achiral column. Toyo'oka and co-workers have developed various chiral derivatization reagents targeting amine-³⁶⁻³⁸ and carboxyl-^{36, 39} chiral metabolites. In general procedure for chiral metabolomics, samples are derivatized by only one enantiomer (S or R) of chiral derivatization reagents. Then the labeled chiral metabolites in samples can be separated using RPLC column and quantified by MS. The authors also reported a method of using a pair of enantiomers of chiral derivatization reagents to specifically extract chiral metabolomics from complex metabolome and performing untargeted profiling.⁴⁰ In this strategy, sample is aliquoted to two parts. One part is labeled with an enantiomer of the derivatization reagent, whereas the other part is derivatized with the opposite enantiomer. The resulting diastereomers from different enantiomers labeled chiral metabolites have different elution position on the chromatograms. In contrast, achiral metabolites have the same retention times even after derivatization with different reagent enantiomers. By this way, signal from only chiral species can be extracted from complex system. Using two chiral reagents, DMT-(S or R)-Pro-OSu and DMT-3(S or R)-Apy, the authors successfully analyzed chiral amine submetabolome and chiral carboxyl submetabolome, respectively. The submetabolome

alteration was also studied in saliva of diabetic patients and brain homogenate of Alzheimer's disease.

1.3.3. From Detection Perspective

The detection of chemical species in mass spectrometry relies on the ionization of molecules in the source part of mass spectrometer. Although in conventional LC-MS based metabolomics approach, combined use of positive ion mode and negative ion mode can provide complementary information for metabolites profiling within a complex biological sample (Figure 1.1A), there are still a large part of metabolites that lack chemical moiety which can be easily ionized in the source area. Usually those metabolites are the “blind spot” of general LC-MS based metabolomics method. At the meantime using both polarity modes doubles the analysis time.

The detection improvement of derivatization can be attributes to several aspects. Firstly, as discussed above, the introduction of hydrophobic moiety into metabolites increases the retention of tagged metabolites on RPLC column. The labeled metabolites will then elute out at higher composition of organic phase, leading to more efficient desolvation process and less surface tension of droplets in ESI interface. At the meantime, increased hydrophobicity assists metabolite to reside on the surface of the droplets. Both of them benefit the ionization process. Secondly, the added tags also increase the mass of the analytes, avoiding severe interference in low m/z region. This provides significantly improvement for detecting low-molecular-weight compounds, such as amino acids.

Last but most importantly, many derivatization reactions will introduce permanently charged moiety or easily ionizable part to metabolites to enhance the

ionization of molecules. For positive ion mode, quaternary ammonium,⁴¹ pyridinium⁴² and phosphonium⁴³ salts are often used as charged moiety. Girard's Reagent P, which contains a pyridinium structure, has been widely used in derivatize carbonyl functional groups. Guo et al. reported that the detection sensitivity of steroid hormones increase 4-504 folds using Girard's Reagent P as derivatization reagent.⁴² Wang et al. presented a method using pyridine to rapidly derivatize cholesterols and fatty alcohols forming pyridinium cationic center.⁴⁴ Different from conventional reagents that already have cations on themselves, in this method the quaternary amine cation was created in the derivatization reaction. Thus the charge-neutral reagents would not affect following analytical analysis. This approach greatly improved the detection limits by 10^3 -folds with matrix-assisted laser desorption/ionization-fourier transform mass spectrometry (MALDI-FTMS). In another way, moieties processing nitrogen atoms are often designed as easily ionizable group in reagents, such as dimethylamino structure in dansyl chloride (DnsCl), which can increase proton affinity of labeled compounds, resulting in 10-1000 folds improvement of detection ability for amine/phenol-containing metabolites.³⁰ Similar example is shown as methylpyridine structure in 2-hydrazino-1-methylpyridine for carbonyl labeling (more than 100 fold increase for neurosteroids).⁴⁵

1.3.4. From Stability Perspective

Some categories of metabolites in biological samples are highly reactive and not stable during conventional sample preparation and analysis steps, such as thiol-containing metabolites⁴⁶. The sulfhydryl group is very easy to be oxidized and interact with many enzymes. The accurate quantification of primary thiol-containing metabolites, including cysteine, homocysteine, glutathione, etc., is adversely affected by the spontaneous redox

reaction in the biological system. Ortmayr et al. reported a workflow implementing thiol group protection with N-ethyl maleimide in the initial stages of sample preparation (e.g. metabolites extraction).⁴⁷ In this methods derivatization reagent N-ethyl maleimide was added into metabolites extraction solvent (e.g. methanol) when analyzing cell extracts, which can rapidly react with sulfhydryl group in three minutes. The combination of metabolites extraction and thiol protection step can avoid unwanted oxidation. Using HILIC-MS/MS platform, accurate and robust quantification of 12 sulfur pathway intermediates was successfully achieved.

1.3.5. From Metabolite Picking Perspective

Another benefit of using CIL internal standards for metabolomics is the metabolites candidates picking. Since in the MS spectra, metabolites are shown as peak pair, which can be easily distinguished from singlet noise signal or singlet unlabeled metabolites signal. Thus extraction of peak pair information can achieve selectively and high-confidence interested metabolite candidates picking. Zhou et al. developed an automatic R-based script, IsoMS, to extract peak pair information from MS spectra by batch mode.⁴⁸

Instead of directly extracting peak pairs from MS spectra, using common MS/MS fragmentation pattern is another approach to recognize metabolite candidates. Feng and co-workers developed stable isotope labeling combined with double precursor ion scan⁴⁹/double neutral loss scan⁵⁰ in MS to selectively analyze a particular group of metabolites.

1.4. Derivatization Reagents for CIL LC-MS Metabolomics

1.4.1. Reagents for amine-containing metabolites derivatization

Although the amino group in amine-containing metabolites is relatively easy to protonate in acidic mobile phase and to ionize in the ESI source, the analysis of amine submetabolome without derivatization is still troublesome. The problem arises from the high polarity and water solubility of amine-containing compounds, resulting in poor retention on RPLC column and strong matrix interference. Derivatization that can increase the hydrophobicity of molecules and provide better separation of the generated compounds is used. Common reagents that are used for derivatize amine-containing metabolites were summarized in Table 1.1. Since phenol functional groups also behave nucleophilicity in basic condition, many reagents are also suitable for phenol-containing metabolites derivatization.

The reaction between sulfonyl chloride and amine group is commonly employed in the CIL LC-MS based metabolomics. The sulfonamide can be rapidly generated in the pH range of 9 to 10 through nucleophilic addition / elimination mechanism. The rapidity of reaction and universality make these compounds suitable as derivatization reagents in metabolomics analysis. Dansyl chloride (5-(dimethylamino)-naphthalene-1-sulfonyl chloride, DnsCl) is one of the most commonly used reagents to derivatize primary, secondary amines and phenol now.³⁰ It provides all the features mentioned above that improve analytical power for amine/phenol submetabolome analysis. The tertiary amine provides ionizable moiety and aromatic naphthyl part increase the hydrophobicity of the metabolite, resulting in 10-1000 fold improvement of sensitivity and better retention for hydrophilic compounds on RPLC column. In addition the two methyl groups on tertiary

amine can introduce ^{13}C isotope into the labeled metabolite, served as internal standard for quantification. Thus only using one instrument platform (i.e. RPLC-ESI-MS in positive ion mode), many amine/phenol-containing metabolites can be detected and quantified simultaneously. The protocols of using DnsCl for metabolites profiling and biomarker discovery in various samples have been reported, including urine,⁵¹⁻⁵² blood,⁵³ serum,²⁷ sweat,⁵⁴ cerebrospinal fluid,⁵⁵ faces,^{52, 56} and cell extracts.⁵⁷⁻⁵⁸ For need of identification, 315 common endogenous metabolites were individually labeled and their accurate mass, retention times and MS/MS spectra were collected to build a labeled standard library.⁵⁹ In addition, a dipeptide specific database containing 361 dansyl labeled dipeptide was constructed and used for evaluating abundance changes of dipeptides in lung cancer.⁶⁰

Another dansyl-based reagent, 5-diethylamino-naphthalene-1-sulfonyl chloride (DensCl) was developed, in which the two methyl groups attached to the tertiary amine in DnsCl were replaced by ethyl groups.⁶¹ This change results in higher detection sensitivity since the ethyl chains offer higher hydrophobicity and stronger electron-donating propensity, resulting in higher chargeability and increased surface activity of labeled metabolites. Also it provides three differential isotope-encoded reagents, i.e., $^{12}\text{C}_4$ -, $^{12}\text{C}_2^{13}\text{C}_2$ -, and $^{13}\text{C}_4$ -DensCl, respective. In this way, it essentially increases the analysis speed in comparative metabolomics since more groups can be analyzed at the same time. The combined using of DnsCl and DensCl was also reported.⁶² In this study, ^{12}C -DnsCl labeled samples were mixed with ^{12}C -DensCl labeled standards which served as structurally analogous internal standards for quantification. Although there were retention time differences, it still minimized the quantitation deviation due to the matrix

effect and avoided using relatively expensive isotope standards or isotope labeling reagents (i.e. ^{13}C -DnsCl and ^{13}C -DensCl). Using this approach, the authors studied the tryptophan metabolism changes in vincristine induced paralysis ileus of rat model. Song and coworkers developed 10-methylacridone-2-sulfonyl chloride (MASC) and its deuterated counterpart d_3 -MASC to analyze various amine- and phenol-containing compounds, including biogenic amines,⁶³ amino acids,⁶⁴ estrogenic compounds,⁶⁵ parabens⁶⁶ and bromophenols.⁶⁷

Sharing the similar reaction mechanism with sulfonyl chloride, some acyl chloride compounds are also utilized for derivatizing amine groups in metabolomics study. Kennedy, Robert's group developed a method targeting 70 neurochemicals using HPLC-MS/MS and benzoyl chloride (BzCl) derivatization.⁶⁸ The commercially available ^{13}C -BzCl was used for chemical isotope labeled internal standards. The reagent offers very fast reaction (seconds at room temperature). This approach has been proved to be effective in various matrix, including tissue, serum, CSF, and microdialysate.

Although sulfonyl chloride or acyl chloride allows for rapid and accurate quantification of amine/phenol-containing metabolites, using MS/MS spectra of labeled metabolites to identify unknown compounds is usually a challenge. The fragmentation behavior shows similar pattern: the aromatic ring-sulfur atom bond was cleaved and release the same highest product ions. Thus some labeled metabolites lack structural information in MS/MS spectra.

Another important category compounds for amine derivatization is NHS ester-based reagent. NHS esters are highly reactive electrophiles which will be attacked easily by amine group in metabolites and then forms a stable amide bond.

The isobaric tag for relative and absolute quantification (iTRAQ) reagents has been widely used in proteomics before and is applied for analyzing amino acids⁶⁹ and other amine-containing metabolites now. Unlike mass-difference isotope coding reagent described above, the chemical isotope labeled internal standards was achieved by different way. iTRAQ reagent consists of a reporter group, a mass balance group and an amine-reactive group which could be NHS ester. Two iTRAQ reagents use different isotope coding patterns in both reporter group and balance group. The overall mass of whole reagent keeps constant. In derivatization for absolute quantification, metabolites from test samples will be labeled by reagent containing light reporter group and heavy mass balance group. While analytes standards are labeled by reagent containing heavy reporter group and light mass balance group, which was spiked into labeled test samples as internal standards for LC-MS analysis. The labeled standards and corresponding metabolites in test samples elute out at the same time and have the same m/z in mass spectra. Upon MS/MS fragmentation, the reporter group will be cleaved from the labeled compounds and produce unique product ions, which can be quantified in the SRM mode as indicator of metabolites amount in samples.⁷⁰

Other NHS ester based reagents were developed in recent years, such as N-benzoyloxysuccinimide,⁷¹ 3-aminopyridyl-N-hydroxysuccinimidyl carbamate,⁷² DIPP-L-Ala-NHS.⁷³ Zhou et al. introduced two NHS ester-based chemical isotope labeling reagents (i.e. ¹²C₂/¹³C₂-4-dimethylamino-benzoylamido acetic acid N-hydroxysuccinimide ester, ¹²C₂/¹³C₂-DBAA-NHS and ¹²C₂/¹³C₂-4-methoxybenzoylamido acetic acid N-hydroxysuccinimide ester, ¹²C₂/¹³C₂-MBAA-NHS) to profile amine submetabolome in urine sample.⁷⁴ Both of the reagents can rapidly react

with amine groups (within 10 min) and significantly improve detection. Pseudo MS³ (in-source CID plus MS/MS) spectra of 20 MBAA labeled amino acids were investigated, showing that labeled amino acids can provide almost all the fragmentation patterns as the unlabeled ones. Thus it provides a possible way to conduct structure analysis or metabolite identification based on a MS/MS spectra search.

Amine as a nucleophile can also attack the electrophilic carbon in isothiocyanates forming a stable thiourea derivative. Several isothiocyanates based reagents were applied to derivatize amine-containing metabolites. Santa examined the feasibility of 3-pyridyl isothiocyanate (Py-NCS), p-(dimethylamino)phenyl isothiocyanate (DMAP-NCS) and m-nitrophenyl isothiocyanate (NP-NCS) for derivatization.⁷⁵ All three candidates have chargeable and hydrophobic moieties. The author found that Py-NCS and DMAP-NCS can react with wide range of amines and produce intense signals, though Py-NCS has a little bit higher reactivity. Yuan et al. further used DMAP-NCS and its deuterated form (d₄-DMAP-NCS) to perform fecal metabolome profiling.²² After derivatization, the detection sensitivities of standards improved by 50-367 folds. 1057 amine metabolite candidates were found in feces of mice. Using a chemically labeled standards library constructed by the authors, 46 amine metabolites were positively identified by matching information with the MSⁿ tree (MS¹, MS², and MS³) of 118 amine compound standards. The metabolome alteration in feces between Alzheimer's disease (AD) model mice and wild type mice was comprehensively studied combined with CIL techniques that target other functional groups.

1.4.2. Reagents for carboxyl-containing metabolites derivatization

The detection of underivatized carboxylic acids using LC-MS platform is usually conducted in negative ion mode, which suffers from less sensitivity and more interference than positive ion mode, especially in a complex matrices of biological samples.⁷⁶ In order to get better retention on reversed phase column, acidifying the mobile phase is required. However the acidic reagent may introduce other background to the MS spectra. In addition, applying high electrospray capillary voltage in the negative ionization mode can lead to discharge, which will hurt system stability and reproducibility. To alleviate this, lower spray voltages are often utilized, which however decreases the detection sensitivity. Therefore derivatization techniques are used for carboxylic acids to enable more sensitive detection. Some reagents for carboxylic acids derivatization are reviewed in Table 1.1.

Many reactions for carboxylic acids derivatization are based on condensation reaction with amines. In this type of reaction, carboxyl groups in metabolites are activated by a condensation reagent (e.g. carbodiimide), followed by reacting with amine groups in derivatization reagents, generating a stable amide bond. N,N'-dimethylethylenediamine (DMED) is one of the simplest reagents which is developed in 1997.⁷⁷ It can be used together with different condensation reagents, including N,N'-dicyclohexylcarbodiimide (DCC),⁷⁷ mukaiyama reagent (2-chloro-1-methylpyridinium iodide, CMPI).⁵⁰ The dimethylamino moiety in the derivatized metabolites is readily ionized in ESI source under positive ion mode. Deuterated reagent d₄-DMED was employed together as an isotope labeling reagent.²² Feng and co-workers established a standard library containing 184 labeled carboxyl compounds. The detection sensitivities of those metabolites are increased 5-817 folds.

Other reagents that share similar mechanism have been reported. Marquis, Bryce J. reported 4-bromo-N-methylbenzylamine (4-BNMA) together with 1-ethyl-3-dimethylaminopropylcarbodiimide (EDC) to derivatize and analyze TCA intermediates.⁷⁸ The bromine's isotope pattern offers the ability to distinguish carboxylic acids. Similar to the combined use of twin labeling reagent: DnsCl and DensCl for amine labeling, another twin reagents, 5-(dimethylamino) naphthalene-1-sulfonyl piperazine (Dns-PP) and (diethylamino) naphthalene-1-sulfonyl piperazine (Dens-PP), were reported to analyze and quantify 38 free fatty acids.⁷⁹ The method of using ¹²C/¹³C-3-nitrophenylhydrazine (3-NPH) to analyzed carboxylic acids was reported. The labeled metabolites gain significant enhancement of detection in negative ion mode. It has been successfully applied in the detection and quantitation of many important categories of acids, including carboxylic acids in central carbon metabolism,⁸⁰ short-chain fatty acids⁸¹ and bile acids.⁸²

Esterification reaction was also used to derivatize carboxyl group in various metabolites. One example is 3.0 N HCl in n-butanol. It reacts with carboxylic acid to form butyl ester after adding into dried serum or plasma metabolite extract.⁸³ The alkyl butyl chain increases the hydrophobicity of metabolites, resulting in higher ionization efficiency. This reagent previously has been used to detect and quantify amino acids and other small molecules for newborn screening. Deuterated (d₉) butanol was used for isotopic butylation of human serum/plasma to achieve untargeted metabolome profiling.⁸⁴

In basic condition, the carboxylic acids will deprotonate to give more nucleophilic carboxylate anion, which can react with electrophilic group, such as alpha carbon in alkyl halide. Kevin and Li⁸⁵ reported using ¹²C-/¹³C- p-dimethylaminophenacyl (DmPA) bromide as chemical isotope labeling reagent to derivatize carboxyl-containing

metabolites and further profile the carboxylic acid submetabolome. This labeling method can enhance ESI efficiency by 2-4 orders of magnitude and also introduce an isotope tag for accurate metabolite quantification. A standard library of 113 carboxylic acid-containing metabolites was constructed by labeling standards individually and collecting retention time and MS/MS spectra.

1.4.3. Reagents for carbonyl-containing metabolites derivatization

Unlike amine or carboxyl groups, carbonyl is a neutral functional group that has low ionization efficiency in the source interface. At the meantime, many important ketones or aldehydes have very low abundance (e.g. hormones) and/or very weak retention on RPLC column (e.g. carbohydrates). Derivatization as a technique that can increase the detection limit and introduce hydrophobic moiety is used in carbonyl-containing metabolites analysis. Table 1.1 lists common reagents that used in carbonyl-containing metabolites derivatization.

Various hydrazine or hydrazide reagents have been proven to be effective on labeling ketones or aldehydes. Girard reagents, a family of hydrazides that contain a quaternary ammonium moiety, were widely employed in many carbonyl-containing metabolites. The introduced pre-charged quaternary nitrogen in the molecules can produce high-abundant gas-phase ions in the ESI process, hence improving sensitivity greatly. Griffiths et al. applied Girard reagent P (1-(carboxymethyl)pyridinium chloride hydrazide) for analyzing oxysterol combined with reversed phase solid phase extraction to remove excess of cholesterol⁸⁶. Bawazeer et al. revised the protocol by allowing the analysis in the presence of excess of cholesterol.⁸⁷ The signal enhancement factor after Girard derivatization was more than 30-folds. Girard's reagent P was also used in

quantitative glycomics with its pentadeuterated (d_5 -) counterpart.⁸⁸ In this method, reducing glycans were labeled with either nondeuterated (d_0 -) or deuterated (d_5 -) Girard's reagent P, followed by online HILIC-MS analysis to achieve rapid and sensitive relative quantitation of reducing glycans between two comparative groups.

Although compared with hydrazine, amine group that attached to an alkyl group or aryl group has relatively weaker nucleophilicity, it can still rapidly react with carbonyl groups under certain conditions. Thus several amine-containing reagents are also developed. Aniline, which is the simplest aryl amine, has been used in many metabolites profiling studies. Yang et al. first used $^{12}C_6/^{13}C_6$ -aniline to detect and quantify carbonyl-, phosphoryl-, and carboxyl-containing metabolites involved in central carbon and energy metabolism.⁸⁹ 33 metabolites were analyzed simultaneously. Bawazeer et al. used d_5 -aniline to derivatize a number of hexose and pentose isomers and achieve good separation using a ZIC-HILIC column.³⁴ Although the benzyl group in aniline can improve the hydrophobicity of derivatized metabolites, lack of pre-charged moiety limits its signal enhancement ability. 4-(2(trimethylammonio)ethoxy)benzenaminium halide (4-APC) was developed, which contained an aniline moiety for reaction with carbonyl and a quaternary ammonium group for MS sensitivity enhancement.⁹⁰ The reaction has high selectivity to aliphatic aldehydes. Tandem MS/MS spectra enable screening for aldehydes. Yuan et al. applied this reagent and deuterated 4-APC- d_4 to profile aldehyde-containing compounds in urine,⁹¹ beer,⁹¹ and wine.⁹²

1.4.4. Reagents for hydroxyl-containing metabolites derivatization

Alcoholic hydroxyl groups are generally neutral moiety in the common pH range of LC gradient. The ionization efficiency is very low that several alcohols are solvents

used in electrospray ionization. Therefore derivatization is usually needed targeting hydroxyl groups to enhance MS detection. Similar as amine group, hydroxyl group is also a nucleophile and can react with electrophile, such as sulfonyl chloride. However the less nucleophilicity make it harder to react with many labeling reagents that are used in amine-containing metabolites derivatization. For some amine labeling reagents, reaction conditions need to be adjusted to allow reaction for hydroxyl. Brief summary is shown in Table 1.1.

Reaction with DnsCl was optimized for hydroxyl metabolites labeling.⁹³ Compared with amine/phenol labeling condition, the reaction with hydroxyl need to be conducted in an aprotic solvent (e.g. CH₂Cl₂, ACN) and need other basic reagent (e.g. 4-(dimethylamino)-pyridine, N,N-Diisopropylethylamine) to facility the nucleophilic attack. Tang and Guengerich first used this reagent to characterize and quantify P450 enzyme oxidation products in human liver extracts. Generally the signal of labeled standards can be increased by 1000 fold.⁹³

Esterification with carboxylic acids was also employed for hydroxyl labeling. Xu's group developed a new derivatization reagent, 4-(dimethylamino)-benzoic acid (DMBA), together with its deuterated compound to profile hydroxyl-containing steroid hormone.⁹⁴ With this reagent, 24 steroid hormones can be detected at sub-ng/mL levels as the detection sensitivity enhanced by 10³- to 10⁴-fold. The method was successfully applied in quantifying hormones in urine samples. Woo et al. used Tris(2,4,6,-trimethoxyphenyl)phosphonium acetic acid (TMPP-AA) to profile hydroxyl-containing lipids from a human serum extract.⁴³

Unlike conventional pre-charged reagents, Cao et al. used N-alkylpyridinium quaternization approach to derivatize fatty alcohols with non-charged pyridine-d₀/d₅. In this reaction, the reagents attach N-cationic pyridinium tag onto hydroxyl compounds, resulting in satisfactory enhancement of detection sensitivity. The approach was applied in profiling fatty alcohols in thyroid tissues.⁹⁵

1.4.5. Reagents for thiol-containing metabolites derivatization

Unlike other functional groups derivatization, techniques to label thiol group need to consider not only detection enhancement and separation improvement, but also stabilization of thiol group. Thiol groups are easily oxidized either by autoxidation and disulfide formation during the sample preparation process.⁴⁶ Some reagents were reported to be used in thiol submetabolome profiling, showing both increasing ionization efficiency and stabilizing thiols ability. The reagents are listed in Table 1.1.

N-ethyl maleimide (NEM) can rapidly react with sulfhydryl group, forming irreversible thioether linkages. Ortmayr et al. developed the protocol to implement the derivatization in the very beginning of the sample preparation (i.e. metabolite extraction).⁴⁷ Combined with HILIC-MS/MS detection, sulfur pathway intermediates can be successfully detected and quantified.

A pair of isotope labeling reagents, ω-bromoacetylquinolinium bromide and ω-bromoacetylquinolinium-d₇ bromide, were synthesized and applied in untargeted thiol-containing metabolites profiling.⁹⁶ The pre-charged nitrogen significantly enhances ESI ionization efficiency. Combined with double precursor ion scan of the common product ions from labeling tag, the thiol-containing metabolites can be easily distinguished from other unlabeled metabolites signal. If using the MRM detection mode, better

quantification would be achieved.⁹⁷ This method has been used in profiling thiol submetabolome in urine,⁹⁷ beer,⁹² and fecal samples.²²

Table 1.1 Summary of common reagents for CIL LC-MS based metabolomics method.

Reagent	Isotope Reagent	Targets	Applications	Reference
DnsCl	¹³ C ₂ -DnsCl	Amines/ phenols	- Constructed library containing 273 dansylated metabolites - Profiled and positively identified metabolites in urine, blood, serum, fecal, CSF, cells - Targeted absolute quantification of 19 metabolites as asthma/CPOD potential biomarkers	Guo and Li ³⁰ Chen et al. ⁵³ Han et al. ²⁷ Xu et al. ⁵⁶ Guo et al. ⁵⁵ Khamis et al. ⁵¹
	¹³ C ₂ -DnsCl	Hydroxyls	- Characterized the P450 7A1 oxidation product in human liver extracts	Tang and Guengerich. ⁹³
DensCl	¹³ C ₂ / ¹³ C ₄ -DensCl	Amines/phenols		Zhou et al. ⁶¹
MASC	d ₃ -MASC	Amino acids	- Absolute quantification of amino acids and monoamine neurotransmitters using MRM mode	Song et al. ⁶⁴ Zheng et al. ⁹⁸
Benzoyl chloride	¹³ C ₆ -BzCl	Neurochemicals	- Absolute quantification of 70 neurochemicals	Wong et al. ⁶⁸
MBAA-NHS	¹³ C ₂ -MBAA-NHS	Amines	- Untargeted profiling of urinary amine submetabolome - Absolute quantification of amino acids in human urine	Zhou et al. ⁷⁴
DBAA-NHS	¹³ C ₂ -DBAA-NHS	Amines	- Untargeted profiling of urinary amine submetabolome - Absolute quantification of amino acids in human urine	Zhou et al. ⁷⁴
BZ-NHS	¹³ C ₆ -BZ-NHS	Amines/thiols /phenols	- Untargeted metabolome profiling of cell extracts - Positively identified 10 metabolites	Wagner et al. ⁷¹
DIPP-L-Ala-NHS	¹⁸ O ₂ -DIPP-L-Ala-NHS	Amines	- Determination of 20 L-amino acids and 10 D-amino acids	Zhang et al. ⁷³
iTRAQ	iTRAQ	Amines	- Analysis of 44 amino acids in plasma, urine and tissue	Takach et al. ⁶⁹
DiLeu	4-plex DiLeu	Amines	- Profiling and relative quantification of amine submetabolome of mouse urine	Hao et al. ⁷⁰
DMAP	d ₄ -DMAP	Amines	- Constructed library containing 118 amine compounds - Positively identified 46 amine metabolites in fecal sample	Yuan et al. ²²
Cyanuric chloride/ methylamine	Methyl-d ₃ -amine	Amines	- Determined concentrations of 27 metabolites in HepG2 cells	Lee and Chang ⁹⁹
Acetone	d ₆ -acetone	Phosphatidyl- ethanolamine	- Identified and quantified 45 PE species in rat livers using double neutral loss scan.	Wang et al. ¹⁰⁰
MPBS	d ₃ -MPBS	Amino acids	- Analysis of amino acids in newborn bloodspot	Johnson ⁴¹
DMABS	d ₃ /d ₆ -DMABS	Amino acids	- Analysis of amino acids in newborn bloodspot	Johnson ⁴¹

C _n -NA-NHS	C ₄ d ₉ -NA-NHS	Amines		Yang et al. ¹⁰¹
Methyl Acetimidate	¹³ C ₂ -Methyl Acetimidate	Amines	- Relative quantification of primary and secondary amines in Arabidopsis seed extracts	Shortreed et al. ¹⁰²
Formaldehyde	¹³ C-Formaldehyde	Amines	- Analysis of 20 amino acids and 15 amines - Profiling human urine amine-containing metabolites.	Guo et al. ¹⁰³
Acetaldehyde	d ₄ -Acetaldehyde	monoamine neurotransmitters	- Determination of neurotransmitters in brain microdialysate	Ji et al. ¹⁰⁴
PEG-OPFP	¹³ C-PEG-OPFP	Primary amines	- quantification of intracellular amino acids	Abello et al. ¹⁰⁵
TAHS	d ₃ -TAHS	Amino acids	- Determination of amino acids in rat plasma	Shimbo et al. ¹⁰⁶
L-PGA-OSu	L-PGA(d ₅)-OSu	Chiral amines	- Differential analysis of DL-amino acids in serum and yogurt	Mochizuki et al. ³⁷
DMED	d ₄ -DMED	Carboxylic acids	- Constructed library containing 184 carboxyl metabolites - Positively identification of 83 carboxyl metabolites	Yuan et al. ²²
Cholamine	d ₉ -cholamine	Carboxylic acids	- Relative quantification of fatty acids from hydrolyzed egg lipid using nanoLC	Lamos et al. ¹⁰⁷
	¹⁵ N-cholamine	Carboxylic acids	- Analysis of 48 carboxylic acids	Tayyari et al. ¹⁰⁸
3-NPH	¹³ C ₆ -3NPH	Short-chain fatty acids	- Quantification of short-chain fatty acids in human feces..	Han et al. ⁸¹
butanolic HCl	d ₉ -butanol	Carboxylic acids	- Profiling and relative quantification of human plasma metabolites	O'Maille et al. ⁸⁴
Aniline	¹³ C ₆ -Aniline	Carbonyl, phosphoryl, and carboxyl	- Quantification 33 intermediate metabolites in central carbon and energy metabolism	Yang et al. ⁸⁹
BAMP/HAMP	d ₉ -BMAP	Carboxylic acids	- Metabolome profiling of rat urine sample and positively identified 32 metabolites	Yang et al. ¹⁰⁹
BMP/CMP	d ₃ -CMP	Fatty acids	- Metabolome profiling, relative quantification and absolute quantification of human serum	Yang et al. ¹¹⁰
DmPA Bromide	¹³ C ₂ -DMPA	carboxylic acids	- Constructed library containing 113 carboxylic acid - Positively identified 51 metabolites in urine	Guo and Li ⁸⁵
HMEP	d ₅ -HMEP	Fatty acids	- Monitored changes of metabolite levels in plasma of individuals.	Koulman et al. ¹¹¹
DBD-PZ-NH ₂	d ₆ -DBD-PZ-NH ₂	Carboxylic acid	- Determination and relative quantification of fatty acids in plasma.	Tsukamoto et al. ¹¹²
DMPP	d ₆ -DMPP	carboxylic acid	- Determine trace free fatty acids in human urine and thyroid tissues.	Leng et al. ¹¹³ Leng et al. ¹¹⁴
T3	d ₂₀ -T3	Fatty acids	- Relative quantification of FAs with general MRM conditions. - Discovered FA species related to the ageing process	Tie et al. ¹¹⁵
HIQB	d ₇ -HIQB	Carbonyls	- Constructed library containing 147 carbonyl compounds - Untargeted profiling carbonyl submetabolome of human serum and fecal sample using double precursor ion scan. 12 and 50 metabolites were positively identified, respectively	Guo et al. ⁴⁹ Yuan et al. ²²

4-APC	d ₄ -4-APC	Aldehydes	Profiling aldehyde submetabolome using double neutral loss scan of urine, beer and wine samples	Zheng et al., ⁹² Yu et al. ⁹¹
Aniline	¹³ C ₆ -Aniline	Carbonyl, phosphoryl, and carboxyl	Quantification 33 intermediate metabolites in central carbon and energy metabolism	Yang et al. ⁸⁹
Girard P	d ₅ -GP	Steroid hormones	Quantified steroid hormones in human follicular fluid	Guo et al. ⁴²
	d ₅ -GP/ isobaric mass	Sterols/oxysterols	Profiling plasma sterols/oxysterols to identify inborn errors	Crick et al. ¹¹⁶
HMP	d ₃ -HMP	neurosteroid	Relative and absolute quantification of allopregnanolone and pregnenolone levels in brain	Higashi et al. ⁴⁵
T3	D3 (d ₂₀ -T3)	Fatty Aldehydes	Globally profiling of fatty aldehyde in plasma and brain tissue	Tie et al. ¹¹⁷
QAO	d ₃ - QAO	Ketosterol	Absolute quantification of ketosterol in very small volumes of plasma,	DeBarber et al. ¹¹⁸
DMBA	d ₄ -DMBA	hydroxyl-containing steroid hormone	Measurement of 17 derivatized free steroid hormones in urine	Dai et al. ⁹⁴
MDMAES	¹³ C ₄ -MDMAES	steroids	Quantitative and comparative analysis of SIRS and sepsis clinical samples.	O'Maille et al. ⁸⁴
acetone	d ₆ -acetone	Ribonucleosides	Profiled urinary metabolome and positively identified 56 ribonucleosides. Metal oxide-based dispersive SPE applied for enrichment of ribonucleosides.	Li et al., ³³ Chu et al. ¹¹⁹
pyridine and T ₂ O	d ₅ -pyridine	Steroids, fatty alcohols and carbohydrates	ten pairs of d ₀ /d ₅ ion peaks were identified as cholesterol and fatty alcohols	Wang et al., ⁴⁴ Wang et al., ¹²⁰
BQB	d ₇ -BQB	Thiols and oxidized thiols	Constructed library containing 27 thiol metabolites Profiled thiol submetabolome in urine, beer and fecal samples. Positively identified 14 and 8 thiol metabolites in fecal and urine, respectively	Yuan et al., ²² Huang et al., ¹²¹ Liu et al. ⁹⁶

1.5. Limitations and Future Directions of CIL LC-MS approach

The extra time and economic cost that are required in derivatization step is one major weakness in CIL LC-MS approach. Although the reactions that are chosen for derivatization have relatively rapid reaction speed, the required time for many reactions is still in hour-scale to ensure that the complex targeted submetabolome can be fully analyzed. One way to address this is to develop and employ reaction conditions that can achieve very fast derivatization. Bian et al used cholamine derivatization coupled with LC-MS to determine long chain-free fatty acids in complex biological samples.¹²² The derivatization step can be finished within 1 min at room temperature. 2000-fold increase of sensitivity was obtained and the limits of detection of femtogram level can be achieved. The feasibility of one-minute derivatization has been validated using both targeted quantification and untargeted profiling approach with serum samples. Further the authors discovered several metabolites that have significant differences between healthy and asthma groups.

The second solution is to use robotic liquid handling system to automatically perform derivatization to achieve high throughput. An autosampler-in-needle-derivatization technique was reported by Siegel et al.¹²³ The authors used p-toluenesulfonylhydrazine to derivatize aldehydes and ketones in a UHPLC autosampler. The labeling reagents and samples were consecutively drawn into the autosampler needle and mixed. The complete derivatization can be finished within 10 min and the solution is ready to be injected for analysis. This automatic derivatization approach has been proved to simultaneously quantify and identify molecules containing carbonyl groups.

Although many chemical isotope labeling methods can improve the detection ability of labeled metabolites, some metabolites of low abundant are still very hard to detect, especially when sample amount is limited. Therefore detection sensitivity need to be further improved. Luo et al. reported a method using CIL combined with nanoflow LC-MS for high-coverage metabolomic analysis of small numbers of MCF-7 breast cancer cells (i.e., <10000 cells).⁵⁷ Even when analyzing about 100 cells, thousands of metabolites can be detected.

In conventional LC-MS/MS based quantification approach, internal standards, either SIL or chemical structural analogues, are added to the sample in the very beginning of sample pretreatment. Therefore the internal standards and analytes share the identical preparation steps and, theoretically, the same recovery rate. Thus matrix-effect and analytes loss during sample preparation can be well corrected. A calibration curve only using standards solution (simple matrix) can still fulfill the quantitation requirement. However internal standards created by CIL approach are usually added to the samples just before instrumental analysis (LC-MS or LC-MS/MS), lacking the ability to correct recovery rates during sample preparation. To overcome it, sample pretreatment procedures need be optimized to get high recovery rate. Another approach is processing the analyte standards of known concentration in an identical or similar matrix with test samples.

Another disadvantage of using CIL internal standards is the isotope effect in chromatography when using deuterium as isotope.¹²⁴⁻¹²⁵ The retention of deuterium derivatized species is generally less than its hydrogen coded counterpart on RPLC column due to weaker hydrophobic interactions with the stationary phase. The lack of co-

elution leads to different matrix for ionization, which may be detrimental for quantification. While the isotope effect of ^{13}C -, ^{18}O - or ^{15}N - is negligible. Therefore reagents with these isotopic atoms can be perfectly used to overcome matrix effect for MS analysis.

1.6. Overview of Thesis

My research focuses on 1) improving CIL LC-MS analysis power to eventually realize high-performance and comprehensive profiling of entire metabolome and 2) enhancing high-confidence metabolite identification for metabolomics.

In the first part of my thesis, three novel CIL LC-MS methods for comparative metabolomics study were developed targeting hydroxyl submetabolome (Chapter 2), carbonyl submetabolome (Chapter 3) and carboxyl submetabolome (Chapter 4), respectively. The high performance of these methods was validated in the analysis of urinary submetabolomes.

In the second part, the idea that using multichannel CIL LC-MS methods to perform high coverage entire metabolome profiling was validated (Chapter 5) and applied for biomarker discovery (Chapter 6). In Chapter 5, human plasma metabolome was analyzed to determine the performance of this technique. In Chapter 6, the technique was demonstrated to successfully differentiate Alzheimer's disease, cerebral amyloid angiopathy and healthy controls.

In the third part, a high-resolution MS/MS-RT library was constructed and convenient metabolite identification solution was developed for improving metabolite

identification in metabolomics. The performance and portability were validated by analyzing various biological samples in different laboratories. The approach was proved to be a useful and powerful tool for endogenous metabolite identification with high confidence.

Chapter 2

Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics

2.1. Introduction

Metabolomics attempts to characterize and quantify all the small molecules found in a biological system. However, profiling all the metabolites or the metabolome is hampered by the great variety of chemical and physical properties that exist within the system to be analyzed. One strategy to address this major challenge is to classify the metabolites into several subgroups based on the presence of common functional moieties (i.e., amine, hydroxyl, etc.) and then perform in-depth analysis of the individual chemical-group-submetabolomes.^{26, 30, 85, 126} The combined data from the submetabolomes would allow the analysis of the entire metabolome with high coverage. This divide-and-conquer strategy does not require many analytical measurements. More than 95% of the chemical structures of over 8000 endogenous human metabolites in the Human Metabolome Database (HMDB)¹²⁷ contain one or more of the four functional groups: amine, carboxyl, hydroxyl and carbonyl. Thus, in principle, if we could analyze all the metabolites within these four submetabolomes, a near-complete metabolomic profile could be generated.

To analyze a chemical-group-submetabolome, one approach is to apply a chemical labeling reagent to react with a common functional group to form metabolite derivatives, followed by MS detection. In this approach, the chemical structure and reactivity of the labeling reagent are critical for sensitive detection of all the metabolites

within the submetabolome. For accurate relative quantification, a differential isotope labeling strategy can be applied, which requires the use of isotope reagents. Research on rational design of chemical labeling reagents is being actively pursued and a variety of reagents and labeling chemistries have been developed for targeted and untargeted metabolite analysis.^{37, 44, 65, 68, 70-71, 94, 99, 108, 113, 119, 128-138}

In our previous studies, we have reported the use of ^{12}C -/ ^{13}C -dansyl chloride (DnsCl) for profiling the amine/phenol submetabolome³⁰ and ^{12}C -/ ^{13}C -dimethylaminophenacyl (DmPA) bromide for profiling the carboxylic acid submetabolome.⁸⁵ However, because of low detectability in MS, the hydroxyl submetabolome is far more difficult to analyze than the other groups (e.g., several alcohols are common solvents used in electrospray ionization). For detecting hydroxyl compounds, several derivatization methods have been developed,¹³⁶ including reaction with DnsCl^{128, 139} and 4-(dimethylamino)-benzoic acid.⁹⁴ These reported studies focused on the analysis of a few compounds of interest where optimization of experimental conditions was more targeted to these compounds. In this report, we present an entire workflow based on base-activated dansylation isotope labeling LC-MS for comprehensive profiling of the hydroxyl submetabolome.

2.2. Experimental Section

2.2.1. Workflow

In our work, differential chemical isotope labeling (CIL) is used for relative metabolite quantification where a control sample is prepared by mixing small aliquots of

individual samples to form a pool, followed by ^{13}C -dansyl labeling. This ^{13}C -labeled control is spiked into all the ^{12}C -labeled individual samples and thus serves as a global internal standard. Figure 2.1 illustrates the overall workflow of the method for profiling the hydroxyl submetabolome for metabolomics. It involves the following steps: 1) liquid-liquid extraction to enrich the hydroxyl metabolites from complex samples, 2) base-activated dansyl labeling in acetonitrile, 3) LC-UV quantification of labeled metabolites for sample amount normalization, 4) mixing of equal amounts of ^{12}C -labeled individual samples and ^{13}C -labeled control, 5) high resolution LC-MS analysis of ^{12}C -/ ^{13}C -mixtures, 6) data processing including peak pair picking, peak ratio measurement and statistical analysis, and 7) metabolite identification based on the use of hydroxyl standards library for positive identification and the use of other compound libraries for putative identification. Some experimental conditions used in the workflow are described below.

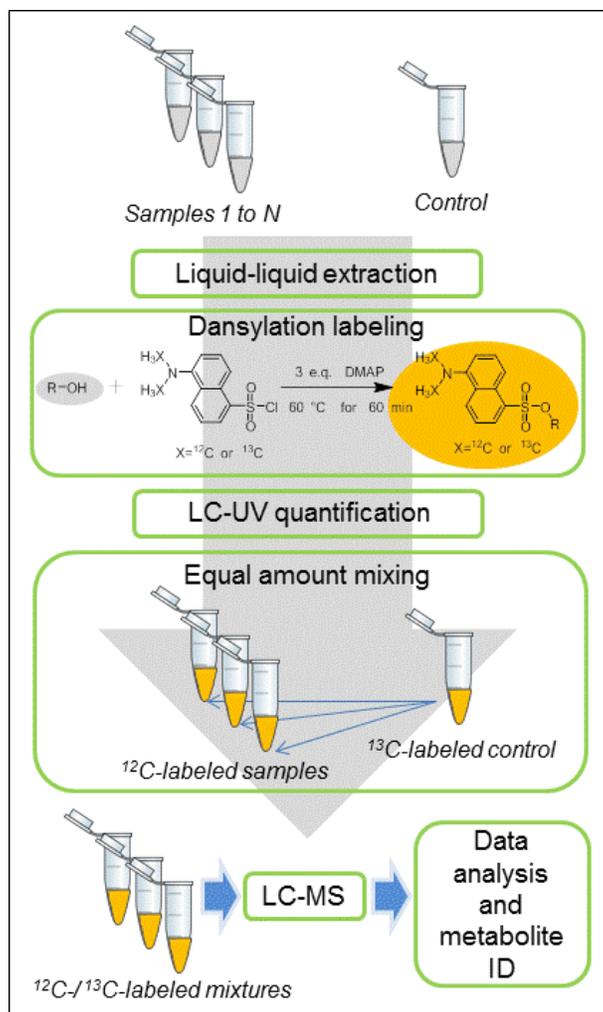


Figure 2.1 Workflow of the base-activated dansylation isotope labeling LC-MS technique for in-depth relative quantification of the hydroxyl submetabolome for metabolomics.

2.2.2. Chemicals and Reagents

All chemicals and reagents were purchased from Sigma-Aldrich Canada Ltd. (Markham, ON, Canada), except those specifically stated. LC-MS grade solvent (water and acetonitrile) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada). The isotope labeling reagent, ¹³C-dansyl chloride, is available from mcid.chem.ualberta.ca.

2.2.3. Standard Solution Preparation

85 hydroxyl standards were obtained from the laboratories of Professors Todd Lowary, Derrick Clive and John Vederas in the Department of Chemistry at the University of Alberta. Each compound was dissolved individually in acetonitrile (ACN) to a concentration of 1 mM. The solution was stored in -20 °C.

2.2.4. Urine Sample Collection and Preparation

Urine samples were collected from a healthy individual. An informed consent was obtained from the volunteer before study. The ethics approval was obtained from the University of Alberta Ethics Approval Board. After collection, the urine samples were stored at 4 °C immediately. The samples then were centrifuged at 14000 rpm for 10 min. The supernatant was filter through a 0.2- μ m filter, which was then aliquoted and stored in -80 °C for further use.

2.2.5. Dansylation Labeling Reaction

For labeling a standard compound, 25 μ L of standard solution was mixed with 25 μ L of freshly prepared 12 C-dansyl chloride (18 mg/mL in ACN) and 25 μ L of 4-dimethylaminopyridine (DMAP) (24.5 mg/mL in ACN). After vortexing and spinning down, the mixture was incubated at 60 °C for 60 min. Then 5 μ L of 250 mM NaOH solution was added to quench the reaction. The solution was vortexed and spun down again. After incubating at 60 °C for 10 min, 25 μ L of 425 mM formic acid solution in 50/50 ACN/water was added to consume the excess NaOH. The combined solution was centrifuged for 10 min at 10000 rpm and diluted with ACN before injecting into LC-UV for quantification or LC-MS for analysis.

For labeling the urine samples, 50 μL of urine was mixed with 10 μL of saturated NaCl solution and 5 μL of 6 M HCl solution. Then the metabolites in urine were extracted with ethyl acetate twice. Each time 150 μL of ethyl acetate was used. The extracted solutions were combined, dried down and then re-dissolved in 50 μL of ACN. To prepare the ^{12}C -/ ^{13}C -labeled urine mixture for LC-MS analysis, 25 μL of the extracted urine was labeled by ^{12}C -dansyl chloride (light labeling) and 25 μL of the extracted urine was labeled by ^{13}C -dansyl chloride (heavy labeling) using the same procedure described above for labeling standards. After labeling, the total concentration of the labeled metabolites was quantified by LC-UV (see below). Then an equal amount of the ^{12}C -labeled urine sample and the ^{13}C -labeled urine sample was mixed. The mixture was injected into LC-MS for analysis (see below).

2.2.6. LC-UV Quantification

A Waters ACQUITY UPLC system combined with a photodiode array detector was used for measuring the total concentration of labeled metabolites. 5 μL of labeled standard solution or urine was injected into a C18 column (Phenomenex Kinetex C18, 2.1 mm \times 5 cm, 1.7 μm particle size, 100 \AA pore size). Mobile phase A was 0.1% (v/v) formic acid and 5% (v/v) ACN in water. Mobile phase B was 0.1% (v/v) formic acid in ACN. A fast gradient was applied to elute all labeled metabolites together (flow rate: 0.45 mL/min; 0 min, 0% B; 1 min 0% B; 1.01 min, 95% B; 2.5 min, 95% B; 3.0 min, 0% B; 6.0 min 0% B). The UV detector was operated at 347 nm.

2.2.7. LC-MS Analysis

Before analysis, the labeled standard solution was diluted with ACN by 10,000-fold. This solution or a ^{12}C -/ ^{13}C -labeled urine sample was analyzed using a Thermo

Dionex Ultimate 3000 UPLC combined with a Bruker Impact HD Quadrupole Time-of-flight (Q-TOF) mass spectrometer (Billerica, MA) with electrospray ionization (ESI). Reversed phase (RP) separation was carried out using an Agilent Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size). Mobile phase A was 0.1% (v/v) formic acid and 5% (v/v) ACN in water. Mobile phase B was 0.1% (v/v) formic acid in ACN. The gradient for the separation was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 9.2 min, 65% B; t = 21.2 min, 99% B; t = 31.2 min, 99% B. The flow rate was 180 μL/min. All MS spectra were collected in positive ion mode at a spectral acquisition rate of 1 Hz. The MS/MS spectra of labeled standards were obtained under collision energy of 20-50 eV in a stepping mode. The width of mass window for precursor ion isolation in the MS/MS experiments was 6 Da.

2.2.8. Data Processing

A set of programs developed by our group were used in sequence to process the raw LC-MS data in batch mode. IsoMS⁴⁸ was used to extract peak pairs from mass spectra, filter the extracted pairs to remove redundant pairs such as those of adduct ions and dimers to retain only a protonated molecule for one labeled metabolite (i.e., one peak pair is generally corresponding to one metabolite), calculate the peak-pair intensity ratio of individual metabolites, and align multiple pairs of the same individual metabolites from different runs according to retention time and accurate mass. The missing ratio values in the aligned file were filled by the Zerofill program.¹⁴⁰ Finally, the chromatographic peak ratios for peak pairs were determined by IsoMS-Quant.¹⁴¹ Putative metabolite identification was done based on accurate mass matched with metabolites in

the human metabolome database (HMDB) (www.hmdb.ca) and MyCompoundID (MCID) (<http://www.mycompoundid.org/>).¹⁴²

2.3. Results and Discussion

2.3.1. Reaction Condition Optimization with Standard

Alcoholic hydroxyl is not good nucleophilic functional group, compared to amine or phenol group that can be more readily labeled by dansylation.³⁰ To label the hydroxyl group, the reaction condition is different from that of labeling amines or phenols. It needs more basic condition to facilitate nucleophilic attack of alcohol to form dansylated derivative.^{128, 139} In this work, 4-dimethylaminopyridine (DMAP) was used to activate the dansylation reaction.¹²⁸

To optimize the labeling condition, a standard, 1-propanol, was labeled under different reaction conditions and then injected into LC-UV for analysis. Peak area of the UV chromatogram was used to compare the relative labeling efficiencies under different reaction conditions. Figure 2.2A shows the chromatogram of dansylated 1-propanol in LC-UV detected at 347 nm. The area under the peak around 1.56 min reflects the amount of the labeled 1-propanol in the sample. Figure 2.2B shows the effects of reaction solvent, time, temperature, DnsCl/DMAP mole ratio and quenching methods (from left to right) on labeling efficiency for 1-propanol labeling. For reaction solvent screening, ACN, ACN/H₂O (1:1, v/v) and CH₂Cl₂ were tested. The results shown in Figure 2.2B indicate that water in the incubation solvent would dramatically decrease the labeling efficiency. In aprotic solvents, CH₂Cl₂ performed better than ACN. However, three practical

considerations led us not to use CH_2Cl_2 in the labeling reaction for metabolomic profiling. Firstly, the low boiling point of CH_2Cl_2 (39.6 °C) causes severe evaporation when reaction conducted at a relatively high temperature, leading to a difficulty of controlling the reaction process. Secondly, when CH_2Cl_2 is used, an extra drying procedure is required after labeling to overcome the problem of volume variations due to inconsistent evaporations in multiple samples, which would prolong the overall labeling process. Finally, the high vapor pressure of CH_2Cl_2 (57.3 kPa vs. 12.2 kPa of ACN at 25 °C) causes a difficulty of transferring solvents using a pipette. Thus, in our work, ACN was selected as the reaction solvent.

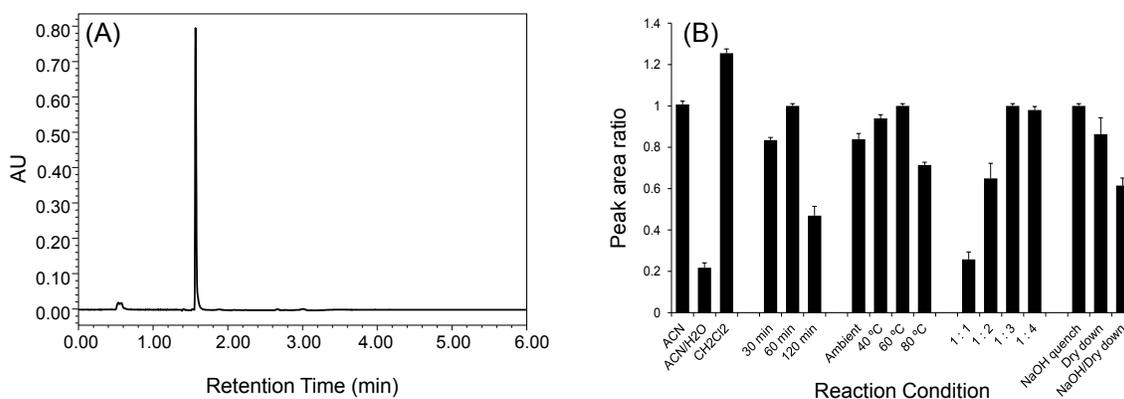


Figure 2.2 (A) LC-UV chromatogram of dansyl labeled 1-propanol. (B) Comparison of labeling efficiencies under different labeling reaction conditions: from left to right: incubation solvent, reaction time, reaction temperature, DnsCl/DMAP mole ratio, and reaction quenching method. Data were presented as mean \pm S.D. from three independent experiments (n=3).

As Figure 2.2B shows, reaction temperature and time have some effects on labeling efficiency. The optimal conditions were 60 min at 60 °C. Different mole ratios of DnsCl/DMAP were also tested. The mole ratio of 1:3 was optimal. For the development

of a proper quenching condition, three methods were compared: 1) adding NaOH to quench the reaction, 2) drying down the sample and then re-dissolving it, and 3) NaOH quenching combined with drying down. Among them, NaOH quenching showed the best labeling result in terms of both efficiency and reproducibility.

2.3.2. Reaction Condition Optimization with Urine Sample

Human urine was used as a representative complex metabolomic sample to evaluate the labeling efficiency under different reaction conditions. In a metabolomic sample, water is always present, but would adversely affect the dansylation reaction targeting the hydroxyl metabolites as discussed above. One option to deal with this issue is to perform complete drying of an aqueous sample before labeling. However, this would result in a complex sample from which dansyl labeling would occur in many groups of metabolites including amines. Amine labeling is best done in a solution containing water in order to solubilize the ionic or very hydrophilic metabolites (e.g., amino acids).³⁰ Thus, in our approach, a metabolite extraction step is used to extract the hydroxyl metabolites into an organic layer while leaving the amines in the aqueous phase. The amines can be labeled using the optimal condition tailored for amine labeling for profiling the amine submetabolome. Separating the amines from the hydroxyl metabolites also simplifies the compositions of both submetabolomes, thereby less co-eluting metabolites would be detected in LC-MS analysis of an individual submetabolome. Ion suppression effect should be reduced.

In our approach, an acidic liquid-liquid extraction (LLE) procedure is used to extract the hydroxyl metabolites. By adding HCl to an aqueous sample such as urine, most amines would be in ionic form and stay in water.¹⁴³ The hydroxyl metabolites,

unless they also contain other functional groups that can be ionized in HCl, would stay in the organic phase. After extraction, the metabolite extract is dried and then re-dissolved in ACN for optimal labeling of the hydroxyl metabolites. For drying the metabolite extract in the organic solvent, the use of SpeedVac is sufficient and convenient; there is no need of using a freeze-dry equipment more commonly used for rapid drying of an aqueous solution.

To examine the effects of reaction conditions on urine labeling, ^{12}C -/ ^{13}C -dansyl labeled urine mixtures were prepared. The peak pair number detected was used as the indicator of the labeling efficiency under different reaction conditions. Figure 2.3A and Figure 2.3B show the effects of reaction time and temperature on labeling, respectively. The results indicate that 60 min and 60 °C were the optimized conditions, which are consistent with those for labeling 1-propanol. On the reagent amount optimization, 1:3 mole ratio of DnsCl/DMAP would achieve the highest labeling efficiency (Figure 2.3C). Although there was no significant difference between the 1:2 group and the 1:3 group, considering the result of labeling 1-propanol, we chose 1:3 as the optimal condition. Three quenching methods stated previously were also tried on urine samples. NaOH quenching method also gave the highest peak pair number (Figure 2.3D).

The above results indicate that dansyl labeling achieved the optimal efficiency when it was carried out at 60°C for 60 min in acetonitrile (ACN) using DnsCl/DMAP mole ratio of 1:3. NaOH was found to be well-suited for quenching the reaction after metabolite labeling.

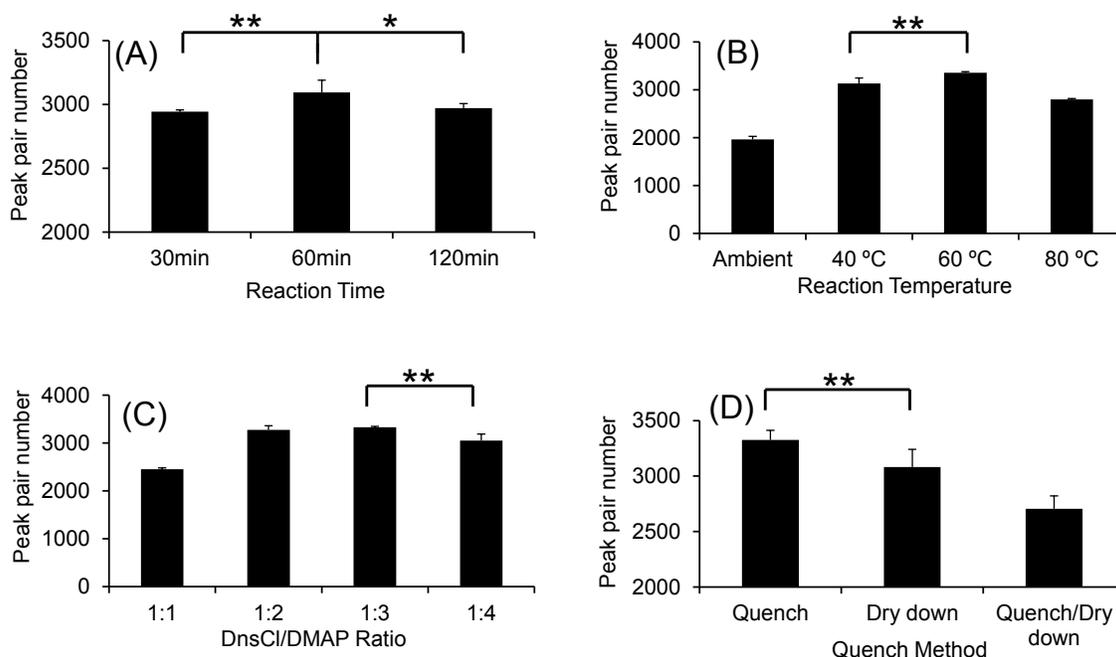


Figure 2.3 Peak pair numbers detected from ^{12}C -/ ^{13}C -dansyl labeled urine samples prepared under different labeling reaction conditions: (A) effect of reaction time, (B) effect of reaction temperature, (C) effect of DnsCl to DMAP mole ratio, and (D) effect of quenching method. Data are presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6). * $p < 0.05$; ** $p < 0.01$.

2.3.3. Stability Test of Labeled Samples

In many studies, labeled samples cannot be analyzed right away. This is particularly true for a metabolomics project involving the use of many samples. Sample storage over a period of time such as a few months is often required. To assess the stability of labeled samples, individually labeled urine samples were stored under different temperature for a given period of time: 4 °C for overnight, -20 °C for a week, -20 °C for a month, and -80 °C for three months. The labeled samples in each group after storage were injected into LC-MS for analysis. Peak pair numbers detected were used to

gauge the sample stability. Figure 2.4A shows the results. Only storing the samples under -20 °C for a month would decrease the peak pair numbers slightly (<5%). These results indicate that the labeled samples are in general very stable and could be stored at -80 °C for an extended period.

Considering water adversely affects the labeling reaction, we also tested the stability of labeled samples in water. Labeled urine samples were dried down after derivatization and then re-dissolved with ACN or ACN/H₂O (1:1, v/v). After LC-MS analysis, peak pair numbers were used to assess the water stability. As it is shown in Figure 2.4B, after labeling water would not affect the dansylated metabolites.

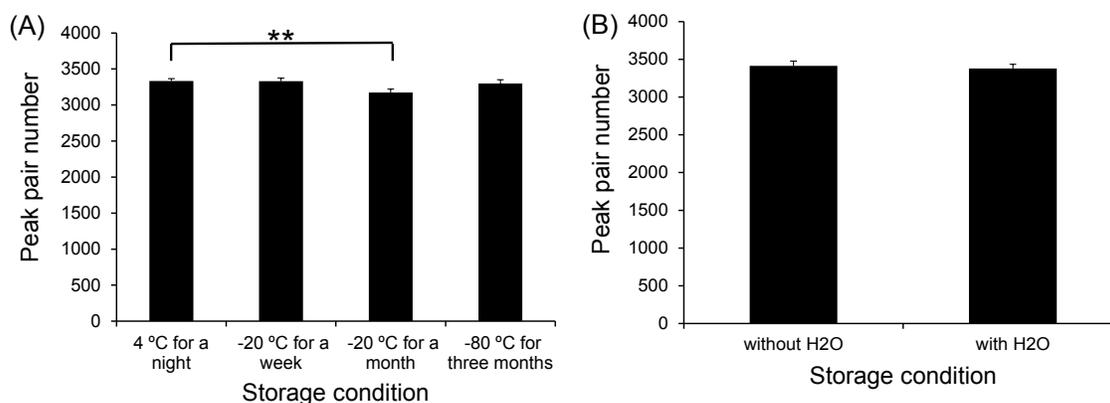


Figure 2.4 (A) Sample storage stability. Labeled urine samples were stored under different temperatures for different periods of time. (B) Water stability. Labeled urine samples were dissolved with or without water. Data were presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6). ** p<0.01.

2.3.4. Sample Quantification and Normalization.

Because of great variations in total concentration for many samples such as urine and cell or tissue extracts, sample normalization is required in quantitative metabolomics.

In this work, we propose to use a LC-UV method¹⁴⁴ to measure the total concentration of dansyl labeled hydroxyl metabolites in a sample. Depending on the concentration, a different volume of sample is taken from each individual sample to ensure that the same amount of sample is used for metabolomic comparison of multiple samples.

In LC-UV, a fast step-gradient LC was used to remove the quenched product (Dns-OH) first using high aqueous solvent and then elute all the labeled metabolites together into one peak using high organic solvent.¹⁴⁴ To optimize the gradient condition and select the optimal wavelength for detecting the labeled hydroxyl metabolites, 10 dansyl-labeled alcohol standards (Table 2.1) were analyzed by LC-UV individually.

Table 2.1 List of 10 labeled alcohols analyzed by LC-UV.

#	HMDB ID	Compound	Retention time (min)
1	HMDB00131	Glycerol	1.46
2	HMDB32985	4-Methyl-5-thiazoleethanol	1.56
3	HMDB31175	Tetrahydrofurfuryl alcohol	1.52
4	HMDB40348	1,8-Octanediol	1.70
5	HMDB00820	1-Propanol	1.56
6	HMDB41796	2,2,2-Trichloroethanol	1.58
7	HMDB35818	Borneol	1.68
8	HMDB36078	Isopulegol	1.70
9	HMDB35094	β -Citronellol	1.73
10	HMDB03352	Menthol	1.78

The absorption spectrum of each labeled metabolite is showed in Figure 2.5A. All labeled alcohols have similar absorption behavior. Three sets of bands originated from $\pi \rightarrow \pi^*$ transitions in the dansyl aromatic rings are observed, which is in agreement with dansyl labeled amines or amino acids.¹⁴⁴ However, the highest wavelength bands of dansylated alcohols are between 343 nm and 353 nm, instead of from 326 to 349 nm for

dansylated amines. It can be attributed to the differences between O-dansyl bonds and N-dansyl bonds. Thus, in our LC-UV method, we chose the median value, 347 nm, as the detection wavelength for the quantification of labeled alcohols. As Table 2.1 shows, the retention time of labeled metabolites is from 1.4 min to 1.8 min, indicating that the gradient used can elute all the labeled metabolites together as one big peak.

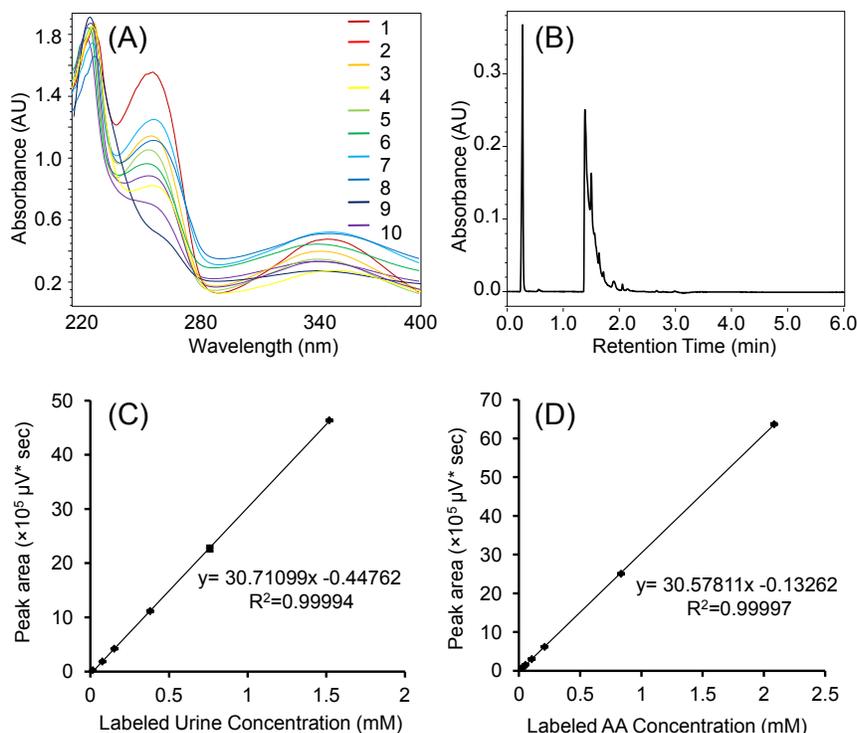


Figure 2.5 (A) Absorption spectra of 10 dansyl labeled alcohols. (B) UV chromatogram of labeled urine sample. (C) Calibration curve of labeled urine sample. (D) Calibration curve of a mixture of 17-dansyl labeled amino acids.

The LC-UV method was applied to urine sample quantification. Figure 2.5B shows the UV chromatogram obtained at 347 nm. In the 6-min gradient, the peaks between 1.3 and 2.0 min are from the labeled metabolites. Figure 2.5C shows the calibration curve for quantification of the labeled metabolites in urine. The data points

were collected from a labeled urine sample after serial dilutions. The undiluted labeled urine concentration was calculated from a mixture of 17-dansyl labeled amino acids (Figure 2.5D). The rationale of this calculation was from the assumption that the average absorptivity of many dansyl labeled metabolites would be similar, which has been proved in a previous study.¹⁴⁴ No significant difference between the slopes of the labeled amino acids curve and the labeled urine curve was found, confirming this assumption. Thus, the mixture of labeled amino acids can be used to build a calibration curve for measuring the concentration of labeled urine or other biological samples in quantitative hydroxyl submetabolome profiling.

2.3.5. Hydroxyl Submetabolome Profiling.

Knowing the concentration of a labeled sample is also very useful for optimizing the injection amount in LC-MS analysis and keeping the injection amount consistent for all the LC-MS runs for hydroxyl submetabolome profiling. Figure 2.6A shows a plot of the number of peak pairs or metabolites detected as a function of injection amount. A plateau of peak pair number was reached when 8.64 nmol of sample was injected. This amount was deemed to be the optimal amount for LC-MS analysis of the hydroxyl submetabolome for the instrumental setup used. For other types of samples or instrumental setups, this same method can be applied for determining the optimal amount for sample injection.

One of the major advantages of dansyl labeling is that it allows the retention and separation of a great variety of hydroxyl metabolites in a RP column, averting the need of switching to different columns for separating metabolites with different ionic or hydrophilic properties. Figure 2.6B show an ion chromatogram (IC) obtained from the

$^{12}\text{C}/^{13}\text{C}$ -labeled human urine sample. Many peaks are detected across the entire RP elution window, indicating a great diversity of hydroxyl metabolites in a complex sample.

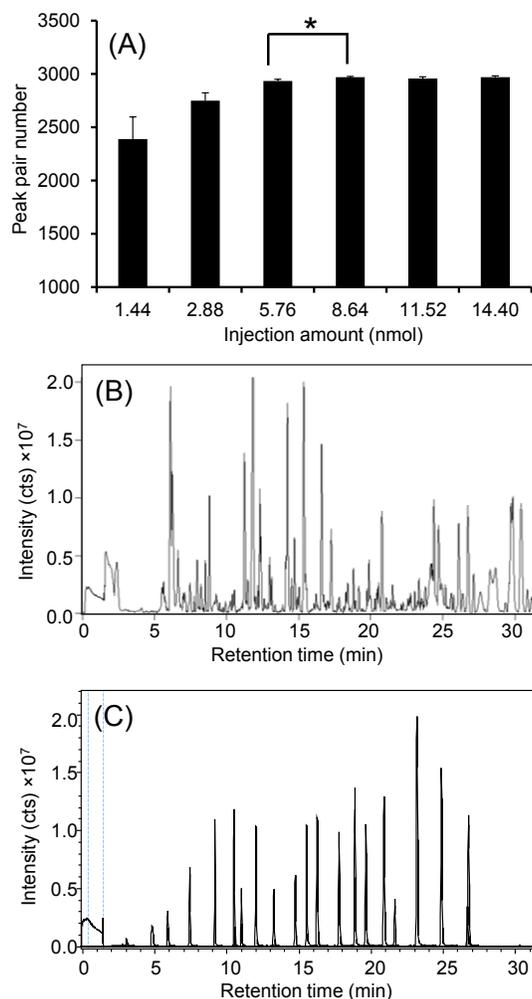


Figure 2.6 (A) Peak pair number detected as a function of injection amount of a $^{12}\text{C}/^{13}\text{C}$ -labeled human urine ($n=3$). (B) Ion chromatogram (IC) of labeled urine. (C) IC of a mixture of 22 labeled standards containing amine or hydroxyl group. * $p<0.05$.

Another advantage of dansyl labeling is that it improves the metabolite detection sensitivity significantly. Based on the analysis of 85 hydroxyl standards (see below), it was found that the hydroxyl compounds without the presence of other functional groups

gave no or very little signals in ESI. However, after dansyl labeling, they could be ionized as efficiently as the labeled amines. For example, Figure 2.6C shows IC of a mixture of 22 labeled metabolites containing amine or hydroxyl group which is also used for retention time calibration (see below). Table 2.2 shows the list and the amounts of the 22 labeled metabolites in the mixture. The detected chromatographic peak area for each metabolite is also shown in Table 2.2. Despite different structures, their detection sensitivity becomes more unified after dansylation. Thus, the net result of dansyl labeling is that all the metabolites will have a better chance to be detected in a mixture.

Table 2.2 List of RT calibrants consisting of 20 labeled standards.

#	Name	m/z	RT (min)	Injection amount (pmol)	Injection mole ratio	Peak area
1	Dns-Arginine	408.1700	2.86	5.29	4	1.752E+06
2	Dns-Serine	339.1009	5.03	5.29	4	9.535E+06
3	Dns-Threonine	353.1165	6.11	5.29	4	1.068E+07
4	Dns-Alanine	323.1060	7.55	5.29	4	1.780E+07
5	Dns-Proline	349.1216	9.28	5.29	4	2.381E+07
6	Dns-Phenylalanine	399.1373	10.6	5.29	4	2.752E+07
7	Dns-Cystine (2 tags)	354.0702	11.16	2.65	2	1.017E+07
8	Dns-2-Methoxyethanol	310.1107	12.16	3.31	2.5	4.235E+07
9	Dns-Ethanol	280.1000	13.4	3.31	2.5	2.073E+07
10	Dns-1-Propanol	294.1158	14.84	6.61	5	2.827E+07
11	Dns-Tyrosine (2 tags)	324.5952	15.62	5.29	4	3.156E+07
12	Dns-Pentenol	320.1315	16.35	3.31	2.5	5.545E+07
13	Dns-3-Phenyl-1-propanol	370.1471	17.85	3.31	2.5	4.992E+07
14	Dns-4-Phenyl-1-butanol	384.1628	18.91	3.31	2.5	6.799E+07
15	Dns-beta-Estradiol	506.2359	19.68	3.31	2.5	5.387E+07
16	Dns-1-Heptanol	350.1784	20.92	1.32	1	6.303E+07
17	Dns-Isopulegol	388.1941	21.67	6.61	5	3.990E+07
18	Dns-Cholic acid methyl ester	656.3615	23.19	6.61	5	1.071E+08
19	Dns-Decyl alcohol	392.2254	24.89	3.31	2.5	1.204E+08
20	Dns-Cholic acid methyl ester (2 tags)	445.2099	25.63			2.266E+05
21	Dns-Deoxycholic acid methyl ester	640.3666	26.77	4.41	3.33	6.525E+07
22	Dns-Deoxycholic acid methyl ester (2 tags)	437.2125	30.34			1.222E+05

To demonstrate the detectability of the workflow shown in Figure 2.1 for analyzing the hydroxyl submetabolome, Figure 2.7A shows the Venn diagram of the numbers of peak pairs detected from experimental triplicate of ^{12}C -/ ^{13}C -labeled urine samples; each sample was also injected in triplicate. An average of 3759 ± 45 pairs per run and 3538 ± 71 pairs per sample were detected with 3093 pairs in common, indicating that peak pair detection is highly reproducible. Figure 2.7B-C shows the number distribution of peak pairs detected as a function of the averaged measured peak ratio and their RSD (n=9). Most of the peak pairs gave the ratio value close to the expected ratio of 1.0, demonstrating high accuracy. This also illustrates that there were no isotopic effects on derivatization efficiency, LC separation and MS ionization, which is not surprising as ^{13}C -isotope reagents, instead of deuterium reagents, were used. The RSD values are less than 18% for 95% of the pairs with an average RSD of 7.1% and thus the analytical precision was also very high.

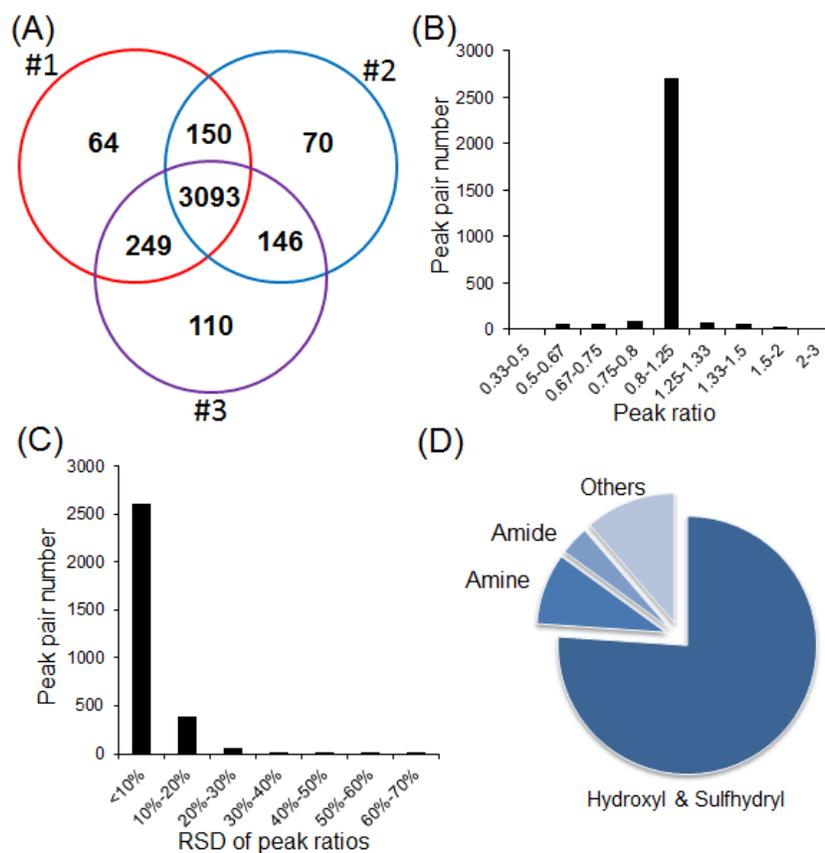


Figure 2.7 (A) Venn diagram of the peak pair numbers detected in triplicate analysis of labeled urine. Number distribution of peak pairs as a function of (B) averaged peak ratio and (C) RSD (n=9). (D) Classes of labeled metabolites.

2.3.6. Hydroxyl Metabolite Identification.

After metabolite detection and quantification, metabolite identification is crucial for linking the metabolomic data to biology and for biomarker discovery. Positive identification of metabolites requires the use of authentic standards. To this end, we have constructed a hydroxyl standards library consisting of 85 unique compounds (Table 2.3), which is expandable in the future. To construct this library, each standard was separately labeled, followed by RPLC-MS and MS/MS measurements to determine the retention

time, accurate molecular mass and MS/MS spectrum. These triple parameters were entered into the library. Figure 2.8 shows two examples of the entries (IC and MS/MS spectra) in the library. To identify a dansyl labeled metabolite in a sample, accurate mass and retention time matches and/or MS/MS match against this library can be performed. In the examples shown in Figure 2.8, the two isomers, 1-phenyl-1-propanol and 3-phenyl-1-propanol, have the same mass. However, their retention time and MS/MS spectra are very different, allowing high-confidence identification based on mass and RT matches, mass and MS/MS matches or all three matches.

Table 2.3 List of 85 compounds in the current dansyl hydroxyl standards library.

Metabolites	RT (min)	Metabolites	RT (min)
1-Propanol	14.54	2-Amino-2-methyl-1,3-propanediol	5.54
2-Propanol	14.14	beta-Estradiol	19.37
tert-Butyl alcohol	16.03	(1S,2S)-trans-1,2-Cyclohexanediol	12.13
Benzyl alcohol	15.64	(S)-3-Benzyloxy-1,2-propanediol	13.64
Decyl alcohol	24.6	1,16-Hexadecanediol	24.40
1,2-Propanediol	9.93	1,7-Heptanediol	12.37
Glycerol	7.31	Triethylene Glycol	8.96
Ethanol	13.13	(+)-1,2,4-Butanetriol	7.77
1-Heptanol	20.59	1-Octanol	22.06
1,5-Pentanediol	10.64	2,2,2-Trichloroethanol	16.79
1,4-Butynediol	10.25	Ethylene glycol	8.89
Piperonyl alcohol	9.42	Propargyl alcohol	12.67
2-Aminobenzyl alcohol	11.13	1-Bromo-2-propanol	15.25
Tetrahydrofurfuryl alcohol	12.81	3-Chloro-1,2-propanediol	11.26
Furfuryl alcohol	8.35	1-Pentanol	19.57
Pentenol	16.04	2-Methyl-1-butanol	17.38
2-Methoxyethanol	11.93	3-Dimethylamino-1-propanol	11.45
Crotyl alcohol	14.82	3-(o-Tolyloxy)-1,2-propanediol	14.55
Anisyl alcohol	9.55	(S)-(-)-β-Citronellol	22.30
1-Adamantanol	21.08	Menthol	22.92
Phenethyl alcohol	11.06	Deoxycholic acid methyl ester	26.55
2-Butene-1,4-diol	14.03	Cholic acid methyl ester	22.96
Cyclopentanol	16.06	3-Octyn-1-ol	18.76
Borneol	21.50	Isopulegol	21.38

Diethylene glycol monoethyl ether	12.87	1,6-Hexanediol	11.42
Butynol	13.06	1,8-Octanediol	13.45
Benzhydrol	9.13	4-Phenyl-1-butanol	18.60
9-Fluorenamethanol	18.69	2-Phenoxyethanol	15.50
4-Hydroxybenzyl alcohol	12.04	N-acetylethanolamine	9.35
4-Hydroxy-3-methoxybenzyl alcohol	11.82	2-Chloroethanol	13.37
2-Fluorobenzylalcohol	15.28	1-Indanol	11.43
3-Phenoxybenzyl alcohol	18.76	2-Cyanoethanol	11.27
4-Hydroxypiperidine	8.67	1-Phenyl-1-propanol	6.60
3-Buten-1-ol	14.77	3-Phenyl-1-propanol	17.64
2-Pentanol	17.08	4-tert-Butylcyclohexanol	22.88
2-Methyl-1-propanol	15.96	3-Methyl-2-buten-1-ol	9.42
3-Butene-2-ol	7.62	4-Methyl-5-thiazoleethanol	11.26
7-Octen-1-ol	20.21	Cyclohexylmethanol	19.64
2,2,2-Trifluoroethanol	14.47	5-Norbornene-2-methanol	18.59
1,12-Dodecanediol	18.82	2-Mercaptoethanol	10.11
1,4-Cyclohexanediol	10.48	Iso amyl Alcohol	17.49
cis-4-Cyclopentene-1,3-diol	10.55	2-(2-thienyl)ethanol	15.77
1,10-Decanediol	15.96		

On MS/MS spectra of dansylated metabolites, we could observe unique fragment ions from the metabolite moiety itself, although the intensities of the fragment ions from the dansyl group were very high. For example, in the MS/MS spectra shown in Figure 2.8, the two labeled isomers gave the fragment ions from the metabolite moiety. However, the number of such fragment ions might not be sufficiently high for *de novo* structural interpretation of an unknown metabolite. Thus, the use of the dansyl library for MS/MS spectral match as proposed in this work becomes very important for metabolite identification. As one can see from the MS/MS spectra of the two isomers, the fragment ion patterns are quite different even though they contain several intense fragment ions of dansyl group itself.

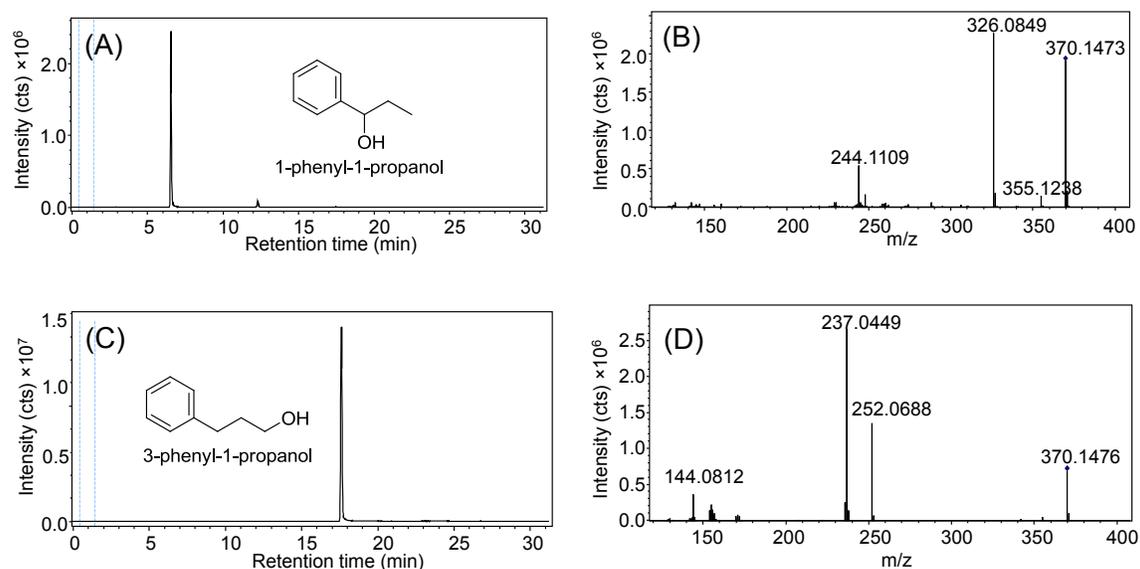


Figure 2.8 Examples of entries in the labeled hydroxyl standards library: (A) IC and (B) MS/MS spectrum of dansyl labeled 1-phenyl-1-propanol, (C) IC and (D) MS/MS spectrum of dansyl labeled 3-phenyl-1-propanol. Each entry contains accurate mass, retention time, ion chromatogram, and MS/MS spectrum obtained using RPLC-QTOF-MS.

One of the triple parameters in the library is retention time (RT) of a labeled compound which is an important parameter that can be used to increase the confidence of metabolite identification. However, RT can be varied depending on the instrumental setup and experimental conditions used for running LC-MS. To use the RT information to assist in metabolite identification, we have recently reported a retention time normalization method to account for the RT shift between a user's setup and the library RT data.⁵⁹ In this method, the retention time values of the analyte peaks obtained in a user's setup are normalized to those of a set of RT calibrants by a correction algorithm. The normalized RT values of unknown metabolites can be compared to those of library standards, as the

same RT calibrants were used for RT normalization when the library was constructed. For the comparison of the RT values of hydroxyl compounds, we have developed a set of RT calibrants consisting of 20 standards as shown in Table 2.2. Among them two metabolites can be labeled by two tags, resulting in a total of 22 peaks from running the RT calibration mixture. As it is shown in Figure 2.6D, these 22 peaks spread in the whole gradient and the RT difference between two peaks is less than 2 min, allowing for linear RT correction within a short elution widow bracketed by two RT calibrants.⁵⁹

The labeled hydroxyl standards library, along with a retention time calibration program to correct the RT shifts between a user's setup and the library RT as well as a search program for metabolite identification, is available on the MyCompoundID website (www.mycompoundid.org). The library and search program are freely accessible by all users.

For the urine sample, we positively identified 20 peak pairs based on accurate mass and retention time matches (Table 2.4). For the remaining peak pairs, putative identification can be made by searching accurate mass against the MyCompoundID (MCID) library composed of 8,021 known human endogenous metabolites and their predicted metabolic products from one metabolic reaction (375,809 compounds).¹⁴² We matched 643 and 1641 metabolites in the zero- and one-reaction libraries, respectively. Thus, out of 3093 peak pairs detected, 2304 pairs (75%) could be positively identified or putatively matched. We also searched the accurate masses of these peak pairs against the HMDB library. 1730 pairs (56%) could be matched and, among them, 1316 pairs (76%) were matched to metabolites containing hydroxyl group including 23 sulfhydryl compounds, 155 (9%) to amines, 63 (4%) to amides and the remaining ones (11%) to

other groups (see Figure 2.7D). It is clear that the majority of metabolites detected are those containing hydroxyl groups. Some amines were detected as they could still be extracted in LLE even after acidification.

Table 2.4 List of hydroxyl metabolites identified based on accurate mass and retention time matches to the dansyl hydroxyl standards library.

Peak pair #	Peak pair information				Identification result			
	Corrected RT (s)	mz_light	mz_heavy	mz	Name	Accurate mass	mz_light	Library RT (s)
376	494.4	340.1230	342.1269	106.0646	(+)-1,2,4-Butanetriol	106.0630	340.1213	466.2
667	575.2	337.1216	339.1276	103.0633	N-acetyethanolamine	103.0633	337.1216	561.0
703	583.4	310.1108	312.1168	76.0525	1,2-Propanediol	76.0524	310.1107	595.8
1485	729.6	310.1076	312.1135	152.0986	2-Methoxyethanol	76.0524	310.1107	715.8
1856	803.6	280.1005	282.1069	46.0422	Ethanol	46.0417	280.1000	787.8
2284	866.1	294.1155	296.1219	60.0572	1-Propanol	60.0575	294.1158	872.4
	866.1	294.1155	296.1219	60.0572	2-Propanol	60.0575	294.1158	848.4
2304	868.2	306.1156	308.1223	72.0573	Crotyl alcohol	72.0575	306.1158	889.2
	868.2	306.1156	308.1223	72.0573	3-Buten-1-ol	72.0575	306.1158	886.2
2507	889.5	416.1509	418.1582	182.0925	3-(o-Tolyloxy)-1,2-propanediol	182.0943	416.1526	873.0
2857	951.0	372.1261	374.1316	138.0678	2-Phenoxyethanol	138.0681	372.1264	930.0
2868	952.1	342.1193	344.1235	108.0610	Benzyl alcohol	108.0575	342.1158	938.4
2979	970.9	320.1311	322.1377	86.0728	Pentenol	86.0732	320.1315	962.4
	970.9	320.1311	322.1377	86.0728	Cyclopentanol	86.0732	320.1315	963.6
2999	982.2	308.1305	310.1377	74.0722	tert-Butyl alcohol	74.0732	308.1315	961.8
	982.2	308.1305	310.1377	74.0722	2-Methyl-1-propanol	74.0732	308.1315	957.6
3390	1062.9	265.0755	267.0823	62.0343	Ethylene glycol	62.0368	265.0767	1039.8
3440	1073.7	322.1457	324.1525	88.0873	Iso amyl Alcohol	88.0888	322.1471	1049.4
3576	1109.8	272.0835	274.0900	76.0503	1,2-Propanediol	76.0524	272.0845	1093.8
4377	1259.6	300.1134	302.1187	132.1101	1,7-Heptanediol	132.1150	300.1158	1276.8
4730	1316.5	364.1935	366.2003	130.1352	1-Octanol	130.1358	364.1941	1323.6
5181	1396.5	390.2086	392.2162	156.1502	Menthol	156.1514	390.2097	1375.2
	1396.5	390.2086	392.2162	156.1502	4-tert-Butylcyclohexanol	156.1514	390.2097	1372.8

5487	1447.5	392.2243	394.2307	158.1660	Decyl alcohol	158.1671	392.2254	1476.0
5768	1494.8	392.2245	394.2308	158.1662	Decyl alcohol	158.1671	392.2254	1476.0

2.4. Conclusions

In summary, we have developed a technique to perform relative quantification of the hydroxyl submetabolome with unprecedentedly high coverage for metabolomics. We envisage the application of this technique for in-depth profiling of the hydroxyl submetabolome which include many important classes of compounds such as hormones in all types of biological samples. This technique should also be readily integrated with other labeling chemistries targeting other chemical-group-submetabolomes to produce a very comprehensive profile of a metabolome, thereby increasing our ability of characterizing the overall metabolome for systems biology and biomarker discovery research.

Chapter 3

Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome

3.1. Introduction

Metabolomics has been proved to be very useful in a number of research areas ranging from disease biomarker discovery to systems biology studies.¹⁴⁵ In metabolomics, comparative metabolomic analysis is carried out to determine the metabolic differences between two or more groups of samples. To do this, high-coverage profiling with accurate relative quantification of individual metabolites is crucial. However, profiling complex samples (e.g., bio-fluids, cells and tissues) is a huge analytical challenge, as biological samples contain thousands of metabolites with diverse structures and physical/chemical properties and with a very wide range of concentrations (e.g., pM to mM).¹⁴⁶

To increase coverage and/or improve quantification accuracy and precision, many enabling methods and strategies, mainly based on NMR and mass spectrometry (MS), have been reported.¹⁴⁷⁻¹⁴⁹ A common approach is to group the metabolites according to their physical properties (e.g., solubility and hydrophobicity) and then analyze them using conditions optimized for each group (e.g., reversed phase (RP) LC for separating hydrophobic metabolites and hydrophilic interaction LC for separating more polar compounds).¹⁴⁹ An alternative approach is to combine high-performance chemical isotope labeling (CIL) of metabolites with liquid chromatography-mass spectrometry (LC-MS) for metabolomic analysis.³⁰ In this approach, the whole metabolome is divided

into several chemical-group-based submetabolomes, followed by CIL LC-MS analysis of each submetabolome. Judging from the structures of more than 8000 endogenous human metabolites in the Human Metabolome Database (HMDB),¹²⁷ more than 95% of these metabolites contain one or more of the four functional groups: amine, carboxyl, hydroxyl and carbonyl. Thus, if these four groups of submetabolomes can be separately analyzed with complete coverage, the combined result of the four submetabolomic profiles would represent over 95% of the whole metabolome.

There are a number of labeling chemistries that have been reported for targeted and untargeted analysis of different groups of metabolites with varying degrees of performance.^{25, 37, 68, 70-71, 94, 108, 113, 119, 128-129, 131, 133, 135, 137-138, 150-151} To detect as many metabolites as possible, a rationally designed labeling reagent can be used to react with the common functional group of a submetabolome (e.g., the amine group in all amine-containing metabolites) to alter the chemical and physical properties of the metabolites to an extent that the labeled metabolites can be efficiently separated by RPLC and effectively ionized in electrospray ionization (ESI) in a positive ion mode.^{30, 61, 74} In this way, the LC-MS setup does not need to be changed in the metabolomic analysis workflow. However, this chemical-group-based submetabolome profiling approach requires the use of robust and effective labeling chemistries targeting the individual chemical groups.

In previous studies, three high-performance labeling methods have been reported for submetabolome profiling: ¹²C-/¹³C-dansyl chloride (DnsCl) labeling for the amine submetabolome,³⁰ ¹²C-/¹³C-dimethylaminophenacyl (DmPA) bromide labeling for the carboxyl submetabolome,¹⁵² and ¹²C-/¹³C-DnsCl with base-activation for the hydroxyl

submetabolome.³¹ Each labeling allows the profiling of thousands of metabolites within a group with high accuracy and precision for relative quantification of metabolomic changes in bio-samples. However, to our knowledge, there is no available method capable of profiling the carbonyl submetabolome with very high coverage (i.e., detecting thousands of metabolites). Metabolites containing carbonyl group (ketones and aldehydes) include several important classes of molecules that often have low ionization efficiency (e.g., sugars) and/or are present at very low concentrations (e.g., steroids in plasma).¹⁴⁶ GC-MS methods have been used for analyzing ketones and aldehydes,¹⁵³⁻¹⁵⁵ while a number of derivatization chemistries^{90, 123, 156-161} including dansylhydrazine labeling¹⁶²⁻¹⁶⁷ have been reported for labeling ketones/aldehydes to improve MS detectability; however, these methods were often used for the analysis of a limited number of analytes. In this report, we present a method using ¹²C-/¹³C-dansylhydrazine labeling for analyzing the carbonyl submetabolome with high coverage and high accuracy and precision of relative quantification. Optimal labeling conditions tailored for high-coverage profiling work are described, followed by the demonstration of the analytical performance of the workflow.

3.2. Experimental Section

3.2.1. Principle and Workflow

The method presented herein complements to the three previously reported labeling methods for relative quantification of chemical-group-based-submetabolomes. In order to generate accurate and precise relative quantification results, differential chemical isotope labeling (CIL) is used for untargeted profiling of the carbonyl submetabolome.

Figure 3.1 shows the schematic of the workflow, including the labeling reaction scheme using the ^{12}C - or ^{13}C -dansylhydrazine (DnsHz) reagent. In this workflow tailored to perform comparative metabolomics of two or more groups of samples, the individual samples are separately labeled by ^{12}C -DnsHz (light labeling), while a pooled sample prepared from mixing aliquots of individual samples is labeled by ^{13}C -DnsHz (heavy labeling). The ^{13}C -labeled pool serves as an internal standard or control for all the ^{12}C -labeled individual samples; the same amount of the ^{13}C -labeled pool is spiked into each ^{12}C -labeled individual sample to form a $^{12}\text{C}/^{13}\text{C}$ -mixture for LC-MS analysis. All the labeled metabolites are detected as peak pairs in the mass spectra, making them easily distinguished from the singlet peaks originated from noise or unlabeled chemical backgrounds introduced in LC-MS. Using IsoMS software (see below), peak pairs are picked and their intensity ratios are calculated. The intensity of a ^{12}C -labeled peak from a ^{12}C -labeled sample is measured with reference to that of the ^{13}C -labeled peak of the same metabolite from a ^{13}C -labeled pool. Since the same pool is added to all individual samples, the intensity ratio values determined from LC-MS analyses of the prepared $^{12}\text{C}/^{13}\text{C}$ -mixtures are related to the concentrations of a given metabolite in these individual samples. Thus, the peak intensity ratios can be used for relative quantification of metabolites in different samples. While it is not shown in this work, absolute quantification of a metabolite can also be performed by adding a known concentration of ^{13}C -labeled standard to a ^{12}C -labeled sample. To identify the metabolites, a carbonyl standard library has been constructed (see below). Experimental conditions used for some steps are described below.

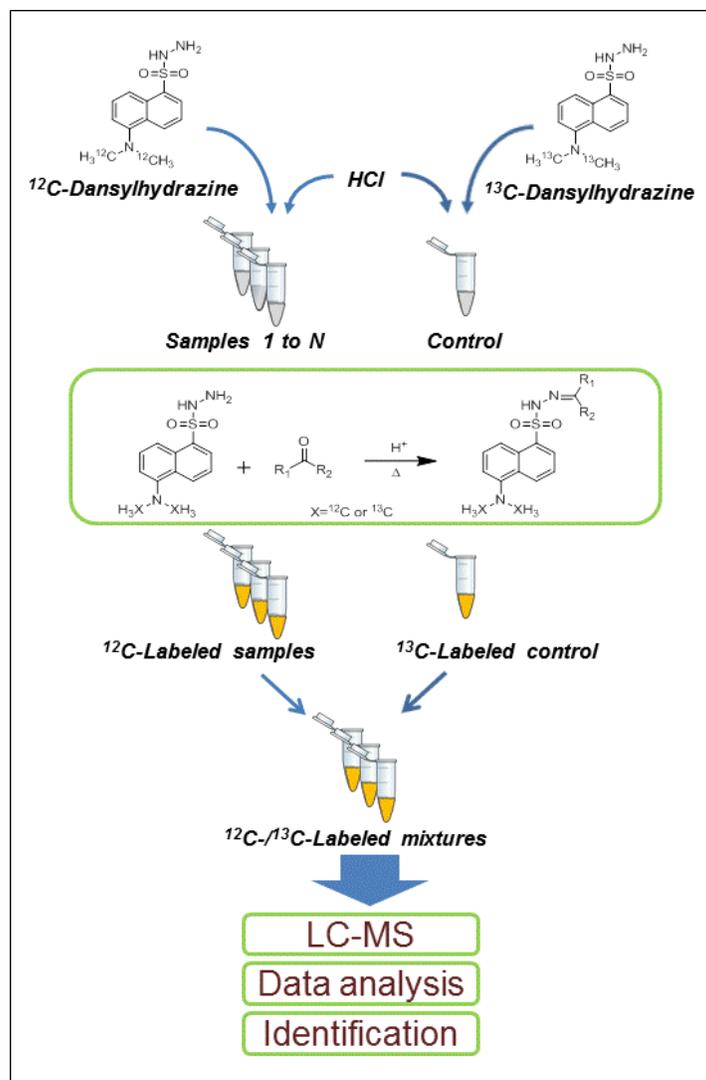


Figure 3.1 Workflow of ^{12}C -/ ^{13}C -dansylhydrazine labeling for relative quantification of the carbonyl submetabolome differences in comparative samples. The dansylhydrazine labeling reaction scheme is also shown.

3.2.2. Chemicals and Reagents

All the chemicals and LC-MS grade solvents (water, methanol and acetonitrile) were purchased from Sigma-Aldrich (Oakville, ON, Canada), except those specifically stated.

3.2.3. Synthesis of Dansylhydrazine

Dansylhydrazine (DnsHz) was prepared by treating dansyl chloride with hydrazine following a literature procedure with some modification.¹⁶⁸⁻¹⁶⁹ 10 mL of 0.1 M dansyl chloride solution in tetrahydrofuran were slowly added into 20 mL of 0.25 M hydrazine monohydrate aqueous solution with stirring in ice-bath. The reaction was monitored by thin-layer chromatography. After reaction was completed (about 30 min), 20 mL of ethyl acetate were added to the reaction mixture, followed by washing with 10 mL of distilled water for three times. The organic phase was dried by sodium sulfate and then evaporated under reduced pressure. The resulting light yellow powder was dansylhydrazine. The product was sufficiently pure that there was no need of performing further purification using column chromatography. To prepare ¹³C-DnsHz, ¹³C-dansyl chloride was used, which is available from mcid.chem.ualberta.ca. The yield of the reaction from ¹³C-dansyl chloride to ¹³C-DnsHz was high (>85%).

The final product was analyzed using NMR and high resolution QTOF-MS for structure confirmation and purity assessment. In the ¹H NMR spectra (Figure 3.2) shown below, the only difference between synthesized ¹³C-dansylhydrazine and ¹²C-dansylhydrazine standard (Sigma-Aldrich, Canada) was from the ¹H-¹³C spin-spin coupling, which split the peak at δ 2.817 (s, 6H) in ¹²C-dansylhydrazine standard to two doublet peaks at δ 2.945 and δ 2.605.

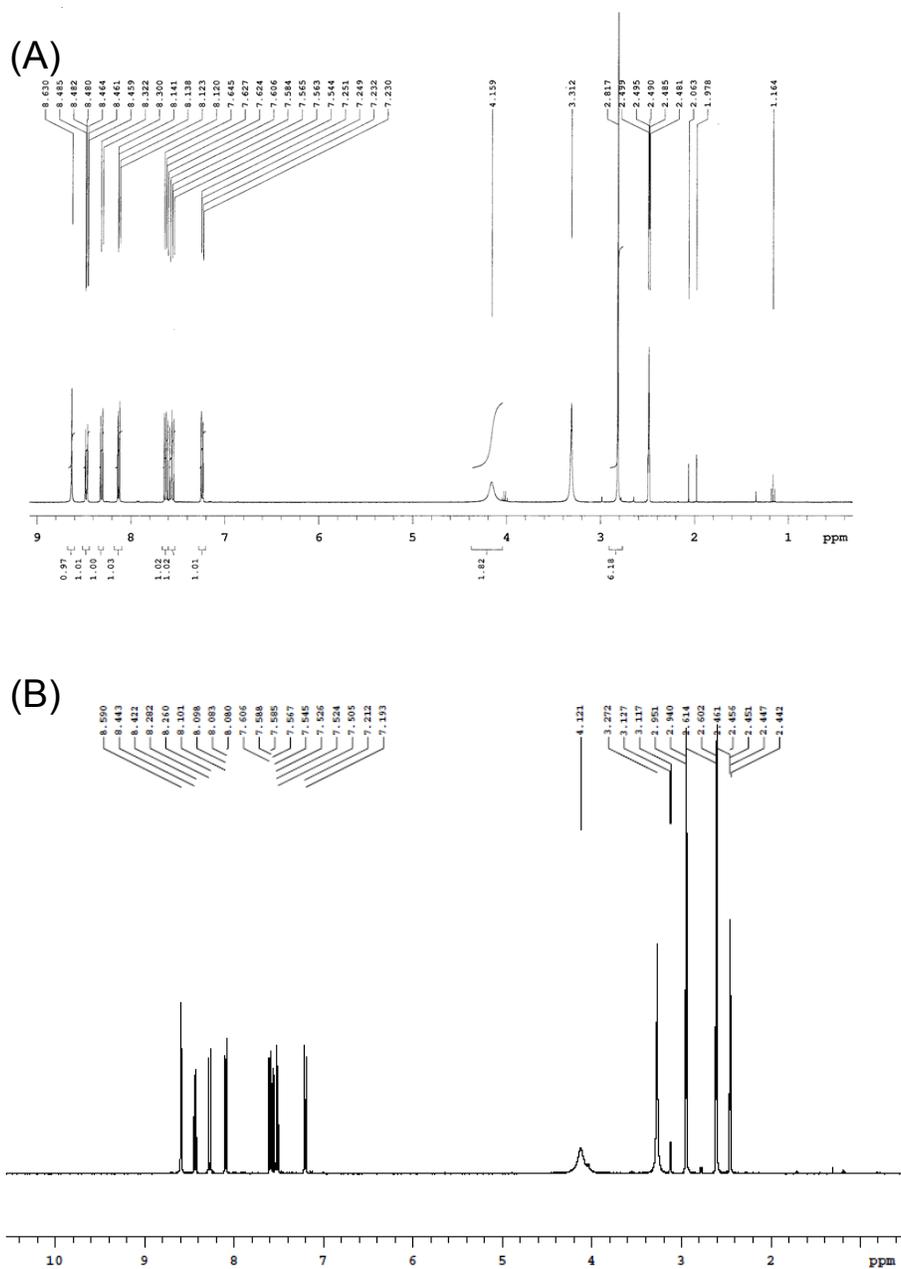


Figure 3.2 ¹H NMR spectrum of (A) ¹²C-dansylhydrazine standard and (B) synthesized ¹³C-dansylhydrazine.

The LC-MS results (Figure 3.3) shown below also confirmed the structure and purities of the product. The synthesized ¹³C-dansylhydrazine shared the same retention time with ¹²C-dansylhydrazine standard at 5.83 min. Their MS spectra also showed the

same pattern of three peaks, i.e., fragment ion peak at m/z 238.0747, $[M+H]$ peak at m/z 268.0960 and $[M+Na]$ peak at m/z 290.0769, compared to m/z 236.0738, 266.0955 and 288.0744, respectively, for the ^{12}C -dansylhydrazine standard).

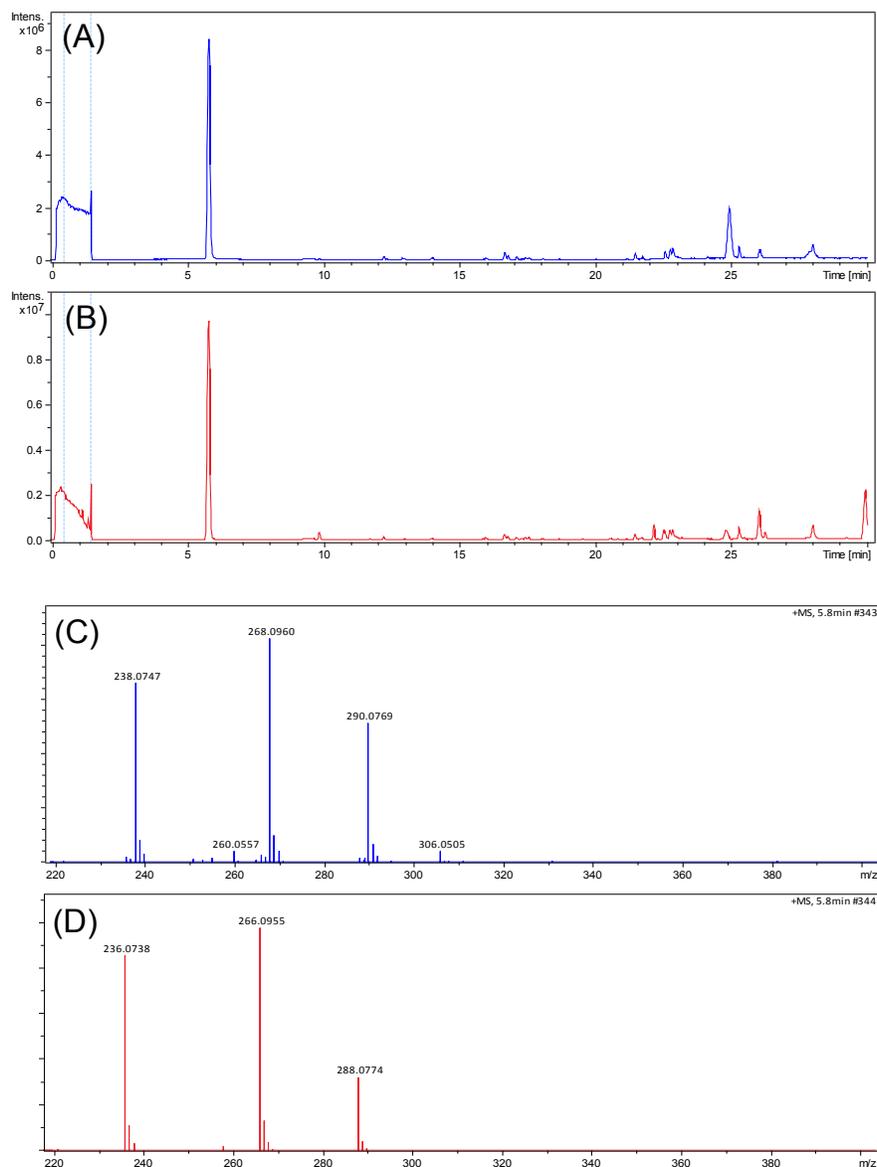


Figure 3.3 LC-MS analysis of ^{12}C -dansylhydrazine standard and synthesized ^{13}C -dansylhydrazine: (A) LC ion chromatogram of ^{13}C -dansylhydrazine, (B) LC ion chromatogram of ^{12}C -dansylhydrazine, (C) MS spectrum of ^{13}C -dansylhydrazine, and (D) MS spectrum of ^{12}C -dansylhydrazine.

3.2.4. Solution Preparation.

78 standards of carbonyl metabolites were obtained from the Human Metabolome Database (HMDB) compound library at the University of Alberta. Each compound was dissolved in methanol/water mixture to a concentration of 5 mM. If the solubility of a compound was too low, the supernatant of a saturated solution was used. All the prepared solutions were stored in -20 °C. The labeling reagent, DnsHz, was dissolved in methanol. Hydrochloric acid (HCl) was prepared by diluting a commercial 12 M HCl solution to 144 mM with methanol.

3.2.5. Urine Sample Collection and Preparation.

Urine samples were collected from a healthy individual. An informed consent was obtained from the volunteer before study. The ethics approval was obtained from the University of Alberta Ethics Approval Board. After collection, the urine samples were stored at 4 °C immediately. The samples then were centrifuged at 20817 g for 10 min. The supernatant was filtered through a 0.2 µm filter, which was then aliquoted and stored in -80 °C for further use.

3.2.6. Dansylhydrazine Labeling Reaction.

For standard labeling, 20 µL of a standard solution were mixed with 20 µL of 144 mM HCl solution and 20 µL of 20 mM ¹²C-DnsHz. After vortexing and spinning down, the mixture was incubated at 40 °C for 60 min. Then the mixture was cooled down in -80 °C freezer for 10 min to stop the labeling reaction. After that, the cooled solution was dried down using SpeedVac to remove the acid catalyst, followed by re-dissolving in 80

μL of acetonitrile/water mixture (ACN/ H_2O , 50:50, v/v). The solution was diluted and centrifuged at 15294 g for 10 min before injecting into LC-MS for analysis.

To prepare the 1:1 ^{12}C -/ ^{13}C -labeled urine sample for method development and analytical performance evaluation, 40 μL of human urine were divided into two parts: 20 μL was labeled by ^{12}C -DnsHz and the other 20 μL was labeled by ^{13}C -DnsHz following the same procedure described above for standard labeling. After labeling and re-dissolving, the ^{12}C -labeled urine sample was mixed with the same volume of the ^{13}C -labeled urine. The mixture was injected into LC-MS for analysis.

3.2.7. LC-MS and MS/MS

The labeled sample was analyzed using UltiMate 3000 UHPLC (Thermo Scientific, MA) combined with an Impact HD Quadrupole Time-of-flight (QTOF) mass spectrometer (Bruker, Billerica, MA). An RP column (Agilent Eclipse Plus C18 column, 2.1 mm \times 10 cm, 1.8 μm particle size, 95 \AA pore size) was used to separate labeled metabolites in a sample. Mobile phase A (MPA) was 0.1% (v/v) formic acid and 5% (v/v) ACN in water. Mobile phase B (MPB) was 0.1% (v/v) formic acid in ACN. The gradient for metabolite separation was: t = 0 min, 1% B; t = 3 min, 25% B; t = 23 min, 99% B; t = 34 min, 99% B. Between two sample runs, 11 min washing was used to wash and re-equilibrate the column. The flow rate was 180 $\mu\text{L}/\text{min}$. The column temperature was kept at 30 $^\circ\text{C}$. All MS spectra were collected in positive ion mode at a spectral acquisition rate of 1 Hz.

The MS/MS spectra of labeled standards and for urine metabolite identification were obtained under the collision energy of 20-50 eV in the stepping mode in the QTOF. The mass window width used in MS/MS experiments was 1 Da.

3.2.8. Data Processing and Analysis

A set of software programs developed in house as reported previously and also freely available from www.mycompoundid.org were used to process the CIL LC-MS data. In the 1st step, IsoMS⁴⁸ was used to extract peak pairs from mass spectra, including filtering out the redundant pairs (e.g., those of adduct ions and dimers). In this way, only a protonated ion of a peak pair was retained for a ¹²C-/¹³C-labeled metabolite. IsoMS also calculated the intensity ratio of each peak pair, and aligned multiple pairs of the same individual metabolites from different runs according to retention time and accurate mass. In the 2nd step, the missing values of intensity ratios in the aligned file were filled by the Zerofill program.¹⁴⁰ Finally, the chromatographic peak ratio of each peak pair was determined by IsoMS-Quant¹⁴¹ and used to produce the final metabolite-intensity table for relative quantification. Positive identification of the labeled metabolites was done using the newly constructed DnsHz standard library (see below). For putative metabolite identification, accurate masses of peak pairs were searched against those of the metabolite entries in metabolome databases by using MyCompoundID (MCID) at www.mycompoundid.org with zero-reaction and one-reaction libraries.¹⁹

3.3. Results and Discussion

3.3.1. Standard-Mixture Labeling

In order to optimize the DnsHz labeling conditions, six aldehyde/ketone standards with very different structures (Figure 3.4) were mixed and labeled together under different conditions with ¹²C-DnsHz. These standards were chosen to represent several

important classes of carbonyl metabolites: acetaldehyde and 2-butanone for low-molecular-weight aldehydes and ketones, cyclohexanone for cyclic ketones, benzaldehyde for aromatic carbonyl compounds, androsterone for steroids and galactose for carbohydrates or sugars. The ^{12}C -labeled standard-mixture was mixed with a ^{13}C -labeled standard-mixture that served as an internal standard, followed by LC-MS analysis. Six experimental replicate analyses were performed for each condition. The peak area ratio of a ^{12}C -/ ^{13}C -labeled standard-mixture was used to evaluate the efficiencies of different labeling conditions and the relative standard deviation (RSD) of the peak area ratio ($n=6$) was used as the indicator of reproducibility (Figure 3.5). Using this approach, the optimal conditions for reaction solvents, acids, time and temperature were screened.

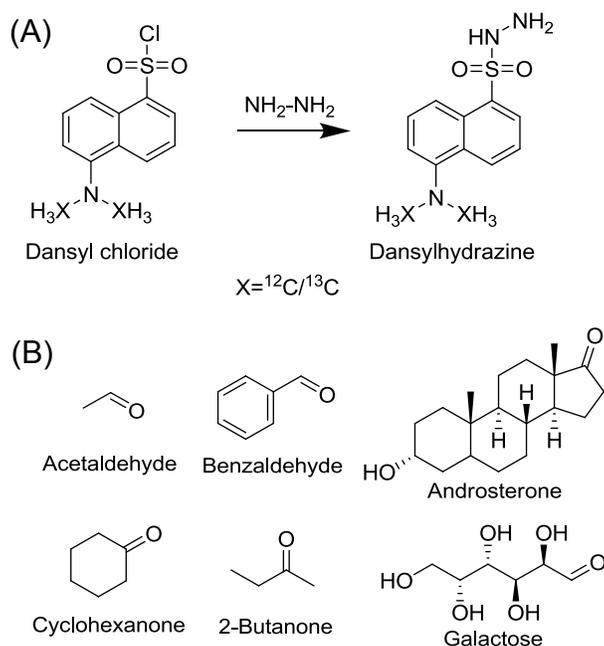


Figure 3.4 (A) Scheme of ^{12}C - or ^{13}C -dansylhydrazine synthesis. (B) Structures of six carbonyl-containing metabolites used to prepare a standard-mixture for method development and evaluation.

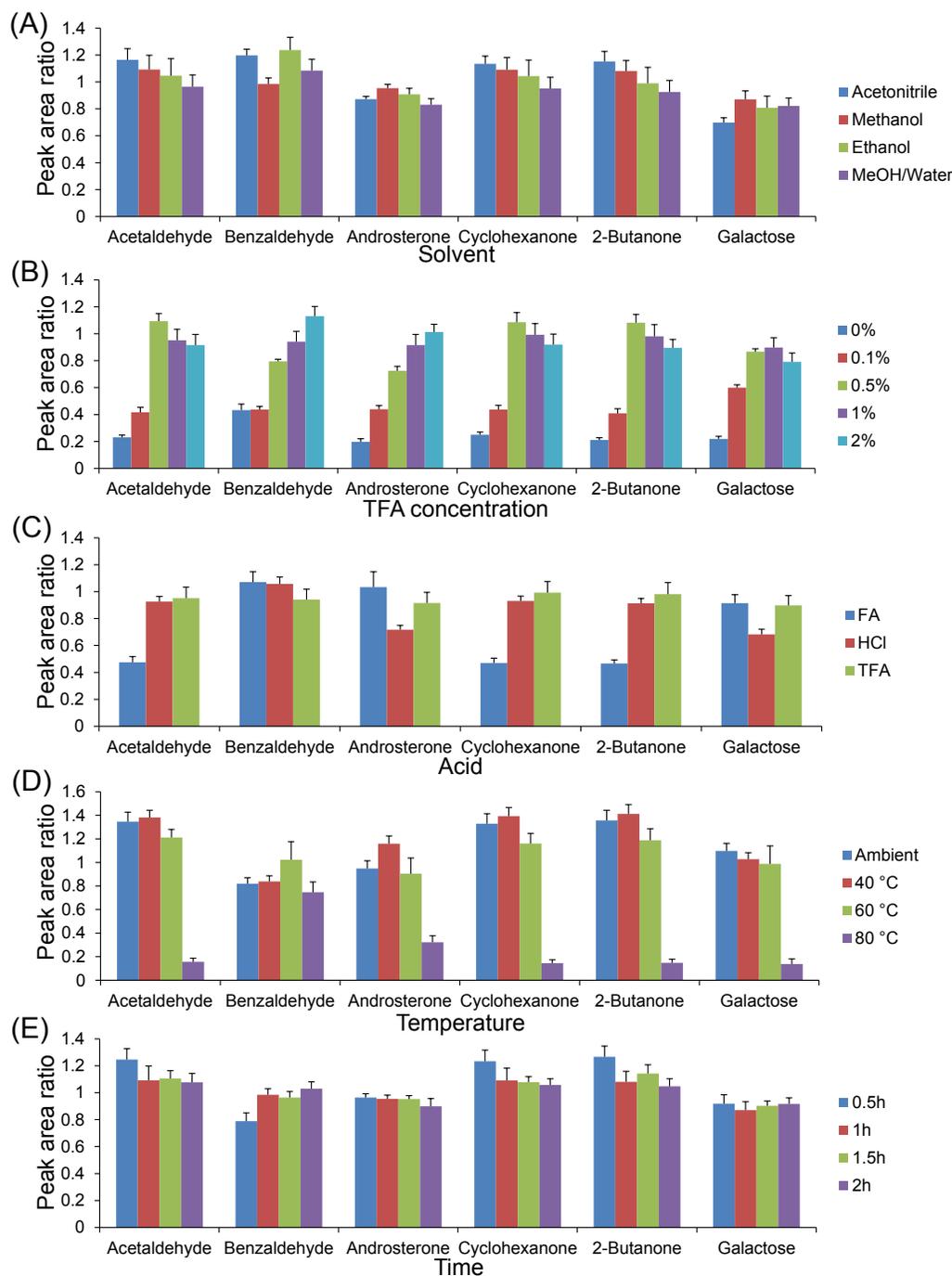


Figure 3.5 Comparison of efficiency and reproducibility for labeling a standard-mixture under different reaction conditions: (A) effect of solvent type, (B) effect of acid concentration, (C) effect of acid type, (D) effect of reaction temperature, and (E) effect of reaction time. Data are presented as the mean \pm S.D. of six replicates (n=6).

Figure 3.5A shows the effect of solvent type on labeling efficiency. For all six standards, the labeling efficiency was similar in the tested solvents, namely acetonitrile, methanol, ethanol and methanol/water (1:1, v/v). In theory, in the formation of hydrazones, decreasing water (also a product of the labeling reaction) would promote the forward reaction. However, our results indicated that the presence of water in the reaction solution did not adversely affect the labeling efficiency. This is good for practical applications as water is often present in biological samples and direct labeling of water-containing samples can simplify the experimental procedure (i.e., no need to remove water). As Figure 3.5A shows, the RSD of peak ratios calculated from repeated labeling experiments was generally less than 10% when acetonitrile, methanol and methanol/water were used, showing good labeling reproducibility. In the case of ethanol, the RSD was up to 12%.

In addition to the selection of a solvent, acid was added to serve as a catalyst which also plays an important role in the labeling reaction. Acid or proton makes the carbonyl group more electrophilic to aid the attack by the nucleophilic -NH_2 group in dansylhydrazine and also helps the elimination of water from the intermediate to form the final product. The results from acid concentration screening confirmed that the reaction needs protons to catalyze the reaction; 0.5% trifluoroacetic acid (TFA) was found to be optimal in terms of both efficiency and reproducibility (Figure 3.5B). Within the similar pH range, HCl and TFA shared similar labeling efficiency and reproducibility (Figure 3.5C). For several standards in the mixture, formic acid (FA) was not a good acid for the labeling and the RSD was relatively larger than that of HCl and TFA. The other two

parameters affecting the reaction process are reaction temperature and time. The optimal conditions were found to be 40 °C for 1 h (Figure 3.5D and 3.5E).

The benefit of using DnsHz labeling for MS detection is shown in Table 3.1 where the signals (i.e., the areas of chromatographic peaks) of unlabeled and labeled standards from injecting the same amount of each compound are compared. For acetaldehyde, the unlabeled compound could not be detected, but the labeled compound could be readily analyzed. For the other 5 standards, signals were significantly increased by DnsHz labeling. The signal enhancement factor ranged from ~15-fold increase for androsterone to ~940-fold for 2-butanone.

Table 3.1 Signal comparison between unlabeled metabolites and labeled metabolites.

Compound	Unlabeled		DnsHz Labeled		Enhancement Factor
	m/z	Peak area	m/z	Peak area	
Acetaldehyde	45.0335	n.a.*	292.1114	4.07E+06	-
2-Butanone	73.0648	4.16E+04	354.1271	3.89E+07	937
Benzaldehyde	107.0491	4.47E+04	320.1427	3.82E+07	854
Cyclohexnone	99.0804	5.30E+04	346.1584	4.23E+07	798
Galactose	181.0707	1.03E+05	428.1486	4.64E+06	45
Androsterone	291.2319	4.77E+06	538.3098	6.99E+07	15

*n.a. indicates that the metabolite is not detectable.

In this work, we did not determine the absolute efficiency of DnsHz labeling for individual compounds. Derivatization efficiency is very much compound-dependent. It is also dependent on the compound concentration and sample matrix. Fortunately, the information on derivatization efficiency of a given metabolite is not needed for relative quantification of the metabolite in comparative samples. For relative quantification using a pooled sample as the control, derivatization efficiency should be very similar for labeling a metabolite in an individual sample vs. labeling the same metabolite in the

pooled sample as this metabolite has similar concentrations and matrices in these samples. As long as the derivatization efficiency is constant, the peak intensity ratio of the ^{12}C -labeled metabolite in a sample vs. the ^{13}C -labeled metabolite in the control should reflect the concentration difference and thus accurate relative quantification can be obtained (see below for urine sample labeling results).

3.3.2. Urine Sample Labeling

While the six standards used for optimizing the experimental conditions represent different classes of carbonyl metabolites, a real biological sample contains many more different compounds with different concentrations. To determine whether the optimized condition is applicable to analyze complex samples, human urine was chosen to test the labeling method. Urine was aliquoted to two parts. One was labeled by ^{12}C -DnsHz and the other aliquot was labeled by ^{13}C -DnsHz using the same condition. Two labeled samples were then mixed by equal volume to form a 1:1 ^{12}C -/ ^{13}C -labeled urine sample for LC-MS analysis. The number of peak pairs detected was used to gauge the performance of a tested experimental condition. For each condition, six LC-MS runs were performed from experimental triplicates of sample processing and injection duplicate of each processed sample.

Figure 3.6A shows the study of the effect of water in a urine sample on labeling. Two methods were applied to remove water from urine before labeling. One was to dry down the urine and then re-dissolve it with methanol, followed by labeling. The other way was to use ethyl acetate to extract the urine metabolites, followed by drying, re-dissolving and labeling the extract. Figure 3.6A clearly shows that direct labeling of urine generated a much higher number of peak pairs. The differences can be attributed to the

metabolite loss in the two drying methods (e.g., loss of volatile compounds during solvent evaporation and loss of low-abundance metabolites from adsorption to the vials). The results shown in Figure 3.6A also confirmed that the existence of water in the labeling condition was acceptable.

Figure 3.6B shows the effect of TFA concentration in the urine labeling solution on the number of peak pairs detected. As in the case of the analysis of the standard mixture, 0.5% TFA gave the optimal result, i.e., the highest peak pair number in this case. There were no significant differences between labeling with TFA (0.5% v/v) and HCl (48 mM) at the same pH value (Figure 3.6C). However, if TFA was used as the acid catalyst, a large peak from the labeled TFA products was observed in the ion chromatogram. For targeted analysis of a few metabolites, this peak may not interfere with the analysis of the metabolites of interest. However, for a real sample metabolomic profiling work, due to the presence of so many metabolites, the labeled TFA peak can interfere with the detection of co-eluting metabolites at the retention window of the labeled TFA (from 12.6 to 13.3 min). Thus, 48 mM HCl was decided to be the optimal catalyst for labeling complex samples such as urine.

The results of reaction time and temperature screening were also matched with those of labeling the standard mixture (Figure 3.6D and 3.6E). Thus, incubating at 40 °C for 1 h was chosen as the optimal condition for labeling complex samples. For labeling a complex sample, using a sufficient amount of reagent is also important to drive the labeling reaction to yield the maximum amounts of products for all the metabolites. To determine how the reagent amounts affect the labeling, different amounts of DnsHz reagent were added to a urine sample to contain 5 mM, 10 mM, or 20 mM

dansylhydrazine. The results shown in Figure 3.6F indicate that the use of 10 or 20 mM DnsHz reached the optimal number of peak pairs detected.

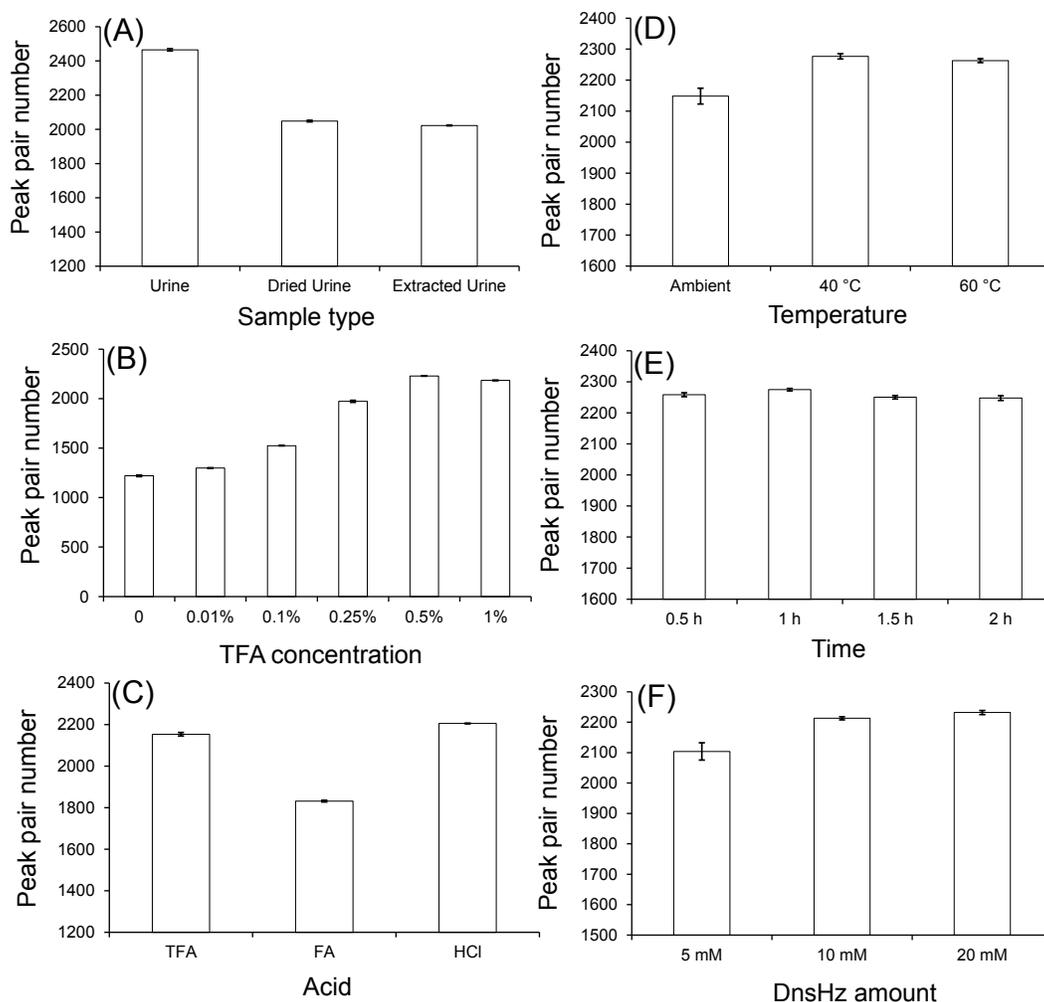


Figure 3.6 Peak pair numbers detected from ^{12}C -/ ^{13}C -dansylhydrazine labeled urine samples prepared under different labeling conditions: (A) effect of sample preparation, (B) effect of acid concentration, (C) effect of acid type, (D) effect of temperature, (E) effect of time, and (F) effect of dansylhydrazine amount. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).

Taken together from the results obtained from the analyses of the standard mixture and human urine, we determined that the optimal conditions for DnsHz labeling for profiling the carbonyl submetabolome should include the use of 48 mM HCl as a catalyst and incubation of labeling solution at 40 °C for 1 h. Water in a sample is acceptable. For urine sample analysis, 20 mM dansylhydrazine in the labeling solution is recommended. For other type of biological samples, the optimal reagent amount to be used for mixing with a sample should be optimized to ensure that the labeling reaction is driven to a maximal extent of completion.

3.3.3. Accuracy and Precision

As Figure 3.1 shows, the entire workflow of DnsHz labeling LC-MS involves several steps with each step potentially affecting the accuracy and precision of the method. We evaluated the performance of our method using the ^{12}C -/ ^{13}C -labeled urine samples. In the intra-day comparison experiments, urine was labeled three times and analyzed by LC-MS with duplicate for each sample in one day. Figure 3.7A shows the Venn diagram of the peak pair numbers obtained from triplicate labeling. An average of 2030 ± 39 pairs per sample ($n=6$) were detected with 1737 pairs in common, indicating high reproducibility of peak pair detection. Figure 3.7B-C shows the distribution of peak pair numbers as a function of the averaged peak area ratio and the RSD of the ratio ($n=6$). The ratios of most peak pairs were close to the expected value of 1.0. For example, 95.5% of the pairs had values of between 0.67 and 1.5, i.e., within $\pm 50\%$ of the expected value, indicating that high accuracy could be achieved. The average RSD of the entire data set was 7.6% and 95.6% of the peak pairs had a RSD value of less than 20%, demonstrating that high precision could also be achieved.

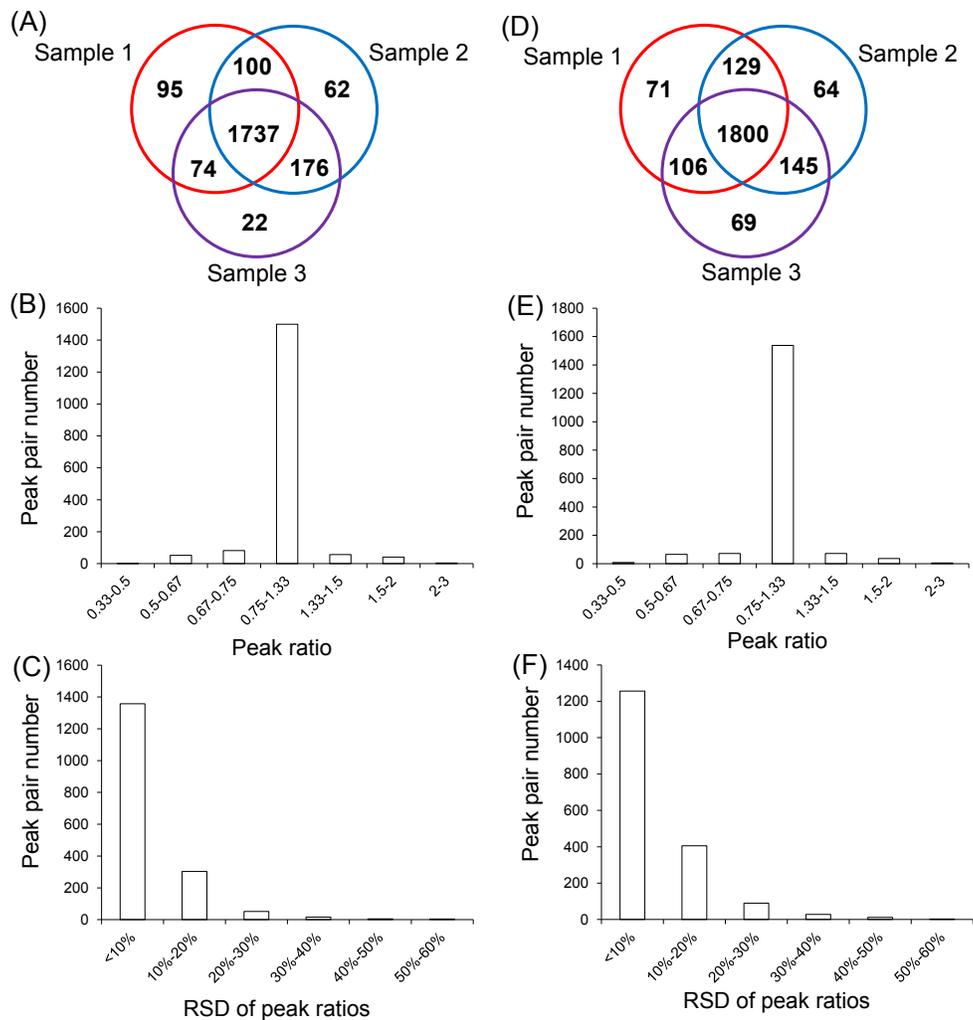


Figure 3.7 Venn diagrams of the peak pair numbers detected in triplicate labeled urine from (A) the intra-day experiments and (D) the inter-day experiments. Distributions of peak pair numbers as a function of averaged peak ratio of (B) intra-day results and (E) inter-day results and RSD of (C) intra-day results and (F) inter-day results. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).

Similar evaluation was also conducted with inter-day experiments (Figure 3.7D-F). Urine was labeled three times in different days. The Venn diagram of the peak pair numbers obtained from the experimental triplicate with injection duplicate also shows

high overlap of the peak pairs detected (Figure 3.7D). A total of 1800 peak pairs were detected in all runs with an average of 2121 ± 16 peak pairs detected per run ($n=6$). The results of peak pair number distributions (Figure 3.7E-F) were in agreement with those of the intra-day experiments.

In this work, we did not determine the linear range for relative quantification; we will do this in future work for real world applications where a wide range of metabolite concentration changes among comparative samples are expected.

3.3.4. Product Stability

The stability of labeled samples was also tested since analyzing a large batch or multiple batches of samples may take days or a month to complete in metabolomics. In this work, a $^{12}\text{C}/^{13}\text{C}$ -labeled urine sample mixture was aliquoted and stored under different conditions for different periods of time, including 4 °C for overnight, -20 °C for a week, -20 °C for a month, and -80 °C for a month. The results of the analyses of these samples are shown in Figure 3.8. As it shows, there is no significant difference of the peak pair numbers detected among these storage conditions. Thus, the labeled samples are very stable and can meet the need of month-long storage for completing a batch analysis of many samples in a large-scale metabolomic profiling project (e.g., with the current setup, a monthly throughput is $30 \text{ days} \times 32 \text{ samples/day} = 960 \text{ samples per month}$).

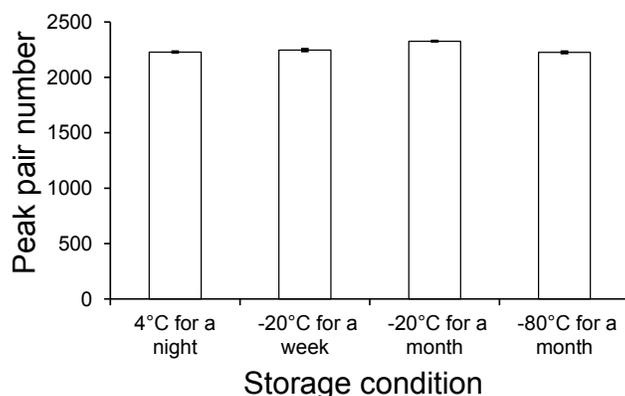


Figure 3.8 Storage condition test.

3.3.5. Dansylhydrazine-labeled Standard Library

The final step of the workflow shown in Figure 3.1 is metabolite identification. Positive or definitive identification requires the use of an authentic standard as a reference for comparing the measured parameters with those of an unknown metabolite. To this end, we have constructed a DnsHz-labeled standard library that currently consists of 78 endogenous human metabolites (Table 3.2). This library is expandable in the future by adding more standards when they become available. To construct the library, each standard was individually labeled and analyzed using LC-MS and MS/MS to generate “triple-parameters” information: accurate molecular mass, retention time (RT) and MS/MS fragmentation spectrum. This library, along with a database search program (DnsHz-ID), is freely available from a publicly accessible website at www.mycompoundid.org. A tutorial for the use of this resource for metabolite identification based on MS, RT and MS/MS matches and an example of metabolite identification is shown in Appendix. These notes are also provided on the website.

Table 3.2 List of 90 entries from 78 compounds for the DnsHz-labeled standard library.

HMDB #	Metabolites	Class	RT (min)
HMDB01426	Formaldehyde	Aldehydes	10.66
HMDB00990	Acetaldehyde	Aldehydes	11.20
HMDB03366	Propanal	Aldehydes	13.28
HMDB03543	Butanal	Aldehydes	14.83
HMDB05994	Hexanal	Aldehydes	17.91
HMDB01140	Octanal	Aldehydes	20.72
HMDB11623	Decanal	Aldehydes	23.26
HMDB01051	Glyceraldehyde	Aldehydes	7.59
HMDB06478	3-Methylbutanal	Aldehydes	16.15
HMDB06115	Benzaldehyde	Aldehydes	16.23
HMDB06236	Phenylacetaldehyde	Aldehydes	16.35
HMDB01358	Retinal	Aldehydes	27.16
HMDB59965	3,4-Dihydroxybenzaldehyde	Aldehydes	11.05
HMDB03441	Cinnamic Aldehyde	Aldehydes	17.23
HMDB01659	Acetone	Ketones	12.01
HMDB00474	2-butanone	Ketones	14.24
HMDB03407	Diacetyl	Ketones	13.97
HMDB03315	Cyclohexanone	Ketones	15.29
HMDB01184	Methyl propenyl ketone	Ketones	15.25
HMDB04487	Carvone	Ketones	20.79
HMDB02670	Naringenin	Ketones	14.03
HMDB01545	Pyridoxal	Ketones	7.69
HMDB00183	Kynurenine	Ketones	7.57
HMDB01935	3-(alpha-Acetylbenzyl)-4-hydroxycoumarin	Ketones	18.27
HMDB00015	Cortexolone	Ketones	17.06
HMDB03075	Flavone	Ketones	21.27
HMDB02927	Naringin	Ketones	9.49
HMDB01892	Vitamin K3	Ketones	22.27
HMDB03066	Chalcone	Ketones	21.15
HMDB00005	2-Ketobutyric	Keto-Acids	11.10
HMDB01259	Succinic acid semialdehyde	Keto-Acids	9.34
HMDB00019	alpha-Ketoisovaleric acid	Keto-Acids	16.78
HMDB00720	Levulinic acid	Keto-Acids	9.59
HMDB00491	3-Methyl-2-oxovaleric acid	Keto-Acids	13.39
HMDB00408	2-Methyl-3-ketovaleric acid	Keto-Acids	13.39
HMDB01864	2-Ketohexanoic acid	Keto-Acids	13.79
HMDB00208	Oxoglutaric acid	Keto-Acids	10.47
HMDB01149	Aminolevulinic acid	Keto-Acids	6.27
HMDB00127	glucuronic acid	Carbohydrates	6.27

HMDB00186	Lactose	Carbohydrates	5.86
HMDB00143	Galactose	Carbohydrates	5.98
HMDB00122	Glucose	Carbohydrates	6.08
HMDB00660	Fructose	Carbohydrates	6.44
HMDB00646	Arabinose	Carbohydrates	6.44
HMDB00283	D-Ribose	Carbohydrates	6.68
HMDB00174	Fucose	Carbohydrates	6.86
HMDB00124	Fructose-6-phosphate	Carbohydrates	6.46
HMDB02545	Galacturonic acid	Carbohydrates	6.32
HMDB00825	Sialyllactose	Carbohydrates	6.03
HMDB00467	Nutriacholic acid	Steroids	18.03
HMDB00503	7 α -hydroxy-3-oxo-5 β cholanoic acid	Steroids	20.57
HMDB00031	Androsterone	Steroids	20.71
HMDB01830	Progesterone	Steroids	22.40
HMDB00374	17-Hydroxyprogesterone	Steroids	18.27
HMDB00037	Aldosterone	Steroids	14.47
HMDB00145	Estrone	Steroids	17.60
HMDB00234	Testosterone	Steroids	19.06
HMDB00899	5 α -Androstane-3,17-dione	Steroids	20.16
HMDB00053	Androstenedione	Steroids	20.05
HMDB00063	Cortisol	Steroids	15.19
HMDB00077	Dehydroepiandrosterone	Steroids	17.88
HMDB00995	16-Dehydroprogesterone	Steroids	20.99
HMDB01547	Corticosterone	Steroids	17.55
HMDB00016	11-Deoxycorticosterone	Steroids	19.50
HMDB00903	Tetrahydrocortisone	Steroids	17.39
HMDB00628	Epitestosterone	Steroids	20.20
HMDB00920	11 α -Hydroxyprogesterone	Steroids	18.72
HMDB00546	Epitiocholanolone	Steroids	19.05
HMDB01939	Medroxyprogesterone	Steroids	20.61
HMDB00502	3-Oxocholic acid	Steroids	17.33
HMDB00490	5 β -Androsterone	Steroids	20.57
HMDB01867	4-Aminohippuric	Acyl Glycines	8.07
HMDB00840	Salicyluric acid	Acyl Glycines	11.73
HMDB03269	Nicotinuric acid	Acyl Glycines	7.19
HMDB00714	Hippuric Acid	Acyl Glycines	10.56
HMDB00821	Phenylacetyl glycine	Acyl Glycines	11.07
HMDB00422	2-Methylglutaric acid	Dicarboxylic Acids	8.76
HMDB01167	Pyruvaldehyde*	Aldehydes/Ketones	17.66
HMDB03407	Diacetyl*	Ketones	18.59
HMDB01830	Progesterone*	Steroids	24.66
HMDB00374	17-Hydroxyprogesterone*	Steroids	22.76
HMDB00899	5 α -Androstane-3,17-dione*	Steroids	22.94

HMDB00053	Androstenedione*	Steroids	22.67
HMDB00063	Cortisol*	Steroids	21.64
HMDB00995	16-Dehydroprogesterone*	Steroids	24.65
HMDB01547	Corticosterone*	Steroids	23.13
HMDB00016	11-Deoxycorticosterone*	Steroids	23.83
HMDB00903	Tetrahydrocortisone*	Steroids	21.79
HMDB00920	11a-Hydroxyprogesterone*	Steroids	22.89
HMDB01939	Medroxyprogesterone*	Steroids	23.45

* Indicate the metabolites are labeled with 2 tags

It should be noted that retention time is not often used as a match parameter, as its value can be easily affected by a number of experimental conditions used for LC-MS, including minor variations in LC equipment used, connecting tube length from LC to MS, solvent composition for running LC, temperature, etc. To facilitate the comparison of user-generated RT values with those of labeled standards in the library, a RT normalization method was used to correct the RT shift due to differences in instrument setups and running conditions between a user's system and our library system.⁵⁹ In this approach, all RT values in the library were normalized to a set of RT calibrants containing 17 DnsHz-labeled standards (Table 3.3). The ion chromatogram of this mixture is shown in Figure 3.9A. The RT difference between two adjacent peaks is less than 2 min. A user needs to run the same mixture of RT calibrants and then upload the LC-MS data to the DnsHz-ID program on the website for RT correction. A built-in program in DnsHz-ID compares the RT values of the calibrants in the user's file with those of the library calibrants and determines the correction parameters. These parameters are then applied, automatically, to the real sample LC-MS data to produce a corrected RT value for each unknown metabolite detected in the sample. These corrected RT values are compared to the library RT values of the standards for matching. The RT calibrants are available upon request from the authors and can also be made by the user.

Table 3.3 List of RT calibrants consisting of 17 labeled standards.

#	HMDB #	Metabolite	m/z	RT
1	HMDB00186	Lactose	590.2014	5.86
2	HMDB00174	Fucose	412.1537	6.86
3	HMDB01545	Pyridoxal	415.1434	7.69
4	HMDB01259	Succinic acid semialdehyde	350.1169	9.34
5	HMDB01426	Formaldehyde	278.0958	10.66
6	HMDB01659	Acetone	306.1271	12.01
7	HMDB03366	Propanal	306.1271	13.28
8	HMDB00474	2-Butanone	320.1427	14.24
9	HMDB03315	Cyclohexanone	346.1584	15.29
10	HMDB06115	Benzaldehyde	354.1271	16.23
11	HMDB00145	Estrone	518.2472	17.60
12	HMDB00546	Epietiocholanolone	538.3098	19.05
13	HMDB01140	Octanal	376.2053	20.72
14	HMDB00063	Cortisol (2 tags)	429.1899	21.64
15	HMDB11623	Decanal	404.2366	23.26
16	HMDB00995	16-Dehydroprogesterone (2 tags)	404.1897	24.65
17	HMDB01358	Retinal	532.2992	27.16

3.3.6. Urine Carbonyl Submetabolome Profiling

To demonstrate the applicability and performance of the developed method for profiling complex samples, the urinary carbonyl submetabolome was analyzed. Figure 3.9B shows a base-peak ion chromatogram of ^{12}C -/ ^{13}C -labeled urine. Many peaks can be detected, indicating the presence of various carbonyl compounds in urine. To detect as many peak pairs as possible, sample injection amount for LC-MS should be optimized. In our work, we normally generate a plot of the peak pair number as a function of injection volume (see Figure 3.9C). A plateau was achieved when the injection volume was 10 μL for urine carbonyl submetabolome profiling. For analyzing other type of samples, a similar approach can be used to find the optimal injection amount for LC-MS.

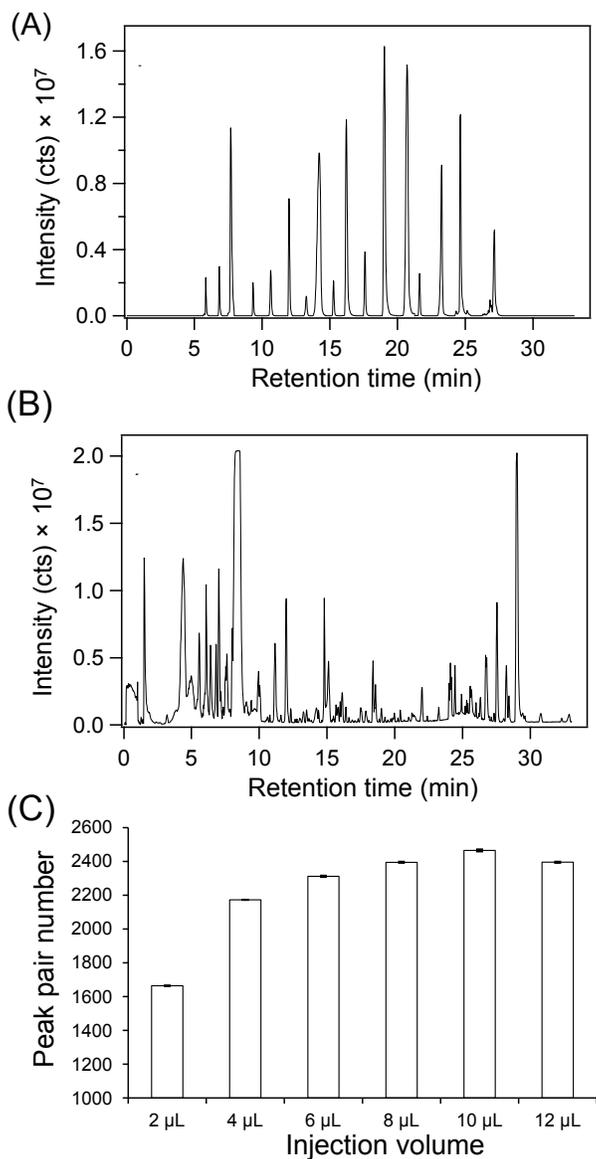


Figure 3.9 (A) LC-MS ion chromatogram of RT calibrants consisting of 17 labeled standards. (B) LC-MS ion chromatogram of ¹²C-/¹³C-dansylhydrazine labeled urine. (C) Peak pair number detected as a function of injection amount of ¹²C-/¹³C-dansylhydrazine labeled urine. Data are presented as mean ± S.D. (n=3).

As discussed earlier, replicate analyses of the ¹²C-/¹³C-labeled urine resulted in the detection of 1737 common peak pairs in all 6 runs. These peak pairs were identified

using three steps at different levels of confidence. In the first step, highly confident and positive metabolite identification was done using the DnsHz labeled standard library. 33 peak pairs could be identified based on both accurate mass and retention time matches with those of the library standards (Table 3.4). It is clear that only a very small fraction of the detected peak pairs were positively identified. Increasing the number of standards in the library is needed. However, the availability of standards is currently limited. In the future, we plan to use chemical synthesis or biologically derived compounds to expand the library.

In the 2nd step of metabolite identification, the remaining peak pairs were matched to some metabolites by searching accurate masses of peak pairs against the MyCompoundID (MCID) zero-reaction library containing 8,021 known human endogenous metabolites (i.e., zero-reaction library). A total of 406 peak pairs were matched to the zero-reaction library. The structures of matched metabolites were manually checked. Over 96.3% of these peak pairs were matched to metabolites containing C=O moiety. In the last step, the remaining unmatched peak pairs were searched against the MCID one-reaction library that contains 375,809 predicted compounds after applying one round of metabolic reactions to the 8,021 human endogenous metabolites. A total of 927 peak pairs were matched. Thus, out of the 1737 peak pairs detected, 78.6% (1366 pairs) could be either identified or matched.

Table 3.4 List of metabolites identified based on accurate mass and retention time matches to the DnsHz-labeled standard library.

Peak pair information							Identification result			
Peak pair #	RT (min)	Corrected RT (min)	mz_light	mz_heavy	monoisotopic mass (Da)	HMDB.No.	Name	Accurate mass	mz_light	library RT (min)
127	5.52	5.52	590.1994	592.2067	342.1142	HMDB00186	Lactose	342.1162	590.2014	5.86
153	5.86	5.86	590.2009	592.2073	342.1157	HMDB00186	Lactose	342.1162	590.2014	5.86
177	6.10	6.05	428.1482	430.1546	180.0630	HMDB00143	Galactose	180.0634	428.1486	5.98
						HMDB00122	Glucose	180.0634	428.1486	6.08
						HMDB00660	Fructose	180.0634	428.1486	6.44
182	6.17	6.10	442.1274	444.1339	194.0422	HMDB00127	Glucuronic acid	194.0427	442.1279	6.27
						HMDB02545	Galacturonic acid	194.0427	442.1279	6.32
232	6.67	6.49	398.1371	400.1436	150.0519	HMDB00646	Arabinose	150.0528	398.138	6.44
						HMDB00283	Ribose	150.0528	398.138	6.68
252	6.90	6.67	428.1452	430.1534	180.0600	HMDB00660	Fructose	180.0634	428.1486	6.44
280	7.14	6.86	412.1510	414.1582	164.0658	HMDB00174	Fucose	164.0685	412.1537	6.86
309	7.62	7.52	338.1150	340.1218	90.0298	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59
324	7.74	7.69	415.1415	417.1485	167.0563	HMDB01545	Pyridoxal	167.0582	415.1434	7.69
641	10.46	10.42	427.1419	429.1480	179.0567	HMDB00714	Hippuric Acid	179.0582	427.1434	10.56
746	11.01	10.99	350.1150	352.1218	102.0298	HMDB00005	2-Ketobutyric	102.0317	350.1169	11.10
856	11.54	11.54	350.1142	352.1218	102.0290	HMDB00005	2-Ketobutyric	102.0317	350.1169	11.10
964	11.99	12.01	306.1264	308.1329	58.0412	HMDB01659	Acetone	58.0419	306.1271	12.01
1045	12.42	12.43	306.1252	308.1305	58.0400	HMDB01659	Acetone	58.0419	306.1271	12.01
1181	13.11	13.10	378.1464	380.1533	130.0612	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39
						HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39
1212	13.30	13.28	306.1261	308.1328	58.0409	HMDB03366	Propanal	58.0419	306.1271	13.28
1215	13.31	13.29	378.1460	380.1533	130.0608	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39
						HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39
						HMDB01864	2-Ketohexanoic acid	130.0630	378.1482	13.79
1249	13.51	13.49	378.1476	380.1540	130.0624	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39

						HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39
						HMDB01864	2-Ketohexanoic acid	130.0630	378.1482	13.79
1381	14.23	14.24	320.1416	322.1482	72.0564	HMDB00474	2-butanone	72.0575	320.1427	14.24
1487	14.84	14.85	332.1405	334.1463	84.0553	HMDB01184	Methyl propenyl ketone	84.0575	332.1427	15.25
1528	15.12	15.13	320.1419	322.1485	72.0567	HMDB03543	Butanal	72.0575	320.1427	14.83
1548	15.28	15.29	346.1560	348.1620	98.0708	HMDB03315	Cyclohexanone	98.0732	346.1584	15.29
1594	15.52	15.52	346.1566	348.1631	98.0714	HMDB03315	Cyclohexanone	98.0732	346.1584	15.29
1627	15.82	15.80	334.1575	336.1641	86.0723	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
1664	16.10	16.07	334.1577	336.1642	86.0725	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
1691	16.37	16.32	368.1407	370.1472	120.0555	HMDB06236	Phenylacetaldehyde	120.0575	368.1427	16.35
1694	16.43	16.38	364.1305	366.1372	116.0453	HMDB00019	Alpha-Ketoisovaleric acid	116.0473	364.1325	16.78
1817	17.35	17.27	612.3079	614.3145	364.2227	HMDB00903	Tetrahydrocortisone	364.2250	612.3102	17.39
1914	18.28	18.20	578.3023	580.3093	330.2171	HMDB00374	17Hydroxyprogesterone	330.2195	578.3047	18.27
						HMDB00920	11a-Hydroxyprogesterone	330.2195	578.3047	18.72
1961	18.95	18.89	578.3017	580.3115	330.2165	HMDB00920	11a-Hydroxyprogesterone	330.2195	578.3047	18.72
2001	19.52	19.48	578.3031	580.3095	330.2179	HMDB00016	11-Deoxycorticosterone	330.2195	578.3047	19.50
2081	20.56	20.55	538.3048	540.3139	290.2196	HMDB00031	Androsterone	290.2246	538.3098	20.71
						HMDB00490	5b-Androsterone	290.2246	538.3098	20.57
2088	20.72	20.72	376.2047	378.2111	128.1195	HMDB01140	Octanal	128.1201	376.2053	20.72

3.4. Conclusions

We have developed a high-performance chemical isotope labeling LC-MS method for profiling the carbonyl submetabolome with high coverage and accuracy and precision. This method, used alone, should be useful for profiling important groups of metabolites such as hormones and sugars in complex biological samples. This method, in combination with previously reported three labeling methods for amine, carboxyl and hydroxyl submetabolomes, can potentially be used to profile over 95% of the entire metabolome, based on the grouping of metabolites using the four functional groups. The real world applications of this method for quantitative metabolomics will be reported in the future.

Chapter 4

Dansylhydrazine Isotope Labeling LC-MS for Comprehensive Carboxylic Acid Submetabolome Profiling

4.1. Introduction

Liquid chromatography-mass spectrometry (LC-MS) is one of the most commonly used analytical platforms for metabolome analysis.¹³ In a conventional LC-MS approach, a combination of two chromatographic conditions [i.e., reversed phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC)] and two ionization modes (i.e., positive and negative ion detection) is often used to increase the number of detected metabolites, thereby expanding the overall metabolome coverage.¹⁷⁰⁻¹⁷¹ To increase quantification accuracy, choosing appropriate internal standards for analytes can correct for sample loss, instrument drift and matrix effect in LC-MS. However, a great variety of metabolites with wide concentration ranges can present a major challenge for detecting and quantifying many analytes simultaneously in complex biological samples.

A different approach to address the challenge is to perform multichannel chemical isotope labeling LC-MS (mCIL LC-MS) analysis. In this approach, metabolites containing a certain chemical functional group (i.e., submetabolome) are derivatized using a pair of chemical isotope labeling reagents and then analyzed by LC-MS. During the derivatization, isotopic moiety can be introduced into the labeled metabolites, thereby creating stable isotopic internal standards for all the labeled metabolites for accurate quantification. In addition, the chemical and physical properties of metabolites are altered

to improve their separation and detection. If multiple CIL LC-MS methods are used for analyzing different chemical-group-based submetabolomes, the combined results would produce a near-complete metabolome profile.

There are a number of labeling reagents reported for targeted or untargeted metabolite analysis with varying degrees of performance. In our previous studies, several high-performance labeling methods have been developed, including $^{12}\text{C}/^{13}\text{C}$ -dansyl chloride (DnsCl) labeling for amine/phenol submetabolome analysis³⁰, base-activated $^{12}\text{C}/^{13}\text{C}$ -DnsCl labeling for hydroxyl submetabolome analysis³¹, $^{12}\text{C}/^{13}\text{C}$ -dansylhydrazine (DnsHz) for carbonyl submetabolome analysis¹⁷² and $^{12}\text{C}/^{13}\text{C}$ -dimethylaminophenacyl (DmPA) bromide labeling for carboxyl submetabolome analysis⁸⁵.

Carboxyl-containing metabolites have many important functions in biological systems. For examples, tricarboxylic acid (TCA) cycle is the key component in the process of energy production and biosynthesis, short-chain fatty acids as signalling molecules play a pivotal role in maintenance of gut and immune homeostasis,¹⁷³ and bile acids are critical for both absorption and elimination.¹⁷⁴ Several chemical derivatization methods have been reported for analyzing carboxylic acid using different MS platforms.^{89, 109-111, 175-177} Most of the reactions for carboxylic acid derivatization are based on condensation reaction with amines^{22, 50, 77-82} or esterification reaction.⁸³⁻⁸⁵ Our previously reported DmPA method has been used for profiling carboxyl submetabolome in metabolomics applications. However, based on our experience of using DmPA for metabolome profiling of various types of sample, we found that this method has two drawbacks. DmPA labeling was found to be optimally done in non-aqueous solution,

requiring an extraction step from an aqueous biological sample. DmPA labeling also produces a relatively higher background, compared to dansyl-based labeling methods for profiling amines, hydroxyls and carbonyls, potentially causing ion suppression of labeled analytes.

We have recently reported a differential ^{12}C -/ ^{13}C -DnsHz labeling method for analyzing carbonyl-containing metabolites and demonstrated very high performance of DnsHz labeling for profiling the carbonyl submetabolome.¹⁷² Encouraged by this development, we examined the possibility of using DnsHz for labeling carboxylic acids with a hope that the benefits of DnsHz labeling could be similarly expanded to carboxylic acid analysis. In principle, the hydrazide moiety in DnsHz can be linked with carboxylic acids using a coupling reagent. In this report, we present our study of developing DnsHz labeling for LC-MS profiling of the carboxyl submetabolome. The reaction conditions were optimized with labeling of several carboxylic acid standards covering different acid classes and structures. A robust workflow including a library of 193 DnsHz-labeled acid standards for positive metabolite identification was developed. The performance of this method for real biological samples was validated using human urine as an example of complex metabolomic systems.

4.2. Experimental Section

4.2.1. Chemicals and Urine Sample

The 193 metabolite standards (Table 4.1) were obtained from the Human Metabolome Database (HMDB).¹⁷ ^{12}C -DnsHz and ^{13}C -DnsHz are available from

mcid.chem.ualberta.ca. Other chemicals were purchased from Sigma-Aldrich (Edmonton, AB, Canada), except those specifically stated.

Urine samples were collected from a healthy volunteer after obtaining the informed consent. The ethics approval was acquired from the University of Alberta Ethics Approval Board. After collection, the urine samples were stored at 4 °C immediately or -80 °C for long term storage. The samples were centrifuged at 20817 g for 10 min at 4 °C. The supernatant was filter through a 0.22 µm filter, which was then aliquoted and stored in -80 °C for further use.

4.2.2. Preparation of Solution

The 193 standards were individually dissolved in water to a concentration of 10 mM. If solubility was too low, the supernatant of a saturated solution was used. The standard solutions were stored in -20 °C. Labeling reagents, ¹²C- or ¹³C-DnsHz, were dissolved in acetonitrile to 10 mg/mL. 2-(N-morpholino)ethanesulfonic acid (MES) and Na₂HPO₃/ NaH₂PO₃ were dissolved in water to generate an MES buffer or phosphate buffer solutions (PBS), respectively. The buffer solutions were used to prepare fresh coupling reagents solutions, including 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and other additives (1-hydroxy-7-azabenzotriazole, 4-dimethylaminopyridine, or N-hydroxysuccinimide; see Results and Discussion). Copper chloride (CuCl₂) was dissolved in water as a quenching reagent.

4.2.3. Chemical Isotope Labeling

For metabolite standard labeling, 20 µL of standard solution in water was placed in a centrifuge vial, followed by adding 20 µL of EDC and 20 µL of additives in buffer

solution. After mixing, 20 μL of 10 mg/mL ^{12}C -DnsHz was added. After vortexing and spinning down, the mixture was shaken at 20 $^{\circ}\text{C}$ for 90 min. Then 20 μL of 100 mM CuCl_2 solution was added to quench the labeling reagents and incubated at 40 $^{\circ}\text{C}$ for 30 min. The solution was centrifuged at 20817 g for 10 min and diluted before injecting into LC-MS for analysis.

To prepare the ^{12}C -/ ^{13}C -DnsHz labeled urine sample for method validation, 40 μL of urine was divided to two aliquots: 20 μL of urine was labeled by ^{12}C -DnsHz (light labeling) and the other 20 μL was labeled by ^{13}C -DnsHz (heavy labeling) using the same reaction conditions following procedures described above. After labeling, the ^{12}C -DnsHz labeled urine was mixed with the same volume of the ^{13}C -DnsHz labeled urine. The mixture was injected into LC-MS for analysis. The ^{12}C -/ ^{13}C - DnsHz labeled blank samples were prepared by labeling LC-MS grade water with the same protocol.

4.2.4. LC-MS Analysis

LC-MS and MS/MS experiments were done using Ultimate 3000 UHPLC system (Thermo Scientific, USA) linked to Impact HD Quadrupole Time-of-flight (Q-TOF) mass spectrometry (Bruker, Billerica, MA). A reversed phase column (Agilent Eclipse Plus C18 column, 2.1 mm \times 10 cm, 1.8 μm particle size, 95 \AA pore size) was used to separate labeled metabolites in a sample. Mobile phase A (MPA) was 0.1% (v/v) formic acid and 5% (v/v) ACN in water. Mobile phase B (MPB) was 0.1% (v/v) formic acid in ACN. The gradient for metabolites separation was: t = 0 min, 1% B; t = 3 min, 25% B; t = 23 min, 99% B; t = 34 min, 99% B. Between each run, 11 min washing was used to equilibrate the column. The flow rate was 180 $\mu\text{L}/\text{min}$. The column temperature was kept at 30 $^{\circ}\text{C}$. All MS spectra were collected in positive ion mode at a spectral acquisition rate of 1 Hz.

The MS/MS spectra of labeled standards were obtained with collision energy of 20-50 eV in stepping mode. The mass window width used in MS/MS experiments was 1 Da.

4.2.5. Data Processing

A set of R language-based programs developed in our group (freely available from www.mycompoundid.org) were used to process the CIL LC-MS data of labeled urine samples and labeled blank samples. In the first step, peak pairs which represent ^{12}C -/ ^{13}C -labeled metabolites were extracted from MS spectra by IsoMS⁴⁸. In this step, the program filtered out the redundant pairs (e.g., adduct ions, multimers, etc.) and noise or background signal (shown as single peak) and retained only a protonated ion of a peak pair for one metabolite. Thus, the number of peak pairs detected largely reflects the number of metabolites. IsoMS also calculated the intensity ratio of each peak pair, and aligned multiple pairs of the same individual metabolites from different runs according to retention time and accurate mass. In the next step, the missing values of intensity ratios in the aligned file were filled by the Zerofill program.¹⁴⁰ After that the chromatographic peak ratio of each peak pair was determined by IsoMS-Quant¹⁴¹. Finally the peak pairs that were also present in labeled blank samples with comparable intensities were removed from peak pair list, producing the final metabolite-intensity table for relative quantification. Positive identification of the labeled metabolites was done using the newly constructed DnsHz-labeled carboxyl standard library (see below). For putative identification, accurate masses of peak pairs were searched against the metabolite entries in metabolome databases by using MyCompoundID (MCID) at www.mycompoundid.org with zero-reaction and one-reaction libraries.¹⁹

4.3. Results and Discussion

4.3.1. Principle of DnsHz Labeling LC-MS Method

The method presented here was developed for profiling the carboxyl submetabolome and performing relative quantification of metabolites in comparative samples. A general workflow of applying this method for metabolomics study is shown in Figure 1. Accurate and precise relative quantification is realized by differential chemical isotope labeling. In the first step, after sample pre-treatment (e.g., filtration, protein precipitation, metabolite extraction, etc., depending on the sample type), individual samples from different groups are labeled by ^{12}C -DnsHz (light labeling), while a control sample or a pooled sample prepared by mixing aliquots of individual samples is labeled by ^{13}C -DnsHz (heavy labeling) using the optimized labeling protocol. In the next step, the same mole amount of ^{13}C -DnsHz labeled control which serves as an internal standard reference is spiked into each ^{12}C -labeled individual sample. The generated ^{12}C -/ ^{13}C -labeled mixtures are analyzed using LC-MS. Since a light labeled derivative and a heavy labeled derivative have nearly identical properties, they elute out at the same time. In mass spectra, the two derivatives are shown as a peak pair, which is easily distinguished from the singlet peaks originated from noise or unlabeled background chemicals. IsoMS software is used to extract peak pairs and calculate the peak area ratio between the ^{12}C -labeled and ^{13}C -labeled metabolites. Since the same amount of ^{13}C -labeled pool is added to all individual samples, relative quantification of a certain metabolite in different samples can be determined from the ratios. Metabolite identification is performed using a DnsHz-labeled carboxyl standard library (see below). Although absolute quantification is

not the focus of this work, CIL LC-MS method can also be used for this purpose by spiking a known concentration of labeled standard(s).

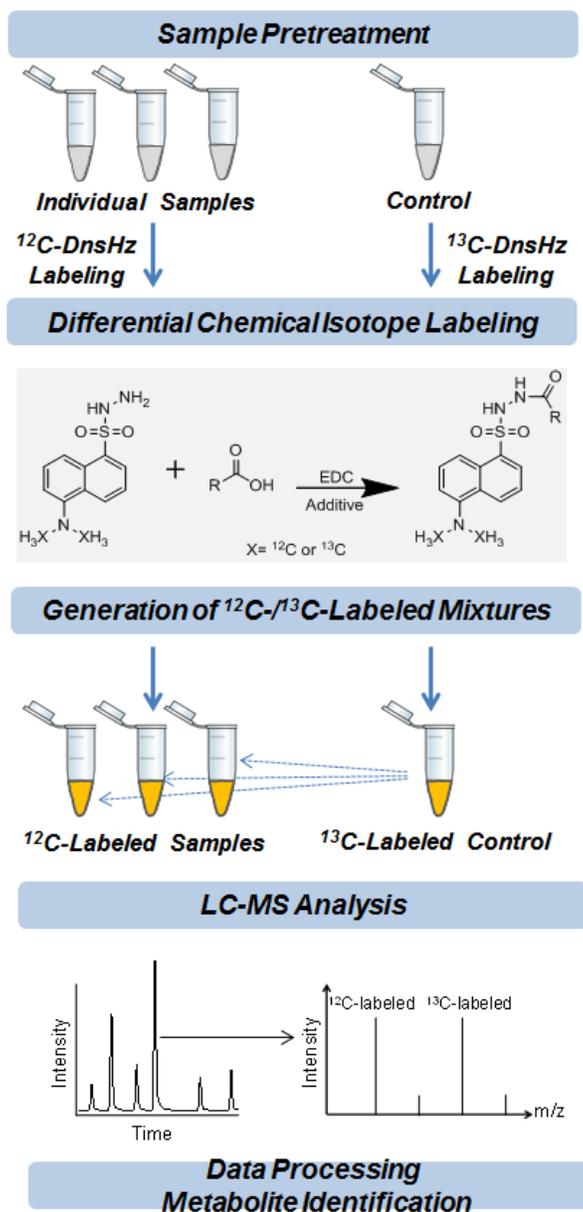


Figure 4.1 Workflow of differential CIL LC-MS method using ^{12}C -/ ^{13}C -DnsHz for carboxyl submetabolome analysis.

4.3.2. Optimization of Labeling Reaction

Carbodiimide-mediated coupling reaction is one of the most popular chemistries for crosslinking carboxyl group to amine group, which has been widely used for peptide synthesis¹⁷⁸, chemical probe design¹⁷⁹ and modification of biomolecules.¹⁸⁰ In this reaction, the carboxyl group present in a compound is activated by a condensation reagent (i.e., carbodiimide), followed by reacting with an amine group in a labeling reagent, resulting in a stable amide bond. 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) is chosen in our work as carbodiimide, because it is water-soluble and allows labeling reaction to be carried out in an aqueous solution, which is good for metabolomics since water is often present in biological samples. Direct labeling of a metabolome sample without removing water is more convenient. For the reaction mechanism,¹⁸¹ the carboxyl group is first activated by EDC to form an O-acylisourea intermediate that contains a better leaving group and can be displaced by nucleophilic attack from an amine group. Although the reaction can be carried out rapidly in mild condition, the hydrolysis and rearrangement of the intermediate limit the reaction efficiency and introduce side reactions. Thus, other additives such as N-hydroxysuccinimide (NHS) are employed to assist the reaction and reduce side reactions.¹⁸²

To optimize the labeling reaction for various carboxyl-containing metabolites in a metabolome sample, seven carboxylic acid standards from different categories (Figure 4.2) were mixed to generate a standard mixture, including propionic acid representing short-chain fatty acids, pyruvic acid representing keto-acids, benzoic acid representing aromatic acids, succinic acid representing polycarboxylic acids, cholic acid representing

bile acids, glucuronic acid representing sugar acids and hydroxyoctanoic acid representing hydroxyl acids. The standard mixture was labeled with ^{12}C -DnsHz under different reaction conditions. The same mixture was labeled with ^{13}C -DnsHz under one condition, as internal standards, and then spiked to ^{12}C -labeled samples with the same amount. The labeled samples were analyzed by LC-MS and the peak area ratio of each ^{12}C -/ ^{13}C -labeled metabolite was used to evaluate the relative efficiencies of different labeling conditions. Using this approach, we optimized the reaction buffer, reagents, time and temperature (Figure 4.3).

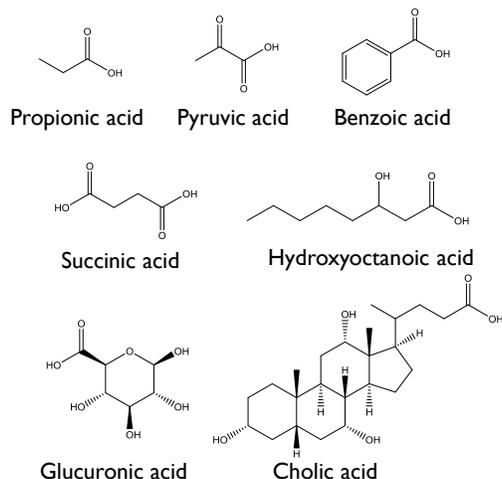


Figure 4.2 Structures of seven carboxyl-containing metabolites used to prepare a standards mixture for method development and evaluation.

Slightly acidic pH is required for the formation of O-acylisourea intermediate.¹⁸¹ Thus it is important to control the pH for carbodiimide-mediated coupling reaction. In this study, two types of buffer solution were tested with different pH values, including PBS with pH 3.5, 5.5, 7 and 8 and MES with pH 3.5 and 5.5. Figure 4.3A shows the effect of buffer solutions and pH on labeling efficiency. For all seven metabolites, reaction in MES solution behaved better than that in PBS. In the pH values tested, the

results of using PBS as buffer solution indicated that pH 3.5 and 5.5 provided higher labeling efficiency than pH 7 and 8, confirming that the reaction needs relatively acidic condition. The results of using MES buffer indicated that reactions of hydrophilic metabolites performed better at pH 3.5 (i.e., labeling glucuronic acid, succinic acid and pyruvic acid), while more hydrophobic ones were better at 5.5 (i.e., labeling propionic acid and cholic acid). Although there was no consistent trend for all seven metabolites, the labeling results of solution at pH 3.5 were found to be still acceptable even for those hydrophobic acids. Therefore, MES solution at pH 3.5 was considered to be an optimal buffer solution in our method.

We examined the effect of EDC concentration on the labeling reaction (Figure 4.3B). As expected, without EDC (0 mM), little if any labeled metabolites could be detected (except for glucuronic acid). Within the concentration range tested, 100 mM of EDC solution provided good labeling results for all seven metabolites and therefore was chosen as the optimal condition. In addition to carbodiimide, other additives can be used in coupling reaction to enhance labeling efficiency and reduce side reactions. In this work, we tested three additives: 1-hydroxy-7-azabenzotriazole (HOAT), 4-dimethylaminopyridine (DMAP) and N-hydroxysuccinimide (NHS). We found that, for all metabolites, the use of HOAT generated consistently the highest labeling efficiency (Figure 4.3C). We then studied the effect of HOAT concentration on the labeling reaction (Figure 4.3D). The results suggested that, for a majority of the tested standards, HOAT was necessary for the reaction. However, the concentration of HOAT did not affect the labeling efficiency dramatically, except for pyruvic acid. Taken all together, we chose 10 mM HOAT solution for the labeling reaction. However, it should be noted that, for

analyzing a special type of sample (e.g., cell extracts of small numbers of cells, sweat, etc.), it is best to fine-tune the reagent concentrations of both EDC and HOAT to produce an optimal workflow.

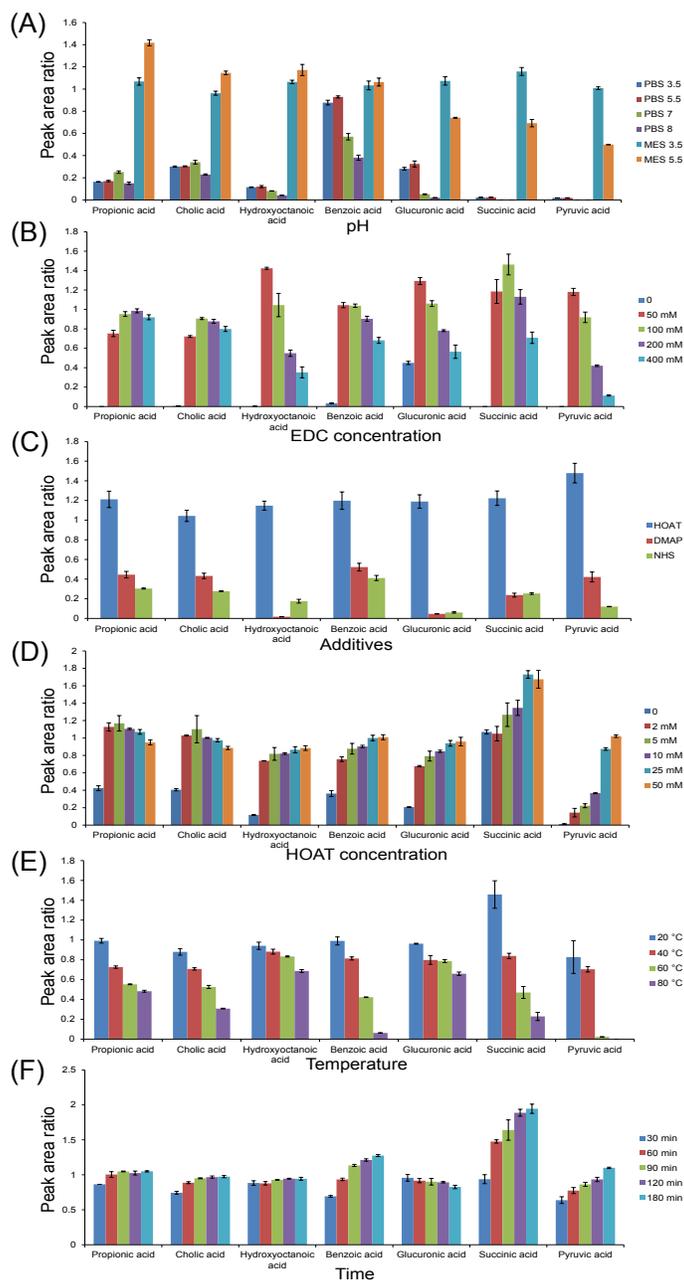


Figure 4.3 Comparison of efficiency for labeling a standards mixture under different reaction conditions: (A) effect of buffer solution and pH, (B) effect of EDC

concentration, (C) effect of additive type, (D) effect of HOAT concentration, (E) effect of reaction temperature, and (F) effect of reaction time. Data are presented as the mean \pm S.D. of six replicates (n=6).

The other two parameters affecting the reaction process were reaction temperature and time (Figures 4.3E and 4.3F). For temperature screening, we found that reaction at 20 °C was best for all the tested metabolites. Using this mild condition is beneficiary to protect thermally unstable metabolites and prevent loss of volatile acids. At 20 °C, reaction for 90 min was suitable for most of the standards. Although the amounts of labeled metabolites slightly increased after 90 min for some metabolites, considering both labeling efficiency and sample throughput, we chose 20 °C for 90 min as the optimal labeling conditions.

At last, a quench method was developed for DnsHz labeling of carboxylic acids in order to remove the remaining reagents that may interfere with the detection of labeled metabolites. DnsHz is a stable compound and can produce an intense signal in MS detection in positive ion mode. According to a recent report of using DnsHz labeling for developing a Cu^{2+} -selective fluorescence signaling probe, copper (II) was found to speed up the hydrolysis of DnsHz in water without affecting labeled compounds.¹⁸³ In our study, we found that adding copper (II) chloride could effectively remove the extra DnsHz in the labeling reaction solution. Figure 4.4A shows the LC-UV chromatograms of DnsHz and CuCl_2 mixtures prepared at different incubation times at 40 °C. It is clear that CuCl_2 could significantly reduce the DnsHz signal after incubating for 30 min. This result was confirmed using LC-MS. Figure 4.4B shows the LC-MS chromatograms of 20 mM DnsHz incubated with or without CuCl_2 . Without adding CuCl_2 , a broad, saturated

and tailed peak from DnsHz could be observed around 9 min, which may cause ion suppression for the detection of co-eluting labeled metabolites. However, this problem could be reduced by adding CuCl_2 and incubating at 40 °C for 30 min. Note that there are several additional peaks detected in LC-MS chromatograms. They were from background reactions which were largely suppressed in a real sample containing analytes (see, for example, labeled urine samples as shown below). These blank peaks could be filtered out in data processing.

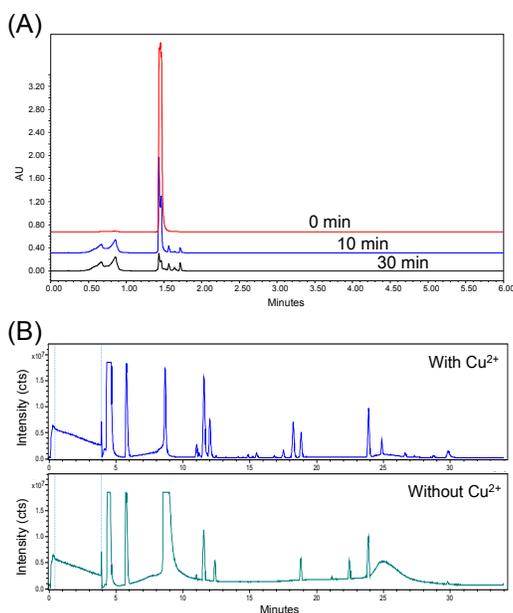


Figure 4.4 (A) LC-UV chromatograms of DnsHz and CuCl_2 mixture for different incubation times at 40 °C. (B) LC-MS chromatograms of DnsHz incubated with or without CuCl_2 .

4.3.3. Validation of Labeling Conditions with Urine Sample

Urine was chosen as a representative complex metabolomic system to validate whether the optimized method is applicable for real biological samples. LC-MS analysis of 1:1 mixture of ^{12}C -/ ^{13}C -labeled urine was used to determine and compare the number of peak pairs detected to evaluate the performance of the labeling conditions. For each

condition, data from six replicates were collected, including experimental triplicate of sample labeling and injection duplicate for each labeled sample.

Figure 4.5A shows the validation of the buffer solution selection. We examined the performance of using MES solution, PBS solution and water (no buffering) for labeling. For the MES buffer, combinations of two pH values (pH 3.5 and pH 5.5) and two concentrations (0.1 M and 0.4 M) were tested. For the PBS buffer, three pH values of 0.1 M solution were tested (pH 3.5, 5.5 and 7). The results confirmed the finding from the standard mixture experiments that the MES buffer solution performed better than the PBS buffer. However, we also found that the concentration and pH of the MES buffer solution had no significant effects on urine labeling, in terms of the number of peak pairs detected. Even without adding any buffer (the H₂O group), the peak pair number was similar to that of the MES buffer group. This can be attributed to the internal urinary buffering system, which keeps normal urine pH values ranging from 4.5 to 8. We tested the urine samples in this study and found that the pH value was about 5.0, which is in the optimal pH range for DnsHz carboxyl labeling. Thus, even without adding external buffer solutions, comparable results of labeling reaction were generated. However, for practical purpose of analyzing different urine samples (i.e., in case of encountering urine with higher pH in a metabolomics study), the use of 0.1 M MES buffer solution is recommended.

We examined the effects of reagent type and concentration for urine sample labeling. The results were in agreement with those of standard mixture experiments, i.e., 100 mM of EDC solution produced the highest peak pair numbers (Figure 4.5B). Among all the additives tested, adding HOAT into the reaction mixture increased the peak pair

numbers (Figure 4.5C). And within the concentrations we tested, the concentration of HOAT had little effect on the peak pair numbers (Figure 4.5D). Figures 4.5E and 4.5F show the results of reaction time and temperature validation, respectively. They were also in agreement with those of standard mixture labeling. Incubating at 20 °C for 1.5 h was therefore chosen as an optimized condition

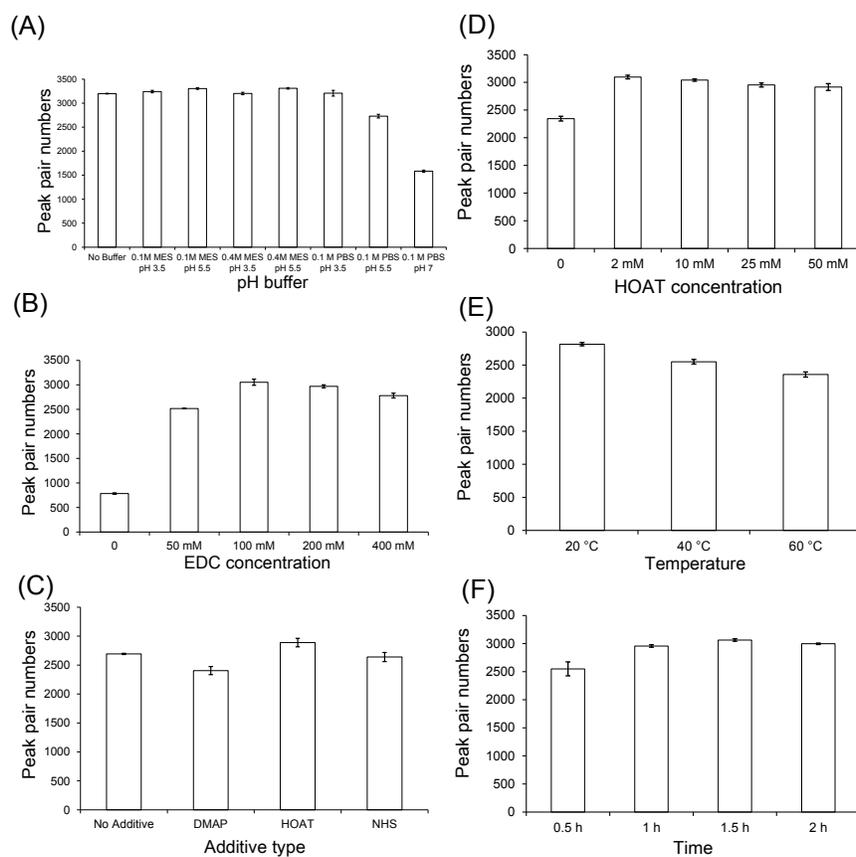


Figure 4.5 Peak pair numbers detected from ^{12}C -/ ^{13}C -DnsHz labeled urine samples prepared under different conditions: (A) effect of buffer solution and pH, (B) effect of EDC concentration, (C) effect of additives type, (D) effect of HOAT concentration, (E) effect of temperature and (F) effect of time. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=6).

Overall, based on the results of standard mixture and urine labeling experiments, it can be concluded that the optimal labeling condition is to use 100 mM of EDC as a coupling reagent and 10 mM of HOAT as the additive. Both reagents are dissolved in 0.1 M MES buffer solution at pH 3.5. Reaction mixtures are shaken at 20 °C for 90 min, followed by adding 100 mM of CuCl₂ as a quenching reagent and incubating at 40 °C for another 30 min. The entire labeling process can be completed in about 2 hr.

4.3.4. Metabolite Detectability

To demonstrate the detectability of the developed method for profiling complex samples, the urinary carboxyl submetabolomes from experimental triplicates of labeled urine samples were analyzed. Figure 4.6A shows a typical base-peak ion chromatogram of ¹²C-/¹³C-DanHz labeled human urine. Many peaks were detected within the whole gradient. We also optimized the sample injection amount in order to detect as many labeled metabolites as possible. To do this, different volumes of ¹²C-/¹³C-DanHz labeled urine were injected into LC-MS for analysis and the peak pair numbers were then compared. Figure 4.6B shows the number of peak pairs as a function of injection volume. A plateau was observed at 8 μL injection. We note that, for different instrumental setup with different detection sensitivity or for analyzing other types of samples, a similar strategy can be applied to find the optimized injection amount.

Venn diagram was used to evaluate the labeling reproducibility among experimental triplicates of sample processing (Figure 4.6C). An average of 2597±28 pairs per run were detected with 2266 pairs in common, indicating high reproducibility of peak pair detection. The distribution of absolute intensities of these common peak pairs is presented in Figure 4.6D (in black). As the peak intensity decreased, more peak pairs

could be detected, which suggests that the developed method could effectively detect metabolites with very low abundance or low signal intensity, leading to high metabolome coverage.

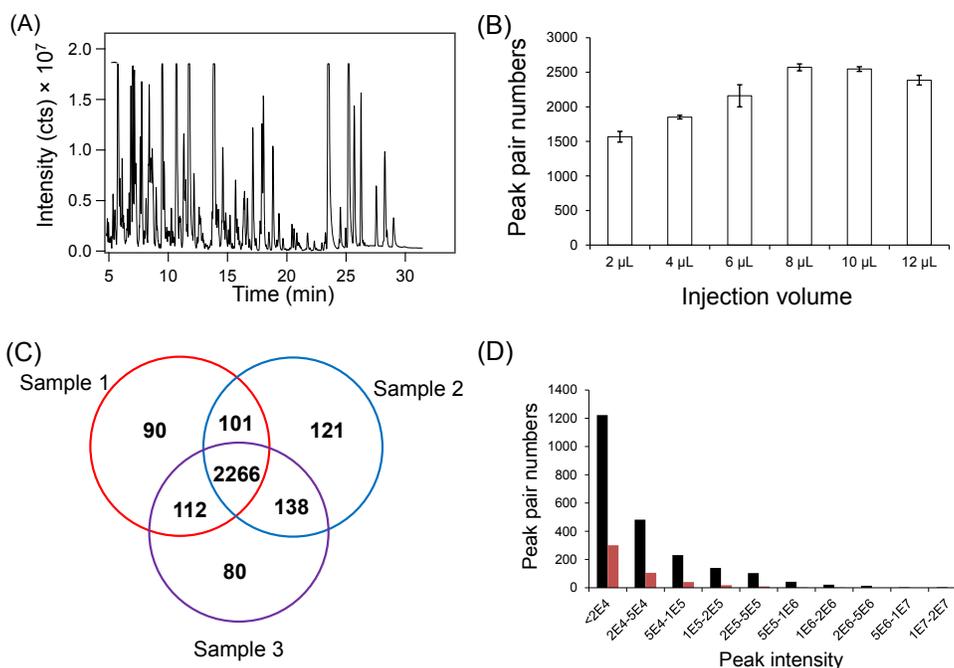


Figure 4.6 (A) LC-MS ion chromatogram of $^{12}\text{C}/^{13}\text{C}$ -DnsHz labeled urine. (B) Peak pair number detected as a function of injection amount of $^{12}\text{C}/^{13}\text{C}$ -DnsHz labeled urine. Data are presented as mean \pm S.D. ($n=3$). (C) Venn diagrams of the peak pair numbers detected in triplicate labeled urine. (D) Peak pair number detected as a function of peak intensity (black: all peak pairs; red: peak pairs with peak ratio >1.25 or <0.8).

4.3.5. Quantification

In CIL LC-MS method, relative quantification of metabolites is realized by measuring peak ratio between the ^{12}C -labeled metabolite from individual samples and its ^{13}C -labeled counterpart in the pooled sample. To evaluate the overall quantification

ability of the workflow for profiling the carboxyl submetabolome, accuracy and precision of six replicate analyses (sample processing triplicates and instrumental injection duplicates) of 1:1 ^{12}C -/ ^{13}C -DanHz labeled urine samples were examined. Figure 4.7A-B shows the distribution of peak pair numbers as a function of the averaged measured peak area ratio and the RSD of the ratio (n=6). Out of the 2266 peak pairs detected in all runs, the ratios of most pairs were close to the expected value of 1.0. 94.6% of the pairs had peak ratio values between 0.67 and 1.5, showing high accuracy of the method. The average RSD of peak ratios was 5.07%, while 98.8% of peak pairs had RSD values of less than 20% and 93.6% of peak pairs had RSD of less than 10%, demonstrating that very reproducible peak ratios could be generated.

We noticed that there were some peak pairs with ratios of not close to the expected value of 1.0. For example, out of the 2266 common peak pairs, 123 pairs had peak ratios that exceeded the $\pm 50\%$ accuracy range (i.e., peak ratio < 0.67 or > 1.5). We plotted the peak ratios of all pairs against their absolute signal intensities (Figure 4.7C). To make the plot clearer, logarithm transformation was used to convert the peak ratios and intensity values. Most of the pairs with large ratio errors were the low intensity peak pairs [i.e., absolute intensity of less than 2×10^5 or $\log(\text{peak intensity}) < 5.3$]. By plotting the distributions of the absolute intensities of peak pairs that have ratios of larger than 1.25 or smaller than 0.8 (Figure 4.6D, in red), we can observe more clearly that low intensity pairs had higher percentage of values outside the $\pm 20\%$ range. This might be caused by poor peak shape of low intensity peak, which made peak area calculation more difficult, resulting in larger errors in ratio calculation. This observation suggests that, for low intensity peaks, great care should be exercised in relative quantification. This is

actually true even for targeted analysis of metabolites; quantification limit is often defined as the lowest signal at $S/N=10$, while detection limit is defined at $S/N=3$. Future work of investigating the relation between signal intensity and quantification accuracy for different ratios of ^{12}C -labeled sample vs. ^{13}C -labeled pool is warranted. Nevertheless, the above results already indicate that good accuracy and precision could be obtained for most of the peak pairs.

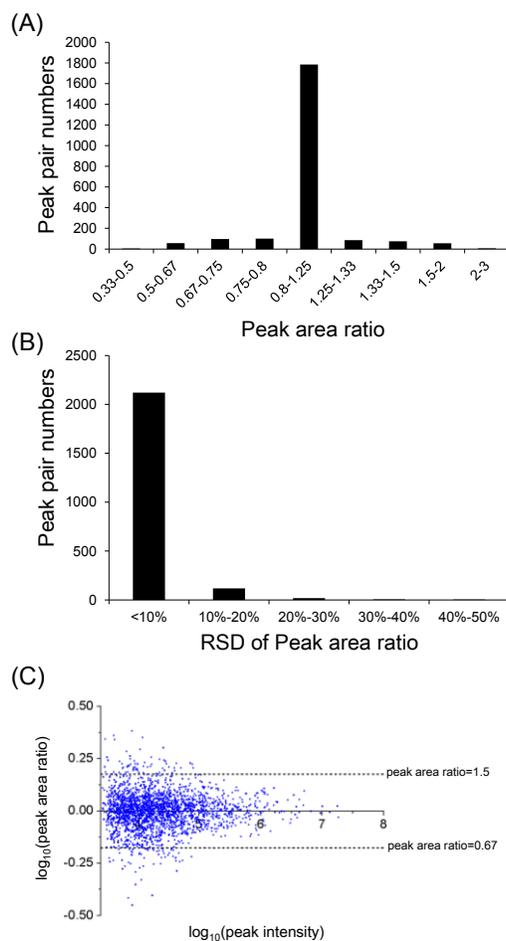


Figure 4.7 Distributions of peak pair numbers as a function of (A) averaged peak ratio and (B) RSD of peak ratio. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate ($n=6$). (C) Peak ratios of all pairs as a function of their absolute signal intensities.

4.3.6. Stability of the Labeled Samples

The stability of labeled sample was tested for both short term storage and long term storage. For short term stability test, one ^{12}C -/ ^{13}C -DanHz labeled urine sample was placed in LC autosampler at 4 °C for 24 hr. The sample was injected every 3 hr for LC-MS analysis. The peak pair numbers of each time point were compared (Figure 4.8A). It shows that very similar peak pair numbers were detected within 24 hr, indicating good short term stability of labeled samples. In addition, long term stability test was performed to demonstrate whether the method can be used for analyzing large batches of samples. One ^{12}C -/ ^{13}C -DnsHz labeled urine mixture was aliquoted and stored in different conditions for different periods of time, including just after labeling, 4 °C for overnight, -20 °C for three days and -80 °C for a month. The results are shown in Figure 4.8B. It shows that there was no significant difference of the peak pair numbers detected among these storage conditions. These results indicated that the labeled samples were very stable and could meet the needs of long term studies.

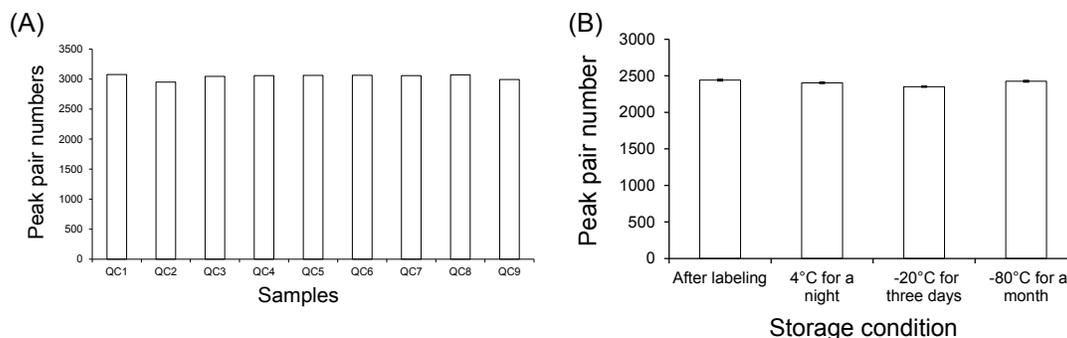


Figure 4.8 (A) Short term stability test of labeled urine samples. (B) Long term stability test of labeled urine samples. Data are presented as mean \pm S.D. from experimental triplicate, injection duplicate (n=6).

4.3.7. Metabolite Identification

Metabolite identification is another important aspect in metabolome analysis. To perform high confidence metabolite identification, usually at least two independent properties of unknown metabolites from samples should be matched to those of authentic standards.¹⁸⁴ We constructed a DnsHz-labeled carboxyl standard library which currently contains 193 human endogenous carboxyl-containing metabolites. These metabolites cover 55 metabolic pathways. Each standard was labeled with DnsHz individually and then analyzed by LC-MS and LC-MS/MS. Triple-parameters information for each standard was collected: accurate mass, retention time (RT) and MS/MS fragmentation spectrum (Table 4.1). To perform identification, the measured mass, RT and/or MS/MS of an unknown labeled metabolite from a sample could be used to search against the library using a website-based program DnsHz-ID.^{59, 172} Both of the library and search program are freely available from a publicly accessible website at www.mycompoundid.org. Accurate mass and MS/MS matches can be performed directly after uploading the corresponding information of unknown samples to the website. However, RT is easily affected by many possible variations in LC-MS conditions, such as instrument or column brand, connection tubing length, etc. To use RT as an identification parameter, a RT normalization method reported previously⁵⁹ was used to correct the RT shift caused by these variations. A set of DnsHz labeled compounds, which was initially developed for identification of DnsHz-labeled carbonyl-containing metabolites¹⁷², was used as RT calibration mixture since the same LC gradient was used for the analysis of the two submetabolomes. The RTs of 193 standards collected for the library were all normalized to this RT calibration mixture. When performing metabolome analysis using the

presented method in a different instrument, the same RT calibration mixture needs to be run on the new instrumental setup. Then the RTs of unknown metabolites will be corrected automatically using the DnsHz-ID program based on the local RT calibration mixture, followed by identification using RT.

Table 4.1 List of 193 compounds in the current labeled DnsHz-carboxyl standard library.

HMDB	Metabolites	RT (min)	HMDB	Metabolites	RT (min)
HMDB00005	2-Ketobutyric acid	11.08	HMDB00744	Malic acid	7.06
HMDB00008	2-Hydroxybutyric acid	8.77	HMDB00746	Hydroxyisocaproic acid	11.85
HMDB00020	p-Hydroxyphenylacetic acid	9.64	HMDB00748	L-3-Phenyllactic acid	12.55
HMDB00039	Butyric acid	10.98	HMDB00749	Mesaconic acid	12.06
HMDB00042	Acetic acid	8.11	HMDB00750	m-Hydroxymandelic acid	8.60
HMDB00060	Acetoacetic acid	8.65	HMDB00752	Methylglutaric acid	8.76
HMDB00094	Citric acid	6.74	HMDB00754	3-Hydroxyisovaleric acid	8.65
HMDB00112	gamma-Aminobutyric acid	6.07	HMDB00755	Hydroxyphenyllactic acid	9.41
HMDB00118	Homovanillic acid	9.87	HMDB00761	Lithocholic acid	23.25
HMDB00121	Folic acid	7.15	HMDB00763	5-Hydroxyindoleacetic acid	8.96
HMDB00123	Glycine	10.62	HMDB00764	Hydrocinnamic acid	14.54
HMDB00128	Guanidineacetic acid	6.26	HMDB00779	Phenyllactic acid	12.55
HMDB00130	Homogentisic acid	7.76	HMDB00784	Azelaic acid	11.32
HMDB00142	Formic acid	8.50	HMDB00792	Sebacic acid	12.48
HMDB00158	Tyrosine	11.60	HMDB00807	3-Phosphoglyceric acid	5.47
HMDB00159	Phenylalanine	12.59	HMDB00842	Quinaldic acid	21.08
HMDB00161	Alanine	10.88	HMDB00858	Monomethyl glutaric acid	10.38
HMDB00162	Proline	7.29	HMDB00883	Valine	11.64
HMDB00167	Threonine	10.60	HMDB00888	Undecanedioic acid	13.63
HMDB00168	Asparagine	10.33	HMDB00892	Valeric acid	12.81
HMDB00172	Isoleucine	12.16	HMDB00929	Tryptophan	9.32
HMDB00176	Maleic acid	5.10	HMDB00930	trans-Cinnamic acid	14.42
HMDB00177	Histidine	5.58	HMDB00933	Traumatic acid	14.23
HMDB00182	Lysine	5.16	HMDB00943	Threonic acid	6.24
HMDB00187	Serine	10.40	HMDB00946	Ursodeoxycholic acid	16.42
HMDB00190	Lactic acid	4.75	HMDB00954	Ferulic acid	6.77
HMDB00191	Aspartic acid	5.95	HMDB00955	3-Hydroxy-4-methoxycinnamic	11.39

				acid	
HMDB00193	Isocitric acid	6.29	HMDB01101	p-Anisic acid	13.05
HMDB00197	Indole-3-acetic acid	12.72	HMDB01259	Succinic acid semialdehyde	8.11
HMDB00202	Methylmalonic acid	8.14	HMDB01336	3,4-Dihydroxyphenylacetic acid	6.19
HMDB00208	Oxoglutaric acid	6.16	HMDB01370	Diaminopimelic acid	5.39
HMDB00209	Phenylacetic acid	13.3	HMDB01392	p-Aminobenzoic acid	9.87
HMDB00210	Pantothenic acid	7.87	HMDB01451	Lipoic acid	15.97
HMDB00223	Oxaloacetic acid	4.88	HMDB01470	Tiglic acid	11.93
HMDB00227	Mevalonic acid	7.32	HMDB01544	m-Chlorobenzoic acid	14.94
HMDB00237	Propionic acid	9.39	HMDB01587	Phenylglyoxylic acid	12.84
HMDB00243	Pyruvic acid	10.02	HMDB01624	2-Hydroxylcaproic acid	12.09
HMDB00254	Succinic acid	7.71	HMDB01713	m-Coumaric acid	11.54
HMDB00263	Phosphoenolpyruvic acid	5.70	HMDB01844	Methylsuccinic acid	8.47
HMDB00267	Pyroglutamic acid	7.20	HMDB01856	Protocatechuic acid	6.05
HMDB00308	3b-Hydroxyl-5-cholenoic acid	20.43	HMDB01864	2-Ketohexanoic acid	13.78
HMDB00329	2-Phenylbutyric acid	15.97	HMDB01870	Benzoic acid	12.81
HMDB00355	3-Hydroxyl-3-methylglutaric acid	7.70	HMDB01873	Isobutyric acid	10.84
HMDB00357	3-Hydroxylbutyric acid	7.71	HMDB01874	threo-Isocitric acid	6.74
HMDB00392	2-Octenoic acid	17.01	HMDB01877	Valproic acid	16.91
HMDB00393	3-Hexenedioic acid	8.41	HMDB01879	Acetylsalicylic acid	12.62
HMDB00402	2-Isopropylmalic acid	9.66	HMDB01891	3-Aminobenzoic acid	9.20
HMDB00407	2-Hydroxyl-3-methylbutyric acid	10.33	HMDB01895	Salicylic acid	13.58
HMDB00415	3a,6b,7b-TriHydroxyl-5b-cholanoic acid	14.54	HMDB01901	Aminocaproic acid	6.35
HMDB00422	2-Methylglutaric acid	8.62	HMDB01923	Naproxen	16.73
HMDB00423	3,4-DiHydroxylhydrocinnamic acid	6.33	HMDB01925	Ibuprofen	19.85
HMDB00426	Citramalic acid	7.65	HMDB01955	3-Phenylbutyric acid	15.44
HMDB00434	3,4-Dimethoxyphenylacetic acid	11.56	HMDB01975	2-Ethyl-2-Hydroxylbutyric acid	11.13
HMDB00440	3-Hydroxylphenylacetic acid	10.09	HMDB01987	2-Hydroxyl-2-methylbutyric acid	9.71
HMDB00444	3-Furoic acid	11.08	HMDB01988	4-Hydroxylcyclohexylcarboxylic acid	8.00
HMDB00448	Adipic acid	8.36	HMDB02001	Dimethylmalonic acid	11.46
HMDB00452	2-Aminobutyric acid	11.29	HMDB02024	Imidazole-4-acetic acid	6.05
HMDB00467	Nutriacholic acid	17.75	HMDB02035	p-Coumaric acid	11.08
HMDB00484	Vanillic acid	6.33	HMDB02043	5-Phenylvaleric acid	16.71
HMDB00491	3-methyl-2-oxo-Valeric acid	15.24	HMDB02059	12-Hydroxyldodecanoic acid	15.36

HMDB00500	4-Hydroxybenzoic acid	10.14	HMDB02072	4-Methoxyphenylacetic acid	13.08
HMDB00502	3-Oxocholeic acid	16.50	HMDB02074	2,2-Dimethylsuccinic acid	9.75
HMDB00503	7a-Hydroxy-3-oxo-5b-choleic acid	19.95	HMDB02085	Syringic acid	6.55
HMDB00505	Allocholeic acid	16.07	HMDB02092	Itaconic acid	8.50
HMDB00511	Decanoic acid	20.64	HMDB02097	4-Ethylbenzoic acid	15.63
HMDB00517	Arginine	5.57	HMDB02107	Phthalic acid	6.07
HMDB00518	Chenodeoxycholeic acid	19.39	HMDB02176	Ethylmethylacetic acid	12.38
HMDB00529	5-Dodecenoic acid	21.96	HMDB02199	3-(4-Hydroxyphenyl)propionic acid	10.77
HMDB00543	4-Phenylbutyric acid	15.61	HMDB02222	3-Methylphenylacetic acid	14.72
HMDB00555	3-Methyladipic acid	9.33	HMDB02229	3-Phenoxypropionic acid	14.40
HMDB00565	Galactonic acid	5.97	HMDB02243	Picolinic acid	20.67
HMDB00574	Cysteine	7.18	HMDB02329	Oxalic acid	5.93
HMDB00576	Ethyl Malonate	10.77	HMDB02359	Phenylpropionic acid	15.82
HMDB00617	Pyromucic Acid	10.65	HMDB02364	Oleanolic acid	25.94
HMDB00619	Cholic acid	16.21	HMDB02390	3-Cresotinic acid	22.31
HMDB00620	Glutaconic acid	7.92	HMDB02428	Terephthalic acid	10.21
HMDB00622	Ethylmalonic acid	9.22	HMDB02466	3-Hydroxybenzoic acid	10.45
HMDB00623	Dodecanedioic acid	14.79	HMDB03070	Shikimic acid	6.43
HMDB00639	Mucic acid	6.06	HMDB03164	Chlorogenic acid	9.23
HMDB00641	Glutamine	5.90	HMDB03355	5-Aminopentanoic acid	6.19
HMDB00661	Glutaric acid	7.90	HMDB03402	Pectin	6.05
HMDB00665	Leucinic acid	11.84	HMDB04110	Phosphonoacetate	5.52
HMDB00666	Heptanoic acid	16.23	HMDB04586	Perillic acid	16.94
HMDB00669	Hydroxyphenylacetic acid	11.65	HMDB04812	2,5-Furandicarboxylic acid	8.11
HMDB00687	Leucine	12.28	HMDB04815	4-Hydroxy-3-methylbenzoic acid	6.45
HMDB00689	Isocaproic acid	14.33	HMDB05807	Gallic acid	6.04
HMDB00691	Malonic acid	7.39	HMDB06331	cis,cis-Muonic acid	8.79
HMDB00694	L-2-Hydroxyglutaric acid	7.17	HMDB11743	2-Phenylpropionic acid	14.65
HMDB00700	Hydroxypropionic acid	7.06	HMDB11753	Iminodiacetic acid	11.24
HMDB00703	Mandelic acid	11.24	HMDB14118	Trifluoroacetic acid	8.65
HMDB00711	Hydroxyoctanoic acid	15.44	HMDB29649	2,4,6-Trihydroxybenzoic acid	11.27
HMDB00714	Hippuric acid	10.65	HMDB31331	Chloroacetic acid	10.28
HMDB00718	Isovaleric acid	12.57	HMDB41604	3-Mercaptopropanoic acid	5.88

HMDB00720	Levulinic acid	8.45	HMDB59916	Tartaric acid	6.41
HMDB00729	Alpha-Hydroxyisobutyric acid	8.36	HMDB59969	3-Methoxyphenylacetate	13.32
HMDB00733	Hyodeoxycholic acid	16.23	HMDB60003	Isovanillic acid	6.39
			HMDB60665	Isonicotinic acid	8.60

Putative identification based on the use of accurate mass match alone to the existing metabolite database is less confident. However, it can still provide valuable identity information of unknown metabolites and hints for further confirmation, such as purchasing or synthesizing standards. Therefore, in this study, we used a two-tier identification approach with positive identification in tier 1 and putative identification in tier 2.

In tier 1, among the common 2266 peak pairs detected in six replicates of labeled urine samples, we positively identified 81 peak pairs based on accurate mass and retention time matches using the constructed labeled metabolite library (Table 4.2). In tier 2, the remaining peak pairs were searched, based on accurate mass match, against the MyCompoundID (MCID) library composed of 8,021 known human endogenous metabolites (zero-reaction library) and their predicted metabolic products from one metabolic reaction (375,809 compounds) (one-reaction library). 517 and 1445 peak pairs were matched in the zero- and one-reaction libraries, respectively. Thus, out of 2266 unique peak pairs detected, 2043 pairs (90.2%) could be positively identified or putatively matched.

It is not surprising that we observed some peak pairs were matched to carbonyl-containing metabolites in putative identification. As reported before, ^{12}C -/ ^{13}C -DnsHz can be used for carbonyl submetabolome profiling using different reaction conditions (40 °C for 60 minutes) and reagents (hydrogen chloride).¹⁷² Although the reaction buffers for

two submetabolomes are different, both of the reactions are carried out in acidic condition. Therefore, it is possible to label some carbonyl-containing metabolites when profiling carboxyl submetabolome using this method. However, achieving the highest selectivity is not the main goal of the CIL LC-MS approach for high-coverage metabolome profiling, as eventually we will combine data from different submetabolomes to generate the entire metabolome results. On the other hand, it also provides an opportunity to combine the two reaction products for a combined LC-MS analysis of the carbonyl and carboxyl submetabolomes which should increase the throughput of sample analysis. We will report a detailed study of the comparison of carbonyl and carboxyl submetabolomes of various types of samples, as well as the comparison of individual submetabolome analysis vs. combined analysis in the future.

Table 4.2 List of carboxylic acid metabolites identified based on accurate mass (10 ppm) and retention time (30 s) matches to the DnsHz labeled carboxyl standard library.

Peak pair #	Peak pair information			Identification result				
	Corrected RT (s)	mz_light	mz_heavy	mz	Name	Accurate mass	mz_light	Library RT (s)
129	5.52	434.0787	436.0852	185.9935	3-Phosphoglyceric acid	185.9929	434.0781	5.47
212	5.92	416.1275	418.1342	168.0423	3,4-Dihydroxyphenylacetic acid	168.0423	416.1275	6.19
					Vanillic acid	168.0423	416.1275	6.33
					Isovanillic acid	168.0423	416.1275	6.39
224	5.96	338.0805	340.0871	89.9953	Oxalic acid	89.9953	338.0805	5.93
267	6.08	442.1281	444.1348	194.0429	Pectin	194.0427	442.1279	6.05
272	6.08	394.1544	396.1612	146.0692	Glutamine	146.0691	394.1543	5.90
278	6.09	458.1228	460.1295	210.0376	Mucic acid	210.0376	458.1228	6.06
341	6.24	394.1067	396.1133	146.0215	Oxoglutaric acid	146.0215	394.1067	6.16
376	6.31	384.1221	386.1288	136.0369	Threonic acid	136.0372	384.1224	6.24
382	6.32	365.1389	367.1455	117.0537	Guanidineacetic acid	117.0538	365.1390	6.26
424	6.42	379.1797	381.1862	131.0945	Aminocaproic acid	131.0946	379.1798	6.35
458	6.50	430.1425	432.1496	182.0573	3,4-DiHydroxyhydrocinnamic acid	182.0579	430.1431	6.33
516	6.64	365.1640	367.1707	117.0788	5-Aminopentanoic acid	117.0790	365.1642	6.19
624	6.90	440.1122	442.1188	192.0270	Citric acid	192.0270	440.1122	6.74
					threo-Isocitric acid	192.0270	440.1122	6.74
724	7.18	338.1169	340.1236	90.0317	Hydroxypropionic acid	90.0317	338.1169	7.06
750	7.24	396.1601	398.1655	148.0749	Mevalonic acid	148.0736	396.1588	7.32
781	7.34	377.1276	379.1345	129.0424	Pyroglutamic acid	129.0426	377.1278	7.20
					Glutamic acid[-H ₂ O]	147.0532	377.1278	7.20
862	7.53	352.0962	354.1027	104.0110	Malonic acid	104.0110	352.0962	7.39
968	7.83	410.1379	412.1456	162.0527	3-Hydroxyl-3-methylglutaric acid	162.0528	410.1380	7.70
987	7.87	352.1326	354.1392	104.0474	3-Hydroxylbutyric acid	104.0473	352.1325	7.71

1007	7.94	467.1961	469.2026	219.1109	Pantothenic acid	219.1107	467.1959	7.87
1045	8.02	394.1432	396.1497	146.0580	Adipic acid	146.0579	394.1431	8.36
1088	8.11	392.1638	394.1704	144.0786	4-Hydroxycyclohexylcarboxylic acid	144.0786	392.1638	8.00
1129	8.20	416.1276	418.1342	168.0424	Homogentisic acid	168.0423	416.1275	7.76
					m-Hydroxylmandelic acid	168.0423	416.1275	8.6
1150	8.23	350.1169	352.1240	102.0317	Acetoacetic acid	102.0317	350.1169	8.65
					Succinic acid semialdehyde	102.0317	350.1169	8.11
1157	8.26	366.1120	368.1185	118.0268	Methylmalonic acid	118.0266	366.1118	8.14
1159	8.26	308.1066	310.1132	60.0214	Acetic acid	60.0211	308.1063	8.11
1258	8.49	352.1325	354.1391	104.0473	Alpha-Hydroxyisobutyric acid	104.0473	352.1325	8.36
					2-Hydroxybutyric acid	104.0473	352.1325	8.77
1259	8.49	394.1422	396.1494	146.0570	2-Methylglutaric acid	146.0579	394.1431	8.62
					Methylglutaric acid	146.0579	394.1431	8.76
1298	8.63	294.0909	296.0976	46.0057	Formic acid	46.0055	294.0907	8.50
1343	8.77	366.1482	368.1548	118.0630	3-Hydroxyisovaleric acid	118.0630	366.1482	8.65
1463	9.02	430.1431	432.1497	182.0579	Hydroxyphenyllactic acid	182.0579	430.1431	9.41
1677	9.53	322.1218	324.1285	74.0366	Propionic acid	74.0368	322.1220	9.39
1732	9.73	394.1430	396.1499	146.0578	2,2-Dimethylsuccinic acid	146.0579	394.1431	9.75
1769	9.83	366.1480	368.1547	118.0628	2-Hydroxy-2-methylbutyric acid	118.0630	366.1482	9.71
1774	9.84	400.1324	402.1390	152.0472	p-Hydroxyphenylacetic acid	152.0473	400.1325	9.64
1827	9.98	430.1430	432.1496	182.0578	Homovanillic acid	182.0579	430.1431	9.87
1838	10.01	348.1011	350.1077	100.0159	Fumaric acid[-O]	116.0110	348.1012	9.80
1889	10.11	336.1016	338.1086	88.0164	Pyruvic acid	88.0160	336.1012	10.02
1936	10.22	400.1321	402.1387	152.0469	3-Hydroxyphenylacetic acid	152.0473	400.1325	10.09
2032	10.42	342.0674	344.0722	93.9822	Chloroacetic acid	93.9822	342.0674	10.28
2141	10.67	362.1167	364.1234	114.0315	Glutaric acid[-H2O]	132.0423	362.1169	10.24
2178	10.78	427.1437	429.1502	179.0585	Hippuric acid	179.0582	427.1434	10.65
2184	10.83	323.1172	325.1238	75.0320	Glycine	75.0320	323.1172	10.62
2225	10.93	380.1271	382.1342	132.0419	Ethyl Malonate	132.0423	380.1275	10.77
2275	11.02	468.6546	470.6616	220.5694	Folic acid- 2 tags 2 charges	441.1397	468.6550	10.64

2294	11.06	414.1482	416.1547	166.0630	3-(4-Hydroxylphenyl)propionic acid	166.0630	414.1482	10.77
2326	11.13	336.1375	338.1443	88.0523	Butyric acid	88.0524	336.1376	10.98
					Isobutyric acid	88.0524	336.1376	10.84
2374	11.22	418.1066	420.1134	170.0214	2,4,6-Trihydroxylbenzoic acid	170.0215	418.1067	11.27
2451	11.36	414.1480	416.1546	166.0628	m-Coumaric acid[+H2]	164.0473	414.1481	11.22
2471	11.41	362.1167	364.1235	114.0315	Methylsuccinic acid[-H2O]	132.0423	362.1169	11.25
2501	11.45	436.1900	438.1966	188.1048	Azelaic acid	188.1049	436.1901	11.32
2692	11.95	380.1631	382.1702	132.0779	Hydroxylisocaproic acid	132.0786	380.1638	11.85
					2-Hydroxylcaproic acid	132.0786	380.1638	12.09
					Leucinic acid	132.0786	380.1638	11.84
2933	12.56	406.1430	408.1498	158.0578	2-Isopropylmalic acid[-H2O]	176.0685	406.1431	12.19
2952	12.61	450.2059	452.2126	202.1207	Sebacic acid	202.1205	450.2057	12.48
2968	12.66	400.1325	402.1394	152.0473	p-Anisic acid	152.0473	400.1325	13.05
3037	12.86	423.1485	425.1551	175.0633	Indole-3-acetic acid	175.0633	423.1485	12.72
3048	12.88	414.1472	416.1545	166.0620	Phenyllactic acid	166.0630	414.1482	12.55
					4-Methoxyphenylacetic acid	166.0630	414.1482	13.08
					L-3-Phenyllactic acid	166.0630	414.1482	12.55
3063	12.92	370.1214	372.1283	122.0362	Benzoic acid	122.0368	370.1220	12.81
3091	12.98	350.1533	352.1599	102.0681	Valeric acid	102.0681	350.1533	12.81
					Isovaleric acid	102.0681	350.1533	12.57
3204	13.25	414.1477	416.1546	166.0625	3-Methoxyphenylacetate	166.0630	414.1482	13.32
3261	13.39	321.0962	323.1029	146.0221	Oxoglutaric acid- 2 tags 2 charges	146.0215	321.0960	13.25
3289	13.43	384.1375	386.1443	136.0523	Phenylacetic acid	136.0524	384.1376	13.30
3321	13.53	314.1067	316.1127	132.0430	Glutaric acid- 2 tags 2 charges	132.0423	314.1063	13.41
3365	13.65	307.0988	309.1055	118.0272	Succinic acid- 2 tags 2 charges	118.0266	307.0985	13.53
3371	13.68	300.0911	302.0977	104.0118	Malonic acid- 2 tags 2 charges	104.0110	300.0907	13.58
3381	13.71	378.1485	380.1551	260.1266	2-Ketohexanoic acid	130.0630	378.1482	13.78
3394	13.74	464.2214	466.2282	216.1362	Undecanedioic acid	216.1362	464.2214	13.63
3497	14.11	476.2214	478.2280	228.1362	Traumatic acid	228.1362	476.2214	14.23
3628	14.49	656.3730	658.3796	408.2878	3a,6b,7b-TriHydroxyl-5b-cholanoic	408.2876	656.3728	14.54

acid								
3656	14.55	414.1481	416.1549	166.0629	3-Phenoxypropionic acid	166.0630	414.1482	14.40
3701	14.67	364.1690	366.1756	116.0838	Isocaproic acid	116.0837	364.1689	14.33
3758	14.79	478.2373	480.2440	230.1521	Dodecanedioic acid	230.1518	478.2370	14.79
3787	14.86	326.0884	328.0951	156.0065	2,5-Furandicarboxylic acid- 2 tags 2 charges	156.0059	326.0881	15.34
3928	15.26	378.1483	380.1548	130.0631	3-methyl-2-oxo-Valeric acid	130.0630	378.1482	15.24
4005	15.52	408.1952	410.2019	160.1100	Hydroxyoctanoic acid	160.1099	408.1951	15.44
4224	16.30	378.1848	380.1915	130.0996	Heptanoic acid	130.0994	378.1846	16.23
4225	16.30	656.3734	658.3798	408.2882	Cholic acid	408.2876	656.3728	16.21
					Allocholic acid	408.2876	656.3728	16.07
4293	16.60	654.3564	656.3636	406.2712	3-Oxochoolic acid	406.2719	654.3571	16.50
4414	17.37	414.1847	416.1912	166.0995	Perillic acid	166.0994	414.1846	16.94
4456	17.65	306.1093	308.1161	116.0481	Levulinic acid- 2 tags 2 charges	116.0473	306.1089	17.54
4561	18.40	362.1534	364.1605	228.1364	Traumatic acid- 2 tags 2 charges	228.1362	362.1533	18.07

4.4. Conclusions

We have developed a high performance CIL LC-MS method to analyze the carboxyl submetabolome. $^{12}\text{C}/^{13}\text{C}$ -DnsHz reagents were used for differential isotope labeling. The reported method is ready to be incorporated in a multichannel CIL LC-MS platform where different chemical-based-submetabolomes are profiled to generate a comprehensive metabolome profile. The comparisons of this method with other labeling methods will be carried out in the future, as well as the application of using this method for various metabolomics applications.

Chapter 5

Multichannel Chemical Isotope Labeling Mass Spectrometry for High-Coverage Quantitative Metabolomics

5.1. Introduction

As the development of systems biology, “omics” studies become more and more important to understand biological processes and to discover mystery behind disease and life. Among all the “omics” approaches, metabolomics, the comprehensive and simultaneous analysis of many small molecules in biological specimens (i.e. metabolome), plays an increasingly important role in disease biomarker discovery and biological studies.¹⁸⁵ Compared to genome, transcriptome and proteome, metabolome is the final downstream product of biological processes. It is also the direct reflection of organism to the environmental exposures, such as diet¹¹ and contaminants.¹² Thereby it is closest to the phenotype of a biological system. However high-coverage metabolome analysis is an analytical challenge due to great diversity of physicochemical properties of metabolites and wide ranges of metabolite concentrations in biological samples

Traditional approach of increasing coverage is to combine lists of metabolites detected by several complementary analytical methods based on nuclear magnetic resonance (NMR) and/or mass spectrometry (MS).^{147, 186} For example, combined using reversed phase liquid chromatography (RPLC) for separating hydrophobic metabolites and hydrophilic interaction liquid chromatography (HILIC) for separating more polar compounds is employed. The detection of metabolites using MS is conducted in both positive ion mode and negative ion mode. This traditional approach requires heavy

investment in equipment and a large amount of samples, and has low sample throughput, particularly for accurate metabolite quantification. Even with this time-consuming approach, accurate quantification of thousands of metabolites at the same time is still difficult.

To overcome these shortcomings, a high-performance multichannel chemical isotope labeling liquid chromatography mass spectrometry (mCIL-LC-MS) technique has been developed, which is different from the traditional approach. In mCIL-LC-MS, the whole metabolome is divided into several submetabolomes based on chemical functional groups of the metabolites. The in-depth analysis of each submetabolome is performed and the combined datasets from the submetabolomes would allow the analysis of the entire metabolome. To analyze chemical-group-based submetabolome, a number of chemical derivatization methods, in combination with LC-MS, have been reported.^{68, 94, 108, 133, 138,}

151

Previously our group has developed several high-performance CIL LC-MS methods for different submetabolomes, including dansylation for analyzing amines/phenols,³⁰ base-activated dansylation for hydroxyls,³¹ DmPA bromide labeling for carboxylic acids,⁸⁵ and dansylhydrazine (DnsHz) labeling for carbonyl metabolites.¹⁷² Here we present a 4-channel chemical labeling approach by combined using the CIL LC-MS methods listed above to achieve very high coverage of metabolites for metabolomics. The rationale of this approach is based on the analysis of chemical group diversity of several common metabolite databases. Using human plasma as an example, the analytical performance of this approach is demonstrated.

5.2. Experimental Section

5.2.1. Chemical Group Diversity Analysis

The workflow of chemical group diversity analysis contains six steps, shown in Figure 5.1. First, four selected metabolite libraries, i.e. MyCompoundID (<http://www.mycompoundid.org/>), Human Metabolome Database (HMDB), Yeast Metabolome Database (YMDB) and E. coli Metabolome Database (ECMDB) were downloaded from corresponding website. Compound information, including compound name, database ID, structure, origin and compound class, was extracted for further analysis. Then substructure patterns were constructed based on SMILES structure to recognize different chemical functional groups. To avoid human-induced error and to increase throughput, we wrote SubstrcMatch, a Java based program that can match compound structure to different substructure patterns. SubstrcMatch used compound SMILES structure file and all functional group patterns as input, and generated a .txt file to indicate what functional groups the compound has. The next step was to remove unrelated compounds. Compounds containing hydrocarbon chain longer than seven were filtered as lipids. For HMDB, based on origin and class information we further filtered the non-endogenous compounds (e.g. synthesized drugs) and metal related compounds. Finally, the chemical group diversity of one library was summarized using remaining “true” metabolites.

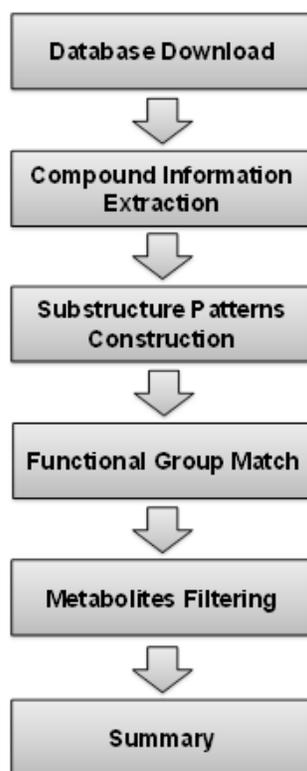


Figure 5.1 Workflow of chemical group diversity analysis

5.2.2. Principle and Workflow of CIL LC-MS Methods

The principle of this approach is based on differential chemical isotope labeling for relative quantification using LC-MS. As an example, Figure 5.2 shows the workflow for analyzing the plasma metabolome which includes the following steps: 1) protein precipitation using three volumes of methanol, 2) applying four isotope labeling chemistries targeting different submetabolomes, 3) LC-UV quantification of dansyl-labeled metabolites for sample amount normalization, 4) mixing of equal moles of ^{12}C -labeled samples and ^{13}C -labeled control or pooled sample, 5) high-resolution reversed-phase (RP) LC-MS analysis of ^{12}C -/ ^{13}C -mixtures, 6) data processing including peak pair peaking and peak ratio measurement, 7) statistical analysis, and 8) metabolite identification based on the use of labeled standards library for positive identification and

the use of other compound libraries for putative identification. The details of experimental procedures are described below.

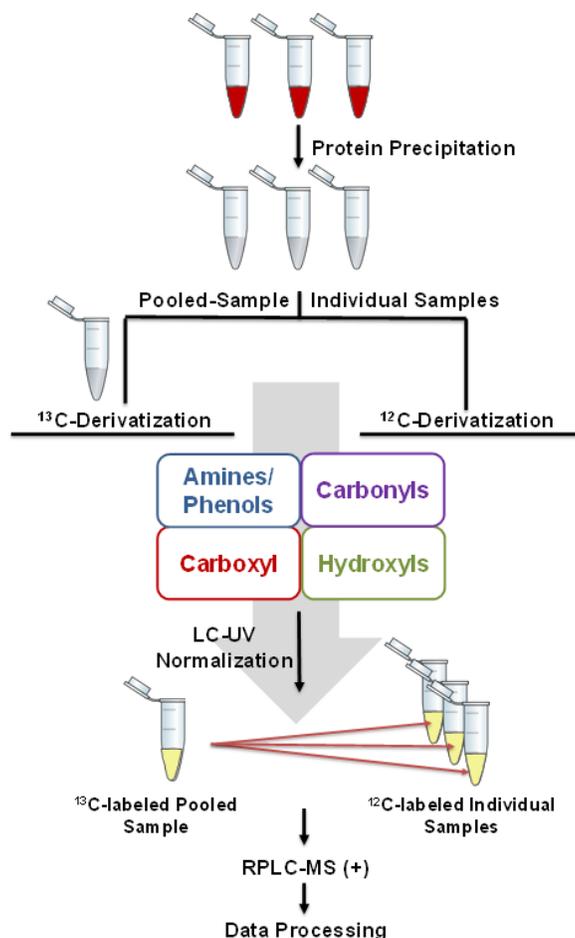


Figure 5.2 Workflow of 4-channel CIL LC-MS for in-depth relative quantification of the plasma metabolome.

Proteins in the plasma samples were first precipitated using three volumes of methanol, followed by incubating at $-20\text{ }^{\circ}\text{C}$ for 30 min, drying down and re-dissolving in the same volume of water. A control sample, i.e., the pooled-samples, prepared by mixing small aliquots of individual samples is labeled by ^{13}C -reagent. This ^{13}C -labeled control is the internal standard and spiked into all ^{12}C -labeled individual samples according to the

LC-UV quantification result, followed by LC-MS analysis. Relative quantification can be done using the peak ratio of the individual peak pairs. For profiling different submetabolomes, four chemical labeling reactions are applied: dansylation for amines/phenols (Figure 5.3A), DmPA bromide labeling for carboxylic acids (Figure 5.3B), base-activated dansylation for hydroxyls (Figure 5.3C) and dansylhydrazine (DnsHz) labeling for carbonyl metabolites (Figure 5.3D). After LC-MS analysis, a set of programs developed in our group are used for data processing and a labeled standard library constructed by our group is used for positive metabolite identification.

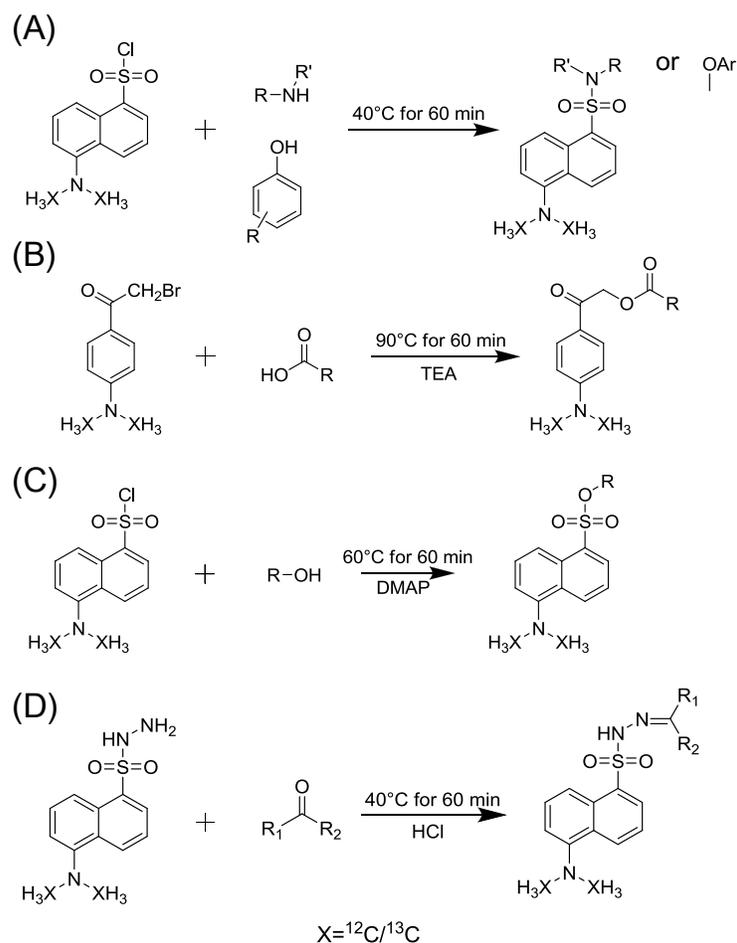


Figure 5.3 Reaction schemes of (A) dansylation labeling for amine/phenol-containing metabolites; (B) DmPA bromide labeling for carboxylic acid-containing metabolites; (C)

base-activated dansylation labeling for hydroxyl-containing metabolites; (D) dansylhydrazine labeling for carbonyl-containing metabolites.

5.2.3. Amine/Phenol Submetabolome Profiling

25 μL of sample was mixed with 12.5 μL of 250 mM $\text{Na}_2\text{CO}_3/\text{NaHCO}_3$ buffer, 12.5 μL of ACN and 25 μL of 18 mg/mL ^{12}C -dansyl chloride (DnsCl) in ACN (for light labeling) or 18 mg/mL ^{13}C -dansyl chloride in ACN (for heavy labeling). After vortexing and spinning down, the mixture was incubated at 40 $^\circ\text{C}$ for 60 min, 5 μL of 250 mM sodium hydroxide solution was used to quench the reaction at 40 $^\circ\text{C}$ for 10 min. Then 25 μL of 425 mM formic acid (in 50/50 ACN/ H_2O) was added to consume the excess NaOH. The combined solution was centrifuged for 10 min at 10000 rpm before injecting into LC-UV for quantification or LC-MS for analysis.

5.2.4. Carboxyl Submetabolome Profiling

Liquid-liquid extraction was applied to enrich carboxyl-containing metabolites in plasma. 60 μL of plasma sample was mixed with 10 μL of saturated NaCl solution, 5 μL of 6 M HCl and 180 μL of ethyl acetate. After vortexing and spinning down, the organic phase was transferred to a new vial. 20 μL of trimethylamine solution (180 mg/mL in ACN) was added. Then the solution was dried down and re-dissolved in 60 μL of 20 mg/mL TEA solution (in ACN). 30 μL of the solution was mixed with 30 μL ^{12}C -DMPA (20 mg/mL in ACN) for light labeling or ^{13}C -DMPA (20 mg/mL in ACN) for heavy labeling. The vial was again vortexed and spun down, followed by incubating in at 80 $^\circ\text{C}$ for 60 min. Then the solution was centrifuged at 10000 rpm for 10 min before injecting into LC-MS for analysis.

5.2.5. Hydroxyl Submetabolome Profiling

To label hydroxyl-containing metabolites in plasma, a method containing liquid-liquid extraction and base-activated dansylation was applied. After protein precipitation, 60 μL of plasma sample was mixed with 10 μL of saturated NaCl solution and 5 μL of 6 M HCl solution. Then extract metabolites with 180 μL of ethyl acetate twice. The organic phase were then combined, dried down and then re-dissolved in 60 μL of ACN. For light labeling, 25 μL of the extracted sample was mixed with 25 μL of 4-dimethylaminopyridine (DMAP) (24.5 mg/mL in ACN) and 25 μL of ^{12}C -DnsCl (18 mg/mL in ACN). For heavy labeling ^{13}C -DnsCl (18 mg/mL in ACN) was used instead of ^{12}C -DnsCl. After incubated at 60 $^{\circ}\text{C}$ for 60 min, 5 μL of 250 mM sodium hydroxide solution was used to quench the reaction. The solution was incubated at 60 $^{\circ}\text{C}$ for 10 min after vortexed and spun down. Then 25 μL of 425 mM formic acid solution in 50/50 ACN/water was added to consume the excess NaOH. The combined solution was centrifuged for 10 min at 10000 rpm before injecting into LC-UV for quantification or LC-MS for analysis.

5.2.6. Carbonyl Submetabolome Profiling

20 μL of plasma sample was mixed with 20 μL of 144 mM HCl solution (diluted with ACN) and 20 μL of 20 mM ^{12}C -DnsHz in ACN (for light labeling) or 20 mM ^{13}C -DnsHz in ACN (for heavy labeling). After vortexing and spinning down, the solution was incubated at 40 $^{\circ}\text{C}$ for 60 min. Then the mixture was cooled down in -80 $^{\circ}\text{C}$ freezer for 5 min to stop the labeling reaction followed by drying down and re-dissolving in 80 μL of acetonitrile/water mixture (ACN/H₂O, 50:50, v/v). The solution was centrifuged at 10000 rpm for 10 min before injecting into LC-MS for analysis.

5.2.7. LC-UV Quantification and Normalization

Waters ACQUITY UPLC system combined with a photodiode array (PDA) detector was used for measuring the total concentration of labeled metabolites. 5 μ L of labeled sample was injected into a C18 column (Phenomenex Kinetex C18, 2.1 mm \times 5 cm, 1.7 μ m particle size, 100 Å pore size). Mobile phase A was 0.1% (v/v) formic acid and 5% (v/v) ACN in water. Mobile phase B was 0.1% (v/v) formic acid in ACN. A fast gradient was applied to elute all labeled metabolites together (flow rate: 0.45 mL/min; 0 min, 0% B; 1 min 0% B; 1.01 min, 95% B; 2.5 min, 95% B; 3.0 min, 0% B; 6.0 min 0% B). The UV detector was operated at 338 nm.

5.2.8. LC-MS Analysis

The labeled ^{12}C -/ ^{13}C -mixture was analyzed using a Bruker Impact HD Quadrupole Time-of-flight (QTOF) mass spectrometer (Bruker, Billerica, MA) linked to UltiMate 3000 UHPLC (Thermo Scientific, MA). An RP column (Agilent Eclipse Plus C18 column, 2.1 mm \times 10 cm, 1.8 μ m particle size, 95 Å pore size) was used to separate labeled metabolites. Mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile. Mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The gradient for amine/phenol submetabolome profiling was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 98% B; t = 34 min, 98% B. The gradient for hydroxyl submetabolome profiling was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 9.2 min, 65% B; t = 21.2 min, 98% B; t = 31.2 min, 98% B. The gradient for carbonyl submetabolome profiling was: t = 0 min, 1% B; t = 3 min, 25% B; t = 23 min, 98% B; t = 34 min, 98% B. The gradient for carboxyl submetabolome profiling was: t = 0 min, 20% B; t = 9 min, 50% B; t = 22 min, 65% B; t = 26 min, 80% B; t = 29 min, 98% B; t = 40 min, 98% B. The flow rate was 180

$\mu\text{L}/\text{min}$. The column temperature was kept at 30 °C. All MS spectra were collected in positive ion mode at a spectral acquisition rate of 1 Hz.

5.2.9. Data Processing

A set of R language-based programs developed in our group (freely available from www.mycompoundid.org) were used to process CIL LC-MS data. In the first step, peak pairs which represent $^{12}\text{C}/^{13}\text{C}$ -labeled metabolites were extracted from MS spectra by IsoMS.⁴⁸ In this step program filtered out the redundant pairs (e.g., those of adduct ions and dimers) and noise signal (shown as single peak), only retaining a protonated ion of a peak pair for one true metabolite. IsoMS also calculated the intensity ratio of each peak pair, and aligned multiple pairs of the same individual metabolites from different runs according to retention time and accurate mass. The second step was filling the missing values of intensity ratios in the aligned file by the Zerofill program.¹⁴⁰ Finally IsoMS-Quant program determine the chromatographic peak ratio of each peak pair and generate the final metabolite-intensity table for further analysis.¹⁴¹ Positive identification of the labeled metabolites was done using the labeled standard library (www.mycompoundid.org).⁵⁹ For putative identification, accurate masses of peak pairs were searched against metabolite entries in metabolome databases, such as Human Metabolome Database (HMDB),¹⁷ MyCompoundID (MCID, www.mycompoundid.org)¹⁹ zero-reaction and one-reaction libraries, etc.

5.3. Results and Discussion

5.3.1. Chemical Group Diversity Analysis

The chemical group classifications of compounds in several metabolome databases were investigated. The results showed that most of the known metabolites belong to one or more of these groups. For example, in the MyCompoundID library,¹⁹ there are 8004 compounds derived from the original Human Metabolome Database (HMDB)¹²⁷ comprised of mainly the known human endogenous metabolites and lipids. After removing the lipids and inorganic compounds from the list of 8004 compounds, there are 2683 metabolites remaining which can be classified according to their chemical groups. About 95% of these metabolites contain one or more of the four functional groups (Figure 5.4A-B). Similarly for the *E. Coli* Metabolome Database (ECMDB),²¹ among 1462 non-lipid metabolites (filtered from 3760 entries in the library), 1401 metabolites (95.8%) can be potentially analyzed using 4-channel CIL-LC-MS approach (Figure 5.4C-D). For the Yeast Metabolome Database (YMDB),²⁰ there are 1107 remaining metabolites after removing all lipids from 2004 entries. Among them, 949 metabolites (85.7%) belong to the four groups (Figure 5.4E-F). The recent version of HMDB (v4.0)¹⁷ has a list of 74,462 compounds and, among them, 8715 compounds can be considered to be non-lipid human metabolites and 86% of them contain one or more of the four functional groups (Figure 5.4G-H). Thus, if we can perform in-depth analysis of the four submetabolomes separately, the combined datasets from the submetabolomes would allow the analysis of the entire metabolome with very high coverage.

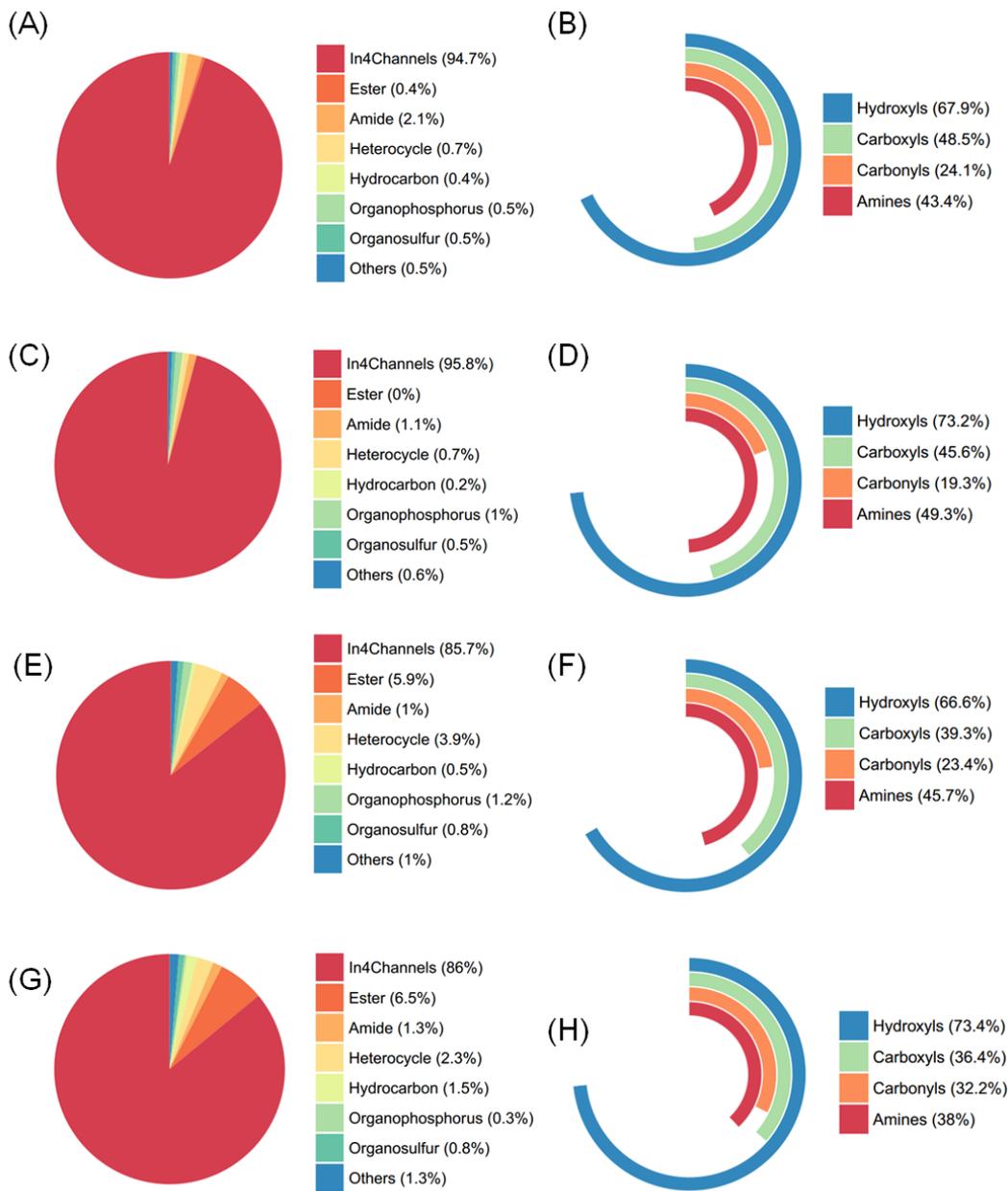


Figure 5.4 Classification of chemical groups of (A) MyCompoundID library, (C)ECMDB, (E)YMDB and (G) HMDB; Percent distributions of metabolites belonging to the four groups that are analyzed using the 4-channel CIL-LC-MS approach in (B) MyCompoundID library, (D) ECMDB, (F) YMDB and (H) HMDB.

5.3.2. Benefits of high-performance CIL LC-MS methods for metabolomics

For biomarker discovery and cellular metabolomics, accurate measurement of metabolite concentration changes in comparative samples (i.e., relative quantification) is required. Using a stable isotope labeled standard (SIL) as a reference for analyte quantification is commonly used in LC-MS to overcome matrix effects, ion suppression effects and instrument sensitivity drifts.¹⁸⁷ Because of lack of SILs for all metabolites, in our approach, we use differential isotope labeling to create references for individual metabolites for accurate relative quantification. A pooled sample is first produced by mixing small aliquots of individual samples and then labeled using a heavy-reagent (e.g., ^{13}C -form of a labeling reagent). Another aliquot is taken from an individual sample and labeled using a light-reagent (e.g., ^{12}C -reagent). Each ^{12}C -labeled sample is spiked with the same mole amount of the ^{13}C -labeled pool. This mixture is subjected to LC-MS analysis and the resultant mass spectra contain metabolite peaks in pairs with mass difference of a peak pair corresponding to the mass difference of the heavy and light labeling reagents multiplying the number of labeling tags attached to the metabolite. Only the differentially labeled metabolites from the samples are detected in pairs and background chemicals are detected as singlet peaks, facilitating the identification of true metabolite peaks and removal of redundant peaks such as those of adduct ions, intra- and inter-molecular dimer and multimers. We have developed the IsoMS software to pick up the peak pairs and calculate the peak intensity ratio of ^{12}C -labeled peak vs. ^{13}C -labeled pool for each peak pair.⁴⁸ Because the same amount of ^{13}C -pool is added to each ^{12}C -sample, the peak ratio values of a labeled metabolite determined from LC-MS analysis of ^{12}C -/ ^{13}C -mixtures measures the relative concentrations of the metabolite in all samples.

These ratio values can be uploaded to statistical tools for further analysis such as discovering metabolites with significant concentration changes in comparative samples.

Another key consideration in metabolomic analysis is to detect as many metabolites as possible. Metabolite detectability can be altered significantly using rationally designed labeling chemistries. Four labeling chemistries have been separately developed with each targeting the analysis of a submetabolome. These labeling chemistries improve analyte separation by changing the hydrophobicity of unlabeled metabolites to an extent that all the labeled metabolites are retained on a reversed phase (RP) column and thus switching columns of different types is no longer needed. They also enhance the analyte ionization significantly in positive ion mode, resulting in highly sensitive detection of even low abundance metabolites in a sample. As a consequence of applying these labeling chemistries, the four submetabolomes can be analyzed using only one experimental setup: RPLC-MS with positive ion detection in a high-resolution instrument. The use of a simple instrumental setup is particularly important for non-expert laboratories.

5.3.3. Metabolite detection

To demonstrate the performance of the combined analysis of the four submetabolomes, human plasma samples were labeled using the four chemistries in experimental triplicates. In this work, a plasma sample was divided into two aliquots. One was ^{12}C -labeled and the other was ^{13}C -labeled, followed by mixing and LC-MS analysis. The number of peak pairs or metabolites detected in each submetabolome was determined and their intensity ratios were measured. Figure 5.5 shows the ion chromatograms of the four submetabolome analyses. Many peaks across the entire

separation window in RPLC were detected, showing both the chemical diversity and enhanced detectability. These labeling chemistries allow the enhancement of ionization efficiency from ~ 10 to ~ 1000 folds. One example is shown in Figure 5.6 where 22 dansyl labeled metabolites with different structures were detected with more similar efficiencies, compared to their unlabeled counterparts; several of the unlabeled metabolites cannot even retain on RP column due to their low hydrophobicity. Thus, applying the four labeling chemistries has the net effect of equalizing the ionization efficiencies of individual metabolites to a large extent, despite of their great chemical and physical differences for the unlabeled metabolites.

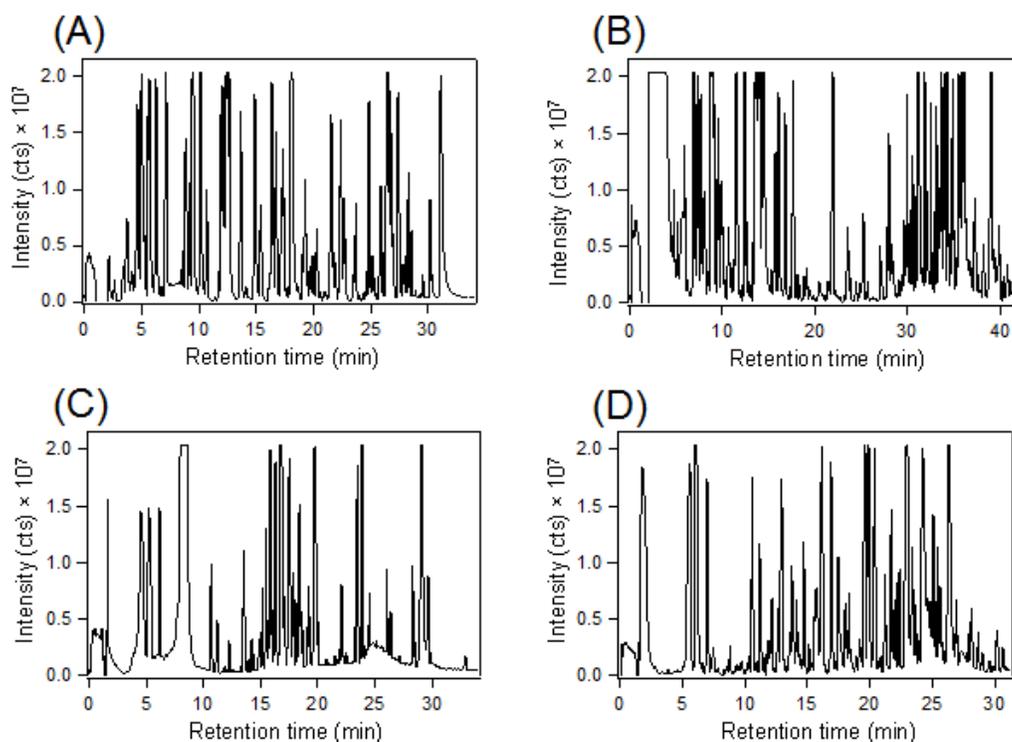


Figure 5.5 Base peak ion chromatograms of CIL LC-MS method profiling (A) Amine/phenol submetabolome; (B) Carboxylic acid submetabolome; (C) Carbonyl submetabolome; (D) Hydroxyl submetabolome.

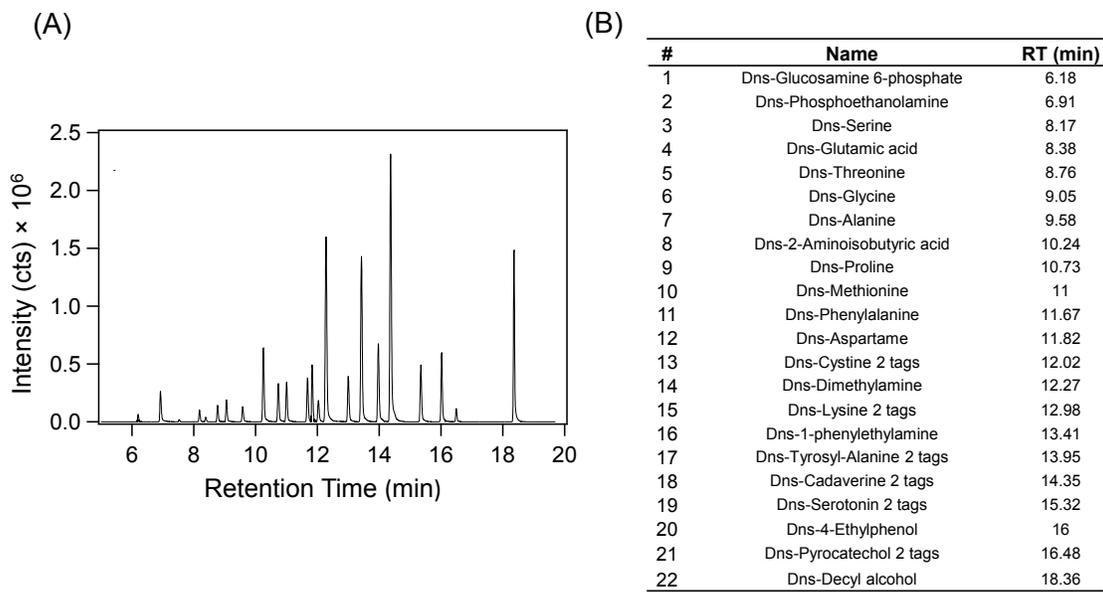


Figure 5.6 (A) Ion chromatograms of 22 dansyl labeled metabolites with different structures. (B) List of 22 dansyl labeled metabolites.

By plotting the number of peak pairs detected in LC-MS as a function of injection volumes or amounts measured by LC-UV (Figure 5.7), the optimal injection amount and the maximum number of peak pairs detectable can be determined. For amine/phenol submetabolome analysis, 1937 ± 4 ($n=3$) peak pairs could be detected when $4 \mu\text{L}$ of labeled sample was injected. For carboxylic acid submetabolome, 2293 ± 5 ($n=3$) peak pairs could be detected when $3 \mu\text{L}$ of sample was injected. For aldehyde/ketone submetabolome, 1660 ± 5 ($n=3$) peak pairs could be detected with $8 \mu\text{L}$ injection. For alcoholic hydroxyl submetabolome, 1428 ± 6 ($n=3$) peak pairs could be detected. The distributions of the absolute intensities of these peak pairs are shown in Figure 5.8A. Within the dynamic range of the instrument, there is a clear trend of increasing numbers of peak pairs detectable as the peak intensity decreases. Thus, using highly sensitive methods can increase the metabolome coverage more rapidly by detecting an increasing

number of lower concentration metabolites. Interestingly, the four submetabolomes have similar distributions, indicating un-biased detection by the four methods as well as similar concentration distributions of these different groups of metabolites in plasma.

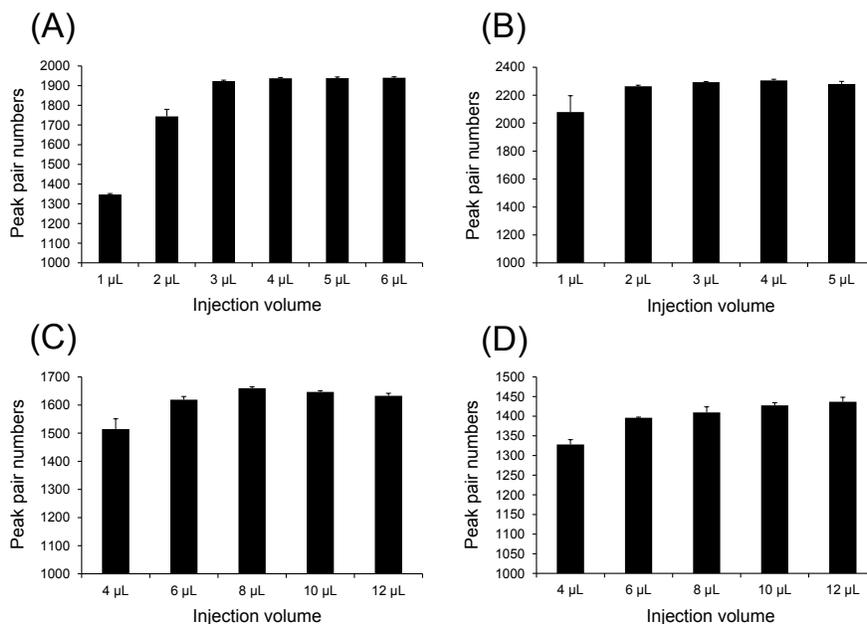


Figure 5.7 Peak pair number detected as a function of injection amount of ^{12}C -/ ^{13}C -labeled plasma of (A) Amine/phenol submetabolome; (B) Carboxylic acid submetabolome; (C) Carbonyl submetabolome; (D) Hydroxyl submetabolome. Data are presented as mean \pm S.D. (n=3).

To further compare the numbers of metabolites detected in the four methods, Figure 5.8B shows the Venn diagram of the number of peak pairs detected in each method. Because of lack of structure identities for many of the metabolites (see below), in this comparison, we assumed that the same metabolite was detected in two methods if the same accurate mass of the intact metabolite (i.e., mass of a labeled metabolite minus the mass of labeling tag(s)) was found in the two methods. There are 1452, 1634, 1198, 916 unique peak pairs detected in amine/phenol, carboxyl, carbonyl, and hydroxyl

submetabolome, respectively. Out of a combined total of 7431 peak pairs detected from the four methods, there are 6206 unique peak pairs remaining after filtering out the overlap peak pairs based on mass matches.

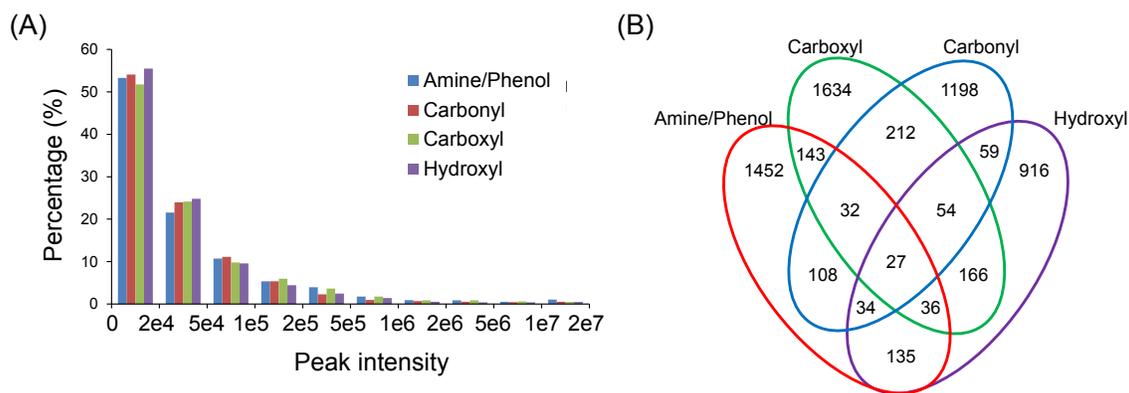


Figure 5.8 (A) Percentage of peak pair detected in four submetabolome profiling methods as a function of peak intensity. (B) Venn diagram of the numbers of peak pairs or metabolites detected in four methods.

5.3.4. Metabolite identification

To identify the labeled metabolites, we use a two-tier approach with positive identification in tier 1 and putative identification or structure match in tier 2. For positive identification, we have constructed a labeled standards library consisting of 664 unique human endogenous metabolites, including 315 amines/phenols, 187 carboxyls, 85 hydroxyls and 77 carbonyls. Each labeled standard was produced and analyzed separately by RPLC-MS and MS/MS measurements to generate triple parameters: retention time, molecular mass and MS/MS spectrum. This library is freely available at www.mycompoundid.org and can be used to identify a metabolite in a sample by using accurate mass and retention time matches and/or MS/MS match against this library. For the plasma sample, we positively identified 223 peak pairs based on accurate mass and

retention time matches (Table 4.1). From the identification result, we found that a few of peak pairs could be matched to one metabolite (e.g., Carbonyl_594, Carbonyl_635 and Carbonyl_657 matched to butanal). These matches have been manually checked from ion chromatogram. These are likely structural isomers of one chemical formula. It is clear that expanding the standard compound library is urgently needed to increase the number of positively identified metabolites. However, in many metabolomics studies, some putative identification can still be helpful such as guiding the synthesis of a standard for eventual positive identification.

In tier 2, the remaining peak pairs are searched, based on accurate mass match, against the MyCompoundID (MCID) library composed of 8,021 known human endogenous metabolites and their predicted metabolic products from one metabolic reaction (375,809 compounds). 1242 and 2797 peak pairs were matched in the zero- and one-reaction libraries, respectively. Thus, out of 6206 unique peak pairs detected, 4262 pairs (68.7%) could be positively identified or putatively matched.

5.3.5. Metabolite quantification

Although studying relative quantification capability of mCIL LC-MS is not the focus of this work, the peak intensity ratios measured in the triplicate analysis of 1:1 ^{12}C -/ ^{13}C -labeled plasma can be used to gauge the accuracy and precision. Figure 5.9 shows the number distribution of peak pairs detected as a function of the average peak ratio and their RSD (n=3). Most of the peak pairs in four submetabolome profiling gave the ratio value close to the expected ratio of 1.0, demonstrating high accuracy. The RSD values are less than 20% for more than 95% of the pairs with an average RSD of 5.1% and thus the analytical precision was also very high.

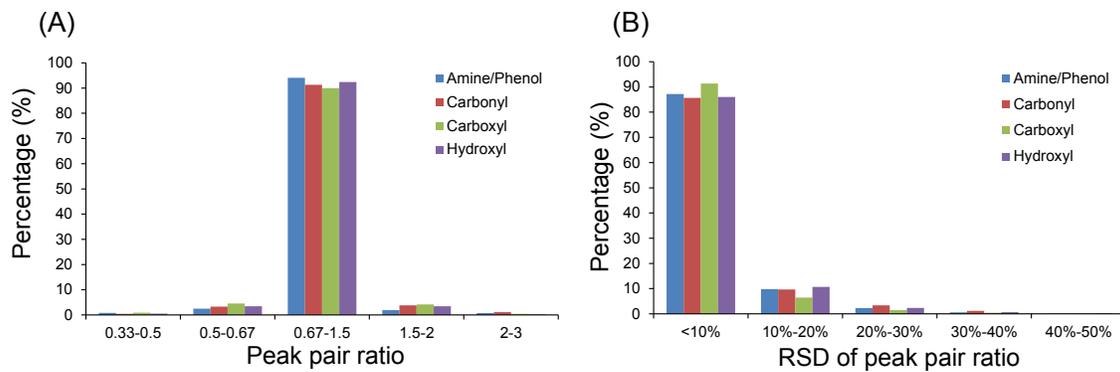


Figure 5.9 Distributions of peak pair numbers as a function of (A) averaged peak ratio and (B) RSD. Data are presented as mean \pm S.D. from experimental triplicate and injection duplicate (n=3)

Table 5.1 List of metabolites identified based on accurate mass and retention time matches to the labeled standard library.

Peak pair #	Peak pair information					Identification result				
	RT (min)	Corrected RT (min)	mz_light	mz_heavy	monoisotopic mass (Da)	HMDB.No.	Name	Accurate mass	mz_light	library RT (min)
Amine_28	2.35	2.32	403.1422	405.1490	169.0839	HMDB00001	1-Methylhistidine	169.0851	403.1434	2.17
						HMDB00479	3-Methylhistidine	169.0851	403.1434	2.01
Amine_50	2.60	2.56	408.1702	410.1768	174.1119	HMDB00517	Arginine	174.1117	408.17	2.44
Amine_101	3.14	3.06	422.1843	424.1909	188.1260	HMDB00670	Homoarginine	188.1273	422.1856	3.00
Amine_116	3.33	3.24	436.1995	438.2062	202.1412	HMDB03334	Symmetric dimethylarginine	202.1430	436.2013	3.05
Amine_122	3.39	3.29	366.1123	368.1189	132.0540	HMDB00168	Asparagine	132.0535	366.1118	3.00
Amine_146	3.71	3.59	380.1278	382.1342	146.0694	HMDB00641	Glutamine	146.0691	380.1275	3.32
Amine_164	3.88	3.75	409.1540	411.1606	175.0957	HMDB00904	Citrulline	175.0957	409.154	3.74
Amine_168	3.95	3.81	399.1041	401.1103	165.0457	HMDB02005	Methionine Sulfoxide	165.0460	399.1043	3.72
						HMDB02005	Methionine Sulfoxide - Isomer	165.0460	399.1043	4.20
Amine_171	3.97	3.83	510.1531	512.1593	276.0948	HMDB11737	Gamma Glutamylglutamic acid	276.0958	510.1541	3.44
Amine_204	4.32	4.15	353.1148	355.1211	119.0565	HMDB00719	Homoserine	119.0582	353.1166	4.05
Amine_214	4.42	4.24	399.1042	401.1106	165.0459	HMDB02005	Methionine Sulfoxide - Isomer	165.0460	399.1043	4.20
Amine_215	4.43	4.26	423.1682	425.1740	189.1099	HMDB00679	Homocitrulline	189.1113	423.1697	4.47
Amine_225	4.60	4.41	339.1008	341.1075	105.0425	HMDB00187	Serine	105.0426	339.1009	4.40
Amine_263	4.99	4.98	381.1114	383.1180	147.0531	HMDB00148	Glutamic Acid	147.0532	381.1115	5.05
Amine_269	5.07	5.11	365.1161	367.1209	131.0578	HMDB00725	Trans-4-Hydroxyl-Proline	131.0582	365.1166	5.17
Amine_280	5.16	5.24	367.0961	369.1026	133.0377	HMDB00191	Aspartic Acid	133.0375	367.0958	5.16
Amine_306	5.39	5.58	422.1728	424.1794	188.1145	HMDB00206	N6-Acetyl-Lysine	188.1161	422.1744	5.71
Amine_320	5.53	5.80	353.1166	355.1232	119.0583	HMDB00167	Threonine	119.0582	353.1166	5.79
Amine_322	5.57	5.84	395.1249	397.1316	161.0666	HMDB00510	Aminoadipic acid	161.0688	395.1271	5.97
Amine_353	5.87	6.20	295.1093	297.1157	61.0510	HMDB00149	Ethanolamine	61.0528	295.1111	6.00
Amine_388	6.21	6.58	309.0899	311.0966	75.0316	HMDB00123	Glycine	75.0320	309.0903	6.59

Amine_402	6.37	6.76	422.1724	424.1794	188.1141	HMDB00446	N-Alpha-acetyllysine	188.1161	422.1744	6.79
Amine_431	6.56	6.98	364.1665	366.1734	130.1082	HMDB02064	N-Acetylputrescine	130.1106	364.1689	7.25
Amine_449	6.64	7.08	406.1416	408.1482	172.0833	HMDB00721	Glycylproline	172.0848	406.1431	7.17
Amine_491	7.02	7.51	478.1261	480.1324	244.0678	HMDB00296	Uridine	244.0695	478.1279	7.84
Amine_497	7.10	7.60	323.1059	325.1126	89.0476	HMDB00056	Beta-Alanine	89.0477	323.106	7.24
						HMDB00161	Alanine	89.0477	323.106	7.57
Amine_503	7.16	7.67	337.1195	339.1261	103.0612	HMDB00112	Gamma-Aminobutyric acid	103.0633	337.1216	7.79
Amine_534	7.53	8.08	453.1675	455.1754	219.1092	HMDB00210	Pantothenic acid	219.1107	453.169	8.37
Amine_556	7.71	8.29	492.1430	494.1477	258.0847	HMDB00884	Ribothymidine - Isomer	258.0852	492.1435	8.54
Amine_562	7.77	8.35	460.1165	462.1231	226.0582	HMDB00296	Uridine - H2O	244.0695	460.1173	8.67
Amine_568	7.83	8.42	370.0957	372.1017	136.0373	HMDB00157	Hypoxanthine - multi-tags	136.0385	370.0968	8.73
Amine_584	8.04	8.65	386.0905	388.0969	152.0322	HMDB00292	Xanthine	152.0334	386.0917	8.95
Amine_587	8.07	8.69	351.1340	353.1420	117.0757	HMDB03355	5-Aminopentanoic acid	117.0790	351.1373	8.68
Amine_593	8.15	8.78	337.1201	339.1282	103.0618	HMDB00452	Alpha-aminobutyric acid	103.0633	337.1216	9.13
						HMDB01906	2-Aminoisobutyric acid	103.0633	337.1216	8.91
						HMDB03911	3-Aminoisobutanoic acid	103.0633	337.1216	8.67
Amine_610	8.33	8.97	408.1568	410.1641	174.0985	HMDB28854	Glycyl-Valine	174.1004	408.1588	9.19
Amine_613	8.35	9.00	376.0946	378.1012	142.0363	HMDB00469	5-Hydroxymethyluracil	142.0378	376.0962	8.87
Amine_625	8.47	9.13	474.1312	476.1369	240.0729	HMDB00884	Ribothymidine - H2O	258.0852	474.1329	9.39
Amine_630	8.52	9.19	337.1211	339.1277	103.0628	HMDB00452	Alpha-aminobutyric acid	103.0633	337.1216	9.13
						HMDB01906	2-Aminoisobutyric acid	103.0633	337.1216	8.91
Amine_646	8.63	9.32	452.1835	454.1899	218.1252	HMDB29043	Seranyl-Leucine	218.1267	452.185	8.9
Amine_655	8.72	9.42	370.0956	372.1022	136.0373	HMDB00157	Hypoxanthine - Isomer	136.0385	370.0968	9.65
Amine_660	8.79	9.49	363.1011	365.1077	129.0428	HMDB00148	Glutamic Acid - H2O	147.0532	363.1009	9.46
Amine_682	8.98	9.71	369.0926	371.0991	135.0342	HMDB02108	Methylcysteine	135.0354	369.0937	9.37
Amine_685	9.01	9.74	466.1989	468.2053	232.1405	HMDB29065	Threoninyl-Leucine	232.1423	466.2006	10.18
Amine_709	9.22	9.98	399.1268	401.1338	165.0685	HMDB00897	7-Methylguanane	165.0651	399.1234	10.32
						HMDB03282	1-Methylguanane	165.0651	399.1234	9.57
Amine_719	9.31	10.07	514.1632	516.1706	280.1049	HMDB00706	Aspartyl-phenylalanine	280.1059	514.1642	10.07
Amine_725	9.42	10.19	349.1227	351.1291	115.0644	HMDB00162	Proline	115.0633	349.1216	10.18

Amine_726	9.43	10.21	466.1975	468.2043	232.1391	HMDB29065	Threoninyl-Leucine	232.1423	466.2006	10.18
Amine_777	9.82	10.60	470.1726	472.1786	236.1143	HMDB28988	Phenylalanyl-Alanine	236.1161	470.1744	10.58
Amine_785	9.88	10.66	365.1505	367.1578	131.0922	HMDB01901	Aminocaproic acid	131.0946	365.1529	10.21
						HMDB03640	Beta-Leucine	131.0946	365.1529	10.78
Amine_792	9.92	10.70	426.1197	428.1244	384.1229	HMDB00939	S-Adenosylhomocysteine	384.1216	426.1191	10.52
Amine_798	9.94	10.72	361.1306	363.1367	127.0722	HMDB03464	4-Guanidinobutanoic acid - H2O	145.0851	361.1329	11.00
Amine_804	9.97	10.75	422.1726	424.1789	188.1143	HMDB00759	Glycyl-Leucine	188.1161	422.1744	11.22
						HMDB28844	Glycyl-Isoleucine	188.1161	422.1744	10.78
Amine_818	10.07	10.85	351.1384	353.1448	117.0801	HMDB00883	Valine	117.0790	351.1373	10.81
Amine_820	10.07	10.85	360.1001	362.1051	126.0418	HMDB02024	Imidazoleacetic acid	126.0429	360.1012	11.12
Amine_828	10.15	10.94	383.1074	385.1137	149.0491	HMDB00696	Methionine	149.0510	383.1094	10.89
Amine_843	10.28	11.07	422.1725	424.1791	188.1142	HMDB00759	Glycyl-Leucine	188.1161	422.1744	11.22
						HMDB28844	Glycyl-Isoleucine	188.1161	422.1744	10.78
Amine_849	10.33	11.12	346.0837	348.0903	112.0254	HMDB00300	Uracil	112.0273	346.0856	11.34
Amine_856	10.40	11.19	495.1679	497.1752	261.1095	HMDB28852	Glycyl-Tryptophan	261.1113	495.1697	11.19
Amine_874	10.53	11.32	436.1884	438.1942	202.1301	HMDB28691	Alanyl-Leucine	202.1317	436.1901	11.36
Amine_888	10.65	11.45	438.1480	440.1545	204.0897	HMDB00929	Tryptophan	204.0899	438.1482	11.44
Amine_917	10.86	11.66	456.1570	458.1640	222.0987	HMDB28848	Glycyl-Phenylalanine	222.1004	456.1588	11.65
Amine_942	11.04	11.85	401.1154	403.1217	167.0571	HMDB01545	Pyridoxal	167.0582	401.1166	12.01
Amine_1068	11.93	12.77	399.1372	401.1437	165.0789	HMDB00159	Phenylalanine	165.0790	399.1373	12.74
Amine_1088	12.09	12.96	462.2024	464.2090	228.1441	HMDB28937	Leucyl-Proline	228.1474	462.2057	12.99
Amine_1099	12.16	13.03	365.1527	367.1592	131.0944	HMDB00172	Lsoleucine	131.0946	365.1529	13.06
						HMDB00557	Alloisoleucine	131.0946	365.1529	13.20
						HMDB00687	Leucine	131.0946	365.1529	13.36
Amine_1123	12.31	13.22	372.0994	374.1055	138.0411	HMDB00301	Urocanic acid	138.0429	372.1012	13.52
Amine_1134	12.40	13.32	365.1541	367.1605	131.0958	HMDB00172	Lsoleucine	131.0946	365.1529	13.06
						HMDB00557	Alloisoleucine	131.0946	365.1529	13.20
						HMDB00687	Leucine	131.0946	365.1529	13.36
Amine_1155	12.72	13.69	498.2042	500.2106	264.1459	HMDB29008	Phenylalanyl-Valine	264.1474	498.2057	13.62

Amine_1178	12.96	13.97	315.1065	317.1137	162.0963	HMDB00450	5-Hydroxylysine	162.1004	315.1085	13.88
Amine_1215	13.31	14.35	416.1150	418.1216	182.0567	HMDB00755	Hydroxyphenyllactici acid	182.0579	416.1162	14.39
Amine_1252	13.74	14.78	425.1152	427.1217	191.0568	HMDB00763	5-Hydroxyindoleacetic acid	191.0582	425.1166	15.09
Amine_1268	13.88	14.94	551.2327	553.2405	317.1743	HMDB29087	Tryptophyl-Leucine	317.1739	551.2323	14.60
Amine_1380	14.75	15.83	585.2175	587.2232	351.1592	HMDB29090	Tryptophyl-Phenylalanine	351.1583	585.2166	15.36
Amine_1381	14.77	15.85	512.2209	514.2268	278.1626	NA	Phenyl-Leucine	278.1631	512.2214	15.90
Amine_1398	15.01	16.10	319.1093	321.1159	85.0509	HMDB03911	3-Aminoisobutanoic acid - H2O	103.0633	319.111	16.29
Amine_1402	15.03	16.12	385.1207	387.1266	151.0624	HMDB01859	Acetaminophen	151.0633	385.1216	16.35
Amine_1404	15.06	16.15	512.2206	514.2281	278.1623	HMDB13243	Leucyl-phenylalanine	278.1630	512.2214	16.59
						NA	Phenyl-Leucine	278.1631	512.2214	15.90
Amine_1415	15.19	16.28	416.1146	418.1209	182.0563	HMDB00118	Homovanillic acid	182.0579	416.1162	16.51
Amine_1422	15.29	16.39	386.1043	388.1107	152.0460	HMDB00440	3-Hydroxyphenylacetic acid	152.0473	386.1057	16.72
						HMDB00669	Ortho-Hydroxyphenylacetic acid	152.0473	386.1057	16.42
						HMDB02390	3-Cresotinic acid	152.0473	386.1057	16.80
Amine_1431	15.40	16.50	300.1022	302.1089	132.0877	HMDB00214	Ornithine	132.0899	300.1033	16.58
Amine_1434	15.43	16.53	546.2052	548.2118	312.1469	HMDB13302	Phenylalanylphenylalanine	312.1474	546.2057	16.55
Amine_1438	15.47	16.57	512.2191	514.2253	278.1608	HMDB13243	Leucyl-phenylalanine	278.1630	512.2214	16.59
Amine_1445	15.52	16.62	371.1047	373.1103	137.0464	HMDB01123	2-Aminobenzoic acid	137.0477	371.106	16.62
Amine_1455	15.58	16.69	386.1049	388.1113	152.0466	HMDB00020	p-Hydroxyphenylacetic acid	152.0473	386.1057	16.91
						HMDB00440	3-Hydroxyphenylacetic acid	152.0473	386.1057	16.72
						HMDB00669	Ortho-Hydroxyphenylacetic acid	152.0473	386.1057	16.42
						HMDB02390	3-Cresotinic acid	152.0473	386.1057	16.80
Amine_1484	15.79	16.91	460.1635	462.1701	226.1052	HMDB28878	Histidinyl-Alanine	226.1066	460.1649	16.69
Amine_1521	16.08	17.21	402.0978	404.1052	168.0395	HMDB00484	Vanillic acid	168.0423	402.1006	17.34
Amine_1547	16.25	17.37	372.0883	374.0948	138.0300	HMDB00500	4-Hydroxybenzoic acid	138.0317	372.09	17.57
Amine_1554	16.35	17.48	307.1097	309.1163	146.1028	HMDB00182	Lysine	146.1055	307.1111	17.47
Amine_1574	16.59	17.72	400.1201	402.1264	166.0618	HMDB02199	Desaminotyrosine	166.0630	400.1213	18.04
Amine_1648	17.17	18.31	389.1272	391.1337	155.0689	HMDB00177	Histidine	155.0695	389.1278	18.09

Amine_1716	17.79	18.92	393.1818	395.1886	159.1235	HMDB00991	2-aminooctanoic acid	159.1259	393.1842	19.20
Amine_1724	17.86	19.00	395.1039	397.1107	161.0456	HMDB03320	Indole-3-carboxylic acid	161.0477	395.106	19.27
Amine_1887	19.41	20.55	421.1200	423.1260	187.0617	HMDB00734	Indoleacrylic acid	187.0633	421.1216	20.61
Amine_1902	19.59	20.73	360.1116	362.1191	252.1065	HMDB29098	Tyrosyl-Alanine	252.1110	360.1138	20.86
Amine_1922	19.77	20.91	278.1056	280.1127	88.0946	HMDB01414	1,4-Diaminobutane	88.1000	278.1083	21.27
Amine_1972	20.31	21.45	353.1028	355.1108	238.0889	HMDB28853	Glycyl-Tyrosine	238.0954	353.106	21.63
Amine_2072	21.51	22.66	324.5937	326.6004	181.0707	HMDB00158	Tyrosine	181.0739	324.5953	22.65
						HMDB06050	o-Tyrosine	181.0739	324.5953	22.38
Amine_2121	22.01	23.16	328.0982	330.1047	94.0399	HMDB00228	Phenol	94.0419	328.1002	23.16
Amine_2222	23.04	24.19	381.1350	383.1416	294.1534	HMDB29109	Tyrosyl-Leucine	294.1580	381.1373	23.77
Amine_2519	25.28	26.43	396.0202	398.0241	161.9619	HMDB04811	2,4-Dichlorophenol	161.9639	396.0222	26.30
Carbonyl_109	5.53	5.56	428.1491	430.1556	180.0639	HMDB00143	Galactose	180.0634	428.1486	5.98
Carbonyl_119	5.68	5.71	590.2018	592.2084	342.1166	HMDB00186	Lactose	342.1162	590.2014	5.86
Carbonyl_130	5.86	5.89	590.2014	592.2084	342.1162	HMDB00186	Lactose	342.1162	590.2014	5.86
Carbonyl_152	6.14	6.17	428.1492	430.1557	180.0640	HMDB00143	Galactose	180.0634	428.1486	5.98
						HMDB00122	Glucose	180.0634	428.1486	6.08
						HMDB00660	Fructose	180.0634	428.1486	6.44
Carbonyl_161	6.25	6.28	442.1267	444.1341	194.0415	HMDB00127	Dlucuronic acid	194.0427	442.1279	6.27
						HMDB02545	Galacturonic acid	194.0427	442.1279	6.32
Carbonyl_177	6.40	6.45	412.1512	414.1585	164.0660	HMDB00174	Fucose	164.0685	412.1537	6.86
Carbonyl_221	6.83	6.91	412.1512	414.1575	164.0660	HMDB00174	Fucose	164.0685	412.1537	6.86
Carbonyl_249	7.19	7.29	338.1156	340.1223	90.0304	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59
Carbonyl_279	7.56	7.68	338.1162	340.1230	90.0310	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59
Carbonyl_349	9.07	9.18	394.1414	396.1484	146.0562	HMDB00422	2-Methylglutaric acid	146.0579	394.1431	8.76
Carbonyl_364	9.35	9.47	350.1161	352.1227	102.0309	HMDB01259	Succinic acid semialdehyde	102.0317	350.1169	9.34
Carbonyl_376	9.59	9.71	364.1317	366.1382	116.0465	HMDB00720	Levulinic acid	116.0473	364.1325	9.59
Carbonyl_384	9.85	9.96	364.1321	366.1386	116.0469	HMDB00720	Levulinic acid	116.0473	364.1325	9.59
Carbonyl_398	10.27	10.38	427.1419	429.1480	179.0567	HMDB00714	Hippuric Acid	179.0582	427.1434	10.56
Carbonyl_408	10.62	10.73	278.0947	280.1013	30.0095	HMDB01426	Formaldehyde	30.0106	278.0958	10.66
Carbonyl_414	10.84	10.95	350.1150	352.1218	102.0298	HMDB00005	2-Ketobutyric	102.0317	350.1169	11.10

Carbonyl_429	11.19	11.29	292.1112	294.1177	44.0260	HMDB00990	Acetaldehyde	44.0262	292.1114	11.20
Carbonyl_468	12.02	12.13	306.1261	308.1327	58.0409	HMDB01659	Acetone	58.0419	306.1271	12.01
Carbonyl_519	13.13	13.25	378.1472	380.1536	130.0620	HMDB00491	3-Methyl-2-oxovaleric acid	130.063	378.1482	13.39
						HMDB00408	2-Methyl-3-ketovaleric acid	130.063	378.1482	13.39
Carbonyl_533	13.30	13.41	306.1259	308.1326	58.0407	HMDB03366	Propanal	58.0419	306.1271	13.28
Carbonyl_594	14.25	14.33	320.1405	322.1477	72.0553	HMDB03543	Butanal	72.0575	320.1427	14.83
						HMDB00474	2-Butanone	72.0575	320.1427	14.24
Carbonyl_635	14.85	14.91	320.1423	322.1488	72.0571	HMDB03543	Butanal	72.0575	320.1427	14.83
Carbonyl_657	15.15	15.21	320.1420	322.1486	72.0568	HMDB03543	Butanal	72.0575	320.1427	14.83
Carbonyl_658	15.17	15.23	610.2921	612.2985	362.2069	HMDB00063	Cortisol	362.2093	610.2945	15.19
Carbonyl_663	15.24	15.30	332.1399	334.1488	84.0547	HMDB01184	Methyl propenyl ketone	84.0575	332.1427	15.25
Carbonyl_674	15.43	15.49	346.1558	348.1629	98.0706	HMDB03315	Cyclohexanone	98.0732	346.1584	15.29
Carbonyl_684	15.61	15.67	334.1560	336.1628	86.0708	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
Carbonyl_732	16.16	16.22	334.1581	336.1647	86.0729	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
Carbonyl_755	16.33	16.39	354.1246	356.1314	106.0394	HMDB06115	Benzaldehyde	106.0419	354.1271	16.23
Carbonyl_766	16.40	16.47	368.1416	370.1475	120.0564	HMDB06236	Phenylacetaldehyde	120.0575	368.1427	16.35
Carbonyl_771	16.42	16.49	334.1577	336.1645	86.0725	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
Carbonyl_886	17.61	17.68	348.1734	350.18	100.0882	HMDB05994	Hexanal	100.0888	348.1740	17.91
Carbonyl_895	17.7	17.77	284.0937	286.0997	72.0171	HMDB01167	Pyruvaldehyde - 2 tags 2 charges	72.0211	284.0958	17.66
Carbonyl_919	17.82	17.9	536.2939	538.3004	288.2087	HMDB00234	Testosterone	288.2089	536.2941	17.88
						HMDB00077	Dehydroepiandrosterone	288.2089	536.2941	17.88
Carbonyl_923	17.94	18.02	348.1739	350.1805	100.0887	HMDB05994	Hexanal	100.0888	348.1740	17.91
Carbonyl_925	18	18.09	536.2947	538.2997	288.2095	HMDB00077	Dehydroepiandrosterone	288.2089	536.2941	17.88
						HMDB00234	Testosterone	288.2089	536.2941	17.88
Carbonyl_1098	19.89	19.92	536.2938	538.3001	288.2086	HMDB00899	5a-Androstane-3,17-dione	288.2089	536.2941	20.16
						HMDB00628	Epitestosterone	288.2089	536.2941	20.20
Carbonyl_1197	20.68	20.71	538.3098	540.3165	290.2246	HMDB00031	Androsterone	290.2246	538.3098	20.71
						HMDB00490	5b-Androsterone	290.2246	538.3098	20.57
Carbonyl_1210	20.75	20.78	376.2051	378.2116	128.1199	HMDB01140	Octanal	128.1201	376.2053	20.72

Carbonyl_1517	23.03	23.05	404.2359	406.2429	156.1507	HMDB11623	Decanal	156.1514	404.2366	23.26
Carbonyl_1561	23.29	23.3	404.2373	406.2439	156.1521	HMDB11623	Decanal	156.1514	404.2366	23.26
Carboxyl_74	6.65	7.10	252.1222	254.1289	90.0309	HMDB00700	Hydroxypropionic acid	90.0317	252.1230	7.02
Carboxyl_112	7.61	8.18	266.1366	268.1434	104.0453	HMDB00357	3-Hydroxybutyric acid	104.0473	266.1387	8.56
Carboxyl_142	8	8.64	266.1366	268.1433	104.0453	HMDB00357	3-Hydroxybutyric acid	104.0473	266.1387	8.56
Carboxyl_149	8.13	8.79	208.0987	210.1054	46.0074	HMDB00142	Formic acid	46.0055	208.0968	9.26
Carboxyl_157	8.17	8.84	266.1365	268.1432	104.0452	HMDB00729	Hydroxyisobutyric acid	104.0473	266.1387	9.21
						HMDB00357	3-Hydroxybutyric acid	104.0473	266.1387	8.56
Carboxyl_179	8.62	9.36	306.1681	308.1743	144.0755	HMDB01988	4-Hydroxycyclohexylcarboxylic acid	144.0786	306.1700	9.69
Carboxyl_180	8.62	9.36	266.1378	268.1443	104.0465	HMDB00008	2-Hydroxybutyric acid	104.0473	266.1387	9.83
						HMDB00729	Hydroxyisobutyric acid	104.0473	266.1387	9.21
Carboxyl_218	8.86	9.64	222.1133	224.1202	60.0220	HMDB00042	Acetic acid	60.0211	222.1125	10.00
Carboxyl_233	9	9.81	344.1468	346.1536	182.0555	HMDB00755	Hydroxyphenyllactic acid	182.0579	344.1492	10.23
Carboxyl_253	9.28	10.13	292.1153	294.1223	130.0240	HMDB00634	Citraconic acid	130.0266	292.1179	9.89
Carboxyl_255	9.29	10.14	278.1363	280.1433	116.0450	HMDB00720	Levulinic acid	116.0473	278.1387	10.43
Carboxyl_367	10.67	11.65	280.1524	282.1591	118.0611	HMDB00407	2-Hydroxy-3-methylbutyric acid	118.063	280.1543	11.86
						HMDB01987	2-Hydroxy-2-methylbutyric acid	118.063	280.1543	11.28
Carboxyl_384	10.95	11.93	314.1358	316.1426	152.0440	HMDB00020	Parahydroxyphenylacetic acid	152.0473	314.1387	12.34
Carboxyl_387	11.02	12	236.1276	238.1342	74.0363	HMDB00237	Propionic acid	74.0368	236.1281	12.41
Carboxyl_404	11.19	12.18	344.1487	346.1552	182.0574	HMDB00423	3,4-Dihydroxyhydrocinnamic acid	182.0579	344.1492	11.72
						HMDB00118	Homovanillic acid	182.0579	344.1492	12.64
Carboxyl_425	11.58	12.58	330.1308	332.137	168.0395	HMDB60003	Isovanillic acid	168.0423	330.1336	12.74
						HMDB00484	Vanillic acid	168.0423	330.1336	12.97
Carboxyl_480	12.14	13.04	300.1203	302.1265	138.0290	HMDB02466	3-Hydroxybenzoic acid	138.0317	300.123	13.30
Carboxyl_494	12.23	13.09	294.1675	296.1742	132.0762	HMDB00746	Hydroxyisocaproic acid	132.0786	294.17	13.57
						HMDB01975	2-Ethyl-2-Hydroxybutyric acid	132.0786	294.17	13.56

Carboxyl_528	12.69	13.33	328.1522	330.1589	166.0609	HMDB02199	3-(4 Hydroxyphenyl)propionic acid	166.063	328.1543	13.54
Carboxyl_586	13.33	14.7	250.1431	252.1499	88.0518	HMDB00039	Butyric acid	88.0524	250.1438	14.94
		14.7	250.1431	252.1499	88.0518	HMDB01873	Isobutyric acid	88.0524	250.1438	15.07
Carboxyl_653	14.13	14.19	314.1359	316.1436	152.0446	HMDB04815	4-Hydroxy-3-methylbenzoic acid	152.0473	314.1387	14.41
Carboxyl_657	14.19	14.29	350.1928	352.2001	188.1015	HMDB00784	Azelaic acid	188.1049	350.1962	14.33
Carboxyl_682	14.50	15.96	262.1413	264.1483	100.0500	HMDB01470	Tiglic acid	100.0524	262.1438	16.27
		15.96	262.1413	264.1483	100.0500	HMDB00509	3-Methylcrotonic acid	100.0524	262.1438	16.18
Carboxyl_705	14.76	16.24	441.1999	445.2138	118.0246	HMDB00254	Succinic acid	118.0266	441.202	16.62
Carboxyl_790	15.70	17.19	264.1586	266.1652	102.0673	HMDB00892	Valeric acid	102.0681	264.1594	17.64
						HMDB00718	Isovaleric acid	102.0681	264.1594	17.53
						HMDB02176	Ethylmethylacetic acid	102.0681	264.1594	17.63
Carboxyl_796	15.80	17.37	413.1687	417.1819	89.9934	HMDB02329	Oxalic acid - 2 tags 1 charge	89.9953	413.1707	17.73
Carboxyl_839	16.16	17.75	264.1583	266.165	102.0670	HMDB00892	Valeric acid	102.0681	264.1594	17.64
						HMDB00718	Isovaleric acid	102.0681	264.1594	17.53
						HMDB02176	Ethylmethylacetic acid	102.0681	264.1594	17.63
Carboxyl_852	16.25	17.85	228.1107	230.1173	132.0388	HMDB00661	Glutaric acid - 2 tags 2 charges	132.0423	228.1125	18.10
Carboxyl_865	16.35	17.95	292.1514	294.1582	130.0601	HMDB00695	Methyloxovaleric acid	130.063	292.1543	18.34
Carboxyl_964	17.48	19.16	469.2308	473.2448	146.0555	HMDB00448	Adipic acid - 2 tags 1 charge	146.0579	469.2333	19.37
Carboxyl_982	17.65	19.34	312.1568	314.1638	150.0648	HMDB00764	Hydrocinnamic acid	150.0681	312.1594	19.60
						HMDB02222	3-Methylphenylacetic acid	150.0681	312.1594	19.80
Carboxyl_1046	18.36	20.09	233.1033	235.1117	142.0239	HMDB06331	Cis,cis-muconic acid - 2 tags 2 charges	142.0266	233.1046	20.01
Carboxyl_1057	18.47	20.21	278.1724	280.1792	116.0811	HMDB00689	Isocaproic acid	116.0837	278.1751	20.55
Carboxyl_1068	18.55	20.29	455.2136	459.2284	132.0383	HMDB00622	Ethylmalonic acid - 2 tags 1 charge	132.0423	455.2177	20.74
Carboxyl_1089	18.82	20.58	278.1736	280.1803	116.0823	HMDB00689	Isocaproic acid	116.0837	278.1751	20.55
Carboxyl_1177	19.51	21.33	318.0864	320.0904	155.9951	HMDB01544	m-Chlorobenzoic acid	155.9978	318.0891	21.64
Carboxyl_1185	19.33	21.13	570.3766	572.383	408.2853	HMDB00415	3a,6b,7b-Trihydroxy-5b-cholanoic acid	408.2876	570.3789	20.84

Carboxyl_1219	19.97	21.82	378.2607	380.2671	216.1694	HMDB02059	12-Hydroxydodecanoic acid	216.1725	378.2639	22.22
Carboxyl_1316	20.96	22.87	570.3783	572.3848	408.2870	HMDB00505	Allocholic acid	408.2876	570.3789	22.82
						HMDB00619	Cholic acid	408.2876	570.3789	23.14
Carboxyl_1319	20.98	22.89	312.1563	314.1645	150.0650	HMDB02097	4-Ethylbenzoic acid	150.0681	312.1594	23.09
Carboxyl_1414	22.18	24.16	292.189	294.1956	130.0977	HMDB00666	Heptanoic acid	130.0994	292.1907	24.42
Carboxyl_1451	22.59	24.62	554.3831	556.3899	392.2918	HMDB00733	Hyodeoxycholic acid	392.2927	554.384	24.89
						HMDB00946	Ursodeoxycholic acid	392.2927	554.384	24.88
Carboxyl_1554	23.69	25.85	256.1415	258.1484	188.1005	HMDB00784	Azelaic acid - 2 tags 2 charges	188.1049	256.1438	26.14
Carboxyl_1596	24.15	26.38	328.1887	330.1952	166.0974	HMDB04586	Perillic acid	166.0994	328.1907	26.51
Carboxyl_1680	25.22	27.58	306.2053	308.212	144.1140	HMDB01877	Valproic acid	144.115	306.2064	27.45
Carboxyl_1710	25.68	28.1	552.3667	554.374	390.2754	HMDB00467	Nutriacholic acid	390.277	552.3683	27.81
Carboxyl_1746	26.08	28.56	263.1494	265.1559	202.1163	HMDB00792	Sebacic acid - 2 tags 2 charges	202.1205	263.1516	28.19
Carboxyl_1894	27.50	30.16	270.1579	272.1644	216.1331	HMDB00888	Undecanedioic acid - 2 tags 2 charges	216.1362	270.1594	29.87
Carboxyl_1931	27.85	30.55	276.1568	278.1635	228.1309	HMDB00933	Traumatic acid - 2 tags 2 charges	228.1362	276.1594	30.46
Carboxyl_1944	27.99	30.71	552.3668	554.3769	390.2755	HMDB00503	7a-Hydroxy-3-oxo-5b-cholanoic acid	390.277	552.3683	31.16
Carboxyl_1966	28.14	30.88	554.3842	556.3907	392.2929	HMDB00518	Chenodeoxycholic acid	392.2927	554.384	30.45
Carboxyl_2029	28.65	31.45	277.1646	279.1709	230.1467	HMDB00623	Dodecanedioic acid - 2 tags 2 charges	230.1518	277.1672	31.06
Carboxyl_2068	28.89	31.73	334.2362	336.2431	172.1449	HMDB00511	Capric acid	172.1463	334.2377	31.54
Carboxyl_2089	29.09	31.94	334.2363	336.2431	172.1450	HMDB00511	Capric acid	172.1463	334.2377	31.54
Carboxyl_2402	31.43	34.59	538.3861	540.3939	376.2948	HMDB00761	Lithocholic acid	376.2977	538.3891	34.64
Carboxyl_2413	31.56	34.74	305.1981	307.2025	286.2135	HMDB00672	Hexadecanedioic acid - 2 tags 2 charges	286.2144	305.1985	34.30
Carboxyl_2591	33.20	36.58	319.2113	321.2176	314.2400	HMDB00782	Octadecanedioic acid - 2 tags 2 charges	314.2457	319.2142	36.50
Carboxyl_2616	33.48	36.9	390.2991	392.3058	228.2078	HMDB00806	Myristic acid	228.2089	390.3003	36.63
Carboxyl_2688	34.31	37.83	442.3315	444.3377	280.2402	HMDB00673	Linoleic acid	280.2402	442.3316	37.65
Carboxyl_2701	34.48	38.03	404.3159	406.3228	242.2246	HMDB00826	Pentadecanoic acid	242.2246	404.3159	38.50
Carboxyl_2709	34.58	38.14	442.33	444.3369	280.2387	HMDB00673	Linoleic acid	280.2402	442.3316	37.65

Carboxyl_2710	34.65	38.22	404.316	406.3228	242.2247	HMDB00826	Pentadecanoic acid	242.2246	404.3159	38.50
Carboxyl_2713	34.74	38.32	468.3472	470.3538	306.2559	HMDB02925	Eicosatrienoic acid	306.2559	468.3472	38.38
Carboxyl_2732	34.96	38.54	468.3461	470.3528	306.2548	HMDB02925	Eicosatrienoic acid	306.2559	468.3472	38.38
Carboxyl_2751	35.23	38.81	468.3467	470.3534	306.2554	HMDB02925	Eicosatrienoic acid	306.2559	468.3472	38.38
Hydroxyl_194	6.88	7.29	326.1043	328.1106	92.0460	HMDB00131	Glycerol	92.0473	326.1056	7.31
Hydroxyl_415	9.63	9.90	310.1098	312.1164	76.0515	HMDB01881	1,2-Propanediol	76.0524	310.1107	9.93
Hydroxyl_536	10.98	11.23	344.0706	346.0764	110.0123	HMDB31334	3-Chloro-1,2-propanediol	110.0135	344.0718	11.26
Hydroxyl_621	12.14	12.48	336.1264	338.1329	102.0681	HMDB31175	Tetrahydrofurfuryl alcohol	102.0681	336.1264	12.81
Hydroxyl_684	12.93	13.35	280.1003	282.1066	46.0420	HMDB00108	Ethanol	46.0417	280.1	13.13
Hydroxyl_752	13.69	14.18	294.1154	296.1221	60.0571	HMDB00820	1-Propanol	60.0575	294.1158	14.54
						HMDB00863	2-Propanol	60.0575	294.1158	14.14
Hydroxyl_826	14.69	15.24	306.1163	308.1226	72.0580	HMDB33858	Crotyl alcohol	72.0575	306.1158	14.82
						HMDB31324	3-Buten-1-ol	72.0575	306.1158	14.77
Hydroxyl_835	14.88	15.42	559.1549	563.1691	92.0456	HMDB00131	Glycerol	92.0473	559.1567	15.67
Hydroxyl_843	14.93	15.48	372.1261	374.1327	138.0678	HMDB41607	2-Phenoxyethanol	138.0681	372.1264	15.50
Hydroxyl_895	15.46	16.07	308.1304	310.1367	74.0721	HMDB31456	tert-Butyl alcohol	74.0732	308.1315	16.03
						HMDB06006	2-Methyl-1-propanol	74.0732	308.1315	15.96
Hydroxyl_1003	16.89	17.64	322.1461	324.1538	88.0878	HMDB31527	2-Methyl-1-butanol	88.0888	322.1471	17.38
						HMDB06007	Iso amyl Alcohol	88.0888	322.1471	17.49
Hydroxyl_1065	17.55	18.35	272.0824	274.0889	76.0482	HMDB01881	1,2-Propanediol - 2 tags 2 charges	76.0524	272.0845	18.23
Hydroxyl_1235	19.58	20.45	350.1781	352.1847	116.1198	HMDB31479	1-Heptanol	116.1201	350.1784	20.59
Hydroxyl_1354	20.94	21.87	364.1938	366.2000	130.1355	HMDB01183	1-Octanol	130.1358	364.1941	22.06
Hydroxyl_1397	21.43	22.41	390.2089	392.2154	156.1506	HMDB35094	Citronellol	156.1514	390.2097	22.30
Hydroxyl_1496	22.19	23.24	390.2095	392.2162	156.1512	HMDB03352	Menthol	156.1514	390.2097	22.92
Hydroxyl_1632	23.37	24.40	392.2252	394.2317	158.1669	HMDB11624	Decyl alcohol	158.1671	392.2254	24.60

5.4. Conclusions

In summary, we developed a simple technique of using one LC-MS instrumental setup with 4-channel chemical isotope labeling to perform relative metabolome quantification for metabolomics. The technique has been demonstrated to have very high coverage of metabolites. Using human plasma as an example, we illustrate the possibility of detecting 7431 peak pairs or metabolites (not spectral features) using 4-channel sample labeling and only one LC-MS condition. This relatively simple technique can increase our ability of characterizing the overall metabolome for systems biology and biomarker discovery research. It opens a new venue for detailed characterization of the metabolomes in metabolomics research.

Chapter 6

Comprehensive Profiling of Cerebral Amyloid Angiopathy and Alzheimer's Disease Using In-depth Metabolomic Analysis for Biomarkers Discovery

6.1. Introduction

Cerebral amyloid angiopathy (CAA), characterised by the deposition of beta-amyloid protein ($A\beta$) on the walls of cortical and leptomeningeal vessels, has attracted increasing attention from both clinical diagnosis and mechanism research areas, especially in the recent five years.¹⁸⁸ As a sister disease of Alzheimer's disease (AD), it shares important feature of molecular mechanisms with AD, i.e., the aggregation of $A\beta$. In the case of AD, the deposition of $A\beta$ forms senile plaques in cortex. While in CAA, the aggregation occurs on the walls of intracerebral vessels, resulting in brittle vessels and bleeding in the brain (i.e. intracranial hemorrhage).¹⁸⁹ Although CAA trends to co-occur with AD, it is also found in a large number of cognitively healthy elderly patients.¹⁹⁰⁻¹⁹¹ In addition, differences of molecular mechanisms between two diseases were also reported.¹⁹²⁻¹⁹³

The current standard guideline for diagnosis of CAA is the modified Boston criteria,¹⁹⁴ in which definite diagnosis requires full post-mortem examination. The lack of availability of brain autopsy, as well as the fact that autopsy is limited to deceased persons, means that definite diagnosis of CAA is uncommon. However, a combination of brain imaging, such as magnetic resonance imaging (MRI), plus clinical criteria can allow a diagnosis of "probable CAA", which correlates well with autopsy and is the

highest level of diagnosis certainty without using brain tissue. Obviously, incorporation of additional biomarkers and using other confirmation approach can improve the non-invasive diagnosis confidence. Reliably identifying CAA at early stage could be useful for selecting patients for clinical trails for this disease which currently has no disease-modifying treatments.¹⁹⁵ Additionally, finding biomarkers may also benefit in understanding the pathogenesis of the disease and differentiating CAA from AD.

Several studies have been reported using different approaches to screen biomarkers, such as assessing CSF A β amount¹⁹⁶ or using proteomics approach.¹⁹⁷ For biomarker discovery, metabolomics is an emerging field, which performs quantification and identification of many small molecules in biological systems. Through metabolome profiling, metabolites that differentiate diseases can be utilized as potential biomarker candidates.^{14, 27, 198} However, the large number of metabolites and their great chemical complexities hamper the comprehensive metabolomic analysis. Multichannel chemical isotope labeling liquid chromatography mass spectrometry (mCIL LC-MS) has been reported as a relatively new technique for biomarker discovery, which can generate comprehensive and in-depth metabolome profile.^{30, 172} The method uses different labeling reagents to analyze chemical-group-based submetabolome. Then the combined datasets from submetabolomes would allow the analysis of the entire metabolome.²²

In this study, we applied four-channel high-performance CIL LC-MS methods to compare the plasma metabolomic profiles of CAA, AD, and controls. The methods include dansylation for analyzing amine/phenol submetabolome,³⁰ base-activated dansylation for hydroxyl submetabolome,³¹ DmPA bromide labeling for carboxyl submetabolome,⁸⁵ and dansylhydrazine (DnsHz) labeling for carbonyl submetabolome.¹⁷²

We performed pair-wise comparisons between two groups to find biomarker candidates for separating 1) AD patients and healthy controls; 2) CAA patients and healthy controls and 3) CAA and AD patients. To our knowledge, this is the first report of using metabolomics approach to obtain a global insight on metabolic alterations of CAA, as well as investigating biomarkers for the two sister diseases.

6.2. Experimental Section

6.2.1. Sample Collection and Metabolite Extraction

Blood samples were collected from donors after obtaining the informed consent. The University of Alberta Research Ethics Board approved this study. Three groups were included in the analysis. 29 CAA patients (mean age 74.4 with standard deviation of 7.4; 11 female) were diagnosed using Boston criteria.¹⁹⁴ 17 AD patients (mean age 70 with standard deviation of 8.7; 7 female) who met the clinical NIA-AA criteria were enrolled.¹⁹⁹ 23 healthy individuals (mean age 65.6 with standard deviation of 9.5; 12 female) in control group were recruited via community advertising without stroke or cognitive impairment. In total of 69 plasma samples were analyzed by duplicating experiment in the study. To prepare plasma samples, whole blood was processed in EDTA-treated tubes. Cells were removed from plasma by centrifuging at 1200 g for 15 minutes at 4 °C. For each individual, 0.5 mL of plasma was sufficient for all the analysis. The collected plasma samples were stored at - 80 °C until further use.

Metabolites were extracted from plasma samples via methanol protein precipitation. Three volumes of cold methanol was added into plasma samples, vortexed

and incubated in -20 °C freezer for 2 hours. The mixtures were then centrifuged at 20817 g for 15 min at 4 °C. The supernatants (3 times of volume of the original sample) were dried down using a SpeedVac. The dried samples were reconstituted in water with the same volume of original sample for chemical labeling or stored in -80 °C freezer for further use.

6.2.2. Multichannel Chemical Isotope Labeling

In the study, multichannel chemical isotope labeling (mCIL) LC-MS techniques were applied to perform comprehensive and in-depth metabolomics analysis. Four chemistries were used to derivatize metabolites from different submetabolomes: dansylation for amine/phenol submetabolome, base-activated dansylation for hydroxyl submetabolome, DmPA bromide labeling for carboxyl submetabolome and dansylhydrazine labeling for carbonyl submetabolome. Samples were labeled either with ¹²C-reagents or ¹³C-reagents (see below). All labeling reagents and ready-to-use kits are available at www.novamt.com.

For the amine/phenol submetabolome analysis, 25 µL of the extracted solution was mixed with 12.5 µL of 250 mM Na₂CO₃/NaHCO₃ buffer and 12.5 µL of ACN. The solution was vortexed, spun down and mixed with 25 µL freshly prepared dansyl chloride (DnsCl) (18 mg/mL in ACN, for light labeling). After vortexing and spinning down, the mixture was incubated at 40 °C for 60 min. After 1 hour, NaOH solution (250 mM, 5 µL) was added to quench the excess DnsCl at 40 °C for another 10 min. Finally 25 µL of 425 mM formic acid (in 50/50 ACN/H₂O) was added to consume the excess NaOH.

DnsCl were also used for hydroxyl submetabolome analysis in different reaction condition combined with liquid-liquid extraction to remove water and interference. After

protein precipitation, 60 μL of sample was mixed with 10 μL of saturated NaCl solution and 5 μL of 6 M HCl. 180 μL of ethyl acetate was used to extract metabolites twice. The organic phase were combined, dried down and then re-dissolved in 60 μL of ACN. 25 μL of the sample was mixed with 25 μL 4-dimethylaminopyridine (DMAP) (24.5 mg/mL in ACN) and 25 μL of DnsCl (18 mg/mL in ACN). The reaction was allowed to proceed for 1 hour at 60 $^{\circ}\text{C}$. Then 5 μL of 250 mM sodium hydroxide solution was added to quench the reaction. The solution was incubated at 60 $^{\circ}\text{C}$ for another 10 min. After that 25 μL of 425 mM formic acid solution in 50/50 ACN/ H_2O was added to consume the excess NaOH.

Liquid-liquid extraction was also applied to enrich carboxyl-containing metabolites in plasma when performing carboxyl submetabolome analysis. 60 μL of plasma sample was mixed with 10 μL of saturated NaCl solution, 5 μL of 6 M HCl and 180 μL of ethyl acetate. After vortexing and spinning down, the organic phase was transferred to a new vial. 20 μL of trimethylamine solution (180 mg/mL in ACN) was added. Then the solution was dried down and re-dissolved in 60 μL of 20 mg/mL TEA solution (in ACN). 30 μL of the solution was mixed with 30 μL DMPA (20 mg/mL in ACN) for labeling. The vial was again vortexed and spun down. It takes 60 min for the reaction at 80 $^{\circ}\text{C}$ incubation.

For carbonyl submetabolome analysis, 20 μL of sample was mixed with 20 μL of 144 mM HCl solution in ACN and 20 μL of 20 mM DnsHz in ACN. After vortexing and spinning down, the solution was incubated at 40 $^{\circ}\text{C}$ for 60 min for labeling. The mixture was then placed in -80 $^{\circ}\text{C}$ freezer for 5 min to stop the labeling reaction, followed by

drying down and resuspending in 80 μL of acetonitrile/water mixture (ACN/ H_2O , 50:50, v/v).

6.2.3. LC-MS Analysis

In mCIL LC-MS approach to study metabolomic changes among groups, individual samples from different groups were labeled with ^{12}C -reagents and a pooled sample which was generated by mixing aliquots of all individual samples was labeled by ^{13}C -reagents. In the study, individual samples were processed as experimental duplicate for each chemical labeling. The same mole amount of ^{13}C -labeled pooled sample was spiked into ^{12}C -labeled individual samples as reference, generating the ^{12}C -/ ^{13}C -labeled mixtures. The mixing of ^{12}C -labeled samples and ^{13}C -labeled pooled was based on LC-UV quantification of dansylated metabolites for sample amount normalization.¹⁴⁴ ^{12}C -/ ^{13}C -labeled mixtures were injected into LC-MS for analysis. Metabolite was detected as a peak pair and the relative quantification of the metabolite was realized by comparing peak ratios between groups. Prior to LC-MS analysis of the entire sample set, quality control (QC) sample for each channel was prepared by equal volume mix of a ^{12}C -labeled and a ^{13}C -labeled pooled sample.

^{12}C -/ ^{13}C -labeled mixtures from each channel were analyzed using a Bruker maXis II QTOF mass spectrometer (Bruker, Germany) linked to an UltiMate 3000 UHPLC system (Thermo Scientific, USA). The samples were injected onto an Agilent Eclipse Plus C18 column (2.1 mm \times 10 cm, 1.8 μm particle size, 95 \AA pore size) for separation. Mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile. Mobile phase B was 0.1% (v/v) formic acid in acetonitrile. The chromatographic conditions for amine/phenol submetabolome profiling was: $t = 0$ min, 20% B; $t = 3.5$ min, 35% B; $t = 18$ min, 65% B;

t = 21 min, 98% B; t = 34 min, 98% B; for hydroxyl submetabolome profiling was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 9.2 min, 65% B; t = 21.2 min, 98% B; t = 31.2 min, 98% B; for carboxyl submetabolome profiling was: t = 0 min, 20% B; t = 9 min, 50% B; t = 22 min, 65% B; t = 26 min, 80% B; t = 29 min, 98% B; t = 40 min, 98% B. and for carbonyl submetabolome profiling was: t = 0 min, 1% B; t = 3 min, 25% B; t = 23 min, 98% B; t = 34 min, 98% B. The flow rate was 180 μ L/min. The column temperature was kept at 30 °C. All MS spectra were collected in positive ion mode with a scan range of 200 to 1000 m/z at a spectral acquisition rate of 1 Hz. The capillary voltage was 4500 V and nebulizer gas was 2.0 bars. The dry gas flow was set to 8 L/min and the temperature was 230 °C. QC samples were injected every 15 sample runs to monitor instrument performance.

6.2.4. Data Processing

The resulting LC-MS data were processed using a set of R language based in-house software. First the $^{12}\text{C}/^{13}\text{C}$ -peak pairs were extracted from each run by the IsoMS software.⁴⁸ The peak intensity ratio was also calculated for each peak pair. Then peak pairs from different samples were aligned by retention time and accurate mass using IsoMS-Align. The missing ratio values were filled back by Zero-fill program.¹⁴⁰ Finally IsoMS-Quant program determine the chromatographic peak ratio of each peak pair and generate the final metabolite-intensity table for each metabolome.¹⁴¹ The combined metabolite-intensity table from four channels was subjected to statistical analysis. Positive metabolite identification was based on accurate mass and retention time search against the labeled standard library containing 664 authentic metabolites.⁵⁹ Putative

identification was done using MyCompoundID zero-reaction and one-reaction libraries (www.mycompoundid.org).¹⁹

6.2.5. Statistical Analysis

Only peak pairs that were shared by more than 80% of the samples within one group were retained for statistical analysis. The data were mean-centered and auto-scaled (unit variance) prior to analysis. Multivariate statistical analysis including principle component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) was carried out using Metaboanalyst (www.metaboanalyst.ca).²⁰⁰ Fold change and p-value were calculated in Metaboanalyst after data processing. The adjusted p-value for the false discovery rate (q-value) was calculated using R and BioConductor.²⁰¹⁻²⁰² ROC analysis and permutation test were performed using biomarker discovery in Metaboanalyst with random forest algorithm. Pathway analysis employed global test as pathway enrichment analysis method and relative-betweenness centrality as pathway topology analysis method in Metaboanalyst.

6.3. Results and Discussion

6.3.1. Comprehensive Metabolomic Analysis of Plasma

The data from four channels of submetabolome analysis were combined to generate comprehensive metabolome profiling. A total of 11691 ¹²C-/¹³C-peak pairs were commonly detected in at least 80% of the samples within one group. An average of 11060 ± 325 peak pairs detected per run for experimental duplicate for each sample. Among them, 2902 peak pairs were detected from amine/phenol labeling channel, 2983

from carboxyl labeling channel, 3736 from carbonyl labeling channel and 2070 from hydroxyl labeling channel. A labeled standard library consisting of 664 unique human endogenous metabolites, including 315 amines/phenols, 187 carboxyls, 77 carbonyls and 85 hydroxyls, was used to perform positive identification. In total of 173 metabolites were identified by matching accurate mass and retention time with the labeled standard library (mass error < 10 ppm and retention time error < 30 s, Table 6.1). The remaining peak pairs were searched based on accurate mass matches, against the MyCompoundID (MCID) library composed of 8,021 known human endogenous metabolites and their predicted metabolic products from one metabolic reaction (375,809 compounds). 2868 and 5916 peak pairs were matched in the zero- and one-reaction libraries, respectively. Thus about 77% of peak pairs could be either identified or matched.

Table 6.1 List of positively identified metabolites based on accurate mass and retention time match to the labeled standard library.

HMDB.No.	Name	Accurate mass	mz_light	library RT (min)
HMDB00001	1-Methylhistidine	169.0851	403.1434	2.17
HMDB00157	Hypoxanthine + H2O	136.0385	388.1098	2.12
HMDB00517	Arginine	174.1117	408.1700	2.44
HMDB00670	Homo-arginine	188.1273	422.1856	3.00
HMDB00168	Asparagine	132.0535	366.1118	3.00
HMDB00641	Glutamine	146.0691	380.1275	3.32
HMDB00904	Citrulline	175.0957	409.1540	3.74
HMDB02005	Methionine Sulfoxide	165.0460	399.1043	3.72
HMDB00719	Homoserine	119.0582	353.1166	4.05
HMDB02005_2	Methionine Sulfoxide - Isomer	165.0460	399.1043	4.20
HMDB00679	Homocitrulline	189.1113	423.1697	4.47
HMDB00187	Serine	105.0426	339.1009	4.40
HMDB00148	Glutamic Acid	147.0532	381.1115	5.05
HMDB00725	Trans-4-HydroxyProline	131.0582	365.1166	5.17
HMDB00191	Aspartic Acid	133.0375	367.0958	5.16
HMDB00206	N6-AcetyLysine	188.1161	422.1744	5.71
HMDB00167	Threonine	119.0582	353.1166	5.79

HMDB00149	Ethanolamine	61.0528	295.1111	6.00
HMDB00123	Glycine	75.0320	309.0903	6.59
HMDB00446	N-Alpha-acetyllysine	188.1161	422.1744	6.79
HMDB00056	Beta-Alanine	89.0477	323.1060	7.24
HMDB00296	Uridine	244.0695	478.1279	7.84
HMDB00161	Alanine	89.0477	323.1060	7.57
HMDB00210	Pantothenic acid	219.1107	453.1690	8.37
HMDB00884_2	Ribothymidine - Isomer	258.0852	492.1435	8.54
HMDB00296_2	Uridine - H2O	244.0695	460.1173	8.67
HMDB00157_2	Hypoxanthine - multi-tags	136.0385	370.0968	8.73
HMDB00292	Xanthine	152.0334	386.0917	8.95
HMDB03355	5-Aminopentanoic acid	117.0790	351.1373	8.68
HMDB00452	Alpha-aminobutyric acid	103.0633	337.1216	9.13
HMDB00469	5-Hydroxymethyluracil	142.0378	376.0962	8.87
HMDB00884_3	Ribothymidine - H2O	258.0852	474.1329	9.39
HMDB01906	2-Aminoisobutyric acid	103.0633	337.1216	8.91
HMDB00157_3	Hypoxanthine - Isomer	136.0385	370.0968	9.65
HMDB00148_2	Glutamic Acid - H2O	147.0532	363.1009	9.46
HMDB02108	Methylcysteine	135.0354	369.0937	9.37
HMDB00164	Methylamine	31.0422	265.1005	9.82
HMDB00897	7-Methylguanine	165.0651	399.1234	10.32
HMDB00162	Proline	115.0633	349.1216	10.18
HMDB00883	Valine	117.0790	351.1373	10.81
HMDB00696	Methionine	149.0510	383.1094	10.89
HMDB00684	Kynurenine	208.0848	442.1431	11.44
HMDB00929	Tryptophan	204.0899	438.1482	11.44
HMDB01392	p-Aminobenzoic acid	137.0477	371.1060	11.52
HMDB00159	Phenylalanine	165.0790	399.1373	12.74
HMDB00291	Vanillylmandelic acid	198.0528	432.1111	12.81
HMDB00172	Isoleucine	131.0946	365.1529	13.06
HMDB00750	3-Hydroxymandelic acid	168.0423	402.1006	12.94
HMDB00687	leucine	131.0946	365.1529	13.36
HMDB00450	5-Hydroxylysine	162.1004	315.1085	13.88
HMDB00755	Hydroxyphenyllactici acid	182.0579	416.1162	14.39
HMDB00087	Dimethylamine	45.0578	279.1162	15.07
HMDB00763	5-Hydroxyindoleacetic acid	191.0582	425.1166	15.09
HMDB29087	TryptophyLeucine	317.1739	551.2323	14.60
HMDB01889	Theophylline	180.0647	414.1230	15.42
HMDB01895	Salicylic acid	138.0317	372.0900	15.62
HMDB02362	2,4-Diaminobutyric acid	118.0742	293.0954	15.80
HMDB94662	PhenyLeucine	278.1631	512.2214	15.90
HMDB29090	TryptophyPhenylalanine	351.1583	585.2166	15.36
HMDB03911_2	3-Aminoisobutanoic acid - H2O	103.0633	319.1110	16.29
HMDB01868	5-Methoxysalicylic acid	168.0423	402.1006	16.38
HMDB00118	Homovanillic acid	182.0579	416.1162	16.51
HMDB00020	p-Hydroxyphenylacetic acid	152.0473	386.1057	16.91
HMDB00214	Ornithine	132.0899	300.1033	16.58

HMDB00440	3-Hydroxyphenylacetic acid	152.0473	386.1057	16.91
HMDB00500	4-Hydroxybenzoic acid	138.0317	372.0900	17.57
HMDB00484	Vanillic acid	168.0423	402.1006	17.34
HMDB00182	Lysine	146.1055	307.1111	17.47
HMDB02199	Desaminotyrosine	166.0630	400.1213	18.04
HMDB00177	Histidine	155.0695	389.1278	18.09
HMDB00991	2-aminooctanoic acid	159.1259	393.1842	19.20
HMDB03320	Indole-3-carboxylic acid	161.0477	395.1060	19.27
HMDB00734	Indoleacrylic acid	187.0633	421.1216	20.61
HMDB29098	TyrosyAlanine	252.1110	360.1138	20.86
HMDB00750_2	3-Hydroxymandelic acid - COOH	168.0423	356.0951	21.64
HMDB00158	Tyrosine	181.0739	324.5953	22.65
HMDB00228	Phenol	94.0419	328.1002	23.16
HMDB01232	4-Nitrophenol	139.0269	373.0853	23.45
HMDB28941	LeucyTyrosine	294.1580	381.1373	23.98
HMDB01858	p-Cresol	108.0575	342.1158	24.54
HMDB00259	Serotonin	176.0950	322.1058	24.65
HMDB00130	Homogentisic acid	168.0423	318.0794	24.84
HMDB29306	4-Ethylphenol	122.0732	356.1315	25.63
HMDB00186	Lactose	342.1162	590.2014	5.86
HMDB00122	Glucose	180.0634	428.1486	6.08
HMDB00660	Fructose	180.0634	428.1486	6.44
HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59
HMDB00422	2-Methylglutaric acid	146.0579	394.1431	8.76
HMDB01259	Succinic acid semialdehyde	102.0317	350.1169	9.34
HMDB00720	Levulinic acid	116.0473	278.1387	10.43
HMDB00208	Oxoglutaric acid	146.0215	394.1067	10.47
HMDB00714	Hippuric Acid	179.0582	427.1434	10.56
HMDB01426	Formaldehyde	30.0106	278.0958	10.66
HMDB01659	Acetone	58.0419	306.1271	12.01
HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39
HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39
HMDB01864	2-Ketohexanoic acid	130.0630	378.1482	13.79
HMDB00474	2-butanone	72.0575	320.1427	14.24
HMDB03543	Butanal	72.0575	320.1427	14.83
HMDB00063	Cortisol	362.2093	610.2945	15.19
HMDB03315	Cyclohexanone	98.0732	346.1584	15.29
HMDB06115	Benzaldehyde	106.0419	354.1271	16.23
HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15
HMDB00077	Dehydroepiandrosterone	288.2089	536.2941	17.88
HMDB05994	Hexanal	100.0888	348.1740	17.91
HMDB00467	Nutriacholic acid	390.2770	638.3622	18.03
HMDB03407	Diacetyl	86.0368	334.1220	13.97
HMDB00628	Epitestosterone	288.2089	536.2941	20.20
HMDB01140	Octanal	128.1201	376.2053	20.72
HMDB00503	7?-hydroxy-3-oxo-5?cholanoic acid	390.2770	638.3622	20.57
HMDB00031	Androsterone	290.2246	538.3098	20.71

HMDB11623	Decanal	156.1514	404.2366	23.26
HMDB00700	Hydroxypropionic acid	90.0317	252.1230	7.02
HMDB00190	L-Lactic acid	90.0317	252.1230	7.72
HMDB00357	3-Hydroxybutyric acid	104.0473	266.1387	8.56
HMDB00142	Formic acid	46.0055	208.0968	9.26
HMDB00008	2-Hydroxybutyric acid	104.0473	266.1387	9.83
HMDB00042	Acetic acid	60.0211	222.1125	10.00
HMDB00407	2-Hydroxy-3-methylbutyric acid	118.0630	280.1543	11.86
HMDB00237	Propionic acid	74.0368	236.1281	12.41
HMDB00423	3,4-Dihydroxyhydrocinnamic acid	182.0579	344.1492	11.72
HMDB00005	2-Ketobutyric acid	102.0317	264.1230	13.22
HMDB00617	Pyromucic Acid	112.0160	274.1074	13.51
HMDB01624	2-Hydroxycaproic acid	132.0786	294.1700	13.72
HMDB00039	Butyric acid	88.0524	250.1438	14.94
HMDB00426	Citramalic acid	148.0372	236.1099	15.05
HMDB00072	cis-Aconitic acid	174.0164	249.0996	15.38
HMDB00019	a-Ketoisovaleric acid	116.0473	278.1387	16.06
HMDB00254	Succinic acid	118.0266	441.2020	16.62
HMDB00209	Phenylacetic acid	136.0524	298.1438	17.44
HMDB01101	p-Anisic acid	152.0473	314.1387	17.51
HMDB01870	Benzoic acid	122.0368	284.1281	17.53
HMDB00718	Isovaleric acid	102.0681	264.1594	17.53
HMDB00892	Valeric acid	102.0681	264.1594	17.64
HMDB00661	Glutaric acid	132.0423	228.1125	18.10
HMDB00749	Mesaconic acid	130.0266	227.1046	17.98
HMDB00695	Methyloxovaleric acid	130.0630	292.1543	18.34
HMDB00689	Isocaproic acid	116.0837	278.1751	20.55
HMDB02097	4-Ethylbenzoic acid	150.0681	312.1594	23.09
HMDB00505	Allocholic acid	408.2876	570.3789	22.82
HMDB00666	Heptanoic acid	130.0994	292.1907	24.42
HMDB00733	Hyodeoxycholic acid	392.2927	554.3840	24.89
HMDB00392	2-Octenoic acid	142.0994	304.1907	26.21
HMDB00784	Azelaic acid	188.1049	256.1438	26.14
HMDB04586	Perillic acid	166.0994	328.1907	26.51
HMDB01877	Valproic acid	144.1150	306.2064	27.45
HMDB00792	Sebacic acid	202.1205	263.1516	28.19
HMDB00847	Nonanoic acid	158.1307	320.2220	30.10
HMDB00888	Undecanedioic acid	216.1362	270.1594	29.87
HMDB00933	Traumatic acid	228.1362	276.1594	30.46
HMDB00511	Capric acid	172.1463	334.2377	31.54
HMDB02327	1,11-Undecanedicarboxylic acid	244.1675	284.1751	31.89
HMDB00529	5-Dodecenoic acid	198.1620	360.2533	32.54
HMDB00638	Dodecanoic acid	200.1776	362.2690	33.86
HMDB00672	Hexadecanedioic acid	286.2144	305.1985	34.30
HMDB01852	Retinoic acid	300.2089	462.3003	36.67
HMDB00806	Myristic acid	228.2089	390.3003	36.63
HMDB00782	Octadecanedioic acid	314.2457	319.2142	36.50

HMDB00673	Linoleic acid	280.2402	442.3316	37.65
HMDB00826	Pentadecanoic acid	242.2246	404.3159	38.50
HMDB02925	Eicosatrienoic acid	306.2559	468.3472	38.38
HMDB00131	Glycerol	92.0473	326.1056	7.31
HMDB00108	Ethanol	46.0417	280.1000	13.13
HMDB00820	1-Propanol	60.0575	294.1158	14.54
HMDB41607	2-Phenoxyethanol	138.0681	372.1264	15.50
HMDB31456	tert-Butyl alcohol	74.0732	308.1315	16.03
HMDB31527	2-Methyl-1-butanol	88.0888	322.1471	17.38
HMDB01881	1,2-Propanediol	76.0524	310.1107	9.93
HMDB31334	3-Chloro-1,2-propanediol	110.0135	289.0650	19.13
HMDB31479	1-Heptanol	116.1201	350.1784	20.59
HMDB01183	1-Octanol	130.1358	364.1941	22.06
HMDB35094	Citronellol	156.1514	390.2097	22.30
HMDB11624	Decyl alcohol	158.1671	392.2254	24.60

6.3.2. Metabolomic Comparisons of All Groups

Principle component analysis (PCA) was applied to visualize the whole dataset. Figure 6.1A shows the PCA scores plots for all groups, including QC samples. The QC samples (in cyan) were clustered close together, indicating stable instrument performance throughout the study. The partial least squares-discriminant analysis (PLS-DA) was then used to investigate the metabolome changes between groups. Two dimensional scores plot and three dimensional scores plot are shown in Figure 6.2B and 6.2C, respectively. The model performance indicators (i.e. R^2 and Q^2) are provided in the corresponding scores plots. Clear separations can be observed among AD group (in red), CAA group (in green) and control group (in blue).

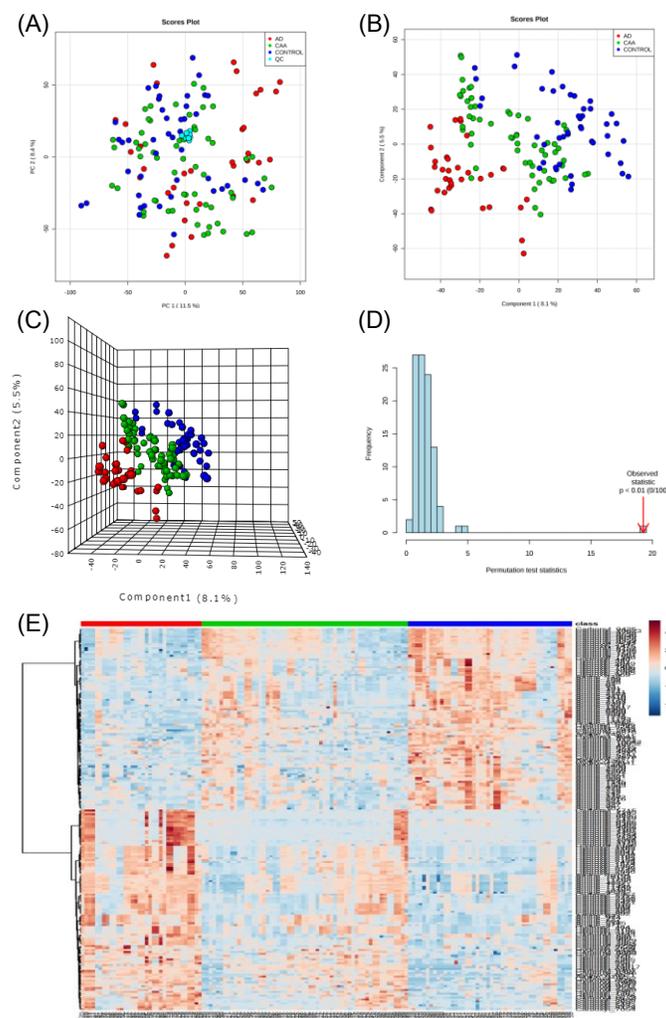


Figure 6.1 (A) PCA, (B) 2D PLS-DA and (C) 3D PLS-DA scores plots of metabolomics dataset from three groups, including Alzheimer’s disease (AD, in red). cerebral amyloid angiopathy (CAA, in green) and control group (in blue). QC samples were shown in PCA as cyan dots. R^2 and Q^2 values given by cross-validation for PLS-DA are: 0.998 and 0.902, respectively. (D) Permutation test of PLS-DA model for the dataset. (E) Clustered heatmap of relative intensity comparison of important peak pairs from three groups.

It is well known that if the the number of variables used for statistical analysis (11619 peak pairs in this study) is much larger than the number of samples (138 runs from 69 samples with experimental duplicate), PLS-DA as a supervised classification

model trends to overfit the data. In this situation, the discrimination or separation among groups may cause by chance. Thus permutation test was employed to check whether the model had this issue. In permutation test, the group labels of the samples are reassigned randomly to build a new classifier. The performance of the new group labels is evaluated and compared with the results from original sample class labels. The process will repeat many times to generate the permutation plot. If the performance of original label is better than any other permuted cases, it confirms the discrimination is not by chance. In this study, permutation test was carried out in Metaboanalyst, which shows that the separation among groups was valid (Figure 6.2D)

Besides, a trend of metabolomic change from control group to CAA group, then to AD group can be observed in PLS-DA scores plot. This trend is confirmed in heatmap, which visualizes the relative intensity change of metabolites among three groups (Figure 6.2E). For each metabolite, deeper red represents higher intensity and deeper blue represents lower intensity.

6.3.3. Comparative Metabolome Analysis of AD and Control

Statistical analysis of pair-wise comparison was carried out. Figure 6.2A shows the PLS-DA scores plot for the comparison between AD and controls. A clear distinction between two groups can be observed. However, this model did not pass the permutation test, which was attributed to the large number of variables and small sample size. We further investigated using intensity filtered peak pairs (absolute intensity less than 15000 were removed) to re-build the model since these peak pairs may have relatively larger quantification error. In this way the number of variables for statistical analysis was reduced to 2766 from 11619. The scores plot of new model is very similar to the original

one (data not shown) and the results of permutation test indicate that the new model is valid (Figure 6.2B). Since PLS-DA model is susceptible to the over-fitting problem, we decided to use random forest algorithm for classifying different groups.

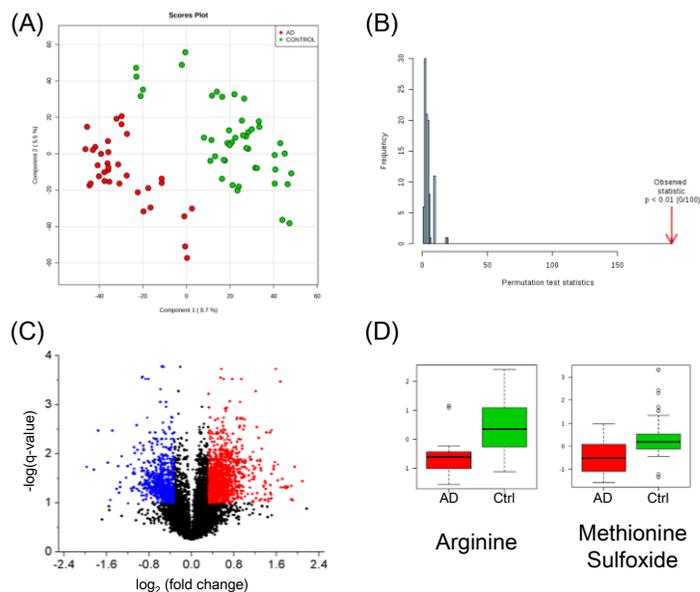


Figure 6.2 (A) PLS-DA scores plot of dataset from AD (in red) and control group (in green). $R^2 = 0.990$; $Q^2 = 0.897$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between AD and control group with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in AD and control group.

The volcano plot as univariate statistical analysis was used to determine the significant metabolites that have different intensities in two groups. It plots the fold-change (FC) of each metabolite against the q-value, which is the false discovery rate analogue of the p-value from a t test. The criterion we used was $FC > 1.25$ or < 0.8 and $q\text{-value} < 0.1$ (corresponding to $p\text{-value} < 0.0456$). The result (Figure 6.2C) shows that 1296

peak pairs with FC > 1.25, q-value <0.1 (in red) and 687 peak pairs with FC < 0.8, q-value < 0.1 (in blue). Among them 27 metabolites were positively identified by labeled standard library (Table 6.2).

Table 6.2 List of positively identified significant metabolites in comparison of AD and control

HMDB #	Compound	VIP score	fold change	p-value	q-value
HMDB02005_2	Methionine Sulfoxide - Isomer	3.0946	0.67906	2.567E-08	0.000168
HMDB02005	Methionine Sulfoxide	3.0051	0.69229	7.751E-08	0.0001691
HMDB00517	Arginine	2.4953	0.76978	1.529E-05	0.0022745
HMDB02199	Desaminotyrosine	1.9189	3.0919	0.0011892	0.0165462
HMDB00357	3-Hydroxybutyric acid	1.8683	0.78512	0.0016317	0.0194916
HMDB00661	Glutaric acid	1.8357	1.5861	0.0019902	0.0209898
HMDB00157_3	Hypoxanthine - Isomer	1.7952	1.3082	0.0025352	0.0236746
HMDB00426	Citramalic acid	1.7434	0.573	0.0034223	0.0268543
HMDB01858	p-Cresol	1.6898	1.5237	0.0046257	0.0314115
HMDB01183	1-Octanol	1.6549	1.3074	0.0055959	0.0347881
HMDB00452	Alpha-aminobutyric acid	1.6158	1.4136	0.0068952	0.0380905
HMDB00157_2	Hypoxanthine - multi-tags	1.6128	1.4674	0.0070039	0.0384519
HMDB00695	Methyloxyvaleric acid	1.6022	0.76896	0.0074062	0.0396138
HMDB03911_2	3-Aminoisobutanoic acid - H2O	1.5638	1.4996	0.0090289	0.0435964
HMDB00108	Ethanol	1.5498	1.3856	0.0096951	0.0449794
HMDB00164	Methylamine	1.5323	1.2899	0.01059	0.0469358
HMDB00072	cis-Aconitic acid	1.5104	1.4693	0.011804	0.0488132
HMDB28941	LeucyTyrosine	1.5015	0.5867	0.012331	0.0497357
HMDB00118	Homovanillic acid	1.4564	1.35	0.01534	0.0554492
HMDB00148	Glutamic Acid	1.4289	0.74393	0.017463	0.0590476
HMDB00500	4-Hydroxybenzoic acid	1.3758	1.672	0.022293	0.067334
HMDB00019	a-Ketoisovaleric acid	1.3496	0.74367	0.025068	0.0716593
HMDB00291	Vanillylmandelic acid	1.3436	1.3368	0.025745	0.0728943
HMDB00020	3-Hydroxyphenylacetic acid	1.3427	1.327	0.025847	0.0730567
HMDB02925	Eicosatrienoic acid	1.2898	1.2812	0.032508	0.0827386
HMDB29098	TyrosyAlanine	1.2726	0.77618	0.03497	0.0862748
HMDB29306	4-Ethylphenol	1.2203	1.3038	0.043419	0.0972721

Receiver operating characteristic (ROC) curve was generated to evaluate whether a combination of metabolites could be used to differentiate AD from control group. Random forest method was used to build the classification model in MetaboAnalysit 4.0

(<http://www.metaboanalyst.ca>). VIP score in PLS-DA model, FC, q-value and identification results for each peak pair were considered as the factors to select the panel of metabolites. The 27 positively identified metabolites with $FC < 1.25$ or > 0.8 and q-value < 0.1 were considered as the best candidates for classification (Table 6.2) since they meet the statistical criteria and were high-confidence identified. They were ranked according to their VIP scores from PLS-DA model. The top five were first chosen to build the ROC curve, including methionine sulfoxide, arginine, desaminotyrosine, 3-hydroxybutyric acid and glutaric acid. Here we excluded the entry “methionine sulfoxide - isomer”, which was created in the labeling as the isomer of methionine sulfoxide. It has nearly the same peak ratios as methionine sulfoxide for all samples (data not shown), confirming that the two peak pairs originate from one metabolite. Using the selected panel of metabolites, the area-under-the-curve (AUC) value was 0.957 (95% confidence interval: 0.865-1) with sensitivity of 88.2% and specificity of 89.1%, indicating good separation power of the metabolite panel for the AD and healthy control (Figure 6.3A). Box plots of two metabolites in the panel are shown in Figure 6.2D as examples of metabolites that have significant differences between two groups. The permutation test was also performed for ROC curve to test whether there was over-fitting of the model. Figure 6.3B shows the results of permutation test. All of the ROC curves that originated from randomized group assignment are near diagonal and their AUC are less than the AUC of the original curve, suggesting no over-fitting issue in the ROC curve.

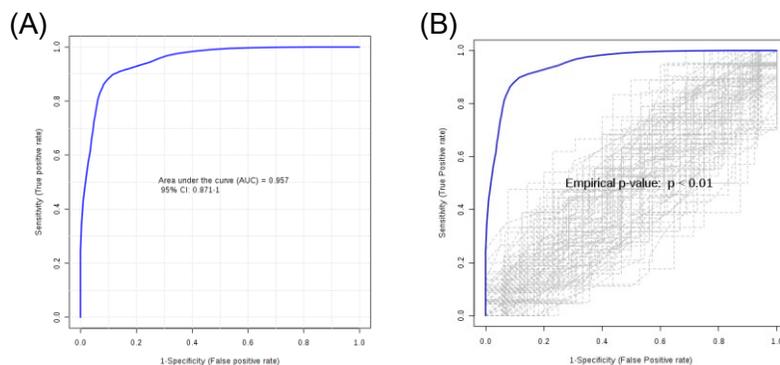


Figure 6.3 (A) ROC curve generated by the random forest model to differentiate AD and control using 5 metabolite biomarker candidates: methionine sulfoxide, arginine, desaminotyrosine, 3-hydroxybutyric acid and glutaric acid. (B) Permutation test result for the ROC curve of AD vs. control.

6.3.4. Comparative Metabolome Analysis of CAA and Control

Comparison of data generated from CAA group and control group were performed using similar approach. Figure 6.4A shows the PLS-DA scores plot of the dataset. Separation between two groups could be observed. The results of permutation test of the model built by intensity filtered peak pairs is shown in Figure 6.4B. The same criteria was used to make the volcano plot ($FC > 1.25$ or < 0.8 and $q\text{-value} < 0.1$, corresponding to $p\text{-value} < 0.0283$), 507 peak pairs were up-regulated and 565 peak pairs were down-regulated in CAA patients compared with healthy control (Figure 6.4C). 21 metabolites were positively identified among 1072 significant peak pairs, including 9 up-regulated and 12 down-regulated (Table 6.3). Box plots of two examples are shown in Figure 6.4D.

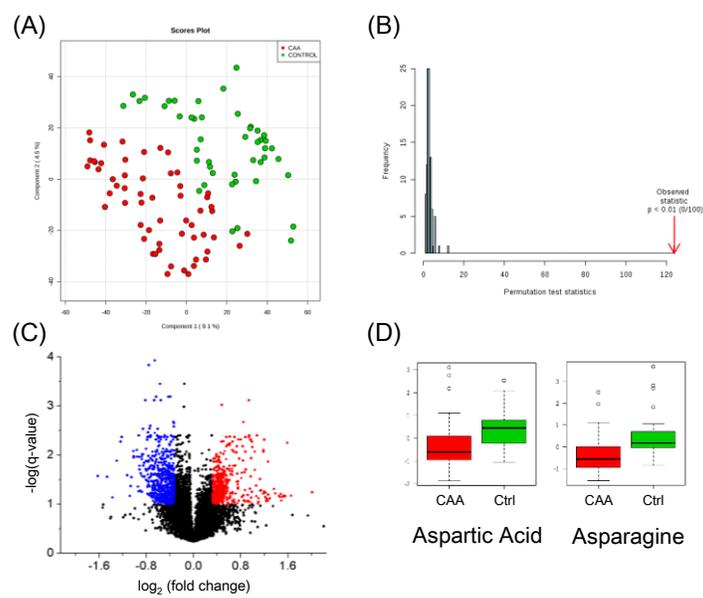


Figure 6.4 (A) PLS-DA scores plot of dataset from CAA (in red) and control group (in green). $R^2 = 0.978$; $Q^2 = 0.760$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between CAA and control group with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in CAA and control group.

Table 6.3 List of positively identified significant metabolites in comparison of CAA and control.

HMDB #	Compound	VIP score	fold change	p-value	q-value
HMDB02005_2	Methionine Sulfoxide - Isomer	3.4029	0.72148	1.758E-08	8.644E-05
HMDB02005	Methionine Sulfoxide	3.3691	0.72861	2.566E-08	8.644E-05
HMDB00517	Arginine	3.052	0.76286	6.685E-07	0.0006435
HMDB00168	Asparagine	2.7763	0.79401	7.783E-06	0.0026222
HMDB00191	Aspartic Acid	2.4798	0.75482	7.752E-05	0.0059358
HMDB00148	Glutamic Acid	2.4621	0.6704	8.798E-05	0.0062997
HMDB00087	Dimethylamine	2.2233	1.3568	0.0004389	0.0138185
HMDB00148_2	Glutamic Acid - H2O	2.0778	0.71956	0.0010665	0.0214524
HMDB03911_2	3-Aminoisobutanoic acid - H2O	2.0217	1.5802	0.0014757	0.0252383
HMDB00164	Methylamine	1.9994	1.2769	0.001675	0.0264535

HMDB00118	Homovanillic acid	1.9727	1.4076	0.0019455	0.0276574
HMDB01858	p-Cresol	1.9431	1.394	0.0022904	0.0298117
HMDB00452	Alpha-aminobutyric acid	1.8544	1.4152	0.0036786	0.0375576
HMDB02108	Methylcysteine	1.7916	1.3471	0.0050784	0.0440046
HMDB01051	Glyceraldehyde	1.7123	0.71669	0.0075068	0.0527412
HMDB02199	Desaminotyrosine	1.6729	2.3298	0.0090577	0.0579192
HMDB00617	Pyromucic Acid	1.6447	0.58278	0.010339	0.061017
HMDB00426	Citramalic acid	1.6312	0.69776	0.011002	0.0622584
HMDB28941	LeucyTyrosine	1.6102	0.67946	0.012113	0.0653504
HMDB00157	Hypoxanthine + H ₂ O	1.4704	0.74145	0.022309	0.0884801
HMDB00423	3,4-Dihydroxyhydrocinnamic acid	1.4204	1.3092	0.027431	0.0980808
HMDB02005_2	Methionine Sulfoxide - Isomer	3.4029	0.72148	1.758E-08	8.644E-05

Figure 6.5A shows the ROC curve using five metabolites as potential biomarker panel: methionine sulfoxide, arginine, asparagine, aspartic acid and glutamic acid, which were picked using the same approach. The ROC curve of the panel produced an AUC value of 0.873, which is within the range of 0.759-0.972 at the 95% confidence interval (CI). The sensitivity and specificity of this panel was 84.5% and 80.4%, respectively. Through the permutation test, no over-fitting issue was observed (Figure 6.5B).

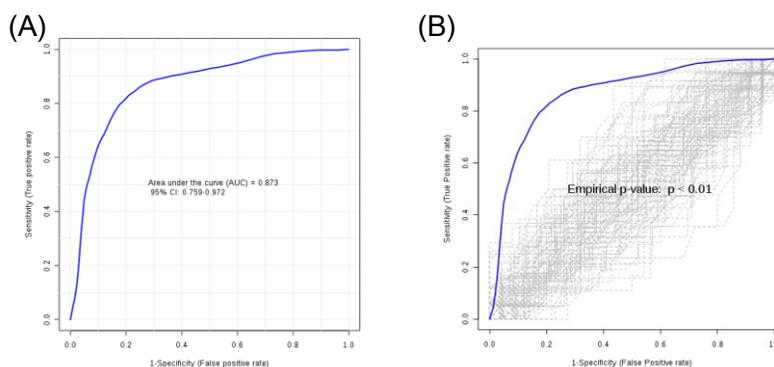


Figure 6.5 (A) ROC curve generated by the random forest model to differentiate CAA and control using 5 metabolite biomarker candidates: methionine sulfoxide, arginine, asparagine, aspartic acid and glutamic acid. (B) Permutation test result for the ROC curve of CAA vs. control.

6.3.5. Comparative Metabolome Analysis of CAA and AD

Metabolomics dataset from two sister disease groups were compared. Figure 6.6A shows the scores plot from PLS-DA model and Figure 6.6B shows the permutation test of model built with intensity filtered peak pairs. Volcano shows 164 peak pairs with $FC > 1.25$, $q\text{-value} < 0.1$ (equivalent to $p\text{-value} < 0.029$, in red) and 697 peak pairs with $FC < 0.8$, $q\text{-value} < 0.1$ (in blue). Among them, 6 peak pairs can be positively identified against the labeled standard library (Table 6.4) and 252 can be putatively identified against the MCID 0-reaction library (data not shown).

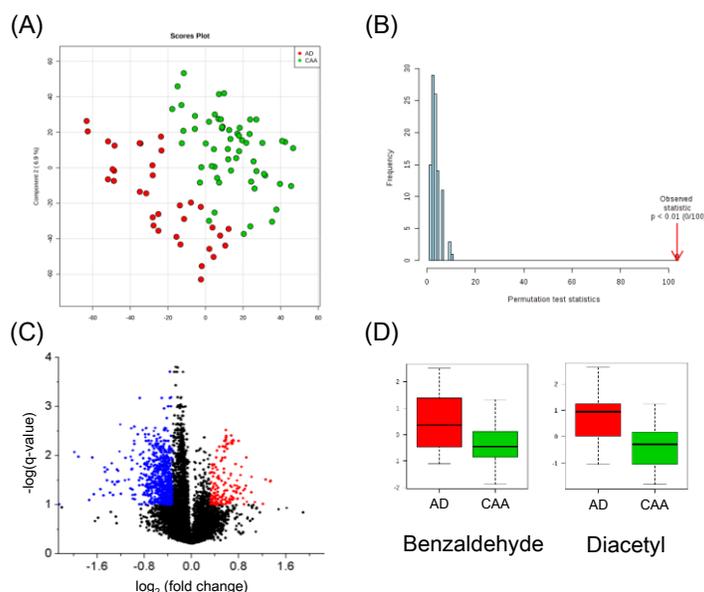


Figure 6.6 (A) PLS-DA scores plot of dataset from CAA (in green) and AD (in red). $R^2 = 0.981$; $Q^2 = 0.802$. (B) Permutation test of PLS-DA model built by intensity filtered peak pairs. (C) Volcano plot of the comparison between CAA and AD with red dots representing variables with $FC > 1.25$, $q < 0.1$ and blue dots representing variables with $FC < 0.8$, $q < 0.1$. (D) Examples of metabolites that have significant differences of intensity in CAA and AD group.

Table 6.4 List of positively identified significant metabolites in comparison of CAA and AD.

HMDB #	Compound	VIP score	fold change	p-value	q-value
HMDB06115	Benzaldehyde	2.5397	0.74404	2.527E-05	0.0027932
HMDB00157_2	Hypoxanthine - multi-tags	2.2841	0.60996	0.0001763	0.0058674
HMDB01051	Glyceraldehyde	1.9475	0.6199	0.0015633	0.0168454
HMDB00500	4-Hydroxybenzoic acid	1.672	0.58315	0.0070402	0.0419516
HMDB00661	Glutaric acid	1.493	0.71791	0.016559	0.0719122
HMDB00157	Hypoxanthine + H2O	1.4027	0.68139	0.024639	0.0907008

To differentiate CAA and AD, five metabolites that were positively identified with higher VIP score were used to build the ROC curve, including benzaldehyde, hypoxanthine, glyceraldehyde, 4-hydroxybenzoic acid and glutaric acid (Figure 6.7A). An AUC value of 0.894 with 95% CI from 0.701-0.986 was generated. The panel provides discrimination with sensitivity of 79.4% and specificity of 91.4%. Figure 6.7B shows the results of permutation test, indicating no significant over-fitting for building the ROC curve.

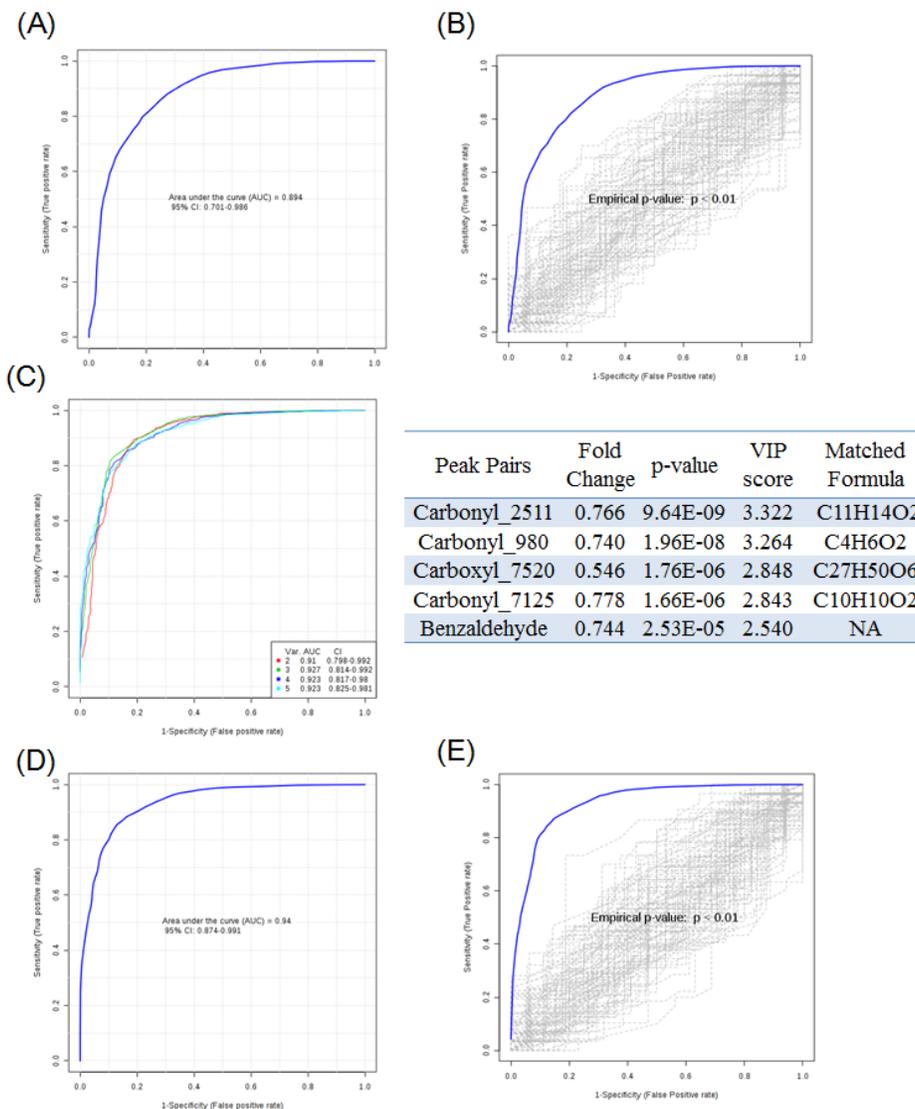


Figure 6.7 (A) ROC curve generated by the random forest model to differentiate CAA and AD using five positively identified candidates: benzaldehyde, hypoxanthine, glyceraldehyde, 4-hydroxybenzoic acid and glutaric acid. (B) Permutation test result for the ROC curve of CAA vs. AD using positively identified metabolites. (C) A series of ROC curves generated with different numbers of putative identified metabolites in the list on the right. (D) ROC curve generated using top three putatively identified metabolites and (E) its permutation test results.

We noticed that only six significant metabolites were positively identified, which limited the selection for biomarker panel. Thereby we incorporate peak pairs that were putatively identified by matching accurate mass with database into the biomarker panel. Although their identities were not definitive, putative match results still provided information for further annotation using authentic standards. In the study, four putatively matched peak pairs were selected as biomarker candidates using the same ranking method (Figure 6.7C). Together with benzaldehyde, the five candidates were subjected to ROC curve analysis. Figure 6.7C shows a series of ROC curves generated with different numbers of metabolites. For instance, the green line represents using the top three variables, which already reaches the highest discrimination power (AUC = 0.927 with 95% of CI from 0.814-0.992). We then only input the top three candidates (Carbonyl_2511, Carbonyl_980 and Carboxyl_7520) to build the ROC curve (Figure 6.7D). The performance of discrimination (AUC = 0.94, 95% of CI: 0.874-0.991) was significantly improved compared with using five positively identified metabolites as panel. The sensitivity and specificity were 85.3% and 87.9%, respectively. No over-fitting issues of the ROC results were detected by validation of permutation test (Figure 6.7E).

6.3.6. Biological Pathway Analysis

A metabolic pathway analysis was performed using all positively identified metabolites and their relative amount to understand the two diseases. It rendered the information of p-values from pathway enrichment analysis and pathway impact values from pathway topology analysis. The pathways that are relevant to the disease should be characterized by low p-value and high pathway impact value.

Figure 6.8A and Figure 6.8B show the overview of all matched pathway in AD vs. control and CAA vs. control, respectively. It shows very similar results for both comparisons. 1) Alanine, aspartate and glutamate metabolism, 2) arginine and proline metabolism and 3) glycine, serine and threonine metabolism are commonly affected pathways, indicating the similarities of two diseases in terms of metabolic change compared with healthy control. Particularly, we observed down-regulation of several amino acids in both AD group and CAA group, including arginine, serine, threonine and asparagine. Aspartic acid was significant in CAA group but not in AD group. The box plots of relative amount of these metabolites are shown in Figure 6.8D. Similar findings have been reported before in biomarker discovery study for Alzheimer's disease.²⁰³ Charged amino acids appear to be decreased in plasma of AD patients,²⁰⁴⁻²⁰⁵ which is in agreement with our results.

Not only similarities were observed between AD and CAA, a pathway, butanoate metabolism, that has significant differences between two diseases was found (Figure 6.8C, 9A). From the upstream metabolites, oxoglutaric acid and butanal, to downstream metabolites including succinic acid semialdehyde, diacetyl and butyric acid, all metabolites show decreased levels in CAA group compared with AD group (Figure 6.9B). Although the fold changes of these metabolites are relatively small (0.834 for diacetyl and around 0.9 for others), the differences are proved to be significant by t-test.

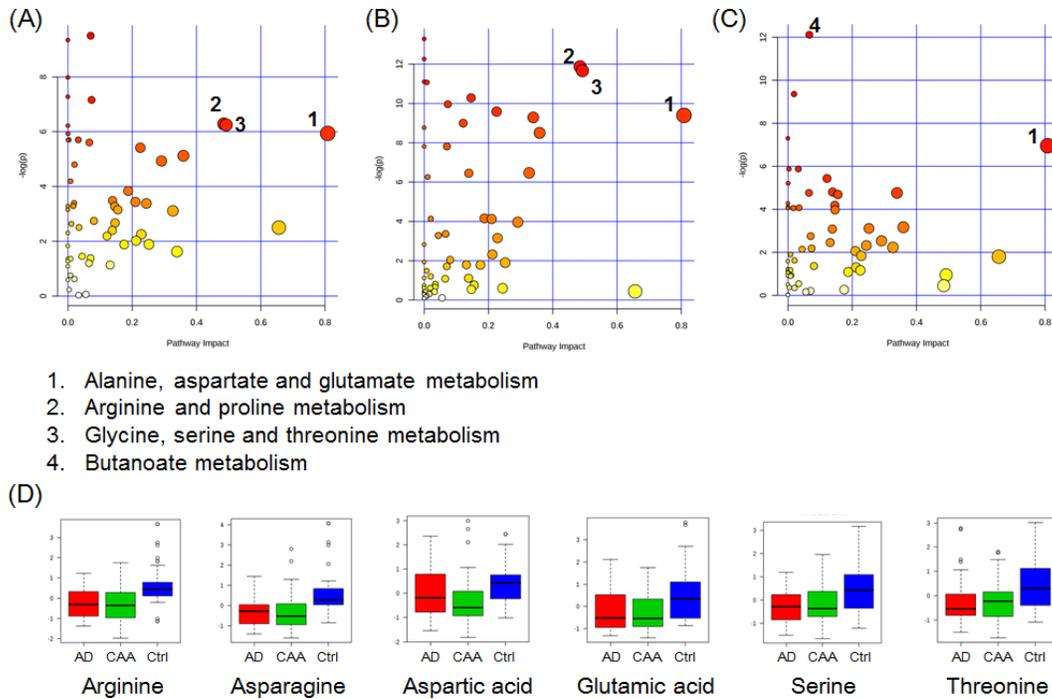


Figure 6.8 Overview of pathway analysis in (A) AD vs. control comparison; (B) CAA vs. control comparison; (C) CAA vs. AD comparison. Each circle represents a matched pathway. The color of circle is based on its p value and the radius is determined by its pathway impact values. Four significant pathways are labeled. (D) Box plots of six significantly altered amino acids. AD: in red; CAA: in green; control group: in blue.

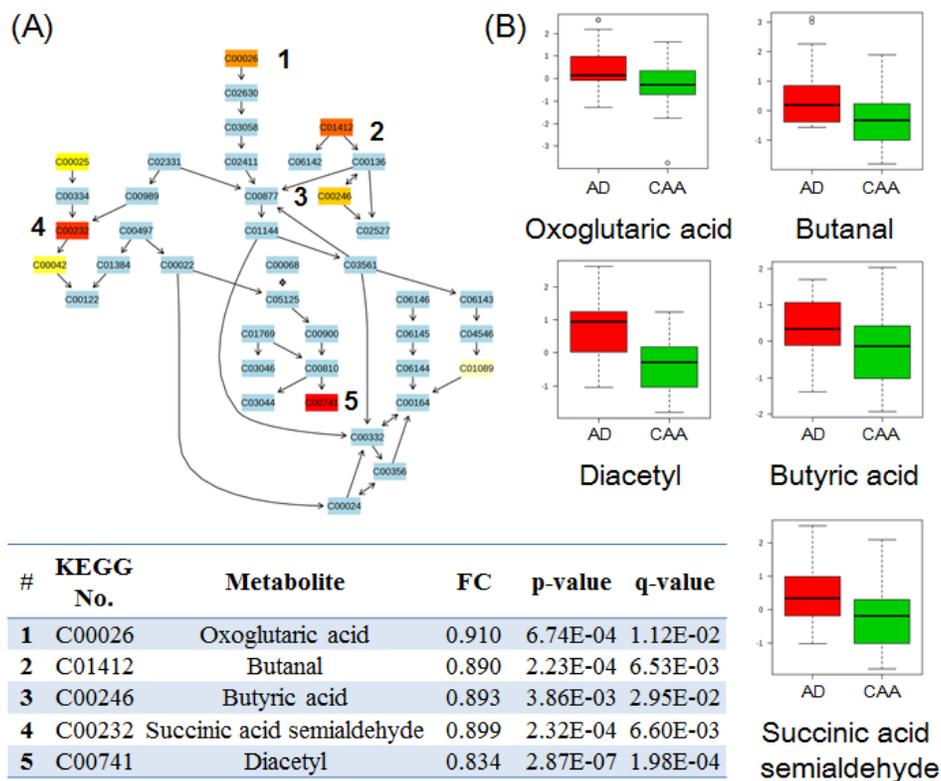


Figure 6.9 (A) Schematic view of butanoate metabolism. Compounds were represented by KEGG numbers. Matched metabolites were shown as nodes with varied heat map colors based on p-values (red indicates lower p-value). The information of matched metabolites was listed in the table below. (B) Box plots of five significant metabolites in butanoate metabolism. AD: in red; CAA: in green.

6.4. Conclusion

In this study, multichannel chemical isotope labeling LC-MS technique was applied in finding biomarkers for Alzheimer's disease (AD) and cerebral amyloid angiopathy (CAA). It is the first time of using this newly developed method for biomarker discovery. The plasma metabolome profiling results demonstrated very high

detectability of metabolites for metabolomics. Using the technique, we first performed metabolomic comparisons between AD, CAA and healthy controls. Statistical analysis suggested that CAA, AD, and controls can be discriminated by metabolomic profile. Then pair-wise comparisons were then carried out to find biomarker candidates for the two diseases and between diseases. For each pair-wise comparison, five positively identified metabolites were selected based on their relative amount to form a panel for differentiating groups. The performance of biomarker panels were examined and validated. All of the panels showed good discriminative ability. The discrimination power can be further improved by incorporating putatively matched metabolites into the panel. Using pathway mapping tool, we further analyzed the altered pathways in both disease.

As sister diseases, CAA and AD have many similarities in pathogenesis and clinical presentations. The similarity of their metabolic alteration was confirmed in this study. As the results shown in comparing with healthy controls, many common metabolites that have significant changes were observed, such as methionine sulfoxide, arginine and glutamic acid. Despite these similarities, we were also able to find a panel of biomarker candidates that discriminate CAA from AD, indicating there are subtle differences between two conditions.

Limitations of current study are the relatively small sample size and the lack of external dataset for validation. In the future study, more samples will be collected to validate the novel biomarker panels established in this study. Additionally analyzing different types of samples, such as cerebrospinal fluid or brain tissues, will also benefit the understanding of disease mechanism.

Chapter 7

Construction and Application of a High-resolution MS/MS-RT Library for Rapid Identification of Endogenous Metabolites in Metabolomics

7.1. Introduction

High-resolution LC-MS has been proved to be an important platform for metabolomics research, featuring high metabolite detectability and accurate quantification ability.²⁰⁶⁻²⁰⁷ In a typical comparative metabolomics study, after data acquisition and processing, the detected metabolite features are used to discriminate between two groups (e.g. healthy controls and patients) using various statistical tools. Important questions related to this step are: which metabolites contribute to the separation of the two groups and what is their identity. To understand the biological process and mechanism for the studied disease, the metabolite features must be identified.

Using LC-MS, the metabolite identification process is usually done by matching one or several properties of these unknown features against the entries in existing databases,²⁰⁸ such as the Human Metabolite Database (HMDB).¹⁷ Depending on the number of parameters that are used for identification, there are different confident levels of the results. Accurate mass or m/z searching is often solely used to perform putative identification. However, due to the existence of isomers with identical masses, it cannot provide unambiguous identification. A consensus in metabolite identification is that if a minimum of two independent and orthogonal data were used to perform identification, the generated results can be treated as positive identification or high confidence identification.¹⁸⁴ For example, using accurate mass plus MS/MS fragmentation pattern or

using accurate mass plus retention time information can significantly improve the identification power of the LC-MS platform.

MS/MS spectra matches are the most popular approach for high confidence identification. The substructure information can be elucidated from the interpretation of fragment pattern.²⁰⁹ Considering that the manual interpretation process is time-consuming, many MS/MS spectral repository have been built worldwide over the last decade or so for rapid MS/MS matches and metabolite identification.^{18, 127, 210-211} Initially many of them were built using low resolution instruments, such as ion-traps and triple quadrupoles mass spectrometers. Now more libraries are expanded using high resolution mass spectrometers (e.g. Q-TOF) to produce spectra with better mass accuracy and higher resolution.²¹² The identification confidence is greatly improved using higher quality data and less error tolerance.

However sometimes MS/MS spectrum matches are not sufficient. The MS/MS spectra from low intensity analytes are usually of poor quality. False results may be generated since many fragment peaks are missing. Moreover, very similar MS/MS spectra might be produced from multiple compounds, such as isomers or structural analogs. In addition, some analytes only produce fragments of one or several dominant ions, which may not be unique to the metabolites. In these situations, other parameters are required for identifying metabolites, such as retention time (RT) in LC separation.²¹³⁻
²¹⁴ The challenge of using RT to perform identification is that RT is very easily affected by variations in experimental conditions like instrument brands, tubing lengths and mobile phase compositions. These may introduce retention time shift between labs when analyzing the same samples.

In this work, we report our study of the construction of an endogenous metabolite library of over 800 compounds containing molecular mass, experimental fragment ion spectrum (MS/MS) and additional retention time (RT) information. We also developed and demonstrated a rapid and high confidence metabolite identification solution using the library for biological samples. A LC normalization method was developed to reduce the retention time shift due to instrument variations. The end user is expected to follow the workflow to perform analysis and consistent results can be generated. The portability of this library and workflow were examined for between different labs.

7.2. Experimental Section

7.2.1. Chemicals and Reagents

Metabolite standards were obtained from the Human Metabolome Database (HMDB)²¹⁵ at the University of Alberta, Canada. LC-MS grade water (H₂O) and acetonitrile (ACN) were purchased from Thermo Fisher Scientific (Edmonton, AB, Canada). Other chemicals and reagents were purchased from Sigma-Aldrich (Markham, ON, Canada), except those otherwise stated.

7.2.2. Standard Solution Preparation

For each metabolite standard, approximate 1-2 mg of solid or liquid were measured in a microcentrifuge vial and dissolved in 1 mL of 1:1 (v/v) H₂O:ACN with 0.1% formic acid (FA) to generate a stock solution. In the case of insoluble solids, supernatant were used.

For MS/MS spectra acquisition in positive ion mode, stock solutions were diluted with 1:1 (v/v) H₂O:ACN with 0.1% FA. For negative ion mode, 1:1 (v/v) H₂O: ACN containing 1 mM acetic acids (HAc) was used to dilute stock solutions. The diluted solutions were then transferred to injection vials for MS/MS analysis. The concentration for each metabolite was adjusted according to the MS signal.

For RT acquisition, metabolite standards were first classified to different categories based on their chemical functional groups and HMDB “Super Class” information (e.g. organic acids, nucleosides, carbohydrate, etc.). Stock solutions of 10 to 24 standards from one category were combined in a microcentrifuge vial by taking 10 µL from each stock solution. 10 µL of the combined solutions were diluted with 800 µL of mobile phase solvents, followed by adding 10 µL of corresponding RT calibrants solution (see below).

7.2.3. Quality Control and Retention Time Calibrants

For MS/MS data collection, a single compound at a low concentration was used as the quality control (QC) sample, which was 1 µM of melatonin in 1:1 (v/v) H₂O:ACN with 0.1% FA for positive ion mode and 1 µM of citric acid in 1:1 (v/v) H₂O:ACN with 1 mM HAc for negative ion mode.

For RT library construction, standard mixtures were made as RT calibration mixture (RT-CalMix) and QC samples for monitoring instrument performance (see Retention Time Acquisition). To prepare the RT-CalMix solution, 100 µL of each calibrants solution were mixed in corresponding mobile phases. 10 µL of RT-CalMix solution was then diluted by 800 µL of mobile phase to generate the final QC sample for RT acquisition in both polarities.

7.2.4. MS/MS Spectra Acquisition

MS/MS spectra were collected using an Impact HD Quadrupole Time-of-flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA). Samples were delivered by flow infusion to electrospray ionization source using Ultimate 3000 UHPLC system (Thermo Scientific, USA) with no column attached. For positive ion mode, delivery mobile phase A (MPA) was H₂O with 0.1% FA and mobile phase B (MPB) was ACN with 0.1% FA; while for negative ion mode, MPA was 1 mM HAc in H₂O and MPB was 1 mM HAc in ACN. Entire run was 2.2 minutes with constant 50% B. The flow rate was: t = 0 min, 100 μ L/min; t = 0.3 min, 100 μ L/min; t = 0.31 min, 30 μ L/min; t = 1.7 min, 30 μ L/min; t = 2.2 min, 500 μ L/min. Mass spectrometer source settings were: nebulizer gas 2.0 bar, dry gas 8.0 L/min, 220 °C, capillary voltage 4.5 kV, capillary offset 0.5 kV, acquisition rate 8 Hz, mass range 20 – 1000 m/z, and CID gas was nitrogen.

For both polarities, there were 13 individual sections in the 2.2 minutes of MS/MS method. The first section was performed MS detection of precursor ion, followed by consecutive 5 sections of MS/MS detection with 6 Da of isolation window and another 5 sections of MS/MS with 1 Da of isolation window. For the five sections, different collision energies were applied, including 10, 20, 30, 20-50, 40 eV, respectively. Then another section of MS mode was employed to ensure there was still enough signal of precursor ion. The last section was used to measure sodium formate signal in MS mode for mass calibration.

7.2.5. Retention Time Acquisition

RT collection was performed twice in two systems. System I is Ultimate 3000 UHPLC system (Thermo Scientific, USA) linked to Impact HD Q-TOF mass

spectrometer (Bruker, Billerica, MA). An Acclaim 120 C₁₈ 100 x 2.1 mm 2.2 μm column (Thermo Scientific) with a Vanguard BEH C₁₈ 1.7 μm guard column (Waters) was used for this system. In System I, RTs was collected using different mobile phases in positive and negative ion mode to achieve high MS/MS signal for both polarities (see Results and Discussion). For positive ion mode, MPA was H₂O with 0.1% FA and MPB was ACN with 0.1% FA. For negative ion mode, MPA was H₂O with 1 mM HAc and MPB was ACN with 1 mM HAc. The gradient was: t = 0 min, 1% B; t = 2 min, 1% B; t = 17 min, 99% B; t = 20 min, 99% B. The flow rate was 250 μL/min. A 10 min washing was used to wash and equilibrate column between each run. The gradient was: t = 0 min, 99% B; t = 4 min, 99% B; t = 4.5 min, 1% B; t = 10 min, 1% B at 350 μL/min flow rate. The column temperature was kept at 30 °C. The injection volume was 5 μL. For MS detection, the same mass spectrometer source settings were used as those in MS/MS Spectra Acquisition above.

To construct a RT library that is more convenient for users, RTs of all metabolite standards were collected in the other system (System II) using one mobile phase for positive and negative ion mode. In this system, samples were delivered using an Elute UHPLC system (Bruker, Germany) through Intensity Solo 2 C₁₈ column (Bruker, Germany) and detected by a Compact Q-TOF mass spectrometer (Bruker, Germany). For both polarities, MPA was H₂O with 0.1% FA and MPB was ACN with 0.1% FA. The column temperature was kept at 35 °C. Other instrumental conditions were the same as those of System I.

7.2.6. Biological Sample Preparation and Analysis

Urine samples were collected after obtaining the informed consent from volunteers and ethics approval from the University of Alberta Ethics Approval Board. After collection, the urine samples were centrifuged at 20817 g for 10 min at 4 °C. Then the samples were filtered by 0.22 µm sterile syringe filters with MCE membrane (Millipore) and stored at -80 °C.

Human plasma samples were purchased from Sigma-Aldrich (Markham, ON, Canada). The samples were aliquoted into microcentrifuge tubes, followed by adding cold methanol at 3:1 v/v (i.e. 90 µL of cold methanol was added for 30 µL of sample). The mixtures were vortexed and placed in -20 °C freezer for 2 hours. The samples were then centrifuged at 20817 g for 15 min at 4 °C. The supernatants (3 times of volume of the original sample) were transferred to new microcentrifuge tubes, which were dried down using Speedvac. The dried samples were stored in -80 °C freezer for further use or dissolved in solvent for LC-MS analysis.

Yeast cells were cultured in proper medium and transferred to a microcentrifuge tube. The media containing cells were centrifuged at 4640 g for 10 min at 4 °C to harvest the cells. After removing the supernatant, the pellets were resuspended in 1 mL of cold water and spun in microcentrifuge at 16000 g for 1 min at 4 °C to wash the cells, followed by removing the water. The washing step was repeated for two more times. Cell pellets are then snap-frozen in liquid nitrogen and store at -80 °C until further use. For cell lysis, 0.5 mL of glass beads and 100 µL of 50% methanol (MeOH: H₂O, 1:1 v/v) were added to the frozen cell pellets. The cells were lysed via five 1 min periods of bead-beating using a vortex alternated with five 1 min incubations in an ice-water bath. After

cell lysis, 800 μL of 50% methanol ($\text{MeOH}:\text{H}_2\text{O}$, 1:1 v/v) was added for metabolite extraction. Then vials were then centrifuged at 16000 g at 4 $^{\circ}\text{C}$ for 10 min to remove the cell debris. The supernatants were transferred to new microcentrifuge tubes, which were then dried down using Speedvac. The dried samples are stored in -80°C freezer for further use or dissolved in solvent for LC-MS analysis.

To analyze biological samples, duplicate LC-MS/MS analysis of both positive ion mode and negative ion mode were performed for each sample (i.e. four LC-MS/MS runs for each sample). The same LC method in Retention Time Acquisition and same MS/MS settings in MS/MS Spectra Acquisition were used.

7.2.7. Data Processing

For MS/MS spectra collection, raw data were first processed by an in-house script based on DataAnalysis (Bruker, Germany). The program automatically produced average MS and MS/MS spectra from all of the scans within one section. The mass calibration was performed based on the sodium formate signal in the last section. The quality of MS spectra in the first section and MS/MS spectra in other sections were manually inspected. Intensities, mass errors and isotopic patterns should meet the defined criteria. Potential noise signals in 6 Da window spectra were removed by comparing with corresponding 1 Da window spectra. Using DataAnalysis' SmartFormula3D tool, chemical formulas were calculated and assigned to all the fragment peaks based on their accurate masses and isotopic peaks. At least two fragments were manually interpreted and structures were assigned to the fragment peaks.

For RT collection, TargetAnalysis (Bruker, Germany) was used to extract RTs of all analytes and RT calibrants in the mixture. The RTs of metabolite standards were normalized to a same QC sample using TASQ software (Bruker, Germany).

For biological sample analysis, Metaboscape 3.0 (Bruker, Germany) was used to automatically calibrate the samples, extract metabolite features, and align features from different samples. Metabolite identification was carried out using the constructed MS/MS-RT library. Statistical analysis was also performed in the software for proof-of-concept study.

7.3. Results and Discussion

7.3.1. Construction of MS/MS Library

825 human endogenous metabolite standards were used to construct the high resolution MS/MS-RT library. In the MS/MS spectra collection, both positive ion mode and negative ion mode acquisition were carried out. The development and optimization of LC-MS/MS method for both polarities were similar. In this report, we use negative ion mode as an example to illustrate the considerations in MS/MS spectra acquisition.

Unlike the predominant use of 0.1% formic acid (FA) as mobile phase additive in positive ion mode detection, the choice of additive in negative ion mode is inconclusive and more complex. Many researches have been reported to explore the optimal additive for various metabolites detection in negative ion mode.²¹⁶⁻²¹⁸ In one study, the authors systematically investigated the performance of untargeted urinary metabolomics analysis using different additives. Their results indicated that 0.1% FA for positive ion mode was

far from the optimum for negative ion mode detection and 1 mM of HAc as additive behaved the best among the tested additives, including FA, ammonium acetate (NH₄Ac) and ammonium fluoride (NH₄F).²¹⁹ Similar conclusions were generated in other studies.²¹⁸ Therefore to collect MS/MS spectra in negative ion mode, we first studied the effect of additives on metabolite detection.

As shown in Figure 7.1, LC-MS signal intensity of several compounds in negative ion mode using different additives were compared, including 0.1% FA, 1 mM HAc, 1 mM NH₄Ac of pH 4, 1 mM NH₄Ac of pH 7 and 1 mM NH₄Ac of pH 8. In the testing metabolites, citric acid and traumatic acid are the representative compounds of hydrophilic and hydrophobic carboxylic acids, respectively, which are mainly detected in negative ion mode (Figure 7.1A and 7.1B). One nucleotide, adenosine 3',5'-diphosphate, was also inspected (Figure 7.1C). The MS/MS fragments of nucleotides in positive ion mode are mainly derived from nucleobases, while negative ion mode fragmentation renders the important structural information of phosphoribosyl moieties.²²⁰ Diacetyl was examined as small neutral compound (Figure 7.1D). Maltose was the representative of carbohydrates (Figure 7.1E). The result clearly shows that for all compounds, 1 mM of HAc as additive in the mobile phase provided the highest LC-MS signal. Although for some compounds, other additive may provide comparable signal (e.g. 1 mM NH₄Ac of pH 4 for diacetyl), the performance of 1 mM of HAc was more consistent.

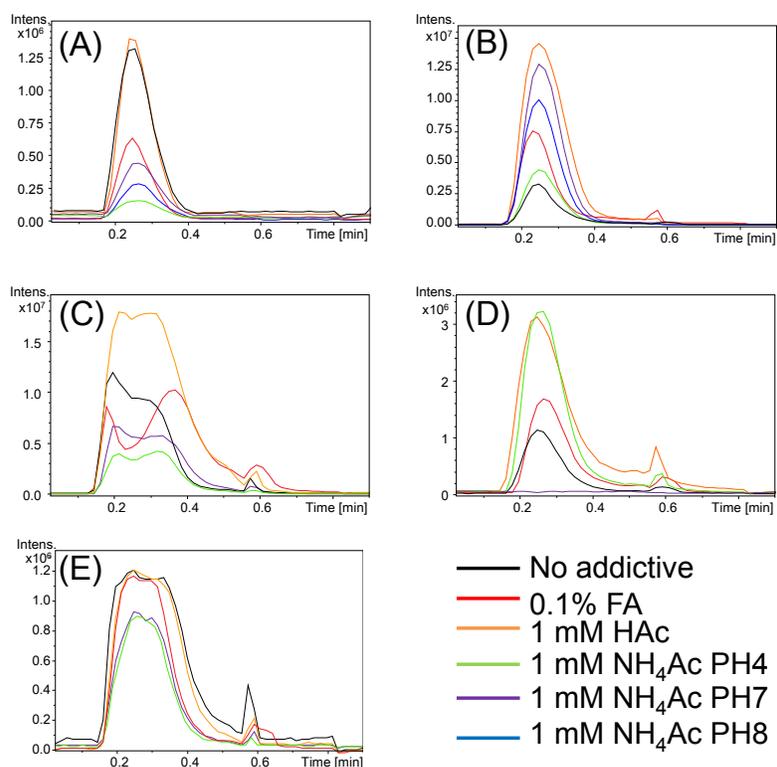


Figure 7.1 Effect of different additives on signal intensity of (A) citric acid, (B) traumatic acid, (C) adenosine 3',5'-diphosphate, (D) diacetyl and (E) maltose in negative ion mode detection.

MS/MS fragmentation pattern was then inspected using different flow injection buffers. Figure 7.2 shows the results of citric acid. Very similar MS/MS spectra were generated with different additives, indicating that using different additives did not change the MS/MS fragmentation pattern. This is good for library construction since the MS/MS spectra collected in one buffer system would be usable to any other systems. The same results were also found for other compounds (data not shown). Hence, we decided to use 1 mM of HAc as additive in MS/MS spectra collection for negative ion mode to achieve better detection. While for the positive ion mode, we stuck with the 0.1 % FA as the most commonly used additive.

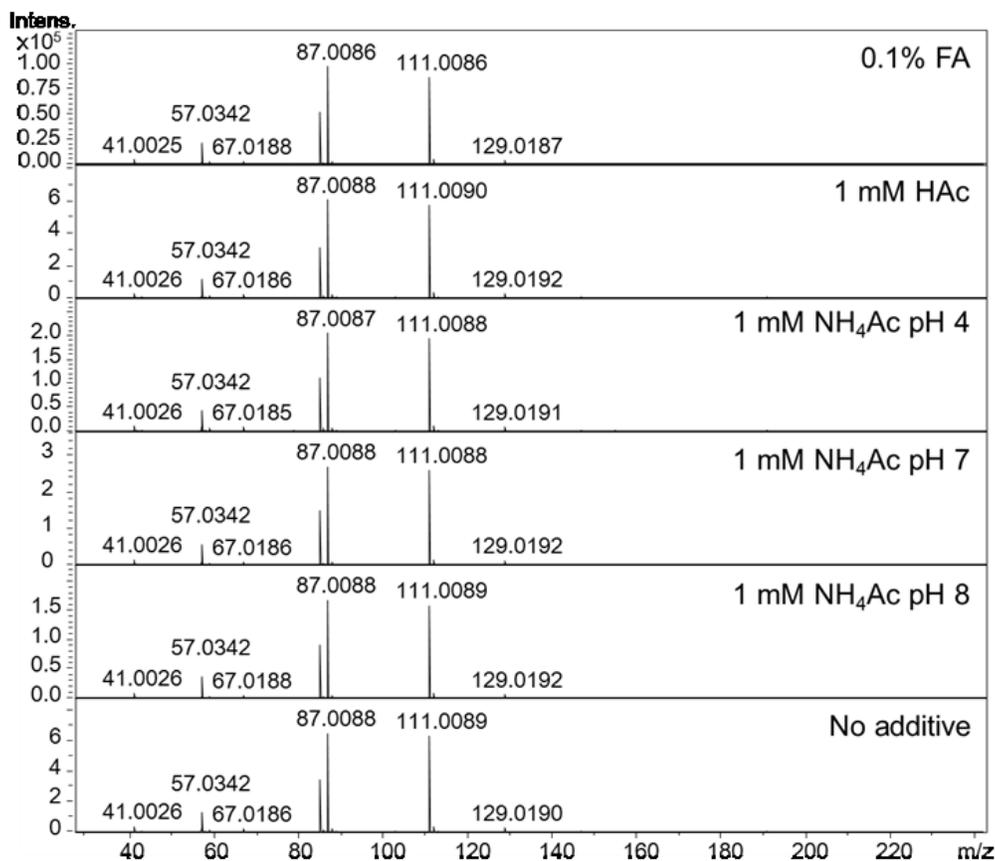


Figure 7.2 MS/MS spectra of citric acid using different additives in mobile phase.

Figure 7.3 shows an example of generated LC-MS/MS data for MS/MS library construction. Several features in MS/MS method and data processing were designed to improve the quality of the library. First, five different collision energy conditions were applied to collect MS/MS spectra for each metabolite, including 10, 20, 30, 40 and stepping mode of 20-50 eV (Figure 7.3). These different levels of collision energy ensured that metabolites of various structures and different fragility of collision-induced dissociation (CID) fragmentation were able to produce at least one high quality MS/MS spectrum. Using multiple collision energy levels also generated more structural information for each compound, which benefits the unknown metabolite identification.

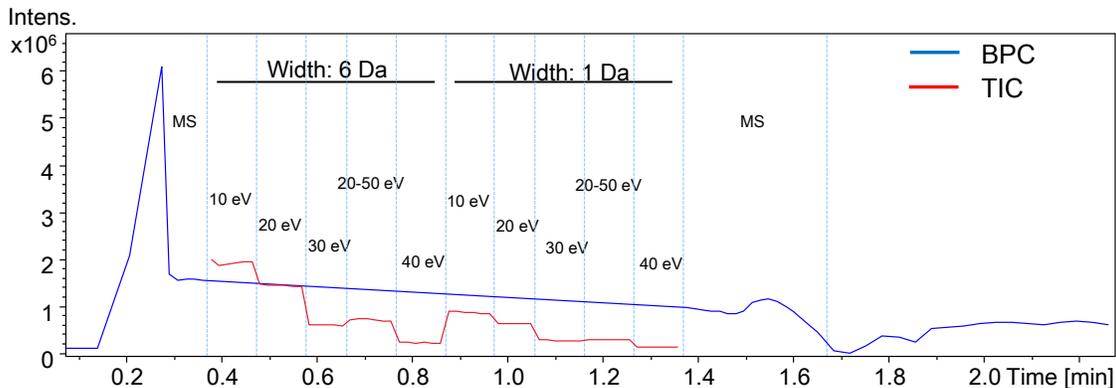


Figure 7.3 A typical LC-MS/MS data of one metabolite for MS/MS library construction.

Second, in order to collect high quality mass spectra, two different isolation windows were employed. One is wide isolation window of 6 Da and the other one is narrow isolation window of 1 Da. When performing MS/MS acquisition, spectra from 6 Da isolation window contained complete isotopic peaks, which can be used for determination of chemical formula of fragment peaks. It also provided higher signal intensity because of more transmission of ions in the quadrupole of mass spectrometer. However, the wide isolation window may result in signals from other different ion species in the fragment spectrum (i.e. lower specificity). Thus after acquisition using 6 Da isolation window, another round of acquisition using 1 Da isolation window was carried out. The spectra generated in this condition were cleaner and contained less contaminated peaks from other species. In data processing, the corresponding spectra from two isolation windows were manually compared and combined in software. A peak in spectrum of wide isolation window were treated as contaminated peaks if 1) it was not an isotopic peak of the other peaks and 2) not presented in narrow window acquisition. It was then removed from the spectrum. In this way, clean final library spectra were yielded.

In addition, as mentioned above, the chemical formulas of all fragment peaks were first calculated based on accurate mass and isotopic pattern. Then at least two fragment peaks were manually interpreted and the structures of the peaks were annotated in the library. The formulas and structures are very useful in identification and interpretation of spectra from unknown metabolites.

Using the developed method, in total, 6023 MS/MS spectra from 743 substances were included in the current library.

7.3.2. Construction of RT Library

The retention times (RTs) of the same 825 metabolite standards on reversed phase LC (RPLC) were measured in both positive ion mode and negative ion mode to build the RT library. Two versions of the RT library were constructed using different instrument setups and conditions. As mentioned above, compared with 0.1% FA, 1 mM HAC behaved better as additive for negative ion mode detection of many metabolites. Therefore in the first version of RT library, we used different additives (0.1 % FA for positive ion mode and 1 mM HAC for negative ion mode) in the mobile phase to achieve better detection sensitivity. However, when performing identifications using the library, samples also need to be run using different mobile phases for different polarities, which is relatively inconvenient. Thus we constructed the second version of RT library using the same additive, 0.1% FA, for both positive and negative ion mode. Although some sensitivity was sacrificed, it would be more convenient and practical to use the library. The principle and method to construct two versions of library were similar. The details in construction of second version of library will be discussed as an example.

RT is not a commonly used parameter for metabolite identification in LC-MS based metabolomics. This is because RT can be easily affected by a number of experimental conditions, including minor variations in LC equipment, connecting tube lengths from LC to MS, solvent composition, temperature, etc. To address this issue, a RT normalization method and RT calibration mixture (RT-CalMix) were used to correct RT shift in different instruments. In this approach, we first created a RT-CalMix containing 28 components (see the list in Figure 7.4). The majority of the components were dansyl chloride labeled metabolites. Dansyl-labeled compounds are much more stable than unlabeled ones and have better peak shape on RPLC, which makes them suitable as RT calibration mixture components. The RTs of 28 calibrants distributed within the whole gradient and the interval between two calibrants are less than one or two minutes (Figure 7.4). This RT-CalMix was used as QC sample to monitor the stability of instrument. It was also spiked into all metabolite mixtures for retention time normalization. The principle of this approach is shown in Figure 7.5A. When constructing the RT library, the measured RTs of metabolites were normalized according to a same RT-CalMix using a linear RT correction method reported before⁵⁹. The normalized RTs were included in the library. When users perform metabolite identification using the library in a different instrumental setup, they need to run the same RT-CalMix on the local instrument. Then all RTs in the library would be corrected according to the new RT-CalMix to generate a local version of the RT library. All the processing can be automatically performed in TASQ software (Bruker, Germany). In this way RT was able to be used as a parameter for metabolite identification.

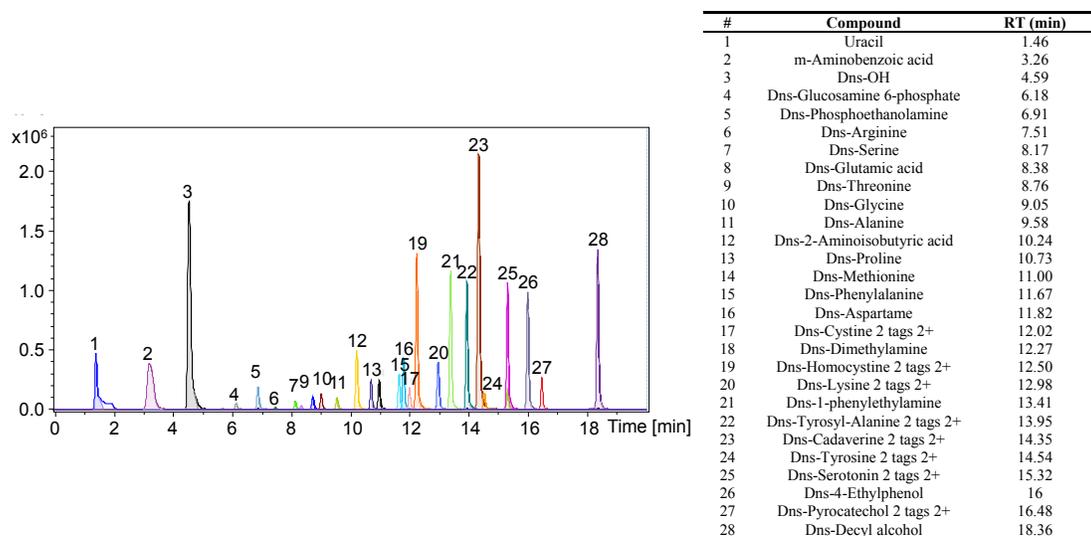


Figure 7.4 LC ion chromatogram and list of RT-CalMix containing 28 components.

To acquire RTs, metabolite standards were mixed together according to their chemical classifications to prevent unwanted interference between each other. Overall 53 standard mixtures were generated for RT collection. All samples were measured in triplicates in both positive and negative ion modes. The averaged RT from triplicate experiments was used as the final RT in the library. The standard deviation (SD) of RT was used to evaluate the reproducibility of the results. In the current library, retention times of 658 metabolites were acquired. The average SD of RT was 0.015 min with maximum 0.053 min, indicating good reproducibility of RT collection. The distribution of RTs and SD of RT for all measured metabolites is plotted in Figure 7.5B.

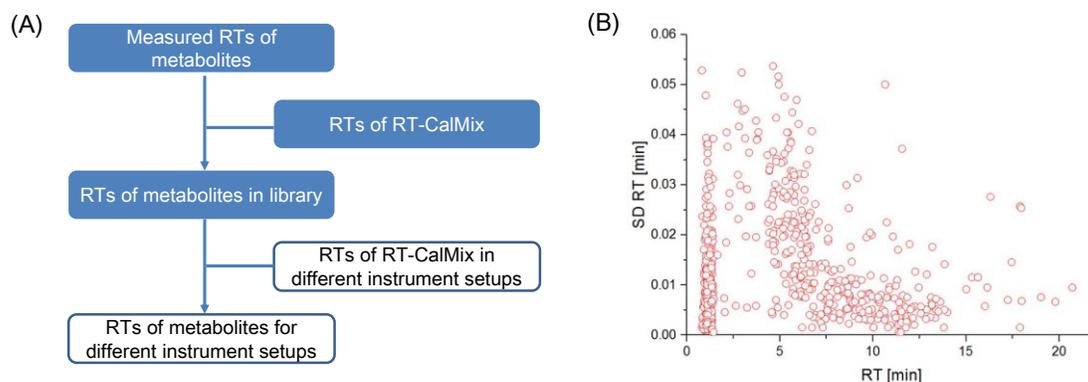


Figure 7.5 (A) Principle of retention time normalization approach. (B) Distribution of RTs and their standard deviations of metabolites in the library.

7.3.3. Metabolite Identification Solution

Based on the high-resolution MS/MS library and the RT library we constructed, a comprehensive workflow for performing metabolite identification of biological samples was developed (Figure 7.6). It involves the following steps: 1) sample preparation following the standard operation procedures, 2) RPLC-MS/MS analysis of samples, 3) data processing using Metaboscape software and 4) metabolite identification with multiple parameters, including accurate mass of precursor ion, MS/MS spectra and retention time information.

So far standard operation procedures for preparing three types of biological samples have been developed, including plasma, urine and yeast cells. For plasma samples, methanol was first used to remove proteins in the samples. After protein precipitation, samples were dried down and reconstituted in proper solvent for LC-MS/MS analysis. Since a very small amount of proteins in urine samples would not affect analysis, only a filtration step was applied to prepare the urine samples. For yeast cells, a highly efficient glass beads-assisted cell lysis method developed previously⁵⁸ was

constructed MS/MS library and RT library were combined for identification. The results from identification using only the accurate mass of precursor ion were considered as putative identification or low confidence identification. If two orthogonal and independent parameters were used, the results were treated as high confidence identification, such as using accurate mass plus MS/MS spectra matches, or accurate mass plus RT information.¹⁸⁴ The approach using RT as a parameter for metabolite identification is described above (Figure 7.5A). Figure 7.7 shows two examples of the identified results. For MS/MS matches, a fit score will be generated to indicate the quality of the matching results. The higher value represents better matches with maximum value of 1000. A common criteria for identification is 5 mDa of precursor mass tolerance, 0.5 min of retention time tolerance and MS/MS fit score larger than 500.

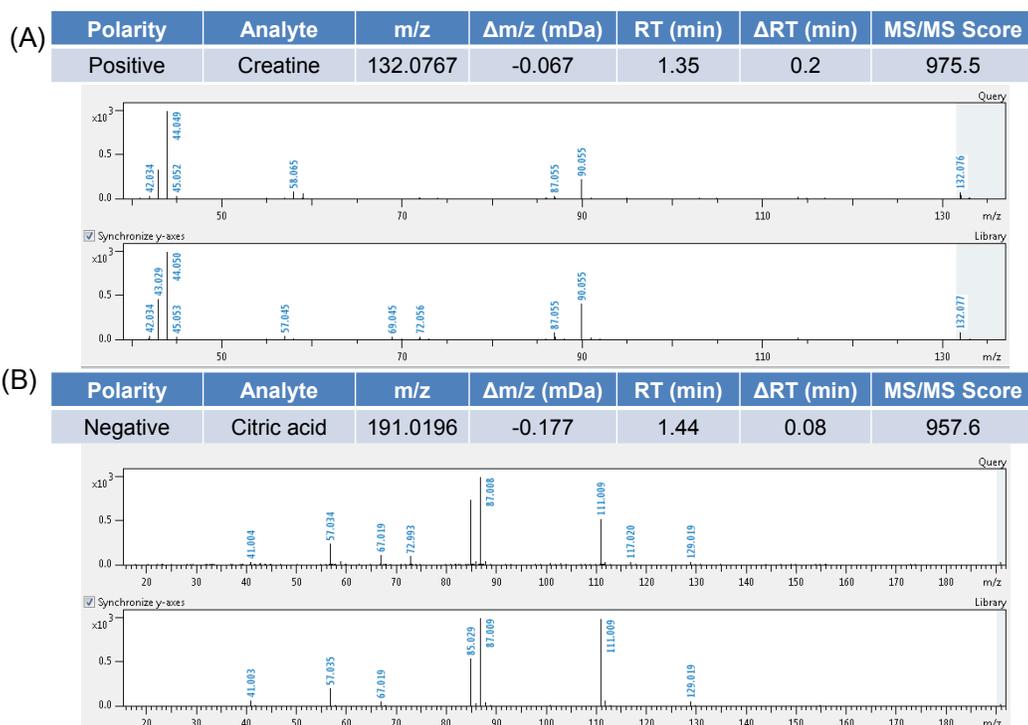


Figure 7.7 Metabolite identification using constructed MS/MS-RT library in Metaboscape of (A) creatine in positive ion mode and (B) citric acid in negative ion mode.

In the last step of the solution, identification results together with the feature table were subjected to statistical analysis. One major advantage of Metaboscape software is its integration of many functions for entire metabolomics workflow. Thereby it allows us to perform data processing, metabolite identification and different statistical analysis in one software without any data file format conversions, which significantly reduces the entire analysis time.

7.3.4. Analysis of Biological Samples

Human urine, human plasma and yeast cells were analyzed following the developed workflow above. Figure 7.8A shows the LC-MS/MS chromatograms of human urine sample and identification results in positive and negative ion modes. Metabolite identification was carried out using solely accurate mass (m/z), combined m/z and MS/MS spectra or combined m/z and RT information. For example, in negative ion mode, 239 features were identified with relatively low confidence if only using m/z ; 68 features could be identified using m/z and MS/MS spectra; 137 features could be identified using m/z and RT information. If combined results of two parameters (i.e. m/z +MS/MS and m/z +RT), 155 high confidence identified features were generated, among them 121 metabolites were identified using at least two parameters. At last, we combined the results from both positive ion mode and negative ion mode, a total of 177 metabolites can be identified with high confidence for human urine samples (Table 7.1). Similarly, metabolite identification was performed for human plasma and yeast cells samples. 127 and 140 metabolites were identified for the two samples, respectively (Figure 7.8B-C).

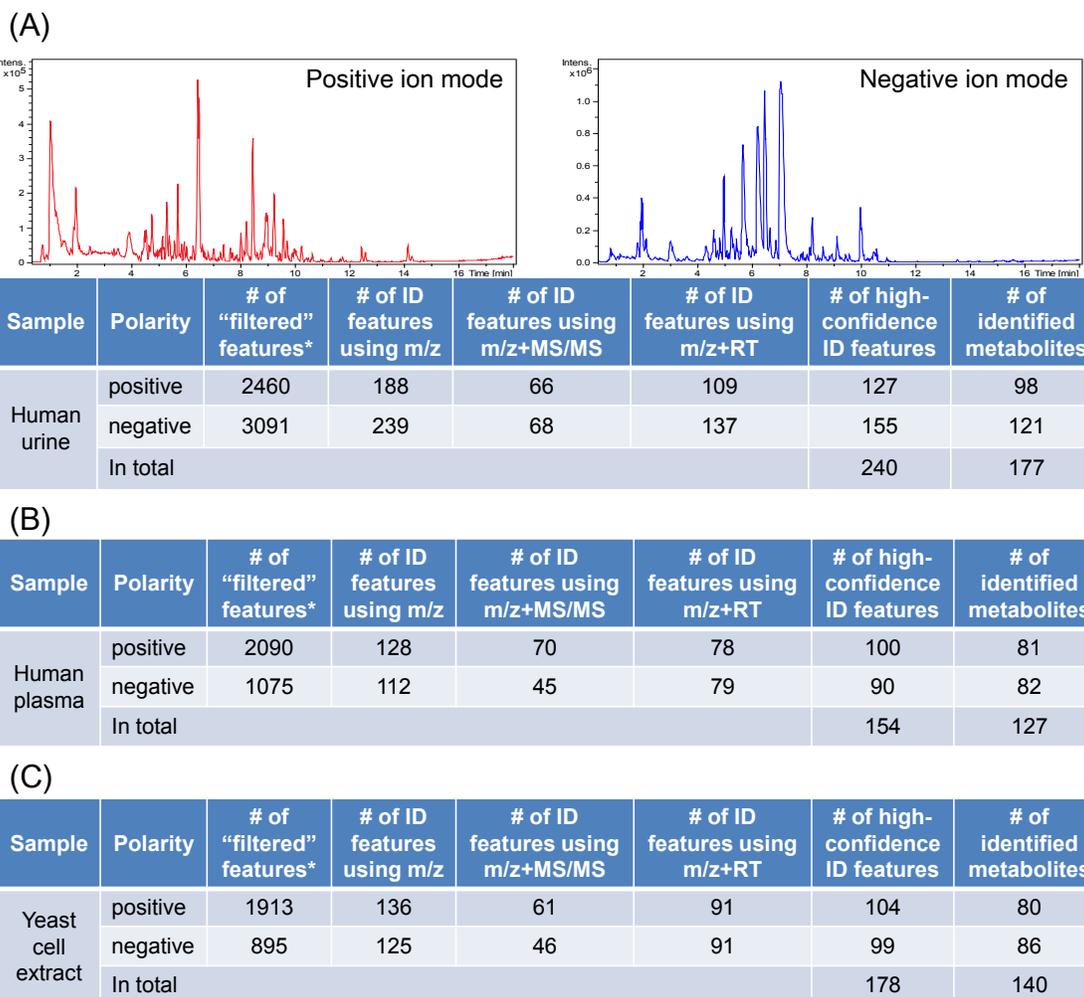


Figure 7.8 LC-MS/MS chromatograms and identification results of (A) human urine sample, (B) human plasma sample and (C) yeast cell extracts in both positive and negative ion mode. “Filtered” features are the ones after removing redundant spectral features (adduct ions, multimers, etc.).

It should be noted that there was a difference between the number of identified features and the number of identified metabolites. We further looked into the difference and found that one major reason is the existence of isomers. As illustrated in Figure 7.9. When plotting extracted ion chromatogram of one metabolite xanthosine, two clear separate peaks were observed with the same m/z. And their MS/MS spectra were so

similar that both of them could be identified as xanthosine if using m/z and MS/MS spectra matches. But, only the second peak was confirmed by RT information, which indicated that the second peak came from the true metabolite while the first peak had a high chance to be an isomer of xanthosine. Though we cannot get the identity of the first peak at this stage due to the limited size of our library, the speculation about the isomer still provided information for further identification or analysis. This example also shows the beauty of adding RT as a parameter to further improve the identification confidence.

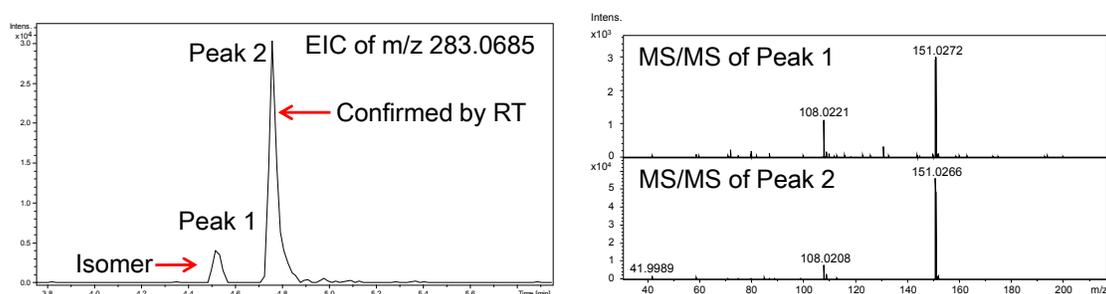


Figure 7.9 Identification of xanthosine in human urine using MS/MS and RT matches.

Table 7.1 List of 177 metabolites that were identified in human urine

Name	Polarity	m/z (Da)	RT (min)	$\Delta m/z$ (mDa)	ΔRT (min)*	MS/MS score*
1,11-Undecanedicarboxylic acid	Neg	243.1602	11.45	0.057	0.10	840.9
1,3-Dimethyluric acid	Pos	197.0665	5.13	-0.378	-0.03	938.3
	Neg	195.0526	5.18	-0.021	0.01	938.9
1-Methylhistidine	Pos	170.0924	0.96	0.004	0.03	714.7
	Neg	168.0789	0.99	1.085	0.07	NA
2,3-Dihydroxybenzoic acid	Neg	153.0195	6.80	0.183	NA	629.7
2,5-Furandicarboxylic acid	Pos	139.0020	4.64	-0.561	-0.10	NA
	Neg	154.9990	4.73	0.913	-0.01	NA
2-Furoic acid	Neg	111.0096	5.68	0.808	0.13	NA
2-Furoylglycine	Pos	170.0444	5.04	-0.401	-0.06	976.9
	Neg	168.0310	5.11	0.746	0.01	907.1
2-Hydroxy-2-methylbutanedioic acid	Neg	147.0299	2.55	-0.038	NA	789.1
2-Ketobutyric acid	Pos	103.0387	2.99	1.331	-0.35	NA
	Neg	101.0246	3.21	-0.007	-0.13	NA
2-Pyrrolidinone	Pos	86.0582	3.43	-1.815	-0.04	NA
	Neg	84.0459	3.91	0.402	0.43	NA

3,4,5-Trimethoxycinnamic acid	Neg	237.0773	9.05	0.407	-0.20	NA
3,4-Dihydroxybenzaldehyde	Neg	137.0256	6.22	1.190	0.12	NA
3,7-Dimethyluric acid	Neg	195.0531	4.83	0.713	0.01	931.6
3-Aminoisobutanoic acid	Pos	104.0712	1.22	0.555	0.10	NA
3-Hexenedioic acid	Neg	143.0358	5.85	0.767	0.41	998.4
3-Hydroxyanthranilic acid	Pos	154.0498	4.81	0.348	-0.49	988.3
3-Hydroxyisovaleric acid	Neg	117.0557	4.72	-0.01	0.03	NA
3-Hydroxymandelic acid	Neg	167.0385	6.39	3.551	-0.40	NA
3-Hydroxymethylglutaric acid	Neg	161.0454	3.20	-0.499	-0.15	958.6
3-Methyl-2-oxovaleric acid	Neg	129.0546	6.11	-1.16	-0.05	NA
3-Methyladipic acid	Neg	159.0661	6.48	-0.236	-0.03	883.2
3-Pyridylacetic acid	Pos	138.0553	1.24	0.340	-0.07	787.3
4-Aminohippuric acid	Pos	195.0768	3.35	0.330	-0.49	NA
4-Deoxytetronic acid	Neg	85.0299	2.89	0.401	0.14	NA
4-Hydroxy-3-methylbenzoic acid	Pos	153.0534	7.77	-1.197	0.33	996.9
4-Hydroxycyclohexylcarboxylic acid	Neg	143.0719	5.87	0.511	0.11	NA
4-Methoxyphenylacetic acid	Pos	167.0700	8.53	-0.293	-0.12	863.9
	Neg	165.0548	8.19	-0.875	-0.46	NA
4-Methylcatechol	Neg	123.0454	6.94	0.213	NA	959.9
4-Pyridoxic acid	Pos	184.0602	4.10	-0.476	-0.24	949.8
	Neg	182.0458	4.29	-0.059	-0.04	968.1
5-Hydroxyindoleacetic acid	Pos	192.0650	6.26	-0.323	-0.03	995.5
5-Hydroxy-L-tryptophan	Pos	221.0906	5.00	-1.479	0.34	NA
5-Hydroxymethyl-4-methyluracil	Pos	157.0605	1.29	-0.243	-0.07	NA
	Neg	155.0462	1.15	0.020	-0.20	NA
5-Methoxytryptophan	Pos	235.1069	5.92	-0.806	-0.18	NA
5-Methoxytryptophol	Neg	190.0871	8.32	-0.274	0.02	NA
7-Methylxanthine	Pos	167.0557	4.56	-0.657	-0.09	952.1
Acetaminophen	Pos	152.0708	5.34	0.149	-0.09	943.4
	Neg	150.0561	5.02	0.012	-0.41	998.2
Acetaminophen glucuronide	Neg	326.0879	4.70	-0.206	0.07	893.9
Acetylcysteine	Pos	164.038	5.59	0.365	0.28	NA
	Neg	162.0233	5.62	0.214	0.31	NA
Acrylamide	Pos	72.0444	1.13	0.051	-0.29	NA
Aldosterone	Pos	361.1969	8.43	-4.010	-0.36	NA
Alpha-hydroxyhippuric acid	Pos	196.0599	5.64	-0.535	0.02	NA
	Neg	194.0454	5.69	-0.394	0.07	NA
Alpha-hydroxyisobutyric acid	Neg	103.0410	3.06	0.878	-0.06	999.5
Alpha-ketoisovaleric acid	Neg	115.0405	4.62	0.392	0.15	NA
Alpha-N-phenylacetyl-L-glutamine	Pos	265.1172	6.53	-0.943	-0.05	977.6
	Neg	263.1037	6.56	-0.003	-0.02	970
Amino adipic acid	Pos	162.0760	2.03	-0.057	NA	675.5
Aniline	Neg	92.0519	7.73	1.286	0.30	NA
Asymmetric dimethylarginine	Pos	203.1501	1.23	-0.120	0.09	907.3
	Neg	201.1343	1.25	-1.427	0.10	NA
Azelaic acid	Pos	171.1013	8.24	-0.262	-0.03	NA
	Neg	187.0977	8.26	0.069	-0.01	841.3

Benzaldehyde	Pos	107.0478	8.31	-1.342	NA	879.7
Benzamide	Neg	120.0457	6.29	0.217	-0.01	NA
Benzoic acid	Neg	121.0293	7.99	-0.175	-0.45	NA
Betaine	Pos	118.0852	1.05	-1.078	0	661.3
Caffeic acid	Neg	179.0355	6.36	0.419	-0.23	610.1
Cis-Aconitic acid	Pos	175.0213	1.48	-2.449	0.15	NA
Citric acid	Pos	175.0240	2.00	0.250	-0.04	866
	Neg	191.0194	2.05	-0.289	0.02	987
Cortisol	Pos	363.2172	9.22	0.768	-0.01	700.4
Creatine	Pos	132.0761	1.21	-0.683	0.10	986.6
Creatinine	Pos	114.0654	1.10	-0.822	0.03	993.9
Daidzein	Pos	255.0642	8.84	-1.129	-0.02	781.9
D-Alanine	Pos	90.0546	1.23	-0.160	0.22	NA
	Neg	88.0406	1.10	0.154	0.10	NA
Delta-hexanolactone	Pos	115.0747	5.52	-0.686	-0.06	514.9
	Neg	113.0612	5.83	0.403	0.24	NA
D-Glucuronic acid	Neg	193.0344	0.99	-1.004	-0.06	NA
D-Glutamine	Neg	145.0624	1.00	0.574	-0.01	NA
Dihydrothymine	Pos	129.0662	3.03	0.306	-0.200	795.1
Dimethyl sulfone	Pos	95.0127	1.55	-3.407	0.19	NA
D-Lactic acid	Neg	89.0245	5.36	0.274	-0.03	NA
Dodecanedioic acid	Neg	229.1456	10.44	1.079	-0.19	599.7
D-Phenyllactic acid	Pos	167.0708	7.41	0.519	-0.18	NA
	Neg	165.0552	7.57	-0.522	-0.02	565
D-Ribulose	Neg	149.0456	1.16	0.005	0.09	821.4
Ethylmalonic acid	Neg	131.0358	5.07	0.544	-0.19	999
Gamma-Caprolactone	Neg	113.0609	6.98	-0.453	-0.22	NA
Genistein	Neg	269.0463	9.83	0.760	-0.01	NA
Gentisic acid	Neg	153.0195	6.25	0.194	-0.02	NA
Glucaric acid	Neg	209.0315	1.17	1.215	0.08	NA
Gluconic acid	Neg	195.0514	1.14	0.594	0.05	511.4
Gluconolactone	Neg	177.0409	1.15	0.455	0.07	NA
Glutaconic acid	Neg	129.0187	4.70	-0.651	-0.18	NA
Glutaric acid	Pos	133.0489	4.70	-0.683	-0.37	NA
Glycine	Pos	76.0390	1.10	-0.282	0.09	NA
Glycylproline	Pos	173.0916	1.77	-0.473	0.43	788.9
Guaiacol	Pos	125.0596	5.61	-0.105	-0.14	NA
	Neg	123.0453	5.47	0.187	-0.28	NA
Guaifenesin	Pos	221.0782	7.92	-0.249	0.37	NA
Guanidoacetic acid	Pos	118.0601	1.06	-1.051	0.03	NA
Hippuric acid	Pos	180.0653	6.49	-0.220	-0.06	937
	Neg	178.0508	6.52	-0.152	-0.03	888.6
Homovanillic acid	Neg	181.0513	6.78	0.248	-0.01	NA
Homoveratric acid	Pos	197.0804	7.98	-0.595	0.09	NA
	Neg	195.0655	7.78	-0.831	-0.11	NA
Hydroxyindoleacetic acid	Neg	190.0514	6.30	1.192	NA	982.7
Hydroxykynurenine	Pos	225.0895	1.75	2.532	0.41	NA

Hydroxyoctanoic acid	Neg	159.1015	9.73	-1.574	-0.04	NA
Hydroxyphenyllactic acid	Neg	181.0507	5.91	0.094	0.1	629.5
Hydroxypropionic acid	Neg	89.0244	1.83	0.083	0.47	NA
Imidazoleacetic acid	Pos	127.0493	1.55	-0.879	0.43	NA
Indoleacetic acid	Neg	174.0561	8.53	0.046	-0.03	988.3
Indolelactic acid	Pos	206.0799	7.82	-1.246	-0.04	869.8
	Neg	204.0663	7.83	-0.344	-0.03	900.3
Indoxyl sulfate	Pos	214.0178	6.25	0.177	-0.13	NA
	Neg	212.0019	6.27	-0.367	-0.11	892.4
Inosine	Neg	267.0742	4.47	0.707	0.03	994.8
Iso-Valeraldehyde	Neg	85.0670	5.57	1.075	-0.01	NA
Isovalerylcarnitine	Pos	246.1691	6.33	-0.926	-0.15	983.6
Isovalerylglycine	Pos	160.0967	5.93	-0.079	-0.08	953.1
	Neg	158.0824	5.97	0.122	-0.03	NA
Ketoleucine	Neg	129.0551	6.47	-0.633	-0.09	NA
Kynurenic acid	Neg	188.0355	6.03	0.129	NA	998.8
L-3-Phenyllactic acid	Neg	165.0545	7.79	-1.249	0.19	NA
L-Aspartic acid	Neg	132.0301	1.14	-0.152	0.08	NA
L-Aspartyl-L-phenylalanine	Pos	281.1119	5.91	0.017	0.01	746.9
	Neg	279.0986	5.96	1.549	0.06	NA
L-Carnitine	Pos	162.1126	1.08	-0.133	0.04	582
L-Fucose	Neg	163.0610	3.56	-0.185	NA	516.5
L-Glutamic acid	Neg	146.0461	1.35	0.216	0	NA
L-Histidine	Pos	156.0774	0.94	0.603	0.02	951.1
L-Isoleucine	Pos	132.1007	2.78	-1.241	0.03	805.8
	Neg	130.0877	2.93	0.362	0.18	NA
L-Kynurenine	Pos	209.0922	4.73	0.130	-0.06	531.3
L-Leucine	Pos	132.1012	2.50	-0.694	-0.24	866.5
L-Malic acid	Neg	133.0146	1.28	0.304	-0.02	NA
L-Norleucine	Neg	130.0865	2.97	-0.899	-0.16	NA
L-Phenylalanine	Pos	166.0845	4.98	-1.786	0.08	NA
	Neg	164.0724	4.90	0.712	0	NA
L-Sorbose	Neg	179.0562	1.16	0.128	0.09	819.5
L-Threonine	Pos	120.0647	1.39	-0.862	0.38	NA
L-Thyronine	Pos	274.1091	6.44	1.752	-0.47	NA
L-Tryptophan	Pos	205.0967	5.77	-0.501	-0.03	807
	Neg	203.0832	5.81	0.317	0.02	911.6
m-Coumaric acid	Neg	163.0403	7.59	0.249	-0.29	NA
Mesaconic acid	Neg	129.0190	4.48	-0.318	-0.16	NA
Methylglutaric acid	Neg	145.0510	5.67	0.412	-0.04	795.8
Methylhippuric acid	Pos	194.0802	7.26	-0.952	-0.36	531.8
	Neg	192.0662	7.27	-0.395	-0.35	581.4
N,N-Dimethylformamide	Pos	74.0608	2.03	0.728	-0.22	NA
	Neg	72.0459	2.13	0.431	-0.12	NA
N6-Acetyl-L-lysine	Pos	189.1231	1.80	-0.275	0.43	578
	Neg	187.1068	1.81	-2.011	0.44	834.6
N-Acetylglutamic acid	Neg	188.0574	1.78	0.988	0.43	649.7

N-Acetyl-L-phenylalanine	Pos	208.0967	7.61	-0.090	0	NA
	Neg	206.0820	7.61	-0.295	0	NA
N-Acetyl-L-tyrosine	Pos	224.0910	5.87	-0.769	-0.02	NA
	Neg	222.0782	5.91	0.988	0.02	NA
N-Acetylneuraminic acid	Neg	308.0991	1.18	0.649	0.09	NA
N-Acetylserotonin	Pos	219.1123	6.51	-0.557	0.07	NA
Niacinamide	Pos	123.0542	1.79	-1.081	0.45	NA
N-Methylhydantoin	Pos	115.0488	1.82	-1.384	0.43	940.6
Ortho-hydroxyphenylacetic acid	Neg	151.0399	7.19	0.176	-0.02	989.2
Oxoglutaric acid	Neg	145.0145	1.69	0.217	0.25	NA
p-Anisic acid	Neg	151.0400	8.31	-0.120	-0.39	974.3
Pantothenic acid	Pos	220.1169	5.15	-1.061	NA	713.9
	Neg	218.1038	5.20	0.374	NA	591
Paracetamol sulfate	Neg	230.0128	6.78	-0.066	NA	937
p-Cresol	Pos	109.0640	7.25	-0.762	-0.19	915.5
Phenol	Neg	93.0343	8.30	-0.300	-0.4	NA
Phenylacetyl glycine	Pos	194.0793	6.98	-1.858	-0.04	NA
	Neg	192.0663	7.01	-0.332	-0.02	NA
Phenylglyoxylic acid	Pos	151.0383	5.80	-0.659	0.43	NA
	Neg	149.0239	5.37	-0.525	0	NA
Phosphoric acid	Pos	98.9855	0.82	1.284	-0.27	NA
Phthalic acid	Pos	167.0345	6.39	0.644	-0.18	NA
p-Hydroxyphenylacetic acid	Pos	153.0543	6.04	-0.209	-0.43	NA
	Neg	151.0400	6.46	-0.059	-0.01	851.7
Pipecolic acid	Pos	130.0862	1.68	-0.085	0.34	NA
Protocatechuic acid	Neg	153.0192	4.99	-0.092	-0.40	NA
Purine	Neg	119.0352	1.72	-1.152	0.37	NA
Pyrocatechol	Pos	111.0435	5.84	-0.543	-0.22	NA
Pyrrolidonecarboxylic acid	Pos	130.0489	6.53	-0.957	NA	961.7
Quinic acid	Neg	191.0561	1.12	-0.020	-0.01	858.7
Riboflavin	Pos	377.1442	6.56	-1.328	0	808.7
	Neg	375.1320	6.61	0.972	0.04	NA
Ribonolactone	Neg	147.0298	1.16	-0.143	-0.18	NA
Salicyluric acid	Pos	196.0607	7.40	0.230	-0.04	993.4
Sebacic acid	Neg	201.1132	9.01	0.126	-0.09	912.3
Shikimic acid	Pos	175.0622	1.04	1.749	-0.29	NA
Sorbitol	Neg	181.0703	1.10	-1.457	0.06	NA
Spermine	Pos	203.2226	0.89	-0.405	0.08	NA
Stearic acid	Neg	283.2621	19.73	-2.113	-0.06	NA
Succinic acid	Neg	117.0197	2.48	0.521	0.18	988.7
Tartaric acid	Neg	149.0094	1.17	0.211	0.03	NA
Terephthalic acid	Pos	149.0233	19.00	-0.074	-0.02	NA
Theophylline	Pos	181.0716	5.66	-0.394	-0.06	958.3
Threonic acid	Neg	135.0298	1.16	-0.063	0.06	904
Thymine	Pos	127.0494	4.85	-0.812	0.22	NA
Tiglylglycine	Pos	158.0810	5.75	-0.207	-0.07	NA
Trans-Aconitic acid	Neg	173.0093	2.07	0.157	NA	995.9

Trigonelline	Pos	138.0547	1.17	-0.218	-0.15	655.9
Trimethylamine	Pos	60.0797	0.96	-1.093	-0.07	NA
Tryptophanol	Neg	160.0763	8.18	-0.530	-0.44	NA
Tyramine	Pos	138.0894	2.83	-1.981	NA	824
Undecanedioic acid	Neg	215.1280	9.73	-0.885	-0.15	963.1
Urea	Pos	61.0380	1.09	-1.658	NA	999.7
Ureidopropionic acid	Pos	133.0601	1.62	-0.683	0.27	NA
Uric acid	Pos	169.0355	2.14	-0.129	NA	927.6
	Neg	167.0206	2.00	-0.468	NA	926.2
Uridine	Neg	243.0623	1.88	-0.010	NA	706.3
Vanillylmandelic acid	Neg	197.0462	4.34	1.121	NA	690.4
Xanthosine	Pos	285.0826	4.70	-0.414	-0.02	913.9
	Neg	283.0687	4.75	0.246	0.03	996
Xanthurenic acid	Pos	206.0447	5.68	-0.134	-0.02	932.5
	Neg	204.0305	5.72	0.270	0.01	NA

* NA indicates that the MS/MS spectra of the feature were not acquired or the retention time of the metabolite was not included in the library.

7.3.5. Portability Test

The portability of the library and workflow was tested in another laboratory in Bremen, Germany. The retention time information in the constructed library was first transferred using the same RT-CalMix to generate the local version of the MS/MS-RT library. Then the same urine sample was processed and analyzed following the same procedures in the two labs. Figure 7.10 shows the identification results comparison in two labs. Very similar identification results were produced, indicating good inter-lab portability of the solution.

Polarity	Laboratory	# of "filtered" features	# of ID using m/z	# of ID using m/z+MS/MS	# of ID using m/z+RT	# of high-confidence ID features	# of identified metabolites
Positive	Canada	2460	188	66	109	127	98
	Germany	2986	210	70	114	139	115
Negative	Canada	3091	239	68	137	155	121
	Germany	2412	190	69	126	139	121
In total	Canada					240	177
	Germany					232	186

Figure 7.10 Inter-lab portability test of the metabolite identification solution.

7.3.6. Proof-of-concept Study

A proof-of-concept experiment was conducted to demonstrate the use of the entire solution for comparative metabolomics study. Urine samples were collected from a healthy volunteer before and after coffee intake. Sample preparation, LC-MS/MS analysis of samples, data processing and metabolite identification were step-by-step carried out following the workflow. Generated feature table and identification results were analyzed using the multivariate analysis tools (e.g. principle component analysis, PCA) and univariate analysis tools (e.g. volcano plot) in the Metaboscape software. Figure 7.11A-B show that PCA was able to fully separate the samples before and after drinking coffee. Volcano plot was used to find metabolites that have significant differences after coffee intake (Figure 7.11C). Some of them were able to be identified using the MS/MS-RT library with high confidence. Box plots of some examples were presented in Figure 7.11D. This proof-of-concept study proved that the new MS/MS-RT library and entire solution allowed rapid features identification and facilitated biological studies through metabolomics approach.

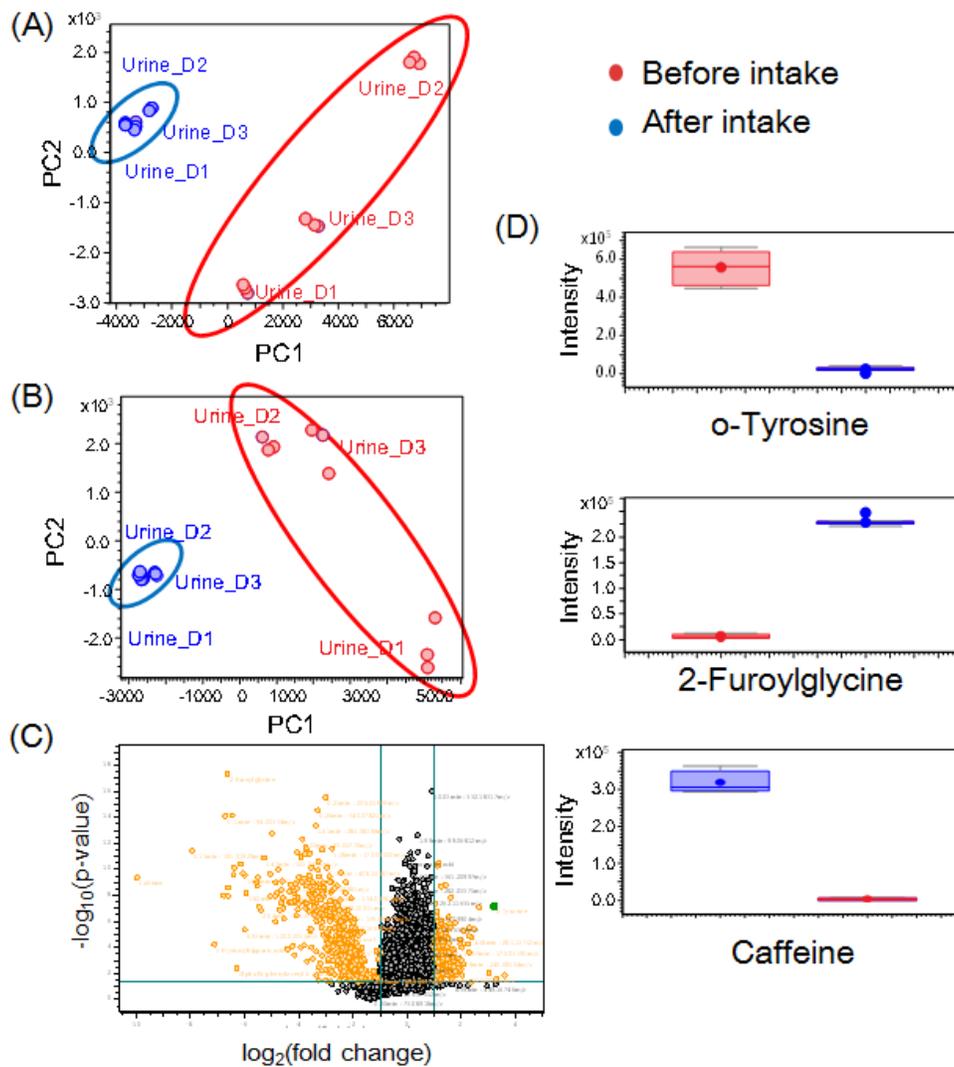


Figure 7.11 Application of the developed solution in a proof-of-concept experiment. The study compared the metabolome differences of urine samples before and after coffee intake. (A) PCA scores plot of LC-MS/MS data in positive ion mode. (B) PCA scores plot of LC-MS/MS data in negative ion mode. (C) Volcano plot showing significant metabolites in two groups. (D) Box plots of three significant metabolites in two groups as example.

7.4. Conclusions

In this work, a high-resolution MS/MS-RT library was constructed using RPLC-ESI-Q-TOF-MS instruments. 825 human endogenous metabolites were used to build the library. The current library does not contain exogenous compounds such as synthetic drugs. Thus this library is particularly useful for cellular metabolomics for studying metabolic perturbations and discovery metabolomics for finding endogenous metabolites as biomarkers of diseases. One important new feature of this library is the inclusion of RT values for RPLC. We have developed a strategy for correcting the RT shifts between the library RTs and experimental RTs using a set of compound standards and a correction algorithm. This method was shown to be effective to correct RT shifts caused by various experimental condition changes.

Based on the library, a high confidence and rapid metabolite identification solution was developed for metabolomics study. The entire solution contains optimized methods for sample preparation, RPLC-MS/MS analysis, data processing and metabolite identification. The performance of the solution was validated by analyzing real biological samples. The portability was also examined to ensure that end users can generate similar and consistent results as the solution developers. A proof-of-concept experiment was conducted to illustrate the workflow of using this solution for real comparative metabolomics study.

However, as shown in the library, nearly 50% of measured compounds were too polar to be retained on C18 reversed phase column. The next step is to introduce the hydrophilic interaction liquid chromatography (HILIC) to improve the analysis of polar metabolites. Combined both RPLC and HILIC data, the library and solution would be

able to provide more comprehensive and useful information for analyzing human metabolome.

Chapter 8

Conclusion and Future Work

8.1. Thesis Summary

LC-MS based metabolomics technology has become an emerging tool with broad applications in different areas such as systems biology, biomarker discovery, medical research, etc. However, because of the great diversity of chemical species and the wide concentration range of different metabolites, high coverage detection, accurate quantification and high-confidence identification of the entire metabolome are still challenges. To address these issues, a common approach is to combine several instrument setups and/or techniques (e.g., using RPLC for hydrophobic metabolites and HILIC for more polar compounds) for analysis. A number of compound repositories have been built to assist metabolite identification.

An alternative approach, the chemical isotope labeling (CIL) LC-MS method, has been developed. In this approach, the entire metabolome is divided into several submetabolomes consisting of compounds with common functional groups. Prior to LC-MS analysis of each submetabolome, metabolites in one group are derivatized with a pair of chemical isotope labeling reagents (e.g., $^{12}\text{C}/^{13}\text{C}$ -dansylation for amine/phenol submetabolome analysis). Integration of data from all submetabolomes can generate the result of the whole metabolome. Successfully applying this approach in metabolomics studies relies on 1) choosing the proper and high-performance labeling method for each submetabolome; 2) combining submetabolomes to form the entire or nearly the entire metabolome.

My research focuses on expanding CIL LC-MS methods to more submetabolomes to eventually achieve a comprehensive analysis of the entire metabolome and improving the metabolite identification power for metabolomics. The thesis work contains three parts. The first part carries out method development of expanding CIL LC-MS methods for different submetabolomes. The second part studies integration of multichannel CIL LC-MS methods for comprehensive metabolome profiling and clinical biomarker discovery. The third part aims at improving the identification power by constructing a high-resolution MS/MS-RT metabolite library. The main achievements of each project are summarized below.

Chapter 2-4 describe the development of novel CIL LC-MS methods for profiling three submetabolomes. In Chapter 2, DnsCl, a reagent that was commonly used in the CIL LC-MS method for amine/phenol-containing metabolites, was developed for performing relative quantification of the hydroxyl submetabolome with unprecedentedly high coverage. The dansylation makes poorly ionizable hydroxyls to be ionized as efficiently as the labeled amines/phenols, resulting in significantly improved and more unified detection sensitivity. Urinary hydroxyl submetabolome was analyzed, demonstrating the high reproducibility, accuracy and precision of the method. A hydroxyl standards library consisting of 85 unique compounds was constructed for metabolite identification.

In Chapter 3, using a similar approach, $^{12}\text{C}/^{13}\text{C}$ -DnsHz was developed as chemical isotope labeling reagents for carbonyl submetabolome. The developed method features easy to carry out, high accuracy and precision, and improved detectability. A DnsHz- labeled carbonyl standard library that currently consists of 78 endogenous human

metabolites was built to enhance metabolite identification confidence. This method, used alone, should be useful for profiling important groups of metabolites such as hormones and sugars in complex biological samples. It can also be used in combination with other methods to provide more comprehensive metabolome coverage.

DnsHz labeling was then migrated for analyzing carboxyl submetabolome in Chapter 4. Using different labeling reaction conditions, high coverage and in-depth carboxyl submetabolome profiling can be successfully performed with $^{12}\text{C}/^{13}\text{C}$ -DnsHz, which has been proved by urinary submetabolome analysis. A DnsHz-labeled carboxyl standard library was constructed, consisting of 193 authentic standards.

These methods, in combination with previously reported labeling methods, including amine/phenol labeling using $^{12}\text{C}/^{13}\text{C}$ -DnsCl and another carboxyl labeling method using $^{12}\text{C}/^{13}\text{C}$ -DmPA, can form multichannel chemical isotope labeling LC-MS (mCIL LC-MS) method system for comprehensively analyzing entire metabolome.

The integration, validation and application of mCIL LC-MS method was studied in Chapter 5-6. Chapter 5 described the integration and validation of the mCIL LC-MS approach for metabolomics. Four CIL LC-MS methods were integrated: dansylation for amines/phenols (reported previously), base-activated dansylation for hydroxyls (Chapter 2), DnsHz labeling for carbonyls (Chapter 3), and DmPA bromide labeling for carboxyls (reported previously and will be replaced by the method in Chapter 4 in the future). This 4-channel CIL LC-MS method was justified by group diversity analysis of common metabolite databases, which showed that the combination of the four submetabolomes would allow the analysis of the entire metabolome with very high coverage (85%-95%

according to different databases). Plasma metabolome were analyzed as an example for validating the performance of this approach.

The newly developed 4-channel CIL LC-MS approach was used for an application of biomarker discovery in Chapter 6. In this study, plasma metabolome of Alzheimer's disease (AD), cerebral amyloid angiopathy (CAA) and healthy controls were profiled and compared. In total of 69 plasma samples were analyzed with experimental duplicates using 4-channel CIL LC-MS method. A total of 11691 peak pairs were detected in common within 80% of the samples and 173 of them were positively identified using the labeled standard library, demonstrating high metabolome coverage and high-confidence identification power. Multivariate analysis showed that the approach successfully discriminated the metabolomic profile of three groups. Pair-wise comparisons were also carried out to find biomarker candidates for the two diseases and between diseases. A panel of five positively identified metabolites was selected as biomarkers for each comparison, all showing good discriminative ability. Pathway mapping was also performed to understand the metabolome alterations of two diseases. This study demonstrates that 4-channel CIL LC-MS approach is a powerful tool for biomarker discovery.

At last, Chapters 7 marks a shift from the CIL LC-MS approach to the conventional unlabeled LC-MS approach for metabolomics. In this project, a high-resolution MS/MS-RT library was constructed using 825 human endogenous metabolites. A RT calibration method was applied for correcting the RT shifts between the library RTs and experimental RTs, which improved RT matching accuracy. Based on the library, a high confidence and rapid metabolite identification solution was developed for

metabolomics study. The performance and portability were validated by analyzing various biological samples.

8.2. Future Work

First of all, although the mCIL LC-MS approach has been proven to be able to provide high metabolome coverage for metabolomics, there are still some metabolites that cannot be covered by this approach, such as amide-containing metabolites and ester-containing metabolites. To further improve the metabolic coverage, novel CIL LC-MS methods targeting uncovered submetabolomes need to be developed. Besides, special techniques are sometimes required to be combined with the CIL LC-MS method to improve the analytical power for a few categories of metabolites. For example, many phosphorylated metabolites are very hydrophilic and has low concentrations in bio-samples. Although most of them contain one or more of other functional groups that can be covered by mCIL LC-MS, it is very difficult to extract them from samples using conventional liquid-liquid extraction with organic solvents. Also, they have very poor retention on RPLC. Even after labeling, the current RPLC conditions are still not suitable for analyzing them. Therefore, a phosphate-specific enrichment method and optimal LC-MS condition need to be developed for targeting phosphorylated metabolites with CIL LC-MS method. In a pilot study, I applied immobilized metal affinity chromatography to enrich the phosphorylated metabolites, followed by CIL LC-MS analysis. The method displayed high metabolite recovery and high detection reproducibility.

Another approach to improve the metabolic coverage is to perform two dimensional LC-MS (2D LC-MS) analysis, which can embrace different separation techniques. Co-eluting metabolites in the first dimension can be further separated in the

second dimensional column, which can significantly reduce the ion suppression effect of the ionization process and improve metabolite detection. For example, preliminary results showed that dansylated metabolites have different retention behavior in mobile phases with high pH and low pH, suggesting that 2D LC-MS can be applied using different mobile phases to improve the detection of labeled metabolites.

Secondly, when applying the mCIL LC-MS method for analyzing large batches of samples in metabolomics studies, reducing analysis time for each sample can significantly increase the throughput. In the currently developed CIL LC-MS methods, although with different reaction conditions, DnsCl is used for both amine/phenol and hydroxyl submetabolome analysis. Thus, it provides an opportunity to combine the two submetabolome analysis in one single LC-MS run. In a proof-of-concept study, dansylated urine sample (for amine/phenol submetabolome) was mixed with the same sample processed by the base-activated dansylation reaction (for hydroxyl submetabolome), followed by LC-MS detection. The result of the combined mixture showed that it covered the majority of the labeled metabolites that can be detected in the two separate channels. A similar approach might be also applicable for DnsHz, which is developed for both carbonyl and carboxyl submetabolome profiling with different labeling reactions. If so, sample processed by 4-channel CIL method can be analyzed in two injection runs, which can reduce instrumental analysis time by half and significantly increase throughput.

Thirdly, hydrophilic interaction LC (HILIC), as a valuable complementary technique to reversed phase LC, has been widely used in metabolomics to improve metabolite coverage. Therefore, the next step of the MS/MS-RT library construction

project is to introduce the HILIC separation to improve the identification of polar metabolites. By combining RPLC and HILIC data together, the library and the identification solution should be able to provide more comprehensive and useful information for analyzing human metabolome.

Reference

1. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F., Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **1998**, *16* (9), 373-378.
2. Folin, O., On the determination of creatinine and creatine in urine. *J. Biol. Chem.* **1914**, *17* (4), 469-473.
3. Horning, E. C.; Horning, M. G., Metabolic Profiles - Gas-phase Methods for Analysis of Metabolites. *Clin. Chem.* **1971**, *17* (8), 802-&.
4. Pauling, L.; Robinson, A. B.; Teranish.R; Cary, P., Quantitative Analysis of Urine Vapor and Breath by Gas-liquid Partition Chromatography. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68* (10), 2374-&.
5. Thomas, G. H., Metabolomics breaks the silence. *Trends Biotechnol.* **2001**, *19* (4), 126-127.
6. Peng, B.; Li, H.; Peng, X. X., Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* **2015**, *6* (9), 628-637.
7. Khamis, M. M.; Adamko, D. J.; El-Aneed, A., Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom. Rev.* **2017**, *36* (2), 115-134.
8. Kumar, B.; Prakash, A.; Ruhela, R. K.; Medhi, B., Potential of metabolomics in preclinical and clinical drug development. *Pharmacol. Rep.* **2014**, *66* (6), 956-963.
9. Caldwell, G. W.; Leo, G. C., Can Untargeted Metabolomics Be Utilized in Drug Discovery/Development? *Current Topics in Medicinal Chemistry* **2017**, *17* (24), 2716-2739.
10. Burdisso, P.; Rasia, R. M.; Vila, A. J., METABOLOMICS AND MEDICAL PRECISION. *Rev. Med. Rosario* **2016**, *82* (2), 75-76.
11. Garcia-Canas, V.; Simo, C.; Herrero, M.; Ibanez, E.; Cifuentes, A., Present and Future Challenges in Food Analysis: Foodomics. *Analytical Chemistry* **2012**, *84* (23), 10150-10159.
12. Holland, N., Future of environmental research in the age of epigenomics and exposomics. *Rev. Environ. Health* **2017**, *32* (1-2), 45-54.
13. Aretz, I.; Meierhofer, D., Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *Int. J. Mol. Sci.* **2016**, *17* (5), 14.
14. Zhang, A. H.; Sun, H.; Yan, G. L.; Wang, P.; Wang, X. J., Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed. Chromatogr.* **2016**, *30* (1), 7-12.
15. Scalbert, A.; Brennan, L.; Manach, C.; Andres-Lacueva, C.; Dragsted, L. O.; Draper, J.; Rappaport, S. M.; van der Hoof, J. J. J.; Wishart, D. S., The food metabolome: a window over dietary exposure. *Am. J. Clin. Nutr.* **2014**, *99* (6), 1286-1308.
16. Liu, L.; Duff, K., A technique for serial collection of cerebrospinal fluid from the cisterna magna in mouse. *Journal of visualized experiments : JoVE* **2008**, (21).
17. Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vazquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C. R.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y. J.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y. F.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach,

- C.; Scalbert, A., HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D608-D617.
18. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27* (6), 747-751.
19. 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.
20. Jewison, T.; Knox, C.; Neveu, V.; Djoumbou, Y.; Guo, A. C.; Lee, J.; Liu, P.; Mandal, R.; Krishnamurthy, R.; Sinelnikov, I.; Wilson, M.; Wishart, D. S., YMDB: the Yeast Metabolome Database. *Nucleic Acids Res.* **2012**, *40* (D1), D815-D820.
21. Guo, A. C.; Jewison, T.; Wilson, M.; Liu, Y. F.; Knox, C.; Djoumbou, Y.; Lo, P.; Mandal, R.; Krishnamurthy, R.; Wishart, D. S., ECMDB: The E-coli Metabolome Database. *Nucleic Acids Res.* **2013**, *41* (D1), D625-D630.
22. Yuan, B. F.; Zhu, Q. F.; Guo, N.; Zheng, S. J.; Wang, Y. L.; Wang, J.; Xu, J.; Liu, S. J.; He, K.; Hu, T.; Zheng, Y. W.; Xu, F. Q.; Feng, Y. Q., Comprehensive Profiling of Fecal Metabolome of Mice by Integrated Chemical Isotope Labeling-Mass Spectrometry Analysis. *Analytical Chemistry* **2018**, *90* (5), 3512-3520.
23. Zhu, Y. T.; Deng, P.; Zhong, D. F., Derivatization methods for LC-MS analysis of endogenous compounds. *Bioanalysis* **2015**, *7* (19), 2557-2581.
24. Deng, P.; Zhan, Y.; Chen, X. Y.; Zhong, D. F., Derivatization methods for quantitative bioanalysis by LC-MS/MS. *Bioanalysis* **2012**, *4* (1), 49-69.
25. Higashi, T.; Ogawa, S., Isotope-coded ESI-enhancing derivatization reagents for differential analysis, quantification and profiling of metabolites in biological samples by LC/MS: A review. *J. Pharm. Biomed. Anal.* **2016**, *130*, 181-193.
26. 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.
27. Han, W.; Sapkota, S.; Camicioli, R.; Dixon, R. A.; Li, L., Profiling Novel Metabolic Biomarkers for Parkinson's Disease Using In-depth Metabolomic Analysis. *Mov. Disord.* **2017**, *32* (12), 1720-1728.
28. Jessome, L. L.; Volmer, D. A., Ion suppression: A major concern in mass spectrometry. *Lc Gc N. Am.* **2006**, *24* (5), 498-+.
29. Buszewski, B.; Noga, S., Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique. *Anal. Bioanal. Chem.* **2012**, *402* (1), 231-247.
30. 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. *Analytical Chemistry* **2009**, *81* (10), 3919-3932.
31. Zhao, S.; Luo, X.; Li, L., Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics. *Analytical Chemistry* **2016**, *88* (21), 10617-10623.
32. Zhu, Q. F.; Hao, Y. H.; Liu, M. Z.; Yue, J.; Ni, J.; Yuan, B. F.; Feng, Y. Q., Analysis of cytochrome P450 metabolites of arachidonic acid by stable isotope probe labeling coupled with ultra high-performance liquid chromatography/mass spectrometry. *J. Chromatogr. A* **2015**, *1410*, 154-163.

33. Li, S. F.; Jin, Y. B.; Tang, Z.; Lin, S. H.; Liu, H. X.; Jiang, Y. Y.; Cai, Z. W., A novel method of liquid chromatography-tandem mass spectrometry combined with chemical derivatization for the determination of ribonucleosides in urine. *Anal. Chim. Acta* **2015**, *864*, 30-38.
34. Bawazeer, S.; Ali, A. M.; Alhawiti, A.; Khalaf, A.; Gibson, C.; Tusiimire, J.; Watson, D. G., A method for the analysis of sugars in biological systems using reductive amination in combination with hydrophilic interaction chromatography and high resolution mass spectrometry. *Talanta* **2017**, *166*, 75-80.
35. Perez, S.; Barcelo, D., Applications of LC-MS to quantitation and evaluation of the environmental fate of chiral drugs and their metabolites. *Trac-Trends Anal. Chem.* **2008**, *27* (10), 836-846.
36. Nagao, R.; Tsutsui, H.; Mochizuki, T.; Takayama, T.; Kuwabara, T.; Min, J. Z.; Inoue, K.; Todoroki, K.; Toyo'oka, T., Novel chiral derivatization reagents possessing a pyridylthiourea structure for enantiospecific determination of amines and carboxylic acids in high-throughput liquid chromatography and electrospray-ionization mass spectrometry for chiral metabolomics identification. *J. Chromatogr. A* **2013**, *1296*, 111-118.
37. Mochizuki, T.; Todoroki, K.; Inoue, K.; Min, J. Z.; Toyo'oka, T., Isotopic variants of light and heavy L-pyroglutamic acid succinimidyl esters as the derivatization reagents for DL-amino acid chiral metabolomics identification by liquid chromatography and electrospray ionization mass spectrometry. *Anal. Chim. Acta* **2014**, *811*, 51-59.
38. Mochizuki, T.; Takayama, T.; Todoroki, K.; Inoue, K.; Min, J. Z.; Toyo'oka, T., Towards the chiral metabolomics: Liquid chromatography-mass spectrometry based DL-amino acid analysis after labeling with a new chiral reagent, (S)-2,5-dioxopyrrolidin-1-yl-1-(4,6-dimethoxy-1,3,5-triazin-2-yl)pyrrolidine-2-carboxylate, and the application to saliva of healthy volunteers. *Anal. Chim. Acta* **2015**, *875*, 73-82.
39. Takayama, T.; Kuwabara, T.; Maeda, T.; Noge, I.; Kitagawa, Y.; Inoue, K.; Todoroki, K.; Min, J. Z.; Toyo'oka, T., Profiling of chiral and achiral carboxylic acid metabolomics: synthesis and evaluation of triazine-type chiral derivatization reagents for carboxylic acids by LC-ESI-MS/MS and the application to saliva of healthy volunteers and diabetic patients. *Anal. Bioanal. Chem.* **2015**, *407* (3), 1003-1014.
40. Takayama, T.; Mochizuki, T.; Todoroki, K.; Min, J. Z.; Mizuno, H.; Inoue, K.; Akatsu, H.; Noge, I.; Toyo'oka, T., A novel approach for LC-MS/MS-based chiral metabolomics fingerprinting and chiral metabolomics extraction using a pair of enantiomers of chiral derivatization reagents. *Anal. Chim. Acta* **2015**, *898*, 73-84.
41. Johnson, D. W., Free amino acid quantification by LC-MS/MS using derivatization generated isotope-labelled standards. *J. Chromatogr. B* **2011**, *879* (17-18), 1345-1352.
42. Guo, N.; Liu, P.; Ding, J.; Zheng, S. J.; Yuan, B. F.; Feng, Y. Q., Stable isotope labeling - Liquid chromatography/mass spectrometry for quantitative analysis of androgenic and progestagenic steroids. *Anal. Chim. Acta* **2016**, *905*, 106-114.
43. Woo, H. K.; Go, E. P.; Hoang, L.; Trauger, S. A.; Bowen, B.; Siuzdak, G.; Northen, T. R., Phosphonium labeling for increasing metabolomic coverage of neutral lipids using electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23* (12), 1849-1855.

44. 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.
45. Higashi, T.; Aiba, N.; Tanaka, T.; Yoshizawa, K.; Ogawa, S., Methods for differential and quantitative analyses of brain neurosteroid levels by LC/MS/MS with ESI-enhancing and isotope-coded derivatization. *J. Pharm. Biomed. Anal.* **2016**, *117*, 155-162.
46. Comini, M. A., Measurement and meaning of cellular thiol:disulfide redox status. *Free Radic. Res.* **2016**, *50* (2), 246-271.
47. Ortmayr, K.; Schwaiger, M.; Hann, S.; Koellensperger, G., An integrated metabolomics workflow for the quantification of sulfur pathway intermediates employing thiol protection with N-ethyl maleimide and hydrophilic interaction liquid chromatography tandem mass spectrometry. *Analyst* **2015**, *140* (22), 7687-7695.
48. Zhou, R.; Tseng, C. L.; Huan, T.; Li, L., IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform. *Analytical Chemistry* **2014**, *86* (10), 4675-4679.
49. Guo, N.; Peng, C. Y.; Zhu, Q. F.; Yuan, B. F.; Feng, Y. Q., Profiling of carbonyl compounds in serum by stable isotope labeling Double precursor ion scan - Mass spectrometry analysis. *Anal. Chim. Acta* **2017**, *967*, 42-51.
50. Zhu, Q. F.; Zhang, Z.; Liu, P.; Zheng, S. J.; Peng, K.; Deng, Q. Y.; Zheng, F.; Yuan, B. F.; Feng, Y. Q., Analysis of liposoluble carboxylic acids metabolome in human serum by stable isotope labeling coupled with liquid chromatography-mass spectrometry. *J. Chromatogr. A* **2016**, *1460*, 100-109.
51. Khamis, M. M.; Adamko, D. J.; El-Aneed, A., Development of a validated LC-MS/MS method for the quantification of 19 endogenous asthma/COPD potential urinary biomarkers. *Anal. Chim. Acta* **2017**, *989*, 45-58.
52. Su, X. L.; Wang, N.; Chen, D. Y.; Li, Y. N.; Lu, Y. F.; Huan, T.; Xu, W.; Li, L.; Li, L. J., Dansylation isotope labeling liquid chromatography mass spectrometry for parallel profiling of human urinary and fecal submetabolomes. *Anal. Chim. Acta* **2016**, *903*, 100-109.
53. Chen, D. Y.; Han, W.; Su, X. L.; Li, L.; Li, L. J., Overcoming Sample Matrix Effect in Quantitative Blood Metabolomics Using Chemical Isotope Labeling Liquid Chromatography Mass Spectrometry. *Analytical Chemistry* **2017**, *89* (17), 9424-9431.
54. Hooton, K.; Han, W.; Li, L., Comprehensive and Quantitative Profiling of the Human Sweat Submetabolome Using High-Performance Chemical Isotope Labeling LC-MS. *Analytical Chemistry* **2016**, *88* (14), 7378-7386.
55. 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.
56. Xu, W.; Chen, D. Y.; Wang, N.; Zhang, T.; Zhou, R. K.; Huan, T.; Lu, Y. F.; Su, X. L.; Xie, Q.; Li, L.; Li, L. J., Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Human Fecal Metabolome. *Analytical Chemistry* **2015**, *87* (2), 829-836.

57. Luo, X.; Li, L., Metabolomics of Small Numbers of Cells: Metabolomic Profiling of 100, 1000, and 10000 Human Breast Cancer Cells. *Analytical Chemistry* **2017**, *89* (21), 11664-11671.
58. Luo, X.; Zhao, S.; Huan, T.; Sun, D. F.; Friis, R. M. N.; Schultz, M. C.; Li, L., High-Performance Chemical Isotope Labeling Liquid Chromatography Mass Spectrometry for Profiling the Metabolomic Reprogramming Elicited by Ammonium Limitation in Yeast. *Journal of Proteome Research* **2016**, *15* (5), 1602-1612.
59. Huan, T.; Wu, Y. M.; Tang, C. Q.; Lin, G. H.; Li, L., DnsID in MyCompoundID for Rapid Identification of Dansylated Amine- and Phenol-Containing Metabolites in LC-MS-Based Metabolomics. *Analytical Chemistry* **2015**, *87* (19), 9838-9845.
60. Wu, M. H.; Xu, Y.; Fitch, W. L.; Zheng, M.; Merritt, R. E.; Shrager, J. B.; Zhang, W. R.; Dill, D. L.; Peltz, G.; Hoang, C. D., Liquid chromatography/mass spectrometry methods for measuring dipeptide abundance in non-small-cell lung cancer. *Rapid Commun. Mass Spectrom.* **2013**, *27* (18), 2091-2098.
61. Zhou, R. K.; Guo, K.; Li, L., 5-Diethylamino-naphthalene-1-sulfonyl Chloride (DensCl): A Novel Triplex Isotope Labeling Reagent for Quantitative Metabolome Analysis by Liquid Chromatography Mass Spectrometry. *Analytical Chemistry* **2013**, *85* (23), 11532-11539.
62. Guo, H. M.; Jiao, Y.; Wang, X.; Lu, T.; Zhang, Z. J.; Xu, F. G., Twins labeling-liquid chromatography/mass spectrometry based metabolomics for absolute quantification of tryptophan and its key metabolites. *J. Chromatogr. A* **2017**, *1504*, 83-90.
63. Cai, Y. P.; Sun, Z. W.; Chen, G.; Liu, X. M.; You, J. M.; Zhang, C. Q., Rapid analysis of biogenic amines from rice wine with isotope-coded derivatization followed by high performance liquid chromatography-tandem mass spectrometry. *Food Chem.* **2016**, *192*, 388-394.
64. Song, C. H.; Zhang, S. J.; Ji, Z. Y.; Li, Y. P.; You, J. M., Accurate Determination of Amino Acids in Serum Samples by Liquid Chromatography-Tandem Mass Spectrometry Using a Stable Isotope Labeling Strategy. *J. Chromatogr. Sci.* **2015**, *53* (9), 1536-1541.
65. 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.
66. Che, D. D.; Sun, Z. W.; Cheng, J.; Dou, K.; Ji, Z. Y.; Chen, G.; Li, G. L.; You, J. M., Determination of parabens in domestic sewage by isotope-coded derivatization coupled with high performance liquid chromatography-tandem mass spectrometry. *Microchem J.* **2017**, *130*, 420-427.
67. Zhang, S. J.; Yu, Q. H.; Sheng, C. C.; You, J. M., Gas Purge Microextraction Coupled with Stable Isotope Labeling-Liquid Chromatography/Mass Spectrometry for the Analysis of Bromophenols in Aquatic Products. *J. Agric. Food Chem.* **2016**, *64* (49), 9452-9458.
68. Wong, J. M. T.; Malec, P. A.; Mabrouk, O. S.; Ro, J.; Dus, M.; Kennedy, R. T., Benzoyl chloride derivatization with liquid chromatography-mass spectrometry for targeted metabolomics of neurochemicals in biological samples. *J. Chromatogr. A* **2016**, *1446*, 78-90.

69. Takach, E.; O'Shea, T.; Liu, H. L., High-throughput quantitation of amino acids in rat and mouse biological matrices using stable isotope labeling and UPLC-MS/MS analysis. *J. Chromatogr. B* **2014**, *964*, 180-190.
70. Hao, L.; Zhong, X. F.; Greer, T.; Ye, H.; Li, L. J., Relative quantification of amine-containing metabolites using isobaric N,N-dimethyl leucine (DiLeu) reagents via LC-ESI-MS/MS and CE-ESI-MS/MS. *Analyst* **2015**, *140* (2), 467-475.
71. Wagner, M.; Ohlund, L. B.; Shiao, T. C.; Vezina, A.; Annabi, B.; Roy, R.; Sleno, L., Isotope-labeled differential profiling of metabolites using N-benzoyloxysuccinimide derivatization coupled to liquid chromatography/high-resolution tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2015**, *29* (18), 1632-1640.
72. Shimbo, K.; Oonuki, T.; Yahashi, A.; Hirayama, K.; Miyano, H., Precolumn derivatization reagents for high-speed analysis of amines and amino acids in biological fluid using liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2009**, *23* (10), 1483-1492.
73. Zhang, S. S.; Shi, J. W.; Shan, C. K.; Huang, C. T.; Wu, Y. L.; Ding, R.; Xue, Y. H.; Liu, W.; Zhou, Q.; Zhao, Y. F.; Xu, P. X.; Gao, X., Stable isotope N-phosphoryl amino acids labeling for quantitative profiling of amine-containing metabolites using liquid chromatography mass spectrometry. *Anal. Chim. Acta* **2017**, *978*, 24-34.
74. Zhou, R.; Huan, T.; Li, L., Development of versatile isotopic labeling reagents for profiling the amine submetabolome by liquid chromatography-mass spectrometry. *Anal. Chim. Acta* **2015**, *881*, 107-116.
75. Santa, T., Isothiocyanates as derivatization reagents for amines in liquid chromatography/electrospray ionization-tandem mass spectrometry. *Biomed. Chromatogr.* **2010**, *24* (9), 915-918.
76. Hiraoka, K.; Kudaka, I., NEGATIVE-MODE ELECTROSPRAY-MASS SPECTROMETRY USING NONAQUEOUS SOLVENTS. *Rapid Commun. Mass Spectrom.* **1992**, *6* (4), 265-268.
77. Bartok, T.; Komoroczy, R.; Boresok, G.; Sagi, F.; Hegyes, P.; Bartok, M., A new derivatization reaction for the identification and determination of organic acids by positive-ion electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997**, *11* (1), 133-135.
78. Marquis, B. J.; Louks, H. P.; Bose, C.; Wolfe, R. R.; Singh, S. P., A New Derivatization Reagent for HPLC-MS Analysis of Biological Organic Acids. *Chromatographia* **2017**, *80* (12), 1723-1732.
79. Jiang, R. Q.; Jiao, Y.; Zhang, P.; Liu, Y.; Wang, X.; Huang, Y.; Zhang, Z. J.; Xu, F. G., Twin Derivatization Strategy for High-Coverage Quantification of Free Fatty Acids by Liquid Chromatography-Tandem Mass Spectrometry. *Analytical Chemistry* **2017**, *89* (22), 12223-12230.
80. Han, J.; Gagnon, S.; Eckle, T.; Borchers, C. H., Metabolomic analysis of key central carbon metabolism carboxylic acids as their 3-nitrophenylhydrazones by UPLC/ESI-MS. *Electrophoresis* **2013**, *34* (19), 2891-2900.
81. Han, J.; Lin, K.; Sequeira, C.; Borchers, C. H., An isotope-labeled chemical derivatization method for the quantitation of short-chain fatty acids in human feces by liquid chromatography-tandem mass spectrometry. *Anal. Chim. Acta* **2015**, *854*, 86-94.
82. Han, J.; Liu, Y.; Wang, R. X.; Yang, J. C.; Ling, V.; Borchers, C. H., Metabolic Profiling of Bile Acids in Human and Mouse Blood by LC-MS/MS in Combination with

Phospholipid-Depletion Solid-Phase Extraction. *Analytical Chemistry* **2015**, *87* (2), 1127-1136.

83. Harder, U.; Koletzko, B.; Peissner, W., Quantification of 22 plasma amino acids combining derivatization and ion-pair LC-MS/MS. *J. Chromatogr. B* **2011**, *879* (7-8), 495-504.

84. O'Maille, G.; Go, E. P.; Hoang, L.; Want, E. J.; Smith, C.; O'Maille, P.; Nordstrom, A.; Morita, H.; Qin, C.; Uritboonthai, W.; Apon, J.; Moore, R.; Garrett, J.; Siuzdak, G., Metabolomics relative quantitation with mass spectrometry using chemical derivatization and isotope labeling. *Spectr.-Int. J.* **2008**, *22* (5), 327-343.

85. 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.

86. Griffiths, W. J.; Hornshaw, M.; Woffendin, G.; Baker, S. F.; Lockhart, A.; Heidelberger, S.; Gustafsson, M.; Sjoall, J.; Wang, Y. Q., Discovering oxysterols in plasma: A window on the metabolome. *Journal of Proteome Research* **2008**, *7* (8), 3602-3612.

87. DeBarber, A. E.; Sandler, Y.; Pappu, A. S.; Merckens, L. S.; Duell, P. B.; Lear, S. R.; Erickson, S. K.; Steiner, R. D., Profiling sterols in cerebrotendinous xanthomatosis: Utility of Girard derivatization and high resolution exact mass LC-ESI-MSⁿ analysis. *J. Chromatogr. B* **2011**, *879* (17-18), 1384-1392.

88. Wang, C. J.; Wu, Z. Y.; Yuan, J. B.; Wang, B.; Zhang, P.; Zhang, Y.; Wang, Z. F.; Huang, L. J., Simplified Quantitative Glycomics Using the Stable Isotope Label Girard's Reagent P by Electrospray Ionization Mass Spectrometry. *Journal of Proteome Research* **2014**, *13* (2), 372-384.

89. Yang, W. C.; Sedlak, M.; Regnier, F. E.; Mosier, N.; Ho, N.; Adamec, J., Simultaneous Quantification of Metabolites Involved in Central Carbon and Energy Metabolism Using Reversed-Phase Liquid Chromatography-Mass Spectrometry and in Vitro C-13 Labeling. *Analytical Chemistry* **2008**, *80* (24), 9508-9516.

90. Eggink, M.; Wijtmans, M.; Ekkebus, R.; Lingeman, H.; de Esch, I. J. P.; Kool, J.; Niessen, W. M. A.; Irth, H., Development of a Selective ESI-MS Derivatization Reagent: Synthesis and Optimization for the Analysis of Aldehydes in Biological Mixtures. *Analytical Chemistry* **2008**, *80* (23), 9042-9051.

91. Yu, L.; Liu, P.; Wang, Y. L.; Yu, Q. W.; Yuan, B. F.; Feng, Y. Q., Profiling of aldehyde-containing compounds by stable isotope labelling-assisted mass spectrometry analysis. *Analyst* **2015**, *140* (15), 5276-5286.

92. Zheng, S. J.; Wang, Y. L.; Liu, P.; Zhang, Z.; Yu, L.; Yuan, B. F.; Feng, Y. Q., Stable isotope labeling-solid phase extraction-mass spectrometry analysis for profiling of thiols and aldehydes in beer. *Food Chem.* **2017**, *237*, 399-407.

93. Tang, Z.; Guengerich, F. P., Dansylation of Unactivated Alcohols for Improved Mass Spectral Sensitivity and Application to Analysis of Cytochrome P450 Oxidation Products in Tissue Extracts. *Analytical Chemistry* **2010**, *82* (18), 7706-7712.

94. 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.

95. Cao, Y. J.; Guan, Q.; Sun, T. Q.; Wang, H.; Leng, J. P.; Guo, Y. L., N-alkylpyridinium quaternization combined with liquid chromatography-electrospray ionization-tandem mass spectrometry: A highly sensitive method to quantify fatty alcohols in thyroid tissues. *Anal. Chim. Acta* **2014**, *849*, 19-26.
96. Liu, P.; Huang, Y. Q.; Cai, W. J.; Yuan, B. F.; Feng, Y. Q., Profiling of Thiol-Containing Compounds by Stable Isotope Labeling Double Precursor Ion Scan Mass Spectrometry. *Analytical Chemistry* **2014**, *86* (19), 9765-9773.
97. Liu, P.; Qi, C. B.; Zhu, Q. F.; Yuan, B. F.; Feng, Y. Q., Determination of thiol metabolites in human urine by stable isotope labeling in combination with pseudo-targeted mass spectrometry analysis. *Sci Rep* **2016**, *6*, 12.
98. Zheng, L. F.; Zhao, X. E.; Zhu, S. Y.; Tao, Y. D.; Ji, W. H.; Geng, Y. L.; Wang, X. A.; Chen, G. A.; You, J. M., A new combined method of stable isotope-labeling derivatization-ultrasound-assisted dispersive liquid-liquid microextraction for the determination of neurotransmitters in rat brain microdialysates by ultra high performance liquid chromatography tandem mass spectrometry. *J. Chromatogr. B* **2017**, *1054*, 64-72.
99. Lee, D. Y.; Chang, G. D., Quantitative liquid chromatography-electrospray ionization-mass spectrometry analysis of amine-containing metabolites derivatized with cyanuric chloride and methylamine isotopologues. *J. Chromatogr. A* **2015**, *1388*, 60-68.
100. Wang, X.; Wei, F.; Xu, J. Q.; Lv, X.; Dong, X. Y.; Han, X. L.; Quek, S. Y.; Huang, F. H.; Chen, H., Profiling and relative quantification of phosphatidylethanolamine based on acetone stable isotope derivatization. *Anal. Chim. Acta* **2016**, *902*, 142-153.
101. Yang, W. C.; Mirzaei, H.; Liu, X. P.; Regnier, F. E., Enhancement of amino acid detection and quantification by electrospray ionization mass spectrometry. *Analytical Chemistry* **2006**, *78* (13), 4702-4708.
102. Shortreed, M. R.; Lamos, S. M.; Frey, B. L.; Phillips, M. F.; Patel, M.; Belshaw, P. J.; Smith, L. M., Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry. *Analytical Chemistry* **2006**, *78* (18), 6398-6403.
103. Guo, K.; Ji, C. J.; Li, L., Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. *Analytical Chemistry* **2007**, *79* (22), 8631-8638.
104. Ji, C. J.; Li, W. L.; Ren, X. D.; El-Kattan, A. F.; Kozak, R.; Fountain, S.; Lepsy, C., Diethylation Labeling Combined with UPLC/MS/MS for Simultaneous Determination of a Panel of Monoamine Neurotransmitters in Rat Prefrontal Cortex Microdialysates. *Analytical Chemistry* **2008**, *80* (23), 9195-9203.
105. Abello, N.; Geurink, P. P.; van der Toorn, M.; van Oosterhout, A. J. M.; Lugtenburg, J.; van der Marel, G. A.; Kerstjens, H. A. M.; Postma, D. S.; Overkleeft, H. S.; Bischoff, R., Poly(ethylene glycol)-Based Stable Isotope Labeling Reagents for the Quantitative Analysis of Low Molecular Weight Metabolites by LC-MS. *Analytical Chemistry* **2008**, *80* (23), 9171-9180.
106. Shimbo, K.; Yahashi, A.; Hirayama, K.; Nakazawa, M.; Miyano, H., Multifunctional and Highly Sensitive Precolumn Reagents for Amino Acids in Liquid Chromatography/Tandem Mass Spectrometry. *Analytical Chemistry* **2009**, *81* (13), 5172-5179.
107. Lamos, S. M.; Shortreed, M. R.; Frey, B. L.; Belshaw, P. J.; Smith, L. M., Relative quantification of carboxylic acid metabolites by liquid chromatography - Mass

- spectrometry using isotopic variants of cholamine. *Analytical Chemistry* **2007**, *79* (14), 5143-5149.
108. Tayyari, F.; Gowda, G. A. N.; Gu, H. W.; Raftery, D., N-15-Cholamine-A Smart Isotope Tag for Combining NMR- and MS-Based Metabolite Profiling. *Analytical Chemistry* **2013**, *85* (18), 8715-8721.
109. Yang, W. C.; Regnier, F. E.; Adamec, J., Comparative metabolite profiling of carboxylic acids in rat urine by CE-ESI MS/MS through positively pre-charged and H-2-coded derivatization. *Electrophoresis* **2008**, *29* (22), 4549-4560.
110. Yang, W. C.; Adamec, J.; Regnier, F. E., Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding. *Analytical Chemistry* **2007**, *79* (14), 5150-5157.
111. Koulman, A.; Petras, D.; Narayana, V. K.; Wang, L.; Volmer, D. A., Comparative High-Speed Profiling of Carboxylic Acid Metabolite Levels by Differential Isotope-Coded MALDI Mass Spectrometry. *Analytical Chemistry* **2009**, *81* (18), 7544-7551.
112. Tsukamoto, Y.; Santa, T.; Yoshida, H.; Miyano, H.; Fukushima, T.; Hirayama, K.; Imai, K.; Funatsu, T., Synthesis of the isotope-labeled derivatization reagent for carboxylic acids: 7-(N,N-dimethylaminosulfonyl)-4-(aminoethyl)piperazino-2,1,3-benzoxadiazole (d(6)) DBD-PZ-NH₂ (D), and its application to the quantification and the determination of relative amount of fatty acids in rat plasma samples by high-performance liquid chromatography/mass spectrometry. *Biomed. Chromatogr.* **2006**, *20* (4), 358-364.
113. Leng, J. P.; Wang, H. Y.; Zhang, L.; Zhang, J.; Wang, H.; Guo, Y. L., A highly sensitive isotope-coded derivatization method and its application for the mass spectrometric analysis of analytes containing the carboxyl group. *Anal. Chim. Acta* **2013**, *758*, 114-121.
114. Leng, J. P.; Guan, Q.; Sun, T. Q.; Wang, H. Y.; Cui, J. L.; Liu, Q. H.; Zhang, Z. X.; Zhang, M. Y.; Guo, Y. L., Direct infusion electrospray ionization-ion mobility-mass spectrometry for comparative profiling of fatty acids based on stable isotope labeling. *Anal. Chim. Acta* **2015**, *887*, 148-154.
115. Tie, C.; Hu, T.; Zhang, X. X.; Zhou, J.; Zhang, J. L., HPLC-MRM relative quantification analysis of fatty acids based on a novel derivatization strategy. *Analyst* **2014**, *139* (23), 6154-6159.
116. Crick, P. J.; Bentley, T. W.; Abdel-Khalik, J.; Matthews, I.; Clayton, P. T.; Morris, A. A.; Bigger, B. W.; Zerbinati, C.; Tritapepe, L.; Iuliano, L.; Wang, Y. Q.; Griffiths, W. J., Quantitative Charge-Tags for Sterol and Oxysterol Analysis. *Clin. Chem.* **2015**, *61* (2), 400-411.
117. Tie, C.; Hu, T.; Jia, Z. X.; Zhang, J. L., Derivatization Strategy for the Comprehensive Characterization of Endogenous Fatty Aldehydes Using HPLC-Multiple Reaction Monitoring. *Analytical Chemistry* **2016**, *88* (15), 7762-7768.
118. DeBarber, A. E.; Luo, J.; Star-Weinstock, M.; Purkayastha, S.; Geraghty, M. T.; Chiang, J.; Merckens, L. S.; Pappu, A. S.; Steiner, R. D., A blood test for cerebrotendinous xanthomatosis with potential for disease detection in newborns. *J. Lipid Res.* **2014**, *55* (1), 146-154.
119. Chu, J. M.; Qi, C. B.; Huang, Y. Q.; Jiang, H. P.; Hao, Y. H.; Yuan, B. F.; Feng, Y. Q., Metal Oxide-Based Selective Enrichment Combined with Stable Isotope Labeling-

Mass Spectrometry Analysis for Profiling of Ribose Conjugates. *Analytical Chemistry* **2015**, *87* (14), 7364-7372.

120. Wang, H.; Wang, H. Y.; Zhang, L.; Zhang, J.; Leng, J. P.; Cai, T. T.; Guo, Y. L., Improvement and extension of the application scope for matrix-assisted laser desorption/ionization mass spectrometric analysis-oriented N-alkylpyridinium isotope quaternization. *Anal. Chim. Acta* **2011**, *707* (1), 100-106.

121. Huang, Y. Q.; Ruan, G. D.; Liu, J. Q.; Gao, Q.; Feng, Y. Q., Use of isotope differential derivatization for simultaneous determination of thiols and oxidized thiols by liquid chromatography tandem mass spectrometry. *Anal. Biochem.* **2011**, *416* (2), 159-166.

122. Bian, X. Q.; Sun, B. Q.; Zheng, P. Y.; Li, N.; Wu, J. L., Derivatization enhanced separation and sensitivity of long chain-free fatty acids: Application to asthma using targeted and non-targeted liquid chromatography-mass spectrometry approach. *Anal. Chim. Acta* **2017**, *989*, 59-70.

123. Siegel, D.; Meinema, A. C.; Permentier, H.; Hopfgartner, G.; Bischoff, R., Integrated Quantification and Identification of Aldehydes and Ketones in Biological Samples. *Analytical Chemistry* **2014**, *86* (10), 5089-5100.

124. Turowski, M.; Yamakawa, N.; Meller, J.; Kimata, K.; Ikegami, T.; Hosoya, K.; Tanaka, N.; Thornton, E. R., Deuterium isotope effects on hydrophobic interactions: The importance of dispersion interactions in the hydrophobic phase. *J. Am. Chem. Soc.* **2003**, *125* (45), 13836-13849.

125. Szarka, S.; Prokai-Tatrai, K.; Prokai, L., Application of Screening Experimental Designs to Assess Chromatographic Isotope Effect upon Isotope-Coded Derivatization for Quantitative Liquid Chromatography-Mass Spectrometry. *Analytical Chemistry* **2014**, *86* (14), 7033-7040.

126. Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R., Isotopic labeling-assisted metabolomics using LC-MS. *Anal. Bioanal. Chem.* **2013**, *405* (1), 27-33.

127. 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.

128. Tang, Z. M.; Guengerich, F. P., Dansylation of Unactivated Alcohols for Improved Mass Spectral Sensitivity and Application to Analysis of Cytochrome P450 Oxidation Products in Tissue Extracts. *Anal. Chem.* **2010**, *82* (18), 7706-7712.

129. Yang, W. C.; Regnier, F. E.; Jiang, Q.; Adamec, J., In vitro stable isotope labeling for discovery of novel metabolites by liquid chromatography-mass spectrometry: Confirmation of gamma-tocopherol metabolism in human A549 cell. *J. Chromatogr. A* **2010**, *1217* (5), 667-675.

130. 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.

131. Yuan, W.; Zhang, J. X.; Li, S. W.; Edwards, J. L., Amine Metabolomics of Hyperglycemic Endothelial Cells using Capillary LC-MS with Isobaric Tagging. *Journal of Proteome Research* **2011**, *10* (11), 5242-5250.
132. Escrig-Domenech, A.; Simo-Alfonso, E. F.; Herrero-Martinez, J. M.; Ramis-Ramos, G., Derivatization of hydroxyl functional groups for liquid chromatography and capillary electroseparation. *J. Chromatogr. A* **2013**, *1296*, 140-156.
133. Yuan, W.; Edwards, J. L.; Li, S. W., Global profiling of carbonyl metabolites with a photo-cleavable isobaric labeling affinity tag. *Chem. Commun.* **2013**, *49* (94), 11080-11082.
134. Roy-Lachapelle, A.; Sollicec, M.; Sauve, S., Determination of BMAA and three alkaloid cyanotoxins in lake water using dansyl chloride derivatization and high-resolution mass spectrometry. *Anal. Bioanal. Chem.* **2015**, *407* (18), 5487-5501.
135. Arrivault, S.; Guenther, M.; Fry, S. C.; Fuenfgeld, M.; Veyel, D.; Mettler-Altmann, T.; Stitt, M.; Lunn, J. E., Synthesis and Use of Stable-Isotope-Labeled Internal Standards for Quantification of Phosphorylated Metabolites by LC-MS/MS. *Analytical Chemistry* **2015**, *87* (13), 6896-6904.
136. Baghdady, Y. Z.; Schug, K. A., Review of in situ derivatization techniques for enhanced bioanalysis using liquid chromatography with mass spectrometry. *J. Sep. Sci.* **2016**, *39* (1), 102-114.
137. Liu, P.; Qi, C. B.; Zhu, Q. F.; Yuan, B. F.; Feng, Y. Q., Determination of thiol metabolites in human urine by stable isotope labeling in combination with pseudo-targeted mass spectrometry analysis. *Sci Rep* **2016**, *6*, 21433; doi:10.1038/srep21433.
138. Yu, L.; Ding, J.; Wang, Y. L.; Liu, P.; Feng, Y. Q., 4-Phenylaminomethyl-Benzeneboric Acid Modified Tip Extraction for Determination of Brassinosteroids in Plant Tissues by Stable Isotope Labeling-Liquid Chromatography-Mass Spectrometry. *Analytical Chemistry* **2016**, *88* (2), 1286-1293.
139. Anari, M. R.; Bakhtiar, R.; Zhu, B.; Huskey, S.; Franklin, R. B.; Evans, D. C., Derivatization of ethinylestradiol with dansyl chloride to enhance electrospray ionization: Application in trace analysis of ethinylestradiol in rhesus monkey plasma. *Anal. Chem.* **2002**, *74* (16), 4136-4144.
140. Huan, T.; Li, L., Counting Missing Values in a Metabolite-Intensity Data Set for Measuring the Analytical Performance of a Metabolomics Platform. *Analytical Chemistry* **2015**, *87* (2), 1306-1313.
141. Huan, T.; Li, L., Quantitative Metabolome Analysis Based on Chromatographic Peak Reconstruction in Chemical Isotope Labeling Liquid Chromatography Mass Spectrometry. *Analytical Chemistry* **2015**, *87* (14), 7011-7016.
142. Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G., MyCompoundID: using an evidence-based metabolome library for metabolite identification. *Anal. Chem.* **2013**, *85* (6), 3401-8.
143. Peng, J.; Li, L., Liquid-liquid extraction combined with differential isotope dimethylaminophenacyl labeling for improved metabolomic profiling of organic acids. *Anal. Chim. Acta* **2013**, *803*, 97-105.
144. Wu, Y. M.; 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. *Analytical Chemistry* **2012**, *84* (24), 10723-10731.

145. Newgard, C. B., Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metab.* **2017**, *25* (1), 43-56.
146. *Normal Laboratory Values for Blood, Urine, CSF, Stool, and Other Fluids*, Merck Manual Homepage. <http://www.merckmanuals.com> (accessed March 20, 2017).
147. Gowda, G. A. N.; Raftery, D., Recent Advances in NMR-Based Metabolomics. *Anal. Chem.* **2017**, *89* (1), 490-510.
148. 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.
149. Ortmayr, K.; Causon, T. J.; Hann, S.; Koellensperger, G., Increasing selectivity and coverage in LC-MS based metabolome analysis. *Trac-Trends Anal. Chem.* **2016**, *82*, 358-366.
150. Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. J., Mass Defect-Based N,N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells. *Analytical Chemistry* **2017**, *89* (2), 1138-1146.
151. Fan, R. J.; Guan, Q.; Zhang, F.; Leng, J. P.; Sun, T. Q.; Guo, Y. L., Benzylic rearrangement stable isotope labeling for quantitation of guanidino and ureido compounds in thyroid tissues by liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chim. Acta* **2016**, *908*, 132-140.
152. 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.
153. Dong, J. Z.; Moldoveanu, S. C., Gas chromatography-mass spectrometry of carbonyl compounds in cigarette mainstream smoke after derivatization with 2,4-dinitrophenylhydrazine. *J. Chromatogr. A* **2004**, *1027* (1-2), 25-35.
154. Moreira, N.; Meireles, S.; Brandao, T.; de Pinho, P. G., Optimization of the HS-SPME-GC-IT/MS method using a central composite design for volatile carbonyl compounds determination in beers. *Talanta* **2013**, *117*, 523-531.
155. Biladeau, S. K.; Richmond, W. N.; Laulhe, S.; Nantz, M. H., Isotopically coded N-methoxy amide reagents for GC-MS profiling of carbonyl compounds via mass spectral tag generation. *Anal. Methods* **2016**, *8* (18), 3704-3710.
156. Richardson, S. D.; Caughran, T. V.; Poiger, T.; Guo, Y. B.; Crumley, F. G., Application of DNPH derivatization with LC/MS to the identification of polar carbonyl disinfection by-products in drinking water. *Ozone-Sci Eng* **2000**, *22* (6), 653-675.
157. Jakober, C. A.; Charles, M. J.; Kleeman, M. J.; Green, P. G., LC-MS analysis of carbonyl compounds and their occurrence in diesel emissions. *Anal. Chem.* **2006**, *78* (14), 5086-5093.
158. Yuan, W.; Li, S. W.; Edwards, J. L., Extraction and Quantitation of Ketones and Aldehydes from Mammalian Cells Using Fluorous Tagging and Capillary LC-MS. *Analytical Chemistry* **2015**, *87* (15), 7660-7666.
159. Bereman, M. S.; Comins, D. L.; Muddiman, D. C., Increasing the hydrophobicity and electrospray response of glycans through derivatization with novel cationic hydrazides. *Chem. Commun.* **2010**, *46* (2), 237-239.

160. Mattingly, S. J.; Xu, T.; Nantz, M. H.; Higashi, R. M.; Fan, T. W. M., A carbonyl capture approach for profiling oxidized metabolites in cell extracts. *Metabolomics* **2012**, *8* (6), 989-996.
161. Flinders, B.; Morrell, J.; Marshall, P. S.; Ranshaw, L. E.; Clench, M. R., The use of hydrazine-based derivatization reagents for improved sensitivity and detection of carbonyl containing compounds using MALDI-MSI. *Anal. Bioanal. Chem.* **2015**, *407* (8), 2085-2094.
162. Appelblad, P.; Ponten, E.; Jaegfeldt, L. H.; Backstrom, T.; Irgum, K., Derivatization of steroids with dansylhydrazine using trifluoromethanesulfonic acid as catalyst. *Anal. Chem.* **1997**, *69* (23), 4905-4911.
163. Hidalgo, F. J.; Navarro, J. L.; Delgado, R. M.; Zamora, R., Determination of alpha-keto acids in pork meat and Iberian ham via tandem mass spectrometry. *Food Chem.* **2013**, *140* (1-2), 183-188.
164. Al-Dirbashi, O. Y.; Rashed, M. S.; Jacob, M.; Al-Ahaideb, L. Y.; Al-Amoudi, M.; Rahbeeni, Z.; Al-Sayed, M. M.; Al-Hassnan, Z.; Al-Owain, M.; Al-Zeidan, H., Improved method to determine succinylacetone in dried blood spots for diagnosis of tyrosinemia type 1 using UPLC-MS/MS. *Biomed. Chromatogr.* **2008**, *22* (11), 1181-1185.
165. Appelblad, P.; Ahmed, A.; Ponten, E.; Backstrom, T.; Irgum, K., Perfluorosulfonated ionomer-modified polyethylene. A material for simultaneous solid-phase enrichment and enhanced precolumn dansylation of C-21 ketosteroids in human serum. *Anal. Chem.* **2001**, *73* (15), 3701-3708.
166. Herrington, J.; Zhang, L.; Whitaker, D.; Sheldon, L.; Zhang, J., Optimizing a dansylhydrazine (DNSH) based method for measuring airborne acrolein and other unsaturated carbonyls. *J. Environ. Monit.* **2005**, *7* (10), 969-976.
167. Tomono, S.; Miyoshi, N.; Ohshima, H., Comprehensive analysis of the lipophilic reactive carbonyls present in biological specimens by LC/ESI-MS/MS. *J. Chromatogr. B* **2015**, *988*, 149-156.
168. Backes, G. L.; Neumann, D. M.; Jursic, B. S., Synthesis and antifungal activity of substituted salicylaldehyde hydrazones, hydrazides and sulfohydrazides. *Bioorg. Med. Chem.* **2014**, *22* (17), 4629-4636.
169. Lee, H. J.; Cho, M. J.; Chang, S. K., Ratiometric Signaling of Hypochlorite by the Oxidative Cleavage of Sulfonhydrazide-Based Rhodamine-Dansyl Dyad. *Inorg. Chem.* **2015**, *54* (17), 8644-8649.
170. Contrepolis, K.; Jiang, L. H.; Snyder, M., Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)-Mass Spectrometry. *Mol. Cell. Proteomics* **2015**, *14* (6), 1684-1695.
171. Haggarty, J.; Oppermann, M.; Dalby, M. J.; Burchmore, R. J.; Cook, K.; Weidt, S.; Burgess, K. E. V., Serially coupling hydrophobic interaction and reversed-phase chromatography with simultaneous gradients provides greater coverage of the metabolome. *Metabolomics* **2015**, *11* (5), 1465-1470.
172. Zhao, S.; Dawe, M.; Guo, K.; Li, L., Development of High-Performance Chemical Isotope Labeling LC-MS for Profiling the Carbonyl Submetabolome. *Analytical Chemistry* **2017**, *89* (12), 6758-6765.
173. Tan, J.; McKenzie, C.; Potamitis, M.; Thorburn, A. N.; Mackay, C. R.; Macia, L., The Role of Short-Chain Fatty Acids in Health and Disease. In *Advances in Immunology*,

- Vol 121, Alt, F. W., Ed. Elsevier Academic Press Inc: San Diego, 2014; Vol. 121, pp 91-119.
174. Vitek, L.; Haluzik, M., The role of bile acids in metabolic regulation. *J. Endocrinol.* **2016**, *228* (3), R85-R96.
175. Bollinger, J. G.; Thompson, W.; Lai, Y.; Oslund, R. C.; Hallstrand, T. S.; Sadilek, M.; Turecek, F.; Gelb, M. H., Improved Sensitivity Mass Spectrometric Detection of Eicosanoids by Charge Reversal Derivatization. *Analytical Chemistry* **2010**, *82* (16), 6790-6796.
176. Kloos, D.; Derks, R. J. E.; Wijtmans, M.; Lingeman, H.; Mayboroda, O. A.; Deelder, A. M.; Niessen, W. M. A.; Giera, M., Derivatization of the tricarboxylic acid cycle intermediates and analysis by online solid-phase extraction-liquid chromatography-mass spectrometry with positive-ion electrospray ionization. *J. Chromatogr. A* **2012**, *1232*, 19-26.
177. Williams, J.; Pandarinathan, L.; Wood, J.; Vouros, P.; Makriyannis, A., Endocannabinoid metabolomics: A novel liquid chromatography-mass spectrometry reagent for fatty acid analysis. *Aaps Journal* **2006**, *8* (4), E655-E660.
178. Al-Warhi, T. I.; Al-Hazimi, H. M. A.; El-Faham, A., Recent development in peptide coupling reagents. *J. Saudi Chem. Soc.* **2012**, *16* (2), 97-116.
179. Metaferia, B.; Wei, J. S.; Song, Y. K.; Evangelista, J.; Aschenbach, K.; Johansson, P.; Wen, X. Y.; Chen, Q. R.; Lee, A.; Hempel, H.; Gheeya, J. S.; Getty, S.; Gomez, R.; Khan, J., Development of Peptide Nucleic Acid Probes for Detection of the HER2 Oncogene. *Plos One* **2013**, *8* (4), 7.
180. Davis, M. T. B.; Preston, J. F., A SIMPLE MODIFIED CARBODIIMIDE METHOD FOR CONJUGATION OF SMALL-MOLECULAR-WEIGHT COMPOUNDS TO IMMUNOGLOBULIN-G WITH MINIMAL PROTEIN CROSSLINKING. *Anal. Biochem.* **1981**, *116* (2), 402-407.
181. Nakajima, N.; Ikada, Y., MECHANISM OF AMIDE FORMATION BY CARBODIIMIDE FOR BIOCONJUGATION IN AQUEOUS-MEDIA. *Bioconjugate Chem.* **1995**, *6* (1), 123-130.
182. Staros, J. V.; Wright, R. W.; Swingle, D. M., ENHANCEMENT BY N-HYDROXYSULFOSUCCINIMIDE OF WATER-SOLUBLE CARBODIIMIDE-MEDIATED COUPLING REACTIONS. *Anal. Biochem.* **1986**, *156* (1), 220-222.
183. Choi, M. G.; Kim, J.; Hong, J. M.; Chang, I. J.; Ahn, S.; Chang, S. K., Cu²⁺-selective ratiometric fluorescence signaling probe based on the hydrolysis of dansylhydrazine. *Tetrahedron Lett.* **2016**, *57* (9), 975-978.
184. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reilly, M. D.; Thaden, J. J.; Viant, M. R., Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007**, *3* (3), 211-221.
185. Johnson, C. H.; Ivanisevic, J.; Siuzdak, G., Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (7), 451-459.
186. Vuckovic, D., Improving metabolome coverage and data quality: advancing metabolomics and lipidomics for biomarker discovery. *Chemical Communications* **2018**, *54* (50), 6728-6749.

187. Maurer, H. H., Current role of liquid chromatography-mass spectrometry in clinical and forensic toxicology. *Anal. Bioanal. Chem.* **2007**, *388* (7), 1315-1325.
188. Banerjee, G.; Carare, R.; Cordonnier, C.; Greenberg, S. M.; Schneider, J. A.; Smith, E. E.; van Buchem, M.; van der Grond, J.; Verbeek, M. M.; Werring, D. J., The increasing impact of cerebral amyloid angiopathy: essential new insights for clinical practice. *J. Neurol. Neurosurg. Psychiatry* **2017**, *88* (11), 982-994.
189. Smith, E. E., Cerebral amyloid angiopathy as a cause of neurodegeneration. *J. Neurochem.* **2018**, *144* (5), 651-658.
190. Brenowitz, W. D.; Nelson, P. T.; Besser, L. M.; Heller, K. B.; Kukull, W. A., Cerebral amyloid angiopathy and its co-occurrence with Alzheimer's disease and other cerebrovascular neuropathologic changes. *Neurobiol. Aging* **2015**, *36* (10), 2702-2708.
191. Kovari, E.; Herrmann, F. R.; Hof, P. R.; Bouras, C., The relationship between cerebral amyloid angiopathy and cortical microinfarcts in brain ageing and Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **2013**, *39* (5), 498-509.
192. Shinohara, M.; Murray, M. E.; Frank, R. D.; Shinohara, M.; DeTure, M.; Yamazaki, Y.; Tachibana, M.; Atagi, Y.; Davis, M. D.; Liu, C. C.; Zhao, N.; Painter, M. M.; Petersen, R. C.; Fryer, J. D.; Crook, J. E.; Dickson, D. W.; Bu, G. J.; Kanekiyo, T., Impact of sex and APOE4 on cerebral amyloid angiopathy in Alzheimer's disease. *Acta Neuropathol.* **2016**, *132* (2), 225-234.
193. Thal, D. R.; Griffin, W. S. T.; de Vos, R. A. I.; Ghebremedhin, E., Cerebral amyloid angiopathy and its relationship to Alzheimer's disease. *Acta Neuropathol.* **2008**, *115* (6), 599-609.
194. Greenberg, S. M.; Charidimou, A., Diagnosis of Cerebral Amyloid Angiopathy Evolution of the Boston Criteria. *Stroke* **2018**, *49* (2), 491-497.
195. Charidimou, A.; Gang, Q.; Werring, D. J., Sporadic cerebral amyloid angiopathy revisited: recent insights into pathophysiology and clinical spectrum. *J. Neurol. Neurosurg. Psychiatry* **2012**, *83* (2), 124-137.
196. van Etten, E. S.; Verbeek, M. M.; van der Grond, J.; Zielman, R.; van Rooden, S.; van Zwet, E. W.; van Opstal, A. M.; Haan, J.; Greenberg, S. M.; van Buchem, M. A.; Wermer, M. J. H.; Terwindt, G. M., beta-Amyloid in CSF Biomarker for preclinical cerebral amyloid angiopathy. *Neurology* **2017**, *88* (2), 169-176.
197. Hondius, D. C.; Eigenhuis, K. N.; Morrema, T. H. J.; van der Schors, R. C.; van Nierop, P.; Bugiani, M.; Li, K. W.; Hoozemans, J. J. M.; Smit, A. B.; Rozemuller, A. J. M., Proteomics analysis identifies new markers associated with capillary cerebral amyloid angiopathy in Alzheimer's disease. *Acta Neuropathol. Commun.* **2018**, *6*, 19.
198. Armitage, E. G.; Barbas, C., Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *J. Pharm. Biomed. Anal.* **2014**, *87*, 1-11.
199. Frisoni, G. B.; Winblad, B.; O'Brien, J. T., Revised NIA-AA criteria for the diagnosis of Alzheimer's disease: a step forward but not yet ready for widespread clinical use. *Int. Psychogeriatr.* **2011**, *23* (8), 1191-1196.
200. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J., MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **2018**, *46* (W1), W486-w494.
201. Storey, J. D.; Taylor, J. E.; Siegmund, D., Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **2004**, *66*, 187-205.

202. Storey, J. D.; Bass, A. J.; Dabney, A.; Robinson, D. qvalue: Q-value estimation for false discovery rate control. R package version 2.12.0. <http://github.com/jdstorey/qvalue>.
203. Griffin, J. W. D.; Bradshaw, P. C., Amino Acid Catabolism in Alzheimer's Disease Brain: Friend or Foe? *Oxidative Med. Cell. Longev.* **2017**, 15.
204. Wang, G.; Zhou, Y.; Huang, F. J.; Tang, H. D.; Xu, X. H.; Liu, J. J.; Wang, Y.; Deng, Y. L.; Ren, R. J.; Xu, W.; Ma, J. F.; Zhang, Y. N.; Zhao, A. H.; Chen, S. D.; Jia, W., Plasma Metabolite Profiles of Alzheimer's Disease and Mild Cognitive Impairment. *Journal of Proteome Research* **2014**, 13 (5), 2649-2658.
205. 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.
206. Pyke, J. S.; Callahan, D. L.; Kanojia, K.; Bowne, J.; Sahani, S.; Tull, D.; Bacic, A.; McConville, M. J.; Roessner, U., A tandem liquid chromatography-mass spectrometry (LC-MS) method for profiling small molecules in complex samples. *Metabolomics* **2015**, 11 (6), 1552-1562.
207. Ortmayr, K.; Causon, T. J.; Hann, S.; Koellensperger, G., Increasing selectivity and coverage in LC-MS based metabolome analysis. *Trac-Trends Anal. Chem.* **2016**, 82, 358-366.
208. Bingol, K.; Bruschiweiler-Li, L.; Li, D. W.; Zhang, B.; Xie, M. Z.; Bruschiweiler, R., Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis* **2016**, 8 (6), 557-573.
209. Dias, D. A.; Jones, O. A. H.; Beale, D. J.; Boughton, B. A.; Benheim, D.; Kouremenos, K. A.; Wolfender, J. L.; Wishart, D. S., Current and Future Perspectives on the Structural Identification of Small Molecules in Biological Systems. *Metabolites* **2016**, 6 (4), 29.
210. 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.; Tanaka, K.; 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.
211. Go, E. P., Database Resources in Metabolomics: An Overview. *J. Neuroimmune Pharm.* **2010**, 5 (1), 18-30.
212. Vinaixa, M.; Schymanski, E. L.; Neumann, S.; Navarro, M.; Salek, R. M.; Yanes, O., Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. *Trac-Trends Anal. Chem.* **2016**, 78, 23-35.
213. Zhang, X.; Li, J.; Wang, C.; Song, D. Q.; Hu, C. Q., Identification of impurities in macrolides by liquid chromatography-mass spectrometric detection and prediction of retention times of impurities by constructing quantitative structure-retention relationship (QSRR). *J. Pharm. Biomed. Anal.* **2017**, 145, 262-272.
214. Zhao, X. J.; Zeng, Z. D.; Chen, A. M.; Lu, X.; Zhao, C. X.; Hu, C. X.; Zhou, L. N.; Liu, X. Y.; Wang, X. L.; Hou, X. L.; Ye, Y. R.; Xu, G. W., Comprehensive Strategy to Construct In-House Database for Accurate and Batch Identification of Small Molecular Metabolites. *Analytical Chemistry* **2018**, 90 (12), 7635-7643.

215. 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.
216. Creydt, M.; Fischer, M., Plant Metabolomics: Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Nontargeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Analytical Chemistry* **2017**, *89* (19), 10474-10486.
217. Liang, Y.; Guan, T. Y.; Zhou, Y. Y.; Liu, Y. N.; Xing, L.; Zheng, X.; Dai, C.; Du, P.; Rao, T.; Zhou, L. J.; Yu, X. Y.; Hao, K.; Xie, L.; Wang, G. J., Effect of mobile phase additives on qualitative and quantitative analysis of ginsenosides by liquid chromatography hybrid quadrupole-time of flight mass spectrometry. *J. Chromatogr. A* **2013**, *1297*, 29-36.
218. Monnin, C.; Ramrup, P.; Daigle-Young, C.; Vuckovic, D., Improving negative liquid chromatography/electrospray ionization mass spectrometry lipidomic analysis of human plasma using acetic acid as a mobile-phase additive. *Rapid Commun. Mass Spectrom.* **2018**, *32* (3), 201-211.
219. Zhang, X. M.; Clausen, M. R.; Zhao, X. L.; Zheng, H.; Bertram, H. C., Enhancing the Power of Liquid Chromatography-Mass Spectrometry-Based Urine Metabolomics in Negative Ion Mode by Optimization of the Additive. *Analytical Chemistry* **2012**, *84* (18), 7785-7792.
220. Hua, Y. S.; Wainhaus, S. B.; Yang, Y. N.; Shen, L. X.; Xiong, Y. S.; Xu, X. Y.; Zhang, F. G.; Bolton, J. L.; van Breemen, R. B., Comparison of negative and positive ion electrospray tandem mass spectrometry for the liquid chromatography tandem mass spectrometry analysis of oxidized deoxynucleosides. *J. Am. Soc. Mass Spectrom.* **2001**, *12* (1), 80-87.

Appendix

Putative Identification

Putative identification results can be found from supplemental folder, which is available from Professor Liang Li (liang.li@ualberta.ca).

DnsHz-ID User Tutorial

1. Workflow. The workflow for metabolite identification using dansylhydrazine (DnsHz) labeled metabolites library is shown below (Figure 1).

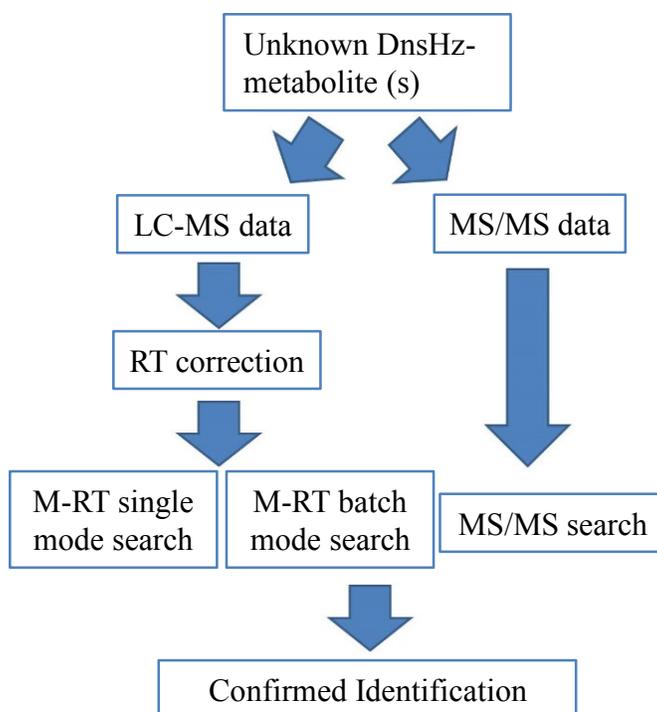


Figure 1. Workflow for M-RT search and MS/MS search.

2. DnsHz library database. The current DnsHz library consists of 78 unique metabolites with a total of 90 entries. The DnsHz library view on the sidebar lists all these DnsHz-

metabolites with their m/z and normalized RT information. Figure 2 shows a screenshot of the DnsHz library database. The user can view the HMDB number, monoisotopic molecular mass, m_z_light, normalized or corrected RT for each of the DnsHz-standards from the table. In addition, the hyperlinks for each DnsHz-metabolite to HMDB and KEGG databases are provided. These databases provide detailed biological information about the metabolite.

#	HMDB No.	Name	Monoisotopic molecular mass	m _z _light	Corrected RT	HMDB link	KEGG link	Show Detail
1	HMDB01426	Formaldehyde	30.0106	278.0958	10.66	Link	Link	Detail
2	HMDB00990	Acetaldehyde	44.0262	292.1114	11.20	Link	Link	Detail
3	HMDB03366	Propanal	58.0419	306.1271	13.28	Link	Link	Detail
4	HMDB03543	Butanal	72.0575	320.1427	14.83	Link	Link	Detail
5	HMDB05994	Hexanal	100.0888	348.1740	17.91	Link	Link	Detail
6	HMDB01140	Octanal	128.1201	376.2053	20.72	Link	Link	Detail
7	HMDB11623	Decanal	156.1514	404.2366	23.26	Link	Link	Detail
8	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59	Link	Link	Detail
9	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15	Link	Link	Detail
10	HMDB06115	Benzaldehyde	106.0419	354.1271	16.23	Link	Link	Detail
11	HMDB06236	Phenylacetaldehyde	120.0575	368.1427	16.35	Link	Link	Detail
12	HMDB01358	Retinal	284.2140	532.2992	27.16	Link	Link	Detail
13	HMDB01659	Acetone	58.0419	306.1271	12.01	Link	Link	Detail
14	HMDB00474	2-butanone	72.0575	320.1427	14.24	Link	Link	Detail
15	HMDB03407	Diacetyl	86.0368	334.1220	13.97	Link	Link	Detail
16	HMDB03315	Cyclohexanone	98.0732	346.1584	15.29	Link	Link	Detail
17	HMDB01184	Methyl propenyl ketone	84.0575	332.1427	15.25	Link	NA	Detail
18	HMDB00005	2-Ketobutyric	102.0317	350.1169	11.10	Link	Link	Detail
19	HMDB01259	Succinic acid semialdehyde	102.0317	350.1169	9.34	Link	Link	Detail
20	HMDB00019	alpha-Ketoisovaleric acid	116.0473	364.1325	16.78	Link	Link	Detail

Figure 2. Screenshot of a partial DnsHz library table.

The user can click the “Show Detail” button, which guides the user to a page with more detailed information about the DnsHz labeled metabolite (Figure 3). An LC-MS chromatogram and MS/MS spectrum are provided on this page. These data were collected using pure standard compound and can be used to compare with the user's experimental data. Details on the preparation of the DnsHz-standards can be found in the materials and methods part of the paper.

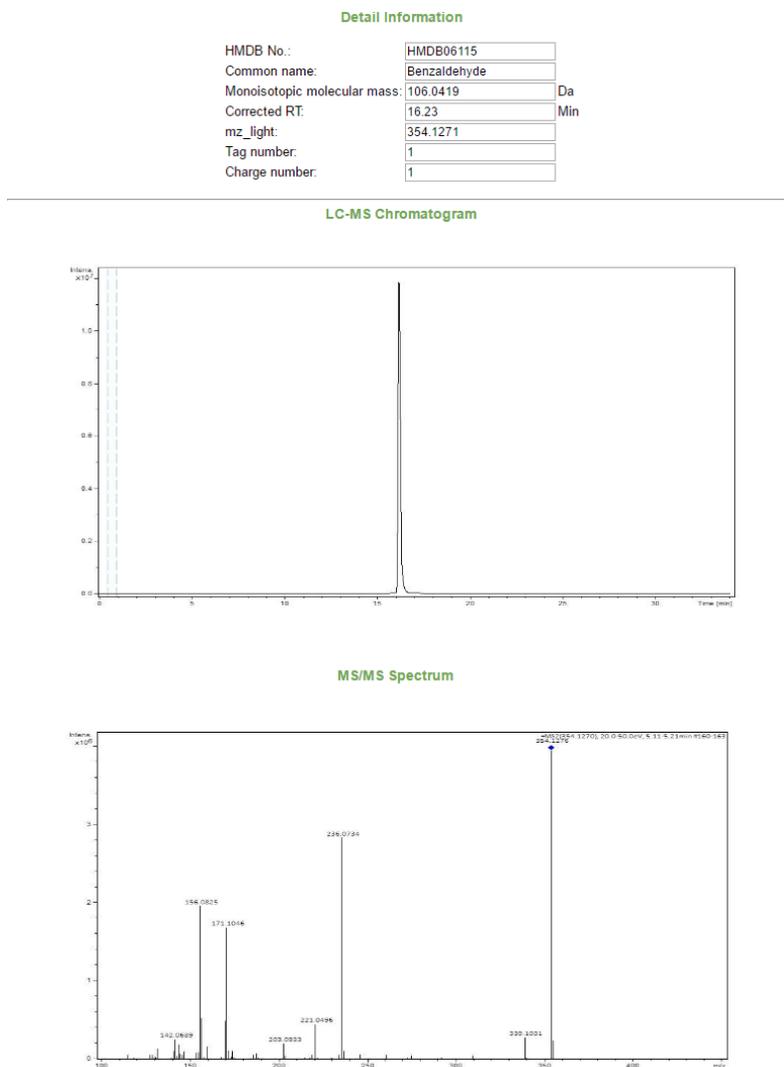


Figure 3. Screenshot of the "Show detail" page.

3. M-RT single mode search. M-RT single mode search allows a user to search the DnsHz library by submitting a single metabolite feature with its RT and m/z (M+H). Also, a calibration file needs to be submitted to correct the retention time of the single metabolite feature. Figure 4 shows the screenshot of the single mode search.

Mass and Retention Time (M-RT) Single Search

Precursor mass	<input type="text" value="354.1271"/>	
Mass tolerance	<input type="text" value="5"/>	ppm
Retention time	<input type="text" value="970"/>	Second
RT tolerance	<input type="text" value="15"/>	Second
Calibration file	<input type="button" value="Choose File"/> Sample Calibration.csv	
Calibration file type	<input checked="" type="radio"/> RTcal	
	<input type="button" value="Submit Query"/>	

Figure 4. M-RT single mode search parameters.

There are six search parameters.

- 1) **Precursor Mass.** The user needs to input the precursor mass of the metabolite feature.
- 2) **Mass tolerance.** The user needs to define a mass tolerance for the precursor mass search. 5 ppm is normally used for data collected using high resolution MS such as TOF and FT (10 ppm or higher may be used for very low abundance peaks). If the experiment is performed using a low resolution MS instrument, a larger mass tolerance should be considered.
- 3) **Retention time.** The user needs to input the retention time of the metabolite feature.
- 4) **RT tolerance.** The user needs to define a retention time tolerance for the M-RT search; 15 seconds is normally used. If no close matches are found, a wider retention time window should be considered with caution. For LC with lower retention time precision, a larger RT tolerance may be used.
- 5) **Calibration file.** A calibration file needs to be uploaded to adjust the retention time of the metabolite feature to match the retention time of metabolites in the DnsHz library. The template of the calibration file is shown in the "**user**

example". The user needs to download the template and fill in the retention time information for each of the calibration standards used in the calibration file. The retention time has a unit of second.

6) **Calibration file type.** In the current DnsHz library RT correction method, a 17 DnsHz-standards file is used (RTcal). We will include other types of the calibration files for different applications in the future.

7) **Submit query.** Once all the parameters have been set, the user can click on the “submit query” to start the M-RT single mode search.

4. **M-RT batch mode search.** M-RT batch search mode allows a user to search the DnsHz library using the entire DnsHz-labeled LC-MS file. Figure 5 shows the screenshot of the batch mode search.

The screenshot shows a web interface titled "Batch Search". It contains several input fields and buttons. The "Mass tolerance" field is set to "5" with the unit "ppm" to its right. The "RT tolerance" field is set to "15" with the unit "Second" to its right. The "Sample file" field has a "Choose File" button and the text "Sample example.csv". The "Calibration file" field has a "Choose File" button and the text "Sample Calibration.csv". The "Calibration file type" field has a radio button selected next to "RTcal". At the bottom, there is a "Submit Query" button.

Figure 5. M-RT batch mode search parameters.

The parameters include:

1) **Mass tolerance.** The user needs to define a mass tolerance for the precursor mass search. 5 ppm is normally used for data collected using high resolution MS such as TOF and FT (10 ppm or higher may be used for very low abundance peaks). If the experiment is performed using a low resolution MS instrument, a larger mass tolerance should be considered.

2) **RT tolerance.** The user needs to define a retention time tolerance for the M-RT search; 15 seconds is normally used. If no close matches are found, a wider retention time window should be considered with caution. For LC with lower retention time precision, a larger RT tolerance may be used.

3) **Sample file.** A sample file needs to be uploaded onto the website for batch mode search. The sample file is the metabolite-intensity matrix after processing the raw LC-MS data in IsoMS, Iso-Align, Zero-fill and IsoQuant.

4) **Calibration file.** A calibration file needs to be uploaded for adjusting the retention time of the metabolite feature to match with the retention time of the metabolites in the DnsHz library. The template of the calibration file is shown in the "**user example**". The user needs to download this template and fill in the retention time information for each of the calibration standards used in the calibration file. The retention time has a unit of second.

5) **Calibration file type.** In the current DnsHz library RT correction method, a 17 DnsHz-standards file is used (RTcal). We will include other types of the calibration files for different applications in the future.

6) **Submit query.** Once all the parameters have been set, the user can click on the "submit query" to start the M-RT batch mode search.

5. MS/MS search. The MS/MS search function allows a user to identify a DnsHz labeled metabolite using MS/MS information. Figure 6 shows the screenshot of the MS/MS search function.

Precursor Mass:

Neutral or Ion:
 Neutral
 [M+H]⁺
 [M+Na]⁺
 [M+K]⁺
 [M+NH₄]⁺
 [M-H]⁻

MS/MS list

156.0811	58.7
157.088	14.2
160.0763	10.6
170.0961	14.4

MS/MS tolerance:
 In ppm (default: ± 5 ppm): ppm
 In Da (default: ± 0.005 Da): Da

Match precursor ion
 No
 Yes

Precursor mass tolerance:
 In ppm (default: ± 5 ppm): ppm
 In Da (default: ± 0.005 Da): Da

Match retention time
 No
 Yes

Retention time Second

Calibration file Sample Calibration.csv

Calibration file type RTcal

RT tolerance Second

Figure 6. MS/MS search parameters.

The parameters include:

- 1) **Precursor mass.** The user needs to input the precursor mass of the metabolite feature.
- 2) **Neutral or ion.** The user can define the type of the precursor mass. It can be either an M+H ion or a neutral mass.
- 3) **MS/MS list.** The user needs to input a list of MS/MS fragment ion masses with their associated intensities.
- 4) **MS/MS tolerance.** The user needs to set a mass tolerance for the MS/MS fragment ions to perform the matching with the MS/MS information in the DnsHz library.

- 5) Match precursor ion.** The user has the option of defining the precursor ion mass for MS/MS search. If this option is enabled, only the DnsHz-metabolites that match with the precursor mass will be further used to compare the MS/MS fragment ions. If this option is disabled, the MS/MS match is performed on all DnsHz-metabolites.
- 6) Precursor mass tolerance.** The user needs to define a mass tolerance for the precursor mass search. 5 ppm is normally used for data collected using high resolution MS such as TOF and FT (10 ppm or higher may be used for very low abundance peaks). If the experiment is performed using a low resolution MS instrument, a larger mass tolerance should be considered.
- 7) Match retention time.** The user has the option of including RT for MS/MS search. If this option is on, only the DnsHz-metabolites that match with the retention time will be further used to compare the MS/MS fragment ions. If this option is off, the MS/MS match is performed on all DnsHz-metabolites.
- 8) RT tolerance.** The user needs to define a retention time tolerance for the M-RT search; 15 seconds is normally used. If no close matches are found, a wider retention time window should be considered with caution. For LC with lower retention time precision, a larger RT tolerance may be used.
- 9) Calibration file.** A calibration file needs to be uploaded for adjusting the retention time of the metabolite feature to be consistent with the retention time of metabolites in the DnsHz library. The template of the calibration file is shown in the "**user example**". The user needs to download that template and fill in the retention time information for each of the calibration standard used in the calibration file. The retention time has a unit of second.

10) Calibration file type. In the current DnsHz library RT correction method, a 17 DnsHz-standards file is used (RTcal). We will include other types of the calibration files for different applications in the future.

6. M-RT search result display. Figure 7 shows the screenshot of the M-RT search result.

The search result table is similar to the DnsHz library table with several extra columns.

Search Result													
#	Input mass	Input rt	Calibrated RT	HMDB No.	Name	Monoisotopic molecular mass	mz_light	RT	Mass error	RT error	HMDB link	KEGG link	Show Detail
1	590.2009	5.86	5.86	HMDB00186	Lactose	342.1162	590.2014	5.86	0.0005	0.00	Link	Link	Detail
2	428.1482	6.10	6.10	HMDB00143	Galactose	180.0634	428.1486	5.98	0.0004	0.12	Link	Link	Detail
3	428.1482	6.10	6.10	HMDB00122	Glucose	180.0634	428.1486	6.08	0.0004	0.02	Link	Link	Detail
4	442.1274	6.17	6.17	HMDB00127	D-glucuronic acid	194.0427	442.1279	6.27	0.0005	0.10	Link	Link	Detail
5	442.1274	6.17	6.17	HMDB02545	Galacturonic acid	194.0427	442.1279	6.32	0.0005	0.15	Link	Link	Detail
6	398.1371	6.67	6.67	HMDB00646	L-Arabinose	150.0528	398.1380	6.44	0.0009	0.23	Link	Link	Detail
7	398.1371	6.67	6.67	HMDB00283	D-Ribose	150.0528	398.1380	6.68	0.0009	0.01	Link	Link	Detail
8	415.1415	7.74	7.74	HMDB01545	Pyridoxal	167.0582	415.1434	7.69	0.0019	0.05	Link	Link	Detail
9	427.1419	10.46	10.46	HMDB00714	Hippuric Acid	179.0582	427.1434	10.56	0.0015	0.10	Link	Link	Detail
10	306.1264	11.99	11.99	HMDB01659	Acetone	58.0419	306.1271	12.01	0.0007	0.02	Link	Link	Detail
11	306.1261	13.30	13.30	HMDB03366	Propanal	58.0419	306.1271	13.28	0.0010	0.02	Link	Link	Detail
12	378.1476	13.51	13.51	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39	0.0006	0.12	Link	Link	Detail
13	378.1476	13.51	13.51	HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39	0.0006	0.12	Link	Link	Detail
14	320.1416	14.23	14.23	HMDB00474	2-butanone	72.0575	320.1427	14.24	0.0011	0.01	Link	Link	Detail
15	334.1577	16.10	16.10	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15	0.0007	0.05	Link	Link	Detail
16	612.3079	17.35	17.35	HMDB00903	Tetrahydrocortisone	364.2250	612.3102	17.39	0.0023	0.04	Link	Link	Detail
17	578.3031	19.52	19.52	HMDB00374	17Hydroxyprogesterone	330.2195	578.3047	19.64	0.0016	0.12	Link	Link	Detail
18	578.3031	19.52	19.52	HMDB00016	11-Deoxycorticosterone	330.2195	578.3047	19.5	0.0016	0.02	Link	Link	Detail
19	376.2047	20.72	20.72	HMDB01140	Octanal	128.1201	376.2053	20.72	0.0006	0.00	Link	Link	Detail

[Export as CSV](#)

Figure 7. Screenshot of M-RT search result.

7. MS/MS search result display. Figure 8 shows the screenshot of the MS/MS search result. The search result table is similar to the DnsHz library table with several extra columns.

Search Result														
#	Input mass	Input RT	Calibrated RT	HMDB No.	Name	Monoisotopic molecular mass	mz_light	Library RT	Mass error	RT error	HMDB link	KEGG link	MS/MS score ▲	Show detail
8	338.1164	7.65	7.65	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59	0.0005	0.06	Link	Link	1.00	Detail
31	338.1164	7.65	7.65	HMDB00143	Galactose	180.0634	428.1486	5.98	90.0322	1.67	Link	Link	0.66	Detail
17	338.1164	7.65	7.65	HMDB01259	Succinic acid semialdehyde	102.0317	350.1169	9.34	12.0005	1.69	Link	Link	0.66	Detail
18	338.1164	7.65	7.65	HMDB00720	Levulinic acid	116.0473	364.1325	9.59	26.0161	1.94	Link	NA	0.66	Detail
32	338.1164	7.65	7.65	HMDB00122	Glucose	180.0634	428.1486	6.08	90.0322	1.57	Link	Link	0.66	Detail
35	338.1164	7.65	7.65	HMDB00174	L-Fucose	164.0685	412.1537	6.86	74.0373	0.79	Link	Link	0.63	Detail

Figure 8. Screenshot of MS/MS search

Examples of M-RT and MS/MS Search for DnsHz-ID

1. An example of using M-RT to do single mode search.

1). For the M-RT single mode search, the user enters a precursor mass (354.1271) and retention time (990.2 seconds), together with their mass tolerance (5 ppm) and RT tolerance (15 seconds) (see Figure 1). A calibration file with 17 calibration standards (RTcal) also needs to be uploaded. The template of the calibration file can be found below. The user needs to download it and change the retention time according to the calibration file performed with the metabolite feature. After filling out the retention time, click the “Submit Query” to start the M-RT single mode search.

Mass and Retention Time (M-RT) Single Search

Precursor mass	<input type="text" value="354.1271"/>	
Mass tolerance	<input type="text" value="5"/>	ppm
Retention time	<input type="text" value="990.2"/>	Second
RT tolerance	<input type="text" value="15"/>	Second
Calibration file	<input type="button" value="Choose File"/> Sample Calibration.csv	
Calibration file type	<input checked="" type="radio"/> RTcal	
	<input type="button" value="Submit Query"/>	

Figure 1. Single mode search parameter

2). The search result is shown in Figure 2.

Search Result

#	Input mass	Input rt	Calibrated RT	HMDB No.	Name	Monoisotopic molecular mass	mz_light	RT	Mass error	RT error	HMDB link	KEGG link	Show Detail
1	354.1271	16.50	16.31	HMDB06115	Benzaldehyde	106.0419	354.1271	16.23	0.0000	0.08	Link	Link	Detail

Figure 2. Single search result.

2. An example of using M-RT to do batch mode search.

1). For the M-RT batch mode search, the user enters a mass tolerance (5 ppm) and RT tolerance (15 seconds) (see Figure 1). The user also needs to upload a sample file and a calibration file. The template of the sample file and calibration file can be found in Introduction page. For the calibration file, the user needs to download it and change the retention time according to the calibration file. After it's all done, click the "Submit Query" to start the M-RT single mode search.

Batch Search

Mass tolerance ppm

RT tolerance Second

Sample file Sample example.csv

Calibration file Sample Calibration.csv

Calibration file type RTcal

Figure 3. Batch mode search parameters.

2). The search result is shown in Figure 4.

Search Result

#	Input mass	Input rt	Calibrated RT	HMDB No.	Name	Monoisotopic molecular mass	mz_light	RT	Mass error	RT error	HMDB link	KEGG link	Show Detail
1	590.2009	5.86	5.80	HMDB00186	Lactose	342.1162	590.2014	5.86	0.0005	0.06	Link	Link	Detail
2	428.1482	6.10	6.03	HMDB00143	Galactose	180.0634	428.1486	5.98	0.0004	0.05	Link	Link	Detail
3	428.1482	6.10	6.03	HMDB00122	Glucose	180.0634	428.1486	6.08	0.0004	0.05	Link	Link	Detail
4	442.1274	6.17	6.09	HMDB00127	D-glucuronic acid	194.0427	442.1279	6.27	0.0005	0.18	Link	Link	Detail
5	442.1274	6.17	6.09	HMDB02545	Galacturonic acid	194.0427	442.1279	6.32	0.0005	0.23	Link	Link	Detail
6	398.1371	6.67	6.54	HMDB00646	L-Arabinose	150.0528	398.1380	6.44	0.0009	0.10	Link	Link	Detail
7	398.1371	6.67	6.54	HMDB00283	D-Ribose	150.0528	398.1380	6.68	0.0009	0.14	Link	Link	Detail
8	415.1415	7.74	7.59	HMDB01545	Pyridoxal	167.0582	415.1434	7.69	0.0019	0.10	Link	Link	Detail
9	306.1264	11.99	11.88	HMDB01659	Acetone	58.0419	306.1271	12.01	0.0007	0.13	Link	Link	Detail
10	306.1261	13.30	13.19	HMDB03366	Propanal	58.0419	306.1271	13.28	0.0010	0.09	Link	Link	Detail
11	378.1476	13.51	13.36	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39	0.0006	0.03	Link	Link	Detail
12	378.1476	13.51	13.36	HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39	0.0006	0.03	Link	Link	Detail
13	320.1419	15.12	14.77	HMDB03543	Butanal	72.0575	320.1427	14.83	0.0008	0.06	Link	Link	Detail
14	334.1577	16.10	15.90	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15	0.0007	0.25	Link	Link	Detail
15	376.2047	20.72	20.50	HMDB01140	Octanal	128.1201	376.2053	20.72	0.0006	0.22	Link	Link	Detail

Figure 4. Batch mode search result.

3). At the end of the search result table, there is an “Export as CSV” button (Figure 5).

By clicking this button, the user can export the search results into a CSV table shown in

Figure 6

11	378.1476	13.51	13.36	HMDB00491	3-Methyl-2-oxovaleric acid	130.0630	378.1482	13.39	0.0006	0.03	Link	Link	Detail
12	378.1476	13.51	13.36	HMDB00408	2-Methyl-3-ketovaleric acid	130.0630	378.1482	13.39	0.0006	0.03	Link	Link	Detail
13	320.1419	15.12	14.77	HMDB03543	Butanal	72.0575	320.1427	14.83	0.0008	0.06	Link	Link	Detail
14	334.1577	16.10	15.90	HMDB06478	3-Methylbutanal	86.0732	334.1584	16.15	0.0007	0.25	Link	Link	Detail
15	376.2047	20.72	20.50	HMDB01140	Octanal	128.1201	376.2053	20.72	0.0006	0.22	Link	Link	Detail

Export as CSV

Figure 5. “Export as CSV” button.

search parameters:														
Search Result*														
search result:														
#	Input mas	Input rt	Calibratec	HMDB No.	Name	Monoisotz	mz	light	RT	Mass erro	RT error	HMDB link	KEGG link	S
1	590.2009	5.86	5.8	HMDB001	Lactose	342.1162	590.2014	5.86	0.0005	0.06				HMDB Lin
2	428.1482	6.1	6.03	HMDB001	Galactose	180.0634	428.1486	5.98	0.0004	0.05				HMDB Lin
3	428.1482	6.1	6.03	HMDB001	Glucose	180.0634	428.1486	6.08	0.0004	0.05				HMDB Lin
4	442.1274	6.17	6.09	HMDB001	D-glucuro	194.0427	442.1279	6.27	0.0005	0.18				HMDB Lin
5	442.1274	6.17	6.09	HMDB025	Galacturoi	194.0427	442.1279	6.32	0.0005	0.23				HMDB Lin
6	398.1371	6.67	6.54	HMDB006	L-Arabino	150.0528	398.138	6.44	0.0009	0.1				HMDB Lin
7	398.1371	6.67	6.54	HMDB002	D-Ribose	150.0528	398.138	6.68	0.0009	0.14				HMDB Lin
8	415.1415	7.74	7.59	HMDB015	Pyridoxal	167.0582	415.1434	7.69	0.0019	0.1				HMDB Lin
9	306.1264	11.99	11.88	HMDB016	Acetone	58.0419	306.1271	12.01	0.0007	0.13				HMDB Lin
10	306.1261	13.3	13.19	HMDB033	Propanal	58.0419	306.1271	13.28	0.001	0.09				HMDB Lin
11	378.1476	13.51	13.36	HMDB004	3-Methyl-	130.063	378.1482	13.39	0.0006	0.03				HMDB Lin
12	378.1476	13.51	13.36	HMDB004	2-Methyl-	130.063	378.1482	13.39	0.0006	0.03				HMDB Lin
13	320.1419	15.12	14.77	HMDB035	Butanal	72.0575	320.1427	14.83	0.0008	0.06				HMDB Lin
14	334.1577	16.1	15.9	HMDB064	3-Methyl	86.0732	334.1584	16.15	0.0007	0.25				HMDB Lin
15	376.2047	20.72	20.5	HMDB011	Octanal	128.1201	376.2053	20.72	0.0006	0.22				HMDB Lin

Figure 6. Exported CSV search result.

3. An example of performing M-RT and MS/MS search.

1). For the MS/MS search, the user inputs a precursor mass (338.1164) and select the ion type as [M+H]. Also, a MS/MS list needs to be uploaded. The MS/MS tolerance is defined at a default of 0.005 Da or 5 ppm. The match precursor ion and match retention time functions can be set as on or off. If they are all turned off, the MS/MS search is only based on the match of MS/MS fragments with the MS/MS standards. After all the parameters are set, click “Submit Query” to start the MS/MS search.

Precursor Mass:

Neutral or Ion:
 Neutral
 [M+H]⁺
 [M+Na]⁺
 [M+K]⁺
 [M+NH₄]⁺
 [M-H]⁻

MS/MS list

156.0811	58.7
157.088	14.2
160.0763	10.6
170.0961	14.4

MS/MS tolerance:
 In ppm (default: ± 5 ppm): ppm
 In Da (default: ± 0.005 Da): Da

Match precursor ion
 No
 Yes

Precursor mass tolerance:
 In ppm (default: ± 5 ppm): ppm
 In Da (default: ± 0.005 Da): Da

Match retention time
 No
 Yes

Retention time Second

Calibration file Sample Calibration.csv

Calibration file type RTcal

RT tolerance Second

Figure 7. MS/MS search parameters.

2). The MS/MS search result is shown in Figure 8.

Search Result														
#	Input mass	Input RT	Calibrated RT	HMDB No.	Name	Monoisotopic molecular mass	mz_light	Library RT	Mass error	RT error	HMDB link	KEGG link	MS/MS score ▲	Show detail
8	338.1164	7.65	7.65	HMDB01051	Glyceraldehyde	90.0317	338.1169	7.59	0.0005	0.06	Link	Link	1.00	Detail
31	338.1164	7.65	7.65	HMDB00143	Galactose	180.0634	428.1486	5.98	90.0322	1.67	Link	Link	0.66	Detail
17	338.1164	7.65	7.65	HMDB01259	Succinic acid semialdehyde	102.0317	350.1169	9.34	12.0005	1.69	Link	Link	0.66	Detail
18	338.1164	7.65	7.65	HMDB00720	Levulinic acid	116.0473	364.1325	9.59	26.0161	1.94	Link	NA	0.66	Detail
32	338.1164	7.65	7.65	HMDB00122	Glucose	180.0634	428.1486	6.08	90.0322	1.57	Link	Link	0.66	Detail
35	338.1164	7.65	7.65	HMDB00174	L-Fucose	164.0685	412.1537	6.86	74.0373	0.79	Link	Link	0.63	Detail

Figure 8. MS/MS search result.