Development of Chemical Isotope Labeling (CIL) Liquid Chromatography-Mass Spectrometry (LC-MS) for Single-Cell Metabolomics

by

Wan Chan

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

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

Single-cell metabolomics (SCM) strives to identify, quantify and characterize all metabolites in a single-cell. The study of the metabolome enhances our understanding of the cellular interactions within and in response to environmental influences on a molecular level. As cells are being analyzed individually, this leads to a more accurate representation of cell-to-cell variations that would otherwise be masked by bulk population measurements. In this respect, SCM is important for illuminating cellular diversity and heterogeneity, and its development has the potential to shed light on, for example, improving the diagnosis and treatment of cancer, which has been recognized as a heterogeneous disease.

The most prevalent limit of SCM is the comprehensive analysis of the metabolome in a single-cell. Given that metabolites can have very different chemical and physical properties, high coverage metabolic profiling with only one analytical platform is difficult to achieve. It is also worth noting that metabolites amount in a single-cell is extremely limited, and additionally, metabolites cannot be amplified. Therefore, a highly sensitive detection method is necessary. Chemical isotope labeling (CIL) liquid chromatography mass spectrometry (LC-MS) provides a way for comprehensive metabolic profiling of single-cell with one platform – positive ion mode reversed phase liquid chromatography mass spectrometry (RPLC-MS). In brief, the metabolome is first being divided into four different groups based on their chemical properties and hydrophobicity. CIL is followed to attach an isotope-mass-encoded tag to metabolites, which has been proved to improve their separation and mitigate ion suppression in RPLC. Moreover, CIL enables quantification metabolomics in which absolute and relative quantification can be performed.

In this work, the integration of efficient single-cell preparation and CIL LC-MS method were developed for single-cell metabolomics. *Xenopus laevis* oocyte is used as a model system. A

sample preparation and processing protocol involving cells extraction from the *Xenopus laevis* and cell lysis to extract the metabolites was developed. Dansylation labeling was first used to profile the amines and phenols containing metabolites in cells. We were able to determine the amine and phenol submetabolome of cells comprehensively and quantitatively. At the same time, the behavior of each single-cell was revealed.

One of the goals in metabolomics is to quantify metabolomics changes induced by one or more effectors, so as to study the perturbations of metabolic networks. Hence, the metabolic responses of cells to heat stress was also studied by applying CIL LC-MS. The short-term and long-term effect of heat stress, as well as the recovery from heat stress were investigated and determined.

CIL LC-MS was also used to study cells at different locations of the *Xenopus laevis* ovary comprehensively by applying four labeling chemistries with proper design of sample collection. Cellular amine and phenol, hydroxyl, carbonyl and carboxylic acid submetabolome were studied. And by comparing the metabolome of cells at different locations, I was able to show that some submetabolome of cells have large variations at different points of the ovary, which demonstrates the importance of applying four labeling chemistries to elucidate the metabolome of cells comprehensively and systematically. This study can provide improvements in experimental design using *Xenopus* oocytes.

When performing metabolic profiling of single-cell, I understood and realized the importance of expanding the library for metabolite identification. Therefore, I constructed two MS/MS-retention time (RT) libraries including the molecular mass, MS/MS spectrum and RT information. RT calibrants were generated and the multipoint RT calibration method was used to transfer RTs from the instrumental setup in one laboratory to the same setup in another laboratory. Moreover, metabolite identification in human urine samples using the two libraries were demonstrated.

Preface

This thesis is an original work by Wan Chan. The research project in Chapter 2 to Chapter 4 of this thesis, received research ethics approval from the Biosciences Animal Care and Use Committee of the University of Alberta.

Some of the research conducted in this thesis, Chapter 2 to Chapter 4, are collaborated with Dr. Michael Schultz, from the Department of Biochemistry at the University of Alberta.

The cell samples used in Chapter 2 to Chapter 4 were obtained by me and Dr. Schultz. Dr. Michael Schultz was involved in the extraction of the ovary from *Xenopus laevis* and the removal of follicle cells. The sample preparation, sample analysis, data processing and statistical analysis are my original work. Dr. Liang Li and Dr. Michael Schultz were involved in the experimental design of these studies.

All Supplemental Tables in Chapter 4 can be obtained from Dr. Liang Li's Lab. Contact Dr. Liang Li (liang.li@ualberta.ca) for information.

Acknowledgements

First and foremost, I would like to express my sincere and heartfelt gratitude to my supervisor, Dr. Liang. Li and collaborator, Dr. Michael C. Schultz from the Department of Biochemistry. Thank you for the guidance, encouragement and support you have given me these past five years, the experience and knowledge I have obtained broaden my horizons in metabolomics and cell biology.

I am thankful to my supervisory committee, Dr. Michael C. Schultz and Dr. Michael Serpe for their inspiring comment and suggestion. I would like to thank Dr. Michael C. Schultz for introducing me to the cell biology and taking time to talk to me in my difficult times. Thank you for participating in every stage of my degree. I would also like to thank Dr. Michael Serpe for participating in my candidacy exam and oral defense; reviewing my annual progress reports and thesis.

I would also like to extend my deepest appreciation to all the current members, particularly Dr. Xian. Luo, Dr. Shuang. Zhao and Dr. Monica Li. I would like to thank Xian and Shuang for giving me very great amount of assistance and suggestion with your valuable experience throughout my research. I would like to thank Monica for reviewing and editing my thesis. In addition, I thank Ms. Chelsea Gates and Ms. Helen Tang from Dr. Michael C. Schultz lab for giving me a lot of help and suggestion in Biology.

Special thanks to the funding agency, Canadian Institutes of Health Research (CIHR), for supporting the works presented in this thesis.

Last but not least, I take this opportunity to thank my family and friends for their support and care in my PhD journey. My mom, Mandy, H.L. Ying, thank you for your unconditional and infinite love and support. Thank you for always being here for me, backing me up so that I am brave and tough enough to cope with my life's challenges. My precious friends, Kai Ling Ching,

Kathy Hui, Lily Wong, RongJin Wang who provided me happy moments so that I can have mental detox, I thank you, I am so beyond lucky to have you all. Finally, thank you God for all the blessings.

Table of Contents

List of Figures xii

List of Tables xix

List of Abbreviations xx

List of Symbols xxiii

Chapter 1 Introduction 1

1.1 The "Omics" Science 1

1.2 Metabolomics 4

1.2.1 Importance of Metabolomics 4

1.2.2 Challenges of Metabolomics 5

1.2.3 Platforms for Metabolomics 6

1.2.4 Workflow for Metabolomics 7

1.2.5 Mass Spectrometry for Metabolomics 9

1.2.6 High-Performance Chemical Isotope Labeling (CIL) LC-MS for Quantitative Metabolomics 14

1.2.6.1 Quantitative Metabolomics 14

1.2.6.2 CIL in Large-Scale Metabolite Quantification 16

1.2.6.3 Divide-and-Conquer Approach 18

1.2.6.4 Sample Normalization 20

1.2.6.5 Optimization of Injection Amount or Volume 22

1.2.6.6 Data Processing and Analysis 22

1.2.6.7 Metabolite Identification 24

1.2.6.8 Statistical Analysis 25

1.2.6.9 Applications 28

1.3 Single-Cell Metabolomics 29

1.3.1 Cell-to-Cell Heterogeneity 29

1.3.2 Advantage and Importance of Single-Cell Metabolomics 30
1.3.3 Challenges of Single-Cell Metabolomics 32
1.3.4 *Xenopus laevis* in Single-Cell Metabolomics 33
1.3.5 Workflow for Single-Cell Metabolomics 34
1.3.5.1 Sample Collection and Metabolite Extraction 34
1.3.5.2 CIL LC-MS 37
1.4 Scope of the Thesis 38
1.5 Literature Cited 38
Chapter 2 – High-Coverage Quantitative Single-Cell Metabolomics: Chemical Isotope Labeling LC-MS Metabolome Analysis of Individual *Xenopus laevis* Oocytes 43
2.1 Introduction 43
2.2 Experimental 45
2.2.1 Chemical and Reagents 45
2.2.2 Sample Collection and Processing 45
2.2.3 Method Optimization 47
..... 52
2.2.4 Chemical Isotope Labeling for Amine and Phenol Metabolomics 52
2.2.5 LC-MS 53
2.2.6 Data Processing and Statistical Analysis 54
2.2.7 Metabolite Identification 55
2.3 Results and Discussion 55
2.4 Conclusions and Future Work 63
2.5 Literature Cited 64
Chapter 3 – Chemical Isotope Labeling LC-MS for Studying the Metabolic Response of Single Cells to Heat Stress 66

3.1 Introduction 66
3.2 Experimental 70
3.2.1 Chemical and Reagents 70
3.2.2 Sample Collection and Preparation 70
3.2.3 Chemical Isotope Labeling for Amine and Phenol Metabolomics 71
3.2.4 LC-MS 72
3.2.5 Data Processing 72
3.3 Results and Discussion 74
3.3.1 Amine and phenol metabolomics 74
3.3.2 Heat stress response of cells 76
3.3.3 Recovery of cells after heat stress 81
3.3.4 Cellular GSH levels under heat stress 82
3.3.5 Cell-to-cell metabolite variation in response to heat stress 83
3.4 Conclusions 84
3.5 Literature Cited 84
Chapter 4 – High-Coverage Single-Cell Metabolomics for Studying Cell-to-Cell Variations from Different Locations of the *Xenopus laevis* Ovary 87
4.1 Introduction 87
4.2 Experimental 90
4.2.1 Chemical and Reagents 90
4.2.2 Sample Collection and Preparation 91
4.2.3 Chemical Isotope Labeling 93
4.2.4 LC-MS 95
4.2.5 Data Processing and Metabolite Identification 96
4.3 Results and Discussion 97

4.3.1 Amine and phenol metabolomics 97
4.3.2 Hydroxyl metabolomics 106
4.3.3 Carbonyl metabolomics 114
4.3.4 Carboxylic acid metabolomics 122
4.3.5 Cell-to-Cell Variability of Metabolite in Different Metabolic Pathway 130
4.3.6 Importance of 4-Channel Profiling 135
4.4 Conclusions 135
4.5 Literature Cited 136
Chapter 5 – Construction and Application of MS/MS Retention Time Library in Metabolomics 138
5.1 Introduction 138
5.2 Experimental 141
5.2.1 Construction and Application of MS/MS-RT Library for Endogenous Metabolite Identification 141
5.2.1.1 Method 141
5.2.1.2 LC-MS 141
5.2.1.3 Library Data Acquisition and Data Processing 142
5.2.2 Construction and Application of HILIC MS/MS-RT Library for Metabolite Identification 143
5.2.2.1 Method 143
5.2.2.2 LC-MS 143
5.2.2.3 Library Data Acquisition and Data Processing 144
5.3 Results and Discussion 144
5.3.1 MS/MS-RT Library 144
5.3.1.1 Overview 144
5.3.1.2 Multiple point RT Normalization 144
5.3.1.3 Validation in Human Urine Sample 146

5.3.2 HILIC MS/MS-RT Library 149
5.3.2.1 Overview 149
5.3.2.2 Parameter Optimization for HILIC MS/MS-RT library 150
5.3.2.3 Multiple point RT Normalization 152
5.3.2.4 Validation in Human Urine Sample 154
5.4 Conclusions 156
5.5 Literature Cited 156
Chapter 6 - Conclusions and Future Work 157
6.1 Thesis Summary 157
6.2 Future Work 159
Bibliography 160
Appendix 168

List of Figures

Figure 2. 1 General workflow of CIL LC-MS for single-cell metabolomics. 46

Figure 2. 2 M1, M2 and M3 workflows of labeling small amounts of metabolites in the single-cell. 48

Figure 2. 3 BPCs (A) of M1 and M2 and (B) BPCs of M1 and M3. 49

Figure 2. 4 Heatmap comparing the absolute peak intensity of M1, M2 and M3 (n=3). 50

Figure 2. 5 EIC of a selected peak in M1, M2 and M3. 51

Figure 2. 6 (A) Number of detected metabolites in M1, M2 and M3 and their (B) peak pair distribution. 52

Figure 2. 7 Content variability of amino acids between cells and variability of signal between...56

Figure 2. 8 Correlation of cystine and asparagine in cellular peak pair ratio. 57

Figure 2. 9 Examples of metabolites with comparable cell-to-cell variation at the beginning and the end of the culture (A) uridine, (B) pantothenic acid. 57

Figure 2. 10 PCA score plot of single cells with different culture times with the QC samples. 58

Figure 2. 11 Volcano plot for comparison of cells in 0 h and 24 h of culture. 59

Figure 2. 12 Box plots showing the cell-to-cell variability of (A) leucine, (B) glycine, (C) serine, (D) alanine, (E) aspartate with RSD values indicated at the bottom of each group. 60

Figure 2. 13 Box plots showing the cell-to-cell variability of (A) TMS and (B) 3-aminobenzoic acid. 62

Figure 3. 1 Cell collection workflow for the investigation of heat stress effects on the cellular metabolome. 71

Figure 3. 2 PCA score plot of single-cell samples in different groups with QC data. 75

Figure 3. 3 PLS-DA score plot of single-cell samples in different groups. 75

Figure 3. 4 Volcano plot comparing cells in (A) 18℃ overnight and 18℃ to 34℃ 15min population. (B) 18℃ overnight and 18℃ to 34℃ 2h population and (C) 18℃ overnight and 18℃ 2h population.

The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q = 0.007 for (A), q = 0.02 for (B) and q = 0.05 for (C)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point). 78

Figure 3. 5 Box plots of (A) proline (B) leucine (C) arginine (D) asparagine (E) glutamine (F) serine (G) aspartic acid (H) glutamic acid with RSD values indicated. 79

Figure 3. 6 Box plots of (A) tyrosine (B) leucine (C) isoleucine (D) proline (E) valine (F) lysine (G) glycine (H) histidine (I) phenylalanine with RSD values indicated. 80

Figure 3. 7 Box plots of (A) asparagine (B) arginine (C) isoleucine (D) proline (E) phenylalanine (F) alanine (G) glutamic acid. 82

Figure 3. 8 Box plots of glutathione in (A), (B) and (C). 83

Figure 4. 1 Workflow of organ zonation study. 90

Figure 4. 2 Photography of the ovaries with punch locations indicated. 92

Figure 4. 3 Workflow of single cells collection in organ zonation study. 93

Figure 4. 4 PCA score plots of amine and phenol submetabolome of cells in (A) animal 1 and (B) animal 2. 100

Figure 4. 5 PLS-DA score plots of amine and phenol submetabolome of cells in (A) animal 1 and (B) animal 2. 101

Figure 4. 6 Volcano plots comparing cells in (A) lobe A (B) lobe B and (C) lobe C with the random cell population in amine and phenol metabolome of animal 1, (D) lobe A (E) lobe B and (F) lobe C with the random cell population in amine and phenol metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.04 for (A) and (B) and q < 0.06 for (C)). FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point). 102

Figure 4.7 Venn diagrams showing significantly altered amines and phenols containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of (A) animal 1 and (B) animal 2. 104

Figure 4. 8 ANOVA for amine and phenol metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). (A) In animal 1, the 1369 significant metabolites ($p < 0.05$) are shown in red, while 1857 non-significant metabolites ($p > 0.05$) are shown in green. (B) In animal 2, the 1131 significant metabolites ($p < 0.05$) are shown in red, while 3340 non-significant metabolites ($p > 0.05$) are shown in green. 105

Figure 4. 9 PCA score plots of hydroxyl submetabolome of cells in (A) animal 1 and (B) animal 2. 108

Figure 4. 10 PLS-DA score plots of hydroxyl submetabolome of cells in (A) animal 1 and (B) animal 2. 109

Figure 4. 11 Volcano plots comparing cells in (A) lobe A (B) lobe B and (C) lobe C with the random cell population in hydroxyl metabolome of animal 1, (D) lobe A (E) lobe B and (F) lobe C with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to $q < 0.02$ for (A), (B) (C)), FC $\geq$ 1.5 was determined as increase (red points) and FC $\leq$ 0.67 as decreased (blue point). 110

Figure 4. 12 Venn diagrams showing significantly altered hydroxyl containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of (A) animal 1 and (B) animal 2. 112

Figure 4. 13 ANOVA for hydroxyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). (A) In animal 1, the 3609 significant metabolites ($p < 0.05$) are shown in red, while 2928 non-significant metabolites ($p > 0.05$) are shown in green. (B) In animal 2, the 3328 significant metabolites ($p < 0.05$) are shown in red, while 2396 non-significant metabolites ($p > 0.05$) are shown in green. 113

Figure 4. 14 PCA score plots of carbonyl submetabolome of cells in (A) animal 1 and (B) animal 2. 116

Figure 4. 15 PLS-DA score plots of carbonyl submetabolome of cells in (A) animal 1 and (B) animal 2. 117

Figure 4. 16 Volcano plots comparing cells in (A) lobe A (B) lobe B and (C) lobe C with the random cell population in carbonyl metabolome of animal 1, (D) lobe A (E) lobe B and (F) lobe C

with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.03 for (A) and (C), q < 0.05 for (B)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point). 118

Figure 4. 17 Venn diagrams showing significantly altered carbonyl containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of (A) animal 1 and (B) animal 2. 120

Figure 4. 18 ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). (A) In animal 1, the 2006 significant metabolites (p <0.05) are shown in red, while 2410 non-significant metabolites (p >0.05) are shown in green. (B) In animal 2, the 2891 significant metabolites (p <0.05) are shown in red, while 2067 non-significant metabolites (p >0.05) are shown in green. 121

Figure 4. 19 PCA score plots of carboxylic submetabolome of cells in (A) animal 1 and (B) animal 2. 124

Figure 4. 20 PLS-DA score plots of carboxylic submetabolome of cells in (A) animal 1 and (B) animal 2. 125

Figure 4. 21 Volcano plots comparing cells in (A) lobe A (B) lobe B and (C) lobe C with the random cell population in carboxylic metabolome of animal 1, (D) lobe A (E) lobe B and (F) lobe C with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.03 for (A), q < 0.05 for (B) and q < 0.07 for (C)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point). 126

Figure 4. 22 Venn diagrams showing significantly altered carboxylic containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of (A) animal 1 and (B) animal 2. 128

Figure 4. 23 ANOVA for carboxylic acid metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). (A) In animal 1, the 442 significant metabolites (p <0.05) are shown in red, while 3189 non-significant metabolites (p >0.05) are shown in green. (B) In

animal 2, the 228 significant metabolites ($p < 0.05$) are shown in red, while 3041 non-significant metabolites ($p > 0.05$) are shown in green. 129

Figure 4. 24 Box plots of showing cell-to-cell variability of (A) glucose and (B) pyruvate in different cell populations of animal 1 and 2. 131

Figure 4. 25 Box plots showing cell-to-cell variability of (A) glyceraldehyde, (B) D-arabino-hex-3-ulose 6-phosphate and (C) sedoheptulose 7-phosphate in different cell populations of animal 1 and 2. 132

Figure 4. 26 Box plots showing cell-to-cell variability of (A) oxaloacetate and (B) succinic acid in different cell populations of animal 1 and 2. 133

Figure 5. 1 Four-level of metabolite identification confidence. 140

Figure 5. 2 Extracted ion chromatograms of RT calibrants. 145

Figure 5. 3 Base peak chromatograms of urine sample with triplicate injection in (A) positive ion mode and (B) negative ion mode. 147

Figure 5. 4 RTs distribution in HILIC MS/MS-RT Library. 149

Figure 5. 5 (A) Gradient elution profile of gradient I and II (B) RTs distribution of 270 compounds using two gradients. 151

Figure 5. 6 Extracted ion chromatograms of RT calibrants. 153

Figure 5. 7 Multi-point linear regression of calibrants RT for both systems. 154

Figure 5. 8 Base peak chromatograms of urine sample with triplicate injection in (A) positive ion mode and (B) negative ion mode. 155

Appendix Figure A3. 1 PLS-DA model validation result of the single-cell metabolome data set in heat stress study with 100 permutations. 168

Appendix Figure A4. 1 PLS-DA model validation result of the amine and phenol metabolome of single cells in animal 1 with 100 permutations. 169

Appendix Figure A4. 2 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for amine and phenol metabolomics of cells of group A, B and C in animal 1. 170

Appendix Figure A4. 3 PLS-DA model validation result of the amine and phenol metabolome of single cells in animal 2 with 100 permutations. 171

Appendix Figure A4. 4 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for amine and phenol metabolomics of cells of group A, B and C in animal 2. 172

Appendix Figure A4. 5 PLS-DA model validation result of the hydroxyl metabolome of single cells in animal 1 with 100 permutations. 173

Appendix Figure A4. 6 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for hydroxyl metabolomics of cells of group A, B and C in animal 1. 174

Appendix Figure A4. 7 PLS-DA model validation result of the hydroxyl metabolome of single cells in animal 2 with 100 permutations. 175

Appendix Figure A4. 8 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for hydroxyl metabolomics of cells of group A, B and C in animal 2. 176

Appendix Figure A4. 9 PLS-DA model validation result of the carbonyl metabolome of single cells in animal 1 with 100 permutations. 177

Appendix Figure A4. 10 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carbonyl metabolomics of cells of group A, B and C in animal 1. 178

Appendix Figure A4. 11 PLS-DA model validation result of the carbonyl metabolome of single cells in animal 2 with 100 permutations. 179

Appendix Figure A4. 12 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carbonyl metabolomics of cells of group A, B and C in animal 2. 180

Appendix Figure A4. 13 PLS-DA model validation result of the carboxylic acid metabolome of single cells in animal 1 with 100 permutations. 181

Appendix Figure A4. 14 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carboxylic acid metabolomics of cells of group A, B and C in animal 1. 182

Appendix Figure A4. 15 PLS-DA model validation result of the carboxylic acid metabolome of single cells in animal 2 with 100 permutations. 183

Appendix Figure A4. 16 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carboxylic acid metabolomics of cells of group A, B and C in animal 2. 184

List of Tables

Table 5. 1 List of 28 metabolites in RT calibrants. 146

Table 5. 2 Identification result of human urine samples using the MS/MS-RT library. 148

Table 5. 3 Comparative results of metabolite identification of human urine sample in two different laboratories. 148

Table 5. 4 HILIC mobile phase buffer optimization. 150

Table 5. 5 List of 14 metabolites in RT calibrants. 153

List of Abbreviations

%RSD	Percent relative standard deviation
ACN	Acetonitrile
ANOVA	Analysis of variance
BPC	Base peak chromatogram
$CaCl_2$	Calcium chloride
CAWG	Chemical analysis working group
CI	Chemical ionization
CID	Collision-induced dissociation
CIL	Chemical isotope labeling
CYP2E1	Cytochrome P450 2E1
DmPA	P-dimethylaminophenacyl
DNMs	De novo mutations
EI	Electron ionization
EIC	Extracted ion chromatogram
EM	Electron multiplier
ESI	Electrospray ionization
FA	Formic acid
FC	Fold change
FTICR	Fourier transfer ion cyclotron resonance
GC	Gas Chromatography
GSH	Glutathione
HCC	Hepatocellular carcinoma
Hepes-NaOH	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid-sodium hydroxide

HER	Human epidermal growth factor receptor
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human metabolome database
IS	Internal standard
KCI	Potassium chloride
LAESI-MS	Laser ablation electrospray ionization-mass spectrometry
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
MCP	Microchannel plate
MeOH	Methanol
$MgCl_2$	Magnesium chloride
MS	Mass spectrometry
MS-222	Tricaine methanesulfonate
MSI	Metabolomics standards initiative
m/z	Mass-to-charge ratio
NaCl	Sodium chloride
Na_2CO_3	Sodium carbonate
$NaHCO_3$	Sodium hydrogen carbonate
Na_2HPO_4	Sodium hydrogen phosphate
Nano-DESI	Nanospray desorption electrospray ionization-mass spectrometry
NGS	Next generation sequencing
NH_4Ac	Ammonium acetate
NH_4FA	Ammonium formate
NMR	Nuclear magnetic resonance

OSCC	Oral squamous cell carcinoma
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PLS-DA	Partial least square discriminant analysis
PMT	Photomultiplier tube
QC	Quality control
Q-TOF	Quadrupole time-of-flight
RCF	Relative centrifugal force
RPLC	Reversed phase liquid chromatography
RT	Retention time
SAGE	Serial analysis of gene expression
SD	Standard deviation
SILA	Stable isotope-labeled analog
SIM	Selected ion monitoring
S/N	Signal-to-noise ratio
SRM	Selected-reaction monitoring
TIC	Total ion chromatogram
TLF	Time-lag focusing
TMS	Tricaine methanesulfonate
TOF	Time-of-flight
UIS	Universal internal standard
UPLC-MS	Ultra-performance mass spectrometry
WHO	World Health Organization

List of Symbols

Å	Angstrom
℃	Degrees Celsius
μL	Microliter
μL/min	Microliters per minute
μm	Micrometer
cm	Centimeter
Da	Dalton
eV	Electron-volt
h	Hour
Hz	Hertz
L/min	Liters per minute
mg/mL	Milligrams per milliliter
min	Minute
mL	Milliliter
mm	Millimeter
mM	Millimolar
nm	Nanometer
ppm	Parts per million
rpm	Revolutions per minute
s	Second
V	Volt
v/v	Volume per volume

Chapter 1 Introduction

1.1 The "Omics" Science

"Omics" is a branch of study in biological sciences aimed at identifying, characterizing and quantifying all the biological molecules in a cell, tissue, or organism holistically. It is a high-dimensional biology including four levels of "omics", genomics, transcriptomics, proteomics and metabolomics[1, 2].

Genomics was first coined by Thomas Huston Roderick, an American geneticist in 1986. It is a field of biology which involves the characterization and quantification of the entire set of genetic material, *i.e.*, genomes, in an organism. Genome is the full DNA set, including all the genes in an organism. DNA microarrays is one of the most common methods used in genomics, it is capable of measuring the expression of thousands of genes simultaneously. The principle of microarrays is the binding of the cDNA or cRNA sample, *i.e.*, the target, to the probe, which can be a specific DNA sequence or a short section of a gene, on the array. Incorporation of fluorescently labeled nucleotides in the cDNA or cRNA is usually involved so that the relative concentration of nucleic acid species in the sample solution can be determined by detecting the fluorescence signal at each spot. Thus, the signal intensity on each spot of microarrays is used as a measure of the expression level of the corresponding gene. This technology has been used to reveal the effects of certain treatment on gene expression. The microarray gene expression profiling can be used to spot out genes whose expression level changed significantly in response to drugs by comparing the gene expression of cells in treatment group to that in the control group. There are several limitations of DNA microarrays, for example, it is hard to design arrays that can detect only a particular gene in the genome if related genes are present. Moreover, it is an indirect

way to measure the relative concentration of genes so the signal is linear only for a limited concentration range[3].

Transcriptomics is the study of the complete complement of mRNA molecules, *i.e.*, transcriptome, generated by a cell type or an organism. It was first proposed by European scientist Charles Auffray in 1996. In this field, scientists focus on the profiling of transcriptomes in different biological systems. Various technologies are involved, from Serial Analysis of Gene Expression (SAGE) and microarrays at early stage to more recent revolution, RNA-seq[4]. RNA-seq is a technique for revealing the sequences and quantity of RNA in a sample using next generation sequencing (NGS), it enables the analysis of the transcriptome of gene expression patterns encoded within the RNA. An example of the application of RNA-seq is alternative splicing. Alternative splicing is a process that occurs during gene expression, which allows one gene to code for multiple proteins, thus it is essential to increase the diversity and functional capacity of a gene at the post-transcriptional level[5]. However, in some situations, mutations would occur and disrupt the regulation of alternative splicing, and this can lead to human disease such as cancer[6]. Therefore, the progress of RNA-seq to target alternative splicing has led to the development of novel therapeutic, which is beneficial to human health. The major limitation of the technique is the difficulty in quantifying and controlling artifacts and biases in certain situations.

Proteomics was coined by an Australian scientist, Marc R. Wilkins in 1994. It aims to characterize and quantify all sets of proteins, *i.e.*, proteome, in a biological system. There are various methods to study proteins, an example is immunoassays, which use antibodies for protein detection. Mass spectrometry has become the most common technology for current large-scale and high-throughput proteomics. For the practical application, it can be used in the discovery of novel drugs. By identifying proteins associated with a disease, drugs can be designed to interfere their

functions based on the structure information of these proteins. It can also be used in protein biomarkers discovery for disease diagnosis[7]. However, the limitation of proteomics is the complexity in analysis owing to the highly dynamic nature of the proteome. The separation of proteins as well as the characterization of post-translational modifications can also increase the difficulties of proteomics study.

Metabolomics was first proposed by Steven Oliver in 1998, and it is the endpoint in the omics cascade. It is the comprehensive study of endogenous and exogenous low-molecular-weight (< 1500 Da) small molecules, *i.e.*, metabolome, in a biological system. Metabolome as the downstream products of transcriptome and proteome, is highly associated with the phenotype of the biological system being studied. Because the changes and interactions of gene and protein expression with the environment are reflected in the metabolome directly, metabolomics is more complex and time sensitive compare to other "omics" approaches. This put metabolomics at a distinct advantage of being the best representative of the molecular phenotype of health and disease over other approaches. Since it can monitor the real-time biochemical activity and dynamics of a biological system, the number of metabolic works has grown considerably in recent years. More details about metabolomics will be provided in Section 1.2.

To summarize, the beauty of the "omics" science is that each "omic" technology provides unique information at different levels, which together gives a complex and comprehensive biological picture of the cell, tissue, or organism. Therefore, we can have a better understanding about how individual cells or different types of tissue are organized and controlled through a holistic multi-omics study.

1.2 Metabolomics

1.2.1 Importance of Metabolomics

Metabolite measurements are increasing in importance because it has made notable progresses in different areas of life sciences, such as biomarker discovery and biological study.

In biomarker discovery, the objectives are to detect and quantify metabolites in order to discover metabolite biomarkers for potential applications in, for example, clinical diagnosis, disease treatment and therapy. Given that pathological conditions can change various metabolic pathways, metabolic profiles of biofluids are different in disease and control groups. Therefore, metabolomics study always involves the investigation of metabolites in these two groups, and aims to find and identify specific biomarkers for disease diagnosis and even prediction. For instance, work presented by Qihui and colleagues identified five salivary biomarkers with high accuracy, sensitivity and specificity for early diagnosis of oral squamous cell carcinoma (OSCC), which is the most common oral cancer[8]. Another study by Guozhu and colleagues discovered that five metabolites are favorable to the prediction of early recurrence of hepatocellular carcinoma (HCC), one of the common types of primary liver cancer, with urinary metabolic profiling using GC-TOF MS[9].

In biological study, the main purpose is to quantify metabolomics changes induced by one or more effectors, so to study the perturbations of metabolic networks. The change and interaction of gene and protein expression with the environment are also reflected in the metabolome pathways. For instance, a study performed by Hong and colleagues investigated the metabolites and biological pathways that can distinguish children with anorexia nervosa from healthy children with the use of ultra-performance mass spectrometry (UPLC-MS) together with multivariate data analysis and pathway analysis[10]. In this study, rat model is used. Rats in the healthy group, which

is regarded as the control group were fed conventional food for 28 days, while rats in the model group were fed with a special feed to achieve the anorexia nervosa condition. With the perturbation induced in the model group, which is the change in normal diet, blood samples in two groups of rats were then collected and used in metabolomics analysis. The results of their study show that there are 26 metabolites changed significantly in the model rats compared to the control rats. Further analysis with the use of MetaboAnalyst platform indicated those differentiated biomarkers were involved in different metabolic networks such as D-glutamine and D-glutamate metabolism, alanine, aspartate and glutamine metabolism and the TCA cycle. And the study of metabolic pathways is also an important way to elucidate pathogenic mechanisms of anorexia nervosa.

Taken all the above together, it is obvious that metabolomics plays an important role in the understanding of diseases. It could provide insights into the underlying causes of diseases, and if these can be controlled, it could lead to disease prevention eventually. Thus, metabolomics is important to improve the quality of life.

1.2.2 Challenges of Metabolomics

Metabolomics is an increasingly important area in 'omics" sciences. However, the study of metabolites has six major challenges which are noteworthy[11]. First of all, fast dynamics of metabolites. Metabolites can interact with the environment on a very short time scale of seconds or even less. Thus, it is challenging to measure the metabolome in a biological system accurately. Secondly, wide diversity of metabolites. In contrast to genes and proteins, metabolites have very different chemical and physical structures, which make them difficult to study in one single analysis. Usually, more than one method has to be employed for comprehensive study of the metabolome as no one method can detect different classes of metabolites in once. Therefore, only

limited metabolome coverage can be achieved with most of the current technologies. Thirdly, metabolites cannot be amplified. Unlike gene or DNA which can be amplified by polymerase chain reaction (PCR), we cannot make copies of metabolites with low concentrations. This makes the analysis of metabolites, particularly those with low abundance challenging. Fourthly, large dynamic range of metabolites. They can range from a couple molecules to 10^{10} molecules per cell. It would be hard to use a method which can cover a wide range. Fifthly, it is not desirable to give metabolites fluorescent tags in order to make them detectable, as this could alter the activity of metabolites. Last but not least, it is difficult to identify unknown metabolites. The metabolites in the current library only represents part of the metabolites in the biological system, so continuous works to expand the library is necessary in order to achieve high coverage in metabolomics study.

1.2.3 Platforms for Metabolomics

Currently, major technologies for metabolomics, including single-cell metabolomics, the major work presented in this thesis, are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).

NMR spectrometry is an analytical tool widely used in chemistry for identifying and quantifying the chemical composition of a sample and for structural elucidation on the basis of monitoring local magnetic fields around the atomic nuclei. MS is a technique mainly applied in the identification and quantification of molecules in a sample based on measuring the mass-to-charge ratio of ions. Both technologies aim to detect, identify and quantify the metabolites in a given biological system of interest. For NMR-based metabolomics, NMR spectroscopy is a nondestructive technique with high reproducibility, so sample can be recovered and stored for a longer period of time. It requires minimal sample preparation such as sample derivatization and

separation before analysis, which is more efficient, compared to MS-based works in general[12, 13]. However, the sensitivity of NMR has limited its applications in metabolomics owning to the fact that metabolites cannot be amplified and methods with low detection limit is very important in analysis. Thus, researches attempting to improve the sensitivity of NMR have become more popular in order to address the challenges[14, 15].

MS-based metabolomics provides a powerful platform for different metabolomics studies. MS offers a number of benefits compared to NMR spectrometry, with the most advantage being its high sensitivity and selectivity. Detection limit of MS can reach nanomolars which is particularly good with samples of limited size and volume, such as single cells. Moreover, MS can be used for both selective (targeted analysis) and nonselective (non-targeted analysis) works. Furthermore, there are different MS technologies available, with different combination of ionization and detection methods, which can potentially increase the coverage of detected metabolites. In short, MS is a powerful tool for studying metabolomics, especially when combined with different separation techniques such as liquid chromatography (LC) and gas chromatography (GC)[16-18], to separate the complex and diverse metabolite mixtures before MS analysis.

1.2.4 Workflow for Metabolomics

Generally, metabolomics workflow involves a biological or clinical study design, sample preparation, metabolomic analysis, data processing and analysis, metabolite identification and quantification, and finally, interpretation of results.

For example, a workflow on fecal metabolomics was conducted by Wei and colleagues in 2014[19]. Human fecal samples contain many endogenous human metabolites, gut microbiota metabolites together with other compounds. The profiling of the fecal metabolome can produce

metabolic information that would be useful not only in disease biomarker discovery, but also could provide insights about the relationship between the gut microbiome and human health. The design of this study was to apply the differential chemical isotope labeling liquid chromatography mass spectrometry (LC-MS) to profile the amine and phenol metabolome of human fecal samples. For sample preparation, fecal samples were collected from three families, and each family in this study is made up of one male and one female in their late 20s, and an infant of 1-2 months. Water (H_2O), methanol (MeOH) and acetonitrile (ACN) were used to extract metabolites. Dansylation labeling[20, 21] was performed followed by LC-UV for sample quantification[22], and finally sample was injected to electrospray ionization (ESI) time-of-flight mass spectrometer. In data processing and analysis, IsoMS was used to process the raw data generated from LC-MS runs and peak-pair intensity ratio calculation was performed for quantitative metabolomics[23]. In metabolites identification, a total of 6200 peak pairs were detected from 243 LC-MS runs of all the fecal samples, and 67 of them were positively identified based on the mass and retention time match to an expandable dansyl standard library. In addition to this, 582 and 3197 peak pairs were putatively identified based on mass match using MyCompoundID against a Human Metabolome Database and an Evidence-based Metabolome Library, respectively. At the end, for the results interpretation, the conclusion of this study is that diet might affect the metabolome. However, the variations among families or even individuals exerted greater effects than day-to-day variations.

The workflow of all metabolomics studies always follows the routine described above with different methods and platforms being used depending on the sample properties and availabilities of the instrumentation.

The workflow for single-cell metabolomics, the main work presented in this thesis, will be discussed in detail in Section 1.3.5.

1.2.5 Mass Spectrometry for Metabolomics

As mentioned in Section 1.2.3, NMR and MS are two common analytical platforms for metabolomics. Here, the discussion will be focused on the details of MS, which is the technique employed in all the work presented in this thesis.

MS is an analytical technique that provides excellent sensitivity, specificity and speed for qualitative and quantitative analysis of metabolites. Mass spectrometer is an instrument used for measuring the mass of an ion. Basically, it is composed of five different components, which are sample introduction, ionization, mass analyzer, ion detector and finally a data system. A vacuum system operating at $< 10^{-5}$ torr is required in the ionization chamber, mass analyzer and ion detector in order to minimize the chance of collisions between ions and sample molecules. Moreover, operating in high vacuum can reduce background spectra, and ions generated in the ionization chamber can have a long free path without hitting air molecules. The measurement is shown in a mass spectrum, which is a plot of detector response, *i.e.*, the signal intensity, as a function of mass-to-charge ratio (m/z). The peak area in the plot is related to the abundance of the ions. In general, a mass spectrum of a pure compound consists of two peaks types, which are the molecular ion peak, providing the molecular mass of the compound, and fragment ion peaks used for structural analysis of the compound.

Depending on the sample state (solid, liquid, gas) and properties (non-volatile or volatile), there are different ways to introduce the sample into the mass spectrometer. For volatile sample, which is referring to chemicals that can evaporate at room temperature or below to form vapour, Gas Chromatography (GC) is commonly used to vaporize and inject the sample to a chromatographic column, where separation of compounds in the mixture occurs in the presence of inert carrier gas mobile phase, such as helium. When coupled with MS, the GC-MS instrument

can separate the chemical mixtures (by the GC component) and identify the components (by the MS component). On the other hand, for non-volatile sample, if it is thermally stable, GC/MS involving high temperature can be used. However, if the sample is thermally instable, other approaches will need to be considered. For example, Liquid Chromatography (LC) can be used for the analysis of thermally unstable compounds. In LC, liquid sample would be injected into a stream of mobile phase flowing through a column packed with stationary phase, sample components are then separated from one another by differential migration as they pass through the column. And when coupled with MS, LC-MS, samples can be introduced, separated and analyzed by the MS.

Ionization is the next important stage, as ions must be formed and transferred into gas phase before they can be analyzed by MS. Ionization occurs in the ion source, and depending on the properties of samples and applications of the study, different ion sources can be used. Here, ionization source is discussed in two forms: hard ionization and soft ionization. For hard ionization, it is a process in which high energy is given to the molecule, resulting in extensive fragmentation. The most common example is the Electron Ionization (EI). In EI, the ionization source is made of a filament and an electron receiving plate. Current is applied to the filament so that its temperature rises and gives off electrons. A small voltage is then applied across the filament and the receiving plate, so that the electrons are accelerated to the receiving plate. And when neutral molecules are introduced into the ionization sources region, electrons will collide with the molecules and some of the molecules will absorb enough energy from the electrons to be ionized. These ions can be moved to the mass analyzer by applying a suitable voltage. Because of the high excitation energy, the EI spectrum is composed of many fragment ions peaks and sometimes, no molecular ion information is generated. The fragment ions are particularly useful for chemical identification as the spectrum serves as a fingerprint of a molecule. It can be easily done by comparing the EI

spectrum of the unknown compound with the known compounds in a library. On an entirely different ground, soft ionization is a process which imparts little energy to the molecule, resulting in limited fragmentation. Examples of this include Chemical Ionization (CI), and Electrospray Ionization (ESI). CI and EI are always complementary to each other and thus commonly used together for analyzing a compound. EI generates many fragment ions which are useful for structural analysis and compound identification, whereas CI generates the protonated molecule which is useful for determining the molecular mass of the compound. The principle of CI is that current is applied so that electrons are given off from the filament and are accelerated to the receiving plate. However, different from EI, electrons will collide mainly with the reagent gas here, ions are thus produced from the electron ionization of the reagent gas molecules. Some form of the reagent ions can then react with the neutral analyte molecules to form analyte ions, and this reaction does not transfer as much excess energy to the analyte ions as in EI. Another soft ionization method to be discussed in detail here is ESI since it is used in the studies presented in this thesis. ESI was first described by Zeleny in 1917, and the coupling of it to MS was performed in 1984 by Yamashita and Fenn[24]. The mechanism of ESI, in particular the positive ionization mode, as described by Kebarle and Tang[25], involves four major processes. These are 1) the production of charged droplets at the ESI capillary tip by applying a high voltage, 2) shrinkage of charged ESI droplet by flowing a dry nitrogen gas, 3) repeated droplet disintegrations which happened because the repulsion force on the droplet surface is larger than the surface tension force and 4) generation of gas phase ions, in which two mechanisms, single ion in droplet theory[26] and ion evaporation theory[27] have been proposed. ESI is a great ionization method for LC-MS owing to the fact that it converts the neutral molecules in the solution into gas phase ions without applying any heat to the molecules. In this way, molecular ions generated are usually free of thermal-

induced decomposition. In other words, the molecular ions are stable and usually do not undergo dissociation to give fragment ions. Because of this, ESI is particularly popular for thermally-labile samples.

After ionization, mass analyzer is used to separate the ions according to their mass-to-charge ratio. It maximizes the resolved ion intensities and outputs them to the detector where they can be detected and later converted to a digital output. There are various types of mass spectrometers and they are usually different only in mass analyzers while the other parts are very similar. When deciding which analyzer is suitable for a study, a number of factors would be taken into consideration. First of all, mass range, which is the maximum and minimum m/z a particular mass analyzer can detect. Second, resolution, which is the ability of an analyzer to separate ions of two adjacent mass peaks. Depending on the samples, mass analyzer of different resolutions could be needed. Third, mass measurement accuracy. This depends on the use of a proper mass scale for calibration, *i.e.*, internal or external calibration method. Fourth, ion transmission, which is related to the sensitivity of ion detection. Fifth, spectral recording speed. This is the rate at which a mass spectrum can be acquired and is generally given in mass units per unit time. In general, the higher the speed, the lower the resultant resolution and sensitivity. Lastly, versatility. This is how easy it is to interface the mass analyzer to different separation methods such as GC, LC and CE. Examples of the mass analyzer are quadrupole, time-of-flight (TOF), ion trap, fourier transfer ion cyclotron resonance (FTICR), linear ion trap and orbitrap. As the studies presented in this study mainly involved the use of TOF mass spectrometer, thus it will be discussed in detail in the following. The principle of TOF analyzer is ions which are accelerated by electric field are adjusted in different ways in which they all have the same initial kinetic energy if they have the same charge. Their velocities are thus depending on the m/z, *i.e.*, ions of the same charge which are heavier

would reach the detector later than those with smaller m/z. The time taken for the ions to reach the detector at a known distance is then measured and become a measure of the ions m/z provided that standards are used to calibrate the mass and time scale in prior. There are two ways to increase the resolution of TOF, they are used to correct for the initial velocity and energy spread in the ions respectively. In this way, ions of same m/z can arrive at the detector simultaneously and thus improve the resolution of the peak in the mass spectrum. The first one is known as the time-lag focusing (TLF) TOF[28], which is very effective to focus the axial velocity distribution of ions. To be more specific, amplitude of the pulse can be adjusted so that the initially less energetic ions can catch up to the initially more energetic ions when they both arrive at the detector. The second one is the reflectron TOF[29] in which ion reflectors are used to compensate for the initial energy distribution and focus ions having the same m/z value to the detector. This can be achieved because the faster ions will penetrate deeper into the field created by the reflector and thus take a longer time to return than the slower ions. As a result, the fast and slow ions are focused in time at the detector. The resolution of TOF can be very high with the use of reflectron TOF. Moreover, good mass accuracy can be achieved with good calibration. Furthermore, it can be readily combined with LC and a pulsed ionization method such as laser ionization which broadens its applications. Taken together, it has become one of the most popular analyzers in MS.

For ion detection, there are several types of ion detectors available for mass spectrometer. The electron multiplier (EM) detector is very commonly used, examples of EM are the discrete dynode multiplier and the continuous dynode multiplier. The principle of these multiplier depends on the production of secondary electrons when ions impact onto the surface of dynode, the electrons gain amplifies the signals generated from incident ions of analytes, and the detector output after amplification is traced on a recorder. Many improvements have been made on the EM,

the development of microchannel plate (MCP) multiplier is an example. Since MCP multiplier has many separate channels and each channel is a channel electron multiplier, it has the added advantage of providing spatial resolution. Another type of ion detector is the scintillation detector. In this detector, ions or electrons are first converted to photons at a scintillator such as phosphor. The photons released are then detected by an electron light sensor such as the photomultiplier tube (PMT).

The last component of a mass spectrometer is the data system. A wide range of data systems have been used, ranging from microcomputers and programmable desk calculators to powerful computers. Here, data is shown as mass spectra, and depending on the purpose of study, qualification and quantitation can be done. For example, in full scan mode, total ion chromatogram (TIC) is produced by adding ion intensities of all m/z's in each spectrum recorded at a specific retention time. It is shown as the total ion intensity as a function of retention time, and is particularly useful in qualitative study. While in selected ion monitoring (SIM) mode in which one or more selected m/z peaks are monitored, the signal-to-noise ratio (S/N) is generally 10-100 times higher than full scan mode as more time is allowed for the integration of the chromatographic peak specified. Owing to the higher sensitivity on the selected m/z peaks, it is particularly useful for quantitative purpose.

1.2.6 High-Performance Chemical Isotope Labeling (CIL) LC-MS for Quantitative Metabolomics

1.2.6.1 Quantitative Metabolomics

Quantitative MS in metabolite measurements has received considerable critical attention. People not only care about the identity of the metabolites, but also the concentrations. In

quantitative MS, selected-ion monitoring (SIM) and selected-reaction monitoring (SRM) are extensively used during data acquisition. For SIM, only one or a few selected ions are monitored, thus it provides better sensitivity when compared to full scan. In general, the molecular ion is selected for higher sensitivity in quantitation as it is the most abundant ion in most cases. It is important to note that a separation technique such as HPLC is always used to reduce interfering species. SRM involves the use of tandem MS. Firstly, MS 1 is used to mass-select the analyte ion of interest, and here, a lot of interfering ions having the same m/z may be selected at the same time. Secondly, collision-induced dissociation (CID) of this ion will generate different fragment ions. Finally, in MS 2, a m/z value corresponding to the mass of a specific fragment ion from CID of the analyte ion is selected and this fragment ion is detected. SRM is very specific for monitoring an analyte of interest with very little interference from interfering ions since the coincidence of any fragment ions from the interfering ions at the same m/z value as the analyte fragment is very rare.

The accuracy of mass measurements in quantitative MS depends on the use of a proper mass scale for calibration. Calibration is to find out the correlation between a known concentration of the analyte and the resulting MS signal. There are three common methods for calibration, which are the external standard method, standard addition method and internal standard method. In external standard method, standard solutions with known concentrations are first prepared and analyzed by MS, signal intensity is then plotted against the concentrations to establish a calibration curve. Least square analysis is used to figure out the relation between the MS response and standard concentrations. Concentration of the unknown can thus be determined by using this equation. However, this method has low accuracy and precision as the matrix of the calibration standards is difficult to match with the sample matrix. Also, as the standards and samples are run

at two different times, the MS response may be different. In standard addition, known concentrations of standard solution of the analyte are added to the unknown sample and the signals are measured. Calibration curve is plotted and the unknown concentration can be determined by extrapolating the linear equation to y=0. This method can overcome the matrix effect and instrumental signal response fluctuations but is very time consuming and tedious. In internal standard method, a constant amount of internal standard (IS) solution with known concentration is added to all calibration solutions of known concentrations of the analyte, in this way, the IS and the analyte are detected by the MS at the same time. The ratio of their signals is plotted against the analyte concentration to establish a calibration curve and this can be used to find the analyte concentration in the unknown. This method provides high accuracy and precision as the internal standard can compensate for any fluctuations in the MS response and the samples losses during sample preparation and chromatographic steps. In other words, the standards are affected to the same extent as the analyte. Therefore, it is good to use IS method for metabolite quantification. There are two types of internal standards: structural homologous and stable isotope-labeled analog (SILA). In the next section, a CIL LC-MS method for quantitative metabolomics will be discussed in detail.

1.2.6.2 CIL in Large-Scale Metabolite Quantification

In the last section, we mentioned that SIL analog is an ideal internal standard. Yet, it may not be feasible for large-scale and comprehensive metabolome quantification. The reason for this is that synthesizing SIL analogues of all metabolites is not practical. Chemical Isotope Labeling (CIL) method has emerged as a good alternative to acquire data related to large number, wide and

diverse range of metabolites. In this section, the use of CIL LC-MS in quantitative metabolomics will be discussed thoroughly.

CIL can be considered as a chemical derivation reaction, in which a chemical reaction is used to attach an isotope-mass-encoded tag to metabolites. CIL could mitigate the ion suppression effect on the premise that the isotope is properly chosen. For example, when deuterium is used as the isotope tag, isotopic effect is often found. To be more specific, in RP LC-MS, the deuterium-tagged standard elutes at different retention time from the analyte. This is due to the fact that the IR oscillation frequency of the C-D bond (2334 cm^{-1}) is lower than that of the C-H bond (3300 cm^{-1}), and this could induce less attraction forces between the C-D bond and the stationary phase and thus the deuterium-tagged standard always elute before the nondeuterated one[30, 31]. The problem of this is that as they elute at different retention time, they experience different extent of ion suppression effect compared to other co-eluting compounds. As a result, these two peaks would have different signal response and thus lead to inaccurate quantification. On the contrary, when ^{13}C-tagged labeling reagent is used, the ^{12}C-tagged analyte and the ^{13}C-tagged standard would coelute perfectly, *i.e.*, resulting in no isotopic effect and more accurate metabolite quantification.

Quantification is one of the important works in metabolomics which provides information on metabolite concentrations. It opens up the door for subsequent statistical analysis of the samples being studied which is essential for biological interpretation. The quantification workflow with CIL is that analytes in individual samples are labeled with ^{12}C-labeling reagent, while another mass-difference isotope tag, *i.e.*, ^{13}C-labeling reagent is added to the same analyte in another comparative sample (for relative quantification) or standard (for absolute quantification). In most of the studies, the comparative sample is usually a pooled sample, which is produced by mixing equal aliquots of all individual samples in the study. The two labeled samples are then mixed for

mass spectrometric analysis. The peak pair ratio of the isotope labeled analyte pair represents the relative or absolute concentration of metabolites in the sample.

The first development of CIL in large-scale metabolite quantification is the dansylation chemistry for labeling the amine and phenol containing metabolites. As amines and phenols are highly polar and hydrophilic in nature, they have serious ion suppression issues in ESI-MS, commonly originated from matrix molecules or co-eluting compounds in LC-MS runs[20]. It has been demonstrated that with CIL dansylation reactions by attaching ^{12}C- and ^{13}C-dansyl chloride to individual samples and pooled sample respectively, ion suppression issues are reduced, leading to increased coverage when compared with methods without labeling. Apart from reducing ion suppression, CIL dansylation has several more benefits. Firstly, dansylation labeling changes the chromatographic behavior of polar metabolites such that they can now be retained and separated by RPLC. Secondly, dansylation labeling can enhance the ESI signal compared with the underivatized metabolites. Thirdly, it can enhance the ESI response since the labeling tag gives metabolites stronger chargeability and increases hydrophobicity, making the metabolites elute in higher percentage of organic solvent. It is desirable for metabolites to elute later in the gradient run because the sensitivity of ESI-MS detection near the initial void may be significantly reduced as a result of inadequate ESI desolvation. Taken together, CIL gives improved separation and enhanced MS detection which is critical for increasing the coverage.

1.2.6.3 Divide-and-Conquer Approach

The rationale of CIL is the divide-and-conquer approach, which is also known as the chemical-group-based submetabolome profiling. The idea is to divide the whole metabolome of a biological system into four different groups based on their properties such as chemical structures

and hydrophobicity. These four different groups, or more precisely, four different submetabolome are the amine/phenol, carboxylic acid, carbonyl and alcoholic hydroxyl. A specific CIL method, each tailored to a group of metabolites is then applied, followed by the optimized LC-MS methods, in attempting to systematically describe the complete set of all metabolites in a biological sample. Apart from the dansylation for the amine/phenol submetabolome, the other three CIL reactions are: base-activated dansylation for hydroxyl submetabolome[32], p-dimethylaminophenacyl (DmPA) bromide labeling for carboxylic acid submetabolome[33] and dansylhydrazine labeling for carbonyl submetabolome[34].

Owing to the fact that metabolites has great diversity in their chemical and physical properties, and in order to achieve comprehensive metabolic profiling, conventional targeted or untargeted LC-MS workflows always involve multiple experimental conditions to increase the coverage. In addition, the detection of various type of metabolites, *i.e.*, polar, non-polar and ionic depend on the use of ionization mode in the MS, so for the sake of increasing the coverage, different ionization modes are used independently. For example, to target the polar metabolites, Hydrophilic Interaction Liquid Chromatography (HILIC) coupled to MS is needed, so HILIC LC, positive ion mode MS and HILIC LC, negative ion mode MS are required. Whereas nonpolar metabolites need Reversed Phase Liquid Chromatography (RPLC) – MS, therefore, RPLC, positive ion mode MS and RPLC, negative ion mode MS are required. Thus, conventional LC-MS approach would require four experimental conditions and LC-MS runs. Obviously, the whole process is inconvenient and cumbersome in practice. One more point is that not all laboratories have access to all these analytical platforms. On an entirely different ground, if we divide the metabolites into groups and perform CIL chemical derivation to alter the chemical properties of the metabolites in a way that they can all be separated with high efficiency in one mode of

separation such as RPLC, which is commonly interfaced to a mass spectrometer, or one ionization mode, such as positive ion mode in electrospray ionization (ESI), this problem can be circumvented. In other words, CIL LC-MS approach allows comprehensive, metabolic profiling in a biological system, with the use of four labeling chemistry and LC-MS runs under one experimental condition, *i.e.*, RPLC-MS, positive ion mode.

To sum up, high-performance CIL LC-MS is a relatively simple platform for quantitative metabolomics with high metabolome coverage. It has four major features summarized in the following. First of all, the chemical derivation reaction, for example, dansylation mentioned above, is simple and robust to perform. It is applicable to a range of amines and phenols with very limited or even no side reaction products. Secondly, CIL is universally applicable to all kinds of samples. Thirdly, it is useful for MS-based metabolome quantification, either relative or absolute depending on the needs with high accuracy and precision. Last but not least, the chemical-group based sub-metabolome profiling which fractionates the metabolome into different groups followed by thorough analysis of each, serves as an important tier in efforts to achieve comprehensive, great depth of chemical coverage in metabolic profiling.

1.2.6.4 Sample Normalization

In quantitative metabolomics, the main purpose is to determine the concentration differences of individual metabolites in two or more comparative samples. In order to generate reliable metabolome profiles for comparison, it is of utmost importance that the amount of starting materials in each sample is comparable. In reality, the concentration of metabolites can be very different from one biological sample to another. For example, total concentration of metabolites

in urine samples can differ by more than 14-fold[35]. Therefore, it is crucial to eliminate the effect of total sample amount variation on quantification of individual metabolites, especially in urinary metabolome profiling.

Sample normalization is a way to account for variations of the overall sample concentrations, or in other words, it can eliminate the effect on metabolome data caused by concentration differences in the samples. In CIL LC-MS, a fast step-gradient LC-UV method has been developed and optimized for quantification of dansyl labeled metabolites, and is readily applicable to many biological samples[22]. The principle of LC-UV quantification is that in dansylation, amines and phenols containing metabolites are labeled with dansyl groups. And since the dansyl group is a good chromophore, LC-UV can be use in quantifying the total labeled metabolites in a biological sample. The detection is optimally performed at the detection wavelength of 338 nm as it has the least absorption compared to other wavelengths, which means 338 nm can provide the least interference from other chromophores. In this fast LC-UV gradient, the labeled metabolites are eluted out together, with the quenching reagent separated in the chromatogram. The UV absorbance is proportional to the sample concentration.

The sample normalization described above, in which the total concentration of metabolites is first measured and the volume of the biological samples to be injected to LC-MS for analysis is adjusted accordingly so that the total sample concentration of all samples is equalized, is known as the pre-acquisition normalization. Another method of normalization is named post-acquisition normalization[36]. In this approach, the sample amounts being injected for metabolomic profiling are different. The individual metabolite signals are then normalized using the total ion signal intensity. Each of this method has its pros and cons, for example, in pre-acquisition normalization, the instrumental responses obtained for all samples are very similar. The results can be used to

facilitate an optimal sample injection amount for metabolomic profiling. On the other hand, post-acquisition normalization is more convenient without using additional experimental steps such as LC-UV. Therefore, depending on the study needs, applications and conditions, different methods can be selected.

1.2.6.5 Optimization of Injection Amount or Volume

It is very important to achieve optimal detectability along with avoiding the sample carryover problem, therefore, optimizing the injection amount or volume is essential in the CIL LC-MS workflow. The optimization workflow is straight forward, after the mixing of ^{12}C- and ^{13}C-labeled samples, different injection amounts and volumes are used in LC-MS runs. The peak pair number is plotted as a function of the injection amount/volume, and the amount/volume that gives the largest peak pair number would be the optimal injection amount/volume[37].

1.2.6.6 Data Processing and Analysis

Metabolomics data processing and analysis are essential to convert data into knowledge, and is a significant part of the metabolomics workflow. In CIL, there is a standard way to process the data.

To begin, all the raw LC-MS data are exported to a CSV file by Bruker Daltonics Data Analysis. Here, total ion chromatogram (TIC) and extracted ion chromatogram (EIC) are extracted from the raw data, de-noised and baseline corrected. Then, a software tool, IsoMS Pro, developed by Nova Medical Testing Inc. (Edmonton, Alberta) is used for the rest of the data analysis, including data quality check, data processing, data cleansing, metabolite identification and statistical analysis. The first step in IsoMS Pro is data quality check, in which mass accuracy check

and retention time check are performed. Mass accuracy check is an indicator of MS performance. To be more specific, a background m/z value is continuously detected in the entire LC-MS run, *i.e.*, the background m/z is checked every second. Measured values of this background peak will be extracted from the whole dataset and if the MS performance is stable throughout the period of data acquisition, the distribution of the detected m/z should be within a narrow range. Similarly, for retention time check, a mixture of retention time calibrants is intermittently injected throughout the LC-MS runs, and the data containing the calibrant peaks are then collected. With stable LC performance, the detected retention time of each calibrant should be within a narrow range.

The second step is data processing. Here, peak-pair picking, peak alignment and zero-filling are carried out. In peak-pair picking, peak pairs resulted from CIL are extracted from the CSV raw files. The filter of the peak pairs is done by removing reductant peaks like adduct ions and dimers, thus retaining only the $[M + H]^+$ pairs. One peak pair detected is corresponding to one individual labeled metabolite. The peak-pair intensity ratios of these are then calculated here. In peak alignment, all the LC-MS runs are aligned together into an integrated matrix of peak pair ratio values (in row) across samples analyzed (in column). And at the end, in zero-filling, missing values in the aligned files are filled from the raw LC-MS data using algorithm.

The third step is data cleansing, which comprises of blank subtraction, missing value treatment and sample-wise normalization. For blank subtraction, it is only after blank filter wherein peak pairs that exist only in method blanks or have a large portion of signal originating from the blank are being excluded. Usually, blank subtraction is not recommended because it can enhance the effects of random variations and also exaggerate the statistical differences between study groups, especially when the blank signal is very strong. For missing value treatment, peak pairs with too many missing values are filtered. As a general rule of thumb, the algorithm used here

requires the peak pairs to be retained have larger than 50% existence among all samples (not including QC) or larger than 80% existence in at least one study group. After this, the remaining missing values are estimated by a ratio-based imputation algorithm, which is specifically designed for the CIL LC-MS data, providing putative values with high accuracy and confidence. Afterward, sample-wise normalization, which is also regarded as post-acquisition normalization is performed. Here, all peak pair ratios in samples are divided by the sample's ratio of total useful signal[36].

The steps mentioned above are important to prepare a complete dataset of the samples being analyzed in a study for the next stage – metabolite identification.

1.2.6.7 Metabolite Identification

IsoMS Pro uses a three-tier metabolite identification approach to identify the metabolites. For Tier 1, it is high-confidence positive identifications matched by retention time and accurate mass to an expandable library of labeled authentic standards. The default retention time window is given for each experimental setting, and the default m/z window is 10 ppm. For Tier 2, it is high-confidence putative identifications based on retention time and m/z matching to a LI library, which consists of various metabolites covering all common metabolism pathways. The default retention time window for the LI library search is given for each experimental setting, and the default m/z window is 10 ppm. And for Tier 3, it is a putative identification based on m/z match against a more comprehensive and complete metabolite database. To be more specific, zero-reaction mode searches is against a library of 8021 known human endogenous metabolites; one-reaction mode searches is against a library involves over 375,000 predicted human metabolites with one biological reaction; and two-reaction mode searches is against a library of over 10,580,000

predicted human metabolites with two biological reactions. The default m/z window for this mass-based database search is 10 ppm.

Information about the CIL LC-MS library is available on the MyCompoundID website: http://www.mycompoundid.org/mycompoundid_IsoMS/index.jsp .

1.2.6.8 Statistical Analysis

There are a lot of statistical tools available for the analysis of data. Broadly speaking, univariate and multivariate statistics are the common methods to extract useful information from the complex metabolomics datasets[38]. Univariate analysis involves only one variable for statistical analysis. Theoretically, in metabolomics, univariate analysis compares the statistic values such as the mean values, between two groups of samples. An example of univariate statistical tool for metabolomics data is volcano plot. Volcano plot is a type of scatterplot that displays significance level (y-axis) versus fold change (x-axis), and each dot in the plot represents a metabolite. It is used for binary comparison, that is to compare data in two populations. For the significance level, both p-value and q-value would be considered. Statistically, p-value can be understood as the probability that the pattern of data in the sample could be produced by randomness, which means there is no real difference between the two tested study groups. In other words, the smaller the p-value, the higher the significance, that is the likelihood that there is a real difference. And in general, a commonly used cut-off of the statistical significance is 0.05. Actually, p-value is calculated from a two samples t-test, which is used in statistic to check if the two means are reliably different from each other or is due to chance. Each t-value has a p-value and t-value would tell if there is a difference and p-value would tell if it is reliable. However, one of the main limitations of p-value, in particular in metabolomics datasets, is the large number of hypothesis tests due to multiple

measured metabolites, and this would result in high possibility of getting false positive results. A way to mitigate it is to consider the q-value, that is the FDR adjusted p-value. IsoMS Pro provides two FDR-based adjustment methods to convert original p-value to q-value, they are the Storey[39] and Benjamin and Hochberg approach[40, 41]. In IsoMS Pro, when the FDR adjustment is enabled and the significance level is set at 0.05, the false discovery rate of the output list of significant metabolites generated would be below 5%. For the fold change, it is the ratio of average values of two groups and the default value is 1.5. This value may need to be adjusted according to the expected extent of metabolic changes in a specific study. The distinct advantage of univariate analysis is it is a simple and robust way which allows the discovery of those metabolites that varies between two populations. Because it visualizes and analyzes each valuable separately, it is a good tool for preliminary screening of significantly altered metabolites. However, the down side of it is it cannot illustrate a statistically significant difference between two groups if there are more than one variable. Moreover, it requires prior knowledge of the measured variable.

Multivariate analysis, on the other hand, involves the study and visualization of multiple variables together. It provides a way to reduce the dimensionality and uncover the coexistent relationship between the variables of metabolomics datasets. It is particularly useful in untargeted metabolomics, in which the number of metabolites is far more than the number of samples. The main advantage of multivariate analysis is it gives a straightforward interpretation of the large-scale complex datasets. Examples of the commonly used multivariate methods in metabolomics studies are the principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA). PCA is an unsupervised method which gives a simple and graphical overview of data without referring to prior knowledge or assumption about if the samples are come from different groups or not. Each group in PCA is bounded by an ellipse, namely the 95% confidence interval

ellipse. This ellipse is the defined region for predicting a new observation in the population. Another point to mention about PCA is principal components (PC), particularly PC1 and PC2, which represents the x and y-axis of PCA respectively. PCA is built based on the principle that the redundancy within the complex datasets can be decomposed in an orthogonal way, into a few linear weighed variables presented as PC[38]. The values of PC are in percentage which indicates the overall variance it can explain. For example, PC1 usually has a largest percentage value, and is the axis that explain the greatest variability among the samples in the analysis. Likewise, PC2 is the second axes explaining more variability and so on. It is worth noting that there are as many PCs as there are characteristics. Leaving out PCs would lead to lose of some information. Though, if the first two or three PCs have capture most of the variance, we can leave the other PCs without losing important information. In contrast, PLS-DA is a supervised method that take the group assignment of the samples into consideration. The key of PLS-DA lies in the PLS. It is used to find the fundamental relations between two matrices. In metabolomics work, matrix 1 is the metabolite information, to be more specific, the peak pairs ratio in CIL LC-MS. While matrix 2 is the categorical variables, for example, disease vs healthy control group. PLS finds the variations in common and then constructs a model which look up the covariance between these two matrices. Covariance is a measure of the common variability of these two matrices, it shows if there is a linear relationship between two variables. A positive covariance is deemed if the increase in one variable agree with the increase of the other variable. Indeed, PLS-DA is used to sharpen the separation between groups and to understand which variables bear the class separating information. However, by virtue of the supervised feature of PLS-DA, which leads to separations between variables including even the variations not related to the categorical variables, overfitting is thus a major challenge. To overcome this problem, cross-validation and permutation test are essential

when performing PLS-DA. In cross-validation, two values, R^2 and Q^2 which indicated the goodness of fitting and goodness of prediction respectively are obtained. In general, a high R^2 and Q^2 indicate the model can better describe the set of explanatory variables and also predict the response ones. That is to say, more desirable is the PLS-DA model. In IsoMS Pro, when building the PLS-DA model, five PCs are set to be checked in default. After cross-validation, operating from one PC to five PC, the number of PCs that gives the highest Q^2 value is used for building the final PLS-DA model. And for permutation test, the default number of tests performed in IsoMS Pro is 100. It is used to examine the chance of over-fitting especially when the number of variables is much larger than the number of observations. The result of the test shows the original Q^2 value and the distribution of the Q^2 values from the permutations. Meanwhile, an empirical p-value is calculated to evaluate whether the original Q^2 value is significantly different from the null distribution.

The above are only some examples of the statistical analysis in metabolomics. Depending on the purpose of studies and the results that researchers want to emphasize, different methods can be chosen and used.

1.2.6.9 Applications

High-performance CIL LC-MS has been emerging as a fresh quantitative metabolome profiling technique in different areas of biological, biomedical and clinical research. This technique has been applied in blood metabolomics[42], urine[43] and fecal[19] metabolomics. It is also being employed in studying the metabolome in human sweat[44], breast milk[45] and even cells[46-48].

Development of CIL LC-MS has now led to the application of standardized four isotope labeling methods to study the metabolome of biological samples comprehensively. In the future, more rapid analysis is needed to enable high-throughput study in metabolomics.

1.3 Single-Cell Metabolomics

1.3.1 Cell-to-Cell Heterogeneity

Single-cell metabolomics is the study of a complete set of cellular metabolites in single cells, in a non-targeted and non-biased manner. It is an increasingly important area in life science. Cell is the fundamental unit of life, which makes up all living organisms on Earth, ranging from unicellular organism such as bacteria to multicellular organism, for instance, human beings. Cells are very dynamic. More importantly, they divide, adapt, communicate and react to their surroundings non-synchronously[49]. The meaning of single-cell analysis stems from better understanding of reasons behind the high variability of cells in the same population, and this is known as the cell-to-cell heterogeneity.

Cellular heterogeneity can be attributed to four main sources in general[50]. First, genetic factor. Acquired genetic change occurs in single cells. For example, De novo mutations (DNMs) in the DNA sequence usually originate during cell division. This can be considered as an evolution by random mutations and resulted in phenotypic change. There is also pre-existing genetic change in single cells, which could also contribute to heterogeneity. When cells undergo rapid divisions, chances of replication errors exist and this can be the potential for prior genetic mutations. Second, epigenetic factor, which arises from nongenetic influences on gene expression. Differential DNA demethylation and differential chromatin modification are examples of epigenetic determinants of heterogeneity. These modifications play important roles in cellular development and

differentiation. Moreover, it can also control the gene expression. However, the differences between cells in these modifications such as the timing of these epigenetic reprogramming could lead to heterogeneity. Third, non-genetic factor, which is also known as stochastic. Heterogeneities related to this factor can be ascribed to the interaction of stochastic fluctuations in molecules within the cell. Temporal noise is an example of it. To be more specific, the protein levels within a cell can actually fluctuate over a short period of time. This change is not permanent in which the daughter cells may not be affected. However, this still can cause variability in the cells. Therefore, stochasticity can be considered as the induced phenotypic heterogeneity through gene and protein expression. Finally, extrinsic sources of heterogeneity. External environment factors may also contribute to cell-to-cell differences. A specific environment condition is necessary and crucial for the proper development of cells, therefore change in the environmental variables, such as temperature, pH and even exposure to toxins can also alter the cells behavior.

1.3.2 Advantage and Importance of Single-Cell Metabolomics

The distinct advantage of single-cell analysis, in particular, when compared to population-level studies is that it gives a more accurate depiction of the cell-to-cell variations. Population-level studies could overlook important differences among individual cells. On account of the fact that cell populations are not homogeneous and thus behave non-synchronously, single cell metabolomics is therefore important for the purpose of getting a more accurate, unbiased metabolic profile of individual cells being studied.

Single cell metabolomics is capable of providing a more immediate and dynamic indicator of the phenotypic diversity of single cells and a picture of how cells react to and interact with environmental influences at a molecular level. It also provides a window into the cell-to-cell

heterogeneity, and allows the measurements of the extent of variation in resulting phenotypes of single cells. This opens up possibilities for better understanding of the biological system. For example, cellular heterogeneity explains the exist of variant subpopulations that could withstand the perturbations such as chemical or environmental stress better and survive[51]. It could provide insights on the mechanisms behind cells coping with stress. Another potential application of single-cell metabolomics that should deserve our attention is oncology – the study of cancer. According to World Health Organization (WHO), cancer is a leading cause of death globally, accounting for nearly 10 million deaths in 2020. It is a group of diseases which can initiate in almost any organ or tissue of the body when abnormal cells grow uncontrollably and has the potential to metastasize, *i.e.*, to invade and spread to other parts of the body. It is worth noting that our body contains 30 trillion cells and only one abnormal cell is enough to cause cancer[49]. Research into understanding and characterizing what and how the cells are different can ultimately shed light on the contribution of cellular heterogeneity in cancer diagnosis and treatment[52]. In cancer diagnosis, studying single-cell heterogeneity can help the selection and use of suitable disease biomarkers. Particularly, in metabolomics work, by detecting, identifying and quantifying the metabolites in two different groups, the cancer cells and normal cells, it is possible to reveal the differences between these two group of cells. And the cancer related metabolites could be used as biomarkers in the diagnosis of specific cancer types. In cancer treatment, understanding single-cell heterogeneity can facilitate the choice of proper drugs which could definitely improve the efficiency of treatment. To elaborate more here, some cells in a tumor can develop drug resistance to a treatment because cells can have heterogeneity in their growth rate and treatment response. The point is treatment aiming only at part of the cells in a tumor will lead to cancer relapse readily. For example, in human epidermal growth factor receptor (HER) family-dependent epithelial

tumors, scientists found that when handling it, it would be more effective to target three receptors of HER family, which are the EGFR, HER2 and HER3 simultaneously owing to single-cell heterogeneity[52].

Altogether, single-cell metabolomics is important as it provides a way to highlight cellular heterogeneity. It holds a great potential to elevate our understanding of the cellular physiology and give biological insights in health and diseases that is not accessible from population-level metabolomics data.

1.3.3 Challenges of Single-Cell Metabolomics

In Section 1.2.2, challenges of metabolomics are mentioned. In single-cell metabolomics, apart from the challenges listed before, there are some distinctive challenges. The first one is the limited sample volume and low abundance of metabolites. Depending on the size of single cells, sample volume can be as small as femtoliters or less. This greatly increases the difficulty of cell collection, preparation and analysis. As mentioned before, metabolites signals cannot be amplified, therefore, the sensitivity of the analytical platform plays a very important role in single cell metabolomics. And because of this, MS-based methods are dominant in the field of single-cell metabolomics.

Another worth noting challenge is it is hard to perform replicate analysis[53], especially in MS-based single-cell metabolomics. MS is a destructive technique, *i.e.*, in order to detect the sample, it has to be ionized and cannot be recovered. Hence, when the sample volume is low, as in single cell, it is very difficult to do replicate analysis. Thus, each single cell analysis cannot be technical repeated.

MS plays a very important role in single-cell metabolomics, and there are many MS-based techniques to study the single-cell metabolome. Apart from the commonly used LC-MS and GC-MS, continuous advancement in developing devices and techniques has been made. For example, the single-probe, which is a miniaturized multifunctional sampling and ionization device directly coupled to the mass spectrometer, enabling live single-cell MS analysis in real time[54]. Laser ablation electrospray ionization mass spectrometry (LAESI-MS), in which the integration of laser ablation from a mid-infrared laser with a secondary electrospray ionization process is used[55, 56]. Nanospray desorption electrospray ionization (nano-DESI) MS, in which the entire cell contents are desorbed into the liquid in a few seconds for subsequent ionization[53]. These are the ambient ionization of single cells and allowing in situ metabolomics. The challenge here is that high vacuum condition is not compatible with living cells analysis and mass spectrometers need to operate at high vacuum in order to minimize the collision of ions with other molecules in the mass analyzer.

To summarize this part, no technique is flawless in all regards. Each single-cell metabolomics method has its own challenges and these are the research goals peoples in the field targeting.

1.3.4 *Xenopus laevis* in Single-Cell Metabolomics

When it comes to single cell study, egg is one of the popular single cells used in research. Embryos of *Xenopus laevis* have been studied to understand embryonic development through single-cell metabolomics[57]. *Xenopus laevis* is an African clawed frog, and is a desirable animal model widely used in cell, molecular and developmental biology research. The major reason for this is because *Xenopus* has conserved important cellular and developmental processes as well as

significant extent of genomic synteny with mammals[58]. This makes it a powerful model to understand human development and disease. One of the ultimate goals of scientific research is to improve the quality of human lives, so with the use of tools which has close evolutionary relationships with humans, this could provide insights into human diseases, which is definitely beneficial to human beings.

In the works presented in this thesis, single cells employed are the *Xenopus laevis* oocytes. The main reason highlighted the use of *Xenopus* oocyte is its large size, with an average diameter of 1.3 mm[59]. In single-cell metabolomics, it is of utmost importance that sample volumes and analyte concentrations are large and high enough for analysis. Sensitivity in smaller cell is always the main limiting factor in analysis so the large size of *Xenopus* oocytes can potentially provide enough, abundant source of material for metabolic analysis.

1.3.5 Workflow for Single-Cell Metabolomics

The general workflow for single-cell metabolomics carried out in the studies presented in this thesis includes the following: single cells isolation from *Xenopus laevis*, cell lysis and metabolite extraction, CIL LC-MS and finally data processing and analysis. Basically, data processing and analysis are the same as metabolomics, as discussed in Section 1.2.6.6 to 1.2.6.8. Thus, only sample collection and metabolite extraction, together with CIL LC-MS will be covered in this section.

1.3.5.1 Sample Collection and Metabolite Extraction

Sample preparation includes the extraction of target cells from the animal and cell lysis to extract the metabolites. It is vital to isolate the target cell without affecting its metabolism and thus the metabolome, in a high-throughput manner.

Single-cell isolation is the first step in sample preparation. Owing to the fact that cellular metabolome is very sensitive to external perturbations such as temperature changes, enzymes treatments and mechanical damages, it is very crucial to minimize the intracellular biochemical activities along with analyte loss in this step. Low temperature is a common approach to reduce metabolic activity and is used in the related studies presented in this thesis. Moreover, the integrity of cell membrane is preserved during cells isolation to reduce metabolite leakage. In the works presented in Chapter 2 to 4 of this thesis, single cells were isolated from the ovary of *Xenopus laevis*. Amphibian skin acts as a semipermeable membrane that allows for respiration and absorption of substance through the skin. Therefore, the frog was first anesthetized by immersion in an anesthetic solution, a mixture of ethyl 3-aminobenzoate methanesulfonate and sodium bicarbonate, for about 1 hour. Ethyl 3-aminobenzoate methanesulfonate is a chemical which can suppress the sensory system of the frog, and sodium bicarbonate is used to adjust the pH of the solution so as to protect the skin of the frog. It should be noted that depending on the weight and the age of frogs, the time needed for anesthesia varies. The next step is dissection to collect the frog oocytes. The spinal cord of the frog was broken using a knife and the abdomen was opened up to extract the whole ovary tissue out of the frog. The ovary is a large piece tissue in the abdomen and it was being torn into smaller pieces in a petri dish. After that, collagenase was added to digest the follicle cell layer that encloses the oocytes. At the same time, blood capillaries surrounded the cells were also removed. The time for collagenase digestion depends on the age of the frog, as older frog would have thicker connective tissue and thus more time would be needed to remove all the follicle cell layer. Sandpaper rolling was then used to make sure all the follicle cells and blood capillaries were removed before oocyte selection. In oocyte selection, stage V oocytes were picked under the microscope. Stage V oocytes appear as distinct hemispheres and the pigment half

fades to brown. Finally, the oocytes were being washed with growth medium OR-2 twice before cell lysis.

After isolation, metabolites need to be extracted from cells for metabolic studies. A method of high efficiency to extract the analytes of interest is an important factor to increase the measurement accuracy. In the meantime, the enzymatic activity of the cells has to be quenched to stop further changes of the metabolome during metabolite extraction. In the context of these, cell lysis, when carried out carefully, can achieve highly efficient extraction and as a result, contribute to accurate, precise and reliable measurements of metabolites in single cells. Cell lysis is a process to break or destroy the cell membrane of the cells in order to release the inter-cellular biomolecules from the cells. In general, cell lysis methods can be divided into physical disruption and chemical cell lysis[60]. Examples of physical lysis includes freeze-thaw cycles, osmotic shock, grinding and sonication. The advantages of physical method are that it can be used with a wide range of materials and the amount of force can be easily adjusted according to a particular type of cells. However, the major shortcoming is that since the cells are being disrupt at different times, the released molecules may be subject to disruptive forces repeatedly. Moreover, it might not be compatible when working with small volumes and high-throughput. For chemical-based methods, such as enzymatic digestion, chemical agent is used to make cellular membranes permeable to analytes. It is a gentle and efficient way to extract the analytes, but high concentrations of chemical could interfere downstream analysis, *e.g.*, the ionization of analyte for MS detection. For this reason, the condition in chemical lysis has to be carefully optimized and controlled. For cell lysis in single-cell metabolomics works presented in this thesis, we employed a well-established method for metabolite extraction from Marc Kirschner's group in the Department of Systems Biology at Harvard Medical School[61]. The lysis solution for metabolite extraction is ACN:MeOH:H_2O in

ratio of 2:2:1. Each cell was deposited onto the side of an Eppendorf tube, ready on an ice bucket. Then each of them was lysed by vortexing on a high speed vortexer for 5 seconds, in a 4 ℃ cold room. It is important to lyse the cell and extract the metabolites at low temperature as this can minimize the hydrolysis and degradation of metabolites.

1.3.5.2 CIL LC-MS

In CIL of single-cell samples, the well-established methods mentioned in Section 1.2.6.3 were applied. The only thing is the starting sample volume is much less compared to other biological samples, such as urine and serum. The labeling protocol has to be scaled down and there is an extra need to exercise great care in performing CIL of single-cell samples, since the labeling reagent added can be as low as 2 μL.

Another point is the reference sample in single-cell metabolomics is not the pooled sample, as in urine and serum metabolomics. The reason for this is single cell sample has very limited volume and is not enough to generate the pooled sample. Hence, universal internal standard (UIS) was used. UIS was made by collecting and lysing a large number of oocytes from the same type of *Xenopus laevis*. The idea is that as the metabolites in the UIS are the same as those in individual single cell samples, the performance of it is similar to the pooled sample. Therefore, for CIL in single-cell metabolomics, UIS is labeled with a ^{13}C-lableing reagent, while individual single cell sample is labeled with ^{12}C-lableing reagent.

1.4 Scope of the Thesis

The research objective is to develop and apply CIL LC-MS techniques for single-cell metabolomics.

In chapter 2, CIL LC-MS is used to study the amine and phenol submetabolome of individual *Xenopus laevis* oocytes.

In chapter 3, CIL LC-MS is used to study the metabolic responses of single cells to heat stress. Short-term and long-term responses of cells, along with recovery from heat stress are investigated.

In chapter 4, CIL LC-MS is used to study the cell-to-cell variations from different locations of the *Xenopus laevis* ovary. Four labeling chemistries are applied to study the complete set of metabolites in single cells.

Finally, in chapter 5, construction of two MS/MS retention time libraries are presented.

1.5 Literature Cited

1. Horgan, R. P.; Kenny, L. C. *Obstet. Gynecol.* **2011,** *13* (3), 189-195.

2. Vailati-Riboni, M.; Palombo, V.; Loor, J. J., What Are Omics Sciences? In *Periparturient Diseases of Dairy Cows: A Systems Biology Approach*, Ametaj, B. N., Ed. Springer International Publishing: Cham, 2017; pp 1-7.

3. Bumgarner, R. *Curr Protoc Mol Biol* **2013,** *Chapter 22*, Unit-22.1.

4. McGettigan, P. A. *Curr Opin Chem Biol* **2013,** *17* (1), 4-11.

5. Zhao, S. *Drug Discov. Today* **2019,** *24* (6), 1258-1267.

6. Ward, A. J.; Cooper, T. A. *J. Pathol.* **2010,** *220* (2), 152-163.

7. Aslam, B.; Basit, M.; Nisar, M. A.; Khurshid, M.; Rasool, M. H. *J. Chromatogr. Sci* **2017,** *55* (2), 182-196.

8. Wang, Q.; Gao, P.; Wang, X.; Duan, Y. *Sci. Rep.* **2014,** *4* (1), 6802.

9. Ye, G.; Zhu, B.; Yao, Z.; Yin, P.; Lu, X.; Kong, H.; Fan, F.; Jiao, B.; Xu, G. *J. Proteome Res.* **2012,** *11* (8), 4361-4372.

10. Yao, H.; Yu, P.-C.; Jiang, C.-M. *RSC Adv.* **2020,** *10* (9), 4928-4941.

11. Zenobi, R. *Science* **2013,** *342* (6163), 1243259.

12. Emwas, A. H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. *Metabolites* **2019,** *9* (7).

13. Markley, J. L.; Brüschweiler, R.; Edison, A. S.; Eghbalnia, H. R.; Powers, R.; Raftery, D.; Wishart, D. S. *Curr. Opin. Biotechnol.* **2017,** *43*, 34-40.

14. Chandra, K.; Al-Harthi, S.; Sukumaran, S.; Almulhim, F.; Emwas, A.-H.; Atreya, H. S.; Jaremko, Ł.; Jaremko, M. *RSC Adv.* **2021,** *11* (15), 8694-8700.

15. Nagana Gowda, G. A.; Raftery, D. *J. Magn. Reson.* **2015,** *260*, 144-60.

16. Liao, Z.; Zhang, S.; Liu, W.; Zou, B.; Lin, L.; Chen, M.; Liu, D.; Wang, M.; Li, L.; Cai, Y.; Liao, Q.; Xie, Z. *J. Chromatogr. B* **2019,** *1133*, 121848.

17. He, Z.; Wang, M.; Li, H.; Wen, C. *Sci. Rep.* **2019,** *9* (1), 3872.

18. Ibáñez, C.; Simó, C.; Palazoglu, M.; Cifuentes, A. *Anal. Chim. Acta* **2017,** *986*, 48-56.

19. Xu, W.; Chen, D.; Wang, N.; Zhang, T.; Zhou, R.; Huan, T.; Lu, Y.; Su, X.; Xie, Q.; Li, L.; Li, L. *Anal. Chem.* **2015,** *87* (2), 829-836.

20. Guo, K.; Li, L. *Anal. Chem.* **2009,** *81* (10), 3919-3932.

21. Zhou, R.; Li, L. *Methods Mol. Biol.* **2014,** *1198*, 127-36.

22. Wu, Y.; Li, L. *Anal. Chem.* **2012,** *84* (24), 10723-10731.

23. Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. *Anal. Chem.* **2014,** *86* (10), 4675-4679.

24. Yamashita, M.; Fenn, J. B. *J. Phys. Chem. A* **1984,** *88* (20), 4451-4459.

25. Kebarle, P.; Tang, L. *Anal. Chem.* **1993,** *65* (22), 972A-986A.

26. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968,** *49* (5), 2240-2249.

27. Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976,** *64* (6), 2287-2294.

28. Wiley, W. C.; McLaren, I. H. *Review of Scientific Instruments* **1955,** *26* (12), 1150-1157.

29. Cornish, T.; Bryden, W. *Johns Hopkins Apl Technical Digest* **1999,** *20*, 335-342.

30. Iyer, S. S.; Zhang, Z. P.; Kellogg, G. E.; Karnes, H. T. *J Chromatogr Sci* **2004,** *42* (7), 383-7.

31. Wade, D. *Chem Biol Interact* **1999,** *117* (3), 191-217.

32. Zhao, S.; Luo, X.; Li, L. *Anal. Chem.* **2016,** *88* (21), 10617-10623.

33. Guo, K.; Li, L. *Anal. Chem.* **2010,** *82* (21), 8789-8793.

34. Zhao, S.; Dawe, M.; Guo, K.; Li, L. *Anal. Chem.* **2017,** *89* (12), 6758-6765.

35. Warrack, B. M.; Hnatyshyn, S.; Ott, K. H.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. *J. Chromatogr. B* **2009,** *877* (5-6), 547-52.

36. Wu, Y.; Li, L. *J Chromatogr A* **2016,** *1430*, 80-95.

37. Zhou, R.; Li, L. *J. Proteomics* **2015,** *118*, 130-9.

38. Lamichhane, S.; Sen, P.; Dickens, A. M.; Hyötyläinen, T.; Orešič, M., Chapter Fourteen - An Overview of Metabolomics Data Analysis: Current Tools and Future Perspectives. In *Comprehensive Analytical Chemistry*, Jaumot, J.; Bedia, C.; Tauler, R., Eds. Elsevier: 2018; Vol. 82, pp 387-413.

39. Storey, J. D. *Ann Stat* **2003,** *31* (6), 2013-2035.

40. Benjamini, Y.; Hochberg, Y. *J R Stat Soc Series B Stat Methodol* **1995,** *57* (1), 289-300.

41. Benjamini, Y. *J R Stat Soc Series B Stat Methodol* **2010,** *72* (4), 405-416.

42. Han, W.; Li, L. *Methods Mol. Biol.* **2018,** *1730*, 213-225.

43. Peng, J.; Chen, Y.-T.; Chen, C.-L.; Li, L. *Anal. Chem.* **2014,** *86* (13), 6540-6547.

44. Hooton, K.; Han, W.; Li, L. *Anal. Chem.* **2016,** *88* (14), 7378-7386.

45. Mung, D.; Li, L. *Anal. Chem.* **2017,** *89* (8), 4435-4443.

46. Wu, Y.; Li, L. *Anal. Chem.* **2013,** *85* (12), 5755-63.

47. Luo, X.; Gu, X.; Li, L. *Anal. Chim. Acta* **2018,** *1037*, 97-106.

48. Luo, X.; Li, L. *Anal. Chem.* **2017,** *89* (21), 11664-11671.

49. Emara, S.; Amer, S.; Ali, A.; Abouleila, Y.; Oga, A.; Masujima, T. *Adv. Exp. Med. Biol.* **2017,** *965*, 323-343.

50. Nguyen, D. H.; Jaszczak, R. G.; Laird, D. J. *Curr. Top. Dev. Biol.* **2019,** *135*, 155-201.

51. Bishop, A. L.; Rab, F. A.; Sumner, E. R.; Avery, S. V. *Mol. Microbiol.* **2007,** *63* (2), 507-20.

52. Qian, M.; Wang, D. C.; Chen, H.; Cheng, Y. *Semin. Cell Dev. Biol.* **2017,** *64*, 143-149.

53. Duncan, K. D.; Fyrestam, J.; Lanekoff, I. *Analyst* **2019,** *144* (3), 782-793.

54. Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z. *Anal. Chem.* **2014,** *86* (19), 9376-80.

55. Shrestha, B.; Sripadi, P.; Reschke, B. R.; Henderson, H. D.; Powell, M. J.; Moody, S. A.; Vertes, A. *PLoS One* **2014,** *9* (12), e115173.

56. Taylor, M. J.; Lukowski, J. K.; Anderton, C. R. *J. Am. Soc. Mass Spectrom.* **2021,** *32* (4), 872-894.

57. Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (21), 6545-6550.

58. Harland, R. M.; Grainger, R. M. *Trends Genet.* **2011,** *27* (12), 507-15.

59. Zeng, S. L.; Sudlow, L. C.; Berezin, M. Y. *Expert Opin Drug Discov* **2020,** *15* (1), 39-52.

60. Raddassi, K.; T, W.; Sf, N.; Av, W. *Drug Designing: Open Access* **2017,** *06*.

61. Vastag, L.; Jorgensen, P.; Peshkin, L.; Wei, R.; Rabinowitz, J. D.; Kirschner, M. W. *PLoS One* **2011,** *6* (2), e16881.

Chapter 2 – High-Coverage Quantitative Single-Cell Metabolomics: Chemical Isotope Labeling LC-MS Metabolome Analysis of Individual *Xenopus laevis* Oocytes

2.1 Introduction

Cell is the fundamental structural and functional unit of life, which makes up all living organisms on Earth, ranging from unicellular organism such as bacteria to multicellular organism, for instance, human beings. Cells are highly dynamic themselves. More importantly, they divide, adapt, communicate and react to their surrounding non-synchronously. The meaning of single-cell analysis stems from better understanding of the reasons for the variability between genetically identical cells. To be more specific, individual cells can differ widely in responsiveness to uniform physiological stimuli, even within a homogeneous population, and this is known as the cell-to-cell heterogeneity[1, 2]. The distinct advantage of single-cell analysis, in particular, when compared to population-based study is that it gives a more accurate depiction of cell-to-cell variations. Population-based study involves bulk measurements of cells and thus result in stochastic average of the cell population in the study. For example, RNA-seq experiment provides information about the average gene expression in a cell population. In other words, the mRNA level is estimated across all cells in each subpopulation. The problem of this is that it could miss important biology of individual cells, since an individual cell could have mRNA expression patterns that vary a lot from the average expression of the cell population[3]. Therefore, single-cell analysis is crucial to illustrate cellular diversity and heterogeneity.

In single-cell metabolomics, the complete set of single-cell metabolome is studied in a non-targeted and non-biased manner. It provides a more immediate and dynamic indicator of the phenotypic diversity of single cells. By detecting, identifying and quantifying the metabolome in a single-cell, it provides a window into the cell-to-cell heterogeneity and can open up possibilities

for better understanding of life science. For example, cellular heterogeneity explains the exist of variant subpopulations that could withstand the perturbation such as chemical or environmental stress better and survive[4]. It can also shed light on the contribution of phenotypic heterogeneity in cancer genesis, progression and therapies[4, 5].

Individual oocytes of the African clawed frog *Xenopus laevis* were used in this study. *Xenopus laevis* has been widely used as a model organism in cell biology. This animal is used because it has a good combination of experimental tractability and close evolutionary relationship with humans. One of the reasons for its popularity is the large size of the cell, it has an average diameter of 1.3 mm[6]. Therefore, each single oocyte can potentially provide enough material for metabolite analysis with high coverage. The groups of Marc Kirschner and Joshua Rabinowitz have developed a metabolite extraction procedure for *Xenopus* eggs that is expected to be suitable for oocytes[7].

Recent advances in high-performance chemical isotope labeling (CIL) liquid chromatography-mass spectrometry (LC-MS) have opened the possibility of carrying out high-coverage quantitative metabolomics research at the single-cell level. CIL LC-MS uses a strategy of divide-and-conquer to perform in-depth analyses of several chemical-group-based submetabolomes and then combine the data sets to represent the near-complete set of the metabolome in a biological system. Using a rationally designed labeling chemistry, metabolite detectability is significantly improved, allowing the analysis of thousands of metabolites from a submetabolome. For example, dansylation labeling LC-MS can be used to analyze amines and phenols containing metabolites with high coverage[8]. In addition, using differential isotope labeling of individual samples and a reference sample (*e.g.*, a pooled sample), accurate quantification of concentration differences of individual metabolites in comparative samples can

be readily performed[9]. The objectives of this study are to develop a robust workflow for optimized detection of cellular metabolites and apply the workflow to study cells with different culture periods.

2.2 Experimental

2.2.1 Chemical and Reagents

In single-cell collection, tricaine methanesulfonate (MS-222) and Collagenase Type IA were purchased from Sigma-Aldrich Canada. OR-2 was prepared in the lab using 5 mM Hepes-NaOH (pH 7.8), 82.5 mM NaCl, 2.5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM Na_2HPO_4[10]. Lysis solution was prepared from ACN, H_2O and MeOH, all purchased from Sigma-Aldrich Canada.

All the chemicals and reagents used in CIL LC-MS, unless otherwise stated, were purchased from Sigma-Aldrich Canada. For dansylation labeling, the ^{12}C-labeling reagent, *i.e.*, dansyl chloride, was purchased from Sigma-Aldrich and the ^{13}C-labeling reagent was synthesized according to the previously published method[8].

2.2.2 Sample Collection and Processing

All the single-cell samples were collected and prepared in Dr. Michael C. Schultz lab (Department of Biochemistry, University of Alberta) and the animal experiments were done according to protocols approved by the Biosciences Animal Care and Use Committee of the University of Alberta.

The general workflow of single-cell metabolomics performed in this study is depicted in Figure 2.1. In detail, oocytes, after being collected from *Xenopus laevis*, were defolliculated and washed. Stage V oocytes were selected under the microscope. They were then being lysed to extract the metabolites either immediately or cultured for different periods of time (2, 4, 8, 16 and 24 hours).

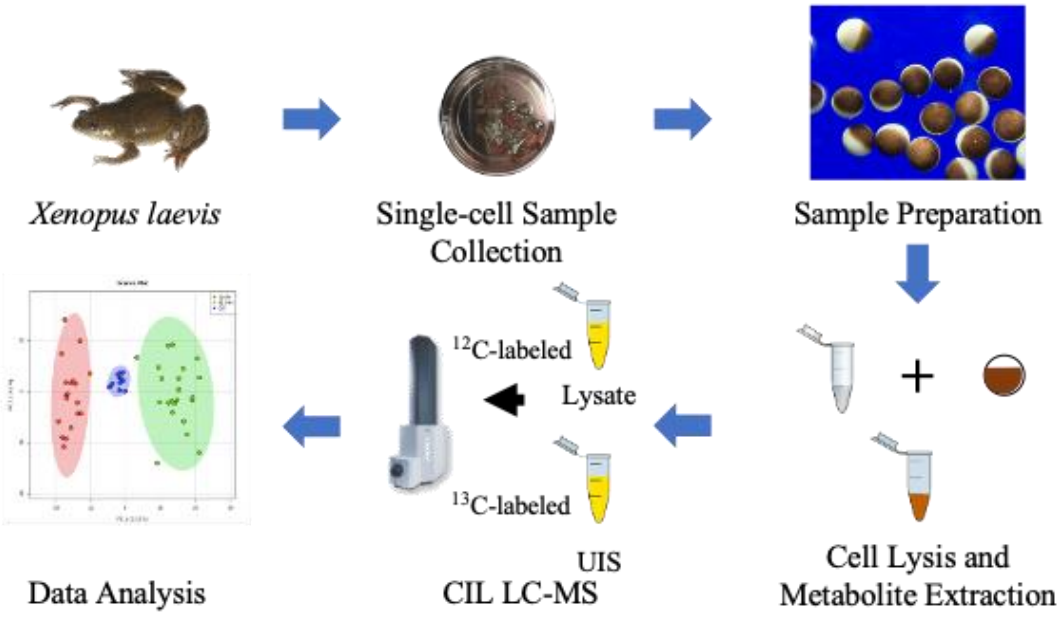


Figure 2. 1 General workflow of CIL LC-MS for single-cell metabolomics.

For cell lysis, 80 μL of lysis solution (ACN: MeOH: H_2O in ratio of 2:2:1) was dispensed into the bottom of 0.6 mL centrifuge tube ready in an ice bucket. This is to prevent evaporation of lysis solution. Each cell was deposited onto the side of the tube at the 0.3 mL mark, this can be done by holding the tube with its long axis about 45 degrees to the horizontal. Single-cell at this stage should be refrained from lysis solution, otherwise, proteins will precipitate immediately. The cells were then carried into the 4 ℃ cold room for cell lysis as low temperature can help minimize

hydrolysis and degradation of the metabolites. The cell was lysed by putting the tube on a high speed vortexer for 5 s, and then incubated on ice for 10 min. After that, the cell was vortexed at full speed for 2 s using a standard laboratory vortexer and incubated on ice for another 10 min. Lastly, the cell was centrifuged under 4℃ at RCF 20,000 for 5 min and the supernatant (about 80 μL) was transferred into a 0.6 mL eppendorf. The lysate was frozen in liquid nitrogen immediately and then stored in -80℃ freezer before CIL.

In differential CIL, cell samples were dried down, resuspended in Na_2CO_3/$NaHCO_3$ buffer and labeled with ^{12}C-dansyl chloride. Universal internal standard (UIS), which is considered as the pooled sample in this study was labeled with ^{13}C-dansyl chloride. ^{12}C- and ^{13}C-labeled samples were mixed in 1:1 ratio prior to injection to LC-MS for analysis. Finally, data analysis was performed to study the metabolome of *Xenopus* oocytes.

2.2.3 Method Optimization

In order to apply CIL LC-MS in single-cell analysis, the dansylation workflow was optimized to profile the amine and phenol submetabolome in single cells. Three labeling protocols denoted as M1, M2 and M3 shown in Figure 2.2 were compared in order to develop an optimized workflow of labeling the small amounts of metabolites present in a single-cell. In M1, 425 mM FA was prepared in 50% ACN. 12 μL of 1:1 mix ^{12}C-labeled single-cell and ^{13}C-labled UIS was injected to LC-MS for analysis. In M2, 425 mM FA was prepared in 100% H_2O, whereas in M3, sample after mixing was diluted with mobile phase A (0.1% (v/v) formic acid in 5% (v/v) acetonitrile).

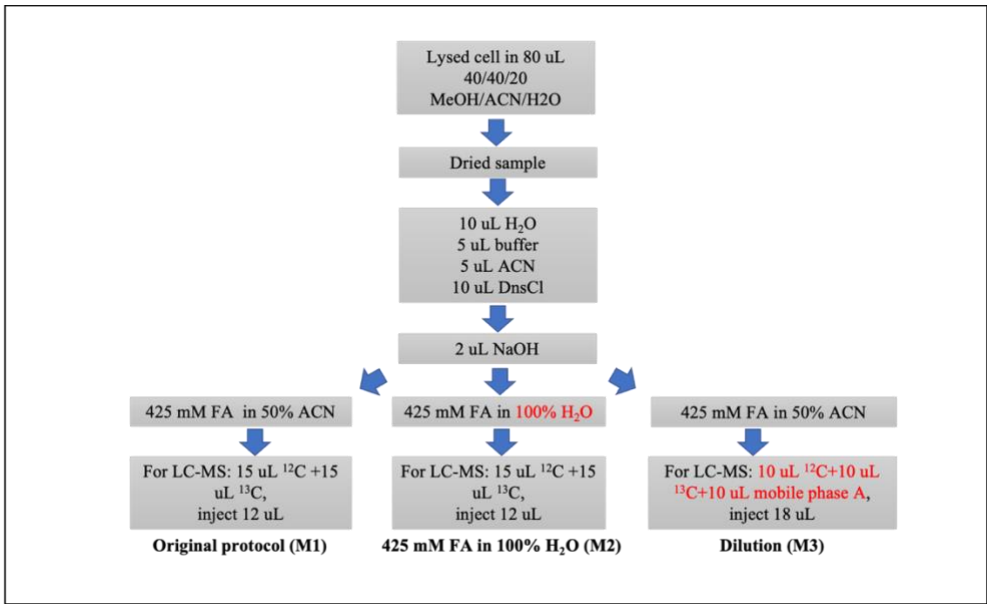


Figure 2. 2 M1, M2 and M3 workflows of labeling small amounts of metabolites in the single-cell.

Figure 2.3A compares the base peak chromatograms (BPCs) of M1 (shown in blue) and M2 (shown in red). In comparison to M1, M2 resulted in less peak broadening and splitting for the early peaks owning to the lower organic percentage. To be more specific, for the quenched DnsCl peak, there were less tailing in M2 and thus less ion suppression of the co-eluting peaks. In addition, M2 gives better peak shape, higher peak intensity and the peaks have less splitting due to reduce peak broadening. In the mid-region, the peak distribution of M1 and M2 is similar. However, for M2, the later peaks have lower intensity than M1. The possible reason here is that for these compounds which eluted out at higher organic phase, the solubility is limited when the aqueous percentage is increased.

Figure 2.3B shows the BPCs for M1 (blue) and M3 (orange). It can be seen that M3 and M2 are similar, with less tailing of the quenched DnsCl peak, higher peak intensity for front peaks, and lower peak intensity for later peaks compared to M1.

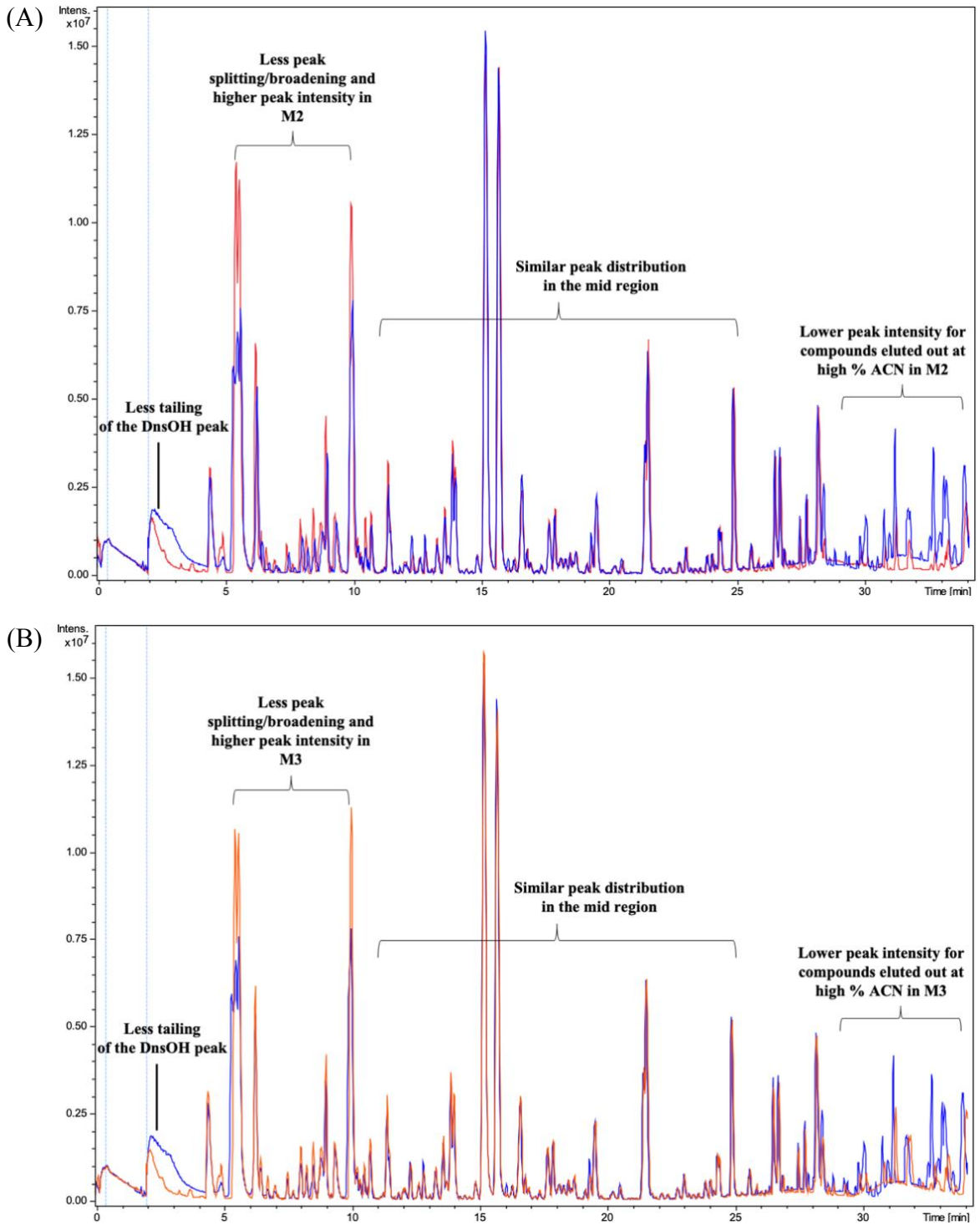


Figure 2. 3 BPCs **(A)** of M1 and M2 and **(B)** BPCs of M1 and M3.

Figure 2.4 shows a heatmap comparing the absolute peak intensity of the three methods, each with three replicates. Red color represents higher peak intensity, and the columns are aligned according to increasing retention time. The data here again suggested that for M2 and M3, the early peaks have higher intensity, while for M1, later peaks have greater intensity.

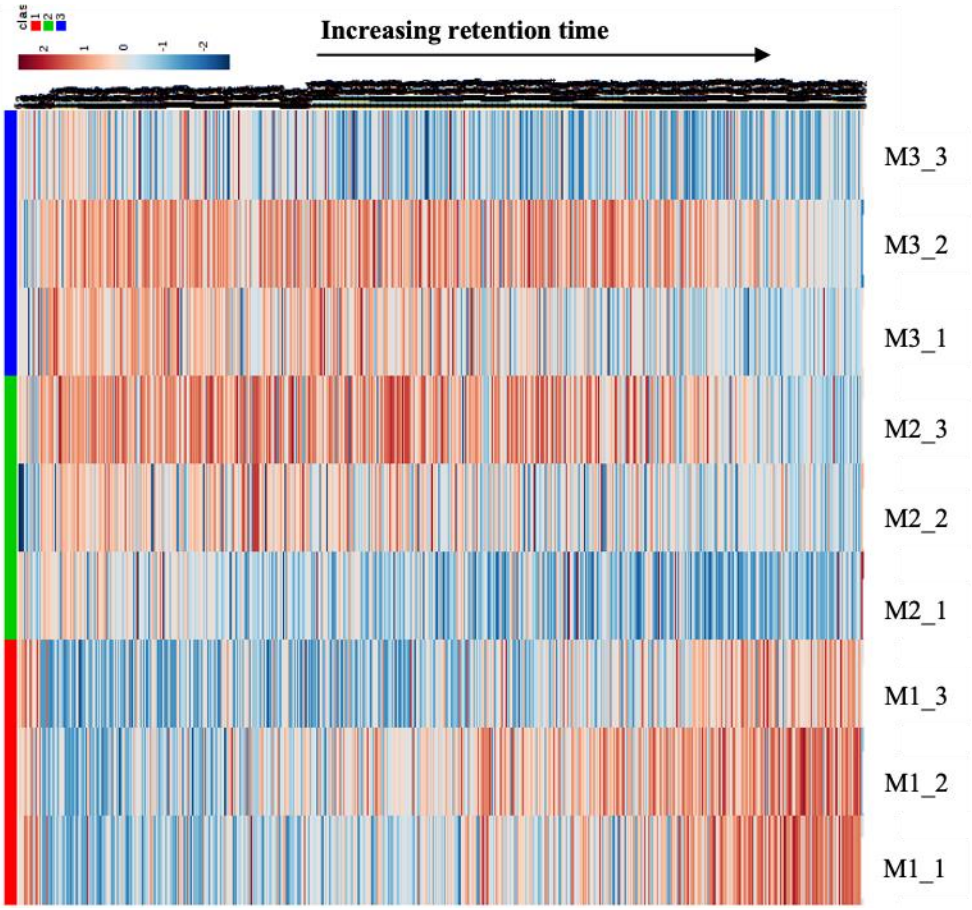


Figure 2. 4 Heatmap comparing the absolute peak intensity of M1, M2 and M3 (n=3).

M2 and M3 provide a distinct advantage in addressing the ion suppression effects when compared to M1 as illustrated in Figure 2.5. From the extracted ion chromatogram (EIC) of a selected peak pair in the three methods, it can be observed that the intensity of the selected peak pair in M1 is significantly lower than that in M3 because of the serious tailing of the quenched

Dns peak. This is because the Dns peak tails to about 5 minutes in M1 and most of the peaks before this time were suppressed by the tailed peak.

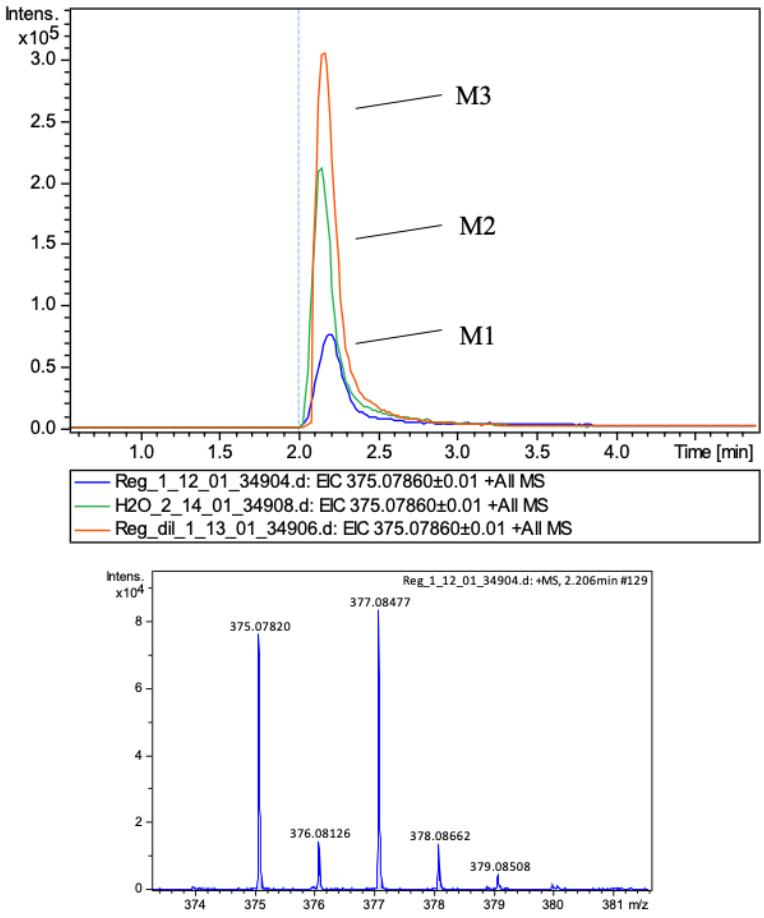


Figure 2. 5 EIC of a selected peak in M1, M2 and M3.

For the number of detected metabolites, M1, M2 and M3 are very close, with an average of 2700 peak pairs being detected after zero-filing as shown in Figure 2.6A. Moreover, 94% of the peak pairs can be commonly detected in these three methods, which means the peak pair distribution of the three methods is comparable as illustrated in Figure 2.6B.

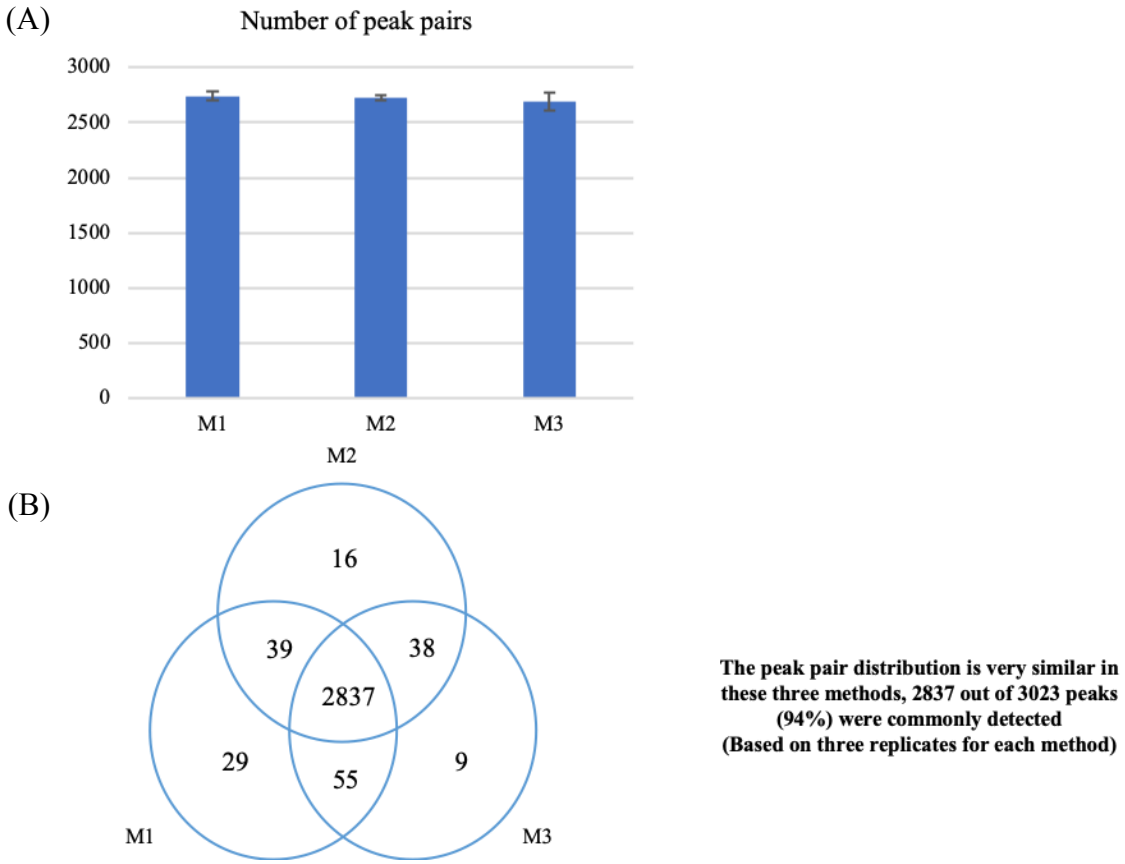


Figure 2. 6 (A) Number of detected metabolites in M1, M2 and M3 and their **(B)** peak pair distribution.

2.2.4 Chemical Isotope Labeling for Amine and Phenol Metabolomics

To analyze the amine and phenol submetabolome of a single-cell, a CIL LC-MS method reported previously was applied[8]. However, because of the limited sample volume of single-cell, the labeling protocol was being scaled down accordingly here. 80 μL of a single-cell sample obtained after cell lysis was first dried down using SpeedVac. The dried sample was then re-suspended in 10 μL of water and mixed with 5 μL of 250 mM sodium bicarbonate buffer

followed by 5 μL acetonitrile and 10 μL of freshly prepared 18 mg/mL ^{12}C-dansyl chloride solution. After vortexing and spinning down, the reaction mixture was incubated at 40 ℃ for 1 hr. 2 μL of 250mM NaOH solution was then added to quench the reaction by reacting with excess dansyl chloride. The mixture was vortexed, spun down and incubated at 40 ℃ for another 10 min. After incubation, 10 μL of 425mM formic acid solution was added to the reaction mixture to acidify the sample followed by dilution with aqueous mobile phase. At the end, the sample solution was vortexed and centrifuged at 10,000 rpm under 4 ℃ for 5 min. When analyzing biofluids, for example urine, a pooled sample is usually generated from all the individual samples being analyzed in the study and is labeled with ^{13}C-dansyl chloride. However, in single-cell analysis, because of the limited sample volume of each individual cell, it would be hard to generate a pooled sample. Therefore, universal internal standard (UIS) was used here. To be more specific, a total of 780 cells were collected from two *Xenopus laevis* and were then being lysed and combined together to generate the UIS. This sample was divided into 0.5 mL aliquots in cold room and were stored in -80 ℃ freezer before use. To label the UIS, 80 μL of UIS was transferred to a 0.6 mL eppendorf, which is the starting volume of the individual single-cell sample. Then, the same procedures used to label individual cell described above were followed, with ^{13}C-dansyl chloride being used instead of ^{12}C-dansyl chloride.

2.2.5 LC-MS

Analysis of single cells was performed on Bruker Impact HD Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA) coupled to Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). Sample was injected to Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Mobile phase for

the gradient elution was solvent A: 0.1% (v/v) formic acid in 5% (v/v) acetonitrile and solvent B: 0.1% (v/v) formic acid in acetonitrile. The gradient elution profile was (i) 0-3.5 min, A+B (20-35% gradient); (ii) 3.5-18 min, A+B (35-65% gradient); (iii) 18-21 min, A+B (65-95% gradient); (iv) 21-37min, A+B (95% isocratic) with flow rate 180 μL/min and injection volume of 12 μL. Mass spectra were recorded from 220 to 1000 m/z with a spectra rate of 1.0 Hz for positive ionization mode. The nebulizer was set to 1.0 bar, and the dry temperature was 230 ℃, with 8 L/min of drying gas. The capillary and end plate offset voltage were set as 4500 V and 500 V respectively.

2.2.6 Data Processing and Statistical Analysis

The raw LC-MS data was exported in Compass DataAnalysis to CSV file. Data in CSV format was ready to process in the following way using IsoMS Pro software developed by Nova MEDICAL TESTING Inc. (Edmonton Alberta): (i) data quality check including mass accuracy and retention time check. (ii) data processing in which peak-pair picking from raw files, peak-pair filtering, peak alignment, peak-pair intensity ratio calculation and zero-filling are performed. (iii) data cleansing which is comprised of missing value treatment and sample-wise normalization. (iv) metabolite identification with a three-tiered metabolite identification system (v) statistical analysis. For univariate analysis, volcano plot was generated using Origin 9.0. MetaboAnalyst website (www.metaboanalyst.ca) was used for multivariate statistical analysis, such as principal component analysis (PCA).

2.2.7 Metabolite Identification

We profiled the amine and phenol submetabolome of 124 single cells scattered in 6 different periods of culture times. LC-MS analysis of 124 samples resulted in the detection of 1531 peak pairs. We positively identified 80 metabolites by retention time and m/z search against CIL standard libraries, putatively identified 146 metabolites with high confidence by retention time and m/z match against the LI library and putatively identified 1173 metabolites by m/z match against the mass-based database.

2.3 Results and Discussion

In this experiment, two animals were used. Animal 1 oocytes were used for optimizing the CIL LC-MS method in single-cell metabolomics. Animal 2 oocytes were cultured for different periods of times, and the amine and phenol submetabolome of cells were studied.

We analyzed metabolite abundance in isolated oocytes at stage V of their development. 24 animal 1 oocytes were individually extracted immediately after isolation. 124 animal 2 oocytes were divided into groups and processed immediately, or after culture for 2 to 24 hours in medium without growth-promoting blood components. The amine and phenol submetabolome were profiled because stage V oocytes rely on protein catabolism to generate amino acids for energy production[11]. Since healthy oocytes were used and there are no cell subtypes in the oocyte lineage, low cell-to-cell variability of the metabolome was expected. However, for freshly extracted cells, *i.e.* t = 0, the content variability of amino acids between cells is substantially higher than the variability of signal between matched quality control replicates as shown in Figure 2.7. An oocyte can rank near the top of the content range for one amino acid and near the bottom for another, which is clearly shown with the arrow in Figure 2.8. Therefore, single cell variability of the

metabolome is not related to cell size. The existence of oocytes that are metabolic outliers was further evident in single cell analysis of the metabolome during primary culture in animal 2. Cell-to-cell spread of abundance of some metabolites is the same at the start and end of the culture, for example, uridine and pantothenic acid shown in Figure 2.9A and 2.9B respectively. Figure 2.10 shows the PCA score plot of single cells with different culture times. High dispersion of whole metabolome PCA scores compared to the QC replicates also persists during culture. Thus, single cell variability of the amine and phenol submetabolome is not erased by in vitro maintenance of oocytes. Remarkably the persistence of single cell variability does not reflect stasis of overall metabolism; the amine and phenol submetabolome is globally remodeled during culture. Specifically, oocytes harvested at 0 and 2 hr cluster together and away from those harvested at 8, 16 and 24 hr; and 4 hr oocytes group lie between the 0/2 and 8/16/24 hr batches. We conclude that the oocyte amine and phenol submetabolome evolves during culture to reach a state of modest stability at 8 hours.

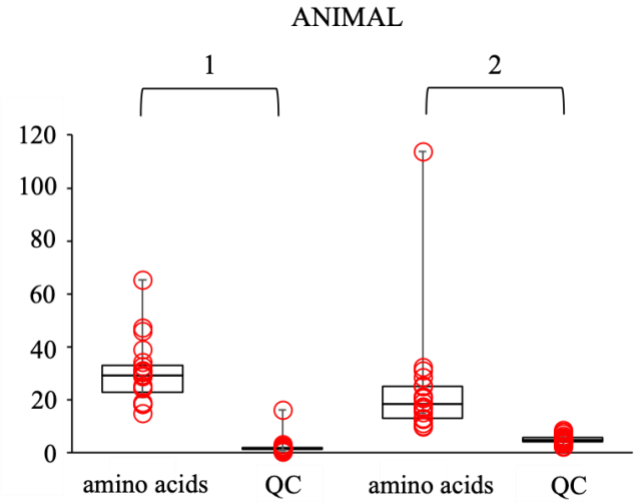


Figure 2. 7 Content variability of amino acids between cells and variability of signal between matched quality control replicates in animal 1 and 2.

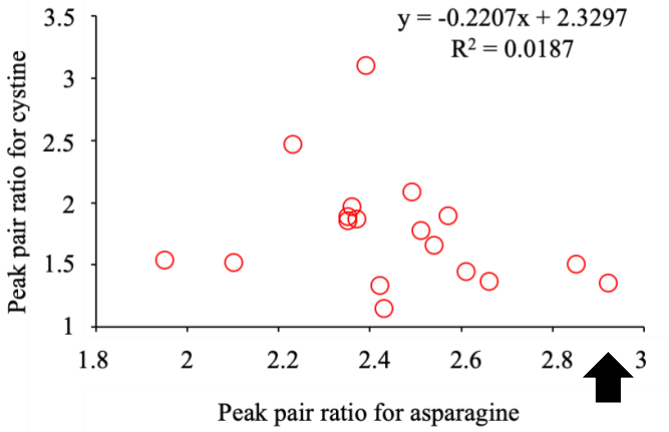


Figure 2. 8 Correlation of cystine and asparagine in cellular peak pair ratio.

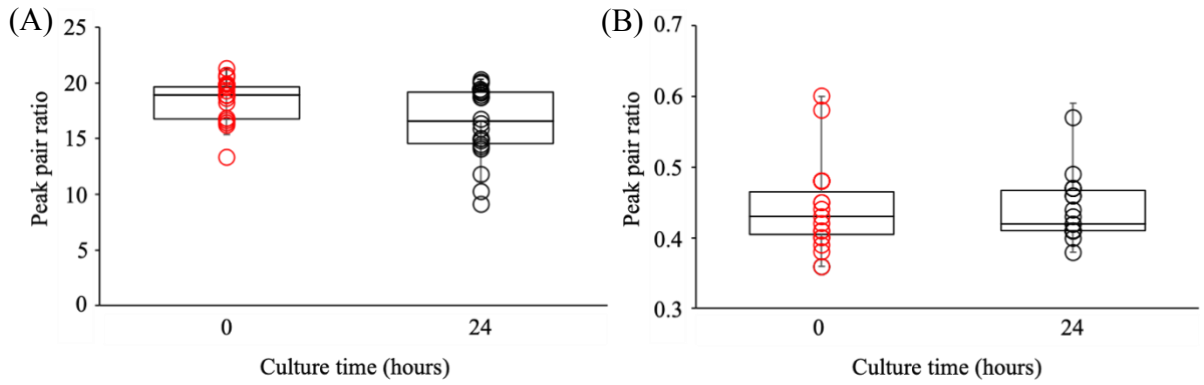


Figure 2. 9 Examples of metabolites with comparable cell-to-cell variation at the beginning and the end of the culture **(A)** uridine, **(B)** pantothenic acid.

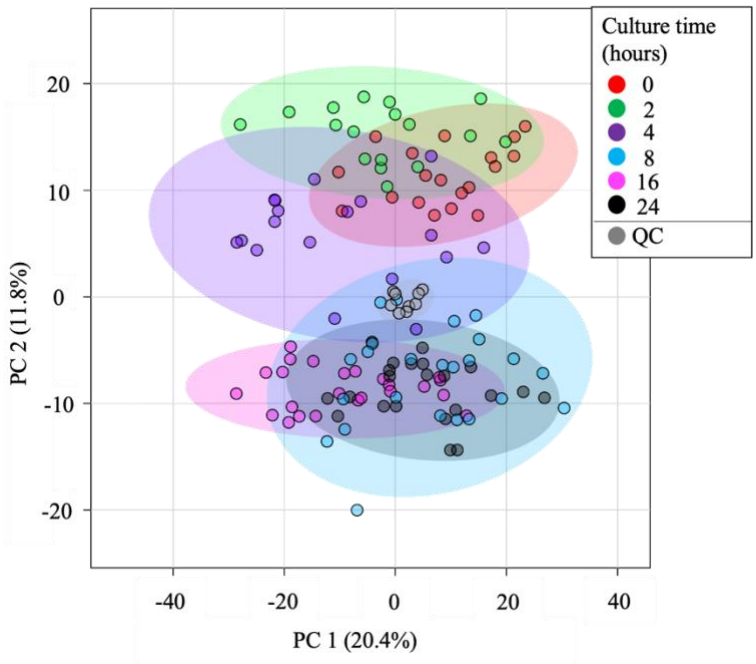


Figure 2. 10 PCA score plot of single cells with different culture times with the QC samples.

Figure 2.11 shows the volcano plot comparing the 0 h and 24 h cell population with the criterion of FC $\geq$ 1.5 and FC $\leq$ 0.67, q-value < 0.03. 228 metabolites (red dots) were up-regulated and 218 metabolites (blue dots) were down-regulated in response to 24 hours of culture. This reconfiguration includes induction of some amino acids and repression of others. For example, up-regulation of proline and methionine, down-regulation of alanine were found, as indicated in the plot.

For cell cohorts, the temporal pattern of abundance change is not always monotonic. For instance, the median abundance of leucine increases during the first 16 hours of culture and then decreases in 24 h as revealed in Figure 2.12A. In addition to this, the amount of cell-to-cell variability of abundance of individual amino acids, like the variability of median amino acid abundance for cell cohorts, can fluctuate over time shown in Figure 2.12A-E. Oscillatory fluctuation is evident for aspartate. Specifically, Figure 2.12E shows the cell-to-cell variability of

aspartate content increases between 0 and 4 hours, declines between 4 and 8 hours and increases again after 8 hours This oscillation is not associated with cell growth rate fluctuation, since maintenance medium does not support growth or substantial changes in the average amount of aspartate per oocyte. What is worth noting is that CIL LC-MS has revealed temporal oscillation in single-cell variability of metabolite abundance for the first time. The literature on such topics as circadian control of metabolism, oscillation of cell growth rate, fluctuation of cell to cell variability of RNA expression during a somatic cell developmental program[12] and stochasticity of biochemical reactions is expected to inform future work on the mechanistic underpinning of this new phenomenon.

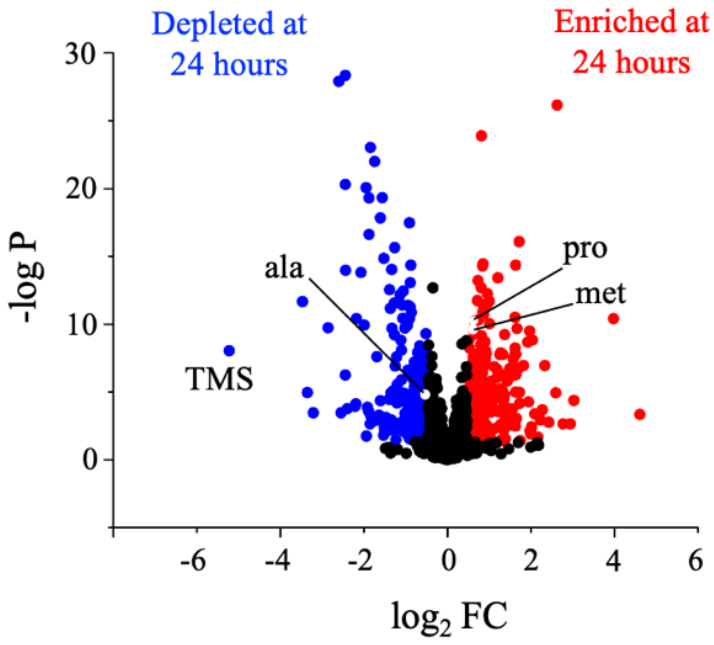


Figure 2. 11 Volcano plot for comparison of cells in 0 h and 24 h of culture.

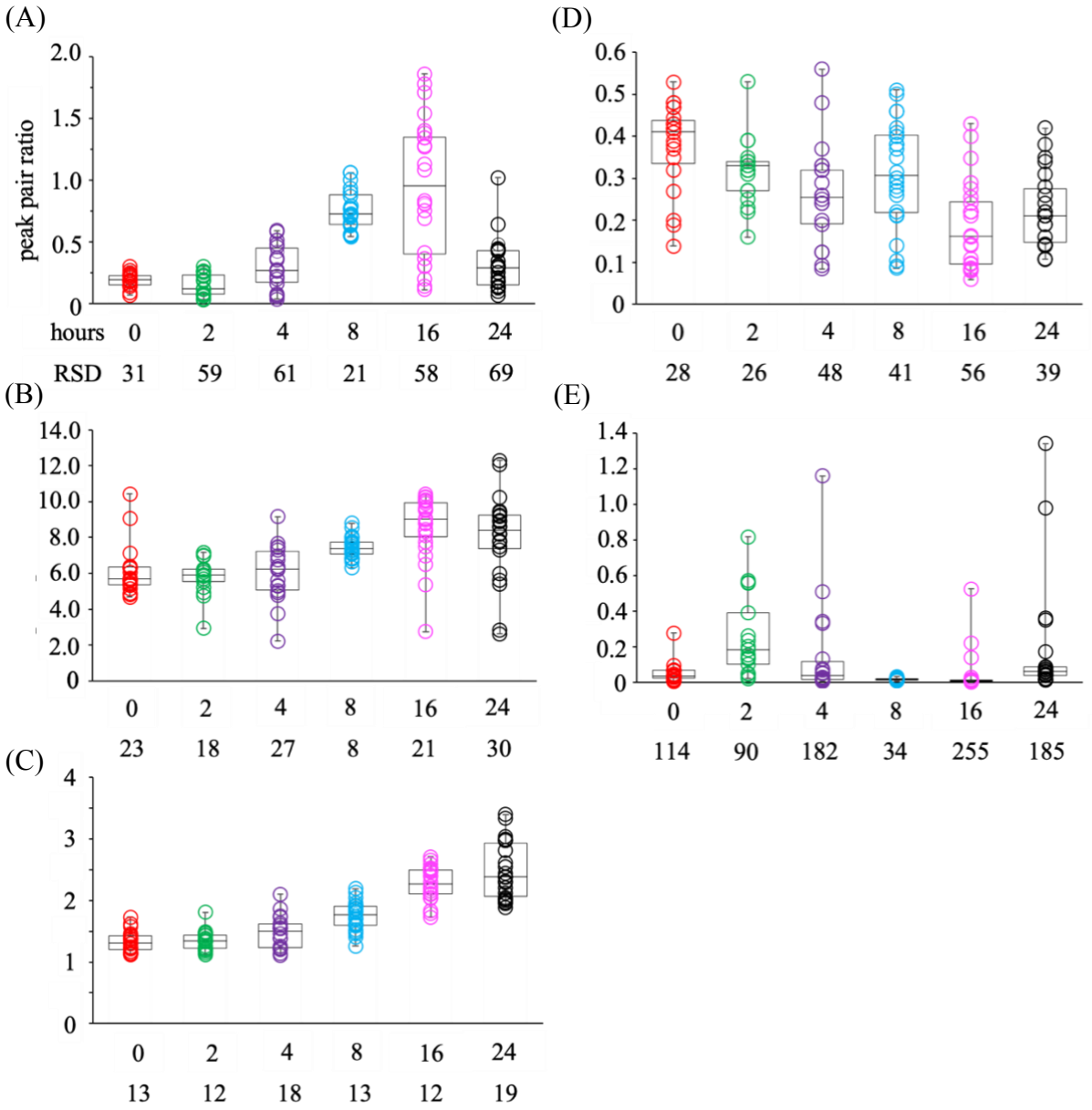


Figure 2. 12 Box plots showing the cell-to-cell variability of **(A)** leucine, **(B)** glycine, **(C)** serine, **(D)** alanine, **(E)** aspartate with RSD values indicated at the bottom of each group.

CIL LC-MS suggested the presence of the drug benzocaine in freshly isolated oocytes, as well as a biotransformation product, 3-aminobenzoic acid of the closely related anesthetic tricaine methanesulfonate (TMS)[13] used in this study. Subsequent comparison to a dansylation standard revealed that the peak pair tentatively assigned to benzocaine in fact corresponds to TMS. Therefore, CIL LC-MS has revealed uptake and metabolism of TMS, consistent with expression of the biotransforming enzyme cytochrome P450 2E1 (CYP2E1)[14] in oocytes[15]. Cell-to-cell variability of TMS and 3-aminobenzoic acid content is similarly high for freshly isolated oocytes from animals 1 and 2, even though they differ in the average starting amount of TMS/oocyte as shown in the t = 0 h group in Figure 2.13A. At the population level, the average concentration of TMS and 3-aminobenzoic acid declines during culture as can be seen in Figure 2.13A and 2.13B respectively. Nonetheless, cell-to-cell variability of TMS and 3-aminobenzoic acid content is high at all time points. These findings suggest that cell-to-cell variability of the capacity to transform drugs should be taken into account when assessing their effects on cells. Overall, our results encourage refinement of CIL LC-MS for metabolomics of single somatic cells in order to support drug development[16] and research on fundamental mechanisms of metabolic regulation in metazoans.

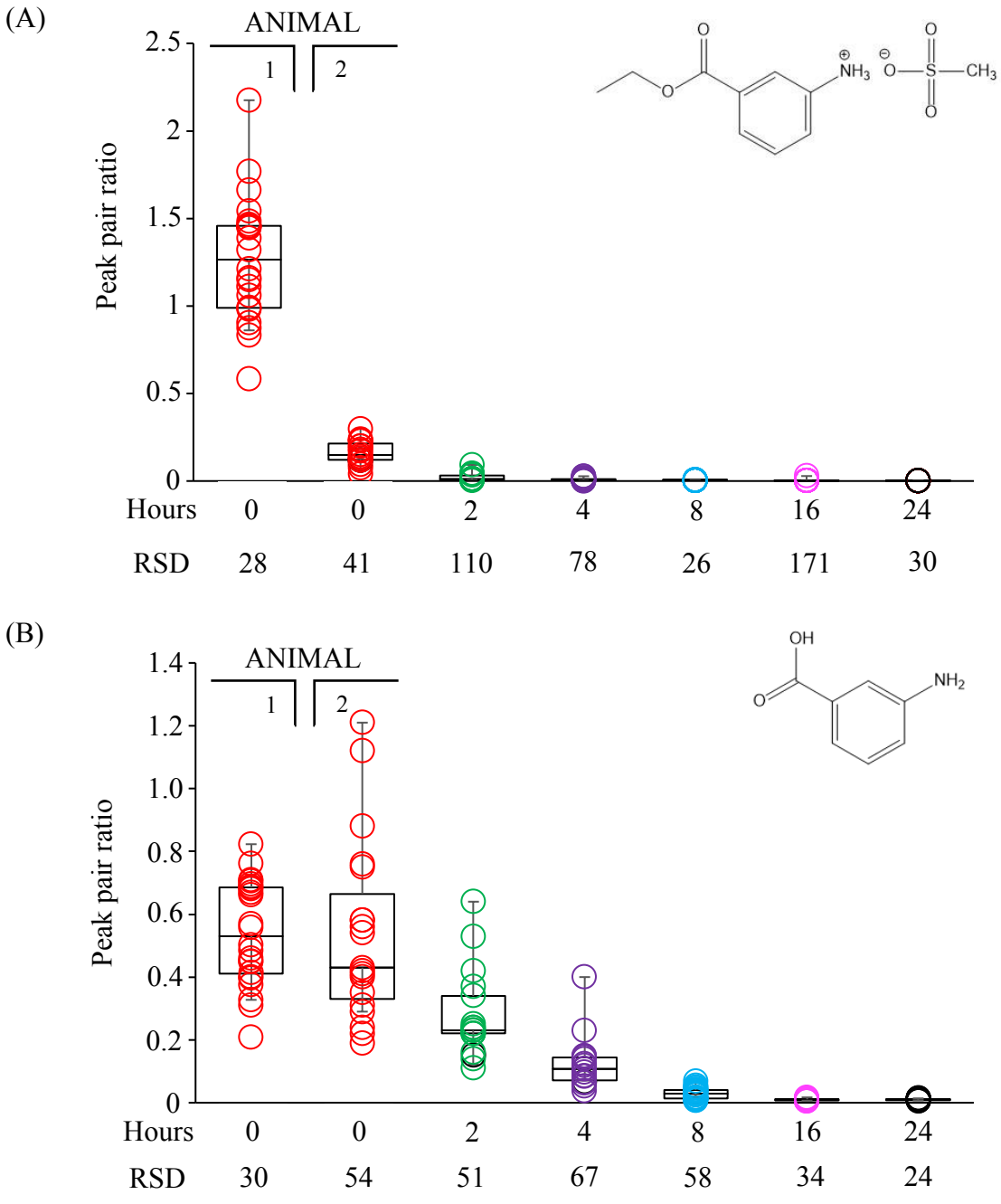


Figure 2. 13 Box plots showing the cell-to-cell variability of (A) TMS and (B) 3-aminobenzoic acid.

2.4 Conclusions and Future Work

In summary, we have developed a CIL LC-MS workflow to profile the amine and phenol submetabolome of single cells comprehensively and quantitatively. This involves cells extraction from *Xenopus laevis*, cell lysis to extract the metabolites, chemical isotope labeling of amines and phenols containing metabolites and subsequent LC-MS analysis. By applying CIL LC-MS, the amine and phenol submetabolome of cells can be analyzed with deep coverage in this work. We believe this is a big leap in single-cell metabolomics profiling and it has the potential to improve the understanding of cell heterogeneity and programming of cell metabolism. The work presented here may not have immediate biomedical applications, but it does illuminate a fundamental way in which cells work.

In the future, application of this workflow in other biological study of single cells, for example, the perturbation of cells is expected. Moreover, the use of other labeling chemistries to study different submetabolome of cells would also be performed in order to elevate our understanding of single-cell metabolome.

2.5 Literature Cited

1. Wang, D.; Bodovitz, S. *Trends Biotechnol.* **2010,** *28* (6), 281-290.

2. Fessenden, M. *Nature* **2016,** *540* (7631), 153-155.

3. Gupta, R. K.; Kuznicki, J. *Cells* **2020,** *9* (8), 1751.

4. Emara, S.; Amer, S.; Ali, A.; Abouleila, Y.; Oga, A.; Masujima, T. *Adv. Exp. Med. Biol.* **2017,** *965*, 323-343.

5. Sender, R.; Fuchs, S.; Milo, R. *PLOS Biol.* **2016,** *14* (8), e1002533.

6. Zeng, S. L.; Sudlow, L. C.; Berezin, M. Y. *Expert Opin Drug Discov* **2020,** *15* (1), 39-52.

7. Vastag, L.; Jorgensen, P.; Peshkin, L.; Wei, R.; Rabinowitz, J. D.; Kirschner, M. W. *PLoS One* **2011,** *6* (2), e16881.

8. Guo, K.; Li, L. *Anal. Chem.* **2009,** *81* (10), 3919-3932.

9. Zhou, R.; Li, L. *Methods Mol. Biol.* **2014,** *1198*, 127-36.

10. Crane, R. F.; Ruderman, J. V. *Methods Mol. Biol.* **2006,** *322*, 435-43.

11. Dworkin, M. B.; Dworkin-Rastl, E. *Dev. Biol.* **1989,** *132* (2), 512-23.

12. Richard, A.; Boullu, L.; Herbach, U.; Bonnafoux, A.; Morin, V.; Vallin, E.; Guillemin, A.; Papili Gao, N.; Gunawan, R.; Cosette, J.; Arnaud, O.; Kupiec, J.-J.; Espinasse, T.; Gonin-Giraud, S.; Gandrillon, O. *PLOS Biol.* **2016,** *14* (12), e1002585.

13. Rombough, P. J. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2007,** *148* (2), 463-9.

14. Oesch, F.; Fabian, E.; Guth, K.; Landsiedel, R. *Arch. Toxicol.* **2014,** *88* (12), 2135-2190.

15. Kırlı, K.; Karaca, S.; Dehne, H. J.; Samwer, M.; Pan, K. T.; Lenz, C.; Urlaub, H.; Görlich, D. *eLife* **2015,** *4*.

16. Vinegoni, C.; Dubach, J. M.; Thurber, G. M.; Miller, M. A.; Mazitschek, R.; Weissleder, R. *Drug Discov. Today* **2015,** *20* (9), 1087-1092.

Chapter 3 – Chemical Isotope Labeling LC-MS for Studying the Metabolic Response of Single Cells to Heat Stress

3.1 Introduction

Stress in cellular network can be defined as any sudden and unpredictable change to cells. There are many different types of cellular stress, including intrinsic and extrinsic stressors. Intrinsic means it is originating within a cell, such as genetic and endoplasmic reticulum stress. On the other hand, extrinsic is driven by external or environmental perturbations such as temperature change, exposure to toxin, and mechanical damages. Depending on the type, duration and severity of stress, cells response mainly in two ways: activation of pathways to protect cells against the stress and induction of programmed cell death to remove damaged cells[1]. More precisely, when the stress induced is reversible, protective responses would be initiated, which can help defend a cell against damage. In contrast, if it is irreversible, the damaged cell would be removed from the organism through different mechanisms leading to cell death, such as apoptosis and necrosis.

Temperature plays an important role in the survival of living organisms on Earth. It is because it can disturb, for example, the encoding of DNA and the structure of proteins in cells. Moreover, it can also affect all physical features of the environment, such as pH, gas solubility and viscosity[2]. Although the thermal limit of three major domains of life: bacteria, archaea and eukaryotes have a wide range, it should be noted that temperatures fairly beyond the optimal growth temperature of a particular living organism could have adverse effects on their survival. A sudden temperature increases off the optimal growth temperature of an organism is referring to heat stress here. It is well-known that heat stress can resulted in denaturation of protein, protein entanglement and non-specific protein aggregation, which can therefore lead to destruction of

cellular structures and thus interfere with essential functions[3]. One example of the damaging effect of elevated temperature to cell, eukaryotes particularly, is the cytoskeleton defect. Cytoskeleton is a system of filaments or fibers that helps the cell to maintain its shape and internal organization. In addition to this, it also provides mechanical support that enables the cell to conduct essential functions. Heat stress can cause disruption of the cytoplasmic actin, a family of globular multi-functional proteins that form microfilaments of the cytoskeleton, which has crucial role in many cellular processes, ranging from maintenance of cellular shape and motility, cytokinesis and to signal transduction[4, 5]. Besides, heat stress can affect cellular organelle. For example, structural changes within the mitochondria, disruption and fragmentation of the Golgi complex[6].

Many studies have reported that heat stress response is associated with an increased production of heat shock proteins, also well-known as molecular chaperone, in order to maintain protein homeostasis and thus normal cellular functions in almost all eukaryotic cells[3, 7]. Briefly, in response to elevated temperature, heat stress proteins can prevent undesirable protein aggregation by controlling the binding and liberate of nonnative proteins. They also manage the refolding to repair damaged proteins or degradation of damaged cellular proteins. If we consider the heat stress-induced change in different cellular networks[8], comprising of gene transcription[9, 10], protein interaction[11], organelle[12], cytoskeleton[13], signaling[14, 15] and the metabolic, we know that heat stress response in transcriptomic and proteomics have been extensively explored. Metabolomic analysis of heat stress response has been considered in detail in plant cells[16-19]. However, the study on the metabolomics side in animal cells, at the single-cell level specifically, is relatively lagging behind due mainly to limited sample volume, structural diversity and broad concentration range of metabolites. Furthermore, very few studies have made attempts to consider the cell-to-cell

variation in responding to heat stress, which is hard to achieve in classic cell averaging measurements.

In metabolic profiling, Mass Spectrometry (MS) provides a technology platform for highly sensitive, accurate and reproducible measurements of cellular metabolites. It thus has a distinct advantage compared to Nuclear Magnetic Resonance (NMR) spectrometry, with a main drawback of low sensitivity. One of the major challenges in metabolomics is the great diversity in chemical and physical properties of metabolites, owing to this fact, multiple analytical platforms are always needed in order to account for the complete metabolome in a biological system. For example, to target the polar metabolites, Hydrophilic Interaction Liquid Chromatography (HILIC) coupled to MS is needed, whereas nonpolar metabolites require Reversed Phase Liquid Chromatography (RPLC) – MS. These would greatly lengthen the overall analysis time. In consideration of this, CIL provides a way to address the challenge, in which the metabolome is grouped into submetabolomes based on their functional groups, and a labeling method is tailored to the analysis of a specific group of metabolites. In this way, the chromatographic behavior of metabolites would be altered and can be retained and separated on a reverse phase column. With CIL, differentially isotope tagged chemical reagent; *i.e.*,^{12}C and ^{13}C is used to react with a particular group of metabolites. To be more specific, samples are labeled with ^{12}C-reagent and a reference sample is labeled with ^{13}C-reagent. They are then mixed and used in LC-MS analysis, peak pairs differing by the mass between two isotope tags can be identified. There are numerous advantages of CIL, for instance, improved retention and thus separation of polar metabolites on a RP column and better ionization efficiency on MS[20]. Additionally, it also enables high precision and accuracy quantitative metabolome analysis. As can be easily understood, the ratio of peak pairs; *i.e.*, the light and heavy peak, can be used for relative or absolute quantification of metabolites[21].

The study here was designed to determine the effect of heat stress on cellular amine and phenol metabolome by CIL LC-MS. Cells were extracted from *Xenopus laevis*, the large size of oocytes can potentially provide enough material for metabolomics study. A labeling technique, dansylation, is used to elucidate the amine and phenol containing metabolites[20]. We shifted cells from 18℃, their optimal temperature to 34℃, and studied the short-term and long-term effects after the induction of heat stress. We also looked into the recovery of cells after stress by bringing them to 18℃. The significantly changed metabolites upon heat stress can be revealed in this study, and at the same time, the cell-to-cell variation in the same population can also be seen.

Elucidating the metabolome changes induced by heat stress in individual cells holds a great potential to elevate our understanding of the cellular physiology that regulates temperature perturbations in their environment. Macroscopically, heat stress responses in liver have been broadly studied, as liver is a crucial regulator of metabolism which controls many physiological processes affected by prolonged heat stress. Many of these studies focused on the molecular mechanism in response to heat stress in different animal livers, such as sheep, cow and chicken[22-24]. It is because heat stress can adversely affect the productivity and reproducibility of different animal breeds. Therefore, understanding the mechanism can help those animals to develop resistance to elevated temperature. In the study here, we look into the effect of heat stress microscopically; *i.e.*, at the single-cell level. This is important as we can see how individual cells react to heat stress, how they are different in regulating an increased temperature. When considering a clinical application, liver transplantation is a good example. In human liver transplant, after cold perfusion to preserve the liver to be transplanted, a period of rewarming when vascular anastomoses are being constructed is usually followed, this can be considered as a heat stress and many metabolic activities can happen during this period of elevated temperature[25]. That

is to say, our study has the potential to make a leap from the proof of concept on bench to answering important biological questions in bedside.

3.2 Experimental

3.2.1 Chemical and Reagents

In single-cell collection, tricaine methanesulfonate (MS-222) and Collagenase Type IA were purchased from Sigma-Aldrich Canada. OR-2 was prepared in the lab using 5 mM Hepes-NaOH (pH 7.8), 82.5 mM NaCl, 2.5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM Na_2HPO_4[26]. Lysis solution was prepared with ACN, H_2O and MeOH purchased from Sigma-Aldrich Canada.

All chemicals and reagents used in CIL LC-MS, unless otherwise stated, were purchased from Sigma-Aldrich Canada. For dansylation labeling, the ^{12}C-labeling reagent, dansyl chloride, was purchased from Sigma-Aldrich and the ^{13}C-labeling reagent was synthesized according to the previously published method[27].

3.2.2 Sample Collection and Preparation

Cells were recruited from *Xenopus laevis*, and the animal experimental procedures are according to protocols approved by the Biosciences Animal Care and Use Committee of the University of Alberta. Figure 3.1 illustrates the cell collection workflow. Concisely, a group of cells were placed on a cell culture plate, and being cultured overnight at 18℃ water bath. Cells were first collected and labeled "18℃ overnight". The plate of cells was then being moved to a 34℃ chamber for 15 minutes, and the second batch of cell samples was collected and labeled "34℃ 15min". 105 minutes later, we collected the third batch of cell samples and labeled "34℃ 2h".

Lastly, we brought cells back to 18℃ water bath and collected them after 2 hours, labeled them "18℃ 2h".

After cell collection, each of them was then being dispensed onto the side of an individual tube containing 80 μL of lysis solution; 2:2:1 acetonitrile:methanol:water mixture [28], ready on an ice bucket. Lysis solution is necessary for quenching the cellular metabolism, in others words, it stops the enzymatic activity of cells. After that, the cell was being lysed to extract metabolites by putting the tube on a high speed vortexer for 5 s in 4℃ cold room. Finally, we froze the lysed cell in liquid nitrogen immediately.

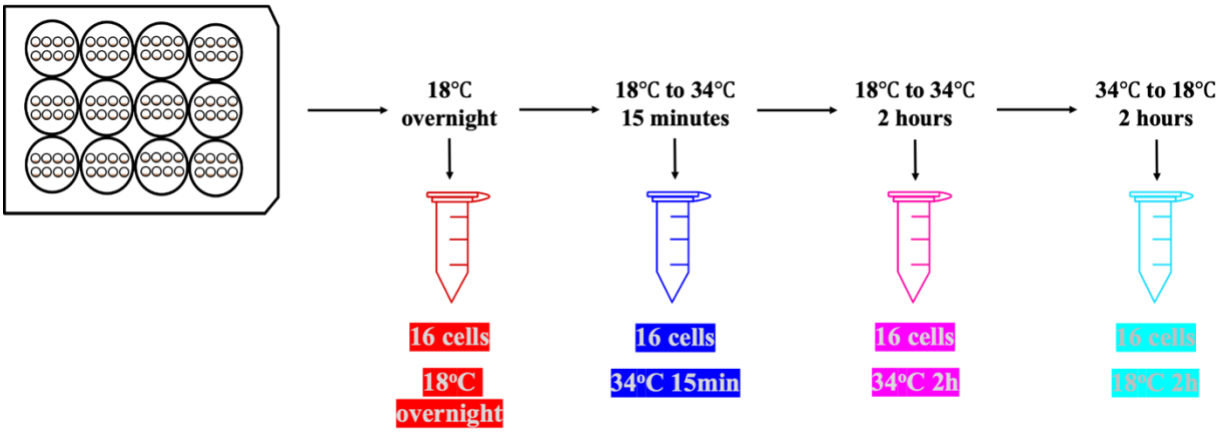


Figure 3. 1 Cell collection workflow for the investigation of heat stress effects on the cellular metabolome.

3.2.3 Chemical Isotope Labeling for Amine and Phenol Metabolomics

Lysates were thawed, dried down using a SpeedVac and resuspended in 80 uL of water. Individual cell sample was labeled with ^{12}C-DnsCl, using the approach we reported previously[27]. For the control sample, we employed universal internal standard (UIS), which is prepared by collecting and lysing a large number of cells from *Xenopus laevis*. The control sample was labeled

with ^{13}C-DnsCl. Individual sample was then mixed with the control sample in 1:1 volume ratio to generate ^{12}C/^{13}C-labeled mixture for LC-MS analysis.

3.2.4 LC-MS

Sample mixtures were analyzed by Bruker Impact HD Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA) coupled to Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). Sample was injected to Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Mobile phase for the gradient elution was solvent A: 0.1% (v/v) formic acid in 5% (v/v) acetonitrile and solvent B: 0.1% (v/v) formic acid in acetonitrile. The gradient elution profile was (i) 0-3.5 min, A+B (20-35% gradient); (ii) 3.5-18 min, A+B (35-65% gradient); (iii) 18-21 min, A+B (65-95% gradient); (iv) 21-37min, A+B (95% isocratic) with flow rate 180 μL/min and injection volume 12 μL. Mass spectra were recorded from 220 to 1000 m/z with a spectra rate of 1.0 Hz for positive ionization mode. The nebulizer was set to 1.0 bar, and the dry temperature was 230 ℃, with 8 L/min of drying gas. The capillary and end plate offset voltage were set as 4500 V and 500 V respectively. ^{12}C- and ^{13}C- labeled metabolites appeared as peak pairs and coeluted perfectly because there are only two isotope-labeled carbon atoms different between them. This differential CIL approach has been previously evaluated to be desirable for relative quantification of metabolites[27].

3.2.5 Data Processing

MS spectral peaks data generated from LC-MS was extracted using Bruker DataAnalysis software 4.4. The metabolomics data was then being processed with IsoMS Pro, which is a software platform for the interpretation of data obtained from CIL LC-MS based metabolomics

studies. The main steps in metabolomics data analysis here included data quality check, data processing, data cleansing, metabolite identification and statistical analysis. Data quality check ensures high quality data by performing mass accuracy and retention time check. Data processing involves peak pair picking from raw files, alignment and calculation of peak pair ratios among all samples and zero-filling[29, 30], which retrieves missing values by searching through the raw data. Data cleansing involves blank treatment, missing value treatment and sample wise normalization[31] so as to make sure the data is complete for metabolite identification and statistical analysis. Metabolite identification was performed using a three-tier metabolite identification approach. Definitive identification depends on Tier 1 - CIL library search, which is based on RT and m/z by searching against CIL standard libraries on the basis of m/z tolerance window 10 ppm and retention time window of 30 s. High-confidence putative identification is based on retention time and m/z searches against the Tier 2 - LI library, which covers all common metabolism pathways containing endogenous metabolites and their derivatives. The m/z window is 10 ppm, while the experimental retention time window is 30 s. Finally, putative identification is based on m/z matching against Tier 3 - mass-based database with monoisotopic mass window of 10 ppm. This library is composed of human endogeneous metabolites and their predicted metabolic products after one or two common metabolic reactions. Statistical analysis, including univariate analysis such as volcano plot and multivariate analysis, for instance, PCA and PLS-DA, were also completed in the IsoMS Pro software.

3.3 Results and Discussion

3.3.1 Amine and phenol metabolomics

The profiling of 64 single-cell samples resulted in the detection of 2876 peak pairs. Among them, 2486 peak pairs were identified with the three-tier metabolite identification approach, which means we identified 86% of detected features. Herein, 122 metabolites were positively identified by retention time and m/z search against the CIL standard libraries. 233 metabolites were putatively identified with high confidence by retention time and m/z match against the LI library, and 2131 metabolites were putatively identified by m/z match against the mass-based database.

Figure 3.2 shows the principal component analysis (PCA) plot of the metabolome data generated from samples analyzed. There is a close clustering of the quality control (QC) samples, indicating that CIL LC-MS did not introduce instrument bias in sample analysis. The 18℃ overnight population cannot be separated from the 34℃ 2h and 18℃ 2h population. However, there is an apparent visual separation between the 18℃ overnight group and the 34℃ 15min group on the principal component 1 (PC1). In a 2-dimensional PCA plot, PC1 represents the direction that contains the most amount of variation between metabolites, while principal component 2 (PC2) describes the direction that contains the second most variation. Figure 3.3 displays the partial least square discriminant analysis (PLS-DA) plot of the data, there are separations between the 18℃ overnight group and the 34℃ 2h and 18℃ 2h group on PC2. The separation of 18℃ overnight group and 34℃ 15min group on PC1 can also be observed. The plot had a R^2 (goodness of fit) and Q^2 (predictability) value of 0.9347 and 0.8396, respectively. Using a 100-permutation test, the model passed the validation with original Q^2 much larger than the permutated data as shown in Appendix Figure A3.1 which means there is no issue of over-fitting. Overall, the pattern and directionality of these plots show that the amine and phenol metabolome of cells does change after

the onset of heat stress, particularly during the first 15 minutes after cells being shifted from 18℃ to 34℃.

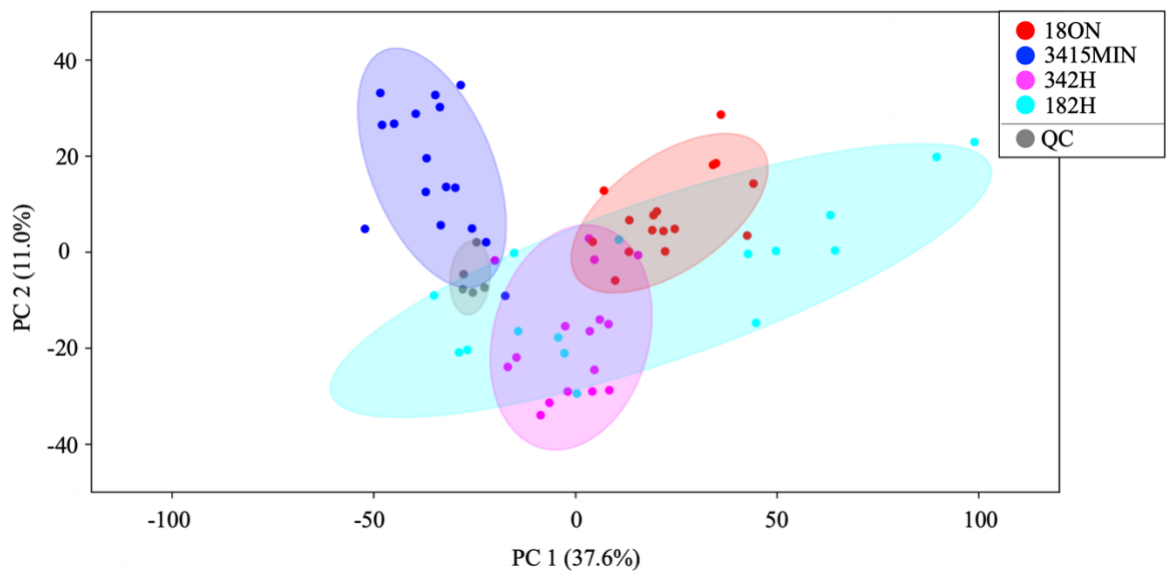


Figure 3. 2 PCA score plot of single-cell samples in different groups with QC data.

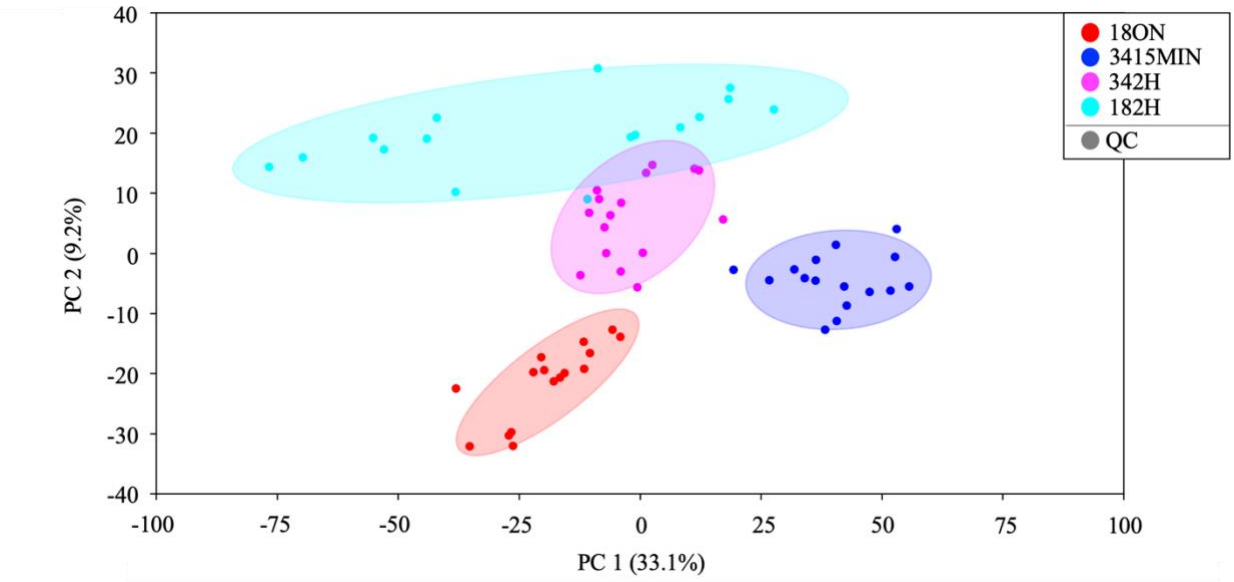
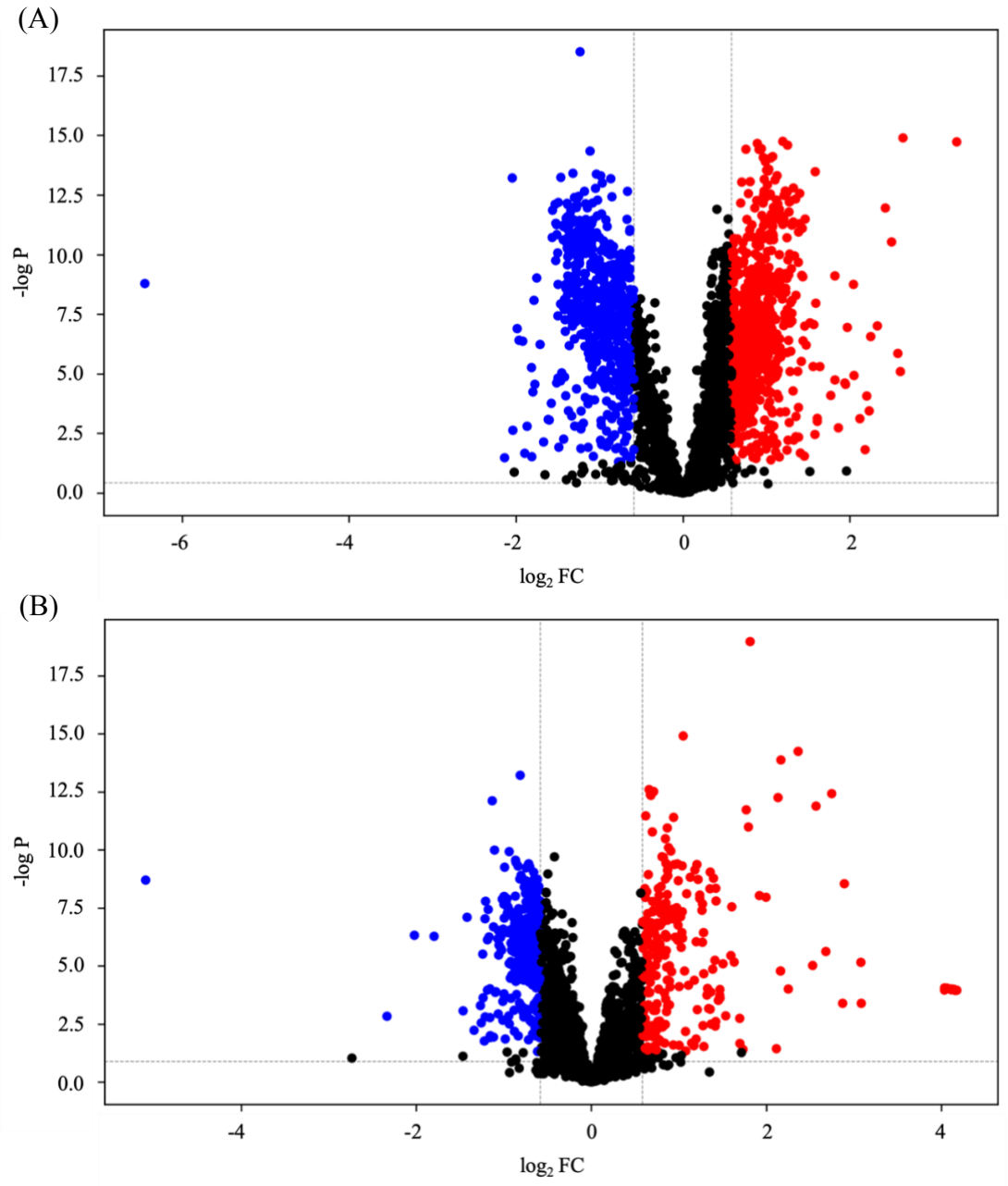


Figure 3. 3 PLS-DA score plot of single-cell samples in different groups.

3.3.2 Heat stress response of cells

We first cultured the cells at their optimal temperature (18℃), overnight. We then induced a heat stress of 34℃ to the cells. Cellular responses at two stages after the induction of heat stress were investigated, which are heat exposure of 15 minutes and 2 hours. We use volcano plots to analyze the metabolome data in this study. Volcano plot allows the visualization of multiple univariate analyses comprehensively, it is for binary comparison between two groups which is presented as a scatter plot with statistical significance (p-value) verses fold change (FC). Metabolites with significant concentration changes can be observed very clearly from the plot. The significant changed metabolites between two groups were identified and plotted on volcano plots, with the following criteria: p-value < 0.05, FC ≥ 1.5 for significantly increased metabolites and FC ≤ 0.67 for significantly decreased metabolites. Figure 3.4A shows the volcano plot of 18℃ overnight vs. 18℃ to 34℃ 15min, with a q-value of 0.007. 737 metabolites were significantly increased and 660 metabolites were significantly decreased after cultured at 34℃ for 15 minutes. While Figure 3.4B displays the volcano plot of 18℃ overnight vs. 18℃ to 34℃ 2h with a q-value of 0.02, there were 250 increases and 374 decreases. For those significant metabolites, 151 increases and 335 decreases were in common at 15 minutes and 2 hours of heat stress. And from the number of metabolites that significantly changed their concentration, we can again conclude that short-term heat stress showed a greater impact on the cellular amine and phenol metabolome. We think that a possible reason for this is after 2 hours at 34℃, cells may have adjusted steady-state metabolite pools back to a similar composition as cells at 18℃. In other words, the enzymes system of cells may have adapted to high temperature in response to long-term heat stress. When we looked into amino acids at different stages of heat stress, we observed that at 15 minutes after shifting cells to 34℃, expression of proline and leucine were significantly higher, while expression

of arginine, asparagine, glutamine, serine, aspartic acid and glutamic acid were significantly lower, as summarized in Figure 3.5A-H. And at 2 hours after cultured at 34℃, tyrosine, leucine, isoleucine, proline, valine, lysine, glycine, histidine and phenylalanine expression were much higher, whereas no essential amino acid expression was dramatically declined, as illustrated in Figure 3.6A-I.



(C)

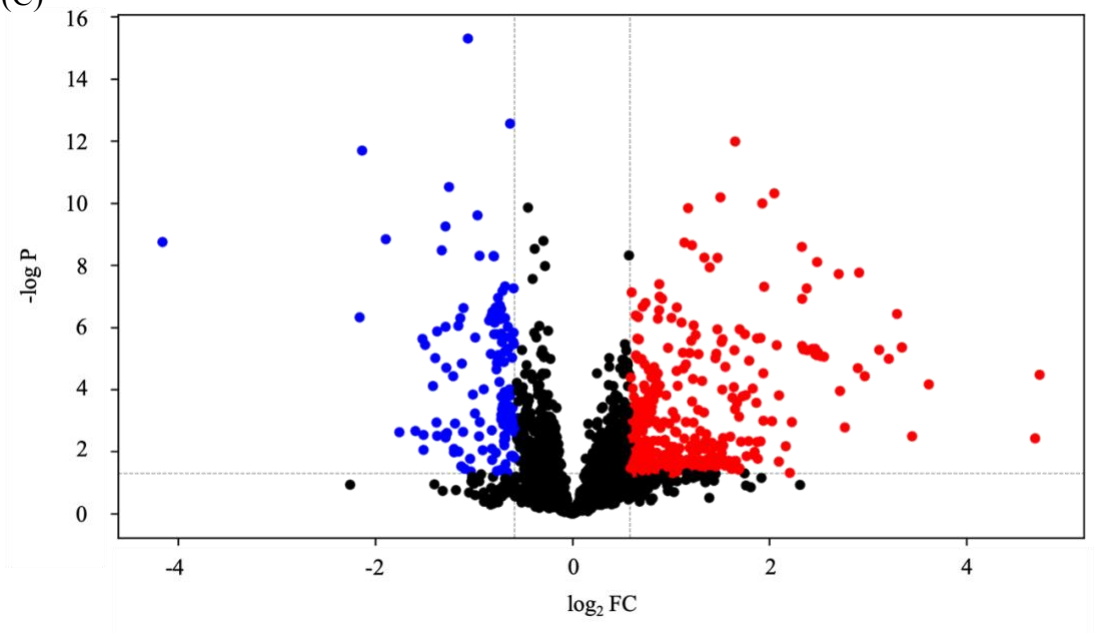


Figure 3. 4 Volcano plot comparing cells in **(A)** 18 ℃ overnight and 18 ℃ to 34 ℃ 15min population. **(B)** 18℃ overnight and 18℃ to 34℃ 2h population and **(C)** 18℃ overnight and 18℃ 2h population. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q = 0.007 for (A), q = 0.02 for (B) and q = 0.05 for (C)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

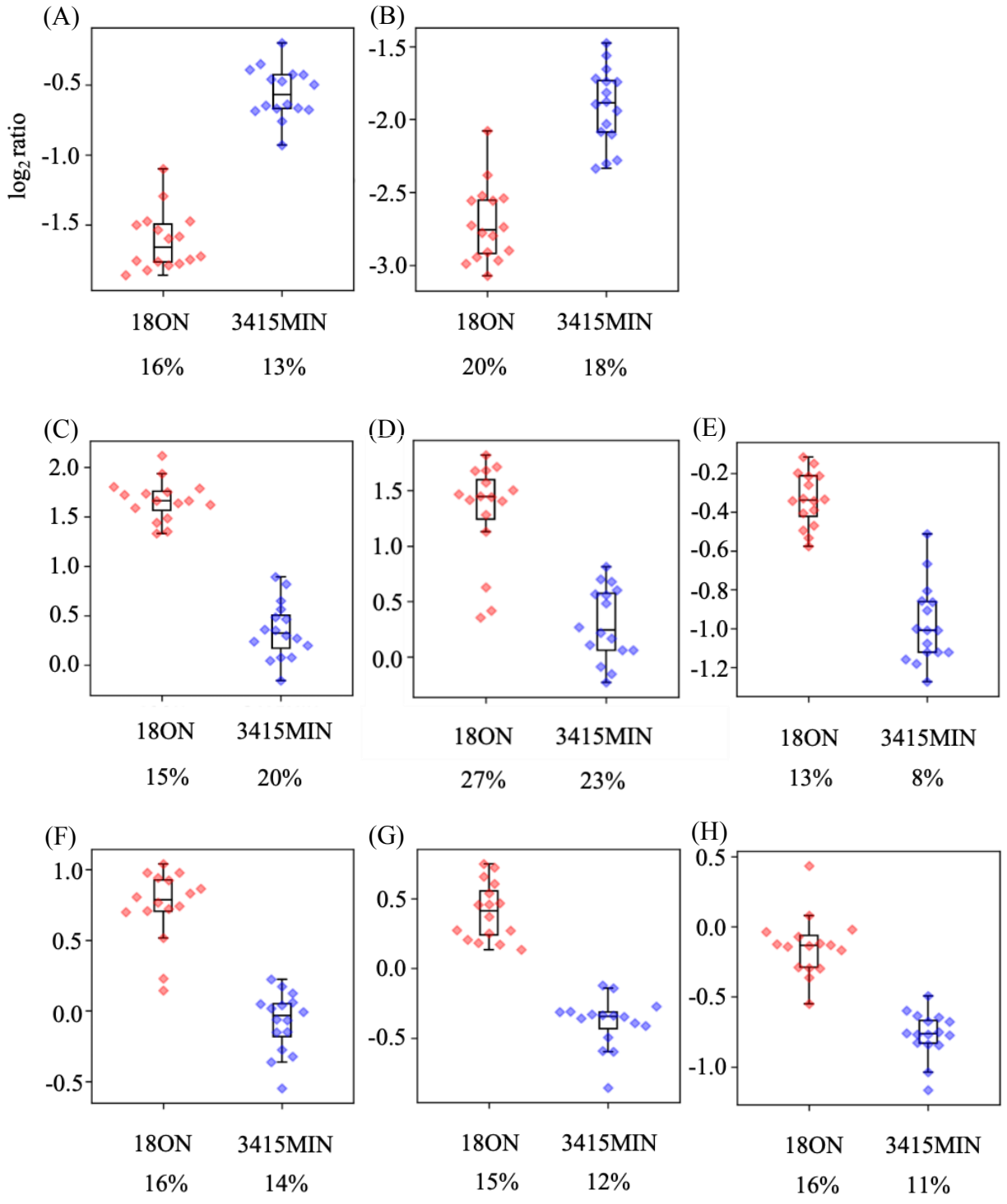


Figure 3. 5 Box plots of **(A)** proline **(B)** leucine **(C)** arginine **(D)** asparagine **(E)** glutamine **(F)** serine **(G)** aspartic acid **(H)** glutamic acid with RSD values indicated.

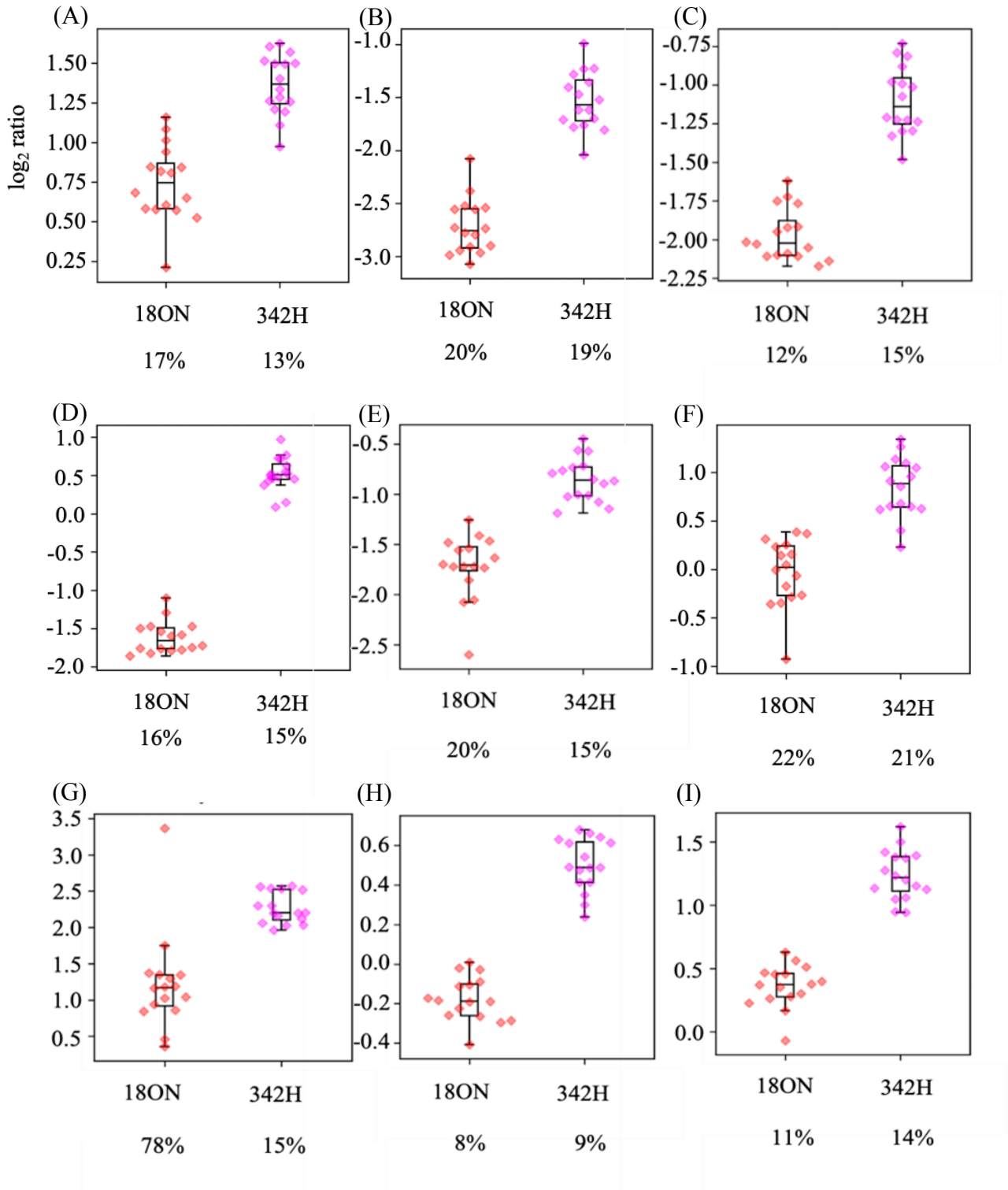
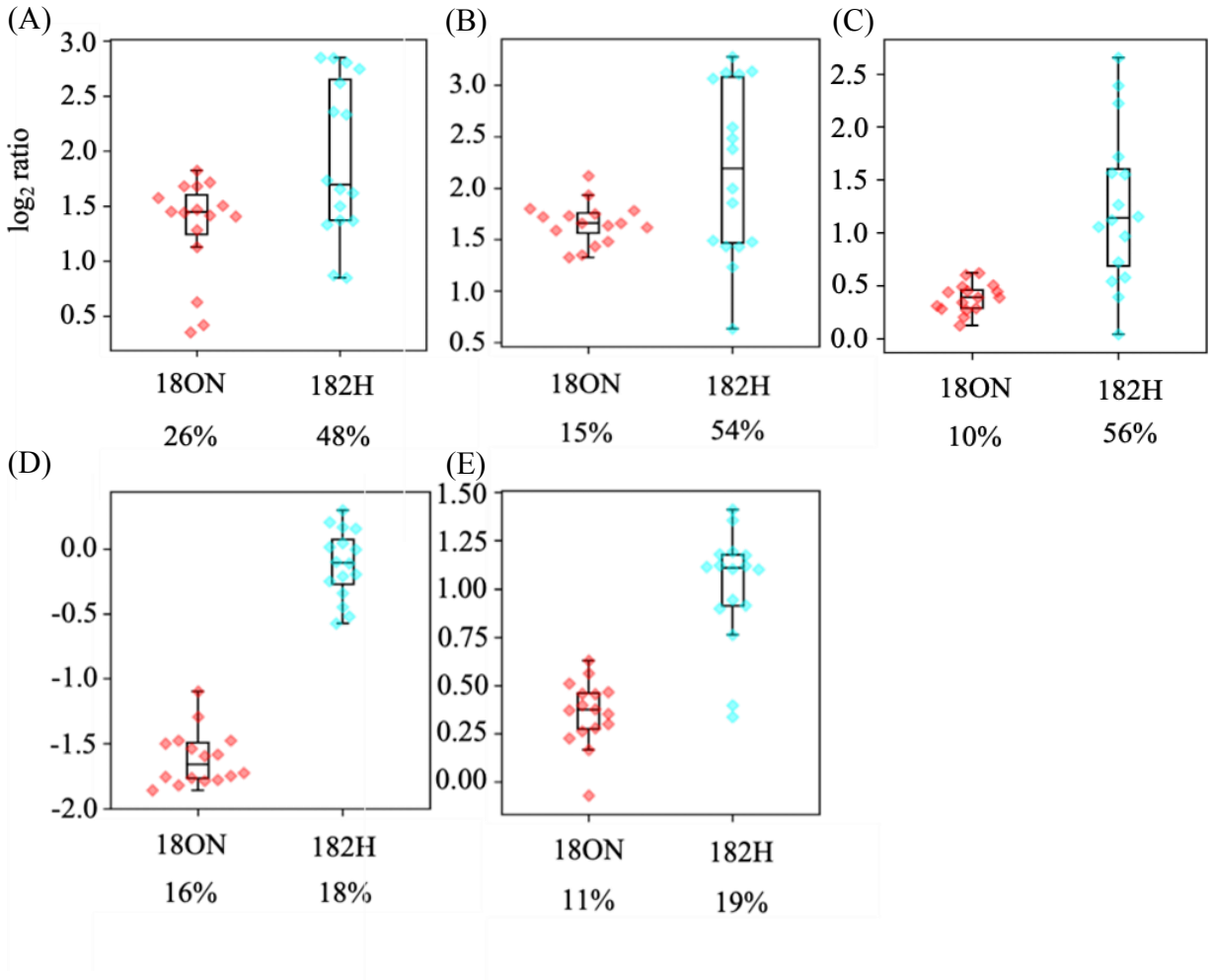


Figure 3. 6 Box plots of **(A)** tyrosine **(B)** leucine **(C)** isoleucine **(D)** proline **(E)** valine **(F)** lysine **(G)** glycine **(H)** histidine **(I)** phenylalanine with RSD values indicated.

3.3.3 Recovery of cells after heat stress

After the induction of heat stress at 34℃ for 2 hours, we brought cells back to 18℃ for another 2 hours to investigate if recovery from heat stress is accompanied by changes in metabolite abundance. Figure 3.4C reveals the volcano plot of 18℃ overnight vs. 18℃ 2h group with a q-value of 0.05. 403 metabolites were significantly increased while 177 metabolites were significantly decreased. Asparagine, arginine, isoleucine, proline and phenylalanine were significantly increased, while alanine and glutamic acid were significantly decreased, as shown in Figure 3.7A-G. These results suggest the cellular amine, phenol metabolome changed after heat stress, even they were in their optimal temperature again.



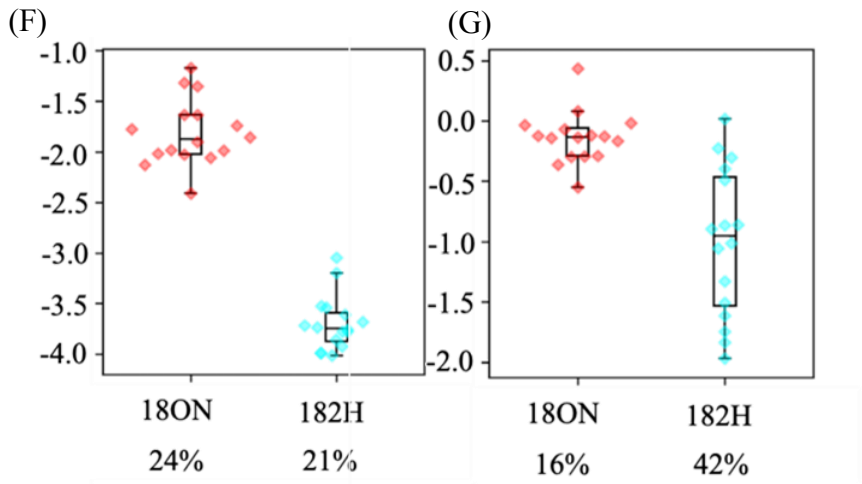


Figure 3. 7 Box plots of **(A)** asparagine **(B)** arginine **(C)** isoleucine **(D)** proline **(E)** phenylalanine **(F)** alanine **(G)** glutamic acid.

3.3.4 Cellular GSH levels under heat stress

Glutathione (GSH) is an antioxidant in eukaryotic cells, it involves in protecting the cells from harmful effects of excess oxidant stress, for example, free radicals, peroxides and heavy metals[32, 33]. Our results indicated at 15 minutes after the onset of heat stress, GSH level in cells was significantly decreased. And at 2 hours, there was no significant change in GSH concentration. These suggested that heat stress is associated with oxidative stress in some extent, short-term heat stress can greatly decrease the GSH level in cells and thus adversely affect important cellular functions in cells. However, for long-term heat stress, as mentioned before, cell adaptation may occur so that GSH level was not being significantly affected in the 342H population. For the recovery of cells after heat stress, when we shifted cells from 34℃ to 18℃ for 2 hours, GSH level in cells was significantly increased, to a much higher level compared to undisturbed cells. The level of GSH in different cell populations were summarized in Figure 3.8A-C.

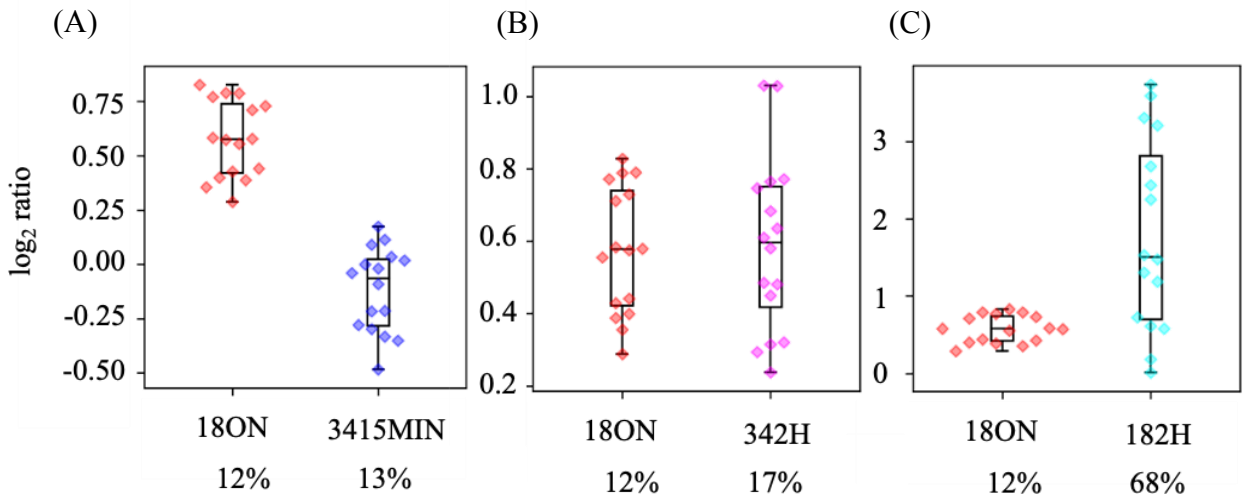


Figure 3. 8 Box plots of glutathione in **(A)**, **(B)** and **(C)**.

3.3.5 Cell-to-cell metabolite variation in response to heat stress

The most advantageous point of single-cell analysis is differences between single cells response to heat stress, which have not been technological discernable by cell averaging measurement thus far, has now become possible. In this work, we also reveal the cell-to-cell variation of some significantly changed amino acids upon heat stress. From Figure 3.5 to 3.8, the number under each group represents the percentage relative standard deviation (%RSD) of all cells within a group of a specific amino acid expression. For example, we can see at 15 minutes after the induction of heat stress, all the significantly changed amino acids have RSD of ≤20%, with the exception of asparagine, in which RSD is slightly higher, 27% and 23% in 18ON and 3415MIN population respectively. However, in 18℃ 2h group, cellular expression of some amino acids has relatively high RSDs. For instance, asparagine, arginine, isoleucine and glutamic acid, with RSDs ranging from 42% to 56% in 18℃ 2h group. These suggest that cellular expression of some metabolites can be very different even in the same population, after cells were recovered from heat stress.

3.4 Conclusions

We have developed and applied a CIL LC-MS method to study the metabolic responses of cells to heat stress. Dansylation was used to elucidate and quantify the change of amine and phenol submetabolome in cells. Short-term heat stress has great effects on the cellular metabolome and the enzymes system of cells may have adapted to high temperature in response to long-term heat stress. Recovery from heat stress was accompanied by changes in metabolite abundance, suggesting that the cellular amine, phenol metabolome changed after heat stress. With CIL LC-MS, the behavior of each single-cell can also be revealed, and we can see that the cellular response to the same environment was heterogeneous in certain situations.

3.5 Literature Cited

1. Fulda, S.; Gorman, A. M.; Hori, O.; Samali, A. *Int J Cell Biol* **2010,** *2010*, 214074.
2. Clarke, A. *Int. J. Astrobiology* **2014,** *13*, 141.
3. Richter, K.; Haslbeck, M.; Buchner, J. *Mol. Cell* **2010,** *40* (2), 253-66.
4. Gu, J.; Emerman, M.; Sandmeyer, S. *Mol. Cell. Biol.* **1997,** *17* (7), 4033-42.
5. Mounier, N.; Arrigo, A. P. *Cell Stress Chaperones* **2002,** *7* (2), 167-76.
6. Welch, W. J.; Suhan, J. P. *J. Cell Biol.* **1985,** *101* (4), 1198-211.
7. Barna, J.; Csermely, P.; Vellai, T. *Cell Mol Life Sci* **2018,** *75* (16), 2897-2916.
8. Szalay, M. S.; Kovács, I. A.; Korcsmáros, T.; Böde, C.; Csermely, P. *FEBS Lett.* **2007,** *581* (19), 3675-80.
9. Garbuz, D. G. *Mol Biol (Mosk)* **2017,** *51* (3), 400-417.
10. Gomez-Pastor, R.; Burchfiel, E. T.; Thiele, D. J. *Nat Rev Mol Cell Biol* **2018,** *19* (1), 4-19.

11. Luo, M.; Meng, Z.; Moroishi, T.; Lin, K. C.; Shen, G.; Mo, F.; Shao, B.; Wei, X.; Zhang, P.; Wei, Y.; Guan, K. L. *Nat Cell Biol* **2020,** *22* (12), 1447-1459.

12. Ruan, L.; Zhou, C.; Jin, E.; Kucharavy, A.; Zhang, Y.; Wen, Z.; Florens, L.; Li, R. *Nature* **2017,** *543* (7645), 443-446.

13. Pan, Z.; Shao, Y.; Geng, Y.; Chen, J.; Su, L. *Zhonghua Wei Zhong Bing Ji Jiu Yi Xue* **2015,** *27* (8), 639-42.

14. Ganesan, S.; Pearce, S. C.; Gabler, N. K.; Baumgard, L. H.; Rhoads, R. P.; Selsby, J. T. *J. Therm. Biol.* **2018,** *72*, 73-80.

15. Ganesan, S.; Volodina, O.; Pearce, S. C.; Gabler, N. K.; Baumgard, L. H.; Rhoads, R. P.; Selsby, J. T. *Physiol Rep* **2017,** *5* (16).

16. Abdelrahman, M.; Burritt, D. J.; Gupta, A.; Tsujimoto, H.; Tran, L. P. *J Exp Bot* **2020,** *71* (2), 543-554.

17. Thomason, K.; Babar, M. A.; Erickson, J. E.; Mulvaney, M.; Beecher, C.; MacDonald, G. *PLoS One* **2018,** *13* (6), e0197919.

18. Feng, Z.; Ding, C.; Li, W.; Wang, D.; Cui, D. *Food Chem* **2020,** *310*, 125914.

19. Wada, H.; Hatakeyama, Y.; Nakashima, T.; Nonami, H.; Erra-Balsells, R.; Hakata, M.; Nakata, K.; Hiraoka, K.; Onda, Y.; Nakano, H. *Sci Rep* **2020,** *10* (1), 2013.

20. Guo, K.; Li, L. *Anal Chem* **2009,** *81* (10), 3919-32.

21. Zhou, R.; Li, L. *Methods Mol. Biol.* **2014,** *1198*, 127-36.

22. Li, Y.; Kong, L.; Deng, M.; Lian, Z.; Han, Y.; Sun, B.; Guo, Y.; Liu, G.; Liu, D. *Genes* **2019,** *10* (5).

23. Ma, L.; Yang, Y.; Zhao, X.; Wang, F.; Gao, S.; Bu, D. *PLoS One* **2019,** *14* (1), e0209182.

24. Hubbard, A. H.; Zhang, X.; Jastrebski, S.; Singh, A.; Schmidt, C. *BMC Genomics* **2019,** *20* (1), 502.

25. Hertl, M.; Howard, T. K.; Lowell, J. A.; Shenoy, S.; Robert, P.; Harvey, C.; Strasberg, S. M. *Liver Transplant.* **1996,** *2* (2), 111-7.

26. Vinegoni, C.; Dubach, J. M.; Thurber, G. M.; Miller, M. A.; Mazitschek, R.; Weissleder, R. *Drug Discov. Today* **2015,** *20* (9), 1087-1092.

27. Guo, K.; Li, L. *Anal. Chem.* **2009,** *81* (10), 3919-3932.

28. Vastag, L.; Jorgensen, P.; Peshkin, L.; Wei, R.; Rabinowitz, J. D.; Kirschner, M. W. *PLoS One* **2011,** *6* (2), e16881.

29. Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. *Anal. Chem.* **2014,** *86* (10), 4675-4679.

30. Huan, T.; Li, L. *Anal. Chem.* **2015,** *87* (2), 1306-1313.

31. Wu, Y.; Li, L. *J Chromatogr A* **2016,** *1430*, 80-95.

32. Pompella, A.; Visvikis, A.; Paolicchi, A.; De Tata, V.; Casini, A. F. *Biochem. Pharmacol.* **2003,** *66* (8), 1499-503.

33. Jeong, E. M.; Yoon, J. H.; Lim, J.; Shin, J. W.; Cho, A. Y.; Heo, J.; Lee, K. B.; Lee, J. H.; Lee, W. J.; Kim, H. J.; Son, Y. H.; Lee, S. J.; Cho, S. Y.; Shin, D. M.; Choi, K.; Kim, I. G. *Stem Cell Rep.* **2018,** *10* (2), 600-614.

Chapter 4 – High-Coverage Single-Cell Metabolomics for Studying Cell-to-Cell Variations from Different Locations of the *Xenopus laevis* Ovary

4.1 Introduction

The ovary is regarded as the largest organ in the adult female African clawed frog *Xenopus laevis*, which occupied a large portion of its abdominal cavity. During dissection, if we get the ovary out in one piece and spread it out, it appears as a large piece of leave with finger-like structure, known as lobes, pointing outward. In general, every lobe is made up of hundreds of follicles, each contains an oocyte at any developmental stage, *i.e.*, stage I to stage VI[1]. *Xenopus* oocytes have been widely used for studies of oogenesis and steroidogenesis[2-4] owing to their large size and high abundance. Moreover, its conserved cellular, developmental and genomic organization with mammals also make it amenable to be used in research ranging from the understanding of human development and diseases[5-8] to drug discovery[9].

We know that metabolomics is a good representation of the molecular phenotype, as the metabolites, or more precisely the level or concentration changes of metabolites can directly reflect the underlying biochemical activity and the state of cells or tissues. Given that *Xenopus* oocytes provide an abundant source of material, the use of them in single-cell metabolomics studies has been growing exponentially in recent years[10-13].

It is a well-known fact that metabolites have wide and diverse range of chemical and physical properties. As a result, multiple LC and MS types are always required to increase the coverage of metabolic analysis. For example, hydrophilic interaction liquid chromatography (HILIC) column is needed for the separation of polar and hydrophilic metabolites, whereas reverse phase liquid chromatography (RPLC) column is necessary for separating non-polar, hydrophobic metabolites. In addition to this, samples have to be run in both positive and negative ion modes so

as to detect various types of metabolites. The downside of these is it greatly increases the overall analysis time, which is not desirable in high-throughput analysis, especially in clinical, forensic and doping laboratories where high speed analysis is important. Furthermore, low throughput for quantitative metabolome analysis with high coverage is another major problem in metabolomics. Because of the high sensitivity and specificity of MS, it is extensively used in quantitative metabolomics. In order to quantify metabolites accurately, calibration is of utmost importance. Among the three common calibration methods, external standard method, standard addition method and internal standard (IS) method, IS method provides two distinct advantages. First, it can overcome the matrix effect and the instrumental signal response fluctuations. Second, it compensates for sample losses which is possible during sample preparation and processing, and in various chromatographic steps. Because of these, a technique that could provide high-throughput qualitative and quantitative analysis, with a great depth of chemical coverage is imperative in metabolic research.

CIL LC-MS is an emerging field of metabolomics in attempting to describe the metabolome systematically. The rationale is centered on the chemical-group-based submetabolome profiling, *i.e.*, the whole metabolome in a biological system is divided into four different chemical groups, amine and phenol submetabolome, hydroxyl submetabolome, carboxylic acid submetabolome and carbonyl submetabolome. Each group of metabolites is then labeled with a reagent before LC-MS analysis. The CIL is designed to enhance separation and increase MS sensitivity, thus the metabolites detectability. Therefore, it ultimately leads to higher metabolome coverage[14]. For quantification, the concept of differential isotope labeling allows accurate relative quantification of all detected labeled metabolites. To be more specific, individual samples are labeled with ^{12}C-reagent, they are then spiked with a ^{13}C-reagent labeled pooled

sample. After that, they are mixed together and the mixture was used for LC-MS analysis. The ratio of the light- (^{12}C) and heavy-labeled (^{13}C) metabolite peak area is used to represent the relative concentration difference in two comparative samples[15].

It this work, we report the application of CIL LC-MS to study cells taken from different regions of the ovary. The region is the tips of the ovary lobe, since they are very distinct anatomical parts and are easy to cut away from the rest of the organ. We hypothesized that CIL LC-MS can be used to analyze the cell-to-cell variability of the metabolome based on cell locations in the organ. This study makes attempt to consider the whole metabolome of single cells by analyzing the samples in four-channels CIL LC-MS separately[16], these are dansylation for amine and phenol submetabolome[14], base-activated dansylation for hydroxyl submetabolome[17], p-dimethylaminophenacyl (DmPA) bromide labeling for carboxylic acid submetabolome[18] and dansylhydrazine labeling for carbonyl submetabolome[19]. The workflow of the study is shown in Figure 4.1. We then combine the result which would represent the whole metabolome of single cells, and evaluate the cell-to-cell differences from different locations of the ovary, which allows a holistic understanding of metabolic differences of cells in an organ.

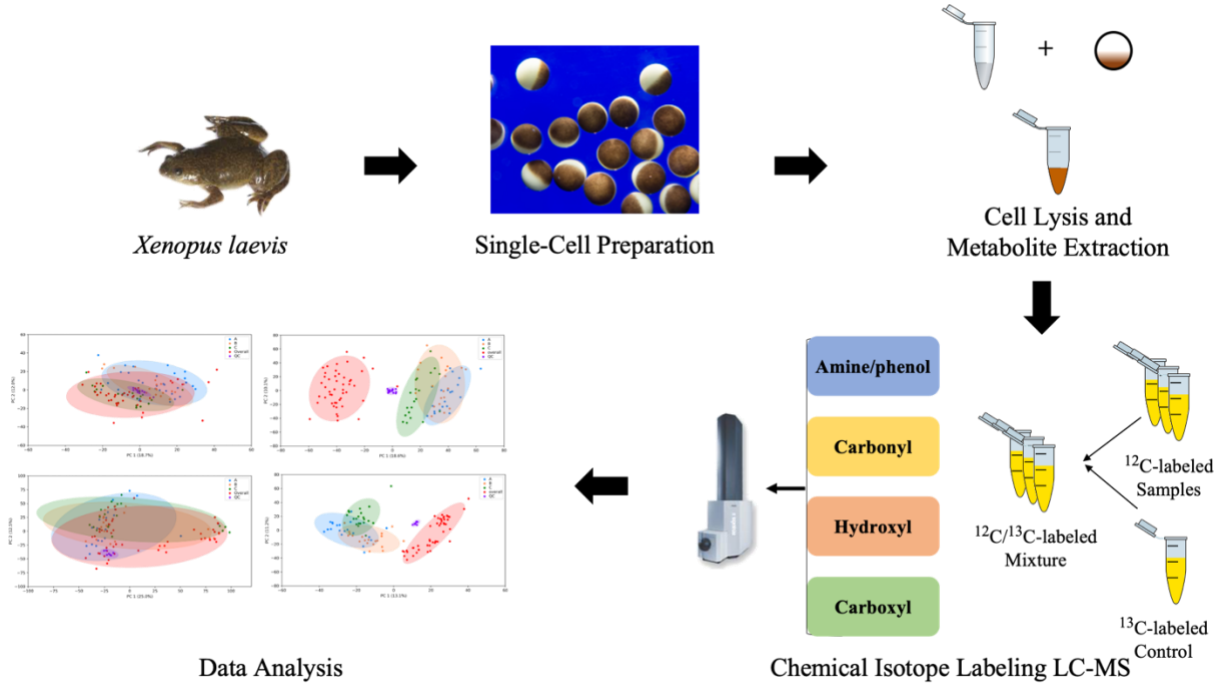


Figure 4. 1 Workflow of organ zonation study.

4.2 Experimental

4.2.1 Chemical and Reagents

In single-cell collection, tricaine methanesulfonate (MS-222) and Collagenase Type IA were purchased from Sigma-Aldrich Canada. OR-2 was prepared in the lab using 5 mM Hepes-NaOH (pH 7.8), 82.5 mM NaCl, 2.5 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM Na_2HPO_4[20]. Lysis solution was prepared with ACN, H_2O and MeOH purchased from Sigma-Aldrich Canada.

All chemicals and reagents used in CIL LC-MS, unless otherwise stated, were purchased from Sigma-Aldrich Canada. For dansylation labeling in amine and phenol metabolome profiling and base-activated dansylation in hydroxyl metabolome profiling, the ^{12}C-labeling reagent, dansyl chloride, was purchased from Sigma-Aldrich and the ^{13}C-labeling reagent was synthesized according to the published method previously[14]. For DmPA bromide labeling in carboxylic acid

metabolome profiling, the ^{12}C- and ^{13}C-labeling reagent, DmPA-Br, were synthesized according to the method published previously[18]. For DnsHz labeling in carbonyl metabolome profiling, the ^{12}C- and ^{13}C-labeling reagent were synthesized according to the method published previously[19].

4.2.2 Sample Collection and Preparation

We collected cells from two animals in this experiment. Ovary was first dissected out of the animal and arranged, so that anatomy is clear; *i.e.*, all lobes are evident. Cells were taken from the tip of the ovary lobe and from random locations in each animal. To be more specific, we first took photo of the ovary and then labeled locations we were going to collect cells as illustrated in Figure 4.2. We chose three lobes on each of the ovary and collected cells at the tip of the lobe using a Biopunch (core diameter: 7 mm, Agar Scientific, Essex, UK). Cells were taken from the tip of the ovary lobe because they are very distinct anatomical parts and easy to cut away from the rest of the organ. Two tips were cut from each lobe. For random sample, cells were collected from all locations; *i.e.*, cells from lobe tips, center of lobes and regions near the organ trunk. The random tissue pieces collected were pooled and processed together. Punches of tips and random tissue samples went into different tubes for collagenase digestion. Then, they were transferred to different tissue culture dishes for the sandpaper rolling. These two steps can help remove follicle cells. For each animal, 108 cells were collected. In order to make sure all cells were being lysed and homogenized at about the same time; we have four individuals involved in preparing cells. More specifically, three people were responsible for collecting and processing cells from three different lobes, one person was responsible for the random sample as depicted in Figure 4.3. This is necessary for minimizing the delay between processing the first cell and the last cell in this study. 10 cells were selected from each tip and 24 cells were selected for the random sample population

under the microscope. The cells were then being washed with OR-2. Each group of cells was photographed, using a camera set up on the dissecting microscope so the diameter of individual cell can be measured if needed later. Every single-cell was then being dispensed onto the side of an individual tube containing 80 µL of lysis solution; *i.e.*, 2:2:1 acetonitrile:methanol:water, ready on an ice bucket. After that, the cell was being lysed to extract metabolites by putting the tube on a high speed vortexer for 5 s in 4 ℃ cold room. Finally, we froze the lysed cell in liquid nitrogen immediately.

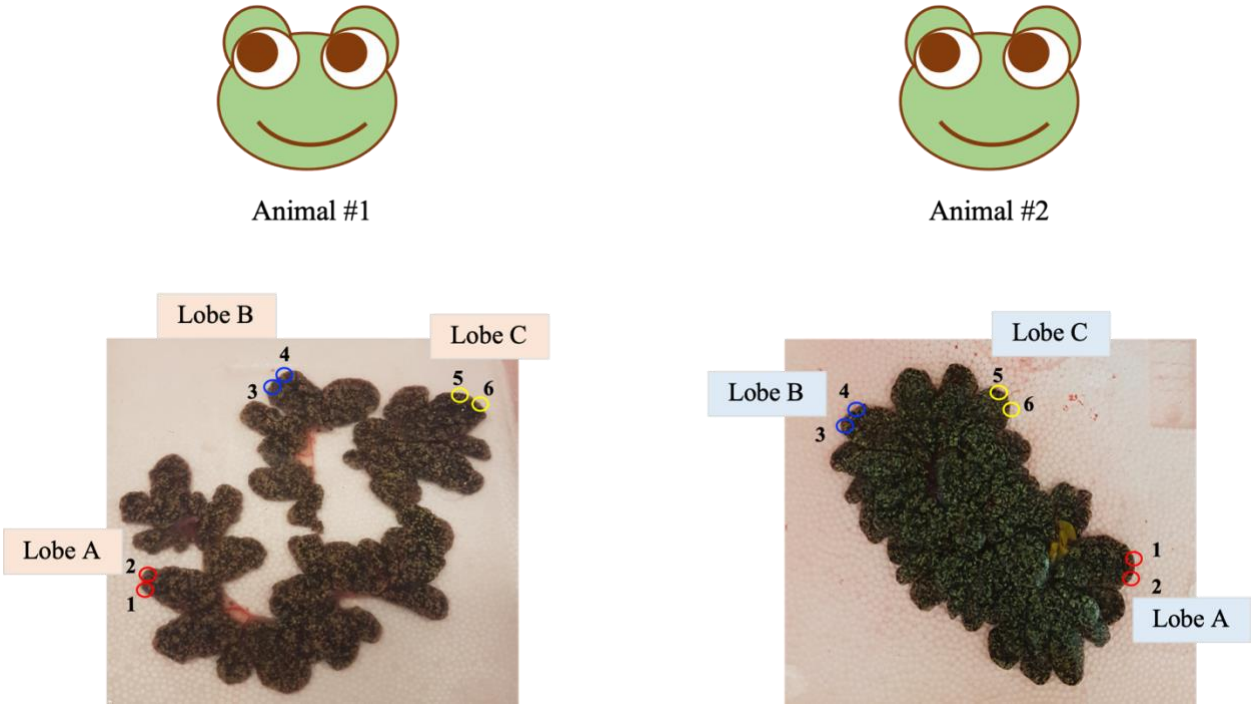


Figure 4. 2 Photography of the ovaries with punch locations indicated.

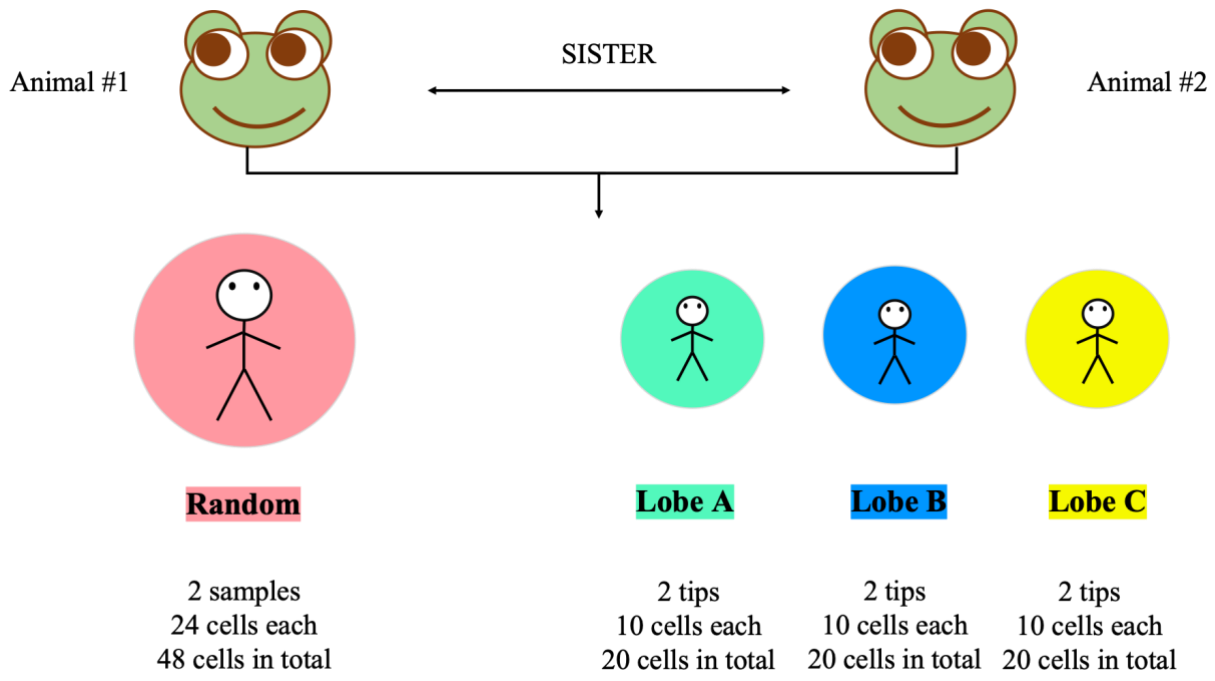


Figure 4. 3 Workflow of single cells collection in organ zonation study.

4.2.3 Chemical Isotope Labeling

Lysates were thawed and divided into four equal portions, 18 µL aliquots in each tube. Each aliquot was used for one submetabolome profiling. For the control sample, we employed universal internal standard (UIS), which was prepared by collecting and lysing a large number of cells from *Xenopus laevis*. Details of CIL in each submetabolome profiling are as follows:

For amine and phenol submetabolome profiling, sample was first dried down and resuspended in 5 µL of water. 2.5 µL of Na_2CO_3/$NaHCO_3$ buffer (250 mM in H_2O) and 2.5 µL of ACN were then added. The solution mixture was vortexed, spun down and mixed with 5 µL of freshly prepared ^{12}C-DnsCl (18 mg/mL in ACN). After being vortexed and spun down, the mixture was incubated at 40 ℃ for 1 h. 1 µL of NaOH (250 mM in H_2O) was then added and reacted at 40 ℃ for another 10 min to quench the excess DnsCl. Finally, 5 µL of formic acid (425 mM in 50/50

ACN/water) was added to consume the excess NaOH. For the labeling of UIS sample, the procedures described above was applied, with ^{13}C-DnsCl (18 mg/mL dissolve in ACN) being used instead.

For hydroxyl submetabolome profiling, sample was first dried down and resuspended in 5 µL of 3:1 ACN:H_2O. 5 µL of 4-dimethylaminopyridine (DMAP) (24.5 mg/mL in ACN) and 8 µL of ^{12}C-DnsCl (18 mg/mL in ACN) were then added. The solution mixture was vortexed, spun down and allowed to react at 60 ℃ for 1 h. 1 µL of NaOH (250 mM in H_2O) was then added and reacted at 60 ℃ for another 10 min to quench the excess DnsCl. Finally, 5 µL of formic acid (425 mM in 50/50 ACN/ H_2O) was added to consume the excess NaOH. ^{13}C-DnsCl (18 mg/mL dissolve in ACN) was used for the labeling of UIS sample.

For carbonyl submetabolome profiling, sample was first dried down and resuspended in 5 µL of water. 5 µL of HCl (144 mM in MeOH) and 5 µL of ^{12}C-DnsHz (20 mM dissolved in MeOH) were added, and the reaction mixture was vortexed and spun down before reacted at 40 ℃ for 1 h. The sample mixture was then being put in -80 ℃ freezer for 10 min to stop the labeling reaction. Finally, the mixture was dried down and re-dissolved in 16.7 µL of 50/50 ACN/ H_2O. ^{13}C-DnsHz (20 mM dissolved in MeOH) was used for the labeling of UIS sample.

For carboxylic acid submetabolome profiling, sample was first dried down and resuspended in 5 µL of 3:1 ACN:H_2O. 2 µL of TEOA (100 mM in ACN) and 5 µL of ^{12}C- DmPA-Br (10 mg/mL in ACN) were added. The mixture was vortexed, spun down and allowed to react at 80℃ for 1 h. 8 µL of triglycine (100 mM in H_2O) was added and reacted at 80℃ for 30 min, in order to quench the excessive labeling reagent. ^{13}C- DmPA-Br (10 mg/mL in ACN) was used to label the UIS sample.

The individual sample was mixed with the UIS sample in 1:1 volume ratio to generate the $^{12}C/^{13}C$-labeled mixture for LC-MS analysis. The final mixture was centrifuged for 10 min at 10,000 rpm under 4 ℃, before injecting to LC-MS.

4.2.4 LC-MS

LC-MS analysis was carried out on an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA) coupled to Bruker Impact HD Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA). Samples were injected to Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Mobile phase for the gradient elution was solvent A: 0.1% (v/v) formic acid in 5% (v/v) acetonitrile and solvent B: 0.1% (v/v) formic acid in acetonitrile.

For amine and phenol submetabolome profiling, the gradient elution profile 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, the gradient elution profile was: t = 0 min, 20 % B; t = 3.5 min, 35 % B; t = 9 min, 65 % B; t = 21 min, 98 % B; t = 34 min, 98 % B.

For carbonyl submetabolome profiling, the gradient elution profile was: t = 0 min, 1 % B; t = 3 min, 25 % B; t = 23 min, 98 % B; t = 34 min, 98 % B.

For carboxylic acid submetabolome profiling, the gradient elution profile 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 = 42 min, 98 % B.

The flow rate in these four channels was 180 μL/min. All mass spectra were collected in positive ion mode and were recorded from 220 to 1000 m/z with a spectra acquisition rate of 1.0

Hz. The nebulizer was set to 1.0 bar, and the dry temperature was 230 ℃, with 8 L/min of drying gas. The capillary and end plate offset voltage were set as 4500 V and 500 V respectively.

^{12}C- and ^{13}C- labeled metabolites appeared as peak pairs and coeluted perfectly because there are only two isotope-labeled carbon atoms different between them. This differential chemical isotope labeling approach has been previously evaluated to be desirable for relative quantification of metabolites[15].

4.2.5 Data Processing and Metabolite Identification

All the raw LC-MS data was exported to a CSV file by Bruker Daltonics DataAnalysis software 4.3. Here, total ion chromatogram (TIC) and extracted ion chromatogram (EIC) were extracted from the raw data, de-noised and baseline corrected. The metabolomics data was then processed with IsoMS Pro, which is a software platform for interpretation of data obtained from CIL LC-MS based metabolomics studies. Data quality check ensures high quality data by performing mass accuracy and retention time check. Data processing involves peak pair picking from raw files, alignment of peak pair ratios among all samples and zero-filling, which retrieves missing values by searching through the raw data. Data cleansing involves blank treatment, missing value treatment and sample wise normalization so as to make sure the data is complete for metabolite identification and statistical analysis. Metabolite identification was performed using a three-tier metabolite identification approach. Definitive identification depends on Tier 1 - CIL library search. This is based on retention time and m/z by searching against CIL standard libraries on the basis of m/z tolerance window 10 ppm and retention time window of 30 s. High-confidence putative identification is based on retention time and m/z searches against the Tier 2 - LI library, which covers all common metabolism pathways containing endogenous metabolites and their

derivatives. The m/z window is 10 ppm, while the experimental retention time window is 30 s. Finally, putative identification is based on m/z matching against Tier 3 - mass-based database with monoisotopic mass window of 10 ppm. This library is composed of human endogeneous metabolites and their predicted metabolic products after one or two common metabolic reactions. Statistical analysis including univariate and multivariate analysis were also performed in IsoMS Pro.

4.3 Results and Discussion

4.3.1 Amine and phenol metabolomics

The chemical isotope labeling of amines and phenols containing metabolites resulted in the detection of 3227 and 4472 peak pairs in animal 1 and 2, respectively. We positively identified 132 metabolites by retention time and m/z search against CIL standard libraries, putatively identified 200 metabolites with high confidence by retention time and m/z match against the LI library and putatively identified 2359 metabolites by m/z match against the mass-based database in animal 1. While in animal 2, we positively identified 120 metabolites, putatively identified 240 metabolites with high confidence and putatively identified 3228 metabolites. Taken together, we identified 83% of detected metabolites in animal 1, 80% in animal 2.

For animal 1, the principal component analysis (PCA) score plot in Figure 4.4A shows very close clustering of quality control (QC) samples, this indicates high stability of the mass spectrometer during data acquisition and thus reliability and reproducibility of the measured data. PCA here shows lobe A and B group overlay on both first principal component (PC1) and second principal component (PC2). There is a slight separation between the lobe C and lobe A, B groups, also on both PC1 and PC2. In 2-dimensional PCA plot, PC1 represents the direction that contains

the most amount of variation between the metabolites, while PC2 describes the direction that contains the next most variation. The random group is overlay with lobe A, B and C groups in PCA. Figure 4.5A displays the partial least square-discriminant analysis (PLS-DA) plot generated from the amine and phenol metabolome data set of cells in animal 1. There are some overlaps between three lobe groups and a small separation between lobe groups and the random group, with R^2 (goodness of fit) and Q^2 values (predictability value) of 0.8314 and 0.6535 respectively. Using a 100-permuation test, the PLS-DA model generated here passed the validation test with the original Q^2 significantly larger than permuted values as shown in Appendix Figure A4.1, which indicates there is no issue of over-fitting. These findings indicate similarities between cells of different lobes, and small differences between cells of lobes and from random locations of the ovary. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.6A-C. With the criterion of p-value < 0.05 (corresponding to q-value < 0.04 for both A/Random, B/Random and < 0.06 for C/Random), FC $\geq$ 1.5 and FC $\leq$ 0.67, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 1 and 2, respectively. And in order to find the unique metabolites of each lobe, Venn diagram was used to reveal the details. Figure 4.7A shows the Venn diagram of animal 1, the numbers in bold indicate the unique metabolites of each lobe, that is, there are 110 metabolites uniquely found in lobe A (Supplemental Table 3), 116 in lobe B (Supplemental Table 4) and 147 in lobe C (Supplemental Table 5). 36 metabolites were commonly found in three lobes. To further investigate the relationships of cells in three lobes, nonparametric analysis of variance (ANOVA) was applied. Figure 4.8A illustrates the ANOVA for amine and

phenol metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 1369 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 1857 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.2.

For animal 2, the PCA score plot in Figure 4.4B indicates close clustering of QC samples. Lobe A, B, and C groups overlap each other and are slightly separated from the random group on PC2. In the PLS-DA plot presented in Figure 4.5B, there is a separation between lobe groups and the random group on PC1, with R^2 and Q^2 values of 0.8895 and 0.7778, respectively. Using a 100-permuation test, the PLS-DA model generated here passed the validation test as presented in Appendix Figure A4.3. These results again suggest the likeness of cellular amine and phenol metabolome of different lobes and their small differences from random cells in animal 2. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.6D-F. With the criterion of p-value < 0.05 (corresponding to q-value < 0.03 for both A/Random and B/Random < 0.04 for C/Random), $FC \geq 1.5$ and $FC \leq 0.67$, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 6 and 7 respectively. Figure 4.7B shows the Venn diagram of animal 2, there are 254 metabolites uniquely found in lobe A (Supplemental Table 8), 60 in lobe B (Supplemental Table 9) and 155 in lobe C (Supplemental Table 10). 163 metabolites were commonly found in three lobes, these metabolites were compared with those in animal 1, and the shared features are shown in Supplemental Table 11. Figure 4.8B illustrates the ANOVA for amine

and phenol metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 1131 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 3340 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.4. The highlighted metabolite is the one also found to be most significant of three lobes in animal 1.

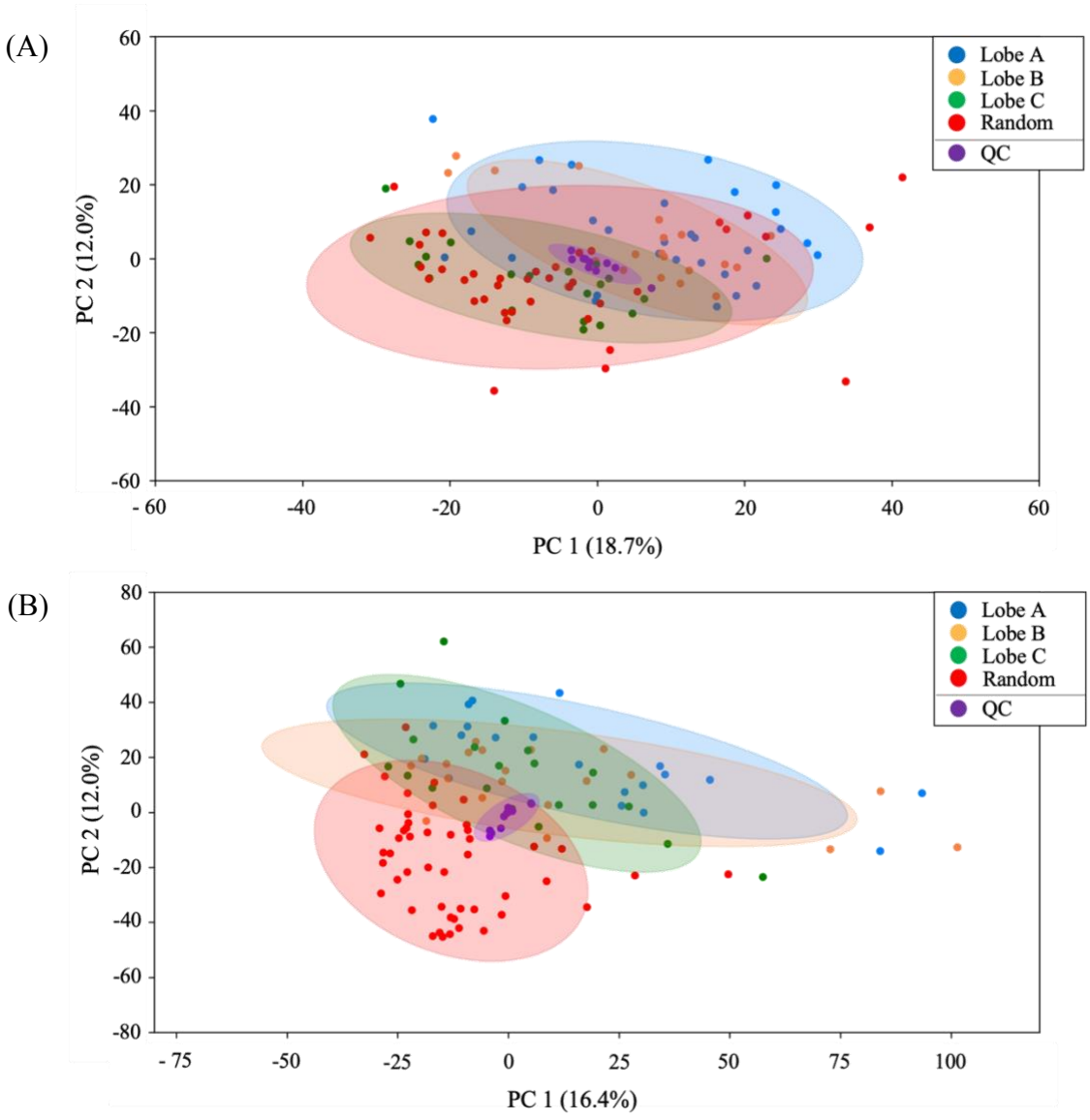


Figure 4. 4 PCA score plots of amine and phenol submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

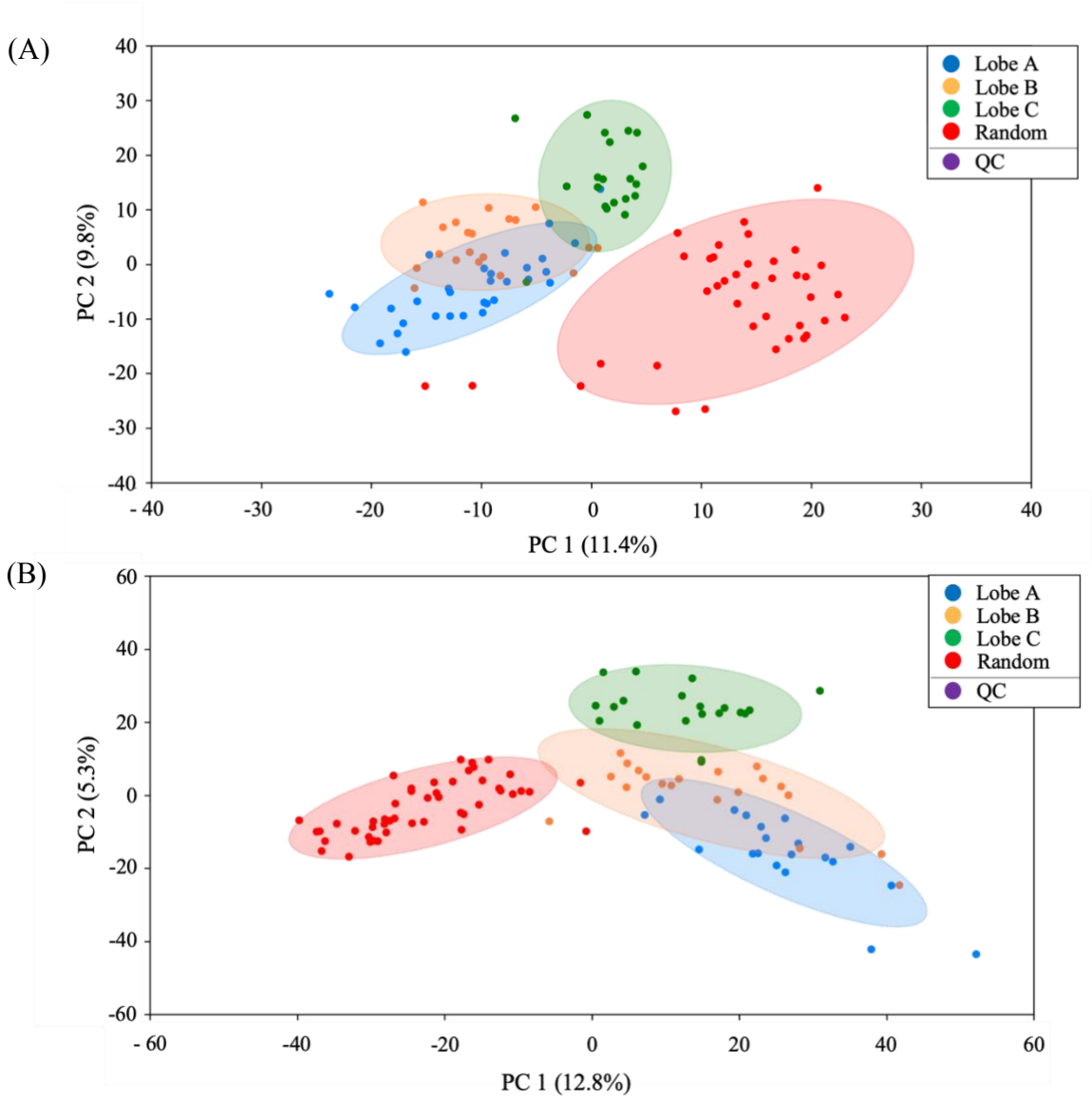


Figure 4. 5 PLS-DA score plots of amine and phenol submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

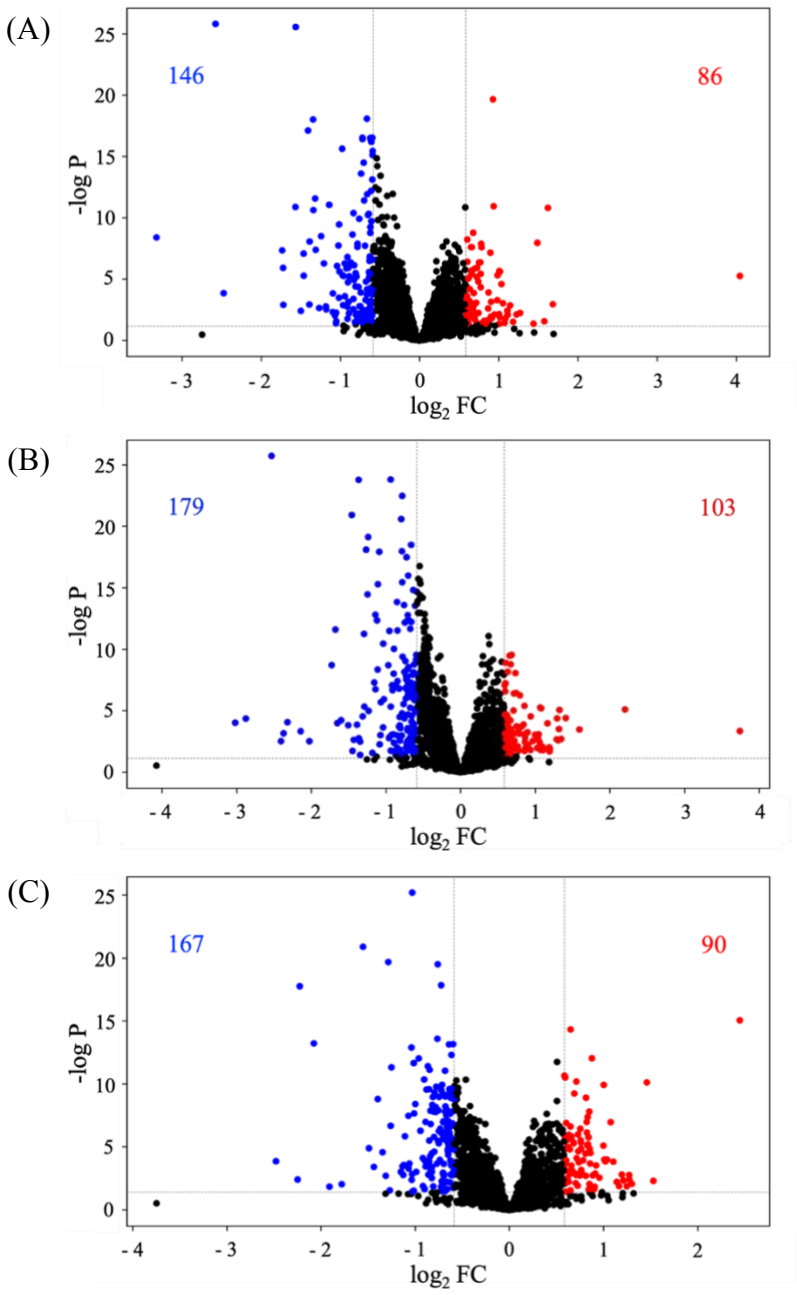


Figure 4. 6 Volcano plots comparing cells in **(A)** lobe A **(B)** lobe B and **(C)** lobe C with the random cell population in amine and phenol metabolome of animal 1, **(D)** lobe A **(E)** lobe B and **(F)** lobe C with the random cell population in amine and phenol metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.04 for (A) and (B) and q < 0.06 for (C)). FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

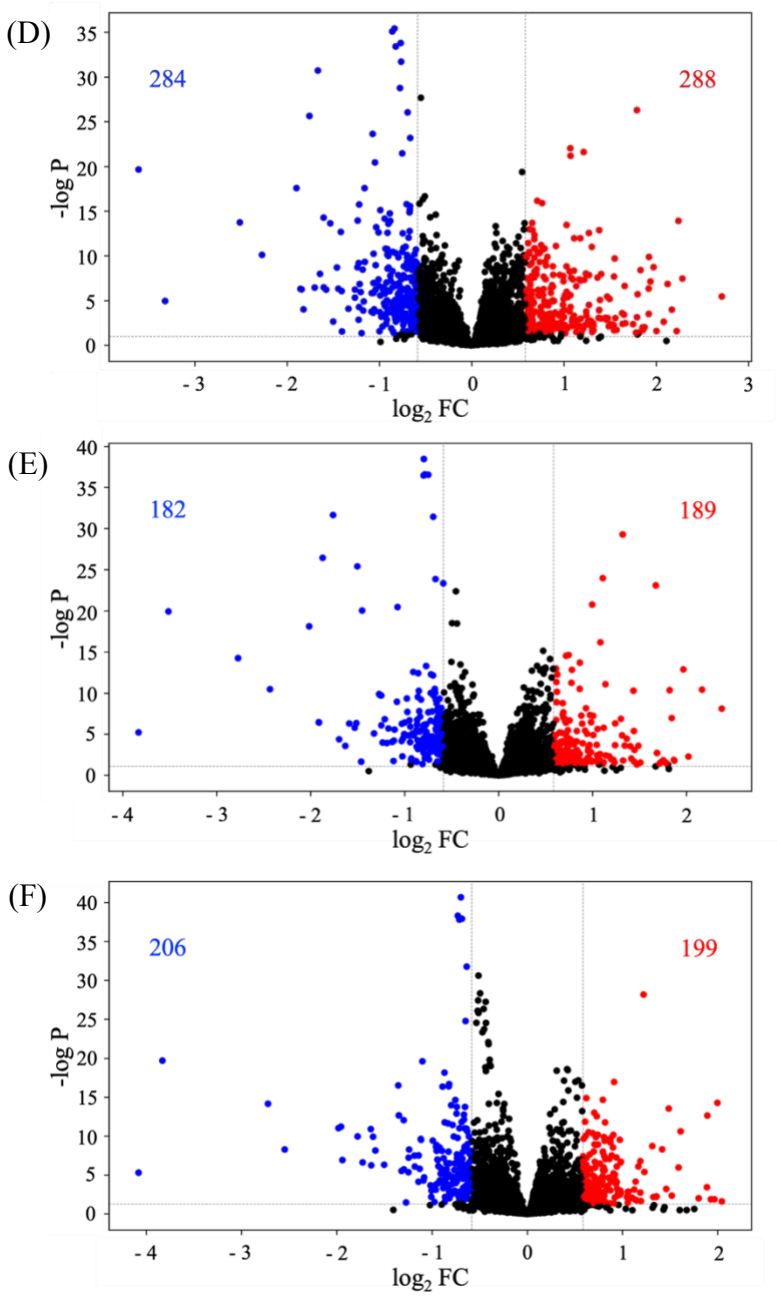
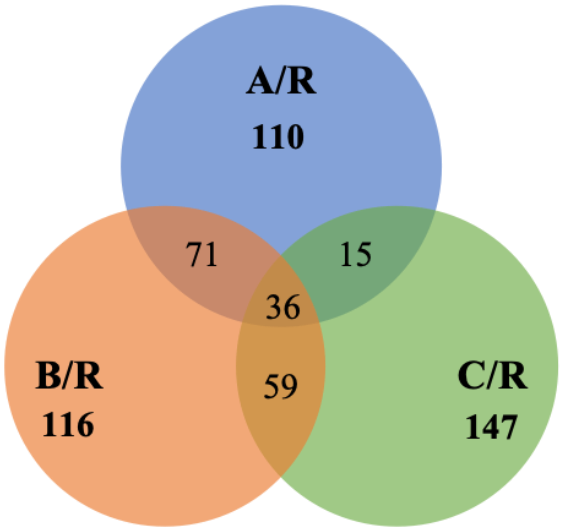


Figure 4.6 The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to $q < 0.03$ for (D) and (E) and $q < 0.04$ for (F)), FC $\geq$ 1.5 was determined as increase (red points) and FC $\leq$ 0.67 as decreased (blue point).

(A)



(B)

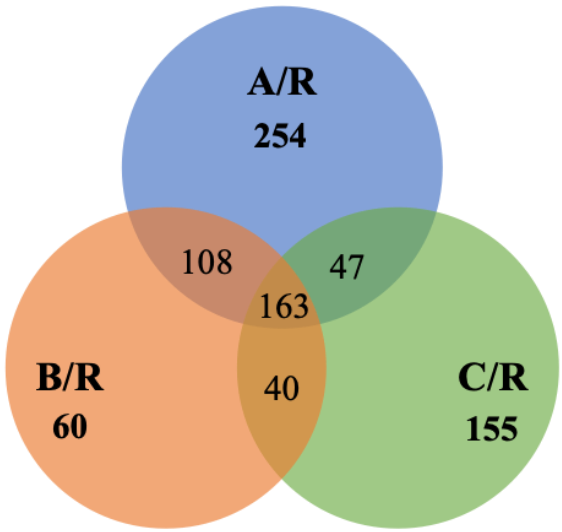


Figure 4.7 Venn diagrams showing significantly altered amines and phenols containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of **(A)** animal 1 and **(B)** animal 2.

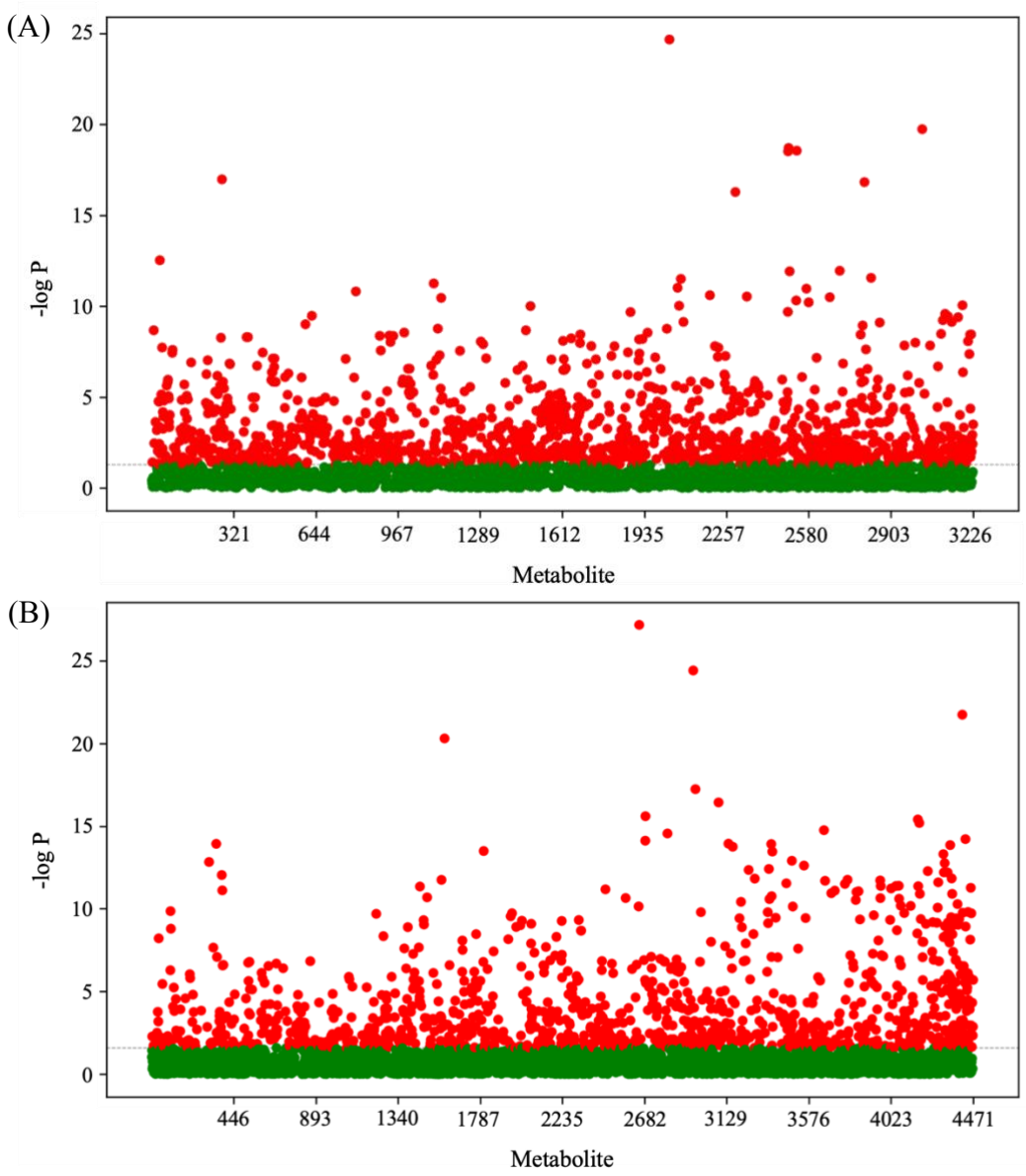


Figure 4. 8 ANOVA for amine and phenol metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). **(A)** In animal 1, the 1369 significant metabolites ($p < 0.05$) are shown in red, while 1857 non-significant metabolites ($p > 0.05$) are shown in green. **(B)** In animal 2, the 1131 significant metabolites ($p < 0.05$) are shown in red, while 3340 non-significant metabolites ($p > 0.05$) are shown in green.

4.3.2 Hydroxyl metabolomics

The chemical isotope labeling of hydroxyl containing metabolites resulted in the detection of 6538 and 5725 peak pairs in animal 1 and 2 respectively. We positively identified 21 and 23 metabolites, putatively identified 418 and 390 metabolites with high confidence and putatively identified 5058 and 4148 metabolites in animal 1 and 2 respectively. Taken together, we identified 84% of detected metabolites in animal 1, 80% in animal 2.

For animal 1, the PCA score plot in Figure 4.9A indicates very close clustering of QC samples. Lobe groups are overlay and there is an apparent visual separation between them and the random group on PC1. The PLS-DA plot in Figure 4.10A also shows a clear separation with R^2 and Q^2 values of 0.9162 and 0.8100, respectively. Moreover, there is also a separation between the lobe C group and A, B groups. With a 100-permutation test, the PLS-DA model passed the validation test with the original data much larger than the permutated data as displayed in Appendix Figure A4.5. The important finding here is that the cellular hydroxyl metabolome of different lobes is similar, however, they are very different from those cells collected at different points of the ovary. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.11A-C. With the criterion of p-value < 0.05 (corresponding to q-value < 0.02 for A/Random, B/Random and C/Random), FC $\geq$ 1.5 and FC $\leq$ 0.67, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 12 and 13 respectively. Figure 4.12A shows the Venn diagram of animal 1, the numbers in bold indicate the unique metabolites of each lobe. There are 295 metabolites uniquely found in lobe A (Supplemental Table 14), 262 in lobe B (Supplemental

Table 15) and 187 in lobe C (Supplemental Table 16). 753 metabolites were in common in three lobes. Figure 4.13A illustrates the ANOVA for hydroxyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 3609 significant altered metabolites ($p < 0.05$) of the three lobes are shown in red, while 2928 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.6.

For animal 2, the PCA score plot in Figure 4.9B indicates clustering of QC samples. There is an overlap between lobe A and B groups and a separation between them and the lobe C group on the PC1. The random group is sitting in the middle of A, B and C. The PLS-DA in Figure 4.10B shows a similar trend but with a separation between the random group and lobe groups. The plot had a R^2 and Q^2 values of 0.8997 and 0.7977, respectively. The model generated passed the validation test as indicated in Appendix Figure A4.7, with a 100-permutation. These results mainly reveal variances of cellular hydroxyl metabolome of lobe A, B with C in animal 2. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.11D-F, with the criterion of p-value < 0.05 (corresponding to q-value < 0.04 for both A/Random and C/Random < 0.03 for B/Random), FC ≥ 1.5 and FC ≤ 0.67, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 17 and 18 respectively. Figure 4.12B shows the Venn diagram of animal 2, there are 140 metabolites uniquely found in lobe A (Supplemental Table 19), 305 in lobe B (Supplemental Table 20) and 262 in lobe C (Supplemental Table 21). 102 metabolites are in common among three lobes, and these metabolites were compared with the those in animal 1, with

the shared feature summarized in Supplemental Table 22. Figure 4.13B illustrates the ANOVA for hydroxyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 3328 significant altered metabolites ($p < 0.05$) of the three lobes are shown in red, while 2396 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.8. The highlighted metabolite is the one also found to be most significant of three lobes in animal 1.

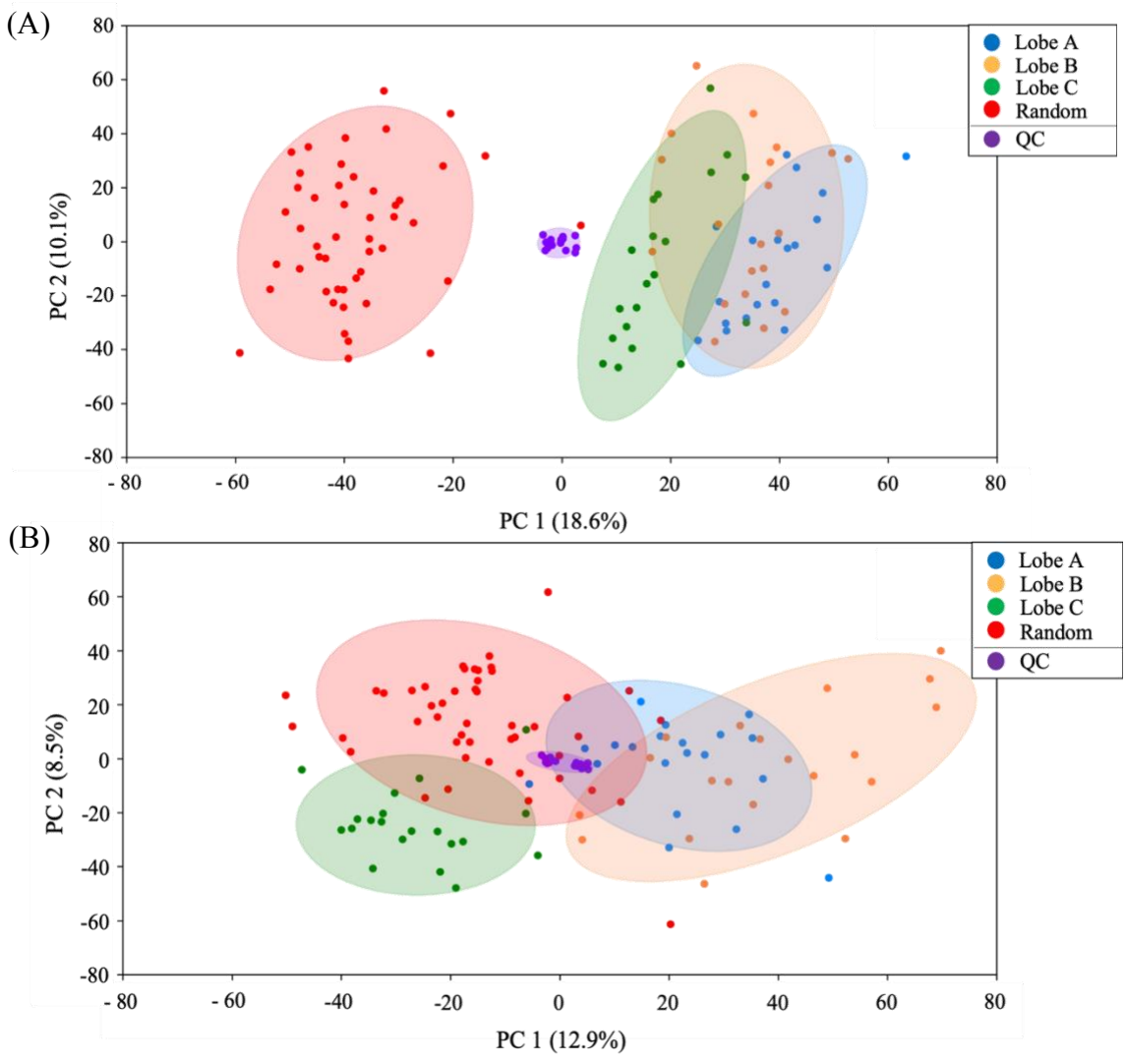


Figure 4. 9 PCA score plots of hydroxyl submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

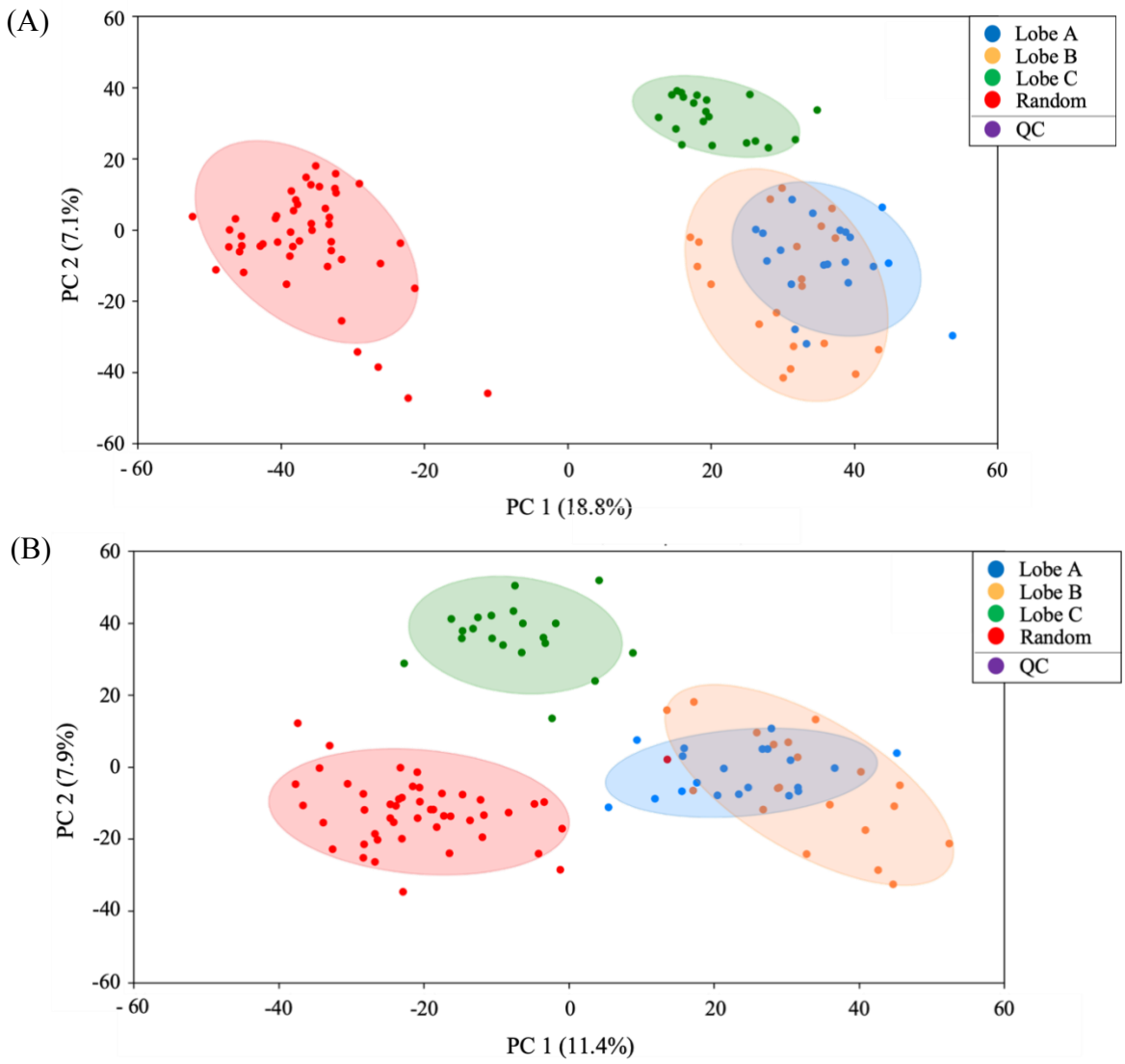


Figure 4. 10 PLS-DA score plots of hydroxyl submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

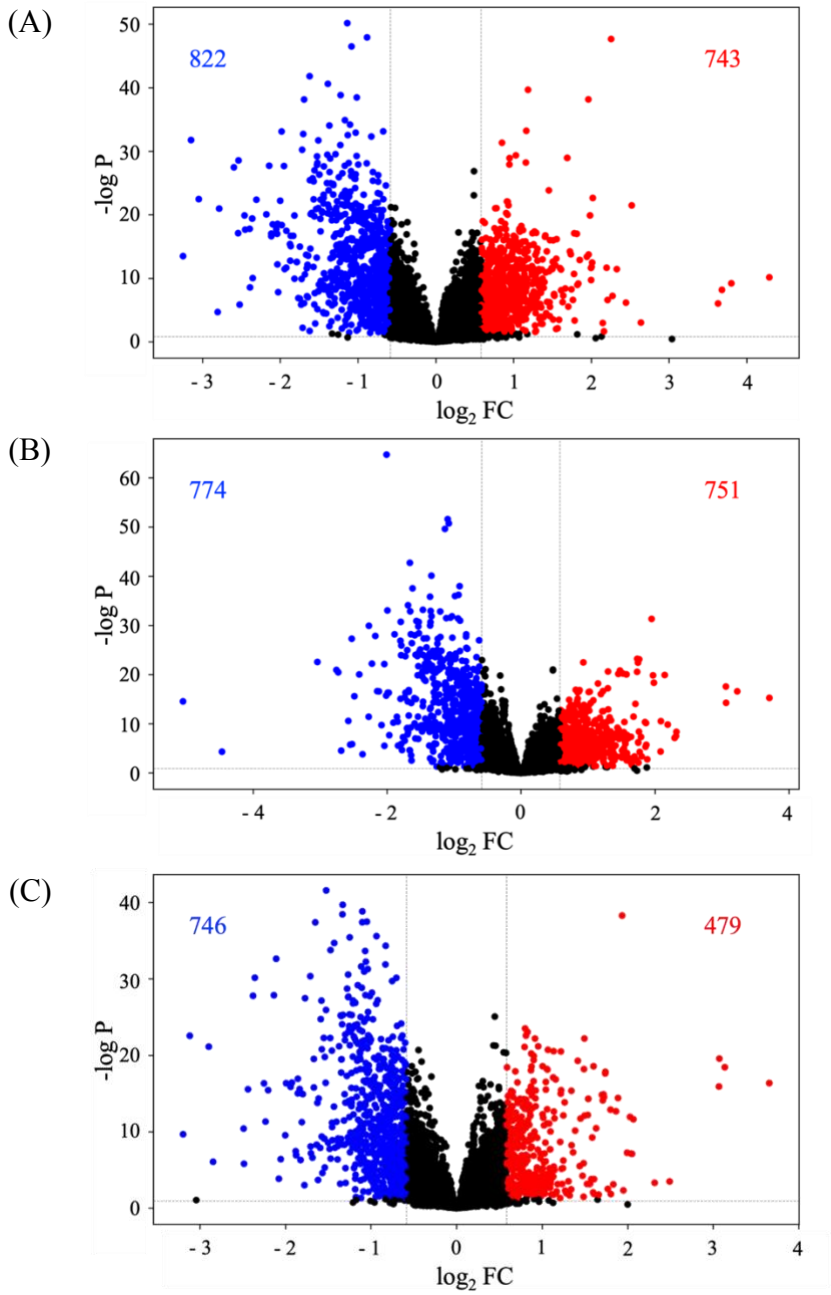


Figure 4. 11 Volcano plots comparing cells in **(A)** lobe A **(B)** lobe B and **(C)** lobe C with the random cell population in hydroxyl metabolome of animal 1, **(D)** lobe A **(E)** lobe B and **(F)** lobe C with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.02 for (A), (B) (C)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

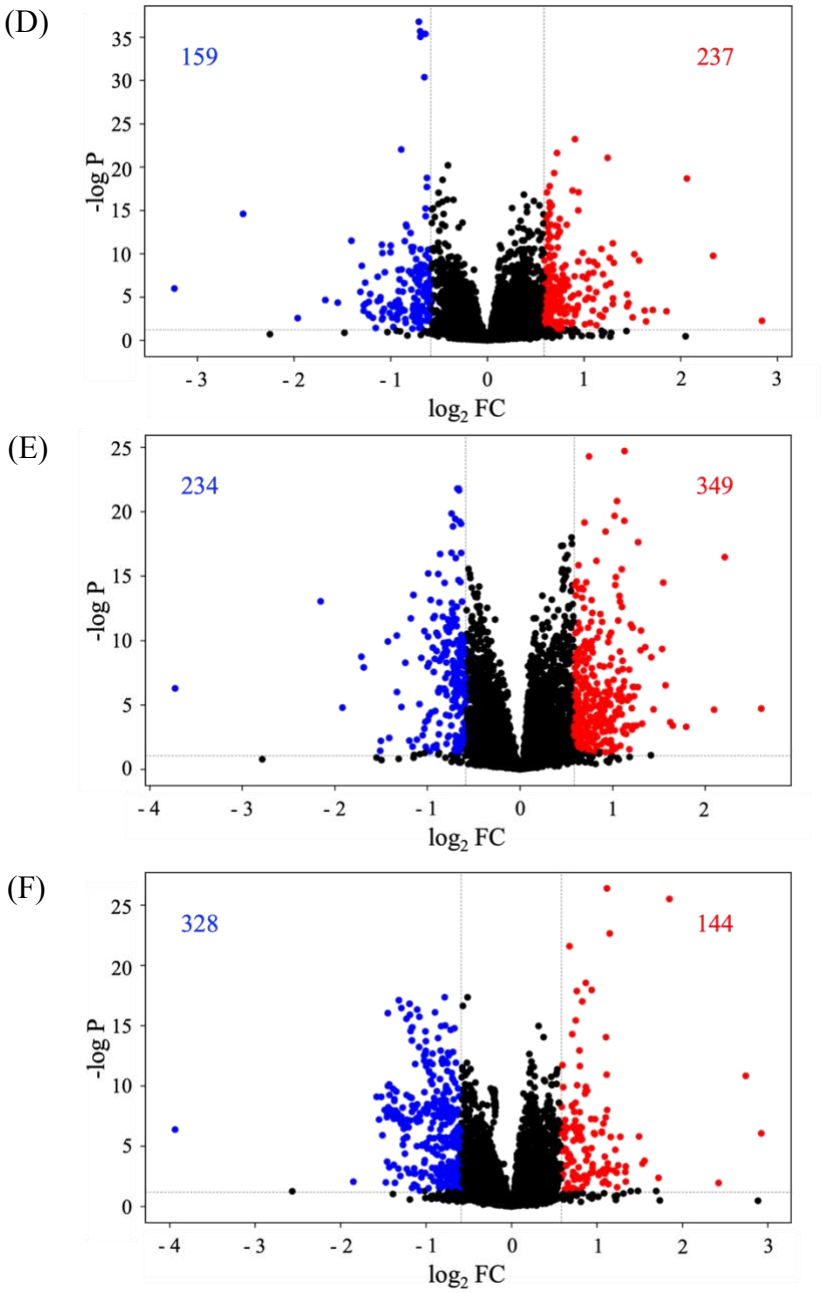


Figure 4.11 The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.04 for (D) and (F) and q < 0.03 for (E)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

(A)

(B)

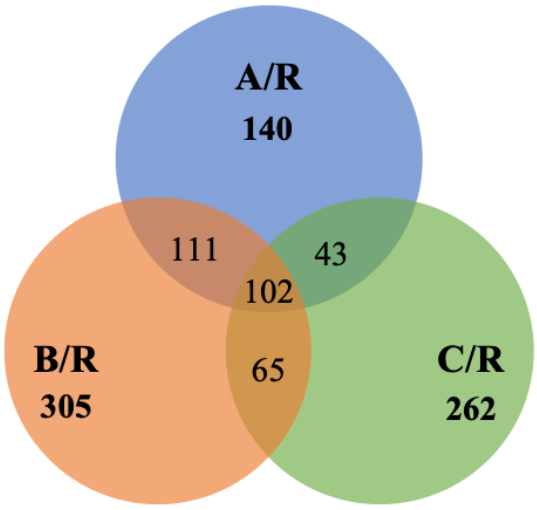


Figure 4. 12 Venn diagrams showing significantly altered hydroxyl containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of **(A)** animal 1 and **(B)** animal 2.

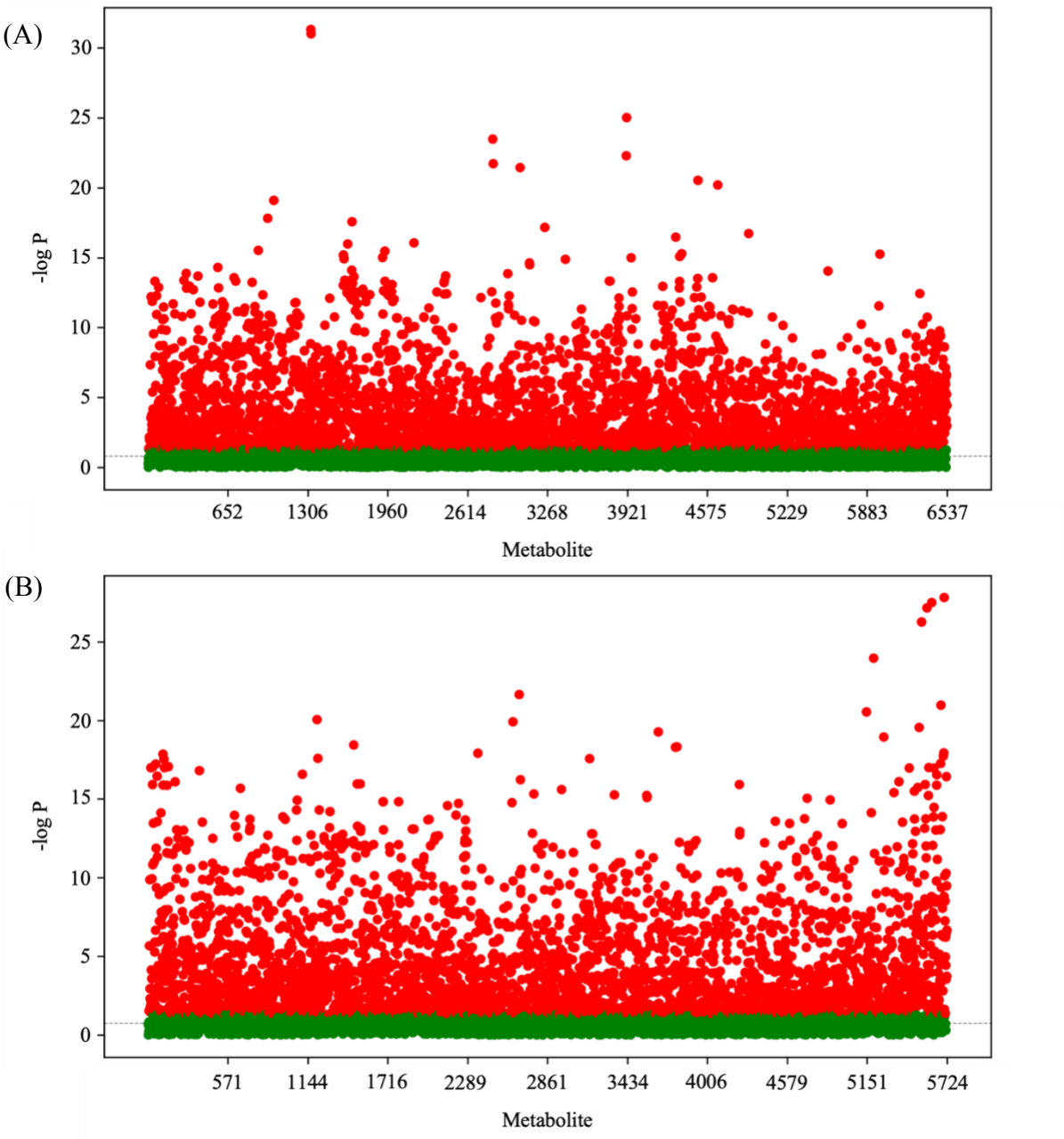


Figure 4. 13 ANOVA for hydroxyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). **(A)** In animal 1, the 3609 significant metabolites ($p < 0.05$) are shown in red, while 2928 non-significant metabolites ($p > 0.05$) are shown in green. **(B)** In animal 2, the 3328 significant metabolites ($p < 0.05$) are shown in red, while 2396 non-significant metabolites ($p > 0.05$) are shown in green.

4.3.3 Carbonyl metabolomics

The chemical isotope labeling of carbonyl containing metabolites resulted in the detection of 4417 and 5027 peak pairs in animal 1 and 2, respectively. We positively identified 53 and 55 metabolites, putatively identified 298 and 304 metabolites with high confidence and putatively identified 3311 and 3629 metabolites in animal 1 and 2, respectively. Taken together, we identified 83% of detected metabolites in animal 1, 80% in animal 2.

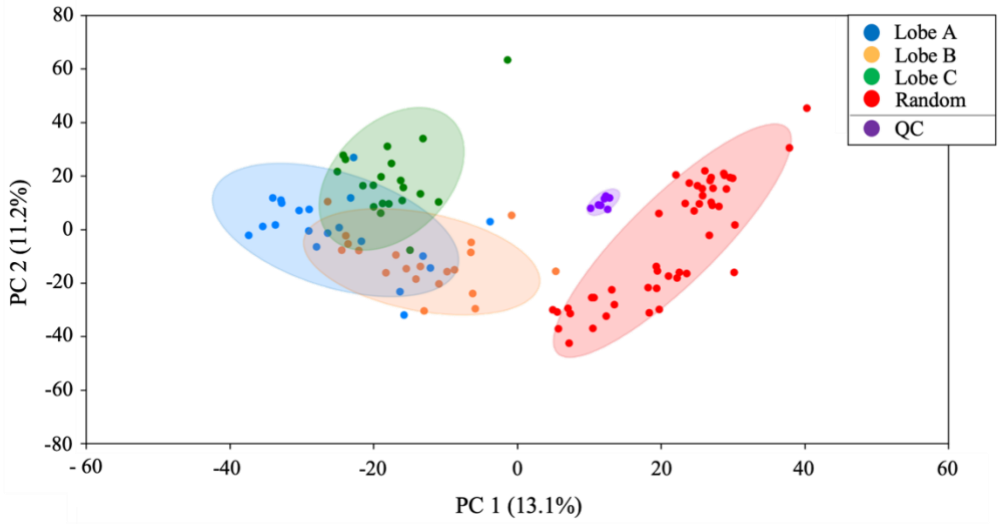
For animal 1, the PCA score plot in Figure 4.14A shows very close clustering of QC samples. There are overlaps of three lobe groups, with a clear separation between them and the random group on the PC1. The PLS-DA plot in Figure 4.15A also displays a clear separation with R^2 and Q^2 values of 0.9173 and 0.8053 respectively. Furthermore, there is also a separation between the lobe C group and A, B group. With a 100-permutation test, the PLS-DA model passed the validation test as shown in Appendix Figure A4.9. It is interesting to note that these results reveal similar findings observed in hydroxyl metabolomics in animal 1, in which the carbonyl metabolome of cells of lobe A, B and C are also similar and different greatly from the random population. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.16A-C. With the criterion of p-value < 0.05 (corresponding to q-value < 0.03, for both A/Random and C/Random < 0.05 for B/Random), FC $\geq$ 1.5 and FC $\leq$ 0.67, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 23 and 24, respectively. Figure 4.17A shows the Venn diagram of animal 1, the numbers in bold indicate the unique metabolites at each lobe. There are 183 metabolites uniquely found in lobe A (Supplemental Table 25), 88 in lobe B (Supplemental

Table 26) and 204 in lobe C (Supplemental Table 27). 235 metabolites were commonly found in three lobes. Figure 4.18A illustrates the ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 2006 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 2410 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.10.

For animal 2, the PCA score plot in Figure 4.14B indicates clustering of QC samples. There is overlap between lobe A and B and a separation between them with the lobe C group. The PLS-DA in Figure 4.15B shows a similar trend but with a separation between the random group and lobe groups. The plot had a R^2 and Q^2 values of 0.8926 and 0.6875, respectively. The model here passed the validation test as presented in Appendix Figure A4.11, with a 100-permutation. These results here reveal variances of cellular carbonyl metabolome between lobe A, B with C, similar to the observation of hydroxyl metabolome in animal 2. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.16D-F, with the criterion of p-value < 0.05 (corresponding to q-value < 0.05 for A/Random and < 0.03 for B/Random and C/Random), the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 28 and 29, respectively. Figure 4.17B shows the Venn diagram of animal 2, there are 129 metabolites uniquely found in lobe A (Supplemental Table 30), 158 in lobe B (Supplemental Table 31) and 347 in lobe C (Supplemental Table 32). 49 metabolites were commonly found in these lobes. These common metabolites were compared with those in animal 1, and the shared features in both animals are

summarized in Supplemental Table 33. Figure 4.18B illustrates the ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 2891 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 2067 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA here are summarized in box plots shown in Appendix Figure A4.12. The highlighted metabolites are those also found to be most significant of three lobes in animal 1.

(A)



(B)

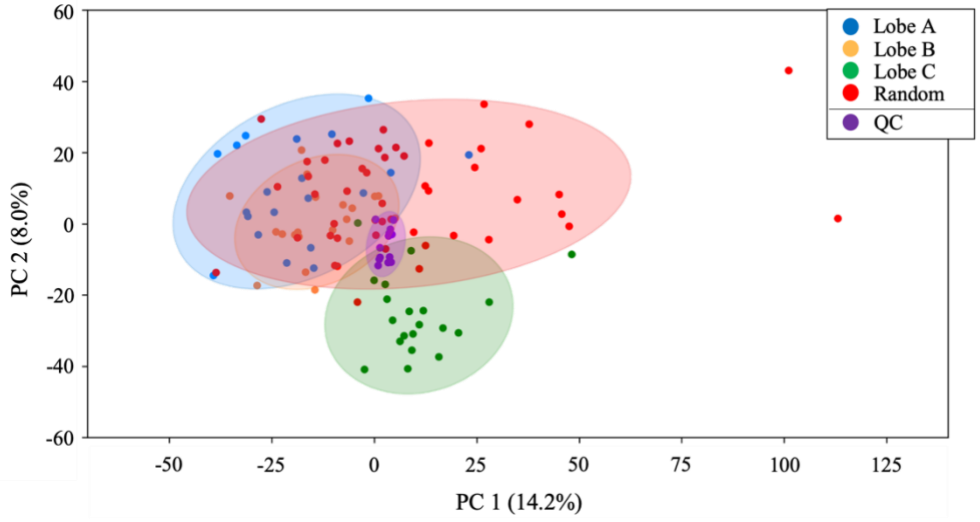


Figure 4. 14 PCA score plots of carbonyl submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

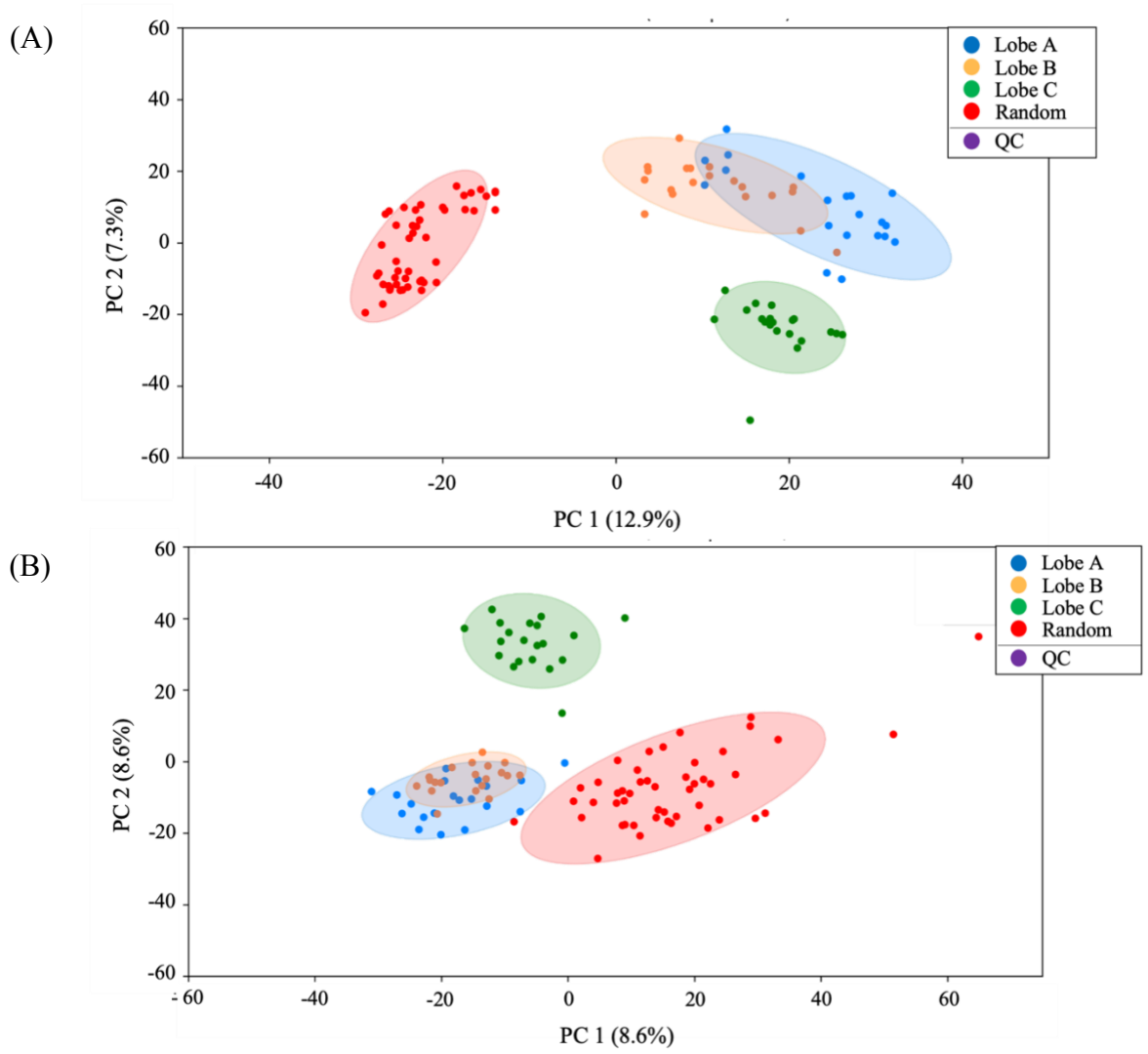


Figure 4. 15 PLS-DA score plots of carbonyl submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

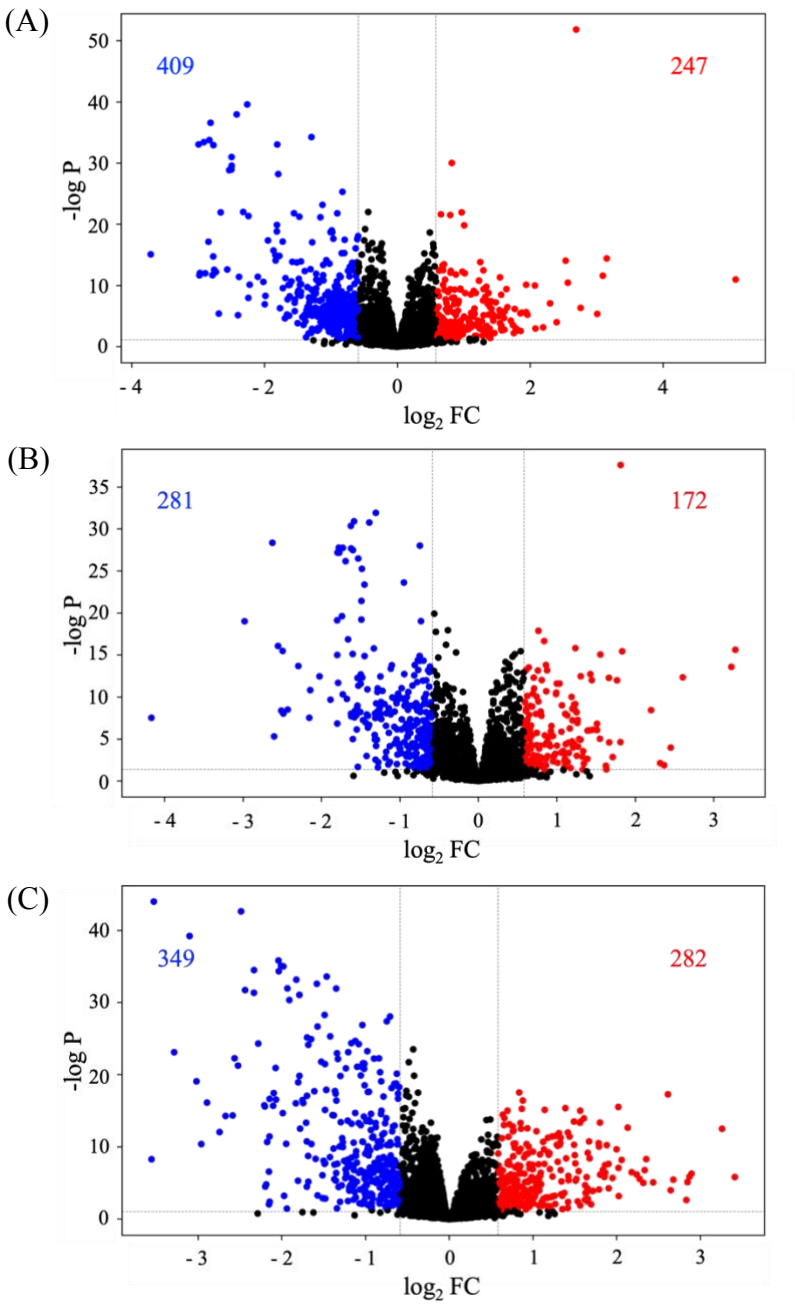


Figure 4. 16 Volcano plots comparing cells in **(A)** lobe A **(B)** lobe B and **(C)** lobe C with the random cell population in carbonyl metabolome of animal 1, **(D)** lobe A **(E)** lobe B and **(F)** lobe C with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.03 for (A) and (C), q < 0.05 for (B)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

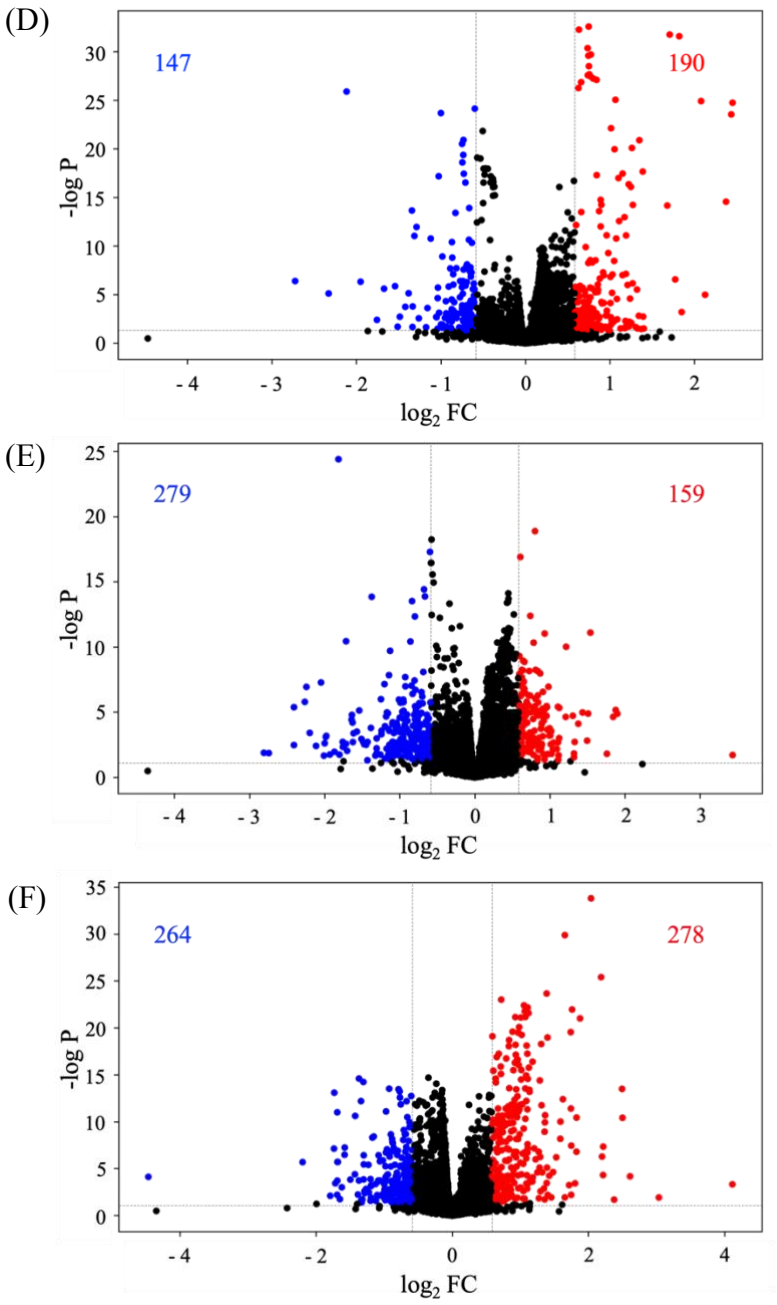
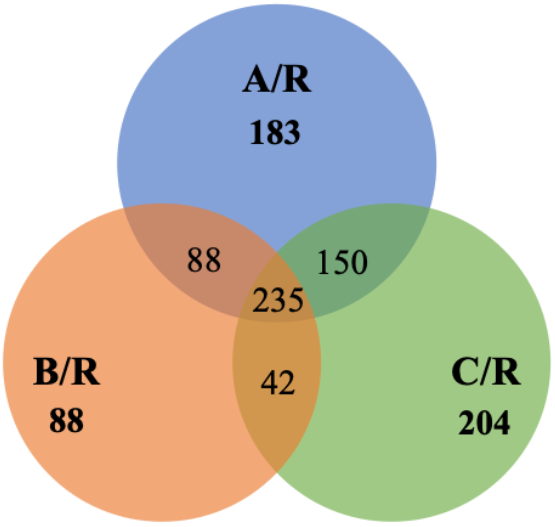


Figure 4.16 The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.05 for (D), q < 0.03 for (E) and (F)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

(A)



(B)

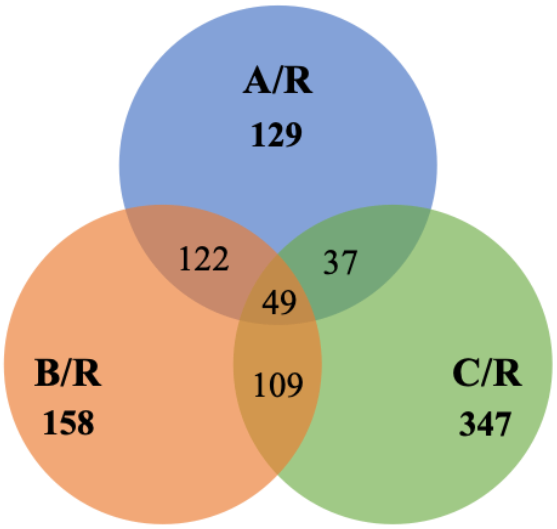


Figure 4. 17 Venn diagrams showing significantly altered carbonyl containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of **(A)** animal 1 and **(B)** animal 2.

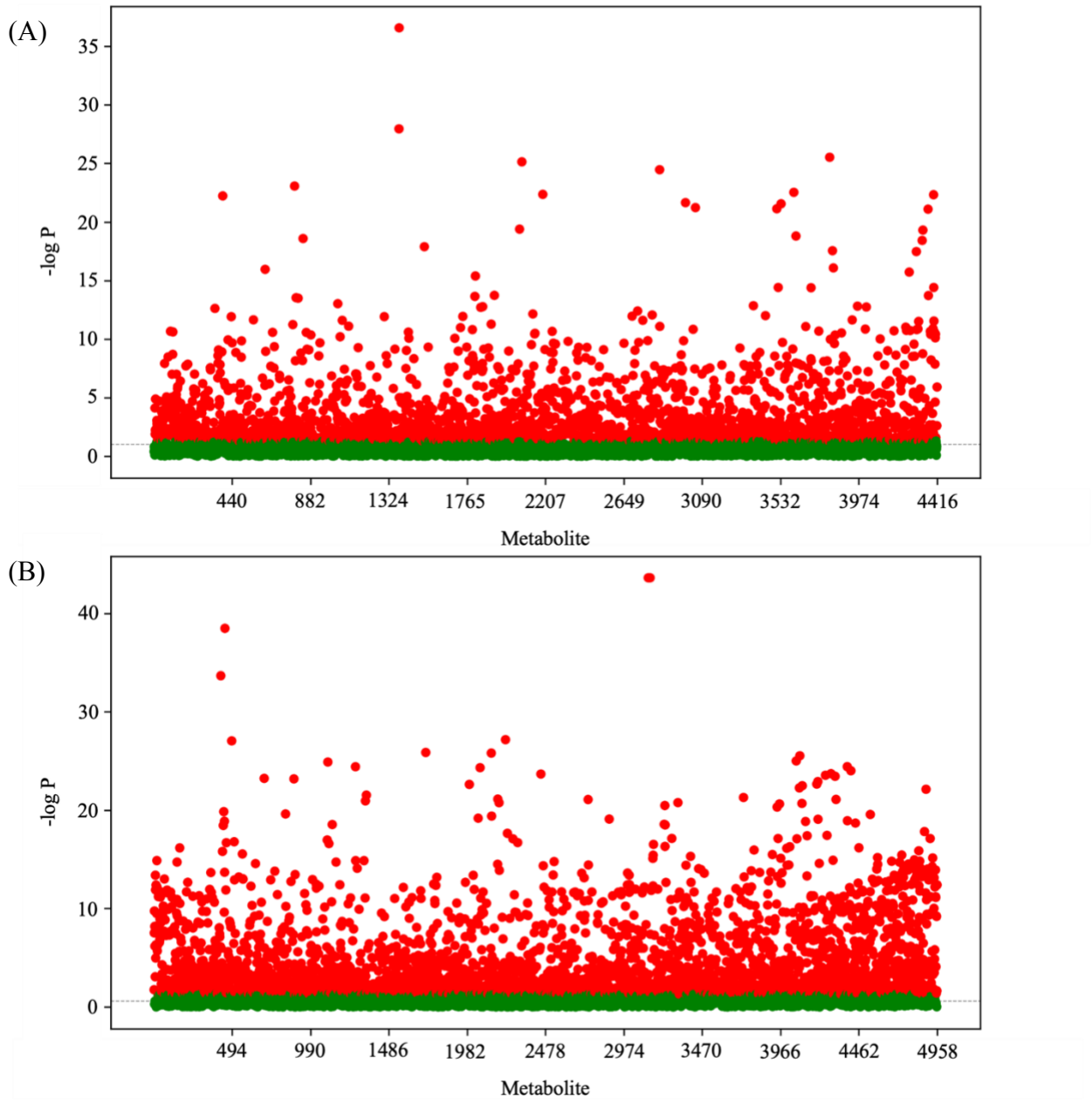


Figure 4. 18 ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). **(A)** In animal 1, the 2006 significant metabolites ($p < 0.05$) are shown in red, while 2410 non-significant metabolites ($p > 0.05$) are shown in green. **(B)** In animal 2, the 2891 significant metabolites ($p < 0.05$) are shown in red, while 2067 non-significant metabolites ($p > 0.05$) are shown in green.

4.3.4 Carboxylic acid metabolomics

The chemical isotope labeling of carboxylic acid containing metabolites resulted in the detection of 3672 and 3223 peak pairs in animal 1 and 2, respectively. We positively identified 72 and 74 metabolites, putatively identified 95 and 123 metabolites with high confidence and putatively identified 2775 and 2349 metabolites in animal 1 and 2, respectively. Taken together, we identified 80% of detected metabolites in animal 1, 79% in animal 2.

For animal 1, the PCA score plot in Figure 4.19A shows very close clustering of QC samples. There are overlaps of three lobe groups, and there are no separations between them with the random group. The PLS-DA plot in Figure 4.20A displays stacking of lobe groups and limited separation of them from the random population with R^2 and Q^2 values of 0.7387 and 0.5898 respectively. With a 100-permutation test, the PLS-DA model passed the validation test as shown in Appendix Figure A4.13. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.21A-C. With the criterion of p-value < 0.05 (corresponding to q-value < 0.03 for A/Random , q-value < 0.05 for B/Random and q-value < 0.07 for C/Random), FC $\geq$ 1.5 and FC $\leq$ 0.67, the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 34 and 35, respectively. Figure 4.22A shows the Venn diagram of animal 1, the numbers in bold indicate the unique metabolites at each lobe. There are 454 metabolites uniquely found in lobe A (Supplemental Table 36), 121 in lobe B (Supplemental Table 37) and 192 in lobe C (Supplemental Table 38). 340 metabolites were commonly found in three lobes. Figure 4.23A illustrates the ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and

in lobe C (group C). 442 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 3189 non-significant metabolites ($p > 0.05$) are shown in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA are summarized in box plots shown in Appendix Figure A4.14.

For animal 2, the PCA score plot in Figure 4.19B indicates clustering of QC samples. There are overlaps of three lobe groups and a slight separation between them with the random group on the PC2. The PLS-DA in Figure 4.20B shows a similar trend with the separation between the random group and lobe groups being more obvious. The plot had a R^2 and Q^2 values of 0.7944 and 0.5928, respectively. The model here passed the validation test as presented in Appendix Figure A4.15, with a 100-permutation. Binary comparisons between cells of the lobe and random locations; *i.e.*, A/Random, B/Random and C/Random are presented in volcano plots shown in Figure 4.21D-F, with the criterion of p-value < 0.05 (corresponding to q-value < 0.04 for both A/Random and B/Random < 0.03 for C/Random), the number of significantly increased (shown in red dots) and decreased (shown in blue dots) metabolites are indicated in the plots. The common significantly increased and decreased metabolites in the comparison of A/Random, B/Random and C/Random are listed in Supplemental Table 39 and 40, respectively. Figure 4.22B shows the Venn diagram of animal 2, there are 65 metabolites uniquely found in lobe A (Supplemental Table 41), 180 in lobe B (Supplemental Table 42) and 371 in lobe C (Supplemental Table 43). 511 metabolites were commonly found in these lobes. These common metabolites were compared with those in animal 1, and the shared features in both animals are summarized in Supplemental Table 44. Figure 4.23B illustrates the ANOVA for carbonyl metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). 228 significant altered metabolites ($p < 0.05$) of three lobes are shown in red, while 3041 non-significant metabolites ($p > 0.05$) are shown

in green. Ten of the most significant metabolites, *i.e.*, with the lowest p-values found by ANOVA here are summarized in box plots shown in Appendix Figure A4.16. The highlighted metabolite is the one also found to be most significant of three lobes in animal 1.

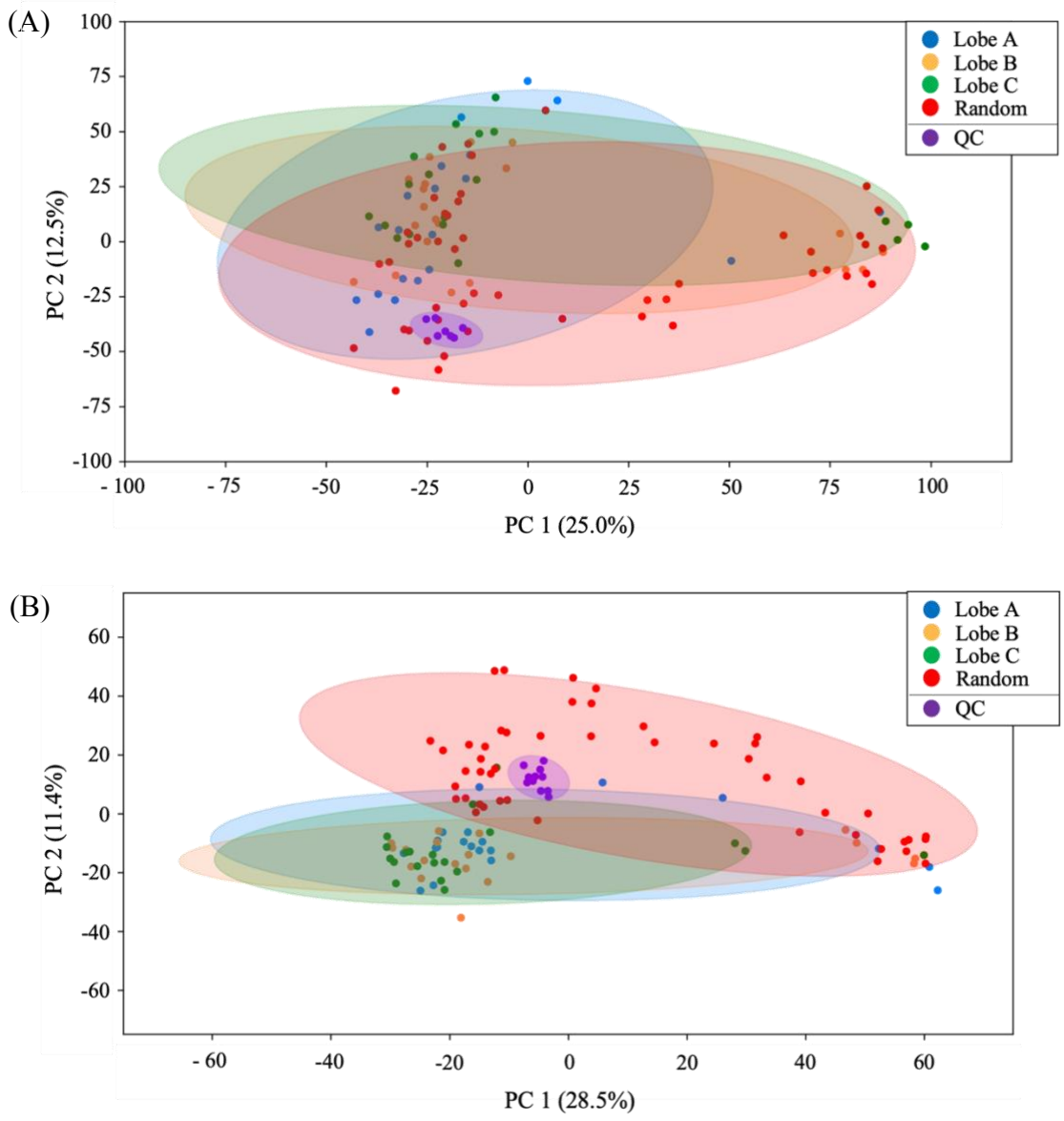


Figure 4. 19 PCA score plots of carboxylic submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

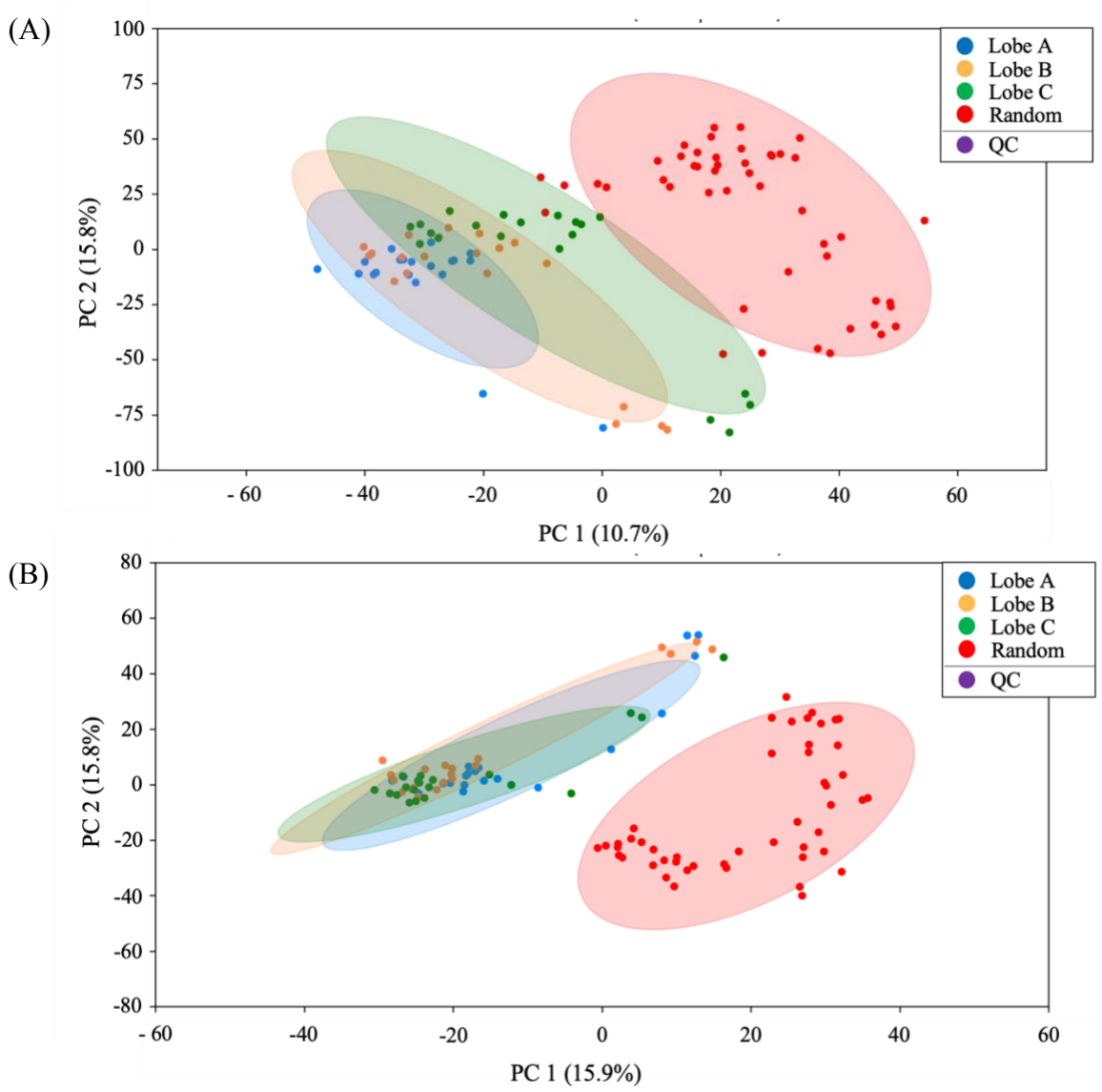


Figure 4. 20 PLS-DA score plots of carboxylic submetabolome of cells in **(A)** animal 1 and **(B)** animal 2.

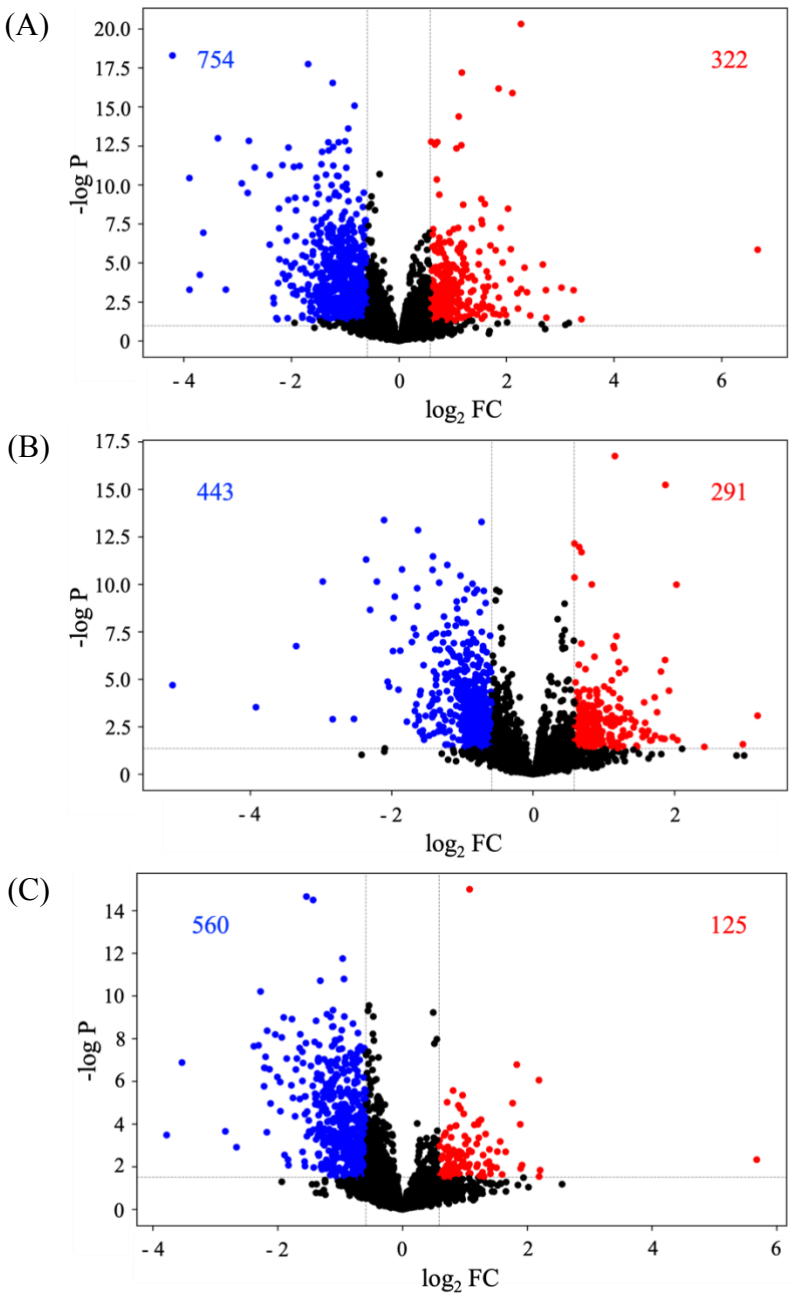


Figure 4. 21 Volcano plots comparing cells in **(A)** lobe A **(B)** lobe B and **(C)** lobe C with the random cell population in carboxylic metabolome of animal 1, **(D)** lobe A **(E)** lobe B and **(F)** lobe C with the random cell population in hydroxyl metabolome of animal 2. The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.03 for (A), q < 0.05 for (B) and q < 0.07 for (C)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point)

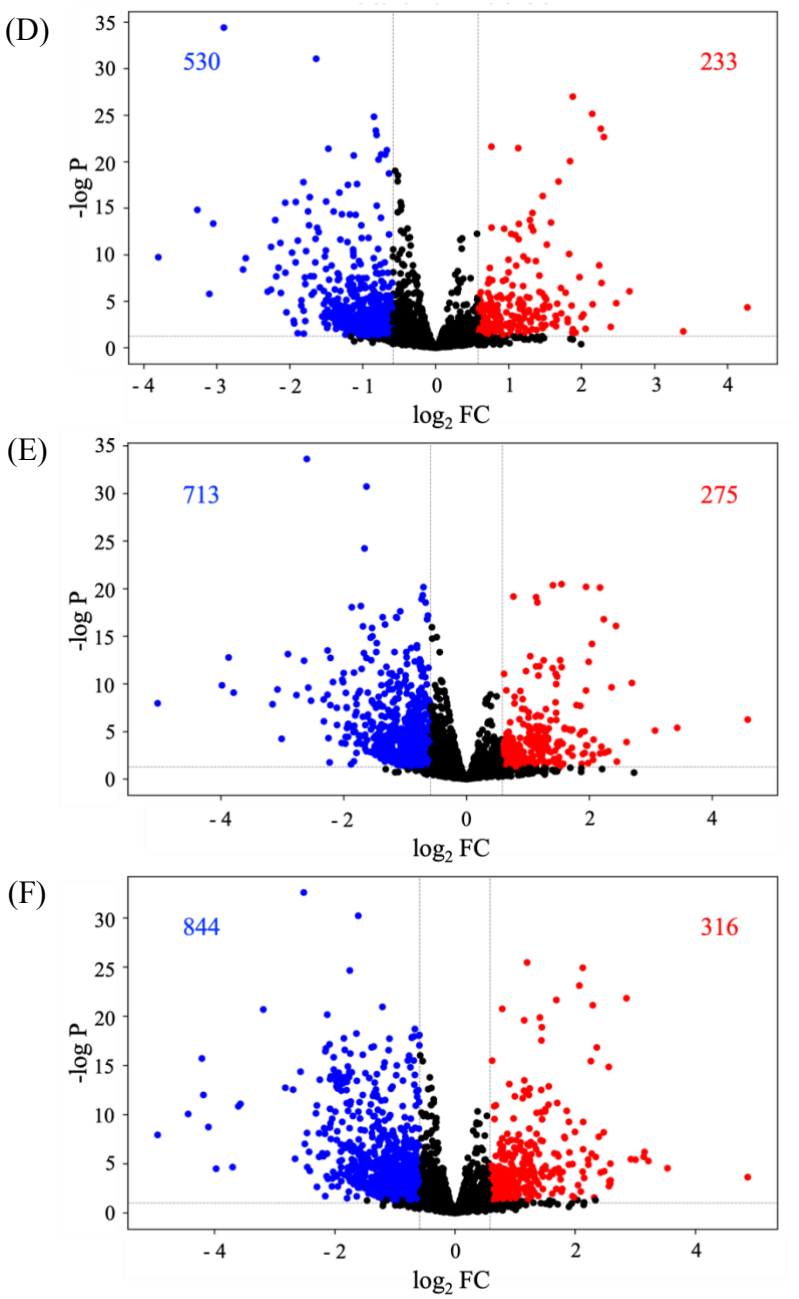
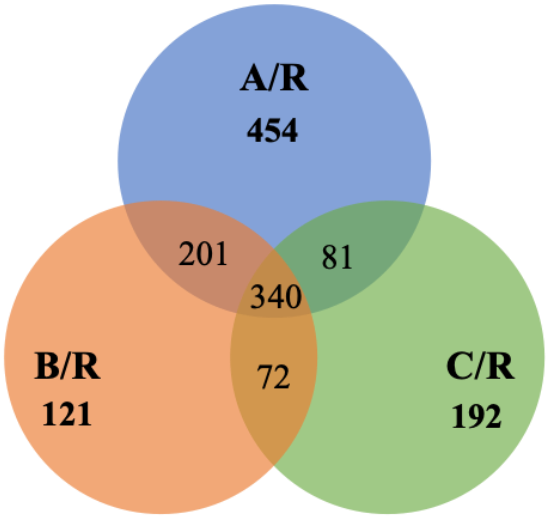


Figure 4.21 The criteria for determination of significant metabolites were as follows: p-value < 0.05 (corresponding to q < 0.04 for (D) and (E), q < 0.03 for (F)), FC ≥ 1.5 was determined as increase (red points) and FC ≤ 0.67 as decreased (blue point).

(A)



(B)

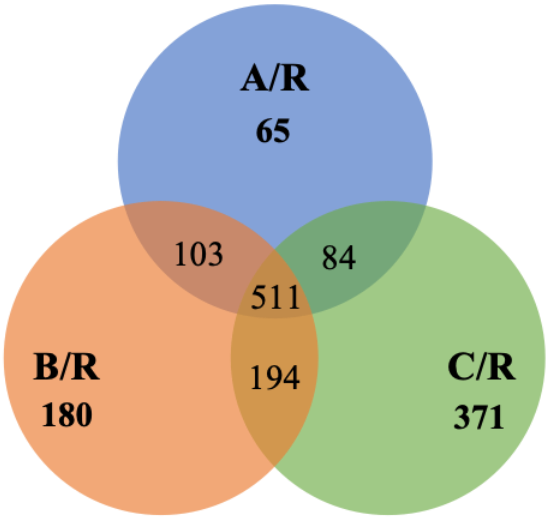


Figure 4. 22 Venn diagrams showing significantly altered carboxylic containing metabolites in lobe A verses random cell population (A/R), in lobe B verses random cell population (B/R) and in lobe C verses random cell population (C/R) of **(A)** animal 1 and **(B)** animal 2.

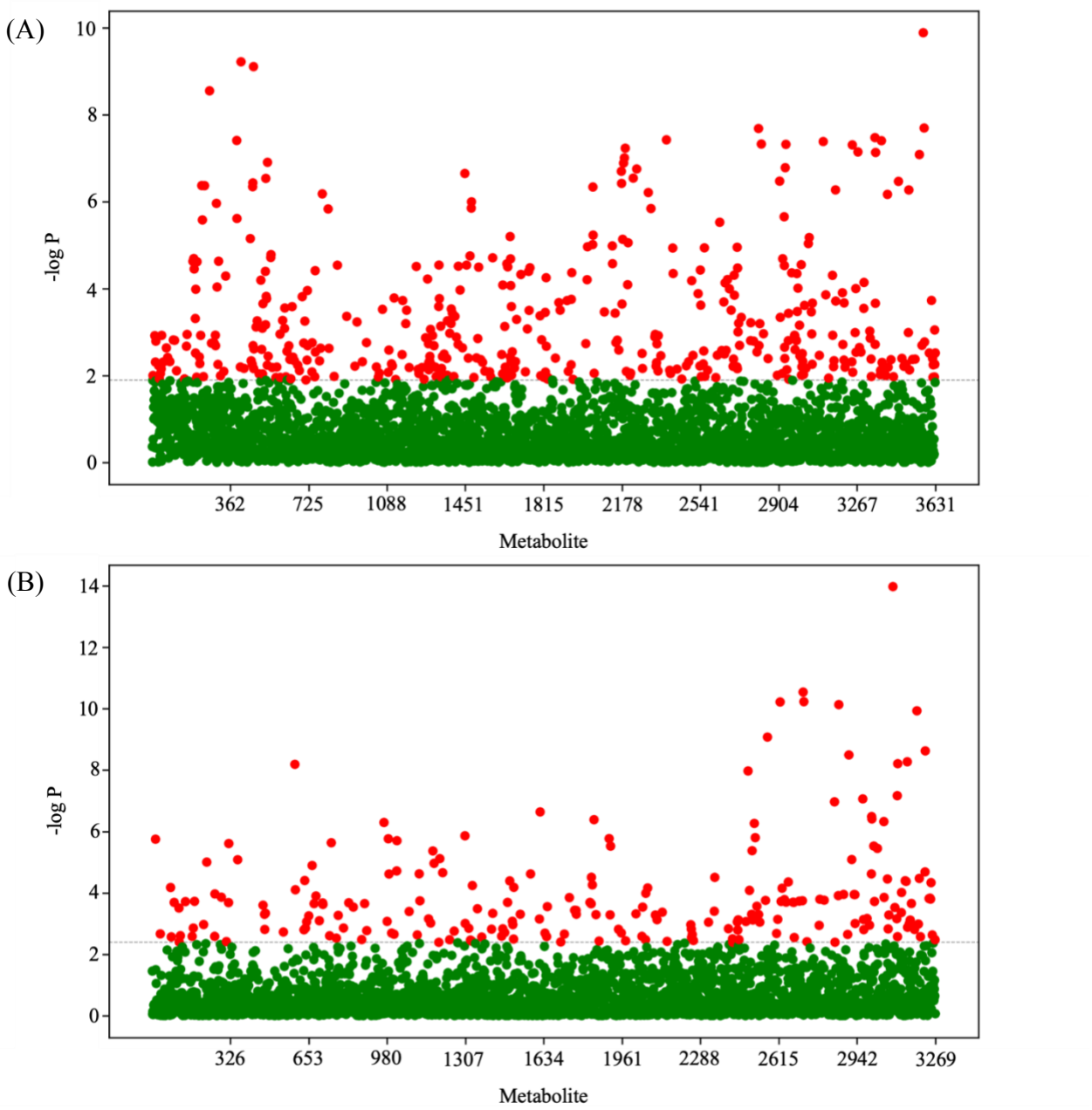


Figure 4. 23 ANOVA for carboxylic acid metabolomics of single cells in lobe A (group A); in lobe B (group B); and in lobe C (group C). **(A)** In animal 1, the 442 significant metabolites ($p < 0.05$) are shown in red, while 3189 non-significant metabolites ($p > 0.05$) are shown in green. **(B)** In animal 2, the 228 significant metabolites ($p < 0.05$) are shown in red, while 3041 non-significant metabolites ($p > 0.05$) are shown in green.

4.3.5 Cell-to-Cell Variability of Metabolite in Different Metabolic Pathway

As being emphasized before, the beauty of single-cell metabolomics is that the individual cell behavior can be revealed. Here, we looked into some important metabolic pathway and presented the variations of individual cells in some of the related metabolites. First of all, for glycolysis (https://www.genome.jp/pathway/map00010+C00031), two related metabolites including glucose and pyruvate are shown in Figure 4.24. The box plot illustrated the cell-to-cell variability of cells in each lobe and the random population, and the number below each group represents the percent relative standard deviation (%RSD). The starting metabolite of the glycolytic pathway, glucose, has RSD of 18% to 37% in animal 1 in the lobes and 9%-16% in animal 2. However, cell-to-cell variability of its downstream product, pyruvate, become very low, all below 10%. This can be very important, because this may provide insights on what is happening during glycolysis so that the cell-to-cell variability of the downstream metabolite become very low. It is also worth noting that the two subpopulations of cells in lobe A of animal 1 is originated from two tips of the lobe, but moving to pyruvate, these subpopulations does not exist.

Secondly, cell-to-cell variation of some related metabolites in pentose phosphate pathway (https://www.genome.jp/pathway/map00030+C00577) is displayed in Figure 4.25. The results showed that glyceraldehyde and D-arabino-hex-3-ulose 6-phosphate have large extent of variation between cells at different lobes and random locations, but one of the metabolites, sedoheptulose 7-phosphate is much less compared to the other two identified metabolites.

Lastly, the TCA cycle (https://www.genome.jp/pathway/map00020+C00036) shown in Figure 4.26, the cell-to-cell variation of all the identified metabolites is high. Although these insights do not have immediate biomedical applications, it does illuminate a fundamental way in which cell works. The purpose of presenting the cell-to-cell variation of metabolites in different

metabolic pathways here is not intended to highlight the fact the well-known fact that heterogeneity exists in cells of the same population; rather it focuses on the themes that CIL LC-MS is capable and important for revealing the single-cell differences in metabolomics.

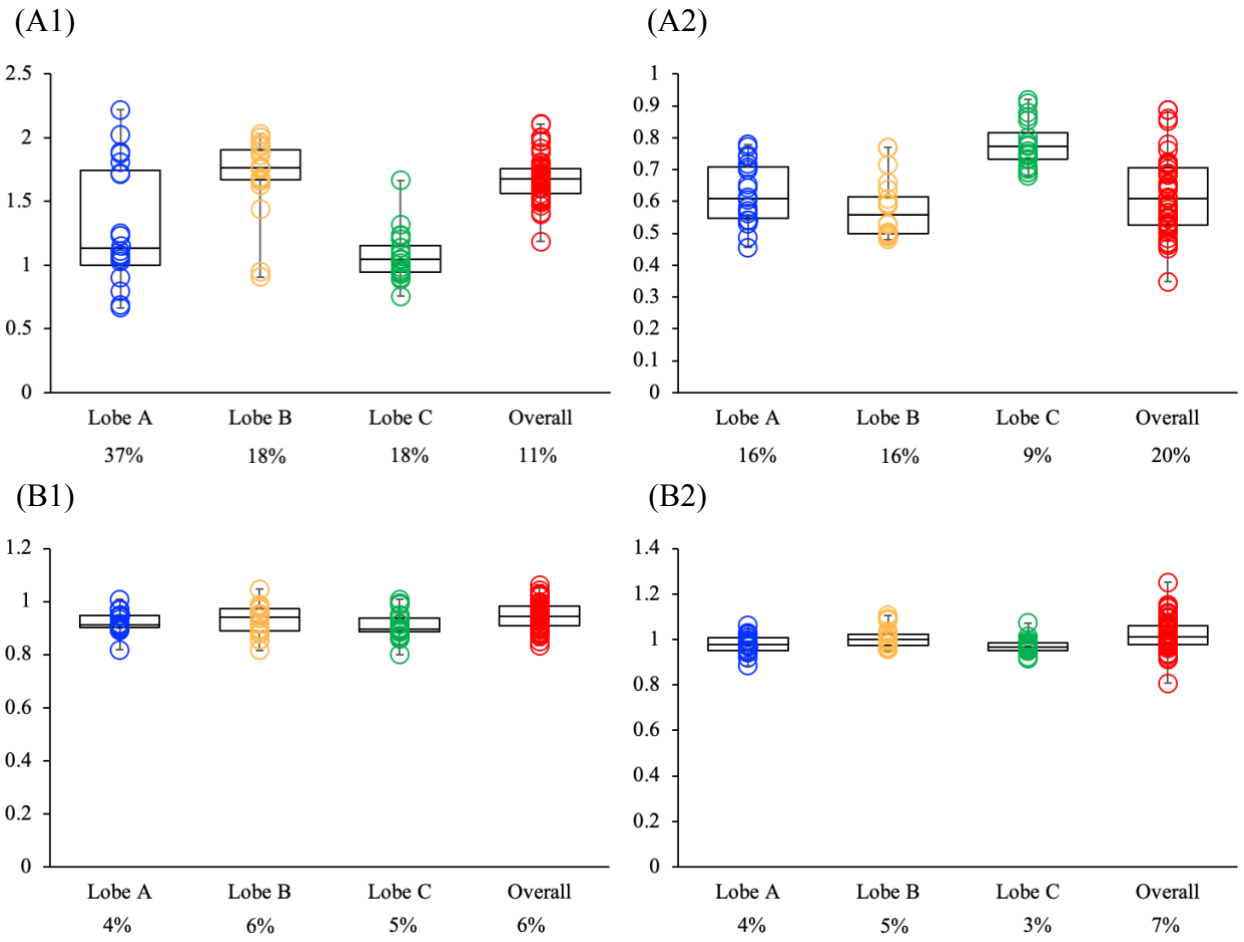


Figure 4. 24 Box plots of showing cell-to-cell variability of **(A)** glucose and **(B)** pyruvate in different cell populations of animal 1 and 2.

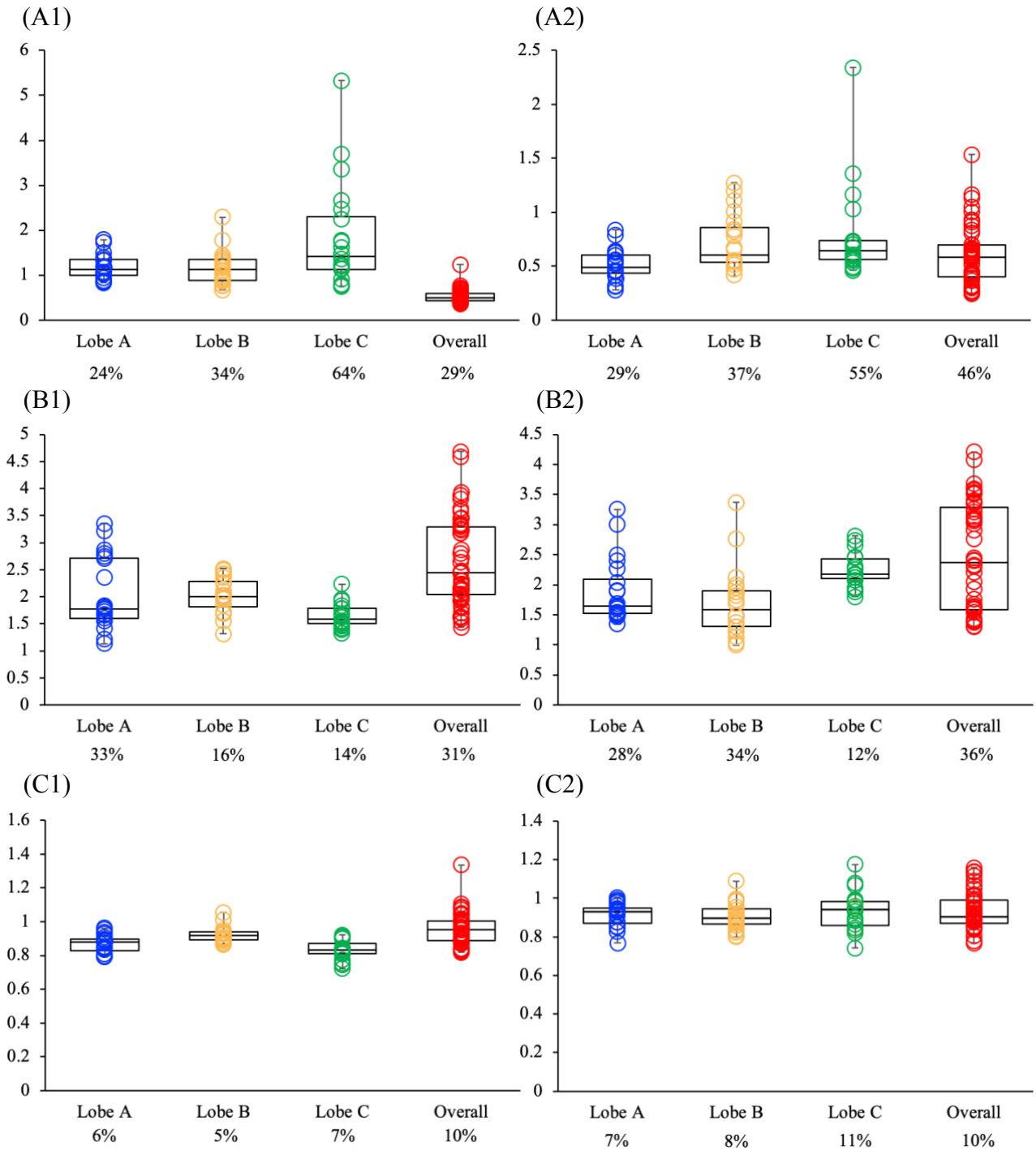


Figure 4. 25 Box plots showing cell-to-cell variability of **(A)** glyceraldehyde, **(B)** D-arabino-hex-3-ulose 6-phosphate and **(C)** sedoheptulose 7-phosphate in different cell populations of animal 1 and 2.

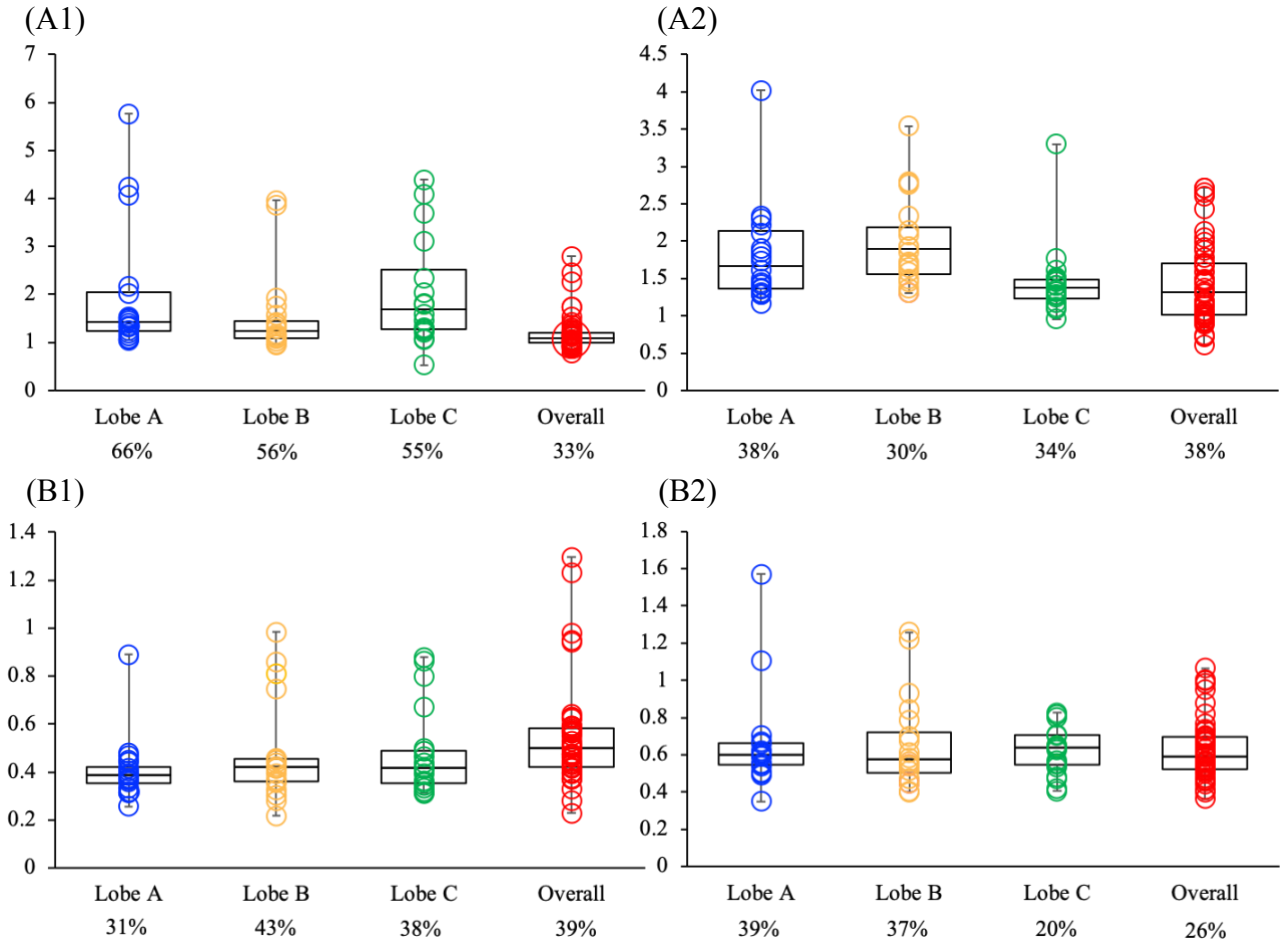


Figure 4. 26 Box plots showing cell-to-cell variability of **(A)** oxaloacetate and **(B)** succinic acid in different cell populations of animal 1 and 2.

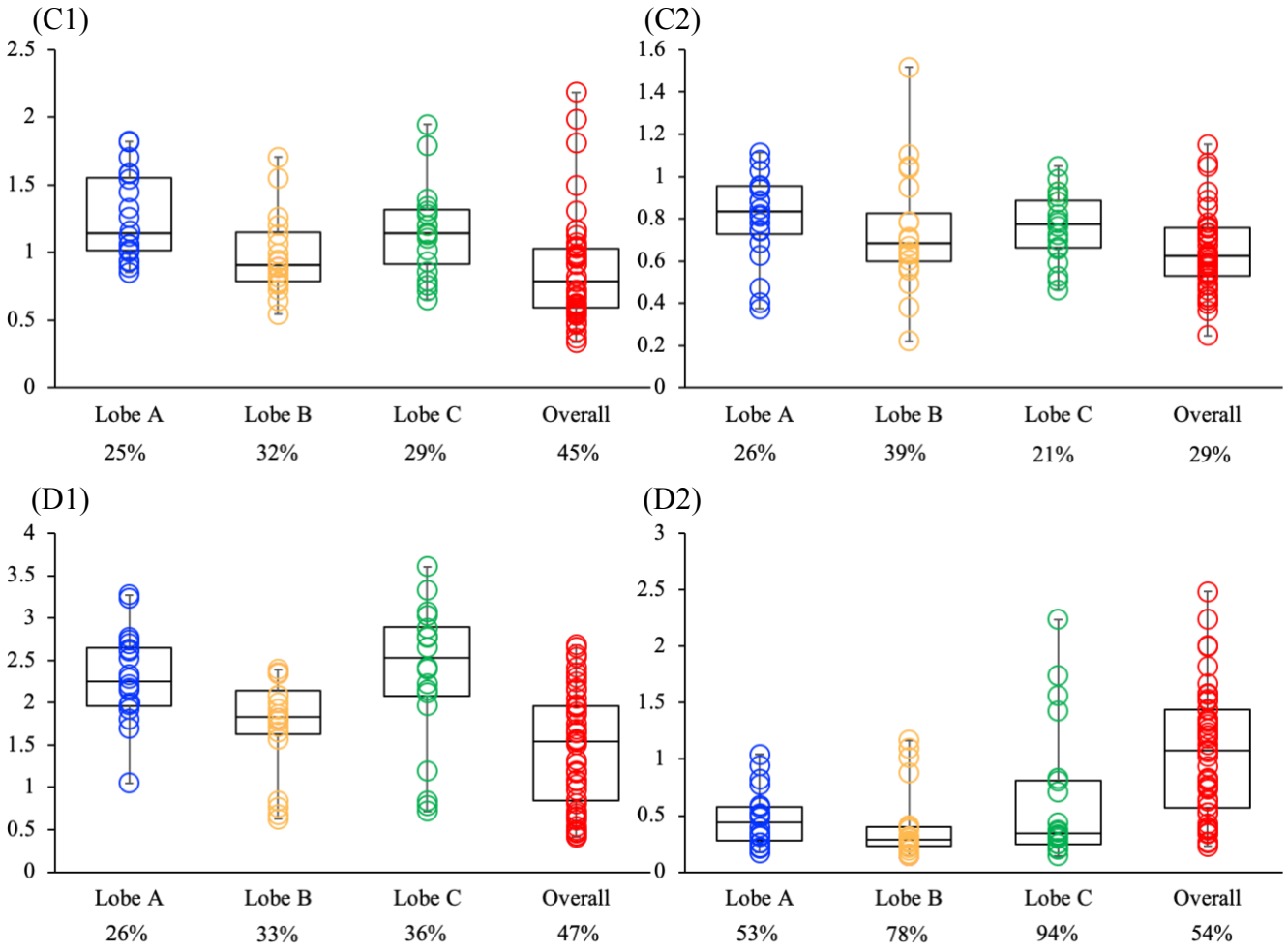


Figure 4.26 Box plots showing cell-to-cell variability of **(C)** fumaric acid and **(D)** malic acid in different cell populations of animal 1 and 2.

4.3.6 Importance of 4-Channel Profiling

The results of this study indicate that the cellular metabolite composition depends on their locations in the ovary. The amine and phenol, carboxylic acid containing metabolites of cells look very similar in cells collected at lobes and at random locations. However, it is interesting to note that the hydroxyl and carbonyl containing metabolites in cells are very different, depend on the locations in the organ we collected. This demonstrates the importance and need to apply four labeling chemistries to elucidate the metabolome of cells comprehensively and systematically.

4.4 Conclusions

To summarize, an analytical workflow for comprehensive and quantitative metabolomic analysis of cells at different locations of the *Xenopus laevis* ovary has been developed. The application of CIL LC-MS in this work demonstrates its general applicability for single-cell metabolomics. Moreover, with a holistic understanding of the oocytes in the ovary, we believe our findings can ultimately shed light on the use of *Xenopus* oocytes in research and herald novel investigative possibilities in cell biology.

4.5 Literature Cited

1. Rasar, M. A.; Hammes, S. R. *Methods Mol. Biol.* **2006,** *322*, 17-30.

2. Cicirelli, M. F.; Smith, L. D. *Dev. Biol.* **1986,** *113* (1), 174-81.

3. Sato, K. I.; Tokmakov, A. A. *Methods Mol. Biol.* **2019,** *1920*, 59-73.

4. Sretarugsa, P.; Wallace, R. A. *Dev. Growth Differ.* **1997,** *39* (1), 87-97.

5. Harland, R. M.; Grainger, R. M. *Trends Genet.* **2011,** *27* (12), 507-15.

6. Cameron, I. L.; Lum, J. B.; Nations, C.; Asch, R. H.; Silverman, A. Y. *Biol. Reprod.* **1983,** *28* (4), 817-22.

7. Palma, E.; Reyes-Ruiz, J. M.; Lopergolo, D.; Roseti, C.; Bertollini, C.; Ruffolo, G.; Cifelli, P.; Onesti, E.; Limatola, C.; Miledi, R.; Inghilleri, M. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (11), 3060-5.

8. Boggavarapu, R.; Hirschi, S.; Harder, D.; Meury, M.; Ucurum, Z.; Bergeron, M. J.; Fotiadis, D. *Methods Mol. Biol.* **2016,** *1432*, 223-42.

9. Zeng, S. L.; Sudlow, L. C.; Berezin, M. Y. *Expert Opin Drug Discov* **2020,** *15* (1), 39-52.

10. Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (21), 6545-6550.

11. Onjiko, R. M.; Plotnick, D. O.; Moody, S. A.; Nemes, P. *Anal. Methods* **2017,** *9* (34), 4964-4970.

12. Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P. *J. Vis. Exp.* **2017,** (130).

13. Portero, E. P.; Nemes, P. *Analyst* **2019,** *144* (3), 892-900.

14. Guo, K.; Li, L. *Anal. Chem.* **2009,** *81* (10), 3919-3932.

15. Huan, T.; Li, L. *Anal. Chem.* **2015,** *87* (14), 7011-7016.

16. Zhao, S.; Li, H.; Han, W.; Chan, W.; Li, L. *Anal. Chem.* **2019,** *91* (18), 12108-12115.

17. Zhao, S.; Luo, X.; Li, L. *Anal. Chem.* **2016,** *88* (21), 10617-10623.

18. Guo, K.; Li, L. *Anal. Chem.* **2010,** *82* (21), 8789-8793.

19. Zhao, S.; Dawe, M.; Guo, K.; Li, L. *Anal. Chem.* **2017,** *89* (12), 6758-6765.

20. Vinegoni, C.; Dubach, J. M.; Thurber, G. M.; Miller, M. A.; Mazitschek, R.; Weissleder, R. *Drug Discov. Today* **2015,** *20* (9), 1087-1092.

Chapter 5 – Construction and Application of MS/MS Retention Time Library in Metabolomics

5.1 Introduction

Metabolomics is the systematic study of small molecules, < 1500 Da in a biological system. The goal of this work is to convert raw data into biological knowledge, and chemical identification of metabolites plays a very significant role in this process. There are two major platforms in metabolomics, which are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). No matter which platform is applied in metabolic studies, it is of utmost importance for us to know the identity of the detected features. By identifying metabolites in a biological sample, it is possible for us to give insights and answer questions of biological significance. Herein, we focus on the metabolite identification using a sensitive detection tool – MS.

In metabolite profiling, particularly in untargeted metabolomics, that is the analysis of all detectable metabolites in a sample. Metabolite identification is a significant bottleneck in drawing biological conclusions from metabolomics data. The reason is that in untargeted profiling, thousands of metabolites are usually detected with limited or even no prior knowledge of the metabolite makeup of samples. And owing to the fact that identification is generally based on database searches, it is challenging to identify the unknown or unreported metabolites[1]. It is also worth noting that the identification process should be able to differentiate metabolites of same nominal mass but different molecular structure and monoisotopic mass; and metabolites with the same nominal and monoisotopic masses but different chemical structures. In addition, correct assignment of the detected metabolite to the "parent" metabolite is difficult since a single metabolite is often detected as multiple different derived species, for example adducts, which are

related species of the metabolite[2]. As a result, immense, unambiguous and universal metabolite identification in high-throughput large scale metabolomics still remained challenging.

The Chemical Analysis Working Group (CAWG) as part of the Metabolomics Standards Initiative (MSI) has defined four different levels of metabolite identification confidence[2, 3]. These are definitive identification (level 1), putative identification (level 2 and 3), and unknown (level 4). Level 1 is based on comparing two or more orthogonal properties, such as m/z and fragmentation pattern, with an authentic chemical standard analyzed under the same analytical conditions. Level 2 and 3 are based on comparing one or two properties, but not with an authentic chemical standard. Instead, it usually compares data collected in different laboratories or even acquired with different analytical methods. Level 4 is based on using the spectral data, such as accurate mass. The identification confidence of the unidentified and unclassified metabolites is the lowest. Figure 5.1 illustrates the four-level of metabolite identification confidence.

Obviously, the expansion of the library, that is adding more information to it, is one of the major works in metabolite identification. In this work, the construction and application of two MS/MS retention time (RT) libraries are reported. Liquid chromatography-mass spectrometry (LC-MS) has been widely used in metabolomics mainly due to its high metabolite detectability and accurate quantification ability. First, we construct an endogenous metabolite library of over 800 compounds containing molecular mass, experimental fragment ion spectrum (MS/MS) and additional RT information. We then use this library to demonstrate a rapid and universal metabolite identification in human biofluid samples. Second, we expand the MS/MS-RT library by introducing HILIC based RTs for more comprehensive and high-confidence metabolite identification. We optimize the parameters for HILIC-MS acquisition and also demonstrate the metabolite identification using the HILIC MS/MS-RT library.

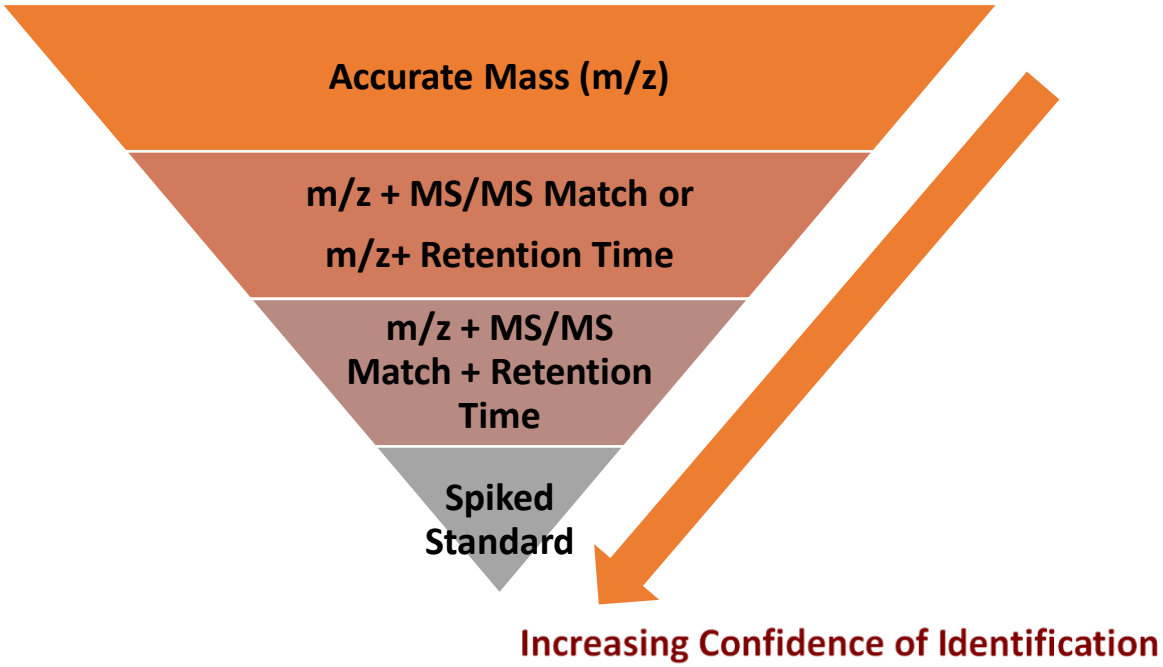


Figure 5. 1 Four-level of metabolite identification confidence.

5.2 Experimental

5.2.1 Construction and Application of MS/MS-RT Library for Endogenous Metabolite Identification

5.2.1.1 Method

831 endogenous metabolite standards were obtained from the Human Metabolome Database (HMDB). Each standard was injected into an Intensity Solo 2 C18 reversed phase column via an Elute UHPLC system and detected by an Impact QTOF-MS (all from Bruker Daltonics). High-resolution MS and MS/MS spectra in both positive and negative ion modes were acquired for each metabolite. The RT of each standard was determined using defined experimental LC conditions and inspected manually. A rapid identification method of real samples was developed in the MetaboScape software by matching of multiple parameters which include precursor mass accuracy, precursor isotopic pattern, RT, and MS/MS spectrum quality. By using RT calibrants and multipoint RT calibration, RTs of metabolites in the library can be corrected for different instrument setups.

5.2.1.2 LC-MS

Mobile phase for the gradient elution was solvent A: H_2O + 0.1% formic acid (FA) and solvent B: acetonitrile (ACN) + 0.1% FA.

The gradient elution profile 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 0.25 mL/min. All mass spectra were collected in both positive and negative ion modes and were recorded from 20 to 1000 m/z with a spectra acquisition rate of 8 Hz. The nebulizer was set to 1.0 bar, and the dry temperature was 200 ℃, with 8 L/min of drying gas.

5.2.1.3 Library Data Acquisition and Data Processing

The MS/MS-RT library contains multiple layers of information for metabolite identification, including accurate masses, high-resolution MS/MS spectra in both positive and negative ion modes and normalized retention time in RPLC separation.

Accurate mass data was acquired on a high-resolution Impact QTOF-MS mass spectrometer (Bruker Daltonics) with flow injection analysis. High-resolution MS/MS spectra were collected using five different collision energy levels, which are 10, 20, 30, 40 and 20-50 eV. For RT, metabolite standards were pooled into several mixtures. Each standard mixture was spiked with a RT calibrant and separated on a RPLC column using a 20-minute linear gradient. RTs of standards were extracted from calibrated LC-MS data using TargetAnalysis (Bruker). After manual inspection, the RTs were entered into the existing Bruker HMDB MS/MS Library. Taken together, different layers of information were integrated together to the Bruker HMDB library.

Human biofluids such as urine samples were analyzed and data were collected by RP-UHPLC-MS/MS. The raw data from LC-MS were uploaded to the MetaboScape software and searched against the library. There is no need to transform data into different formats, such as a csv file containing mass list. In data processing, feature was first extracted and the reductant spectral features such as adduct ions and multimers were removed. The spectral library search was performed using the Bruker HMDB Metabolite Library in MetaboScape.

5.2.2 Construction and Application of HILIC MS/MS-RT Library for Metabolite Identification

5.2.2.1 Method

829 human endogenous metabolite standards from the Human Metabolome Database (HMDB) were used to construct the HILIC MS/MS-RT library. Standards and samples were separated on an YMC-Triart Diol-HILIC column (1.9 μm, 100×2.1 mm) in an Elute UHPLC system and detected by Impact II HD QTOF-MS system (all from Bruker Daltonics). An optimized 20-minute linear gradient was used to separate metabolite standards and samples. Five types of additives were screened to find the best mobile phase buffer for data collection. High-resolution MS and MS/MS spectra in both positive and negative ion modes were acquired for each metabolite. RTs were determined using optimized LC conditions and inspected manually. A multipoint RT calibration strategy was used for transferring RTs from the instrumental setup in one laboratory to the same setup in a different laboratory.

5.2.2.2 LC-MS

Mobile phase for the gradient elution was solvent A: 10mM ammonium acetate (NH_4Ac) + 0.1% FA and solvent B: ACN + 5% H_2O.

The gradient elution profile was: t = 0 min, 100 % B; t = 2 min, 100 % B; t = 17 min, 50 % B; t = 20 min, 50 % B.

The flow rate was 0.25 mL/min. All mass spectra were collected in both positive and negative ion mode and were recorded from 20 to 1000 m/z with a spectra acquisition rate of 8 Hz. The nebulizer was set to 1.0 bar, and the dry temperature was 200 ℃, with 8 L/min of drying gas.

5.2.2.3 Library Data Acquisition and Data Processing

Library data acquisition and data processing in HILIC MS/MS-RT library is the same as described in Section 5.2.1.3.

5.3 Results and Discussion

5.3.1 MS/MS-RT Library

5.3.1.1 Overview

Metabolite standards were analyzed in triplicates. High-resolution MS/MS spectra of 635 compounds were generated in positive ion mode and those of 474 metabolites were acquired in negative ion mode. Quality control (QC) sample was injected and analyzed intermittently for the duration of this library construction work to assess the variance observed in the data throughout the sample preparation and data acquisition steps. Up to 5 collision energy levels were applied for each standard. For each metabolite, manual interpretation of the fragmentation was performed and fragments were labeled with structure in the library, which is very useful in chemical structure identification and interpretation.

5.3.1.2 Multiple point RT Normalization

To achieve unambiguous identification, RT information for each standard was added, which significantly improved the identification confidence, especially for low intensity or isobaric compounds and isomers. However, RT can be easily affected by variations in experimental conditions such as instrument brands, LC columns and gradients. To address this issue, RT calibrants and RT normalization method were used to correct RT shift in an effective and universal

manner. RT calibrant, which is a mixture containing 28 metabolites, was first generated. The majority of the components are dansyl cholride labeled metabolites, since dansyl-labeled compounds are more stable than unlabeled metabolites. Moreover, they also have better peak shape in reverse phase liquid chromatography. The RTs of metabolites in the RT calibrant spread out through the 20-minutes gradient and the interval between two metabolites is less than 1 or 2 minutes as shown in Figure 5.2. The list of metabolites in the RT calibrants is summarized in Table 5.1.

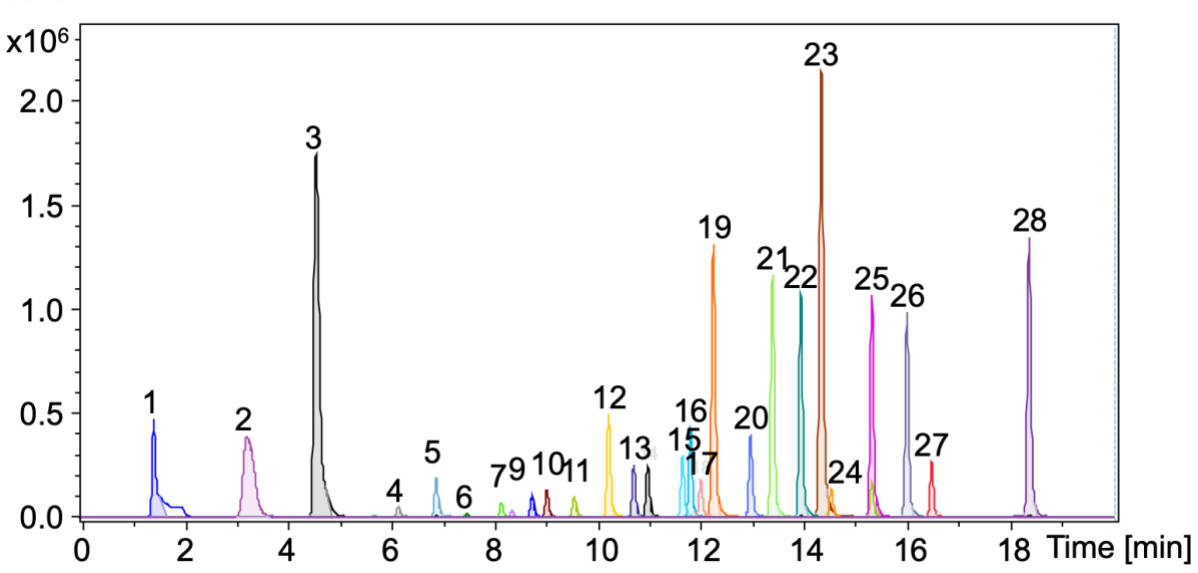


Figure 5. 2 Extracted ion chromatograms of RT calibrants.

#	Compound	RT (min)
1	Uracil	1.46
2	m-Aminobenzoic acid	3.26
3	Dns-OH	4.59
4	Dns-Glucosamine 6-phosphate	6.18
5	Dns-Phosphoethanolamine	6.91
6	Dns-Arginine	7.51
7	Dns-Serine	8.17
8	Dns-Glutamic acid	8.38
9	Dns-Threonine	8.76
10	Dns-Glycine	9.05
11	Dns-Alanine	9.58
12	Dns-2-Aminoisobutyric acid	10.24
13	Dns-Proline	10.73
14	Dns-Methionine	11.00
15	Dns-Phenylalanine	11.67
16	Dns-Aspartame	11.82
17	Dns-Cystine 2 tag 2+	12.02
18	Dns-Dimethylamine	12.27
19	Dns-Homocystine 2 tags 2+	12.50
20	Dns-Lysine 2 tags 2+	12.98
21	Dns-1-phenylethylamine	13.41
22	Dns- Tyrosyl-Alanine 2 tags 2+	13.95
23	Dns-Cadaverine 2 tags 2+	14.35
24	Dns-Tyrosine 2 tags 2+	14.54
25	Dns-Serotonin 2 tags 2+	15.32
26	Dns-4-Ethylphenol	16
27	Dns-Pyrocatechol 2 tags 2+	16.48
28	Dns-Decyl alcohol	18.36

Table 5. 1 List of 28 metabolites in RT calibrants.

5.3.1.3 Validation in Human Urine Sample

The MS/MS-RT Library constructed has been applied in metabolite identification of human urine samples. Figure 5.3A and 5.3B shows the base peak chromatograms of urine sample with triplicate injection in positive and negative ion modes, respectively. The identification detail is shown in Table 5.2.

The portability of this library for different instrumental conditions and for different laboratories was also assessed using human urine samples. That is to say, the urine samples were analyzed in two independent laboratories, one in Canada, another in Germany. The comparative results of metabolite identification were shown in Table 5.3.

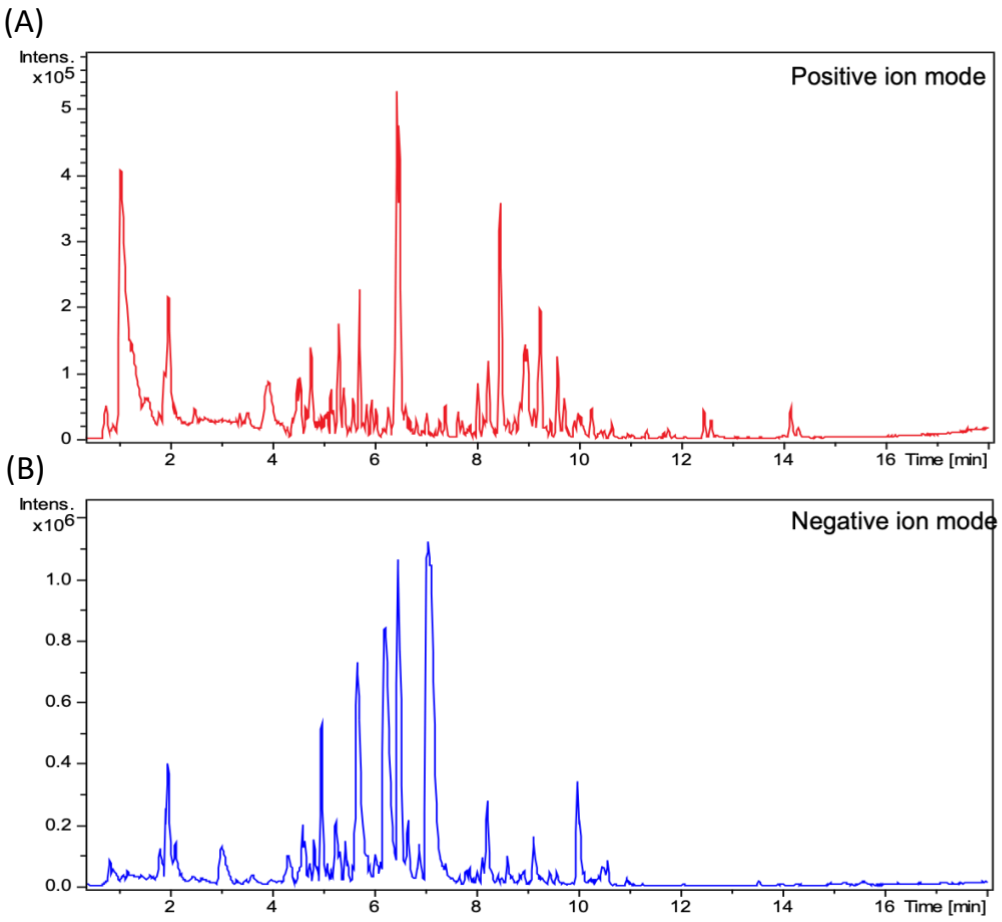


Figure 5. 3 Base peak chromatograms of urine sample with triplicate injection in **(A)** positive ion mode and **(B)** negative ion mode.

Sample	Polarity	# of "filtered" features	# of ID features using m/z	# of ID features using m/z+MS/MS	# of ID features using m/z+RT	# of high-confidence ID features	# of identified metabolites
Human urine	positive	2460	188	66	109	127	99
	negative	3091	239	68	137	155	122
	In total					240	177
* Features after removing redundant spectral features (adduct ions, multimers, etc.)							

Table 5. 2 Identification result of human urine samples using the MS/MS-RT library.

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	99
	Germany	2986	210	70	114	139	115
Negative	Canada	3091	239	68	137	155	122
	Germany	2412	190	69	126	139	121
In total					Canada	240	177
					Germany	232	186

Table 5. 3 Comparative results of metabolite identification of human urine sample in two different laboratories.

5.3.2 HILIC MS/MS-RT Library

5.3.2.1 Overview

829 human endogenous metabolites from HMDB were used as metabolite standards in constructing the HILIC MS/MS-RT library. These metabolite standards were analyzed in triplicates in both positive and negative ion modes. Quality control (QC) sample was injected and analyzed intermittently for the duration of this library construction to assess the variance observed in the data throughout the sample preparation and data acquisition steps. Retention times of metabolites, with standard deviation less than 6 seconds were determined. The retention time distribution of metabolites in HILIC MS/MS-RT library is shown in Figure 5.4.

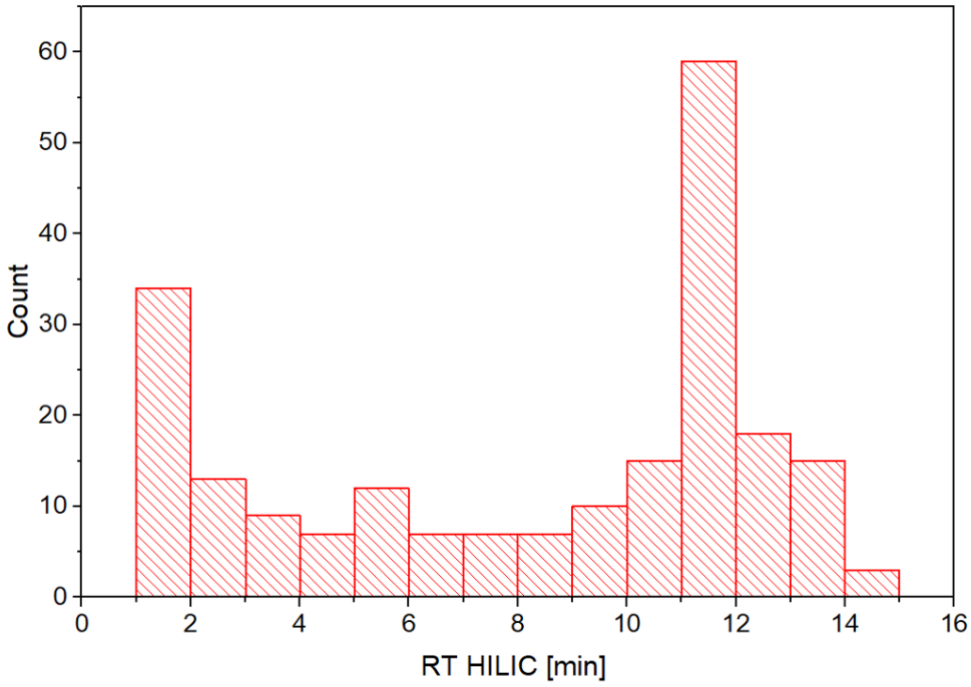


Figure 5. 4 RTs distribution in HILIC MS/MS-RT Library.

5.3.2.2 Parameter Optimization for HILIC MS/MS-RT library

For mobile phase buffer optimization, it is important to consider that the performance of HILIC may not be as stable as RPLC separation. For example, it has poorer retention and reproducibility, therefore, the mobile phase buffer has to be optimized for optimal elution of polar/ionic metabolites. Five types of additives were screened to find the best mobile phase buffer, these are (i) 0.1% FA; (ii) 10mM NH_4Ac; (iii) 10mM NH_4FA; (iv) 10mM NH_4Ac with 0.1% FA; and (v) 10mM NH_4FA with 0.1% FA. After analyzing 24 amino acids, aromatic compounds and carbohydrates, 10 mM NH_4Ac+0.1% FA was found to be the optimal buffer for mobile phase, in which most of the metabolites were detected as a reproducible peak with good peak shape. The standard deviation (SD) of retention time and the relative standard deviation (RSD) of peak width were being considered as shown in Table 5.4.

	SD of Retention Time (s)					RSD of Peak Width				
Compound	FA	NH4Ac	NH4FA	NH4Ac+FA	NH4FA+FA	FA	NH4Ac	NH4FA	NH4Ac+FA	NH4FA+FA
Dimethylglycine	8.0	1.8	0.2	1.6	0.4	18.05%	9.40%	0.80%	2.76%	3.68%
gamma-Aminobutyric acid	10.2	1.3	1.3	1.0	0.4	21.97%	7.20%	5.47%	4.09%	3.68%
L-Isoleucine	23.8	1.4	0.1	1.2	0.4	16.75%	4.92%	2.35%	1.03%	1.83%
L-Histidine	1.3	11.1	4.8	1.6	1.8	4.66%	26.06%	31.73%	4.11%	1.36%
3-Methylhistidine	0.1	0.1	2.7	0.4	0.5	4.58%	2.66%	9.70%	2.41%	0.83%
Kynurenine	1.3	1.2	0.4	1.0	0.4	6.90%	5.58%	2.54%	0.38%	2.08%
L-Aspartyl-L-phenylalanine	0.4	0.4	0.5	0.1	0.7	0.62%	2.82%	6.18%	4.94%	0.42%
Glycyl-L-leucine	0.6	0.7	0.2	1.1	0.5	3.76%	2.49%	0.96%	2.61%	1.00%
N-Acetyl-L-alanine	4.8	5.5	0.2	1.9	0.7	4.62%	2.30%	0.93%	3.36%	3.04%
Quinaldic acid	4.7	0.6	0.9	0.7	9.3	0.78%	7.84%	4.89%	2.80%	0.90%
Methylhippuric acid	0.6	0.5	0.2	0.3	0.6	5.62%	3.20%	0.47%	2.20%	2.67%
Biopterin	0.9	0.8	1.7	0.4	0.5	2.34%	0.29%	1.58%	3.85%	3.12%
Hematoporphyrin	0.2	1.2	1.3	0.3	0.5	0.24%	5.85%	0.65%	3.72%	5.01%
p-Aminobenzoic acid	0.2	0.7	0.0	0.2	0.2	0.89%	8.43%	10.06%	1.22%	2.59%
Famotidine	2.0	0.8	12.2	10.3	11.0	8.39%	12.24%	1.51%	5.99%	16.66%
Clotrimazole	0.6	0.2	0.4	0.7	0.3	0.87%	0.71%	2.38%	1.56%	0.92%
Levofloxacin	1.3	1.8	6.2	0.1	6.6	7.33%	2.14%	3.04%	6.15%	2.17%
Triamterene	6.1	5.9	26.0	5.9	16.7	10.45%	1.04%	7.04%	6.33%	0.57%
Phenylpropanolamine	2.0	1.6	5.3	1.6	12.0	6.01%	9.76%	21.28%	9.76%	18.38%
Pseudoephedrine	8.1	4.3	18.8	6.4	18.6	13.30%	11.13%	32.11%	1.92%	10.44%
Lactose	1.4	1.4	1.6	0.2	0.7	3.77%	7.27%	3.23%	1.45%	3.77%
N-Acetyl-D-glucosamine	1.1	0.8	1.3	1.1	0.9	2.39%	0.88%	1.26%	3.26%	2.06%
Glucosamine 6-sulfate	0.1	1.0	1.0	0.4	0.8	3.91%	3.31%	2.59%	2.30%	1.33%
Sialyllactose	0.0	0.1	0.8	0.2	0.4	2.34%	0.59%	0.40%	0.45%	1.05%

Table 5. 4 HILIC mobile phase buffer optimization.

For gradient elution profile optimization, two gradients were compared. Both of them are 20-minutes linear gradient. Gradient I ends at 50% of mobile phase B, while gradient II ends at 60% mobile phase B as depicted in Figure 5.5A. The RTs distribution of 270 compounds using the two gradients were compared and shown in Figure 5.5B. Gradient I is used for eluting the polar/ionic metabolites.

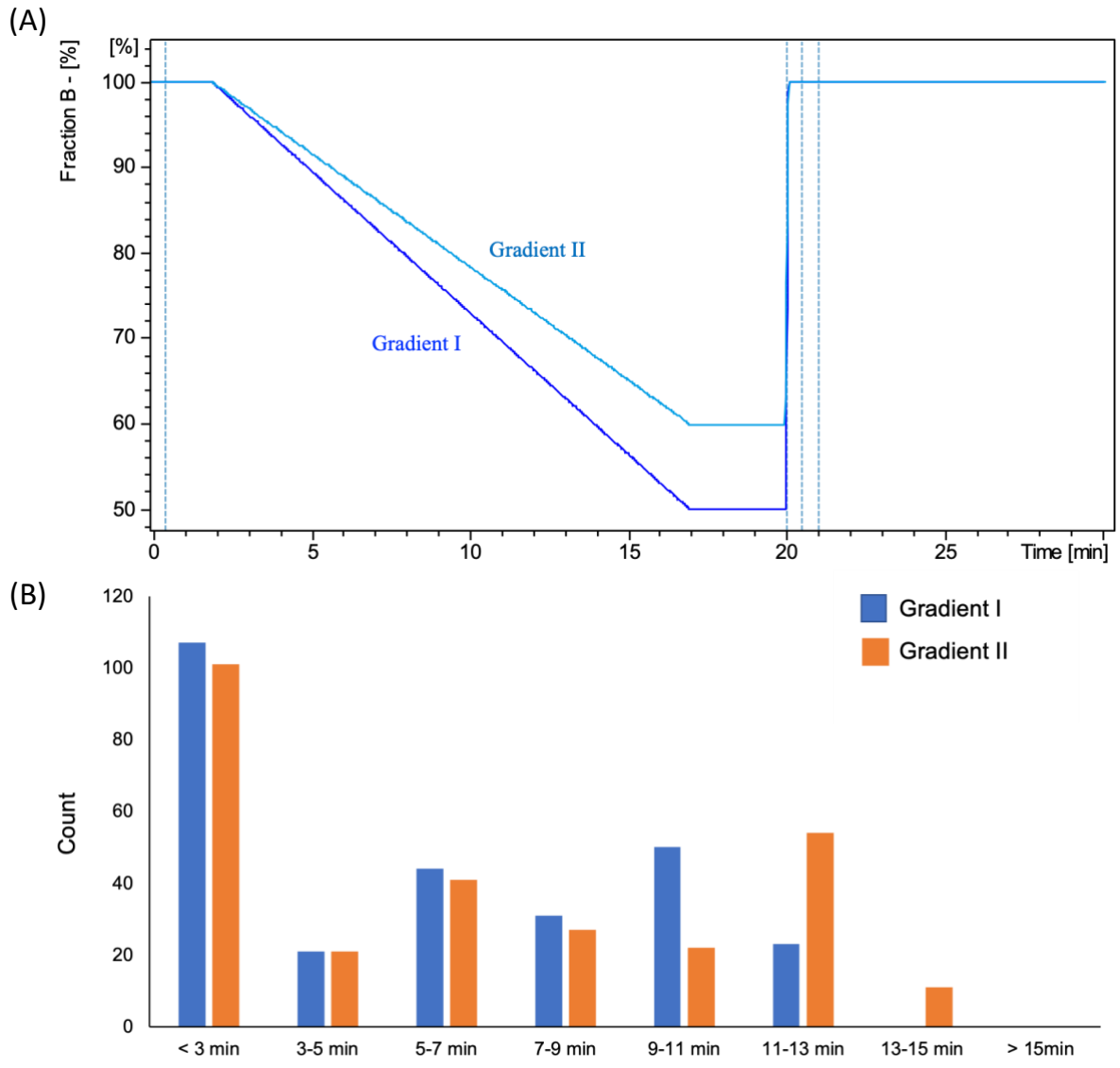


Figure 5. 5 (A) Gradient elution profile of gradient I and II **(B)** RTs distribution of 270 compounds using two gradients.

5.3.2.3 Multiple point RT Normalization

Since the RT of HILIC separation can be easily affected by minor variations in experimental conditions, a multipoint RT calibration strategy was used to transfer RTs from the instrumental setup in one laboratory to the same setup in a different laboratory. In other words, that is the portability of the library for different laboratories. This step is important and essential to ensure unambiguous and confident metabolite identification, as well as to verify the general applicability of the library.

Two instrument systems are involved here, HILIC-QTOF system 1 is located in Alberta, Canada, while HILIC-QTOF system 2 is located in Bremen, Germany. A RT calibrant contains a list of metabolites that can be separated and distributed on the chromatogram evenly as shown in Figure 5.6. The list of metabolites in the RT calibrants is summarized in Table 5.5. HILIC-QTOF system 1 was used to generate the RT calibrants here. The RT calibrants was then sent to Germany and the RTs for the RT calibrants were determined experimentally using HILIC-QTOF system 2. The RTs in the HILIC MS/MS-RT Library for system 2 were then corrected using a multipoint RT calibration algorithm. In this way, correction of RTs for local version of HILIC MS/MS-RT Library is achieved. Figure 5.7 shows the multipoint linear regression of RT calibrants in HILIC-QTOF systems 1 and 2. This demonstrates that the multipoint RT calibration method with the use of RT calibrants can be used to correct RT shift between the library RTs and the experimental RTs. This is effective to overcome various experimental variations, such as running samples in different LC-MS systems.

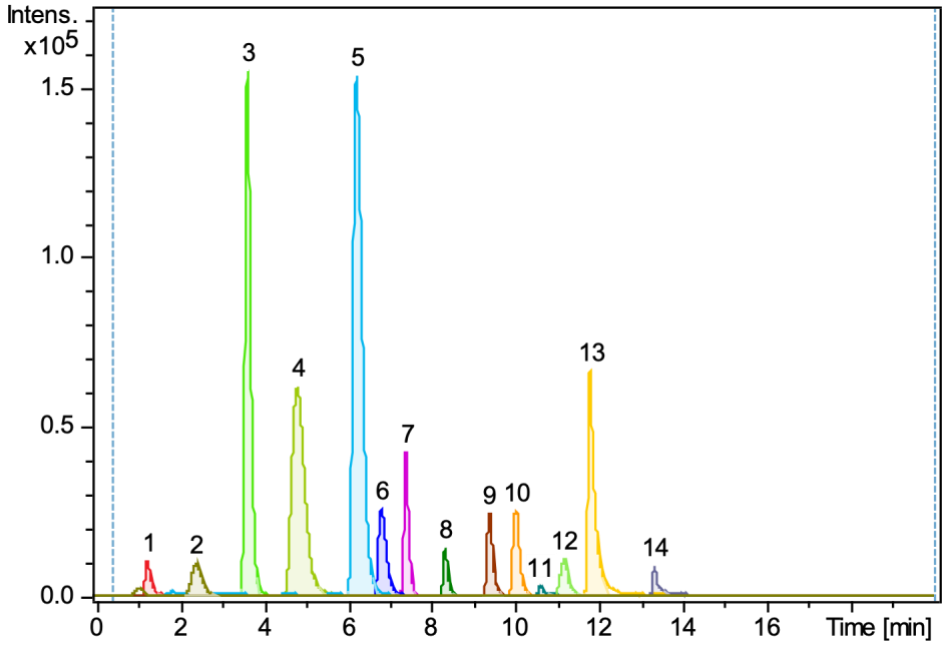


Figure 5. 6 Extracted ion chromatograms of RT calibrants.

#	Compound	RT (min)
1	Caffeine	1.39
2	3-Methylxanthine	2.43
3	1,3-Dimethyluric acid	3.63
4	Hypoxanthine	5.31
5	Riboflavin	6.96
6	Cytosine	7.4
7	Tiglylglycine	7.92
8	Nicotinic acid	8.71
9	L-Phenylalanine	9.63
10	Alloisoleucine	10.23
11	Neopterin	10.75
12	Histamine	11.37
13	3-Methylhistamine	12.07
14	Carnosine	13.56

Table 5. 5 List of 14 metabolites in RT calibrants.

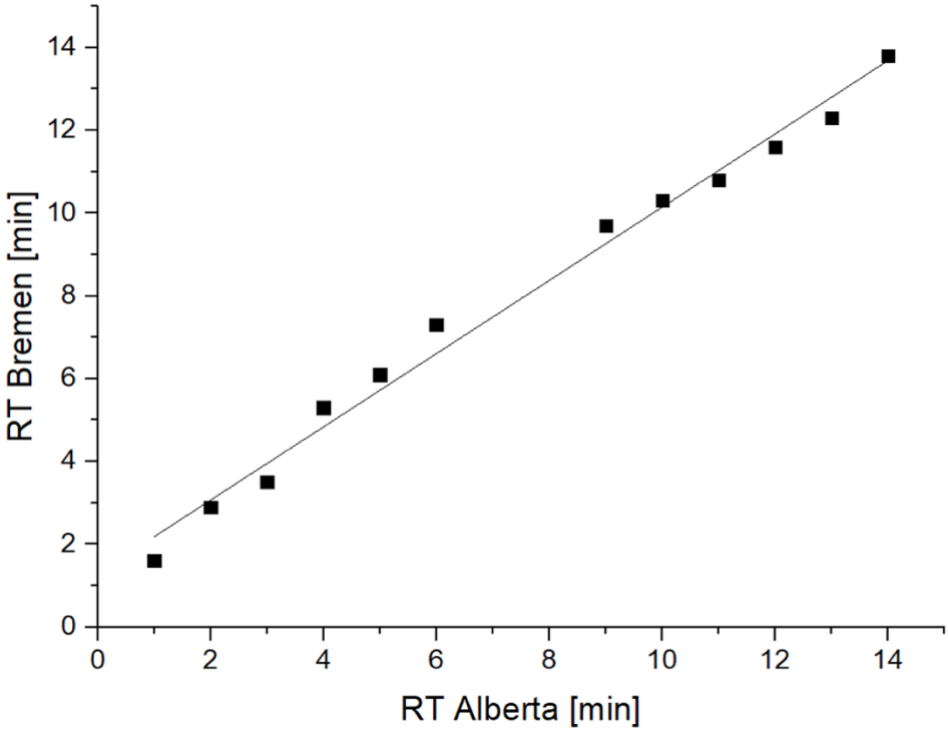


Figure 5. 7 Multi-point linear regression of calibrants RT for both systems.

5.3.2.4 Validation in Human Urine Sample

In order to illustrate the utility of the library in analyzing real-world samples, human urine samples were analyzed by LC-MS and subsequent metabolite identification was done by using the HILIC MS/MS-RT library constructed.

Urine samples were prepared by diluting the filtered urine with threefold of acetonitrile so that the sample composition is close to the initial mobile phase. HILIC-MS/MS data was acquired on system 2 (in Bremen, Germany), in both positive and negative ion modes, followed by metabolite identification. Highly reproducible LC-MS data across replicates were generated as shown in Figure 5.8.

To determine the number of identified features, the RT correction workflow described in Section 5.3.2.3 was applied. The annotation was conducted using the MetaboScape software (Bruker Daltonics) by automatically matching precursor mass, precursor isotopic pattern (mSigma value), corrected retention time, as well as fragment spectrum information. The results show that 130 metabolites were identified in positive ion mode and 98 metabolites were identified in negative ion mode.

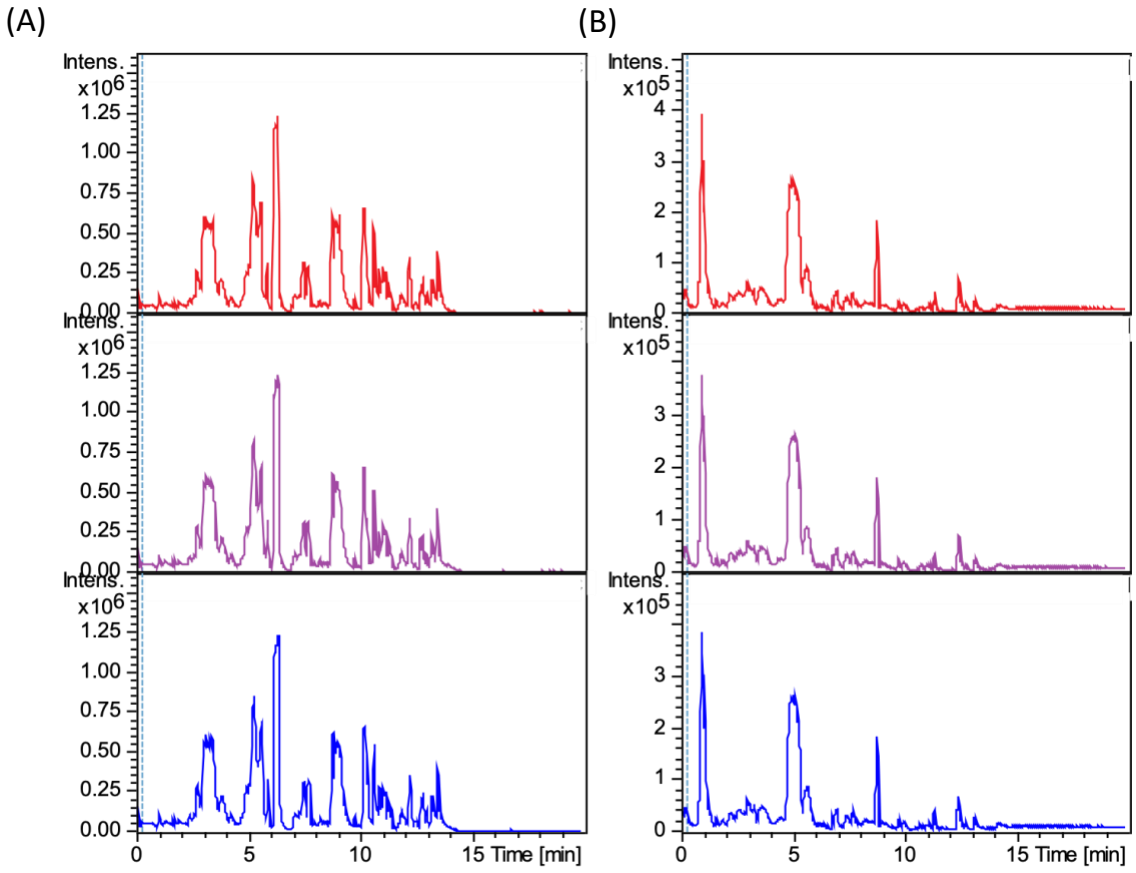


Figure 5. 8 Base peak chromatograms of urine sample with triplicate injection in **(A)** positive ion mode and **(B)** negative ion mode.

5.4 Conclusions

Two MS/MS-RT libraries were constructed. Both of them contain over 800 endogenous metabolites, with the second library being HILIC based. The applicability of these libraries in real sample analysis was demonstrated using human urine sample. For both libraries, RT calibrants were generated and RT normalization method was used to correct RT shift in different instrument setups. In this way, inter-lab portability of the libraries and workflow for metabolite identification can be demonstrated.

5.5 Literature Cited

1. Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (12), 1897-1905.

2. Dunn, W. B.; Erban, A.; Weber, R. J. M.; Creek, D. J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M. R. *Metabolomics* **2013,** *9* (1), 44-66.

3. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; 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.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007,** *3* (3), 211-221.

Chapter 6 - Conclusions and Future Work

6.1 Thesis Summary

Studies of single-cell metabolomics using *Xenopus* oocytes as a model system with chemical isotope labeling (CIL) liquid chromatography mass spectrometry (LC-MS) platform were performed. The studies revealed the cell-to-cell heterogeneity, which holds a great potential to elevate our understanding of the cellular physiology and give biological insights in health and diseases.

In chapter 2, preparation of *Xenopus* oocytes and CIL LC-MS workflow were developed to profile the amine and phenol submetabolome of single cells comprehensively and quantitatively. These involve cells extraction from the *Xenopus laevis*, cell lysis to extract the metabolites, CIL of amines and phenols containing metabolites and subsequent LC-MS analysis. In the metabolomic analysis of 124 single cells with 6 different periods of culture times, 1531 peak pairs were detected. Among them, 80 metabolites were positively identified by retention time and m/z search against CIL standard libraries, 146 metabolites were putatively identified with high confidence by retention time and m/z match against the LI library and 1173 metabolites were putatively identified by m/z match against the mass-based database. In other words, 89% of detected features are identified. We believe this is a big leap in single-cell metabolomics profiling and it has the potential to improve the understanding of cell heterogeneity and programming of cell metabolism. The work presented in this chapter may not have immediate biomedical applications, but it does illuminate a fundamental way in which cells work.

In chapter 3, cellular response to perturbation was investigated with the purpose of studying the behavior and molecular mechanisms underlying cellular systems. A study plan was developed and CIL LC-MS method was applied to study the metabolic responses of cells to heat stress.

Dansylation was used to elucidate and quantify the change of amine and phenol submetabolome in cells. The results suggest that short-term heat stress has great effects on the cellular metabolome and the enzymes system of cells may have adapted to high temperature in response to long-term heat stress. Recovery from heat stress was accompanied by changes in metabolite abundance, which indicates that the cellular amine, phenol metabolome changed after heat stress. Again, the behavior of each single-cell is revealed.

In chapter 4, an analytical workflow for comprehensive and quantitative metabolomic analysis of cells at different locations of the *Xenopus laevis* ovary has been developed, with the use of four labeling chemistries. The results indicate that the cellular metabolite composition depends on their locations in the ovary. The amine and phenol, carboxylic acid containing metabolites of cells look very similar in cells collected at lobes and at random locations. However, the hydroxyl and carbonyl containing metabolites in cells are very different, depending on locations in the organ being collected. This study demonstrates the importance to apply four labeling chemistries to elucidate the metabolome of cells comprehensively and systematically. We believe the findings here can ultimately shed light on the use of *Xenopus* oocytes in medical research.

Finally, in chapter 5, two MS/MS-RT libraries were constructed for metabolite identification. Molecular mass, MS/MS spectrum and RT information were collected. RT calibrants and multipoint RT calibration were used to transfer RTs from the instrumental setup in one laboratory to the same setup in another laboratory. Moreover, metabolite identification in human urine samples was demonstrated using the two libraries.

6.2 Future Work

For chapter 2 and 3, I hope to expand the coverage by using other three chemical isotope labeling methods. In this way, a better understanding of the single-cell metabolome can be achieved.

For chapter 5, I hope to include more metabolites and provide more information to the library for high-confidence metabolite identification.

Bibliography

1. Abdelrahman, M.; Burritt, D. J.; Gupta, A.; Tsujimoto, H.; Tran, L.-S. P. *J. Exp. Bot.* **2019,** *71* (2), 543-554.

2. Ali, A.; Abouleila, Y.; Shimizu, Y.; Hiyama, E.; Emara, S.; Mashaghi, A.; Hankemeier, T. *Trends Analyt Chem* **2019,** *120*, 115436.

3. Alseekh, S.; Fernie, A. R. *Plant J.* **2018,** *94* (6), 933-942.

4. Aslam, B.; Basit, M.; Nisar, M. A.; Khurshid, M.; Rasool, M. H. *J. Chromatogr. Sci* **2017,** *55* (2), 182-196.

5. Barker, M.; Rayens, W. *J. Chemom* **2003,** *17* (3), 166-173.

6. Barna, J.; Csermely, P.; Vellai, T. *Cell. Mol. Life Sci.* **2018,** *75* (16), 2897-2916.

7. Benjamini, Y. *J R Stat Soc Series B Stat Methodol* **2010,** *72* (4), 405-416.

8. Benjamini, Y.; Hochberg, Y. *J R Stat Soc Series B Stat Methodol* **1995,** *57* (1), 289-300.

9. Bishop, A. L.; Rab, F. A.; Sumner, E. R.; Avery, S. V. *Mol. Microbiol.* **2007,** *63* (2), 507-20.

10. Boggavarapu, R.; Hirschi, S.; Harder, D.; Meury, M.; Ucurum, Z.; Bergeron, M. J.; Fotiadis, D. *Methods Mol. Biol.* **2016,** *1432*, 223-42.

11. Bumgarner, R. *Curr Protoc Mol Biol* **2013,** *Chapter 22*, Unit-22.1.

12. Cameron, I. L.; Lum, J. B.; Nations, C.; Asch, R. H.; Silverman, A. Y. *Biol. Reprod.* **1983,** *28* (4), 817-22.

13. Chandra, K.; Al-Harthi, S.; Sukumaran, S.; Almulhim, F.; Emwas, A.-H.; Atreya, H. S.; Jaremko, Ł.; Jaremko, M. *RSC Adv.* **2021,** *11* (15), 8694-8700.

14. Chen, D.; Han, W.; Su, X.; Li, L.; Li, L. *Anal. Chem.* **2017,** *89* (17), 9424-9431.

15. Clarke, A. *Int. J. Astrobiology* **2014,** *13*, 141.

16. Cornish, T.; Bryden, W. *Johns Hopkins Apl Technical Digest* **1999,** *20*, 335-342.

17. Date, S.; Mizuno, H.; Tsuyama, N.; Harada, T.; Masujima, T. *Anal Sci* **2012,** *28* (3), 201-3.

18. Duncan, K. D.; Fyrestam, J.; Lanekoff, I. *Analyst* **2019,** *144* (3), 782-793.

19. Dunn, W. B.; Erban, A.; Weber, R. J. M.; Creek, D. J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M. R. *Metabolomics* **2013,** *9* (1), 44-66.

20. Emara, S.; Amer, S.; Ali, A.; Abouleila, Y.; Oga, A.; Masujima, T. *Adv. Exp. Med. Biol.* **2017,** *965*, 323-343.

21. Fazakas, J.; Doros, A.; Smudla, A.; Tóth, S.; Nemes, B.; Kóbori, L. *Transplant. Proc.* **2011,** *43* (4), 1275-1277.

22. Feng, Z.; Ding, C.; Li, W.; Wang, D.; Cui, D. *Food Chem.* **2020,** *310*, 125914.

23. Fessenden, M. *Nature* **2016,** *540* (7631), 153-155.

24. Fulda, S.; Gorman, A. M.; Hori, O.; Samali, A. *Int J Cell Biol* **2010,** *2010*, 214074.

25. Ganesan, S.; Pearce, S. C.; Gabler, N. K.; Baumgard, L. H.; Rhoads, R. P.; Selsby, J. T. *J. Therm. Biol.* **2018,** *72*, 73-80.

26. Ganesan, S.; Volodina, O.; Pearce, S. C.; Gabler, N. K.; Baumgard, L. H.; Rhoads, R. P.; Selsby, J. T. *Physiol. Rep.* **2017,** *5* (16), e13397.

27. Gao, J.; Shen, W. *Environ. Pollut.* **2021,** *268* (Pt B), 115809.

28. Garbuz, D. G. *J. Mol. Biol.* **2017,** *51* (3), 352-367.

29. Gomez-Pastor, R.; Burchfiel, E. T.; Thiele, D. J. *Nat. Rev. Mol. Cell Biol.* **2018,** *19* (1), 4-19.

30. Gu, J.; Emerman, M.; Sandmeyer, S. *Mol. Cell. Biol.* **1997,** *17* (7), 4033-42.

31. Guo, K.; Li, L. *Anal. Chem.* **2009,** *81* (10), 3919-3932.

32. Guo, K.; Li, L. *Anal. Chem.* **2010,** *82* (21), 8789-8793.

33. Gupta, R. K.; Kuznicki, J. *Cells* **2020,** *9* (8), 1751.

34. Han, W.; Li, L. *Metabolomics* **2015,** *11* (6), 1733-1742.

35. Harland, R. M.; Grainger, R. M. *Trends Genet.* **2011,** *27* (12), 507-15.

36. Hassan, A. H. A.; Hozzein, W. N.; Mousa, A. S. M.; Rabie, W.; Alkhalifah, D. H. M.; Selim, S.; AbdElgawad, H. *Sci. Rep.* **2020,** *10* (1), 15076.

37. Hertl, M.; Howard, T. K.; Lowell, J. A.; Shenoy, S.; Robert, P.; Harvey, C.; Strasberg, S. M. *Liver Transplant.* **1996,** *2* (2), 111-7.

38. Hiyama, E.; Ali, A.; Amer, S.; Harada, T.; Shimamoto, K.; Furushima, R.; Abouleila, Y.; Emara, S.; Masujima, T. *Anal Sci* **2015,** *31* (12), 1215-7.

39. Horgan, R. P.; Kenny, L. C. *Obstet. Gynecol.* **2011,** *13* (3), 189-195.

40. Huan, T.; Li, L. *Anal. Chem.* **2015,** *87* (14), 7011-7016.

41. Huan, T.; Li, L. *Anal. Chem.* **2015,** *87* (2), 1306-1313.

42. Hubbard, A. H.; Zhang, X.; Jastrebski, S.; Singh, A.; Schmidt, C. *BMC Genom.* **2019,** *20* (1), 502.

43. Iyer, S. S.; Zhang, Z. P.; Kellogg, G. E.; Karnes, H. T. *J Chromatogr Sci* **2004,** *42* (7), 383-7.

44. Jeong, E. M.; Yoon, J. H.; Lim, J.; Shin, J. W.; Cho, A. Y.; Heo, J.; Lee, K. B.; Lee, J. H.; Lee, W. J.; Kim, H. J.; Son, Y. H.; Lee, S. J.; Cho, S. Y.; Shin, D. M.; Choi, K.; Kim, I. G. *Stem Cell Rep.* **2018,** *10* (2), 600-614.

45. Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2016,** *17* (7), 451-9.

46. Khamis, M. M.; Adamko, D. J.; El-Aneed, A. *Mass Spectrom. Rev.* **2017,** *36* (2), 115-134.

47. Li, Y.; Kong, L.; Deng, M.; Lian, Z.; Han, Y.; Sun, B.; Guo, Y.; Liu, G.; Liu, D. *Genes* **2019,** *10* (5).

48. Li, Y.; Li, L. *International Journal of Mass Spectrometry* **2018,** *434*, 202-208.

49. Li, Y.; Li, L. *J Am Soc Mass Spectrom* **2019,** *30* (9), 1733-1741.

50. Li, Y.; Li, L. *Rapid Communications in Mass Spectrometry* **2020,** *34* (S1), e8643.

51. Liao, Z.; Zhang, S.; Liu, W.; Zou, B.; Lin, L.; Chen, M.; Liu, D.; Wang, M.; Li, L.; Cai, Y.; Liao, Q.; Xie, Z. *J. Chromatogr. B* **2019,** *1133*, 121848.

52. Luo, M.; Meng, Z.; Moroishi, T.; Lin, K. C.; Shen, G.; Mo, F.; Shao, B.; Wei, X.; Zhang, P.; Wei, Y.; Guan, K.-L. *Nat. Cell Biol.* **2020,** *22* (12), 1447-1459.

53. Luo, X.; Gu, X.; Li, L. *Anal. Chim. Acta* **2018,** *1037*, 97-106.

54. Luo, X.; Li, L. *Anal. Chem.* **2017,** *89* (21), 11664-11671.

55. Luo, X.; Zhao, S.; Huan, T.; Sun, D.; Friis, R. M.; Schultz, M. C.; Li, L. *J Proteome Res* **2016,** *15* (5), 1602-12.

56. Ma, L.; Yang, Y.; Zhao, X.; Wang, F.; Gao, S.; Bu, D. *PLoS One* **2019,** *14* (1), e0209182.

57. Marcén, M.; Cebrián, G.; Ruiz-Artiga, V.; Condón, S.; Mañas, P. *Food Res. Int.* **2019,** *121*, 806-811.

58. Masujima, T. *Anal Sci* **2009,** *25* (8), 953-60.

59. Mattsson, A.; Kärrman, A.; Pinto, R.; Brunström, B. *PLoS One* **2015,** *10* (12), e0143780.

60. McGettigan, P. A. *Curr Opin Chem Biol* **2013,** *17* (1), 4-11.

61. Mounier, N.; Arrigo, A.-P. *Cell Stress Chaperones* **2002,** *7* (2), 167-176.

62. Mowry, K. L. *Cold Spring Harb. Protoc.* **2020,** *2020* (4), 095844-095844.

63. Nagana Gowda, G. A.; Raftery, D. *J. Magn. Reson.* **2015,** *260*, 144-60.

64. Nguyen, D. H.; Jaszczak, R. G.; Laird, D. J. *Curr. Top. Dev. Biol.* **2019,** *135*, 155-201.

65. Nicholson, J. K.; Lindon, J. C. *Nature* **2008,** *455* (7216), 1054-6.

66. Núñez, O.; Lucci, P. *Foods* **2020,** *9* (9), 1277.

67. Oesch, F.; Fabian, E.; Guth, K.; Landsiedel, R. *Arch. Toxicol.* **2014,** *88* (12), 2135-2190.

68. Onjiko, R. M.; Moody, S. A.; Nemes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (21), 6545-6550.

69. Onjiko, R. M.; Plotnick, D. O.; Moody, S. A.; Nemes, P. *Anal. Methods* **2017,** *9* (34), 4964-4970.

70. Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P. *J. Vis. Exp.* **2017,** (130).

71. Ortmayr, K.; Dubuis, S.; Zampieri, M. *Nat. Commun.* **2019,** *10* (1), 1841.

72. Palma, E.; Reyes-Ruiz, J. M.; Lopergolo, D.; Roseti, C.; Bertollini, C.; Ruffolo, G.; Cifelli, P.; Onesti, E.; Limatola, C.; Miledi, R.; Inghilleri, M. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (11), 3060-5.

73. Pan, N.; Rao, W.; Kothapalli, N. R.; Liu, R.; Burgett, A. W.; Yang, Z. *Anal. Chem.* **2014,** *86* (19), 9376-80.

74. Pompella, A.; Visvikis, A.; Paolicchi, A.; De Tata, V.; Casini, A. F. *Biochem. Pharmacol.* **2003,** *66* (8), 1499-503.

75. Portero, E. P.; Nemes, P. *Analyst* **2019,** *144* (3), 892-900.

76. Qian, M.; Wang, D. C.; Chen, H.; Cheng, Y. *Semin. Cell Dev. Biol.* **2017,** *64*, 143-149.

77. Raddassi, K.; T, W.; Sf, N.; Av, W. *Drug Designing: Open Access* **2017,** *06*.

78. Rai, V.; Mukherjee, R.; Ghosh, A. K.; Routray, A.; Chakraborty, C. *Arch. Oral Biol.* **2018,** *87*, 15-34.

79. Rasar, M. A.; Hammes, S. R. *Methods Mol. Biol.* **2006,** *322*, 17-30.

80. Richter, K.; Haslbeck, M.; Buchner, J. *Mol. Cell* **2010,** *40* (2), 253-66.

81. Ruan, L.; Zhou, C.; Jin, E.; Kucharavy, A.; Zhang, Y.; Wen, Z.; Florens, L.; Li, R. *Nature* **2017,** *543* (7645), 443-446.

82. Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. *J. Am. Soc. Mass Spectrom.* **2016,** *27* (12), 1897-1905.

83. Sender, R.; Fuchs, S.; Milo, R. *PLOS Biol.* **2016,** *14* (8), e1002533.

84. Shrestha, B. *Methods Mol. Biol.* **2020,** *2064*, 219-223.

85. Shrestha, B. *Methods Mol. Biol.* **2020,** *2064*, 1-8.

86. Shrestha, B.; Sripadi, P.; Reschke, B. R.; Henderson, H. D.; Powell, M. J.; Moody, S. A.; Vertes, A. *PLoS One* **2014,** *9* (12), e115173.

87. Sretarugsa, P.; Wallace, R. A. *Dev. Growth Differ.* **1997,** *39* (1), 87-97.

88. Storey, J. D. *Ann Stat* **2003,** *31* (6), 2013-2035.

89. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; 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.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007,** *3* (3), 211-221.

90. Sun, Y.; Jia, L. L.; Yu, W. L.; Yu, H. L.; Sheng, M. W.; Du, H. Y. *HBPD INT* **2018,** *17* (6), 496-501.

91. Szalay, M. S.; Kovács, I. A.; Korcsmáros, T.; Böde, C.; Csermely, P. *FEBS Lett.* **2007,** *581* (19), 3675-80.

92. Taylor, M. J.; Lukowski, J. K.; Anderton, C. R. *J. Am. Soc. Mass Spectrom.* **2021,** *32* (4), 872-894.

93. Thomason, K.; Babar, M. A.; Erickson, J. E.; Mulvaney, M.; Beecher, C.; MacDonald, G. *PLoS One* **2018,** *13* (6), e0197919.

94. Vailati-Riboni, M.; Palombo, V.; Loor, J. J., What Are Omics Sciences? In *Periparturient Diseases of Dairy Cows: A Systems Biology Approach*, Ametaj, B. N., Ed. Springer International Publishing: Cham, 2017; pp 1-7.

95. Van, Q. N.; Veenstra, T. D.; Issaq, H. J. *Curr. Urol. Rep.* **2011,** *12* (1), 34-40.

96. Vastag, L.; Jorgensen, P.; Peshkin, L.; Wei, R.; Rabinowitz, J. D.; Kirschner, M. W. *PLoS One* **2011,** *6* (2), e16881.

97. Wada, H.; Hatakeyama, Y.; Nakashima, T.; Nonami, H.; Erra-Balsells, R.; Hakata, M.; Nakata, K.; Hiraoka, K.; Onda, Y.; Nakano, H. *Sci. Rep.* **2020,** *10* (1), 2013.

98. Wade, D. *Chem Biol Interact* **1999,** *117* (3), 191-217.

99. Wang, D.; Bodovitz, S. *Trends Biotechnol.* **2010,** *28* (6), 281-290.

100. Wang, Q.; Gao, P.; Wang, X.; Duan, Y. *Sci. Rep.* **2014,** *4* (1), 6802.

101. Wang, R.; Zhao, H.; Zhang, X.; Zhao, X.; Song, Z.; Ouyang, J. *Anal. Chem.* **2019,** *91* (5), 3667-3674.

102. Wang, X.; Han, W.; Yang, J.; Westaway, D.; Li, L. *Analytica Chimica Acta* **2019,** *1050*, 95-104.

103. Ward, A. J.; Cooper, T. A. *J. Pathol.* **2010,** *220* (2), 152-163.

104. Welch, W. J.; Suhan, J. P. *J. Cell Biol.* **1985,** *101* (4), 1198-211.

105. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. *Metabolomics* **2008,** *4* (1), 81-89.

106. Wiley, W. C.; McLaren, I. H. *Review of Scientific Instruments* **1955,** *26* (12), 1150-1157.

107. Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S. *J. Chromatogr. A* **2015,** *1382*, 136-164.

108. Worley, B.; Powers, R. *Curr Metabolomics* **2013,** *1* (1), 92-107.

109. Wu, Y.; Li, L. *Anal. Chem.* **2012,** *84* (24), 10723-10731.

110. Wu, Y.; Li, L. *J Chromatogr A* **2016,** *1430*, 80-95.

111. Xu, W.; Chen, D.; Wang, N.; Zhang, T.; Zhou, R.; Huan, T.; Lu, Y.; Su, X.; Xie, Q.; Li, L.; Li, L. *Anal. Chem.* **2015,** *87* (2), 829-836.

112. Yao, H.; Yu, P.-C.; Jiang, C.-M. *RSC Adv.* **2020,** *10* (9), 4928-4941.

113. Ye, G.; Zhu, B.; Yao, Z.; Yin, P.; Lu, X.; Kong, H.; Fan, F.; Jiao, B.; Xu, G. *J. Proteome Res.* **2012,** *11* (8), 4361-4372.

114. Zeng, S. L.; Sudlow, L. C.; Berezin, M. Y. *Expert Opin Drug Discov* **2020,** *15* (1), 39-52.

115. Zenobi, R. *Science* **2013,** *342* (6163), 1243259.

116. Zhang, A.; Sun, H.; Wu, X.; Wang, X. *Clin. Chim. Acta* **2012,** *414*, 65-69.

117. Zhang, L.; Hatzakis, E.; Patterson, A. D. *Curr. Pharmacol. Rep.* **2016,** *2* (5), 231-240.

118. Zhao, S.; Dawe, M.; Guo, K.; Li, L. *Anal. Chem.* **2017,** *89* (12), 6758-6765.

119. Zhao, S.; Li, H.; Han, W.; Chan, W.; Li, L. *Anal. Chem.* **2019,** *91* (18), 12108-12115.

120. Zhao, S.; Li, L. *Anal. Chem.* **2018,** *90* (22), 13514-13522.

121. Zhao, S.; Luo, X.; Li, L. *Anal. Chem.* **2016,** *88* (21), 10617-10623.

122. Zhou, R.; Li, L. *Methods Mol. Biol.* **2014,** *1198*, 127-36.

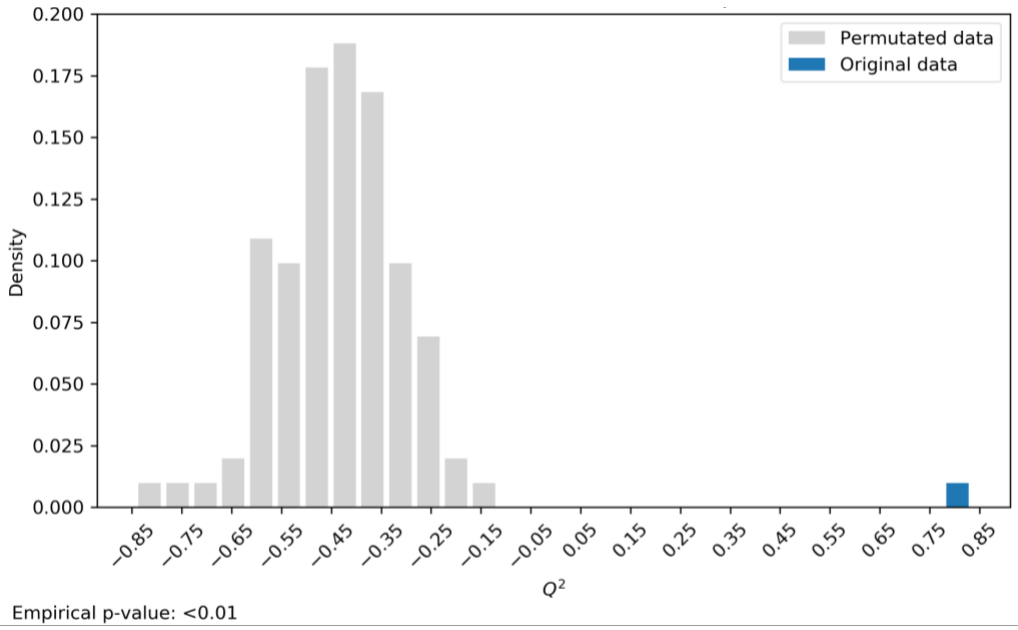
123. Zhou, R.; Li, L. *J. Proteomics* **2015,** *118*, 130-9.

124. Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. *Anal. Chem.* **2014,** *86* (10), 4675-4679.

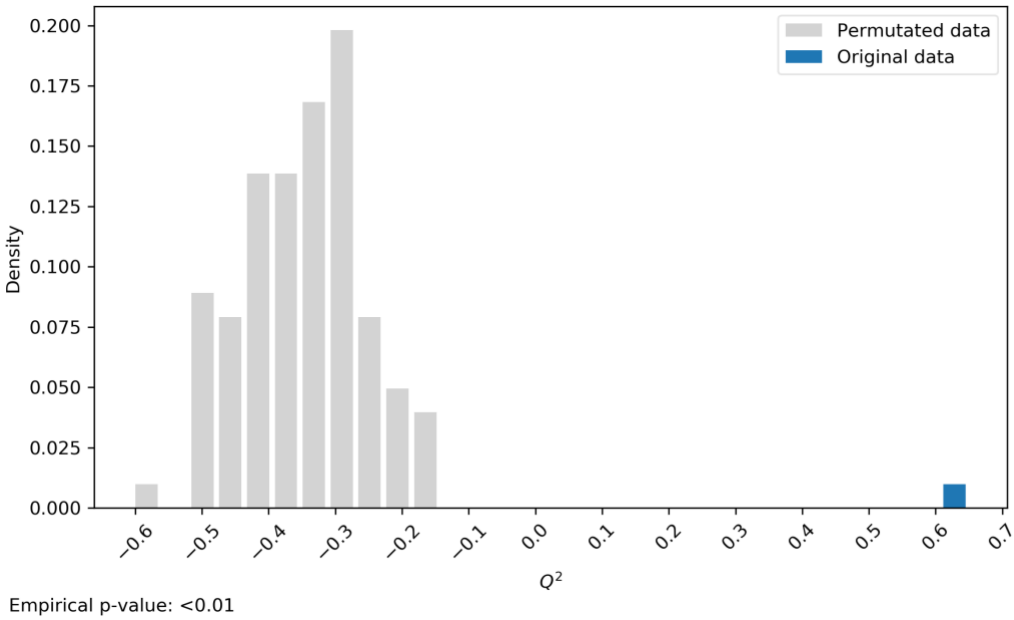
125. Zwart, R.; Mazzo, F.; Sher, E. *Drug Discov. Today* **2019,** *24* (2), 533-543.

Appendix

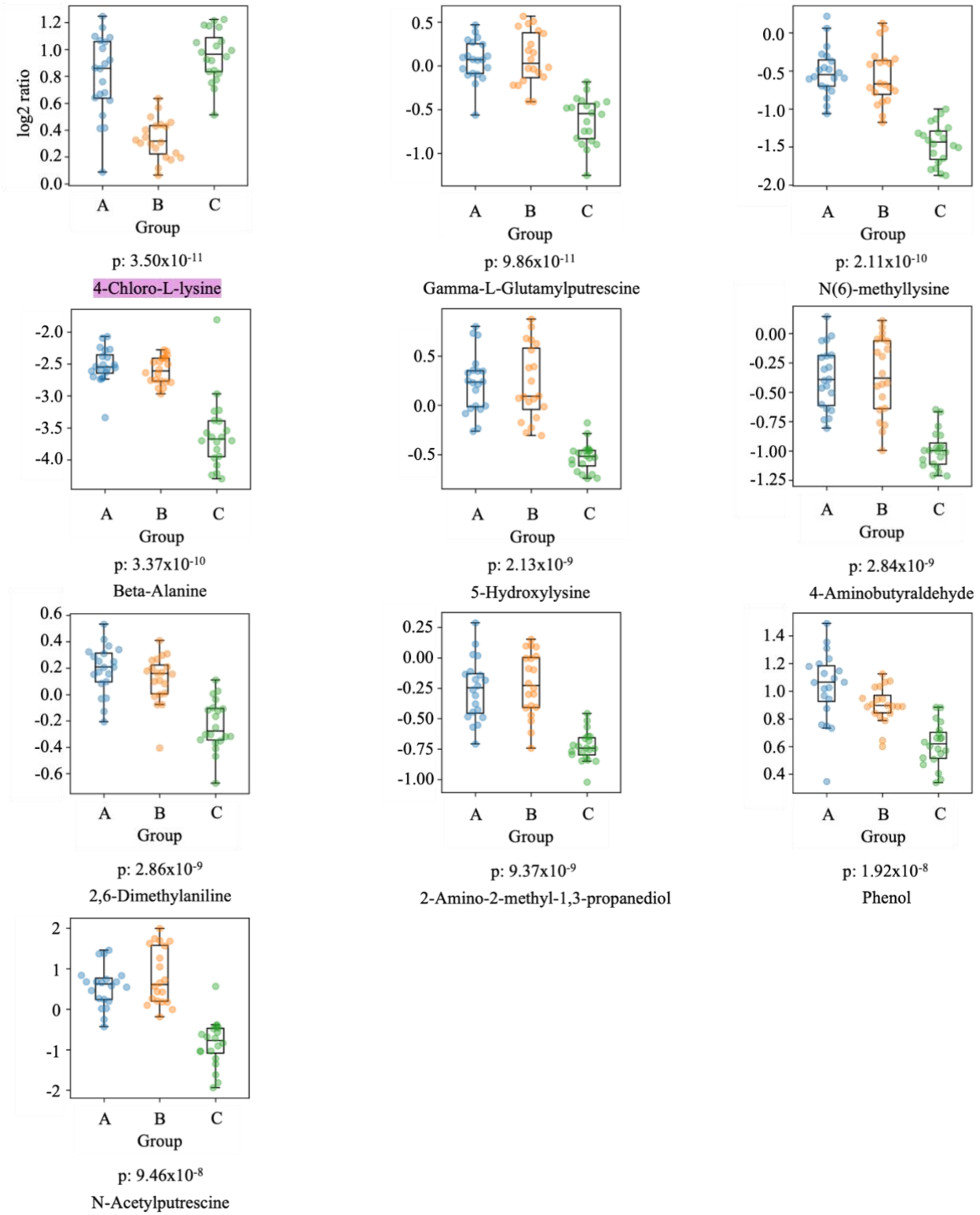
Chapter 3 – Chemical Isotope Labeling LC-MS for Studying the Metabolic Response of Single Cells to Heat Stress



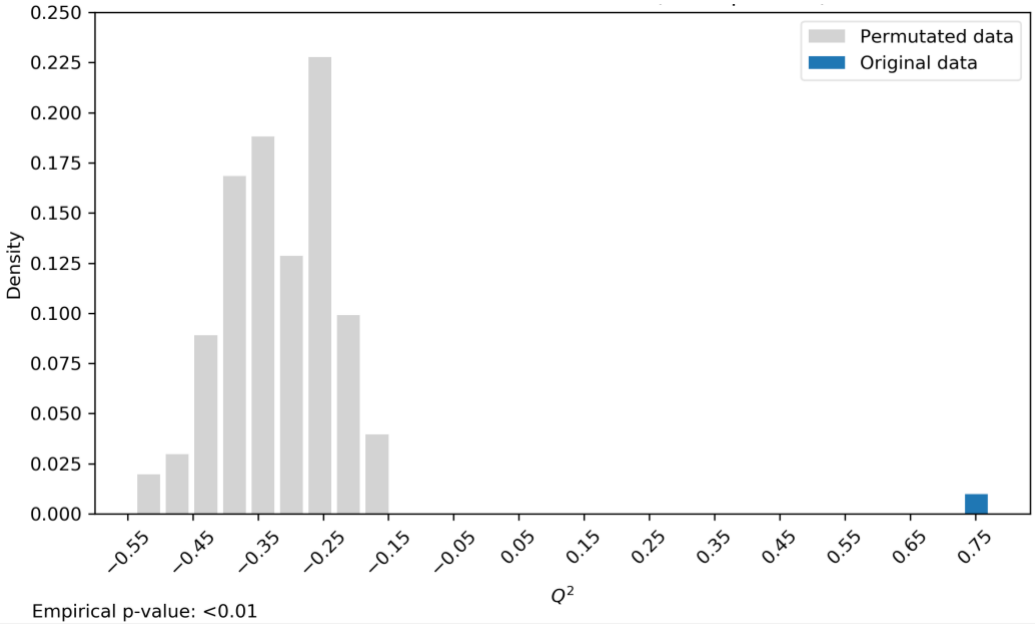
Appendix Figure 3. 1 PLS-DA model validation result of the single-cell metabolome data set in heat stress study with 100 permutations.



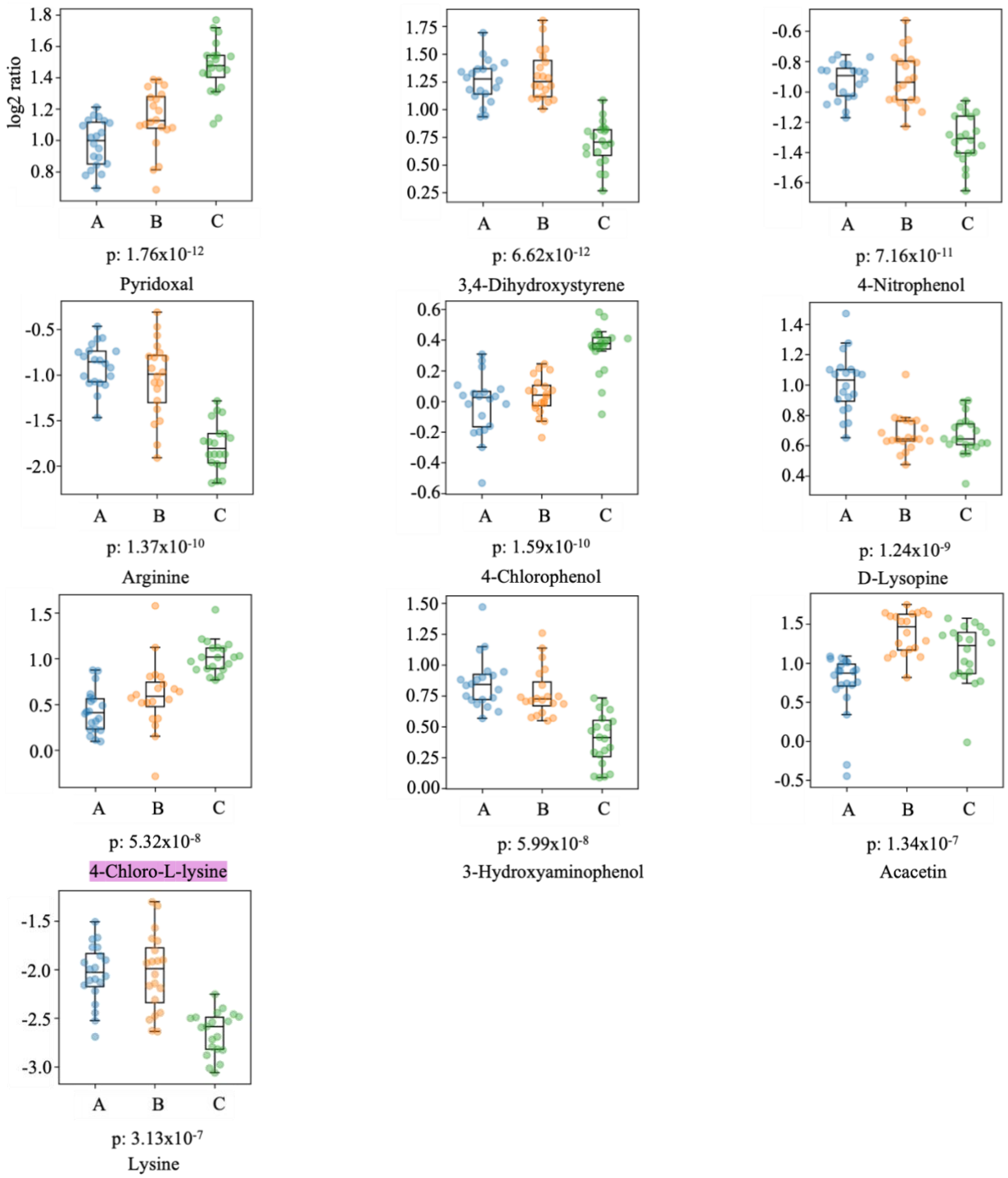
Appendix Figure A4. 1 PLS-DA model validation result of the amine and phenol metabolome of single cells in animal 1 with 100 permutations.



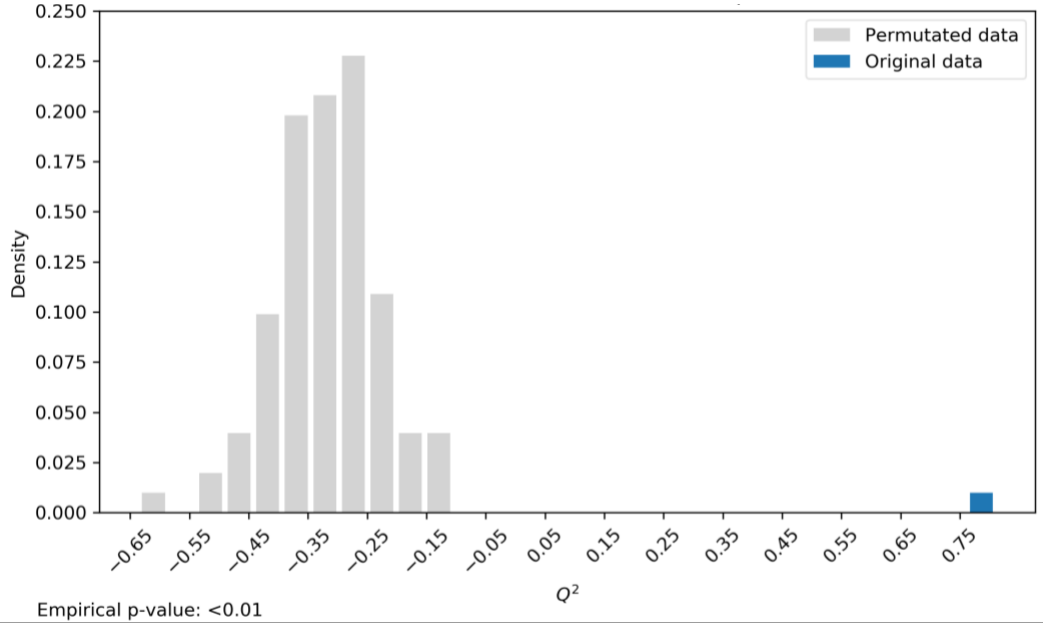
Appendix Figure A4. 2 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for amine and phenol metabolomics of cells of group A, B and C in animal 1.



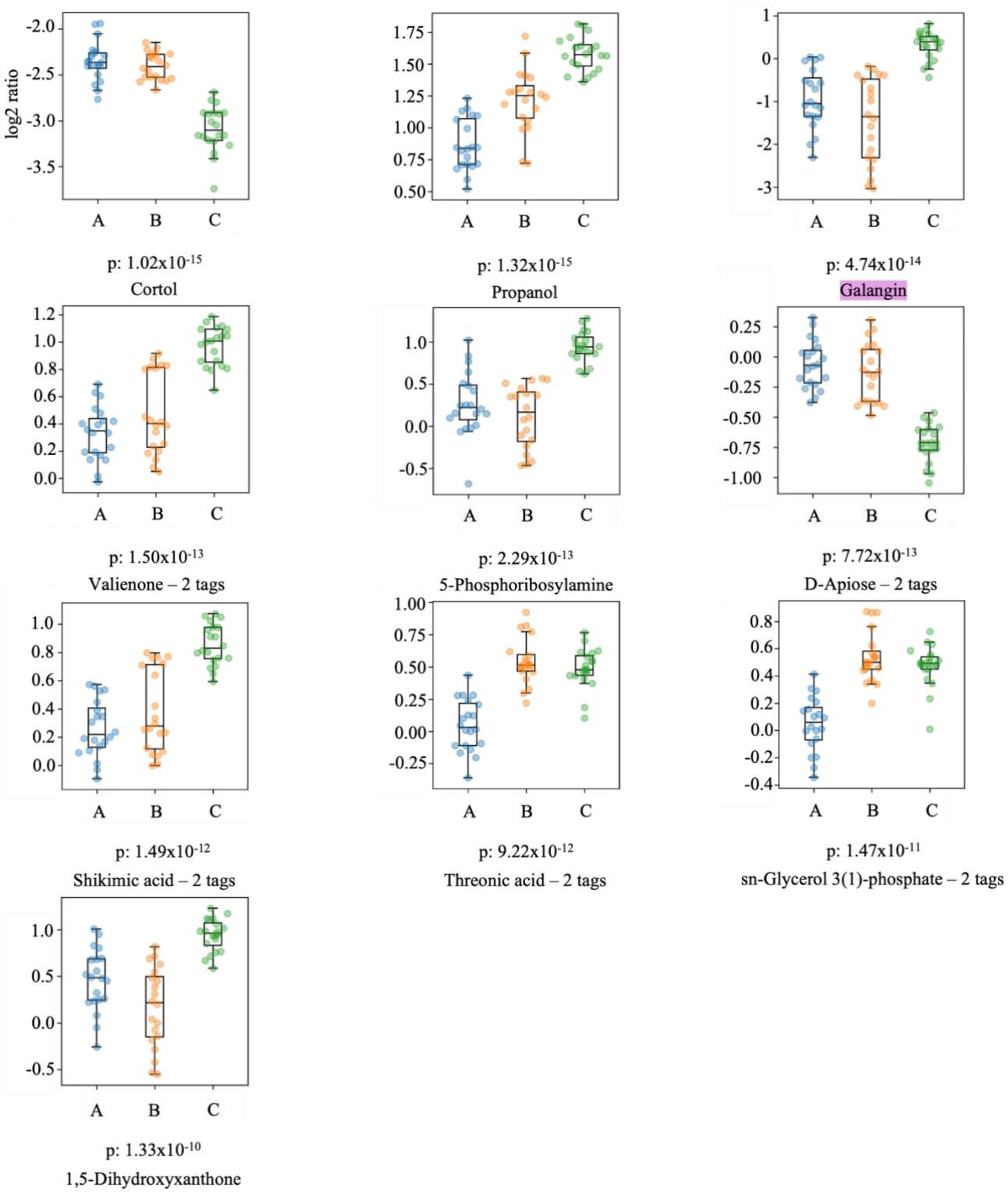
Appendix Figure A4. 3 PLS-DA model validation result of the amine and phenol metabolome of single cells in animal 2 with 100 permutations.



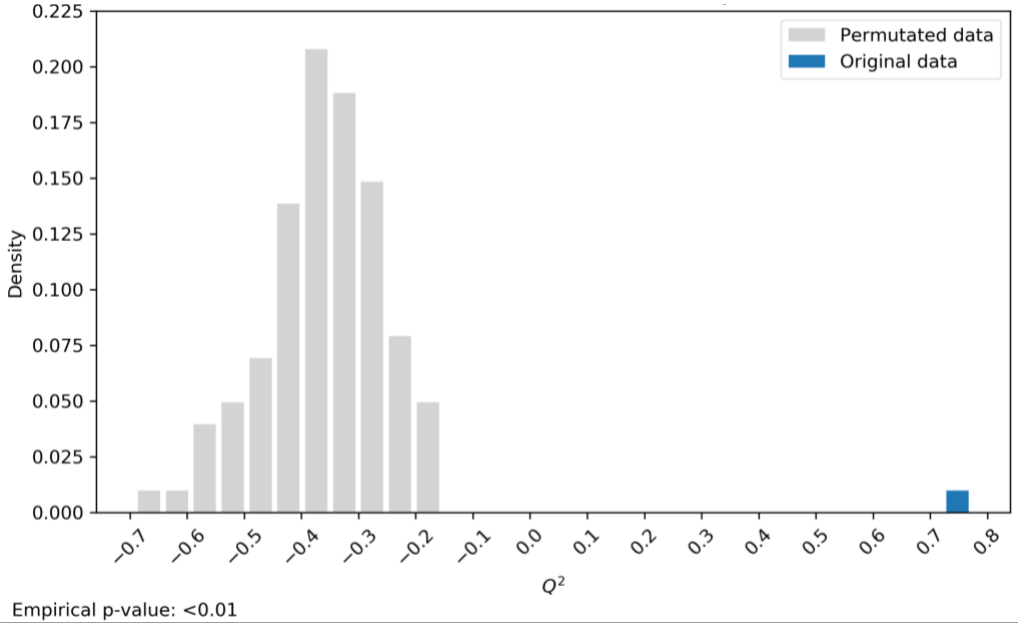
Appendix Figure A4. 4 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for amine and phenol metabolomics of cells of group A, B and C in animal 2.



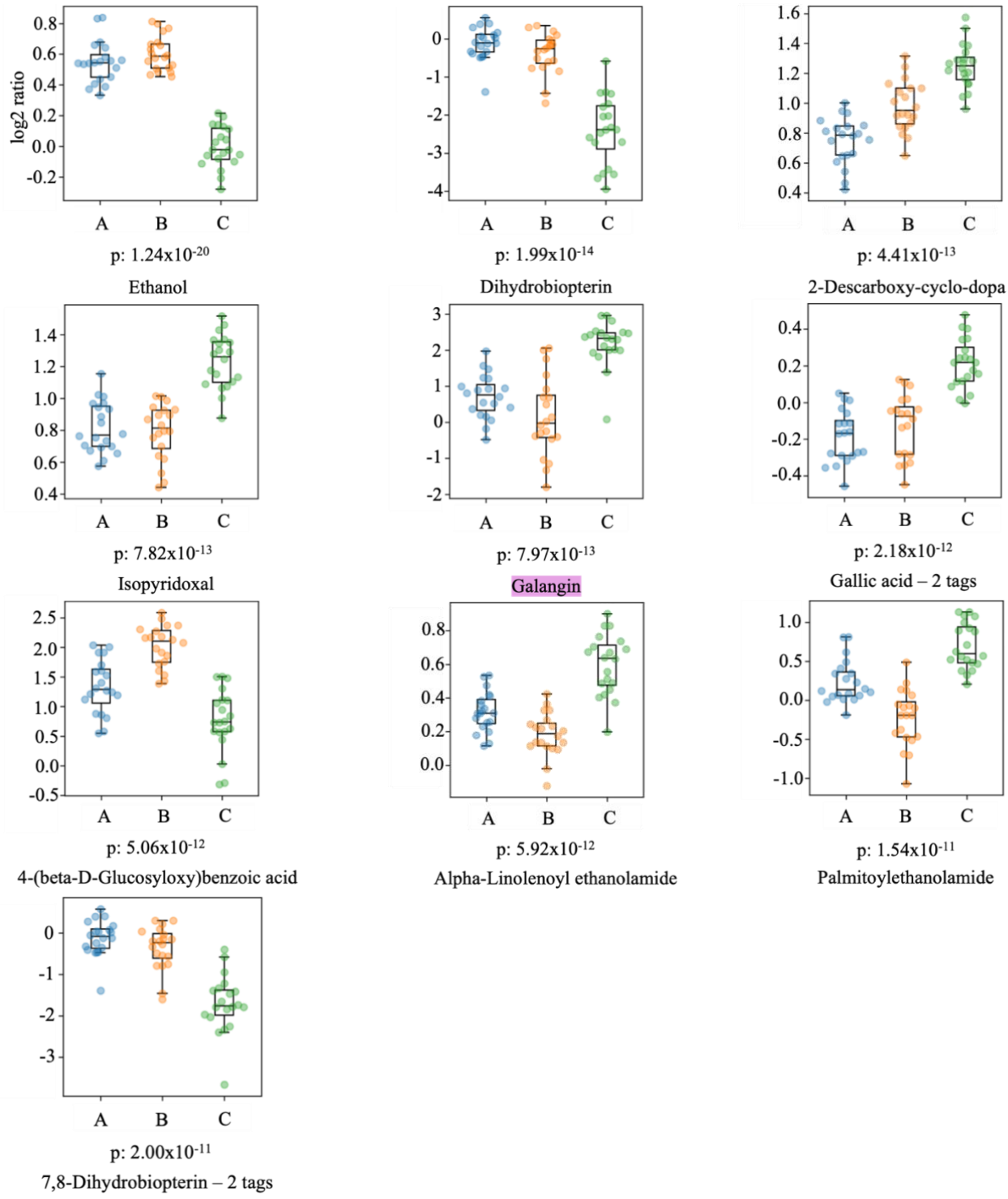
Appendix Figure A4. 5 PLS-DA model validation result of the hydroxyl metabolome of single cells in animal 1 with 100 permutations.



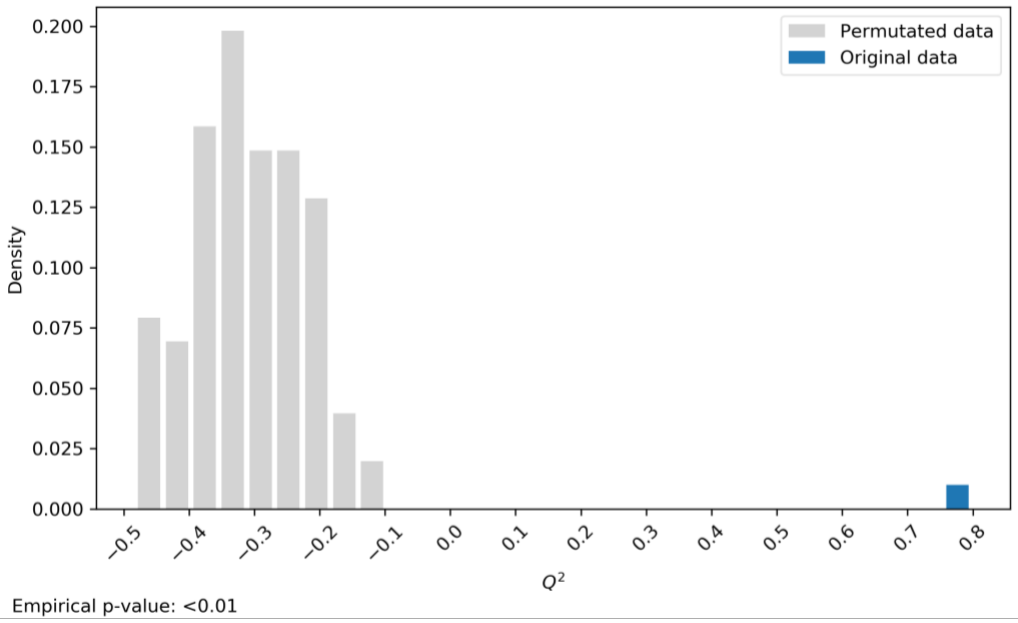
Appendix Figure A4. 6 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for hydroxyl metabolomics of cells of group A, B and C in animal 1.



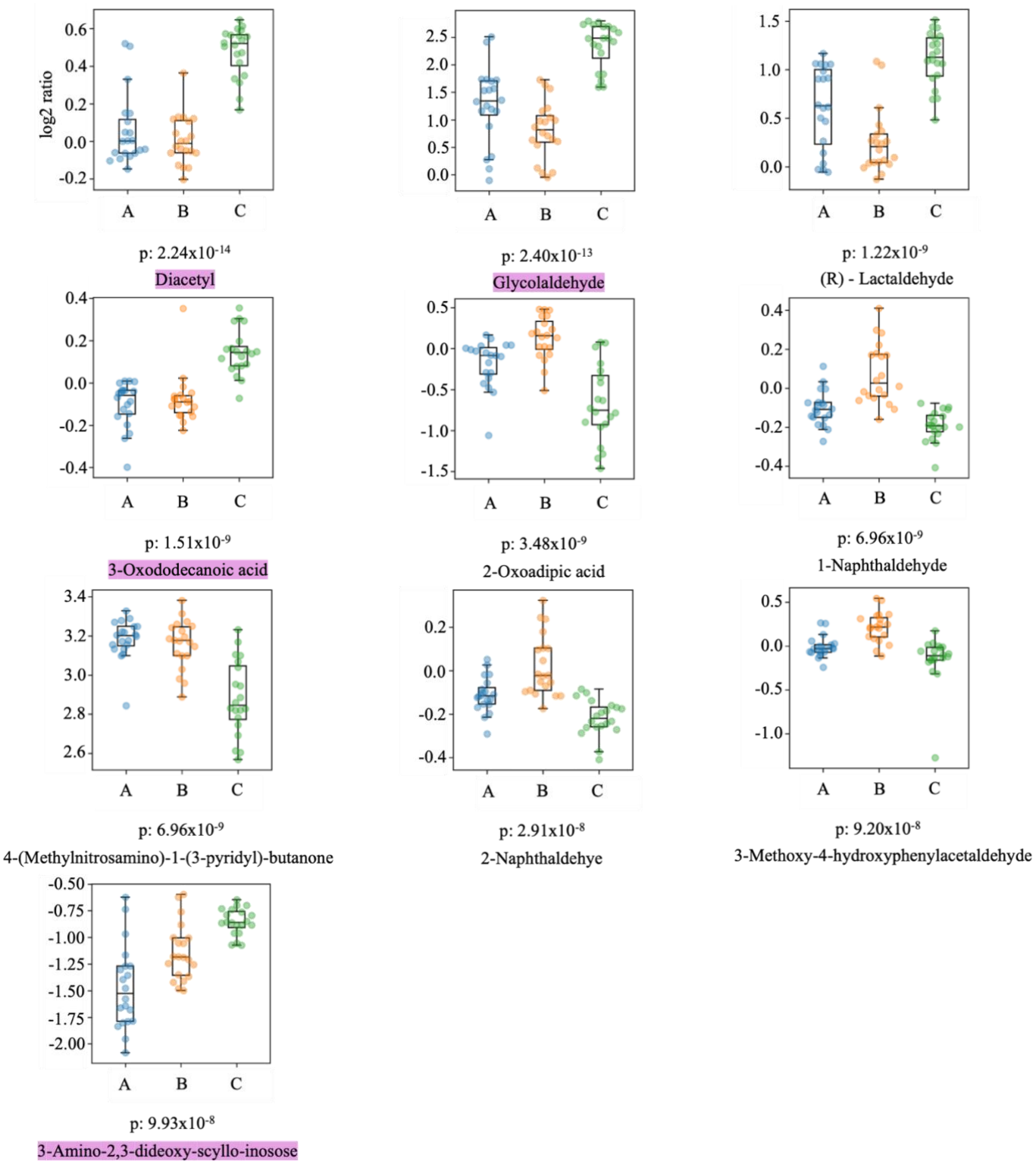
Appendix Figure A4. 7 PLS-DA model validation result of the hydroxyl metabolome of single cells in animal 2 with 100 permutations.



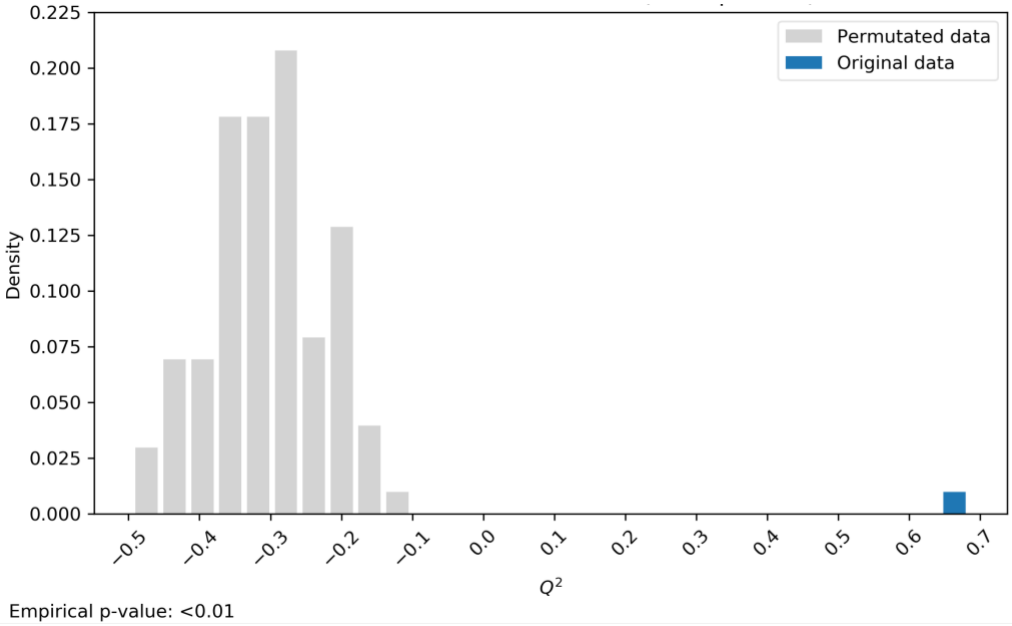
Appendix Figure A4. 8 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for hydroxyl metabolomics of cells of group A, B and C in animal 2.



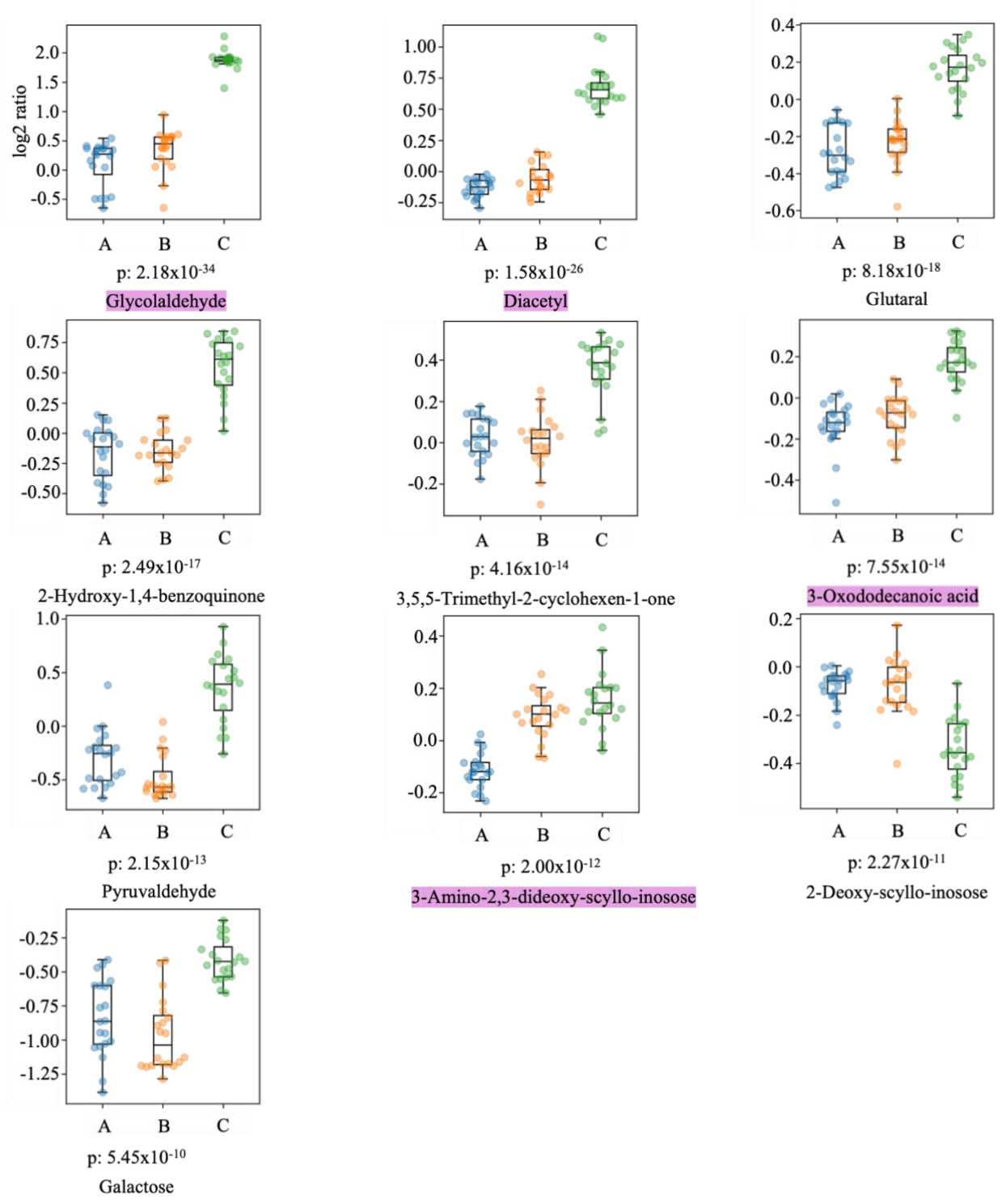
Appendix Figure A4. 9 PLS-DA model validation result of the carbonyl metabolome of single cells in animal 1 with 100 permutations.



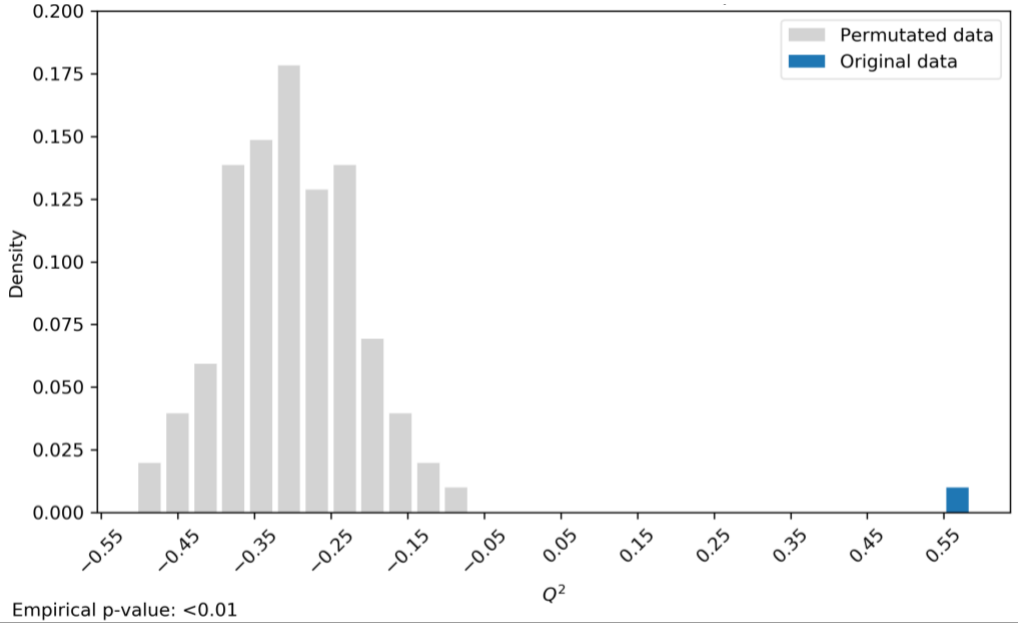
Appendix Figure A4. 10 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carbonyl metabolomics of cells of group A, B and C in animal 1.



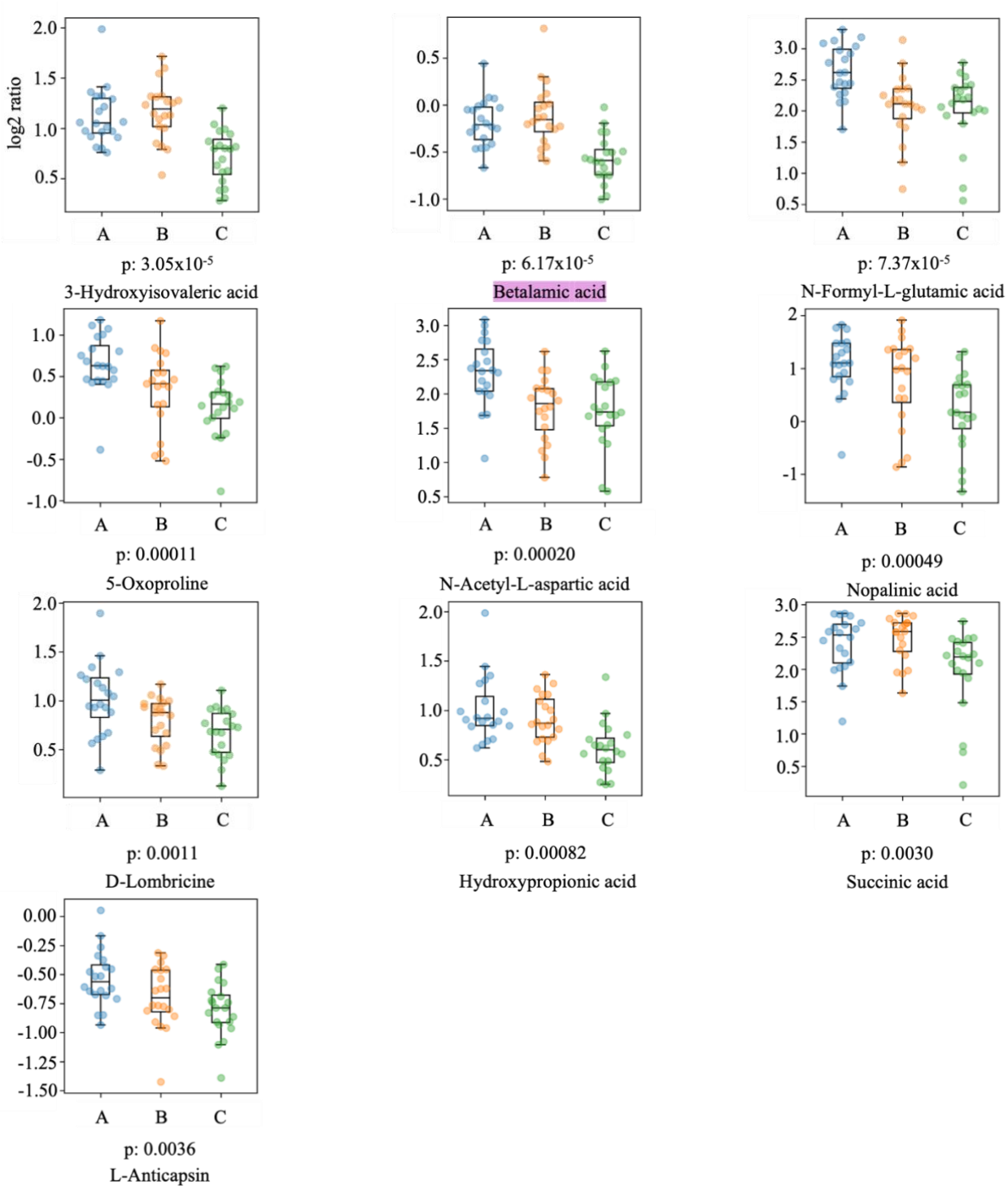
Appendix Figure A4. 11 PLS-DA model validation result of the carbonyl metabolome of single cells in animal 2 with 100 permutations.



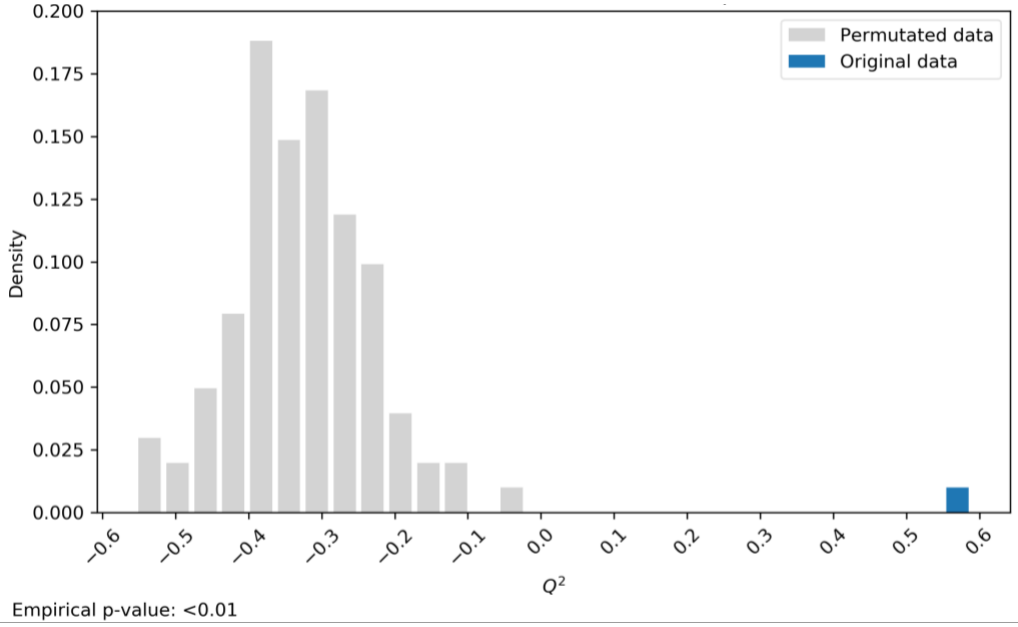
Appendix Figure A4. 12 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carbonyl metabolomics of cells of group A, B and C in animal 2.



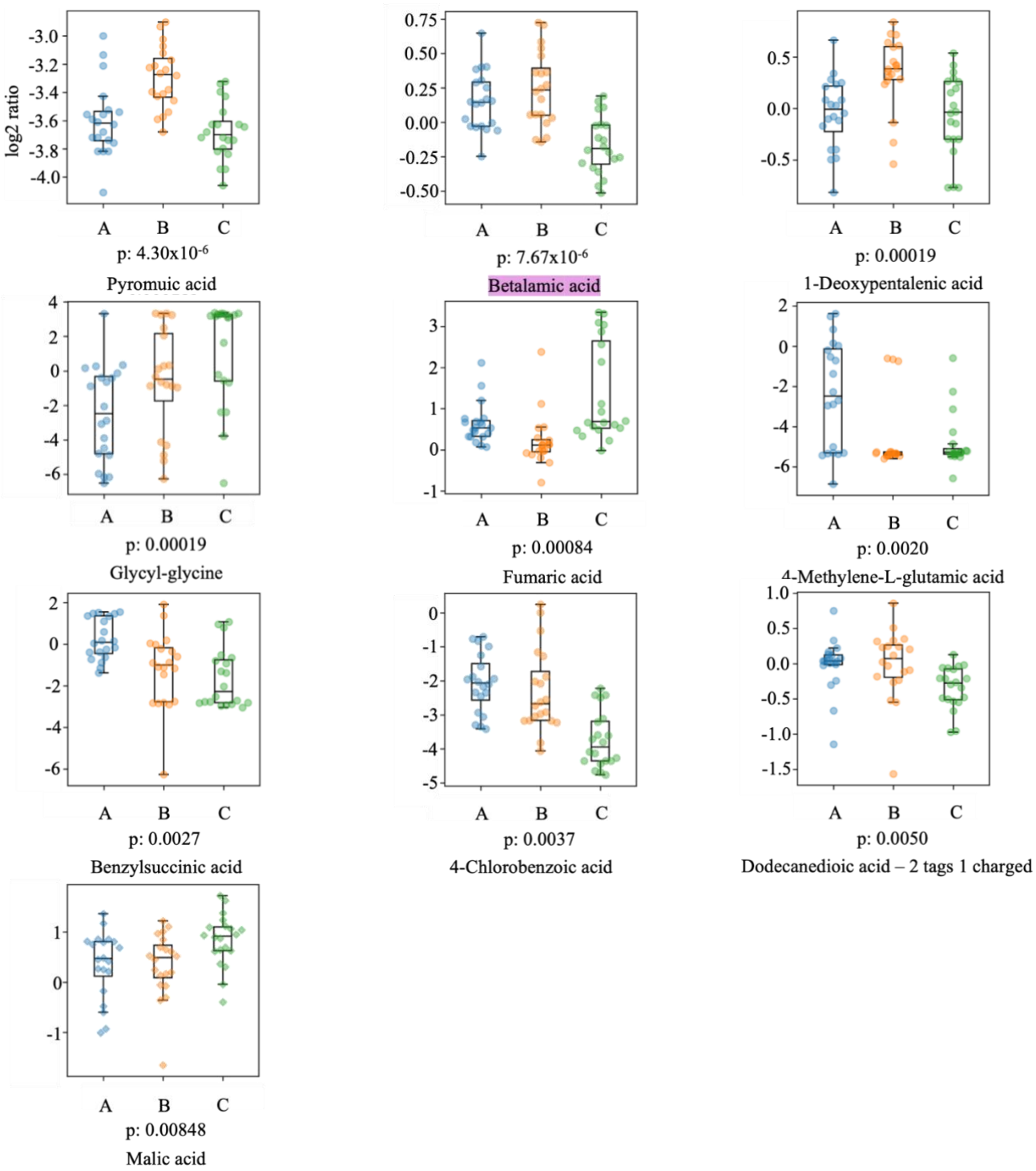
Appendix Figure A4. 13 PLS-DA model validation result of the carboxylic acid metabolome of single cells in animal 1 with 100 permutations.



Appendix Figure A4. 14 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carboxylic acid metabolomics of cells of group A, B and C in animal 1.



Appendix Figure A4. 15 PLS-DA model validation result of the carboxylic acid metabolome of single cells in animal 2 with 100 permutations.



Appendix Figure A4. 16 Box plots for ten of the most significant metabolites (lowest p-values) found by ANOVA for carboxylic acid metabolomics of cells of group A, B and C in animal 2.