Method Development of Chemical Isotope Labeling of Cellular Extract and Biofluids on Metabolomics Profiling and Its Application for Disease Biomarker Discovery

by

Xinyun Gu

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

Master of Science

Department of Chemistry
University of Alberta

Abstract

Metabolomics aims at studying all the small molecules in biological samples. Compared to transcriptome and proteome, metabolome is highly sensitive to diverse individuals and environmental factors. Thus, metabolomics study has been used to understand individual variations caused by genes, exosomes, diseases, and other metabolic activity and became a very useful tool in biomarker discovery. In order to detect and quantify certain biomarkers from the whole metabolome, high coverage profiling and accurate quantification are essential. Conventional approaches combined several complementary methods to improve low coverage. Our chemical isotope labeling (CIL) LC-MS stands out because of a simplified platform and an overall promoted analytical performance of metabolites.

My research focuses on utilizing CIL LS-MS methods to profile an amine/phenol submetabolome to study the impact of diseases on biofluids such as cell extracts and serum samples. In the first part of my thesis, we used CIL LC-MS to evaluate and compare two cell harvest methods (physical scraping and trypsinization) and two cell lysis methods (glass-bead-assisted lysis and freeze-thaw-cycle lysis). In the seconded part, we applied the optimized cell harvest protocol to the cell extract, followed by routine LC-MS analysis. By combining the cell results and serum results, we found several metabolites that could be potential biomarkers for Hyper IgE syndrome. In another biomarker discovery study of metformin treatment of diabetes (Chapter 4), CIL LC-MS was applied to reveal the metabolome differences among control, obese, diabetes, and metformin treated groups.

Preface

A version of Chapter 2 has been published as "Development of a Simple and Efficient Method of Harvesting and Lysing Adherent Mammalian Cells for Chemical Isotope Labeling LC-MS-Based Cellular Metabolomics" Luo, X.; Gu, X.; Li, L. Anal. Chim. Acta. 2018, 1037, 97-106. Samples were obtained from University of Alberta. I was responsible for running samples, performing data processing, and statistical analysis. Xian Luo was involved in the experimental design and writing the manuscript. Dr. Liang Li supervised the project and edited the manuscript.

A version of Chapter 3 has been submitted to The Journal of Immunology as "Metabolomics distinguishes DOCK8 deficiency from atopic dermatitis: A biomarker discovery". This project was collaborated with Dr. Anas M. Abdel Rahman. I was responsible for the experiment design, data collection, and data analysis. Xian Luo was helped with experimental design and data collection. Minnie Jacob was responsible for collecting the samples, conducting the experiments and drafting the manuscript. Dr. Liang Li supervised the project and edited the manuscript.

Chapter 4 was a collaboration work with Dr. Anas M. Abdel Rahman who provided. The samples, while I was responsible for the experiment design, acquiring data, analyzing data, and drafting the manuscript. This project is under the supervision of Dr. Liang Li.

Acknowledgement

I would like to give my upmost appreciation to my supervisor, Dr. Liang Li, for his immense guidance and support throughout my entire graduate studies in leading me to the field of omics studies and LC-MS analysis. His critical thinking and scientific insight have been invaluable in cultivating me to become an analytical scientist. His advice and support paved the way for my future career.

I would like to extend my thanks to my supervisory committee members, Dr. James Harynuk and Dr. John Vederas, for their guidance and help.

I cannot express enough appreciation to Dr. Anna Jordan who gave me enormous help during my thesis editing.

Furthermore, I would like to especially thank my collaborator, Dr. Anas M. Abdel Rahman, for his advice and help in all the biomarker discovery work. His deep and broad knowledge in the clinical area has benefited my research work greatly.

I would like to thank all my group members for their encouragement and selfless help over three years. In particular, I would like to thank Wei Han, Xian Luo, Shuang Zhao, for their extreme patience and mentorship in my early years. In no particular order, thanks to: Drothea Mung, Wan Chan, Hao Li, Yunong Li, Kevin Hooton, Xiaohang Wang, Adriana Zardini Buzatto, Minglei Zhu, Barinder Bajwa, and Yingwen Wang.

In addition, I would like to express my appreciation fot the support from all the staff in the Department of Chemistry. I'm especially thankful to Gareth Lambkin of the Biological Services Facility for his kind assistance in the use of bio-hazard level 2 fume hood. I also want to acknowledge Dr. Randy Whittal, Jing Zheng and Bela Reiz of the Mass Spectrometry Facility for their professional support.

Finally, I would like to thank my family and friends for their support, reassurance, and understanding through the course of my degree. The utmost thanks go to the some of the most important people in my life, my parents Jianghong Gu and Linghong Gou. Without all of you, I would not have been able to get this far in my academic career. To my boyfiend, Xiaoyu Zhang, for your unwavering patience and support throughout the stressful times.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

%RSD	Percent relative standard deviation
ACN	Acetonitrile
AD	Atopic dermatitis
AUC	Area-under-the-curve
CE	Capillary electrophoresis
CIL	Chemical Isotope Labeling
CSF	Cerebrospinal fluid
DCM	Dichloromethane
EML	Evidence-based Metabolome Library
ESI	Electrospray ionization
FA	Formic acid
FC	Fold change
GC	Gas chromatography
HPLC	High performance liquid chromatography
HMDB	Human metabolome database
HIES	Hyper IgE syndrome
HILIC	Hydrophilic interactions liquid chromatography (HILIC)
LC	Liquid chromatography
MeOH	Methanol
min	Minutes
MPA	Mobile phase A
MPB	Mobile phase B

MS	Mass spectrometry
NMR	Nuclear magnetic resonance spectroscopy
PBS	Phosphate buffered saline
PC1	First principal component
PC2	Second principal component
PCA	Principal component analysis
PDA	Photodiode array
PLS-DA	Partial least squares discriminant analysis
QC	Quality control
QTOF	Quadrupole time of flight
RP	Reversed phase
RT	Retention time
TIC	Total ion count
T2D	Type two diabetes
USD	United States Dollar
UV	Ultraviolet
WHO	World Health Organization

Chapter 1

Introduction

1.1 Metabolomics

Metabolomics is an emerging science, which studies the chemical processes involving metabolites by analyzing their characterization as well as their metabolism in biological systems. (http://metabolomicssociety.org). Unlike genomics and proteomics, which focus on characterizing the profiles of genes and proteins, respectively, metabolomics mainly emphasize profiling small-molecule metabolites (metabolic profiling).[1] Compared to genes, transcripts, and proteins, metabolites are considered as a "spoken language"; it then combines information from both genetics and environmental influences while directly related to the phenotype.[2] Metabolites as the downstream products of pathways. The difference between metabolites and proteins or genes is that metabolites are more species independent, that is they are less varied among different organisms. Thus, one established metabolomics methods for a certain organism can be applied to another organism.[3] In addition, alternations at the genes and transcripts levels may not affect the protein translation step. Plus, changes that happen in proteins are not translated always into cell biochemistry and phenotype level. Therefore, protein concentrations are not related necessarily to activities.[4] Compared to the alternations in transcriptome and proteome, metabolome changes are amplified, resulting in increased sensitivity.[5] Subsequently, a highly dynamic metabolome has the ability to reflect continuous fluctuations of both metabolic and

signaling pathways. Besides, a metabolome is highly sensitive to diverse individuals and environmental factors. All of these advantages enable metabolomics studies to detect subtle changes, evaluate a variety of complex pathways simultaneously, and allow invisible changes to be measurable (for example, morphological changes).[6]

1.2 Metabolomics Profiling and Biomarker Discovery

Metabolites profiling tries to analyze the whole metabolome, relating to their chemical nature or metabolic pathways simultaneously. The huge profile of the large number of metabolites enables the possibility of exploring unknown biochemical pathways and biological features.[7] By profiling the huge number of compounds, metabolites that are dysregulated in a specific biological status can be detected and measured.[7] This advantage makes untargeted metabolomics one of the most powerful methods for the study of biomarker discovery.[7] Biomarkers are broadly considered as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention".[8] The basis of biomarker discovery is that due to the disturbed biochemical pathways caused by disease, the concentration of certain metabolites are affected accordingly.[9,10] One of the most well-known biomarkers is glucose, which is a biomarker for type two diabetes patients. According to Diabetes Canada, a higher than 7.0 mmol/L of fasting glucose level would be considered an indication of diabetes. Furthermore, biomarkers often are related directly to the onset of the disease, and studying the biological processes behind them may deepen our understanding of the disease mechanisms. Importantly, diagnostic biomarkers sometimes can demonstrate detectable changes before the disease symptoms become noticeable, enabling early

diagnosis of the disease. With preventive interventions, the onset and progression of the disease may be delayed dramatically.[11,12]

1.3 Major Metabolomics Platforms for Metabolomics

Most common platforms used for metabolomics analysis, including Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS), are chosen due to their relatively high metabolome coverage and abundant database resources. NMR has been one of the very first platforms for metabolomics studies[13] because it can provide detailed structural information[14,15] and requires relatively simple sample preparation.[16,17] The NMR technique has been highly robust and reproducible because of low instrument drift.[18] However, the drawbacks of NMR analysis are low sensitivity, typically only detecting metabolites as low as the micromolar range, requiring a larger sample volume, plus the complexity of interpreting spectral information for complex mixtures.[16,19] MS has much more sensitivity than NMR and can provide information about the exact mass as well as fragmentation patterns for metabolites.[20]

Ion source and mass analyzer are two main parts of a mass spectrometer. Different ionization techniques are applied for different purposes. For instance, electrospray ionization (ESI) is used often for acquiring molecular information on metabolites, peptides, and proteins with minimal fragmentation.[21] The major ionization procedures in ESI are: (1) eluent from LC flows through a capillary with a high voltage applied and then breaks into highly charged droplets; (2) with heated dry gas blowing, the solvent in the droplets is evaporated, further shrinking the droplet; (3) the droplet disintegrations are repeated, and small “offspring” droplets are formed; (4) the electrostatic force of the droplets becomes high, and the solute ions “escape” from the surface of the droplets to the gas phase.

Various mass analyzers also can be used for the detection of metabolites, and they mainly differed in mass range, sensitivity and resolution. Quadrupole Time of Flight (QTOF) stands out due to its high resolution, over 20,000.

Separation techniques are incorporated usually in a MS-based metabolomics study to reduce the complexity of the sample and improve the detection ability. Gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE) are techniques commonly incorporated with MS. Among all the platforms, the usage of liquid chromatography mass spectrometry (LC-MS) to perform metabolomics studies expanded rapidly due to several unique advantages, including high sensitivity, good compatibility with the majority of samples types, and accurate quantification ability. With LC-MS, different column chemistries can be used for different separation requirements. For instance, reversed phase liquid chromatography (RPLC) mainly serves for the separation of non-polar compounds, and hydrophilic interactions liquid chromatography (HILIC) mostly serves for the analysis of polar compounds. To obtain better coverage, both RPLC and HILIC separation are acquired to avoid losing metabolites information. Chemical isotope labeling (CIL) LC-MS is one of the metabolomics platforms that uses designed chemical labeling reagents to react with metabolites, forming metabolite derivatives.[22] By altering the properties of metabolites, higher coverage, less ion suppression, and ability for quantities analysis can be achieved.

1.4 Chemical Isotope Labeling in LC-MS based Metabolomics

The ultimate goal of metabolomics is to provide qualitative and quantitative information for as many metabolites as possible. To achieve this goal, we need methods with high metabolite coverage, sensitive detection, accurate quantification, and a confident ability of unknown

metabolite identification. However, several obstacles must be solved in order to obtain the ultimate goal. First, metabolites in the biological samples have a large diversity in both physical and chemical properties. There is no single technique or analytical platform that can be used for all these metabolites.[23] The complementary use of many platforms for a comprehensive profiling of a metabolome is required (such as combining RPLC and HILIC for the separation step). Besides, ion suppression is always an issue in the detection of complex mixtures using MS.[7] As mentioned above, the ionization efficiency is the major factor determining the intensity, thus affecting the concentration. In ESI, during the ionization step (4), ions on the surface of the shrinking droplet get easier to ionize than those in the center of the droplet. This heterogeneous ionization efficiency leads to ion suppression. However, with the help of a proper separation method, ion suppression can be reduced. To resolve these two obstacles, a conventional LC-MS approach would combine RPLC LC-MS and HILIC LC-MS together; but double instrument time will be needed as well as sample preparation.

In our lab, a strategy called "divide and conquer" is applied to divide the metabolome into four major sub-metabolomes, based on functional groups. Chemical Isotope Labeling (CIL) is used to target a particular functional group (sub-metabolome) of interest. For a sub-metabolome containing an amine/phenol functional group, dansyl chloride (5-(dimehylamino)-napthalene-1-sulfonyl chloride, DnsCl) is chosen as the labeling reagent. With the derivatization, the complexity of the metabolome is decreased dramatically. The aromatic naphthyl part increases the hydrophobicity of the labeled metabolites, thus only RPLC is needed for the separation step. Additionally, its tertiary amine enhances chargeability, further enhancing ionization efficiency. DnsCl results in a 10–1000 fold of improvement in sensitivity and better retention for hydrophilic compounds on the RPLC column.

In addition, the two methyl groups on the tertiary amine of DnsCl can introduce a 13C isotope into the labeled metabolite, serving as an internal standard for quantification. In the CIL system, experiment samples will be labeled by 12C-DnsCl, whereas an internal reference sample will be labeled by 13C-DnsCl. After mixing and LC-MS analysis, peak pairs containing a lighter peak and a heavier peak will be observed in the chromatogram. The ratio between the two peaks can be used for relative or absolute quantification of metabolites. The protocols of using DnsCl for metabolites profiling and biomarker discovery in various samples have been reported, including cell extracts, urine, serum, sweat, and cerebrospinal fluid.

1.5 Workflow for Metabolomics

1.5.1 Sample Preparation

In our lab, a typical LC-MS metabolomics study requires sample preparation, chemical isotope labeling, normalization by LC-UV, LC-MS analysis, and data processing. Sample preparation is heavily dependent on sample types. Common sample types are cell extract, blood, urine, cerebrospinal fluid, saliva, etc. For cellular metabolomics, additional sample handling steps, including cell harvest and metabolism quenching, are required usually. Other biofluids such as blood which is protein enriched, needed to go through protein precipitation using solvents such as acetonitrile, methanol, or acetone. Besides protein precipitation, ACN and methanol also can be used for metabolites extraction.

1.5.2 Chemical Isotope Labeling

As described above, each individual sample is labeled by the 12C-DnsCl, and the pooled sample, which is the mixture of all the sample aliquots, is labeled by the 13C-DnsCl. Then, the mixed sample is analyzed by LC-MS. For each metabolite, a peak pair is detected instead of a single mass peak. The light peak of the pair represents the 12C-labeled metabolite from the individual sample, and the heavy peak of the pair is the 13C-labeled metabolite from the pooled sample. The distance between the two peaks equals to 2.00671 Da. The relative concentration is measured by calculating the intensity ratio of the two peaks in a pair. Consequently, every metabolite has a corresponding internal reference to accurately measure its relative concentration. Although the quantification is relative, the information is adequate enough for metabolomics analysis to find the metabolites with significant changes. Absolute quantification of confirmed biomarker candidates can be conducted afterward. The dansyl-labeling platform detects metabolites in forms of peak pairs instead of single mass peaks, making it easier to differentiate metabolites from the background noise peaks.

1.5.3 Sample Normalization

Sample normalization is an essential step for quantitative analysis. Especially in biomarker discovery study, concentration differences between samples should be excluded, and only the variation caused by biological differences should be kept. For instance, urine can have large concentration differences because of factors such water intake, dehydration, diet, exercise activity, and sweating. Thus, injecting the same amount of sample may not give an accurate quantitative analysis. Generally, there are two types of sample normalization methods: pre-acquisition normalization and post-acquisition normalization. In pre-acquisition normalization,

the total concentration of metabolites of each sample is determined, and by adjusting the injection volume, equal amount of samples will be analyzed by analytical platforms.[24–26] In post-acquisition normalization, samples are loaded on analytical platforms without controlling the sample amount and statistical adjustments are done later .[27] In most situations, the total signal intensity is proportional to the total concentration of the metabolite. Thus, the concentration of each individual metabolite is normalized, based on the total intensity of the signal. Creatinine is a well-known pre-acquisition normalization reference for adjusting urine concentration.[28]

1.5.3 Data Analysis

Two common statistical strategies for biomarker discovery are: uni-variate analysis and multi-variate analysis. The uni-variate analysis studies focus on individual metabolites, while the emphasis of the multi-variate analysis focus more on the entire sample. The two approaches complement each other and usually are used together in biomarker discovery studies.

In a biomarker discovery study, we usually have a control group and a disease group. In order to state that the two groups are different, we need a test to show that the difference between disease and control is statistically significant. The t-test is used widely to determine whether the two populations are statistically equal by calculating the p-value. In the metabolomics field, if a p-value is less than 0.05, it indicates that the two groups are statistically different. However, with a large coverage of metabolomics profiling, the false positive (e.g., two groups are not statistically different but with a small p-value) issue will be more severe. In this case, a false discovery rate (FDR) adjusted p-value is applied to reduce false positive errors. With the gauge of q-value, the chance of having false positive is reduced. Volcano plots are used often to visualize both the fold change and the p-value. In the volcano plot, -log (FDR-adjusted-p-value)

is plotted against log 2 (fold change), making a volcano-shaped scatter plot. Each point in the volcano plot represents a metabolite, and those whose fold changes and p-value pass the criterion, are referred as “significant metabolites”.

The limitation of uni-variate analyses is that simply counting the number of significant metabolites cannot tell us how different two groups are. Multi-variate analyses, which treat the data matrix as a whole, can show us the inter-group differences in a broader perspective. Principal Component Analysis (PCA) and Partial Least Squares-Discriminant Analysis (PLS-DA) are the most widely used multi-variate tools in metabolomics. PCA is an unbiased, high-confidence dimensionality reduction method. It finds the principal components (PC) by the linear combination of a set of variables and projects the high-dimensional data onto two or three axes. PCs are supposed to account for as much of the variability in the data as possible. When there is a statistically significant difference between two study groups, there should be a clear spatial separation between these two groups on a PCA plot. PLS-DA is a supervised method that considers the group assignment of the observations. Due to the group assignment information, PLS-DA provides a more focused view on the useful variations.

1.6 Overview of Thesis

My research started from developing and optimizing a chemical isotope labeling (DnsCl labeling) LC-MS platform for cellular extracts and the analysis of a serum metabolome for biomarker discovery.

Chapter 2 focuses on the development of a simple and rapid method for adherent cell harvesting and lysis. In this chapter, the efficiency of different cell harvesting and lysis methods

were evaluated by LC-UV. Based on the findings, physical scarping and frozen-thaw cycles were suggested to be used for cell harvesting and lysis in CIL LC-MS metabolomics.

In chapter 3, CIL LC-MS is used to profile the amine and phenol submetabolome of serum samples and find a potential biomarker for hyper IgE syndrome.

In Chapter 4, CIL LC-MS is used to study metabolomics changes in diabetes patients and how medical treatments, such as metformin, can change the metabolome.

Chapter 2

Development of a Simple and Efficient Method of Harvesting and Lysing Adherent Mammalian Cells for Chemical Isotope Labeling LC-MS-Based Cellular Metabolomics

2.1 Introduction

Cellular metabolomics uses analytical techniques to detect and quantify a whole set of metabolites or the metabolome in cells. Even with the analysis of a subset of the metabolome using techniques currently available, cellular metabolomics has become an important tool in biological research[29,30]. Compared to analyzing the metabolomes of biofluids such as urine and blood, cellular metabolomics requires additional sample handling steps, i.e., cell harvest and lysis. In order to profile the metabolome properly, a robust and reproducible method for cell harvest and lysis is required. Several studies have shown that improper handling of the cell harvest and lysis process could alter the metabolite concentrations artificially due to sample loss as well as residual metabolic reactions. For instance, it has been shown that the use of trypsinization for cell harvest could cause substantial metabolite leakage[30]. Other studies have illustrated significant differences in metabolomic results produced from different harvest and lysis methods[29].

Because current analytical techniques can cover only a fraction of the entire metabolome and metabolite detectability varies from one technique to another, the extent of any effects on the metabolome data caused by cell harvest and lysis methods is technique dependent. The reported studies are based mainly on the use of NMR, GC-MS, and LC-MS methods for metabolome analysis. Among these methods, LC-MS provides higher sensitivity for metabolite detection. However, conventional LC-MS techniques, even with the use of multiple methods (e.g., various combinations of reversed phase LC separation, hydrophilic interaction LC separation, positive ion detection, and negative ion detection), do not offer high-coverage metabolome analysis with high quantification accuracy. One alternative approach of metabolome analysis is to use chemical isotope labeling (CIL) to alter the chemical and physical properties of the metabolites for improving detection sensitivity and quantification accuracy. There are a number of labeling reagents that have been reported for targeted and untargeted metabolite analysis with varying degrees of success.

Our laboratory has been involved in developing a "divide-and-conquer" approach based on CIL LC-MS for comprehensive and quantitative metabolomics. We have reported four rationally designed isotope labeling reagents for analyzing the amine/phenol[31], carboxyl[32], hydroxy[33], and carbonyl[34] submetabolomes separately. The combined results of the four submetabolomes offer a high-coverage analysis of the whole metabolome. In addition, the labeled metabolites can be separated efficiently using reversed phase (RP) LC and ionized effectively as mainly protonated ions, rendering the possibility of using a single setup, RPLC-MS with positive ion mode detection for metabolite analysis. We have developed CIL LC-MS workflows and demonstrated their applications for metabolomic profiling of various types of biological samples.

In this study, we report a simple and efficient method of harvesting and lysing adherent mammalian cells tailored to CIL LC-MS-based cellular metabolomics. Being able to detect thousands of cellular metabolites with high quantification accuracy, 12C-/13C-dansylation LC-MS was employed to examine the amine/phenol submetabolomes of cell extracts prepared using different cell harvest and lysis methods. Using MCF-7 cells and HeLa cells as representatives of cultured adherent cells widely used in biological studies, we examined and compared the performance of the trypsinization method vs. the physical scrapping method for cell harvest, and the glass-bead-assisted method vs. the freeze-thaw-cycle method for cell lysis.

2.2 Experimental

2.2.1. Overall Workflow

Figure 2.1 shows the overall workflow of this study. MCF-7 cells and HeLa cells were cultured in 6-well plates in replicates with the same cell number. Cells were harvested by two different methods: trypsinization or physical scraping. The cell pellets were then treated by two lysis methods: freeze-thaw-cycle lysis or glass-bead-assisted lysis. The cell lysates were extracted and subjected to chemical labeling using 12C-dansyl chloride (DnsCl). A pooled sample from aliquots of individual samples was prepared and labeled by 13C-dansyl chloride. The total concentration of the labeled metabolites in each sample was measured by LC-UV. The 12C-labeled sample and the 13C-labeled pool were mixed by equal mole amounts. The mixture was injected into LC-MS for analysis. The peak pairs detected in MS were extracted by IsoMS, and individual peak-pairs from different LC-MS runs were aligned together, based on accurate mass and retention time, to produce a metabolite peak ratio table. Multivariate data analysis was

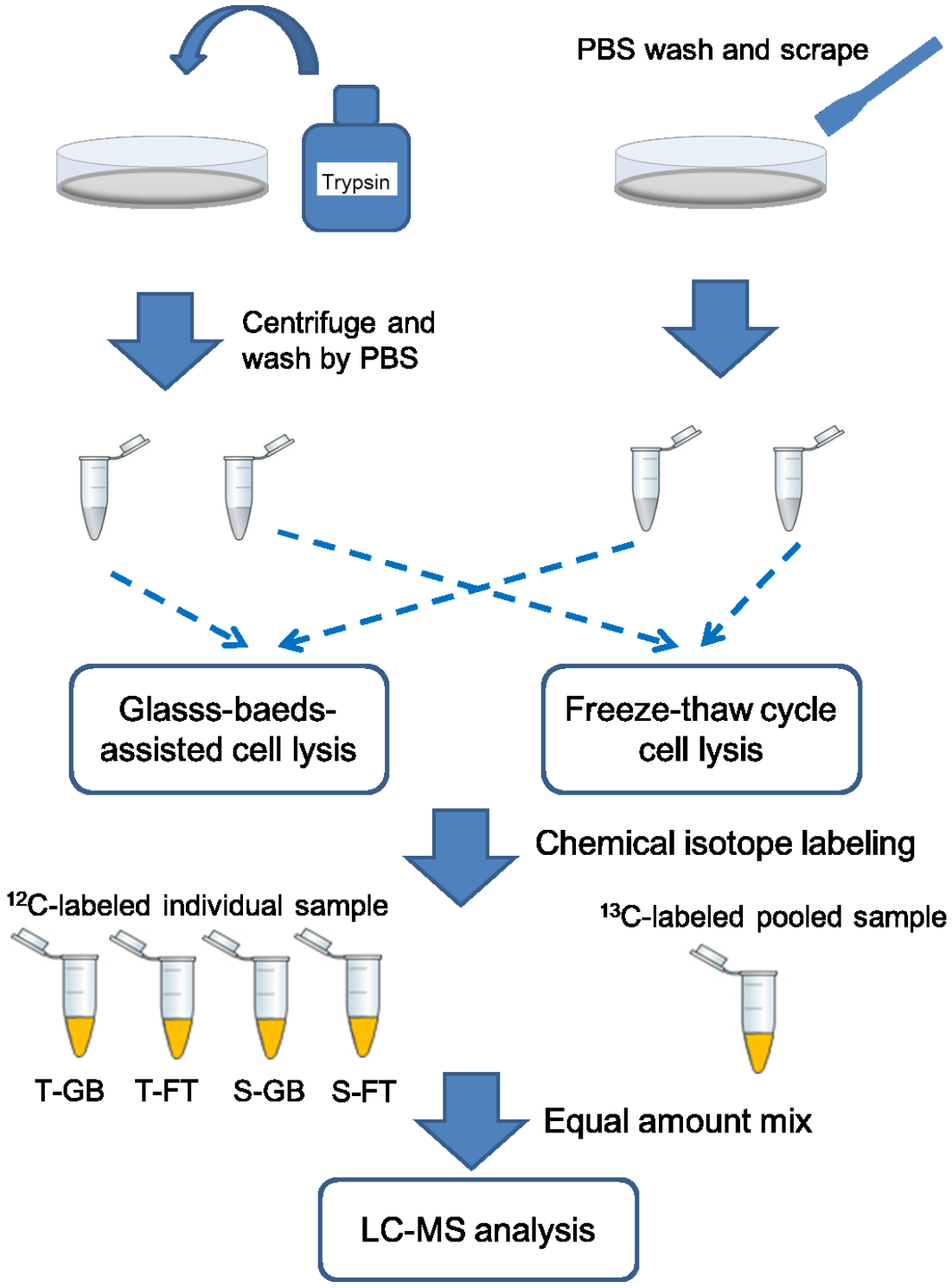


Figure 2.1. Workflow for comparing different methods to develop a simple and efficient cell harvest and lysis method for CIL LC-MS metabolomics of adherent mammalian cells.

performed by MetaboAnalyst (http://www.metaboanalyst.ca). The metabolites were identified by searching against MyCompoundID library (http://www.mycompoundid.org).

2.2.2. Chemicals and Reagents

The LC-MS grade reagent, including water, acetonitrile, methanol, and formic acid, were purchased from Fisher Scientific (Ottawa, ON), and 0.5-mm-diameter glass beads were purchased from Biospec Products (Bartlesville, OK). 13C-dansyl chloride was available from the University of Alberta (http://mcid.chem.ualberta.ca).

2.2.3. Cell Culture

Two types of cell lines, HeLa (ATCC CCL-2) and MCF-7 (ATCC HTB-22), were selected in this study. The growth medium for HeLa cell was Hyclone Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum (FBS). For the MCF-7 cell culture, additional 0.01 mg/mL human recombinant insulin was supplemented, as suggested by the American Type Culture Collection (ATCC). The same number of cells was cultured in Falcon 6-well plates. The cultures were incubated at 37 °C in a humidified atmosphere with 5% CO_2. The growth medium was renewed every two days.

2.2.4. Cell Harvest

Cells were harvested by either trypsinization or physical scraping. For trypsinization, the cells were washed with cold phosphate buffer saline (PBS), and 0.5 mL of 0.25% trypsin/EDTA (Hyclone, Logan, Utah) was added and incubated with the cell cultures at 37 °C. The trypsinization process was monitored under an inverted microscope and quenched by growth

medium when cells appeared rounded. The cultures were then transferred into 15-mL centrifuge tubes. The trypsin and growth medium were removed by 7-min 125-g centrifugation at 4 °C. The cell pellets were suspended in 5 mL of PBS and centrifuged at 125 g for 7 min at 4 °C. The cells were washed three times and were then snap-frozen in liquid nitrogen and stored in a -80 °C freezer. For physical scraping, the growth medium was removed, and the cell cultures were washed with cold PBS three times, then 1 mL of cold methanol was added for metabolism quenching. The cells were detached by scraping and transferred into 1.5-mL vials. After methanol was removed using Savant SC110A Speed Vac, the sample vials were stored in a -80 °C freezer for further use.

2.2.5. Cell Lysis

Cell lysis by using the glass-bead-assisted lysis method followed a previously published protocol[35]. In brief, the cell pellets were suspended in 100 µL of 50% MeOH and 50% water, and 0.5 mL of glass beads were added. Cells were lyzed via five 1-min periods of bead-beating at 3200 rpm alternated with five 1-min incubations in an ice-water bath. After cell lysis, 800 µL of 50% MeOH and 50% water were added for metabolite extraction. Glass beads, cell debris, and unbroken cells were removed by centrifugation at 16000 g for 10 min at 4 °C. Then, the supernatant was transferred into another vial and dried down in Speed Vac.

For freeze-thaw-cycle lysis, 300 µL of 50% MeOH and 50% water were added into the cell pellets. The vial was placed in liquid nitrogen for 2 min and thawed in water for 2 min with vortex. The freeze-thaw cycle was repeated for four more times. Then, the vial was centrifuged at 16000 g for 10 min, and the supernatant was transferred to another vial and dried down. The dried metabolites were re-dissolved in water and stored in a -80 °C freezer.

2.2.6. Dansylation Labeling

The labeling protocol was the same as that previously reported. In brief, 25 µL of cell extract were mixed with 12.5 µL of sodium carbonate/sodium bicarbonate buffer and 12.5 µL of 12C-dansyl chloride (18 mg/mL in ACN) or 13C-dansyl chloride (18 mg/mL in ACN). The reaction vial was incubated at 40 °C for one hour, then 5 µL of 250 mM NaOH were added and incubated for another 10 min to quench the excess DnsCl. Finally, 25 µL of 425 mM formic acid in 1:1 ACN/H_2O were added to the reaction mixture to acidify the solution.[31]

2.2.7. LC-UV

The total concentration of dansyl labeled metabolites was measured by a step-gradient LC-UV method.[24] A 5-µL labeled sample was injected into a Phenomenex Kinetex C18 column (2.1 mm × 5 cm, 1.7 µm particle size, 100 Å pore size) connected to a Waters ACQUITY UPLC system (Waters, Milford, MA). Mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The LC gradient was as follows: t = 0, 0% B; t =1 min, 0% B%; t = 1.1 min, 95% B; t = 2.6 min, 95% B; t = 3.1 min 0% B; t = 6.5 min, 0% B. The flow rate was 0.45 mL/min. The PDA detector was operated at 338 nm.

2.2.8. LC-MS

Each 12C-labeled individual sample was mixed with a 13C-labeled pool sample by equal mole amounts. LC-MS was done using a Thermo Scientific Dionex Ultimate 3000 UHPLC System (Sunnyvale, CA) linked to a Bruker Maxis II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA). The LC column was an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 µm particle size, 95 Å pore size). The mobile phases

were the same as those used for LC-UV. The LC gradient was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 99% B; t = 34 min, 99% B. The flow rate was 0.18 mL/min. The MS conditions were as follows: polarity, positive; dry temperature, 230 °C; dry gas, 8 L/min; capillary voltage, 4500V; nebulizer, 1.0 bar; end plate offset, 500V; spectra rate, 1.0 Hz.

2.2.9. Data Analysis

All the spectra were converted first to .csv files by Bruker Daltonics Data Analysis 4.3 software. The peak pairs were extracted from .csv files by IsoMS.[36] Data generated from the multiple runs were aligned together based on each peak's accurate mass and retention time. The missing values in the aligned file were filled by Zerofill software[37]. The principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed by MetaboAnalyst (www.metaboanalyst.ca).[38] The metabolites were identified positively by searching against DnsID Library, which contains retention time, MS, and MS/MS information of 275 unique amine/phenol-containing metabolite standards[39] (www.mycompoundid.org). Putative identification or match was performed by searching the accurate mass against the MyCompoundID library, which contains 8,021 known human metabolites and 375,809 predicted metabolites[40] (www.mycompoundid.org).

2.3. Results and Discussion

2.3.1. LC-UV Quantification for Cell Harvest and Lysis Efficiency Comparison

Cellular metabolomics involves the comparison of the metabolomes of different groups of cells. To compare the concentration differences of individual metabolites in different cell samples, it is critical to normalize the sample amounts before performing LC-MS analysis of samples. We have reported a step-gradient LC-UV method to measure the total concentration of dansyl labeled metabolites in a sample and use the total concentration for sample amount normalization.[24] In this work, we applied this approach to gauge the relative performance of the cell harvest and lysis methods. Briefly, we started with the seeding of the same number of cells in individual wells of a 6-well plate for replicate culturing, which ensured that the same number of cultured cells was used as the starting material from each well for cell harvest, cell lysis, and cell-extract labeling. We performed the LC-UV analysis of labeled metabolites from the processed samples and then compared their LC-UV quantification results, which should reflect the differences in efficiencies of cell harvest and cell lysis done by different methods.

In our study, cells were harvested by trypsinization (abbreviated as T) or physical scraping (abbreviated as S) and lysed by glass-bead-assisted lysis (GB) or freeze-thaw cycle lysis (FT). In total, there were four combinations for comparison: T-GB, T-FT, S-GB, and S-FT (see Figure 1). Two commonly used cell lines in biological studies, HeLa cells and MCF-7 cells, were selected for our study to represent adherent mammalian cells. Figures 2.2A and 2.2B show plots of the average concentration of labeled metabolites determined in each of the four combination methods for HeLa and MCF-7 cells, respectively. For the HeLa cells, the total concentration in T-GB, T-FT, S-GB and S-FT was found to be 0.52 ± 0.13, 0.52 ± 0.19, 0.90 ± 0.16 and 1.15 ±

0.23 mM, respectively. The standard deviation for each concentration measurement was the result of combined variations in biological triplicate and experimental duplicate (n=6). The total metabolite concentration in the physical scraping group is about 1.8-fold higher than that of the trypsinization group, with either GB lysis or FT lysis. Thus, for the HeLa cells, the scraping harvest method was more efficient than trypsinization. The same finding was obtained for the MCF-7 cells. In this case, the total metabolite concentration of T-GB, T-FT, S-GB, and S-FT was 0.71 ± 0.25, 1.1 ± 0.24, 1.30 ± 0.09 and 1.79 ± 0.20 mM, respectively. The concentration of the scraping group was also about 1.8-fold higher than that of the trypsinization group. The reduced concentration might be caused by metabolite loss during the trypsinization process through cell membrane damage and metabolites leakage.

The concentration plots shown in Figure 2.2 also can be used to gauge the differences in cell lysis efficiencies. For both HeLa and MCF-7 cells, the total metabolite concentrations of the FT groups were higher than those of the GB groups, except for the case of T-GB and T-FT groups of Hela cells where the total concentrations of the two groups had no significant difference. In the GB lysis method, to ensure that we could recover most of the metabolites, a relatively larger volume (800 μL) of extract solvent was used to rinse the beads, followed by drying. During the drying process, some relatively volatile metabolites might be lost, while other metabolites might adsorb onto the container walls and could not be re-dissolved, resulting in sample loss.

The above results obtained from LC-UV measurements of labeled metabolites indicate that the combination of physical scraping for the cell harvest and freeze-thaw-cycle for cell lysis gave the highest efficiencies. However, the LC-UV data only gauges the total metabolite amount difference, not the metabolite composition difference. Moreover, the experimental conditions

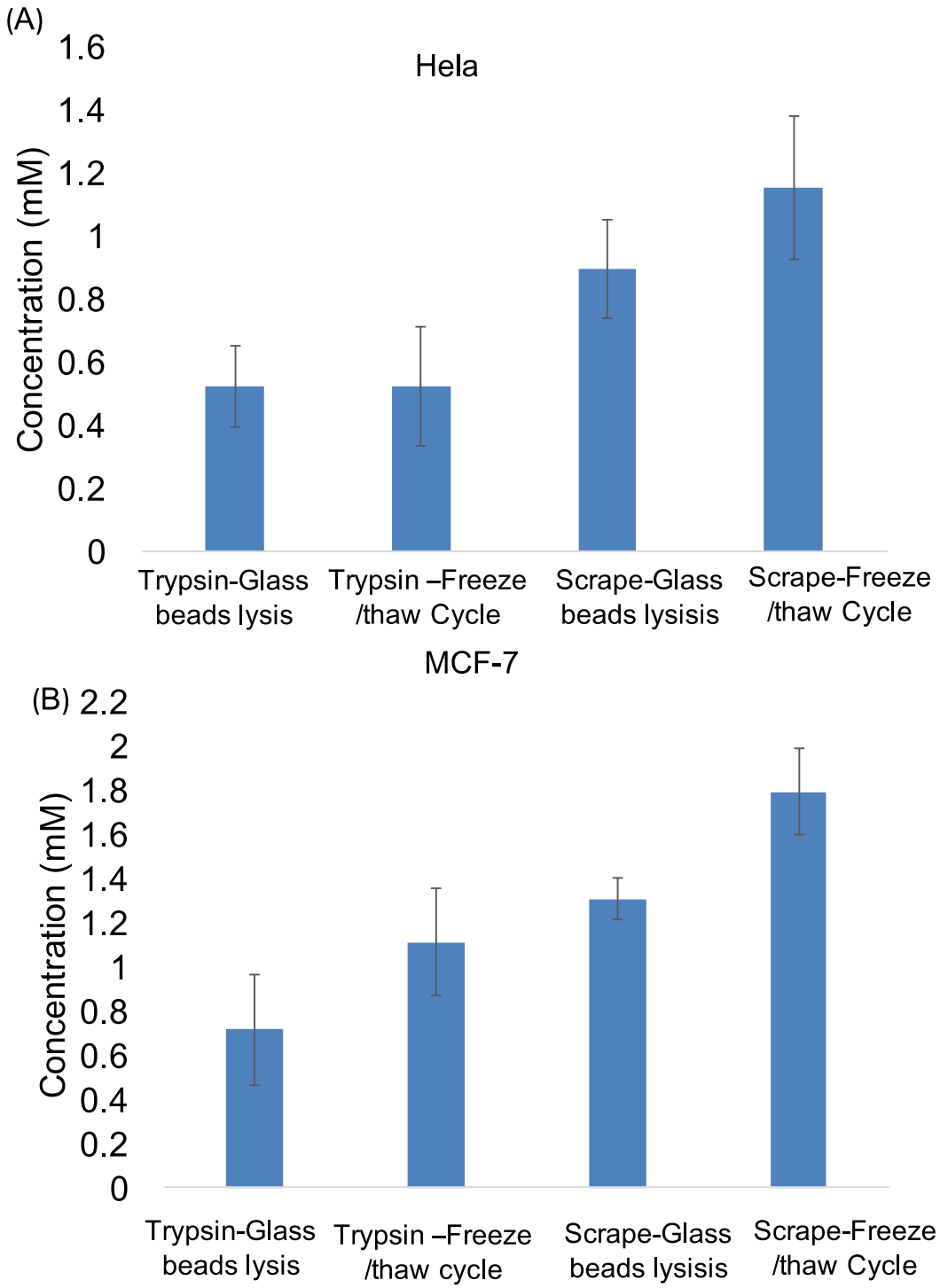


Figure 2.2. Average concentrations of dansyl labeled metabolites in cell extracts (n=6) prepared using different combinations of harvest and lysis methods from (A) HeLa and (B) MCF-7 cells. T-GB = trypsinization cell harvest followed by glass-bead lysis; T-FT = trypsinization cell harvest followed by freeze-thaw-cycle lysis; S-GB = scraping cell harvest followed by glass-bead lysis; S-FT = scraping cell harvest followed by freeze-thaw-cycle lysis.

used in harvest and cell lysis may affect the downstream process and analysis. Thus, from the metabolomic profiling point of view, we need to determine which combination method generates the optimal metabolomic result. We proceeded to use LC-MS and statistical analysis to examine the differences of the metabolome profiles generated from different combination methods.

2.3.2. LC-MS Results

In our LC-MS analysis, the 13C-labeled pool served as a global internal standard and was mixed with the 12C-dansyl labeled individual sample by equal mole amounts. The same amount of mixtures prepared from all individual samples was injected into a LC-MS. On average, for the HeLa cells, 3079 ± 50, 3033 ± 71, 3045 ± 68 and 3016 ± 73 peak pairs were detected from T-GB, T-FT, S-GB, and S-FT, respectively (Figure 2.3A). There was no significant difference among the four groups prepared by different harvest and lysis combinations. These results show that, if the same injection amount was used in LC-MS, the harvest and lysis methods would not affect the number of peak pairs detected. These results were confirmed in the analysis of MCF-7 cells. An average of 2768 ± 127, 2773 ± 49, 2641 ± 16, and 2604 ± 72 peak pairs was detected from T-GB, T-FT, S-GB, and S-FT, respectively (Figure 2.3B). It is interesting to note that we detected about 400 peak pairs less from the MCF-7 cell lysates, compared to the HeLa cells. Judging from the number of peak pairs or metabolites detected alone, these results indicate that the amine/phenol submetabolome profiles of MCF-7 and HeLa cells are different. Nevertheless, in both cases, thousands of peak pairs or metabolites were detected, illustrating the high metabolomic coverage achievable by the dansylation LC-MS method.

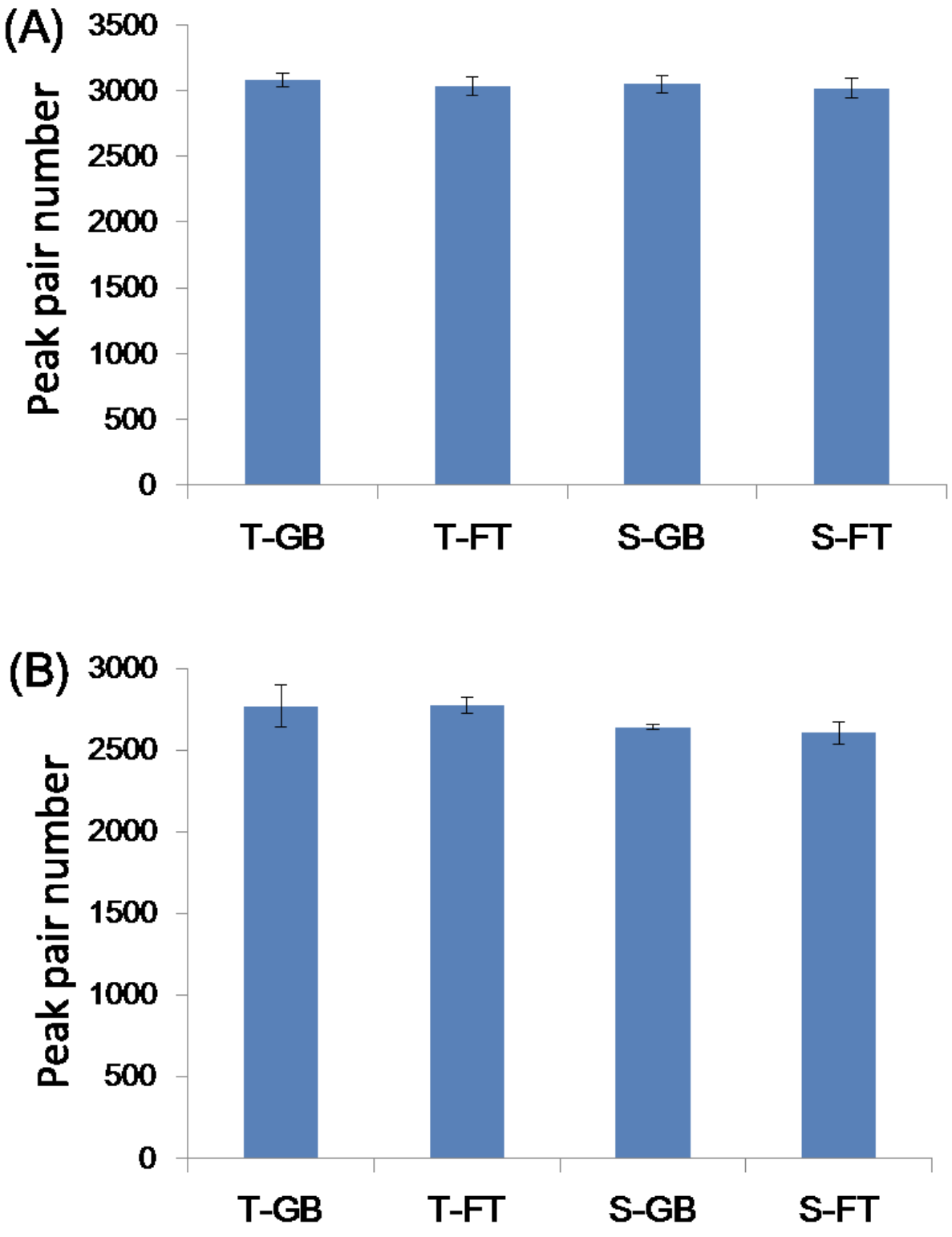


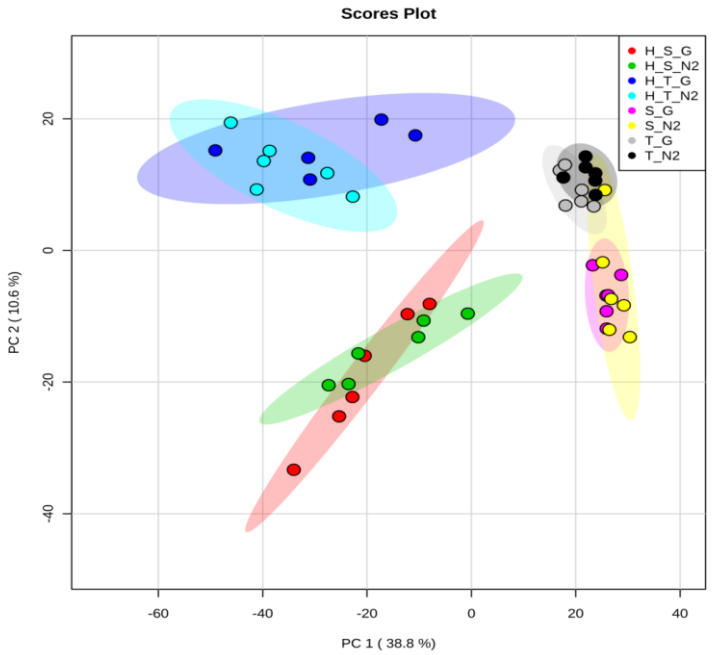
Figure 2.3. Average number of peak pairs (n=6) detected from (A) HeLa cell extracts and (B) MCF-7 cell extracts prepared using different combinations of harvest and lysis methods.

2.3.3. Multivariate Statistical Analysis

We applied multivariate statistical analysis to the metabolome data set obtained from the samples prepared using different harvest and lysis methods in order to examine the overall metabolome profile differences and similarities (i.e., the number and type of metabolites detected as well as their relative concentration differences in different samples). The score plot from the unsupervised PCA analysis is shown in Figure 2.4A. In this plot, 38.8% of the data were captured by the first principal component (PC), and 10.6% of the data were captured by the second PC. Overall, ~50% of the data could be captured by the 1st and 2nd PCs, indicating an excellent model. As Figure 2.4A shows, in both Hela and MCF-7 cells, the samples of the trypsinization group (T) were separated from those of the scraping group (S), whiles the glass-bead-assisted lysis group (GB) and the freeze-thaw-cycle lysis group (FT) overlapped. These results suggest that the trypsinization process might cause not only metabolite leakage or sample loss but also cause concentration changes for some of the detectable metabolites. The lysis method (GB or FT) had minor impact on the cellular metabolome, although the GB group gave a lower total metabolite concentration, as was discussed earlier. In the PCA plot, no matter which harvest and lysis methods were used, HeLa cells (H) and MCF-7 cells (M) are separated clearly. Thus, the metabolomes of the two different cell lines are significantly different.

Supervised PLS-DA analysis was applied also to the metabolome data set, and the score plot is shown in Figure 2.4B. The two cell lines are separated on component 1, and the two cell harvest groups are separated on component 2, with R2 = 0.968 and Q2 = 0.932 from cross-validation test. The high scores of R2 and Q2 confirm the robustness of the model. These PLS-DA analysis results confirmed the findings of the PCA analysis.

(A)



(B)

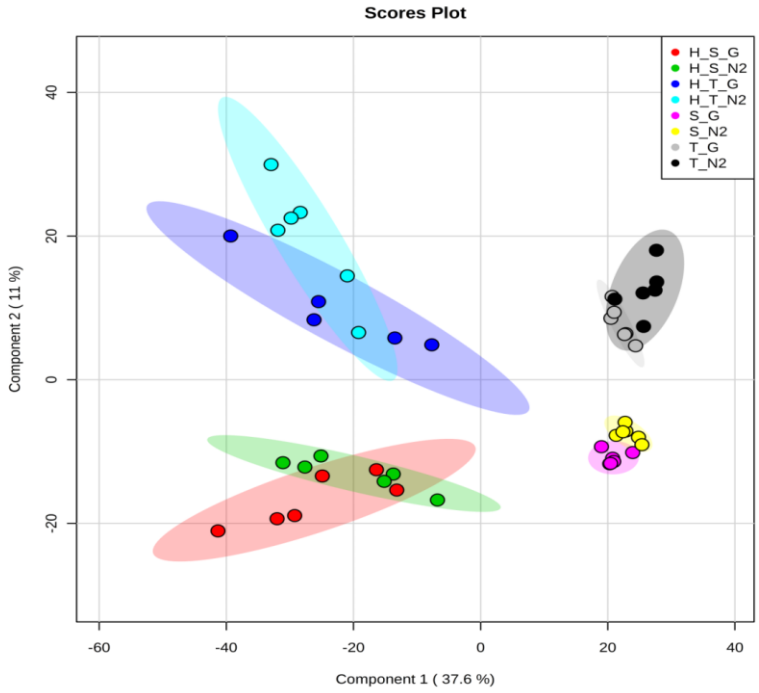


Figure 2.4. (A) PCA and (B) PLS-DA plots of the amine/phenol submetabolomes of HeLa and MCF-7 cells from cell extracts prepared using different combinations of harvest and lysis methods. H=HeLa cells. M=MCF-7 cells, with other abbreviations shown in the Figure 2.4 caption.

2.3.4. Impact of Different Harvest Methods on Cellular Metabolome

To analyze the impact of different harvest methods on cellular metabolome, univariate analysis using volcano plots was performed on the metabolome data set (Figure 2.5). In the volcano plot, the x-axis is the fold change (FC) of trypsinization/scraping groups, and the y-axis is the p-value from the t-test of the two groups. For HeLa cells (Figure 2.5A), there were 429 metabolites with significantly higher fold changes and 305 metabolites with significantly lower fold changes using the criteria of p-value < 0.01 and FC > 1.5. For MCF-7 cells (Figure 2.5B), there were 131 metabolites with higher fold changes and 88 metabolites with lower fold changes. These results show that there was a large number of metabolites having significantly different concentrations in samples prepared using the two different harvest methods. The metabolites with significant changes are listed in Tables 2.1 and 2.2, along with the metabolite identification results shown in Tables 2.3 and 2.4.

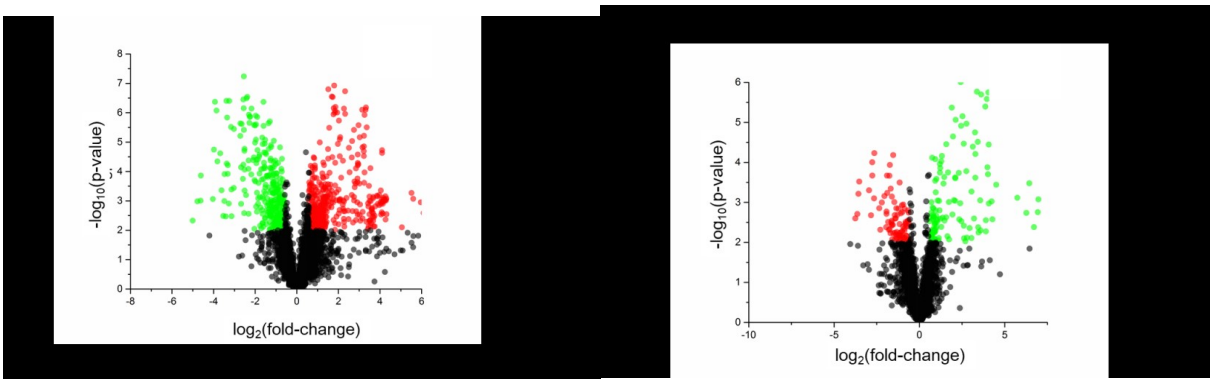


Figure 2.5. Volcano plots of the amine/phenol submetabolomes of (A) HeLa and (B) MCF-7 cells harvested by different methods. The p-value of each metabolite was calculated from the t-test, and the fold change (FC) was calculated from the peak ratios of the trypsinization group divided by the peak ratios of the scraping group (i.e., trypsinization/scraping). Using a cut-off value of $p < 0.01$ and $FC > 1.5$ or < 0.67, the red points represent the metabolites with higher concentrations in the trypsinization group, and the green points represent the metabolites with lower concentrations in the trypsinization group. The black points represent the metabolites with no significant differences.

Table 2.1. Metabolites with Significantly Different Concentrations in HeLa Cell Extracts Prepared Using Different Harvest Methods (trypsinization/physical-scraping).

No.	HMDB	Compound Name	fold change	p-value	Level
59	HMDB01397	Guanosine monophosphate	2.93	1.47E-03	1
68	HMDB00001	1-Methylhistidine	0.46	2.67E-04	1
101	HMDB00045	Adenosine monophosphate	9.56	1.18E-06	1
105	HMDB00133	Guanosine	4.96	2.25E-03	1
107	HMDB12114	(3S)-3,6-Diaminohexanoate	1.92	1.69E-03	2
144	HMDB00517	L-Arginine	1.77	1.27E-03	1
146	HMDB00052	Argininosuccinic acid	0.55	7.64E-03	2
196	HMDB01410	2-Amino-4-oxo-6-(1',2'-dioxoprolyl)-7,8-dihydroxypteridin	2.17	5.16E-03	2
202	HMDB01325	N6,N6,N6-Trimethyl-L-lysine	1.69	7.14E-05	2
206	HMDB00045	Adenosine monophosphate	9.90	7.91E-07	1
215	HMDB03331	1-Methyladenosine	2.34	1.15E-04	2
247	HMDB03276	Hydrogen sulfide	2.83	1.72E-05	2
255	HMDB00299	Xanthosine	2.55	1.25E-03	2
256	HMDB00095_2	Cytidine monophosphate - Isomer	4.08	1.85E-06	1
264	HMDB00195	Inosine	62.38	1.12E-03	2
275	HMDB03334	Symmetric dimethylarginine	2.64	3.74E-04	1
364	HMDB00641	L-Glutamine	0.25	1.04E-05	1
364	HMDB03423	D-Glutamine	0.25	1.04E-05	1
371	HMDB00904	Citrulline	0.56	6.85E-03	1
380	HMDB00856	N-a-Acetylcitrulline	1.57	7.65E-03	2
381	HMDB11737	Gamma Glutamylglutamic acid	0.60	2.57E-03	1
429	HMDB00187	L-Serine	0.66	7.30E-04	1
503	HMDB00125	Glutathione	1.83	4.02E-03	2
515	HMDB02005	Methionine Sulfoxide	0.59	8.66E-03	1
529	HMDB00187	L-Serine	0.51	1.14E-04	1
542	HMDB00191	L-Aspartic Acid	2.95	3.24E-06	1
560	HMDB12326	L-Gulose	6.74	9.64E-04	2
577	HMDB00288	Uridine 5'-monophosphate	4.81	7.24E-07	2
581	HMDB00191	L-Aspartic Acid	2.79	6.65E-04	1
622	HMDB04437	Diethanolamine	0.55	6.81E-03	1

707	HMDB02335	Aspartyl-L-proline	0.43	1.47E-05	2
763	HMDB06555	dIMP	1.51	2.35E-04	2
811	HMDB01263	Allysine	6.25	3.31E-05	2
850	HMDB00174	L-Fucose	0.59	2.94E-03	2
854	HMDB00721	Glycylproline	0.35	8.35E-06	1
907	HMDB00079	Dihydrothymin	0.29	3.06E-05	2
978	HMDB00296	Uridine	10.74	1.22E-03	1
984	HMDB00056	Beta-Alanine	0.64	8.47E-03	1
986	HMDB00585	Glucosylgalactosyl hydroxylysine	1.82	8.47E-05	2
1001	HMDB00721	Glycylproline	0.50	3.69E-04	1
1015	HMDB00323	3-Amino-2-piperidone	1.92	1.66E-04	2
1138	HMDB02284	N-Acetylcadaverine	1.87	3.17E-03	2
1155	HMDB00576	Monoethyl malonic acid	4.39	1.62E-04	2
1202	HMDB00296_2	Uridine - H2O	11.63	7.67E-05	1
1208	HMDB12136	1-Amino-propan-2-ol	3.18	4.40E-04	2
1251	HMDB11166	L-beta-aspartyl-L-leucine	2.29	3.11E-04	2
1268	HMDB00292	Xanthine	2.70	5.17E-03	1
1310	HMDB11170	L-gamma-glutamyl-L-isoleucine	1.93	9.44E-03	2
1354	HMDB03911	3-Aminoisobutanoic acid	0.44	7.35E-05	1
1407	HMDB01080	4-Aminobutyraldehyde	2.85	3.65E-05	2
1430	HMDB00600	Galactosylhydroxylysin	1.79	3.53E-05	2
1431	HMDB03609	2-Aminoacrylic acid	0.57	9.40E-04	2
1435	HMDB01257	Spermidine	2.09	2.43E-03	1
1492	HMDB00594	Glutamylphenylalanine	2.23	9.90E-03	2
1549	HMDB02201	N-Carboxyethyl-g-aminobutyric acid	0.15	2.30E-06	2
1575	HMDB28691	Alanyl-Leucine	4.19	3.02E-05	1
1661	HMDB00883	L-Valine	0.66	1.94E-03	1
1677	HMDB00759	Glycyl-L-Leucine	1.99	2.14E-03	1
1677	HMDB28844	Glycyl-Isoleucine	1.99	2.14E-03	1
1695	HMDB00300	Uracil	48.47	8.47E-04	1
1736	HMDB28691	Alanyl-Leucine	7.62	2.64E-05	1
1799	HMDB28848	Glycyl-Phenylalanine	2.02	2.96E-03	1
1846	HMDB01545	Pyridoxal	0.49	4.22E-05	1
1893	HMDB03581	Dethiobiotin	0.53	3.78E-04	2

1929	HMDB00159	L-Phenylalanine	0.59	1.40E-03	1
1959	HMDB00243	Pyruvic acid	2.20	1.45E-03	2
2034	HMDB02248	Gamma glutamyl ornithine	1.52	4.74E-03	2
2080	HMDB28937	Leucyl-Proline	0.17	1.97E-05	1
2084	HMDB00687	L-leucine	0.46	5.70E-04	1
2091	HMDB00159	L-Phenylalanine	0.57	2.52E-04	1
2133	HMDB03869	Epsilon-(gamma-Glutamyl)-lysine	0.47	2.98E-04	2
2151	HMDB01491	Pyridoxal 5'-phosphate	1.93	2.27E-03	2
2216	HMDB00450	5-Hydroxylysine	2.43	1.68E-04	1
2451	HMDB11162	L-beta-aspartyl-L-alanine	0.58	6.72E-04	2
2463	HMDB00339	2-Methylbutyrylglycine	0.26	8.22E-06	2
2523	HMDB01889	Theophylline	0.24	3.29E-05	1
2569	HMDB12230	Gamma-glutamyl-L-putrescine	12.15	1.90E-03	2
2580	HMDB06045	Dityrosine	0.33	5.02E-05	2
2609	HMDB00656	Cysteineglutathione disulfide	14.81	6.97E-04	2
2647	HMDB01256	Spermine	2.15	1.02E-05	2
2667	HMDB00214	Ornithine	3.42	6.98E-07	1
2733	HMDB03454	4-Pyridoxolactone	0.39	2.28E-06	2
2760	HMDB00955	Isoferulic acid	0.52	3.46E-03	1
2819	HMDB02135	S-(3-oxo-3-carboxy-n-propyl)cysteine	1.51	6.19E-03	2
2827	HMDB02107	Phthalic acid	0.53	1.26E-03	2
2888	HMDB12134	1,2-Dihydroxy-3-keto-5-methylthiopentene	1.83	6.19E-05	2
2956	HMDB00500	4-Hydroxybenzoic acid	0.32	7.04E-06	1
2960	HMDB00701	Hexanoylglycine	0.19	2.87E-07	2
2993	HMDB00512	N-Acetyl-L-phenylalanine	0.21	2.22E-06	2
3038	HMDB03227	Methanethiol	1.72	6.52E-03	2
3041	HMDB01276	N1-Acetylspermidine	1.66	2.58E-03	2
3097	HMDB11686	p-Cresol glucuronide	4.79	4.07E-03	2
3106	HMDB00132	Guanine	4.04	2.70E-03	2
3143	HMDB00177	L-Histidine	0.64	1.28E-03	1
3266	HMDB06524	3-Indoleacetonitrile	0.55	7.32E-03	2
3282	HMDB01526	S-Acetyldihydrolipoamide	7.81	5.62E-05	2
3316	HMDB03320	Indole-3-carboxylic acid	0.50	8.81E-04	1

3362	HMDB00209	Phenylacetic acid	0.62	3.76E-03	2
3528	HMDB29105	Tyrosyl-Glycine	0.50	4.53E-03	1
3622	HMDB02044	8-Hydroxyguanosine	1.65	1.80E-03	2
3643	HMDB12286	S-Prenyl-L-cysteine	0.45	6.02E-03	2
3751	HMDB29098	Tyrosyl-Alanine	2.01	8.69E-03	1
3793	HMDB01414	1,4-diaminobutane	1.80	1.80E-03	1
3953	HMDB02043	5-Phenylvaleric acid	2.13	2.04E-03	2
4027	HMDB02322	Cadaverine	1.91	4.46E-03	1
4095	HMDB11176	L-phenylalanyl-L-hydroxyproline	0.51	4.39E-03	2
4097	HMDB00107	Galactitol	0.40	1.91E-03	2
4101	HMDB00158	L-Tyrosine	0.37	8.20E-04	1
4127	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	0.06	8.65E-04	2
4299	HMDB05809	Eugenol	2.00	6.63E-03	2
4342	HMDB04586	Perillic acid	1.84	1.94E-03	2
4558	HMDB04072	4-Hydroxystyrene	0.62	2.09E-03	2
4612	HMDB00866	N-Acetyl-L-tyrosine	0.10	1.46E-06	2
4884	HMDB04058	5,6-Dihydroxyindole	0.59	9.77E-04	2
4926	HMDB01387	N-Methylphenylethanolamine	0.28	2.44E-04	2
5010	HMDB03905	Imidazole-4-acetaldehyd	0.61	1.00E-04	2
5024	HMDB11687	Phenylbutyrylglutamine	0.38	2.43E-03	2
5054	HMDB00226	Orotic acid	0.54	9.36E-03	2
5104	HMDB00030	Biotin	0.15	6.12E-06	2
5272	HMDB11718	4-Hydroxybenzaldehyde	0.46	3.83E-03	2
5341	HMDB12219	Dopamine quinone	0.23	4.07E-05	2
5503	HMDB06954	2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate	0.18	6.35E-04	2
5521	HMDB01430	L-Dopachrome	0.60	5.62E-03	2
5667	HMDB12182	8-Hydroxypurine	0.38	1.35E-04	2
5791	HMDB00472	5-Hydroxy-L-tryptophan	0.12	3.62E-04	2
6264	HMDB06779	Indole-5,6-quinone	0.48	5.05E-03	2
6916	HMDB00175	Inosinic acid	0.49	5.19E-03	2

Table 2.2 Metabolites with Significantly Different Concentrations in MCF-7 Cell Extracts Prepared Using Different Harvest Methods (trypsinization/physical-scraping).

No.	HMDB No.	Compound Name	Fold change	p-value	Level
59	HMDB01397	Guanosine monophosphate	2.75	2.32E-04	1
83	HMDB01410	2-Amino-4-oxo-6-(1',2'-dioxoprolyl)-7,8-dihydroxypteridine	1.70	6.02E-03	2
84	HMDB00229	Nicotinamide ribotide	2.00	2.59E-04	2
101	HMDB00045	Adenosine monophosphate	49.20	4.15E-07	1
105	HMDB00133	Guanosine	9.73	6.16E-05	1
114	HMDB01397	Guanosine monophosphate	5.35	9.85E-07	2
127	HMDB02022	Glycineamideribotide	1.81	2.26E-03	2
144	HMDB00517	L-Arginine	2.44	9.35E-05	1
146	HMDB00052	Argininosuccinic acid	0.42	6.48E-03	2
206	HMDB00045	Adenosine monophosphate	45.43	3.70E-07	1
264	HMDB00195	Inosine	104.89	4.15E-03	2
288	HMDB11168	L-beta-aspartyl-L-serine	1.68	7.63E-05	2
365	HMDB00912	Succinyladenosine	0.66	9.05E-03	2
380	HMDB00856	N-a-Acetylcitrulline	0.33	3.79E-03	2
383	HMDB02335	Aspartyl-L-proline	0.48	7.16E-03	2
392	HMDB00802	Pterin	0.48	7.82E-03	2
532	HMDB02278	2-(acetylamino)-1,5-anhydro-2-deoxy-3-O-b-D-galactopyranosyl-D-arabino-Hex-1-enitol	3.83	2.01E-03	2
563	HMDB00167	L-Threonine	0.59	3.93E-03	1
563	HMDB00719	L-Homoserine	0.59	3.93E-03	1
568	HMDB05765	Ophthalmic acid	0.57	3.55E-03	2
577	HMDB00288	Uridine 5'-monophosphate	5.45	1.22E-05	2
619	HMDB00149	Ethanolamine	10.37	1.70E-06	1
850	HMDB00174	L-Fucose	0.57	3.31E-03	2
892	HMDB00854	Formiminoglutamic acid	1.65	6.24E-03	2
970	HMDB01263	Allysine	3.73	4.27E-06	2
978	HMDB00296	Uridine	10.71	3.07E-05	1
983	HMDB00056	Beta-Alanine	0.64	2.79E-03	1
983	HMDB00161	L-Alanine	0.64	2.79E-03	1
1034	HMDB03338	Hydroxylamine	1.71	5.83E-03	2

1046	HMDB12201	Cis-zeatin-7-N-glucoside	1.78	4.91E-03	2
1202	HMDB00296_2	Uridine - H2O	14.54	4.01E-06	1
1407	HMDB01080	4-Aminobutyraldehyde	1.93	1.59E-03	2
1575	HMDB28691	Alanyl-Leucine	0.37	2.08E-03	1
1761	HMDB11105	5-Acetylamino-6-formylamino-3-methyluracil	0.50	1.74E-03	2
1846	HMDB01545	Pyridoxal	0.26	1.93E-03	1
2080	HMDB28937	Leucyl-Proline	0.22	1.00E-03	1
2100	HMDB01263	Allysine	0.28	3.10E-03	2
2216	HMDB00450	5-Hydroxylysine	2.77	2.85E-03	1
2314	HMDB00130	Homogentisic acid	0.41	7.87E-03	2
2956	HMDB00500	4-Hydroxybenzoic acid	0.39	6.28E-03	1
3159	HMDB00195	Inosine	125.60	8.41E-04	2
3165	HMDB01904	3-Nitrotyrosine	12.29	2.01E-06	2
3190	HMDB05199	(R)-Salsolinol	0.08	1.95E-03	2
3192	HMDB01488	Nicotinic acid	3.20	1.78E-04	2
3928	HMDB12176	5-Aminopentanamide	1.76	9.89E-03	2
4127	HMDB00375	3-(3-Hydroxyphenyl)propanoic acid	0.16	7.91E-04	2
4187	HMDB00656	Cysteineglutathione disulfide	1.64	9.15E-03	2
4303	HMDB01257	Spermidine	2.22	9.17E-04	2
4382	HMDB03747	Resveratrol	1.91	8.31E-05	2
4409	HMDB00206	N6-Acetyl-L-lysine	0.31	4.53E-04	2
4807	HMDB11150	Deoxyhypusine	1.75	4.60E-03	2
4898	HMDB01084	D-1-Piperideine-2-carboxylic acid	3.92	2.25E-05	2
4915	HMDB00555	3-Methyladipic acid	0.46	8.04E-03	2
4926	HMDB03633	N-Methyltyramine	0.57	3.07E-03	2
5225	HMDB02338	Biochanin A	0.16	5.86E-05	2
5243	HMDB00132	Guanine	6.47	9.37E-03	2
5320	HMDB00299	Xanthosine	0.20	4.82E-03	2
5374	HMDB05199	(R)-Salsolinol	5.24	1.05E-03	2
5503	HMDB06954	2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate	0.30	2.16E-04	2
5719	HMDB04089	Formylanthranilic acid	0.44	6.59E-03	2
5733	HMDB01314	Cyclic GMP	0.54	4.53E-03	2

5791	HMDB00472	5-Hydroxy-L-tryptophan	0.34	3.63E-03	2
5911	HMDB02393	N-Methyl-D-aspartic acid	0.33	7.17E-04	2
6442	HMDB00252	Sphingosine	0.35	5.95E-03	2
6571	HMDB00269	Sphinganine	0.30	1.16E-04	2

Table 2.3. List of Metabolites from Hela Cell Lysates Positively Identified by Searching Against the DnsCl Labeled Standard Library.

Peak Pair Information						**Identification Result**	
Peak Pair#	T_R(min)	Corrected T_R(min)	mz_light	mz_heavy	monoisotopic mass (Da)	HMDB.No.	Name
68	2.16	2.12	403.1438	405.1505	169.0854	HMDB00001	1-Methylhistidine
68	2.16	2.12	403.1438	405.1505	169.0854	HMDB00479	3-methyl-histidine
101	2.29	2.24	581.1216	583.1276	347.0632	HMDB00045	Adenosine monophosphate
105	2.30	2.26	517.1504	519.1570	283.0921	HMDB00133	Guanosine
130	2.40	2.35	375.0777	377.0843	141.0193	HMDB00224	O-Phosphoethanolamine
144	2.42	2.37	408.1702	410.1766	174.1118	HMDB00517	L-Arginine
153	2.47	2.42	510.1910	512.1973	276.1327	HMDB00279	Saccharopine
154	2.47	2.43	388.1077	390.1137	154.0494	HMDB00157	Hypoxanthine + H2O
213	2.75	2.68	422.1862	424.1925	188.1279	HMDB00670	Homo-L-arginine
256	2.97	2.89	557.1126	559.1211	323.0543	HMDB00095_ 2	Cytidine monophosphate - Isomer
275	3.09	3.01	436.2016	438.2082	202.1433	HMDB03334	Symmetric dimethylarginine
313	3.28	3.19	366.1118	368.1184	132.0535	HMDB00168	L-Asparagine
364	3.62	3.51	380.1276	382.1342	146.0693	HMDB00641	L-Glutamine
364	3.62	3.51	380.1276	382.1342	146.0693	HMDB03423	D-Glutamine

371	3.79	3.67	409.1545	411.1611	175.0961	HMDB00904	Citrulline
381	3.82	3.70	510.1555	512.1616	276.0971	HMDB11737	Gamma Glutamylglutamic
421	3.94	3.82	501.1551	503.1611	267.0968	HMDB00050	Adenosine
443	4.07	3.93	399.1050	401.1111	165.0466	HMDB02005	Methionine Sulfoxide
443	4.07	3.93	399.1050	401.1111	165.0466	HMDB02005_2	Methionine Sulfoxide - Isomer
494	4.25	4.11	353.1167	355.1234	119.0584	HMDB00719	L-Homoserine
499	4.29	4.15	339.1008	341.1078	105.0425	HMDB00187	L-Serine
574	4.92	5.06	381.1114	383.1184	147.0531	HMDB00148	L-Glutamic Acid
579	5.01	5.11	365.1160	367.1222	131.0577	HMDB00725	Trans-4-Hydroxyl-L-Proline
581	5.11	5.18	367.0959	369.1027	133.0376	HMDB00191	L-Aspartic Acid
591	5.32	5.51	422.1744	424.1816	188.1161	HMDB00206	N6-Acetyl-L-Lysine
611	5.40	5.64	492.1806	494.1870	258.1223	HMDB00279_2	Saccharopine - H2O
616	5.49	5.78	353.1167	355.1235	119.0584	HMDB00167	L-Threonine
618	5.52	5.82	395.1274	397.1339	161.0691	HMDB00510	Aminoadipic acid
619	5.63	5.94	295.1111	297.1177	61.0528	HMDB00149	Ethanolamine
622	5.65	5.97	339.1375	341.1439	105.0791	HMDB04437	Diethanolamine
740	6.19	6.58	309.0912	311.0976	75.0329	HMDB00123	Glycine
812	6.49	6.96	364.1693	366.1758	130.1109	HMDB02064	N-Acetylputrescine
854	6.60	7.09	406.1435	408.1501	172.0852	HMDB00721	Glycylproline
862	6.63	7.13	323.1060	325.1125	89.0476	HMDB00161	L-Alanine
978	6.99	7.57	478.1282	480.1348	244.0699	HMDB00296	Uridine

984	7.04	7.64	323.1067	325.1127	89.0484	HMDB00056	Beta-Alanine
987	7.08	7.67	337.1219	339.1285	103.0636	HMDB00112	Gamma-Aminobutyric acid
1143	7.52	8.16	453.1690	455.1754	219.1107	HMDB00210	Pantothenic acid
1178	7.64	8.29	492.1444	494.1498	258.0861	HMDB00884_2	Ribothymidine - Isomer
1202	7.73	8.38	460.1178	462.1243	226.0595	HMDB00296_2	Uridine - H2O
1218	7.79	8.45	370.0973	372.1040	136.0390	HMDB00157_2	Hypoxanthine - multi-tags
1268	7.99	8.67	386.0922	388.0989	152.0339	HMDB00292	Xanthine
1272	8.01	8.69	337.1220	339.1286	103.0637	HMDB00452	L-Alpha-aminobutyric acid
4101	21.48	22.64	324.5955	326.6023	181.0744	HMDB06050	o-Tyrosine
4112	21.65	22.81	374.1302	376.1368	280.1437	HMDB29118	Tyrosyl-Valine
4170	21.92	23.08	328.1011	330.1073	94.0428	HMDB00228	Phenol
4221	22.17	23.33	373.0859	375.0925	139.0276	HMDB01232	4-Nitrophenol
4331	22.64	23.80	381.1380	383.1440	294.1593	HMDB29109	Tyrosyl-Leucine
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB01858	p-Cresol
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB02048	m-Cresol
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB02055	o-Cresol
4601	23.56	24.72	322.1044	324.1115	176.0922	HMDB00259	Serotonin
4846	24.27	25.43	356.1316	358.1383	122.0733	HMDB29306	4-Ethylphenol
5026	24.61	25.77	302.6005	304.6074	137.0844	HMDB00306	Tyramine
5287	25.55	26.71	289.0790	291.0837	110.0414	HMDB00957	pyrocatechol

Table 2.4. List of Metabolites from MCF-7 Cell Lysates Positively Identified by Searching Against the DnsCl Labeled Standard Library

Peak Pair Information						Identification Result	
Peak Pair #	T_R (min)	Corrected T_R (min)	mz_light	mz_heavy	monoisotopic mass (Da)	HMDB.No.	Name
68	2.16	2.12	403.1438	405.1505	169.0854	HMDB00001	1-Methylhistidine
						HMDB00479	3-methyl-histidine
101	2.29	2.24	581.1216	583.1276	347.0632	HMDB00045	Adenosine monophosphate
105	2.30	2.26	517.150	519.1570	283.0921	HMDB00133	Guanosine
130	2.40	2.35	375.077	377.0843	141.0193	HMDB00224	O-Phosphoethanolamine
144	2.42	2.37	408.1702	410.1766	174.1118	HMDB00517	L-Arginine
153	2.47	2.42	510.1910	512.1973	276.1327	HMDB00279	Saccharopine
154	2.47	2.43	388.1077	390.1137	154.0494	HMDB00157	Hypoxanthine + H2O
213	2.75	2.68	422.186	424.1925	188.1279	HMDB00670	Homo-L-arginine
256	2.97	2.89	557.1126	559.1211	323.0543	HMDB00095_ 2	Cytidine monophosphate - Isomer
275	3.09	3.01	436.2016	438.2082	202.1433	HMDB03334	Symmetric dimethylarginine
313	3.28	3.19	366.1118	368.1184	132.0535	HMDB00168	L-Asparagine
364	3.62	3.51	380.1276	382.1342	146.0693	HMDB00641	L-Glutamine
						HMDB03423	D-Glutamine

371	3.79	3.67	409.1545	411.1611	175.0961	HMDB00904	Citrulline
381	3.82	3.70	510.1555	512.1616	276.0971	HMDB11737	Gamma Glutamylglutamic acid
4020	20.97	22.14	324.5954	326.6024	181.0742	HMDB06050	o-Tyrosine
4027	21.00	22.17	285.1164	287.1231	102.1163	HMDB02322	Cadaverine
4101	21.48	22.64	324.5955	326.6023	181.0744	HMDB00158	L-Tyrosine
4112	21.65	22.81	374.1302	376.1368	280.1437	HMDB29118	Tyrosyl-Valine
4170	21.92	23.08	328.1011	330.1073	94.0428	HMDB00228	Phenol
4221	22.17	23.33	373.0859	375.0925	139.0276	HMDB01232	4-Nitrophenol
4331	22.64	23.80	381.1380	383.1440	294.1593	HMDB29109	Tyrosyl-Leucine
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB01858	p-Cresol
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB02048	m-Cresol
4525	23.30	24.46	342.1164	344.1230	108.0581	HMDB02055	o-Cresol
5026	24.61	25.77	302.600	304.6074	137.0844	HMDB00306	Tyramine
5287	25.55	26.71	289.0790	291.0837	110.0414	HMDB00957	pyrocatechol

To illustrate the concentration differences of individual metabolites, we selected some of the metabolites commonly detected in both HeLa and MCF-7 cells that also were identified positively with very large concentration differences between the trypsinization group and the scraping group to produce heap maps (Figure 2.6).

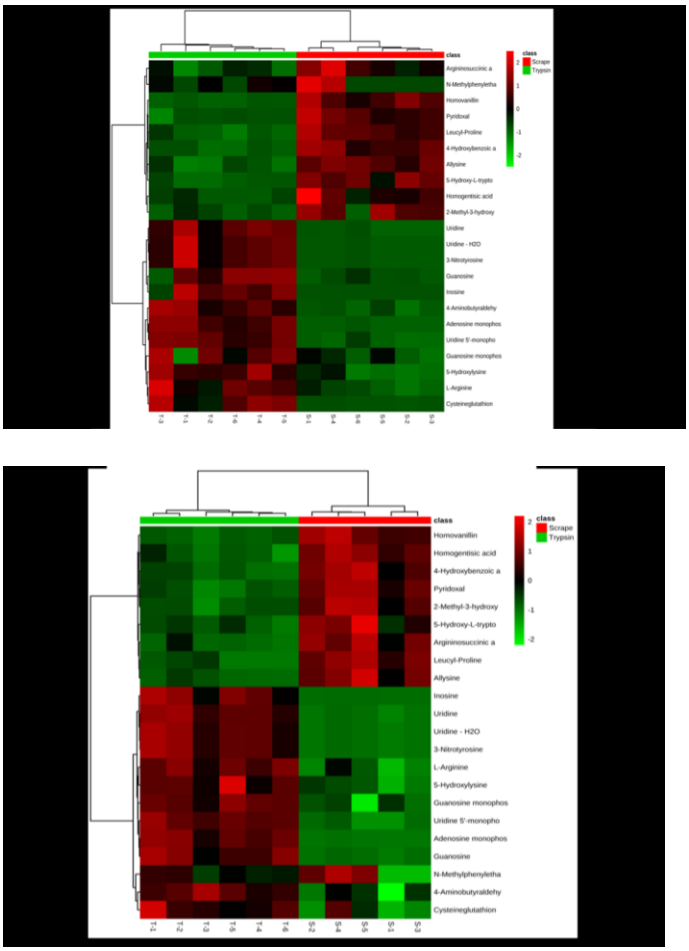


Figure 2.6. Heat maps showing 22 selected metabolites with significant concentration differences in cell extracts prepared using different harvest methods (scraping and trypsinization) from (A) HeLa and (B) MCF-7 cells. The metabolites in (A) are 1. Argininosuccinic acid; 2. N- Methylphenylethanolamine; 3. Homovanillin; 4. Pyridoxal; 5. Leucyl-Proline; 6. 4-Hydroxybenzoic acid; 7. Allysine; 8. 5-Hydroxy-L-tryptophan; 9. Homogentisic acid; 10. 2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate; 11. Uridine; 12. Uridine-H2O; 13. 3-Nitrotyrosine; 14. Guanosine; 15. Inosine; 16. 4-Aminobutyraldehyde; 17. Adenosine monophosphate; 18. Uridine 5'-monophosphate; 19. Guanosine monophosphate; 20. 5- Hydroxylysine; 21. Arginine; 22. Cysteineglutathion. The metabolites in (B) are 1. Homovanillin; 2. Homogentisic acid; 3. 4-Hydroxybenzoic acid; 4. Pyridoxal; 5. 2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate; 6. 5-Hydroxy-L-tryptophan; 7. Argininosuccinic acid; 8. Leucyl- Proline; 9. Allysine; 10. Inosine; 11. Uridine; 12. Uridine-H2O; 13. 3-Nitrotyrosine; 14. L- Arginine; 15. 5-Hydroxylysine; 16. Guanosine monophosphate; 17. Uridine 5'-monophosphate; 18. Adenosine monophosphate; 19. Guanosine; 20. N-Methylphenylethanolamine; 21. 4- Aminobutyraldehyde; 22. Cysteineglutathion.

As Figure 2.6 shows, there is a variety of metabolites with different structures having significant differences in concentration in the two groups of samples. The observed concentration differences could be attributed to the residual enzyme activity or metabolism in cells harvested by trypsinization. In the scraping harvest method, the cellular metabolism should stop immediately after MeOH was added in. In contrast, cell metabolism still could take place during trypsinization till the three time cell-washing procedure finished (about 30 min). During the trypsinization and washing process, cells would have quick and multiple changes in enzyme levels and metabolic activities, resulting in changes in concentration of some metabolites.[41] Our results are consistent with those of a previous study, which concluded that trypsinization is a more suitable technique for sub-culturing the cells but not for metabolomics study, as it was observed some metabolites related with oxidative stress changed significantly by the trypsinization process.[29]

To examine the impact of trypsinization on metabolic pathways in our study, we uploaded the identified/matched metabolites onto the Pathway Enrichment Analysis tool in Metaboanalyst. Figure 2.7 shows the enrichment analysis results. The x-axis represents the pathway impact, and the y-axis represents the negative logarithm of the p-value. Figure 2.7 shows several amino-acid and purine related metabolic pathways were significantly affected by trypsinization. One interesting pathway affected was the glutathione pathway. This finding is not surprising, as glutathione is an important antioxidant in cells. The arginine and proline pathway have the least p-value and the most impact in enrichment analysis. As an example, we mapped the detected metabolites into this pathway, and the results are shown in Figure 2.7. In this pathway, the levels of some upstream metabolites, such as glutamine, citrulline, and argininosuccinate, were decreased, while the downstream metabolite such as arginine, was

increased after trypsinization. All the polyamines, including spermine, spermidine and putrescine, were increased. Taken together, the above results show that there were metabolite level differences observed in cell samples prepared using trypsinization and scraping methods. The differences likely were caused by the trypsinization process where cell metabolism was not stopped immediately.

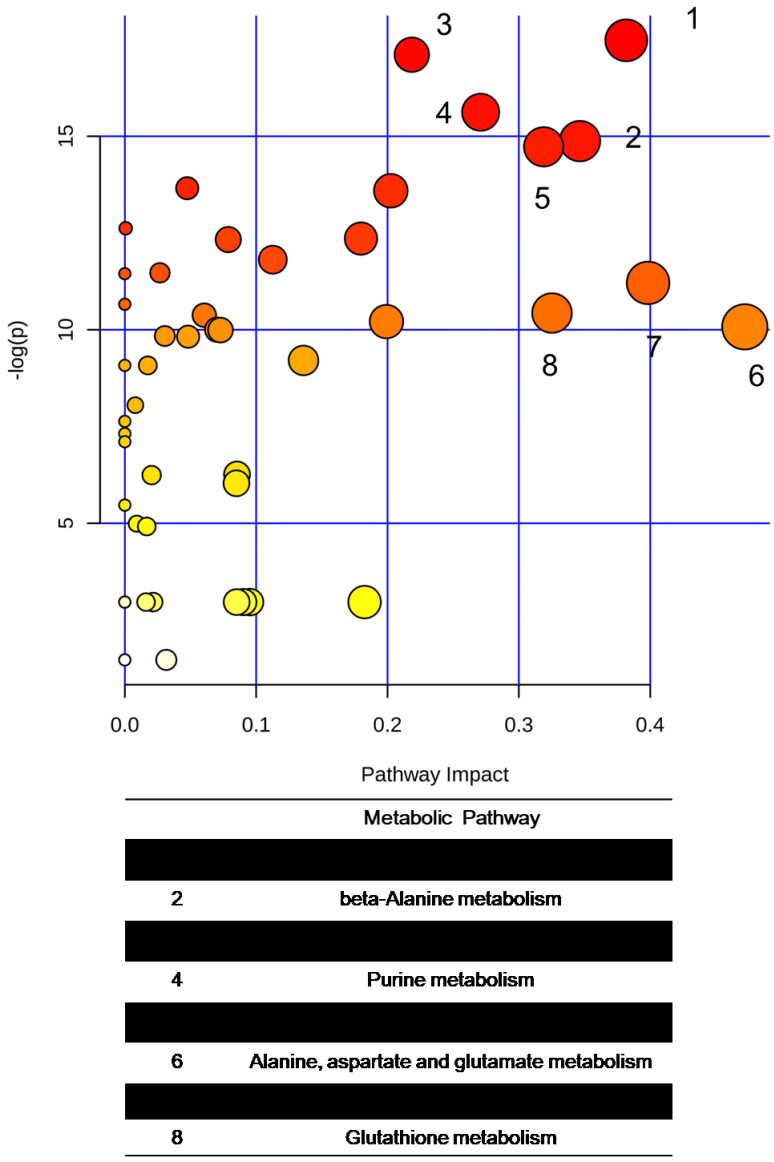


Figure 2.7. Metabolic pathways enrichment analysis. The x-axis represents the impact of the pathway, and the y-axis represents the p-value.

2.3.5. Impact of Different Lysis Methods on Cellular Metabolome

The PCA and PLS-DA analyses shown in Figure 2.3 indicate that the cellular metabolomes of samples prepared by FT and GB lysis methods do not different as significantly as those from the two harvest methods. We used the volcano plots to examine the impact of lysis methods on the cellular metabolomes of HeLa and MCF-7 cells (Figure 2.8) further. For the HeLa cells (Figure 2.9A), there are only 70 metabolites with significantly higher fold changes and 77 metabolites with significantly lower fold changes found in the two lysis methods with scraping for cell harvest, compared to 429 metabolites with higher fold changes and 305 metabolites with lower fold changes found in the two harvest methods. Similarly, for the MCF-7 cells (Figure 2.9B), only 85 metabolites with higher fold changes and 37 metabolites with lower fold changes were

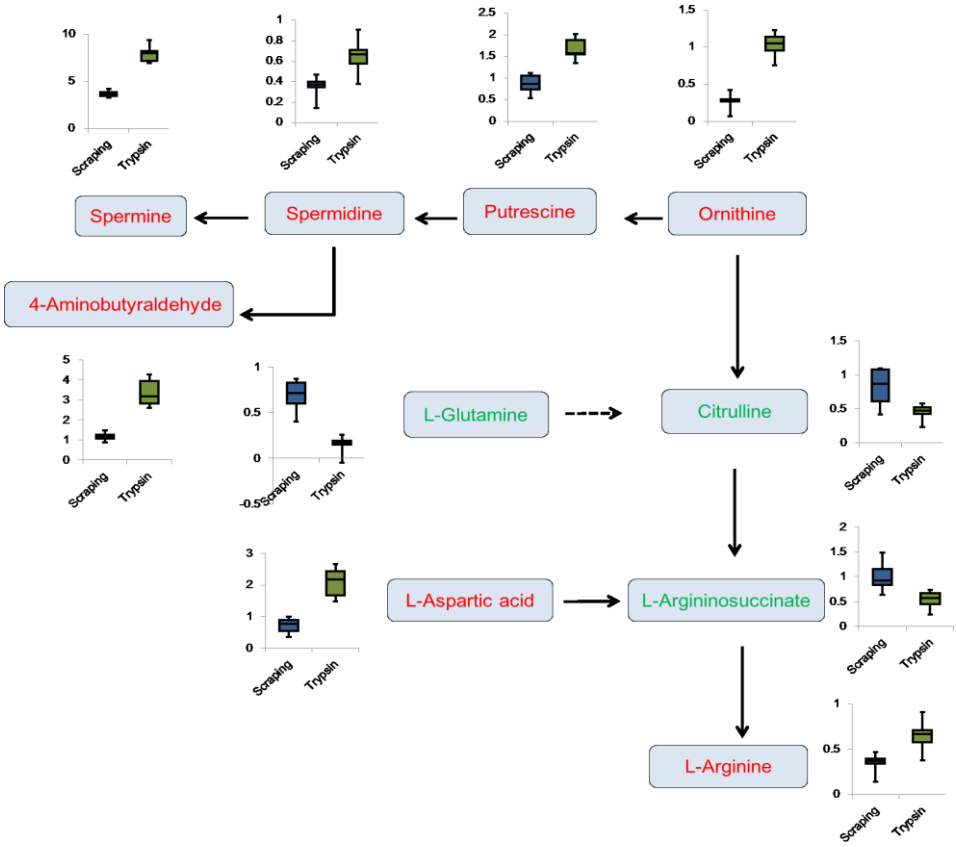


Figure 2.8. Metabolite changes in selected metabolic pathways. The box plots show the relative metabolite abundances in different harvesting groups.

detected with scraping harvest. This observation of smaller impact by the lysis method is not surprising, considering that the cellular metabolism had been quenched already in the cell harvest step, and the cellular metabolite levels should not change without active enzymes. Some differences in metabolite levels were observed in the samples prepared by the GB and FT methods. These differences could be attributed to the variations in lysis efficiencies and the extent of metabolite loss in these two methods. These results suggest that using the same method for cell lysis is important for comparing the metabolomes of different groups of cells.

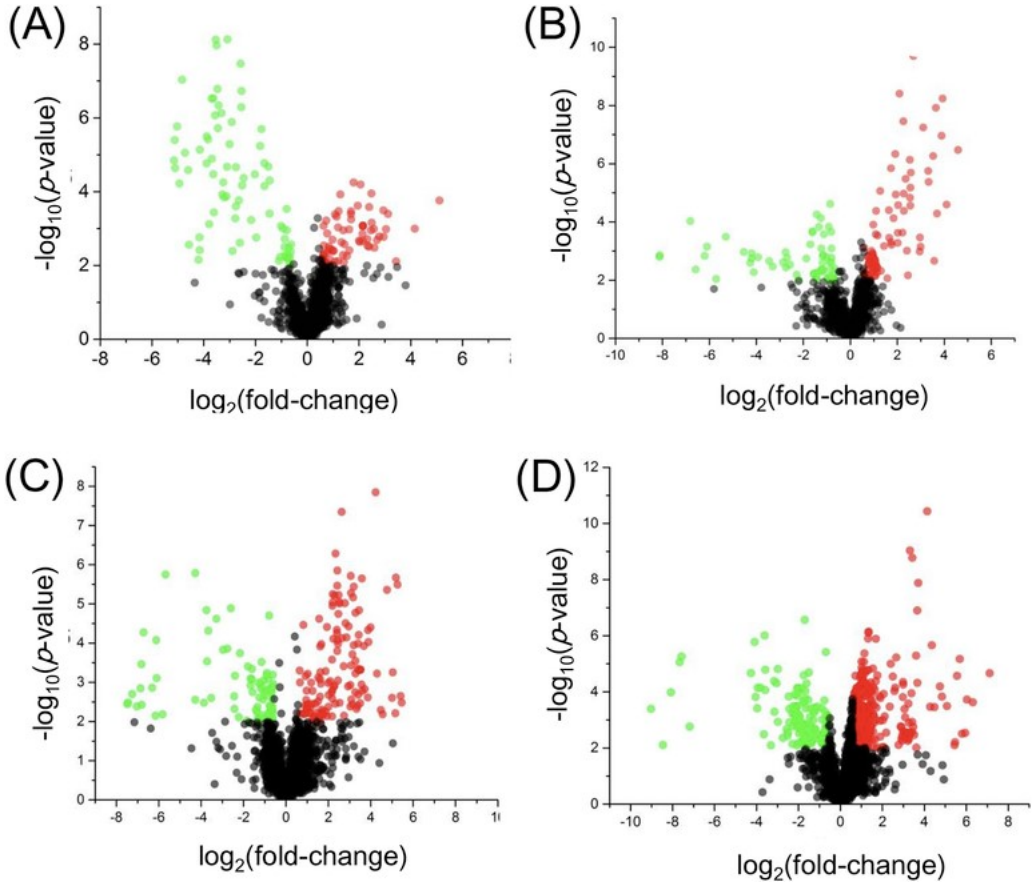


Figure 2.9. Volcano plots for comparison of the amine/phenol submetabolomes of cell extracts prepared using different lysis methods: (A) HeLa and (B) MCF-7 cells harvested by scrapping; (C) HeLa and (D) MCF-7 cells harvested by trypsinization. The p-value was from the t-test, and the fold change was calculated from the glass-bead/freeze-thaw-cycle. The red points represent the metabolites with higher concentrations in glass-bead lysed samples, and the green points represent the metabolites with lower concentrations in glass-bead lysed samples and the black points represent the metabolites with no significant differences in the two lysis methods.

We detected more significantly changed metabolites between the FT and GB lysis methods from the cells with trypsinization harvest. For the HeLa cells (Figure 2.9C), 185 metabolites with higher concentrations and 81 metabolites with lower concentrations metabolites were found in the two lysis methods. For the MCF-7 cells (Figure 2.9D), 341 metabolites with higher concentrations and 134 metabolites with lower concentrations were detected. This larger difference may be caused by the cell membrane damage or metabolite leak during the trypsinization process. However, even with a larger number of significantly changed metabolites, the cells harvested by trypsinization treated with different lysis methods still could not be separated on the PCA and PLS-DA plots, as was shown in Figure 2 and discussed in Section 2.3.4. Thus, the impact of cell lysis methods was relatively small.

In choosing the lysis method for cellular metabolomics, both GB lysis and FT lysis use physical disruption to lyse the cells with no chemical or surfactant added and thus are compatible with the downstream sample processing and analysis in CIL LC-MS. However, based on the LC-UV quantification results, FT lysis gave higher lysis efficiency. In addition, the freeze-thaw-cycle is easy to perform, although liquid nitrogen is required for fast processing. We conclude that, if liquid nitrogen is readily available, the FT lysis is preferred for lysis of adherent mammalian cells. If liquid nitrogen is not available, the GB lysis method can be used. It should be noted that for some bacteria cells and yeast cells that have tough cell walls, more aggressive lysis such as ultrasonication lysis[42] or glass-bead-assisted lysis[35] should be applied.

2.4. Conclusions

We have examined two cell harvest methods (trypsinization and scraping) and two cell lysis methods (freeze-thaw-cycle and glass-bead-assisted) to evaluate the effects of their combinations

on cellular metabolome results. Based on the data obtained from LC-UV measurement of the total concentration of dansyl labeled metabolites in each cell extract and 13C-/12C-dansylation LC-MS analysis of the amine/phenol submetabolome, we concluded that the combination of the scraping and the freeze-thaw-cycle methods is a simple and effective method for harvesting and lyzing adherent mammalian cells for CIL LC-MS metabolomics. We envisage a wide use of this protocol for cellular metabolomics where comprehensive and quantitative analysis of the chemical-group-based submetabolomes is done using multiple chemical labeling LC-MS.

Chapter 3

Metabolomics Distinguishes DOCK8 Deficiency from Atopic Dermatitis: A Biomarker Discovery

3.1 Introduction

Hyper Immunoglobulin E syndrome (HIES), initially known as Job syndrome, now represents a constellation of primary immunodeficiency (PID) disorders, resulting in markedly elevated serum IgE levels, eczema, and predisposition to Staphylococcal and sinopulmonary infections. David et al. (1966) first described Job syndrome in two girls who exhibited atopy, recurrent sinopulmonary and Staphylococcal skin infection.[43] HIESs are a group of genetic disorders, characterized by an increased susceptibility to specific infections, and in severe cases, an increased incidence of malignancy, leading to premature death.

Depending upon the clinical features and genetics, there exist two common forms of HIES: autosomal recessive type (AR-HIES) and autosomal dominant type (AD-HIES). AR-HIES is mainly caused by mutations in the dedicator of cytokinesis 8 (DOCK8) and less likely by mutations in phosphoglucomutase 3 (PGM3) and interleukin 6-signalling transducer (IL6ST) genes. AD-HIES normally is a mutated signal transducer activator of transcription 3 (STAT3). [43-46] DOCK8 is a cytoskeletal protein, which contains two related conserved protein domains, DHR1 and DHR2 with homozygous and heterozygous DOCK8 mutations; they have been reported with frequent large deletions and point mutations, leading to protein loss of function.[47,48] The production of DOCK8 protein in the immune system, especially in

lymphocytes, is elevated. Moreover, its expression also is found in the placenta, kidney, lung, and pancreas.[49] DOCK8-deficient patients develop atopic dermatitis, asthma, and severe allergies to food and environmental antigens in early infancy.[50] In addition, chronic viral infections also are distinctive features, with the common pathogens being herpes simplex virus (HSV), human papillomavirus (HPV) molluscum contagiosum virus (MCV), and varicella-zoster virus (VZV).

More recently, AR-HIES has been reported in a smaller subset of patients as a result of mutations in PGM3 and IL6ST.[51,52] In contrast, STAT3 was identified as the only causative gene of AD-HIES.[53-56] AD-HIES is also a multisystem disorder with skin, skeletal, vascular, connective tissue and immune involvement. STAT3 deficient patients develop a distinctive craniofacial profile (from childhood towards teenage), which includes facial asymmetry, prominent forehead, broad nose, deep set eyes, rough facial skin, and retention of primary teeth. They also are prone to developing pulmonary infections complicated by pneumatoceles and are at high risk of hematological malignancies.[55-57] In general, all HIES patients are susceptible to recurrent sinopulmonary infections caused by a wide variety of pathogens, including Streptococcus pneumoniae, Haemophilus influenzae, Pneumocystis jirovecii, Histoplasma capsulatum, and Legionella pneumophila[58].

Atopic dermatitis (AD) or eczema is a prevalent pediatric skin disease with chronic inflammatory, and specific food allergens and nutrients are related closely to the development and severity of this disease. AD is characterized by intense pruritus and occurs primarily in infants and children, with approximately 70% of cases starting before the age of 5 years. The eczematous lesions classically involve the face, scalp, and extensor surfaces of extremities. Impaired innate and adaptive immunity, environmental changes, and alterations in genes involved in epidermal barrier functions all contribute towards the clinical manifestations of this

disease.[59] These patients are susceptible to superficial infections with Staphylococcus aureus, but deep-seated infections rarely occur in the AD, unlike DOCK8 deficient patients. Treatment of AD is directed mainly towards prevention and management of infection and immunomodulation to control the associated rash and pruritus. Topical corticosteroids, systemic antibiotics, and antifungal agents are used for both prophylactic and symptomatic treatment in conjunction with topical therapy.

Atopic dermatitis and HIES share similar clinical symptoms, including eczema, eosinophilia, and characteristic elevated levels of serum IgE. Metabolomics is a rapidly growing and promising discipline, which quantifies the group of small molecules involved in intermediary metabolism encoded by genomic DNA. Over the last decade, several relevant biomarkers have been identified through both targeted and untargeted metabolomics studies, and they have been involved in complex clinical phenotypes in diverse biological systems. Significant environmental and clinical disturbances can be monitored at the metabolomic level by examining an array of different pathways that are crucial for cellular homeostasis.[60,61] Since the metabolome is complex and very dynamic, newer and more reliable quantitative technologies have enabled the discovery of biomarkers specific enough to distinguish patients in various health states from healthy subjects.[62] Chemical isotope labeling liquid chromatography-mass spectrometry (CIL LC-MS) is a robust and emerging analytical platform used in biomarker discovery, where different labeling reagents are used to target functional groups based sub-metabolomes.[63,64]

Apart from biomarkers capable of distinguishing DOCK8-deficient from AD patients,[65] definitive metabolomics biomarkers have not been identified yet. Therefore, we aimed to employ in-depth metabolomics technologies to study the metabolomics profiles of a cohort of patients

with different forms of HIES and explore biomarkers that potentially reflect disease pathogenesis and may contribute towards improved disease monitoring and, ultimately, novel clinical interventions. We, therefore, applied CIL LC-MS, targeting the amine/phenol sub- metabolomes to find novel differentially expressed biomarkers in hereditary HIES (DOCK8, PGM3, STAT3) and AD patient groups.

3.2 Material and Methods

3.2.1. Chemicals

The LC-MS grade reagents, including water, acetonitrile (ACN), methanol, and formic acid, were purchased from Fisher Scientific (Ottawa, ON) and 13C-dansyl chloride was available from the University of Alberta) (http://mcid.chem.ualberta.ca). For the cell culture, Rosewell Park Memorial Institute (RPMI) medium, penicillin, and glutamine was obtained from Sigma Chemicals,(St. Louis, MO), and Fetal bovine serum (FBS) from Gibco, Life Technologies, (Saint-Aubin, France)

3.2.2. Characteristics of The Study Population

Through the Allergy/Immunology clinics at KFSHRC, children and adults with a genetically confirmed diagnosis of hereditary (DOCK8, PGM3, STAT3) HIES syndrome and atopic dermatitis (AD) meeting the Hanifin and Rajka clinical criteria, together with healthy controls, were consented to participate in this study.[66] Patients who received bone marrow transplantation, enrolled in another clinical study, unwilling to provide informed consent, or whose sample amount was not sufficient were excluded from the study. A baseline questionnaire, including

clinical symptoms, allergies, and family history was collected. This study was approved by the Research Ethics Committee at the Office of Research Affairs of King Faisal Specialist Hospital and Research Center (KFSHRC) (RAC No. 2160 015).

3.2.3. Cell Culture

Lymphoblastoid cell lines were obtained by transformation of peripheral blood mononuclear cells (PBMCs) with Epstein-Barr virus through density gradient centrifugation according to the manufacturer's instructions, as published in our previous work.[67]

3.2.4. LC-MS

In this CIL LC-MS metabolomics workflow (Figure 3.1), each sample was labeled by 12C-dansyl chloride (DnsCl), while a pooled sample was generated by mixing all individual samples, then labeled by 13C-DnsCl.[63] The 13C-labeled pooled sample served as a reference for all the 12C- labeled individual samples. Each sample was normalized prior to LC-MS analysis. LC-UV quantitation was performed to determine the total concentration of dansyl-labeled metabolites. Each 12C-labeled sample was mixed with the same molar amount of 13C-labeled pooled sample and injected into the LC-MS. All labeled metabolites were identified as peak pairs on mass spectra, and the peak area ratios were used for quantitative metabolomic analysis.

The serum and cell lysates processed samples were analyzed using a Thermo Fisher Scientific Dionex Ultimate 3000 UHPLC System (Sunnyvale, CA) linked to a Bruker Maxis II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA). The LC column was an Agilent reversed phase Eclipse plus C18 column (2.1 mm × 10 cm, 1.8-µm particle size, 95-Å pore size), while the mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and

solvent B was 0.1% (v/v) formic acid in acetonitrile. The LC gradient was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 99% B; t = 34 min, 99% B, with a flow rate of 0.18 mL/min. The MS conditions were as follows: polarity, positive; dry temperature, 230 °C; dry gas, 8 L/min; capillary voltage, 4500V; nebulizer, 1.0 bar; end plate offset, 500V; spectra rate, 1.0 Hz.

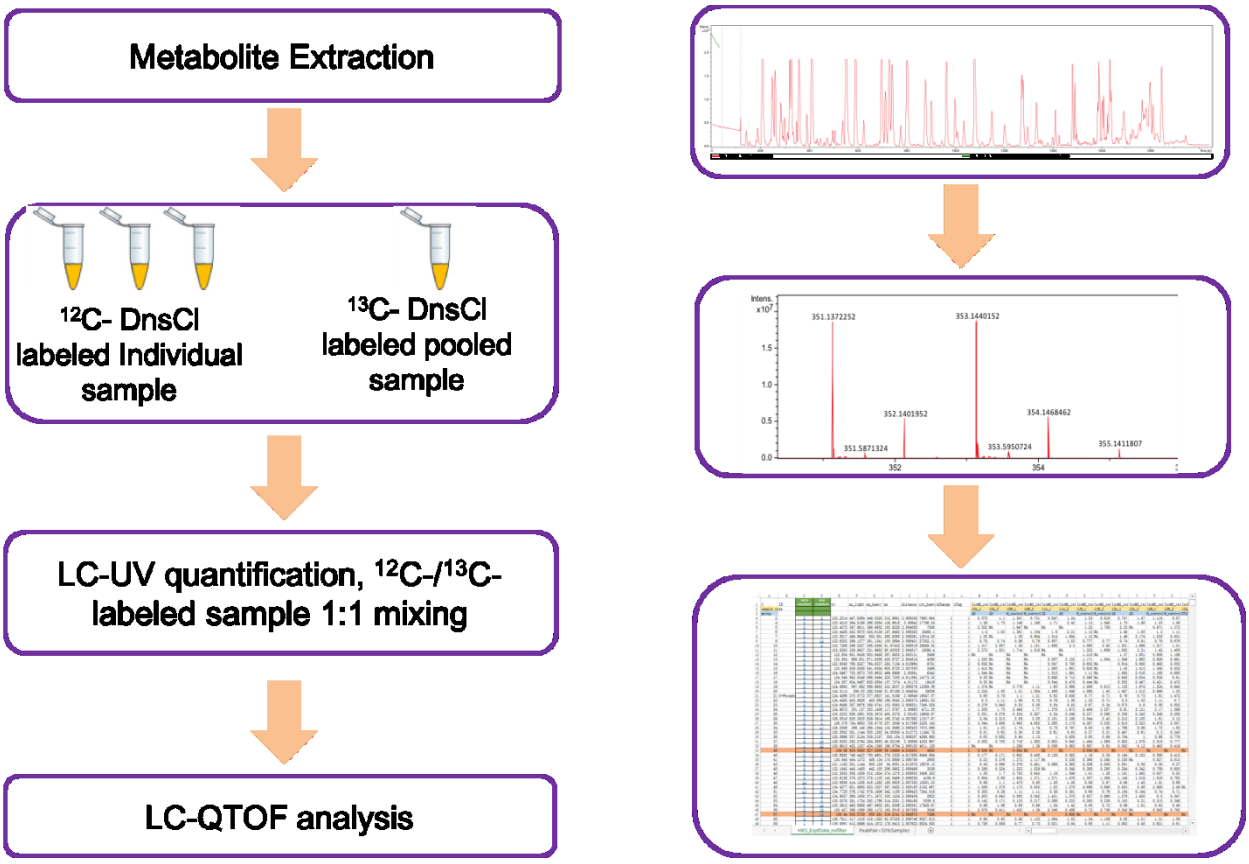


Figure 3.1. Flowchart of metabolomics workflow.

3.2.5. Data Collection, Processing, and Analysis

The LC-MS spectra were converted first to .csv files by Bruker Daltonics Data Analysis 4.3 software. The peak pairs were extracted from .csv files by IsoMS. Meanwhile the redundant pairs (e.g., those of Na^+, NH_3^+ adduct ions and their dimers) were filtered out.[68] All data generated from multiple runs were aligned together based on the peak's accurate mass and

retention time. The missing values in the aligned file were filled by Zero fill software.[69] A univariate analysis (volcano plot) was performed for each binary comparison to identify significantly differentially expressed metabolites. Here, we used a criterion of fold-change of greater than 1.5 or less than 0.67 with a q-value (false discovery rate) less than 0.05. The q-value is calculated by R script based on a p-value from a t-test. In the volcano plot, the x-axis represents the fold change (FC) between two comparison groups, and the y-axis represents the p-value. The principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed using MetaboAnalyst (www.metaboanalyst.ca).[70] The metabolites were identified positively by searching against DnsID Library (www.mycompoundid.org) using retention time and accurate mass.[71] Putative identification was performed by searching accurate mass against My Compound ID library, which contains 8,021 known human metabolites and 375,809 predicted metabolites (www.mycompoundid.org).[72]

Statistical analysis among the three groups was performed by Analysis of Variance (ANOVA) using post-hoc Tukey's method of analysis, with multiplicity adjusted p-values for each comparison. This method was chosen not only because of the unequal group sizes among the experimental and the control groups but also because it reduces the probability of making a type 1 error and supports testing of pairwise differences. Further analysis was performed on GraphPad Prism (version 6.0, Graph Pad software, LA Jolle, CA).

The Receiver Operating Characteristic (ROC) curves were constructed using random forest method MetaboAnalyst software version 3.0 (McGill University, Montreal, Canada) (http://www.metaboanalyst.ca) for global analysis. The raw data was normalized, transformed, and scaled by a median, log, and Pareto, respectively, to make sure all the data are visualized under Gaussian distribution.

3.3. Results

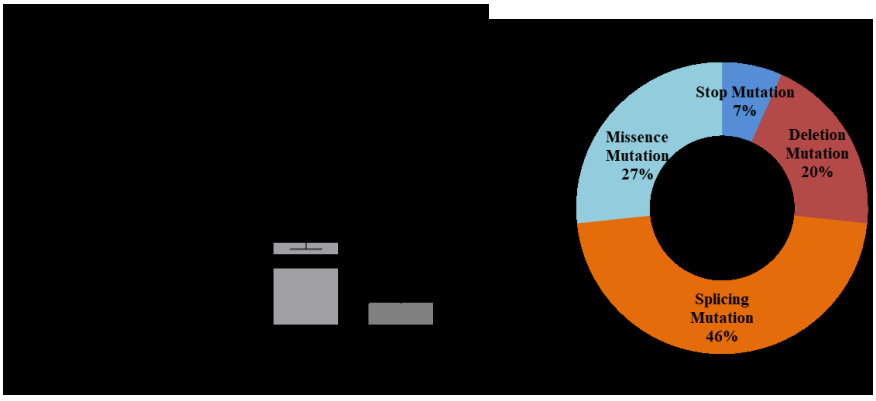
3.3.1. Clinical Characterizations of DOCK8-Deficient and AD Patients

The clinical and laboratory characteristics of the study cohorts are represented in Table 3.1. The mean age of the DOCK8-deficient and AD cohort were 13.2 ± 5.9 and 10.8 ± 1.4 years, respectively, where the Ctrl was 23 ± 1.03 (Table 3.1). Comparatively, the CD4/CD8 ratio in the DOCK8 cohort (2.8 ± 0.99) was higher than the PGM3 and STAT3 cohorts, whereas in the AD cohort the ratio was 1.43 ± 0.14. Eosinophilia was present in all patients but did not correlate with elevated IgE levels. The mean RBCs and WBCs counts in HIES patients were 4.6 ± 0.4(10^ 12/L) and 6.7 ± 1.7 (10^ 9/L), whereas in AD patients they were 5.3 ± 0.16 (10^ 12/L) and 6.74 ± 0.9 (10^9/L), respectively. The Severity Scoring of Atopic Dermatitis (SCORAD) and the Visual Analogue Scale (VAS) pruritus scores were calculated for both DOCK8 deficient and atopic dermatitis groups (Table 3.1).[73]

The most commonly seen clinical presentations in our HIES cohort were atopic dermatitis, food allergies, pneumonia, and staphylococcal infections, whereas in AD patients, pneumonia or deep-seated staphylococcus infections were not observed. As anticipated, total IgE levels in both HIES and AD groups were elevated when compared to the control group, with HIES patients showing significantly higher serum IgE levels (p-value < 0.05) compared to AD patients and controls (5-500 KU/L), as shown in Figure 3.2A. Among the HIES, cohort splicing mutations were the most common (46%), followed by missense mutations (27%), deletion mutations (20%), and stop codon mutations (7%).(Figure 3.2B)

Table 3.1. Clinical Scores (VAS and SCORAD) and Laboratory Findings in HIES and Atopic Dermatitis Cohorts. Abbreviations: AD atopic dermatitis, DOCK8 dedicator of cytokinesis, F,female, M male, PGM3 phosphoglucomutase-3, SCORAD (severity scoring of atopic dermatitis), SEM standard error of the mean, STAT3 signal transducer and activator of transcription 3, VAS (visual analogue score).

Diagnosis	Patient code	Mutation	Age (Y)	Sex (M/F)	IgE Levels (KU/L)	RBC 10^ 12/L	WBC 10^9/L	Eosinophils 10^9/L	CD4/CD8 Ratio	VAS	SCORAD
DOCK8	P1	NM_203447.3:c.[2606-1G>A; c.405_827del]	21	F	25000	0.01	9.2	24.8	2.1	15	68.7
	P2	NM_203447.3:c.[2606-1G>A; c.405_827del]	16	M	9140	4.17	3.72	35	0.8	13	67
	P3	NM_203447.3:c.5625T>G:p.Y1875	16	F	1690	4.2	4.63	50.3	1	14	59.8
	P4	NM_203447:c.827+6T>C	16	F	11240	5.9	4.9	11.4	0.73	15	77.4
	P5	NM_203447.3:c.5962_6068del	7	M	10170	5.29	9.93	26.4	0.7	17	79.5
	P6	NM_001193536:c.3949+1G>T	14	M	15550	6.82	4.46	11.7	1.3	15	62.5
	P7	NM_001193536:c.1593+1G>T	4	M	86940	5.3	24.9	31.4	9	17	70.6
	P8	NM_001193536:c.1593+1G>T	9	F	44630	5.26	12.49	34.8	8.1	9	58.2
	P9	NM_001190458:c.1905_1905+1delGG	8	F	26340	3.81	21.42	35.1	3.2	14	58.8
	P10	NM_203447.3:c.5625T>G:p.Y1875	21	M	265	4.7	9.7	12	1.1	12	70.1
Average±SEM			13.2±5.9	5/5 (M/F)	19817.30±4772.6	4.5±0.57	10.53±2.3	27.29±4.02	2.8±0.99	14.1±0.75	67.26 ±2.37
ATOPIC DERMATITIS	P11	DOCK8 mutation negative	6	M	6540	5.2	10.6	16.5	0.9	16	64.4
	P12		15	M	16500	6.39	3.72	15.6	1.4	14	71.3
	P13		10	M	1612	5.48	8.01	13.9	1.2	16	73.3
	P14		3	M	839	5.3	6.59	6.7	_	13	56.1
	P15		8	M	1221	4.9	5.09	13	1.5	9	40.8
	P16		13	M	659	5.2	5.15	13	1.2	10	54.9
	P17		15	F	1387	4.71	9.63	12.3	1.4	15	75
	P18		16	M	365	5.15	2.71	7.9	2.3	14	65.1
	P19		12	M	4587	5.66	9.2	5.1	1.5	16	70.6
Average±SEM			10.8±1.4	8/1 (M/F)	5288.20±1736.3	5.33±0.16	6.74±0.92	27.29±1.3	1.43±0.14	13.6±0.86	63.5 ± 3.702
PGM3	P20	NM_001199918:c.325G>A:p.A109T	16	M	492	5.15	2.71	0.201	0.9		
	P21	NM_001199918:c.325G>A:p.A109T	2	M	2095	4.58	5.7	0.064	2.3		
Average±SEM			9±7	2/0 (M/F)	1293.5±801.5	4.865±0.26	4.205±1.49	0.1325±0.06	1.6±0.7		



A **B**

Figure 3.2: Serum IgE levels and mutations in HIES and DOCK8 deficient patients. (A) Serum IgE levels in patients with HIES (Hyper IgE syndrome), atopic dermatitis (AD), and healthy controls (Ctrl) measured at collection point (One way ANOVA, Post hoc Tukey's method, ** p-value <0.001). (B) Distribution of mutations in HIES patients.

3.3.2. Metabolomics Profiling

Serum metabolomics profiles were generated for HEIS (DOCK8-deficient, PGM3, and STAT3deficient), AD, and Ctrl groups that showed cluster separation among these groups (Figure 3.3). Pathway analysis (Figure 3.4A) identified nitrogen (global) amino acid metabolism pathways to be the most perturbed, followed by an amino acyl-tRNA biosynthesis when DOCK8 deficiency was compared with AD and Ctrl, as shown in Figure 3.4B. The global metabolomics profile was dissected in several binary analyses for a better understanding of the distinctive contribution of each gene in the HIES group compared to either the AD or Ctrl groups. Partial least square discriminant analysis (PLS-DA) score plot demonstrates a significant separation between the DOCK8- deficient and Ctrl groups (Figure 3.4C). The univariate and volcano plot analyses also were performed, and a total of 3442 metabolites features were detected, among which a group of metabolites (n=481) was expressed differentially and visualized in the volcano plot (Figure 3.4C). The cutoff p-value has a corresponding q-value of less than 0.05and a fold change cutoff value of 1.5. Among the 481 dysregulated metabolites, 274 metabolites were up-regulated, while 207 metabolites were down-regulated in the DOCK8-deficient group (Figure 3.4C). However, only 40 metabolites were identified positively using the dansyl standard library based on exact mass and retention time match for the metabolite and its labeled internal standard.

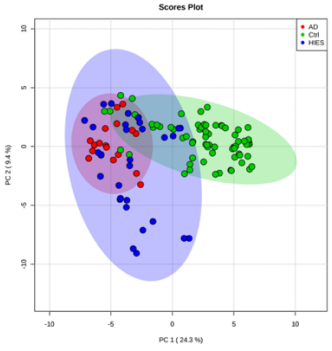


Figure 3.3. PCA score plots: HIES vs AD vs Controls.

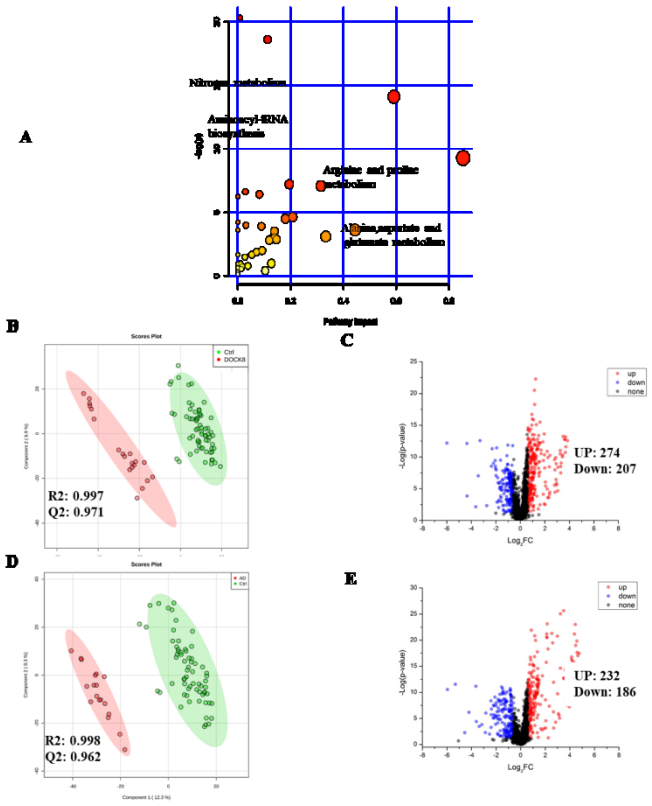


Figure 3.4. Pathway analysis and binary comparisons. (A) Pathway analysis for HIES vs Ctrl comparison. (B) DOCK8 (n=10 in duplicates) vs Control (n=33 in duplicates): PLS-DA score plot with a calculated space Q2=0.971 and R2=0.997. (C) Volcano plots (DOCK8 deficiency vs. control) with fold change >1.5 (up-regulated=274 metabolites) and <0.67 (down-regulated=207 metabolites); q=0.049, p=0.107, 40 metabolites were positively identified. (D) AD(n=9) vs Control(n=33): PLS-DA score plot, with a calculated space Q2=0.962 and R2=0.998. (E) Volcano plots with fold change >1.5 (up regulated=232) and <0.67 (down regulated=186), q=0.050, p=0.055, total 37 metabolites were identified positively. Abbreviations: DOCK8-Dedicator of cytokinesis 8, AD-Atopic dermatitis.

Similarly, the binary comparison between AD patients and Ctrl groups (Figure 3.4D), showed a clear cluster separation between the two groups (Q2 = 0.962), and a total of 418 metabolites were dysregulated, including 232 up-regulated and 186 down-regulated metabolites (Figure 3.4E). In this group, only 37 metabolites were identified positively using the dansyl standard library. Seven metabolites were positively identified using the dansyl standard library after a binary comparison between DOCK8 and AD cohorts (Figure 3.5A, B), while a total of 147 metabolites were dysregulated (118 and 29 metabolites were up- and down-regulated in the DOCK8 deficient group, respectively). The seven positively identified metabolites are presented

in Figure 3.5 C–I. Among those, aspartic acid and 3-hydroxyanthranillic acid were elevated significantly in DOCK8-deficient patients, whereas the dipeptides leucyl-phenylalanine and glycyl- phenylalanine were down-regulated compared to the AD patients. Hypotaurine, guanosine, and 2- aminooctanoic acid were not found to be significantly differentially expressed in DOCK8- deficiency compared to AD after using one way ANOVA / post-Tukey's method (Figure 3.5G-I). The patients PBMC cell line metabolomics profile was generated using the same mass spectrometry platform.

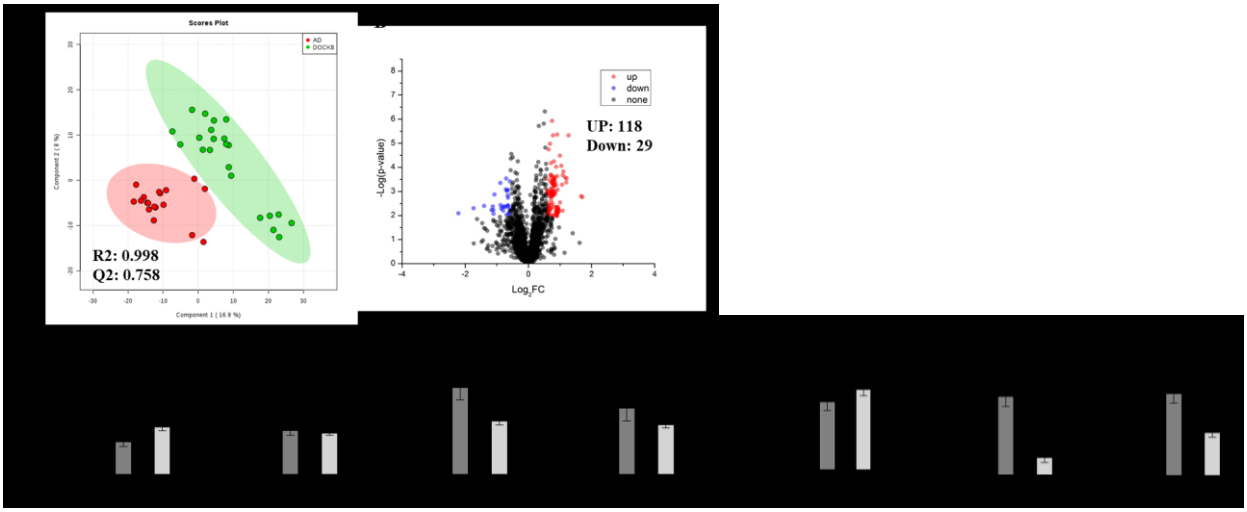


Figure 3.5. Positively identified serum metabolites in DOCK8 vs AD vs Ctrls. (A) PLS-DA score plot for binary comparison between DOCK8 and AD, with a calculated space Q2=0.758 and R2=0.998. (B) Volcano plot with fold change >1.5 (up-regulated=118) and <0.67 (down-regulated=29), total 7 metabolites were positively identified. (C) L-Aspartic acid is up-regulated in DOCK8 deficient patients compared to AD patients. (D) 3-Hydroxyxanthranillic acid is up-regulated in DOCK8 patients. Dipeptides leucyl-phenylalanine and glycyl-phenylalanine are up-regulated in AD patients compared to DOCK8 patients (E, F respectively). (G) Hypotaurine is down-regulated in DOCK8 compared to Ctrl. Guanosine is up-regulated in DOCK8 and AD patients, while 2-aminooctanoic acid is up-regulated in AD patients only (I). For paired analysis, a combination of t-test and fold change analyses is represented in this volcano plot, where the x-axis (FDR-corrected p-value) and the y-axis is a true positive. Statistical analysis was performed using one way ANOVA, post hoc Tukey's, where * Indicates significance with p-value < 0.05, ** p-value < 0.001, and otherwise not significant (ns). Abbreviations: DOCK8-Dedicator of cytokines8, AD-Atopic dermatitis, Ctrl-healthy controls.

Pathway analysis revealed some perturbations. PLS-DA analysis showed cluster separation in the binary comparisons: DOCK8 (n=7) vs Ctrl (n=4), AD (n=4) vs Ctrl (n=4), DOCK8 vs AD (Figure 3.6). 4-Hydroxybenzoic and 3-hydroxymandelic acids were the common differentially expressed metabolites in both the PMBC and serum samples of DOCK8 deficient

patients when compared with controls. When AD cohort samples were compared to Ctrls, glutamic acid, ornithine, serotonin, 1, 4-diaminobutane and aniline (a primary aromatic amine) were in common among the serum and cell line samples. No common metabolites were observed among the serum and cell lines samples for the binary comparison between DOCK8 and AD patients.

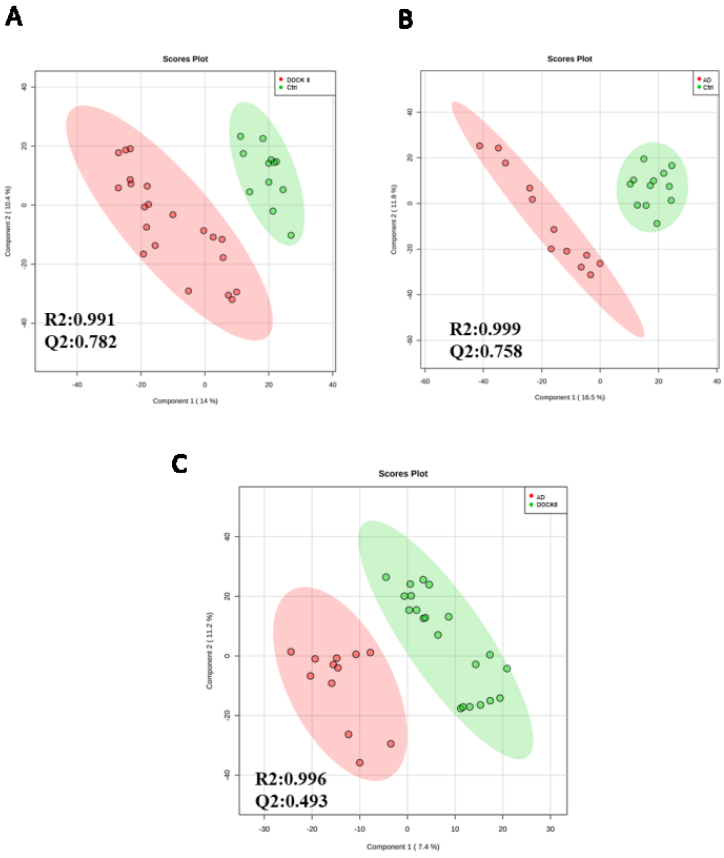


Figure 3.6. Binary comparisons of cell lines lysates metabolomic profiles (run in triplicates) (A) PLS-DA score plot in DOCK8 (n=7) vs Ctrl (n=4). (B) PLS-DA score plot in AD (n=4) vs Ctrl. (n=4) (C) PLS-DA score plot in DOCK8 vs AD. Abbreviations: DOCK8-Dedicator of cytokinesis 8, AD-Atopic dermatitis, Ctrl-healthy controls.

3.3.3. Biomarker Evaluation

As a result of the binary comparisons between DOCK8 vs Ctrl, AD vs Ctrl, and DOCK8 vs AD groups, receiver operating characteristics (ROC) exploring curves were generated (Figure 3.7A). The 95% confidence interval was calculated for these curves using 500 bootstrappings, and the optimal cutoff was determined using the furthest to diagonal line (Youden) to evaluate the sensitivity and specificity of the potential metabolites for being differentiating potential biomarkers mainly between DOCK8 deficient and AD patients. The combination of the top metabolites in ROC curves show AUCs ranging from 0.85 to 0.93 (Figure 3.7A). The significant features of the positively identified metabolites (Figure 3.7B) show the aspartic acid and 3-hydroxyanthranillic acid are being up-regulated whereas hypotaurine, leucyl-phenylalanine, glycyl-phenylalanine, guanosine, and 2-aminooctanoic acid were down-regulated in DOCK8 deficiency. The combination of all seven analytes gave the maximum confidence of differentiation and detection of DOCK8 deficiency from the AD with (AUC=0.922).

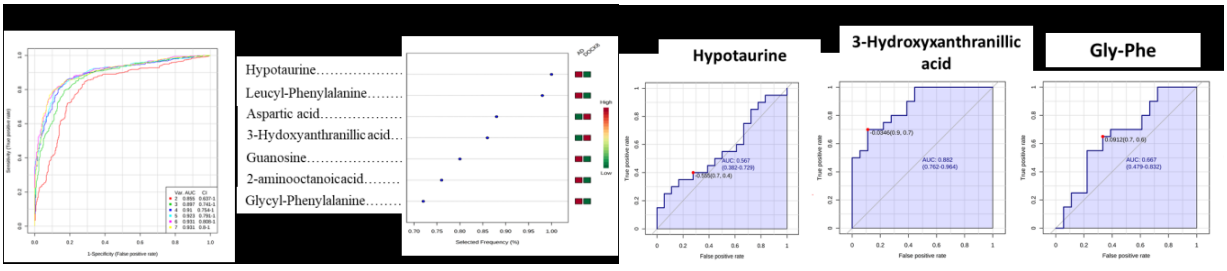


Figure 3.7. Receiver operating characteristics (ROC) curve and loading plots for positively identified metabolites in comparison between DOCK8 vs AD. (A) ROC generated by random forest model shows area under the curve (AUC) =0.931. (B) Loading plots with seven positively identified metabolites. (C) Hypotaurine is not significantly expressed in DOCK8 patients, AUC-0.597 and p value of 0.41537. (D) 3-Hydroxyxanthranillic acid is up-regulated in DOCK8 patients, AUC: 0.884 and p-value of 4.4491E5. (E) Glycyl-phenylalanine is down-regulated in DOCK8 patients compared to AD patients, AUC: 0.677 and p value of 0.04766. Data was normalized, transformed, and scaled by median, log, and Pareto scaling, respectively to make sure all the data are under Gaussian distribution. For paired analysis, a combination of t-test and fold change analyses is represented, where the x-axis (FDR-corrected p-value), and the y-axis is true positive. Abbreviations: DOCK8-Dedicator of cytokines8, AD-Atopic dermatitis, Ctrl-healthy controls.

3.4. Discussion

It is critical to recognize DOCK8 deficiency and to differentiate its various clinical and molecular forms before severe life-threatening complications arise. Differentiating HIES patients from AD can be difficult in infants and young children because of overlapping clinical and laboratory findings. The DOCK8 protein regulates intracellular signaling networks, proliferation, differentiation, migration, synapsis formation, adhesion, and survival of cells affecting innate and adaptive immunity reflecting complex function.[74-76]

The identification of predictive biomarkers to distinguish DOCK8 deficiency from AD, based on serum metabolite changes, requires a highly sensitive platform to allow the detection of very low abundant (pmol to fmol) metabolites. Chemical isotope labeling LC-MS represents a robust method for metabolomics profiling and biomarker discovery, as the 13C-labeled pool served as an internal standard and compensated for the fluctuations in MS response.[77] In this study, seven metabolic features were found to differentiate significantly between DOCK8 deficiency and AD. The expression of these potential metabolites biomarkers was studied also in PGM3 and STAT3 deficiency for further validation of DOCK8 specificity, as shown in Figure 3.8 and 3.9, respectively. Taken together, these seven differentially expressed metabolites paint a distinctive metabolomics profile in the various HIES and AD (Table 3.2). Up-regulation of 3-hydroxyanthranilic acid appears to be specific for DOCK8 deficiency compared to Ctrl and AD, while aspartic acid was only up-regulated in DOCK8 and STAT3 compared to Ctrl. Hypotaurine was down-regulated in DOCK8 deficiency compared to the AD.

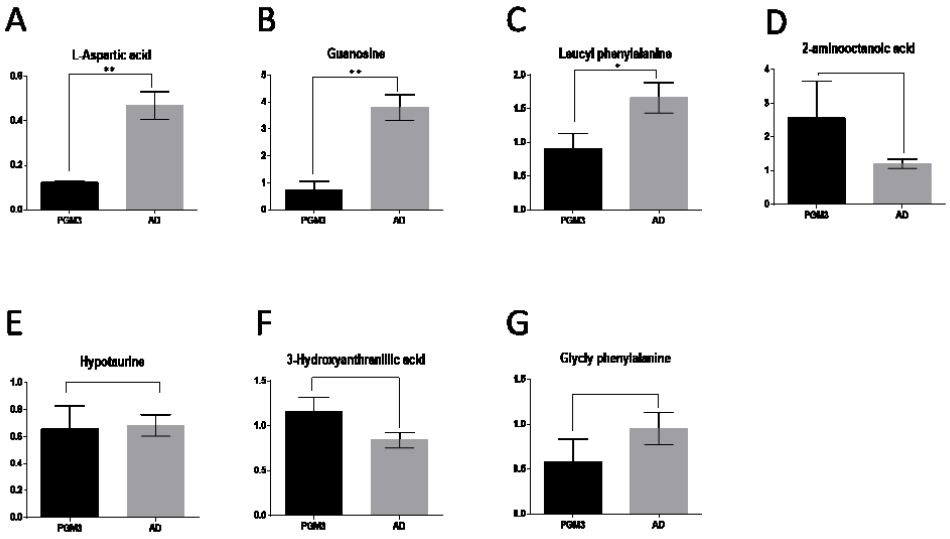


Figure 3.8. Binary comparison of positively identified serum metabolites between PGM3 deficient vs AD patients. L-Aspartic acid (A), Guanosine (B), and Leucyl-phenylalanine (C) were significantly down-regulated while 2-aminooctanoic acid (D) was up-regulated in PGM3 deficient compared to AD patients. Hypotaurine (E), 3-Hydroxyxanthranillic (F) Glycyl-phenylalanine (G) expression was not significant. Statistical analysis was performed using student t test, where * indicates significance with p-value < 0.05 and, otherwise, not significant (ns). Abbreviations: PGM3-Phosphoglucomutase3, AD-Atopic dermatitis

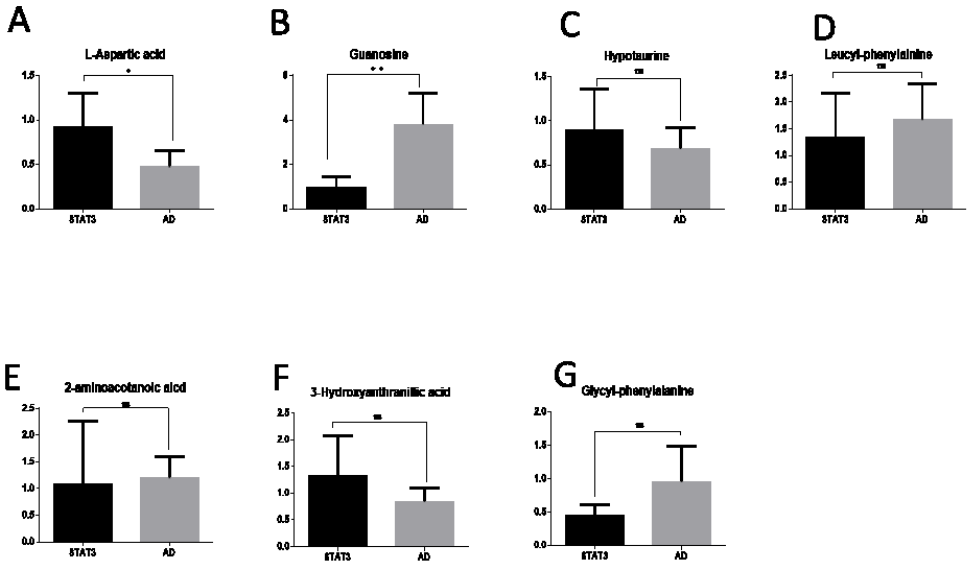


Figure 3.9. Analysis of seven positively identified metabolites in STAT3 deficient vs AD patients. Only L-aspartate (A) was significantly up-regulated while Guanosine (B) was down-regulated in STAT3 deficient compared to AD patients. The statistical analysis was performed using a student t-test, where * indicates significance with a p-value < 0.05 and, otherwise, are not significant (ns). Abbreviations: STAT3-Signal transducer and activator of transcription 3, AD-atopic dermatitis.

Table 3.2. Summary of Metabolomics Profiles in HIES and AD.

Metabolite	HMDB #	DOCK8 d f.	PGM3 def.	STAT3 def.
Aspartic acid	HMDB0000191	↑	↓	↑
3OH-anthranilic acid	HMDB0001476	↑	NS	NS
Hypotaurine	HMDB0000965	↓	NS	NS
Guanosine	HMDB0000133	↓	↓	↓
Leucyl-Phenylalanine	HMDB0013243	↓	↓	NS
Glycyl-Phenylalanine	HMDB0028848	↓	NS	NS
2-aminooctanoic acid	HMDB0000991	NS	NS	NS

The binary analyses between DOCK8 deficiency with and without various clinical complications (asthma, bronchiectasis, molluscum contagiosum, sclerosing cholangitis, candidiasis, warts, sinusitis, or malignancy) failed to demonstrate a secondary role for these phenotypes on the overall DOCK8 deficiency specific metabolites (Figure 3.10), which suggests that these metabolites are due primarily to the underlying genetic deficiency, rather than its secondary medical complications.

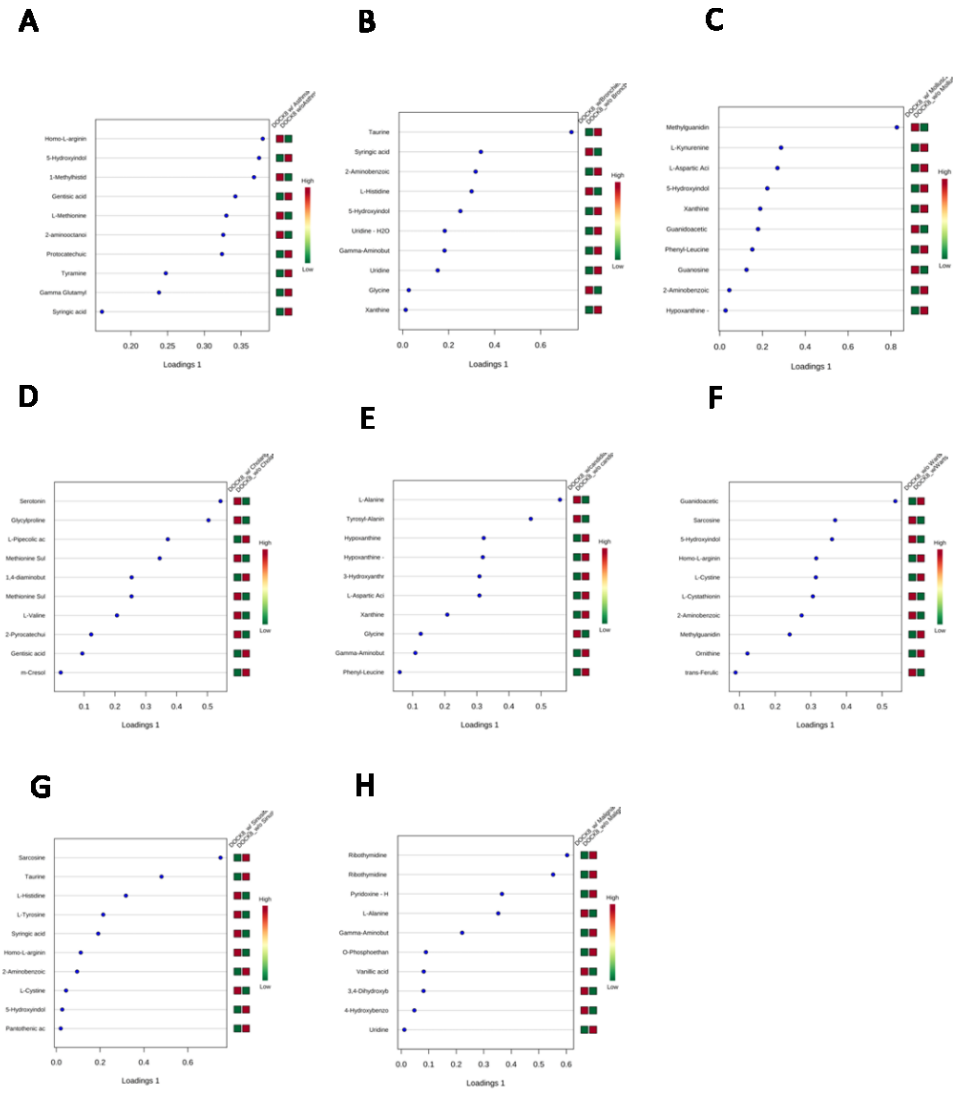


Figure 3.10. PLS-DA loading plots based on binary comparisons in DOCK8 deficient patients with/without various clinical phenotypes including (A) asthma, (B) bronchiectasis, (C) molluscum contagiosum, (D) sclerosing cholangitis, (E) candidiasis, (F) warts, (G) sinusitis, (H) malignancy.

Mammals oxidize hypotaurine to taurine using trace amounts of hydrogen peroxide (H_2O_2) produced by cellular metabolism, which is likely to be more frequent in the brain than the liver.[78] Holopainen et al.(1982) demonstrated the rapid uptake of hypotaurine into neuroblastoma cells suggesting that hypotaurine may have a function in the regulation of neuronal activity.[79] Other studies suggested a role for hypotaurine as an antioxidant and protective agent under physiological conditions.[73,80] Peng et al. (2016) also showed that under hypoxic signaling, hypotaurine behaves as an oncometabolite, promoting tumor progression.[81]

3-Hydroxyanthranillic acid (3-HAA), a tryptophan catabolism molecule produced through the kynurenine pathway, suppresses antitumor immunity in human malignancy[82]. It can inhibit Th1 and Th2 cells as an immune regulator, by increasing the percentage of regulatory T-cells, and regulate leucocyte infiltration and plaque formation.[83] It is found in the human epidermis where it participates in multiple enzymatic reactions.[84,85] Also, 3-HAA appears to play an essential role in the pathogenesis of several inflammatory, infectious, and degenerative diseases.[86] The increased tryptophan catabolism in relation to infections during the course of the disease, may lead to the increased levels of 3HAA, as seen in our DOCK8 patients (Figure 3.5B). More recently, Hongjun et al. (2017) showed that homozygous germline deficiency in 3-hydroxyanthranilic acid 3,4-dioxygenase (HAAO) causes niacin deficiency, responsible for a complex human multiple congenital (cardiac, vertebral, renal) anomalies syndrome, which can be averted in affected mice via prenatal niacin supplementation.[87]

Perturbations in amino acid metabolism also had been observed in some cancers as well as neurodegenerative disorders, such as Parkinson diseases and Alzheimer disease.[88-90] Aspartic acid, one of the major excitatory neurotransmitters, decreased in patients with depression and brain atrophy yet was found to have an elevated level in some epileptic and stroke patients. In

contrast, guanosine, a nucleoside that exerts important neuroprotective and neuromodulatory roles in the central nervous system, became a leading role to inhibit the glutamatergic neurotransmission activity. Glycyl-phenylalanine, a dipeptide produced by incomplete protein catabolism, consists of glycine and phenylalanine and is known to play an essential role in cell signaling by impacting specific amino acid degradation pathways.[91] It is transported intact by a cation independent facilitative diffusion mechanism during which the dipeptide is hydrolyzed to its component amino acids.[91] Some dipeptides, although they have not been detected in human tissues or biofluids, are simply short-lived intermediates involved in specific amino degradation pathways and proteolysis, later on affecting cell-signaling. These dipeptides also are classified as an 'expected' metabolites. 2-Aminooactanoic acid is shown to be perturbed in human colorectal cancers.[92] Taken together, these findings call for further analysis of the perturbed amino acid pathways for additional insight into its significance.

3.5. Conclusion

DOCK8 deficiency appears to be associated with a distinctive metabolomics profile characterized by significant differential overexpression of 3-HAA and aspartic acid coupled with underexpression of hypotaurine, guanosine, leucyl-phenylalanine, glycyl-phenylalanine, which together seem to contribute to some of the immune and malignancy related phenotypes observed in this disease. The complex nature of these diseases suggests that no single biomarker will be sufficient to meet the clinical needs of such patients; instead a larger panel of biomarkers will be required.

Chapter 4

Untargeted Metabolomics Profile of Metformin Effect on Type 2 Diabetes in Obese Patients

4.1. Introduction

Diabetes mellitus is a chronic condition that occurs when the blood glucose raises because of the inefficient hormone insulin production or ineffective insulin usage.[93] In 2017, diabetes affected more than 420 million people worldwide and became one of the leading causes of cardiovascular disease (CVD), blindness, kidney failure, and lower-limb amputation.[94] There are three major types of diabetes: type 1, type 2, and gestational diabetes. Among them, type 2 diabetes (T2D) is the most common diagnosis, accounting for around 90% of all cases.[93-95] In type 2 diabetes, insulin resistance is the leading cause of hyperglycemia, mainly blamed on the resistance to poor insulin production and on the inefficiency of the body's response to insulin. The onset of T2D is relatively long, and symptoms of T2D can occur at a very late stage without acute metabolic disturbance, making it difficult for diagnosis.[96] Most of the T2D patients are coupled with obesity issues, and obesity itself leads to some degree of insulin resistance but is not severe enough to cause hyperglycemia.[97] However, obesity is considered the most crucial factor in the growth of metabolic diseases. Obesity's prevalence in the Middle East is increasing; for people above 20, 70% of men are overweight in Saudi Arabia, while 74% of females have the same issue.[98] Obesity-associated metabolic conditions, including diabetes, cardiovascular diseases,

sleep apnea, and idiopathic intracranial hypertension (IIH), have not been studied explicitly in the Middle East.

In obese individuals, increased amounts of non-esterified fatty acids (NEFAs) have been released by adipose tissue, resulting in insulin resistance and β-cell dysfunction.[99] An acute increase of plasma NEFA levels will initiate glucose transport/phosphorylation inhibition and further reduce the rate of muscle glycogen synthesis and glucose oxidation. Together, insulin resistance is developed.[100] Researchers also reported that an activation of serine/threonine kinase cascade resulted in phosphorylation of insulin receptor substrates.[101] This could be caused by the NEFA delivery enhancement or reduced intracellular metabolism of fatty acids.[101] Further, the phosphorylated substrates caused diminished activity in the downstream of insulin-receptor signaling.[101]

In addition to elevated NEFAs, excess body weight also affects insulin resistance. Researchers found that body fat distribution is one of the important factors that affect insulin. Individuals with more peripherally distributed body fat tend to be more sensitive to insulin than individuals with more central fat distribution.[99] Compared to subcutaneous fat, abdominal fat is more lipolytic, and its response to the antilipolytic action of insulin is less sensitive. Therefore, intra-abdominal fat is considered one of the causes for insulin resistance and, thus, diabetes.[102]

Metformin, the drug of choice for T2D treatment that reduces glucose levels during the hyperglycemia, has other phenotypic effects, such as reduces insulin resistance, improving the tissue glucose uptake, and lowering the LDL cholesterol.[103] Lowering the LDL cholesterol is due to activation of the AMPK in the liver disturbing several pathways, including nitric oxide production by endothelial Nitric oxide synthase (eNOS). NOS is stimulated by metformin, where the mechanism of action is not understood completely.[104] Metformin suppresses hepatic

gluconeogenesis, and reduces the glucose absorption in the intestine, which alternates the composition of gut microbiota and controls the muscle glucose metabolism.[105,106] Metformin has shown associations with cancer such as breast[107], prostate[108], colorectal[109], endometrial[110], and ovarian[1, 2]. Mitochondria utilizes glucose, amino acids, and fatty acids for ATP production, redox balance, and biomass production. Metformin is believed to play a role in disrupting mitochondrial function by partially inhibiting NADH dehydrogenase or by inhibiting the hepatic glycerol phosphate dehydrogenase.[113] A metabolomics based study on tumors from ovarian cancer patients (stage III/IV) on metformin compared to patients non-exposed to metformin reveals that TCA intermediates and short-chain acylcarnitines (mitochondrial metabolism) were suppressed in the tumor of metformin-treated patients.[114] As metformin selectively kills breast cancer stem cells, a metabolomic study showed its role in depleting the TCA and glycolytic intermediates during the cell transformation, showing the evidence of inhibiting complex I of the mitochondria in cancer cell lines[115]. The metabolomics profile of Li-Fraumeni Syndrome (LFS) patients showed an increase in TCA cycle intermediates (i.e., aconitate, malate, and fumarate), and signature of fatty acid oxidation (i.e., acylcarnitines, long chain fatty acids, and 3 hydroxy fatty acids) due to metformin treatment.[116]

BCAA, aromatic amino acids (AAA), Glu/Gln, Met, and C3 and C5 acylcarnitines were found to be associated strongly with insulin resistance in a targeted metabolomics platform.[117,118] HF/BCAA feeding induced insulin resistance was accompanied by chronic phosphorylation of mTOR, JNK, and IRS1Ser307, as well as multiple acylcarnitines accumulation in muscle. Leu, Ile, Val, Phe, and Tyr levels in plasma were found to be associated with future development of Type 2 diabetes.[118,119] A dramatic drop in BCAA and C3and C5 acylcarnitines was observed in obese cases with type 2 diabetes had surgical (i.e. gastric bypass (GBP), gastric sleeve) versus

dietary interventions, which suggests, that the glucose homeostasis improvement is more dramatic in surgical than life stile methods.[120,121] The plasma levels of BCAA and AAA were found to be lowered in insulin-sensitive compared to insulin-resistant subjects treated with a sulfonylurea drug. For the same subjects, the profile was switched when they were treated with metformin.[122]

A population-based targeted metabolomics study reported decreased levels of three amino acid clusters, phenylalanine/tyrosine, citrulline/arginine, and lysine/α-aminoadipic acid in 12 metformin-treated patients as an effect of insulin sensitizer therapy.[123] The non targeted metabolomics studies on the effect of metformin on T2D patients reveals only two metabolites: citrulline and an unknown compound coded X-21365, where they were down- and up-regulated in patient groups, respectively.[124] Triglycerols (TGs) were found to be associated with diabetes risk, whereas the TGs containing fatty acids with a lower number of carbon and double-bonds lowered the risk of diabetes.[125] 2-Amino adipic acid (2-AAA) was thought to be derived from lysine metabolism. However, in diabetes, it was found to be correlated with incidents of type 2 diabetes and in agreement with an amino acid profile, which strongly suggests an independent diseases risk mechanism to produce 2-AAA.[126] Glycine was found to be a negatively-associated metabolite in a relationship of insulin sensitivity to cardiovascular risk study, compared to BCAA using a combined approach between GCMS and LCMS.[127] Oleic acid, linoleoyl-glycerophosphocholine, and α-hydroxybutyrate were correlated with impaired glucose tolerance and with a better insulin resistance state than BCAAs or glycine.[127,128]

An integrated-network between transcriptomics and metabolomics profiles generated from H295R cells treated with metformin reveals the effect of metformin in energy metabolism, sex steroid biosynthesis, the cell cycle, and immunity.[129]

However, due to the limitation of current tools for detecting a large number of metabolites, the above studies only analyzed a small fraction of the metabolites, for example, the conventional LC-MS method detected a few hundreds of metabolites semi-quantitatively. In order to identify the metabolomics profile of metformin in diabetic obese human individuals, high-performance chemical isotope labeling (CIL) liquid chromatography-mass spectrometry (LC-MS) was used in this study. CIL is used to alter the chemical and physical properties of metabolites for much-improved separation and enhanced detection sensitivity, thereby increasing the number of detectable metabolites[31]. Also, using differential isotope labeling, accurate and precise quantification of metabolite concentration differences in comparative samples (i.e., relative quantification) can be achieved[31].

4.2. Methodology

4.2.1. Metabolomics Profiling Workflow

Supplemental Figure S1 shows the schematic of the overall workflow. In this study, each sample was labeled by 12C-Dansyl Chloride (DnsCl), while a pooled sample, which was generated by mixing aliquots of all the individual samples and was labeled by 13C-DnsCl. The 13C-labeled pool served as an internal standard for all the 12C-labeled individual samples. Each sample amount of each individual sample was normalized using the LC-UV method[25]. The 12C-labeled individual sample was mixed with the same mole amount of 13C-labeled pool sample, and the mixture was injected onto a LC-MS. All the labeled metabolites were detected as peak pairs on mass spectra. The peak area ratios were used for quantitative metabolomic analysis; the same 13C-labeled pool was spiked into all 12C-labeled individual samples, thus the peak ratio values of a labeled metabolite in different samples reflected the concentration differences of this

metabolite in these samples. In other words, every 12C-labeled metabolite from an individual sample had its corresponding 13C-labeled metabolite in the pooled sample as a reference, resulting in high accuracy for relative quantification.

4.2.2. Chemicals and Reagents

The LC-MS grade reagents, including water, acetonitrile, methanol, and formic acid, were purchased from Fisher Scientific (Ottawa, ON). 13C-dansyl chloride was available from NovaMT Inc. (http://www.novamt.com) with the procedures published previously[31].

4.2.3. Serum Samples and Dansylation Labeling

Serum samples, including control (n=30), obese (n=26), and T2D (n=16) were collected and stored at -80℃. A 15-μL aliquot was taken out from each sample and serum metabolites extracted, followed by protein precipitation with 45 μL of methanol. After two hours incubation in a -20 ℃ freezer, 45 μL of supernatant were taken out and dried down, then mixed with 25 μL of water, 12.5 μL of ACN, 12.5 μL of sodium carbonate/sodium bicarbonate buffer, and 25 μL of 12C-dansyl chloride or 13C-dansyl chloride (18 mg/mL in ACN). After vortex and spinning down, the mixture was incubated at 40 ℃ for 45 min. A 5-μL sample of 250 mM NaOH was added to quench the reaction for 10 min at 40 ℃. A 25-μL aliquot of 425 mM formic acid in 1:1 ACN/H2O was added to consume excess NaOH.

4.2.4. LC-UV

Before LC-MS injections, sample normalization was performed to minimize variations in the total sample amount of individual samples when comparing samples. A step-gradient LC-UV

method[24] measured the total concentration of dansyl labeled metabolites. In brief, 5 μL of labeled sample were injected into a PhenomenesKinetes C18 column (2.1 mm×5 cm, 1.7 μm particle size, 100 Å pore size) connected to a Waters ACQUITY UPLC system (Waters, Milford, MA). Mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and mobile phase B was 0.1% (v/v) formic acid in 5% (v/v) ACN. The 6.5-min LC gradient was: t = 0 min, 0% B; t = 1 min, 0% B; t = 1.1 min, 95% B; t = 2.6 min, 95% B; t = 3.1 min, 0% B. The flow rate was 0.45 mL/min. The PDA detector was operated at 338 nm. The peak area, which represents the total concentration of dansyl-labeled metabolites, was integrated using the Waters Empower (6.00).

4.2.5. LC-MS

Each sample was labeled by 12C-DnsCl and mixed in equal amounts with a 13C-labeled pool sample, based on the quantification results from LC-UV analysis. The mixtures were analyzed by a Thermo Scientific Dionex Ultimate 3000 UHPLC System (Sunnyvale, CA) linked to a Bruker Maxis II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA). The LC column was an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size). The mobile phase A was 0.1% (v/v) formic acid in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in acetonitrile. The LC gradient was: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 99% B; t = 34 min, 99% B. The flow rate was 0.18 mL/min. The MS conditions were as follows: polarity, positive; dry temperature, 230 °C; dry gas, 8 L/min; capillary voltage, 4500V; nebulizer, 1.0 bar; end plate offset, 500V; spectra rate, 1.0 Hz.

4.2.6. Data Analysis

Bruker Daltonics Data Analysis 4.3 software was used first to convert MS spectra information into .cvs files. An in-house software IsoMS[36] was used to process the raw data generated from multiple LC-MS runs by peak picking, peak pairing, and peak-pair filtering to remove redundant peaks. IsoMS files from each injection were aligned together based on the peak's accurate mass and retention time to generate the aligned files. The missing peak pair information in aligned file were re-extracted from raw data by Zerofillsoftware.[37] The final metabolite-intensity data file can be used for statistical analysis after normalization and/or scaling. The principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed by MetaboAnalyst (www.metaboanalyst.ca). The metabolites were identified positively by searching against DnsID Library, which contains retention time, MS, and MS/MS information of 311 amine/phenol-containing metabolite standards (www.mycompoundid.org).[39] Putative identification or match was performed by searching accurate mass against MyCompoundID library, which contains 8,021 known human metabolites and 375,809 predicted metabolites (www.mycompoundid.org).[40]

4.3 Results and Discussion

4.3.1. Metabolomics Results

In this study, based on our unique characteristics of peak pair detection offered by the CIL LC-MS method, we detected a total of 3632 peak pairs. The IsoMS software filtered out redundant peak pairs, such as those from adduct ions, dimers, multimers, etc., to retain only one peak pair for each metabolite. By removing redundant information, the number of peak pairs detected

represents the number of metabolites detected. From the detected peak pairs, 216 metabolites were identified positively using both retention time and accurate mass searching against the MycompoundID (MCID) database, and 611 and 1296 metabolites were identified putatively with the zero-reaction and the one-reaction library, respectively, using accurate mass only. Thus, 62.2% of the 3632 peak pairs detected were either identified or matched, which shows the significant coverage of the submetabolome using the dansylation labeling LC-MS method for the serum samples analyzed in this study.

4.3.2. Obese versus Lean

Multivariate statistical analysis was performed to analyze the serum metabolome dataset. Figure 4.1A shows the PCA plot of three groups and quality controls (QC), while in Figure 4.1B, the QCs were excluded. In the plots, the clustered QC data ensured the reproducibility as well as the robustness of the CIL technique.

Partial least squares discriminant analysis (PLS-DA) was performed first to reveal the distinct separation between groups visually. We first analyzed the metabolome dataset to see the separation between lean and obese (Obs (-T2D/-Met)); from Figure 4.1A, the clusters of both groups were separated with Q2 = 0.737.

To analyze the metabolome change further, univariate analysis using volcano plots was performed on the metabolome set. In the volcano plot, the x-axis is the fold change (FC) of obese group over control group, and the y-axis is the p-value from the t-test comparing two groups. To determine metabolites with significance, we used the criteria of q-value (false discovery rate) less than 0.05, and FC > 1.5 (or FC < 0.67). The q-value is calculated by R script based on the p-value from the t-test. The fold-change criterion chosen was based on the technical accuracy and

reproducibility, i.e., for dansylation LC-MS, the errors and RSD values are less than ±25%. Thus, we conservatively used a ±5 0% change as the criterion. In Figure 4.1B, a total of 189 metabolites were dysregulated. Among them, 78 metabolites were up-regulated (FC > 1.5) and 111 metabolites were down-regulated (FC < 0.67) comparing obese to lean groups. By searching against our dansyl standard library using these 189 metabolites, 30 of them were identified positively. The identification results of significantly changed metabolites with their FC and p-value information are shown in Table 4.1.

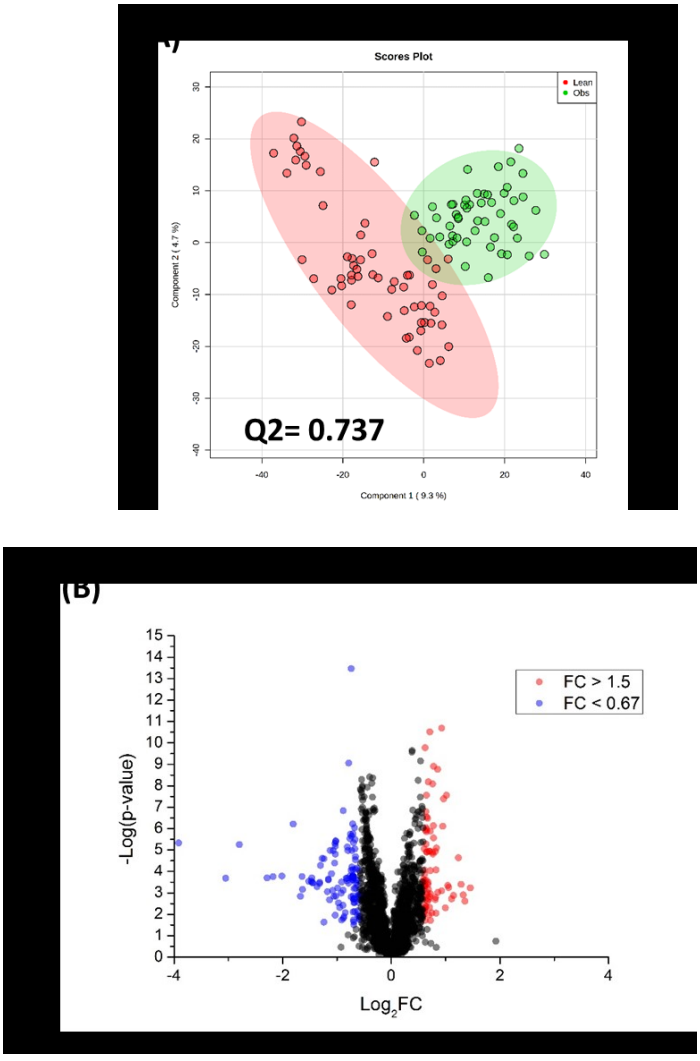


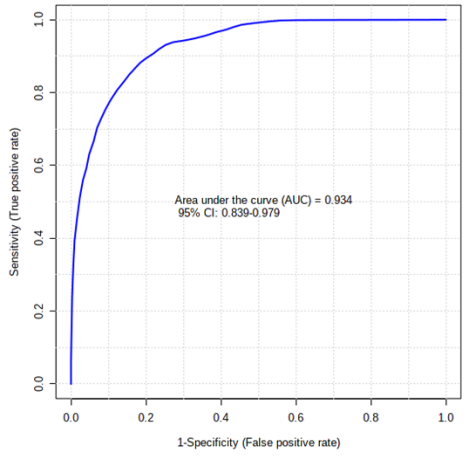
Figure 4.1. Healthy lean control and obese patients' metabolomics profile was evaluated using PLSDA analysis (A), where the clusters of both groups were separated (Q2=0.737). (B) Volcano plot of obese versus lean group. 78 metabolites were up- and 111 were down-regulated in the obese group, with fold change and FDR adjusted p-value at the cut-off 1.5(or 0.67) and 0.05, respectively.

Table 4.1 Positive Identification Results of Significantly Changed Metabolites between Lean and Obese Groups.

ID	No	rt	mz	fc	p
Asparaginyl-Asparagine	46	125.6503	246.0956	0.423058	3.86E-05
Glutaminyl-Asparagine	75	131.6984	260.1113	0.476398	0.000198
Glutaminyl-Glutamine	83	134.51	274.1275	0.489987	0.031774
L-Ornithine	119	152.6878	132.0893	0.494231	5.58E-05
Asparaginyl-Aspartic acid	164	168.4914	247.08	0.498545	5.03E-06
Guanidine	197	180.641	59.04718	0.529743	0.000905
Gamma-Glutamylglutamic acid	272	225.1985	276.0951	0.559242	1.24E-06
Alanyl-Serine	359	267.6916	176.0793	0.565238	0.009574
Glutamic Acid	398	287.0138	147.0529	0.589568	2.66E-09
Aspartic Acid	401	290.9654	133.037	0.606353	4.97E-06
Valyl-Asparagine	443	312.7338	231.1216	0.609659	1.19E-05
Tryptophyl-Asparagine	610	385.7995	318.1317	0.612233	0.000199
1-Aminocyclopropane-1-carboxylate	756	440.0404	101.0473	0.633118	1.37E-06
5-Hydroxyectoine	889	479.2398	158.0692	0.634187	3.55E-06
Phenylalanyl-Glutamate	891	479.7834	294.121	0.636742	1.99E-05
Glycyl-Valine	918	489.5974	174.0999	0.637195	0.03186
prolyl-proline	944	504.7034	212.116	0.658986	0.002969
Glutamic Acid - H2O	956	509.0195	129.0421	1.514919	4.01E-10
Serinyl-Leucine	957	511.2347	218.1264	1.547097	3.46E-08
2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate	970	519.9122	181.036	1.550853	0.000322
S-Glutathionyl-L-cysteine	1027	542.6186	426.088	1.556174	1.25E-06
Leucyl-Alanine	1183	570.0081	202.1313	1.56327	5.01E-07
Isoleucyl-Valine	1929	727.3233	230.1647	1.578142	1.13E-08
4,6-Dihydroxyquinoline	2520	901.2758	161.0503	1.626598	2.23E-11
Phenylalanyl-Phenylalanine	2527	911.921	312.1467	1.626791	0.010808
Tyrosyl-Glutamine	2597	933.646	309.1324	1.631172	0.005273
N-Acetylindoxyl	3100	1032.785	175.0663	1.648656	2.23E-05
1,4-Diaminobutane	4205	1175.911	88.1003	1.761737	7.08E-05
Serotonin	6129	1410.734	176.0947	1.765138	3.52E-09
3,4-Dihydroxybenzoate	6321	1454.165	154.0265	1.848377	1.42E-06

With the known identification information, these 30 metabolites were studied further to find a potential biomarker for obese groups. Metaboanalyst 4.0 was used to generate receiver operating characteristics (ROC) curves analysis to differentiate obese from leans. The Random Forest method was used to establish the classification model based on five metabolites with highest areas under the ROC curve (AUC) values. In Figure 4.2A shows the ROC of a model consisting of: 2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate, S-Glutathionyl-L-cysteine, Phenylalanyl-Glutamate, Serotonin, and Glycyl-Valine. Among them, 2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate and S-Glutathionyl-L-cysteine are identified putatively. This model gave an AUC value equal to 0.934, indicating a strong discrimination power between obese and lean groups. The AUC values of each metabolite individually are shown in Figure 4.2B. We also combined unidentified significant metabolites into the model to test if the differentiation power would be improved. However, after adding a few unidentified metabolites, no significant increase in performance was obtained. (Figure 4.3)

(A)



(B)

Name	Independent AUC
2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate	0.844
S-Glutathionyl-L-cysteine	0.827
Phenylalanyl-Glutamate	0.804
Serotonin	0.771
Glycyl-Valine	0.764

Figure 4.2. (A)The ROC curve build up by Random Forest using the top five AUC identified metabolites panel with AUC = 0.934. (B) AUC values of individual metabolites, among them, 2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate and S-Glutathionyl-L-cysteine are putatively identified. Phenylalanyl-Glutamate, Serotonin, and Glycyl-Valine are identified positively.

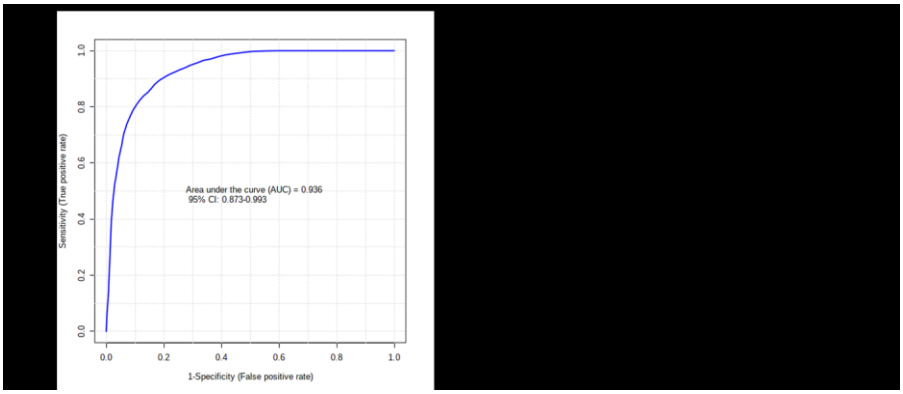


Figure 4.3. (A) The ROC curve build up by the top three unidentified metabolites and two identified metabolites with AUC = 0.936. (B)The individual AUC of each metabolites. 2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate is identified putatively and Phenylalanyl-Glutamate is identified positively.

4.3.3 Diabetes versus Obese

A clear separation was observed with Q2 = 0.885 from the PLS-DA score plot of diabase and obese groups from Figure 4.4A. The clear separation illustrates that diabetes and obese groups experienced some significant metabolome alternation. From the volcano plot shown in Figure 4.4B, 459 metabolites were up-regulated (FC > 1.5) and 166 metabolites were down-regulated (FC < 0.67) in comparing diabetes to obese groups. The cut-off of p-value here is 0.038 (when q-value = 0.05).

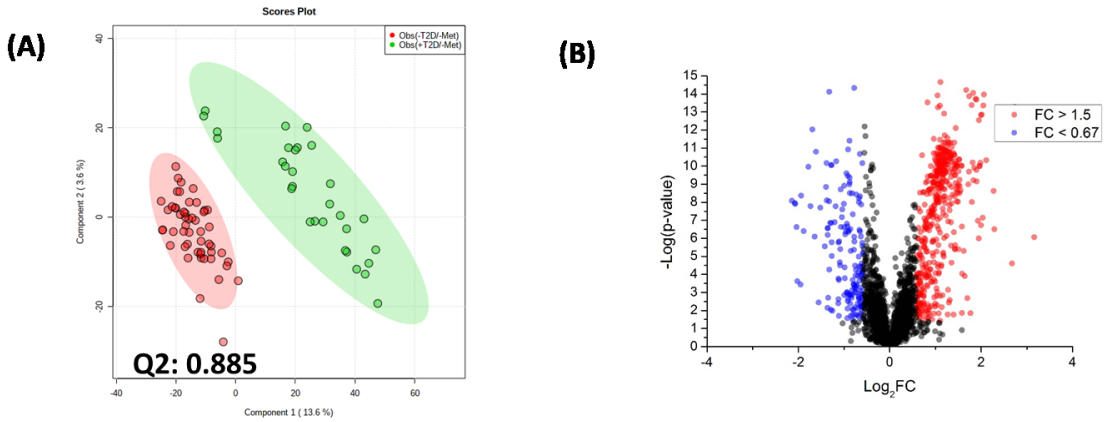


Figure 4.4. obese and diabetes patients' metabolomics profile was evaluated using PLSDA analysis(A), where the clusters of both groups were separated (Q2=0.885). (B) Volcano plot of diabetes versus obese group. 459 metabolites were up- and 166 were down-regulated in diabetes group with fold change and FDR adjusted p-value at the cut-off 1.5(or 0.67), and 0.038, respectively.

The large number of dysregulated metabolites reconfirmed the observation from PLS-DA that the metabolome of diabetes and obese patients was altered. Out of 625 metabolites, 67 were identified positively using the Dnsyl library. These 67 metabolites were studied further by ROC analysis. Their detailed information was provided in Table 4.2.

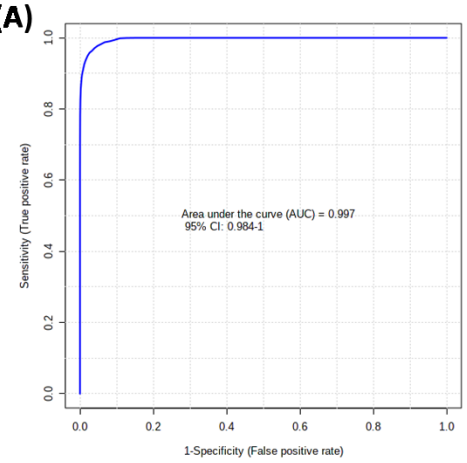
Table 4.2 Positive Identification Results of Significantly Changed Metabolites between Obese and Diabetes Groups.

Name	NO	rt	mz	fc	p
O-Phosphoethanolamine	37	124.185	141.0186	2.869162	6.84E-07
Asparaginyl-Asparagine	46	125.6503	246.0956	0.389027	3.10E-05
Glutaminyl-Asparagine	75	131.6984	260.1113	0.376805	0.001062
Glutaminyl-Glutamine	83	134.51	274.1275	1.628483	1.57E-10
Carnosine	128	155.1351	226.1058	0.618315	0.015118
Asparaginyl-Aspartic acid	164	168.4914	247.08	0.635139	0.000597
Gamma-Glutamylglutamic acid	272	225.1985	276.0951	0.39748	1.26E-15
Glycyl-Serine	299	237.1	162.0629	0.579695	4.07E-05
Alanyl-Serine	359	267.6916	176.0793	2.180799	1.46E-08
Cytidine	367	273.4085	243.0841	2.111074	7.65E-05
Aspartic Acid	401	290.9654	133.037	0.636463	3.83E-10
N-Formimino-L-glutamate	411	299.3845	174.0647	2.563758	8.67E-12
N-Formimino-L-aspartate	425	307.881	160.0504	0.608945	7.85E-06
Valyl-Asparagine	443	312.7338	231.1216	0.390149	2.72E-07
2-Amino-5-oxohexanoate	453	320.7825	145.0735	1.577619	0.012272
7-Carboxy-7-carbaguanine	517	356.2334	194.0421	1.935685	1.00E-09
L-Aspartate	573	379.7011	133.037	0.639808	0.000252
Tryptophyl-Asparagine	610	385.7995	318.1317	2.037326	0.001512
2-Hydroxyethylenedicarboxylate	665	405.1094	132.0057	1.644719	3.78E-06
Asparaginyl-Isoleucine	681	407.8613	245.1369	0.518395	0.005272
Leucyl-Glutamine	690	408.5738	259.1524	0.512163	0.0016
Phenylalanyl-Serine	714	423.775	252.1102	1.515672	0.001417
1-Aminocyclopropane-1-carboxylate	756	440.0404	101.0473	1.615413	0.002409
3,4-Dihydroxymandelate	786	447.7319	184.0336	2.089459	2.90E-09
Glutaminyl-Leucine	805	451.388	259.1529	0.576654	0.000187
4-Amino-4-deoxychorismate	809	451.999	225.0617	2.000493	4.54E-09

Isoleucyl-Threonine	865	470.3305	232.1418	0.222311	2.55E-09
5-Hydroxyectoine	889	479.2398	158.0692	3.161868	1.04E-12
Phenylalanyl-Glutamate	891	479.7834	294.121	0.640988	2.38E-09
Glycyl-Valine	918	489.5974	174.0999	0.312201	2.24E-12
Phenylalanyl-Threonine	919	490.3167	266.1262	1.527314	0.002392
Alpha-aminobutyric acid	934	499.3752	103.0629	1.525817	0.000974
prolyl-proline	944	504.7034	212.116	2.430371	1.27E-07
Serinyl-Leucine	957	511.2347	218.1264	0.295113	8.04E-09
Phenylalanyl-Glycine	964	517.8332	222.1005	2.040276	8.00E-10
Aspartyl-Leucine	968	518.9159	246.121	1.795577	0.001115
Alanyl-Valine	989	524.5014	188.1156	0.434684	4.93E-09
S-Glutathionyl-L-cysteine	1027	542.6186	426.088	2.725341	7.83E-13
Aspartyl-Phenylalanine	1029	545.2254	280.1055	1.880958	2.52E-07
Leucyl-Alanine	1183	570.0081	202.1313	1.682599	1.68E-05
Aspartyphenylalanine	1257	575.9494	280.1055	1.533361	0.000884
Glycyl-Leucine	1482	606.5119	188.1149	0.427325	1.32E-06
5-Amino-6-(5'-phospho-D-ribitylamino)uracil	1549	610.519	356.0649	2.481777	2.79E-08
Cysteinyl-Glycine dimer	1553	610.9304	354.0672	3.325434	2.02E-10
Lysyl-Glutamine	1718	652.7676	274.1636	1.81367	7.01E-05
Histidinyl-Glutamine	1892	710.7142	283.1275	0.256539	9.12E-08
Glutamyl-Lysine	1983	756.999	275.1479	2.047569	8.85E-10
Histidinyl-Serine	2098	786.4957	242.1009	0.553656	0.00046
Histidinyl-Glutamate	2145	798.8732	284.1114	0.400831	3.02E-06
Theophylline	2269	847.2595	180.0642	1.899354	0.017689
Histidinyl-Glycine	2320	859.7841	212.0906	0.254922	1.19E-09
Leucyl-Leucine	2359	872.5397	244.1783	0.420306	3.16E-08
Phenylalanyl-Leucine	2413	875.3997	278.1626	2.157993	1.23E-09
Tyrosyl-Glutamine	2597	933.646	309.1324	3.954746	6.54E-11
Histidinyl-Alanine	2608	934.9733	226.1061	0.448789	0.001355
Pyrimidodiazepine	2840	990.1734	221.0896	1.934437	1.26E-08
Tyrosyl-Serine	2981	1016.983	268.1052	1.964086	1.47E-05
Seryl-Tyrosine	3487	1092.128	268.1051	1.825555	2.26E-05
1,4-Diaminobutane	4205	1175.911	88.1003	2.151932	1.16E-10
3,4-Dihydroxybenzoate	6321	1454.165	154.0265	3.05642	0.000122
Hydroquinone	7213	1575.759	110.0371	2.031765	1.16E-08

Five metabolites of Gamma Glutamylglutamic acid, Serinyl-Leucine, Isoleucyl-Threonine, Leucyl-Leucine, and Histidinyl-Glycine were chosen to build up a ROC model with AUC = 0.997 (Figure 4.5A). The high AUC value indicated the strong discrimination power of the classification model. Their individual AUC values are provided in Figure 4.5B. Unidentified metabolites also were tested to replace metabolites with smaller AUC from above. But no model with significant increasing in discrimination power was found (Figure 4.6).

(A)

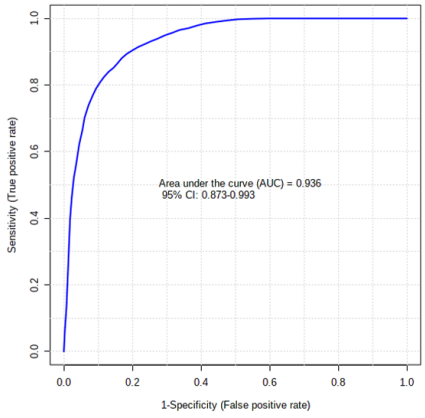


(B)

Name	Independent AUC
Gamma Glutamylglutamic acid	0.971
Serinyl-Leucine	0.965
Isoleucyl-Threonine	0.950
Leucyl-Leucine	0.948
Histidinyl-Glycine	0.934

Figure 4.5. (A)The ROC curve build-up by Random Forest using the top five AUC identified metabolites panel with AUC = 0.997. (B) AUC values of individual metabolites. All five metabolites are identified positively.

(A)



(B)

Name	Independent AUC
7508	0.852
2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate	0.844
Phenylalanyl-Glutamate	0.804
891	0.804
803	0.789

Figure 4.6. (A) The ROC curve build-up by the top three unidentified metabolites and two identified metabolites with AUC = 0.936. (B)The individual AUC of each metabolites. 2-Methyl-3-hydroxyl-5-formylpyridine-4-carboxylate is identified putatively and Phenylalanyl-Glutamate is identified positively.

4.3.4 Metformin Treated versus Diabetes

A PLS-DA plot of metformin treated and diabetes patients is shown in Figure 4.7A, with diabetes patients clustered on the left and metformin treated patients clustered on the right; a clear separation between these two groups was observed with Q2 = 0.752. As shown in Figure 4.8B, 48 metabolites were up- and 174 were down-regulated in the metformin treated group with a FDG adjusted p-value = 0.016, indicating that metformin affected the metabolome of diabetes patients. Out of 222 metabolites, 12 were identified positively, and their detailed information is provided in Table 4.3.

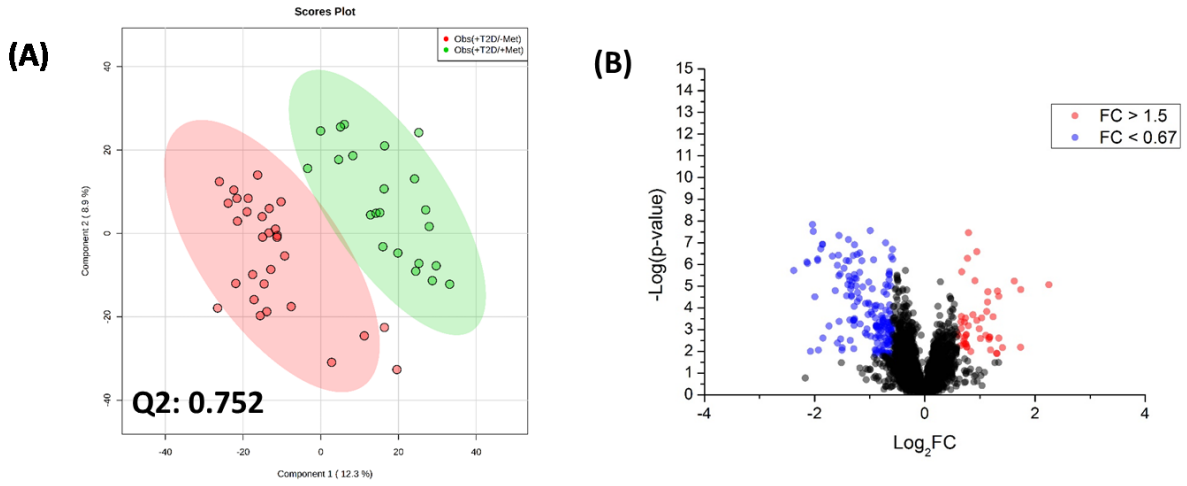


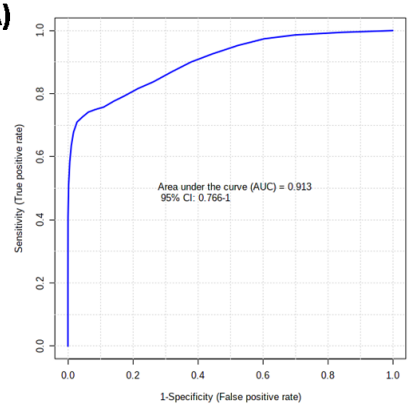
Figure 4.7. Diabetes and metformin treated patients' metabolomics profile was evaluated using PLSDA analysis (A), where the clusters of both groups were separated (Q2=0.752). (B) Volcano plot of metformin treated versus diabetes group, with 48 metabolites up- and 174 down-regulated in metformin treated group, with a fold change and FDR adjusted p-value at the cut-off 1.5(or 0.67) and 0.016, respectively.

Table 4.3. Positive Identification Results of Significantly Changed Metabolites between Diabetes and Metformin Treated Groups.

Name		rt	mz	fc	p
Gamma-Glutamylglutamic acid	272	225.1985	276.0951	1.588964	0.0017
Valyl-Asparagine	443	312.7338	231.1216	1.797321	0.003039
Isoleucyl-Threonine	865	470.3305	232.1418	2.671899	0.000648
Glycyl-Valine	918	489.5974	174.0999	2.028194	0.00093
Serinyl-Leucine	957	511.2347	218.1264	1.77051	0.001698
Aspartyl-Leucine	968	518.9159	246.121	0.501274	0.000268
S-Glutathionyl-L-cysteine	1027	542.6186	426.088	0.576978	0.000324
5-Amino-6-(5'-phospho-D-ribitylamino)uracil	1549	610.519	356.0649	0.652347	0.004825
Cysteinyl-Glycine dimer	1553	610.9304	354.0672	0.610015	0.005231
Histidinyl-Glutamate	2145	798.8732	284.1114	1.559797	0.009398
Histidinyl-Glycine	2320	859.7841	212.0906	2.28754	0.012256
Leucyl-Leucine	2359	872.5397	244.1783	1.707704	1.64E-05

Figure 4.8A shows the ROC curve with AUC = 0.913 consisting of: Leucyl-Leucine, Isoleucyl-Threonine, and S-Glutathionyl-L-cysteine. Among them, S-Glutathionyl-L-cysteine is identified putatively. Figure 4.8B shows the individual AUC of those three metabolites. After adding three unidentified metabolites to the current model, a better classification panel with AUC = 0.99 was obtained (Figure 4.9).

(A)

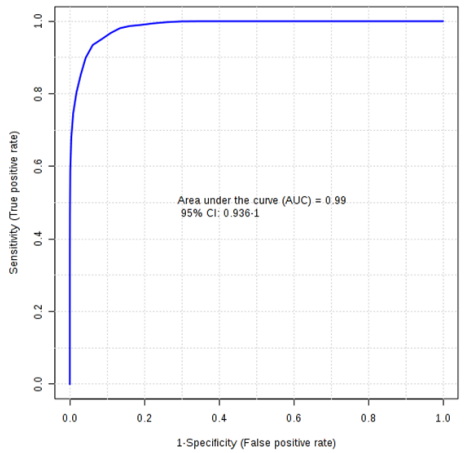


(B)

Name	Independent AUC
Leucyl-Leucine	0.893
Isoleucyl-Threonine	0.843
S-Glutathionyl-L-cysteine	0.819

Figure 4.8. (A)The ROC curve build up by Random Forest using the top five AUC identified metabolites panel with AUC = 0.913. (B) AUC values of individual metabolites. S-Glutathionyl-L-cysteine is identified putatively, and leucyl-Leucine and Isoleucyl-Threonine are identified positively.

(A)



(B)

Name	Independent AUC
2408	0.927
734	0.925
513	0.910
Leucyl-Leucine	0.893
Isoleucyl-Threonine	0.843

Figure 4.9. (A) The ROC curve build up by the top three unidentified metabolites and two identified metabolites with AUC = 0.99. (B)The individual AUC of each metabolites. Two identified metabolites were identified positively.

4.3.5. Pathway Analysis

Based on the identified metabolites, pathway analysis was performed by MetaboAnalyst to reveal the pathways related to these metabolites. There three potential pathways were affected mostly. These three pathways were: alanine, aspartate and glutamate metabolism; glycine, serine and threonine metabolism; and arginine and proline metabolism (Figure 4.10).

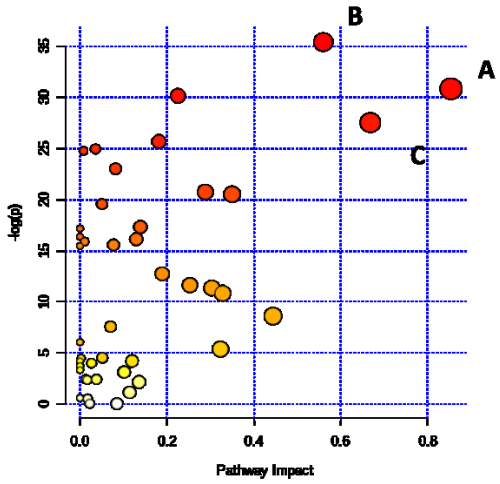


Figure 4.10: Pathway analysis generated using all identified metabolites. Three important pathways were selected based on the p-value and impact number. (A)Alanine, aspartate and glutamate metabolism. (B) Glycine, serine and threonine metabolism. (C) Arginine and proline metabolism.

The alanine, aspartate, and glutamate metabolism represented in Figure 4.10 appeared to be affected strongly by diabetes. The box plots for the detected metabolites related to this pathway, presented in Figure 4.11, revealed there their concentration differences. The significant metabolites related to the arginine, aspartate and glutamate metabolism include but are not limited to L-Aspartic acid, L-Asparagine, L-Alanine, L-Glutamine, L-Glutamic acid, and Gamma-Aminobutyric acid.

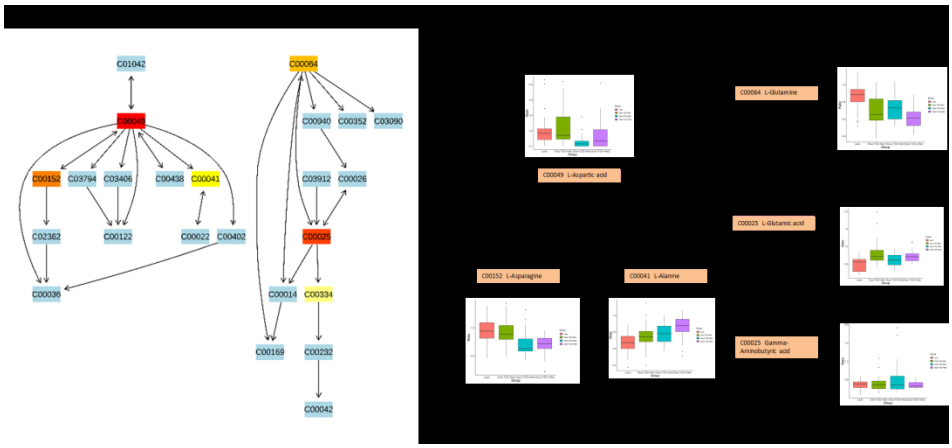


Figure 4.11. Affected metabolites and their box plots for alanine, aspartate, and glutamate metabolism, including L-Aspartic acid, L-Asparagine, L-Alanine, L-Glutamine, L-Glutamic acid, and Gamma-Aminobutyric acid.

Based on the box plots, we observed that alanine was elevated considerably for the diabetes group. Meanwhile, with the metformin treatment, the concentration of alanine is even higher. Alanine plays an important role in the glucose-aniline cycle in liver, as it acts as a gluconeogenic substrate via its metabolism to pyruvate.[130] The alanine level is usually high in diabetes patients' due to insulin deficiency, as insulin is known to possess a suppression effect on tissue proteolysis. Increased alanine concentration in the metformin treated group, suggesting that less alanine was converted under the conditions of metformin treatment. One possible explanation is that metformin may lower aminotransferase (ALT) concentrations, resulting in

building up the serum alanine level. In addition, alanine regulates the expression of genes related to β-cell signal transduction, metabolism, and apoptosis chronically[131,132].

4.4 Conclusions

In this study, our CIL LC-MS technique was applied to analyze serum samples and detected a total of 3632 peak pairs, indicating the high coverage of this method. To reveal the influence of metformin treatment, both uni-variate and multi-variate analyses were carried out to find significantly affected metabolites by metformin in type 2 diabetes patients. From the PLS-DA plot, a clear separation between two groups is observed. Among all the metabolites that could possibly cause the separation, 12 metabolites with significant fold changes were identified. These metabolites could be used for a further study on the influence of metformin. From pathway analysis, we found out that alanine, asparagine, and glutamine levels are affected by metformin.

Chapter 5

Conclusions

The CIL LC-MS technique is applied widely in metabolomics study because of its high coverage and quantification ability. In Chapter 2, CIL LC-MS was used to examine four combinations of cell harvest and lysis methods in order to build a simple and efficient platform for adherent mammalian cells. Physical scraping and trypsinization were treated as cell harvest methods, while glass-bead-assisted and freeze-thaw-cycle were tests for lysis efficiency. HeLa and MCF-7 cells, as common mammalian cells, were tested to cross-validate the findings. Results showing that scraping combined with freeze-thaw-cycle lysis gave the highest total metabolite concentration. Comparing to scraping, trypsinization caused more significant metabolome changes likely due to metabolite leakage and metabolite level changes. The cellular metabolomes obtained from the two lysis methods were found to be similar; however, freeze-thaw-cycle lysis gave a higher lysis efficacy, compared to the glass bead method. We concluded that the combination of scraping and freeze-thaw-cycle was optimal for harvesting and lysing adherent mammalian cells for CIL LC-MS metabolomics.

In Chapter 3, CIL LC-MS was involved in comparative profiling the metabolome of serum and cell line from HIES and AD patients and in identifying potential biomarkers. In this study, two sample types: serum and cell extract were analyzed, and their results were compared and combined. Multi-variate analysis results suggested that there is a significant difference in metabolome between HIES and AD patients, although they share very a close syndrome. Uni-variate analysis was carried out to find metabolites that possibly could cause the metabolome

alternation. Seven metabolites commonly showed up in cell and serum samples and were positively identified in a binary comparison between DOCK8 and AD: Aspartic acid and 3-hydroxyanthranillic acid were up regulated in DOCK8 deficiency, whereas hypotaurine, leucyl phenylalanine, glycly-phenylalanine, and guanosine were down-regulated. Hypotaurine, 3-hydroxyanthranillic acid, and glycyphenyalanine were found to be potential biomarkers for DOCK8 deficiency. The significantly altered metabolites can be used to differentiate DOCK8 deficiency from atopic dermatitis, thus contributing towards improved understanding of HIES. Further validation of these biomarkers in larger cohorts can be used for both discrimination and establishing prognosis.

In Chapter 4, CIL LC-MS was applied for an untargeted metabolomics profile in order to study the metformin effect on type 2 diabetes in obese patients. Binary comparisons were carried out among control, obese, diabetes, and metformin treated samples. From the multi-variate analysis, the metabolome difference between diabetes and metformin treated groups is very clear. Further, volcano analysis reveals dysregulated metabolites that could be responsible for the metabolome differences. Several metabolites were identified for potential biomarker discovery. Metformin affects need to be studied further through pathway analysis.

References:

(1) Yeung, P. K. Metabolomics and Biomarkers for Drug Discovery. *Metabolites* **2018**, *8* (1).
(2) Jewett, M. C.; Hofmann, G.; Nielsen, J. Fungal Metabolite Analysis in Genomics and Phenomics. *Curr. Opin. Biotechnol.* **2006**, *17* (2), 191–197.
(3) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. Metabolomics by Numbers: Acquiring and Understanding Global Metabolite Data. *Trends Biotechnol.* **2004**, *22* (5), 245–252.
(4) Monteiro, M. S.; Carvalho, M.; Pinho, M. L. B. and P. G. de. Metabolomics Analysis for Biomarker Discovery: Advances and Challenges. *Current Medicinal Chemistry*. 2013, pp 257–271.
(5) Hollywood, K.; Brison, D. R.; Goodacre, R. Metabolomics: Current Technologies and Future Trends. *Proteomics* **2006**, *6* (17), 4716–4723.
(6) Davis, V. W.; Bathe, O. F.; Schiller, D. E.; Slupsky, C. M.; Sawyer, M. B. Metabolomics and Surgical Oncology: Potential Role for Small Molecule Biomarkers. *J. Surg. Oncol.* **2011**, *103* (5), 451–459.
(7) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. Expanding Coverage of the Metabolome for Global Metabolite Profiling. *Anal. Chem.* **2011**, *83* (6), 2152–2161.
(8) Naylor, S. Biomarkers: Current Perspectives and Future Prospects. *Expert Rev. Mol. Diagn.* **2003**, *3* (5), 525–529.
(9) Lindon, J. C.; Holmes, E.; Nicholson, J. K. Peer Reviewed: So What's the Deal with Metabonomics? **2003**.
(10) Gomase, V. S.; Changbhale, S. S.; Patil, S. A.; Kale, K. V. Metabolomics. *Curr. Drug Metab.* **2008**, *9* (1), 89–98.
(11) Belinsky, S. A.; Nikula, K. J.; Palmisano, W. A.; Michels, R.; Saccomanno, G.; Gabrielson, E.; Baylin, S. B.; Herman, J. G. Aberrant Methylation of P16(INK4a) Is an Early Event in Lung Cancer and a Potential Biomarker for Early Diagnosis. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (20), 11891–11896.
(12) Hagmar, L.; Bonassi, S.; Strömberg, U.; Brøgger, A.; Knudsen, L. E.; Norppa, H.; Reuterwall, C. Chromosomal Aberrations in Lymphocytes Predict Human Cancer: A Report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). *Cancer Res.* **1998**, *58* (18), 4117–4121.
(13) Manach, C.; Hubert, J.; Llorach, R.; Scalbert, A. The Complex Links between Dietary Phytochemicals and Human Health Deciphered by Metabolomics. *Mol. Nutr. Food Res.* **2009**, *53* (10), 1303–1315.
(14) Wang, J. H.; Byun, J.; Pennathur, S. Analytical Approaches to Metabolomics and Applications to Systems Biology. *Semin. Nephrol.* **2010**, *30* (5), 500–511.
(15) Smolke, C. D. *The Metabolic Pathway Engineering Handbook*; CRC, 2010.
(16) Weiss, R. H.; Kim, K. Metabolomics in the Study of Kidney Diseases. *Nat. Rev. Nephrol.* **2012**, *8* (1), 22–33.
(17) Nicholson, J. K.; Lindon, J. C. Metabonomics. *Nature* **2008**, *455* (7216), 1054–1056.
(18) Young Hae Choi, †; Hye Kyong Kim, †; Huub J. M. Linthorst, ‡; Johan G. Hollander, §; Alfons W. M. Lefeber, §; Cornelis Erkelens, §; Jean-Marc Nuzillard, ⊥ and; Robert Verpoorte*, †. NMR Metabolomics to Revisit the Tobacco Mosaic Virus Infection in Nicotiana Tabacum Leaves. **2006**.
(19) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G. Clinical Applications of Metabolomics in

Oncology: A Review. *Clin. Cancer Res.* **2009**, *15* (2), 431–440.
(20) Claus, S. P.; Swann, J. R. Nutrimetabonomics:Applications for Nutritional Sciences, with Specific Reference to Gut Microbial Interactions. *Annu. Rev. Food Sci. Technol.* **2013**, *4* (1), 381–399.
(21) Gowda, G. A. N.; Djukovic, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. In *Methods in molecular biology (Clifton, N.J.)*; 2014; Vol. 1198, pp 3–12.
(22) Han, W.; Li, L. Chemical Isotope Labeling LC-MS for Human Blood Metabolome Analysis; Humana Press, New York, NY, 2018; pp 213–225.
(23) Horgan, R. P.; Kenny, L. C. 'Omic' Technologies: Genomics, Transcriptomics, Proteomics and Metabolomics. *Obstet. Gynaecol.* **2011**, *13* (3), 189–195.
(24) Wu, Y.; Li, L. Determination of Total Concentration of Chemically Labeled Metabolites as a Means of Metabolome Sample Normalization and Sample Loading Optimization in Mass Spectrometry-Based Metabolomics. *Anal. Chem.* **2012**, *84* (24), 10723–10731.
(25) Wu, Y.; Li, L. Dansylation Metabolite Assay: A Simple and Rapid Method for Sample Amount Normalization in Metabolomics. *Anal. Chem.* **2014**, *86* (19), 9428–9433.
(26) Wu, Y.; Li, L. Sample Normalization Methods in Quantitative Metabolomics. *J. Chromatogr. A* **2016**, *1430*, 80–95.
(27) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* **2007**, *3* (3), 211–221.
(28) Zhou, H.; Yuen, P. S. T.; Pisitkun, T.; Gonzales, P. A.; Yasuda, H.; Dear, J. W.; Gross, P.; Knepper, M. A.; Star, R. A. Collection, Storage, Preservation, and Normalization of Human Urinary Exosomes for Biomarker Discovery. *Kidney Int.* **2006**, *69* (8), 1471–1476.
(29) García-Cañaveras, J. C.; López, S.; Castell, J. V.; Donato, M. T.; Lahoz, A. Extending Metabolome Coverage for Untargeted Metabolite Profiling of Adherent Cultured Hepatic Cells. *Anal. Bioanal. Chem.* **2016**, *408* (4), 1217–1230.
(30) Bi, H.; Krausz, K. W.; Manna, S. K.; Li, F.; Johnson, C. H.; Gonzalez, F. J. Optimization of Harvesting, Extraction, and Analytical Protocols for UPLC-ESI-MS-Based Metabolomic Analysis of Adherent Mammalian Cancer Cells. *Anal. Bioanal. Chem.* **2013**, *405* (15), 5279–5289.
(31) Guo, K.; Li, L. Differential ^{12}C-/^{13}C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. *Anal. Chem.* **2009**, *81* (10), 3919–3932.
(32) Guo, K.; Li, L. High-Performance Isotope Labeling for Profiling Carboxylic Acid-Containing Metabolites in Biofluids by Mass Spectrometry. *Anal. Chem.* **2010**, *82* (21), 8789–8793.
(33) Zhao, S.; Luo, X.; Li, L. Chemical Isotope Labeling LC-MS for High Coverage and Quantitative Profiling of the Hydroxyl Submetabolome in Metabolomics. *Anal. Chem.* **2016**, *88* (21), 10617–10623.
(34) Zhao, S.; Dawe, M.; Guo, K.; Li, L. Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Carbonyl Submetabolome. *Anal. Chem.* **2017**, *89* (12), 6758–6765.
(35) Luo, X.; Zhao, S.; Huan, T.; Sun, D.; Friis, R. M. N.; Schultz, M. C.; Li, L. High-Performance Chemical Isotope Labeling Liquid Chromatography–Mass Spectrometry for Profiling the Metabolomic Reprogramming Elicited by Ammonium Limitation in Yeast. *J.*

Proteome Res. **2016**, *15* (5), 1602–1612.
(36) Zhou, R.; Tseng, C.-L.; Huan, T.; Li, L. IsoMS: Automated Processing of LC-MS Data Generated by a Chemical Isotope Labeling Metabolomics Platform. *Anal. Chem.* **2014**, *86* (10), 4675–4679.
(37) Huan, T.; Li, L. Counting Missing Values in a Metabolite-Intensity Data Set for Measuring the Analytical Performance of a Metabolomics Platform. *Anal. Chem.* **2015**, *87* (2), 1306–1313.
(38) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0—Making Metabolomics More Meaningful. *Nucleic Acids Res.* **2015**, *43* (W1), W251–W257.
(39) Huan, T.; Wu, Y.; Tang, C.; Lin, G.; Li, L. DnsID in MyCompoundID for Rapid Identification of Dansylated Amine- and Phenol-Containing Metabolites in LC–MS-Based Metabolomics. *Anal. Chem.* **2015**, *87* (19), 9838–9845.
(40) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; et al. MyCompoundID: Using an Evidence-Based Metabolome Library for Metabolite Identification. *Anal. Chem.* **2013**, *85* (6), 3401–3408.
(41) Halliwell, B. Oxidative Stress in Cell Culture: An under-Appreciated Problem? *FEBS Lett.* **2003**, *540* (1–3), 3–6.
(42) Wu, Y.; Li, L. Development of Isotope Labeling Liquid Chromatography–Mass Spectrometry for Metabolic Profiling of Bacterial Cells and Its Application for Bacterial Differentiation. *Anal. Chem.* **2013**, *85* (12), 5755–5763.
(43) International Conference On Harmonisation, I. (2005). Validation Of Analytical Procedures: Text And Methodology. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use (pp. 13). USA.
(44) Yong, P. F.; Freeman, A. F.; Engelhardt, K. R.; Holland, S.; Puck, J. M.; Grimbacher, B., An update on the hyper-IgE syndromes. *Arthritis research & therapy* **2012,** *14* (6), 228.
(45) Renner ED, Puck JM, Holland SM, Schmitt M, Weiss M, Frosch M, et al. Autosomal recessive hyperimmunoglobulin E syndrome: a distinct disease entity. *The Journal of pediatrics*. **2004**;144(1):93-9.
(46) Grimbacher B, Holland SM, Gallin JI, Greenberg F, Hill SC, Malech HL, et al. Hyper-IgE syndrome with recurrent infections--an autosomal dominant multisystem disorder. *The New England journal of medicine*. **1999**;340(9):692-702.
(47) Engelhardt, K. R.; McGhee, S.; Winkler, S.; Sassi, A.; Woellner, C.; Lopez-Herrera, G.; Chen, A.; Kim, H. S.; Lloret, M. G.; Schulze, I.; Ehl, S.; Thiel, J.; Pfeifer, D.; Veelken, H.; Niehues, T.; Siepermann, K.; Weinspach, S.; Reisli, I.; Keles, S.; Genel, F.; Kutukculer, N.; Camcioglu, Y.; Somer, A.; Karakoc-Aydiner, E.; Barlan, I.; Gennery, A.; Metin, A.; Degerliyurt, A.; Pietrogrande, M. C.; Yeganeh, M.; Baz, Z.; Al-Tamemi, S.; Klein, C.; Puck, J. M.; Holland, S. M.; McCabe, E. R.; Grimbacher, B.; Chatila, T. A., Large deletions and point mutations involving the dedicator of cytokinesis 8 (DOCK8) in the autosomal-recessive form of hyper-IgE syndrome. *The Journal of allergy and clinical immunology* **2009,** *124* (6), 1289-302 e4.
(48) Cote, J. F.; Motoyama, A. B.; Bush, J. A.; Vuori, K., A novel and evolutionarily conserved PtdIns(3,4,5)P3-binding domain is necessary for DOCK180 signalling. *Nature cell biology* **2005,** *7* (8), 797-807.

(49) Su, H. C., Dedicator of cytokinesis 8 (DOCK8) deficiency. *Current opinion in allergy and clinical immunology* **2010,** *10* (6), 515-20.
(50) Boos AC, Hagl B, Schlesing A, Halm BE, Ballenberg r N, Pinarci M, et al. Atopic dermatitis, STAT3- and DOCK8-hyper-IgE syndroms differ in IgE-based sensitization pattern. Allergy. **2014**.
(51) Sassi, A.; Lazaroski, S.; Wu, G.; Haslam, S. M.; Fliegauf, M.; Mellouli, F.; Patiroglu, T.; Unal, E.; Ozdemir, M. A.; Jouhadi, Z.; Khadir, K.; Ben-Khemis, L.; Ben-Ali, M.; Ben-Mustapha, I.; Borchani, L.; Pfeifer, D.; Jakob, T.; Khemiri, M.; Asplund, A. C.; Gustafsson, M. O.; Lundin, K. E.; Falk-Sorqvist, E.; Moens, L. N.; Gungor, H. E.; Engelhardt, K. R.; Dziadzio, M.; Stauss, H.; Fleckenstein, B.; Meier, R.; Prayitno, K.; Maul-Pavicic, A.; Schaffer, S.; Rakhmanov, M.; Henneke, P.; Kraus, H.; Eibel, H.; Kolsch, U.; Nadifi, S.; Nilsson, M.; Bejaoui, M.; Schaffer, A. A.; Smith, C. I.; Dell, A.; Barbouche, M. R.; Grimbacher, B., Hypomorphic homozygous mutations in phosphoglucomutase 3 (PGM3) impair immunity and increase serum IgE levels. *The Journal of allergy and clinical immunology* **2014,** *133* (5), 1410-9, 1419 e1-13.
(52) Schwerd, T.; Twigg, S. R. F.; Aschenbrenner, D.; Manrique, S.; Miller, K. A.; Taylor, I. B.; Capitani, M.; McGowan, S. J.; Sweeney, E.; Weber, A.; Chen, L.; Bowness, P.; Riordan, A.; Cant, A.; Freeman, A. F.; Milner, J. D.; Holland, S. M.; Frede, N.; Muller, M.; Schmidt-Arras, D.; Grimbacher, B.; Wall, S. A.; Jones, E. Y.; Wilkie, A. O. M.; Uhlig, H. H., A biallelic mutation in IL6ST encoding the GP130 co-receptor causes immunodeficiency and craniosynostosis. *The Journal of experimental medicine* **2017,** *214* (9), 2547-2562.
(53) Holland, S. M.; DeLeo, F. R.; Elloumi, H. Z.; Hsu, A. P.; Uzel, G.; Brodsky, N.; Freeman, A. F.; Demidowich, A.; Davis, J.; Turner, M. L.; Anderson, V. L.; Darnell, D. N.; Welch, P. A.; Kuhns, D. B.; Frucht, D. M.; Malech, H. L.; Gallin, J. I.; Kobayashi, S. D.; Whitney, A. R.; Voyich, J. M.; Musser, J. M.; Woellner, C.; Schaffer, A. A.; Puck, J. M.; Grimbacher, B., STAT3 mutations in the hyper-IgE syndrome. *The New England journal of medicine* **2007,** *357* (16), 1608-19.
(54) Minegishi, Y.; Saito, M.; Tsuchiya, S.; Tsuge, I.; Takada, H.; Hara, T.; Kawamura, N.; Ariga, T.; Pasic, S.; Stojkovic, O.; Metin, A.; Karasuyama, H., Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome. *Nature* **2007,** *448* (7157), 1058-62.
(55) Renner, E. D.; Torgerson, T. R.; Rylaarsdam, S.; Anover-Sombke, S.; Golob, K.; LaFlam, T.; Zhu, Q.; Ochs, H. D., STAT3 mutation in the original patient with Job's syndrome. *The New England journal of medicine* **2007,** *357* (16), 1667-8.
(56) Freeman, A. F.; Holland, S. M., Clinical manifestations, etiology, and pathogenesis of the hyper-IgE syndromes. *Pediatric research* **2009,** *65* (5 Pt 2), 32R-37R.
(57) Freeman, A. F.; Olivier, K. N., Hyper-IgE Syndromes and the Lung. *Clinics in chest medicine* **2016,** *37* (3), 557-67.
(58) Zhang, Q.; Su, H. C., Hyperimmunoglobulin E syndromes in pediatrics. *Current opinion in pediatrics* **2011,** *23* (6), 653-8.
(59) Wollenberg, A.; Rawer, H. C.; Schauber, J., Innate immunity in atopic dermatitis. *Clinical reviews in allergy & immunology* **2011,** *41* (3), 272-81.
(60) Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass spectrometry reviews* **2007,** *26* (1), 51-78.

(61) Jacob, M.; Lopata, A. L.; Dasouki, M.; Abdel Rahman, A. M., Metabolomics toward personalized medicine. *Mass spectrometry reviews* **2017**.
(62) Park YH, Fitzpatrick AM, Medriano CA, Jones DP. High-resolution metabolomics to identify urine biomarkers in corticosteroid-resistant asthmatic children. *The Journal of allergy and clinical immunology*. **2017**;139(5):1518-24 e4.
(63) Guo, K.; Li, L., Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal Chem* **2009,** *81* (10), 3919-32.
(64) Han, W.; Li, L., Chemical Isotope Labeling LC-MS for Human Blood Metabolome Analysis. *Methods in molecular biology* **2018,** *1730*, 213-225.
(65) Jacob M, Bin Khalaf D, Alhissi S, Arnout R, Alsaud B, Al-Mousa H, et al. Quantitative profiling of cytokines and chemokines in DOCK8 deficient and Atopic dermatitis patients. Allergy. **2018**.
(66) Schultz Larsen, F.; Hanifin, J. M., Secular change in the occurrence of atopic dermatitis. *Acta dermato-venereologica. Supplementum* **1992,** *176*, 7-12.
(67) Minnie Jacob, D. B. K., Safa Alhissi, Rand Arnout, Bander Alsaud, Hamoud Al-Mousa , Andreas L Lopata, Anas M Alazami, Majed Dasouki, Anas M. Abdel Rahman, Quantitative profiling of cytokines and chemokines in DOCK8 deficient and Atopic dermatitis patients. *European journal of Allergy and ckinical immunology* **2018**.
(68) Zhou, R.; Tseng, C. L.; Huan, T.; Li, L., IsoMS: automated processing of LC-MS data generated by a chemical isotope labeling metabolomics platform. *Anal Chem* **2014,** *86* (10), 4675-9.
(69) Huan, T.; Li, L., Counting missing values in a metabolite-intensity data set for measuring the analytical performance of a metabolomics platform. *Anal Chem* **2015,** *87* (2), 1306-13.
(70) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* **2015,** *43* (W1), W251-W257.
(71) Huan, T.; Wu, Y.; Tang, C.; Lin, G.; Li, L., DnsID in MyCompoundID for rapid identification of dansylated amine- and phenol-containing metabolites in LC-MS-based metabolomics. *Anal Chem* **2015,** *87* (19), 9838-45.
(72) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G., MyCompoundID: using an evidence-based metabolome library for metabolite identification. *Anal Chem* **2013,** *85* (6), 3401-8.
(73) Reich, A.; Heisig, M.; Phan, N. Q.; Taneda, K.; Takamori, K.; Takeuchi, S.; Furue, M.; Blome, C.; Augustin, M.; Stander, S.; Szepietowski, J. C., Visual analogue scale: evaluation of the instrument for the assessment of pruritus. *Acta dermato-venereologica* **2012,** *92* (5), 497-501.
(74) Mizesko, M. C.; Banerjee, P. P.; Monaco-Shawver, L.; Mace, E. M.; Bernal, W. E.; Sawalle-Belohradsky, J.; Belohradsky, B. H.; Heinz, V.; Freeman, A. F.; Sullivan, K. E.; Holland, S. M.; Torgerson, T. R.; Al-Herz, W.; Chou, J.; Hanson, I. C.; Albert, M. H.; Geha, R. S.; Renner, E. D.; Orange, J. S., Defective actin accumulation impairs human natural killer cell function in patients with dedicator of cytokinesis 8 deficiency. *The Journal of allergy and clinical immunology* **2013,** *131* (3), 840-8.
(75) Zhang, Q.; Davis, J. C.; Lamborn, I. T.; Freeman, A. F.; Jing, H.; Favreau, A. J.; Matthews, H. F.; Davis, J.; Turner, M. L.; Uzel, G.; Holland, S. M.; Su, H. C., Combined immunodeficiency associated with DOCK8 mutations. *The New England journal of medicine* **2009,** *361* (21), 2046-55.

(76) Ham, H.; Guerrier, S.; Kim, J.; Schoon, R. A.; Anderson, E. L.; Hamann, M. J.; Lou, Z.; Billadeau, D. D., Dedicator of cytokinesis 8 interacts with talin and Wiskott-Aldrich syndrome protein to regulate NK cell cytotoxicity. *Journal of immunology* **2013,** *190* (7), 3661-9.
(77) Han, W.; Sapkota, S.; Camicioli, R.; Dixon, R. A.; Li, L., Profiling novel metabolic biomarkers for Parkinson's disease using in-depth metabolomic analysis. *Movement disorders : official journal of the Movement Disorder Society* **2017,** *32* (12), 1720-1728.
(78) Oja, S. S.; Kontro, P., Oxidation of hypotaurine in vitro by mouse liver and brain tissues. *Biochimica et biophysica acta* **1981,** *677* (3-4), 350-7.
(79) Holopainen, I.; Kontro, P.; Frey, H. J.; Oja, S. S., Taurine, hypotaurine, and GABA uptake by cultured neuroblastoma cells. *Journal of neuroscience research* **1983,** *10* (1), 83-92.
(80) Alsum, Z.; Hawwari, A.; Alsmadi, O.; Al-Hissi, S.; Borrero, E.; Abu-Staiteh, A.; Khalak, H. G.; Wakil, S.; Eldali, A. M.; Arnaout, R.; Al-Ghonaium, A.; Al-Muhsen, S.; Al-Dhekri, H.; Al-Saud, B.; Al-Mousa, H., Clinical, immunological and molecular characterization of DOCK8 and DOCK8-like deficient patients: single center experience of twenty-five patients. *Journal of clinical immunology* **2013,** *33* (1), 55-67.
(81) Gao, P.; Yang, C.; Nesvick, C. L.; Feldman, M. J.; Sizdahkhani, S.; Liu, H.; Chu, H.; Yang, F.; Tang, L.; Tian, J.; Zhao, S.; Li, G.; Heiss, J. D.; Liu, Y.; Zhuang, Z.; Xu, G., Hypotaurine evokes a malignant phenotype in glioma through aberrant hypoxic signaling. *Oncotarget* **2016,** *7* (12), 15200-14.
(82) Adams S, Braidy N, Bessede A, Brew BJ, Grant R, Teo C, et al. The kynurenine pathway in brain tumor pathogenesis. Cancer Res. 2012;72(22):5649-57.
(83) Platten, M.; Ho, P. P.; Youssef, S.; Fontoura, P.; Garren, H.; Hur, E. M.; Gupta, R.; Lee, L. Y.; Kidd, B. A.; Robinson, W. H.; Sobel, R. A.; Selley, M. L.; Steinman, L., Treatment of autoimmune neuroinflammation with a synthetic tryptophan metabolite. *Science* **2005,** *310* (5749), 850-5.
(84) Pae, H. O.; Oh, G. S.; Lee, B. S.; Rim, J. S.; Kim, Y. M.; Chung, H. T., 3-Hydroxyanthranilic acid, one of L-tryptophan metabolites, inhibits monocyte chemoattractant protein-1 secretion and vascular cell adhesion molecule-1 expression via heme oxygenase-1 induction in human umbilical vein endothelial cells. *Atherosclerosis* **2006,** *187* (2), 274-84.
(85) Kim, Y. J.; Choi, M. J.; Bak, D. H.; Lee, B. C.; Ko, E. J.; Ahn, G. R.; Ahn, S. W.; Kim, M. J.; Na, J.; Kim, B. J., Topical administration of EGF suppresses immune response and protects skin barrier in DNCB-induced atopic dermatitis in NC/Nga mice. *Scientific reports* **2018,** *8* (1), 11895.
(86) Krause, D.; Suh, H. S.; Tarassishin, L.; Cui, Q. L.; Durafourt, B. A.; Choi, N.; Bauman, A.; Cosenza-Nashat, M.; Antel, J. P.; Zhao, M. L.; Lee, S. C., The tryptophan metabolite 3-hydroxyanthranilic acid plays anti-inflammatory and neuroprotective roles during inflammation: role of hemeoxygenase-1. *The American journal of pathology* **2011,** *179* (3), 1360-72.
(87) Shi, H.; Enriquez, A.; Rapadas, M.; Martin, E.; Wang, R.; Moreau, J.; Lim, C. K.; Szot, J. O.; Ip, E.; Hughes, J. N.; Sugimoto, K.; Humphreys, D. T.; McInerney-Leo, A. M.; Leo, P. J.; Maghzal, G. J.; Halliday, J.; Smith, J.; Colley, A.; Mark, P. R.; Collins, F.; Sillence, D. O.; Winlaw, D. S.; Ho, J. W. K.; Guillemin, G. J.; Brown, M. A.; Kikuchi, K.; Thomas, P. Q.; Stocker, R.; Giannoulatou, E.; Chapman, G.; Duncan, E.

L.; Sparrow, D. B.; Dunwoodie, S. L., NAD Deficiency, Congenital Malformations, and Niacin Supplementation. *The New England journal of medicine* **2017,** *377* (6), 544-552.
(88) Sugimoto, M.; Wong, D. T.; Hirayama, A.; Soga, T.; Tomita, M., Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics* **2010,** *6* (1), 78-95.
(89) Fonteh, A. N.; Harrington, R. J.; Tsai, A.; Liao, P.; Harrington, M. G., Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects. *Amino acids* **2007,** *32* (2), 213-24.
(90) Goedert, J. J.; Sampson, J. N.; Moore, S. C.; Xiao, Q.; Xiong, X.; Hayes, R. B.; Ahn, J.; Shi, J.; Sinha, R., Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* **2014,** *35* (9), 2089-96.
(91) Reshkin, S. J.; Ahearn, G. A., Intestinal glycyl-L-phenylalanine and L-phenylalanine transport in a euryhaline teleost. *The American journal of physiology* **1991,** *260* (3 Pt 2), R563-9.
(92) Brown, D. G.; Rao, S.; Weir, T. L.; O'Malia, J.; Bazan, M.; Brown, R. J.; Ryan, E. P., Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer & metabolism* **2016,** *4*, 11.
(93) Bruno, G.; Runzo, C.; Cavallo-Perin, P.; Merletti, F.; Rivetti, M.; Pinach, S.; Novelli, G.; Trovati, M.; Cerutti, F.; Pagano, G.; Piedmont Study Group for Diabetes, E., Incidence of type 1 and type 2 diabetes in adults aged 30-49 years: the population-based registry in the province of Turin, Italy. *Diabetes care* **2005,** *28* (11), 2613-9.
(94) Holman, N.; Young, B.; Gadsby, R., Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK. *Diabetic medicine : a journal of the British Diabetic Association* **2015,** *32* (9), 1119-20.
(95) Evans, J. M.; Newton, R. W.; Ruta, D. A.; MacDonald, T. M.; Morris, A. D., Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. *Diabetic medicine : a journal of the British Diabetic Association* **2000,** *17* (6), 478-80.
(96) Sas, K. M.; Karnovsky, A.; Michailidis, G.; Pennathur, S., Metabolomics and diabetes: analytical and computational approaches. *Diabetes* **2015,** *64* (3), 718-32.
(97) Al-Goblan, A. S.; Al-Alfi, M. A.; Khan, M. Z., Mechanism linking diabetes mellitus and obesity. *Diabetes, metabolic syndrome and obesity : targets and therapy* **2014,** *7*, 587-91.
(98) Ng, M.; Fleming, T.; Robinson, M.; Thomson, B.; Graetz, N.; Margono, C.; Mullany, E. C.; Biryukov, S.; Abbafati, C.; Abera, S. F.; Abraham, J. P.; Abu-Rmeileh, N. M.; Achoki, T.; AlBuhairan, F. S.; Alemu, Z. A.; Alfonso, R.; Ali, M. K.; Ali, R.; Guzman, N. A.; Ammar, W.; Anwari, P.; Banerjee, A.; Barquera, S.; Basu, S.; Bennett, D. A.; Bhutta, Z.; Blore, J.; Cabral, N.; Nonato, I. C.; Chang, J. C.; Chowdhury, R.; Courville, K. J.; Criqui, M. H.; Cundiff, D. K.; Dabhadkar, K. C.; Dandona, L.; Davis, A.; Dayama, A.; Dharmaratne, S. D.; Ding, E. L.; Durrani, A. M.; Esteghamati, A.; Farzadfar, F.; Fay, D. F.; Feigin, V. L.; Flaxman, A.; Forouzanfar, M. H.; Goto, A.; Green, M. A.; Gupta, R.; Hafezi-Nejad, N.; Hankey, G. J.; Harewood, H. C.; Havmoeller, R.; Hay, S.; Hernandez, L.; Husseini, A.; Idrisov, B. T.; Ikeda, N.; Islami, F.; Jahangir, E.; Jassal, S. K.; Jee, S. H.; Jeffreys, M.; Jonas, J. B.; Kabagambe, E. K.; Khalifa, S. E.; Kengne, A. P.; Khader, Y. S.; Khang, Y. H.; Kim, D.; Kimokoti, R. W.; Kinge, J. M.; Kokubo, Y.; Kosen, S.; Kwan, G.; Lai, T.; Leinsalu, M.; Li, Y.; Liang, X.; Liu, S.; Logroscino, G.; Lotufo, P. A.; Lu, Y.; Ma, J.; Mainoo, N. K.; Mensah, G. A.; Merriman, T. R.; Mokdad, A. H.; Moschandreas, J.; Naghavi, M.;

Naheed, A.; Nand, D.; Narayan, K. M.; Nelson, E. L.; Neuhouser, M. L.; Nisar, M. I.; Ohkubo, T.; Oti, S. O.; Pedroza, A.; Prabhakaran, D.; Roy, N.; Sampson, U.; Seo, H.; Sepanlou, S. G.; Shibuya, K.; Shiri, R.; Shiue, I.; Singh, G. M.; Singh, J. A.; Skirbekk, V.; Stapelberg, N. J.; Sturua, L.; Sykes, B. L.; Tobias, M.; Tran, B. X.; Trasande, L.; Toyoshima, H.; van de Vijver, S.; Vasankari, T. J.; Veerman, J. L.; Velasquez-Melendez, G.; Vlassov, V. V.; Vollset, S. E.; Vos, T.; Wang, C.; Wang, X.; Weiderpass, E.; Werdecker, A.; Wright, J. L.; Yang, Y. C.; Yatsuya, H.; Yoon, J.; Yoon, S. J.; Zhao, Y.; Zhou, M.; Zhu, S.; Lopez, A. D.; Murray, C. J.; Gakidou, E., Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **2014,** *384* (9945), 766-81.

(99) Al-Goblan, A. S.; Al-Alfi, M. A.; Khan, M. Z., Mechanism linking diabetes mellitus and obesity. *Diabetes Metab Syndr Obes* **2014,** *7*, 587-591.

(100) Roden, M.; Price, T. B.; Perseghin, G.; Petersen, K. F.; Rothman, D. L.; Cline, G. W.; Shulman, G. I., Mechanism of free fatty acid-induced insulin resistance in humans. *J Clin Invest* **1996,** *97* (12), 2859-2865.

(101) Shulman, G. I., Cellular mechanisms of insulin resistance. *J Clin Invest* **2000,** *106* (2), 171-176.

(102) Fain, J. N.; Madan, A. K.; Hiler, M. L.; Cheema, P.; Bahouth, S. W., Comparison of the Release of Adipokines by Adipose Tissue, Adipose Tissue Matrix, and Adipocytes from Visceral and Subcutaneous Abdominal Adipose Tissues of Obese Humans. *Endocrinology* **2004,** *145* (5), 2273-2282.

(103) Xu, T.; Brandmaier, S.; Messias, A. C.; Herder, C.; Draisma, H. H.; Demirkan, A.; Yu, Z.; Ried, J. S.; Haller, T.; Heier, M.; Campillos, M.; Fobo, G.; Stark, R.; Holzapfel, C.; Adam, J.; Chi, S.; Rotter, M.; Panni, T.; Quante, A. S.; He, Y.; Prehn, C.; Roemisch-Margl, W.; Kastenmuller, G.; Willemsen, G.; Pool, R.; Kasa, K.; van Dijk, K. W.; Hankemeier, T.; Meisinger, C.; Thorand, B.; Ruepp, A.; Hrabe de Angelis, M.; Li, Y.; Wichmann, H. E.; Stratmann, B.; Strauch, K.; Metspalu, A.; Gieger, C.; Suhre, K.; Adamski, J.; Illig, T.; Rathmann, W.; Roden, M.; Peters, A.; van Duijn, C. M.; Boomsma, D. I.; Meitinger, T.; Wang-Sattler, R., Effects of metformin on metabolite profiles and LDL cholesterol in patients with type 2 diabetes. *Diabetes care* **2015,** *38* (10), 1858-67.

(104) Konopka, A. R.; Esponda, R. R.; Robinson, M. M.; Johnson, M. L.; Carter, R. E.; Schiavon, M.; Cobelli, C.; Wondisford, F. E.; Lanza, I. R.; Nair, K. S., Hyperglucagonemia Mitigates the Effect of Metformin on Glucose Production in Prediabetes. *Cell reports* **2016,** *15* (7), 1394-1400.

(105) Forslund, K.; Hildebrand, F.; Nielsen, T.; Falony, G.; Le Chatelier, E.; Sunagawa, S.; Prifti, E.; Vieira-Silva, S.; Gudmundsdottir, V.; Pedersen, H. K.; Arumugam, M.; Kristiansen, K.; Voigt, A. Y.; Vestergaard, H.; Hercog, R.; Costea, P. I.; Kultima, J. R.; Li, J.; Jorgensen, T.; Levenez, F.; Dore, J.; Meta, H. I. T. c.; Nielsen, H. B.; Brunak, S.; Raes, J.; Hansen, T.; Wang, J.; Ehrlich, S. D.; Bork, P.; Pedersen, O., Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **2015,** *528* (7581), 262-266.

(106) Fullerton, M. D.; Galic, S.; Marcinko, K.; Sikkema, S.; Pulinilkunnil, T.; Chen, Z. P.; O'Neill, H. M.; Ford, R. J.; Palanivel, R.; O'Brien, M.; Hardie, D. G.; Macaulay, S. L.; Schertzer, J. D.; Dyck, J. R.; van Denderen, B. J.; Kemp, B. E.; Steinberg, G. R., Single

phosphorylation sites in Acc1 and Acc2 regulate lipid homeostasis and the insulin-sensitizing effects of metformin. *Nature medicine* **2013,** *19* (12), 1649-54.
(107) Coughlin, S. S.; Calle, E. E.; Teras, L. R.; Petrelli, J.; Thun, M. J., Diabetes Mellitus as a Predictor of Cancer Mortality in a Large Cohort of US Adults. *American Journal of Epidemiology* **2004,** *159* (12), 1160-1167.
(108) Lee, M.-Y.; Lin, K.-D.; Hsiao, P.-J.; Shin, S.-J., The association of diabetes mellitus with liver, colon, lung, and prostate cancer is independent of hypertension, hyperlipidemia, and gout in Taiwanese patients. *Metabolism - Clinical and Experimental* **2012,** *61* (2), 242-249.
(109) Yang, Y.-X.; Hennessy, S.; Lewis, J. D., Type 2 Diabetes Mellitus and the Risk of Colorectal Cancer. *Clinical Gastroenterology and Hepatology* **2005,** *3* (6), 587-594.
(110) Friberg, E.; Orsini, N.; Mantzoros, C. S.; Wolk, A., Diabetes mellitus and risk of endometrial cancer: a meta-analysis. *Diabetologia* **2007,** *50* (7), 1365-74.
(111) Romero, I. L.; McCormick, A.; McEwen, K. A.; Park, S.; Karrison, T.; Yamada, S. D.; Pannain, S.; Lengyel, E., Relationship of type II diabetes and metformin use to ovarian cancer progression, survival, and chemosensitivity. *Obstet Gynecol* **2012,** *119* (1), 61-7.
(112) Kumar, S.; Meuter, A.; Thapa, P.; Langstraat, C.; Giri, S.; Chien, J.; Rattan, R.; Cliby, W.; Shridhar, V., Metformin intake is associated with better survival in ovarian cancer: a case-control study. *Cancer* **2013,** *119* (3), 555-62.
(113) Madiraju, A. K.; Erion, D. M.; Rahimi, Y.; Zhang, X. M.; Braddock, D. T.; Albright, R. A.; Prigaro, B. J.; Wood, J. L.; Bhanot, S.; MacDonald, M. J.; Jurczak, M. J.; Camporez, J. P.; Lee, H. Y.; Cline, G. W.; Samuel, V. T.; Kibbey, R. G.; Shulman, G. I., Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. *Nature* **2014,** *510* (7506), 542-6.
(114) Liu, X.; Romero, I. L.; Litchfield, L. M.; Lengyel, E.; Locasale, J. W., Metformin Targets Central Carbon Metabolism and Reveals Mitochondrial Requirements in Human Cancers. *Cell metabolism* **2016,** *24* (5), 728-739.
(115) Janzer, A.; German, N. J.; Gonzalez-Herrera, K. N.; Asara, J. M.; Haigis, M. C.; Struhl, K., Metformin and phenformin deplete tricarboxylic acid cycle and glycolytic intermediates during cell transformation and NTPs in cancer stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (29), 10574-9.
(116) Walcott, F. L.; Annunziata, C. M.; Sotomayor, E. M.; Fojo, A. T., Effect of metformin chemoprevention on metabolomics profiles in Li-Fraumeni Syndrome (LFS). *Journal of Clinical Oncology* **2017,** *35* (15_suppl), 1556-1556.
(117) Newgard, C. B.; An, J.; Bain, J. R.; Muehlbauer, M. J.; Stevens, R. D.; Lien, L. F.; Haqq, A. M.; Shah, S. H.; Arlotto, M.; Slentz, C. A.; Rochon, J.; Gallup, D.; Ilkayeva, O.; Wenner, B. R.; Yancy, W. S., Jr.; Eisenson, H.; Musante, G.; Surwit, R. S.; Millington, D. S.; Butler, M. D.; Svetkey, L. P., A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell metabolism* **2009,** *9* (4), 311-26.
(118) Chen, T.; Ni, Y.; Ma, X.; Bao, Y.; Liu, J.; Huang, F.; Hu, C.; Xie, G.; Zhao, A.; Jia, W.; Jia, W., Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. *Scientific reports* **2016,** *6*, 20594.
(119) Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C.; O'Donnell, C. J.; Carr, S. A.; Mootha,

V. K.; Florez, J. C.; Souza, A.; Melander, O.; Clish, C. B.; Gerszten, R. E., Metabolite profiles and the risk of developing diabetes. *Nature medicine* **2011,** *17* (4), 448-53.
(120) Laferrere, B.; Reilly, D.; Arias, S.; Swerdlow, N.; Gorroochurn, P.; Bawa, B.; Bose, M.; Teixeira, J.; Stevens, R. D.; Wenner, B. R.; Bain, J. R.; Muehlbauer, M. J.; Haqq, A.; Lien, L.; Shah, S. H.; Svetkey, L. P.; Newgard, C. B., Differential metabolic impact of gastric bypass surgery versus dietary intervention in obese diabetic subjects despite identical weight loss. *Science translational medicine* **2011,** *3* (80), 80re2.
(121) Magkos, F.; Bradley, D.; Schweitzer, G. G.; Finck, B. N.; Eagon, J. C.; Ilkayeva, O.; Newgard, C. B.; Klein, S., Effect of Roux-en-Y gastric bypass and laparoscopic adjustable gastric banding on branched-chain amino acid metabolism. *Diabetes* **2013,** *62* (8), 2757-61.
(122) Walford, G. A.; Davis, J.; Warner, A. S.; Ackerman, R. J.; Billings, L. K.; Chamarthi, B.; Fanelli, R. R.; Hernandez, A. M.; Huang, C.; Khan, S. Q.; Littleton, K. R.; Lo, J.; McCarthy, R. M.; Rhee, E. P.; Deik, A.; Stolerman, E.; Taylor, A.; Hudson, M. S.; Wang, T. J.; Altshuler, D.; Grant, R. W.; Clish, C. B.; Gerszten, R. E.; Florez, J. C., Branched chain and aromatic amino acids change acutely following two medical therapies for type 2 diabetes mellitus. *Metabolism: clinical and experimental* **2013,** *62* (12), 1772-8.
(123) Irving, B. A.; Carter, R. E.; Soop, M.; Weymiller, A.; Syed, H.; Karakelides, H.; Bhagra, S.; Short, K. R.; Tatpati, L.; Barazzoni, R.; Nair, K. S., Effect of insulin sensitizer therapy on amino acids and their metabolites. *Metabolism: clinical and experimental* **2015,** *64* (6), 720-8.
(124) Adam, J.; Brandmaier, S.; Leonhardt, J.; Scheerer, M. F.; Mohney, R. P.; Xu, T.; Bi, J.; Rotter, M.; Troll, M.; Chi, S.; Heier, M.; Herder, C.; Rathmann, W.; Giani, G.; Adamski, J.; Illig, T.; Strauch, K.; Li, Y.; Gieger, C.; Peters, A.; Suhre, K.; Ankerst, D.; Meitinger, T.; Hrabe de Angelis, M.; Roden, M.; Neschen, S.; Kastenmuller, G.; Wang-Sattler, R., Metformin Effect on Nontargeted Metabolite Profiles in Patients With Type 2 Diabetes and in Multiple Murine Tissues. *Diabetes* **2016,** *65* (12), 3776-3785.
(125) Rhee, E. P.; Cheng, S.; Larson, M. G.; Walford, G. A.; Lewis, G. D.; McCabe, E.; Yang, E.; Farrell, L.; Fox, C. S.; O'Donnell, C. J.; Carr, S. A.; Vasan, R. S.; Florez, J. C.; Clish, C. B.; Wang, T. J.; Gerszten, R. E., Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *The Journal of clinical investigation* **2011,** *121* (4), 1402-11.
(126) Wang, T. J.; Ngo, D.; Psychogios, N.; Dejam, A.; Larson, M. G.; Vasan, R. S.; Ghorbani, A.; O'Sullivan, J.; Cheng, S.; Rhee, E. P.; Sinha, S.; McCabe, E.; Fox, C. S.; O'Donnell, C. J.; Ho, J. E.; Florez, J. C.; Magnusson, M.; Pierce, K. A.; Souza, A. L.; Yu, Y.; Carter, C.; Light, P. E.; Melander, O.; Clish, C. B.; Gerszten, R. E., 2-Aminoadipic acid is a biomarker for diabetes risk. *The Journal of clinical investigation* **2013,** *123* (10), 4309-17.
(127) Gall, W. E.; Beebe, K.; Lawton, K. A.; Adam, K. P.; Mitchell, M. W.; Nakhle, P. J.; Ryals, J. A.; Milburn, M. V.; Nannipieri, M.; Camastra, S.; Natali, A.; Ferrannini, E.; Group, R. S., alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PloS one* **2010,** *5* (5), e10883.
(128) Cobb, J.; Eckhart, A.; Motsinger-Reif, A.; Carr, B.; Groop, L.; Ferrannini, E., alpha-Hydroxybutyric Acid Is a Selective Metabolite Biomarker of Impaired Glucose Tolerance. *Diabetes care* **2016,** *39* (6), 988-95.

(129) Udhane, S. S.; Legeza, B.; Marti, N.; Hertig, D.; Diserens, G.; Nuoffer, J. M.; Vermathen, P.; Fluck, C. E., Combined transcriptome and metabolome analyses of metformin effects reveal novel links between metabolic networks in steroidogenic systems. *Scientific reports* **2017,** *7* (1), 8652.
(130) Preiss, D.; Rankin, N.; Welsh, P.; Holman, R. R.; Kangas, A. J.; Soininen, P.; Wurtz, P.; Ala-Korpela, M.; Sattar, N., Effect of metformin therapy on circulating amino acids in a randomized trial: the CAMERA study. *Diabet Med* **2016,** *33* (11), 1569-1574.
(131) Newsholme, P.; Brennan, L.; Bender, K., Amino Acid Metabolism, β-Cell Function, and Diabetes. **2006,** *55* (Supplement 2), S39-S47.
(132) Newsholme, P.; Brennan, L.; Rubi, B.; Maechler, P., New insights into amino acid metabolism, beta-cell function and diabetes. *Clinical science (London, England : 1979)* **2005,** *108* (3), 185-94.