Chemical Isotope Labeling LC-MS for Human Saliva and Urine Metabolomics

by

Tran Thi Ngoc Tran

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

Master of Science

Department of Chemistry

University of Alberta

Abstract

The objective of this work was to employ differential $^{13}C/^{12}C$ dansyl chloride (DnsCl) derivatization strategies for qualitative and quantitative profiling of amine- and phenol-containing metabolites by liquid chromatography electrospray ionization Fourier Transform ion cyclotron resonance mass spectrometry (LC-ESI FT-ICR-MS). A method based on the use of this isotope labeling LC-MS platform was optimized for human salivary metabolome analysis, aiming for the discovery of metabolite biomarkers in mild cognitive impairment (MCI) and Alzheimer's diseases (AD). In addition, a large scale metabolomics study was performed on 100 urine samples taken during a 3 continuous day period. All samples were analyzed using the isotope labelling liquid chromatography mass spectrometry methodology. A pragmatic approach combining well-established statistical tools was developed to process this large data set in order to generate metabolic profiles of those samples.

Acknowledgement

I am especially indebted to my supervisor, Dr. Liang Li, for his guidance, support, and encouragement throughout my time of research and learning under his direction. I would also like to thank my supervisory committee members Dr. Michael Serpe, Dr. Derrick Clive, Dr. John Vederas and Dr. Roger Dixon for their time on reviewing my thesis.

Dr. Randy Whittal and the entire staff of the University of Alberta's Mass Spectrometry Laboratory have provided me valuable advice during my research. I am very grateful to them for their prompt and patient assistance. I would like to thank Tao Huan who has collaborated with me in the work presented in Chapter 3. The members of the Li group have contributed greatly to my growth and experience here at the University of Alberta. I especially thank all the past and current group members whom I had the pleasure to learn from and work with: Dr. Ruokun Zhou, Jiamin Zheng, Chiao-Li Tseng, Jared Curle, Yiman Wu, Wei Han and Zhendong Li.

Last but not least, I would like to express my gratitude to all my friends and family. In particular, I would like to thank my parents and my brother for their love, support and faith in me throughout the years.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-dimensional
Aβ	Beta-amyloid
ACN	Acetonitrile
AD	Alzheimer's disease
APCI	Atmospheric pressure chemical ionization
API	Atmospheric pressure ionization
APPI	Atmospheric pressure photo ionization
AUC	Area under the curve
B	Homogeneous magnetic field
CE	Capillary electrophoresis
CE	Collision energy
CI	Chemical ionization
CID	Collision-induced dissociation
CRM	Charge residue model
CSF	Cerebrospinal fluid
CV	Coefficient of variation
Da	Dalton
dc	Direct current
DIL	Differential isotope labeling
DmPA	p-dimethylaminophenacyl

DnsCl	Dansyl chloride
ESI	Electrospray ionization
ESI-MS	Electrospray ionization-Mass spectrometry
FAB	Fast atom bombardment
FFT	Fast Fourier transformation
FT-ICR-MS	Fourier Transform ion cyclotron resonance mass spectrometry
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
HILIC	Hydrophilic interaction chromatography
HMDB	Human Metabolome Database
HPLC	High performance liquid chromatography
IEM	Ion evaporation model
IS	Internal standard
iTRAQ	Isobaric tag for relative and absolute quantitation
LC	Liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LC-ESI-MS/MS	Liquid chromatography electrospray ionization tandem mass spectrometry
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MCI	Mild cognitive impairment
MTBSTFA	*N-tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide
MRM	Multiple reaction monitoring

MS	Mass Spectrometry
MSn	Tandem mass spectrometry
MS/MS	Tandem mass spectrometry
MVA	Multivariate data analysis
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal projections to latent structures discriminant analysis
PCA	Principal component analysis
PLS	partial least square projection to latent structures
Q-TOF-MS	Quadrupole-time of flight-mass spectrometry
rf	Radio frequency
ROC	Receiver operating characteristic
RP	Reversed phase
RPLC	Reversed phase liquid chromatography
SFC	Supercritical fluid chromatography
SFE	Supercritical fluid extraction
SIL	Stable isotope labeled
SPE	Solid phase extraction
TEM	Temperature
TLC	Thin layer chromatography
TOF	Time of flight
UV	Ultraviolet
VIP	Variable influence on projection

Chapter 1: Introduction

1.1 General Introduction

Metabolomics has emerged as an important discipline in science research, capable of detecting and quantifying hundreds or thousands of metabolites in biological and biochemical systems.[1,2] Together with genomics, transcriptomics and proteomics, metabolomics plays an important role in providing a more comprehensive understanding of the physiological state of living organisms. However due to the underlying complexity of the metabolome, metabolomics requires careful approaches in sample selection, preparation and separation and data analysis.[3]

Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy was used predominantly for metabolomic analysis in the early days.[4,5] Recently, separation tools such as capillary electrophoresis (CE), supercritical fluid chromatography (SFC), gas chromatography (GC) or liquid chromatography (LC) coupled to mass spectrometry (MS) have been more widely used.[6] In particular, LC-MS has become the workhorse of metabolomic investigations. Typically, LC-MS provides metabolite separation on a reversed phase (RP) column followed by electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI).[7] Metabolites are detected in both positive and negative modes, thereby allowing a more extensive metabolome coverage.

However, due to the diverse physiochemical properties of metabolites in biological samples, it remains extremely challenging to elucidate all the metabolites using LC-MS based analyses. Our research group has employed the method of differential isotope labelling to target a group of phenol and amine-containing metabolites. This derivatization strategy allows for relative quantification and comprehensive profiling of metabolites, while achieving separation of

polar and ionic metabolites and improving ESI sensitivity. The goal of this thesis is to develop and apply the $^{12}C/^{13}C$-dansylation labelling technique in human salivary and urine samples for profiling endogeneous phenol and amine-containing submetabolome.

1.1.1 Profiling Metabolomics

Metabolite profiling studies have gained tremendous attention in the past decade as a tool to investigate changes in phenotypes. Metabolomics is the ending stage of the "Omics" cascade (Figure 1-1). It can provide a good biological understanding when integrated with other "Omics" database as metabolome is considered as being the most predictive of phenotype.[8]

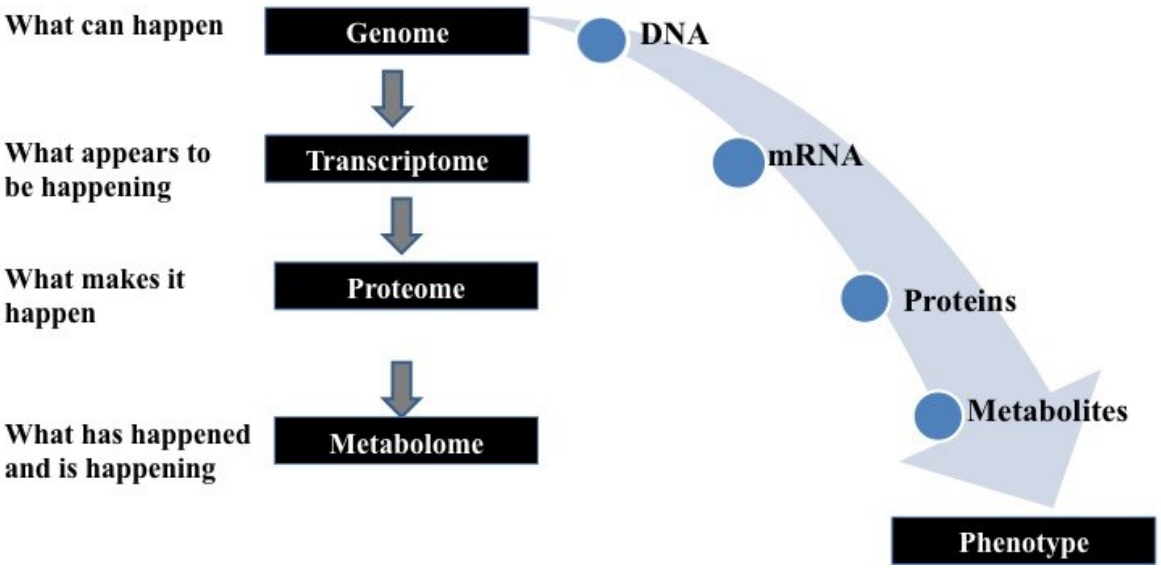


Figure 1-1 The "Omics" cascade (adapted from Dettmer et al., 2007[8])

The most challenging aspect of metabolomics is the chemical complexity and heterogeneity of metabolome, which spans over a diversity of compound classes such as organic acids, alcohols, amines, lipids, nucleotides, etc. For a comprehensive analysis of the metabolome, it is crucial to utilize approaches that provide the widest coverage of compounds as possible. A

combination of different analytical techniques has been popularized in the past decade. NMR-based investigations that are based on the chemical shift signals are extremely useful in detecting and identifying high abundance metabolites with very minimal sample preparation. However due to the low sensitivity of NMR technique, mass spectrometry-based metabolomics has emerged as a more widely applied technology, as it has the advantages of providing a higher sensitivity and selectivity and compound identification capability.[9-11]

For a typical MS-based experiment, careful attention need to be paid to all the steps in method development which include sample preparation, sample analysis, data acquisition, and statistical analysis. The sample preparation is usually carried out by extracting the metabolites from biological samples, such as urine, serum, whole blood, saliva, and cell pellets. To inhibit enzymatic processes, freezing in liquid nitrogen or freeze clamping after sampling and consequent freezing in -80˚C for storage are typically employed.[12] Since any sample preparation could introduce a loss in metabolites, an extraction procedure can be used to include a sample concentration step to help detect the low abundance metabolites. Some examples of sample preparation techniques are solid phase extraction (SPE),[13] liquid-liquid extraction (LLE),[14] supercritical fluid extraction (SFE),[15] protein precipitation,[8] etc. Regardless of whatever methods are used for the experiments, the sample preparation steps need to be streamlined, simple and universal and can be tailored for a large group of compounds for targeting metabolic profiling.

The combination of gas chromatography and mass spectrometry provides a powerful platform to separate, detect and identify non-volatile, high molecular weight metabolites. However, polar metabolites will require chemical derivatization to increase thermal stability and volatility while decreasing the polarity. Another popular system, LC-MS, is mostly employed in reversed phase (RP) or hydrophilic interaction chromatography (HILIC). RP can separate

medium to low polar compounds while HILIC can analyze vey polar compounds.[16,17] RP-HPC separation using a typical C18 column with particle sizes of 3-5 μm often resulted in poor resolution for some metabolites in complex biological samples. Sub-2 μm particles are used in ultra performance liquid chromatography (UPLC) to improve peak capacity and resolution and increase limit of detection.[18] After separation by chromatography, metabolites are detected by electrospray ionization (ESI) typically in both positive and negative modes to obtain a more comprehensive coverage. Time of flight (TOF) mass spectrometer is chosen for their high mass accuracy and acceptable resolution. However in complex samples, excellent mass accuracy and resolution are important to help finding the unknowns; Fourier transform ion cyclotron resonance (FT-ICR) proves useful in this application. FT-ICR has high mass accuracy (<1ppm) and high mass resolution (>100,000).[19] Even though FT-ICR provides higher-throughput analyses and lowers the detection limit, it has the disadvantage of requiring longer time to obtain spectra for the peaks to be properly resolved. Regardless of which instruments to be employed, the key issue here is how to analyze the exorbitant amount of data generated in a scientific and sound way.

Raw data of mass spectra series are usually extracted by vendor or in-house software. Automatic peak picking, noise filtering, and baseline correcting are some of the steps that are performed to convert the data into useful format for biological and statistical analyses. Results are returned as features (molecular entities with corresponding m/z and retention time) and signal abundances. Depending on the objectives of the studies, metabolomics data are often subjected to statistical analysis. Unsupervised or supervised classification may be used to generate sample groupings. Often the aim of the study is to discover novel biomarkers for different diseases. As a standard practice, metabolites are identified by searching against metabolite database using m/z values. Putative search results are confirmed by matching their corresponding retention times

with authentic standards. One of the major bottlenecks in profiling metabolomics is the identification of potential biomakers. Despite the increasing number of public database such as METLIN, HMDB, KEGG, MycompoundID, etc., it is extremely difficult to integrate and compare data collected on different analytical instruments. Furthermore, biochemical pathway analysis and visualization are still underdeveloped, posing a hurdle in understanding multidimentional data.

Despite its limitations, metabolomics profiling has evolved and will continue to grow across all life sciences disciplines, from biomarker discovery to functional genomics. As more and more research is being carried out in this field, it is crucial that data standardization, data analysis and visualization tools, publicly available and well-curated databases need to be better developed in order to transform metabolomic data into useful knowledge.

1.1.2 Chemical Derivatization and Dansylation Labeling

Even though LC/MS is very powerful in elucidating a variety of trace level compounds in living systems, one of its limitations is its poor separation and detection of highly polar metabolites. A compound that is favorably analyzed by LC/ESI-MS should be chargeable and also possess a hydrophobic region. Therefore the analyte is often subjected to chemical derivatization to improve the sensitivity in ESI-MS.[20-22]

Our research group has illustrated a chemical derivatization strategy by labeling metabolites with reagent dansyl chloride in a series of publications.[23-25] Dansyl chloride is a well-studied and well-characterized compound, which has been used for the quantification of phenolic hydroxyls, biogenic amines and amino acid by thin layer chromatography (TCL) and HPLC separation and detection by fluorescence or UV detection.[26-31] Past studies have shown the use of dansylation in chemically derivatizing a group of functional groups that include

fenfluramine and phentermine,[32] ß-estradiol and estrone,[33] and amines and phenols.[20] Dansylation increased the overall ion intensity by introducing a hydrophobic naphathalene moiety to the analyte, thereby allowing the polar compounds to be eluted in a higher percentage of organic mobile phase during a RPLC gradient. At a higher organic solvent composition, both the ionization desolvation process and electrospray stability are enhanced greatly due to decreased surface tension in the droplet. Also, significant improvement in ESI surface activity was observed due to the fact that the naphthalene moiety increased the droplet surface affinity of the analyte. The analyte that has a more hydrophobic and less polar group will exhibit a higher electrospay response.[34] The presence of a more easily protonated dimethylamino moiety improved the chargeability of the analyte, making it easier to compete for the limited amount of charges on the droplet surfaces.[23] Through dansylation labeling, signal-to-noise ratio is also improved due to the shift of mass-to-charge ratio (m/z) from low mass to high mass region in which there will be less background noise from solvent clusters and common contaminants.[35] Finally, the stability of metabolites analyzed by LC/MS increases tremendously. In source fragmentation is less likely to occur and the peaks in a typical mass spectrum can now be more confidently assigned to the metabolite ions. Due to its selective signal enhancement for phenols and amine, dansyl chloride aids the detection of phenol and amine-contaning metabolies, by RPLC-MS; therefore, a more comprehensive sub-metabolome coverage will be achieved.

1.1.3 Differential Stable Isotope Labeling for Metabolomics

Another challenge in ESI technique is the susceptibility to ion suppression from matrix molecules. To address the ion suppression effects[36] that create a different analyte response due to the changes in sample matrix, isotope dilution technique has been employed in a large variety of

applications. The technique uses a stable isotope tag attached to the compound of interest to form a physical-chemical similar analyte as an internal standard (IS) in MS analysis, thereby compensating for the ion supression effects.[37-39]

Stable isotopic labeling (SIL) has been widely used in metabolomics especially in GC-MS and LC-MS. Huang et al. labeled reactive functional groups with isotope-coded *N-tert*-Butyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA) as a mean to relatively quantify metabolites in complex biological samples using GC-MS.[40] Wu et al. reported a successful and robust method for quantitative analysis of the intracellular metabolite concentrations by using a mixture of fully uniformly ^{13}C-labeled metabolites as an IS in LC-ESI-MS/MS analysis.[41]

For metabolomics analysis, differential isotopic labeling has great potentials in quantifying targeted metabolites with higher accuracy and precision. The general workflow for SIL in metabolomics is that the sample is separated into two identical aliquots and is labeled with light or heavy isotopic tags respectively by chemical reaction, followed by mixing for mass spectrometric analysis. The peak ratios of the isotope labeled ion pair allows for relative quantification of the metabolite in comparative samples. Furthermore, if one sample is a standard compound of known concentration, the absolute concentration of the other sample can be quantified.

1.1.4 Statistical Analysis in Metabolomics

Metabolomics studies usually consist of hundreds or thousands of samples, creating a complexity in data analysis. Multivariate data analysis (MVA) has grown to assist with analysis and interpretation of complex data structures as it proves to be useful in visualizing and interpreting the difference that exist between many samples in metabolomics study.[42,43]

Principal component analysis (PCA) is the workhorse in MVA. It is an unsupervised method designed to extract and display the systematic variation in a data matric X by a reduced set of latent variables.[44]

The PCA model can be expressed as

$$X=TP^T+E \qquad \text{(Eq. 1.1)}$$

A model plane is defined by two principal components. Each of the samples is projected onto the plane, thus allowing a more comprehensive visualization of all samples. The coordinates of each sample are called scores T. Therefore, the visualization of all the scores are called score plot, which provides an overview of all samples in X and how they relate to each other. Interpretation of the grouping or interesting patterns is revealed in the corresponding loading (P) plot. Unsystematic information is stored in residual matrix E and important for detection of outliers. PCA has an advantage that the orthogonality between components is a rigid structure, which is useful in extracting the information from the samples but does not identify natural phenomena in the components.

When a sample–specific information (e.g., gender, dietary habits, age groups) is known, partial least square projection to latent structures (PLS) is employed. PLS is a supervised method where it extracts latent variables by maximizing the variation in X and the covariation between the explanatory variables X and the response Y.[45] Similarly to PCA, score vectors and loading vectors are also formed. Novel extension of the PLS method, orthogonal projections to latent structures (OPLS) can be separated into predictive variation T_pP_pT and orthogonal variation T_oP_oT.

$$X = T_pP_p^T + T_oP_o^T + E \qquad \text{(Eq. 1.2)}$$

$$Y = T_p C_p^{T} + F$$

When Y is qualitative, the OPLS is called OPLS discriminant analysis (OPLS-DA).[46] The source of orthogonal variation is often correlated to the study conditions such as gender, age and environmental factors. A variable influence on projection (VIP) analysis is often performed to identify the metabolites that discriminant the separation between two groups.[47] VIP plot is a type of coefficient plot that gives the summary of relationship between X and Y variables. The higher the VIP score, the more significant the variable (typically a metabolite) is in comparing the difference between the two groups. As the number of variables is significantly high in metabolomics, methods such as cross validation or permutation tests are often performed to reduce the probability of finding the correlations by chance.[48]

The application of statistical tools in metabolomics has been gaining popularity over the years. Choice of regression and classification methods becomes a question of robustness, adaptability and the objective of the studies. It is evident that metabolomics can greatly benefit from the development of new and improved analytical tools when it comes to what to utilize and validate on raw untouched-data in order to turn them into useful working knowledge.

1.2 Mass spectrometry

Mass spectrometry has become extremely important and versatile in several branches of natural sciences, e.g., forensics, archaeology and environmental sciences just to name a few. Mass spectrometry is an analytical technique that measures the mass-to-charge ratios of ions to identify and quantify molecules in simple and complex mixtures. The foundation of mass spectrometry was laid by Wilhelm Wien in 1898, who demonstrated that beams of charged particles could be deflected by passing them through superimposed parallel electric and magnetic

field. However the birth of MS is more commonly credited to Sir J.J. Thomson, who discovered the electron and its deflection by an electric field by an electric field.[49] He later constructed the parabola mass spectrograph for the determination of m/z of ions, which is considered as the great grand-father of modern mass spectrometers.[50]

In the 1940s, the potential of mass spectrometry as an analytical tool was further realized and commercial mass spectrometers became to be introduced to the market, with its first application to monitor petroleum refinement processes.[51] By the 1960s, the field of structural analysis revolutionized with the development of tandem mass spectrometry and collision induced decompositions.[52] Mass spectrometry has become a standard analytical tool in the analysis of organic compounds. However, its application to the biosciences was limited because of the lack of suitable methods to ionize fragile and non-volatile compounds. The development of soft ionization methods catapulted the range of MS applications during the 1980s. These include fast atom bombardment (FAB),[53] electrospray ionization (ESI),[54] matrix-assisted laser desorption/ionization (MALDI),[55] atmospheric pressure chemical ionization (APCI)[56] and atmospheric pressure photo ionization (APPI).[57]

Especially with the development of ESI, the use of mass spectrometry in life sciences has exponentially evolved. ESI provides a medium for coupling of solution introduction of compounds for analysis and the capability for ionization of involatile and highly polar compounds. As ESI is the ionization of choice in our work, it will be discussed in detail in the next section.

1.2.1 Electrospray Ionization

A solution containing analytes is introduced to a capillary needle which is held at high positive or negative voltage (2-4kV). The high electric field generates a mist of highly charged droplets, which travel toward the mass analyzer portion of the mass spectrometer. The droplet formation and ionization is still under debate.[58,59] The solution delivered to the tip of the capillary experienced the electric field at a very high potential. As all experiments carried out in this thesis were done in positive mode, positive mode ionization will be the focus of discussion here. Positive ions accumulate at the surface while anions migrate away from the tip, which establishes a "Taylor cone". At high enough field the cone is drawn to a filament, becomes unstable at the tip and starts to produce the positively charged droplets into a fine jet. Solvent from the initially formed charged droplets evaporate as they travel towards the analyser of the mass spectrometer, resulting in a reduction in diameters of the droplets. Fission will occur at the "Rayleigh limit" which is the point where magnitude of the charge is high enough to overcome the surface tension holding the droplets together. Continuous reduction of the droplet size by solvent evaporation and fission will lead to the formation of droplets containing a single ion. That mechanism is called the charge residue model (CRM),[60,61] which is commonly believed to be associated with larger molecular ions (>1000 Da). Another mechanism, ion evaporation model (IEM), has been proposed based on the ideas of Iribane and Thompson.[62,63] Ion evaporation takes place from small highly charged droplets because of the coulombic repulsion between the charged ion and other charges of the droplet. There are three important factors in this process. First of all is the geometric parameter which means ion evaporation becomes significant when the droplet diameter reaches about 20 nm. Second factor is related to the chemical properties of the ion; the ion evaporation rate constant is dependant on the reaction free enthalpy difference when the ion is expelled from the droplet. Lastly, the surface charge density

of the droplet is below the highest possible surface charge density at the Rayleigh limit. IEM explains how smaller ions can be expelled from droplets via evaporation in a solvated state.[64]

Both CRM and IEM are just models to describe experimental data in a simplified way and it remains a challenge to fit all experimental findings within a single model. Regardless, ions released by electrospray process are very stable and not in an excited state. Combined with the coupling capability with other chromatographic separation techniques and its high ionization efficiency, ESI remains one of the most popular ionization techniques in chemical and biochemical analysis.

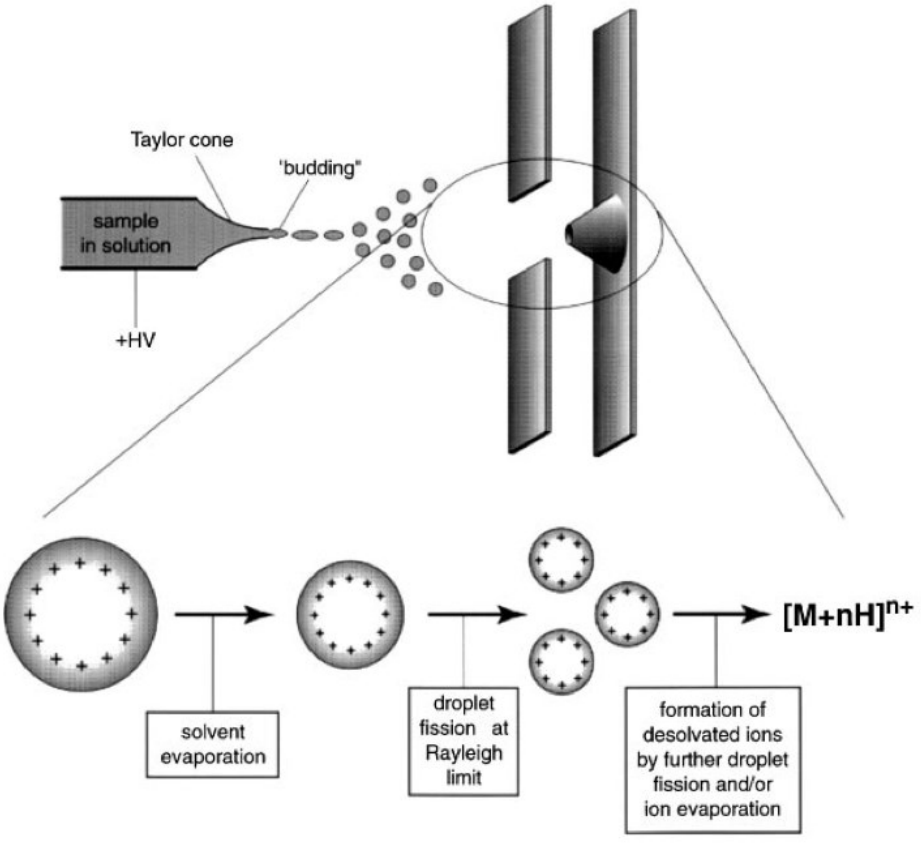


Figure 1-2 Droplet production in the electrospray interface (adapted from Gaskell, 1997[65])

1.2.2 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Of the diverse variety of mass spectrometers, the Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer has become the forefront of research because of its ability to achieve high resolution and mass accuracy. FT-ICR was invented by Melvin Comisarow and Alan Marshall at the university of British Columbia in 1974.[66,67] As the analyte is introduced in solid, liquid or gas forms depending on the type of ionization technique, gas-phase ions are generated and expelled into the mass analyzer portion of the mass spectrometer and are detected. In FT-ICR mass spectrometer, generated ions pass through a series of pumping stages at high vacuum, and then are injected into an "ICR cell". The ICR cell is situated within a strong magnetic field generated by a superconducting magnet. The pressure inside the cell reaches 10^{-10} to 10^{-11} mBar with temperature close to absolute zero. Ions moving in the presence of a magnetic field will experience a force known as the "Lorentz force",[68] given by Eq 1.3

$$F = qvB\sin\theta \qquad \text{(Eq. 1.3)}$$

where q and v are the charge (Coulombs, C) and velocity (m s^{-1}) of the charged particle, B is the magnetic field strength (Telsa, T) and θ is the angle between the axis of the charged particle's moving direction and the axis of the magnetic field. If those axes are perpendicular to each other the equation becomes

$$F = qvB \qquad \text{(Eq. 1.4)}$$

The charged particle then revolves about the center of the magnetic axis, resulting in a "cyclotron motion". The m/z of the ion can be related to the cyclotron frequency as shown in equations 1.4 and 1.5:

$$\omega = \frac{qB}{m} \qquad \text{(E.q. 1.5)}$$

where ω is cyclotron frequency (rad s^{-1}), m is the mass of the ion (kg).

$$q = ze \qquad \text{(E.q. 1.6)}$$

where z is the net number of electronic charges on the ion and e is the charge per electron (Coulombs, C)

The FT-ICR cell consists of the two excitation and two detection plates. Two trapping plates are included to constrict the ion's motion along the axis of the magnetic field by being applied a low potential between them (typically 1V).

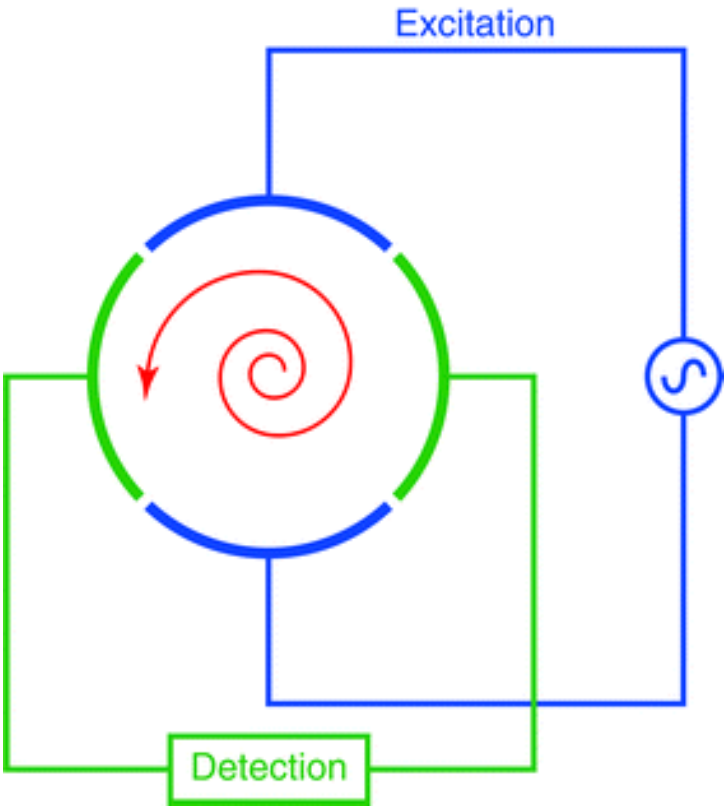


Figure 1-3 A cross section of a cylindrical FT-ICR cell (adapted from Barrow et al., 2005[69]).

A simplified way that the FT-ICR can be summarized is in four steps. The first step that occurs in an FT-ICR-MS experiment is quenching: a large negative voltage is applied to the trapping plates to remove all remaining ions from previous experiments. Secondly, an electron beam or a laser beam is used to ionize the molecules inside the cell. Thirdly, a fast rf sweep of a few milliseconds is applied across the cells to excite the ions of a given m/z as a coherent package to larger cyclotron orbits. That step is crucial in order to produce a measurable signal. Finally, decay of the cyclotron motions induces image current that can be amplified, digitized and stored in memory. Deconvolution of this signal by fast Fourier transformation (FFT) methods results in a mass spectrum in which signal magnitude is plotted as a function of m/z after calibration is carried out.[70]

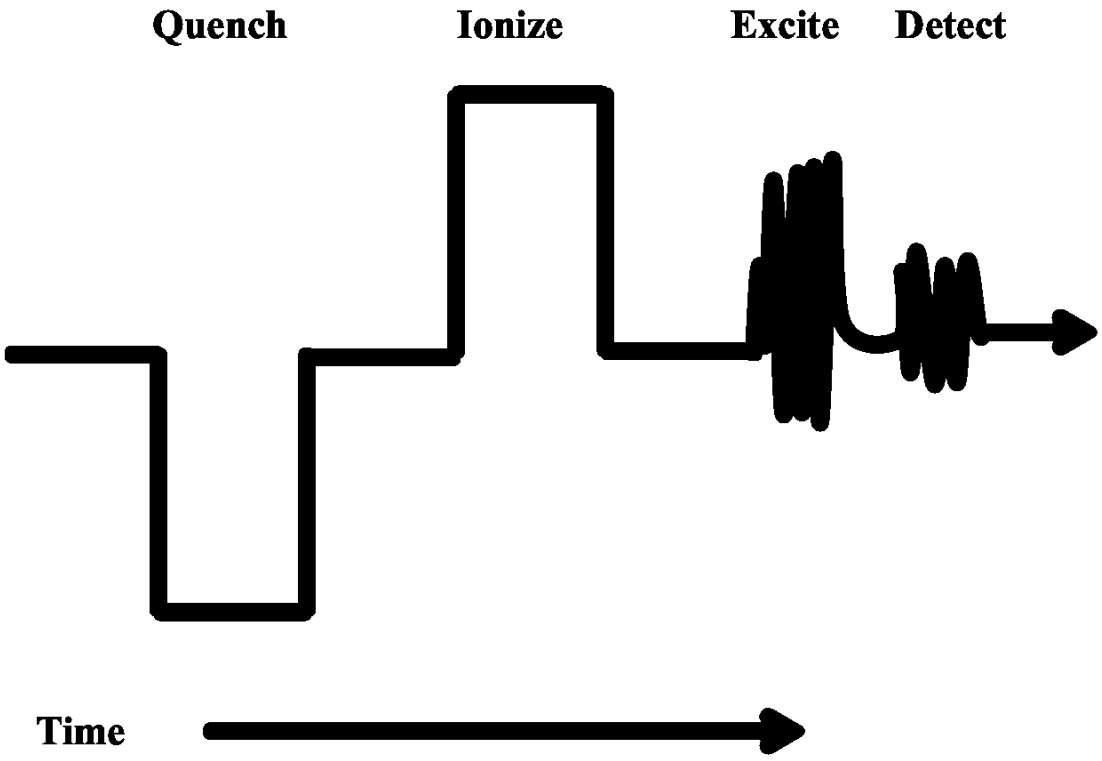


Figure 1-4 A typical experimental sequence events in FT-ICR.

FT-ICR mass spectrometer is popular for its mass accuracy and resolution. Its mass accuracy is typically cited between 2 to 5 ppm. It can reach resolutions of hundreds of thousands in "normal" experimental conditions while a resolution of 10000 is achieved routinely. However, the downsides of FT-ICR are the slow scanning speed, prohibitive cost and large footprint. FT-ICR represents a powerful technique for the study of biological molecules, in particular within proteomics and metabolomics. New improvements in FT-ICR cell design and its capability in tandem mass spectrometry experiment have led to even better mass accuracy and resolution, as such structure elucidation and more comprehensive ranges of samples can be analyzed by FT-ICR-MS.

1.2.3 Liquid Chromatography

The bulk of metabolomics applications and studies have focused on LC-MS based methods, especially those using RP separations. The term "reverse phase chromatography" was used because RP is a form of partition chromatography where chemically bonded phase is non-polar, whereas the starting mobile phase is polar. The mobile phase typically consists of water and an organic solvent, such as acetonitrile (ACN) or methanol (MeOH) and a small amount of additives or buffers depending on the analysis. Conventional column takes the following formats, typically 2.1–4.6 mm in internal diameter, 5–25 cm in length and packed with 3–5 μm particles. Silica is the most common particle material. On small molecule studies, silica's pore size of 100 angstroms is typically employed.[71] The stationary phase is generally made up of alkyl chains that interact with the analytes. The C18 chain length is often used to capture small molecules or peptides.

In RPLC compounds are separated based on their hydrophobicity characters. Once the sample is injected, different analytes interact and partition differently with the mobile and stationary phase. Polar molecules in the sample will interact strongly with the solvent and there won't be much attraction between the stationary phase and the polar molecules. Therefore, they will elute earlier. On the other hand, non-polar molecules will form attraction with the hydrocarbon because of Van der Waals dispersion force. They also interact less with the solvent and this will delay their elution through the column.

Gradient elution is the method of choice in reversed phase chromatography as gradient separations can separate complex mixtures better in shorter amount of time. Such gradient separations are excellent for compounds of low to medium polarity, but they remain unsatisfactory for the polar and ionic compounds. Most importantly, despite the powerful separation capability of LC method and the sensitive detection in MS, a large unknown number of analytes remain undetected. This can be attributed to the complex and varying polarities of the analytes, low concentration, poor ionization and ion suppression issue. Ion suppression is a phenomenon resulting from the coelution of matrix components that affect the detection capability of the analyte of interest. Precision and accuracy of detection as well as the signal intensity also decreased. Therefore, high resolution and reproducible chromatographic separations are important in order to produce reliable high-quality data. Despite those aforementioned limitations, HPLC when coupled with MS, is a very powerful tool for global metabolite profiling. The addition of HILIC or other mixed-mode separation will likely enhance metabolome coverage and doesn't eliminate the need for HPLC. It is crucial that better sample clean up should be ensured and more attention should be paid to the chromatographic separation and calibration and normalization techniques.

1.3 Overview of Thesis

The main objective of this thesis is to develop robust and sensitive LC-MS methods for amine-and phenol-containing submetabolome profiling by using differential isotope labeling technique. Chapter 2 outlines this approach and presents the results of using 82 saliva samples of normal aging controls and MCI and AD patients for metabolite biomarkers discovery. Our objective for this work is to investigate further the role of saliva as a diagnostic tool and to elucidate the disease-related salivary biomarkers in order to help identify and predict the progression of MCI and AD diseases. Chapter 3 is focused on the development of the metabolome database using human urinary samples. This work is part of the large scale collaborative metabolomics study which will examine the metabolome profiles in both China and Canada. Because many factors can affect the metabolome profiles, it is crucial to analyze the metabolome of diverse populations of samples. In this work the urine metabolome of 100 healthy individuals were analyzed. The differences among those samples due to gender, Body Mass Index (BMI), and age were also investigated. The objective of this work is to generate a "baseline" metabolome database that would be eventually employed for metabolite biomarkers discovery for diagnosis and prognosis of diseases using metabolomics. Finally, chapter 4 outlines the conclusion of this thesis and directions for future works.

Chapter 2: Metabolomics Analyses of Salivary Samples from Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease Groups: Toward Clinical Discrimination Using a Noninvasive Biofluid

2.1. Introduction

Mild Cognitive Impairment (MCI) is defined as a clinical state characterized by significant cognitive impairment in the absence of dementia. MCI is known as the transition between normal aging and the prodromal phase of Alzheimer's disease (AD). Early diagnosis of MCI and AD can assist in the management and treatment of the diseases. There is a pressing need for an improved, rapid and sensitive screening method for diagnosing MCI patients as the preclinical period of AD can reveal valuable information to develop interventions that may delay or prevent clinical diagnosis.[72]

To this end, an accumulating literature has focused on discovering and validating biological markers of AD that may provide both early signals that preclinical neurobiological changes are underway.[73] All of these approaches have had considerable success in detecting independent or combined effects of biomarkers associated with MCI or AD status,[74,75] but much remains to be done in this rapidly advancing field. In this work, we adopt the non-targeted large-scale metabolomics approach, which is similar to other OMICS-type methods and techniques (e.g., genomics, proteomics) that consider and sample a whole biological subset.

Metabolomics is a powerful platform to study the metabolic perturbations, which indicates an accurate biochemical phenotype of an organism. Metabolomics study can readily be done with non-invasive biofluid samples, e.g., blood, urine, saliva. In particular, human saliva is an attractive medium because of its inexpensive cost, low evasiveness, and easy sample collection and processing.[76] Saliva, often considered as "the mirror of the body", is secreted from

three pairs of major salivary glands and many salivary glands lying beneath the oral mucosa.[77] Saliva metabolomics has emerged as an attractive approach to screen for potential diagnostic and prognostic biomarkers and to distinguish different states of diseases. Recently, comprehensive salivary metabolic profiles of healthy controls and of patients with oral, breast or pancreatic cancer has been investigated.[78] Despite still being in its infancy, the potential role of saliva metabolomics in early disease diagnoses and prognoses is clear.

Isotope labeling LC-MS approach has been applied to saliva metabolomics to study the effects of MCI on metabolome changes with great success.[25] By employing dansylation chemistry in our isotope labeling technique,[79] the amine- and phenol-containing metabolites can be targeted, thus opening the possibility of profiling and quantifying a large number of putative metabolites in saliva samples.

In this work, we apply the strategy of differential $^{12}C/^{13}C$-dansylation labeling of human salivary samples, in conjunction with Liquid Chromatography-Mass Spectrometry (LC-MS), for comprehensive and quantitative profiling of the salivary metabolome with an objective to discover metabolite biomarkers. Herein we present the results of discovering the discriminant metabolites for MCI and AD from using a set of 82 clinical saliva samples from MCI and AD patients and healthy aging patients.

2.2. Experimental

2.2.1 Subjects

The total of 82 participants who were in the present study included *n* = 35 NA adults (age = 64-75 years; 62.9 % female), *n* = 25 MCI adults classified as MCI (age: 64-75 years; 60 % female), and *n* = 22 AD patients (age = 52-91 years; 72.7 % female). The saliva samples were obtained from Dr. Roger Dixon's laboratory (Department of Psychology, University of Alberta).

Ethics approval of this work was obtained from the University of Alberta according to the university's health research policy. The demographic characteristics and clinical information of the participants are provided in Table 1-1.

Table 2-1 Baseline clinical characteristics

	NA	**MCI**	**AD**
Clinical Characteristics			
N	35	25	22
Age, years	69.94 (3.80)	70.40 (3.38)	77.09 (11.20)
Women (M/F)	13/22	10/15	6/16

2.2.2 Sample Sampling and Storage

Salivary samples were collected using Oragene®•DNA Self-Collection Kit OG-500 DNA Genotek, Inc., Ottawa, Ontario, Canada). Whole saliva was collected according to the manufacturer's instructions and was placed inside the kit which also contained an Oragene DNA-preserving solution. The complete list of ingredients of Oragene solution include but not limited to proprietary reagents, some of which are Ethyl alcohol (<24 %) and Tris-HCl buffer (pH 8)[80,81]. As provided by established procedures, samples were stored at room temperature before isotope labelling experiments and were preserved at -20°C or -80°C after the experiments for long-term storage and follow-up studies.

2.2.3 Chemicals and Reagents

^{13}C-dansyl chloride was synthesized in-house as described by Guo and Li.[79] ^{12}C-dansyl chloride was purchased from Sigma-Aldrich (Milwaukee, WI). All reagents were of ACS grade or higher with water and organic solvents being of MS grade. Metabolite standards were of analytical grade and were used as specified.

2.2.4 Metabolite extraction and isotope labelling

We have upgraded some aspects of the experimental method reported in our initial publication[25]. Specifically, each 5 μL individual saliva sample was directly labelled with ^{12}C-dansyl chloride and the pooled saliva from all 82 samples (NA, MCI and AD samples) with ^{13}C-dansyl chloride. It was noted that proteins have been subjected to rapid degradation and precipitation by the presence of Ethyl Alcohol in Oragene·DNA saliva collection kits. Therefore, subsequent protein precipitation and concentration steps were therefore omitted in order to prevent any progressive loss of samples. An existing protocol previously published [25] was adapted and adjusted for labeling salivary samples for metabolome profiling. An aliquot of 5 μL of saliva sample was dissolved in 20 μL ACN/ H_2O (50/50) in a screw cap vial. 12.5 μL of $NaHCO_3/NaH_2CO_3$ buffer solution (500 mM, 1:1, v/v) was mixed in the solution and the vial was vortexed and spun down. 36.6 μL of freshly prepared ^{12}C-DnsCl or ^{13}C-DnsCl in acetonitrile (12mg/mL) was added into the vial. The solution was vortexed and spun down again then let to react for 60 minutes in an oven at 60˚C. 5 μL of NaOH (250mM) was then added to quench the excess DnsCl. After another 10 minutes incubation in the 60˚C oven, 25 μL of formic acid in ACN/ H_2O (425 mM, 50/50) was added to neutralize the solution.

The experimental workflow for human salivary analysis is summarized in Figure 2-1.

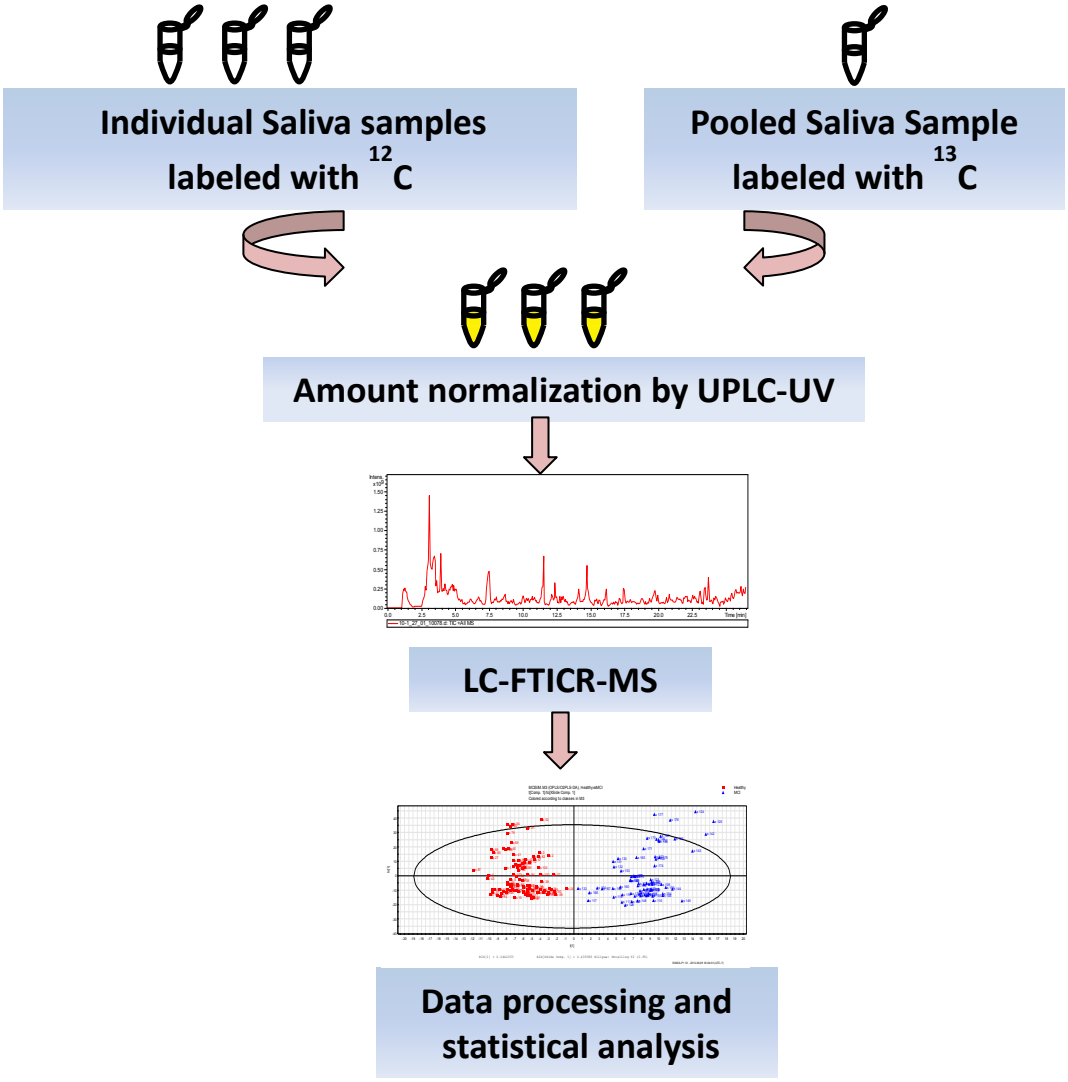


Figure 2-1 Workflow of the high-performance isotope labeling LC–MS method developed for salivary metabolome profiling.

2.2.5 UPLC-UV

After being labelled with $^{12}C/^{13}C$-dansyl chloride, extracted metabolites were quantified by UPLC-UV using an established methodology[24]. An ACQUITY UPLC® system (Waters Corporation, Milford, MA) equipped with photo diode array (PDA) detector, and a Waters ACQUITY UPLCTM BEH (Ethylene Bridged Hybrid) C18 column (2.1 mm × 50 mm, 1.7 μm

particle size) were used for online LC-UV. LC solvent A was 0.1% (v/v) in 5% (v/v) ACN, and solvent B was 0.1% (v/v) formic acid in ACN. The gradient elution profile was as follows: t = 0 min, 0% B; t = 1.00 min, 0% B; t = 1.01 min, 95% B; t = 2.50 min, 95% B; t = 3.00 min, 0% B; t = 6.00 min, 0% B. The flow rate was 450 μL/min, and the sample injection volume was 2 μL. The detection wavelength was set at 338 nm.

2.2.6 LC-FTICR-MS

Metabolomic analyses were performed using an Agilent 1100 series capillary HPLC system (Agilent, Palo Alto, CA, USA) connected to a Bruker 9.4 T Apex-Qe Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA, USA) equipped with an ESI interface operating in positive mode. Reversed phase chromatographic separation was carried out on an Eclipse C18 column (2.1 mm ×100 mm, 1.8 μm), with solvent A being water with 0.1% (v/v) formic acid and 5% acetonitrile (ACN) (v/v), and solvent B being ACN with 0.1% (v/v) formic acid. The flow rate was 180 μL/min and running time was 26.50 min. The gradient was: t = 0 min, 20% B; t = 3.50 min, 35% B; t = 18.00 min, 65% B; t = 21.00 min, 95% B; t = 21.50 min, 95% B; t = 23.00 min, 98% B; t =24.00 min, 98% B; t = 26.50 min, 99% B. The sample injection volume was 6 and the flow was split 1:3 before entering the ESI-MS system.

2.2.7 Data processing and Statistical Analysis

The $^{12}C/^{13}C$ ion pairs of the same metabolites were detected by a sensitive and robust in-house peak peaking program ISOMS. The peak pair data were aligned by retention time and accurate mass and each metabolite feature was converted into a matrix of peak pairs ratio versus metabolite identification. Volcano plot and univariate-fold change analysis were performed for comparisons between three successive sets of two groups (e.g., NA vs. MCI, NA vs. AD, MCI vs.

AD) by Metaboanalyst (www.metaboanalyst.ca)[82,83]. The mean-centered and unit-variance (UV)-scaled data were introduced into SIMCA-P+ 12.0 software (Umetrics, Umea, Sweden) for multivariate analysis to reveal any discrimination of the three study groups (NA, MCI and AD). Receiver Operating Characteristic (ROC) curves of discriminant metabolites were generated by using freely available web-based tool ROCCET (www.roccet.ca)[84].

2.2.8 Metabolite Identification

Resulting significant metabolites were putatively identified against the Human Metabolome Database (HMDB)[85-87], MyCompoundID[88] and METLIN metabolite database using a mass accuracy window of 5 ppm. Isotope patterns of those metabolites were manually inspected to eliminate any possible mismatches. Definite metabolite identification was achieved by matching retention times and accurate mass of observed peaks of metabolites to those of standards.

2.3. Results

2.3.1 Experimental conditions maximization

A set of experiments was carried out to determine the best solvent to dissolve the saliva sample in. A 5µL starting saliva sample was aliquoted out and dissolved in 20 µL acetone (ACE), or 20 µL H_2O, or 20 µL ACN/ H_2O (50/50) at room temperature. Another 5 µL starting saliva sample was dissolved in -20˚C acetone overnight to serve as the control sample. Each solution was then subjected to dansylation chemistry per the above protocol. A Waters ACQUITY UPLC system with binary solvent manager and a photodiode array (PDA) detector was used for the quantification of labeled metabolites by UV absorbance using a fast 6-minutes gradient run [89]. The peak areas of the labeled metabolites were measured and compared among the three solvent system experiments. While the labeling efficiency was consistently comparable among all 3

chosen solvent system experiments and much higher than that of the control, ACN/ H_2O (50/50) was shown to have the best performance as the volume of organic and inorganic solvent were kept equal, thus ensuring a homogenous solution for most efficient dansylation reaction.

It was noted that some sample loss could be encountered during the drying down process by SpeedVac. Therefore the $^{12}C/^{13}C$ mixture was diluted 2 fold , and 6 µL of the diluted solution was directly injected into LC-MS without the concentrating step for analysis. For comparison purposes, the control reaction was dried down to half of the original volume, after which 3 µL was used as the injection amount for LC-MS analysis. 1911 (±15) ion pairs were detected in the ACN/ H_2O (50/50) experiment, showing a marked improve over the old protocol as described in the control experiment (1727 (±16) peak pairs) (Fig. 2-2)

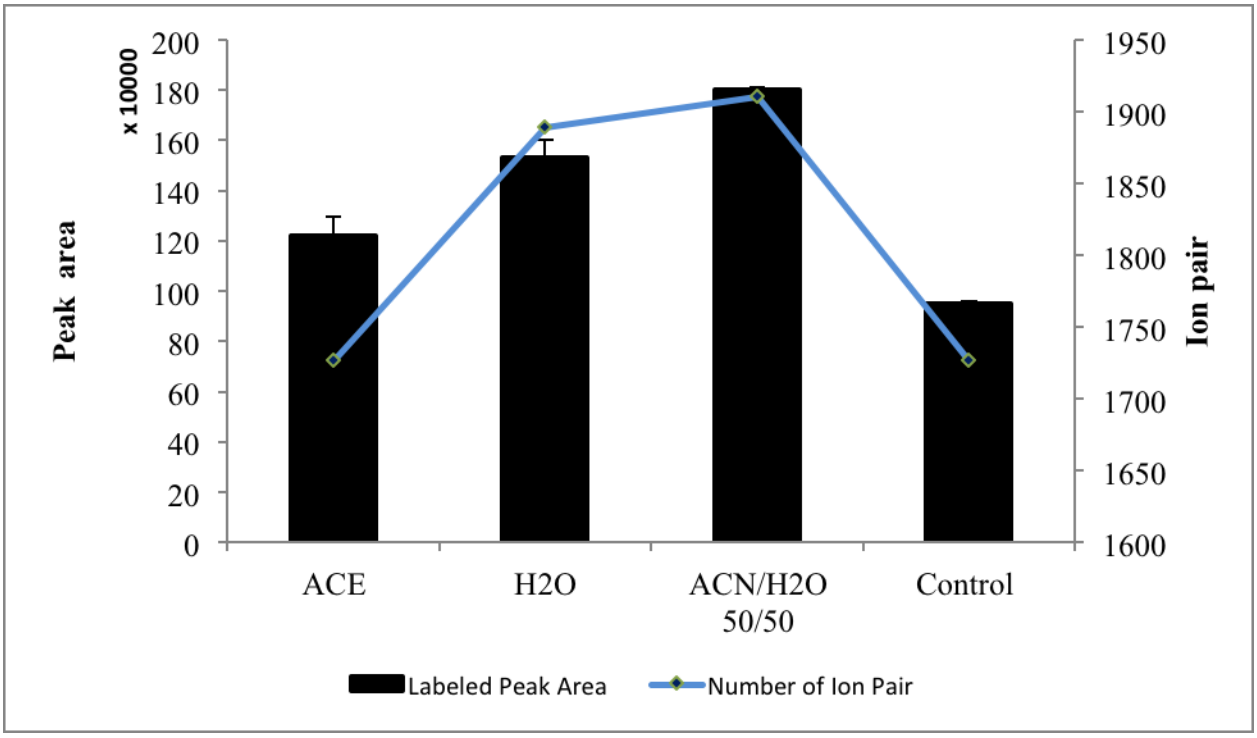


Figure 2-2 Bar graph: Peak area of the dansylated metabolites quantified from UPLC in acetone (ACE), H_2O and ACN/ H_2O (50/50) experiments respectively. Line graph: number of ion pairs detected in each experiment. The experiments were carried out in triplicate.

2.3.2 Determination of optimal amount injection

To determine the amount of volume of sample used in LC-MS analysis, we resorted to the use of dilution curve of a labeled saliva sample for relative quantification of the total concentration of labeled metabolite in individual samples.[89] A dilution curve for the pooled labeled saliva sample was generated as seen in Figure 2-3.

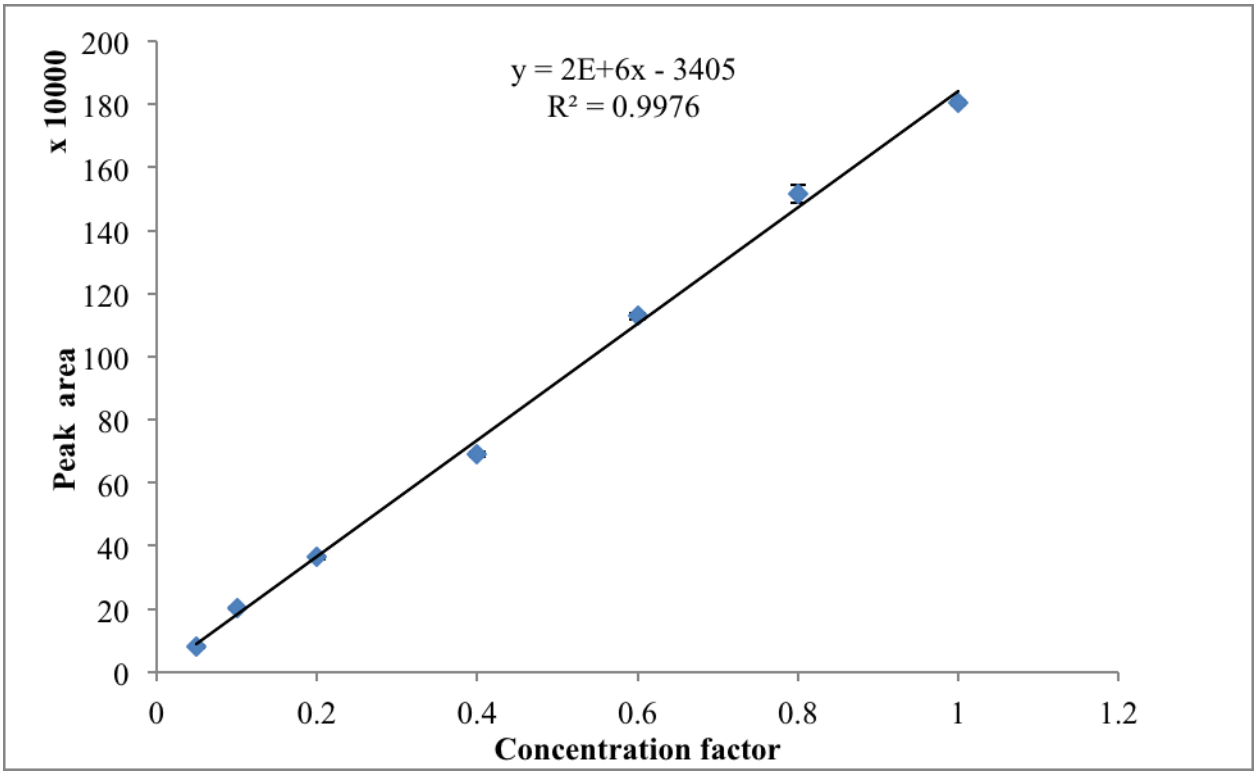


Figure 2-3 Calibration curve built from a series of dilution of ^{12}C-dansyl-labeled pooled saliva sample.

To normalize the total concentration of the ^{12}C- and pooled ^{13}C-labeled samples, the UV peak areas of the individual and the pooled sampled were measured. By comparing the peak area differences, we can use the equation of the calibration curve to determine the mixing volume of the two samples.

2.3.3 Metabolomic profiling by LC-FTICR-MS

Relative standard deviations (RSD) for peak pairs ratio ranged from 0.1% to 15%, with an average of 2% for the triplicate experiments. A set of data that fell within the range of 10-15% was subjected to Grubbs test at 99% confidence level to detect any statistical outlier. After peaking picking by ISOMS and alignment, 7698 unique features (defined as molecular ion m/z coupled with its retention time) were obtained from LC-FTICR-MS analysis. On average, 1515 peak pairs were obtained with signal-to-noise ratio (S/N) threshold of 20 from an individual saliva sample labelled with isotope dansylation chemistry.

2.3.4 Multivariate modeling of NA, MCI and AD

Multivariate regression analysis in terms of orthogonal partial least squares discriminant analysis (OPLS-DA)[90] was applied to study metabolic variation in all three groups, NA, MCI and AD. OPLS-DA, a supervised approach, is utilized to recognize the metabolic changes in different groups of samples while minimizing the influence of inter-group variability. To evaluate the quality of the OPLS-DA models, an internal validation method using a seven-fold cross-validation step was applied, from which the values of Q^2Y (predictive ability of the model) and R^2Y (goodness of fit parameter) were calculated. The score scatter plot shows a very clear separation between the 3 study groups with high validation parameters (R^2Y=0.93 and Q^2Y =0.87), indicating the robustness of the model (Figure 2-4).

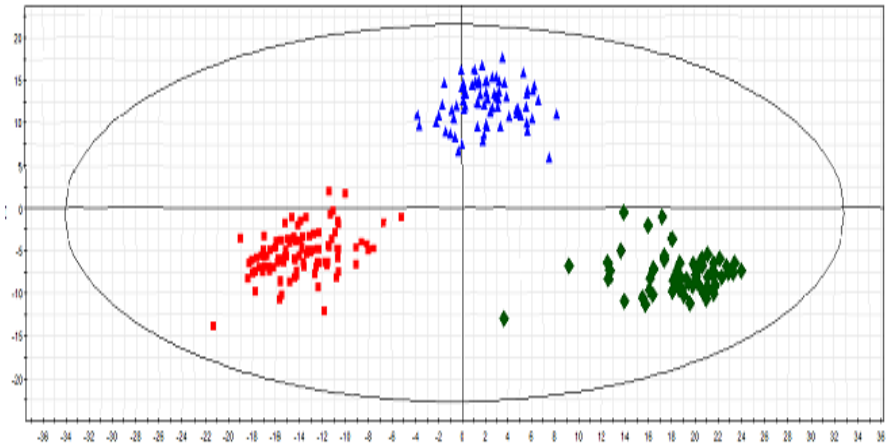


Figure 2-4 The orthogonal partial least squares discriminant analysis score scatter plots for models consisting of three classes : NA (red), MCI (blue) and AD (green).

In addition, OPLS-DA modelling was also applied to compare the metabolic changes between three sets of two classes: NA versus MCI, MCI versus AD, and NA versus AD. The score scatter plots for each statistical analysis are shown in Figure 2-5. Notably, all OPLS-DA models demonstrate clear group separation of groups. The models are also described by high validation metrics, confirming the goodness of fit and good predictive capabilities of the proposed models. Validation parameters calculated for the NA versus MCI models are R^2Y=0.95 and Q^2Y =0.90. Those values are not as high in comparison with those obtained for NA versus AD model (R^2Y=0.99 and Q^2Y =0.96), indicating that the separation between NA and MCI is not as distinguishable as between NA and AD. As anticipated, the most significant metabolites according to the Variable Importance in the Projection (VIP) values generated from NA versus AD and NA versus MCI show an overlapping pattern of perturbation in terms of metabolic profiles, suggesting a definite transition from MCI to AD. All the metabolites that have a VIP value higher than 1.5 were retained for potential biomarker analysis.

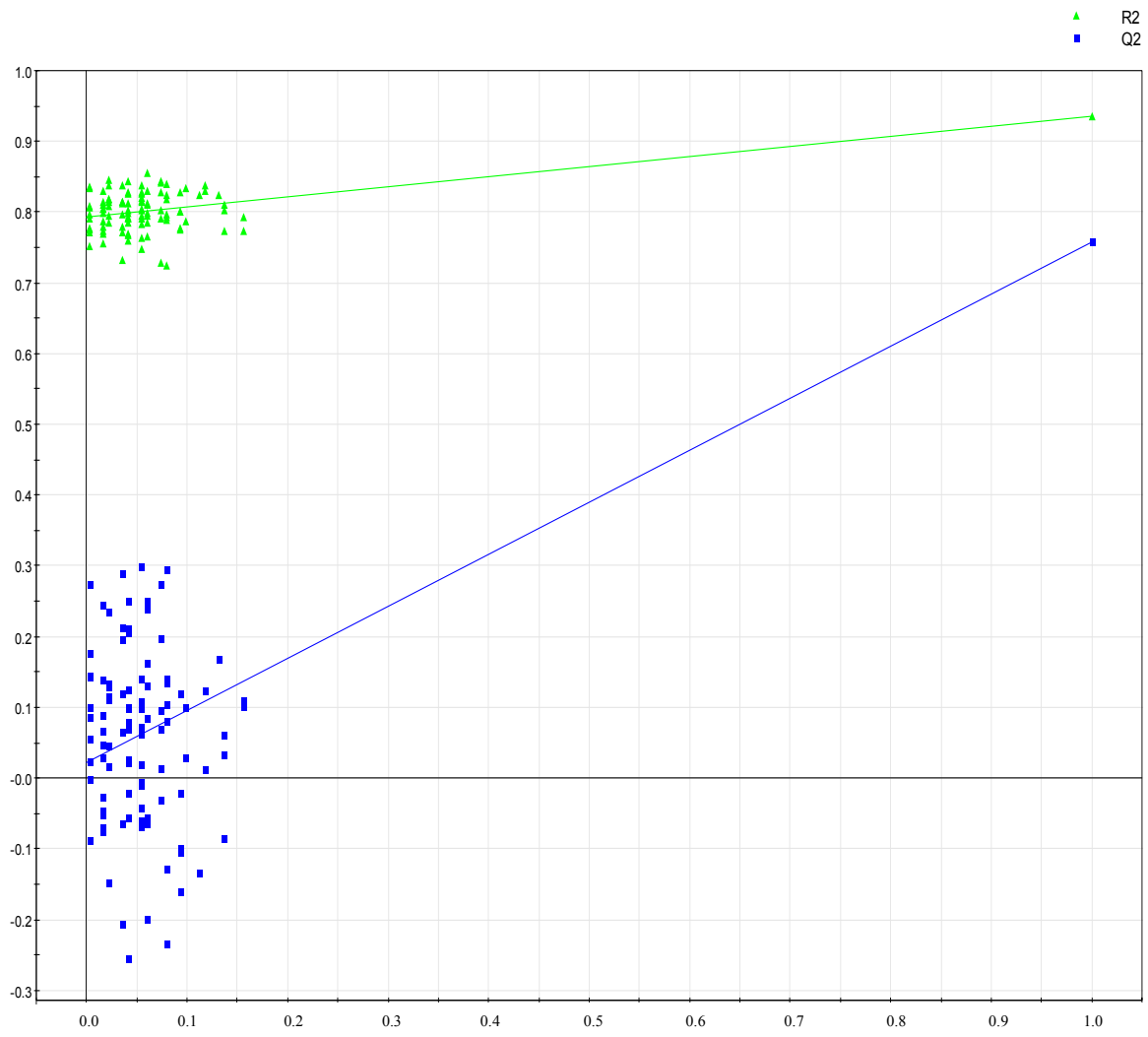


Figure 2-5 Cross validation of the OPLS-DA model of three classes: NA, MCI and AD. Statistical validation of the OPLS-DA model by permutation analysis using 100 different model permutations test built in SIMCA-P+ software.

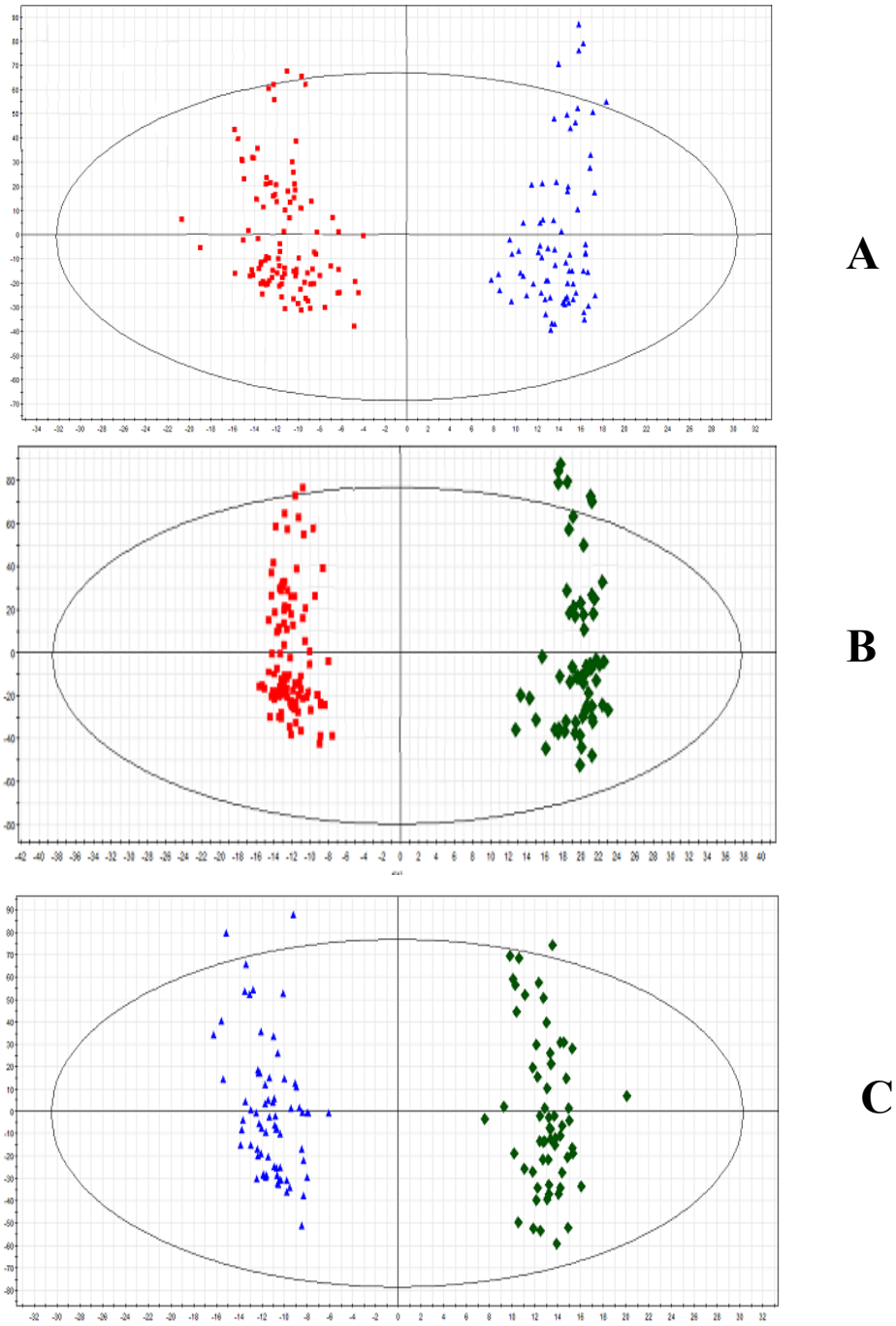


Figure 2-6 The orthogonal partial least squares discriminant analysis score scatter plots for models consisting of two classes: (A) NA versus MCI, (B) NA versus AD, (C) MCI versus AD. Participants in NA group are marked in red, MCI patients in blue and mild AD patients in green.

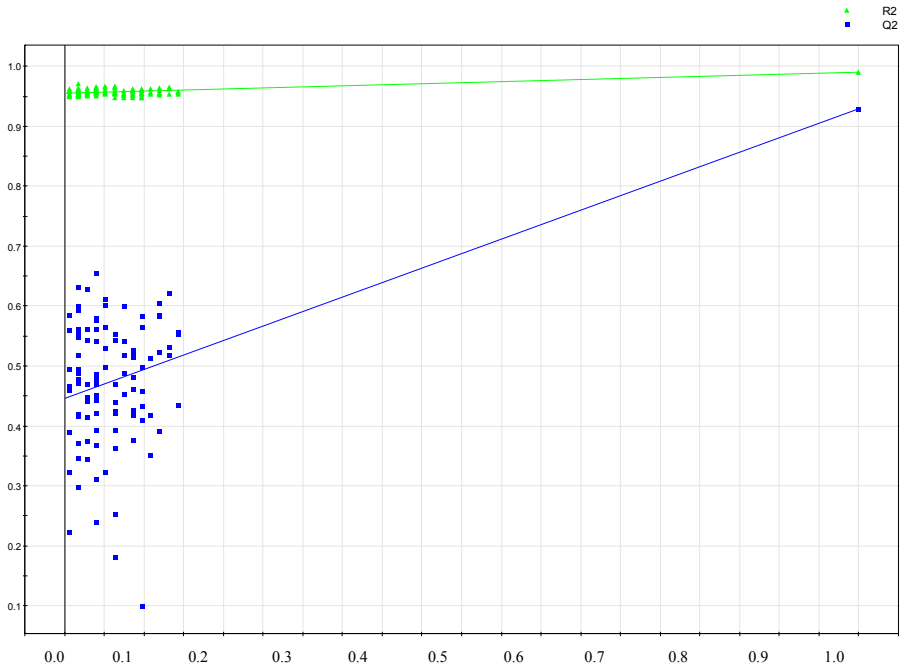


Figure 2-7 Cross validation of the OPLS-DA model derived from (A) NA versus MCI, (B) NA versus AD, (C) MCI versus AD. Statistical validation of the OPLS-DA model by permutation analysis using 100 different model permutations test built in SIMCA-P+ software.

2.3.5 Volcano plot and fold change analysis

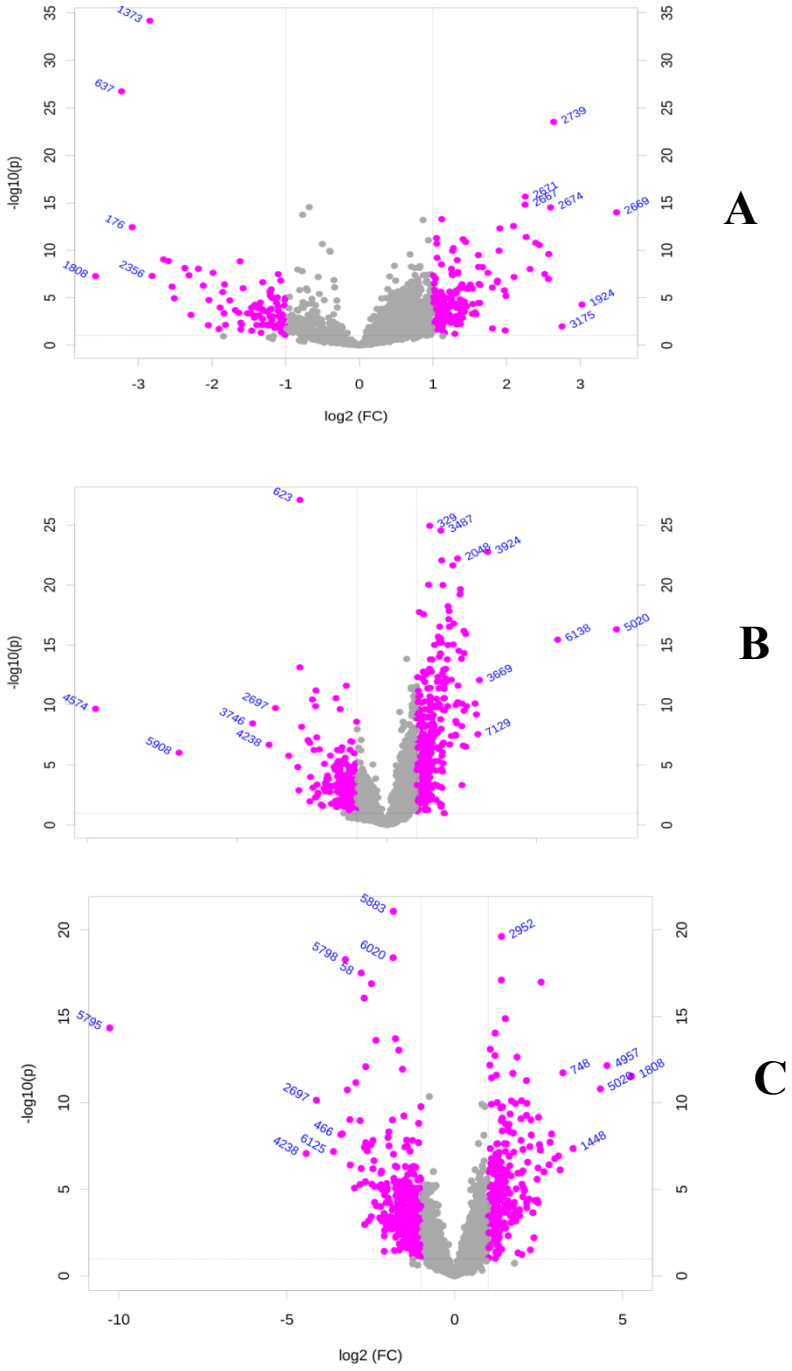


Figure 2-8 Volcano plot of fold change in abundance versus probability value for (A) NA versus MCI, (B) NA versus AD, (C) MCI versus AD. The solid lines indicate cut-off values of fold change of 2.0 (horizontal axis). p = 0.1 (vertical axis). The most significant metabolites (highlighted in pink) are distributed in the top left and right region of the plot.

In order to investigate the metabolic changes due to AD progression, volcano plots were used to visualize abundance differences of metabolites between the selected groups. All significant metabolites were ranked in a volcano plot according to their fold change level and their statistical P-value. A fold change ratio cut-off of 2.0 and threshold p-value of 0.1 were selected to discriminate between the significantly up and down-regulated metabolites between normal aging participants and diseased groups (Fig. 2-6).

2.3.6 Potential biomarkers

We employed a simultaneous combination of two powerful statistical models to determine the significant level of the metabolites under study. This comprehensive strategy allows a more insightful and wider view of the potential biomarkers with high confident information. All the significant metabolites that have a VIP score higher than 1.5 and a fold change cut off of at least 2 ($p<0.1$) are subjected to putative identification through publicly available metabolome database based on direct matching of the accurate mass (m/z). In addition, receiver operating characteristic (ROC) curves were generated for each significant metabolite to judge the combinatory discrimination ability of the aforementioned statistical methods. In a ROC curve, false positive rate (sensitivity) is plotted as a function of false positive rate (1 - specificity). The portion of total area that falls below the ROC curve is defined as the area under the curve, or AUC, which is a well-established validation metric of how well a parameter can discriminate between two studied groups.[84,91] An AUC of 0.5 is equivalent to classifying subjects as having no predictive ability (or random chance), whereas an estimated AUC of 1 indicates perfect discrimination and suggests a very high potential of being a clinically useful biomarker.[92] Therefore, the top discriminant metabolites for each studied groups comparison (NA versus MCI,

NA versus AD, and MCI versus AD) with highest AUC values are depicted in Table 2-2, 2-3, and 2-4 respectively.

Table 2-2 List of potentially important metabolites ranked by AUC values based on the OPLS-DA model for discrimination between normal aging group and mild cognitive impairment group

MW	Putative Match	VIP	AUC	FC	N	NA (M ± SD)	MCI (M ± SD)
168.08		2.71	0.928	4.79	35	.622 ± .209	3.522 ± 4.228
146.071		2.58	0.706	2	35	1.428 ± .780	2.786 ± 1.540
336.158		2.53	0.955	5.24	35	.616 ± .201	3.212 ± 3.156
163.063	3-Methyldioxyindole	2.53	0.795	2.07	35	3.950 ± 2.549	8.168 ± 4.885
170.204	MG(0:0/14:0/0:0); MG(14:0/0:0/0:0)	2.35	0.828	2.17	34	.722 ± 0.489	1.236 ± .168
169.035		2.22	0.804	2.53	35	.387 ± .326	.956 ± .836
147.054	L-Glutamic acid; N-Acetylserine; L-4-Hydroxyglutamate semialdehyde;	2.21	0.799	2.38	35	.437 ± .294	1.021 ± .872
185.009	Phosphoserine; DL-O-Phosphoserine	2.16	0.767	2.53	35	.399 ± .369	1.027 ± .783
127.074		2.03	0.734	2.85	35	.684 ± .857	1.903 ± 2.069
132.054	L-Asparagine;D-Asparagine; N-Carbamoylsarcosine; Ureidopropionic acid	2	0.733	2.04	35	1.46 ± 1.198	3.007 ± 2.178
210.081		1.96	0.758	0.48	35	.984 ± .470	.603 ± .321
234.114		1.96	0.793	2.66	34	.889 ± 1.021	2.227 ± .945
265.096	3,4-Dihydrocinnamic Acid (L-alanine methyl ester) amide	1.89	0.786	2.43	33	.793 ± .534	1.482 ± .781
166.063	Desaminotyrosine	1.85	0.711	2.48	35	.874 ± 1.176	2.219 ± 2.285
240.112	3'-Oxopentobarbitone	1.77	0.711	2.19	35	.708 ± .562	1.511 ± 1.572
101.069		1.75	0.727	2.16	34	.739 ± .499	1.420 ± .847
387.268		1.72	0.774	3.07	34	.858 ± .837	1.947 ± 1.536
270.096	His-Asp; Asp-His	1.68	0.696	2.1	35	3.236 ± 3.809	7.297 ± 8.112
102.106		1.56	0.759	2.08	34	3.128 ± 2.40	6.369 ± 1.199
256.116	His-Thr; Thr-His	1.55	0.697	2.01	35	1.896 ± 2.040	4.004 ± 4.345

Table 2-3 List of potentially important metabolites ranked by AUC values based on the OPLS-DA model for discrimination between normal aging group and Alzheimer disease group.

MW	Putative Match	VIP	AUC	FC)	NA (N)	NA (M ± SD)	AD (N)
74.036		3.07	0.909	1.44	35	.735 ± .240	21
382.149	Tyr-Asn-Ser; Ser-Tyr-Asn; Tyr- Ser -Asn ;AsnSer Tyr	2.94	0.889	1.81	35	.779 ± .469	21
297.107	1-Methylguanosine	2.89	0.861	1.83	35	.773 ± .298	21
383.152	Met- Ala -Tyr; Met -Tyr- Ala; Tyr- Ala -Met	2.84	0.872	2.21	30	.923 ± .479	19
293.1	N'-Formylkynurenine	2.78	0.897	2.36	35	.728 ± .246	20
333.134	TrpGlu; GluTrp	2.72	0.865	1.87	34	.637 ± .282	21
212.091	Gly His; His Gly	2.65	0.853	2.07	35	1.937 ± 1.88	21
254.148		2.62	0.81	1.77	28	.753 ± .230	20
341.208	Melanostatin	2.59	0.872	1.09	35	.904 ± .394	21
295.116	Tyr Asn; Asn Tyr	2.58	0.825	1.23	35	.948 ± .386	21
284.111	His-Glu, Glu-His	2.49	0.836	2.24	33	1.738 ± 1.209	21
253.152		2.41	0.841	2.08	30	1.205 ± .736	20
267.169		2.33	0.811	2.05	30	.953 ± .733	20
201.111	N-a-Acetylcitrulline; N-a-Acetyl-L-arginine; 2-Oxoarginine;	2.31	0.843	1.72	33	.711 ± .336	21
505.254	6-Keto-decanoylcarnitine	2.27	0.858	2.44	24	.835 ± .607	19
252.627		2.23	0.858	2.46	24	.848 ± .634	19
330.19		2.2	0.809	1.19	35	.908 ± .525	21
397.196	Ile Glu His; Glu His Ile; His GluLeu; His LeuGlu; Leu His Glu; GluLeu His; His Glu Ile; Glu His Leu; His Ile Glu; Ile His Glu; LeuGlu His	2.19	0.81	1.97	31	1.477 ± .967	20
155.069	L-Histidine	2.17	0.849	1.97	35	1.937 ± .351	21
323.196		1.93	0.818	1.85	28	.636 ± .391	20

Table 2-4 List of potentially important metabolites ranked by AUC values based on the OPLS-DA model for discrimination between mild cognitive impairment group and Alzheimer disease group.

MW	Putative Match	VIP	AUC	FC	MCI (N)	MCI (M ± SD)	AD (N)	AD (M ± SD)
352.145	Estra-1,3,5(10)-triene-3,17beta-diol 3-phosphate	3.85	0.913	1.40	22	.660 ± .159	22	1.334 ± .430
382.149	Tyr-Asn-Ser; Ser-Tyr-Asn; Tyr- Ser - Asn ;AsnSer Tyr	3.01	0.823	1.24	25	1.138 ± .769	22	2.670 ± 1.373
293.100	N'-Formylkynurenine	2.85	0.834	1.73	23	1.273 ± .903	22	3.265 ± 1.988
333.134	TrpGlu; GluTrp	2.74	0.802	1.37	23	.967 ± .763	22	2.297 ± 1.517
155.069	L-Histidine	2.71	0.803	1.65	25	2.552 ± 2.983	22	7.823 ± 5.735
383.152	Met- Ala -Tyr; Met -Tyr- Ala; Tyr- Ala - Met	2.66	0.878	2.57	17	1.251 ± .766	21	3.229 ± 1.690
234.101	5-Methoxytryptophan	2.26	0.822	-1.21	25	1.889 ± 1.381	22	1.130 ± 1.719
310.132	Taurocyamine	2.22	0.812	-1.47	25	1.738 ± 1.542	21	.933 ± 1.353

Dissecting the list of significant metabolites is of particular interest since the link from neurodegeneration metabolism to cognitive impairment has been found to be closely correlated with pathogenesis of Alzheimer's disease.[93] One of the major adducts of protein oxidation, N'-Formylkynurenine (NFK), formed by the oxidation of tryptophan[94] is identified as one of the most discriminating metabolites between AD patients and NA participants (AUC= 0.897) (Fig 2-7). The concentration of NFK is higher (5-fold) in subjects with Alzheimer's disease compared with cognitively normal aging participants, which is in agreement with the published proteomic results about an elevation of NFK in cerebrospinal fluid (CSF) of AD patients.[95] NFK has been reported to be involved in many important and neurodamaging events in AD pathology[96] and therefore could potentially be a biomarker indicative of oxidative stress in AD participants.

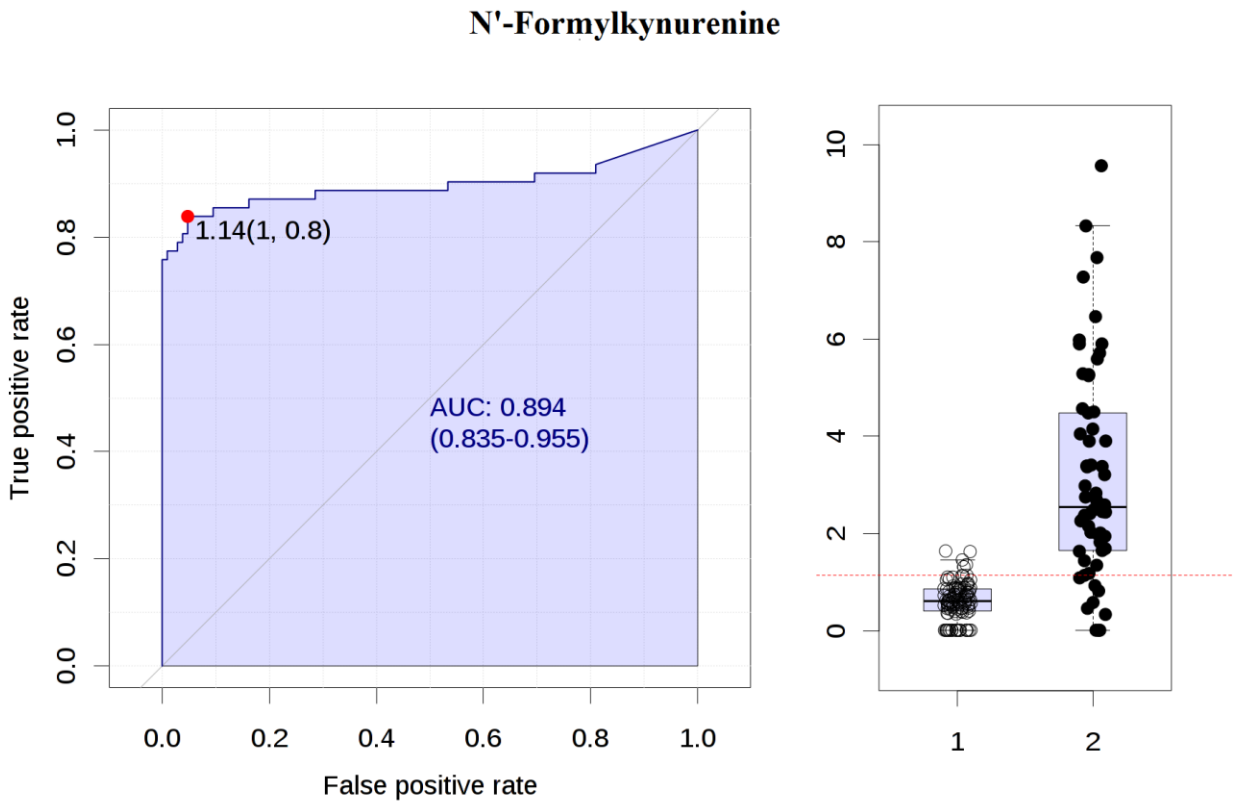


Figure 2-9 The Receiver operating characteristic curve was performed to evaluate the diagnostic performance of the potential biomarker N'-Formylkynurenine.

In this study, we found that AD and MCI patients, compared with NA controls, had higher level of L-histidine and histidine-rich peptides. L-histidine and gly-his peptide have been unambiguously identified by matching their retention time with commercial standards. Similar results have been obtained by Molina et al. (1998) and Schmitt et al. (1984) when they compared a group of demented patients with normally aging controls.[97,98] This significant elevation provides evidence that high levels of histidine could have a pathophysiological implication in dementia diseases that need to be further investigated.

Olfactory dysfunction is present in many individuals with mild cognitive impairment and early Alzheimer's diseases, suggesting that it is one of the first signs of the brain disorder.[99] Degeneration of olfactory system has been proposed as one of the mechanisms of cognitive impairment. The present study has shown that 3-methyldioxyindole, a major metabolite of olfatotoxin 3-methylindol has a 2-fold increase in MCI subjects and ranked as one of the most significant metabolite in discriminating between MCI patients and NA individuals with the ROC AUC value of 0.795. This demonstrates the feasibility of utilizing 3-Methyldioxyindole as a neurobiological biomarker for recognizing the first clinical signs of MCI disease.

Of particular interest is a group of compounds known as phosphomonoester which has been shown to be elevated in AD patient's brains.[100] One of the most abundant of phosphomonoester in human brain is phosphoserine, which has been found to be significantly increased for the MCI patients in the present study (2.5 fold) above normal aging matched controls, but doesn't vary significantly in AD patients. Interestingly, there has been reports that phosphomonoester levels are at its peak at a very early stage of the disease.[101] This evidence, together with our findings, indicates that phosphoserine could be a physiologically highly relevant metabolite marker in detecting MCI disease.

2.4 Conclusion

In this work, an improved isotope labelling LC-FTICR-MS method has been developed for metabolite biomarker discovery using human salivary samples. Even though only a very small starting material (5μL of individual saliva) was used for the experiments, we were able to detect approximately 1500 metabolites. This study demonstrated the power of untargeted metabolite profiling in MCI and AD patients and provides new insights into the analysis of brain metabolism. More importantly, we have identified a group of potentially significant metabolites for MCI and early AD detection. Of course, a second validation phase should be carried out using a larger cohort of demented patients to obtain an unbiased and real estimate of their discriminating performance. Nonetheless, this metabolite profiling strategy could be regarded as a very promising application as a rapid and cost-effective MCI and AD prognosis. Saliva metabolomics, in particular, will propel the field of medial diagnostics forward because of its easy, inexpensive and simple approach for sample collection and disease detection. Development of salivary biomarkers could be used to monitor the disease progresses. Currently available and newly emerging metabolomics techniques will open new possibilities that saliva can eventually be employed as a medium for clinical applications.

Chapter 3: Development of Human Urine Metabolome Database

3.1 Introduction

Metabolomics is the study of the "end products" or metabolites that are the results of numerous biochemical and biological processes. The measurement of metabolite levels and variations is particularly useful as it can provide a great deal of information into disease processes,[102] discovery of drugs[103] and environment-gene interactions.[104] In this work, we utilized differential isotope labeling via dansylation combined with Liquid Chromatography (LC) and Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) to target phenol and amine containing submetabolome in human urinary samples.

Urine has long been regarded as a useful diagnostic biofluid since the ancient Egypt.[105,106] Urine, an amber-colored, transparent liquid secreted by the kidneys, contains a variety of compounds. Since its collection is simple and non-invasive, urine can be sampled in a serial manner, thus providing insights into temporal metabolic perturbation studies as well as key information about changes in response to disease processes.[107,108]

However urine poses several challenges for metabolic profiling because of the huge dynamic range of metabolite concentrations, the diversity of chemical classes as well as other confounding factors such as age, gender, body mass index (BMI), and environment. Several groups have attempted to perform metabolic profiling on urine by nuclear magnetic resonance spectroscopy (NMR),[109] high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS),[110] ultra high performance liquid chromatography (UPLC-RP),[111] and two-dimensional gas chromatography coupled to quadrupole mass spectrometry (GCxGC-MS).[112] To facilitate research into urine metabolome, we employed a parallel application of LC-MS and

differential isotope labeling approach via dansylation for comprehensive metabolic profiling. By sampling urines of 100 healthy volunteers for a 3 continuous-day period, we aim to investigate temporal metabolic changes and generate a comprehensive metabolite profile. Furthermore, significant metabolites contributing to the sample differences based on gender, age, and BMI are being discovered and elucidated. We also envisage to build a reference metabolome database that will benefit the search of metabolite biomarkers for diagnosis and prognosis of different diseases. This database will form the basis of how these reference metabolites will be affected by disease development.

3.2 Experimental

3.2.1 Subjects

100 healthy volunteers who claimed to have met the requirements for the study provided a second void volume (morning), midstream urine sample after 8 hour of fasting (with the only exception of water) once a day for three consecutive days. The eligible participants were non-smokers and between the age of 18 and 40 years. They also did not consume prescription medications, counter pharmaceuticals or natural supplements. Female participants should not be menstruating on the days of sample collection. Identity and demographic information of the volunteers remained anonymous and the information was protected under the Freedom of Information and Protection of Privacy Act. The study was conducted in accordance with the codes of the University of Alberta's Arts, Science, and Law Research Ethics Board. Table 3-1 provides a list of demographic characteristics for the studied population.

Table 3-1 Demographic characteristics of the study population.

Characteristics	No.
Gender	
Male	35
Female	65
Age	
≤ 25 years	52
≥26 years	48
Body Mass Index (BMI)	
Underweight (< 18.5)	9
Normal Weight (18.5-24.9)	69
Overweight (≥25)	22

3.2.2 Sample Sampling and Storage

Once the urine sample was collected in the provided 50 mL BD Falcon polypropylene conical tube, volunteers were directed to store the sample in the 4˚C refrigerator in the contaminant level 2 laboratory. Samples were processed within two hours of receipt. Samples were centrifuged for ten minutes at 4000 rpm. The supernatant was then filtered twice through a 0.2 µm polyvinylidene fluoride (PVDF) filter and collected into a BD Falcon tube. 500 µL aliquots of the double filtered urine were then stored in a -80˚C freezer.

3.2.3 Chemicals and Reagents

^{13}C-dansyl chloride was synthesized in-house as described by Guo and Li.[79] ^{12}C-dansyl chloride was purchased from Sigma-Aldrich (Milwaukee, WI). All reagents were of ACS grade or higher with water and organic solvents being of MS grade.

3.2.4 Labeling Reaction

An aliquot of 12.5 µL of urine sample was dissolved in 62.6 µL of $NaHCO_3/NaH_2CO_3$ buffer solution (166.7 mM, 1:1, v/v) in a screw cap vial. [1]The vial was vortexed, then spun down. 75 µL of freshly prepared ^{12}C-DnsCl or ^{13}C-DnsCl in acetonitrile (12mg/mL) was added into the vial. The solution was vortexed and spun down again then let to react for 60 minutes in an oven at 60˚C. 10 µL of NaOH (50mM) was then added to quench the excess DnsCl. After another 10 minutes incubation in the 60˚C oven, 50 µL of formic acid in ACN/ H_2O (425 mM, 50/50) was added to neutralize the solution. The experimental workflow for human urinary analysis is summarized in Figure 3-1.

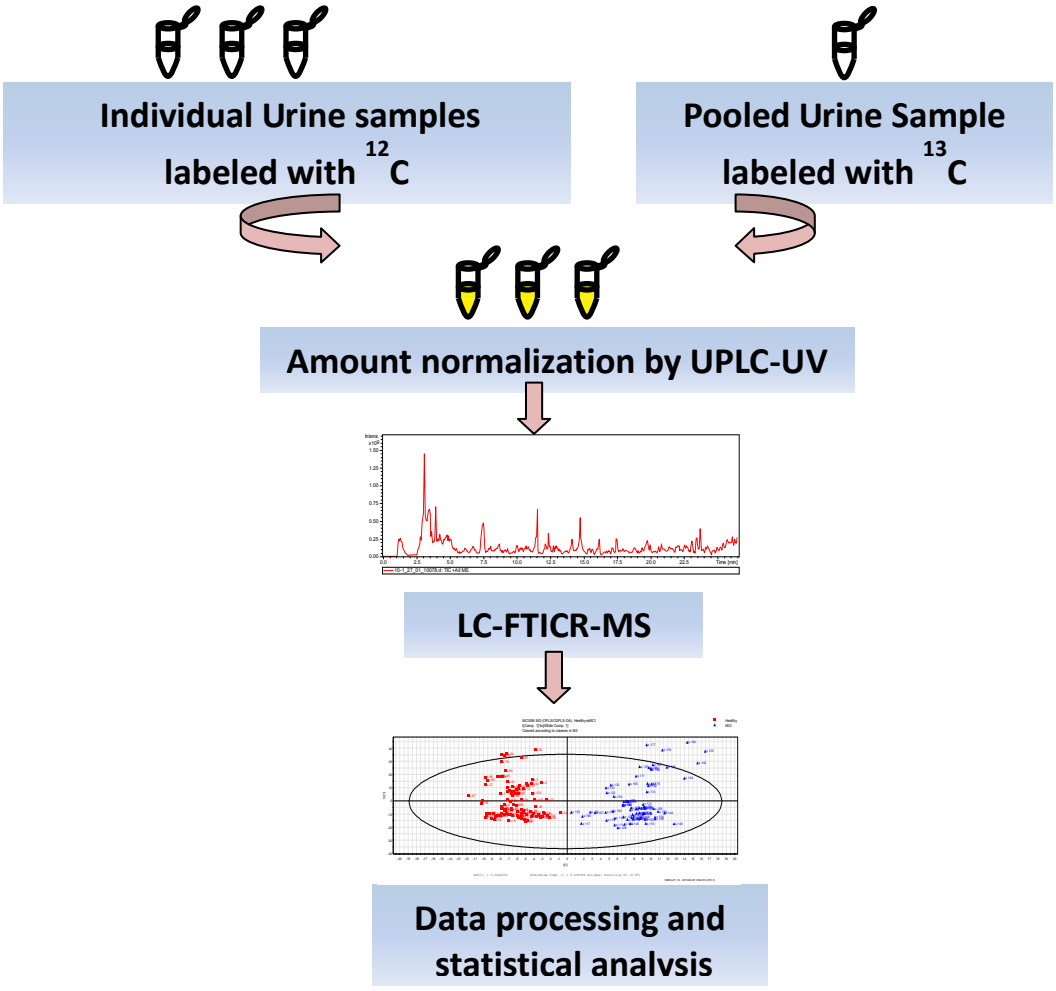


Figure 3-1 The experimental workflow for human urinary analysis

3.2.5 Universal Metabolite Standard

To reduce biological variance and to generate a sample that is representative of the dataset, a pooled sample was prepared by taking equal amount of aliquots from all the urine samples. This pooled sample served as a quality control to monitor the reproducibility of the experiments. Most importantly, when labeled with the heavy-isotope labeling reagent, this pooled sample will become the universal metabolite standard for future work involving metabolome profiling of the same type of sample. For example, when the individual urine sample was labelled with light isotope reagent and mixed with heavy isotope reagent-labeled metabolite standard, the relative concentration of any urine metabolite can be determined. Following the LC-MS analysis, the peak pair ratio of the labeled metabolite will allow the determination of the relative amount of the metabolite in the individual sample to that in the universal metabolite standard. If the concentration of the metabolite in the universal standard is known, the absolute concentration of the metabolite in individual samples can also be calculated.

3.2.6 UPLC-UV

30 μL of each solution was prepared into a UPLC vial. 2 μL was the injection amount. The peak area of each individual urine sample was used to calculate the amount of volume to be mixed with 20 μL ^{13}C-DnsCl pooled urine, according to a predetermined calibration curve:

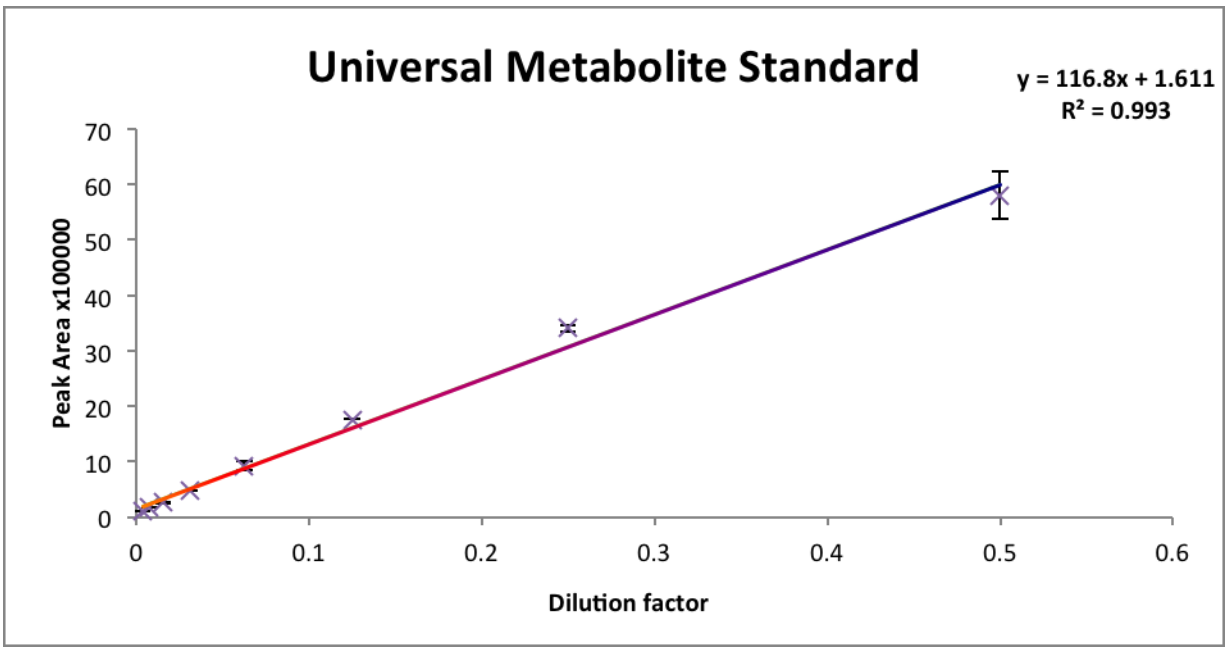


Figure 3-2 Calibration curve built from a series of dilution of ^{12}C-dansyl-labeled universal metabolite standard.

3.2.7 LC-FTICR-MS

The combined ^{12}C/^{13}C-DnsCl labeled urine mixture was prepared for LC-MS analysis. Reversed phase chromatographic separation was carried out on an Eclipse C18 column (2.1 mm ×100 mm, 1.8 µm), with solvent A being water with 0.1% (v/v) formic acid and 5% acetonitrile (ACN) (v/v), and solvent B being ACN with 0.1% (v/v) formic acid. The flow rate was 180 µL/min and running time was 28 min. The gradient was: t = 0 min, 20% B; t = 3.50 min, 35% B; t = 18.00 min, 65% B; t =24.00 min, 99% B; t = 28 min, 99% B. The sample injection volume was 4 µL and the flow was split 1:3 before entering the ESI-MS system.

3.2.8 Data Processing and Statistical Analysis

The ^{12}C/^{13}C ion pairs of the same metabolites were detected by a sensitive and robust in-house peak peaking program ISOMS. The peak pair data were aligned by retention time and

accurate mass and each metabolite feature was converted into a matrix of peak pairs ratio versus metabolite identification. Volcano plot and univariate-fold change analysis were performed for comparisons between samples based on gender, age group and BMI by Origin 9.1 (OriginLab, Northampton, MA). The mean-centered and unit-variance (UV)-scaled data were introduced into SIMCA-P+ 12.0 software (Umetrics, Umea, Sweden) for multivariate analysis to reveal any discrimination of the 3 comparisons based on gender, age group and BMI.

3.2.9 Metabolite Identification

Resulting significant metabolites were putatively identified against the Human Metabolome Database (HMDB)[85-87] and MyCompoundID[88] database using a mass accuracy window of 5 ppm. Isotope patterns of those metabolites were manually inspected to eliminate any possible mismatches.

3.3 Results and Discussion

3.3.1 Reproducibility Analysis

A major challenge with LC-MS based method is the potential of the analytical system to vary with time during analysis. One of the popular approaches to deal with this issue is to employ a quality control (QC) sample. This QC sample would be a pooled urine sample, prepared by mixing aliquots of all samples. Our study has aimed to build an universal metabolite standard, which can be conveniently used for quality control purposes. For a small sample set, this QC sample can be easily prepared and be broadly representative of the whole sample set. This sample was injected at regular intervals (i.e., every 30 samples) throughout the analytical run to generate a set of data from which reproducibility can easily be monitored. As shown in the Principal Component Analysis (PCA) score slot (Fig 3-3), all the QC samples clustered closely,

indicating that the variability in the analytical run was close to none, thus assuring the validity of the experiment.

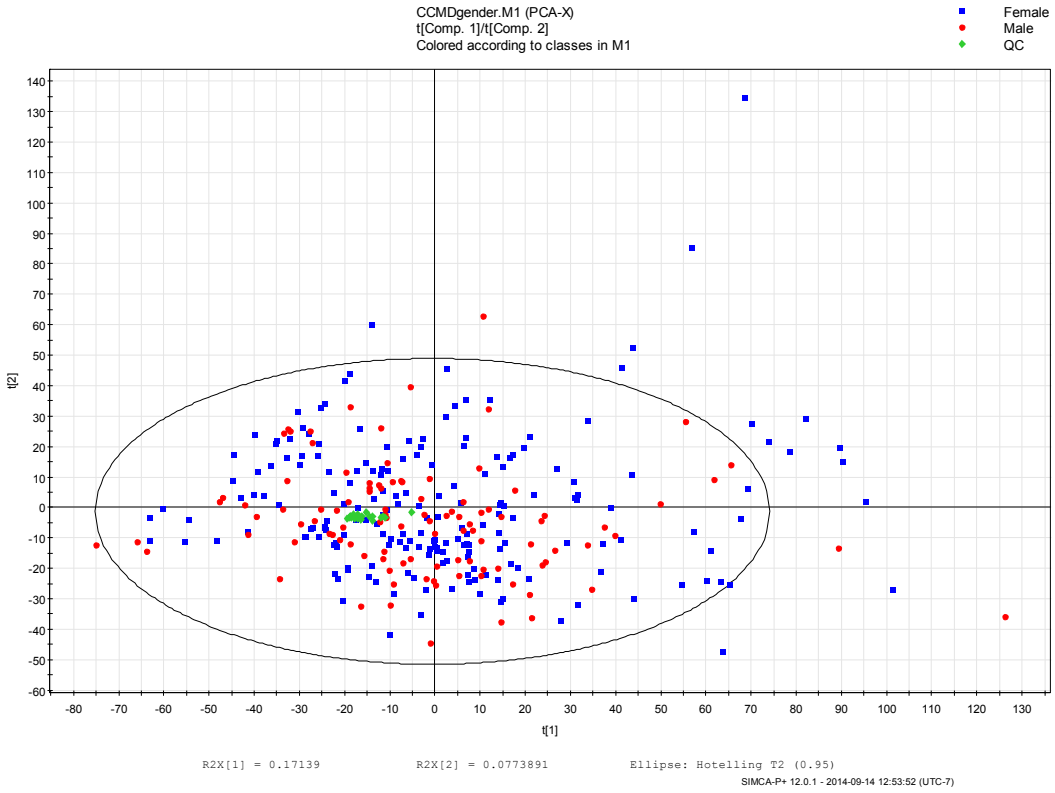


Figure 3-3 A two dimensional PCA scores of human urine female sample (blue), human urine male sample (red) and QCs (green) obtained by LC-ESI-MS in positive mode.

3.3.2 Analysis of Gender Differences

Analysis of the metabolomics profiles from urine samples showed a clear differentiation between male and female participants. An orthogonal partial least squares discriminant (OPLS-DA) score plot) (Fig. 3-4) demonstrated a good separation between the male and female groups (R^2Y=0.89 and Q^2Y =0.83). The Variable Importance in the Projection (VIP) parameter was used to select the metabolites that have the most significant contribution to the separation. Only VIP values >1.5 were selected and used for further data analysis.

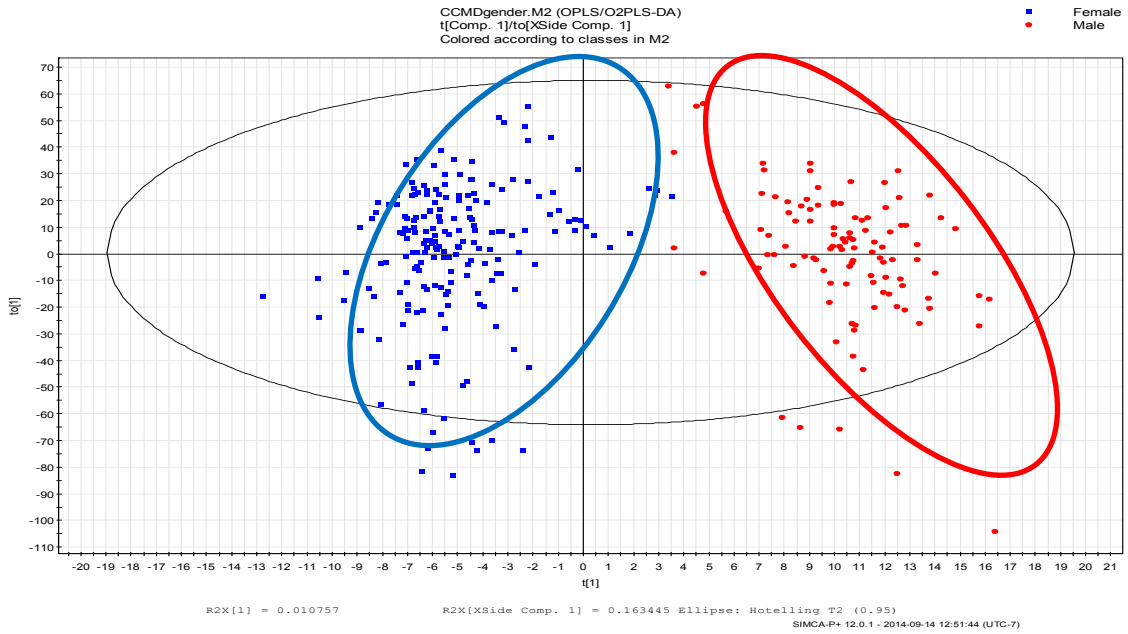


Figure 3-4 OPLS-DA score plot showing a significant separation between female and male volunteers.

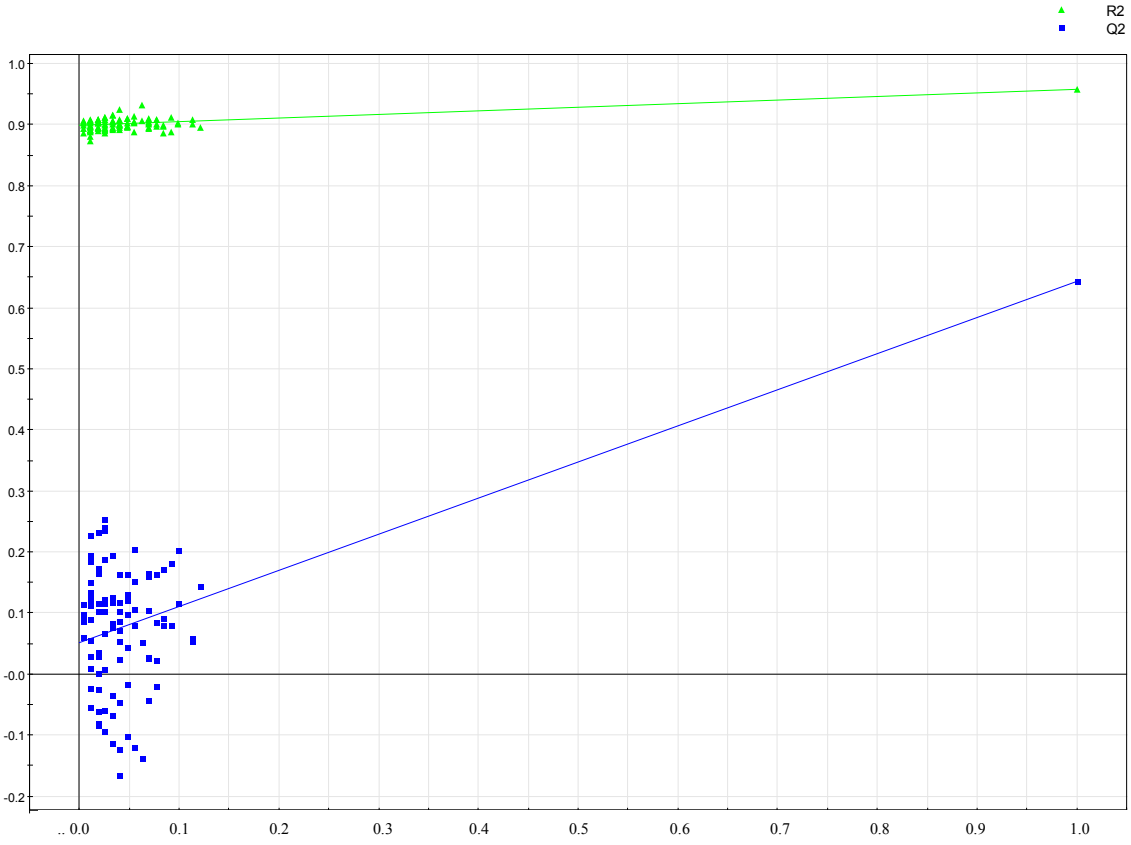


Figure 3-5 Cross validation of the OPLS-DA model derived from male versus female participants. Statistical validation of the OPLS-DA model by permutation analysis using 100 different model permutations test built in SIMCA-P+ software.

In order to investigate the metabolic changes due to gender difference, volcano plots were used to visualize abundance differences of metabolites between the selected groups (Fig. 3-6). Volcano plots were useful in depicting the up-regulated or down-regulated metabolites. The metabolite that has the fold change greater than 1.2 was selected and used in further analysis to study the significantly discriminant metabolites.

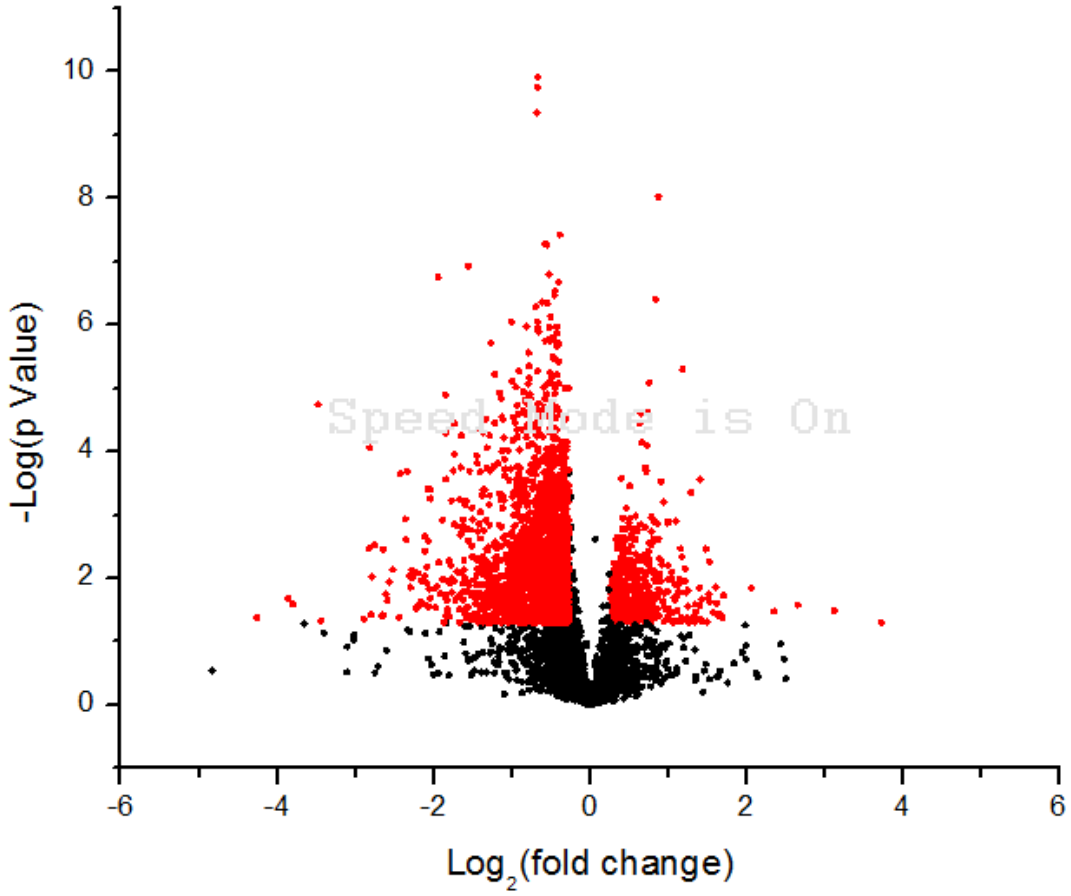
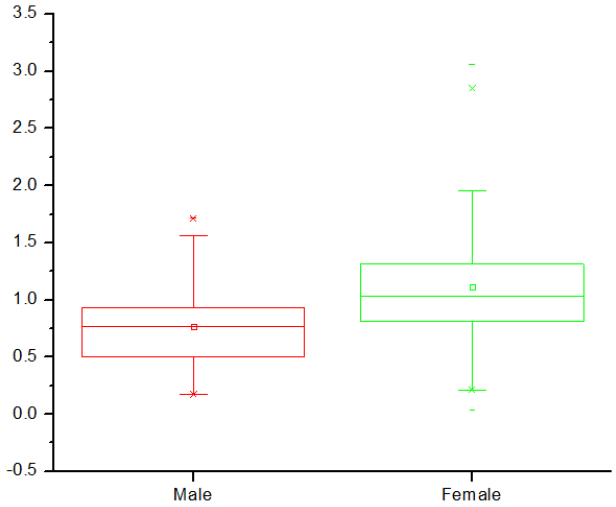


Figure 3-6 Volcano plot of fold change in abundance of metabolites versus probability value for Male versus Female. The most significant metabolites (highlighted) in red are distributed in the top left and right region of the plot.

820 metabolites satisfied both the VIP and fold change requirements in the gender-related metabolomics profiles analysis. Men were found to excrete greater amounts of Diethylthiophosphate and 5-Hydroxykynurenamine. Meanwhile, higher concentrations of 5-Aminopentanamide, L-Serine, Putrescine and 5-Hydroxy-L-tryptophan were present in female urine.

Amino acid excretion was found to vary between genders in our sample. Greater serine excretion was observed in female participants which agreed with some of findings by Fukagawa et. al.[113] Serine is an important component in the transsulferation pathway. This result suggests that male and female differ in their utilization of certain amino acid, and study of metabolic differences can provide insights into human biochemical pathways. A more detailed description of the most significantly discriminations that distinguish male and female urine samples were tabulated in Table 3-2.



5-Aminopentanamide or isomers

Figure 3-7 Box plot of one of the most significantly discriminant metabolites in gender-related metabolome analysis.

Table 3-2 List of potentially important metabolites ranked by VIP values based on the OPLS-DA model for discrimination between male and female groups

VIP	FC>1.2	p-value	m/z	RT (s)	Male	SD	Female	SD	Putative Match (my compound ID 0 rxn)	Putative Match (my compound ID 1rxn)
3.79	↑	9.56E-09	147.0908	494.42	1.47	0.9	0.80	0.6		28
3.36	↓	1.23E-10	129.0429	603.71	0.70	0.4	1.12	0.6	5	2
3.17	↓	1.80E-06	188.1609	939.07	0.84	0.4	1.26	0.7		6
3.10	-	-	240.0067	1335.25	0.98	1.1	0.84	0.6		9
3.01	↓	4.61E-07	116.0956	461.47	0.76	0.4	1.11	0.5	5-Aminopentanamide	0
2.99	↓	1.08E-06	269.5831	734.26	0.92	0.7	1.62	1.3		9
2.98	↓	1.59E-06	174.0303	1314.55	0.82	0.4	1.14	0.5		40
2.97	↑	8.24E-06	147.0891	467.65	1.36	1.0	0.80	0.6		2
2.93	↑	2.40E-03	227.1267	774.14	1.46	1.0	0.98	0.5		36
2.92	↑	1.33E-03	131.0944	698.65	1.55	1.7	0.78	0.7	6	4
2.91	↓	2.30E-03	259.1502	763.39	0.83	0.5	1.07	0.5		30

2.91	↓	1.61E-07	129.0792	580.27	0.81	0.4	1.17	0.6	4	32
2.91	↓	7.44E-07	152.0481	1300.77	0.84	0.4	1.18	0.5	9	33
2.88	↓	1.42E-03	190.0478	677.50	0.81	0.3	1.03	0.5		13
2.83	↓	3.77E-08	215.0582	138.01	0.90	0.2	1.18	0.5	Glycerylphosphorylethanolamine	3
2.80	↑	3.58E-05	170.0145	1310.31	1.10	0.7	0.71	0.5	Diethylthiophosphate	52
2.77	↓	5.46E-08	180.0432	1121.12	0.93	0.4	1.36	0.8	8	21
2.75	↓	6.13E-06	105.0426	232.65	0.78	0.4	1.05	0.4	L-Serine	2
2.75	↓	1.78E-07	144.1620	321.40	0.38	0.3	1.46	2.2		41
2.75	↓	4.49E-06	226.0957	580.31	1.18	1.0	2.03	1.4	Porphobilinogen	2
2.75	↓	1.18E-07	144.1642	354.65	0.56	0.5	1.64	2.2		2
2.74	↓	2.91E-07	152.0477	1288.36	0.83	0.4	1.13	0.5	9	32
2.74	↓	1.10E-06	263.0804	1005.64	0.77	0.3	1.10	0.6		12
2.73	↑	3.48E-04	180.0903	1224.51	1.33	0.8	0.94	0.6	5-Hydroxykynurenamine	22
2.70	↓	1.34E-06	152.0475	1281.75	0.82	0.4	1.10	0.5	9	32
2.70	↓	1.69E-06	180.0446	1132.59	0.99	0.4	1.42	0.8	8	50
2.69	↑	1.55E-02	153.0436	219.82	1.77	3.0	0.75	0.6	Hydroxyanthranilic ; 3-Aminosalicylic acid	24

2.67	↓	2.12E-07	212.9914	471.89	1.11	0.4	1.46	0.7		7
2.67	↑	7.16E-05	188.0814	391.36	1.86	1.4	1.17	0.8	N-Acetylglutamine; L-glycyl-L-hydroxyproline	22
2.66	↓	4.36E-05	161.0488	1481.57	4.37	7.7	11.59	14.1	3	19
2.66	↓	2.64E-04	202.0913	398.59	0.80	0.3	1.04	0.5		19
2.65	↓	9.73E-06	262.0952	658.03	0.85	0.4	1.27	0.8		22
2.65	↓	9.02E-07	263.1388	434.80	1.27	1.0	2.55	2.3		21
2.65	↑	2.49E-05	210.0659	591.80	1.54	1.1	0.98	0.9		40
2.64	↓	3.47E-06	220.0861	513.71	1.05	0.4	1.43	0.7	5-Hydroxy-L-tryptophan	27
2.64	↓	7.55E-03	88.0986	427.47	0.88	0.4	1.07	0.5	Putrescine	5
2.63	↓	3.12E-04	152.0877	454.66	0.83	0.4	1.08	0.5		15
2.62	↓	5.40E-05	439.1873	701.82	0.63	0.3	0.87	0.5		1
2.62	↓	1.82E-06	222.1008	548.50	0.80	0.4	1.15	0.6		24
2.62	↓	6.42E-04	167.0584	571.61	0.97	0.7	1.40	0.9	Pyridoxal	35
2.61	↓	1.08E-06	152.0473	1277.18	0.83	0.4	1.11	0.5	9	32
2.61	↓	1.82E-06	323.0803	634.43	0.99	0.5	1.35	0.6		18
2.61	↓	1.98E-06	152.0476	1268.10	0.83	0.4	1.10	0.5	9	32

2.60	↓	1.28E-06	222.0552	645.32	0.14	0.1	0.22	0.1		15
2.60	↓	4.43E-07	210.0533	1083.86	1.00	0.6	1.53	1.0	Vanilpyruvic acid	31
2.58	↓	5.31E-05	128.0936	557.03	0.71	0.5	1.11	0.9		8
2.58	↑	3.06E-03	197.1156	1221.29	1.37	1.1	0.95	0.6		3
2.57	↓	8.62E-05	130.1069	454.29	0.89	0.4	1.18	0.5	N-Acetylputrescine	10
2.57	↓	3.96E-02	191.0577	883.37	0.82	0.5	0.99	0.6	5-Hydroxyindoleacetic acid; N-Acetyl-L-methionine	31

Key : M:mean. SD: standard deviation. ↑ c indicates an up-regulation that has higher metabolite concentration in the male group, and ↓ represents a down-regulation that has higher metabolite concentration in the female group.

3.3.3 Analysis of Age Effects

To identify the age-related differences in urinary metabolic phenotypes, we set the age of 25 as the cut off point between two age groups since the age of the volunteers ranged from 18 to 36 years and the average age was ~25. OPLS-DA model was again calculated for the age-restricted dataset. A high validation parameter ($R^{2(}Y$)=0.907; Q^2(cum)=0.755) was obtained, thereby indicating that there was indeed a difference in the sample population due to age.

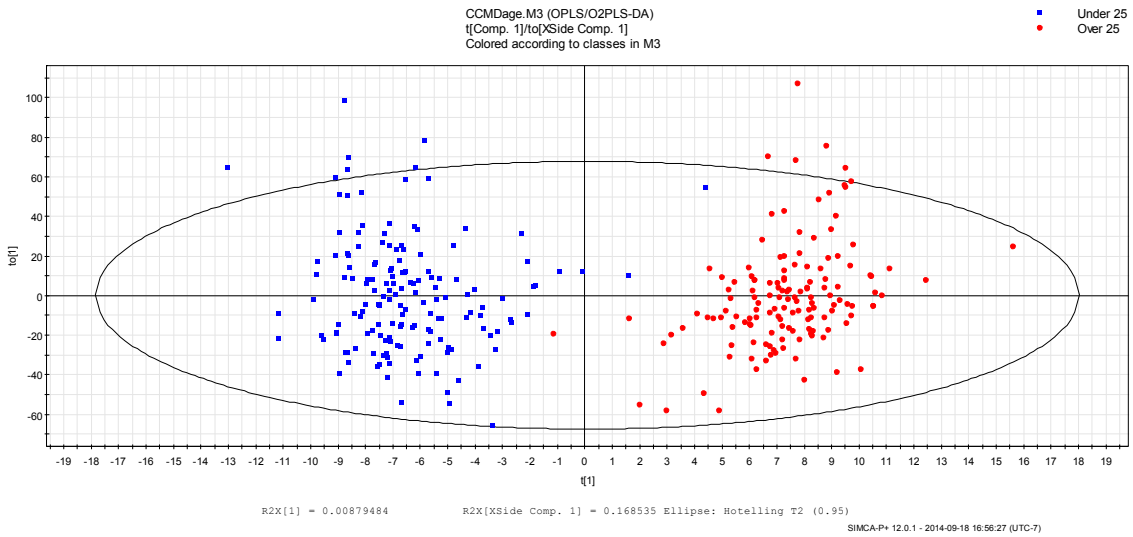


Figure 3-8 OPLS-DA score plot showing a significant separation between two age groups.

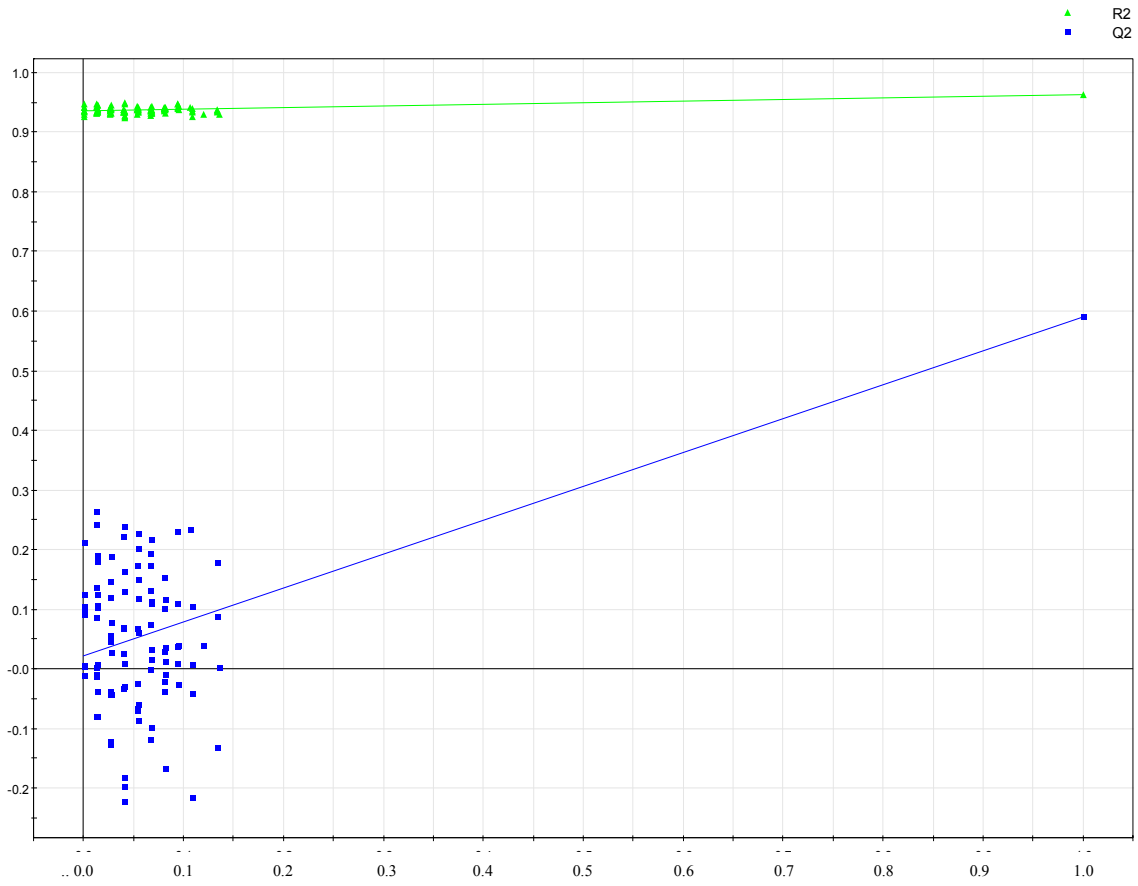


Figure 3-9 Cross validation of the OPLS-DA model derived from two studied age groups. Statistical validation of the OPLS-DA model by permutation analysis using 100 different model permutations test built in SIMCA-P+ software.

Volcano plot analysis was also performed on the age-related metabolic profiles dataset. All the metabolites that had the change greater than 1.2-fold were retained for further analysis. In total, 883 metabolites had both the VIP values higher than 1.5 and fold change higher than 1.2.

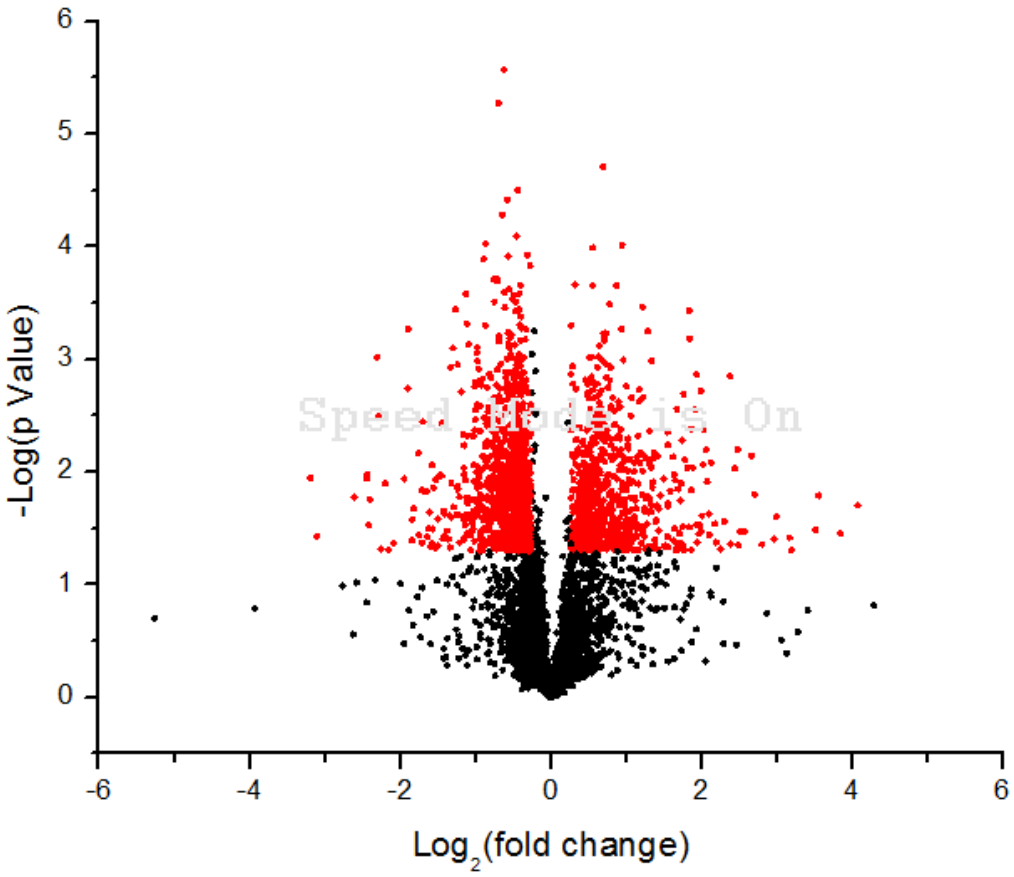


Figure 3-10 Volcano plot of fold change in abundance of metabolites versus probability value for ≤ 25 years vs. ≥26 years. The most significant metabolites (highlighted) in red are distributed in the top left and right region of the plot.

Lower excretion of Cysteinylglycine, 2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate, Formamide, 2-Keto-glutaramic acid, Creatinine, and Methylimidazole acetaldehyde were found to be associated with the older age group, while 3-Methylthiopropionic acid and Prolylhydroxyproline showed a positive relationship with age.

Through this metabolic profiling approach, we have identified numerous potential metabolites that could contribute to age-related progressions. Specifically, creatinine is an

index of musle mass and is likely to be associated with age.[114] When people get older, progressive loss of muscle mass occurs.[115] Differences in creatininine concentrations between two age groups can also be attributed to age-dependent decrease in renal plams flow and glomerular filtration rate.[116]

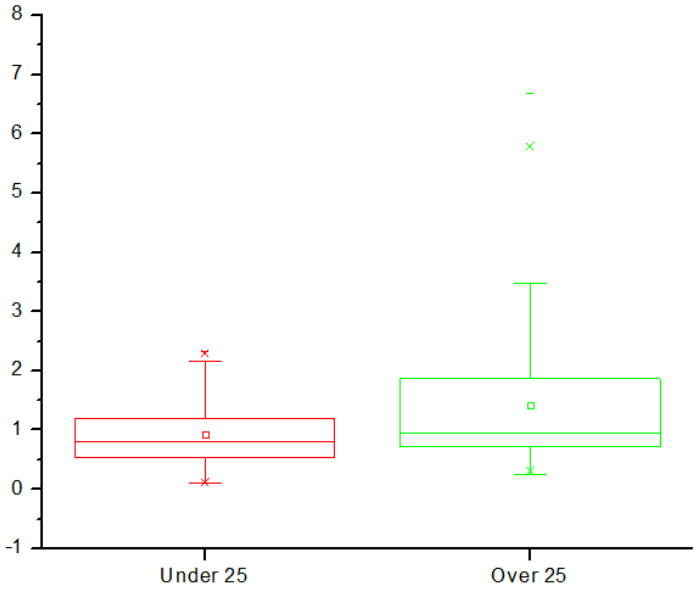


Figure 3-11 Box plot of one of the most significantly discriminant metabolites in age-related metabolome analysis.

Other important metabolites correlated with increasing age were listed in Table 3-3. This list offers great potentials for understanding the biochemical events and possibly for forming the biochemical basis of disease processes. This information can be exploited to help aid with the early diagnostics of different diseases.

Table 3-3 List of potentially important metabolites ranked by VIP values based on the OPLS-DA model for discrimination between different age groups.

VIP	FC>1.2	p-value	m/z	RT (s)	M (≤ 25)	SD	M (≥26)	SD	Putative Match (My compound ID 0 rxn)	Putative Match (Mycompound ID 1 rxn)
3.17	↓	2.69E-06	131.0606	403.59	0.91	0.5	1.41	1.1	8	60
3.04	↓	2.91E-04	98.0482	539.87	0.64	0.3	0.92	0.7		4
2.95	↓	3.62E-04	166.0244	608.45	0.71	0.4	0.95	0.6		18
2.90	↓	7.40E-04	208.0380	876.85	0.57	0.4	0.79	0.6		20
2.87	-	-	61.0506	476.24	0.87	0.3	1.01	0.4	Ethanolamine	15
2.86	↓	6.11E-03	178.0384	807.76	0.77	0.6	1.06	0.8	Cysteinylglycine	14
2.85	↓	3.81E-05	125.0598	689.85	0.75	0.5	1.13	0.8	3	8
2.80	↓	8.02E-05	197.1525	1043.97	0.77	0.5	1.06	0.5		1
2.78	↓	1.62E-03	135.0435	439.16	0.46	0.3	0.64	0.5		
2.70	↑	1.86E-03	361.0687	507.39	1.98	1.7	1.34	1.0		8
2.69	↓	3.14E-05	183.1371	1064.88	0.73	0.4	0.99	0.5		

2.66	↑	1.81E-03	162.0972	451.99	1.28	0.5	1.02	0.5	5-Hydroxylysine	16
2.65	↑	9.67E-05	195.0544	695.01	0.96	1.1	0.50	0.5	6	38
2.64	↓	2.03E-04	181.0383	1257.23	0.61	0.6	1.01	0.9	2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate	27
2.63	↓	3.69E-04	123.0176	1118.99	0.68	0.5	0.95	0.7		1
2.63	↓	5.19E-05	182.0404	587.82	1.02	0.8	1.60	1.4	4	15
2.61	↓	6.77E-04	123.0166	1083.25	0.63	0.5	1.01	0.9		1
2.60	↓	1.47E-04	47.0179	526.59	0.99	0.3	1.20	0.5		
2.58	↑	1.96E-05	120.0214	791.32	1.30	1.2	0.80	0.7	3-Methylthiopropionic acid	13
2.58	↓	1.10E-03	137.0483	1251.83	0.70	0.5	1.11	1.0	5	
2.58	↓	1.77E-03	169.0874	450.75	0.58	0.7	1.04	1.3	1-Methylhistidine; 3-Methylhistidine	
2.55	↓	7.21E-03	155.9557	694.00	1.05	0.7	1.45	1.2		
2.55	↓	2.38E-04	88.0166	904.99	0.89	0.5	1.30	1.0	Pyruvic acid; Malonic semialdehyde	
2.55	↓	4.93E-04	45.0216	530.53	0.82	0.5	1.09	0.7	Formamide	
2.55	↓	1.65E-03	123.0181	1129.87	0.64	0.5	0.88	0.7		1
2.55	↓	2.72E-04	145.0379	667.60	0.78	0.5	1.11	0.9	2-Keto-glutaramic acid	

2.54	↓	9.68E-03	290.0579	313.50	0.76	0.5	1.06	0.9		10
2.52	↑	1.02E-04	182.0608	926.29	1.98	1.6	1.35	0.9	6	-
2.52	↓	4.73E-03	187.0970	891.34	0.62	0.9	1.34	2.2		9
2.51	↓	1.98E-03	163.9963	1143.24	0.88	0.6	1.31	1.3		2
2.50	↓	1.69E-03	146.0463	489.21	0.68	0.3	0.89	0.5		0
2.50	↓	3.07E-04	145.0748	463.80	0.72	0.7	1.21	1.3	6	54
2.49	↓	1.64E-02	113.0597	502.66	1.07	0.6	1.29	0.7	Creatinine	-
2.48	↓	2.19E-04	89.0474	549.08	0.96	0.5	1.27	0.9	4	38
2.47	↓	1.82E-03	124.0647	1014.13	0.99	0.8	1.48	1.3	Methylimidazole acetaldehyde	-
2.46	↓	7.34E-04	187.0976	894.62	0.71	1.0	1.51	2.0		9
2.45	↓	1.11E-03	156.1378	340.53	0.51	0.3	0.69	0.5		-
2.45	↓	3.09E-04	230.0916	370.05	1.14	0.7	1.57	1.0	Aspartyl-L-proline; Naproxen	-
2.44	↓	1.92E-04	181.0394	1232.63	0.64	0.6	1.05	0.9	2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate	-
2.43	↓	1.85E-02	180.0388	840.25	0.54	0.5	0.74	0.6	7	41
2.43	↓	7.89E-04	199.0105	1362.55	0.65	0.5	0.96	0.8		7
2.43	↓	1.36E-02	283.1542	480.62	0.86	0.9	1.40	1.7		7

2.42	↓	2.13E-03	148.0630	909.51	0.95	1.0	1.54	1.6		2
2.42	↓	1.04E-03	139.0994	1043.91	0.79	0.5	1.07	0.7		9
2.41	↓	1.27E-03	224.0807	1406.47	0.83	0.6	1.16	0.9	3	
2.40	↑	5.07E-03	209.0799	1134.17	1.24	1.4	0.77	0.7		14
2.40	↓	5.41E-04	31.0423	647.85	1.10	0.5	1.38	0.7	Methylamine	
2.39	↑	2.18E-04	228.1117	359.42	1.10	0.4	0.88	0.4	Prolylhydroxyproline	
2.38	-	-	87.0875	1232.69	2.19	0.5	2.54	1.0		

Key : M:mean. SD: standard deviation. ↑ indicates an up-regulation that means higher metabolite concentration in the volunteers that were older than 25, and ↓ represents a down-regulation that means higher metabolite concentration in the group that were or younger than 25 years old.

3.3.4 Analysis of BMI Differences

A high body mass index represents a high risk factor for many diseases; however, an explanation for this correlation has not yet been elucidated. We attempt to identify novel BMI-related metabolites to establish a baseline for BMI metabolome. Height and weight measurements of the participants were obtained from the questionnaires. In terms of BMI index, the studied population was quite homogenous, with the majority of the participants fell into the normal BMI range. It is interesting that the OPLS-DA score plot still showed marked differences between the three BMI groups: underweight, normal weight and overweight ($R^{2(}Y)=0.881$; $Q^2(cum)=0.716$)

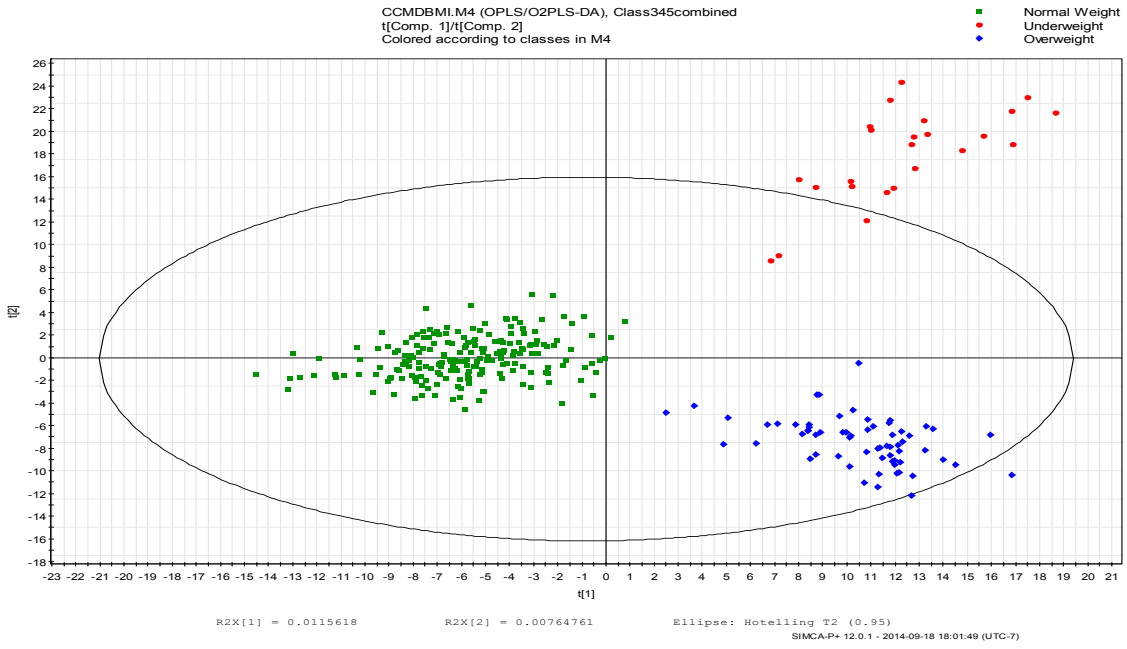


Figure 3-12 OPLS-DA score plot showing a significant separation between three BMI groups.

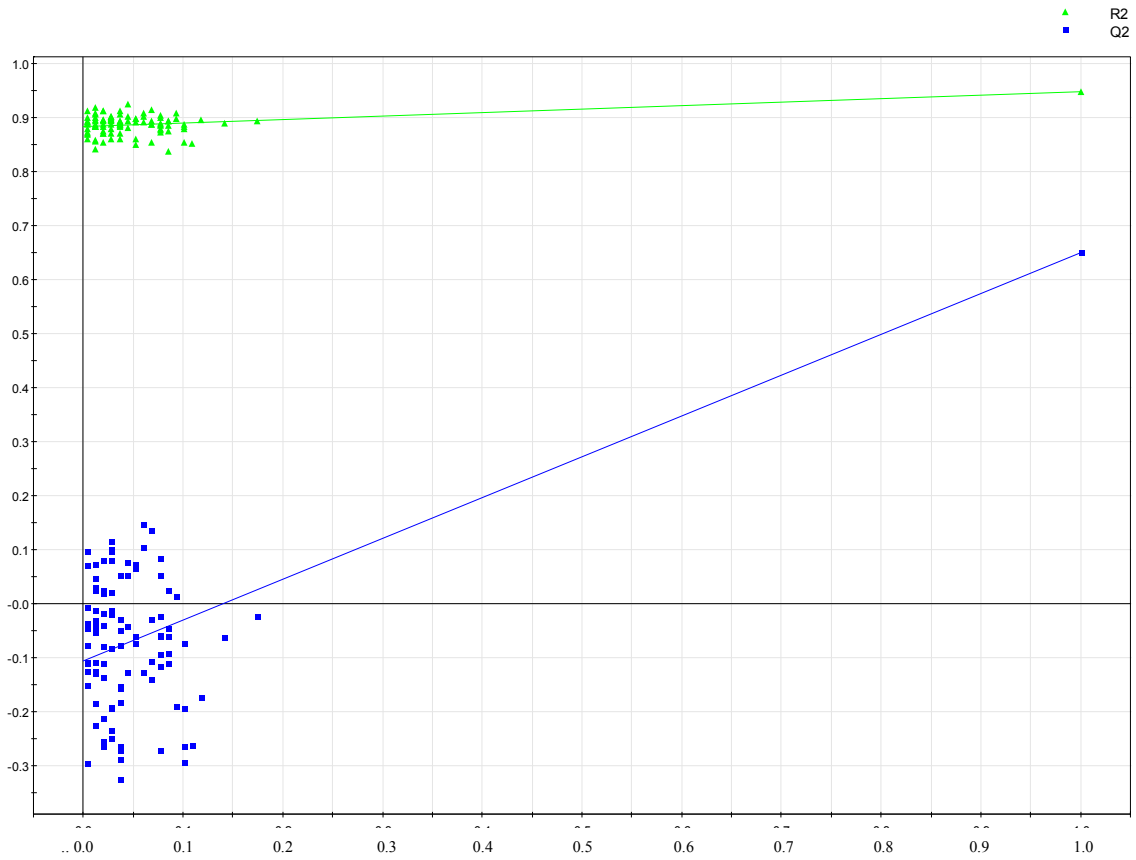
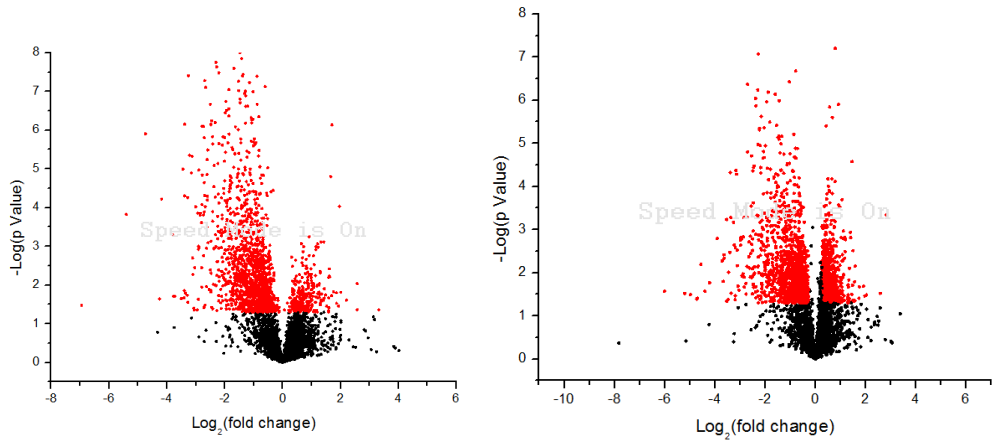


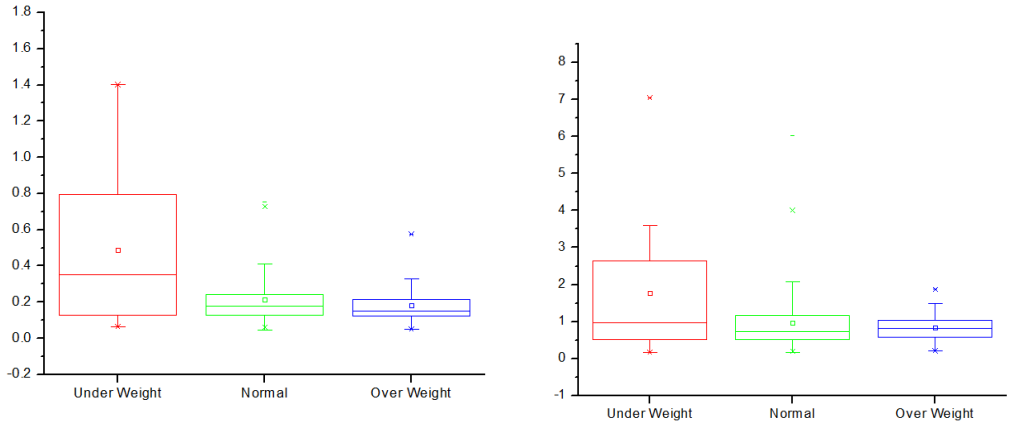
Figure 3-13 Cross validation of the OPLS-DA model derived from three studied BMI groups. Statistical validation of the OPLS-DA model by permutation analysis using 100 different model permutations test built in SIMCA-P+ software.

Our combined VIP analysis and fold change investigation revealed that 683 metabolites can be considered as significantly discriminant (VIP>1.5 and FC>1.2). Out of those metabolites, N-Acetyl-L-tyrosine was found to be highly correlated with the known BMI-associated metabolites as previously reported by Moore et. al.[117] Our analysis also confirmed the link already established between amino acids and BMI, for example, association for glycine and valine.[118,119] This further confirmed that amino acids are biologically important, and their associations with BMI could provide important hints about different diseases (e.g. diabetes).



Underweight vs. Normal Weight Overweight vs. Normal Weight

Figure 3-14 Volcano plot of fold change in abundance of metabolites versus probability value for different BMI groups. The most significant metabolites (highlighted) in red are distributed in the top left and right region of the plot.



Valine and isomers Glycine and isomers

Figure 3-15 Box plot of one of the most significantly discriminant metabolites in age-related metabolome analysis.

Table 3-4 List of potentially important metabolites ranked by VIP values based on the OPLS-DA model for discrimination between different BMI groups.

VIP	FC>1.2 UWvsNW	p-value	FC>1.2 OWvs.NW	p-value	m/z	RT (s)	M (UW))	SD	M (NW)	SD	M (OW)	SD	PutativeMatch (my compound ID 0 rxn)
3.96	↑	1.05E-03	-	-	75.0530	464.58	1.68	1.26	0.72	0.43	0.68	0.32	
3.37	-	-	↑	6.65E-05	189.1372	590.57	0.81	0.45	0.89	0.37	1.41	0.69	
3.29	↑	4.94E-02	↑	2.02E-04	223.0529	307.90	1.26	0.71	0.84	0.55	1.76	1.42	4-(2-Amino-3-hydroxyphenyl)-2,4-dioxobutanoic acid
3.09	↑	9.14E-03	-	-	205.0751	1024.20	4.62	5.73	0.77	0.61	1.01	0.90	3.00
2.99	↑	1.34E-02	-	-	75.0772	460.42	1.45	0.93	0.92	0.53	0.88	0.46	
2.91	↑	1.06E-02	-	-	295.0715	959.93	0.49	0.43	0.21	0.14	0.18	0.10	5.00
2.89	↑	1.71E-02	↓	4.82E-02	61.0178	980.99	6.53	6.52	2.57	2.34	1.93	1.42	Carbamic acid
2.88	-	-	↑	1.59E-03	145.1106	325.71	1.28	1.30	1.06	0.98	2.26	2.22	3-Dehydroxycarnitine
2.84	↑	1.49E-02	↑	1.14E-02	117.0744	1455.57	2.62	2.98	0.77	0.77	1.32	1.48	

2.80	-	-	-	-	75.0302	467.72	1.76	1.90	0.97	0.81	0.84	0.39	Glycine
2.76	-	-	↑	6.15E-04	344.0524	292.02	0.89	0.42	0.97	0.55	1.63	1.16	
2.75	-	-	↑	1.45E-06	217.1683	819.82	0.96	0.44	0.97	0.42	1.44	0.63	
2.72	↑	1.89E-02	-	-	139.0551	988.54	5.47	5.91	1.70	2.86	1.53	1.85	
2.71	↑	7.65E-03	-	-	75.0303	422.63	1.76	1.50	0.88	0.72	0.75	0.54	
2.64	-	-	↑	3.92E-06	229.0735	152.94	0.91	0.53	0.99	0.38	1.34	0.48	
2.63	-	-	↑	7.96E-03	181.0370	1454.32	1.84	2.25	0.76	0.74	1.35	1.40	2-Methyl-3-hydroxy-5-formylpyridine-4-carboxylate
2.56	-	-	↑	9.81E-05	159.0540	329.49	1.15	0.58	1.07	0.51	1.53	0.67	
2.56	-	-	-	-	223.0844	764.09	2.65	3.61	1.05	0.63	1.08	0.93	N-Acetyl-L-tyrosine
2.54	-	-	↑	2.34E-03	173.0153	820.39	0.93	0.60	0.90	0.62	1.55	1.22	
2.53	↑	3.72E-02	↑	8.41E-03	201.1233	212.17	1.38	1.02	0.82	0.37	1.12	0.73	
2.53	↑	2.45E-02	-	-	166.0508	557.85	3.93	5.79	0.85	0.93	1.62	3.13	5.00
2.53	↑	3.52E-02	-	-	182.0449	1420.83	2.74	4.07	0.95	1.07	1.43	1.80	4.00
2.52	↑	1.10E-02	-	-	75.0427	463.11	1.60	1.29	0.86	0.66	0.72	0.56	
2.51	-	-	-	-	272.1092	904.72	3.26	5.41	1.36	1.37	1.57	1.67	5C-aglycone
2.51	↑	1.99E-02	↑	2.43E-02	173.1062	1136.0	1.85	1.66	0.83	0.58	1.14	0.77	4.00

						6							
2.50	-	-	-	-	153.0378	1434.17	2.05	1.91	1.04	1.14	1.05	1.46	3-Hydroxyanthranilic acid; 3-Aminosalicylic acid
2.48	↑	3.84E-02	-	-	184.0392	1460.00	2.83	3.69	0.93	1.01	1.26	1.31	3,4-Dihydroxymandelic acid
2.46	-	-	↑	1.10E-02	177.0483	317.37	1.72	3.05	0.73	0.77	1.60	2.15	N-Formyl-L-methionine
2.46	-	-	↑	1.23E-04	203.1524	658.34	0.96	0.55	0.94	0.45	1.37	0.70	
2.44	-	-	-	-	102.0797	489.22	3.98	6.06	1.61	1.10	1.70	1.17	
2.44	-	-	-	-	135.0533	202.64	2.97	6.22	0.73	0.31	0.65	0.24	Adenine
2.43	-	-	-	-	129.0425	554.56	1.07	0.66	0.76	0.37	0.86	0.43	5.00
2.42	-	-	-	-	147.0876	402.88	2.01	3.00	1.04	0.50	0.92	0.38	
2.42	-	-	-	-	73.0529	812.93	1.70	1.90	0.96	0.53	0.95	0.41	3.00
2.41	-	-	-	-	166.0493	546.37	3.35	5.46	0.93	1.27	1.36	1.51	5.00
2.40	-	-	↑	2.98E-04	245.1038	1069.81	2.12	2.06	1.97	1.70	3.88	3.26	
2.40	↑	2.48E-02	↑	1.20E-02	182.0645	1445.61	1.74	2.14	0.65	0.75	1.07	1.09	
2.39	-	-	-	-	184.0613	245.09	3.13	5.14	0.84	0.98	1.23	1.53	
2.39	↑	4.93E-02	↑	2.86E-02	182.0451	1461.74	2.45	3.85	0.82	0.91	1.33	1.68	4.00

2.39	-	-	↑	6.66E-05	231.1842	895.22	0.94	0.55	0.94	0.46	1.33	0.64	
2.38	-	-	-	-	263.1388	434.80	4.00	4.25	1.93	1.70	1.89	1.32	
2.38	-	-	↓	5.22E-04	166.0626	1048.23	2.46	5.52	0.75	1.48	0.27	0.33	7.00
2.37	-	-	↑	4.43E-03	257.0420	368.70	0.73	0.61	1.09	0.97	1.88	1.84	
2.37	-	-	↑	2.61E-03	426.4157	1343.63	1.13	0.57	0.99	0.66	1.58	1.03	
2.37	-	-	↑	3.80E-02	89.0303	178.15	1.77	0.95	1.43	0.68	1.75	0.83	
2.37	-	-	↑	1.72E-03	173.0155	882.10	1.03	0.66	0.99	0.65	1.65	1.27	
2.36	-	-	-	-	206.1051	922.91	2.59	3.72	1.27	0.62	1.46	0.75	
2.35	-	-	↓	7.80E-03	75.0177	460.42	0.71	0.79	0.51	0.41	0.39	0.19	
2.34	↑	1.67E-02	-	-	151.0307	250.50	2.42	2.66	1.04	0.84	1.32	1.64	

Key:

Key : M:mean. SD: standard deviation. UW: underweight. NW: normal weight. OW: overweight. ↑ indicates an up-regulation that means higher metabolite concentration compared to that of normal weight group, and ↓ represents a down-regulation that means higher metabolite concentration compared to that of normal weight group.

3.4 Conclusion

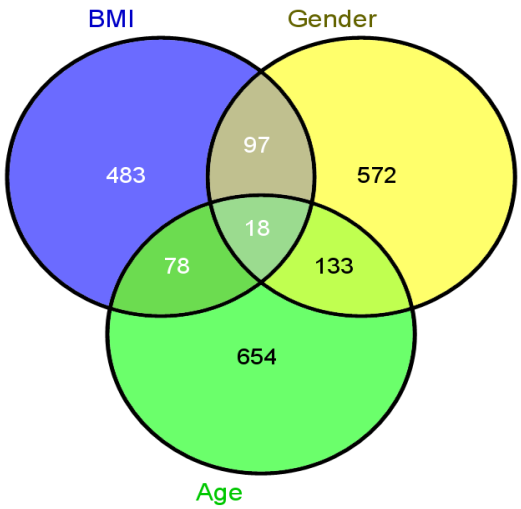


Figure 3-16 Venn diagram showing the overlap of significantly discriminant metabolites of age, gender and BMI-related analysis.

We have shown that results from metabolic profiling using differential isotope labeling approach via LC-MS can provide valuable insights. This work reinforced the potential of our technique to larger-scale epidemiology studies. We have identified a large set of metabolites that differentiate gender, age and BMI. A great advantage of the technique was the large number of metabolites detected, which allowed us to discover new links that might have been neglected by using other analytical tools

This work will contribute to the understanding of the biochemical basis of healthy individuals, and potentially to elucidating the metabolism difference in disease progresses.

Interpretation of the metabolome database will contribute to the development of a comprehensive and accessible dataset of detected metabolites in human urinary samples for the discovery of disease associations and diagnoses in future research.

Chapter 4: Conclusions & Future Directions

This thesis work focused on the development and application of the differential ^{13}C-/^{12}C-isotope dansylation labeling technique targeted at amine and phenol containing compounds. This isotope labeling approach allows a better detection sensitivity and higher precision relative quantification of those target metabolites, thus resulting in a more comprehensive metabolome database.

In Chapter 2, an optimized method for the discovery of biomarkers for MCI and AD diseases using human salivary samples was described. Our work presents our metabolic profiling technique as a promising method for diagnostic and prognostic evaluation of patients of MCI and AD disease. Chemometrics tools such as OPLS-DA and volcano plot showed a clear separation between the diseased groups as the normal aging participants, thus leading to the discovery of the most discriminant metabolites that serve as potential biomarkers. However, it is crucial that this application is further validated using a larger and independent cohort of samples. For this validation phase a similar analysis would be carried out in order to confirm if the biomarkers profile would be consistent with our prior analysis. In the future, other labeling chemistry approaches, which target other functional groups, should be assessed in order to generate an even more comprehensive human salivary metabolome database. One of the limitations of our studies was our sample size was relatively small with all the baseline characteristics being very similar. We were unable to investigate the metabolic differences due to other factor such as race or geography. Because of the noninvasive and simple nature of the analyzed sample, we recommend our technique to be further studied and applied for early diagnosis of MCI and prediction of AD.

In Chapter 3, we employ differential ^{13}C-/^{12}C--isotope dansylation labeling strategy with analysis by LC FT-ICR-MS to establish a human urine metabolome database from 300 urine sample provided by a study group of 100 healthy volunteers over a three consecutive days period. We demonstrated that targeting profiling produces robust statistical models and generates data that can be utilized to understand the metabolic difference due to gender, age and BMI in a healthy population. In subsequent studies, the mechanisms by which these differences occur will be crucial in the discovery of new diagnostic tests. Our results contribute to the establishment of a reference urine metabolome database of Canadian population. It should be noted that a similar study is concurrently carried out in China so that epidemiological investigation can be compared across countries. As a consequence, we can identify metabolic difference due to geographically distinct populations. It is crucial to analyze the metabolome of different geographical regions so that we would be able to determine the presence, or lack there-of, the under- or over-expression of a metabolite that indicates a true representation of a disease state. Future studies should aim to enhance the coverage of the urine metabolome database by employing other isotopic labeling reagents. For example, our research group has developed isotope-coded dansylhydrazine labeling for the determination of carbonyl- containing metabolites and p-dimethylaminophenacyl for the detection of compounds containing carboxylic groups. With the development of new statistical tools and improved positive identification method, our technique is powerful in elucidating a diverse set of metabolites using rapid and simple preparation steps and analytical runs. Future work includes a larger population cohorts and individuals with certain diseases for metabolite biomarker investigations.

References

(1) Fiehn, O. *Plant Molecular Biology* **2002**, *48*, 155-171.

(2) Rochfort, S. *Journal of Natural Products* **2005**, *68*, 1813-1820.

(3) Gika, H. G.; Wilson, I. D.; Theodoridis, G. A. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2014**, *966*, 1-6.

(4) Wilson, I.; Wade, K.; Nicholson, J. *TrAC Trends in Analytical Chemistry* **1989**, *8*, 368-374.

(5) Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D. *Analytica chimica acta* **2012**, *711*, 7-16.

(6) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 251-260.

(7) Lu, W.; Bennett, B. D.; Rabinowitz, J. D. *Journal of Chromatography B* **2008**, *871*, 236-242.

(8) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass spectrometry reviews* **2007**, *26*, 51-78.

(9) Rabinowitz, J. D.; Kimball, E. *Anal. Chem.* **2007**, *79*, 6167-6173.

(10) Want, E. J.; Nordström, A.; Morita, H.; Siuzdak, G. *Journal of proteome research* **2007**, *6*, 459-468.

(11) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends in biotechnology* **2004**, *22*, 245-252.

(12) Theobald, U.; Mailinger, W.; Reuss, M.; Rizzi, M. *Analytical Biochemistry* **1993**, *214*, 31-37.

(13) Mitra, S. *Sample preparation techniques in analytical chemistry*; John Wiley & Sons, 2004; Vol. 237.

(14) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573-3580.

(15) Stashenko, E. E.; Jaramillo, B. E.; Martínez, J. R. *Journal of Chromatography A* **2004**, *1025*, 105-113.

(16) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. *Analytical biochemistry* **2007**, *363*, 185-195.

(17) Horie, K.; Ikegami, T.; Hosoya, K.; Saad, N.; Fiehn, O.; Tanaka, N. *Journal of Chromatography A* **2007**, *1164*, 198-205.

(18) Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S. *Journal of proteome research* **2005**, *4*, 591-598.

(19) Dunn, W. B.; Ellis, D. I. *TrAC Trends in Analytical Chemistry* **2005**, *24*, 285-294.

(20) Quirke, J. M. E.; Adams, C. L.; Van Berkel, G. J. *Anal. Chem.* **1994**, *66*, 1302-1315.

(21) Van Berkel, G. J.; Asano, K. G. *Anal. Chem.* **1994**, *66*, 2096-2102.

(22) Van Berkel, G. J.; Quirke, J. M. E.; Tigani, R. A.; Dilley, A. S.; Covey, T. R. *Anal. Chem.* **1998**, *70*, 1544-1554.

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

(24) Wu, Y.; Li, L. *Anal. Chem.* **2013**, *85*, 5755-5763.

(25) Zheng, J.; Dixon, R. A.; Li, L. *Anal. Chem.* **2012**, *84*, 10802-10811.

(26) Seiler, N.; Deckardt, K. *Journal of Chromatography A* **1975**, *107*, 227-229.

(27) Seiler, N.; Knödgen, B.; Eisenbeiss, F. *Journal of Chromatography B: Biomedical Sciences and Applications* **1978**, *145*, 29-39.

(28) Loukou, Z.; Zotou, A. *Journal of Chromatography A* **2003**, *996*, 103-113.

(29) Minocha, R.; Long, S. *Journal of Chromatography A* **2004**, *1035*, 63-73.

(30) Zezza, F.; Kerner, J.; Pascale, M.; Giannini, R.; Martelli, E. A. *Journal of Chromatography A* **1992**, *593*, 99-101.

(31) Barrett, D.; Shaw, P.; Davis, S. *Journal of Chromatography B: Biomedical Sciences and Applications* **1991**, *566*, 135-145.

(32) Kaddoumi, A.; Nakashima, M. N.; Wada, M.; Kuroda, N.; Nakahara, Y.; Nakashima, K. *Journal of liquid chromatography & related technologies* **2001**, *24*, 57-67.

(33) Xia, Y. Q.; Chang, S. W.; Patel, S.; Bakhtiar, R.; Karanam, B.; Evans, D. C. *Rapid communications in mass spectrometry* **2004**, *18*, 1621-1628.

(34) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.

(35) Wade, D. *Chemico-biological interactions* **1999**, *117*, 191-217.

(36) Annesley, T. M. *Clinical chemistry* **2003**, *49*, 1041-1044.

(37) Grey, L.; Nguyen, B.; Yang, P. *Journal of Chromatography A* **2002**, *958*, 25-33.

(38) Sojo, L. E.; Lum, G.; Chee, P. *Analyst* **2003**, *128*, 51-54.

(39) Kuklenyik, Z.; Ashley, D. L.; Calafat, A. M. *Anal. Chem.* **2002**, *74*, 2058-2063.

(40) Huang, X.; Regnier, F. E. *Anal. Chem.* **2008**, *80*, 107-114.

(41) Wu, L.; Mashego, M. R.; van Dam, J. C.; Proell, A. M.; Vinke, J. L.; Ras, C.; van Winden, W. A.; van Gulik, W. M.; Heijnen, J. J. *Analytical biochemistry* **2005**, *336*, 164-171.

(42) Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2005**, *78*, 567-574.

(43) Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Calmani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton,

J.; Paternostro, G.; Somorjai, R.; Sjostrom, M.; Trygg, J.; Wulfert, F. *Metabolomics* **2007**, *3*, 231-241.

(44) Wibom, C.; Surowiec, I.; Moren, L.; Bergstrom, P.; Johansson, M.; Antti, H.; Bergenheim, A. T. *Journal of Proteome Research* **2010**, *9*, 2909-2919.

(45) Liland, K. H. *TrAC Trends in Analytical Chemistry* **2011**, *30*, 827-841.

(46) Nocairi, H.; Mostafa Qannari, E.; Vigneau, E.; Bertrand, D. *Computational statistics & data analysis* **2005**, *48*, 139-147.

(47) Trivedi, K.; Iles, K. *J Chromatogr Sep Tech* **2012**, *3*, 1-5.

(48) Stone, M. *Journal of the Royal Statistical Society. Series B (Methodological)* **1974**, 111-147.

(49) Thomson, J. J. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1897**, *44*, 293-316.

(50) Thomson, J. J. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1907**, *13*, 561-575.

(51) Sommer, H.; Thomas, H. A.; Hipple, J. A. *Physical Review* **1951**, *82*, 697-702.

(52) Jennings, K. R. *International Journal of Mass Spectrometry and Ion Physics* **1968**, *1*, 227-235.

(53) Barber, M.; Bordoli, R. S.; Sedgwick, R. D.; Tyler, A. N. *Journal of the Chemical Society, Chemical Communications* **1981**, 325-327.

(54) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(55) O'Connor, P. B.; Hillenkamp, F. In *MALDI MS*; Wiley-VCH Verlag GmbH & Co. KGaA, 2007, pp 29-82.

(56) Byrdwell, W. *Lipids* **2001**, *36*, 327-346.

(57) Raffaelli, A.; Saba, A. *Mass Spectrom. Rev.* **2003**, *22*, 318-331.

(58) Kebarle, P.; Peschke, M. *Anal. Chim. Acta* **2000**, *406*, 11-35.

(59) Bleiner, D.; Hametner, K.; Günther, D. *Fresenius' journal of analytical chemistry* **2000**, *368*, 37-44.

(60) Dole, M.; Mack, L.; Hines, R.; Mobley, R.; Ferguson, L.; Alice, M. d. *The Journal of Chemical Physics* **1968**, *49*, 2240-2249.

(61) Wilm, M. S.; Mann, M. *International Journal of Mass Spectrometry and Ion Processes* **1994**, *136*, 167-180.

(62) Iribarne, J.; Thomson, B. *The Journal of Chemical Physics* **1976**, *64*, 2287-2294.

(63) Thomson, B.; Iribarne, J. *The Journal of Chemical Physics* **1979**, *71*, 4451-4463.

(64) Wilm, M. *Molecular & Cellular Proteomics* **2011**, *10*, M111. 009407.

(65) Gaskell, S. J. *Journal of mass spectrometry* **1997**, *32*, 677-688.

(66) Comisarow, M. B.; Marshall, A. G. *Chemical Physics Letters* **1974**, *25*, 282-283.

(67) Comisarow, M. B.; Marshall, A. G. *Chemical Physics Letters* **1974**, *26*, 489-490.

(68) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.

(69) Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. *Analyst* **2005**, *130*, 18-28.

(70) Marshall, A. G.; Hendrickson, C. L. *International Journal of Mass Spectrometry* **2002**, *215*, 59-75.

(71) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Mass Spectrom. Rev.* **2011**, *30*, 884-906.

(72) Twamley, E. W.; Ropacki, S. A. L.; Bondi, M. W. *Journal of the International Neuropsychological Society* **2006**, *12*, 707-735.

(73) Biagioni, M. C.; Galvin, J. E. *Neurodegenerative disease management* **2011**, *1*, 127-139.

(74) Anstey, K. J.; von Sanden, C.; Salim, A.; O'Kearney, R. *American journal of epidemiology* **2007**, *166*, 367-378.

(75) Brainerd, C. J.; Reyna, V. F.; Petersen, R. C.; Smith, G. E.; Taub, E. S. *Neuropsychology* **2011**, *25*, 679.

(76) Bertram, H. C.; Eggers, N.; Eller, N. *Anal. Chem.* **2009**, *81*, 9188-9193.

(77) Spielmann, N.; Wong, D. T. *Oral Diseases* **2011**, *17*, 345-354.

(78) Yan, S.-K.; Wei, B.-J.; Lin, Z.-Y.; Yang, Y.; Zhou, Z.-T.; Zhang, W.-D. *Oral Oncology* **2008**, *44*, 477-483.

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

(80) DNA Genotek , I., 2013.

(81) Birnboim, H. C.; Doly, J. *Nucleic acids research* **1979**, *7*, 1513-1523.

(82) Xia, J.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S. *Nucleic Acids Res.* **2012**, *40*, W127-W133.

(83) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Res.* **2009**, *37*, W652-W660.

(84) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. *Metabolomics* **2013**, 1-20.

(85) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic acids research* **2013**, *41*, D801-807.

(86) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic acids research* **2007**, *35*, D521-526.

(87) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic acids research* **2009**, *37*, D603-610.

(88) Li, L.; Li, R.; Zhou, J.; Zuniga, A.; Stanislaus, A. E.; Wu, Y.; Huan, T.; Zheng, J.; Shi, Y.; Wishart, D. S.; Lin, G. *Anal. Chem.* **2013**, *85*, 3401-3408.

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

(90) Trygg, J.; Wold, S. *Journal of Chemometrics* **2002**, *16*, 119-128.

(91) Baker, S. G. *Journal of the National Cancer Institute* **2003**, *95*, 511-515.

(92) Søreide, K. *Journal of clinical pathology* **2009**, *62*, 1-5.

(93) Uttara, B.; Singh, A. V.; Zamboni, P.; Mahajan, R. *Current neuropharmacology* **2009**, *7*, 65.

(94) Gieβauf, A.; van Wickern, B.; Simat, T.; Steinhart, H.; Esterbauer, H. *FEBS letters* **1996**, *389*, 136-140.

(95) Ahmed, N.; Ahmed, U.; Thornalley, P. J.; Hager, K.; Fleischer, G.; Münch, G. *Journal of neurochemistry* **2005**, *92*, 255-263.

(96) Mangialasche, F.; Polidori, M. C.; Monastero, R.; Ercolani, S.; Camarda, C.; Cecchetti, R.; Mecocci, P. *Ageing research reviews* **2009**, *8*, 285-305.

(97) Martinez, M.; Frank, A.; Diez-Tejedor, E. *Journal of Neural Transmission-Parkinson's Disease and Dementia Section* **1993**, *6*, 1-9.

(98) Schmitt, L.; Escande, M.; Gayral, L.; Charlet, J.; Thouvenot, J. *Journal de psychiatrie biologique et thérapeutique* **1984**, 13-16.

(99) Mesholam, R. I.; Moberg, P. J.; Mahr, R. N.; Doty, R. L. *Archives of neurology* **1998**, *55*, 84.

(100) Brown, G. G.; Levine, S.; Gorell, J.; Pettegrew, J.; Gdowski, J.; Bueri, J.; Helpern, J.; Welch, K. *Neurology* **1989**, *39*, 1423-1423.

(101) Pettegrew, J. W.; Panchalingam, K.; Moossy, J.; Martinez, J.; Rao, G.; Boller, F. *Archives of neurology* **1988**, *45*, 1093.

(102) Brindle, J. T.; Antti, H.; Holmes, E.; Tranter, G.; Nicholson, J. K.; Bethell, H. W.; Clarke, S.; Schofield, P. M.; McKilligin, E.; Mosedale, D. E. *Nature medicine* **2002**, *8*, 1439-1445.

(103) Wishart, D. S. *Drugs in R & D* **2008**, *9*, 307-322.

(104) Bundy, J. G.; Davey, M. P.; Viant, M. R. *Metabolomics* **2009**, *5*, 3-21.

(105) Kouba, E.; Wallen, E. M.; Pruthi, R. S. *The Journal of urology* **2007**, *177*, 50-52.

(106) Echeverry, G.; Hortin, G. L.; Rai, A. J. In *The Urinary Proteome*; Springer, 2010, pp 1-12.

(107) Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E. *International journal of epidemiology* **2008**, *37*, i31-i40.

(108) Lauridsen, M.; Hansen, S. H.; Jaroszewski, J. W.; Cornett, C. *Anal. Chem.* **2007**, *79*, 1181-1186.

(109) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *NMR in Biomedicine* **2005**, *18*, 143-162.

(110) Minkler, P. E.; Hoppel, C. L. *Journal of Chromatography B: Biomedical Sciences and Applications* **1993**, *613*, 203-221.

(111) Jiménez Girón, A.; Deventer, K.; Roels, K.; Van Eenoo, P. *Anal. Chim. Acta* **2012**, *721*, 137-146.

(112) Zhang, Y.; Tobias, H. J.; Auchus, R. J.; Brenna, J. T. *Drug testing and analysis* **2011**, *3*, 857-867.

(113) Fukagawa, N. K.; Martin, J. M.; Wurthmann, A.; Prue, A. H.; Ebenstein, D.; O'Rourke, B. *The American journal of clinical nutrition* **2000**, *72*, 22-29.

(114) Davies, K. M.; Heaney, R. P.; Rafferty, K. *Metabolism* **2002**, *51*, 935-939.

(115) Wang, Y.; Lawler, D.; Larson, B.; Ramadan, Z.; Kochhar, S.; Holmes, E.; Nicholson, J. K. *Journal of proteome research* **2007**, *6*, 1846-1854.

(116) Perrone, R. D.; Madias, N. E.; Levey, A. S. *Clinical chemistry* **1992**, *38*, 1933-1953.

(117) Moore, S. C.; Matthews, C. E.; Sampson, J. N.; Stolzenberg-Solomon, R. Z.; Zheng, W.; Cai, Q.; Tan, Y. T.; Chow, W.-H.; Ji, B.-T.; Liu, D. K.; Xiao, Q.; Boca, S. M.; Leitzmann, M. F.; Yang, G.; Xiang, Y. B.; Sinha, R.; Shu, X. O.; Cross, A. J. *Metabolomics* **2014**, *10*, 259-269.

(118) Gaudet, M. M.; Falk, R. T.; Stevens, R. D.; Gunter, M. J.; Bain, J. R.; Pfeiffer, R. M.; Potischman, N.; Lissowska, J.; Peplonska, B.; Brinton, L. A. *The Journal of Clinical Endocrinology & Metabolism* **2012**, *97*, 3216-3223.

(119) Cheng, S.; Rhee, E. P.; Larson, M. G.; Lewis, G. D.; McCabe, E. L.; Shen, D.; Palma, M. J.; Roberts, L. D.; Dejam, A.; Souza, A. L. *Circulation* **2012**, *125*, 2222-2231.